Dissertation zur Erlangung des Doktorgrades

der Fakultät Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural and biochemical characterization of specific phosphodiesterases in the degradation pathway of c-di-AMP

David Jan Drexler

aus

Koblenz, Deutschland

2019

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn PD Dr. Gregor Witte betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenhändig und ohne unerlaubte Hilfe erarbeitet.

München, den 02.09.19

_____

David Jan Drexler

Dissertation eingereicht am	05.09.2019
1. Gutachter:	PD Dr. rer. nat. Gregor Witte
2. Gutachter:	Prof. Dr. rer. nat. Karl-Peter Hopfner
Mündliche Prüfung am	07.11.2019

This thesis has been prepared from August 2016 to August 2019 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-University Munich.

This is a cumulative thesis based on following publications

Drexler, D. J.*; Müller, M.*; Rojas-Cordova, C. A.; Bandera, A. M.; Witte, G. Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from Thermotoga maritima. *Structure* **2017**, *25* (12), 1887-1897.e4.

Latoscha, A.*; Drexler, D. J.*, Al-Bassam, M. M.; Kaever, V.; Findlay, K. C.; Witte, G.; Tschowri, N. C-di-AMP hydrolysis by a novel phosphodiesterase is crucial for differentiation of antibiotic-producing bacteria. (submitted)

Andreeva, L.; Hiller, B.; Kostrewa, D.; Lässig, C.; De Oliveira Mann, C. C.; Jan Drexler, D.; Maiser, A.; Gaidt, M.; Leonhardt, H.; Hornung, V.; Hopfner, K. P. cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. *Nature* **2017**, *549* (7672), 394–398.

Hopfner, K. P.; Andreeva, L.; Drexler, D. J.; **2019**. A fluorescent cyclic dinucleotide and its use in methods of identifying substances having an ability to modulate the cGAS/STING pathway. European Patent Application EP 17182689.4.

Dialer, C. R.; Stazzoni, S.; Drexler, D. J.; Müller, F. M.; Veth, S.; Pichler, A.; Okamura, H.; Witte, G.; Hopfner, K. P.; Carell, T. A Click-Chemistry Linked 2′3′-cGAMP Analogue. *Chem. - A Eur. J.* **2019** *25* (8), 2089–2095.

*: equal contribution

Parts of this thesis have been presented at international conferences:

Drexler, D. J.; Bandera, A. M.; Schuller, S.; Witte G. The Essential Second Messenger in c-di-AMP in Bacteria: Analyzing the Molecular Framework of Synthesis, Recognition and Decay
Poster presentation at CSSB Opening Symposium: Frontiers in Structural Systems Biology of Host-Pathogen Interactions: 13-15.11.2017 in Hamburg, Germany

Drexler, D. J.; Bandera, A. M.; Schuller, S.; Witte G. Structural and biophysical analysis of the c-di-AMP hydrolyzing phosphodiesterases.
Poster presentation at SPP 1879 International Symposium: Nucleotide Second Messenger Signalling in Bacteria: 30.09.-03.10.2018 in Berlin, Germany

Table of contents

Summary

In recent years, cyclic dinucleotides have been discovered in many important signalling pathways forming a highly interesting field in the research topic of second messengers. They exist in prokaryotes, regulating key functions of the cell, and in eukaryotes, as signalling molecule in the mammalian innate immune system. The main part of the present work focuses on the degradation of the prokaryotic second messenger c-di-AMP, while a minor section deals with applications for the eukaryotic 2′3′-cGAMP.

The recently discovered dinucleotide second messenger c-di-AMP plays an important role in bacteria and was found to be essential in most species including many pathogens, such as *Staphylococcus aureus*, *Listeria monocytogenes* or *Streptococcus pneumoniae*. In the past 10 years, a new research field developed to characterize the functions of c-di-AMP. It is involved in DNA integrity, sporulation, growth, cell wall metabolism, antibiotic resistance, potassium homeostasis, osmolyte uptake, biofilm formation, virulence and gene regulation. Although several c-di-AMP-binding proteins have been discovered, many of the pathways are not understood in detail. Recent studies identified the regulation of osmolyte homeostasis, in particular potassium ion transport, to be a major function of c-di-AMP. For accurate signal transduction in the cell, the levels of c-di-AMP have to be tightly regulated. This can be achieved either by controlling rates of synthesis or degradation of the second messenger. The synthesis from two ATP molecules is catalyzed by diadenylate cyclase domain (DACs) containing proteins, whereas the degradation is performed by specific phosphodiesterases (PDEs). In view of the fact that c-di-AMP has crucial functions in many pathogens, the signalling pathway represents an interesting research field for drug discovery.

The hydrolysis pathways of c-di-AMP are not understood in detail yet. PgpH, GdpP and DhhP represent prototypes of the three major classes of c-di-AMP specific PDEs that have been identified so far. The multi domain membrane PDE GdpP and the single domain soluble DhhP PDE use the same catalytic DHH domain for c-di-AMP hydrolysis. Both DHH-domain containing PDEs are present in parallel in many species, which raises several questions regarding the degradation pathway of c-di-AMP. For a better understanding of c-di-AMP hydrolysis by DHH-type PDEs, a homologue from *Thermotoga maritima* (TmPDE) was analyzed biochemically and structurally. The results include six high-resolution structures of TmPDE in different reaction steps and thus provide a detailed insight into the reaction mechanism. Furthermore, the substrate specificity was described extensively, supporting a two-step degradation pathway for c-di-AMP. However, in some species such as Streptomyces that lack all of the so far described PDEs, c-di-AMP hydrolysis has to be performed differently. In *Streptomyces venezuelae*, recently a novel PDE class was identified, which is distinct from all

known c-di-AMP specific PDEs. In a collaboration project with the Tschowri group (Humboldt-Universität Berlin), the present work reveals that this enzyme specifically hydrolyzes c-di-AMP. It is mainly present in Actinobacteria and therefore described as AtaC (Actinobacterial PDE targeting c-di-AMP). Based on a combination of structure predictions and mutational analysis, the results indicate a manganese ion-dependent hydrolysis mechanism. In contrast to DHH-type PDEs, AtaC was shown to be a monomer. In conclusion, the findings on the two PDEs TmPDE and AtaC described in the present work contribute substantially on the degradation pathway of c-di-AMP.

Apart from the prokaryotic CDNs, the recently identified eukaryotic 2′3′-cGAMP moved into focus of scientific research. 2′3′-cGAMP is a signalling molecule in the cGAS-STING pathway, which is an important innate immune component. Since modifications in this pathway were found to be potential approaches for cancer immunotherapy, it not only raised scientific interest in academia but also in industry. Therefore, methods to describe this pathway are highly required. In the present work, a fluorescent analogue of 2′3′-cGAMP, named fGAMP, was characterized and analyzed in regard to novel applications. It was shown to bind STING *in vitro* and induce its activation cells. Binding to STING decreases the fluorescence of fGAMP, which provides the possibility to develop a fluorescent STING binding assay that could be commercially applied as screening technology to identify therapeutic molecules for STING.

Introduction

For proper environmental adaption, every organism has to sense external signals that derive from changes in temperature, light, nutrients or pH. Subsequently, the information can be transduced to multiple targets by using second messenger molecules, often connected in complex signalling pathways. A major class of these molecules are nucleotide-derived second messengers, which have key functions in a variety of signal pathways in eukaryotes, prokaryotes and archaea. As these signal pathways are often required for many important cellular processes, they display possible targets for the treatment of diseases. The increasing number of antibiotic resistant pathogens is a worldwide concern and alternatives for treatment of bacterial diseases are highly required. A better understanding of the nucleotide second messenger signalling pathways can give important insights for novel drug development [1–3].

1. Cyclic dinucleotide second messengers

Nucleotide second messenger were discovered in many important signalling pathways of eukaryotes and prokaryotes. So far, all physiological types consist of an adenosine monophosphate (AMP) or a guanosine monophosphate (GMP) component. The cyclic nucleotides cyclic 3′,5′-AMP (cAMP) and cyclic 3′,5′-GMP (cGMP) were first identified in eukaryotes in 1957 and 1963 by the group of Sutherland, who reported an important role in mammalian hormone regulation [4–6]. Later, both molecules were also described in prokaryotic signalling pathways. In 1970, (p)ppGpp (GDP 3′-diphosphate and GTP 3′-diphosphate) was found in *Escherichia coli*, representing the first prokaryotic nucleotide second messenger that has been hitherto described [7]. It is produced upon nutrient starvation and coordinates survival in stationary phases, which is termed stringent response [1]. Another type of nucleotide second messengers are cyclic dinucleotides (CDNs), including cyclic di-GMP (c-di-GMP), cyclic di-AMP (c-di-AMP) and two cyclic GMP-AMP types (3′3′-cGAMP and 2′3′-cGAMP). The prokaryotic CDNs are key signalling molecules for a variety of functions in bacterial cells. Particularly c-di-AMP is essential many species. All known prokaryotic CDNs are linked by two 3′5′ phosphodiester bonds between the ribose components. However, in eukaryotes a non-canonical version of cGAMP was discovered as signalling molecule in the innate immune system [8,9]. It is linked by one 3′5′ and one 2′5′ phosphodiester bond and is therefore distinguished as 2′3′-cGAMP [10–12] (described in chapter 4 in more detail).

Besides the characterized CDNs, a series of novel CDNs and one trinucleotide have been discovered as products of nucleotidyltransferases (CD-NTases) in a recent study. These CD-NTases are widely distributed in bacterial genomes and share a high structural homology to well characterized CD-NTases, like OAS1, cGAS or DncV. The analysis comprises major product formation of cyclic di-UMP, cyclic UMP-AMP and cyclic AMP-AMP-GMP. Although

they all have been shown to bind and activate mammalian CDN-receptors of the innate immune system, their physiological existence and function in prokaryotes have not been elucidated yet [13].

The best-characterized CDN is c-di-GMP, which was already discovered in 1987 by the group of Benziman [14]. The major role of c-di-GMP is to switch between motility and biofilm formation of many species, such as *Vibrio cholerae*, *Pseudomonas aeruginosa*, *E. coli* or *Klebsiella pneumoniae* [15–19]. It is synthesised by diguanylate cyclases (DGCs) and degraded by specific phosphodiesterases (PDEs), containing either EAL or HD-GYP domains, which are described in chapter 3.2 in more detail. DGCs synthesize c-di-GMP from two molecules of GTP. Their conserved motif is present in the catalytic active site (A-site) of GGDEF or GGEEF domains, which function as homodimers. The active site is located in the dimer interface with each unit coordinating one GTP. In the first step, the intermediate 5′-pppGpG is formed prior to cyclization to c-di-GMP. In addition to the active site, many GGDEF domains also contain an inhibitory site (I-site), which binds c-di-GMP and leads to allosteric auto-inhibition. Furthermore, GGDEF domains are often linked to regulatory domains, which can activate or inhibit catalysis of c-di-GMP. It is noteworthy that a significant number of GGDEF domains are connected to EAL or HD-GYP domains, thus the respective proteins can contain both synthase and phosphodiesterase at the same time (discussed in more detail in chapter 3.2). The signal transduction of c-di-GMP was described for several receptors, including PiLZ domains, I-sites in catalytic enzymes and degenerate GGDEF, EAL or HD-GYP domains. In addition, c-di-GMP is signalling by gene regulation through binding to transcription factors, repressors or riboswitches. This regulation pathway mainly correlates with biofilm formation. However, some studies also report effects on virulence, regulation of the cell cycle and other processes [1,20].

The prokaryotic 3′3′-cGAMP was discovered in 2012 in *V. cholerae* [21]. It is synthesized from ATP and GTP by the dinucleotide cyclase DncV, which also produces c-di-AMP and c-di-GMP as byproducts. DncV is required for chemotaxis repression in *V. cholerae*. Interestingly, the repression of chemotaxis leads to enhanced intestinal colonialization indicating an important role of DncV on this function [21]. The *V. cholerae* phospholipase CapV was shown to be directly activated by 3′3′-cGAMP and to modify the cell membrane [22]. Besides, 3′3′-cGAMP can also bind to riboswitches, as observed for electrophysiology genes in *Geobacter sulfurreducens* and the control of exoelectrogenesis in *Deltaproteobacteria* [23,24]. Apart from *V. cholerae*, the production of 3′3′-cGAMP by DncV was also described in *E. coli*, where it correlates with biofilm formation and motility [25]. Although DncV is distinct from other prokaryotic CD-NTases, its structure and binding mode are highly similar to the eukaryotic 2′3′-cGAMP synthase cGAS [26]. Recently, 3′3′-cGAMP was also shown to be synthesized by hybrid promiscuous (Hypr) GGDEF enzymes, that can produce all three prokaryotic CDNs [27]. For the

degradation of 3′3′-cGAMP, three specific PDEs have been identified so far [28] (described in chapter 3.2).

In 2008, c-di-AMP was discovered in the crystal structure of the DNA integrity scanning protein DisA by Witte *et al.* [29]. As it is essential for most bacteria, including many pathogens, c-di-AMP expanded the research field of bacterial CDNs significantly. The synthesis, function and degradation of c-di-AMP will be the main topic of the following chapters.

2. The essential second messenger: c-di-AMP

Since the discovery of c-di-AMP, several targets and functions of this second messenger have been reported. It is mainly produced in gram-positive bacteria but also found in a few gram-negative species. The signalling pathway of c-di-AMP has been analyzed intensively in many pathogens, including *S. aureus*, *L. monocytogenes* or *Mycobacterium tuberculosis* and model organisms like *B. subtilis*. The research topic raised particular interest as c-di-AMP was found to be essential under normal growth conditions in almost all tested species, including *B. subtilis*, *S. aureus*, *L. monocytogenes*, *S. pneumoniae* and *Mycoplasma pneumoniae* [30–37]. On the other hand, accumulation of c-di-AMP inhibits growth. In addition, c-di-AMP is secreted from pathogens during infection and induces innate immune response. In bacterial cells, c-di-AMP is involved in DNA integrity, sporulation, growth, cell wall metabolism, antibiotic resistance, potassium homeostasis, osmolyte uptake, biofilm formation, virulence and gene regulation. The synthesis of c-di-AMP from two ATP molecules is catalyzed by diadenylate cyclases (DACs). Specific phosphodiesterases (PDEs) hydrolyze c-di-AMP to avoid accumulation of the second messenger (Figure 1). The synthesis, degradation and the most important functions of c-di-AMP from recent studies are described in the following chapters.

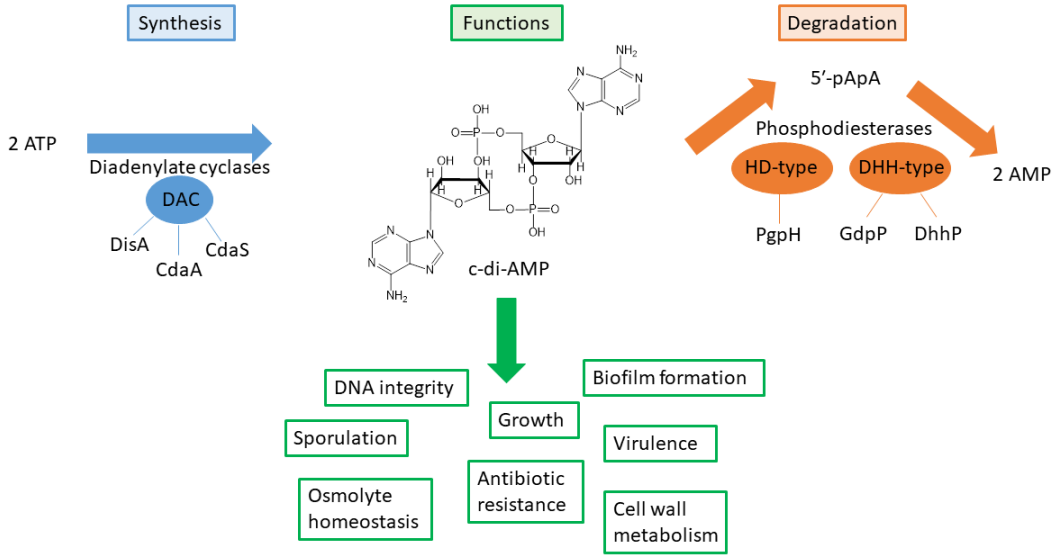


Figure 1: Signalling pathway of c-di-AMP divided into synthesis, functions and degradation. DAC domains, found in different proteins, synthesise c-di-AMP from two molecules of ATP. Receptors bind c-di-AMP for signalling functions. The degradation of c-di-AMP is performed by HD- or DHH- domain containing PDEs, producing AMP via the intermediate 5′-pApA.

2.1. Synthesis

The synthesis of c-di-AMP from two molecules of ATP is catalyzed by diadenylate cyclase domains (DACs). DAC domains are present in all so far analyzed c-di-AMP-producing species, annotated in the Pfam database [38] in 4053 protein sequences from 4909 organisms (Pfam accession PF02457). Five different major classes of proteins containing DAC domains have been characterized so far, classified as DisA, CdaA (also DacA), CdaS (also DacB), CdaM and CdaZ (also DacZ) [29,35–37,39–41]. In some species, these enzymes occur in parallel, as for example DisA, CdaA and CdaS in *B. subtilis*, while most bacteria only possess a single c-di-AMP synthase [42]. Although the single protein classes can differ in additional domains and oligomeric assembly, all DAC domains contain conserved motifs, in particular the DGA (Asp-Gly-Ala) and RHR (Arg-His-Arg) motif (Figure 2B).

The first DAC domain was characterized from the crystal structure of the DNA integrity scanning protein A (DisA) from *T. maritima* [29]. Prior to this study, DisA was reported to trigger a sporulation DNA damage checkpoint response [43]. After the identification of its diadenylate activity, DisA proteins were also described to synthesize c-di-AMP in other species, such as *B. subtilis*, *Mycobacterium smegmatis* or *M. tuberculosis* [29,44,45]. In addition, the determination of crystal structures of DisA in different reaction states allowed the precise characterization of DAC domains [29,39]. DisA consists of an N-terminal DAC domain and a C-terminal DNA-binding helix-hairpin-helix (HhH) domain connected by a helical spine. Two DisA tetramers form an octamer with the DAC domains facing each other, comprising the active

sites (Figure 2A), which is highly specific for adenosine through hydrophobic interactions and steric properties. The synthesis of c-di-AMP is dependent on magnesium or manganese ions, which are coordinated by Asp 75 of the DGA motif of the opposite monomer. The triphosphate of ATP is bent around the metal ion and the conserved RHR motif is additionally coordinating the β- and γ-phosphate. A highly conserved serine is polarizing the γ-phosphate, resulting in a suitable arrangement of the α-phosphate for nucleophilic attack from the 3′-OH group of the ATP located opposite to form c-di-AMP. DisA is constitutively active unless it binds branched DNA, such as holiday junctions, which reduces c-di-AMP synthesis [29,39]. In *B. subtilis*, the induction of DNA damage was shown to result in arrested sporulation dependent on DisA. In addition, the deletion of DisA leads to lower c-di-AMP levels and reduced sporulation [46]. In conclusion, the function of DisA is to scan the DNA for integrity and signal DNA damage via reduced levels of c-di-AMP [29,46]. Additionally, DisA is part of a conserved operon also encoding for RadA (SMS), an ATPase associated to DNA damage repair. RadA inhibits DisA activity in *B. subtilis* and *M. smegmatis* and the interplay of DisA and RadA was shown to be crucial for homologous recombination (HR) [47,48]. A recent study reports the interaction of DisA with Mfd and Uvr in correlation with nucleotide excision repair (NER) [49]. Although further investigation is required to characterize the regulation of DisA, an important function for c-di-AMP in DNA maintenance is evident.

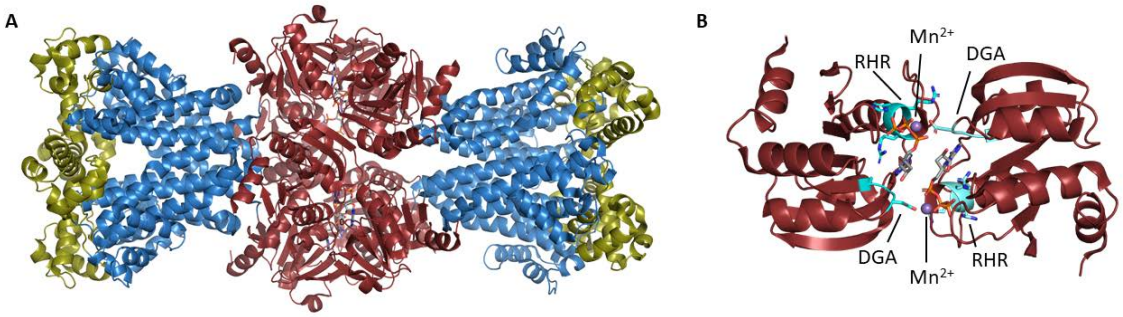


Figure 2: Crystal structure of *T. maritima* DisA. (A) Octameric assembly of *T. maritima* DisA in complex with c-di-AMP; DAC domain (purple), linker domain (green) and HhH DNA-binding domain (blue) (PDB 3C1Y [29]). (B) Close up view on a DAC domain dimer of *T. maritima* DisA in complex with 3′-dATP and Mn^{2+} (purple, PDB 4YVZ [39]). Conserved motifs are annotated (cyan).

Another DAC domain containing protein is the c-di-AMP synthase A (CdaA or DacA), which is the most prominent among bacteria. In many pathogenic species like *S. aureus* or *L. monocytogenes* it is the only DAC domain protein. CdaA consists of a cytosolic DAC domain flanked by coiled coils and is associated to the membrane via three transmembrane helices at the N-terminus. The single DAC domains of CdaA from *L. monocytogenes* and *S. aureus* are structurally highly similar to the DAC domain of DisA (Figure 3). Although some crystal structures are lacking the head-to-head conformation [50,51], the active DAC dimer could be analyzed in a recent crystal structure of *L. monocytogenes* CdaA [52]. Consistent with the DAC

reaction mechanism, multiple dimers of DacA from *S. aureus* have to interact for enzymatic activity [50]. CdaA is part of a conserved operon containing genes for the regulatory protein y*bbR* and the phosphoglucosamine mutase *glmM*. The transmembrane region of CdaA interacts with YbbR (also CdaR) while the cytosolic part of CdaA binds to GlmM. YbbR consists of one transmembrane region and four YbbR domains which are predicted to be extracellular [53]. Although the precise function is unclear, YbbR was shown to negatively regulate CdaA activity in *L. monocytogenes* and to be involved in acid stress resistance in *S. aureus* [54,55]. GlmM is essential for cell wall integrity being required in the synthesis pathway of peptidoglycan [56]. It was shown to inhibit CdaA activity in *Lactococcus lactis* and *S. aureus* most likely by preventing CdaA to form dimers [50,57]. These observations depict an important role of CdaA in cell wall metabolism in agreement with the phenotypes of CdaA mutant strains addressed in chapter 2.3.

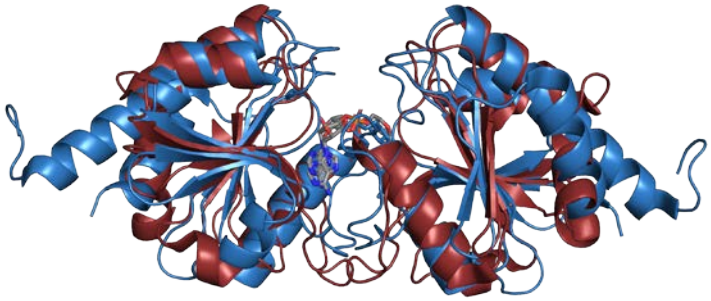


Figure 3: Crystal structure of *L. monocytogenes* CdaA (blue, PDB 6HVL [52]) in superimposition with a DAC domain dimer of *T. maritima* DisA (red, PDB 3C1Y [29]).

A third class of diadenylate cyclases, which is only present in *Bacillus* species, is exclusively expressed during sporulation and is referred to as cyclic di-AMP synthase S (CdaS). As the levels of c-di-AMP are important for spore germination, CdaS is assumed to control these locally in forespores. The deletion of CdaS in *B. subtilis* spores results in disturbed germination. However, the precise regulation of CdaS can only be suggested based on the structural information. In addition to the DAC domain, CdaS contains a YojJ domain, consisting of two α-helices. The crystal structure and SEC experiments indicate a possible hexameric assembly of CdaS, without the DAC domains facing each other as required for enzymatic activity. A mutation in the α-helices leads to a dramatic increase in activity, which indicates an autoinhibitory role of the YojJ domain by holding CdaS in the inactive hexameric form. It is assumed that a signal activates CdaS by disrupting the hexameric assembly to promote spore germination [58,59].

Another DAC domain containing protein is CdaM, which consists of an additional transmembrane helix and is present in species of the *Mycoplasma* genus. The synthesis of c-di-AMP was observed for CdaM from *M. pneumoniae* [37]. DAC domains were identified in several further hypothetical proteins by sequence homology, including CdaZ, which is present in archaea. The analysis of a homologue from *Haloferax volcanii* confirms the existence of c-di-AMP in

archaea. CdaZ is essential in *H. volcanii*, whereas its overexpression leads to cell death [41]. Besides, the production of c-di-AMP was also observed for a CdaZ homologue from *Methanocaldococcus jannaschii* [40].

2.2. DNA integrity and sporulation

Some bacterial species react to nutritional stress by forming highly environmentally resistant dormant cells, called endospores. The spores return to vegetative growth by a process named germination if nutrients become available again [60,61]. Sporulation is mainly observed and characterized in *Bacillus* species but was more recently also detected in *Mycobacteria* [62]. The correlation to c-di-AMP is reported for *B. subtilis* in several studies. During sporulation, the levels of c-di-AMP are increased and reduced levels cause a delay. The DNA integrity scanning protein DisA has an important regulatory function for this process. DisA produces c-di-AMP dependent on an intact DNA, which is required for spore formation. The deletion of DisA results in delayed sporulation. In addition, DNA damaging agents lead to reduced levels of c-di-AMP in a DisA-dependent manner in *B. subtilis* [29,46]. Another correlation is observed for the c-di-AMP synthase CdaS, which is required for spore germination and is only expressed during sporulation [58,59]. However, the targets of c-di-AMP which regulate this process are not yet identified.

2.3. Growth, cell wall metabolism and antibiotic resistance

As c-di-AMP is essential for most bacterial species, basic cell functions are expected to be associated to this second messenger. Several studies report that altered c-di-AMP levels induce growth defects as observed in many species, such as *B. subtilis*, *S. aureus*, *L. monocytogenes*, *M. smegmatis*, *M. tuberculosis*, *M. pneumoniae*, *Streptococcus pyogenes* and *Streptococcus suis* [2,35,37,47,54,63–65].

A direct correlation between c-di-AMP levels and cell wall biosynthesis has been shown in *S. aureus*. The cell wall of such gram-positive bacteria is composed of highly cross-linked peptidoglycan (PG) polymers, wall teichoic acid (WTA) and lipoteichoic acid (LTA). As shown for LTA deficient mutant strains in *S. aureus*, lack of LTA results in growth defects. This phenotype can be compensated by increased PG cross-linking, which correlates with elevated c-di-AMP levels [63]. Accordingly, c-di-AMP has an effect on PG in *B. subtilis*. The deletion of c-di-AMP results in cell lysis and reduced cell growth. This phenotype can be can be supressed by addition of SMM (Spizizen minimal medium), which is known to restore growth of PG deficient cells [35]. A direct link of c-di-AMP to cell wall metabolism was identified for the membrane-associated c-di-AMP synthase CdaA (DacA). The activity of CdaA is inhibited by the phosphoglucosamine mutase GlmM, which is an essential component in the synthesis pathway of

PG [50,57]. Deletion of CdaA in *L. monocytogenes* affects cell wall integrity, cell division and growth and leads to higher sensitivity to cell wall active antibiotics [54].

The role of c-di-AMP in cell wall metabolism is particularly interesting in regard to resistance against cell wall targeting antibiotics. The absence of c-di-AMP in *B. subtilis* results in increased susceptibility to cefixime, aztreonam and moenomycine. The opposite effect, a higher resistance against these antibiotics, is observed at increased levels of c-di-AMP [35]. The correlation between c-di-AMP and antibiotic resistance has also been shown for pathogenic species, such as *S. aureus*, *S. pyogenes* or *L. monocytogenes* [2,54,63,66]. For example, the deletion of c-di-AMP in *L. monocytogenes* results in decreased growth rates and increased susceptibility to cefuroxime, ampicillin and penicillin [67]. A higher sensitivity to antibiotics was also shown for c-di-AMP deficient strains of methicillin resistant *S. aureus* (MRSA) [66].

2.4. Osmolyte homeostasis

In recent years, the regulation of osmolyte transport, in particular potassium ion (K^+) transport has been described as a major role of c-di-AMP in bacteria. K^+ is one of the most important and abundant cations in a cell. It buffers the negative charge of DNA, maintains the turgor and is required for the activity of various enzymes. Bacterial cells possess K^+ concentrations between 200 – 400 mM and have to respond to changes in K^+ levels, as an accumulation is toxic [68]. Recent studies revealed several K^+ transport systems that are regulated by c-di-AMP. These can be either controlled directly by c-di-AMP binding, as it is the case for proteins of the Trk/Ktr/HTK superfamily and proteins of the KupA/B class, or by c-di-AMP dependent gene regulation as described for KimA and Kdp(F)ABC [69–73] (Figure 4). For all these transport systems, the import of K^+ was shown to be inhibited by c-di-AMP. In addition, the cation/proton antiporter CpaA, which exports K^+ and Na^+ in *S. aureus* is more active when bound to c-di-AMP [74].

The direct interaction with c-di-AMP is best described for RCK_C (regulator of conductance of K^+) domains, which are present in the two Ktr family transport systems KtrA/B and KtrC/D, the antiporter CpaA and the transcription factor BusR. Ktr transport systems consist of two components, a dimeric membrane channel (KtrB or KtrD) and a cytoplasmic regulatory octamer (KtrA or KtrC), which is required for activity. The cytoplasmic component consists of one ATP-binding N-terminal RCK_N domain and one C-terminal RCK_C domain, which binds c-di-AMP. ATP induces higher activity while the interaction with c-di-AMP inhibits complex formation of the two channel components and leads to inhibition of K^+ import [69,73,75–77]. The crystal structures of KtrA (Figure 5) and CpaA in complex with c-di-AMP reveal a common dimeric overall fold with c-di-AMP bound in the dimer interface of the RCK_C domain. As both structures are lacking their N-terminal part, it remains unclear what conformational changes from

binding c-di-AMP modulate the activity of the ion channel [74,78]. Further proteins of the Ktr/Trk family that were shown to bind c-di-AMP, are the KtrAB orthologues CabP/SPD_0076 from *S. pneumoniae* and TrkH from *Streptococcus agalactiae* [71,79]. In *L. lactis*, a different K^+ transporter system has been discovered recently. Some strains contain a two-copy operon of the K^+ transporters KupA and KupB, which are both inhibited by binding c-di-AMP [72].

A different way of controlling K^+ transport by c-di-AMP occurs via gene regulation. In *S. aureus*, the expression of the K^+ transport system Kdp(F)ABC is activated by the two component system KdpDE. This activation is inhibited by c-di-AMP binding to the sensor kinase component KdpD [70]. However, for Kdp expressing species that lack the KdpDE component, a different pathway of regulation is required. In *Bacillus thuringiensis*, a riboswitch at the 5′-UTR of the *Kdp* transcript interacts with c-di-AMP and subsequently blocks *Kdp(F)ABC* transcription [80]. This riboswitch-mediated regulation by c-di-AMP is also observed for other K^+ transport systems. Particularly, the *B. subtilis* genome contains two copies of *ydaO*, a riboswitch located upstream of KimA and KtrA/B, respectively [73]. Binding of c-di-AMP to *ydaO* was shown to inhibit downstream gene expression [81]. Interestingly, both high affinity K^+ transporters are expressed at lower levels during high external K^+ concentrations. This coincides with a higher expression of CdaA and elevated c-di-AMP levels under these conditions. However, the mechanism how K^+ levels affect CdaA expression is not yet understood [73].

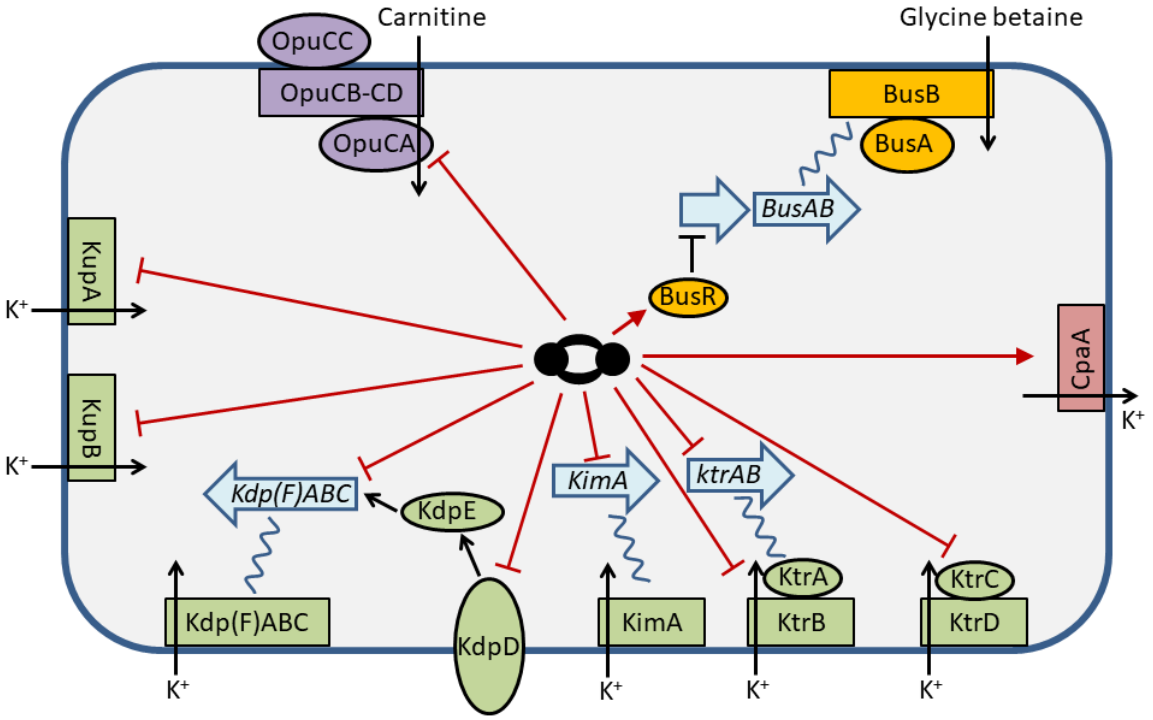


Figure 4: Schematic overview of c-di-AMP (black symbol) mediated osmolyte regulation identified in different species. Inhibition and activation processes are indicated by red arrows. Enzymes involved in K^+ uptake (green), K^+ export (red), glycine betaine uptake (orange) and carnitine uptake (purple) are depicted. Operon structures are shown as blue arrow boxes and expression pathways by waved lines. Adapted from F. M. Commichau, J. Gibhardt, S. Halbedel, J. Gundlach, and J. Stülke, "A Delicate Connection: c-di-AMP Affects Cell Integrity by Controlling Osmolyte Transport," Trends Microbiol., vol. 26, no. 3, pp. 175–185, 2018 with permission from Elsevier [82].

Besides the regulation of K^+ transport, also other osmolyte transport systems are presumably regulated by c-di-AMP. A common feature of these systems is the cystathionine-β-synthase domain (CBS), which in some proteins specifically binds c-di-AMP. For example, the ATPase subunit OpuCA of the carnitine transporter OpuC contains two CBS domains, which bind c-di-AMP. It was shown that c-di-AMP inhibits the osmolyte uptake of OpuC for *L. monocytogenes* and *S. aureus* homologues [83,84]. The crystal structure of the *L. monocytogenes* OpuCA CBS domain reveals a dimeric assembly with c-di-AMP located in the dimeric interface (Figure 5). It is assumed that c-di-AMP inhibits the ATPase activity of OpuCA, which is required for the osmolyte transport [84,85]. Another osmolyte transport system containing a CBS domain is the glycine betaine transporter BusAB. In contrast to OpuCA, the CBS domain of the BusA subunit cannot interact with c-di-AMP. In this case, the regulation by c-di-AMP is bridged by the transcription factor BusR. It binds c-di-AMP with its RCK_C domain and targets a sequence upstream of the *BusAB* operon. Elevated c-di-AMP levels were shown to reduce BusA/B expression and result in lower glycine betaine levels in *L. lactis* and *S. agalactiae* [79,86]. Nevertheless, the precise molecular mechanisms of BusR regulation by c-di-AMP remain elusive.

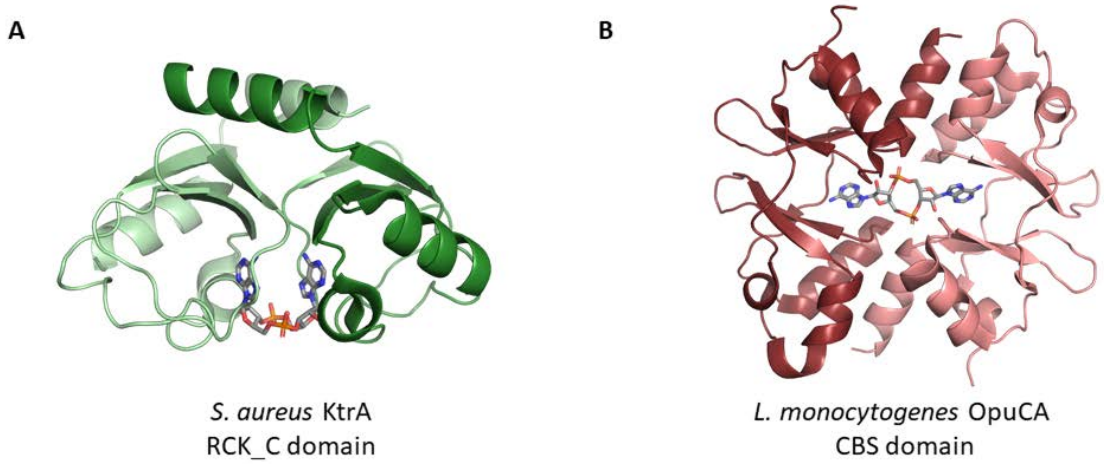


Figure 5: Crystal structures of two common c-di-AMP-binding domains. (A) RCK_C dimer of *S. aureus* KtrA (PDB 4XTT [78]). (B) CBS dimer of *L. monocytogenes* OpuCA in complex with c-di-AMP (PDB 5KS7 [84]).

The signalling and the regulation of K^+ import and other osmolytes is crucial for cell growth and osmoprotection and is most likely a major function of c-di-AMP. This leads to the suggestion that most of the phenotypes in c-di-AMP deficient cells, like perturbations in cell wall, cell growth and survival, could be caused by changes in the cellular turgor [73,82].

2.5. Biofilm formation and virulence

The formation of biofilms is an important ability of many bacteria to colonize on a surface and it often correlates with virulence of pathogens. The switch between motility and biofilm formation is a major function of c-di-GMP [87]. However, several studies have also reported a crucial role of c-di-AMP in this process [63,65,88–90]. For example, *S. aureus* cells with elevated c-di-AMP levels showed 3-fold higher biofilm amounts [63]. Similar phenotypes were observed in *S. mutans*, whereas the accumulation of c-di-AMP reduces biofilm formation in *B. subtilis* and *S. suis* [65,89,90]. Whether these functions are directly regulated by c-di-AMP or result from other cellular defects in c-di-AMP-altered cells, is not fully explored. Only for *S. mutans*, a recent study was able to propose a direct link to c-di-AMP [89]. During the formation of a biofilm, the cells synthesise extracellular polysaccharides (EPS) which are required for adhesion, protection and structure of the matrix [91]. A major polysaccharide synthase of *S. mutans* is GtfB. PDE deletion strains of *S. mutans* show a correlation between increased c-di-AMP levels and higher expression of GtfB. The *gtfB* gene regulator VicR is assumed to interact with the c-di-AMP receptor CabPA. According to the proposed pathway, increased levels of c-di-AMP stimulate the expression of GtfB via CabPA and VicR to induce biofilm formation [89].

In addition, increased levels of c-di-AMP lead to reduced virulence in the pathogens *L. monocytogenes*, *M. tuberculosis*, *S. pneumoniae*, *S. pyogenes* and *S. suis* [32,65,92–94]. Mice, infected with a PDE deletion strain of *L. monocytogenes* (Δ*pdeA* Δ*pgpH*) result in 3-log reduced bacterial growth in the analysed organs compared to wild type infections [94]. Similarly, the

respective *M. tuberculosis* PDE deletion (Δ*cnpB)* strain results in less infection phenotypes and extended survival rates in mouse models [93]. These observations highlight the relevance of the c-di-AMP signalling pathway as a target for antibacterial treatment.

2.6. Other functions

Besides the major pathways of c-di-AMP that have been described in the previous chapters, the second messenger is also found to bind targets, which have hitherto been less characterized.

The first discovered c-di-AMP receptor regulator was the TetR family transcription factor DarR from *M. smegmatis*. TetR family regulators are described to change their target DNA affinity dependent on small molecule binding. The negative gene expression regulator DarR was shown to bind c-di-AMP specifically, thereby increasing the affinity to its palindromic 14 bp target DNA. This sequence is present in three promotor regions, including the genes encoding for one cold shock protein and two proteins involved in fatty acid metabolism and transportation. Deletion of DarR leads to an increased cell length in *M. smegmatis* while overexpression results in inhibited cell growth [95]. However, the precise regulation of DarR by c-di-AMP is not well understood as an independent study and more detailed structural and biochemical investigations are missing.

The pyruvate carboxylase homologues LmPC from *L. monocytogenes*, EfPC from *Enterococcus faecalis* and LlPC from *L. lactis* were also shown to specifically bind c-di-AMP. These enzymes catalyze the carboxylation of pyruvate to oxaloacetate, an important component in the citric acid cycle. They form a tetramer with each monomer consisting of a biotin carboxyl carrier protein (BCCP), a biotin carboxylase domain (BC), a carboxyltransferase domain (CT) and a PC tetramerization domain (PT). In the first step, biotin is carboxylated by the BC. Consequently, the carboxyl group is transferred to pyruvate by the CT. This reaction requires large conformational changes, which are suggested to be inhibited by c-di-AMP. The crystal structures of LmPC and LlPC reveal the binding site of c-di-AMP in the dimer interface of two CT domains distinct from the pyruvate binding site. Compared to the apo structure, the single domains are twisted in the c-di-AMP bound structure, which is assumed to be an inactive conformation. The deletion of c-di-AMP in *L. monocytogenes* strains results in metabolic imbalance, increased synthesis of glutamine and disturbed intracellular growth. Thus, the regulation of PC by c-di-AMP seems to be important for the virulence of *L. monocytogenes* [96,97].

Another class of c-di-AMP-binding proteins are the P_{II}-like signal transduction proteins (PstA or DarA). P_{II} proteins are widely distributed among bacteria, archaea and plants and have a variety of functions and interaction partners. In bacteria, they often bind ATP/ADP and 2-oxoglutarate and are mainly associated with nitrogen metabolism. They share high sequence homology and have a common trimeric structure with a ferredoxin-like fold core domain and two flexible regions, called B-loop and T-loop [98]. The interaction with other proteins is diverse, as direct binding is

reported for the T-loop but also for the core domain [99,100]. The interaction of PstA with c-di-AMP was biochemically and structurally characterized for homologues from *S. aureus*, *L. monocytogenes* and *B. subtilis*. Although PstA has a low sequence homology, the conserved structure and assembly is highly similar to P_{II} proteins. However, a major difference is the swapped length of the B-loop and the T-loop, which are both flexible in the apo structure. The interface between the single monomers represents the binding site for c-di-AMP. Although there are no major conformational changes between the ligand-free and c-di-AMP bound structure, c-di-AMP binding is supposed to stabilize the T-loop. The highly flexible B-loop is not resolved in the structures, wherefore no information on influences from c-di-AMP binding is available. Presumably, a rearrangement of one or both loops is important for association with a downstream interaction partner, which has not been identified. So far, no phenotypes of PstA deletion strains were observed and further investigation is required to elucidate the function of these proteins [101–104].

During the search for c-di-AMP interaction partners, further c-di-AMP-binding proteins were identified but have not been characterized. In particular, a DRaCALA (differential radial capillary action of ligand assay) screening assay with proteins from *B. subtilis* containing either a RCK_C or a CBS domain revealed binding of c-di-AMP for the K^+/H^+ antiporter KhtT, the primary Mg^{2+} transporter MgtE and a protein of unknown function, called DarB [105]. Furthermore, a proteomics search for c-di-AMP-binding proteins in *L. monocytogenes* identified the transcription factor Nrd as potential interaction partner, but so far no characterization except for c-di-AMP binding has been performed [96].

Several functions of c-di-AMP have been identified but the precise signalling pathway and the connection to different phenotypes is still not yet fully understood. The roles of c-di-AMP differ among bacteria and the search for further interaction partners will be a crucial part in future investigations.

3. Degradation of bacterial cyclic dinucleotides

In addition to synthesis, the levels of a second messenger in the cell can be regulated by degradation or export. Cyclic dinucleotides are degraded by cleavage of their phosphodiester linkages by phosphodiesterases (PDEs). The hydrolysis of phosphodiester linkages is an important enzymatic reaction in all kingdoms for the degradation of DNA, RNA or small molecules. In case of second messengers, degradation is crucial to regulate their levels precisely. Therefore, the hydrolyzing PDEs have to be very specific. Cyclic dinucleotide specific PDEs share a two metal ion-dependent hydrolysis mechanism, as common for the cleavage of phosphodiester linkages [106] (Figure 6). However, cyclic dinucleotides are hydrolyzed by different PDE classes, which are described in the following chapters.

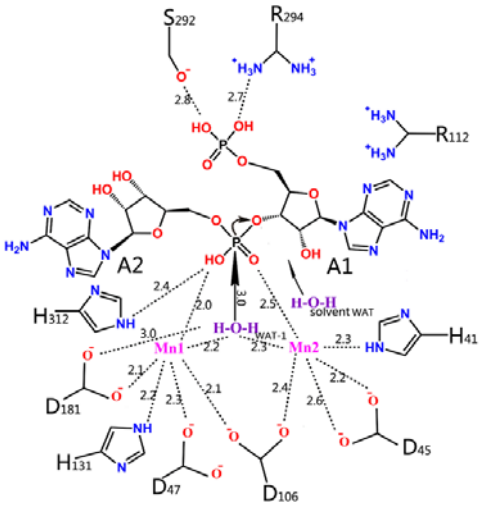


Figure 6: The two metal ion-dependent hydrolysis of 5′-pApA by Rv2837c. The two highly coordinated Mn^{2+} ions bring a water molecule in position for a nucleophilic attack on the phosphate of the phosphodiester linkage of 5′-pApA. Originally published by He *et al.* [107]

3.1. Degradation of c-di-AMP

Two major catalytic domains have been discovered to hydrolyze c-di-AMP specifically. They are classified according to their conserved active site motifs into DHH-type (Asp-His-His) or HD-type (His-Asp) PDEs. DHH domains can be part of the membrane-coupled multidomain PDEs of the GdpP-type or occur as cytosolic standalone domains termed DhhP-type PDEs. However, it was also shown that c-di-AMP can be secreted from *L. monocytogenes* and *M. tuberculosis* during infection, which represents another possibility to regulate c-di-AMP levels [93,108]. More recently, an extracellular c-di-AMP specific ectonucleotidase named CdnP has been discovered in *S. agalactiae*, which is responsible for the hydrolysis of secreted c-di-AMP to reduce innate immune response from the infected host [109].

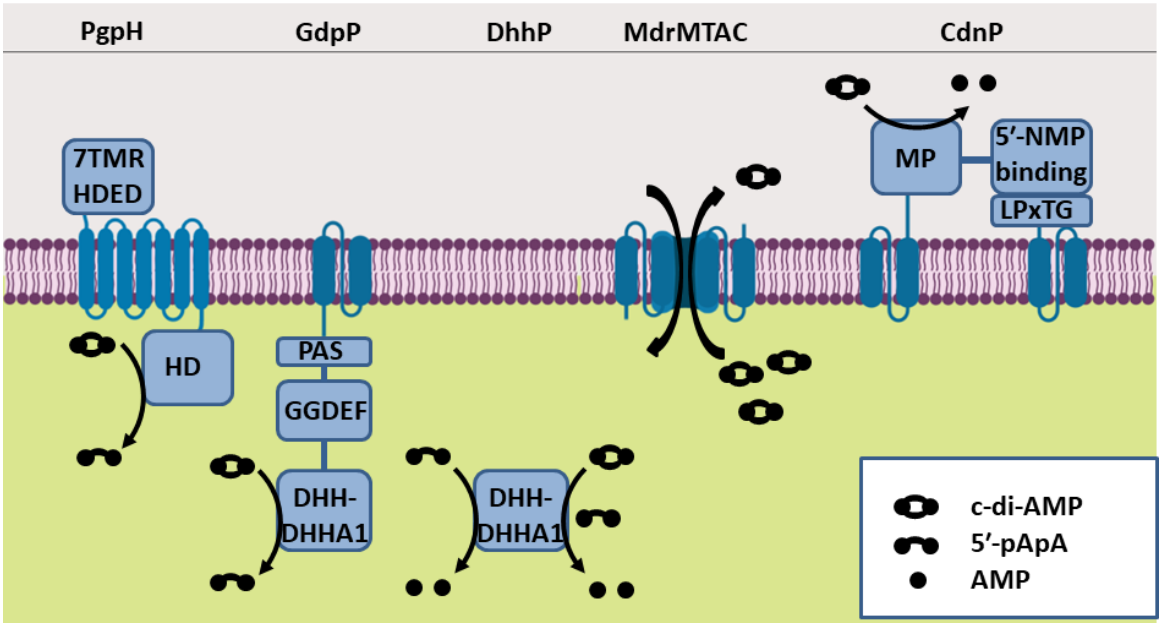


Figure 7: Schematic overview of the degradation pathways of c-di-AMP. The membrane-coupled PDEs PgpH and GdpP catalyse the hydrolysis of c-di-AMP to 5′-pApA. The soluble DhhP PDE hydrolyses the hydrolysis of c-di-AMP to AMP or exclusively the hydrolysis of 5′-pApA to AMP, depending on the species homologue. The MdrMTAC transport systems were are required for secretion of c-di-AMP. CdnP hydrolyses secreted c-di-AMP to AMP. Adapted from T. A. N. Huynh and J. J. Woodward, "Too much of a good thing: Regulated depletion of c-di-AMP in the bacterial cytoplasm," Curr. Opin. Microbiol., vol. 30, pp. 22–29, 2016, with permission from Elsevier [110].

3.1.1. DHH-type PDEs

DHH-type PDEs are members of the DHH-DHHA1 domain superfamily, which comprises exonucleases (RecJ) [111], oligoribonucleases (NrnA) [112] and PDEs (GdpP and DhhP) [113,114]. They share a highly similar structure and a two metal ion-dependent reaction mechanism, which was described for DHH-type PDEs [107,112,113,115,116], RecJ-like proteins [117] and the nanoRNAse NrnA [118]. The c-di-AMP specific DHH-type PDEs occur in two different versions, referred to as GdpP or DhhP, according to their domain architecture and substrate specificity.

The GdpP (GGDEF domain protein containing phosphodiesterase) consists of a transmembrane part, a Per-Arnt-Sim (PAS) domain, a degenerated GGDEF domain and the catalytic DHH-DHHA1 domain. It was the first c-di-AMP specific PDE to be discovered in 2010 by Rao *et al.* [113]. The hydrolysis of c-di-AMP is catalyzed by the DHH-DHHA1 domain and results exclusively in the linear dinucleotide intermediate 5′-pApA. The reaction is metal ion-dependent with the highest activity observed for Mn^{2+} or Co^{2+} ions [113]. The additional domains PAS and GGDEF are supposed to possess a regulatory function. PAS domains occur as small molecule binders with different functions including redox-sensing with O_2 or CO as ligands. For *B. subtilis* and *Geobacillus thermodenitrificans* GdpP, the PAS domains were shown to bind b-type heme and thereby strongly inhibit c-di-AMP hydrolysis [119]. Although lacking the typical PAS-conserved histidine or cysteine residues for heme-binding, the NMR structure of the PAS

domain from *G. thermodenitrificans* GdpP reveals a suitable binding pocket for heme [120]. In addition, the complex with ferrous heme is able to coordinate carbon monoxide (CO), cyanide (CN^-) and nitric oxide (NO), however, reactivation of c-di-AMP hydrolysis was only observed for NO binding. NO is known to be involved in antibacterial host immune response leading to the suggestion that GdpP PDEs are sensing NO [119]. Another possible regulation of GdpP PDEs might be mediated by the degenerate GGDEF domain. The majority of GGDEF homologues catalyse the synthesis of c-di-GMP. On the other hand, several degenerate GGDEF domains that lack the conserved motif residues can only bind or hydrolyse GTP without producing c-di-GMP. They are often present as additional domains in c-di-GMP-degrading proteins. In *Caulobacter crescentus*, a correlation between stimulation of the c-di-GMP hydrolysing EAL domain and binding of GTP by the GGDEF has been observed [121,122]. It is tempting to speculate about an analogous function for the GGDEF domain of GdpP. On the contrary, only weak ATPase activity was observed for *B. subtilis* GdpP (YybT) and no difference in c-di-AMP hydrolysis could be shown [113]. Furthermore, no full-length structure of GdpP is available yet to describe the mechanism of signal transduction between PAS or GGDEF to the catalytic DHH-DHHA1 domain. The precise membrane association and domain organization of these proteins is not characterized either.

The DhhP (also Pde2) PDE is a soluble stand-alone DHH-DHHA1 domain, first described from *S. pneumoniae* by Bai *et al.* in 2013 [32]. In contrast to GdpP, the DhhP PDEs can accept different substrates depending on the species homologue. The DhhP homologues of *Borrelia burgdorferi*, *M. tuberculosis* and *M. smegmatis* hydrolyse c-di-AMP into both, 5′-pApA and the single nucleotide AMP as final product [45,107,114]. On the other hand, DhhP analogues from *S. pneumoniae*, *S. aureus*, *T. maritima* and *S. mutans* have a clear preference for linear dinucleotide substrates, like 5′-pApA, while the hydrolysis of c-di-AMP could not be shown under physiological conditions [32,55,115,123]. As DhhP PDEs can occur in addition to GdpP PDEs, a two-step degradation pathway for c-di-AMP is suggested for these species. In the first step, GdpP hydrolyses c-di-AMP to the linear intermediate 5′-pApA and a DhhP catalyses the second step from 5′-pApA to AMP [115].

Structural information of the DHH-DHHA1 domain is available from different species homologues (Figure 8). Crystal structures were obtained from the DhhP homologues Rv2837c from *M. tuberculosis*, MsPDE (or NrnA) from *M. smegmatis* and TmPDE from *T. maritima* [107,115]. A full-length structure of GdpP is not solved yet but the separate DHH-DHHA1 construct from *S. aureus* could be crystallized in a recent study [116]. DHH-DHHA1 domains form butterfly shaped dimers with each monomer consisting of the two subunits DHH and DHHA1 connected by a flexible loop (Figure 8A and 8C). A cleft between the two subunits represents the active site, including the DHH motif from the DHH part and a

conserved GGG(H) stretch in the DHHA1 part [107,112,115,116,118] (Figure 8B). The highly conserved aspartate and histidine residues of the DHH motif coordinate the metal ions, which are required for activity [107,113,115,118]. In addition to the DHH motif, two aspartate residues (D420 and D499 in *B. subtilis* GdpP, D80 and D154 in *T. maritima* TmPDE) were shown to be crucial for metal ion binding as a mutation to alanine or asparagine resulted in catalytic inactivation [113,115]. For TmPDE it could be shown that these residues do not affect substrate binding [115]. The substrate coordination depends on the dinucleotide ligand and the species homologue. The active site can be divided into 3 nucleoside binding subsites C, R and G. The C site contains a conserved GGG(H) stretch, which is responsible for the main interactions to one of the nucleotide residues. The second nucleotide residue of c-di-AMP and 5′-pApA is coordinated differently (Figure 8D). The catalytic DHH domain of *S. aureus* GdpP binds and hydrolyses c-di-AMP in GC position. In contrast, 5′-pApA is coordinated in the RC site of Rv2837c and TmPDE. It is assumed that the smaller R-site of GdpP does not allow 5′-pApA to enter the RC position and thus prevents its hydrolysis [116].

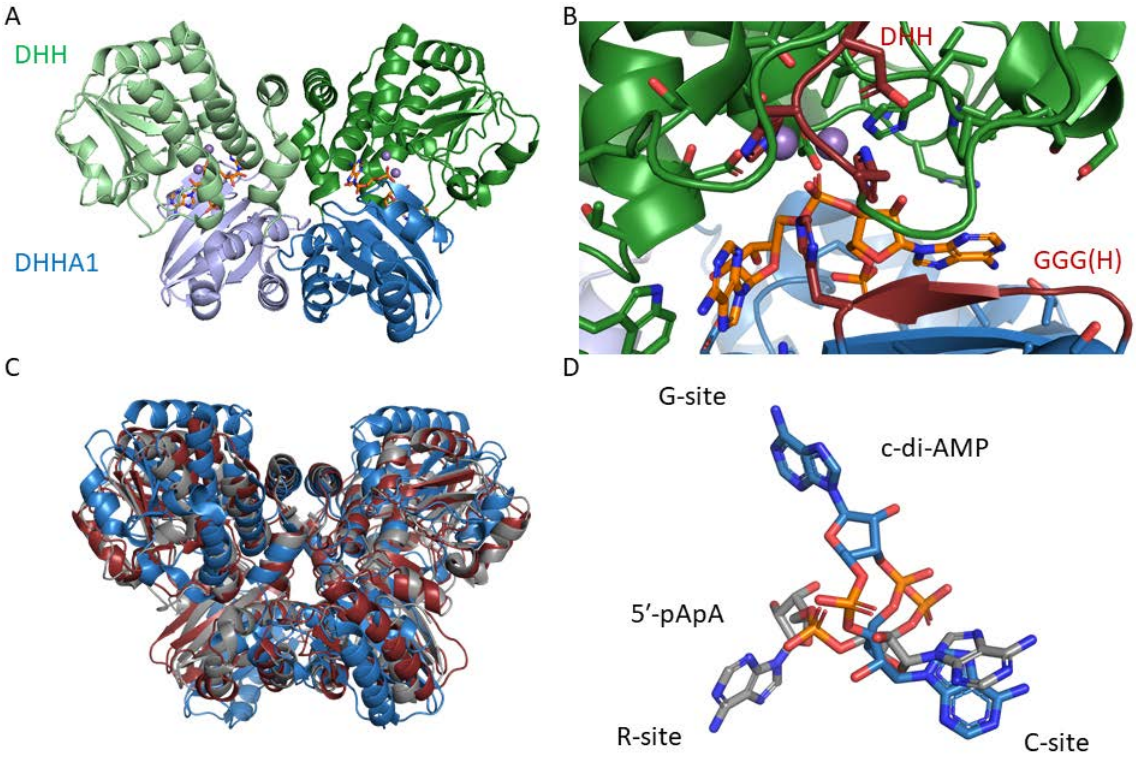


Figure 8: Crystal structures of the DHH domain from different species. (A) Rv2837c from *M. tuberculosis* in complex with 5′-pApA (orange) and two Mn^{2+} ions (purple), divided into the DHH domain (green) and the DHHA1 domain (blue) (PDB 5JJU [107]) (B) Close up view on the active site of Rv2837c in complex with 5′-pApA with depicted conserved motifs (red) (PDB 5JJU [107]). (C) Superimposition of Rv2837c in complex with 5′-pApA (gray, PDB 5JJU [107]), TmPDE in complex with 5′-pApA (red, PDB 5O4Z [115]) and the DHH-DHHA1 domain from *S. aureus* GdpP in complex with c-di-AMP (blue, PDB 5XSN [116]). (D) Superimposition of the conformation of 5′-pApA in Rv2837c (gray, PDB 5JJU [107]) and c-di-AMP in the active site of *S. aureus* GdpP (blue, PDB 5XSN [116]).

DHH-type PDEs have a significant role in the degradation of c-di-AMP. The deletion of these PDEs results in accumulation of c-di-AMP and defective cell functions. For example, the deletion

of the GdpP homologue in *S. aureus* leads to a smaller cell size and an increased PG cross-linking in the cell wall. In addition, a higher resistance against cell wall targeting antibiotics [63]. Furthermore reduced virulence was observed in PDE deficient strains of many pathogens, such as *M. tuberculosis*, *S. pneumoniae*, *S. suis* or *S. pyogenes* [32,65,92,93,124]. The deletion of the DhhP homologue Rv2837 in *M. tuberculosis* results in increased levels of secreted c-di-AMP and a higher innate immune response from the host cell in infection mouse models [93]. In *S. pneumoniae*, the homologues of DhhP and GdpP are important for growth and essential for virulence in infection models [32]. Similar effects have also been observed in GdpP deletion strains of *S. suis* [65]. In conclusion, the DHH-type PDEs play a crucial role for growth and virulence of many bacterial species, as they contribute significantly to the regulation of c-di-AMP levels.

3.1.2. HD-type PDEs

HD-type PDEs belong to a superfamily of phosphohydrolases with different substrates, including HD-GYP domain proteins for hydrolysis of c-di-GMP and 3′3′-cGAMP. One class of HD-type PDEs was also shown to be specific for c-di-AMP hydrolysis. These proteins are widely distributed in many c-di-AMP producing species and are referred to as PgpH, after the first described homologue from *L. monocytogenes*. PgpH PDEs are membrane-associated via seven transmembrane helices and consist of an N-terminal extracellular domain and a C-terminal cytoplasmic HD-domain [94]. This domain architecture is referred to as 7TMR-HD family, which consists of mainly uncharacterized prokaryotic surface receptors [125]. While the structure and the role of the extracellular domain have not been described yet, the single HD domain, containing the conserved His-Asp motif, has been identified to hydrolyze c-di-AMP. Similar to GdpP, PgpH hydrolyses c-di-AMP exclusively to 5′-pApA and has significantly lower activity for c-di-GMP. PgpH is dependent on Mn^{2+} ions, while Fe^{2+} ions inhibit the activity. The crystal structure reveals a similar reaction mechanism to the previously described DHH-type hydrolysis. Two metal ions coordinate the oxygens from one phosphodiester bond and activate a water molecule to perform a nucleophilic attack on the phosphate. At this, the histidine and aspartate residues of the highly conserved HD motif are essential for coordinating the metal ions and thus for hydrolysis activity. The activity of PgpH was shown to be inhibited by ppGpp with an IC_{50} value of approximately 200 – 400 µM. The all-helical overall structure of PgpH (Figure 9) is similar to HD-GYP PDE structures although they only share <20% sequence homology [94].

Homologues of PgpH can be found in many c-di-AMP producing species and are assumed to contribute significantly to the degradation of c-di-AMP. They can exist in parallel to GdpP and DhhP PDEs as found in *L. monocytogenes*. Interestingly, the activity of GdpP and PgpH was shown to be dependent on the cellular state of *L. monocytogenes*. During broth growth, the impact on c-di-AMP is higher for PdeA (GdpP homologue) and during intracellular growth inside host

cells, c-di-AMP levels are affected particularly by PgpH [94]. For *B. subtilis*, a deletion of PgpH results in a 2-fold higher c-di-AMP concentration compared to the deletion of GdpP [126]. This indicates an important role for PgpH in the degradation of c-di-AMP, although more investigations are required to characterize the activity and the role of the extracellular domain.

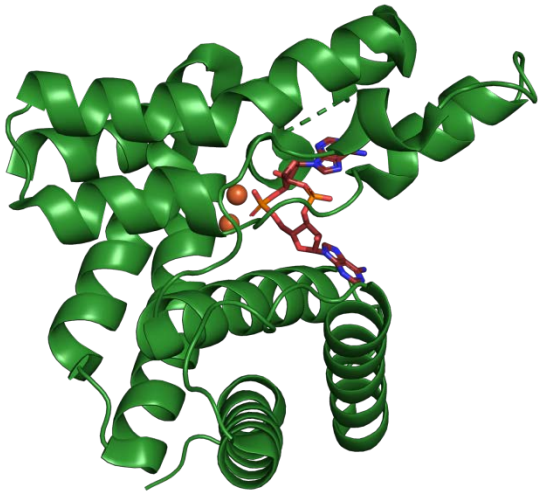


Figure 9: Crystal structure of *L. monocytogenes* PgpH in complex with two Fe^{2+} ions (orange) and c-di-AMP (PDB 4S1B [94]).

3.1.3. Export of c-di-AMP and detection by the host cell

In addition to the hydrolysis by PDEs, the intracellular c-di-AMP levels can also be regulated by export. The secretion of c-di-AMP has been observed in several species including the pathogens *M. tuberculosis* and *L. monocytogenes* [64,108,127–129]. The mechanism of c-di-AMP secretion was described to be dependent on multidrug efflux pumps (MDRs) of the MFS superfamily, as first described by Woodward *et al.* for *L. monocytogenes* [108]. MDRs are important transport systems for a variety of unspecific and structurally unrelated molecules in all kingdoms. Their function in drug export from bacteria makes them particularly interesting in regard to antibiotic resistance [130]. In *L. monocytogenes*, the four MDRs MdrM, MdrT, MdrA and MdrC (MdrMTAC) have been identified to secrete c-di-AMP [108,127,131,132]. This secretion contributes significantly to the degradation of c-di-AMP levels. Increased c-di-AMP levels in ΔMdrMTAC knockout strains result in altered peptidoglycan production and increased susceptibility to vancomycine [127]. Besides, the secretion of c-di-AMP by MDRs was also described in other species. Llmg1210 from *L. lactis* shares 45% sequence homology to MdrT and 36% to MdrM and is involved in c-di-AMP export as well. Accordingly, the overexpression of Llmg1210 together with the membrane protein Llmg1211 results in a complementation of osmoprotection in ΔGdpP strains [86]. Furthermore, the two putative c-di-AMP transporters ycnB and yhcA from *B. subtilis* share over 30% sequence homology to MdrM and MdrT and a double knockout strain leads to a significant decrease in secreted c-di-AMP. In *B. subtilis*, the secretion of c-di-AMP correlates with biofilm formation and plant attachment [88]. MdrMTAC transporters

are also involved in the response to lipoteichoic acid (LTA) stress [132]. LTAs are associated to membrane glycolipids and have important functions for the cell wall of gram positive bacteria [133]. The synthesis of LTAs is dependent on the described MDR pumps and a reduced synthesis of LTAs results in increased IFN-β response in infected cells. Although a link to c-di-AMP is tempting, there is no experimental evidence for a correlation yet [132].

Apart from the contribution to c-di-AMP degradation, the secretion of c-di-AMP has effects in infected host cells. Secreted c-di-AMP can be detected by the mammalian innate immune system, as shown in infection models with *M. tuberculosis*, *L. monocytogenes* or *Chlamydia trachomatis* [64,108,134]. So far, the three different host receptors STING, RECON and ERAdP have been identified to sense c-di-AMP. The ER membrane-coupled adaptor protein STING (stimulator of interferon genes) plays a key role in the mammalian innate immune system (in more detail in chapter 4.3.). Although the major function is signalling of the endogenous second messenger 2′3′-cGAMP, STING can also detect invading bacterial CDNs, such as c-di-AMP, resulting in the induction of a type I IFN response [128,129,134,135]. In a more recent study, the aldo-keto reductase RECON (reductase controlling NF-κB) was shown to have a significantly higher affinity for c-di-AMP than STING. The activity of RECON is dependent on NAD-binding, which is inhibited competitively by c-di-AMP. This inhibition results in the production of pro-inflammatory cytokines and bactericidal gene expression as observed in *L. monocytogenes* infection models of murine macrophages. In addition, the induction of a type I interferon response is diminished by overexpression of RECON, leading to the suggestion of reduced STING activation by c-di-AMP under these conditions [136]. Another ER-located receptor, termed ERAdP (Endoplasmatic Reticulum Adaptor Protein), was also shown to bind c-di-AMP with a higher affinity compared to STING [137]. This adaptor protein is found in the two transcript isoforms *Eradp*, which is mainly expressed in NK cells and cytotoxic T cells, and *Cnep1r1*, which is also present in monocytes and macrophages. Specific binding of c-di-AMP leads to dimerization of ERAdP and association to the kinase TAK1. This results in activation of NF-κB and subsequently expression of pro-inflammatory cytokines. The role of ERAdP for immune response was confirmed in infection models of mice with *L. monocytogenes* [137]. However, more investigations are highly required to elucidate, whether one of these pathways is mainly activated or whether these receptors are abundant in specific cell types. Besides, it was shown that the detection of c-di-AMP can also be prevented by the bacterial cell. In *S. agalactiae*, the secreted c-di-AMP is hydrolyzed by the extracellular ectonucleotidase CdnP to inhibit an innate immune response [109].

3.1.4. The ectonucleotidase CdnP

In a recent study from Andrade *et al.*, the ectonucleotidase CdnP has been discovered to hydrolyse extracellular c-di-AMP in *S. agalactiae* (Group B *Streptococcus*) [109]. CdnP consists

of a metallophosphatase domain including a conserved NHE motif, a 5′-nucleotide-binding site and a transmembrane part at each terminus, one of them containing an LPxTG cell wall anchoring motif. Amongst bacterial nucleotide second messengers, CdnP has a clear preference for hydrolysis of c-di-AMP to AMP. However, also weak activity for c-di-GMP and 3′3′-cGAMP have been observed. The histidine from the conserved NHE motif is essential for hydrolysis activity as a mutation leads to full inactivation. CdnP acts in concert with the ectonucleotidase NudP, which further hydrolyzes AMP to adenosine and phosphate. Inactivation of CdnP in *S. agalactiae* does not affect c-di-AMP related intracellular functions but the levels of extracellular c-di-AMP are increased significantly [109]. The secretion of c-di-AMP is assumed to be an important additional regulatory mechanism of intracellular c-di-AMP levels. However, bacterial nucleotide second messenger are detected by the host and activate innate immune response [64,108,134,135]. CdnP is therefore suggested to be a regulator of extracellular c-di-AMP levels to prevent an innate immune response [109].

3.2. Degradation of c-di-GMP and 3′3′-cGAMP

The degradation of c-di-GMP is catalyzed by EAL (Glu-Ala-Leu) domains or by HD-GYP (His-Asp-Gly-Tyr-Pro) domains [20]. Although both can occur as stand-alone domains, they are often N-terminally linked to various sensory input domains as part of multidomain proteins [20,87,138]. Therefore, the levels of c-di-GMP are suggested to be regulated through synthesis and degradation by different environmental and cellular signals [87].

The major role of c-di-GMP in bacteria is the switch between motility and biofilm formation [20]. In order to ensure a correct signalling, the levels of c-di-GMP have to be regulated by degradation. The first characterized hydrolysis of c-di-GMP was described for EAL domains [139]. *In vitro* experiments confirmed hydrolysis of c-di-GMP to 5′-pGpG whereas no activity for c-di-AMP was observed. The second step of hydrolysis, the conversion of 5′-pGpG to the final product 5′-GMP, is very slow and probably performed by other enzymes [20]. At this, enzymatic activity is dependent on Mn^{2+} or Mg^{2+} ions and is inhibited by Ca^{2+} ions [140]. Structural approaches revealed a dimeric assembly for EAL domains, which appears to be required for activation. The glutamate residue of the conserved EAL motif coordinates one metal ion and is essential for activity, as a mutation to glutamine leads to inactivation. The hydrolysis of c-di-GMP is catalyzed by two metal ions activating a water molecule, which performs a nucleophilic attack on the phosphodiester linkage, which is a common mechanism for metal ion-dependent PDEs. In contrast to the bent conformation in DGCs and other receptors, c-di-GMP is bound in a stretched conformation in the EAL domain to facilitate the nucleophilic attack on the phosphate [20,141,142]. EAL domains are often found in combination with regulatory domains. In particular, ~2/3 of all EAL domains are coupled to GGDEF domains of which ~40% are predicted to lack DGC activity [20,143]. However, the majority of such proteins contain one or

both domains in a catalytically inactive state. Only a small number was shown to be bifunctional [20]. One possible reason for the coexistence of two domains with opposite functions is the possibility to activate only one of them at a time. For example, the membrane-bound regulator MorA from *P. aeruginosa* contains both, GGDEF and EAL domains, in a functional state [144]. In *Vibrio parahaemolyticus*, the activities of GGDEF and EAL of ScrC are suggested to be regulated by the interaction partners SrcA and SrcB [145]. The structure of another bifunctional protein, RbdA from *P. aeruginosa* is available with GGDEF bound to GTP and EAL bound to c-di-GMP (Figure 10). Each monomer of the RbdA dimer consists of two transmembrane helices, a PAS domain, a GGDEF domain and an EAL domain. The crystal structure reveals an autoinhibited state, as the active sites of GGDEF and EAL face each other and are not accessible for substrates. Activity assays and SAXS analysis indicate a conformational change upon GTP-binding and allosteric activation of the EAL domain [146,147].

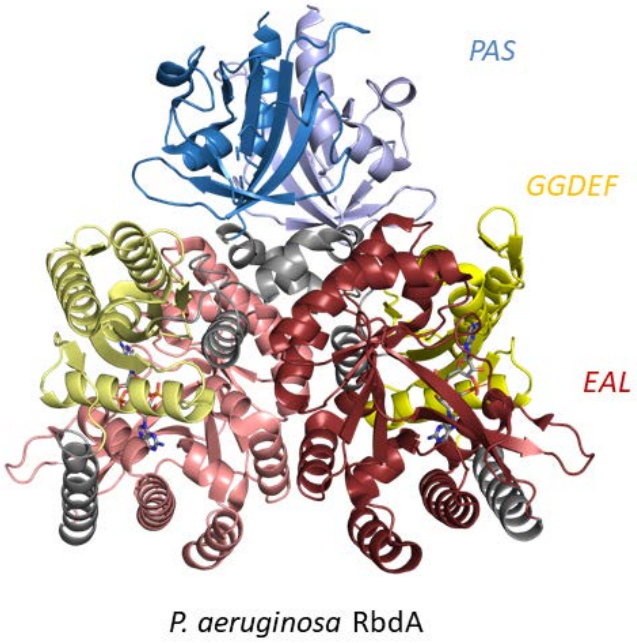


Figure 10: Crystal structure of *P. aeruginosa* RbdA in complex with c-di-GMP, PAS domain dimer (light and dark blue), GGDEF domains (light and dark yellow) and EAL domain dimer (light and dark red) are depicted (PDB 5XGE [147]).

The second type of c-di-GMP-degrading domains are HD-GYP domains, which are a subgroup of the HD-domain family. HD-GYP domains are present in different proteins with structural and functional varieties [148]. The first hydrolysis of c-di-GMP by an HD-GYP domain was characterized for the regulatory protein RpfG from *Xanthomonas campestris*. In contrast to EAL domains, which produce mainly 5′-pGpG, the analyzed reaction product of RpfG was solely GMP [149]. The HD-GYP domain proteins PA4781 and PA4108 from *P. aeruginosa* show 5′-pGpG as intermediate product, which is bound with higher affinity and has a lower turnover compared to c-di-GMP [150]. Deletion of these two proteins leads to increased c-di-GMP levels,

reduced swarming motility and altered virulence [151]. However, HD-GYP domains can also be inactive for c-di-GMP hydrolysis, like the homologue in Bd1817 from *Bdellovibrio bacteriovorus*, which is lacking the conserved tyrosine residue [152]. Structural information is available from crystal structures of *B. bacteriovorus* Bd1817, *Persephonella marina* PmGH and *P. aeruginosa* PA4781. They all share an all helical fold and a comparable active site, which is distinct from EAL domains (Figure 11). However, differences between the HD-GYP domains can be observed in the active site and the metal ion coordination. While Bd1817 and PA4781 coordinate the metal ions in two positions, PmGH has a trinuclear metal active site (M1, M2, M3). The positions M1 and M2 are occupied in Bd1817 whereas M1 and M3 are occupied in PA4781 [148,152–154]. PmGH is the only protein, for which a structure with complexed c-di-GMP could be solved. The dinucleotide is coordinated in a bent conformation to locate the phosphodiester linkage in close proximity to the metal-binding site for hydrolysis [153]. The conserved GYP loop in PmGH and PA4781 creates a kink in a helix (helix 3), which leads to an open conformation of a "lid" loop on top of the active site [148,153]. In Bd1817, this lid is closed, restricting the access to the active site [152]. HD-GYP proteins were also shown to be regulated by their N-terminal domain. For example, the REC domain of PA4781 has to be phosphorylated for activity [150]. In addition, HD-GYP domains are proposed to directly interact with other proteins for regulation of their activity [148,152,153]. For many HD-GYP domains, it remains elusive whether they are degenerated or have to be activated by so far unknown interaction partners.

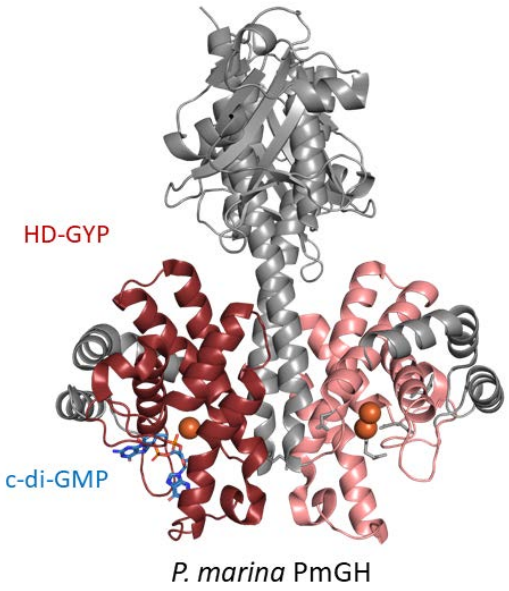


Figure 11: Crystal structure of *P. marina* PmGH in complex with c-di-GMP (blue) and Fe^{2+} ions (orange). HD-GYP domain (red) is depicted (PDB 4MDZ [153]).

In addition to c-di-GMP degradation, HD-GYP domains were also found to hydrolyze 3′3′-cGAMP. Three HD-GYP domain PDEs from *V. cholerae* were shown to be specific for

3′3′-cGAMP and were thus designated as V-cGAP1/2/3. V-cGAP2 and V-cGAP3 are converting 3′3′-cGAMP to the linear intermediate 5′-pApG. V-cGAP1 is additionally cleaving the linkage to the 5′-phosphate in a second step resulting in ApG as final product. All three PDEs are specific for a 3′3′-phosphodiester linkage as they are not able to hydrolyze 2′3′-cGAMP. It was shown that mutations in the HD or the GYP motif abolish activity and the deletion of V-cGAPs *in vivo* results in enhanced bacterial infectivity [28].

4. Cyclic dinucleotide second messengers in eukaryotes

4.1. 2′3′-cGAMP

Besides prokaryotes, signal transduction by a cyclic dinucleotide second messenger was also discovered in the mammalian innate immune system, which raised particular scientific interest [8,9]. The cyclic GMP-AMP synthase (cGAS) produces an isomer of cGAMP that is distinct from bacterial dinucleotide second messengers through its non-canonical phosphodiester linkage. Whereas all known bacterial CDNs are all linked by two 3′5′-phosphodiester bonds, the mammalian cGAMP is connected by one 2′5′ and one 3′5′-phosphodiester linkage and is therefore distinguished as 2′3′-cGAMP [10–12] (Figure 12).

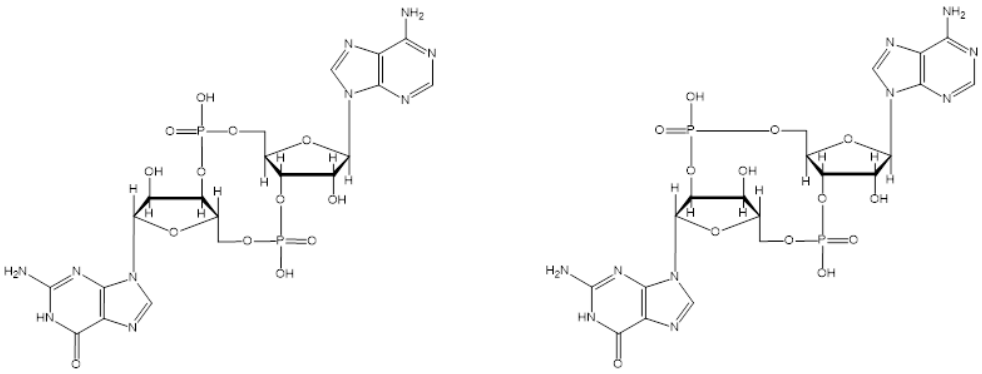


Figure 12: Structure comparison of the prokaryotic 3′3′-cGAMP and the eukaryotic 2′3′-cGAMP.

The innate immune system has to deal with continuous threats. Therefore, several pathways for recognition, signal transduction and response to pathogens have evolved. A crucial step in this system is the detection of foreign molecules, which are described as pathogen-associated molecular patterns (PAMPs) or damage-associated molecular patterns (DAMPs). This is performed by pattern recognition receptors (PRRs) located on the cell membrane, including toll-like receptors (TLRs) or C-type lectin receptors (CLRs), or by cytosolic receptors, such as NOD-like receptors (NLR) or RIG-I-like receptors (RLR) [155]. Another type of pathogen detection and signalling is the cGAS-STING pathway [8]. The synthesis of 2′3′-cGAMP by cGAS is the first step in recognition of foreign DNA or damage induced leakage of self-DNA in the

cytosol. The stimulator of interferon genes (STING) binds 2′3′-cGAMP and induces further signal transduction resulting in antiviral and pro-inflammatory gene expression [10–12]. The cGAS-STING pathway represents an extensive research field (reviewed in e.g. [156,157]), thus the following chapters focus on the structure and function of cGAS and STING.

4.2. Synthesis of 2′3′-cGAMP

The cyclic GMP-AMP synthase cGAS is a member of nucleotidyltransferase (NTase) family proteins, which are involved in diverse mechanisms, including RNA modification, DNA repair or signal transduction [158]. It consists of an unstructured N-terminal part and a C-terminal catalytic male abnormal 21 (Mab21) domain, which contains the NTase fold [8]. Although it has a different signalling pathway, cGAS is structurally and mechanistically highly similar to the 2′-5′-oligoadenylate synthase OAS1 [8,159,160]. OAS proteins sense dsRNA and produce 2′-5′-oligoadenylates to activate RNA degradation by RNase L and thus inhibit virus proliferation [161–163]. In a similar manner, cGAS senses double-stranded DNA (dsDNA) in the cytosol and synthesizes 2′3′-cGAMP from ATP and GTP to signal innate immune response to pathogens. Independent of the sequence, cGAS binds DNA and dimerizes. In addition, the activity of cGAS increases with DNA length. The current model proposes that binding of one cGAS dimer facilitates binding of further cGAS dimers, which is indicated by higher affinities for longer DNA. The crystal structure of a $cGAS_4$:DNA_2 complex supports this model and reveals a slightly bend DNA, which is assumed to be a crucial factor for oligomerization. It was also shown that this orientation of the DNA can be induced by other DNA interacting proteins, like bacterial nucleoid HU proteins, mitochondrial transcription factor A (TFAM) or the high mobility group box 1 protein (HMGB1) [164–166].

The catalytic Mab21 domain of cGAS consists of two lobes connected by a long “spine” helix. DNA interaction occurs on two sites of each cGAS monomer (Figure 13A). The spine helix and a zinc-thumb form “site A” and the second DNA-binding site is referred to as “site B”. Binding to DNA induces a “kink” in the spine helix that closes the lobes and makes the active site accessible for the substrates (Figure 13B). A loop rearrangement in the active site facilitates coordination of Mg^{2+}, ATP and GTP [11,167]. One nucleoside-binding site is highly specific for GTP. The other site coordinates the ATP base mainly through stacking interactions, thus it can also be occupied by GTP. Two Mg^{2+} ions coordinate the triphosphate of ATP to position the α-phosphate for a nucleophilic attack by the 2′-OH group of GTP, which results in the linear intermediate pppGpA. As this reaction can only be catalyzed on one site of the catalytic core, the linear intermediate has to flip or dissociate from the active site and rebind for cyclization. In the flipped position, the triphosphate of the guanosine is coordinated by the two Mg^{2+} ions to facilitate a nucleophilic attack of the 3′-OH group of the adenosine ribose. Since both binding pockets can be occupied by GTP, formation of c-di-GMP is observed as byproduct [10,11,167].

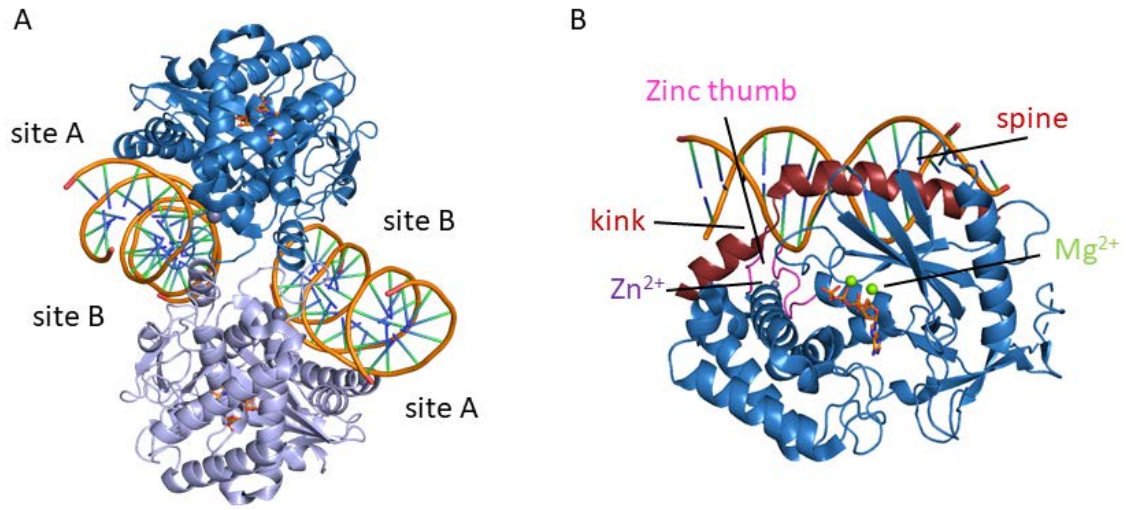


Figure 13: Structure of the Mab21 domain of cGAS. (A) Dimeric assembly of the Mab21 domain of mouse cGAS in complex with 18 bp DNA and 2′3′-cGAMP (PDB 4LEZ [166]). (B) Mab21 domain of mouse cGAS in complex with 16 bp DNA and ATP (PDB 4K97 [11]).

The N-terminus of cGAS (aa 1-161) is unstructured, poorly conserved and controversially described in recent studies. Despite the poor sequence homology, it shares a predominantly positive charge among cGAS homologues [168]. This enables the N-terminus to bind DNA unspecifically and independently from the Mab21 domain [8]. According to electromobility shift assays (EMSAs), cGAS constructs containing the N-terminus have a significantly higher affinity to DNA and have higher enzymatic activity compared to truncated constructs [169]. A recent study reports an accumulation of cGAS 161-522 (ΔN-terminus) and cGAS 1-161 (only N-terminus) in the nucleus, whereas a cGAS 1-212 construct is mainly located in the cytosol. In addition, it was shown that the N-terminus is required for centromeric DNA sensing [170]. Another indicated function of the N-terminus is the localization of cGAS at the plasma membrane. The N-terminus of cGAS binds to PI(4,5)P_2 (phosphatidylinositol 4,5-biphosphate) which is a component of cell membranes. Full-length cGAS expressed in THP1 cells is mainly present at the plasma membrane whereas the N-terminally truncated version accumulates in the cytosol and the nucleus [171]. Although these observations give possible suggestions about the function of the cGAS N-terminus, it remains not fully understood and a structural analysis is missing.

4.3. cGAS-STING signalling

Prior to the discovery of 2′3′-cGAMP, it was suggested that one type of innate immune pathway is mediated through direct sensing of invading bacterial CDNs by the adaptor protein STING. In recent years, several studies have revealed the cGAS-STING pathway and discovered STING activation by endogenous 2′3′-cGAMP. The central role of STING is the activation of the

interferon regulatory factor 3 (IRF3) and NF-κB transcription factors, resulting in the production of type I interferons and other cytokines. Binding of 2′3′-cGAMP induces oligomerization of STING, which is subsequently phosphorylated by the tank-binding kinase I (TBK1). This enables phosphorylation of IRF3 by STING [172] (Figure 14).

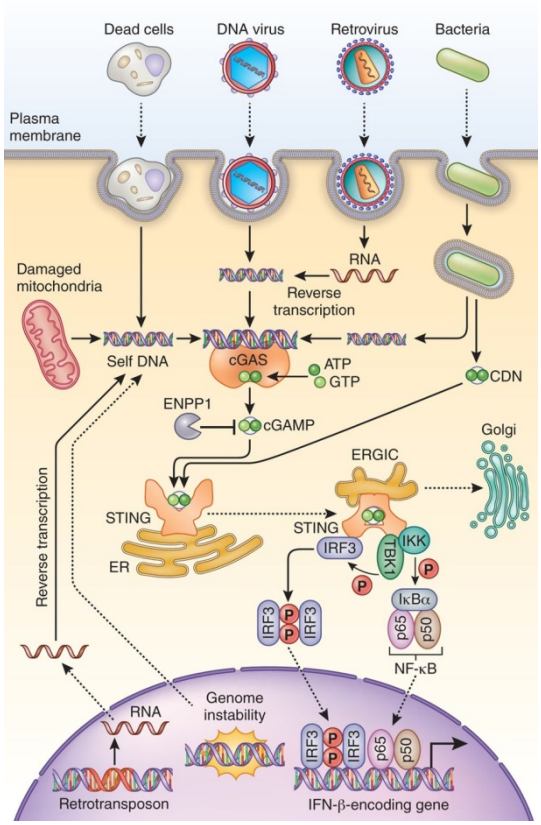


Figure 14: The cGAS-STING signalling pathway. DNA in the cytosol can originate from bacteria, retroviruses, DNA viruses, dead cells, damaged mitochondria or genome instability. The cGAMP synthase cGAS detects DNA in the cytosol and starts producing 2′3′-cGAMP from ATP and GTP. The ER coupled receptor STING can bind 2′3′-cGAMP or bacterial derived CDNs. Subsequently, STING translocates to the Golgi apparatus and phosphorylates IRF3 via TBK1 to induce a type I IFN response. ENPP1 regulates the 2′3′-cGAMP levels by hydrolysis. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Nature Immunology, Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing, Q. Chen, L. Sun, and Z. J. Chen, Copr. 2016 [157].

STING is a 42 kDa protein located at the endoplasmatic reticulum (ER) membrane. It consists of four N-terminal transmembrane helices, a cytosolic ligand-binding domain (LBD) and a C-terminal TBK1-binding motif (TBM). STING forms a dimer, which is twisted between the transmembrane part and the LBD in the inactive state (Figure 15A). The ligand-binding pocket is located between two “wings” of a V-shaped dimer of the LBDs. Binding to 2′3′-cGAMP induces a 180° rotation of the LBDs to the transmembrane part, resulting in a parallel orientation of the dimer (Figure 15B). This conformational change is assumed to be required for STING oligomerization. A cryo-EM structure of a chicken STING tetramer supports this model and reveals possible oligomerization interactions in a LBD loop. In addition, it was shown that STING phosphorylation by TBK1 requires an oligomeric assembly of STING. Independent of 2′3′-cGAMP activation, the TBK1 dimer binds to a C-terminal β-strand like motif of STING,

which is connected to the LBD by a flexible linker. The available complex structure of STING and TBK1 has two kinase domain active sites of TBK1 located too distant from the STING phosphorylation motif *p*LxIS (*p* for hydrophobic residue, x for any residue and S for the phosphorylation site). Thus, the current model proposes, that TBK1 phosphorylates this motif on the tail of a neighbouring STING dimer [173,174]. The phosphorylated *p*LxIS motif is subsequently bound by IRF3, as has also been observed for other adaptor proteins, like MAVS (mitochondrial antiviral signalling) or TRIF (TIR domain containing adapter inducing interferon β) [175]. This brings IRF3 and TBK1 in close proximity for IRF3 phosphorylation [173].

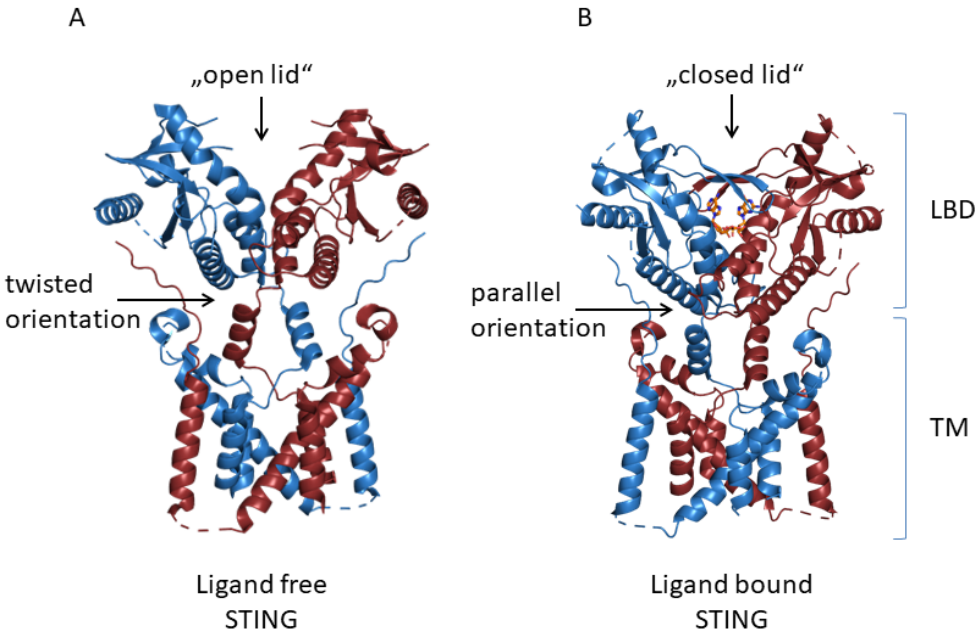


Figure 15: Full-length structure of chicken STING in (A) ligand-free state (PDB 6NT6 [174]) and (B) in complex with 2′3′-cGAMP, LBD = ligand-binding domain. TM = transmembrane part (PDB 6NT7 [174]).

Apart from these major orientational rearrangements, some minor conformational changes can be observed in the LBD upon ligand binding. The V-shaped LBD structure of STING closes upon 2′3′-cGAMP binding and a four stranded β-sheet forms a “lid” covering the binding pocket (Figure 15B). It is assumed that this closure correlates with the activation of STING for different ligands [176]. In addition to 2′3′-cGAMP, the bacterial dinucleotides c-di-AMP, c-di-GMP and 3′3′-cGAMP were also shown to bind and activate STING. However, the binding affinity and activation is significantly higher for 2′3′-cGAMP [177,178]. Mainly open lid conformations are observed in crystal structures of the human STING LBD in complex with c-di-GMP [179–181], except for one structure with a closed lid [182]. In contrast to human STING structures, mouse STING LBD structures share a closed conformation in complex with different CDNs and possess less ligand specificity [183,184]. Furthermore, STING exists in two allele variants, having a R232 or a H232 residue in the LBD that determines CDN binding specificity. According to a human

genome DNA screen for STING from over 1000 individuals, the R232 variant is predominant [185]. Interestingly, it was shown that both STING variants are selective for 2′3′-cGAMP, whereas only STING R232 also responds to bacterial CDNs [12]. Unfortunately, no structure of full-length STING bound by a bacterial CDN is available to elucidate major conformational changes upon bacterial CDN binding. Thus, further investigations are important to get a better understanding of the selective STING activation and its role in CDN sensing.

4.4. Therapeutic potential of the cGAS-STING pathway

In recent years, the cGAS-STING pathway was discovered to be a highly promising target for anti-cancer therapy. One important strategy to treat cancer is immunotherapy, which represents an emerging research field. Immunotherapy uses the body's own immune system to fight tumor cells, in particular by inducing adaptive anti tumor immunity. This can be achieved by different antibody-based approaches, such as bispecific antibodies or immune checkpoint inhibitors (ICI) or modified immune cells, including chimeric antigen receptor T cells (CAR-T). The cGAS-STING pathway has an important role in innate immunity and represents an additional target for immunotherapy. Cytosolic DNA is sensed by cGAS and induces a type I IFN response via STING. This pathway can also be induced by DNA leakage in tumor cells or tumor cell derived DNA in dendritic cells (DCs). The expression of type I IFN in DCs contributes to the induction of tumor specific cytotoxic T cell priming and results in a specific anti tumor immune response to tumor cells [186,187] (Figure 16).

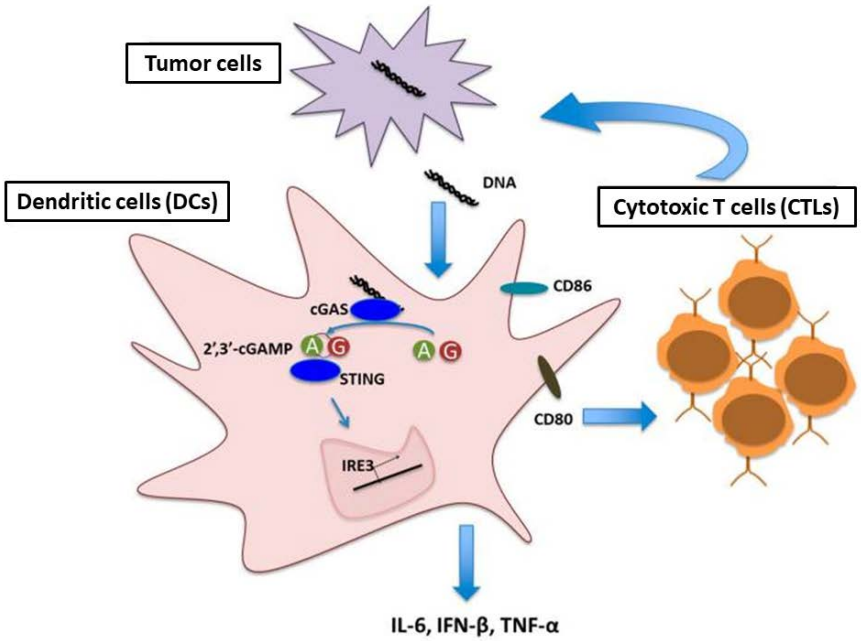


Figure 16: Schematic representation of the role of cGAS-STING pathway in anti-cancer treatment. Tumor cell-derived DNA can enter the cytosol of DCs. Upon detection, cGAS produces 2′3′-cGAMP, which activates STING resulting in type I IFN production. This induces priming of cytotoxic T cells that specifically attack the tumor cells. Adapted from D. Bose, "CGAS/STING pathway in cancer: Jekyll and hyde story of cancer immune response," Int. J. Mol. Sci., vol. 18, no. 11, 2017 [187]

In order to manipulate the induction of type I IFN production for tumor cell treatment, the cGAS-STING pathway could be activated artificially. This approach raised the search for STING agonists, with 5,6-dimethyllxanthenone-4-acetic acid (DMXAA, also Vadimezan or ASA404) (Novartis) as most prominent example. DMXAA was shown to bind and activate mouse STING and possess a high anti-tumor potential in mouse models [188–190]. However, it failed in phase III clinical trials [191] as it does not bind and activate the human homologue of STING [184,190,192]. A similar specificity was observed for the type I interferon inducer 10-carboxymethyl-9-acridanone (CMA) [193], which binds and activates mouse STING but fails to bind human STING [184]. Besides synthesized small molecule binders, also CDNs represent potential agonists. For example, c-di-GMP was shown to enhance anti-tumor effects as adjuvant in several approaches or directly attack tumor cells by activation of caspase-3. STING was shown to discriminate between canonical 3′3′-CDNs and noncanonical 2′3′-CDNs by higher binding affinity for the latter version. Therefore, STING agonists have been developed based on the noncanonical 2′3′-phosphodiester linkage. Noncanonical linked dithio 2′3′-CDNs were shown to bind STING and have a potential in anti-cancer immune response, as for example the derivative ML RR-S2 CDA (also ADU-S100), which was developed by Aduro Biotech [190]. The efficacy of ADU-S100 is currently investigated in phase I clinical trials. One major disadvantage of CDNs is the incapability to cross the cellular membrane. Therefore a carrier molecule is required [186].

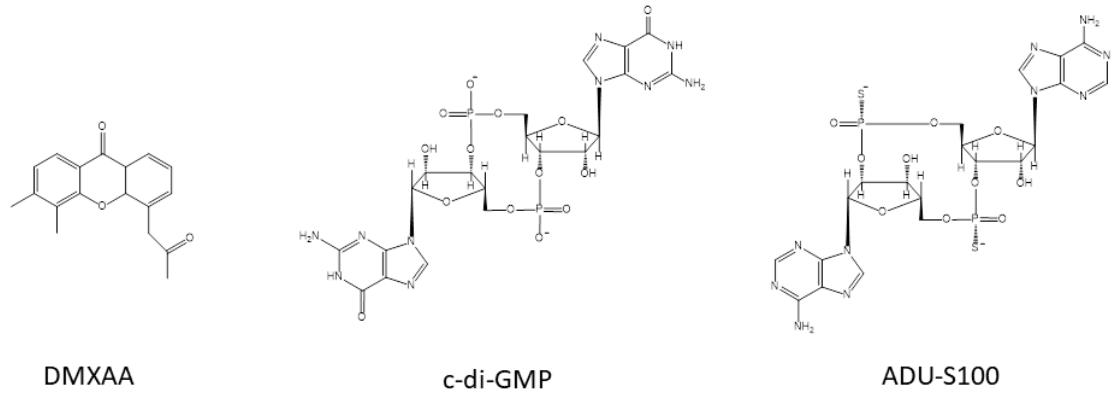


Figure 17: Structures of the STING agonists DMXAA (Novartis), c-di-GMP and ADU-S100 (Aduro Biotech).

Objectives

In recent years, cyclic dinucleotides have been discovered to play a role as second messengers in many important signalling pathways in both, prokaryotes and eukaryotes.

In prokaryotes, c-di-AMP is an important second messenger as it is essential in most species. Since it was discovered only recently, many functions of the c-di-AMP signalling pathway are not yet fully understood. In particular, the degradation pathway raises many open questions and thus represents the major topic of this work. Two of the so far discovered PDE types, GdpP and DhhP, use the same highly conserved DHH-DHHA1 domain for catalysis although they occur in parallel in many species. While GdpP hydrolyzes c-di-AMP exclusively to 5′-pApA, the substrate specificity of DhhP homologues varies in literature. One major aim of this work was to characterize the structure and function of DhhP PDEs based on the homologue TmPDE from *T. maritima*. In order to determine the substrate specificity of TmPDE, activity assays and binding experiments were performed. Different cyclic dinucleotides, linear dinucleotides or short RNAs were tested as possible substrates. Crystallization, small-angle x-ray scattering (SAXS) and right-angle light scattering (RALS) were used for structural characterization. Based on the structural information, an inactive TmPDE D80N D154N was designed for binding experiments. The results were compared to DhhP homologues from other species, such as SpPde2 and Rv2837c, to describe the function of these enzymes. Although these findings contribute to a better understanding of c-di-AMP degradation, the hydrolysis pathway remains unclear in some species, such as Streptomyces, that do not contain any of the known PDEs. The discovery of a novel PDE for c-di-AMP hydrolysis in *S. venezuelae* by the group of Natalia Tschowri (Humboldt-Universität Berlin) gives important insights to unravel this unknown degradation pathway. In a collaborative project with the Tschowri group the precise function of this enzyme should be elucidated. Therefore, the second major part of this work was the biochemical characterization of this novel PDE, named AtaC. The substrate specificity was analyzed using different dinucleotide ligands in activity assays. In addition, several approaches to gain structural information were tested.

Cyclic dinucleotides also play an important role in eukaryotes. The recently discovered 2′3′-cGAMP functions as a second messenger in the cGAS-STING pathway of the innate immune system. It is involved in the detection of pathogens and represents the first signal for cytosolic DNA recognition. The pathway raised particular scientific interest, since it was found to be a potential target for cancer immunotherapy. Therefore, an extensive research on the activity of cGAS or STING is highly interesting. One part of the present work was the characterization of a fluorescent 2′3′-cGAMP analogue, named fGAMP, in regard to a novel application for STING analysis. Thus, it was tested for STING binding by ITC experiments. Furthermore, the fluorescent

properties of fGAMP were analyzed using a fluorescence plate reader. A change of fluorescence intensity upon STING binding is required for the establishment of a fluorescence-based assay.

Publications

1. Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from *Thermotoga maritima*

Drexler, D. J.*; Müller, M.*; Rojas-Cordova, C. A.; Bandera, A. M.; Witte, G. Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from Thermotoga maritima. *Structure* **2017**, *25* (12), 1887-1897.e4.

*: equal contribution

DOI: https://doi.org/10.1016/j.str.2017.10.001

URL: https://www.sciencedirect.com/science/article/pii/S0969212617303295

This publication provides a detailed insight in the degradation of c-di-AMP by DHH-type PDEs, based on the homologue TmPDE from *T. maritima*. TmPDE was characterized structurally and biophysically to elucidate its function. The substrate specificity was determined using HPLC-based ion exchange chromatography activity assays and by binding analysis using surface plasmon resonance spectroscopy and isothermal titration calorimetry. These experiments show a high activity and binding affinity for linear dinucleotides, such as 5′-pApA, 5′-pApG and 5′-pGpG, whereas cyclic dinucleotides are not hydrolyzed and bound under physiological conditions. Furthermore, six high-resolution crystal structures of TmPDE in different reaction states were obtained, allowing a detailed description of the hydrolysis mechanism. In addition, right-angle light scattering and small-angle x-ray scattering data reveal a dimer in solution. In conclusion, the structural and biophysical characterization of TmPDE indicates a clear preference for linear dinucleotides, supporting a two-step degradation pathway for c-di-AMP.

Author contribution

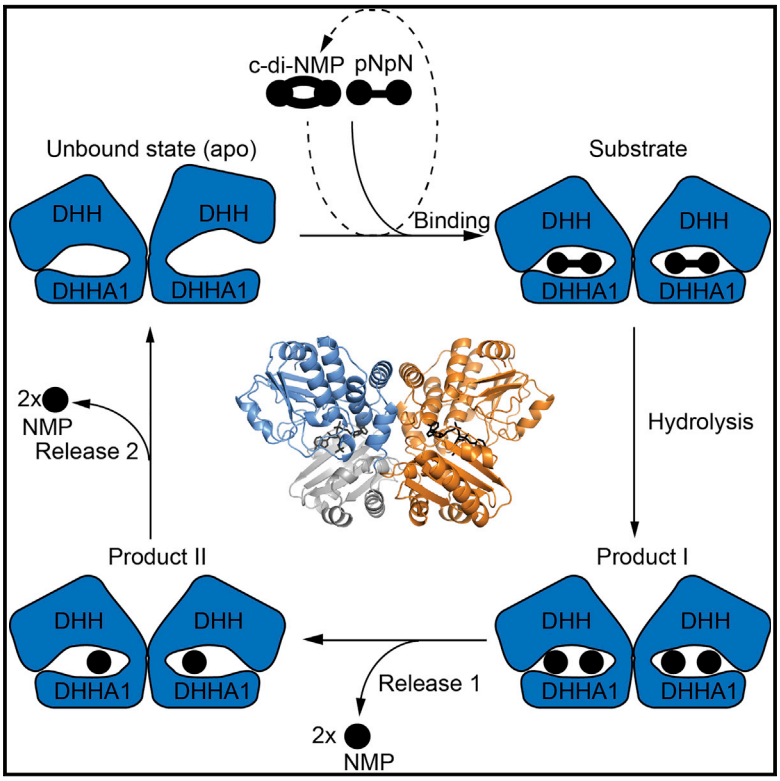
The author of the present thesis expressed and purified TmPDE for structural and biophysical analysis. Together with M. Müller, the substrate specificity and binding was analyzed using ion exchange chromatography activity assays and surface plasmon resonance spectroscopy. Furthermore, the author of this thesis characterized substrate binding affinity by isothermal titration calorimetry. In addition, he crystallized TmPDE in complex with 5′-pApA, 5′-pApG and GMP and performed data collection and evaluation with support of G. Witte. He wrote the manuscript together with G. Witte.

Short Article

Structure

Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from *Thermotoga maritima*

Graphical Abstract



Highlights

- Crystal structures of a *T. maritima* DHH/DHHA1-type PDE in different reaction states
- Biophysical and biochemical characterization of PDE activity and substrate binding
- TmPDE has high affinity for linear dinucleotides consistent with structural data
- The data support a two-step degradation mechanism of c-di-AMP

Authors

David Jan Drexler, Martina Müller, Carlos Alberto Rojas-Cordova, Adrian Maurice Bandera, Gregor Witte

Correspondence

witte@genzentrum.lmu.de

In Brief

In their paper Drexler et al., 2017 describe the biochemical and structural investigation of a soluble DHH-type phosphodiesterase and provide experimental evidence for a probable two-step decay of c-di-AMP with the analyzed phosphodiesterase being responsible for the second step in hydrolysis.

Drexler et al., 2017, Structure *25*, 1887–1897
December 5, 2017
https://doi.org/10.1016/j.str.2017.10.001

Structure
Short Article

Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from *Thermotoga maritima*

David Jan Drexler,[1,4] Martina Müller,[1,2,4] Carlos Alberto Rojas-Cordova,[1,3] Adrian Maurice Bandera,[1] and Gregor Witte[1,5,*]
[1]Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany
[2]Present address: CRELUX GmbH, Am Klopferspitz 19a, 82152 Planegg-Martinsried, Germany
[3]Present address: EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: witte@genzentrum.lmu.de
https://doi.org/10.1016/j.str.2017.10.001

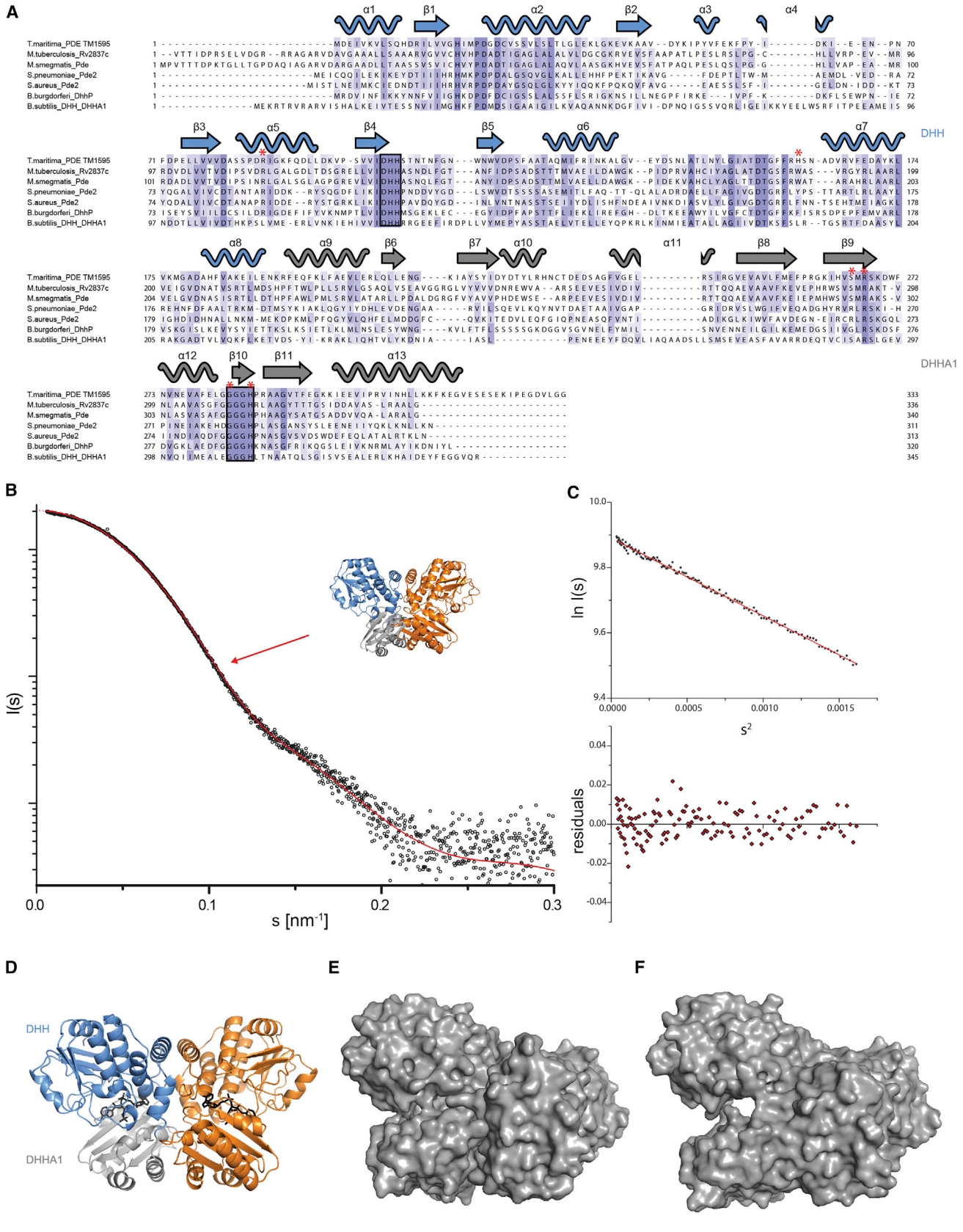
SUMMARY

The concentration of messenger molecules in bacterial cells needs to be tightly regulated. This can be achieved by either controlling the synthesis rate, degradation, or export by specific transporters, respectively. The regulation of the essential second messenger c-di-AMP is achieved by modulation of the diadenylate cyclase activity as well as by specific phosphodiesterases that hydrolyze c-di-AMP in the cell. We provide here structural and biochemical data on the DHH-type phosphodiesterase TmPDE (TM1595) from *Thermotoga maritima*. Our analysis shows that TmPDE is preferentially degrading linear dinucleotides, such as 5′-pApA, 5′-pGpG, and 5′-pApG, compared with cyclic dinucleotide substrates. The high-resolution structural data provided here describe all steps of the PDE reaction: the ligand-free enzyme, two substrate-bound states, and three post-reaction states. We can furthermore show that Pde2 from *Streptococcus pneumoniae* shares both structural features and substrate specificity based on small-angle X-ray scattering data and biochemical assays.

INTRODUCTION

Nucleotide second messengers are key components in cellular signaling that link signals to cellular responses. To achieve robust signaling that allows fast reaction to environmental or metabolic changes, the cellular levels of the messenger molecules are often regulated in both their synthesis and degradation, or they are even exported from the cell. In the case of the bacterial second messenger 3′,5′-cyclic di-adenosine monophosphate (c-di-AMP), the synthesis of the messenger molecule is carried out by DAC (di-adenylate cyclase) domain proteins that selectively synthesize c-di-AMP from two molecules of ATP (Witte et al., 2008). c-di-AMP is involved in a variety of processes in the cell, most of them related to potassium homeostasis or cell wall biogenesis, reviewed for example in Corrigan and Gründling (2013). Remarkably, c-di-AMP is the first second messenger that has been described to be essential for the bacteria synthesizing it, as total knockouts of DAC domain proteins are for example lethal in *Bacillus subtilis* (Mehne et al., 2013), *Staphylococcus aureus* (Corrigan et al., 2013), and *Listeria monocytogenes* (Witte et al., 2013). *In vivo* experiments in different organisms showed that the cellular level of c-di-AMP needs to be tightly regulated, as even small changes lead to drastic effects in the viability of the cells. For quite some time the essentiality of c-di-AMP could not be pinpointed, but a recent study in *B. subtilis* showed that c-di-AMP is, in fact, a key component in controlling potassium homeostasis (Gundlach et al., 2017). The c-di-AMP pathway is also an interesting target for antimicrobial therapies (Müller et al., 2015a), because the pathway influences sensitivity to antibiotic treatment (Dengler et al., 2013; Whiteley et al., 2017) and is found in many pathogenic bacteria such as *L. monocytogenes* (Woodward et al., 2010), *Mycobacterium tuberculosis* (Bai et al., 2012), *Streptococcus pneumoniae* (Kamegaya et al., 2011), and *S. aureus*, including methicillin-resistant *S. aureus*. The regulation of c-di-AMP synthesis is controlled by expression levels or activity of the respective proteins. For example, *B. subtilis* CdaS is expressed prior to sporulation (Nicolas et al., 2012), and also direct regulation of the DAC domains via ligand interactions was observed, as described for the DAC prototype DNA-integrity scanning protein A (DisA) from *Thermotoga maritima* and *B. subtilis.* DisA switches off its c-di-AMP synthesis upon binding to recombination intermediate DNA structures and likely acts as a checkpoint protein that ensures genome integrity prior to sporulation of the bacteria (Bejerano-Sagie et al., 2006; Oppenheimer-Shaanan et al., 2011; Witte et al., 2008). Another example is the regulation of the membrane protein CdaA by CdaR (Mehne et al., 2013). In addition to limited synthesis, a decrease in the total cellular c-di-AMP levels can also be achieved by selective degradation of the messenger molecule by c-di-AMP-specific phosphodiesterases (PDE) that degrade c-di-AMP to AMP and/or 5′-pApA. In bacteria, two types of c-di-AMP PDEs have been identified so far.

The first PDE that was characterized in detail is the multidomain membrane-associated DHH/DHHA1-type phosphodiesterase GdpP, which is encoded by YybT in *B. subtilis* (Figure 1A) (Rao et al., 2010). GdpP carries two additional regulatory



(legend on next page)

domains in addition to the two active PDE domains DHH and DHHA1; a sensory PAS (Per-Arnt-Sim) domain and a degenerated GGDEF domain. Whereas the former has been shown to bind Heme and nitric oxide, thereby affecting the PDE activity (Rao et al., 2011), the latter motif originally derives from proteins involved in c-di-GMP signaling but in this case binds and hydrolyzes ATP. GdpP has been shown to specifically hydrolyze c-di-AMP to 5′-pApA in presence of Mn^{2+} but also shows c-di-GMP phosphodiesterase activity resulting in 5′-pGpG (Rao et al., 2010). DHH/DHHA1-type PDEs also occur as soluble proteins comprising just the DHH/DHHA1 domains without the regulatory domains (Figure 1A), such as in *S. pneumoniae* (Pde2) (Bai et al., 2013), *M. tuberculosis* (CnpB/PDE/Rv2837c) (He et al., 2016), *M. smegmatis* (PDE) (Tang et al., 2015), and *S. aureus* (Pde2) (Bowman et al., 2016). Although these proteins belong to the DHH/DHH1A group, they have interestingly been reported to possess altered product specificity as they degrade c-di-AMP as well as 5′-pApA to AMP. Even though it was shown that *S. aureus* Pde2 might degrade c-di-AMP to 5′-pApA, it has a clear preference for the hydrolysis of the second step, i.e., 5′-pApA to 2 AMP (Bowman et al., 2016). In contrast, the biochemistry of the other PDEs mentioned above remains somewhat uncertain. In addition to the DHH/DHHA1-type PDEs, a second class of c-di-AMP-specific phosphodiesterases has been identified in *L. monocytogenes* (Huynh et al., 2015). *L. monocytogenes* PgpH is a member of the HD superfamily, named after the His-Asp residues that are functionally important as they coordinate the catalytically essential metal ions. PgpH and homologs are integral membrane proteins that belong to the 7TM-7TMR_HD family and consist of a seven transmembrane helix domain and a cytoplasmatic HD domain and (Huynh and Woodward, 2016). PgpH-type PDEs also show preferential hydrolysis of one of the 3′-5′-phosphodiester bonds in c-di-AMP, resulting in the linear product 5′-pApA. Of note, both DHH- and HD-type phosphodiesterases occur in bacteria with c-di-AMP pathways; some species have both of them (e.g., *L. monocytogenes*), whereas others have only one representative, most commonly DHH-type enzymes (Huynh and Woodward, 2016). As PgpH and GdpP are obviously responsible for the first step in c-di-AMP hydrolysis (its linearization to 5′-pApA), it is tempting to speculate that the soluble DHH-type PDEs carry out its subsequent hydrolysis from 5′-pApA to 2 AMP. Structural studies of the intracellular domain of the *L. monocytogenes* PgpH HD domain PDE (Huynh et al., 2015) and the soluble DHH domain phosphodiesterases from *M. tuberculosis* (He et al., 2016) suggest a similar mechanism of hydrolysis in which the two-metal-ion catalytic center facilitates the attack of the phosphodiester by an activated water molecule. To gain a more detailed view on the mechanism and biophysical properties of the second step in c-di-AMP degradation, we solved crystal structures of the soluble DHH-type PDE from *T. maritima* (TM1595) in different nucleotide-bound states and performed biochemical and biophysical assays to determine the kinetic and thermodynamic properties of nucleotide binding.

RESULTS AND DISCUSSION

Protein Characterization

Based on sequence homology searches we identified a putative c-di-AMP phosphodiesterase in *T. maritima*. The 37.5 kDa protein encoded by TM_1595 is highly similar to known DHH/DHHA1-PDEs and thus will be referred to as TmPDE (Figure 1A). We recombinantly expressed TmPDE in *Escherichia coli* and purified it to homogeneity. For comparative reasons we also purified the Pde2 protein from *S. pneumoniae* (SpPde2), which has been described to possess c-di-AMP phosphodiesterase activity (Bai et al., 2013). Both purified PDEs elute as single peaks in size-exclusion chromatography (SEC) with a hydrodynamic radius corresponding to a dimeric form of the PDEs. To unambiguously determine the molecular weight in solution we also performed analytical SEC coupled right-angle laser light scattering (SEC-RALS) and small-angle X-ray scattering (SAXS). The respective molecular weights derived from these methods indicate that TmPDE forms a dimeric assembly in solution (Figures 1B, 1C, and S1).

Product Specificity and Phosphodiesterase Activity

To analyze the phosphodiesterase activity of both TmPDE and SpPde2, we individually incubated the proteins with c-di-AMP, 5′-pApA, 5′-pApG, or 5′-pGpG, and analyzed the products by liquid chromatography (LC; see the STAR Methods). The reaction partners showed well-separated peaks, allowing their integration and Michaelis-Menten-like kinetic evaluation. Interestingly, we could detect robust and fast phosphodiesterase activity of both PDEs when using the linear substrates 5′-pApA (K_M = 204 ± 10 μM; k_{cat} = 0.14 s^{-1}), 5′-pApG (K_M = 355 ± 36 μM, k_{cat} = 0.26 ± 0.02 s^{-1}), and 5′-pGpG (similar activity, but no detectable saturation in the K_M assays). The cyclic dinucleotide c-di-AMP, however, was only degraded to a small extent under the various conditions tested, and only the use of 1,000-fold higher protein concentration yielded detectable

Figure 1. Multiple Sequence Alignment and Overall Structural Characterization of TmPDE

(A) Sequence alignment of PDEs from different species. Secondary structure elements from TmPDE are shown above, with the DHH domain in light blue and DHHA1 domain in gray. Nucleotide-binding residues are marked by asterisks and the highly conserved active site is outlined. For *B. subtilis* GdpP only the C-terminal DHH/DHHA1 domain is shown, the other sequences represent the full-length proteins.

(B) Buffer-corrected SAXS curve of a 7 mg/mL TmPDE wild-type sample shown as black circles with the theoretical scattering curve shown in red of TmPDE (in apo state, inset) obtained using CRYSOL (Svergun et al., 1995) (χ^2 = 2.3).

(C) Guinier plot analysis ln I(s) versus s^2 of the SAXS-data from (C). The radius of gyration was determined from linear regression (red) of data in the Guinier region ($s^*R_G < 1.3$), R_G = 2.67 nm. The lower panel shows the residuals of the linear regression. The linearity of the Guinier plot indicates that the TmPDE sample is not aggregating.

(D) Crystal structure of the biological dimer of TmPDE D80N D154N mutant in the 5′-pApG-bound state shown in cartoon representation. The DHH and DHHA1 domains of one monomer are colored light blue and gray, respectively, whereas the other monomer is colored orange. 5′-pApG is shown as black sticks.

(E) Surface representation of the TmPDE D80 D154N mutant in 5′-pApG-bound state in which both binding sites are closed.

(F) Surface representation of the TmPDE apo structure in the same orientation as shown in (E) with one open binding site.

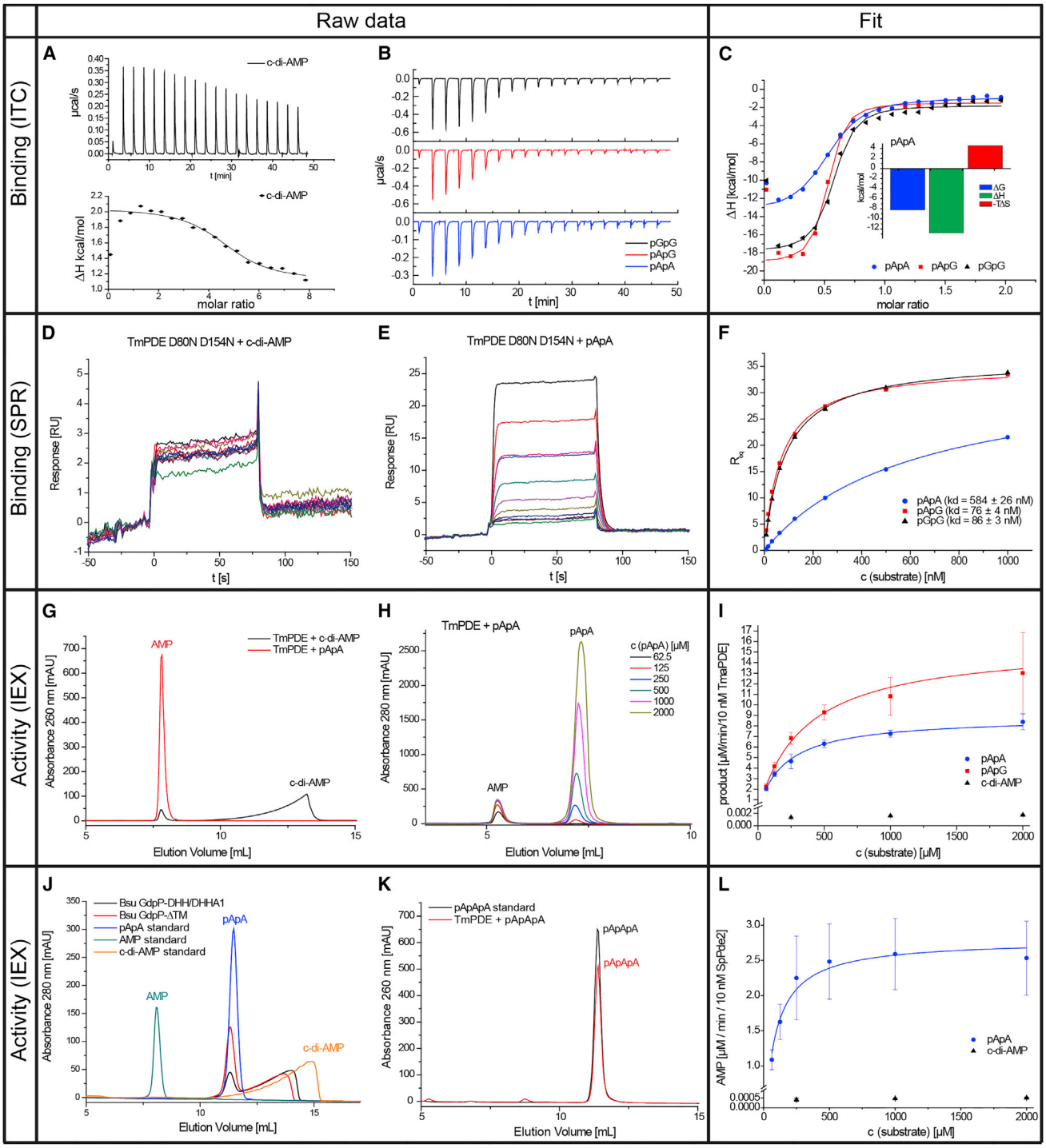


Figure 2. Binding and Activity of TmPDE with Different Ligands

(A) ITC measurement of 20 μM TmPDE D80N D154N mutant titrated with 200 μM c-di-AMP and the respective binding curve fit (K_D = 4.58 ± 1.93 μM).

(B) ITC measurement raw data of 20 μM TmPDE D80N D154N mutant titrated with 250 μM ligand: 5′-pGpG (black), 5′-pApG (red), and 5′-pApA (blue).

(C) Binding curves and fits of ITC titrations of the TmPDE D80N D154N mutant with 5′-pGpG (K_D = 247 ± 59 nM) (black), 5′-pApG (K_D = 191 ± 36 nM) (red), and 5′-pApA (K_D = 860 ± 94 nM) (blue). The inset shows a representative signature plot of the ITC measurement with 5′-pApA with ΔG = −8.26 kcal/mol, ΔH = −12.8 kcal/mol, and –TΔS = 4.55 kcal/mol (similar values were observed for the other linear 5′-pNpNs).

(D) SPR measurement of the TmPDE D80N D154N mutant (on chip) with injections of 8–1,000 μM c-di-AMP.

(E) SPR measurement of the TmPDE D80N D154N mutant (on chip) with injections of 8–1,000 μM 5′-pApA.

(F) Steady-state affinity fits for SPR measurements of the TmPDE D80N D154N mutant titrated with 5′-pGpG (K_D = 86 ± 3 nM) (black), 5′-pApG (K_D = 76 ± 4 nM) (red), and 5′-pApA (K_D = 584 ± 26 nM) (blue).

(legend continued on next page)

amounts of AMP (Figures 2 and S2). This extremely low turnover did not allow for reliable determination of kinetic parameters for either TmPDE or SpPde2. This finding is in agreement with data from Bowman et al. (2016) showing that soluble DHH/DHHA1-type PDEs prefer the hydrolysis of 5′-pApA over the direct hydrolysis of c-di-AMP. Indeed, the turnover of c-di-AMP for both TmPDE and SpPde2 is almost negligible in comparison with the 5′-pApA substrate. In contrast, the DHH/DHHA1-type GdpP-type protein from *B. subtilis* (Figure 2J) and *S. pneumoniae* SpPde1 (not shown) clearly show degradation of c-di-AMP to 5′-pApA, despite the homologous catalytic domains. A truncated *B. subtilis* GdpP construct lacking the regulatory domains (GdpP-DHH/DHHA1) also hydrolyzes c-di-AMP to 5′-pApA, thus excluding the regulatory domains of GdpP as a reason for the ligand specificity (Figure 2J). To elucidate differences between PDE and GdpP phosphodiesterases, we modeled *B. subtilis* GdpP based on sequence alignment and our TmPDE structure using SWISS-MODEL (Biasini et al., 2014). As expected, the GdpP model shows a similar overall structure (Figure S3E), whereas analysis of the active site reveals that the modeled ligand 5′-pApA does not fit into the GdpP model structure due to steric hindrance of residues L191 and N314 (Figure S3F). This could be a reason for the incomplete hydrolysis of c-di-AMP to 5′-pApA in GdpP-like phosphodiesterases, although this assumption is only based on a structure prediction and remains to be verified.

To be able to monitor substrate binding to the protein, we produced an inactive mutant of TmPDE that can bind the substrate(s) but lacks catalytic activity. Site-directed mutagenesis of two acidic residues abolished binding of the catalytically essential active site Mn^{2+} ions in the active site, which are necessary for activating the water molecule that performs the nucleophilic attack on the 3′-5′-phosphodiester (Rao et al., 2010). These respective aspartate residues were identified by homology analysis and mutated to asparagine (i.e., D80N D154N; Figures 3A and 3B) instead of alanine, as described previously (Rao et al., 2010), to have a more conservative mutation. The resulting mutant and wild-type proteins showed identical elution profiles and molecular weight in SEC-RALS, and also show similar SAXS curves, suggesting that the mutations did not alter overall protein conformation or oligomeric state (Figures S1A, S1G, and S1H). In the LC-based assays we were not able to detect any turnover of 5′-pApA or c-di-AMP by TmPDE D80N D154N, confirming that this mutant is completely inactive under the assay conditions used.

We performed surface plasmon resonance (SPR) experiments to rule out mechanistic differences in c-di-AMP or 5′-pApA hydrolysis, which might lead to the apparent inactivity *in vitro* and to further characterize the product binding of TmPDE. The inactive mutant TmPDE D80N D154N was immobilized on the SPR chip, and different concentrations of the respective cyclic and linear nucleotides were injected over the chip's surface. The resulting binding curves determined from the sensorgrams (Figures 2D–2F and S2E–S2G) indicate that TmPDE D80N D154N is not capable of binding c-di-AMP or c-di-GMP under the conditions tested. In contrast, linear products were bound with K_D values in the nanomolar range (Figures 2D–2F and 4B), some of them with high on-off rates beyond the resolution of the SPR instrument, as indicated by the abrupt rise and fall in the respective sensorgrams (Figure 2E). To further determine thermodynamic parameters of the substrate binding, we also performed isothermal titration calorimetry (ITC). Using c-di-AMP as a ligand, the integrated heat curves only showed a low-affinity binding (K_D = 4.58 ± 1.93 μM) in an endothermic reaction. In contrast, the linear substrates clearly show high-affinity binding and parameters indicative of enthalpy-driven substrate binding with 5′-pGpG (K_D = 247 ± 59 nM), 5′-pApG (K_D = 191 ± 36 nM), and 5′-pApA (K_D = 860 ± 94 nM) (Figures 2A–2C and 4B).

These biophysical data, in combination with the *in vitro* activity assays, strongly argue that the native substrate of TmPDE (and SpPde2) is in fact a linear dinucleotide (e.g., 5′-pApA, 5′-pApG, and 5′-pGpG as tested here), and that cyclic derivatives are bound only to a much lower extent. The fast on-off kinetics suggest that the active site is readily accessible by diffusion from the bulk solvent. Because there is only one active site in the DHH-DHHA1 domain phosphodiesterase, the full degradation of c-di-AMP to two molecules of AMP would require either reorientation (rotation) of the intermediate 5′-pApA within the active site or its release and rebinding. With respect to the previously described hydrolysis of c-di-AMP by SpPde2 and Rv2837c, and the binding affinities determined here, we can confirm that, for c-di-AMP hydrolysis, orders of magnitude higher concentrations are needed in the assays. Using higher concentrations of c-di-AMP certainly increases the probability of c-di-AMP binding to TmDPE/SpPde2 and might potentially lead to the linearized product 5′-pApA, which is then degraded much more rapidly to two molecules of AMP due to the higher affinity to the DHH active site as seen in Figure S2A. This might explain why c-di-AMP PDE activity has been described for some of the soluble DHH-type PDEs in literature. To rule out that TmPDE has

(G) Ion-exchange chromatography runs on Mono Q 5/50 GL column of the reaction products after 1 hr incubation from 100 μL reactions containing 2 μM TmPDE + 250 μM 5′-pApA (red) and 250 μM c-di-AMP (black).

(H) Representative chromatograms of the activity assays of TmPDE (10 nM) with substrate 5′-pApG (62.5–2000 μM) after 25 min at 20°C analyzed by ion-exchange chromatography with a Mono Q 5/50 GL column.

(I) Michaelis-Menten kinetics of the reactions from TmPDE (10 nM) with substrates 5′-pApA (62.5–2,000 μM) (blue) and 5′-pApG (62.5–2,000 μM) (red) after 25 min at 20°C. 5′-pApA, K_M = 204 ± 10 μM; 5′-pApG, K_M = 355 ± 36 μM. Also shown are the data points for c-di-AMP (black). Shown are mean values with errors from three independent experiments.

(J) Ion-exchange chromatography runs on Resource Q column of the reaction products from 100 μL reactions containing 1 μM *B. subtilis* GdpP-ΔTM + 250 μM c-di-AMP (red) and 1 μM *B. subtilis* GdpP-DHH/DHHA1 + 250 μM c-di-AMP (black) after 1 hr at 20°C. Also shown are the standards AMP (green), 5′-pApA (blue), and c-di-AMP (orange).

(K) Ion-exchange chromatography runs on Mono Q 5/50 GL column of the reaction products from a 100 μL reaction containing 100 nM TmPDE + 250 μM 5′-pApApA (red) after 1 hr at 20°C and the respective standard 5′-pApApA (shown in black).

(L) Michaelis-Menten kinetics of the reactions from SpPde2 with 5′-pApA (62.5–2,000 μM) (blue). K_M = 97 ± 8 μM after 25 min incubation at 20°C. Also shown are the data points for c-di-AMP (black). Shown are mean values with errors from three independent experiments.

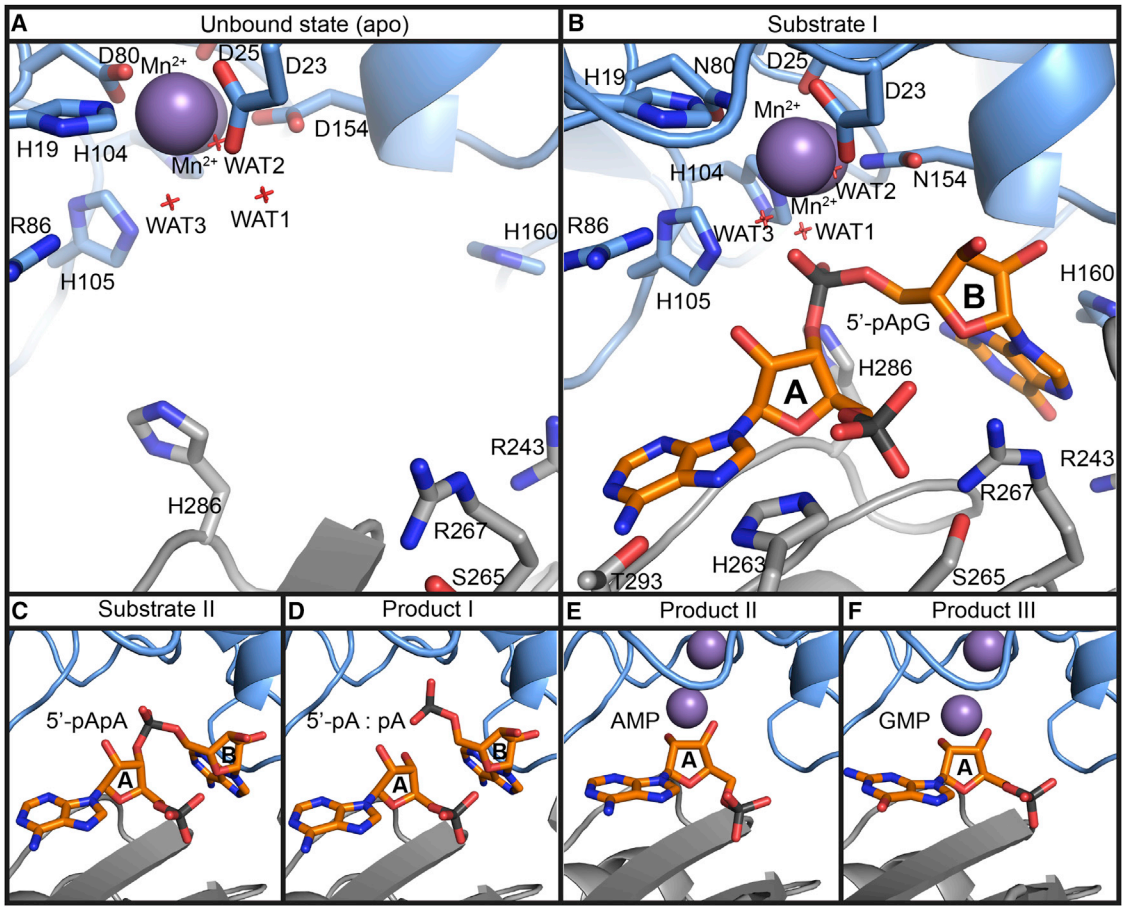


Figure 3. Active Site of TmPDE in Different Ligand-Bound (Reaction Intermediate) States

(A) Detailed view on the empty active site of TmPDE. Substrate-binding residues are shown as sticks and manganese ions are shown as purple spheres.

(B) Detailed view of the active site of TmPDE D80N D154N crystallized with 5′-pApG. The bound ligand and the interacting residues are shown as sticks and the manganese ions (modeled, see main text) as purple spheres.

(C) TmPDE D80N D154N with 5′-pApA in the nucleotide binding cleft. The positions for the 5′- and 3′-nucleotides are labeled with A and B in the sugar ring.

(D) TmPDE D80N D154N with two molecules of AMP in site A and site B resulting from 5′-pApA cleavage. The phosphate of nucleotide in position B has moved up, in agreement with the S_N2-mechanism.

(E) Wild-type TmPDE with AMP and Mn^{2+} ions at site A. The second AMP, c.f., D has already left site B.

(F) TmPDE D80N D154N after hydrolysis of 5′-pGpG and only one GMP (site A) left in the active site.

even higher specificity for longer RNAs, such that it might serve as nanoRNase, we also tested 5′-pApApA as a substrate (Figure 2K). TmPDE showed almost no cleavage of the trinucleotide, despite a doubled reaction time and a ten times higher protein concentration compared with the 5′-pApA experiment.

Structural Determination of TmPDE in Different Ligand-Bound States

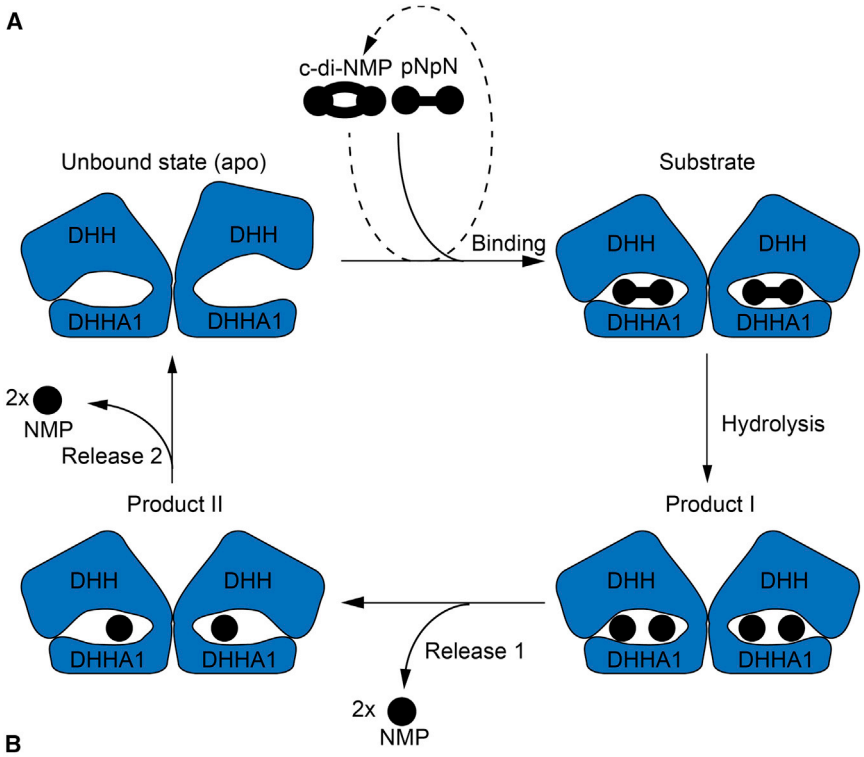
To obtain structural information that allows a description of the putative substrate recognition and PDE mechanism, we crystallized wild-type TmPDE in absence of ligands (apo-) and in the product-bound form II (AMP/Mn^{2+}). We also crystallized the mutant TmPDE D80N D154N in the presence of different linear dinucleotides, trapping the enzyme in different reaction intermediate states (substrates I and II, products I–III).

We used datasets collected from TmPDE product state crystals (containing Mn^{2+}, AMP, or GMP) diffracting to 1.9 Å to phase the data by single-wavelength anomalous diffraction. The obtained map showed readily interpretable electron density for the two molecules in the asymmetric unit. This allowed us to build the almost complete structure of TmPDE with good validation statistics, and the final model comprises residues 1–319 for both chains in the asymmetric unit, 501 water molecules and 6 Mn^{2+} ions (numbers for TmPDE/AMP-Mn^{2+}, i.e., product II) (Table 1). The refined model was subsequently used to determine the structures of apo TmPDE and the substrate and product complexes by molecular replacement. These structures were then refined and yielded similar statistics (Table 1).

Overall Structural Properties of TmPDE

One chain of TmPDE consists of the two clearly separated domains DHH (residues 1–190) and DHHA1 (residues 194–end), connected by a linker between helices α8 and α9 (Figure 1 and Movie S1). Monomeric TmPDE has overall dimensions of

A



B

	K_D (ITC) [nM]	K_D (SPR) [nM]	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_M [μM]	k_{cat} [s^{-1}]
pApA	860 ± 94	584 ± 26	> $5x10^5$	0.33	204 ± 10	0.14
pApG	191 ± 36	76 ± 4	> $10x10^5$	0.07	355 ± 36	0.26 ± 0.02
pGpG	247 ± 59	86 ± 3	> $6x10^5$	0.04	/	/
c-di-AMP	4580 ± 1930	n.b.	n.b.	n.b.	n.t.	n.t.
c-di-GMP	/	n.b.	n.b.	n.b.	n.t.	n.t.

n.b. = not binding
n.t. = no (significant) turnover

Figure 4. Schematic Representation of the Hydrolysis Reaction of TmPDE

(A) The reaction cycle based on available structures: The pre-reaction state of TmPDE shows one open active site in the dimer. 5′-pNpN ligands bind tightly to TmPDE, whereas c-di-NMPs bind much weaker. The bound ligand is hydrolyzed resulting in two NMP molecules in the active site. The two NMPs are then released separately (3′ nucleotide followed by the 5′ nucleotide) prior to binding of the next substrate.

(B) The table summarizes all results of binding experiments and activity assays. K_D values are calculated from ITC and SPR experiments, k_{on} and k_{off} from SPR measurements. The errors represent the error of the individual fit of the binding model to the experimental data. K_M and k_{cat} values from Mono Q activity assays are mean values and errors of three independent experiments. Please note that 5′-pGpG is hydrolyzed with approximately similar activity, but the assay does not allow for determination of parameters.

approximately 55 × 35 × 30 Å. The eponymous active site containing the highly conserved DHH motif is located in the cleft between DHH and DHHA1. The two domains share high structural similarity with the DHH/DHHA1 structures of nanoRNases (Srivastav et al., 2014; Uemura et al., 2013) and also the PDE from *M. tuberculosis* Rv2837c (He et al., 2016). The DHH domain comprising residues 1–192 consists of a five-stranded parallel β sheet domain flanked by helices (α1–α8). It harbors the catalytically active residues (see below) and is responsible for metal ion coordination. The smaller C-terminal DHHA1 domain (residues 195–end) is connected to the DHH domain via a flexible linker domain and has a DHH similar fold comprising a central mixed β sheet domain (β6–β11) surrounded by helices (α9–α13). The dimeric assembly is mainly achieved by a large number of contacts between the α helices, creating a continuous helical central part in the dimer whereas the β sheet domains face outward.

The crystal packing (5′-pApA-structure with one molecule per asymmetric unit in $C222_1$) and the arrangement of the molecules in the asymmetric units of the other structures suggest that TmPDE forms a butterfly-shaped dimer with overall dimensions of 80 × 55 × 40 Å. The interactions of the DHH/DHHA1 domains and the linker region, with many hydrophobic interactions mediated by, e.g., helices α7–α9, results in outward facing clefts of the active sites, with the monomers having an approximately 60° tilt with respect to each other. This arrangement creates a large buried surface area of approximately 2,050 $Å^2$. In contrast, the only other possible dimeric assembly has inward facing active sites and a much lower buried surface area of approximately 800 $Å^2$. To experimentally validate the biological assembly, we collected solution scattering data of wild-type TmPDE and also of the D54N D180N mutant. The shapes of the averaged scattering curves are only compatible with the theoretical scattering curves of the respective butterfly-shaped dimers (Figures 1B, S1G, and S1H). We also collected SAXS data from the homologous *S. pneumoniae* Pde2 (Figures S1C–S1F), which we also used as a reference in our biochemical assays. Lacking a high-resolution structure of SpPde2, we performed *ab initio* modeling using the SpPde2 SAXS data. All *ab initio* models of independent runs (n = 20) were highly similar, and the final averaged shape shows a good fit and similarity to our crystallographic TmPDE dimer (description below, Figure S1E).

Because the linker between DHH and DHHA1 domains allows the cleft to open and consequently change its accessibility to the bulk solvent, the active site of TmPDE is certainly able to harbor and degrade more substrates than tested here. This is in agreement with data on the mycobacterial phosphodiesterase Rv2837c (He et al., 2016) and nanoRNases (Srivastav et al., 2014), which show that these enzymes can degrade various small nucleic acid substrates.

Substrate Coordination and Mn^{2+}-Activated 3′-5′-Phosphodiester Bond Cleavage

To characterize substrate binding, cleavage, and release, we determined the structures of the nucleotide-free state, two substrate-bound, and three product-bound states of TmPDE. The

Table 1. Crystallographic Data and Refinement Statistics

	Unbound State (Apo)	Substrate I 5′-pApA	Substrate II 5′-pApG	Product I pApA	Product II AMP/Mn^{2+}	Product III GMP/Mn^{2+}
Protein	wild-type	D80N D154N	D80N D154N	D80N D154N	wild-type	D80N D154N
Data Collection						
Beamline	EMBL P14	ESRF ID30A MASSIF-1	EMBL P13	EMBL P13	EMBL P14	SLS X06SA
Wavelength (Å)	0.9763	0.9660	1.1999	1.1999	1.5498 (SAD)	1.2999 (SAD)
Space group	C 2	C 2 2 2_1	C 2 2 2_1	C 2 2 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Cell dimensions						
a, b, c (Å)	132.7, 54.9, 108.6	71.7, 87.8, 122.4	70.4, 88.3, 124.3	71.4, 87.6, 121.9	55.3, 106.8, 134.1	55.9, 105.6, 134.5
α, β, γ (°)	90, 118, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution	95–1.7 (1.74–1.70)	50–1.7 (1.74–1.70)	62–1.55 (1.59–1.55)	60.9–1.55 (1.59–1.50)	50–1.9 (1.95–1.90)	50–1.9 (1.95–1.90)
$CC_{1/2}$	99.8 (65.9)	99.9 (75.0)	100 (75.5)	100 (75.8)	99.9 (66.1)	99.9 (74.6)
R_{meas}	11.7 (101)	5.5 (76.6)	3.7 (82.8)	4.5 (121)	10 (85.5)	7.2 (130.5)
I/σI	13.3 (2.0)	21.4 (2.25)	19.2 (1.87)	19.7 (1.67)	14.3 (2.5)	17.4 (1.9)
Completeness (%)	96.5 (95.9)	98.8 (97.8)	98.4 (95.4)	99.5 (98.3)	95.2 (92.1)	99.7 (99.9)
Redundancy	6.9 (6.8)	9.9 (4.5)	4.5 (4.3)	6.6 (6.3)	6.7 (6.6)	6.9 (6.9)
Refinement						
Resolution	20–1.75	44–1.7	50–1.55	41–1.55	49–1.9	46–1.9
No. of reflections	67,550	42,286	106,907	106,661	115,140	62,479
R_{work}/R_{free}	16.4/19.0	17.3/19.6	15.8/18.3	16.8/19.5	16.4/20.4	16.3/20.3
Molecules (asu)	2	1	1	1	2	2
No. of atoms						
Protein	5,118	2,606	2,620	2,639	5,130	5,144
Nucleotides/ligands	-/87	45/44	46/94	46/62	46/11	48/28
Water	527	249	279	295	501	551
RMSD						
Bond lengths (Å)	0.004	0.006	0.006	0.006	0.007	0.010
Bond angles (°)	0.691	0.763	0.811	0.849	0.821	1.004
Ramachandran plot (%)						
Favored	98	99	97	98	98	99
Allowed	2	1	3	2	2	1
Outliers	–	–	–	–	–	–
PDB ID	5O25	5O4Z	5O58	5O70	5O1U	5O7F

RMSD, root-mean-square deviation; asu, asymmetric unit.

mutant TmPDE D80N D154N has reduced activity due to impaired coordination of the catalytically essential Mn^{2+} ions, thus we will describe the metal ion coordination on the basis of the apo state structure. Here, we could clearly identify the Mn^{2+} ions in the active site owing to density in the anomalous difference map, and we used the apo structure to model the Mn^{2+} ions and respective water molecules in the other representations of TmPDE structures. Even though all nucleotide ligands were already clearly visible in the mF_o-DF_c difference density of the ligand-bound states after replacement, we also confirmed the presence and nature of the ligand in the product state I (AMP:AMP) structure by anomalous difference electron density map visualizing the phosphate (Figure S3A), and by calculation of Polder maps (Liebschner et al., 2017) with the ligand atoms omitted (Figure S3B). As mentioned before, the active site mutations do not alter the behavior or structure of TmPDE in solution, suggesting that the mutations only affect the active site.

As described previously for other DHH/DHHA1-phosphodiesterases, the 3′-5′-phosphodiester bond is attacked by a water molecule that is activated by divalent cations (Mn^{2+}) (He et al., 2016; Uemura et al., 2013). The nucleophilic attack for the S_N2-reaction is carried out by the activated water WAT1 from the "backside" of the phosphate and facilitates the planar intermediate state before the final cleavage. Consistent with this, our D80N D154N mutant, crystallized in the presence of 5′-pApA, shows that the 3′-5′-phosphodiester bond is in the ideal orientation for an attack by activated WAT1 (distance 1.7 Å; Figure S3C). The putative reaction mechanism of TmPDE is also in good agreement with data on Rv2837 (He et al., 2016), and also

remarkably similar to the EAL domain described for a light-regulated c-di-GMP phosphodiesterase (Barends et al., 2009).

Unbound State (Apo Form)

Interestingly, our ligand-free structure reveals that the dimeric DHH/DHHA1 assembly has a slightly asymmetric shape, whereas the ligand-bound structures show almost identical conformations in the two chains. Specifically, one-half of the dimer has a more open cleft, whereas the other half of the PDE is rather closed and slightly shifted and rotated (Figures 1E and 1F). This might indicate that TmPDE is dynamic during its reaction cycle, such that the open form of the empty enzyme facilitates substrate binding, and that substrate binding leads to closure of the active site (Movie S1). A readily accessible active site would provide a structural basis for the relatively fast on-rate of substrate binding (Figure 4B). Interestingly, our ITC experiments show a stoichiometry of n = 0.5 for substrate binding to the inactive enzyme (Figure 2C), this raises the possibility that TmPDE might function in a flip-flop-like manner in solution, i.e., only one site is active per cycle, in agreement with the half-open apo structure. In contrast, the crystal structures of ligand- and product-bound states argue for two active sites, but might be influenced by crystallization conditions and/or packing. We cannot exclude an alternating active site in the dimer at this stage but the conclusions drawn for the reaction chemistry and substrate preference for linear dinucleotides (see below) remain untouched by this possibility.

In the active site of TmPDE, two manganese ions are coordinated by a dense network of highly conserved residues; His-19 and Asp-23 coordinate Mn^{2+} 1, Asp-25, His-104, and Asp-154 coordinate Mn^{2+} 2, and Asp-80 coordinates both Mn^{2+} 1 and 2. The ligand binding site is further occupied by bulk solvent and some ordered water molecules. In this state, the enzyme is ready to bind its dinucleotide substrate.

Substrate-Bound Pre-reactive States (Substrates I and II)

The substrate-bound structures of TmPDE D80N D154N with 5′-pApA (substrate I, Figures 3C) and 5′-pApG (substrate II, Figure 3B) reveal that TmPDE dimer is symmetrical and has closed active site clefts compared with the apo state. The dinucleotides are oriented with the 3′-5′-phosphodiester bond facing the position of the two modeled manganese ions, where highly coordinated water molecules are located instead. The bases of substrate molecules are coordinated by a network of polar and stacking interactions (Figure 3B) that correctly position the substrate. In particular, the 3′-base (site B) is stacked by His-160 and the loop comprising residues 154–160 closes the active site pocket. The 5′-base (site A) stacks on top of the GGG stretch sheets (β10 and β11; Figure 3B). In good agreement with SPR and ITC data showing different thermodynamic parameters for 5′-pApG and 5′-pApA binding, we can identify differences in the base stabilization in the structures. The guanine base forms additional interactions compared with the adenine substrate; Arg-243 of the DHHA1 domain contacts O6 and Asn-162 contacts N2 of the base (Figure S3D). These additional guanine-specific interactions manifest in both (1) slightly tighter K_D values for the binding of 5′-pApG and 5′-pGpG compared with 5′-pApA, and (2) a lower turnover (higher K_M) for the 5′-pApG substrate, indicating either a lower overall activity or a slower rate of substrate-product exchange.

Post-reactive State Formation and Product Release (Products I–III)

Directly after the phosphodiester bond cleavage of the 5′-pApA substrate, two molecules of AMP are present in the active site (product state I, Figure 3D). The 5′-phosphate of the nucleotide in site B has moved up, consistent with the directionality of the S_N2-reaction carried out by the activated water WAT1. In the next step (product state II, Figure 3E), only the AMP moiety in site A remains present in the binding cleft, kept in position by the GGG stretch, whereas the nucleotide from site B has already been released from the cleft. The 5′-phosphate of molecule A is tightly bound by His-263 and Arg-267 in the DHHA1 domain. The base is stacked by Arg-86 (DHH domain) and Thr-293 (DHHA1 domain) and is located on the surface of the GGG/AAA-stretch motif of the DHHA1 domain (β10 and β11). The analogous situation resulting from 5′-pGpG-cleavage can be seen in Figure 3F, where only the GMP remained bound in site A (product state III). Of note, our AMP structure contains three Mn^{2+} ions in the active site. Two of them (Mn^{2+} 1, Mn^{2+} 2) are in good agreement with structures of DHH/DHHA1 enzymes described above and also in (Srivastav et al., 2014). In contrast, Mn^{2+} 3 is less coordinated (sugar 2′ and 3′-OH, His-105, His-286), and has a lower occupancy in the structure. We assume that the third Mn^{2+} ion is bound in the active site due to the high Mn^{2+} concentration in the crystallization conditions and thus has no physiological relevance.

Conclusion

The structural and biochemical data provide a mechanistic framework for the phosphodiesterase activity of *T. maritima* PDE (summarized in Figure 4A). After high-affinity binding of the linear dinucleotide substrate to the ligand-free catalytic site, the phosphodiester bond is hydrolyzed in a S_N2 reaction by an activated water molecule resulting in two NMPs. The nucleotide in position B leaves the cleft prior to the one in site A, and the active site is then ready for the next cycle. The reported structures and biophysical data for TmPDE and SpPde2 suggest that linear dinucleotides are the main substrate for the soluble DHH-type PDEs and supports the idea of a two-step degradation of c-di-AMP in bacteria.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Experimental Model and Subject Details
  - Cloning, Expression and Purification
  - TmPde Crystallization
  - Data Collection and Processing
  - Phasing, Model Building and Refinement
  - Small-Angle X-ray Scattering
  - Static Light Scattering
  - Enzymatic Activity Assay
  - Surface Plasmon Resonance
  - Isothermal Titration Calorimetry
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, one table, and one movie and can be found with this article online at https://doi.org/10.1016/j.str.2017.10.001.

AUTHOR CONTRIBUTIONS

D.J.D. and M.M. performed and evaluated most of the experiments. C.A.R.C. purified protein and helped with crystallization. A.M.B. purified GdpP and performed assays. G.W. supervised data evaluation, performed experiments, designed research, and wrote the manuscript with the help of D.J.D. and M.M. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the staff of the EMBL Hamburg Beamlines at PETRA3 P12, P13, and P14 (EMBL/DESY Hamburg, Germany) for outstanding scientific support; data of the TmPDE/5′-pApA complex were collected at the automated MASSIF-1 beamline ID30A (ESRF Grenoble, France) and data on the product state III at SLS X06SA (Swiss Light Source, Switzerland). We thank Karl-Peter Hopfner for valuable discussions and support, and Sandra Schuller, Katja Lammens, Robert Byrne, and Dirk Kostrewa for discussions. This work was funded by grants from the Deutsche Forschungsgemeinschaft (DFG) WI 3717/2-1, WI 3717/3-1 (SPP1879), and GRK1721 (A10) to G.W.; A.M.B., D.J.D., and M.M. received support from the DFG GRK1721 research training group.

Received: July 17, 2017
Revised: September 7, 2017
Accepted: September 29, 2017
Published: October 26, 2017

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Bai, Y., Yang, J., Eisele, L.E., Underwood, A.J., Koestler, B.J., Waters, C.M., Metzger, D.W., and Bai, G. (2013). Two DHH subfamily 1 proteins in *Streptococcus pneumoniae* possess cyclic di-AMP phosphodiesterase activity and affect bacterial growth and virulence. J. Bacteriol. *195*, 5123–5132.

Bai, Y., Yang, J., Zhou, X., Ding, X., Eisele, L.E., and Bai, G. (2012). *Mycobacterium tuberculosis* Rv3586 (DacA) is a diadenylate cyclase that converts ATP or ADP into c-di-AMP. PLoS One *7*, e35206.

Barends, T.R., Hartmann, E., Griese, J.J., Beitlich, T., Kirienko, N.V., Ryjenkov, D.A., Reinstein, J., Shoeman, R.L., Gomelsky, M., and Schlichting, I. (2009). Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. Nature *459*, 1015–1018.

Bejerano-Sagie, M., Oppenheimer-Shaanan, Y., Berlatzky, I., Rouvinski, A., Meyerovich, M., and Ben-Yehuda, S. (2006). A checkpoint protein that scans the chromosome for damage at the start of sporulation in *Bacillus subtilis*. Cell *125*, 679–690.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. *42*, W252–W258.

Blanchet, C.E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C.M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., et al. (2015). Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). J. Appl. Crystallogr. *48*, 431–443.

Bowman, L., Zeden, M.S., Schuster, C.F., Kaever, V., and Gründling, A. (2016). New insights into the cyclic di-adenosine monophosphate (c-di-AMP) degradation pathway and the requirement of the cyclic dinucleotide for acid stress resistance in *Staphylococcus aureus*. J. Biol. Chem. *291*, 26970–26986.

Corrigan, R.M., Campeotto, I., Jeganathan, T., Roelofs, K.G., Lee, V.T., and Gründling, A. (2013). Systematic identification of conserved bacterial c-di-AMP receptor proteins. Proc. Natl. Acad. Sci. USA *110*, 9084–9089.

Corrigan, R.M., and Gründling, A. (2013). Cyclic di-AMP: another second messenger enters the fray. Nat. Rev. Microbiol. *11*, 513–524.

Dengler, V., McCallum, N., Kiefer, P., Christen, P., Patrignani, A., Vorholt, J.A., Berger-Bachi, B., and Senn, M.M. (2013). Mutation in the C-di-AMP cyclase dacA affects fitness and resistance of methicillin resistant *Staphylococcus aureus*. PLoS One *8*, e73512.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Crystallogr. *50*, 1212–1225.

Gundlach, J., Herzberg, C., Kaever, V., Gunka, K., Hoffmann, T., Weiss, M., Gibhardt, J., Thurmer, A., Hertel, D., Daniel, R., et al. (2017). Control of potassium homeostasis is an essential function of the second messenger cyclic di-AMP in *Bacillus subtilis*. Sci. Signal. *10*, https://doi.org/10.1126/scisignal.aal3011.

He, Q., Wang, F., Liu, S., Zhu, D., Cong, H., Gao, F., Li, B., Wang, H., Lin, Z., Liao, J., et al. (2016). Structural and biochemical insight into the mechanism of Rv2837c from *Mycobacterium tuberculosis* as a c-di-NMP phosphodiesterase. J. Biol. Chem. *291*, 3668–3681.

Huynh, T.N., Luo, S., Pensinger, D., Sauer, J.D., Tong, L., and Woodward, J.J. (2015). An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. Proc. Natl. Acad. Sci. USA *112*, E747–E756.

Huynh, T.N., and Woodward, J.J. (2016). Too much of a good thing: regulated depletion of c-di-AMP in the bacterial cytoplasm. Curr. Opin. Microbiol. *30*, 22–29.

Jeffries, C.M., Graewert, M.A., Blanchet, C.E., Langley, D.B., Whitten, A.E., and Svergun, D.I. (2016). Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. Nat. Protoc. *11*, 2122–2153.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kamegaya, T., Kuroda, K., and Hayakawa, Y. (2011). Identification of a *Streptococcus pyogenes* SF370 gene involved in production of c-di-AMP. Nagoya J. Med. Sci. *73*, 49–57.

Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., and Adams, P.D. (2017). Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallogr. D Struct. Biol. *73*, 148–157.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Mehne, F.M., Gunka, K., Eilers, H., Herzberg, C., Kaever, V., and Stülke, J. (2013). Cyclic di-AMP homeostasis in *Bacillus subtilis*: both lack and high level accumulation of the nucleotide are detrimental for cell growth. J. Biol. Chem. *288*, 2004–2017.

Mertens, H.D., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. *172*, 128–141.

Müller, M., Deimling, T., Hopfner, K.P., and Witte, G. (2015a). Structural analysis of the diadenylate cyclase reaction of DNA-integrity scanning protein A (DisA) and its inhibition by 3'-dATP. Biochem. J. *469*, 367–374.

Müller, M., Hopfner, K.P., and Witte, G. (2015b). c-di-AMP recognition by *Staphylococcus aureus* PstA. FEBS Lett. *589*, 45–51.

Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., et al. (2012).

Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. Science *335*, 1103–1106.

Oppenheimer-Shaanan, Y., Wexselblatt, E., Katzhendler, J., Yavin, E., and Ben-Yehuda, S. (2011). c-di-AMP reports DNA integrity during sporulation in *Bacillus subtilis*. EMBO Rep. *12*, 594–601.

Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. J. Appl. Crystallogr. *37*, 843–844.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera – a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. *40*, 191–285.

Rao, F., Ji, Q., Soehano, I., and Liang, Z.X. (2011). Unusual heme-binding PAS domain from YybT family proteins. J. Bacteriol. *193*, 1543–1551.

Rao, F., See, R.Y., Zhang, D., Toh, D.C., Ji, Q., and Liang, Z.X. (2010). YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. J. Biol. Chem. *285*, 473–482.

Schrodinger, LLC. (2015). The PyMOL Molecular Graphics System, Version 1.8.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 479–485.

Srivastav, R., Kumar, D., Grover, A., Singh, A., Manjasetty, B.A., Sharma, R., and Taneja, B. (2014). Unique subunit packing in mycobacterial nanoRNase leads to alternate substrate recognitions in DHH phosphodiesterases. Nucleic Acids Res. *42*, 7894–7910.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. *28*, 768–773.

Tang, Q., Luo, Y., Zheng, C., Yin, K., Ali, M.K., Li, X., and He, J. (2015). Functional analysis of a c-di-AMP-specific phosphodiesterase MsPDE from *Mycobacterium smegmatis*. Int. J. Biol. Sci. *11*, 813–824.

Uemura, Y., Nakagawa, N., Wakamatsu, T., Kim, K., Montelione, G.T., Hunt, J.F., Kuramitsu, S., and Masui, R. (2013). Crystal structure of the ligand-binding form of nanoRNase from *Bacteroides fragilis*, a member of the DHH/DHHA1 phosphoesterase family of proteins. FEBS Lett. *587*, 2669–2674.

Whiteley, A.T., Garelis, N.E., Peterson, B.N., Choi, P.H., Tong, L., Woodward, J.J., and Portnoy, D.A. (2017). c-di-AMP modulates *Listeria monocytogenes* central metabolism to regulate growth, antibiotic resistance and osmoregulation. Mol. Microbiol. *104*, 212–233.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Witte, C.E., Whiteley, A.T., Burke, T.P., Sauer, J.D., Portnoy, D.A., and Woodward, J.J. (2013). Cyclic di-AMP is critical for *Listeria monocytogenes* growth, cell wall homeostasis, and establishment of infection. MBio *4*, e00282-13.

Witte, G., Hartung, S., Büttner, K., and Hopfner, K.P. (2008). Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. Mol. Cell *30*, 167–178.

Woodward, J.J., Iavarone, A.T., and Portnoy, D.A. (2010). c-di-AMP secreted by intracellular *Listeria monocytogenes* activates a host type I interferon response. Science *328*, 1703–1705.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
E. coli BL21 (DE3) Rosetta (expression strain)	Novagene	
T. maritima MSB8 (used for genomic DNA prep.)	DSMZ; DSM 3109	https://www.dsmz.de/
B. subtilis subsp. subtilis	DSMZ; DSM 402	https://www.dsmz.de/ used for colony PCR
S. pneumoniae	DSMZ; DSM 11865	https://www.dsmz.de/ genomic DNA used for PCR
Chemicals, Peptides, and Recombinant Proteins		
5′-pApA	Biolog	Cat #P 033
5′-pApG	Biolog	Cat #P 082
5′-pGpG	Biolog	Cat #P 023
c-di-AMP	Biolog	Cat #C 088
c-di-GMP	Biolog	Cat #C 057
AMP	Sigma-Aldrich	Cat #A1752-1G
Deposited Data		
TmPDE pre-reaction state (apo)	This paper	PDB 5O25
TmPDE D80ND154N + 5′-pApA	This paper	PDB 5O4Z
TmPDE D80ND154N + 5′-pApG	This paper	PDB 5O58
TmPDE D80ND154N + pA : pA	This paper	PDB 5O70
TmPDE D80ND154N + GMP/Mn^{2+}	This paper	PDB 5O7F
TmPDE AMP/Mn^{2+}	This paper	PDB 5O1U
SAXS data (TmPDE wildtype)	This paper	SASBDB SASDCD7
SAXS data (TmPDE D80N D154N)	This paper	SASBDB SASDCC7
SAXS data (SpPde2)	This paper	SASBDB SASDCB7
Oligonucleotides		
RNA sequence: 5′-pApApA_ligand: AAA	Metabion	N/A
Primers for cloning, see Table S1	Metabion	N/A
Recombinant DNA		
Plasmid: pET-M11-SUMO1	EMBL Heidelberg	N/A
Software and Algorithms		
XDS/XSCALE	(Kabsch, 2010)	http://xds.mpimf-heidelberg.mpg.de/
SHELX CDE	(Sheldrick, 2010)	http://shelx.uni-ac.gwdg.de/SHELX/
HKL2MAP	(Pape and Schneider, 2004)	http://webapps.embl-hamburg.de/hkl2map/
PHENIX suite	(Adams et al., 2010)	https://www.phenix-online.org/
CCP4 suite	(Winn et al., 2011)	http://www.ccp4.ac.uk/
PHASER	(McCoy et al., 2007)	(included in CCP4 and PHENIX)
PyMOL	(Schrodinger, 2015)	https://www.pymol.org/
SWISS-MODEL	(Biasini et al., 2014)	https://swissmodel.expasy.org/interactive
UCSF Chimera	(Pettersen et al., 2004)	https://www.cgl.ucsf.edu/chimera/
ATSAS (includes CRYSOL)	(Franke et al., 2017)	https://www.embl-hamburg.de/biosaxs/software.html
PEAQ-ITC Analysis Software	Malvern	Provided with the instrument

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biacore X100+ Data Evaluation Software	GE Lifesciences	Provided with the instrument
Origin Pro	OriginLabs	http://www.originlab.de/
Other		
Columns (IEX, SEC, Affinity)	GE Healthcare Life Sciences	http://www.gelifesciences.com/webapp/wcs/stores/servlet/Home/en/GELifeSciences-de/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Gregor Witte (witte@genzentrum.lmu.de)

METHOD DETAILS

Experimental Model and Subject Details

B. subtilis subsp. subtilis (DSM 402) and *T. maritima* MSB8 (DSM 3109) cells were purchased from the DSMZ (Germany) and used either for genomic DNA preparation or directly for colony PCR. *S. pneumoniae* (DSM 11865) genomic DNA was purchased from DSMZ (Germany).

Cloning, Expression and Purification

T. maritima PDE (TM_1595), *S. pneumoniae* Pde2 (SPD_1153), *B. subtilis* GdpP-ΔTM (GdpP without transmembrane part) and *B. subtilis* GdpP-DHH/DHHA1 (active site domains only) were cloned into separate pET28 M11 SUMO1 vectors (EMBL Heidelberg) via the AgeI/NotI for the first or BamHI/NotI for the latter constructs restriction sites, respectively. A GSG-linker was introduced in between the SenP2 cleavage-site and the start of the TmPDE gene to facilitate tag cleavage. Recombinant expression of all proteins containing a His_6-SUMO tag was performed in *E. coli* Rosetta BL21(DE3) after induction with 0.2 mM IPTG (for TmPDE and BsuGdpP) or 0.1 mM IPTG (for SpPde2) at 18°C. Cells were harvested after 18 hours and lysed in buffer A (50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 5% v/v glycerol, pH 7.5) by sonication. After clarification of the lysate by centrifugation, the supernatant was loaded onto a Ni-NTA column (Qiagen). The column was washed with buffer A and followed by buffer B (50 mM Tris-HCl, 300 mM NaCl, 30 mM imidazole, 5% v/v glycerol, pH 7.5) prior to elution of the proteins with buffer C (50 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, 5% v/v glycerol, pH 7.5). SenP2 was added at a ratio of 1:500 (w:w) to TmPDE and SpPde2 in order to cleave the His_6-SUMO tag during dialysis into buffer D (20 mM Tris-HCl, 150 mM NaCl, 5% v/v glycerol, pH 7.5). After separating the tag from the proteins with the aid of a second Ni-column, TmPDE and SpPde2 were concentrated and loaded onto a HiLoad Superdex 200 column (GE Healthcare) equilibrated with buffer D. The peak fractions were pooled, concentrated, flash frozen in liquid nitrogen and stored at -80°C. TmPDE mutants were created by site-directed mutagenesis using overlap extension PCR. Mutant proteins were expressed and purified as described for the wild-type protein.

TmPde Crystallization

For determination of the product state structure, TmPDE at a concentration of 10 mg/mL was supplemented with 1 mM $MnCl_2$ and 1 mM AMP. Crystals were grown by hanging-drop vapor diffusion at 20°C after mixing 1 μL protein with 1 μL reservoir solution (0.1 M HEPES-NaOH pH 7.8, 25% w/v PEG4000, 0.2 M $CaCl_2$) with a total reservoir volume of 300 μL. Crystals were cryo-protected by transfer into mother liquor supplemented with 25% v/v ethylene glycol prior to flash cooling in liquid nitrogen. Crystals of apo TmPDE were grown similarly, but the protein was concentrated to 15 mg/mL and the reservoir solution contained 0.1 M Tris-HCl pH 7.5, 25% w/v PEG4000, 0.2 M $MgCl_2$. Cryo-protection was achieved by addition of 25% v/v PEG400 to the reservoir solution. Crystals of the inactive mutant TmPDE D80N D154N with 5′-pApA were grown by hanging-drop vapor diffusion with 10 mg/mL protein, 1.5 mM 5′-pApA and 1.5 mM $MnCl_2$ and a reservoir solution of 2.2 M ammonium sulfate and 0.1 M sodium citrate. Crystals were cryo-protected by soaking in reservoir solution supplemented with 25% v/v glycerol. Crystals of the inactive mutant TmPDE D80N D154N with 5′-pApG were obtained by sitting-drop vapor diffusion with 13 mg/mL protein, 1.5 mM 5′-pApG and 5 mM $MnCl_2$ and a reservoir solution of 20% w/v 2-propanol, 0.1 M sodium acetate pH 4.6 and 0.2 M $CaCl_2$. Crystals of the inactive mutant TmPDE D80N D154N with GMP (product III) were obtained by sitting-drop vapor diffusion with 13 mg/mL protein, 1.5 mM 5′-pGpG and 5 mM $MnCl_2$ and a reservoir solution of 30% w/v PEG4000, 0.1 M Tris pH 8.5 and 0.2 M $MgCl_2$. For both the 5′-pApG and 5′-pGpG complexes cryo-protection was achieved by the addition of 50% v/v ethylene glycol to the reservoir solution.

Data Collection and Processing

Diffraction data of the AMP/Mn^{2+}-TmPDE and apo TmPDE crystals were collected at the EMBL P14 beamline at PETRA III (DESY Hamburg, Germany). Diffraction data of the TmPDE D80N D154N mutant with 5′-pApA were measured at the ESRF ID30A MASSIF-1 (ESRF Grenoble, France). TmPDE with 2 AMP (product state II) and 5′-pApG were collected at the EMBL P13 beamline at PETRA III (DESY Hamburg, Germany). Diffraction data of the TmPDE mutant crystallized in presence of 5′-pGpG (GMP state) were collected at SLS X06SA (Swiss Light Source, Villigen, Switzerland). All data were indexed and scaled using XDS and XSCALE, respectively (Kabsch, 2010).

Phasing, Model Building and Refinement

Phases were determined by the use of a single anomalous diffraction dataset collected from TmPDE-AMP/Mn^{2+} crystals. Identification of Mn^{2+} sites, phasing and solvent flattening was performed using SHELXCDE (Sheldrick, 2010) through the HKL2MAP GUI (Pape and Schneider, 2004). The relatively high resolution allowed automatic model building with PHENIX (Adams et al., 2010) followed by manual model building in COOT (Emsley and Cowtan, 2004) and automatic refinement in PHENIX (Afonine et al., 2012). The product state III was phased similarly. The structures of the apo TmPDE and both TmPDE D80N D154N nucleotide complexes were phased by molecular replacement with PHASER (McCoy et al., 2007) within the CCP4 suite (Winn et al., 2011) using the DHH and DHHA1 domains as separate search models. Model building and refinement was performed with Coot and PHENIX as described above. All crystallographic figures were prepared with PyMOL (Schrodinger, 2015). Morphing in Movie S1 was prepared in Chimera (Pettersen et al., 2004) using the apo state and 5′-pApA bound state of TmPDE to illustrate the potential dynamics of the ligand-binding cleft.

Small-Angle X-ray Scattering

Small-angle X-ray scattering experiments were performed at the EMBL P12 beamline at PETRA III (DESY Hamburg, Germany). To remove potential aggregates, all samples were additionally purified by size-exclusion chromatography and centrifuged prior to measurements. Size-exclusion chromatography coupled SAXS was performed for the TmPDE D80N D154N mutant (Blanchet et al., 2015; Jeffries et al., 2016). The running buffer of the chromatography step was used as a buffer reference for buffer subtraction of the protein sample scattering data. Proteins in batch mode were measured at different concentrations and all data sets were analyzed using the ATSAS software package (Franke et al., 2017) as described in (Mertens and Svergun, 2010; Putnam et al., 2007). Calculation of theoretical scattering curves from crystal structures was performed using CRYSOL (Svergun et al., 1995). Figures of *ab initio* models with docked crystal structures were prepared with UCSF Chimera (Pettersen et al., 2004). Scattering data of SpPde2, wild type TmPDE and TmPDE D80N D154N have been deposited in the SASBDB.

Static Light Scattering

Size-exclusion chromatography coupled light scattering was performed using an ÄKTA micro chromatography system equipped with a Superdex 200 10/300 Increase column (GE Healthcare Life Sciences) and a right-angle laser light scattering device and refractive index detector (Malvern/Viscotek). BSA (66 kDa) was used to calibrate the system. Evaluation was performed using the OmniSEC software (Malvern/Viscotek) provided with the instrument.

Enzymatic Activity Assay

To check the activity of TmPDE and SpPde2 we used a LC-based assay to quantify the reaction products. A 100 μL reaction mix containing 50 mM Tris pH 8.5, 20 mM KCl, 0.1 mM $MnCl_2$, 62.5-2000 μM c-di-AMP or 5′-pApA, and 10 nM-10 μM enzyme (monomer concentration) was incubated at room temperature for 15-60 min. This was then quickly diluted with assay buffer A (50 mM Tris-HCl pH 9.0) and the nucleotides were separated from the protein by spin concentrators (30 kDa cutoff). 500 μL of the flow-through was loaded on to a Mono Q 5/50 GL column (GE Lifesciences) equilibrated with buffer A (50 mM Tris-HCl pH 9.0). A linear gradient to 50% buffer B (50 mM Tris-HCl pH 9, 1 M NaCl) over 30 column volumes was used to elute the reaction products. The peaks were then compared to standard nucleotides (AMP, 5′-pApA, c-di-AMP). The individual chromatogram peaks were integrated using Origin (OriginLabs). Please note, for assays shown in Figures 2G, 2J, and S2A–S2D a linear gradient to 35% buffer B over 28 column volumes was used.

Surface Plasmon Resonance

c-di-AMP and 5′-pApA binding was analyzed by SPR using a Biacore X100+ as previously described (Müller et al., 2015b). 5278 RU TmPDE D80N D154N were coupled to flow cell 2 (FC2) of an activated CM-5 (GE Life sciences) chip using amino reactive EDC/NHS coupling chemistry. FC1 was only activated and blocked with ethanolamine to serve as a reference for unspecific binding of the analyte to the chip and the reported data therefore correspond to RU(FC2-FC1). Different concentrations of nucleotides in HBS-EP buffer (150 mM NaCl, 10 mM HEPES-NaOH pH 7.4, 3 mM EDTA, 0.05% v/v surfactant P20) were injected to monitor binding of c-di-AMP and nucleotide substrates. Data were analyzed using the Biacore X100 Evaluation software. All experiments were repeated at least once to confirm the robustness of the assay.

Isothermal Titration Calorimetry

ITC data were collected using a Malvern PEAQ-ITC system with 20 μM TmPDE D80N D154N in HBS-EP buffer in the cell. The respective nucleotide at a concentration of 200 μM in HBS-EP buffer was titrated into the cell by 19 injections of 2 μL, spaced 150 s apart, at

25°C. Data evaluation was performed with the Malvern software package. All experiments were repeated at least once to confirm the robustness of the assay.

DATA AND SOFTWARE AVAILABILITY

Coordinates and structure factors have been deposited in the PDB under ID codes 5O25, 5O4Z, 5O58, 5O70, 5O1U, and 5O7F. SAXS data have been deposited in the SASBDB under ID codes SASDCB7, SASDCC7, and SASDCD7.

Structure, Volume *25*

Supplemental Information

Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from *Thermotoga maritima*

David Jan Drexler, Martina Müller, Carlos Alberto Rojas-Cordova, Adrian Maurice Bandera, and Gregor Witte

Table S1. Primers used in this study, Related to STAR Methods.

Name	Length	Orientation	T_m [°C]	Sequence (5'3')
TmPDE_AgeI_GSG_fwd	51	forward	56	TAATAAACCGGTGGATCCGGCTCTGGCTGGACGAGATCGTCAAAGTGCTC
TmPDE NotI_rev	35	reverse	60	TAATAAGCGGCCGCTCATCCCCCCAGTACGTCTCC
TmPDE D80N_fwd	33	forward	67	CTTCTTGTGGTGGTCAACGCCTCCTCTCCCGAC
TmPDE D80N_rev	33	reverse	69	GTCGGGAGAGGAGGCGTTGACCACCACAAGAAG
TmPDE D154N_fwd	31	forward	64	CTTGGGATCGCAACCAACACAGGGTTTTTCA
TmPDE D154N_rev	31	reverse	63	TGAAAAACCCTGTGTTGGTTGCGATCCCAAG
SpPdeI_Lys51_BamHI_fwd	33	forward	49	TAATAAGGATCCAAGAAACTGAGAGTGCATTAT
SpPdeI_Lys51_NotI_rev	35	reverse	48	TAATAAGCGGCCGCTCATTCTTCTTTCTCCTTTTC
SpPdeII_Lys51_BamHI_fwd	33	forward	51	TAATAAGGATCCATGGAGATTTGCCAACAAATT
SpPdeII_Lys51_NotI_rev	35	reverse	47	TAATAAGCGGCCGCTCAGTTTTTAAGCAAGTTTTT
BsGdpP_K1_BamHI_F	33	forward	46	TAATAAGGATCCATTGGAATCATGCTTTTTAAT
BsGdpP_NotI_R	32	reverse	53	TAATAAGCGGCCGCTCATCTCTGTACGCCTCC
GdpP DHH/DHHA1 B.s. C1 for	31	forward	63	TAAGGATCCATGGAGAAACGAACAAGGGTGC
GdpP DHH/DHHA1 B.s. C1 rev	32	reverse	67	TAAGCGGCCGCTCATCTTTGAACCCCTCCTTC

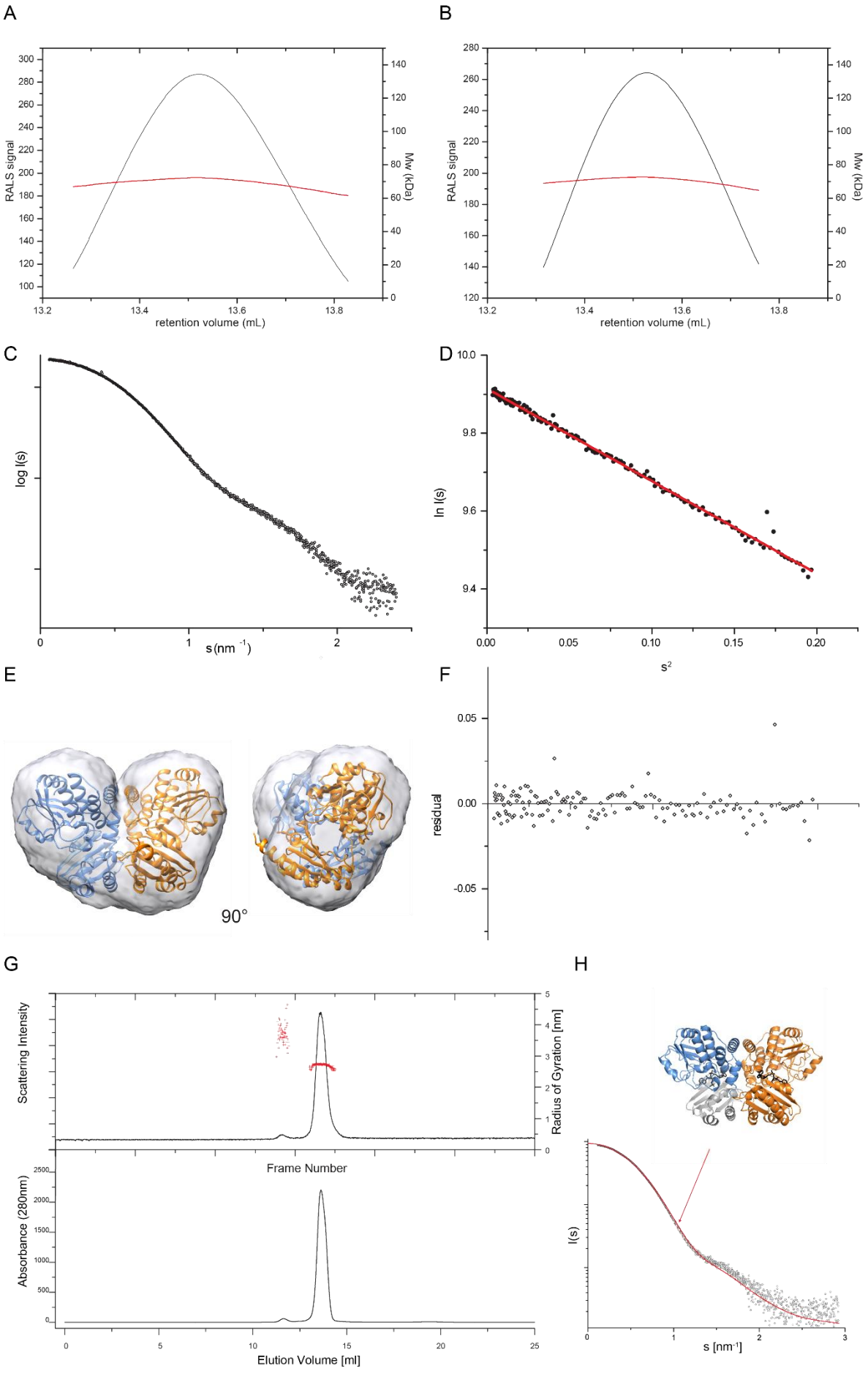

A
RALS signal
Mw (kDa)
retention volume (mL)
B
RALS signal
Mw (kDa)
retention volume (mL)
C
log I(s)
s(nm -1)
D
ln I(s)
s2
E
90°
F
residual
G
Scattering Intensity
Radius of Gyration [nm]
Frame Number
Absorbance (280nm)
Elution Volume [ml]
H
I(s)
s [nm-1]

Figure S1. Molecular weight determination of TmPDE wt (and mutant) and SAXS data, Related to Figure 1.

Molecular weight determination by size-exclusion coupled right-angle light scattering. Panel A and B show the TmPDE peaks of a run using a 24ml 10/300 S200 increase column (GE Healthcare) and their respective molecular weight determined by RALS for wildtype TmPDE (A) and TmPDE D80N D154N (B). Both proteins show molecular weights of approx. 70 kDa and thus TmPDE is dimeric in solution (the monomer Mw of the construct is 37.8 kDa). The mutation does not change the oligomeric state in solution.

(C) buffer corrected small-angle X-ray scattering curve of a 7 mg/mL *S. pneumoniae* Pde2 sample

(D) Guinier plot analysis, data from (C), R_G=2.7nm was determined from the slope of the linear regression for $s*R_G<1.3$ (Guinier approximation).The sample is free of aggregates as judged from linearity of the Guinier region

(E) final averaged *ab initio* shape of *S. pneumoniae* Pde2 calculated from 20 DAMMIF models (NSD 0.81) with docked TmPDE dimer in two orientations.

(F) shows the respective residuals of the linear regression of the Guinier plot (D)

(G) Chromatogram of SEC-coupled SAXS measurements of the TmPDE D80N D154N mutant using a Superdex 200 10/300 column. Shown are scattering intensity and radius of gyration (top) and absorbance at 280nm (bottom) versus elution volume.

(H) Averaged and buffer corrected SAXS curve (grey points) derived from the main peak (RG=2.7nm, panel B). Also shown is the theoretical scattering curve (red) calculated with CRYSOL (Svergun et al., 1995) of the “butterfly” shaped TmPDE D80N D154N dimer shown as cartoon model.

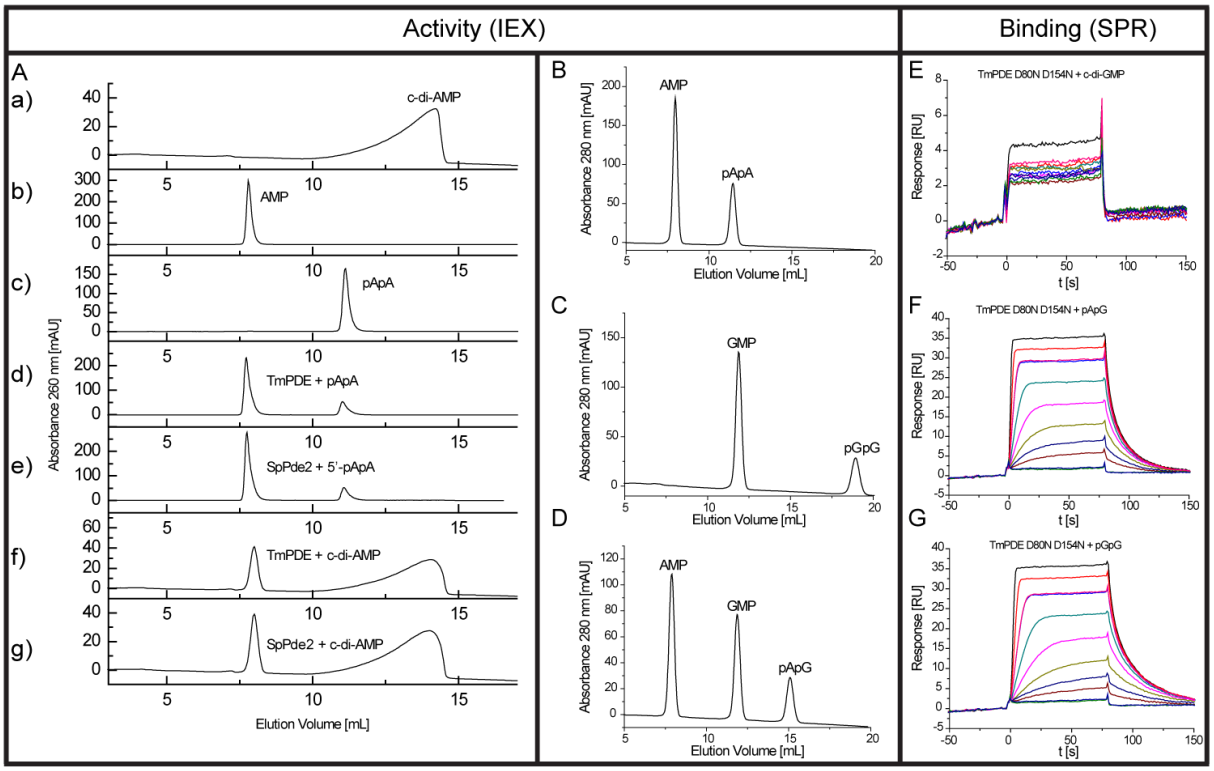


Figure S2. IEX experiments and SPR raw data, Related to Figure 2.

(A) Ion Exchange chromatography on Mono-Q column. From top:
a) 125 µM c-di-AMP standard; b) 250 µM AMP standard; c) 125 µM 5'-pApA standard; d) Reaction of 10 nM TmPDE with 125 µM 5'-pApA; e) Reaction of 25 nM SpPde2 with 125 µM 5'-pApA; f) Reaction of 10 µM TmPDE with 250 µM c-di-AMP; g) Reaction of 10 µM SpPde2 with 250 µM c-di-AMP. (t=25 min reaction time at 20 °C).
(B) Ion exchange chromatography on Mono-Q 5/50 GL column of the reaction products from 100 µL reactions with 10 nM TmPDE + 125 µM 5'-pApA after 25 min incubation.
(C) Ion exchange chromatography on Mono-Q 5/50 GL column of the reaction products from 100 µL reactions with 10 nM TmPDE + 125 µM 5'-pGpG after 25 min incubation.
(D) Ion exchange chromatography on Mono-Q 5/50 GL column of the reaction products from 100 µL reactions with 10 nM TmPDE + 125 µM 5'-pApG after 25 min incubation.
(E) SPR measurements of TmPDE D80N 154N mutant (on chip) with injections of 8 – 1000 µM c-di-GMP analyte concentrations.
(F) SPR measurements of TmPDE D80N 154N mutant (on chip) with injections of 8 – 1000 µM 5'-pApG analyte concentrations.
(G) SPR measuremenst of TmPDE D80N 154N mutant (on chip) with injections of 8 – 1000 µM 5'-pGpG analyte concentrations.

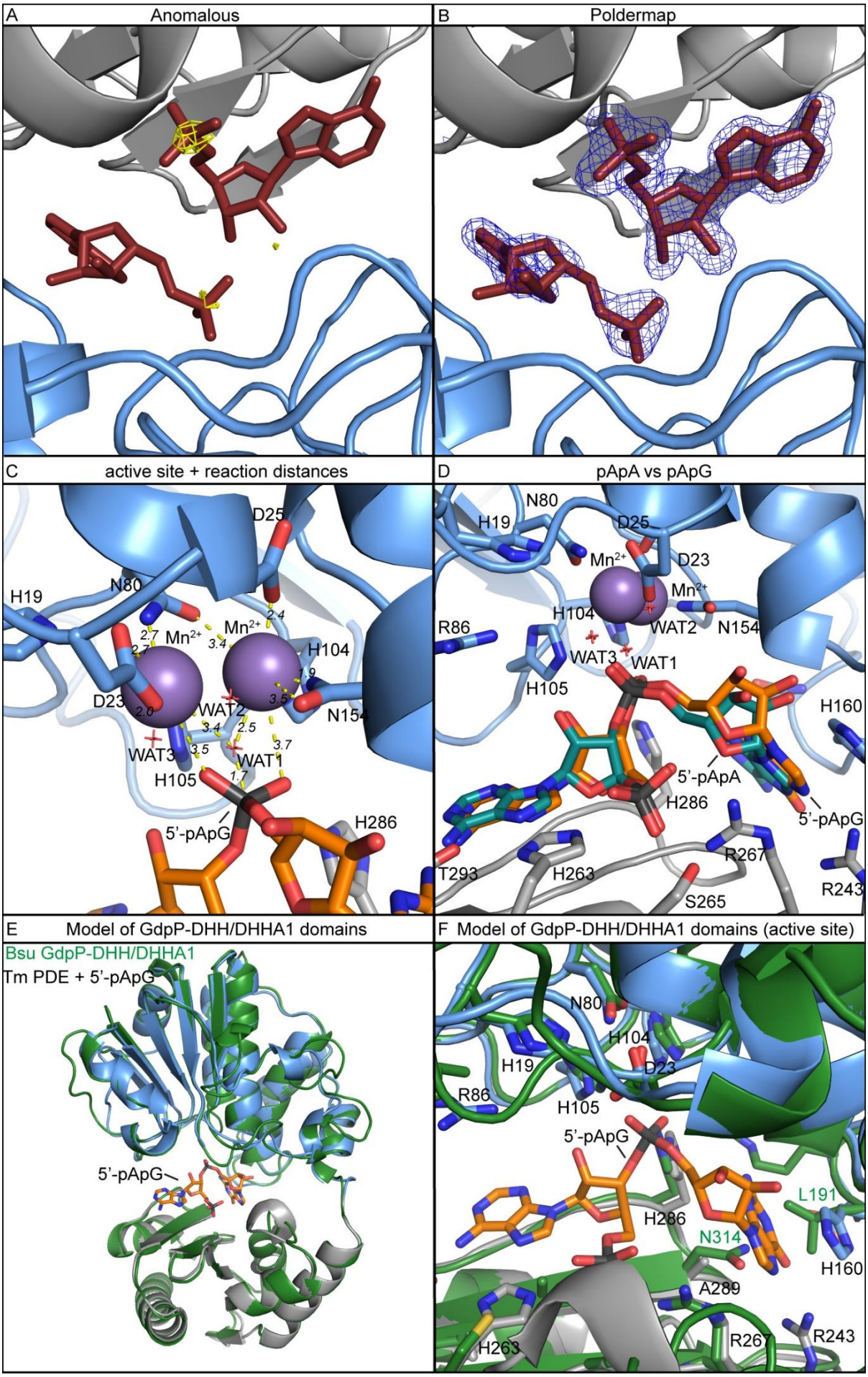

A
Anomalous
B
Poldermap
C
active site + reaction distances
D
pApA vs pApG
E
Model of GdpP-DHH/DHHA1 domains
F
Model of GdpP-DHH/DHHA1 domains (active site)
Bsu GdpP-DHH/DHHA1
Tm PDE + 5'-pApG
D25
N80
H19
Mn2+
Mn2+
H104
D23
WAT2
N154
WAT3
WAT1
H105
H286
5'-pApG
5'-pApA
H160
R86
T293
H263
S265
R267
R243
L191
N314
A289

Figure S3. Detailed view on the active site of TmPDE, Related to Figure 3.

(A) Active site of TmPDE D80N D154N shown in cartoon representation, DHH domain (light blue) and DHHA1 domain (grey) with hydrolyzed 5'-pApA (i.e. 2 AMP in red) and anomalous difference map (yellow) shown at 3.7σ contour level clearly defining the phosphate positions.
(B) Active site of TmPDE D80N D154N shown in cartoon representation, DHH domain (light blue) and DHHA1 domain (grey) with hydrolyzed 5'-pApA (2 AMP, red) and the respective Polder map (dark blue) shown at 4σ contour level.
(C) Close-up view on the active site of TmPDE D80N D154N shown in cartoon representation, DHH domain (light blue), DHHA1 domain (grey), with interacting residues and 5'-pApG shown as sticks, water molecules as red crosses and manganese ions (purple) as spheres. Distances are given in Å and shown as dashed lines (yellow).
(D) Active site superposition of TmPDE D80N D154N with 5'-pApA and 5'-pApG, respectively. TmPDE is shown in cartoon representation, DHH domain (light blue), DHHA1 domain (grey), with interacting residues, 5'-pApA (cyan) and 5'-pApG (orange) shown as sticks, water molecules as red crosses and manganese ions (purple) as spheres.
(E) Overall structure superposition of the modelled GdpP DHH/DHHA1 domains (green, based on the TmPDE D80N D154N structure with 5'-pApG as template using SWISS-MODEL) and the template. Protein structures are shown in cartoon representation, TmPDE DHH domain (light blue), DHHA1 domain (grey) and 5'-pApG (orange) shown as sticks.
(F) Active site close-up of the superposition of the modelled GdpP DHH/DHHA1 domain (green) and TmPDE D80N D154N (with 5'-pApG) illustrate a potential clash between L191 and N314 (site B) of the GdpP-model and the second base of the ligand.

2. C-di-AMP hydrolysis by a novel phosphodiesterase is crucial for differentiation of antibiotic-producing bacteria (submitted)

Latoscha, A.*; Drexler, D. J.*, Al-Bassam, M. M.; Kaever, V.; Findlay, K. C.; Witte, G.; Tschowri, N. C-di-AMP hydrolysis by a novel phosphodiesterase is crucial for differentiation of antibiotic-producing bacteria. (submitted) (online on bioRxiv)

*: equal contribution

DOI: https://doi.org/10.1101/789354

URL: https://www.biorxiv.org/content/10.1101/789354v1

This publication describes the effects of c-di-AMP levels in *S. venezuelae*, including the identification and characterization of a novel phosphodiesterase for c-di-AMP degradation. The hydrolysis of c-di-AMP is unclear in species that lack all of the so far identified c-di-AMP specific phosphodiesterases. The herein identified PDE, termed AtaC, is mainly present in Actinobacteria and contributes to the understanding of c-di-AMP degradation in many species. AtaC was shown to hydrolyze c-di-AMP and 5′-pApA specifically in high turnover rates. A manganese-dependent reaction mechanism was elucidated based on binding analysis and mutational approaches. In addition, right-angle light scattering and small-angle x-ray scattering data indicate AtaC to be a monomer in solution. The deletion of AtaC in *S. venezuelae* revealed significantly increased c-di-AMP levels, which confirmed the physiological relevance of AtaC. Furthermore, this mutation leads to delayed growth and sporulation defects. In contrast, the deletion of the c-di-AMP synthase DisA results in altered growth only under high salt conditions, which indicates an important role of c-di-AMP on osmolyte homeostasis.

Author contribution

The author of the present thesis expressed and purified AtaC for structural and biochemical analysis. He performed ion exchange chromatography activity assays to describe the substrate specificity. Furthermore, he analyzed manganese binding using differential scanning fluorimetry. In addition, he constructed an inactive mutant of AtaC based on structure prediction analysis. He described the binding affinity of c-di-AMP to this AtaC mutant by isothermal titration calorimetry and differential scanning fluorimetry. To elucidate the monomeric state of AtaC, he performed right-angle light scattering and small-angle x-ray scattering. He supported A. Latoscha, N. Tschowri and G. Witte in writing the manuscript.

c-di-AMP hydrolysis by a novel type of phosphodiesterase promotes differentiation of multicellular bacteria

Andreas Latoscha[1§], David Jan Drexler[2§], Mahmoud M. Al-Bassam[3], Volkhard Kaever[4], Kim C. Findlay[5], Gregor Witte[2*] and Natalia Tschowri[1*]

[1]*Department of Biology / Microbiology, Humboldt-Universität zu Berlin, 10115 Berlin, Germany*

[2]*Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany*

[3]*Department of Pediatrics, University of California, San Diego, La Jolla, California 92093, USA*

[4]*Research Core Unit Metabolomics, Medizinische Hochschule Hannover, 30625 Hannover, Germany*

[5]*Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK*

§ These authors contributed equally to this work

* Corresponding author: natalia.tschowri@hu-berlin.de

* Correspondence may also be addressed to: witte@genzentrum.lmu.de

Keywords: c-di-AMP, *Streptomyces*, phosphodiesterase, development, osmostress

ABSTRACT

Antibiotic-producing *Streptomyces* use the diadenylate cyclase DisA to synthesize the nucleotide second messenger c-di-AMP but the mechanism for terminating c-di-AMP signaling and the proteins that bind the molecule to effect signal transduction are unknown. Here, we identify the AtaC protein as a new type of c-di-AMP-specific phosphodiesterase that is also conserved in pathogens such as *Streptococcus pneumoniae* and *Mycobacterium tuberculosis*. AtaC is monomeric in solution and binds Mn^{2+} to specifically hydrolyze c-di-AMP to AMP via the intermediate 5'-pApA. As an effector of c-di-AMP signaling, we characterize the RCK-domain protein CpeA as the first c-di-AMP-binding protein to be identified in *Streptomyces*. CpeA interacts with the predicted cation / proton antiporter, CpeB, linking c-di-AMP signaling to ion homeostasis in actinobacteria. Hydrolysis of c-di-AMP is critical for normal growth and differentiation in *Streptomyces*, connecting osmotic stress to development. Thus, we present the discovery of two novel components of c-di-AMP signaling in bacteria and show that precise control of this second messenger is essential for osmoregulation and coordinated development in *Streptomyces*.

INTRODUCTION

Bacteria use mono-, di-, and trinucleotides as second messengers to control fundamental physiological functions in response to signal sensing (1). Among these molecules, cyclic di-3´,5´-adenosine monophosphate (c-di-AMP) is the only nucleotide messenger that must be precisely balanced, since both, its depletion and overproduction can be toxic (2). Its core function is to control cellular integrity by setting homeostasis of osmolytes that in many bacteria are used for osmoregulation (3, 4). Changes of external osmolarity trigger water fluxes across the membrane, which can lead to cell dehydration or swelling and finally collapse or burst when osmobalance mechanisms fail to respond properly (5). As a key component of these mechanisms, c-di-AMP directly targets transport systems for osmoactive and osmoprotective substances such as potassium ions and low-molecular-weight compatible solutes in many bacteria (6-10).

c-di-AMP also plays a central role in host-pathogen interactions and bacterial virulence (11). Secreted c-di-AMP is recognized by host' innate immunity receptors STING, DDX41 and RECON to regulate type I interferon immune response and NF-kB pathways, respectively (12-15). Modulation of intracellular c-di-AMP has been reported to affect virulence of *Streptococcus pyogenes* (16), *Listeria monocytogenes* (17), *Streptococcus pneumonia* (18) and *Mycobacterium tuberculosis* so that the molecule is considered as an attractive antimicrobial target (19).

c-di-AMP synthesis out of two ATP molecules is catalyzed by the diadenylate cyclase (DAC) activity of the DisA_N domain (Pfam PF02457), which was identified in the structural and biochemical analysis of the DNA-integrity scanning protein A (DisA) of *Thermotoga maritima* (20). DisA is mainly present in sporulating firmicutes and actinobacteria (21) and has a conserved domain organization consisting of a N-terminal DAC domain and a C-terminal DNA-binding helix-hairpin-helix domain separated by a linker region (20). C-di-AMP hydrolysis is mediated by the DHH-DHHA1 domain containing the Asp-His-His motif. The multidomain membrane-associated GdpP protein in *Bacillus subtilis* was the first characterized DHH-DHHA1-type phosphodiesterase (PDE) (22). In addition, HD domains with a catalytic His-Asp motif, which were first identified in the PgpH protein in *L. monocytogenes*, also degrade c-di-AMP (17).

However, most actinobacteria contain DisA for c-di-AMP synthesis but do not encode DHH-DHHA1-domain containing or HD-type c-di-AMP PDEs. Hence, we wondered how actinomycetes balance intracellular c-di-AMP. Within actinobacteria, *Streptomyces* are the most extensively studied mycelial organisms and the richest natural source of antibiotics (23). For growth and reproduction, *Streptomyces* undergo a complex developmental life cycle, which involves the conversion between three morphologically and physiologically distinct forms of cell existence. During exponential growth, they proliferate by extension and branching of vegetative hyphae. The switch to stationary phase and onset of the reproductive phase is marked by the erection of aerial hyphae. These filaments elongate and divide into unigenomic prespore compartments that ultimately mature into chains of spores. Completion of the developmental program is easily visible by eye since mature *Streptomyces* spores accumulate a spore pigment. For example, our model species, the chloramphenicol producer *S. venezuelae*, is characterized by a green spore pigment such that colonies turn green at the end of the life cycle (24, 25). Importantly, antibiotic production and morphological differentiation are co-regulated in *Streptomyces*. Hence, studying their developmental biology also provides a better understanding of the control of their secondary metabolism.

In this work, we identified and characterized the PDE superfamily protein AtaC as the founding member of a novel type of c-di-AMP-specific hydrolases. AtaC is broadly distributed in bacteria and the only known c-di-AMP PDE in most actinomycetes. Among others, pathogens such as the causative agent of pneumonia, *S. pneumoniae*, contain an AtaC homolog that we characterize here to be a functional c-di-AMP hydrolase. Our biochemical and structural analyses show that AtaC is a monomeric Mn^{2+}-dependent PDE with high affinity for c-di-AMP. Moreover, we provide direct biochemical evidence that *Streptomyces* DisA is an active DAC and that c-di-AMP produced by DisA is crucial for survival under ionic stress conditions. Further, we show that accumulation of c-di-AMP in the *S. venezuelae ataC* mutant results in profound developmental and growth defects and report the identification of the RCK_C-domain (RCK for regulator of conductance of K^+) containing protein CpeA as the first c-di-AMP binding protein in *Streptomyces*. Overall, in this study we identified and functionally characterized core components of c-di-AMP signaling in *Streptomyces* and link c-di-AMP regulation with ion homeostasis to control differentiation in multicellular bacteria.

RESULTS

DisA is the major c-di-AMP synthetase in *S. venezuelae*

DisA is the sole DAC protein encoded in the *S. venezuelae* genome and is conserved in all sequenced *Streptomyces* strains. To demonstrate DisA DAC activity, we purified N-terminally his-tagged DisA and $DisA_{D86A}$ that carries an alanine instead of aspartate in the active site. We included his-tagged *B. subtilis* DisA ($DisA_{Bsu}$) as a positive control for enzymatic activity (20). [^{32}P]-labeled ATP was added as substrate for *in vitro* DAC assays and the reactions were separated by thin layer chromatography (TLC). DisA synthesized c-di-AMP whereas the mutated $DisA_{D86A}$ failed, demonstrating that *S. venezuelae* DisA is a functional DAC, which requires the conserved catalytic aspartate D_{86} for activity (Figure 1A).

In vivo, DisA is the major source for c-di-AMP in *S. venezuelae* (Figure 1B) (26), however, we reproducibly detected low c-di-AMP levels in Δ*disA* during vegetative growth (10 and 12 h), which disappeared upon onset of sporulation (14 h), suggesting that *S. venezuelae* might contain a non-DAC-domain enzyme capable of c-di-AMP production (Figure 1B). The presence of c-di-AMP throughout the wild type *S. venezuelae* life cycle suggested that *disA* expression is constitutive. To confirm this, we complemented the *disA* mutant by chromosomal insertion of a C-terminally 3xFLAG-tagged *disA* under control of its native promoter. Using a monoclonal anti-FLAG antibody, we detected constant DisA-3xFLAG expression in all developmental stages, which correlated with c-di-AMP production in the wild type under the conditions tested (Figure 1C).

Altogether, our data show that DisA is a functional DAC *in vitro* and *in vivo* and the major enzyme for c-di-AMP production in *S. venezuelae*.

The phosphodiesterase superfamily protein AtaC (Vnz_27310) degrades c-di-AMP

Streptomycetes do not possess PDEs with a DHH-DHHA1 domain or a PgpH-type HD domain, known to degrade c-di-AMP in other bacteria (17, 22), raising the question as to how *S. venezuelae* removes c-di-AMP from the cytoplasm. To find a potentially novel c-di-AMP PDE,

we used interproscan (http://dx.doi.org/10.7717/peerj.167) to search for Pfam PF01663, which is associated with putative type I phosphodiesterases/nucleotide pyrophosphatases. Among others, we found two proteins (Vnz_27310 and Vnz_31010) belonging to the phosphodiesterase and metallophosphatase superfamilies, respectively, that we selected for *in vitro* PDE activity tests.

Purified N-terminally his-tagged Vnz_27310 and Vnz_31010 were assayed *in vitro* using [^{32}P]-labeled c-di-AMP as substrate. While we could not detect [^{32}P]-c-di-AMP cleavage activity for Vnz_31010, Vnz_27310 clearly degraded c-di-AMP to 5′-pApA and finally to AMP (Figure 2A) so that we named Vnz_27310 AtaC for actinobacterial PDE targeting c-di-AMP. Addition of unlabeled c-di-AMP but not of c-di-GMP or cAMP competed with [^{32}P]-c-di-AMP and led to reduced cleavage of the radiolabeled substrate, showing specificity for c-di-AMP (Figure 2A). We analyzed the kinetics of c-di-AMP hydrolysis activity of Vnz_27310 using anion exchange chromatography assays and determined a k_{cat} of 0.2 s^{-1} (Figure S1 A-B), while only a negligible c-di-GMP hydrolysis activity was detected (Figure S1 C). We also compared Vnz_27310-dependent hydrolysis of the linear dinucleotides 5′-pApG and 5′-pGpG to the hydrolysis of 5′-pApA and observed a high hydrolysis activity for 5′-pApA (k_{cat}= 2.1 s^{-1}), whereas the other substrates tested were only degraded to a small extent (Figures 2B and S1 D-F).

Using the PATRIC database (https://www.patricbrc.org), we examined the distribution of the here discovered c-di-AMP PDE (PGF_00172869) and found at least 5374 prokaryotic species containing homologs to AtaC (Table S2), including pathogens such as *S. pneumoniae* and *M. tuberculosis*. AtaC from *S. pneumoniae* (AtaC$_{SPN}$; sequence ID: CVN04004.1) and from *M. tuberculosis* (AtaC$_{MTU}$; sequence ID: CNE38097.1) share 41 % and 47 %, respectively, identical residues with AtaC from *S. venezuelae*. In agreement with the high degree of protein identity, enzyme assays data shown in Figure 2C demonstrate that AtaC$_{SPN}$ also represents a c-di-AMP PDE and AtaC$_{MTU}$ likely has the same function.

In summary, we identified and functionally characterized the sole c-di-AMP hydrolase in *Streptomyces* and a new c-di-AMP signaling component in pathogens and show that AtaC is a conserved phosphodiesterase that efficiently and specifically hydrolyzes c-di-AMP to AMP via the intermediate 5′-pApA.

AtaC is a monomeric Mn^{2+}-dependent phosphodiesterase

To further characterize the c-di-AMP hydrolysis mechanism of AtaC and to gain some structural insights into this PDE, we used HHpred (27) and found two close structural homologs. The core domain of a phosphonoacetate hydrolase (PhnA) from *Sinorhizobium meliloti* 1021 (28) and PDB code 3SZY) showed highest similarity and served as a template for the structural model of AtaC including the putative active site. The predicted active site comprises three aspartates (D68, D227 and D269), three histidines (H231, H270 and H384) and one threonine (T108) (Figure 3A).

Our size-exclusion chromatography (SEC) coupled multi-angle laser light scattering (MALLS) data show that AtaC is a monomer in solution with a molecular weight of 43.7 kDa (Figure S2A). The calculated *ab initio* shape of AtaC from SEC-SAXS (size-exclusion coupled small-angle X-ray scattering) data superimposes well with the HHpred model structure (Figure 3B) and the measured SAXS curve of AtaC is very similar to the theoretical scattering curve of PhnA (Figure S2 B-D), indicating that AtaC and PhnA have a similar shape in solution.

The enzymatic reaction of the PhnA-class hydrolases is known to be catalyzed by two metal ions in the active site (28) so we tested metal binding for AtaC by thermal unfolding assays using nano differential scanning fluorimetry (nanoDSF) assay and observed protein stabilization upon addition of manganese ions (Mn^{2+}) (Figure 3C). Based on the structural similarity to PhnA, we identified potential metal-binding residues in AtaC and generated a variant, $AtaC_{D269N}$, that we expected to lack Mn^{2+} coordination but retain nucleotide binding, as shown for DHH-DHHA1-type PDEs (22, 29). NanoDSF data confirmed stability of $AtaC_{D269N}$ with a melting temperature comparable to the wild type protein when incubated with ethylenediaminetetraacetic acid (EDTA) (Figure 3D). Moreover, $AtaC_{D269N}$ behaved identically to the wild type protein during purification and final SEC. In line with our predictions, $AtaC_{D269N}$ failed to bind Mn^{2+} (Figure 3E) and did not hydrolyze c-di-AMP, as shown using ion exchange chromatography (IEX) based assays (Figure S3A). However, $AtaC_{D269N}$ was still capable of c-di-AMP binding, as confirmed by nanoDSF experiments that showed a shift in the melting curve with increasing ligand concentration (Figure 3F). Using

isothermal titration calorimetry (ITC) analysis we determined the dissociation constant (K_d) of AtaC$_{D269N}$ for c-di-AMP to be 731 ± 266 nM, whereas binding of c-di-GMP could not be detected (Figures 3G-H, and Figure S3B).

Altogether, our combined structural analysis and biochemical data strongly suggest that AtaC uses the same metal-ion dependent mechanism as its structural homolog PhnA for substrate cleavage.

AtaC hydrolyzes c-di-AMP *in vivo*

We quantified c-di-AMP in cell extracts isolated from wild type *S. venezuelae* and the *ataC* null mutant using LC-MS/MS. Our data show that c-di-AMP levels are elevated in the *ataC* mutant during all developmental stages when compared to the wild type, demonstrating that AtaC degrades c-di-AMP *in vivo* and thus is an important component of c-di-AMP metabolism in *S. venezuelae* (Figure 4A). Western blot analysis showed that AtaC is constitutively expressed across the developmental cycle (Figure 4B).

Inactivation of AtaC delays *S. venezuelae* development

To investigate the physiological functions of *disA* and *ataC* and thus of c-di-AMP in *S. venezuelae*, we first analyzed the developmental phenotypes of mutant strains. Colonies of *S. venezuelae* Δ*disA* became green (Figure 5A) and scanning electron microscopy (SEM) confirmed that the Δ*disA* mutant produced spore chains with identical morphology to those of the wild type (Figure 5B). Thus, neither the DisA protein nor the c-di-AMP produced by DisA is required for differentiation.

In contrast, the *ataC* mutant showed a severe delay in development. After 4 days, the Δ*ataC* strain developed aerial hyphae but did not turn green as the wild type (Figure 5A) and SEM imaging showed mainly undifferentiated aerial hyphae, in contrast to the fully sporulated hyphae seen in the wild type (Figure 5B). Moreover, many of the aerial hyphae of the Δ*ataC* mutant had lysed. After extended incubation (7 days), the aerial hyphae of the Δ*ataC* mutant

had largely sporulated, with sporadic non-differentiated and lysed filaments still detected (Figure 5B).

The lysed hyphae seen in the SEMs led us to analyze the growth the Δ*ataC* strain in liquid MYM. As shown in Figure 5C, the *ataC* mutant grew slower than the wild type in exponential phase but reached a similar final OD_{578} after 20 hours. Notably, deletion of *disA* had no effect on growth (Figure 5C).

We could fully complement the defects of Δ*ataC* in development and growth by introduction of the *ataC* wild type allele under the control of its native promoter from the pIJ10170 vector (30) that integrates into the chromosomal *attB*$_{\Phi BT1}$ site (Figure 5A and S4A). In contrast, expression of the *ataC*$_{D269N}$, which cannot cleave c-di-AMP (Figure S3A) from the same integrative vector did not restore the developmental defects caused by *ataC* deletion (Figure 5A), showing that the cleavage of c-di-AMP by AtaC is crucial for normal development of *Streptomyces*.

Altogether, these results demonstrate that elevated levels of c-di-AMP impair growth and development, whereas reduced levels of c-di-AMP do not affect differentiation under standard growth conditions.

disA mutant is more susceptible to ionic osmostress

Since regulation of osmotic balance is a major function of c-di-AMP in many bacteria (3), we next investigated the osmotic stress resistance of strains with altered c-di-AMP levels due to mutations in either *ataC* or *disA*. We spotted serially diluted spores on nutrient agar (NA) medium plates supplemented with 0.5 M NaCl and a control plate without extra added NaCl. On both plates, the growth of the Δ*ataC* strain was slightly impaired resulting in smaller colony size compared to the wild type (Figure 5D), which likely reflects the growth defect of this strain (Figure 5C). We complemented the growth phenotype of Δ*ataC* with the *ataC* wild type allele expressed *in trans* from the integrative vector pIJ10170 from the *attB*$_{\Phi BT1}$ site under the control of the native promoter (Figure 5D).

In contrast, when grown on NA plates containing 0.5 M NaCl, Δ*disA* and *disA*$_{D86A}$ showed pronounced reduction in growth. Expression of wild type *disA* from pIJ10170 fully

complemented the growth defect of Δ*disA* (Figure 5D). The identical Δ*disA* and *disA*$_{D86A}$ phenotypes demonstrate that c-di-AMP produced by DisA is crucial for osmotic stress resistance in *S. venezuelae* (Figure 5D).

In summary, our data revealed that accumulation of c-di-AMP due to *ataC* inactivation, delays development and slows down *Streptomyces* growth in the exponential phase. On the other hand, depletion of c-di-AMP due to *disA* inactivation renders *S. venezuelae* highly susceptible to ionic osmostress.

The RCK_C domain protein CpeA (Vnz_28055) binds c-di-AMP

RCK_C domains are established direct targets of c-di-AMP that have the I(L)I(L)X$_2$DX$_1$RX$_5$NI(L)I(L) signature for ligand binding (Figure 6A) (31). We found the RCK_C-domain protein Vnz_28055 with a putative c-di-AMP binding motif (Figure 6A-B) in 93 *Streptomyces* species for which complete genome sequences are available (32). We purified N-terminally His-tagged Vnz_28055 and applied differential radial capillary action of ligand assay (DRaCALA) to probe interaction between Vnz_28055 and c-di-AMP. DRaCALA allows visualization of protein-bound radiolabeled ligand as a concentrated ring after the application of the protein-ligand mixture onto nitrocellulose (33). With this assay, we confirmed that Vnz_28055 binds [^{32}P]-labeled c-di-AMP (Figure 6C). Excess unlabeled c-di-AMP but not c-di-GMP competed with [^{32}P]-c-di-AMP for binding to Vnz_28055. Thus, we identified Vnz_28055 as the first c-di-AMP binding protein in the genus *Streptomyces*.

Vnz_28055 forms a conserved operon with *vnz_28050*. Some *Streptomyces* species, such as *S. venezuelae*, contain the small open reading frame *vnz_28045* in the same operon (Figure 6B). Vnz_28050 is a structural homolog of the sodium/proton antiporter NapA (PDB code 5BZ3_A) from *Thermus thermophilus* (34), as predicted with 100 % probability using HHpred (27). To test whether Vnz_28055 and Vnz_28050 form a functional interacting unit, we used a bacterial two-hybrid system in which an interaction between bait and target protein reconstitutes a functional adenylate cyclase (Cya), that allows a *E. coli* Δ*cya* mutant to utilize maltose as a carbon source (35). The two proteins were found to form a complex (Figure 6D), supporting our model that c-di-AMP controls the transport activity of Vnz_28050 by binding

to its interaction partner Vnz_28055. Thus, we connect the c-di-AMP function to ionic balance in *Streptomyces* and renamed Vnz_28055-28045 to CpeABC for cation proton exchange component A, B and C.

DISCUSSION

In this work, using the chloramphenicol-producer *S. venezuelae* as a model and a combination of bioinformatic, biochemical, structural and genetic analyses, we identified AtaC as a novel class of c-di-AMP specific PDEs. AtaC is widely distributed in bacteria and represent the only c-di-AMP PDE in the majority of actinobacteria and an up to now unrecognized c-di-AMP signaling component in pathogens, such as *S. pneumoniae* (Figure 2 and Table S2).

AtaC is a soluble, single-domain phosphodiesterase superfamily protein that is monomeric in solution (Figure S2). In solution, AtaC is structurally similar to the alkaline phosphatase superfamily domain of the C-P bond-cleaving enzyme PhnA from *S. meliloti* 1021 (Figure 3A) (28). As described for DHH-DHHA1 domain-containing proteins GdpP and DhhP, and the HD-domain PDE PgpH, AtaC binds Mn^{2+} to hydrolyze c-di-AMP and we show that residue D269 participates in metal-ion coordination contributing to the active site formation (Figure 3C-E) (17, 22, 36). AtaC has a k_{cat} of 0.2 s^{-1} which is comparable to the reported k_{cat} of GdpP (0.55 s^{-1}). Hydrolytically inactive $AtaC_{D269N}$ has a dissociation constant of 0.7 µM, which is highly similar to the K_d of wild type PgpH (0.3 - 0.4 µM) (Figures 3G-H) (17, 22). Since we determined the AtaC dissociation constant using a protein carrying the D269N mutation lacking Mn^{2+}-coordination, the K_d value represents a lower limit as the metal ions bound by the wild type protein likely contribute to c-di-AMP binding. However, while PgpH- and GgdP-type PDEs hydrolyze c-di-AMP exclusively to the linear 5′-pApA, AtaC cleaves c-di-AMP and the intermediate product 5′-pApA to AMP, which has also been shown for DhhP-type PDEs (Figures 2A-B and S1A-B, D) (17, 22, 36). The substrate specificity of AtaC is strictly dependent on two adenosine bases as it shows only weak hydrolysis activity for 5′-pApG and 5′-pGpG in contrast to the DhhP-type PDE TmPDE, which does not distinguish between different nucleobases (Figures 2B and S1E-F) (29).

In *Streptomyces*, AtaC and the DAC DisA are the major regulators of c-di-AMP (Figures 1B and 4A). However, strikingly, the phenotypes of the Δ*ataC* and Δ*disA* mutants with high and low c-di-AMP, respectively, are not invers. On standard growth medium, elevation of intracellular c-di-AMP in Δ*ataC* interferes with growth and ordered hyphae-to-spores transition, while reduction of the second messenger in Δ*disA* does not have any noticeable consequences on these cell functions. On the other hand, when incubated at high external NaCl concentrations, Δ*disA* is severely inhibited in growth, whereas Δ*ataC* grows similarly to the wild type (Figure 5). We found that the RCK_C-domain protein CpeA senses c-di-AMP signals by direct binding of the ligand (Figure 6C). CpeA interacts with CpeB (Figure 6), a structural homolog of the Na^+/H^+ antiporter NapA from *T. thermophilus* and a member of the large monovalent cation / proton antiporter (CPA) superfamily (34). Sodium / proton antiporters exist in all living cells, where they regulate intracellular pH, sodium levels, and cell volume (37). In some bacteria, Na^+/H^+ antiporters use the proton-motive force to extrude sodium out of the cell and are activated at alkaline pH (38). However, in *Staphylococcus aureus*, the CPA-family transporter CpaA has a cytosolic RCK_C domain that binds c-di-AMP to regulate transport activity (6, 39). Similarly, the regulatory RCK_C-domain proteins KtrA and KtrC bind c-di-AMP to control the activity of the corresponding transport units KtrB and KtrD, respectively (31). Thus, in agreement with this general concept and our data, we propose that c-di-AMP binds to the regulatory RCK_C-domain protein CpeA to activate sodium export via CpeB in *Streptomyces*. At low c-di-AMP, CpeB is presumably inactive allowing accumulation of toxic Na^+-ions in the cell and leading to growth defects of Δ*disA* on NaCl containing medium. However, on the other hand, likely constant activity of CpeB at high c-di-AMP in Δ*ataC* may result in continuous proton influx affecting intracellular pH and thus important cellular functions causing growth and developmental defects.

In summary, in this study we identified AtaC as a new component of c-di-AMP metabolism in bacteria and uncovered CpeA as the link between c-di-AMP and ion balance in multicellular actinomycetes.

MATERIAL AND METHODS

For a full explanation of the experimental protocols, see Extended Experimental Procedures in Supplemental Information.

Bacterial strains and plasmids

All strains, plasmids and oligonucleotides used in this study are listed in Table S1. Plasmids and strains were constructed as described in Extended Experimental Procedures.

Protein overexpression and purification

E. coli BL21 (DE3) pLysS and Rosetta (DE3), respectively, were used for protein overexpression. Cultures were grown in presence of required antibiotics at 37°C and induced with IPTG in the logarithmic phase and transferred for growth at 16°C overnight. Strains overexpressing 6xHis-AtaC, 6xHis-AtaC$_{D269N}$, 6xHis-Vnz_31010, and 6xHis-AtaC$_{Spn}$ were supplemented with $MnCl_2$ (17). Cultures were harvested and lysed using a FrenchPress and the proteins were purified via Ni-NTA chromatography. 6xHis-DisA variants and 6xHis-Vnz_28055 were dialyzed twice against 2 L of DisA cyclase buffer (40), and tested PDEs were dialyzed twice against 2 L PDE buffer with 5-10% glycerol (17) at 4°C. Dialyzed proteins were stored at -20 °C. For characterization of biophysical properties of 6xHis-AtaC and 6xHis-AtaC$_{D269N}$, the protein elution was concentrated prior to size exclusion chromatography, flash frozen in liquid nitrogen and stored at -80°C.

Biochemical characterization of DisA and AtaC variants

Biochemical assays using radioactive-labeled substrates were conducted as described in (32). For diadenylate cyclase (DAC) assays, 5 µM 6xHis-tagged DisA$_{Sven}$, DisA$_{D86A}$ or DisA$_{Bsu}$ were incubated with 83 nM [^{32}P]-ATP (Hartmann Analytic) in DisA cyclase buffer. For phosphodiesterase (PDE) assays, 100 nM 6xHis-AtaC or 8 µM 6xHis-Vnz_31010 were mixed with 2 nM [^{32}P]-c-di-AMP (Hartmann Analytic, synthesized using purified 6xHis-DisA$_{Bsu}$) in PDE buffer. For competition, 100 µM unlabeled c-di-AMP, c-di-GMP or cAMP were added on ice prior to starting the PDE reactions with [^{32}P]-c-di-AMP.

Alternatively, enzymatic activity of 6xHis-AtaC and 6xHis-AtaC$_{D269N}$ was detected by separation of non-labeled reaction products by anion exchange chromatography as described in (29). Reaction solutions contained 50 mM Tris (pH = 7.5), 20 mM NaCl, 100 µM $MnCl_2$, 62.5 – 2000 µM ligand (c-di-NMP, 5'-pNpN; N = A or G), 100 nM - 10 µM of 6xHis-AtaC and were

incubated at 37°C for 1 h. The reaction was stopped by separating the reaction products from the protein by ultrafiltration (Centricon, 30 kDA cutoff). The filtrate was diluted to 500 µl with running buffer A (50 mM Tris, pH 9) and loaded on a 1 ml Resource™ Q anion exchange column (GE Healthcare Life Sciences). A linear gradient to 40% running buffer B (50 mM Tris, 1 M NaCl, pH 9) over 20 column volumes (CV) was used to separate the nucleotides. The product peaks were identified by comparison to nucleotide standards, c-di-NMP, pNpN, N = A or G obtained from BioLog.

Differential radial capillary action of ligand assay

DRaCALAs were performed using 2 µg of purified 6xHis-CpeA (Vnz_28055) as described in Roelofs et al 2011 (33) with minor modifications. Purified HD domain of PgpH from *L. monocytogenes* (17) fused to an N-terminal GST tag was used as a positive control. For competition, reactions were supplemented 100 µM of non-labeled c-di-AMP or c-di-GMP prior to addition of [^{32}P]-c-di-AMP.

Western blotting

For detection of 3xFLAG-tagged DisA, Western blot analysis was performed as described in (32) using 5 µg total protein of *S. venezuelae* Δ*disA* expressing the FLAG-tagged *disA* allele from the Φ_{BT1} integration site under the control of the native promoter. Anti-FLAG primary antibody (Sigma) and the anti-mouse IgG-HRP (Thermo Fisher Scientific) were used for detection. AtaC was detected in the wild type strain (10 µg total protein) using polyclonal rabbit anti-AtaC antiserum as primary antibody (generated by Pineda GmbH using purified 6xHis-AtaC) and donkey anti-rabbit-HRP secondary antibody (GE Healthcare). ECL chemiluminescent detection reagent (Perkin Elmer) was used for visualization.

c-di-AMP extraction and quantification

The nucleotide extraction protocol from (2) was adapted to *Streptomyces*. Wild type, Δ*disA* and Δ*ataC* strains were grown in MYM. Samples for c-di-AMP extraction and for determination of the protein concentration were taken every 2 h after initial growth for 10 h. c-di-AMP was extracted using acetonitrile/methanol from cells disrupted using the BeadBlaster (Biozym). Samples were analyzed using LC-MS/MS as described in (2).

Bacterial Adenylate Cyclase Two-Hybrid (BACTH) assays

BACTH system was used to assay protein-protein interaction of CpeA and CpeB *in vivo* (35). Plasmids expressing C-terminal fusions of CpeA and CpeB to T18 and T25 fragments of *cyaA* from *Bordetella pertussis*, respectively, were transformed into *E. coli* W3110 lacking *cya* (41). Co-transformants were spotted on MacConkey agar supplemented with maltose (1%), ampicillin (100 µg/ml), and kanamycin (50 µg/ml). Red colonies indicate cAMP-dependent fermentation of maltose which occurs upon direct interactions of the proteins fused to the otherwise separate adenylate cyclase domains.

Small-angle X-ray scattering

Size-exclusion chromatography coupled small-angle X-ray scattering data (42, 43) for AtaC were collected at the EMBL Hamburg P12 beamline at PETRA3 (DESY, Hamburg). CHROMIXS of the ATSAS Suite (44) was used for analysis and processing of the chromatogram results. In brief, after choosing an appropriate buffer region and averaging of the respective frames, the protein scattering frames from the elution peak were buffer subtracted and averaged. The final protein scattering data were then analyzed using the ATSAS suite. The theoretical scattering curve of the AtaC model derived from HHpred/MODELLER was obtained using CRYSOL (45). Ab initio models were calculated using DAMMIF and averaged using DAMAVER as described earlier (29).

Nano differential scanning fluorimetry

Thermal unfolding experiments of AtaC were performed with a Tycho NT.6 instrument (NanoTemper Technologies). The samples were heated in a glass capillary at a rate of 30 K/min and the internal fluorescence at 330 nm and 350 nm was recorded. Data analysis, data smoothing and calculation of derivatives was done using the internal evaluation features of the Tycho instrument.

Bioinformatic characterization of AtaC and its abundance in prokaryotes

AtaC was identified as a member of the phosphodiesterase family of proteins by annotation of the *S. venezuelae* genome with interproscan (version 5.27-66.0; http://dx.doi.org/10.7717/peerj.167) and searching for proteins harboring type I phosphodiesterase / nucleotide pyrophosphatase domain (Pfam: PF01663).

Scanning electron microscopy (SEM)

SEM was performed as previously described (46).

ACKNOWLEDGEMENTS

We are grateful to Mark J. Buttner and Fabian M. Commichau for helpful discussion and critical reading of the manuscript and thank Matt Bush for technical support with scanning electron micrographs. We thank the staff of the EMBL-Hamburg beamline P12 at PETRA3 (EMBL/DESY, Hamburg, Germany) for outstanding scientific support. We also acknowledge Anna-Lena Hagemann and Annette Garbe for technical support with LC-MS/MS funded by the DFG Priority Program SPP 1879 (KA 730/9-1). Research in Gregor Witte's lab is funded by DFG GRK1721 and the DFG Priority Program SPP 1879 (WI 3717/3-1). Research in Natalia Tschowri's lab is funded by the DFG Emmy Noether-Program (TS 325/1-1) and the DFG Priority Program SPP 1879 (TS 325/2-1).

AUTHOR CONTRIBUTIONS

N.T. designed the study. All authors designed and interpreted experiments, which were performed by A.L., D.J.D., M.M.A-B, G.W., V.K. and K.C.F. The figures were made by A.L., D.J.D., M.M.A-B, G.W. and N.T. The paper was written by A.L., D.J.D., G.W. and N.T. with input from the other authors.

REFERENCES

1. R. Hengge *et al.*, Recent Advances and Current Trends in Nucleotide Second Messenger Signaling in Bacteria. *Journal of molecular biology* **431**, 908-927 (2019).
2. J. Gundlach *et al.*, An essential poison: Synthesis and degradation of cyclic di-AMP in *Bacillus subtilis*. *Journal of bacteriology* 10.1128/JB.00564-15 (2015).
3. F. M. Commichau, J. Gibhardt, S. Halbedel, J. Gundlach, J. Stulke, A Delicate Connection: c-di-AMP Affects Cell Integrity by Controlling Osmolyte Transport. *Trends in microbiology* **26**, 175-185 (2018).
4. J. Gundlach *et al.*, Control of potassium homeostasis is an essential function of the second messenger cyclic di-AMP in *Bacillus subtilis*. *Sci Signal* **10** (2017).
5. E. Bremer, R. Kramer, Responses of Microorganisms to Osmotic Stress. *Annu Rev Microbiol* 10.1146/annurev-micro-020518-115504 (2019).
6. R. M. Corrigan *et al.*, Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9084-9089 (2013).
7. C. F. Schuster *et al.*, The second messenger c-di-AMP inhibits the osmolyte uptake system OpuC in *Staphylococcus aureus*. *Sci Signal* **9**, ra81 (2016).
8. J. Gundlach *et al.*, Sustained sensing in potassium homeostasis: Cyclic di-AMP controls potassium uptake by KimA at the levels of expression and activity. *The Journal of biological chemistry* **294**, 9605-9614 (2019).

9. I. M. Quintana *et al.*, The KupA and KupB proteins of *Lactococcus lactis* IL1403 are novel c-di-AMP receptor proteins responsible for potassium uptake. *Journal of bacteriology* 10.1128/JB.00028-19 (2019).
10. H. T. Pham *et al.*, Enhanced uptake of potassium or glycine betaine or export of cyclic-di-AMP restores osmoresistance in a high cyclic-di-AMP *Lactococcus lactis* mutant. *PLoS genetics* **14**, e1007574 (2018).
11. L. Devaux, P. A. Kaminski, P. Trieu-Cuot, A. Firon, Cyclic di-AMP in host-pathogen interactions. *Current opinion in microbiology* **41**, 21-28 (2018).
12. J. J. Woodward, A. T. Iavarone, D. A. Portnoy, c-di-AMP secreted by intracellular *Listeria monocytogenes* activates a host type I interferon response. *Science* **328**, 1703-1705 (2010).
13. A. P. McFarland *et al.*, RECON-Dependent Inflammation in Hepatocytes Enhances *Listeria monocytogenes* Cell-to-Cell Spread. *mBio* **9** (2018).
14. K. Parvatiyar *et al.*, The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nat Immunol* **13**, 1155-1161 (2012).
15. J. R. Barker *et al.*, STING-dependent recognition of cyclic di-AMP mediates type I interferon responses during *Chlamydia trachomatis* infection. *mBio* **4**, e00018-00013 (2013).
16. T. Fahmi, S. Faozia, G. C. Port, K. H. Cho, The Second Messenger c-di-AMP Regulates Diverse Cellular Pathways Involved in Stress Response, Biofilm Formation, Cell Wall Homeostasis, SpeB Expression, and Virulence in *Streptococcus pyogenes*. *Infect Immun* **87** (2019).
17. T. N. Huynh *et al.*, An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E747-756 (2015).
18. Y. Bai *et al.*, Two DHH subfamily 1 proteins in *Streptococcus pneumoniae* possess cyclic di-AMP phosphodiesterase activity and affect bacterial growth and virulence. *Journal of bacteriology* **195**, 5123-5132 (2013).
19. R. J. Dey *et al.*, Inhibition of innate immune cytosolic surveillance by an *M. tuberculosis* phosphodiesterase. *Nature chemical biology* **13**, 210-217 (2017).
20. G. Witte, S. Hartung, K. Buttner, K. P. Hopfner, Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**, 167-178 (2008).
21. R. M. Corrigan, A. Grundling, Cyclic di-AMP: another second messenger enters the fray. *Nature reviews. Microbiology* **11**, 513-524 (2013).
22. F. Rao *et al.*, YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. *The Journal of biological chemistry* **285**, 473-482 (2010).
23. K. F. Chater, Recent advances in understanding *Streptomyces*. *F1000Res* **5**, 2795 (2016).
24. M. J. Bush, N. Tschowri, S. Schlimpert, K. Flardh, M. J. Buttner, c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. *Nature reviews. Microbiology* **13**, 749-760 (2015).

25. N. Tschowri, Cyclic Dinucleotide-Controlled Regulatory Pathways in *Streptomyces* Species. *Journal of bacteriology* **198**, 47-54 (2016).
26. R. J. St-Onge *et al.*, Nucleotide second messenger-mediated regulation of a muralytic enzyme in *streptomyces*. *Molecular microbiology* **96**, 779-795 (2015).
27. A. Hildebrand, M. Remmert, A. Biegert, J. Soding, Fast and accurate automatic structure prediction with HHpred. *Proteins* **77 Suppl 9**, 128-132 (2009).
28. V. Agarwal, S. A. Borisova, W. W. Metcalf, W. A. van der Donk, S. K. Nair, Structural and mechanistic insights into C-P bond hydrolysis by phosphonoacetate hydrolase. *Chem Biol* **18**, 1230-1240 (2011).
29. D. J. Drexler, M. Muller, C. A. Rojas-Cordova, A. M. Bandera, G. Witte, Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from *Thermotoga maritima*. *Structure* **25**, 1887-1897 e1884 (2017).
30. S. Schlimpert *et al.*, Two dynamin-like proteins stabilize FtsZ rings during *Streptomyces* sporulation. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E6176-E6183 (2017).
31. M. Schrecker, D. Wunnicke, I. Hanelt, How RCK domains regulate gating of K+ channels. *Biological chemistry* **400**, 1303-1322 (2019).
32. M. M. Al-Bassam, J. Haist, S. A. Neumann, S. Lindenberg, N. Tschowri, Expression Patterns, Genomic Conservation and Input Into Developmental Regulation of the GGDEF/EAL/HD-GYP Domain Proteins in *Streptomyces*. *Front Microbiol* **9**, 2524 (2018).
33. K. G. Roelofs, J. Wang, H. O. Sintim, V. T. Lee, Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15528-15533 (2011).
34. C. Lee *et al.*, A two-domain elevator mechanism for sodium/proton antiport. *Nature* **501**, 573-577 (2013).
35. G. Karimova, J. Pidoux, A. Ullmann, D. Ladant, A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5752-5756 (1998).
36. M. Ye *et al.*, DhhP, a cyclic di-AMP phosphodiesterase of *Borrelia burgdorferi*, is essential for cell growth and virulence. *Infect Immun* **82**, 1840-1849 (2014).
37. E. Padan, M. Venturi, Y. Gerchman, N. Dover, Na(+)/H(+) antiporters. *Biochim Biophys Acta* **1505**, 144-157 (2001).
38. T. A. Krulwich, G. Sachs, E. Padan, Molecular aspects of bacterial pH sensing and homeostasis. *Nature reviews. Microbiology* **9**, 330-343 (2011).
39. K. H. Chin *et al.*, Structural Insights into the Distinct Binding Mode of Cyclic Di-AMP with SaCpaA_RCK. *Biochemistry* **54**, 4936-4951 (2015).
40. M. Christen, B. Christen, M. Folcher, A. Schauerte, U. Jenal, Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *The Journal of biological chemistry* **280**, 30829-30837 (2005).
41. S. Herbst *et al.*, Transmembrane redox control and proteolysis of PdeC, a novel type of c-di-GMP phosphodiesterase. *The EMBO journal* **37** (2018).

42. C. M. Jeffries *et al.*, Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat Protoc* **11**, 2122-2153 (2016).
43. C. E. Blanchet *et al.*, Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr* **48**, 431-443 (2015).
44. D. Franke *et al.*, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212-1225 (2017).
45. D. I. Svergun, C. Barberato, M. H. J. Koch, CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates *J. Appl. Cryst.* **28**, 768-773 (1995).
46. N. Tschowri *et al.*, Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* **158**, 1136-1147 (2014).
47. H. Kim *et al.*, Structural Studies of Potassium Transport Protein KtrA Regulator of Conductance of K+ (RCK) C Domain in Complex with Cyclic Diadenosine Monophosphate (c-di-AMP). *The Journal of biological chemistry* **290**, 16393-16402 (2015).
48. R. Rocha, C. M. Teixeira-Duarte, J. M. P. Jorge, J. H. Morais-Cabral, Characterization of the molecular properties of KtrC, a second RCK domain that regulates a Ktr channel in *Bacillus subtilis*. *J Struct Biol* **205**, 34-43 (2019).

FIGURES WITH LEGENDS

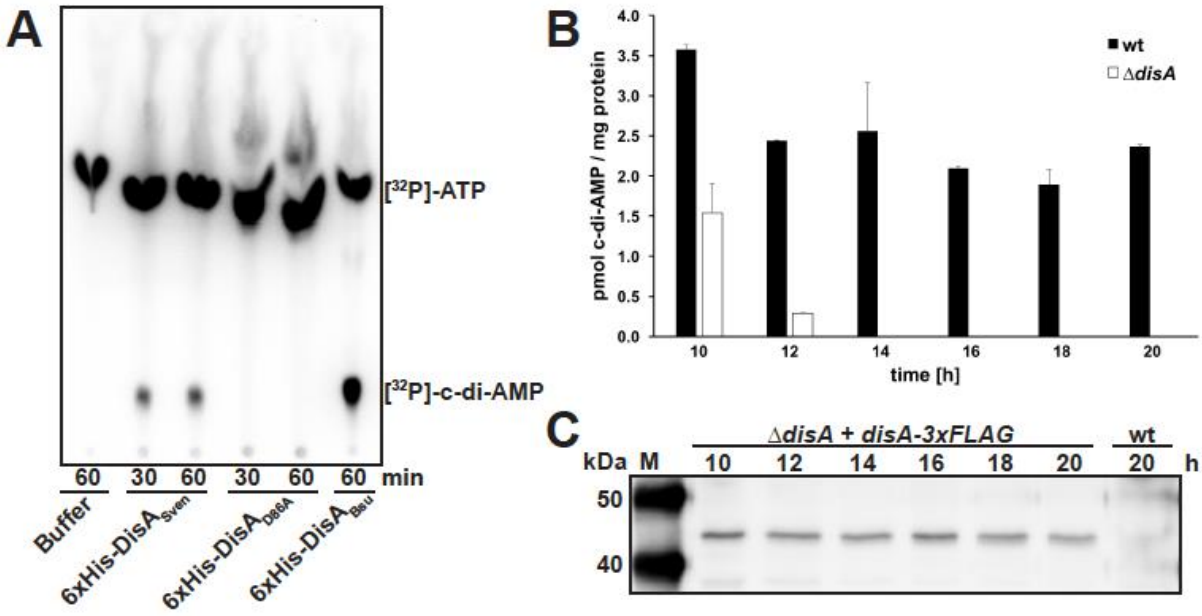
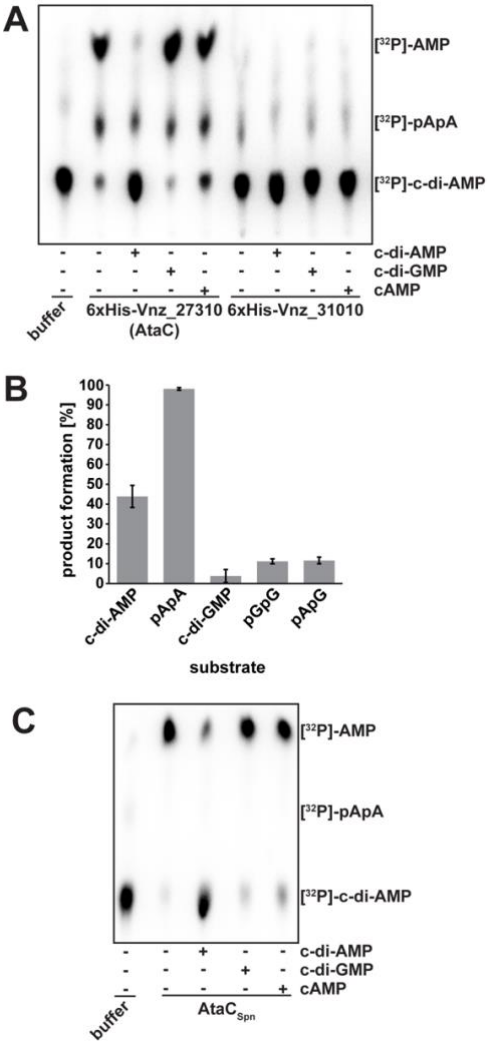


Figure 1. DisA is an active diadenylate cyclase *in vitro* and *in vivo*.

(A) Thin layer chromatography of diadenylate cyclase (DAC) assay with purified 6xHis-DisA$_{Sven}$ and 6xHis-DisA$_{D86A}$, and [$_{32}$P]-ATP as substrate. Migration of [$_{32}$P]-ATP in buffer is shown in lane 1. 6xHis-DisA$_{Bsu}$ served as positive control for DAC activity.

(B) Intracellular c-di-AMP levels in *S. venezuelae* wild type and *ΔdisA* during late vegetative growth (10 to 12 h), early sporulation (14 to 16 h) and sporulation (from 18 h). Data are presented as mean of biological replicates ± standard deviation (n=3).

(C) Expression profile of DisA-3xFLAG in a *disA* mutant complemented with *disA-3xFLAG* under control of *disA* promoter grown in liquid sporulation medium (MYM). DisA-3xFLAG was detected using a monoclonal anti-FLAG antibody. Wild type served as negative control.



592
593

594 **Figure 2. AtaC is a c-di-AMP-specific phosphodiesterase (PDE).**

595 Thin layer chromatography of PDE assay of AtaC and Vnz_31010 from *S. venezueale* (A) and
596 *S. pneumoniae* (AtaCSPN) (C) with [32P]-c-di-AMP. Radioactively labeled c-di-AMP in buffer
597 migrates as shown in lane 1. In samples used for competition, unlabeled c-di-AMP, c-di-GMP
598 or cAMP (indicated by "+") were added in excess before starting the reaction with [32P]-c-di-
599 AMP.

600 (B) AtaC activity assay by ion-exchange chromatography runs on a 1 ml Resource Q column
601 of the reaction products after 1 h incubation from 100 µl reactions containing 100 nM AtaC +
602 250 µM c-di-AMP, 5′-pApA, c-di-GMP, 5′-pGpG or 5′-pApG (n=3).

603

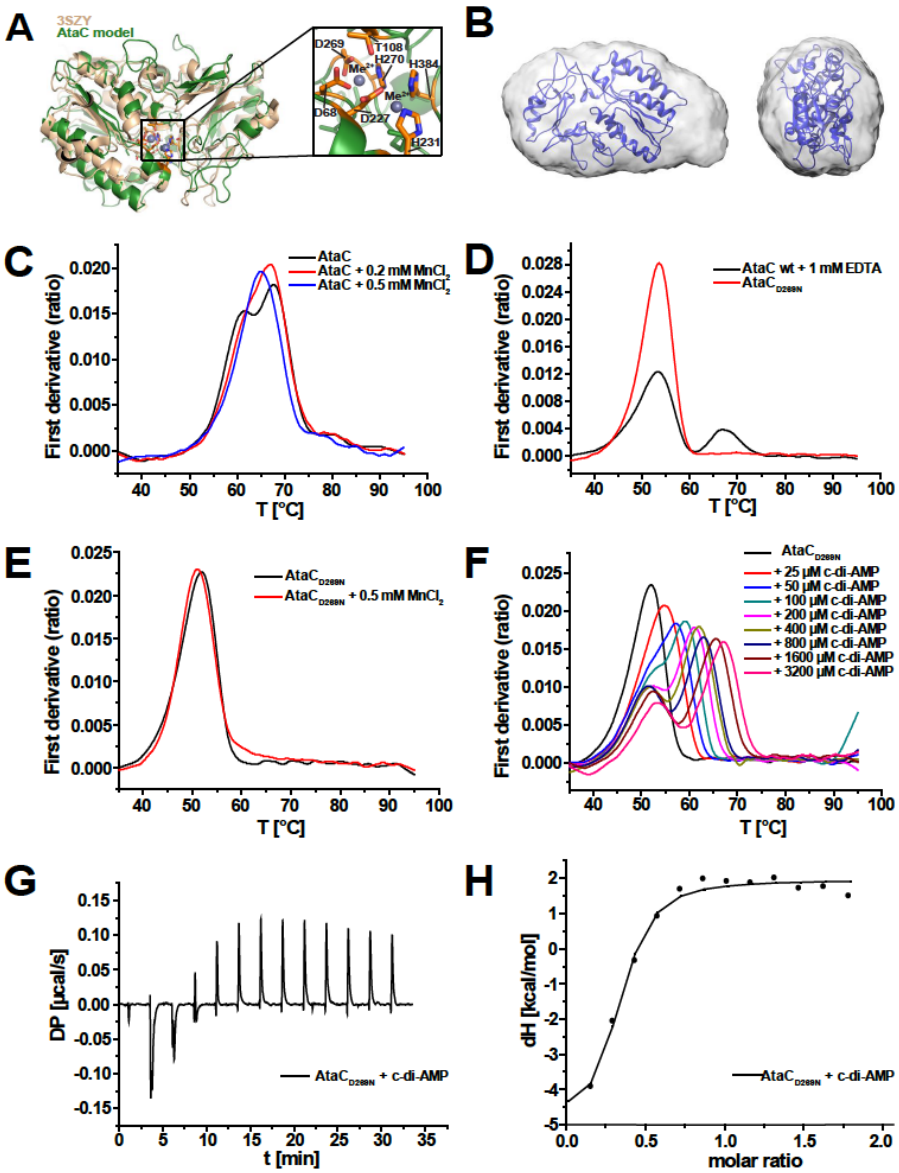


Figure 3. AtaC is a monomeric Mn^{2+}-dependent phosphodiesterase.

(A) Model of AtaC obtained from HHpred/MODELLER (green) superimposed with best match 3SZY (beige), Zoom-In shows the predicted active site, annotated with all most conserved residues.

(B) Modelled structure from (A) superimposed with the final averaged and filtered *ab initio* shape (16 ab initio models averaged) from SEC-SAXS with front view (left) and side view (right).

(C) nanoDSF thermal shift first derivative curves of 10µM apo AtaC (black), 10µM AtaC + 0.2 mM $MnCl_2$ (red) and 10µM AtaC + 0.5 mM $MnCl_2$ (blue).

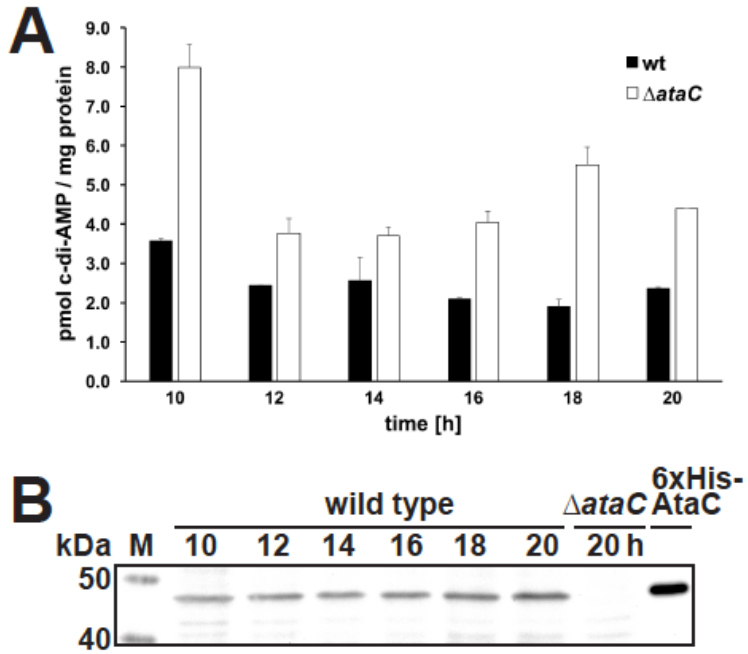
(D) nanoDSF thermal shift first derivative curves of 10 µM AtaC + 1 mM EDTA (black) and 10µM $AtaC_{D269N}$ (red).

(E) nanoDSF thermal shift first derivative curves of 10 µM $AtaC_{D269N}$ (black) and $AtaC_{D269N}$ + 0.5 mM $MnCl_2$ (red).

(F) nanoDSF thermal shift first derivative curves of 10 µM $AtaC_{D269N}$ + c-di-AMP (25 µM - 3200 µM).

(G) ITC measurement raw data of 23 µM $AtaC_{D269N}$ mutant titrated with 231 µM c-di-AMP.

(H) Binding curve and fit of ITC titration of the $AtaC_{D269N}$ mutant with c-di-AMP (K_D = 731 ± 266 nM) (n=3).



643
644

645 **Figure 4. AtaC hydrolyzes c-di-AMP *in vivo* and is constitutively expressed during the life**
646 **cycle of *S. venezuelae*.**

647 (A) Intracellular c-di-AMP levels in *S. venezuelae* wild type and Δ*ataC* during late vegetative
648 growth (10 to 12 h), early sporulation (14 to 16 h) and sporulation (from 18 h). Data are
649 presented as mean of biological replicates ± standard deviation (n=3).

650 (B) Expression profile of AtaC in *S. venezuelae* wild type grown in liquid sporulation medium
651 (MYM). AtaC was detected using a polyclonal anti-AtaC antiserum. Protein samples harvested
652 from Δ*ataC* served as negative control and purified 6xHis-AtaC as positive control,
653 respectively.

654
655
656
657
658

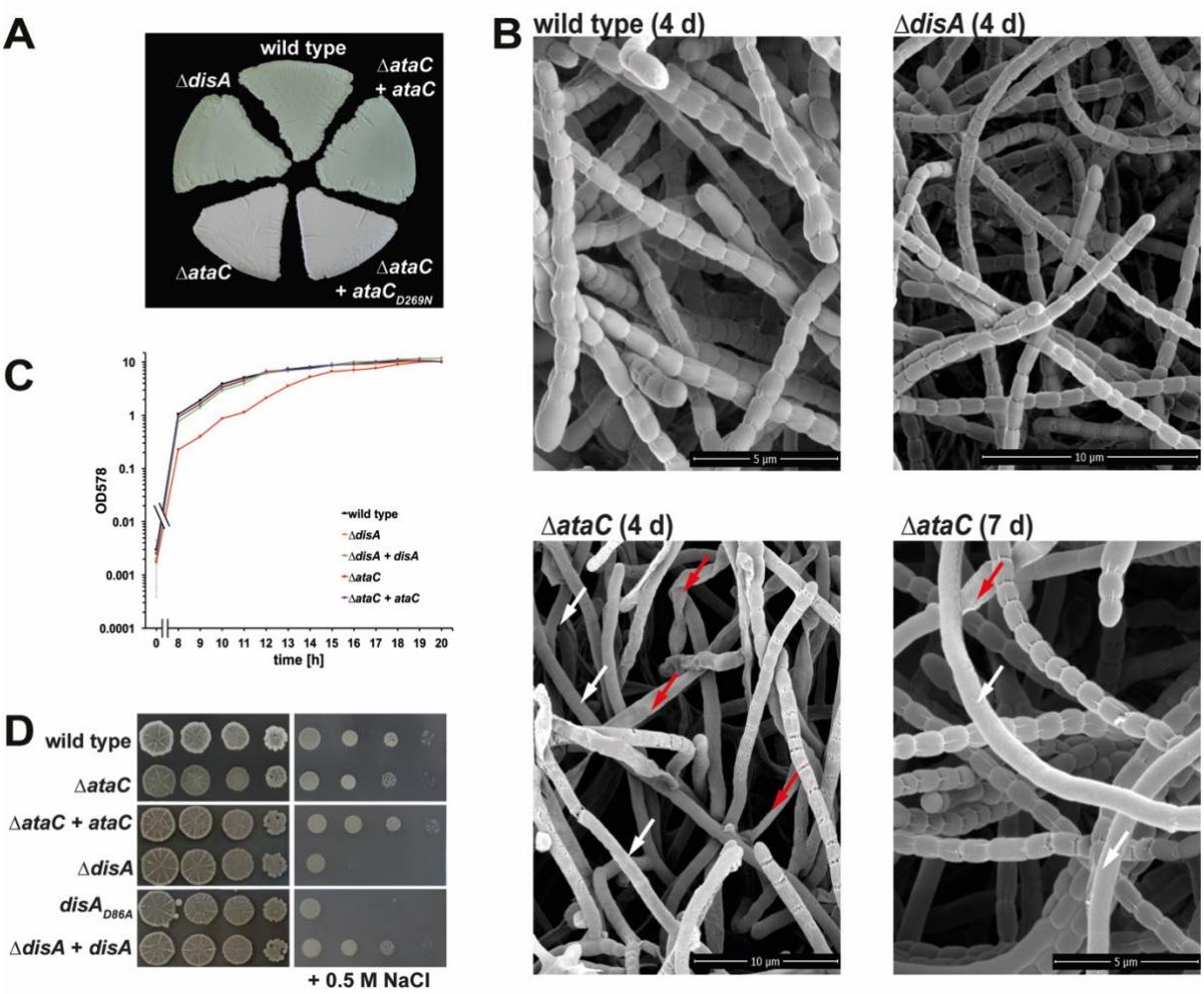


Figure 5. Mutagenesis of c-di-AMP metabolizing enzymes impacts development and ionic stress resistance in *S. venezuelae*.

(A) Green morphologies of *S. venezuelae* wild type and Δ*disA* indicate formation of mature spores after 4 days of growth at 30°C on solid sporulation medium (MYM agar). *S. venezuelae* Δ*ataC* fails to accumulate the spore pigment and remains white after the same incubation time. Wild type *ataC* allele complements the Δ*ataC* phenotype but not the enzymatically inactive variant *ataC*$_{D269N}$, when expressed *in trans* under control of the native promoter.

(B) Scanning electron micrographs showing that after 4 days of incubation on MYM, *S. venezuelae* wild type and Δ*disA* form spores but Δ*ataC* consists predominantly of non-sporulating aerial hyphae (white arrows) and forms flat, likely lysed hyphae (red arrows). After

671 7 days of growth, Δ*ataC* produced wild type-like spore chains but occasional non-differentiated
672 and lysed hyphae were still detectable.
673 (C) Deletion of *ataC* leads to a growth defect in *S. venezuelae*. c-di-AMP mutants were grown
674 in liquid sporulation medium (MYM) at 30°C and optical density was measured at 578 nm.
675 Δ*ataC* growth is delayed by 3 h and can be restored by expression of the wild type allele under
676 control of its native promoter from the *attB*$_{\Phi BT1}$ site.
677 (D) Osmotic stress resistance of c-di-AMP mutants. Serial dilutions of spores were spotted on
678 nutrient agar [NA] without additional salt or supplemented with 0.5 M NaCl and grown at 30°C
679 for ~2 days. Δ*disA* and *disA*$_{D86A}$ (expressing inactive DisA) are hypersensitive to salt stress.
680

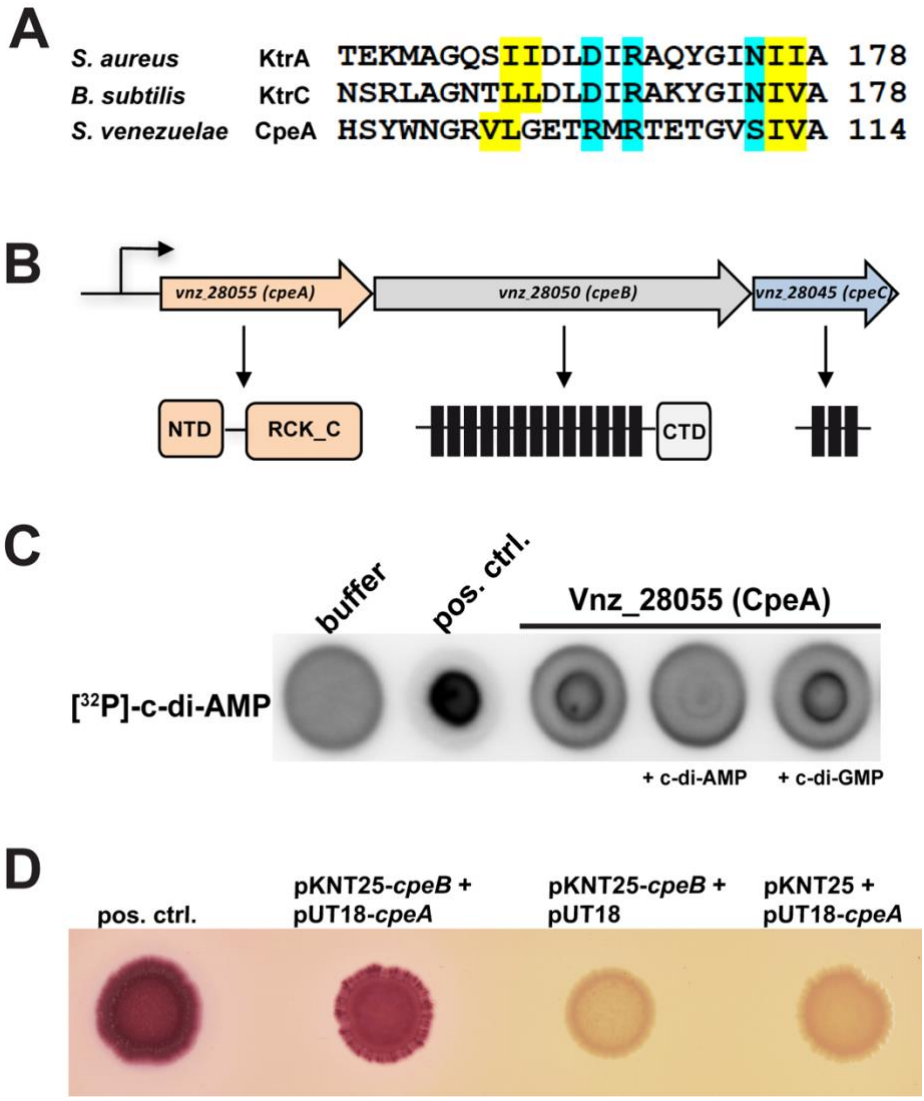


Figure 6. Vnz_28055 (CpeA) binds c-di-AMP.

(A) Alignment of the c-di-AMP binding regions in RCK_C (regulator of conductance of K^+, Carboxy-terminal) domains was generated using Clustal Omega. C-di-AMP binding residues in KtrA (*S. aureus*;(47)), KtrC (*B. subtilis*;(48)) and conserved amino acids in CpeA are highlighted. Amino acids that form the hydrophobic patch are shown in yellow, residues involved in hydrophilic coordination are highlighted in cyan.

(B) *CpeA (vnz_28055)*, *cpeB (vnz_28050)* and *cpeC (vnz_28045)* form an operon in *S. venezuelae*. CpeA has an N-terminal domain (NTD) of unknown function and a C-terminal RCK_C domain. NheB is a predicted structural homolog to the Na^+/H^+ antiporter NapA (34).

692 It consists of 13 transmembrane (TM) domains and a cytosolic fraction at the C-terminus
693 (CTD). NheC is a predicted membrane protein with 3 TM domains.
694 (C) CpeA binds [^{32}P]-c-di-AMP in DRaCALAs. Binding of the radiolabeled ligand is indicated
695 by dark spots centered on the nitrocellulose. In competition assays, excess (100 µM) cold c-di-
696 AMP or c-di-GMP, respectively, was added to the binding reaction containing [^{32}P]-c-di-AMP
697 and 6xHis-CpeA.
698 (D) Adenylate cyclase-based two hybrid assays revealing that CpeA and CpeB interact *in vivo*.
699 Using pKNT25 and pUT18, the T25 and T18 fragments of adenylate cyclase are attached at the
700 C-termini of CpeB and CpeA, respectively. As a positive control, the leucin zipper part of the
701 yeast GCN4 protein was used. Spotted co-transformants were grown for 20 h at 30°C with
702 further incubation at room temperature for ca. 3 days.
703
704
705

SUPPLEMENTARY MATERIAL

c-di-AMP hydrolysis by a novel type of phosphodiesterase promotes differentiation of multicellular bacteria

Andreas Latoscha, David Jan Drexler, Mahmoud M. Al-Bassam, Volkhard Kaever[4], Kim C. Findlay, Gregor Witte* and Natalia Tschowri*

* Corresponding author: natalia.tschowri@hu-berlin.de

* Correspondence may also be addressed to: witte@genzentrum.lmu.de

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Bacterial strains and growth conditions

All strains used in this study are listed in Table S1. *Escherichia coli* strains were grown in LB medium under aerobic conditions at 37°C. If required, LB was supplemented with 100 µg/ml ampicillin (Amp), 50 µg/ml kanamycin (Kan), 50 µg/ml apramycin (Apr) and/or 15 µg/ml chloramphenicol (Cam). If hygromycin B (Hyg) was used, LB agar was replaced by Nutrient Agar (NA; Roth) and LB was substituted by LBon (LB without salt) with addition of 16 µg/ml and 22 µg/ml Hyg, respectively. *S. venezuelae* strains (Table S1) were grown aerobically at 30 °C in liquid Maltose-Yeast Extract-Malt Extract (MYM) medium (1) supplemented with trace element solution (2) or on MYM agar. For growth analysis, 50 ml MYM were inoculated with spores at a final concentration of 10_6 CFU/µl and OD was measured at 578 nm. To study development, 12 µl of 10_5 CFU/ml *S. venezuelae* spores were spread as patches on MYM agar and bacteria were photographed using a Canon EOS 1300D (W) camera after 4 days of growth at 30 °C. For osmostress experiments, 10 µl of serially diluted of *S. venezuelae* spores (10_1 to 10_4 CFU/µl) were dropped on NA medium with or without 0.5 M NaCl, respectively. Plates were incubated at 30°C and pictures were taken using a Canon EOS 1300D (W) camera. For *in vivo* interaction studies using Bacterial Adenylate Cyclase Two-Hybrid (BACTH) assays three corresponding single transformants of *E. coli* W3110 Δ*cya* were suspended in 1 ml of sterile phosphate buffered saline (PBS) and 3 µl of the resulting suspension was spotted on a plate containing MacConkey Base Agar (Difco) supplemented with Amp (100 µg/ml), Kan (50

µg/ml) and maltose (1%). The plate was photographed using a Canon EOS 1300D (W) camera after growth at 30 °C for 20 h with further incubation at room temperature for three days.

Generation and complementation of *S. venezuelae* c-di-AMP mutants

Oligonucleotides and bacterial strains used for mutagenesis are listed Table S1.

Generation of vnz_27310 (ataC) deletion mutant and phage transduction. The *ataC* deletion was conducted using a modified Redirect PCR targeting protocol (3, 4). *E. coli* BW25113/pIJ790 cells induced for λ Red mediated recombination and containing cosmid PI1_F15 (with *ataC*) were transformed with PCR amplified *aac(3)IV-oriT* (*apr-oriT*) cassette extended with regions homologous to the *ataC* locus. After growth of transformants in presence of apramycin only, cosmids were isolated and re-transformed into *E. coli* W3110 with selection for Apr resistance. Correct integration of the Apr resistance cassette was verified by PCR using test primers annealing to the *ataC* flanking region (Table S1). PI1_F15 *ataC::apr* was transformed into strain ET12567/pUZ8002 and conjugated into wild type *S. venezuelae.* Bacteria were plated on SFM agar, incubated overnight at room temperature (RT), overlaid with 20 µl of 25 mg/ml nalidixic acid (Nal) and 50 mg/ml Apr in 2 ml ddH_2O and incubated at 30°C. *S. venezuelae* colonies were selected on NA containing Apr and Nal to remove *E. coli*. Kan^S and Apr^R mutants were confirmed by PCR.

The *ataC::apr* allele was transduced into a new wild type background via SV1 phage using a modified protocol from (5). Briefly, SV1 wild type phages were diluted and mixed with *S. venezuelae* Δ*ataC* spores. After overnight incubation at 30°C, plates were soaked with LB at RT and phages were gathered and filtered using a 0.45 µm filter. 100 µl phages containing the *ataC::apr* allele were plated with wild type spores on MYM agar and incubated overnight at RT. For selection of desired transductants, plates were overlaid with 20 µl 50 mg/ml Apr in 2 ml ddH_2O. The *ataC* deletion was confirmed by PCR.

Generation of disA deletion mutant. The *disA* mutant was generated by transduction of the *disA::apr* allele (6) into the *S. venezuelae* wild type using SV1 phage.

Generation of disA$_{D86A}$ point mutation. The point mutation was introduced using a combination of modified Redirect PCR targeting and single stranded DNA recombineering protocols as described in (7).The Kan resistance cassette in cosmid SV-4-B12 (containing *disA*) was exchanged with *apr-oriT* extended with homologous regions to *neo*. The resulting cosmid was used to generate the *E. coli* strain HME68/SV-4-B12 *neo::apr-oriT* which was induced for

λ Red recombination and electroporated with the mutagenic oligonucleotide disA_D86Achr_rev and oligo100 (8) and plated on MacConkey agar containing 1% galactose and Apr. Red, Apr^R clones were analyzed for the *disA::disA*$_{D86A}$ allele by PCR using a primer pair specific for the D86A mutation and sequencing. Purified cosmid was electroporated into ET12567/pUZ8002 and conjugated into wild type *S. venezuelae*. Single colonies were selected on NA medium containing Apr and Nal followed by growth on NA without antibiotics to achieve loss of the cosmid. Colonies sensitive to Apr were verified for the *disA* D86A mutation by sequencing.

Complementation of the deletion mutants. To complement the *disA* deletion, a DNA fragment containing the wild type allele with 528 bp upstream of the start codon (including *disA* promoter) was cloned into pIJ10170 or its derivative p3xFLAG which allow integration at the *attB*$_{\Phi BT1}$ site in the *S. venezuelae* chromosome. pIJ10170-*disA* and p3xFLAG-*disA*, respectively, were conjugated into *disA::apr* using *E. coli* ET12567/pUZ8002. After selection on NA medium containing Nal and Hyg, Hyg^R colonies were grown on MYM agar for spore stock generation.

The complementation of the *ataC* deletion mutant with the wild type allele was conducted similarly. Here, using an overlap PCR, a DNA fragment corresponding to 200 bp upstream of the start codon of *vnz_27305* was fused to full length *ataC* resulting in the construct *vnz_27305prom-ataC*. The fragment was cloned into pIJ10170 resulting in the plasmid pSVAL11 and introduced into the Δ*ataC* chromosome as described above. For complementation of Δ*ataC* with the D269N allele, pSVAL11 was used as template in a backbone PCR to amplify a circular construct with the primers D269N_backbone_f and D269N_backbone_r. The plasmid pIJ10170-*vnz_27305prom-ataC*$_{D269N}$ was generated, the mutation confirmed by sequencing and conjugated into Δ*ataC* as described above.

Construction of plasmids

Oligonucleotides used for cloning are listed in Table S1. *disA*, *vnz_31010*, *ataC* and *vnz_28055* were amplified from *S. venezuelae* genomic DNA (gDNA). D86A point mutation in *disA* was introduced by following the four-primer/two-step PCR protocol (9). PCR products of all constructs were cloned into the pET15b vector. The *ataC*$_{D269N}$ construct was obtained using quick change site directed mutagenesis using the pET15b-*ataC* plasmid as a template. Codon optimized *ataC*$_{Spn}$ was synthesized *de novo* and cloned into pET15b via the NdeI and BamHI restriction sites by GenScript.

Full-length *cpeA* (*vnz_28055*) and *cpeB* (*vnz_28050*) excluding the respective stop codons were cloned into pUT18 and pKNT25 (Euromedex). The resulting constructs carry in-frame fusions of the sequences encoding the T18 and T25 fragments of *cyaA* from *Bordetella pertussis* to the 3' end of *cpeA* and *cpeB*, respectively. Expression of the fused genes is under control of the *lac* promoter.

Protein overexpression and purification

pET15b constructs were transformed into *E. coli* BL21 (DE3) pLysS. Strain Rosetta (DE3) pET28-*disA*$_{Bsu}$ was directly used for overexpression. 1 L LB containing Amp and Cam (and 0.2% glucose in case of PDEs) was inoculated 1:100 with overnight cultures and grown with shaking at 37°C. For DisA$_{Bsu}$ overexpression Amp was replaced with 25 µg/ml Kan. Cultures were induced with a final concentration of 0.1-0.2 mM IPTG at OD_{578} between 0.5 and 0.7; cultures for PDE overexpression were supplemented with 0.35 mM $MnCl_2$ (10). Proteins were overexpressed overnight at 16°C and shaking. Subsequently, cultures were pelleted and lysed using a FrenchPress. Strains expressing DisA variants and 6xHis-Vnz_28055 were lysed in DisA lysis buffer (20 mM Tris HCl, pH 8; 300 mM NaCl, 10% glycerol, 20 mM imidazole; 0.05% Triton X-100; 0.5 mM DTT; 5 mM $MgCl_2$) supplemented with cOmplete protease inhibitor cocktail tablets, EDTA-free (Roche). Strains expressing 6xHis-Vnz_31010, 6xHis-AtaC and 6xHis-AtaC$_{Spn}$ were lysed in PDE lysis buffer containing cOmplete protease inhibitor cocktail tablets, EDTA-free (similar to DisA lysis buffer but Tris HCl, pH 8 replaced by 20 mM Tris HCl 7.5; $MgCl_2$ replaced by 10 mM $MnCl_2$). Clarified lysate supernatants of 6xHis-tagged proteins were loaded on 0.5-1 ml 50% Ni-NTA SuperFlow (iba) overnight at 4°C. Then, the matrix was washed with respective lysis buffers. DisA protein variants and 6xHis-Vnz_28055 were eluted with the following buffer: 50 mM Tris HCl, pH 8; 300 mM NaCl; 10% glycerol; 250 mM imidazole; 0.5 mM DTT; 5 mM $MgCl_2$. The PDE elution buffer was similar to DisA elution buffer but containing 50 mM Tris HCl, 7.5 instead of Tris HCl, pH 8 and 10 mM $MnCl_2$ instead of $MgCl_2$. Fractions containing eluted proteins (identified by Coomassie staining of 12% polyacrylamide gels) were pooled. Eluates of DisA variants and 6xHis-Vnz_28055 were dialyzed twice against 2 L of DisA cyclase buffer (25 mM Tris HCl, pH 8; 250 mM NaCl, 10 mM MgCl, 5 mM β-mercaptoethanol, 10% glycerol (modified from (11), and tested PDEs were dialyzed twice against 2 L PDE buffer with 5-10% glycerol (20 mM Tris HCl, pH 7.5; 50 mM NaCl; 10 mM $MnCl_2$ (modified from(10) at 4°C under stirring. Dialyzed proteins were stored

at -20°C until further use in diadenylate cyclase (DAC), differential radial capillary action of ligand (DRACaLA) or phosphodiesterase (PDE) assays.

For characterization of biophysical properties of 6xHis-AtaC and 6xHis-AtaC$_{D269N}$, Rosetta (DE3) cell pellets containing pET15b-*ataC* and pET15b-*ataC*$_{D269N}$ constructs, respectively, were resuspended in buffer A (20 mM HEPES, 300 mM NaCl, 20 mM Imidazole, 10% glycerol, 0.5 mM $MnCl_2$, pH 7.5) and lyzed by sonication. After centrifugation, clear supernatant was loaded on Ni-NTA columns. The columns were washed with buffer A and the protein was eluted with buffer B (20 mM HEPES, 100 mM NaCl, 250 mM imidazole, 10% glycerol, 0.5 mM $MnCl_2$, pH 7.5). The protein elution was concentrated prior to size exclusion chromatography on a HiLoad Superdex 200 column (GE Healthcare) equilibrated with buffer C (20 mM HEPES, 100 mM NaCl, 0.5 mM $MnCl_2$, pH 7.5). The pure protein was concentrated, flash frozen in liquid nitrogen and stored at -80°C.

c-di-AMP extraction and quantification

The nucleotide extraction protocol from (12) was adapted to *Streptomyces*. Wild type, Δ*disA* and Δ*ataC* strains were grown in 100 ml MYM. Beginning with 10 h, 5 ml samples for c-di-AMP extraction and two 1 ml samples for protein determination were taken every 2 h. c-di-AMP samples were centrifuged at 4000 rpm and 4°C for 15 min using a swing rotor (Heraeus Megafuge 16R, Thermo Scientific), frozen in liquid nitrogen and stored at -80°C. Protein samples were centrifuged at max. speed and stored at -20°C.

For c-di-AMP extraction, samples were suspended in 800 µl Extraction mixture II (acetonitrile/methanol/water [2:2:1]), transferred into 2 ml screw cap tubes prefilled with 0.1 mm silica beads (Biozym), shock frozen for 15 s in liquid nitrogen and heated for 10 min at 95°C. After cooling on ice, samples were disrupted using the BeadBlaster at 4°C with 2 cycles at 6 m/s for 45 s and 2 min interval. Samples were cooled for 15 min on ice and centrifuged at max. speed and 4°C for 15 min. Supernatants were transferred into a 2 ml reaction tubes. Remaining pellets were suspended in 600 µl Extraction mixture I (acetonitrile/methanol [1:1]), pulsed two times for 30 s at 6 m/s with a 60s interval, incubated on ice and centrifuged as above. The extraction with 600 µl Extraction mixture I was repeated once. All supernatants (~2 ml) were combined and stored for protein precipitation for two days at -20°C. Precipitated proteins were removed by centrifugation and the precipitation step was repeated. Finally, samples were air dried in a SpeedVac Plus SC110A connected to Refrigerated Vapor Trap RVT100 (Thermo Scientific) at low temperature settings and analyzed using LC-MS/MS as described in (12).

Samples for protein quantification were suspended in 800 µl 0.1 M NaOH, transferred into 2 ml screw cap tubes prefilled with 0.1 mm silica beads (Biozym) and heated for 10 min at 98°C. Cell lysis was performed in BeatBlaster with 2 pulses for 30 s at 6 m/s and an interval of 2 min. Lysates were centrifuged at max. speed and 4°C for 15 min. Supernatant was saved and the extraction step was repeated. Supernatants were combined and protein concentration was determined via Bradford using Roti-Quant.

For normalization of c-di-AMP concentration to the protein amount, following formula was used:

$$\frac{c\text{-}di\text{-}AMP\ [nM] \cdot 200}{cV\ [ml] \cdot c590\ [\frac{\mu g}{ml\ cells}]} = \frac{c\text{-}di\text{-}AMP\ [pmol]}{protein\ [mg]}$$

Isothermal Titration Calorimetry

ITC experiments were performed using a Malvern PEAQ-ITC system with 21 µM protein in ITC buffer (20 mM HEPES, pH = 7.5; 100 mM NaCl) in the cell. The respective nucleotides (210 µM) were titrated into the cell by 19 injections of 2 µl, spaced 150 s apart, at 25°C. The data was analyzed using the MicroCal PEAQ-ITC analysis software provided with the instrument. All titrations were repeated to confirm robustness of the assay.

Size-exclusion coupled static light scattering

Determination of the molecular weight of AtaC was performed using a 24 ml Superdex S200 increase size-exclusion coupled to multi-angle laser light scattering and refractive index monitor (WYATT miniDAWN TREOS, WYATT Optilab T-rEX). Data were analyzed using the ASTRA software package provided with the instrument (Wyatt).

Bioinformatic characterization of AtaC and its abundance in prokaryotes

A local PATRIC database was installed and used to determine the conservation of AtaC (PGF_00172869) across prokaryotes (Suppl. Table 2). Duplicated entries with identical genome were removed (177 in total), but keeping the first entry. The same database was also used to determine the conservation of DisA (PGF_00421347), PgpH homologues (PGF_03110657), GdpP-type proteins (PGF_00033444) and DhhP-like proteins (PGF_01833449). An in-house python script was used to extract taxonomic information for each of the AtaC homologues. Specifically, the accession number of each species was used to access NCBI taxonomy and the taxonomic information was integrated with the original PATRIC table. Only phyla with more than 5 genomes were kept, and entries with no taxonomic information were excluded from the analysis. Finally, a total of 5200 entries were used to

generate the AtaC abundance pie chart. Multiple sequence alignment of AtaC proteins (see Suppl. Table S2 for respective PATRIC IDs) from different phyla was performed using CLUSTAL OMEGA (1.2.4) on https://www.ebi.ac.uk/Tools/msa/clustalo/.

SUPPLEMENTARY INFORMATION

Table S1. Strains, plasmids and oligonucleotides used in this study

	Genotype or comments	**Source or reference**
Strains		
S. venezuelae		
NRRL B-65442	Wild type	(NCBI Reference Sequence: NZ_CP018074.1)
disA::apr	ATCC 10712 *SVEN_3211::aac(3)IV*; Apr^R	(6)
SVAL5	Δ*disA::apr*, *attB*$_{\Phi BTI}$::p3xFLAG-*disA*; Apr^R, Hyg^R	This study
SVAL8	*disA::disA*$_{D86A}$	This study
SVAL19	Δ*disA* (SV1-transduction); Apr^R	This study
SVAL20	Δ*vnz_27310::apr* (Redirect); Apr^R	This study
SVAL22	Δ*ataC::apr* (SV1-transduction from SVAL20); Apr^R	This study
SVAL24	Δ*disA::apr*, *attB*$_{\Phi BTI}$::pIJ10170-*disA*; Apr^R, Hyg^R	This study
SVAL26	Δ*ataC::apr*, *attB*$_{\Phi BTI}$::pIJ10170-*vnz_27305prom-ataC*; Apr^R, Hyg^R	This study
SVAL27	Δ*ataC::apr*, *attB*$_{\Phi BTI}$::pIJ10170-*vnz_27305prom-ataC*$_{D269N}$; Apr^R, Hyg^R	This study
E. coli		

W3110	K-12 derivative; *F-, λ- ,rpoS(Am), rph-1, Inv(rrnD-rrnE)*	(13)
W3110 Δ*cya*	W3110 derivative with deleted adenylate cyclase	(14)
ET12567/pUZ8002	*dam*, *dcm*, *hsd*; Kan^R, Cm^R	(15)
BW25113/pIJ790	(Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB-4*), *lacIp-4000*($lacI^Q$), λ-, *rpoS369*(Am), *rph-1*, Δ(*rhaD-rhaB*)568, *hsdR514*; Cm^R	(16)
BL21 (DE3) pLysS	F– *ompT hsdS*(rB– mB–) *gal dcm* λ(DE3), Cm^R	Promega
HME68	W3110 Δ(*argF-lac*)*U169 galKtyr145UAG mutS<>cat*	(17)
Rosetta (DE3) pET28-$disA_{Bsu}$	Overexpression of *Bacillus subtilis* DisA; Kan^R, Cm^R	(18)
Rosetta 2 (DE3)	F- *ompT hsdS*$_B$(r_{B-} m_{B-}) *gal dcm* (DE3) pRARE2 (Cam^R)	Novagen
Plasmids		
pIJ773	Plasmid template for amplification of the *apr-oriT* cassette for 'Redirect' PCR-targeting; Apr^R	(3)
pIJ790	Modified λRED recombination plasmid [*oriR101*] [*repA101*(ts)] *araBp-gam-be-exo*; Cm^R	(3)
pIJ10170	pMS82 derivative; Hyg^R	(19)
pUZ8002	RP4 derivative with defective oriT; Kan^R	(15)
p3xFLAG	pIJ10170 derivative containing *3xFLAG* sequence downstream of MCS; Hyg^R	(20)
pGEX_6P_1	T7 expression vector (modified MCS); Amp^R	GE Healthcare
pET15b	T7 expression vector; Amp^R	Novagen
pKNT25	Low copy vector encoding the T25 fragment of *Bordetella pertussis cyaA* downstream of the MCS; Kan^R	Euromedex

pUT18	High copy vector encoding the T18 fragment of *B. pertussis cyaA* downstream of the MCS; Amp^R	Euromedex
pECAL1	pET15b-*disA*; Amp^R	This study
pECAL4	pET15b-$disA_{D86A}$; Amp^R	This study
pECAL12	pET15b-*ataC*; Amp^R	This study
pECAL13	pET15b-*vnz_31010*; Amp^R	This study
pECAL16	pET15b-$ataC_{Spn}$; Amp^R	This study
pECAL17	pET15b-*vnz_28055 (cpeA)*; Amp^R	This study
pECAL18	pKNT25-*vnz_28050 (cpeB)*; Kan^R	This study
pECAL19	pUT18-*vnz_28055 (cpeA)*; Amp^R	This study
pSVAL6	pIJ10170-*disA*; Hyg^R	This study
pSVAL11	pIJ10170-*vnz_27305prom-ataC* (200 bp upstream of *vnz_27305* start codon fused to *ataC*); Hyg^R	This study
pSVAL12	pIJ10170-*vnz_27305prom-ataC*$_{D269N}$ (pSVAL11 derivative); Hyg^R	This study
pSVNT-10	p3xFLAG-*disA*; Hyg^R	This study
pDD29	pET15b-*ataC*$_{D269N}$ (pECAL12 derivative); Amp^R	This study

Underlined nucleotides indicate restriction sites, nucleotides in bold represent introduced mutations and nucleotides in italics indicate sequences overlapping to other genes

Oligonucleotide	Sequence
Oligonucleotides used for chromosomal *disA* D86A point mutation, PCR verification and sequencing	
disA_D86Achr_rev	TCTTGGTGATGTCCTTGTCGAGGACGAGCGCGCC**CG**CGAGCTTGCAC AGCTCCCGCAGCCGCGTGGCGGC
disA_D86A_check_fwd	GGAGCTGTGCAAGCTCGCG
disA_fwd_NdeI	TATCATATGGTGGCAGCCAAGGAC

disA_rev_XhoI	TATCTCGAGCTAGACGTACCGCTCAAG
Oligonucleotides used for verification of *disA* deletion	
disA_test_f	GTGGTTCACTCACGCCGCATGAACGGTTC
disA_test_r	GGCACGTACCTGGTGGAGGCGAAGGTG
Oligonucleotides used for complementation of *disA* deletion with *disA-3xFLAG*	
3142_NdeI-for	GCTGCATATGGGCCGGCGGGTCG
3142_XhoI-rev	GCAGCCTCGAGGACGTACCGCTCAAGGATC
Oligonucleotides used for complementation of *disA* deletion with wild type *disA*	
3142_NdeI-for	GCTGCATATGGGCCGGCGGGTCG
disA_rev_XhoI	TATCTCGAGCTAGACGTACCGCTCAAG
Oligonucleotides used for *vnz_27310 (ataC)* deletion and PCR verification	
27310_fwd_Apra	CGAAGCGATCGCGGCCACCGCCGCGCCCACCCGCTGATG*ATTCCGGG GATCCGTCGACC*
27310_rev_Apra	GGTCGTGGGGGGGGAGGAGACGGGTGAGGAGTGGGCTCA*TGTAGGC TGGAGCTGCTTC*
27310_test_f	ACACCGTGCGGCAGACCC
27310_test_r	TTCCCGCAGTCCATGGTTCC
Oligonucleotides used for complementation of *ataC* deletion (sequences overlapping to putative promoter in italics)	
27305prom_f_NdeI	TATCATATGGGTGTCCCGGCTCGTCGAC
27305prom_r_OL_ataC	GCTGCACCAT*GGACTCCATCCTACGGGGCT*
ataC_f_OL_27305prom	*GATGGAGTCC*ATGGTGCAGCCGACCGCCGT
5409_rev_XhoI	TATACTCGAGTCAGGTGCGGACTTCGAG
Oligonucleotides used for generation of *pIJ10770-27305prom-ataC*$_{D269N}$	

D269N_backbone_f	CGGCGCTGTACGTCACGGCC**A**ACCACGGCATGGTCGACAT
D269N_backbone_r	ATGTCGACCATGCCGTGGT**T**GGCCGTGACGTACAGCGCCG
Oligonucleotides used for generation of pET15b overexpression constructs	
disA_fwd_NdeI	TATCATATGGTGGCAGCCAAGGAC
disA_rev_XhoI	TATCTCGAGCTAGACGTACCGCTCAAG
disA_D86A_fwd	GCAAGCTCG**CG**GGCGCGCTC
disA_D86A_rev	GAGCGCGCC**CG**CGAGCTTGC
5409_fwd_NdeI	TATCATATGATGGTGCAGCCGACCG
5409_rev_XhoI	TATACTCGAGTCAGGTGCGGACTTCGAG
6143_fwd_NdeI (for *vnz_31010*)	TATCATATGGTGATCGTCATCGCCCATGT
6143_rev_BamHI (for *vnz_31010*)	TATGGATCCTCAGACCGGCACGGTC
28055_fwd_NdeI	TAT CATATG GTGCCTGCTCCACGGATG
28055_rev_XhoI	TATA CTCGAG TCACTCCCGTCCGAGTATGG
Oligonucleotides used for generation of the pET15b-*ataC*$_{D269N}$ construct	
DD60AtaC(D269N)_fwd	GTACGTCACGGCC**A**ACCACGGCATGGTCGA
DD61AtaC(D269N)_rev	GGCCGTGACGTACAGCGCCGAGC
Oligonucleotides used for generation of the constructs for interaction studies	
28050_NT25f_XbaI	TATTCTAGACGTGCATTCCGCTCTGTTCCT
28050_NT25r_KpnI	TATGGTACCCGCGCGCGGCCGGCCGGATCCT
28055_T18f_KpnI	TATGGTACCGCCTGCTCCACGGATGAGC
28055_T18r_EcoRI	TAGCAGAATTCGACTCCCGTCCGAGTATGGAGG

198

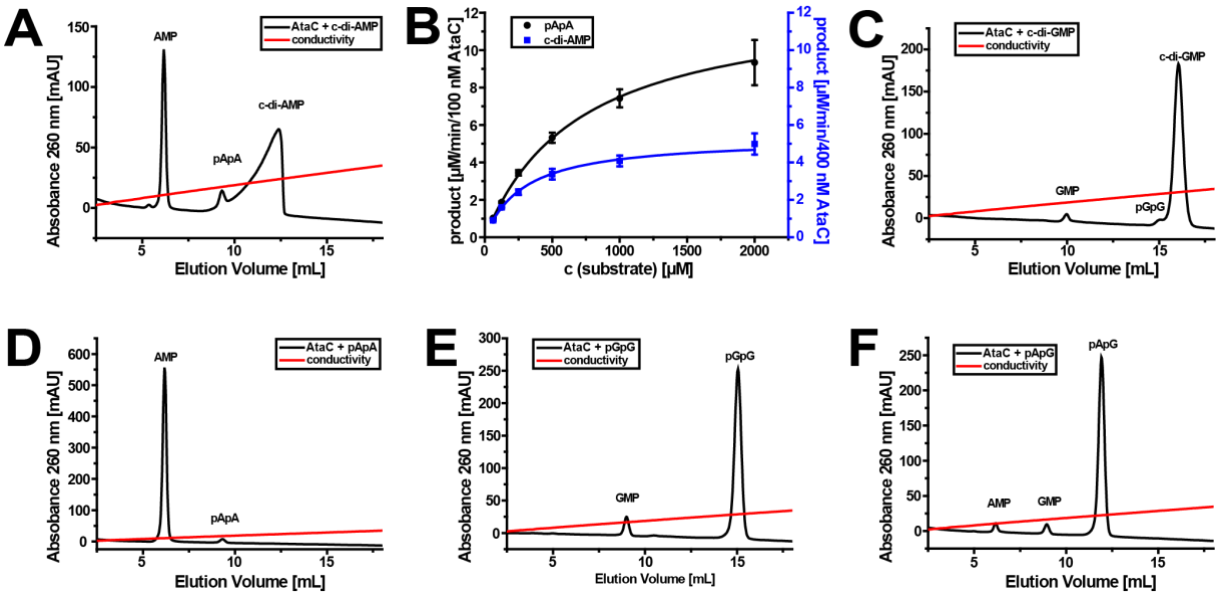


Figure S1. Hydrolysis activity of AtaC is specific for adenosine bases.

(A) Ion-exchange chromatography run on a Resource Q column of the reaction products after 1 h incubation from a 100 µl reaction containing 100 nM AtaC + 250 µM c-di-AMP.

(B) Michaelis-Menten kinetics of the reactions from 400 nM AtaC + c-di-AMP (62.5 – 2000 µM) and 100 nM AtaC + 5′-pApA (62.5 – 2000 µM) after 1 h of incubation at 37°C. c-di-AMP, K_M= 285 ± 32 µM, k_{cat}= 0.2 s^{-1}; 5′-pApA, K_M= 698 ± 32 µM, k_{cat}= 2.1 s^{-1}.

(C) Ion-exchange chromatography run on a Resource Q column of the reaction products after 1 h incubation from a 100 µl reaction containing 100 nM AtaC + 250 µM c-di-GMP.

(D) Ion-exchange chromatography run on a Resource Q column of the reaction products after 1 h incubation from a 100 µl reaction containing 100 nM AtaC + 250 µM c-di-AMP.

(E) Ion-exchange chromatography run on a Resource Q column of the reaction products after 1 h incubation from a 100 µl reaction containing 100 nM AtaC + 250 µM 5′-pGpG.

(F) Ion-exchange chromatography run on a Resource Q column of the reaction products after 1 h incubation from a 100 µl reaction containing 100 nM AtaC + 250 µM 5′-pApG.

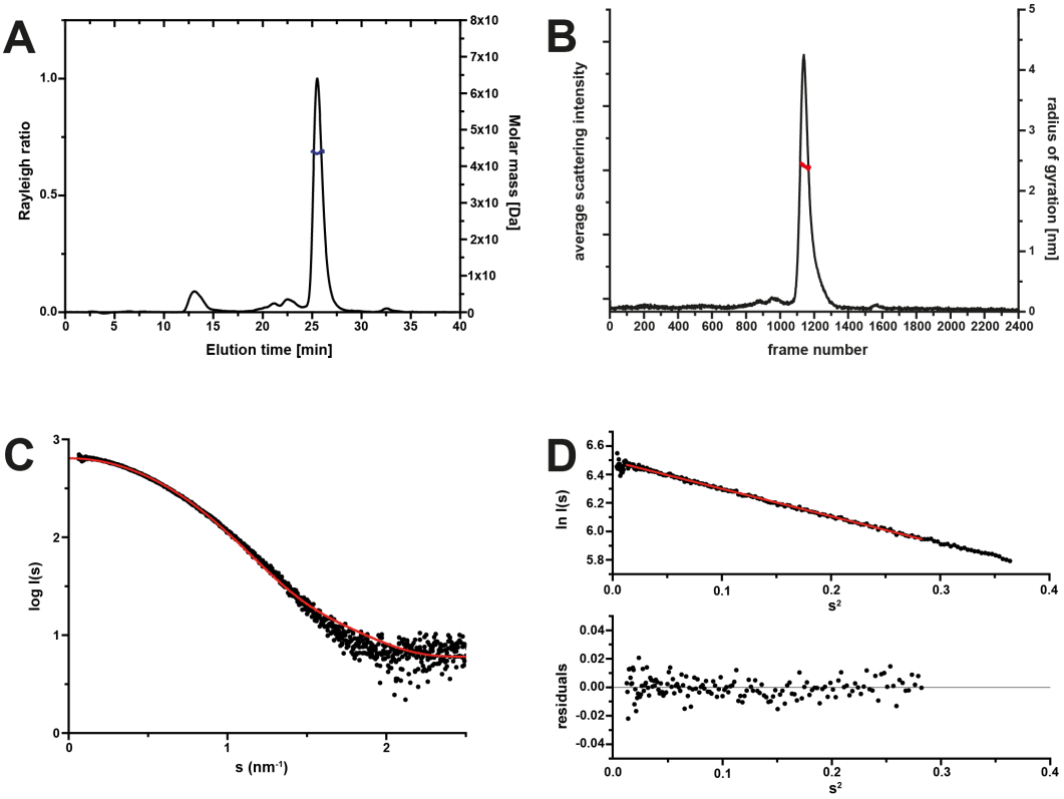


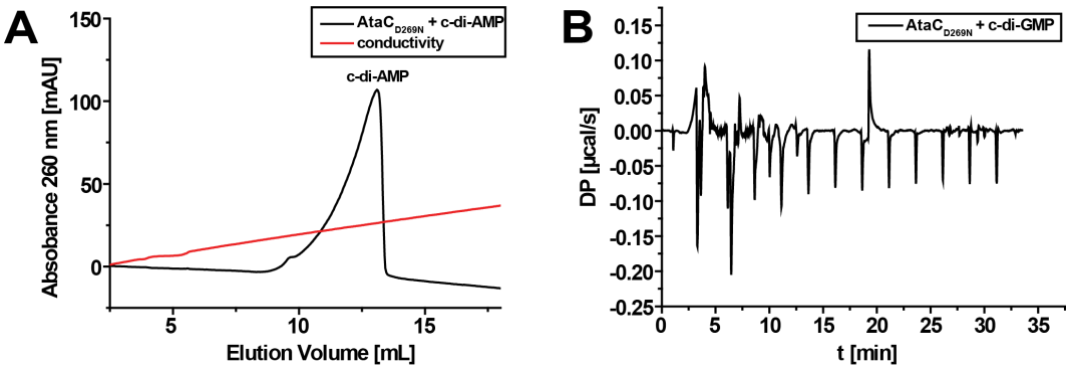
Figure S2. AtaC is a monomer in solution.

(A) Molecular weight determination of AtaC by SEC coupled multi-angle laser light scattering. The obtained molecular weight is 43.7 kDa and stable for the main protein peak at 25 ml.

(B) shows the relative scattering intensity of the sample during a size-exclusion coupled SAXS run at EMBL-P12 using a 24 ml Superdex increase S200 10/300 column (Intensity vs. frame No.). The respective estimated radius of gyration for each frame in the main peak is shown in red (right Y-axis).

(C) Measured SAXS curve of AtaC and a theoretical scattering curve (red) of the model of AtaC using PhnA as template (obtained from HHpred/MODELLER (21), χ_2=3.6).

(D) Guinier plot ln I(s) vs. s^2 (top part) of the averaged buffer corrected scattering data (from B) and the respective residuals of the linear regression (R_G= 2.41 ± 0.1nm). The equally distributed errors of the linear regression (for $s*R_G<1.3$, Guinier approximation) indicates that the sample is not aggregating.



231

232 **Figure S3. AtaC$_{D269N}$ does not cleave c-di-AMP and does not bind c-di-GMP.**

233 (A) Ion-exchange chromatography run on a Resource Q column of the reaction products after

234 1 h incubation from a 100 µl reaction containing 1 µM AtaC$_{D269N}$ + 250 µM c-di-AMP.

235 (B) ITC measurement of 20 µM AtaC titrated with 140 µM c-di-GMP. No binding was

236 detected.

237

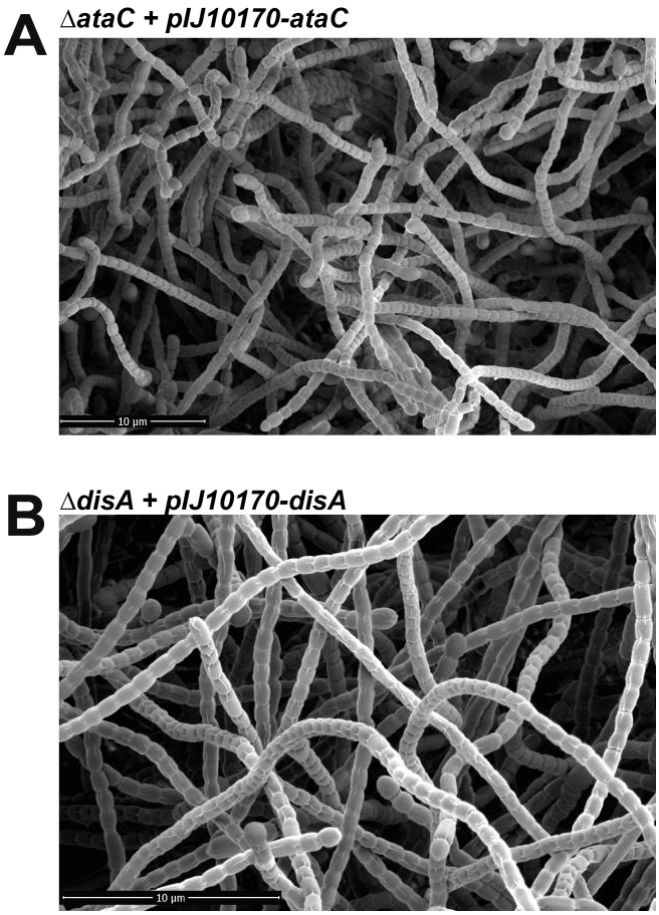


Figure S4. Complementation of the Δ*ataC* and Δ*disA* mutants with the wild type alleles. (A) Scanning electron micrographs show that expression of *ataC* from the *attB*$_{\Phi BT1}$ site under the control of the native promoter from pIJ10170 complemented the delayed developmental phenotype of the Δ*ataC* mutant. (B) Expression of *disA* from pIJ10170 in the Δ*disA* mutant did not alter the wild type phenotype of the mutant. For comparison see also Figure 5B in the main text. Cells were grown on MYM for 4 days at 30 °C.

SUPPLEMENTARY REFERENCES

1. C. Stuttard, Temperate phages of *Streptomyces venezuelae*: lysogeny and host specificity shown by phages SV1 and SV2. *Microbiology* **128**, 115-121 (1982).
2. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood, *Practical Streptomyces Genetics.* (The John Innes Foundation, Norwich, 2000).
3. B. Gust, G. L. Challis, K. Fowler, T. Kieser, K. F. Chater, PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1541-1546 (2003).
4. B. Gust *et al.*, Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Advances in applied microbiology* **54**, 107-128 (2004).
5. N. Tschowri *et al.*, Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* **158**, 1136-1147 (2014).
6. R. J. St-Onge *et al.*, Nucleotide Second Messenger-Mediated Regulation of a Muralytic Enzyme in *Streptomyces*. *Molecular microbiology* **96**, 779-795 (2015).
7. M. A. Schumacher *et al.*, The *Streptomyces* master regulator BldD binds c-di-GMP sequentially to create a functional BldD2-(c-di-GMP)4 complex. *Nucleic acids research* **45**, 6923-6933 (2017).
8. N. Costantino, D. L. Court, Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15748-15753 (2003).
9. J. Germer, G. Becker, M. Metzner, R. Hengge-Aronis, Role of activator site position and a distal UP-element half-site for sigma factor selectivity at a CRP/H-NS-activated sigma(s)-dependent promoter in *Escherichia coli*. *Molecular microbiology* **41**, 705-716 (2001).
10. T. N. Huynh *et al.*, An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E747-756 (2015).
11. M. Christen, B. Christen, M. Folcher, A. Schauerte, U. Jenal, Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *The Journal of biological chemistry* **280**, 30829-30837 (2005).
12. J. Gundlach *et al.*, An essential poison: Synthesis and degradation of cyclic di-AMP in Bacillus subtilis. *Journal of bacteriology* 10.1128/JB.00564-15 (2015).
13. K. Hayashi *et al.*, Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2**, 2006 0007 (2006).
14. S. Herbst *et al.*, Transmembrane redox control and proteolysis of PdeC, a novel type of c-di-GMP phosphodiesterase. *The EMBO journal* **37** (2018).
15. M. S. Paget, L. Chamberlin, A. Atrih, S. J. Foster, M. J. Buttner, Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *Journal of bacteriology* **181**, 204-211 (1999).
16. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640-6645 (2000).
17. L. C. Thomason, J. A. Sawitzke, X. Li, N. Costantino, D. L. Court, Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol* **106**, 1 16 11-39 (2014).

18. G. Witte, S. Hartung, K. Buttner, K. P. Hopfner, Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**, 167-178 (2008).
19. S. Schlimpert *et al.*, Two dynamin-like proteins stabilize FtsZ rings during *Streptomyces* sporulation. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E6176-E6183 (2017).
20. M. M. Al-Bassam, J. Haist, S. A. Neumann, S. Lindenberg, N. Tschowri, Expression Patterns, Genomic Conservation and Input Into Developmental Regulation of the GGDEF/EAL/HD-GYP Domain Proteins in *Streptomyces*. *Front Microbiol* **9**, 2524 (2018).
21. D. I. Svergun, C. Barberato, M. H. J. Koch, CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates *J. Appl. Cryst.* **28**, 768-773 (1995).

3. cGAS senses long and HMGB/TFAM bound U-turn DNA by forming protein-DNA ladders

Andreeva, L.; Hiller, B.; Kostrewa, D.; Lässig, C.; De Oliveira Mann, C. C.; Jan Drexler, D.; Maiser, A.; Gaidt, M.; Leonhardt, H.; Hornung, V.; Hopfner, K. P. cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. *Nature* **2017**, *549* (7672), 394–398.

DOI: http://dx.doi.org/10.1038/nature23890

URL: https://www.nature.com/nature/journal/v549/n7672/full/nature23890.html

In this publication, the mechanism of DNA sensing by the mammalian cyclic GMP-AMP synthase cGAS is characterized in detail. The cGAS-STING pathway is an important component of the innate immune system, thus representing an emerging research topic. Upon DNA detection in the cytosol, cGAS produces 2′3′-cGAMP, which binds the adaptor protein STING to induce a type I interferon response. However, the precise mechanisms of cGAS activation are not fully explored, in particular, regarding the activity of cGAS in correlation with DNA length. This study provides results that indicate a protein-DNA ladder model for cGAS, induced by prearranged U-turned DNA. The crystal structure cGAS in complex with 39 bp DNA reveals a $cGAS_4$:DNA_2 tetramer, with bent DNA. Furthermore, a cooperative cGAS-DNA association was observed in binding studies. These observations lead to the suggestion, that DNA binding by the first cGAS dimer prearranges the DNA to facilitate the association of further cGAS dimers. Of note, proteins that bend the DNA in a similar way were shown to increase cGAS activity significantly, while inhibiting cGAS activity in high concentrations. These proteins include the bacterial nucleoid packaging HU proteins and the eukaryotic TFAM (mitochondrial transcription factor A) and HMGB1 (high-mobility group box 1), which can both occur in the cytosol under mitochondrial stress conditions or DNA transfection. In this study, HMGB1 was shown to co-localize with DNA and cGAS in the cytosol, supporting the physiological relevance of the present model. The DNA bending proteins could allow cGAS to discriminate between short non-pathogenic DNA fragments and damage-derived or foreign DNA.

Author contribution

The author of the present thesis performed luciferase reporter assays. In addition, he analyzed cGAS reaction products and supported isothermal titration calorimetry experiments.

LETTER

doi:10.1038/nature23890

cGAS senses long and HMGB/TFAM–bound U-turn DNA by forming protein–DNA ladders

Liudmila Andreeva[1,2], Björn Hiller[1,2], Dirk Kostrewa[1,2], Charlotte Lässig[1,2], Carina C. de Oliveira Mann[1,2,†], David Jan Drexler[1,2], Andreas Maiser[3], Moritz Gaidt[1,2], Heinrich Leonhardt[3,4], Veit Hornung[1,2,4] & Karl-Peter Hopfner[1,2,4]

Cytosolic DNA arising from intracellular pathogens triggers a powerful innate immune response[1,2]. It is sensed by cyclic GMP–AMP synthase (cGAS), which elicits the production of type I interferons by generating the second messenger 2′3′-cyclic-GMP–AMP (cGAMP)[3–5]. Endogenous nuclear or mitochondrial DNA can also be sensed by cGAS under certain conditions, resulting in sterile inflammation. The cGAS dimer binds two DNA ligands shorter than 20 base pairs side-by-side[6–9], but 20-base-pair DNA fails to activate cGAS *in vivo* and is a poor activator *in vitro*. Here we show that cGAS is activated in a strongly DNA length-dependent manner both *in vitro* and in human cells. We also show that cGAS dimers form ladder-like networks with DNA, leading to cooperative sensing of DNA length: assembly of the pioneering cGAS dimer between two DNA molecules is ineffective; but, once formed, it prearranges the flanking DNA to promote binding of subsequent cGAS dimers. Remarkably, bacterial and mitochondrial nucleoid proteins HU and mitochondrial transcription factor A (TFAM), as well as high-mobility group box 1 protein (HMGB1), can strongly stimulate long DNA sensing by cGAS. U-turns and bends in DNA induced by these proteins pre-structure DNA to nucleate cGAS dimers. Our results suggest a nucleation-cooperativity-based mechanism for sensitive detection of mitochondrial DNA[10] and pathogen genomes[11], and identify HMGB/TFAM proteins as DNA-structuring host factors. They provide an explanation for the peculiar cGAS dimer structure and suggest that cGAS preferentially binds incomplete nucleoid-like structures or bent DNA.

cGAS detects a broad range of intracellular viral and bacterial pathogens[11–16] but also senses mitochondrial (mt)DNA in the cytosol upon mitochondrial stress[10]. cGAS dimers bind two DNA ligands side-by-side, whereby each cGAS monomer directly recognizes fewer than 20 base pairs (bp). In principle, 20 bp DNA should be sufficient to fully activate cGAS; however, it does not activate cGAS in human cells[17]. Longer DNA ligands were even proposed to lead to steric clashes unless cGAS binds near DNA ends[8]. Thus, the current model fails to explain the detection of immunostimulatory DNA. Keeping the overall amount of DNA constant, we stimulated a human monocyte cell line (BLaER1 cells) with DNA of different lengths and observed a striking concentration- and length-dependent activation of cGAS by measuring C-X-C motif chemokine 10 (CXCL10) production as an appropriate surrogate parameter of cGAS activity[18], as well as interferon-β (IFN-β) mRNA expression levels (Fig. 1a and Extended Data Fig. 1a–c). Long herring testis DNA robustly activated cGAS at all tested DNA amounts, while shorter cGAS ligands required increasing amounts of DNA. Consistent with previous studies, DNA of ~45 bp constituted a 'length' threshold, below which no activation was observed.

Since other cellular factors could contribute to the length-sensitive detection, we analysed the intrinsic capability of mouse (m)cGAS and human (h)cGAS to sense DNA length *in vitro*, using a new fluorescent-based assay (Extended Data Fig. 2 and Methods). Both hcGAS and hcGAS^{cd} (catalytic domain) showed a remarkable DNA length-dependent activation, which appears to be independent of the cGAS N-terminal part (Fig. 1b). Like the ~45 bp threshold *in vivo*, robust activation at physiologically relevant cGAS concentrations *in vitro* required DNA >40 bp. Plasmid DNA was the most potent activator. Mouse mcGAS^{cd} exhibited a comparable length-dependent activation, with a gradual increase in activity until about 75 bp (Fig. 1c and Extended Data Fig. 1d). At higher DNA concentrations and short DNA lengths, mcGAS was activated in an almost stepwise fashion, with a first plateau between 30 and 45 bp and a second above 50 bp (Fig. 1c inset), perhaps reflecting stepwise binding along DNA. DNA <20 bp can also activate cGAS but requires 10–20 times higher cGAS and 50–250 times higher DNA amounts to induce activities similar to 50 bp DNA (Extended Data Fig. 1e, f). Finally, circular and linearized plasmid DNA activated cGAS equally well, ruling out the possibility that cGAS needs DNA ends (Extended Data Fig. 1g).

To determine a mechanism for the DNA length sensing, we crystallized mcGAS^{cd} in complex with 39 bp DNA. The crystals diffracted anisotropically to 3.6–4.8 Å resolution and we determined the structure by molecular replacement. In the crystal, two cGAS dimers and two DNA 39-mers form a $cGAS_4$–DNA_2 complex (Fig. 1d and Extended Data Fig. 3a). Each of the two cGAS dimers in our $cGAS_4$–DNA_2 complex is similar to the dimers seen in $cGAS_2$–DNA_2 complexes determined with <20 bp DNA (Extended Data Fig. 3b, c), indicating that long DNA does not induce a substantially different structural state in cGAS than short DNA. The two cGAS dimers are arranged in a 'head-to-head' orientation along the DNA. DNA between the cGAS dimers is slightly curved (Fig. 1d), avoiding the proposed clashes and showing that cGAS does not need DNA ends (Extended Data Fig. 3c, d).

The asymmetric unit contains one full $cGAS_4$–DNA_2 complex along with one half $cGAS_4$–DNA_2 complex situated on a twofold crystallographic symmetry axis (Extended Data Fig. 4a). Remarkably, these $cGAS_4$–DNA_2 complexes are further stacked into fibrils with alternating 'head-to-head'- and 'tail-to-tail'-oriented cGAS dimers (Fig. 1e), forming a DNA–protein ladder with rungs (cGAS dimers) and two ladder sides (DNA). The DNA is continuous between 'head-to-head'-oriented cGAS dimers and quasi-continuous (stacked 3′ to 3′ and 5′ to 5′) between the 'tail-to-tail'-oriented cGAS dimers. A ~5 bp (half helical turn) increased spacing of 'tail-to-tail'-oriented cGAS dimers, however, would allow an energetically favourable continuous assembly of alternatingly oriented cGAS dimers between two long DNA elements (Extended Data Fig. 3e, f).

Neighbouring cGAS dimers along the DNA barely interact in either 'head-to-head' or 'tail-to-tail' orientations in the crystal lattice (Extended Data Fig. 4a). Consistently, mutational analysis *in vitro* and *in vivo* showed that whereas protein–DNA contacts are critical, protein–protein contacts between adjacent cGAS dimers in either orientation

[1]Department of Biochemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. [2]Gene Center, Ludwig-Maximilians-Universität München, 81377 Munich, Germany. [3]Department of Biology, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany. [4]Center for Integrated Protein Science Munich, 81377 Munich, Germany. †Present address: Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115-5418, USA.

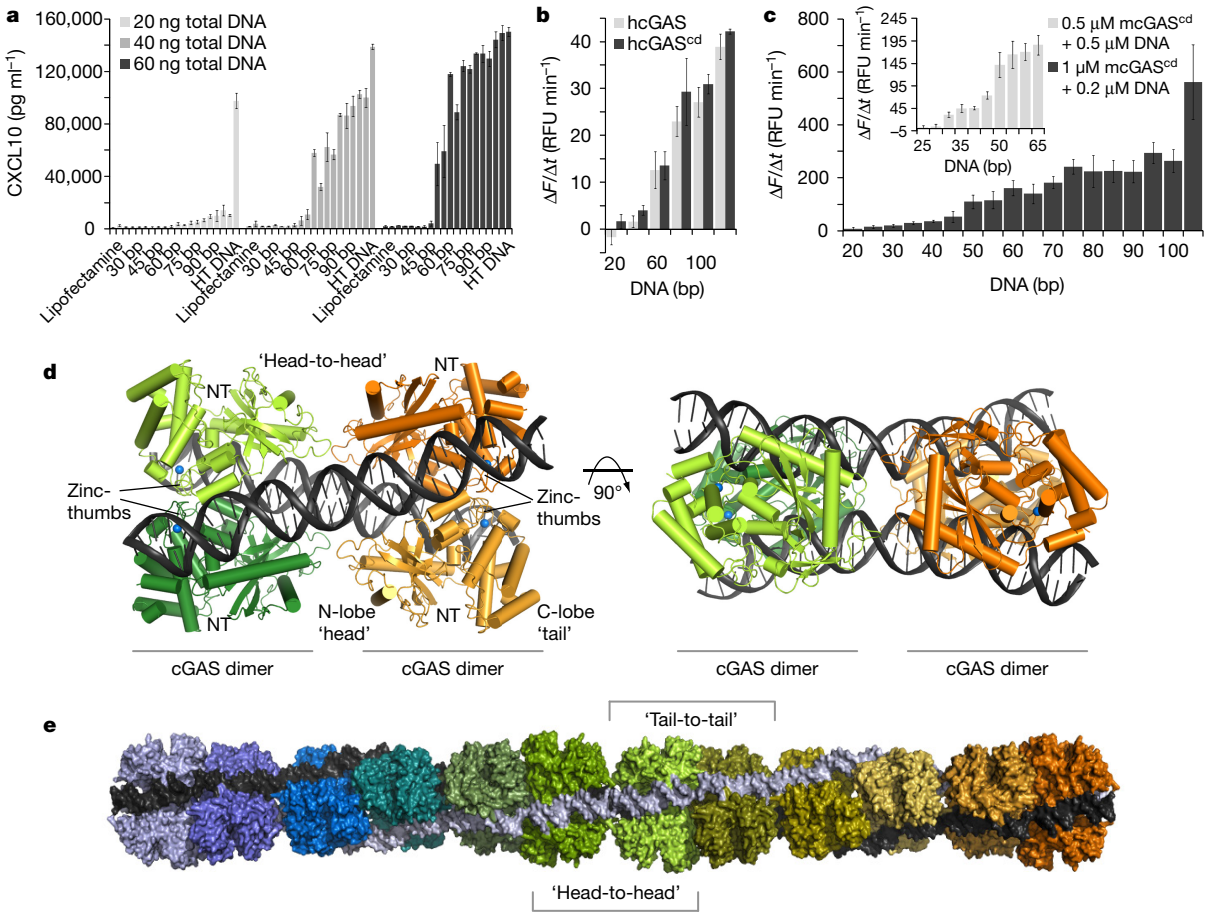


Figure 1 | cGAS shows DNA length-dependent activity and forms DNA–protein ladders. a, cGAS activation in transdifferentiated BLaER1 cells. Cells were stimulated with 20, 40, and 60 ng DNA of different length (20–100 bp in 5 bp intervals) and herring testis (HT) DNA. First two bars each represent unstimulated cells and Lipofectamine controls. Shown are mean values ± s.d., $n = 3$. **b**, *In vitro* activity of hcGAS, hcGAS^{cd}, and **c**, mcGAS^{cd} in the presence of DNA of increasing length (20–100 bp in 5 bp intervals) or plasmid DNA (last bar). Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$, relative fluorescence units (RFU) min^{-1}), measured by the rate of fluorescent ATP analogue (fATP) incorporation into fluorescent cGAS product (fGAMP, see Extended Data Fig. 2) are plotted against DNA length ± s.d., $n = 3$–5. **d**, Crystal structure of cGAS in complex with 39 bp DNA. Two cGAS dimers (green and orange) assemble on DNA in 'head-to-head' orientation. Zinc-thumb dimerization elements and the nucleotidyl-transferase (NT) sites are indicated. **e**, Ladder-like cGAS assembly along a quasi-continuous DNA in the crystal lattice through alternating 'head-to-head' or 'tail-to-tail' orientations of cGAS dimers.

are not (Extended Data Fig. 4b–d)[8,9]. Accordingly, the DNA–protein ladder does not require direct interactions between neighbouring cGAS dimers to be stable, because cGAS–DNA contacts and cGAS–cGAS dimer contacts between the DNA already generate a mesh-like structure.

To see whether ladder-like structures exist in solution, we used isothermal titration calorimetry (ITC) and size-exclusion chromatography coupled to right-angle light scattering (SEC–RALS). ITC showed a DNA length-dependent increase in affinity and at the same time an increase in cGAS:DNA molar ratio (Fig. 2a and Extended Data Fig. 5). DNA of ~45 bp conferred a threshold, resulting in higher affinity and at the same time binding of two cGAS molecules per DNA ligand. The affinity increase was driven by a decrease in binding enthalpy that also compensated a decreasing entropy (Fig. 2b and Extended Data Fig. 6a, b). The latter suggests the formation of ordered structures on DNA >45 bp. Together with ITC, SEC–RALS showed that cGAS forms single homogeneous complexes with DNA of different lengths: a $cGAS_1$–DNA_1 complex for 20-mer DNA, a $cGAS_4$–DNA_2 complex for 50-mer DNA, and $cGAS_6$–DNA_2 complexes for 70/80-mer DNA (Fig. 2c–f and Extended Data Fig. 6c). Of note, $cGAS_2$–DNA_2 complexes were unstable in the presence of 20 bp DNA and only the $cGAS_1$–DNA_1 complex robustly formed with 20 bp DNA that poorly activates cGAS (Fig. 1a–c). Given that activation of cGAS requires additional dimerization[8], this explains why 20 bp DNA is not a good cGAS activator.

The DNA–protein ladders offer an intriguing mechanism for the DNA length-dependent activation of cGAS. Formation of a pioneering $cGAS_2$–DNA_2 complex is highly unfavourable, but once formed, it pre-arranges the two DNA molecules in its vicinity to promote formation of adjacent cGAS dimers with increased affinity (Fig. 3a). Multiple cGAS dimers in $cGAS_{2n}$–DNA_2 ($n \geq 2$) ladders stabilize each other by cooperatively holding together the two DNA ladder sides. An analytical equation for this 'DNA–protein ladder' model can be derived (see Supplementary Methods) and comprises a V_{max} (maximal reaction velocity); three DNA constants K_1, K_2^2, and K_3 that describe interaction of monomeric cGAS with DNA (K_1), interaction of cGAS dimers with two DNA ligands (K_2^2), and the association/dissociation of cGAS–DNA and $cGAS_2$–DNA_2 states (K_3); and a parameter s that is a measure of cooperative binding sites of cGAS dimers along DNA (equivalent to a Hill coefficient).

To test this model, we measured cGAS activity by varying both mcGAS^{cd} concentration and DNA length. We observed a dramatic length-dependent activation that was not due to an increase in V_{max}, consistent with our model, but was due to a DNA length-dependent increase in binding affinity, as seen with ITC (Fig. 3b). The data can be globally fitted with a simplified equation derived from the mathematical

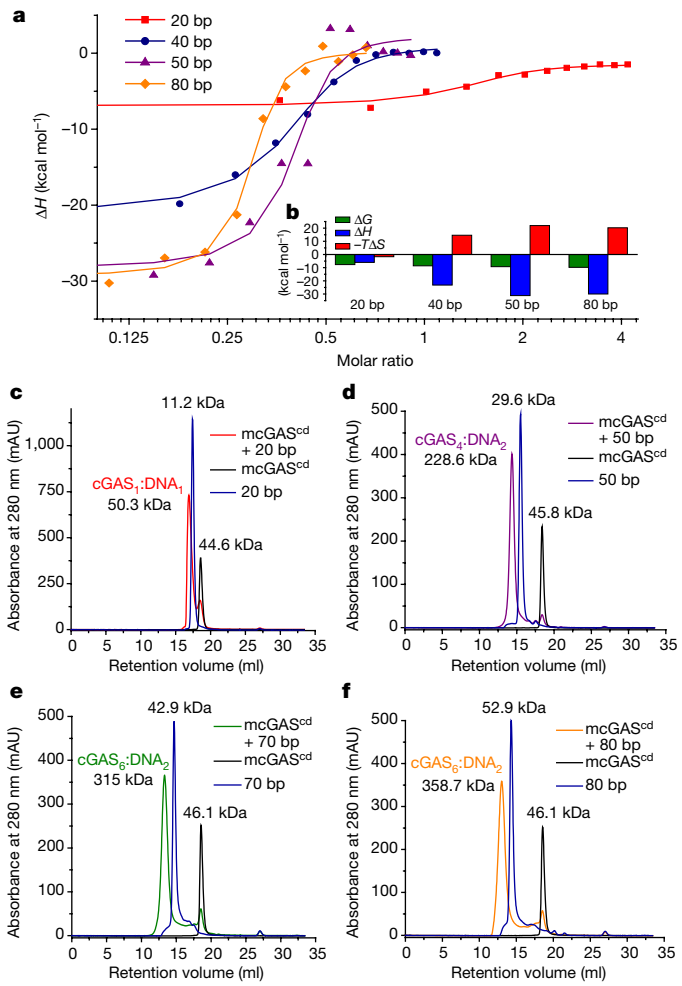


Figure 2 | cGAS and DNA assemble into $cGAS_{2n}$–DNA_2 complexes in solution. **a**, ITC of $mcGAS^{cd}$ and DNA. Enthalpy (ΔH) values are plotted against the molar ratio of DNA:$mcGAS^{cd}$. Obtained dissociation constant (K_d) values for 20, 40, 50, and 80 bp DNA are $2.4 \times 10^{-6} \pm 1.89 \times 10^{-6}$, $507 \times 10^{-9} \pm 128 \times 10^{-9}$, $165 \times 10^{-9} \pm 166 \times 10^{-9}$, and $73.3 \times 10^{-9} \pm 25.5 \times 10^{-9}$ M, respectively. Molar ratios are 1.28 ± 0.125, 0.364 ± 0.01, 0.371 ± 0.027, and 0.274 ± 0.006, respectively. **b**, Thermodynamic parameters obtained with ITC. **c**–**f**, SEC–RALS analysis of $mcGAS^{cd}$ with 20 bp (**c**), 50 bp (**d**), 70 bp (**e**), and 80 bp (**f**) DNA; mAU, milli absorption units. Estimated molecular masses are indicated. Molecular masses 12.2, 30.8, 43.1, and 49.3 kDa were used as reference for 20, 50, 70, and 80 bp DNA, respectively; $mcGAS^{cd}$ was estimated with 43 kDa (see also Extended Data Fig. 6c). Some molecular mass deviation from reference values occurred because of limitations of measurement accuracy.

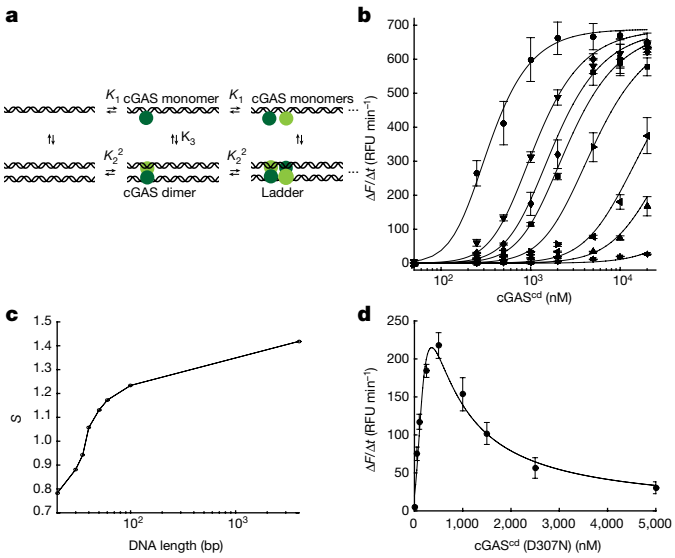


Figure 3 | DNA–protein ladder model. **a**, Kinetic scheme for the DNA–protein ladder model. cGAS binds as monomer to DNA (dissociation constant K_1). These $cGAS_1$–DNA_1 complexes can assemble into $cGAS_2$–DNA_2 structures with parallelized DNA (K_3). Finally, two cGAS molecules can directly bind as dimer to the parallelized DNA ligands (K_2^2). Overall, this scheme describes a cooperative, DNA length-dependent interaction of cGAS with DNA (see Supplementary Methods). **b**, Activity of $mcGAS^{cd}$ on different DNA ligands as a function of protein concentration (symbols: circle, plasmid; downward arrowhead, 100 bp; diamond, 60 bp; square, 50 bp; rightward arrowhead, 40 bp; leftward arrowhead, 35 bp; upward arrowhead, 30 bp; star, 20 bp). The overall concentration of base pairs in each reaction is kept constant. Mean values of cGAS activity (see Fig. 1c legend) are plotted against $mcGAS^{cd}$ concentrations ± s.d., $n = 3$. Superposed is a global least square minimization of the data with an equation describing the DNA–protein ladder model ($R^2 = 0.988$) (see Supplementary Methods). **c**, Plot of the cooperativity parameter s obtained from **b** as a function of DNA length. **d**, Titration of catalytically inactive $mcGAS^{cd}$ (D307N) into a sub-active solution of $mcGAS^{cd}$. Mean values of cGAS activity (see Fig. 1c legend) are plotted against $mcGAS^{cd}$ (D307N) concentrations ± s.d., $n = 3$. Superposed is a least square minimization of the data with the DNA–protein ladder model ($R^2 = 0.96$) (see Supplementary Methods).

model (Supplementary Methods equation 3.6 and sections 4.1 and 4.2) and can be explained by a model in which V_{max} and the DNA binding constants do not depend on the length of DNA, while s increases as a function of DNA length (Fig. 3c). Parameter s measures cooperativity between adjacent cGAS dimers owing to the DNA arranging activity[19]. The transitions of s around 40–50 bp and $2s \sim 2.8$ for plasmid DNA suggest that length sensing can be explained by cooperative binding of two or more adjacent cGAS dimers. Furthermore, titrating catalytically inactive $mcGAS^{cd}$(D307N) to low amounts of active $mcGAS^{cd}$ markedly increased cGAS activity about 200-fold, providing direct evidence for cooperative sequestering of cGAS into oligomeric structures on DNA (Fig. 3d). Higher amounts of the inactive mutant gradually competed $mcGAS^{cd}$ away from DNA, as expected. These data can be fitted well ($R^2 = 0.96$) with the 'DNA–protein ladder' equation (Supplementary Methods) and result in a cooperativity parameter $2s = 3.6$, similar to that described in Fig. 3c. In summary, cGAS length sensing emerges from cooperative formation of $cGAS_{2n}$–DNA_2 ($n \geq 2$) complexes.

Cellular factors could enhance detection of DNA by cGAS if they suitably structure the DNA. We noticed that the HMGB proteins HMGB1/2 and mitochondrial nucleoid organizing protein TFAM could be well suited to nucleate cGAS dimers because they properly prearrange DNA by forming U-turns[20,21] (Extended Data Fig. 7a, b) and because mtDNA activates cGAS[10] and HMGB1/2 facilitate cytosolic nucleic-acid sensing[22]. Bacterial HU proteins, for example from *Listeria monocytogenes* (lHU), also bend DNA to form bacterial nucleoids. Indeed, adding increasing amounts of mTFAM, mHMGB1, and lHU robustly activated $mcGAS^{cd}$ *in vitro* up to ~25-fold (Fig. 4a–c and Extended Data Fig. 7c). A similar activation was also seen for hTFAM and full-length hcGAS (Fig. 4d). cGAS activation required the DNA-bending ability of HMGB1, since mutations that decreased DNA-bending capacity without significantly affecting DNA binding[23] reduced or nearly abolished cGAS activation (Extended Data Fig. 8a–d). TFAM activates cGAS both on circular and on linear plasmid DNA (Extended Data Fig. 8e), works without direct TFAM–cGAS interactions (Extended Data Fig. 8f), and even does not require proteins from the same species (Extended Data Fig. 8g, h). Thus, these proteins presumably activate cGAS not by protein–protein interactions but rather by prestructuring DNA to nucleate or stabilize cGAS dimers. At higher concentrations, TFAM and lHU sharply abolished cGAS activity, presumably because of cooperative formation of nucleoid-like

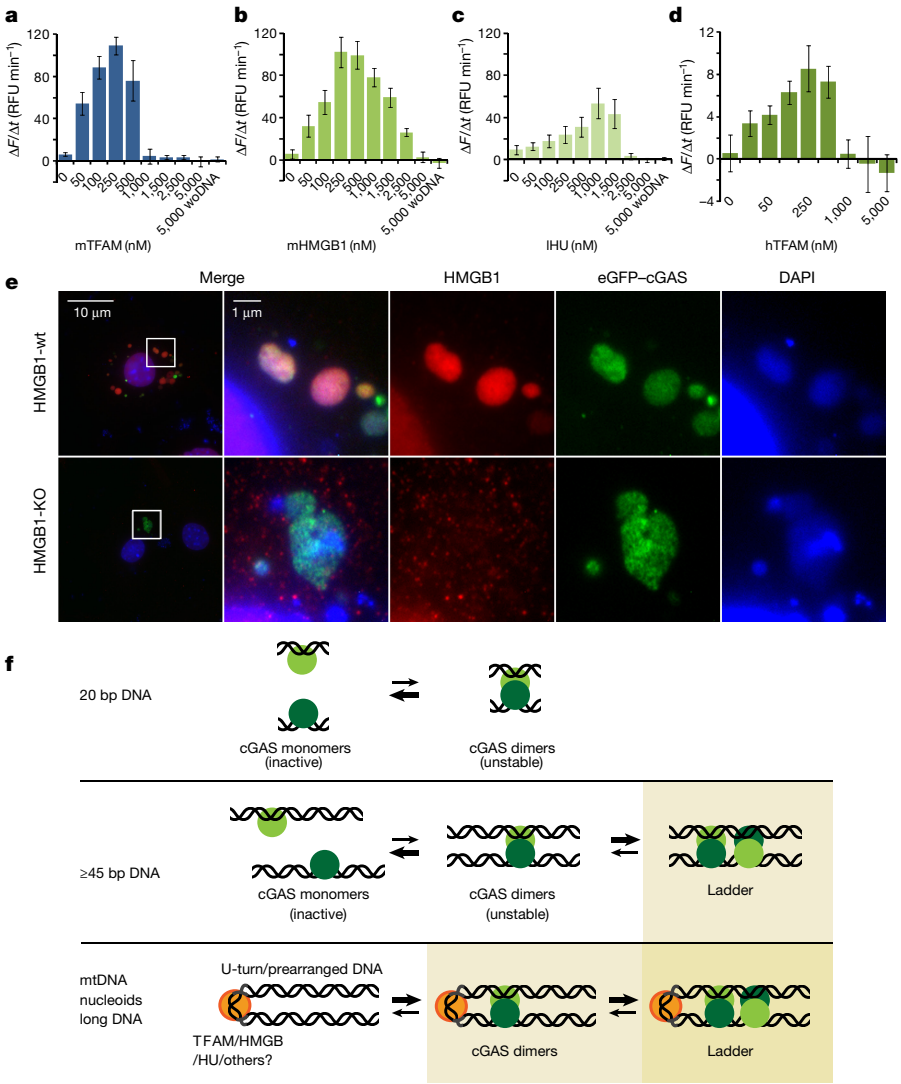


Figure 4 | DNA-bending proteins enhance cGAS activity by prearranging DNA into U-shaped structures. a–d, cGAS activities (see Fig. 1c legend) in the presence of different DNA-bending proteins. The last value corresponds to a control without DNA (woDNA). Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) are plotted against increasing concentrations of DNA-bending proteins ± s.d., $n = 3$–6. mTFAM (**a**), mHMGB1 (**b**) and lHU (**c**) robustly enhance mcGAS^{cd} activity in a dose-dependent manner until they eventually compete cGAS away. **d**, hTFAM activates hcGAS. **e**, HMGB1 (red), cGAS (green), and cytosolic DNA (blue) co-localization. Wild-type (WT) or HMGB1-knockout (KO) MEFs were transfected with an enhanced green fluorescent protein (eGFP)–cGAS-expressing vector and stained for wide-field fluorescence microscopy (pDV) (compare with Extended Data Fig. 10). DNA was stained with 4,6-diamidino-2-phenylindole (DAPI, blue). Fluorescent signal within enlarged images is enhanced until saturation of the nuclear signal to better visualize cytoplasmic structures. **f**, Model for DNA length- and structure-dependent cGAS activation. Activation of cGAS requires the formation of $cGAS_2$–DNA_2 dimers, which are very unstable. Multiple cGAS dimers along DNA ≥ 45 bp stabilize each other by cooperatively holding DNA together, leading to stable, active ladders (shaded panels). DNA-bending/U-turn-inducing proteins prearrange DNA, nucleating the formation of cGAS dimers and ladders (darker shaded panel).

structures[24]. HMGB1 inhibited cGAS more gradually, probably because it does not cooperatively form compact nucleoids. Strong stimulation of cGAS activity by TFAM required DNA >100 bp (Extended Data Fig. 7d), suggesting that robust activation still involves $cGAS_{2n}$–DNA_2 complexes ($n \geq 2$).

TFAM normally resides in mitochondria; however, under mitochondrial stress conditions, induced by a combination of caspase and Bcl-2 inhibitors[25,26], we saw increased TFAM presence in the cytosol (Extended Data Fig. 9a–d) where it can assist detection of leaked mtDNA. Consistently, cytosolic hcGAS and hTFAM co-immunoprecipitated (Extended Data Fig. 9e). Importantly, we found that upon DNA transfection of mouse embryonic fibroblasts (MEFs), endogenous HMGB1 co-localized with cytosolic DNA and cGAS in 85% of the observed cytosolic DNA loci (Fig. 4e and Extended Data Fig. 10). This is consistent with previous findings that HMGB1 is involved in cytosolic DNA sensing[22] even though it normally resides in the nucleus. However, HMGB1 shuttles between nucleus and cytoplasm[27], enabling it to encounter cytosolic DNA where it could assist cGAS.

In summary, we provide a molecular mechanism for the sensitive detection of long DNA by cGAS (Fig. 4f). $cGAS_1$–DNA_1 complexes appear to be inactive and $cGAS_2$–DNA_2 complexes unstable, requiring DNA of sufficient length to form stable $cGAS_{2n}$–DNA_2 ladders ($n \geq 2$). Nucleic-acid-stress HMGB proteins and nucleoid-structuring proteins (TFAM, HU) can additionally nucleate and stabilize cGAS ladders by prearranging DNA. Thus, cGAS preferentially senses structured DNA ligands and DNA with residual nucleoid proteins bound, rather than naked DNA. In terms of recognizing danger or pathogen-associated molecular patterns in the form of mtDNA and bacterial nucleoids, these findings make biological sense and provide a plausible rationale for the evolution of the peculiar cGAS dimer structure. TFAM/HMGB1/HU-enhanced stimulation of cGAS could be especially helpful for the initial detection of long cytosolic DNA with low amounts of cGAS

present. In this context, it should be noted that short ~20-mer DNA can also strongly activate cGAS if it additionally contains G-rich ssDNA Y overhangs[17]. In principle, this flanking G-rich DNA could also stabilize cGAS dimers, but the precise mechanism, as well as that of other postulated host proteins[28,29], remains to be uncovered. In any case, the nucleation- and cooperativity-based mechanism imposes a threshold-like response that conceptually unifies DNA sensing by the cGAS–STING axis with other oligomerization-based nucleic-acid-sensing pathways[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 February; accepted 31 July 2017.
Published online 13 September 2017.

1. Rodero, M. P. & Crow, Y. J. Type I interferon-mediated monogenic autoinflammation: the type I interferonopathies, a conceptual overview. *J. Exp. Med.* **213,** 2527–2538 (2016).
2. Barber, G. N. STING: infection, inflammation and cancer. *Nat. Rev. Immunol.* **15,** 760–770 (2015).
3. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
4. Zhang, X. *et al.* Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51,** 226–235 (2013).
5. Chen, Q., Sun, L. & Chen, Z. J. Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. *Nat. Immunol.* **17,** 1142–1149 (2016).
6. Gao, P. *et al.* Cyclic [G(2′,5′)pA(3′,5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).
7. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498,** 332–337 (2013).
8. Li, X. *et al.* Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39,** 1019–1031 (2013).
9. Zhang, X. *et al.* The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Reports* **6,** 421–430 (2014).
10. West, A. P. *et al.* Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520,** 553–557 (2015).
11. Hansen, K. *et al. Listeria monocytogenes* induces IFNβ expression through an IFI16-, cGAS- and STING-dependent pathway. *EMBO J.* **33,** 1654–1666 (2014).
12. Gao, D. *et al.* Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341,** 903–906 (2013).
13. Reinert, L. S. *et al.* Sensing of HSV-1 by the cGAS–STING pathway in microglia orchestrates antiviral defence in the CNS. *Nat. Commun.* **7,** 13348 (2016).
14. Collins, A. C. *et al.* Cyclic GMP-AMP synthase is an innate immune DNA sensor for *Mycobacterium tuberculosis*. *Cell Host Microbe* **17,** 820–828 (2015).
15. Wassermann, R. *et al. Mycobacterium tuberculosis* differentially activates cGAS- and inflammasome-dependent intracellular immune responses through ESX-1. *Cell Host Microbe* **17,** 799–810 (2015).
16. Watson, R. O. *et al.* The cytosolic sensor cGAS detects *Mycobacterium tuberculosis* DNA to induce type I interferons and activate autophagy. *Cell Host Microbe* **17,** 811–819 (2015).
17. Herzner, A. M. *et al.* Sequence-specific activation of the DNA sensor cGAS by Y-form DNA structures as found in primary HIV-1 cDNA. *Nat. Immunol.* **16,** 1025–1033 (2015).
18. Holm, C. K. *et al.* Virus-cell fusion as a trigger of innate immunity dependent on the adaptor STING. *Nat. Immunol.* **13,** 737–743 (2012).
19. Weiss, J. N. The Hill equation revisited: uses and misuses. *FASEB J.* **11,** 835–841 (1997).
20. Ngo, H. B., Kaiser, J. T. & Chan, D. C. The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat. Struct. Mol. Biol.* **18,** 1290–1296 (2011).
21. Rubio-Cosials, A. *et al.* Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.* **18,** 1281–1289 (2011).
22. Yanai, H. *et al.* HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses. *Nature* **462,** 99–103 (2009).
23. Allain, F. H. T. *et al.* Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. *EMBO J.* **18,** 2563–2579 (1999).
24. Ngo, H. B., Lovely, G. A., Phillips, R. & Chan, D. C. Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation. *Nat. Commun.* **5,** 3077 (2014).
25. Rongvaux, A. *et al.* Apoptotic caspases prevent the induction of type I interferons by mitochondrial DNA. *Cell* **159,** 1563–1577 (2014).
26. White, M. J. *et al.* Apoptotic caspases suppress mtDNA-induced STING-mediated type I IFN production. *Cell* **159,** 1549–1562 (2014).
27. Bonaldi, T. *et al.* Monocytic cells hyperacetylate chromatin protein HMGB1 to redirect it towards secretion. *EMBO J.* **22,** 5551–5560 (2003).
28. Yoh, S. M. *et al.* PQBP1 is a proximal sensor of the cGAS-dependent innate response to HIV-1. *Cell* **161,** 1293–1305 (2015).
29. Jønsson, K. L. *et al.* IFI16 is required for DNA sensing in human macrophages by promoting production and function of cGAMP. *Nat. Commun.* **8,** 14391 (2017).
30. Sohn, J. & Hur, S. Filament assemblies in foreign nucleic acid sensors. *Curr. Opin. Struct. Biol.* **37,** 134–144 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Kremmer for the generation of antibodies, A. Butryn for help with structure determination, K. Lammens and G. Witte for help with crystallization, S. Somarokov for help with protein co-localization studies, S. Bauernfried for help with cell studies, C. Isakaj and O. Fettscher for technical assistance, H. Harz and F. Schüder for advice on staining, F. Civril for cGAS constructs, T. Graf for BLaER1 cells, T. Cremer for fibroblasts, and T. Fujita for the p-125luc reporter plasmid. We thank the Swiss Light Source (Villigen), the European Synchrotron Radiation Facility (Grenoble), and the DESY Petra III (Hamburg) for technical assistance. This work was funded by German Research Foundation grant HO2489/8-1 to K.-P.H., and the Center for Integrated Protein Sciences to K.-P.H., H.L., and V.H. L.A. acknowledges the International Max Planck Research School for Molecular Life Sciences. C.L. and K.-P.H. acknowledge support from BioSysNet (Bavarian Ministry of Education). D.J.D. and C.C.O.M. acknowledge German Research Foundation RTG1721.

Author Contributions L.A. performed crystallographic and biochemical studies. B.H. performed enzyme-linked immunosorbent assay (ELISA) assays and IFN-β mRNA expression analysis. D.K. built and refined the structure. C.L. performed co-immunopurification studies. C.C.O.M. established staining protocols for three-dimensional structured illumination microscopy (3D SIM). D.J.D. performed luciferase reporter assays and analysed cGAS products. A.M. performed microscopy. M.G. generated cGAS-deficient BLaER1 cells. H.L., C.C.O.M., C.L., and L.A. designed and interpreted microscopy experiments. V.H., B.H., and C.L. designed and interpreted cell-based experiments. K.-P.H. designed the study, derived the mathematical model, and analysed data. K.-P.H. and L.A. wrote the paper with contributions from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to K.-P.H. (hopfner@genzentrum.lmu.de).

Reviewer Information *Nature* thanks O. Nureki and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Constructs and cloning. The plasmids encoding full-length (amino acids (aa) 1–522) and truncated (catalytic domain, 'cd') *Homo sapiens* (h) (aa 155–522) and *Mus musculus* (m) (aa 141–507) cGAS for N-terminal His_6–MBP (maltose-binding protein) fusion protein expression were described before[7].

The sequence encoding mTFAM without a mitochondrial localization signal (aa 43–243) optimized for *Escherichia coli* expression was synthesized by Eurofins Genomics and cloned for N-terminal His_6 fusion protein expression into modified pET28a vector (Novagen), where the thrombin cleavage site was exchanged with a tobacco etch virus protease cleavage site.

hTFAM (aa 43–246) sequence was obtained from total cDNA and inserted into pET28a for His_6 fusion protein expression in *E. coli*. Full-length mHMGB1 (aa 1–215), as well as *L. monocytogenes* HU (lHU) (aa 1–121) sequences were purchased from Eurofins Genomics and cloned into pET28a for His_6 fusion protein expression. Truncated mHMGB1 (aa 1–185) (mHMGB1dCTT) was cloned analogously. cGAS and mHMGB1dCTT point mutants were generated by QuikChange site-directed mutagenesis with Pfu-Ultra polymerase (Agilent) followed by DpnI (Fermentas) digestion. For expression in human cells, Flag/haemagglutinin (HA)-tagged hcGAS (aa 1–522) and HA-tagged hTFAM (aa 43–246) were cloned into pcDNA5/FRT/TO (Invitrogen, Thermo Fisher Scientific). For localization studies, hcGAS was cloned into pEGFP-C1vector (Clontech) for N-terminal eGFP-tagged cGAS expression. All protein constructs abbreviations and descriptions are listed in Supplementary Table 1.

Cell lines and reagents. All DNA oligonucleotides were purchased from Metabion. The exact sequences of stimulatory DNAs are listed in Supplementary Table 2. Linearized plasmid DNA was obtained by digestion of pET28M–SUMO1–GFP vector (EMBL) with BamHI (NEB). A 200 bp PCR fragment was amplified from the MBP sequence. Anti-hcGAS (catalytic domain) and a control antibody were produced by The Service Unit Monoclonal Antibodies (German Research Center for Environmental Health, Helmholtz Zentrum München). The following antibodies were purchased commercially: anti-HA–HRP (Cell Signaling, clone 6E2), mouse anti-hTFAM (Abnova), rabbit anti-hTOM20 (FL-145) (Santa Cruz Biotechnology), mouse anti-HMGB1 (Sigma, clone 2F6), goat anti-mouse Alexa Fluor 594 (Life Technologies), goat anti-rabbit Alexa Fluor 488 (Life Technologies), donkey anti-rabbit Alexa Fluor 594 (Life Technologies), and donkey anti-mouse Alexa594 and Alexa Fluor 488 (Life Technologies). The following cell lines were used: HEK293T (American Type Culture Collection, CRL-11268), HEK293T STING-KI[31], primary human fibroblasts (provided by M. Cremer, T. Cremer's group (Ludwig-Maximilians-University Munich, Biocentre Martinsried)), BLaER1[32] (provided by T. Graf's group (Center for Genomic Regulation, Universidad Pompeu Fabra and Institució Catalana de Recerca i Estudis Avançats, Barcelona)), and BLaER1 cGAS-KO, HMGB1-KO and HMGB1-WT MEF (HMGBiotech, HM-221).

For generation of cGAS-deficient BLaER1 cells (BLaER1 cGAS-KO), a single-guide (sg)RNA targeting the sequence GAACTTTCCCGCCTTAGGCA**GGG** (protospacer adjacent motif is in bold type) of the human *MB21D1* gene was cloned via ligation-independent cloning into pR-U6-gRNA to yield pR-U6-MB21D1 as previously described[33]. BlaER1 cells were electroporated with pR-U6-MB21D1 and pCMV-mCherry-T2A-Cas9 (ref. 31) expression plasmids using a GenePulser device (Biorad), and 2 days later FACS-sorted mCherry-positive cells were subcloned by limiting dilution. Monoclonal cell lines were rearranged and duplicated for genotyping. The genomic locus surrounding the sgRNA binding site was PCR amplified (cGAS forward primer: ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCTTTTTGGCGCGGGCCCCAGTTG; cGAS reverse primer: TGACTGGA GTTCAGACGTGTGCTCTTCCGATCTAAGGCCATGCAGAGAGCTTCCGA) and subjected to deep sequencing using a MiSeq platform (Illumina) as previously described[34]. KO cell clones contained all-allelic frame shift mutations without any WT reads. The deficiency for cGAS was not validated at the protein level.

All bought cell lines were kept at low passages to maintain their identity. Non-commercially available HEK293T STING-KI, BLaER1, and BLaER1 cGAS-KO cell lines were not authenticated.

The female primary human fibroblast cell line was authenticated as follows. DNA was isolated separately from the samples. Genetic characteristics were determined by PCR-single-locus-technology. Twenty-one independent PCR-systems (Amelogenin, D3S1358, D1S1656, D6S1043, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433, and FGA) were investigated (PowerPlex 21 PCR Kit, Promega). In parallel, positive and negative controls were performed yielding correct results.

No mycoplasma contamination of the used cells was detected in regular screenings.

Cell culture. WT and cGAS-KO BLaER1 cells were cultivated in RPMI medium containing heat-inactivated 10% FCS, penicillin (100 U ml^{-1}), streptomycin (100 µg ml^{-1}) (Thermo Fisher Scientific), and 1 mM sodium pyruvate (Thermo Fisher Scientific). For transdifferentiation, 5×10^4 cells were seeded per well of a flat bottomed 96-well plate and cultivated in the presence of β-oestradiol (100 nM, Sigma-Aldrich), hr-IL-3 (10 ng ml^{-1}), and M-CSF (10 ng ml^{-1}) (both PeproTech) for 5 days before the experiment[35]. HEK293T, HEK293T STING-KI, and human fibroblasts were cultured in DMEM (Thermo Fisher Scientific or Sigma-Aldrich, respectively) supplemented with 10–20% heat-inactivated FBS (Thermo Fisher Scientific or Biochrom, respectively). All cells were incubated at 37 °C with 5% CO_2.

HMGB1-KO and WT MEFs were cultured in DMEM (Thermo Fisher Scientific) supplemented with 15% heat-inactivated FBS (Thermo Fisher Scientific), 2 mM L-glutamine (Sigma, G7513), 1% non-essential amino acids (Sigma, M7145), 0.1 mM β-mercaptoethanol (Sigma, M3148), and 0.1 mg ml^{-1} penicillin–streptomycin (Sigma, P4333) and incubated at 37 °C with 5% CO_2.

Protein expression and purification. All proteins were overexpressed in *E. coli* Rosetta (DE3) for 16–18 h at 18 °C after induction with 0.2 mM IPTG. Cells were lysed by sonication in 50 mM Tris, 500 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% glycerol, pH 7.5, supplemented with 2 mM β-mercaptoethanol and protease inhibitor cocktail (Sigma-Aldrich) and purified with Ni-NTA agarose resin (Qiagen). For truncated cGAS proteins and human full-length cGAS without MBP-tag, His_6–MBP-tag was removed with tobacco etch virus protease (1:50 mass ratio) during 16 h dialysis against 30 mM Tris, 100 mM NaCl, 2 mM DTT, pH 7.0. cGAS proteins were further purified by cation-exchange chromatography (30 mM Tris, 100 mM/1 M NaCl, 2 mM DTT, pH 7.0) on HiTrap SP HP columns (GE Healthcare) followed by size-exclusion chromatography (SEC) on HiLoad S200 16/60 column (GE Healthcare) equilibrated with 20 mM Tris, 100 mM NaCl, pH 7.5. Full-length hcGAS with or without N-terminal His_6–MBP-tag (hcGAS) and mcGAS proteins was concentrated to 8–12 mg ml^{-1}. Truncated hcGAS was concentrated to 4 mg ml^{-1}. All proteins were flash-frozen in liquid nitrogen and stored at −80 °C. mTfam and mHMGB1 variants were purified as described for full-length hcGAS with His_6–MBP-tag, except the cation-exchange chromatography step was omitted and after dialysis against 20 mM Tris, 300 mM NaCl, pH 7.5 SEC on HiLoad S75 16/60 (GE Healthcare) column was performed. Proteins were concentrated to 10–13 mg ml^{-1}.

lHU was purified as described for full-length hcGAS with His_6–MBP-tag, except after dialysis affinity chromatography on a HiTrap Heparin HP (GE Healthcare) column was performed (20 mM Tris, 100 mM/1 M NaCl, 2 mM DTT, pH 7.5). HiLoad S75 16/60 (GE Healthcare) equilibrated with 20 mM Tris, 100 mM NaCl, pH 7.5 was used for SEC. Protein was concentrated to 7 mg ml^{-1}.

Crystallization of cGAS–DNA complex. For crystallization, purified mcGAS (aa 141–507) 6 mg ml^{-1} was mixed with 39 bp DNA (39 bp-s: AGATCTACTAGTGATCTATGACTGATCTGTACATGATCT; 39 bp-as: AGATCATGTACAGATCAGTCATAGATCACTAGTAGATCT) in a molar ratio of 1:0.6 protein:DNA in a buffer containing 20 mM Tris pH 7.5, 300 mM NaCl and 20 mM $MgCl_2$. Crystals were obtained by hanging-drop vapour diffusion in 0.1 M Tris pH 8, 0.2 M ammonium citrate pH 7, 27.5% w/v PEG3350 after 1 month at 20 °C. The crystals were soaked in 25% glycerol diluted in reservoir solution, flash-frozen, and stored in liquid nitrogen.

Data collection and refinement. X-ray diffraction data were collected at the beamline PXI (X06SA) at the Swiss Light Source, Switzerland, using a Pilatus 6M detector. Data sets were processed with XDS and merged with XSCALE[36]. The STARANISO server[37] was used to generate structure factor amplitudes and their s.d. with a local $I/\sigma(I)$ cut-off of 1.9. The resulting data set showed varying high-resolution cut-offs between 4.8 Å and 3.6 Å and an effective resolution of ~4.2 Å. The crystal structure was solved by molecular replacement with PHASER[38,39] using six copies of a search model based on the published structure of mouse $cGAS^{cd}$ in complex with 18 bp DNA[8] (Protein Data Bank (PDB) accession number 4LEY). Iterative model building and refinement was done with the molecular graphics program MOLOC[40] and the CCP4 suite refinement program REFMAC5 (ref. 41). Owing to the rather low resolution, we could not identify the DNA sequence. Thus, sense and anti-sense strands, as well as the starting nucleotide, were chosen arbitrarily. At the symmetry contact between two neighbouring cGAS dimers, we saw continuous electron density for a 34 bp DNA passing through the crystallographic twofold axis. Since the DNA sequence was not palindromic, we modelled the DNA with two 17 bp long symmetry-halves. Except for removal of a steric clash with the bound DNA by choosing a different rotamer for Arg244, we did not attempt to re-model the cGAS protein. Data collection and refinement statistics are listed in Supplementary Table 3.

Fluorescence-based cGAS activity assays. In fluorescence-based cGAS activity assays, a fluorescent analogue of ATP (2-aminopurine riboside-5′-*O*-triphosphate (Biolog) (fATP)) was used. DNA (6.5 ng μl^{-1}) of different lengths (20–100 bp in 5 bp intervals, and pET28M–SUMO1–GFP vector (EMBL) (6.2 kbp, plasmid)) corresponding to roughly 0.5 µM 20 bp (approximate length of cGAS binding site) was premixed with 0.5 µM cGAS in 40 mM Tris pH 7.5 and 100 mM

NaCl. Alternatively, 2.6 ng μl^{-1} DNA (~0.2 μM binding sites) and 1 μM cGAS, 13 ng μl^{-1} DNA (~1 μM binding sites), and 0.2 μM cGAS or 2.6 ng μl^{-1} DNA (~0.2 μM binding sites) and 0.2 μM cGAS were used. To compare cGAS mutants, the same 2.6 ng μl^{-1} DNA (~0.2 μM binding sites) was mixed with 1 μM cGAS. The reaction was started by adding 5 mM $MgCl_2$ with 500 μM GTP and 50 μM fATP and performed at 32 °C. Fluorescence decrease was measured in 96-well black non-binding PS plates (Greiner Bio-One) on Tecan infinite M1000 (λ_{ex} = 305 nm, λ_{em} = 363 nm, gain 100, settle time 10 ms, kinetic interval 2 min). In cGAS titration experiments, 2.6 ng μl^{-1} DNA (~0.2 μM binding sites) was mixed with increasing cGAS concentrations from 0.05 to 20 μM.

cGAS stimulation assays induced by DNA-bending protein were performed analogously. Briefly, 13 ng μl^{-1} DNA (~1 μM binding sites) was premixed with DNA-bending proteins or cGAS inactive mutant in concentrations of 0–5 μM, after which 100 nM or 50 nM cGAS in 40 mM Tris pH 7.5 and 100 mM NaCl was added. The reaction was started by adding 5 mM $MgCl_2$ together with 500 μM GTP and 50 μM fATP and performed at 32 °C for mcGAS^{cd} or 37 °C for hcGAS. Fluorescence measurement was made as described above.

Data were analysed with OriginPro 8G (OriginLab).

Radiolabelled cGAS activity assays. Radiolabelled cGAS activity assays were performed analogously as previously described[7]. Briefly, 13 ng μl^{-1} DNA (~1 μM binding sites) were mixed with 2 μM mcGAS^{cd} and the reaction started by adding 50 μM ATP, 500 μM GTP, 5 mM $MgCl_2$ in 40 mM Tris pH 7.5, and 100 mM NaCl containing 1:800 [α^{32}P]ATP (3,000 Ci $mmol^{-1}$, Hartman Analytic). Samples were incubated at 35 °C and the reactions were stopped by plotting on PEI-Cellulose F plates (Merck) and analysed by thin-layer chromatography with 1 M $(NH_4)_2SO_4$/1.5 M KH_2PO_4 as running buffer. The radiolabelled products were visualized with a Typhoon FLA 9000 phosphor imaging system. For testing of fATP incorporation into cGAS enzymatic activity product, 1 μM mcGAS^{cd} was mixed with 13 ng μl^{-1} (~1 μM binding sites) 55 bp, 500 μM fATP/ATP, 500 μM GTP, 5 mM $MgCl_2$, and 1:160 [α^{32}P]ATP or [α^{32}P]GTP (3,000 Ci $mmol^{-1}$, Hartman Analytic) in the same condition. For testing TFAM-dependent increase in cGAS activity, 250 nM mcGAS^{cd} was mixed with 13 ng μl^{-1} (~1 μM binding sites) plasmid in the same condition and 1:1,000 instead of 1:160 [α^{32}P]ATP was added. The reaction was incubated at 32 °C and the radiolabelled products were separated and visualized as described previously. For testing cGAS activation by short DNAs, the previously described protocol was used[8]. Briefly, 10 μM or 5 μM cGAS was mixed with 650 ng μl^{-1} or 325 ng μl^{-1} DNA, respectively, in buffer containing 5 mM $MgCl_2$, 2 mM ATP and GTP, and 1:400 [α^{32}P]ATP. The reaction mixture was incubated at 37 °C and the radiolabelled products were separated and visualized as described previously.

Anion-exchange chromatography of cGAS reaction products. Ten micromolar mcGAS^{cd} or hcGAS^{cd} and 195 ng μl^{-1} (~15 μM binding sites) plasmid DNA were incubated at 35 °C for 2 h in buffer containing 40 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM fATP, and 2 mM GTP. Reaction mixtures were centrifuged for 10 min at 16,100 relative centrifugal force, and the supernatant was separated from cGAS by ultrafiltration (30 kDa, Amicon). Resulting flow through was diluted in 50 mM Tris pH 9.0 and loaded on a Mono Q 5/50 GL (GE Healthcare) for anion-exchange chromatography (50 mM Tris, 0 M/1 M NaCl, pH 9.0). Control runs with fATP, GTP, 2′3′- and 3′3′-cGAMP were made analogously to validate the resulting peaks of cGAS reaction products.

Electrophoretic mobility shift assay. DNA (2.6 ng μl^{-1}) of length 200 bp (~0.2 μM binding sites) was incubated with 0–2.5 μM WT or mutant mHMGB1dCTT for 30 min on ice in the buffer containing 50 mM Tris pH 7.5 and 100 mM NaCl. Samples were separated in 0.6% agarose gel prepared with Gel-Red (Biotium) in 40 mM Tris pH 9.2 as running buffer. The gel images were obtained with Gel iX Imager (Intas).

ITC. The calorimetric titration of DNA (60–400 μM in the syringe) to mcGAS^{cd} (20 μM in the reaction cell) was performed with MicroCal PEAQ-ITC (Malvern) in buffer containing 30 mM HEPES, 100 mM NaCl, pH 7.5. The following parameters were used: 25 °C, 1 × 0.4 μl + 13 × 3 μl injections,10 $\mu cal\ s^{-1}$ reference offset, 750 revolutions per minute syringe stirring speed, 60 s initial delay, 6 s injection duration, 150 s spacing between injections, high feedback. To study TFAM direct interaction with cGAS, 50 μM human full-length cGAS without MBP-tag in the reaction cell and 400–500 μM hTFAM in the syringe were used. ITC was performed with the same parameters as above at 25 °C, 35 °C, and 15 °C. K_d, molar ratio, Gibbs free energy, enthalpy, and entropy of the binding were calculated with MicroCal PEAQ-ITC analysis software (Malvern).

SEC–RALS. SEC of cGAS–DNA complexes was performed on a Superose 6 increase 10/300 column (GE Healthcare). mcGAS^{cd} (4 mg ml^{-1}) was mixed with DNA in 1:0.6 molar ratio for 50 bp DNA or 1:0.3 molar ratio for 70 and 80 bp DNA in binding buffer containing 20 mM Tris, 243 mM NaCl, pH 7.5. Alternatively, 6 mg ml^{-1} mcGAS^{cd} and 20 bp DNA in 1:1.1 molar ratio were used. One hundred microlitres of the mix were loaded to Superose 6 increase 10/300 column (GE Healthcare) equilibrated with 20 mM Tris, 100 mM NaCl, pH 7.5 buffer and separate peaks were analysed with RALS using Viscotek 270 Dual Detector (Malvern) and Viscotek VE3580 Refractive Index Detector (Malvern). Molecular masses of the complexes were calculated with OmniSEC 4.7.0 software (Malvern).

ELISA. Fifty thousand trans-differentiated BLaER1 cells were transfected with 20, 40, or 60 ng DNA of different lengths (20–100 bp in 5 bp intervals) and herring testis DNA using 0.5 μl Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific) in 50 μl Opti-MEM Reduced Serum medium (Thermo Fisher Scientific). CXCL10 expression in the supernatants was quantified 8 h after transfection using ELISA (BD OptEIA, human IP-10 ELISA Set). Cells stimulated with transfection reagent only and unstimulated cells served as control.

IFN-β mRNA expression analysis. Fifty thousand trans-differentiated WT and cGAS-KO BLaER1 cells were transfected with 40 ng DNA of different lengths (20, 40, 60, 80, 100 bp and herring testis DNA). Eight hours after stimulation, total RNA was isolated from 7.5 × 10^5 pooled cells using an RNeasy Mini Kit (Qiagen) following the vendor's recommendations. Subsequently, RNA was digested with DNase I (Thermo Fisher Scientific) to remove residual DNA. Five hundred nanograms of RNA were reverse transcribed using poly $(dT)_{18}$ oligonucleotides according to the manufacturer's instructions (RevertAid cDNA Synthesis Kit). IFN-β expression levels were analysed by quantitative PCR, using gene-specific primers (IFN-β forward primer: CAGCATCTGCTGGTTGAAGA; reverse primer: CATTACCTGAAGGCCAAGGA), normalized to GAPDH expression measured analogously (GAPDH forward primer: GAGTCAACGGATTTGGTCGT; reverse primer: GACAAGCTTCCCGTTCTCAG) and the fold change was calculated on the basis of the unstimulated control.

Luciferase reporter assays. All immunostimulatory assays were performed in HEK293T STING-KI cells[31]. Five hundred thousand cells were seeded on 24-well plates and transfected with 100 ng p-125Luc[42], 10 ng pGL4.74 (Promega), 50 ng Flag/HA–cGAS plasmids, and a total of 500 ng DNA per well (filled up with an empty vector pcDNA5) using Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific) as transfection reagent according to the vendor's protocol. After 14 h, cells were lysed in 200 μl passive lysis buffer (Promega). Immunoactivity experiments using a Dual-Glo luciferase assay system (Promega) were performed as previously described[43].

Co-immunopurification of cGAS and TFAM from HEK293T cells. Ten million HEK293T cells were transfected with 10 μg Flag–HA–cGAS and 10 μg HA–TFAM expression vectors and harvested 24 h after transfection. Cell pellets were flash-frozen in liquid nitrogen and stored at −20 °C. For immunoprecipitation, cells were incubated in Nonidet P-40 lysis buffer (50 mM HEPES, 150 mM KCl, 1 mM NaF, 0.5% NP-40, 2 mM DTT, 10 μM $ZnCl_2$, protease inhibitor (Sigma-Aldrich), pH 7.5) for 10 min on ice. Lysates were cleared by centrifugation for 30 min. Proteins were immunoprecipitated for 1.5 h with anti-hcGAS or a control antibody bound to magnetic protein G Dynabeads (Novex, Life Technologies). Beads were washed four times with Nonidet-P40 lysis buffer and subjected to SDS–polyacrylamide gel electrophoresis and immunoblotting.

Fluorescence microscopy. HEK293T cells and primary human fibroblasts were grown at 40–45% and 50–60% confluency, respectively, on coverslips in a six-well plate overnight and treated the next day with the inhibitors Q-VD-OPH (MP Biomedicals) and ABT-737 (Santa Cruz Biotechnology) at a final concentration of 10 μM for 0 h or 6 h. Cells were fixed with 3% paraformaldehyde and 0.1% glutaraldehyde in PBS for 12 min, reduced with 1 mg ml^{-1} $NaBH_4$ for 7 min, permeabilized with 0.25% (v/v) Triton X-100, and blocked with 3% (w/v) BSA in PBS for 3 h.

MEF and MEF HMGB1-KO cells were grown at 40–50% confluency on coverslips in a six-well plate overnight and transfected with 2 μg pcDNA5/FRT/TO (Invitrogen, Thermo Fisher Scientific) or eGFP–cGAS expression vectors overnight. Cells were then fixed with 2% paraformaldehyde in PBS for 10 min at room temperature, permeabilized with 0.5% (v/v) Triton X-100, and blocked with 2% (w/v) BSA in PBST for 10 min.

Immunostaining was performed for 1 h at room temperature with primary antibodies against TFAM (1:50), TOM20 (1:250), or HMGB1 (1:100). The primary antibodies were detected with the secondary anti-mouse or anti-rabbit goat or donkey antibodies (1:500) by incubation for 1 h. After immunostaining, samples were post-fixed in 4% paraformaldehyde for 10 min, stained with 1 $\mu g\ ml^{-1}$ (HEK293T and primary human fibroblasts) or 10 $\mu g\ ml^{-1}$ DAPI (MEF) for 10 min, and mounted in VECTASHIELD (Vector Laboratories). Cells were washed three times with PBST after each step.

Three-dimensional structured illumination microscopy (3D SIM) was performed with a DeltaVision OMX V3 microscope (GE Healthcare), equipped with a ×100/1.40 numerical aperture PlanApo oil-immersion objective (Olympus), Cascade II:512 EMCCD cameras (Photometrics), and lasers for 405 nm, 488 nm,

and 594 nm. Image stacks were first reconstructed and corrected for colour shifts with softWoRx 6.0 Beta 19 (unreleased) software. After establishing composite tiff stacks with a custom-made macro in Fiji, the data were subsequently aligned again and maximum intensity projections were used.

Wide-field fluorescent microscopy was performed with a personal DeltaVision (pDV) microscope (GE Healthcare) equipped with a ×60/1.42 oil-immersion objective PlanApo U (Olympus), Cool-Snap camera (12 bit, 1024 pixels × 1024 pixels, Photometrics) by acquiring one focal plane. Insight SSI LEDs (GE Healthcare) for 405 nm, 488 nm, 594 nm were used.

Statistical analysis. No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments or accessing the outcome.

All experiments were conducted at least three times and the mean values and s.d. of technical or biological replicates (BLaER1 stimulation experiments) were calculated. Statistical significance, if applicable, was calculated on the basis of unpaired one-sided *t*-tests.

For the data fitting in Fig. 3, the description is included in the figure legend and in Supplementary Methods.

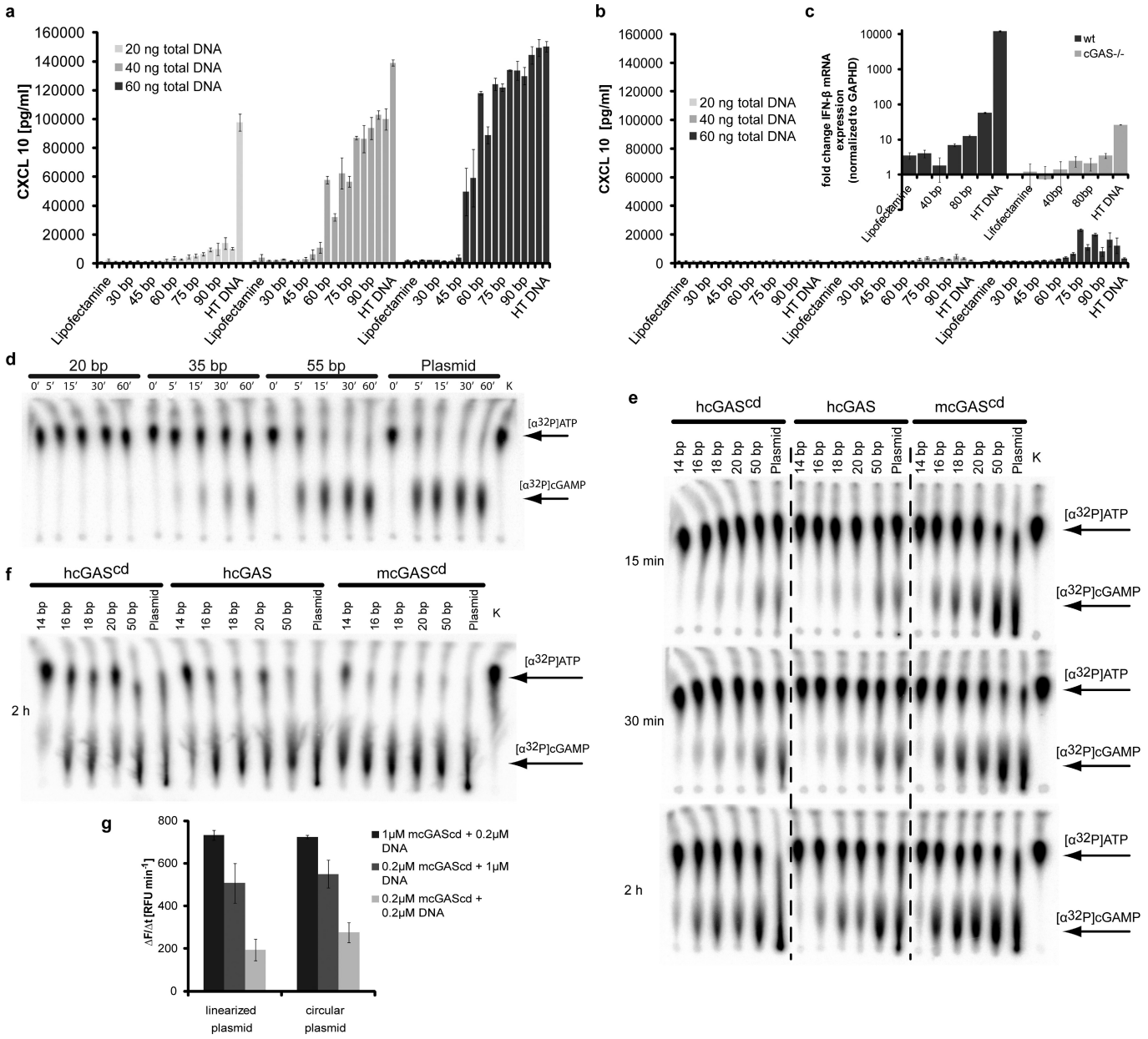
For quantification of TFAM signal in microscopy images presented in Extended Data Fig. 9c, maximum intensity projections were subsequently loaded as RGB into Volocity calculation software (Volocity 6.1.2 (Perkin Elmer)). Areas corresponding to mitochondria, nucleus, and cytosol were obtained, segmented, and measured in all channels by defining intensity threshold and minimum object size on each channel. Cytosolic TFAM signal was quantified according to number of counts and signal area, and additionally corrected and normalized to the cytosolic area in an image. For each calculation, three different cells and three segments per cell were used.

Figure preparation. Figures showing protein structures and electron densities were prepared with PyMOL Molecular Graphic Systems[44]. Sequence alignment (Extended Data Fig. 8a) was prepared with Jalview[45]. All other figures were prepared with Microsoft Excel, OriginPro 8G (OriginLab), or Matlab_R2015a (MathWorks).

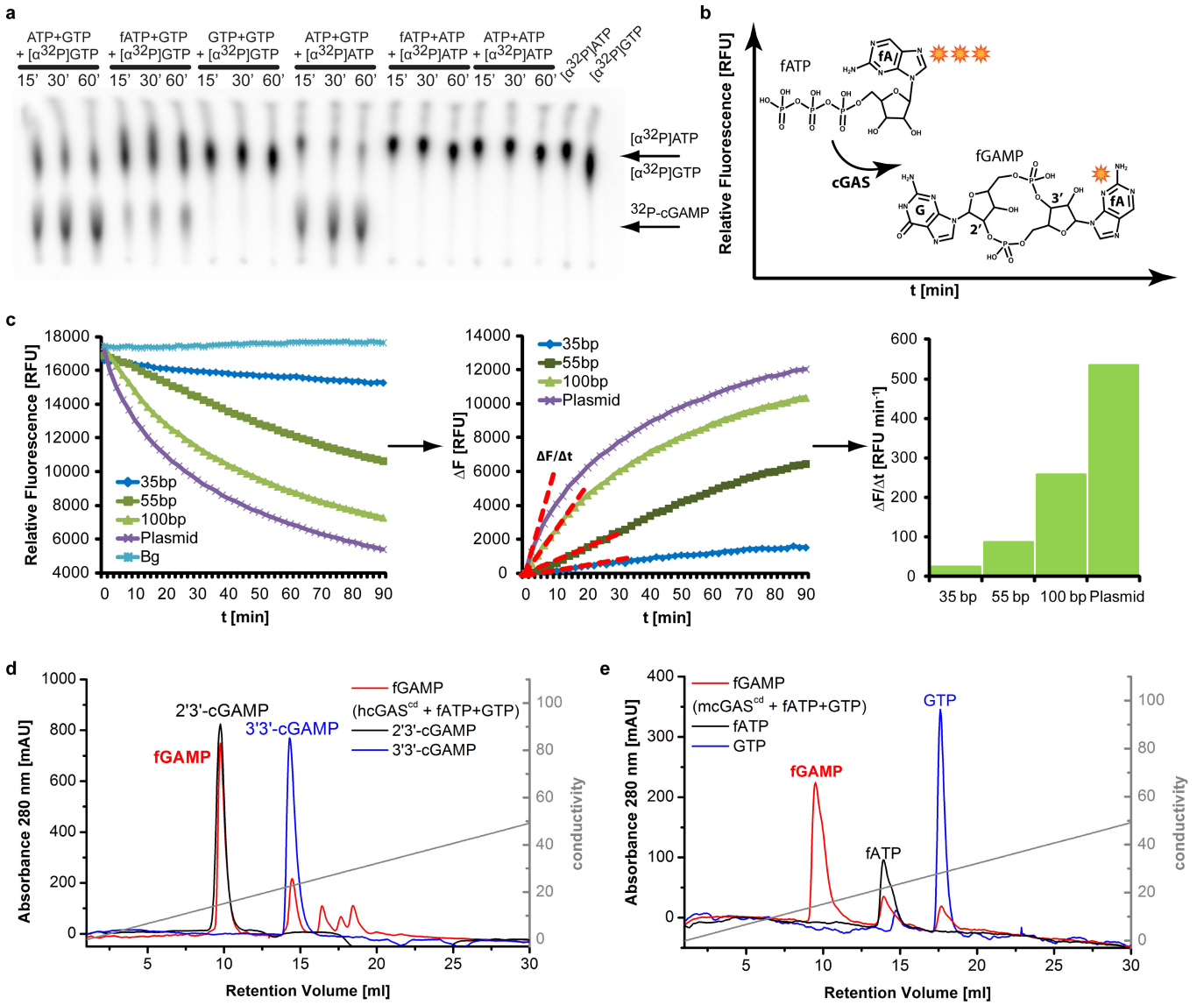
Data availability. The coordinates and structure factors have been deposited in the PDB under accession number 5N6I. All other data are available from the corresponding author upon reasonable request.

Code availability. Matlab code used in this study is available from the corresponding author upon reasonable request.

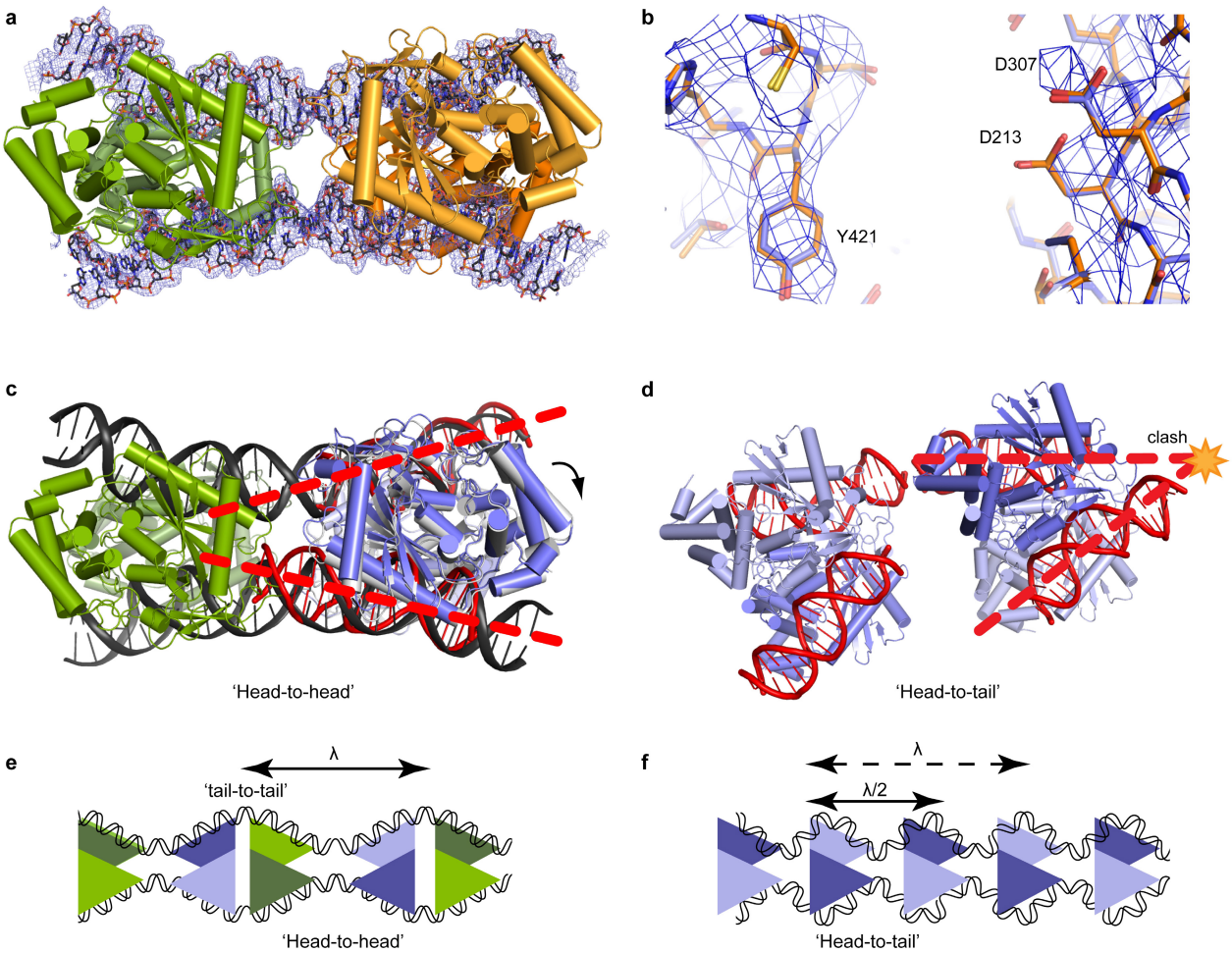
31. Ablasser, A. *et al.* Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature* **503,** 530–534 (2013).
32. Rapino, F. *et al.* C/EBPα induces highly efficient macrophage transdifferentiation of B lymphoma and leukemia cell lines and impairs their tumorigenicity. *Cell Reports* **3,** 1153–1163 (2013).
33. Schmidt, T., Schmid-Burgk, J. L. & Hornung, V. Synthesis of an arrayed sgRNA library targeting the human genome. *Sci. Rep.* **5,** 14987 (2015).
34. Schmid-Burgk, J. L. *et al.* OutKnocker: a web tool for rapid and simple genotyping of designer nuclease edited cell lines. *Genome Res.* **24,** 1719–1723 (2014).
35. Gaidt, M. M. *et al.* Human monocytes engage an alternative inflammasome pathway. *Immunity* **44,** 833–846 (2016).
36. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
37. Tickle, I. J. *et al.* STARANISO (Global Phasing, 2016).
38. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
39. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57,** 1373–1382 (2001).
40. Gerber, P. R. & Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **9,** 251–268 (1995).
41. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
42. Yoneyama, M. *et al.* Autocrine amplification of type I interferon gene expression mediated by interferon stimulated gene factor 3 (ISGF3). *J. Biochem.* **120,** 160–169 (1996).
43. Lässig, C. *et al.* ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *eLife* **4,** e10859 (2015).
44. Schrödinger. The PyMOL molecular graphics system, version 1.8 (2010).
45. Waterhouse, A. M. *et al.* Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).



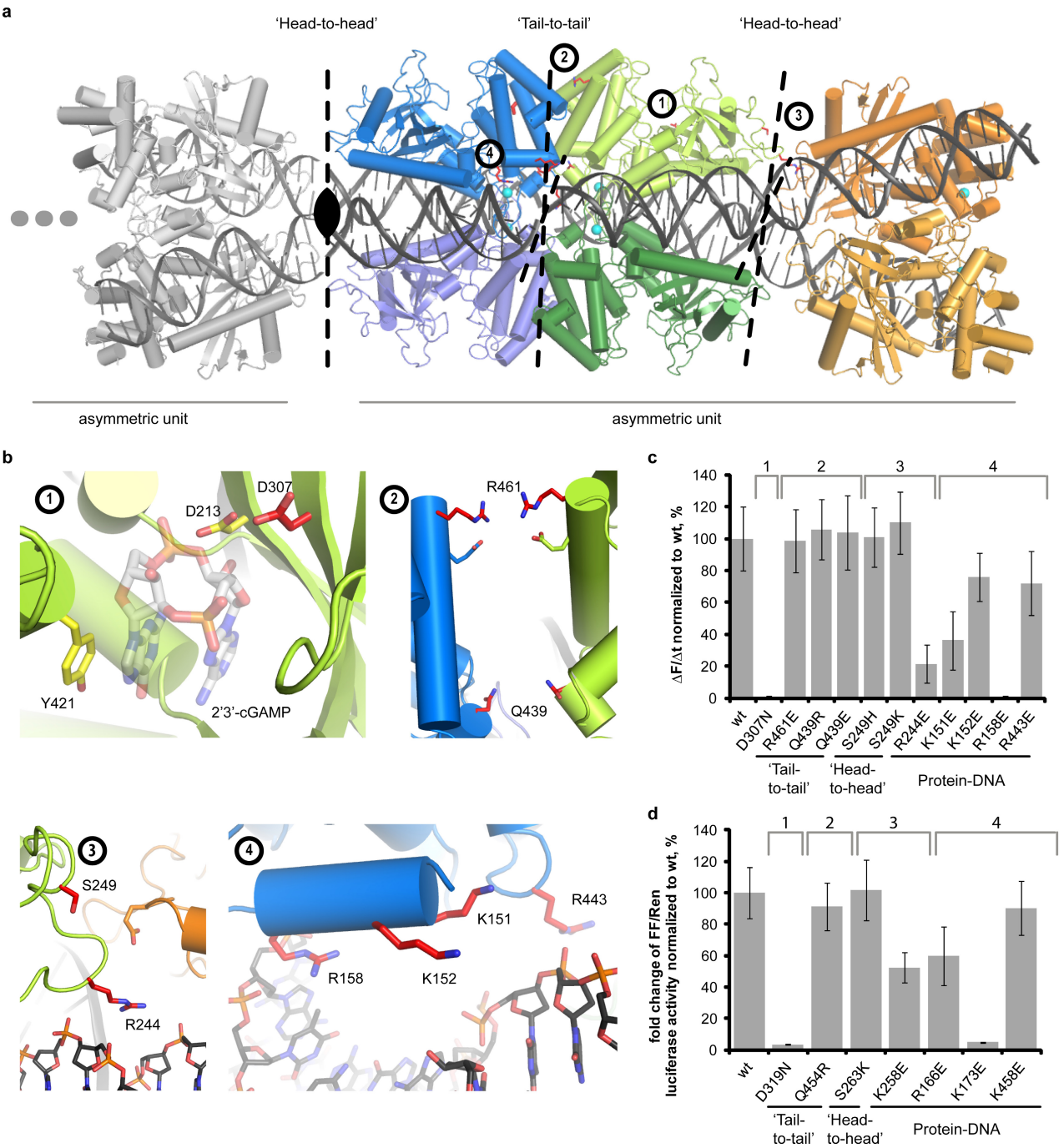
Extended Data Figure 1 | cGAS activity increases with DNA length *in vitro* and *in vivo*. **a**, **b**, CXCL10 cytokine production of WT (**a**) and cGAS-KO (**b**) transdifferentiated BLaER1 cells. Cells were transfected with 20, 40, and 60 ng DNA of increasing length (20–100 bp in 5 bp intervals) and herring testis (HT) DNA, and CXCL10 concentration in the supernatant was measured by ELISA. The first two bars of each series represent unstimulated cells and Lipofectamine controls. Shown are mean values ± s.d., $n = 3$. **c**, Fold change of IFN-β mRNA expression in WT and cGAS-KO BLaER1 cells transfected with 20–100 bp and herring testis DNA. IFN-β mRNA expression levels were normalized to GAPDH mRNA and the fold change was calculated on the basis of the unstimulated control. Shown are mean values ± s.d., $n = 3$. **d**, Radiolabelled cGAMP production of cGAS stimulated with DNA of different lengths (20, 35, 55 bp, and plasmid). cGAS reactions in the presence of [α^{32}P]ATP were stopped at the indicated time points and radiolabelled compounds (shown with black arrows) were visualized. **e**, **f**, Radiolabelled cGAMP production by (left to right) hcGAS^{cd}, hcGAS, and mcGAS^{cd} stimulated with 14, 16, 18, 20, 50 bp and plasmid DNA. cGAS (**e**, 5 μM; **f**, 10 μM) was incubated with 325 ng μl^{-1} (**e**) or 650 ng μl^{-1} (**f**) DNA of indicated length in the presence of ATP, GTP, and [α^{32}P]ATP at 37 °C. The reactions were stopped at the indicated time points and radiolabelled compounds (shown with black arrows) were visualized. **g**, mcGAS^{cd} activity measured by the rate of fATP incorporation into fGAMP (see Extended Data Fig. 2 for the assays) in the presence of linearized or circular plasmid DNA. Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) are plotted against DNA constructs ± s.d., $n = 3$. No significant difference between linearized and circular plasmid could be detected.



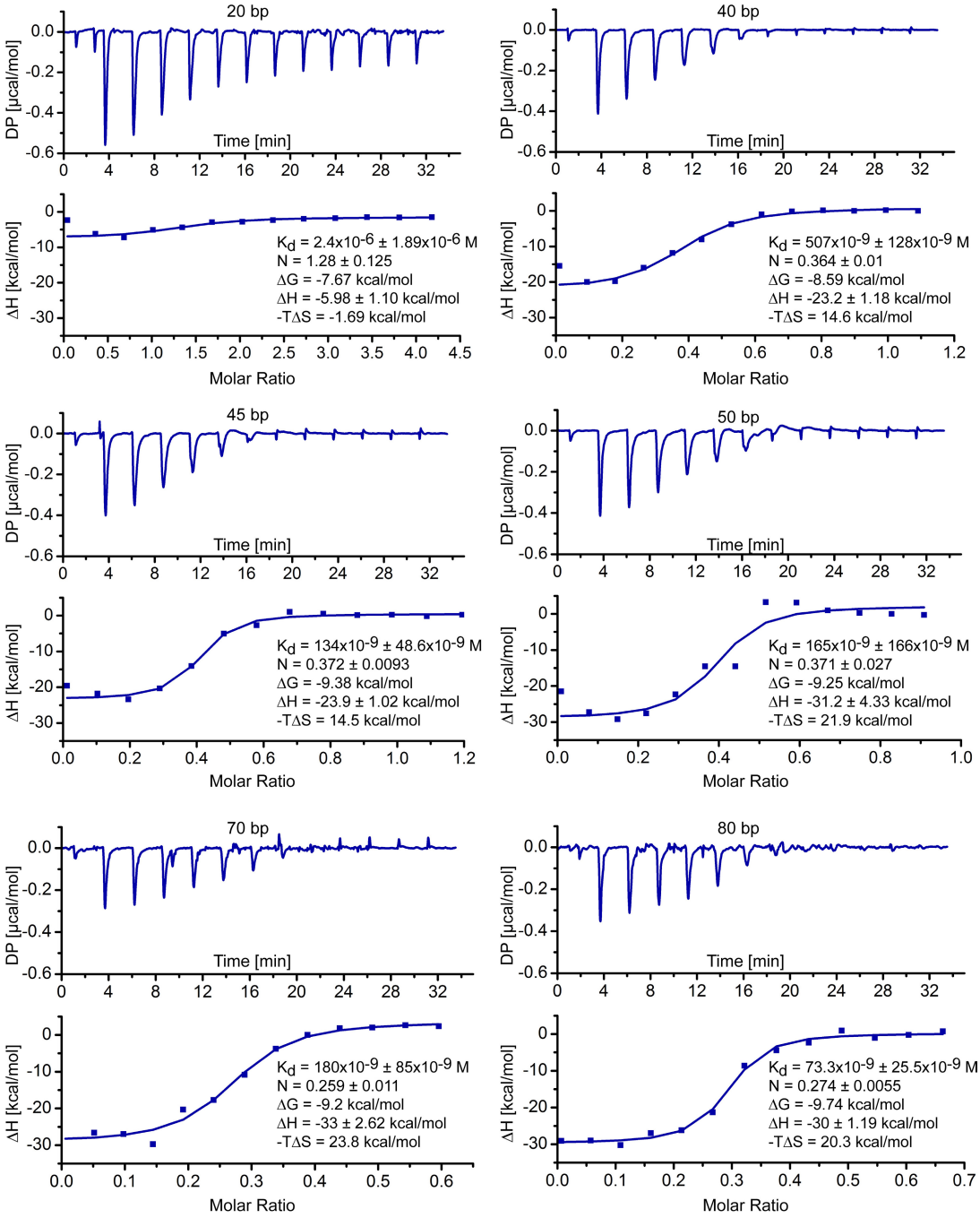
Extended Data Figure 2 | Description of the fluorescence-based cGAS activity assay. a, Radiolabelled cGAS products obtained with different NTP combinations. Reactions were stopped at the indicated time points, and products separated by thin-layer chromatography and visualized by radiography. **b**, The principle of the fluorescence-based cGAS activity assay. cGAS catalyses the conversion of fluorescent ATP analogue fATP (2-aminopurine riboside triphosphate, three orange stars) into less fluorescent fGAMP (fluorescent cGAS product, one orange star), resulting in a gradual decrease in fluorescence intensity during the reaction. **c**, General workflow for calculating the initial cGAS reaction rates. From initial fluorescence curves (left), the background fluorescence was subtracted and the resulting curve was inverted for better visualization (ΔF). Initial rates were calculated as a slope of the linear intervals (red dashed lines) and defined as $\Delta F/\Delta t$ (relative fluorescence units per minute) (right). **d**, **e**, fGAMP mobility in anion-exchange chromatography on a MonoQ 5/50 GL column. **d**, Comparison of fGAMP, produced by cGAS from fATP and GTP (red), 2′3′-cGAMP (black), and 3′3′-cGAMP (blue) mobilities. **e**, Comparison of fGAMP (red), fATP (black), and GTP (blue) mobilities.



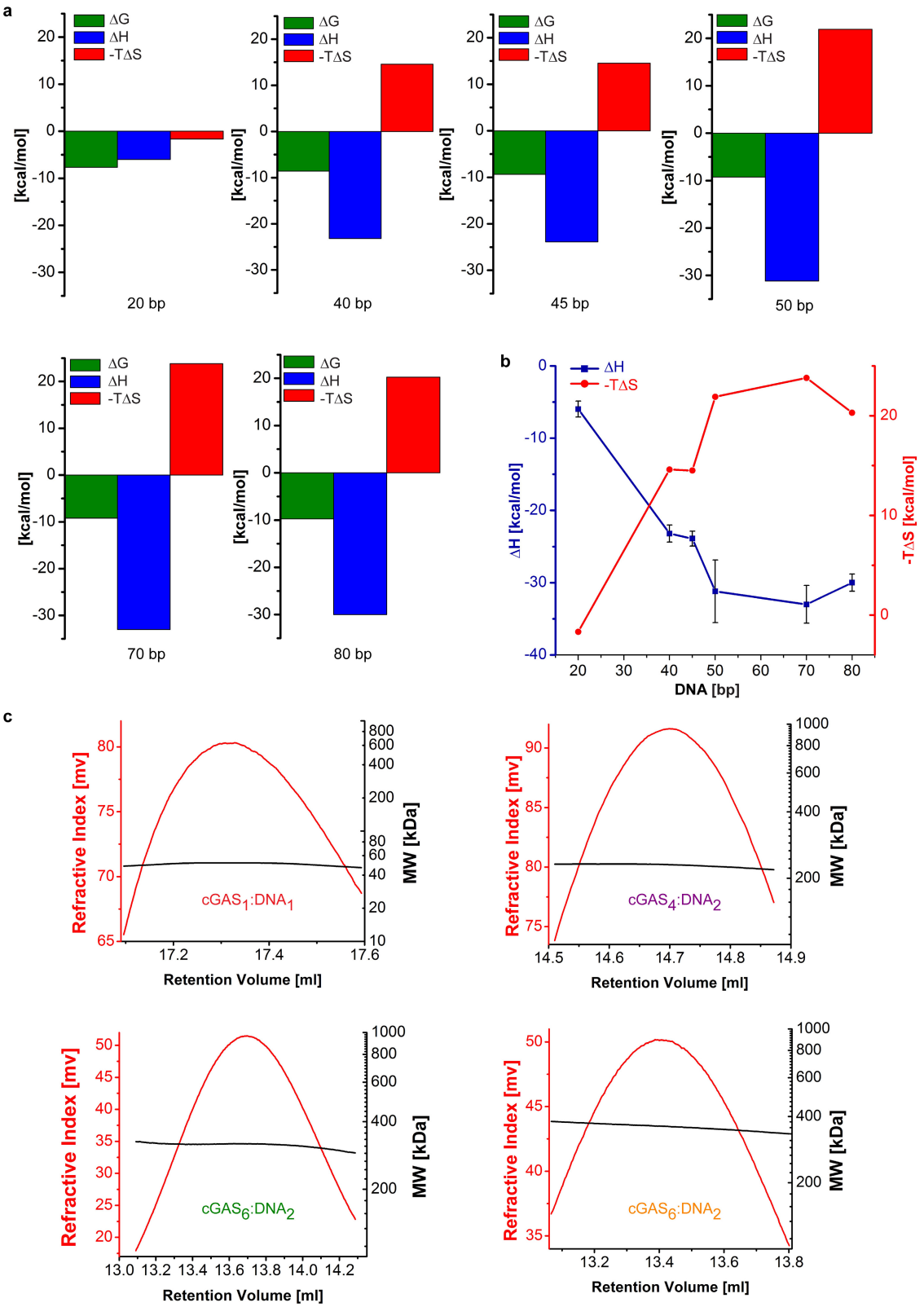
Extended Data Figure 3 | Electron densities and comparison with cGAS in complex with 18 bp DNA. a, $2F_o - F_c$ electron density of 39 bp DNA within the complex at a contour level of 1σ. **b**, Close-up view of the cGAS active site (orange) superposed with a previously published structure of cGAS bound to 18 bp DNA (PDB accession number 4LEY, blue). Shown is the $2F_o - F_c$ electron density of the active-site residues at a contour level of 1σ. **c**, Superposition of the cGAS complex with 39 bp DNA (green and grey) with 18 bp DNA-bound cGAS (PDB accession number 4LEY, blue). A slight rotation of superposed 18 bp-cGAS (blue) relative to 39 bp-cGAS (grey) is shown with a black arrow. The superposition shows the difference between a hypothetical straight (red dashed line) DNA, leading the previously proposed DNA end preference of cGAS, and the curved DNA observed in our crystal structure. **d**, Binding mode of the DNA strands to cGAS within the previously published structure (PDB accession number 4LEY). Two neighbouring crystallographic asymmetric units represent a 'head-to-tail' cGAS dimer orientation. Elongation of both strands (red dashed line) leads to a steric clash (orange star). **e**, Schematic model of cGAS binding to continuous DNA in alternating 'head-to-head' and 'tail-to-tail' arrangements. DNA curves over λ bp (black arrows). **f**, Schematic model of cGAS binding to continuous DNA in a (not observed) 'head-to-tail' arrangement. DNA curves over $\lambda/2$ bp (black arrows), if the same density of cGAS per DNA is assumed. Thus DNA must be bent twice more often than in model **e** to sustain the parallel DNA arrangement and is energetically less favourable.



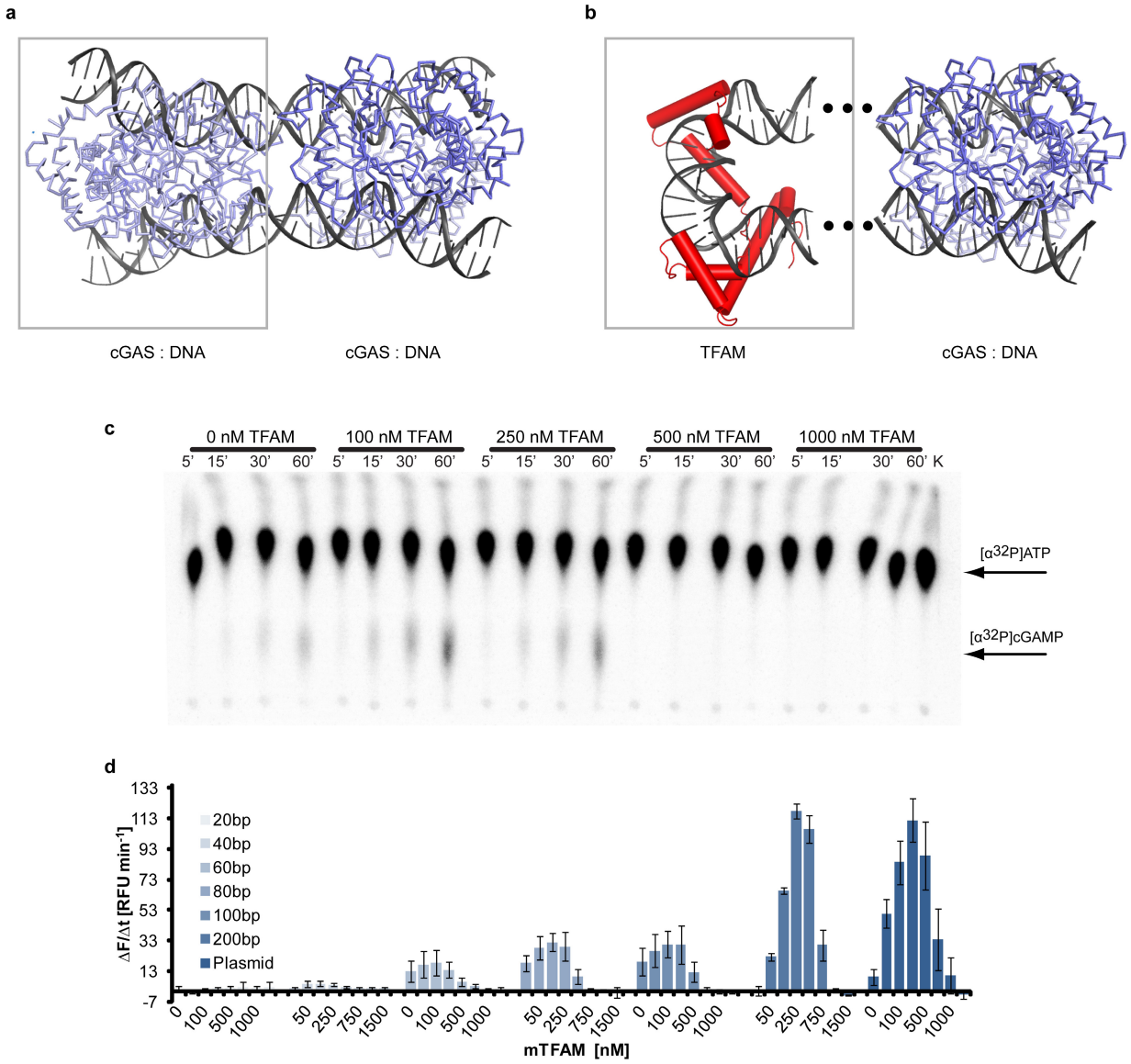
Extended Data Figure 4 | Asymmetric unit of the cGAS–39 bp complex and details of protein–protein and protein–DNA contacts. a, An overview of the asymmetric unit of 39 bp DNA-bound cGAS. Filled oval-like symbol represents a twofold crystallographic symmetry axis; black dashed lines represent non-crystallographic twofold symmetry axes. The asymmetric unit contains one full and one half 'head-to-head'-oriented $cGAS_4$–DNA_2 complex. Residues mutated to examine four areas (encircled 1–4) are shown as red sticks. **b**, Close-up view of mutated interfaces: 1, cGAS active site (yellow) with superimposed cGAMP (from PDB accession number 4LEZ, grey) and labelled active-site residues D213 and D307; 2–4, potential protein–protein or DNA–protein interaction sites, respectively. Mutated residues are in red. **c**, **d**, Mutational analysis of the described regions of $mcGAS^{cd}$ *in vitro* and hcGAS *in vivo*, respectively. D307N and D319N correspond to active-site mutations in mouse and human cGAS, respectively. **c**, cGAS activity measured by the rate of fATP incorporation into fGAMP ($\Delta F/\Delta t$, see Extended Data Fig. 2) and normalized to WT. Mean values represent percentage of WT activity ± s.d., $n = 3$. **d**, Percentage change of IFN-β promoter-driven luciferase activity upon the expression of WT or mutant cGAS. IFN-β response was measured as a proportion of firefly (FF) to *Renilla* (REN) luciferase activity in HEK293T STING-KI cells upon Flag/HA–hcGAS overexpression. All ratios were normalized to WT. Mean values represent the percentage of WT activity ± s.d., $n = 3$. Mutants are named according to their position in hcGAS: D319N (active site), Q454R, S263K, R166E, K173E, K458E, K258E mutants refer to D307N, Q439R, S249K, K151E, R158E, R443E, and R244E in mcGAS.



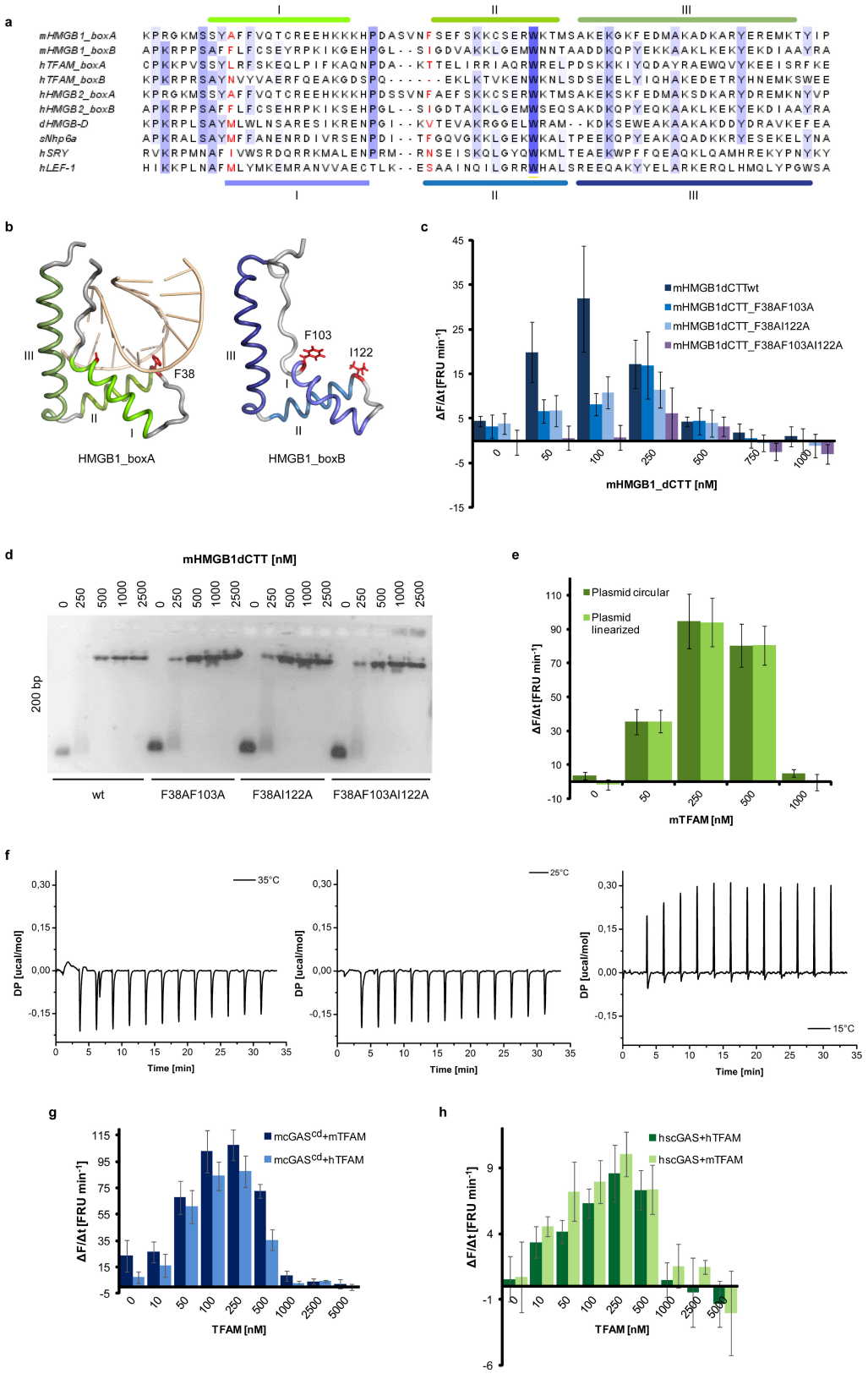
Extended Data Figure 5 | cGAS affinity to DNA increases with DNA length accompanied by increase in number of cGAS binding sites along the DNA. ITC measurements of mcGAS^{cd} binding to 20, 40, 45, 50, 70, and 80 bp DNA. For each DNA, the power differential (DP) is plotted against time and ΔH is plotted against the molar ratio of DNA:cGAS. Calculated binding parameters are given on each graph.



Extended Data Figure 6 | cGAS forms higher-ordered structures on long DNA. a, Signature plots for each ITC measurement (Extended Data Fig. 5) showing ΔG (green), ΔH (blue), and $-T\Delta S$ (red) for each binding reaction. **b**, ΔH and $-T\Delta S$ components of Gibbs free energy measured with ITC for mcGAS^{cd} binding to 20–80 bp DNA are plotted against DNA length ± s.d. of the measured values from the fit. **c**, Molecular mass distribution within SEC peaks (Fig. 2c–f) containing mcGAS$^{cd}_1$–DNA$_1$ (with 20 bp DNA), mcGAS$^{cd}_4$–DNA$_2$ (with 50 bp DNA), and mcGAS$^{cd}_6$–DNA$_2$ (with 70 and 80 bp DNA) calculated with RALS. Refractive index (red) and estimated molecular mass (black) are plotted against retention volume.

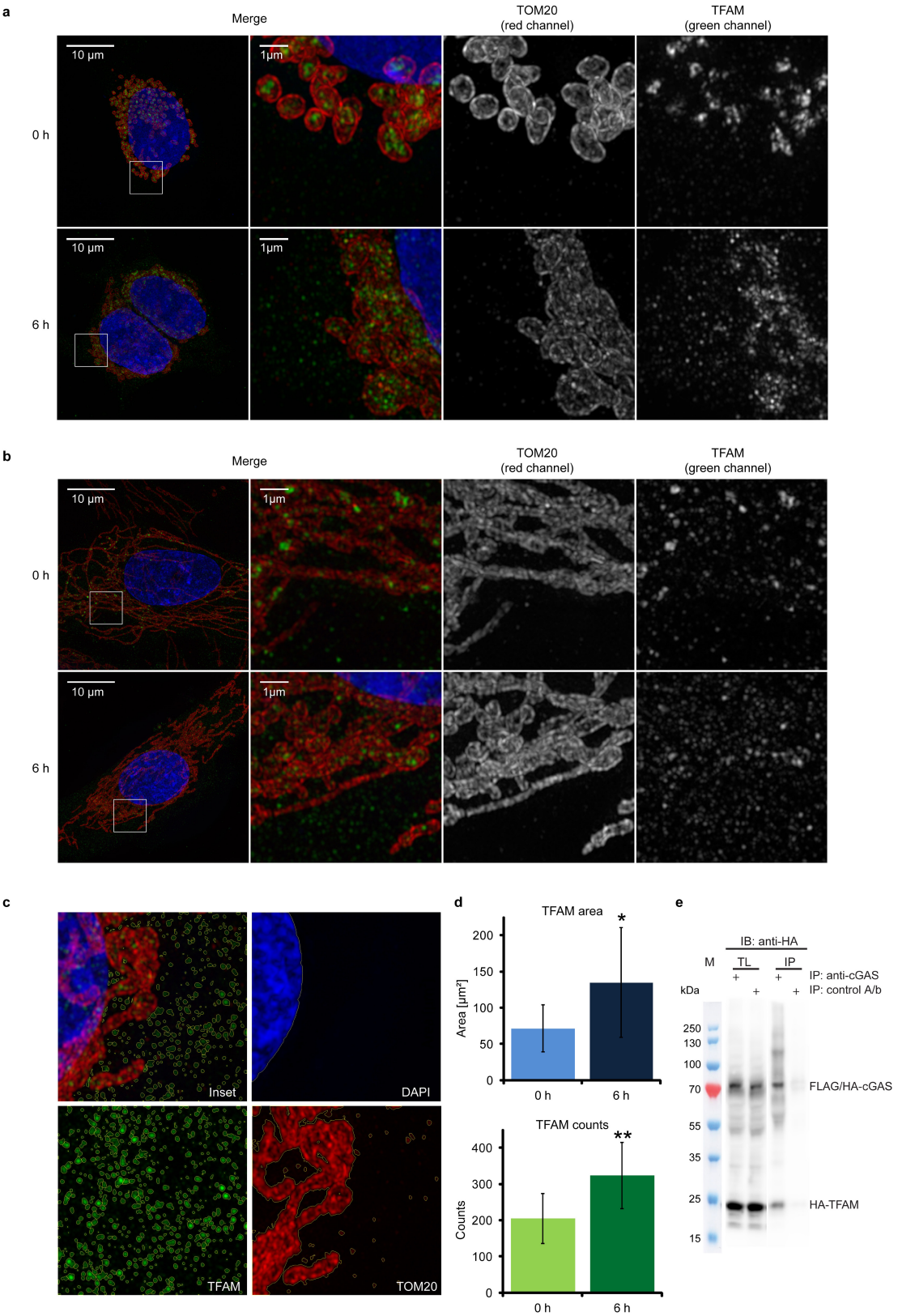


Extended Data Figure 7 | TFAM enhances cGAS activity by prearranging DNA into U-shape. a, DNA ladders (blue) and (**b**) DNA-bound TFAM structure (PDB 3TMM, red) arrange DNA in a remarkably similar fashion. The two similarly spaced DNA strands flanking cGAS ladders or the TFAM U-turn (black dotted lines) provide a possible explanation for the effect of TFAM on cGAS activity. **c**, Radiolabelled cGAMP production in the presence of increasing TFAM concentrations. cGAS reactions with ATP, GTP, and [α^{32}P]ATP were stopped at the indicated time points and the radiolabelled compounds (shown with black arrows) were visualized. **d**, Activation of mcGAS^{cd} by mTFAM and DNA of increasing length (20–200 bp or plasmid DNA). Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$, see Extended Data Fig. 2) are plotted against increasing concentrations of mTFAM ± s.d., $n = 4–8$.



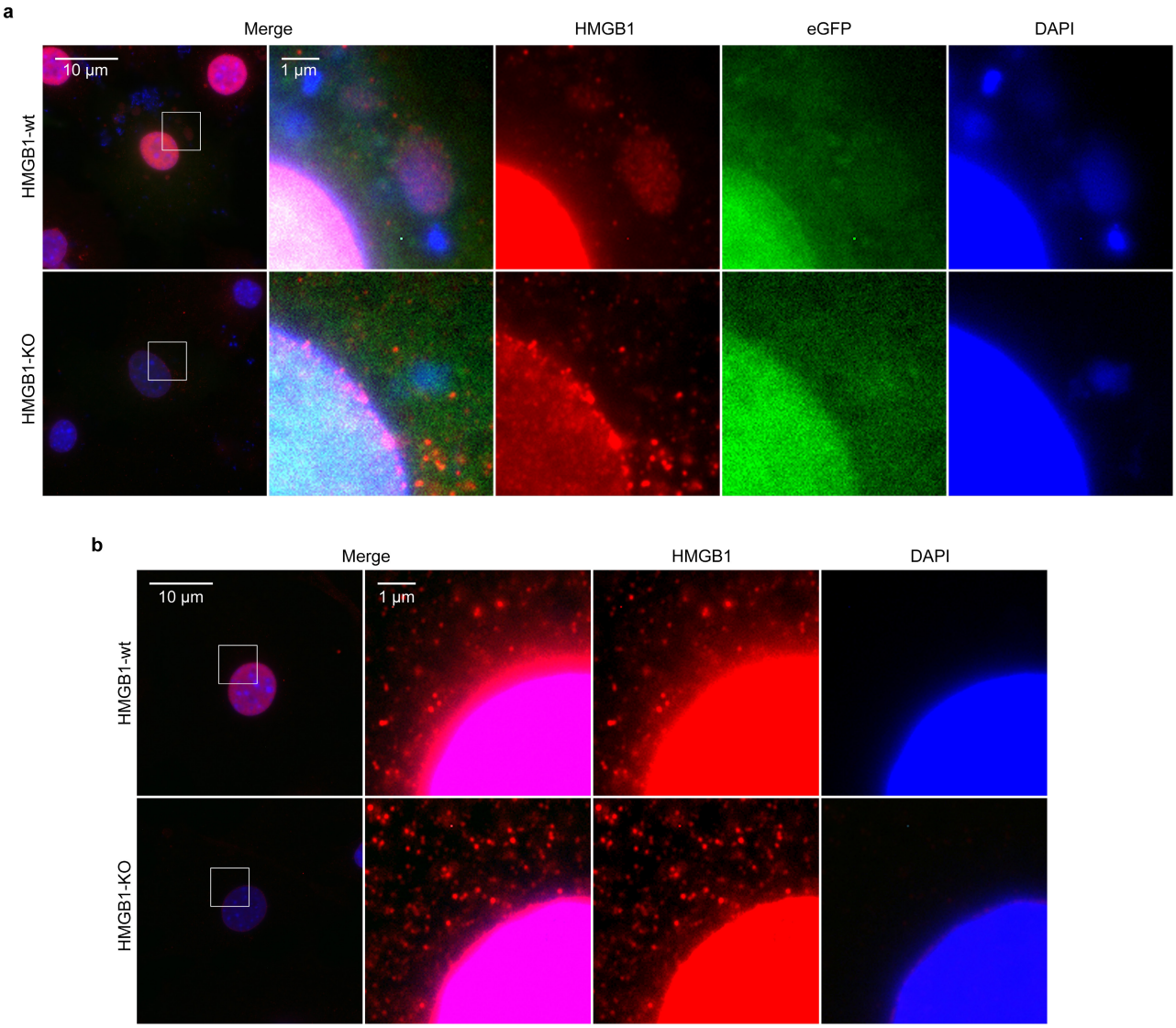
Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | HMGB proteins activate cGAS through DNA bending. a, Sequence alignment of HMG boxes from different HMG proteins. Positions of intercalating residues, responsible for the DNA-bending activity (red), are shown with respect to their location within helices I, II, or III of HMGB1_boxA (green lines) and HMGB1_boxB (blue lines). **b**, Structures of box A in complex with DNA (PDB accession number 4QR9) and box B (PDB accession number 1HME) of rat HMGB1 with indicated intercalating residues (red). **c**, mcGAS^{cd} activity measured by the rate of fATP incorporation into fGAMP (see Extended Data Fig. 2) in the presence of increasing concentration of mHMGB1dCTT WT and intercalating residues mutants. Double mutations (F38A in box A, and F103A or I122A in box B) or triple mutants (F38A in box A, and both F103A and I122A in box B) were used. Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) are plotted against increasing concentrations of mHMGB1dCTT ± s.d., $n = 7$. **d**, Electrophoretic mobility shift assay of mHMGB1dCTT WT and point mutants with 2.6 ng μl^{-1} (~200 nM binding sites) 200 bp DNA. Introduced mutations do not reduce DNA binding of mHMGB1 mutants under conditions used for cGAS activity assays (**c**). **e**, mcGAS^{cd} activity measured by the rate of fATP incorporation into fGAMP (see Extended Data Fig. 2) in the presence of mTFAM with circular or linearized plasmid. Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) are plotted against increasing concentrations of mTFAM ± s.d., $n = 3$. **f**, ITC of hcGAS with hTFAM at 35, 25, and 15 °C. Power differential (DP) is plotted against time of the experiment. hTFAM (~530 μM) was titrated to ~50 μM hcGAS. No binding is observed, indicating $K_d > 100$ μM. **g**, **h**, cGAS activity measured by the rate of fATP incorporation into fGAMP (see Extended Data Fig. 2) in the presence of increasing TFAM concentrations. Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) are plotted against increasing concentrations of TFAM ± s.d., $n = 3$ or 4. mcGAS^{cd} (**g**) and hcGAS (**h**) are activated by both human and mouse TFAM.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | TFAM relocalizes from mitochondria into cytosol during mitochondrial stress. a, **b**, TFAM (green) localization upon mitochondrial stress induction. Cells were incubated with caspase and Bcl-2 inhibitors for indicated time and stained for super-resolution fluorescence microscopy (3D SIM). Mitochondria were visualized by TOM20 (mitochondrial import receptor subunit, red) staining, cell nuclei by DAPI (blue). **a**, TFAM cytosolic localization under mitochondrial stress conditions in HEK293T cells. **b**, TFAM decondensation in mitochondria and its leakage into cytoplasm in human primary fibroblasts in response to mitochondrial stress. **c**, Example of statistical analysis in human primary fibroblasts (**b**). Cells were incubated with caspase and Bcl-2 inhibitors for 6 h and stained for super-resolution fluorescence microscopy (3D SIM). Mitochondria were visualized by TOM20 (mitochondrial import receptor subunit) staining, cell nuclei by DAPI. Areas stained for TFAM (green), TOM20 (red), and nucleus (blue) are defined with yellow line. Cytosolic TFAM spots are depicted on the top left view. **d**, Comparison of cytosolic TFAM signal in control cells (0 h) and under mitochondrial stress conditions (6 h) (**b**, **c**). Plotted are mean values of cytosolic TFAM signal calculated as area (top) or number of spots (bottom) ± s.d. Three cells of each type and three segments per cell were used ($n = 9$). Single asterisk indicates a statistically significant difference ($P < 0.05$, $P = 0.0267$), double asterisk indicates a statistically significant difference ($P < 0.01$, $P = 0.0050$), compared with control. The TFAM signal in the cytosol of control cells (0 h) represents background signal. **e**, Co-immunopurification of cGAS and TFAM. Cell lysates with overexpressed Flag/HA–cGAS and HA–TFAM were incubated with anti-hcGAS or with a control antibody (A/b) and the proteins in total lysates (TL) and elution fractions (IP) were visualized by immunoblotting (IB).



Extended Data Figure 10 | HMGB1 but not eGFP co-localizes with cytosolic DNA. a, **b**, HMGB1 (red), eGFP (green), and DNA (blue) localization in WT or HMGB1-KO MEFs. DNA was stained by DAPI (blue). Images were obtained by wide-field fluorescence microscopy (pDV). Fluorescent signal within enlarged images is enhanced until saturation of the nuclear signal to better visualize cytoplasmic structures. **a**, Cells were transfected with eGFP expression construct. HMGB1 co-localizes with DAPI-staining in WT but not HMGB1-KO cells. eGFP does not co-localize with DNA or HMGB1. **b**, In non-transfected cells, HMGB1 and DAPI stainings are present only in the nucleus. The HMGB1 staining in the cytosol of non-transfected and HMGB1-KO cells represents background signals that become visible with the artificial signal amplification beyond saturation, as shown in the zoom-in images.

nature research

Corresponding author(s): Karl-Peter Hopfner

☐ Initial submission ☒ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample sizes varied between different types of experiments. Sample sizes were at least n=3.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   Experimental findings were reliably reproduced

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   No randomization

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   No blinding

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	OriginPro 8G (OriginLab), MatLab R2015a (MatLab scripts are available upon request), crystallographic software (XDS, XSCALE, STARANISO server, PHASER, REFMAC - see Methods section), OmniSEC 4.7.0 (Malvern), Volocity 6.1.2 (Perkin Elmer, Waltham, MA, USA), MicroCal PEAQ-ITC analysis software (Malvern), Microsoft Excel.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Unique materials such as the protein expression vectors or the cGAS antibody are available upon request. The Tom20 antibody from Santa Cruz Biotechnology is discontinued and cannot be purchased anymore. The source of all other materials is listed in the Methods part and can be purchased as indicated.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	1) The monoclonal anti-human cGAS antibody (raised against the human cGAS Mab21 domain in rat) and a control antibody were produced at the Helmholtz Zentrum München (Monoclonal Antibody Core Facility, German Research Center for Environmental Health) and have been validated in both Western Blots and IPs with purified proteins or human cell lysates containing endogenous or overexpressed human cGAS. 2) The HA antibody coupled to horseradish peroxidase was purchased from Cell Signaling (monoclonal AB raised in mouse , clone 6E2, cat# 2999S, lot# 3) and validated in Western Blots of cell lysates containing overexpressed HA-tagged proteins. 3) The TFAM antibody was purchased from Abnova (polyclonal AB against human full length TFAM produced in mouse, cat# H00007019-B01P, lot# G5251) and validated in Western Blots of human cell lysates containing endogenous and overexpressed TFAM. 4) The Tom20 antibody was purchased from Santa Cruz Biotechnology (polyclonal AB against full length human Tom20 produced in rabbit, cat# sc-11415, lot# K1915) and was validated before (Wurm et al, 2011, PNAS and Kukat et al, 2011, PNAS). 5) The HMGB1 antibody was purchased from Sigma Aldrich (monoclonal AB against human HMGB1 aa 1-91 produced in mouse, clone 2F6, cat# WH0003146M8, lot# FB201-2F6) and was validated in Western Blots of human and mouse cell lysates containing endogenous and/or overexpressed HMGB1 and immunocytofluorescence (M.K. Gunasekaran et al, 2013, Cytokine). 6) All Alexa-coupled secondary antibodies were purchased from Thermo Fisher Scientific (goat anti-mouse Alexa594: cat# A11032; goat anti-rabbit Alexa488 cat# A11034, lot# 1141875; donkey anti-rabbit Alexa594 cat# A21207, lot# 1454437; donkey anti-mouse Alexa488 cat# A21202, lot# 1423052; donkey anti-mouse Alexa594 cat# A21203, lot# 431805) and were validated in fixed cells that were not stained with primary antibodies.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

HEK293T cells were purchased from ATCC (cat# CRL-11268). MEF and MEF HMGB1-KO cells were purchased from HMGBiotech (cat# HM-221).
HEK293T STING-KI cells (knock in of mCherry fused to mouse STING) were generated by transduction of HEK293T cells and are described in Ablasser et al. Nature, 2011.
Female primary human fibroblast cell line is a kind gift from Dr. med. Marion Cremer (T. Cremer group, Ludwig-Maximilians-University Munich, Biocenter Martinsried).
BLaER1 cells were transdifferentiated from a B cell lymphoma cell line (a subclone of the Seraphina Burkitt lymphoma line expressing the transcription factor C/EBPα fused to the estrogen receptor hormone binding domain) which are a kind gift from Thomas Graf, Center for Genomic Regulation, Universidad Pompeu Fabra and Institució Catalana de Recerca i Estudis Avançats, Barcelona. BlaER1 cells are described and functionally validated in Rapino et al. Cell Reports, 2013.
BLaER1 cGAS-KO cells were generated through genomic silencing the cGAS gene in the Seraphina cell line and subsequent transdifferentiation of these cells.
All knock-in or knock-out cell lines were validated by Western Blots of cell lysates and in functional or microscopy assays as written in the manuscript. BLaER1 cGAS-KO cells were further validated by sequencing.

b. Describe the method of cell line authentication used.

All bought cell lines were kept at low passages in order to maintain their identity. Non-commercially available HEK293T STING-KI, BLaER1 and BLaER1 cGAS-KO cell lines have not been authentificated.
Female primary human fibroblast cell line was authentificated as follows. DNA was isolated separately from the samples. Genetic characteristics were determined by PCR-single-locus-technology. 21 independent PCR-systems Amelogenin, D3S1358, D1S1656, D6S1043, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433 and FGA were investigated (Promega, PowerPlex 21 PCR Kit). In parallel, positive and negative controls were carried out yielding correct results.

c. Report whether the cell lines were tested for mycoplasma contamination.

All cell lines were routinely tested for mycoplasm contaminations.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used in this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants.

doi:10.1038/nature23890

Supplementary Methods

Quantitative analysis of cooperative binding of cGAS along DNA

1 Introduction

The catalytic activity of cGAS is activated by binding to DNA, which induces a structural change that properly forms the active site. Initial structural studies showed that two cGAS molecules form a heterotetrameric complex with two DNA ligands. Hereby, the two cGAS molecules sandwich 2, approx. 30° angled DNA duplex molecules. cGAS binds approx. 16-20 base pairs of each DNA ligand. However, dsDNA that in principle fully spans the cGAS dimer (20 bp) does not activate the enzyme to any appreciable amounts *in vitro* and in human cells *in vivo* as well. An activating transition is observed if the DNA length is increased to 40-50 bp. Longer DNA activates cGAS even more robustly, and long plasmid DNA is a "gold standard" for cGAS activation. Even if we account for “DNA end” effects, i.e. the enzyme falling off ends more rapidly than dissociation from internal DNA binding sites through Brownian motion, the evident incapability of short blunt ended DNAs to activate cGAS and the increasing activity with increasing DNA length (keeping the molarity of base pairs constant) is not explained by the cGAS dimer model. Our structural results indicate that once the DNA ligands are long enough to bind two cGAS dimers next to each other, a DNA:protein network is formed that resembles a twisted ladder.

2 DNA-Protein Ladder Model (DPL)

Our new crystal structures suggest that pairs of cGAS dimers mutually stabilize each other via a DNA:protein network: cGAS dimers are positioned like the rungs of a ladder, with the two DNA strands being the beams. A mathematical treatment for the cooperative binding of cGAS dimers to DNA can be formulated along the general ideas of the Monod-Wyman-Changeux (MWC) model for cooperative transitions in proteins: proteins bind as monomers along a single duplex of DNA with a dissociation constant K_1[mol/l]. Along two parallel DNA duplexes, such as those prearranged in the vicinity of a cGAS:DNA heterotretramer, two cGAS molecules bind with an overall microscopic dissociation constant K_2[mol/l]. Since we assume for the model that two cGAS molecules bind simultaneously, the macroscopic dissociation constant for assembly of cGAS dimers with two parallel DNA molecules would be K_2^2, analogous to the empirical Hill equation. Furthermore, the binding of two cGAS:DNA heterodimer into a $cGAS_2$:DNA_2 heterotetramer is characterized by a dissociation constant K_3 [mol/l]. The overall binding scheme is depicted in (Fig. 3a). The model includes simplifications. In particular, we neglect binding of single cGAS molecules to the prearranged "dimeric" DNA lattice and the equilibrium can be formulated as: $2\,cGAS + 2\,DNA \overset{K_2^2}{\Leftrightarrow} cGAS_2{:}DNA_2$. Furthermore, we neglect interactions between DNA molecules with unequal number of bound cGAS molecules. Such an assumption can be justified, if interaction between DNA molecules with equal number of bound cGAS molecules is much stronger than interaction between DNA molecules with unequal numbers of cGAS, because latter has "less" interacting cGAS dimers between the two DNA molecules.

3 Mathematical Modelling

3.1 Nomenclature DNA-Protein Ladder Model (DPL)

In deriving an expression for the activity as a function of the concentration and lengths of DNA ligands and the concentration of protein, we use the following terminology:

p_0: total concentration of protein (cGAS monomer)
l_0: total concentration of DNA molecules
s: number of binding sites per DNA molecule
$p_i l_j$: complex of j DNA with i cGAS molecules
K_1: equilibrium dissociation constant cGAS monomer with DNA
K_2: equilibrium dissociation constant for binding of cGAS to prearranged DNA molecules, forming the $cGAS_2$:DNA_2 complex
K_3: equilibrium dissociation constant for the interaction of two cGAS:DNA complexes, forming the $cGAS_2$:DNA_2 complex
a: a constant relating the concentration of DNA bound cGAS to the observed rate of product formation

3.2 General dimer equilibrium

In general, the formation of a dimer, e.g. the association of two DNA molecules by cGAS, can be described in its simplest form with the following equilibrium scheme:

$$l + l \overset{K}{\Leftrightarrow} l_2 \tag{1.1}$$

with $K = \frac{k_{\text{off}}}{k_{\text{on}}}$. Using the resulting equations

$$l * l = K l_2 \tag{1.2}$$

$$l_0 = l + 2l_2 \tag{1.3}$$

one can derive expression for *l* and l_2 as a function of l_0 and *K*:

$$l = -\frac{1}{4}\left(K - \sqrt{K}\sqrt{K + 8l_0}\right) \tag{1.4}$$

$$l_2 = \frac{1}{8}\left(K + 4l_0 - \sqrt{K}\sqrt{K + 8l_0}\right) \tag{1.5}$$

If *l* has some form of associated activity a_1 (e.g. fluorescence, enzymatic activity) and l_2 a related property $2*a_2$ (since l_2 has two "active" sites), one can combine equations (1.4 and 1.5) to compute the overall activity *A*:

$$A = \frac{1}{4}\left((a_2 - a_1)K + 4a_2 l_0 - (a_2 - a_1)\sqrt{K}\sqrt{K + 8l_0}\right) \tag{1.6}$$

With $a = a_2 - a_1$, $b = a_2$ and $K' = K/4$ one obtains

$$A = aK' + bl_0 - a\sqrt{K'}\sqrt{K' + 2l_0} \tag{1.7}$$

3.3 DNA-Protein Ladder Model (DPL)

We now turn to the scheme outlined in (Fig. 3a) and derive an equation for the concentration of DNA bound cGAS. In doing so, we assume that under steady state conditions, ATP/2APTP, GTP and cGAMP/fGAMP have no substantial influence on K_1, K_2 and K_3. The experimentally observed steady state activity **A** of product formation (in our case measured by (ΔF/Δt: change in fluorescence per time) can be written as

$$A = \sum_{i=1}^{s} a_1 * i * p_i l \; + \; \sum_{i=1}^{s} a_2 * 2\, i * p_{2i} l_2 \tag{2.1}$$

with a_1 and a_2 constants that relate the concentration of DNA bound cGAS monomers (e.g. $p_1 l$) and dimers (e.g. $p_2 l_2$) to the rate of product formation. The factor of 2 in the second sum accounts for the two "active" sites in each $p_{2i} l_2$. For a given number of binding sites **s**, at equilibrium, the scheme in (Fig. 3a) has the following equilibrium equations:

a) conservation of protein and DNA ligand:

$$p_0 \approx p \tag{2.2}$$

$$l_0 = l + 2\, l_2 + \sum_{i=1}^{s} p_i l \; + \; 2\, p_{2i} l_2 \tag{2.3}$$

To derive a mathematical expression, in 2.1 a simplification is used: we designed our experiments with for the most part a surplus of protein over DNA binding sites, assuming that one molecule of cGAS covers ~20 bp of DNA. For the purpose of deriving an analytic mathematical equation, we assume $p_0 >> s * l_0$, i.e. $p \approx p_0$.

b) binding of cGAS to DNA: we assume that cGAS can bind anywhere on DNA. The DNA has s non-overlapping binding sites (e.g. one per 20 bp) all s binding sites are equal.

$$\frac{s * p_0 * l}{p_1 l} = K_1 \tag{2.4}$$

$$\frac{(s-1) * p_0 * p_1 l}{2 * p_2 l} = K_1 \tag{2.5}$$

...

$$\frac{(s-(i-1)) * p_0 * p_{i-1} l}{i * p_i l} = K_1 \tag{2.6}$$

...

$$\frac{p_0 * p_{s-1} l}{s * p_s l} = K_1 \tag{2.7}$$

Likewise for the cGAS dimers bound to two DNA molecules (l_2 denotes the prearranged "DNA dimer"):

$$\frac{s * p_0^2 * l_2}{p_2 l_2} = K_2^2 \tag{2.8}$$

...

$$\frac{(s-(i-1))*p_0^2*p_{2i-2}l_2}{i*p_{2i}l_2} = K_2^2 \quad (2.9)$$

...

$$\frac{p_0^2*p_{2s-2}l_2}{s*p_{2s}l_2} = K_2^2 \quad (2.10)$$

Finally, we have the interaction of two cGAS bound DNA molecules p_1l. We only need one equation; all others are redundant due to thermodynamic linkage.

$$\frac{p_1l*p_1l}{p_2l_2} = K_3 \quad (2.11)$$

For a given s, the equations 2.1a, 2.2, ...2.7 can now be used to eliminate all but one p_il or one $p_{2i}l_2$. Repeating this procedure for each of the p_il and $p_{2i}l_2$, we obtain a set of expressions for the concentrations of the p_il and $p_{2i}l_2$ (i=1...*s*) as functions of s, p_0, l_0, K_1, K_2 and K_3.

For example, let's look at the case **s=3**. In Mathematica code, we formulate the following expressions for equations 2.2, ...2.10:

```
eq1 = l0 - (l + p1l + p2l + p3l +2*l2+2*p2l2 + 2*p4l2 + 2*p6l2)
eq2 = 3*p0*l - k1*p1l
eq3 = 2*p0*p1l - 2*k1*p2l
eq4 = p0*p2l - 3*k1*p3l
eq5 = 3*p0*p0*l2 - k2^2*p2l2
eq6 = 2*p0*p0*p2l2 - 2 k2^2*p4l2
eq7 = p0*p0*p4l2 - 3 k2^2*p6l2
eq8 = p1l*p1l - k3*p2l2
```

In equilibrium, eq0=0, eq1=0, eq2=0 ... We now use these expressions in Mathematica to eliminate l and all p_il but pl:

```
Eliminate[{eq1==0,eq2==0,eq3==0,eq4==0,eq5==0,eq6==0,eq7==0,eq8==0},{l,l2,p2l,p3l,p2l2,p4l2,p6l2
}]
```

This procedure results in the following equation:

```
p1l (k1^3 k2^4 k3 p0 + 3 k1^2 k2^4 k3 p0^2 + 3 k1 k2^4 k3 p0^3 + k2^4 k3 p0^4 + 2 k1^2 k2^6 p1l
+ 6 k1^2 k2^8 p0^2 p1l + 6 k1^2 k2^4 p0^4 p1l + 2 k1^2 p0^6 p1l) == 3 k1^2 k2^4 k3 l0 p0^2
```

The equation can be solved for p_1l with

```
Solve[p1l (k1^3 k2^4 k3 p0 + 3 k1^2 k2^4 k3 p0^2 + 3 k1 k2^4 k3 p0^3 + k2^4 k3 p0^4 + 2 k1^2
k2^6 p1l + 6 k1^2 k2^4 p0^2 p1l + 6 k1^2 k2^2 p0^4 p1l + 2 k1^2 p0^6 p1l) == 3 k1^2 k2^4 k3 l0
p0^2, {p1l}]
```

resulting in a solution that describes the concentration of p_il as a function of s=3,p_0, l_0, K_1, K_2:

```
p1l = (k1^3 k2^4 k3 p0 - 3 k1^2 k2^4 k3 p0^2 - 3 k1 k2^4 k3 p0^3 - k2^4 k3 p0^4 +
Sqrt((k1^3 k2^4 k3 p0 + 3 k1^2 k2^4 k3 p0^2 + 3 k1 k2^4 k3 p0^3 + k2^4 k3 p0^4)^2 + 12 k1^2 k2^4
k3 l0 p0^2 (2 k1^2 k2^6 + 6 k1^2 k2^4 p0^2 + 6 k1^2 k2^2 p0^4 + 2 k1^2 p0^6)))/
(4 (k1^2 k2^6 + 3 k1^2 k2^4 p0^2 + 3 k1^2 k2^2 p0^4 + k1^2 p0^6))

p2l = (-k1^3 k2^4 k3 p0^2 - 3 k1^2 k2^4 k3 p0^3 - 3 k1 k2^4 k3 p0^4 - k2^4 k3 p0^5 +
Sqrt((k1^3 k2^4 k3 p0^2 + 3 k1^2 k2^4 k3 p0^3 + 3 k1 k2^4 k3 p0^4 + k2^4 k3 p0^5)^2 + 12 k1 k2^4
k3 l0 p0^4 (2 k1^3 k2^6 + 6 k1^3 k2^4 p0^2 + 6 k1^3 k2^2 p0^4 + 2 k1^3 p0^6)))/
(4 (k1^3 k2^6 + 3 k1^3 k2^4 p0^2 + 3 k1^3 k2^2 p0^4 + k1^3 p0^6))

p3l = (-k1^3 k2^4 k3 p0^3 - 3 k1^2 k2^4 k3 p0^4 - 3 k1 k2^4 k3 p0^5 - k2^4 k3 p0^6 +
```

```
Sqrt(4 k2^4 k3 l0 p0^6 (6 k1^4 k2^6 + 18 k1^4 k2^4 p0^2 + 18 k1^4 k2^2 p0^4 + 6 k1^4 p0^6) + (k1^3
k2^4 k3 p0^3 + 3 k1^2 k2^4 k3 p0^4 + 3 k1 k2^4 k3 p0^5 + k2^4 k3 p0^6)^2))/(12 (k1^4 k2^6 + 3 k1^4
k2^4 p0^2 + 3 k1^4 k2^2 p0^4 + k1^4 p0^6))
```

Inserting the three expression into $A_1 = \sum_{i=1}^{s} a_1 * i * p_i l$ (first half of (eq. 2.1)), one can formulate the following sum (s=3):

$$A_1 = a_1 * f_1 * \sum_{i=1}^{3} \binom{3}{i} \frac{i p_0^i}{K_1^{i+1}}$$

with (3.1)

$$f_1 = \left(\frac{-K_2^4 K_3 (K_1 + p_0)^3 + \sqrt{K_2^4 K_3} \sqrt{K_2^4 K_3 (K_1 + p_0)^6 + 24 K_1^4 l_0 (K_2^2 + p_0^2)^3}}{12 (K_2^2 + p_0^2)^3} \right)$$

Generalizing s results in the following expression:

$$A_1 = a_1 * f_1 * \sum_{i=1}^{s} \binom{s}{i} \frac{i p_0^i}{K_1^{i+s-2}}$$

with (3.2)

$$f_1 = \left(\frac{-K_2^{2s-2} K_3 (K_1 + p_0)^s + \sqrt{K_2^{2s-2} K_3} \sqrt{K_2^{2s-2} K_3 (K_1 + p_0)^{2s} + 8 s K_1^{4s-4} l_0 (K_2^2 + p_0^2)^s}}{4 s (K_2^2 + p_0^2)^s} \right)$$

The summation has an explicit expression:

$$\sum_{i=1}^{s} \binom{s}{i} \frac{i p_0^i}{K_1^{i+s-2}} = K_1^2 s \frac{p_0}{K_1 + p_0} \left(\frac{K_1 + p_0}{K_1^2} \right)^s$$

With this expression and reformulation, 3.2 can be written as:

(3.3)

$$A_1 = -\frac{a_1}{4} \frac{p_0}{K_1 + p_0} * \left(\frac{K_1^2 K_3}{K_2^2} \left(\frac{K_2^2 (K_1 + p_0)^2}{K_1^2 (K_2^2 + p_0^2)} \right)^s - \sqrt{\frac{K_1^2 K_3}{K_2^2} \left(\frac{K_2^2 (K_1 + p_0)^2}{K_1^2 (K_2^2 + p_0^2)} \right)^s} \sqrt{\frac{K_1^2 K_3}{K_2^2} \left(\frac{K_2^2 (K_1 + p_0)^2}{K_1^2 (K_2^2 + p_0^2)} \right)^s + 8 s l_0} \right)$$

Note that this expression has the general form of (1.4). Using the same procedure, one obtains an expression for A_2:

(3.4)

$$A_2 = \frac{a_2}{4} \frac{p_0^2}{K_2^2 + p_0^2} * \left(\frac{K_1^2 K_3}{K_2^2} \left(\frac{K_2^2 (K_1 + p_0)^2}{K_1^2 (K_2^2 + p_0^2)} \right)^s + 4 s l_0 - \sqrt{\frac{K_1^2 K_3}{K_2^2} \left(\frac{K_2^2 (K_1 + p_0)^2}{K_1^2 (K_2^2 + p_0^2)} \right)^s} \sqrt{\frac{K_1^2 K_3}{K_2^2} \left(\frac{K_2^2 (K_1 + p_0)^2}{K_1^2 (K_2^2 + p_0^2)} \right)^s + 8 s l_0} \right)$$

In our experiments we study the activation of cGAS as function of DNA length and keep the effective concentration of binding sites $l_{eff} = l_0 * s$ constant. Combining A_1 and A_2 results in the final formula for the concentration of cGAS bound to DNA and hence steady state activity **A** of product formation as a function of s:

$$A(s) = aK'(s) + bl_{\text{eff}} - a\sqrt{K'(s)}\sqrt{K'(s) + 2l_{\text{eff}}} \tag{3.5}$$

with

$$l_{eff} = sl_0$$
$$K'(s) = \frac{K_1^2 K_3}{4K_2^2}\left(\frac{K_2^2(K_1 + p_0)^2}{K_1^2(K_2^2 + p_0^2)}\right)^s$$
$$a = \frac{a_2 p_0^2}{K_2^2 + p_0^2} - \frac{a_1 p_0}{K_1 + p_0}$$
$$b = \frac{a_2 p_0^2}{K_2^2 + p_0^2}$$

3.4 DPL Model: Special Cases and Simplifications

a) $K_3 \rightarrow \infty$ (no "dimer" state)

A Taylor series expansion of (3.5) around l_{eff} results around 0 yields

$$A(s) = (-a + b)l_{\text{eff}} + \frac{a}{2K'} l_{\text{eff}}^2 - \frac{a}{2K'^2} l_{\text{eff}}^3 \cdots$$

hence for $K_3 \rightarrow \infty$ (therefore $K' \rightarrow \infty$),

$$A(s) = (-a + b)l_{\text{eff}} = a_1 \frac{l_{\text{eff}}\, p_0}{K_1 + p_0}$$

which describes bind of cGAS monomers on a single DNA ligand with s binding sites and concentration l_0.

b) $K_3 \rightarrow 0$ (no "monomer" state)

Setting $K_3 = 0$ in (3.5) yields

$$A(s) = bl_{\text{eff}} = a_2 \frac{l_{\text{eff}}\, p_0^2}{K_2^2 + p_0^2}$$

with describes a (maximally) cooperative formation of cGAS dimers on (prearranged) DNA with s binding sites and a concentration of l_0. Binding along DNA of cGAS dimers is non-cooperative.

c) $K_1 \gg p_0$

This condition is a reasonable assumption for our studies since the experimentally determined DNA binding affinity for short DNA is only ~20 µM. In this case (3.5) simplifies to:

$$A(s) = a\left(K'(s) + l_{\text{eff}} - \sqrt{K'(s)}\sqrt{K'(s) + 2l_{\text{eff}}}\right) \tag{3.6}$$

with

$$l_{eff} = sl_0$$
$$K'(s) = K_1'\left(\frac{K_2^2}{K_2^2 + p_0^2}\right)^s$$
$$K_1' = \frac{K_1^2 K_3}{4K_2^2}$$
$$a = a_2 \frac{p_0^2}{K_2^2 + p_0^2}$$

Note, in this case, there are only two independent binding constants K_2 and $K_1'=K_1^2K_3/4K_2^2$. Eq. 3.6 is the one used to fit the experimental data in our study.

4 Fitting of experimental data

For global fitting of the data matrix (8 different DNA ligands by 8 different cGAS concentrations), we used the *fminsearch* procedure as implemented in Matlab_R2015a (The MathWorks, Inc).

4.1 Fitting with Hill equations

First, the data were fitted using a set of empirical Hill equations

$$V_{Hill}(V_{max}, s_i, K_i) = V_{max}\frac{l_{\mathrm{eff,i}}\ p_0^{s_i}}{K_i + p_0^{s_i}} \tag{4.1}$$

with V_{max} fitted globally (i.e. it is assumed to be the same for all ligands) and s_i, K_i fitted for each DNA ligand l_i individually. Here, $l_{eff,i}$ is the concentration of DNA ligand i times its number of bps and $l_{eff,i}$=constant in all reactions. *fminsearch* was used to minimize the following function (i ligands, j protein concentrations, r_{ij}: experimentally measured rate for ligand i and protein concentration j, w_{ij}: standard deviation of r_{ij} from three independent experiments):

$$Res = \sum_{i,j}\frac{1}{w_{ij}}\left(V_{Hill}(V_{max}, s_i, K_i) - r_{i,j}\right)^2$$

Minimization of 17 parameters (V_{max}, 8 s_i, 8 K_i) resulted in an R^2=0.991. Fig. S1 (left panel) shows the experimental data along with the fit of the set of Hill equations, the right panel is a plot of the "Hill coefficients" s_i as a function of DNA length. In general, the data show an increase in "cooperativity" for longer DNA as expected.

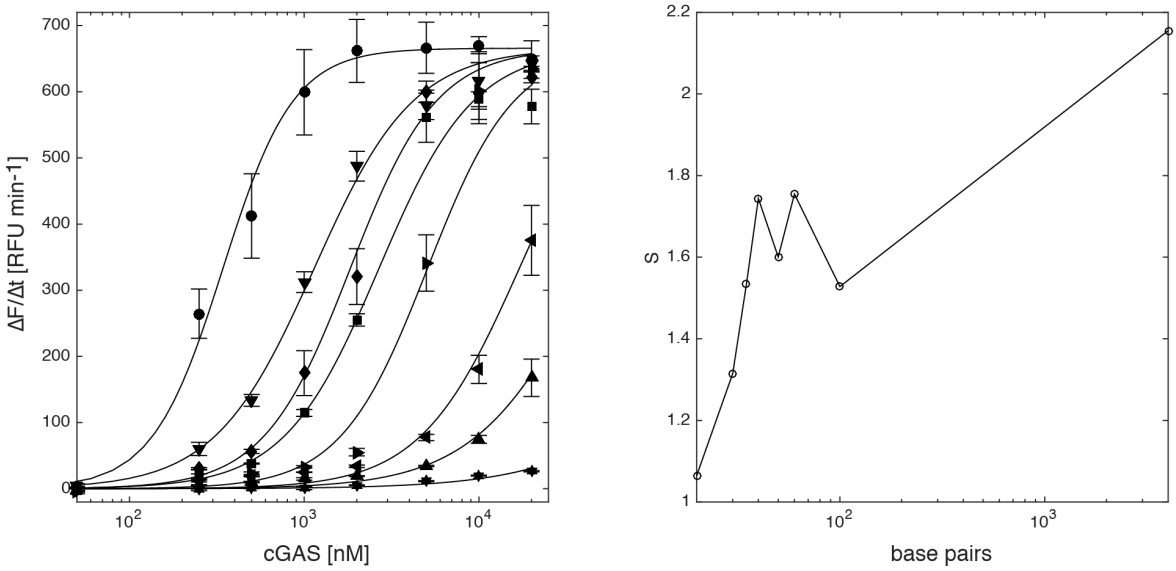


Figure S1. Left panel: Experimental data (see also legend for Fig. 3c) fitted to a set of Hill equations with a global V_{max}, and individual K_i,s_i. Plotted is the rate of substrate turnover (ΔF/Δt [RFU min^{-1}]) as a

function of cGAS concentration. Right panel: Plot of s_i as a function of DNA length (number of base pairs). Longer DNA leads to an increase in s_i, which can be interpreted as an increase in cooperative binding.

The empirical Hill equation, although being able to generally fit the data with an R_2=0.991, has no direct physical interpretation, because it is unclear how K_i and s_i correlate with the underlying molecular events.

4.2 Fitting with DPL equation

Using the same procedure, we fitted the experimental data with the simplified DPL model (eq. 3.6) through least square minimization and the *fminsearch* procedure (Fig. S2):

$$Res = \sum_{i,j} \frac{1}{w_{ij}} \left(V_{DPL}(V_{max}, s_i, K_1', K_2) - r_{i,j} \right)^2$$

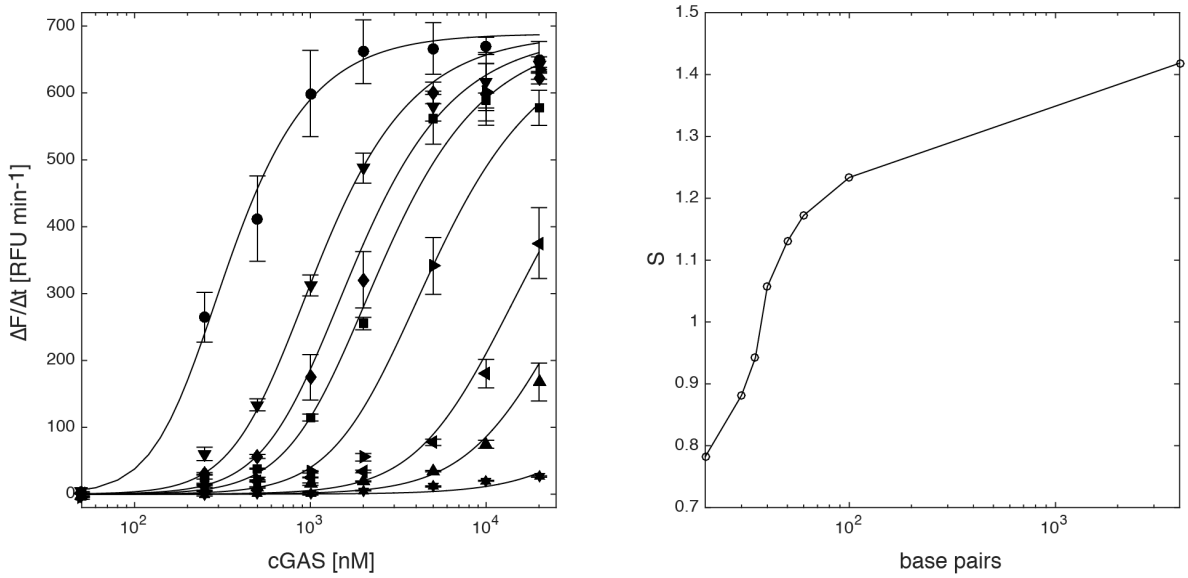


Figure S2. Left panel: Experimental data (see also legend for Fig. 3c) fitted to the DPL equation (3.6) with a global V_{max}, K_1' and K_2, and individual s_i. Plotted is the rate of substrate turnover (ΔF/Δt [RFU min^{-1}]) as a function of cGAS concentration. Right panel: Plot of s_i as a function of DNA length (number of base pairs). Longer DNA leads to an increase in s_i, which can be interpreted as an increase in cooperative binding.

Although here only 11 (V_{max}, 8 s_i, K_1', K_2) compared to the 17 parameters of the Hill equations (V_{max}, 8 s_i, 8 K_i) are optimized, the resulting R_2=0.988 is very close to that obtained using the Hill equations. In the DPL model, K_1' and K_2 have a physical interpretation in the assembly of the DNA-protein ladder (see above). s corresponds to effective number of cooperative binding sites along the ladder. It should be noted, that p_0 appears in the equation already as quadratic form p_0^2, since cGAS is assumed to bind as dimer, so the overall "cooperativity" with respect to cGAS concentration is 2s, the "cooperativity" between adjacent cGAS dimers is s. In the DPL fit, s smoothly increases with increasing DNA length, showing a steeper transition around 40-50 base pairs, corresponding well to the observed in vivo threshold (Fig. 1a). For longer DNA, s is 1.2-1.4, but not higher. Thus, per this modeling, cooperative binding of two cGAS dimers along DNA (short ladder) would be in principle sufficient to account for the cooperativity. The yet higher

activity of very long DNA could originate from more efficient assembly of cGAS dimer in cis because long DNA could easily bend back and thus can help assemble cGAS dimers at low DNA concentrations.

4.3 Titration with inactive cGAS D307N

In the titration experiment, an increasing amount of $cGAS^{cd}$ (D307N), i.e. a mutant that does not turn over substrate GTP and ATP, is titrated into a solution of a fixed amount of $cGAS^{cd}$ and a fixed amount of DNA. The ladder model predicts a hill or bell shaped curve. In the absence of $cGAS^{cd}$ (D307N), the low amount of $cGAS^{cd}$ results in low fGAMP production, because most if not all of $cGAS^{cd}$ is not bound to DNA in catalytically active $cGAS_n$:DNA_2 complexes. Titrating in $cGAS^{cd}$ (D307N) will cooperatively promote ladder formation and therefore help trap $cGAS^{cd}$ in $cGAS_n$:DNA_2 ladders, therefore increasing activity. Increasing amounts of $cGAS^{cd}$ (D307N), however, will more and more compete $cGAS^{cd}$ away from DNA:protein ladders due to the limiting number of protein binding sites on DNA. Thus, after a maximum stimulation, further increase of $cGAS^{cd}$ (D307N) will result in a gradual reduction of the observed rate of fGAMP production.

The observed activity is as follows:

$$A = \frac{p_a}{p_{tot}} * f(p_{tot})$$

Here, p_a is the concentration of catalytic active cGAS and p_i the concentration of catalytic inactive cGAS (D307N). $p_{tot}=p_a+p_i$ is the total concentration of cGAS molecules, both active and inactive. f is the DPL function (3.5 or 3.6) or e.g. an empirical Hill function (4.1). Fitting was done as described for 4.2 using the *fminsearch* function as implemented in Matlab. For fitting, K1' was set fixed to the value obtained from the fit in Fig. S2 (to reduce the number of free paramteter), all other parameters (Vmax, K2 and s) were kept free and optimized. It should be noted that our premise that $p_0>>s*l_0$, i.e. $p\approx p_0$ is not exactly true for the experimental conditions used in Fig. 3d, so both a Hill equation and the DPL equation are only approximations. Nevertheless, we obtain a good fit. The important point here is that a good fit of the data lead to s=1.8. Since the protein concentration scales with p^{2s} in the DPL model, 2*1.8=3.6 indicates substantial cooperativity for the stimulation of active cGAS by titrating in inactive cGAS.

Supplementary Table 1 Protein constructs and their descriptions

Construct Name	Protein Name	Source Organism	Fragment, aa	Modification
mcGAS^{cd}	Cyclic GMP-AMP synthase (cGAS)	*Mus musculus*	141-507	-
hcGAS^{cd}	Cyclic GMP-AMP synthase (cGAS)	*Homo sapiens*	155-522	-
hcGAS	Cyclic GMP-AMP synthase (cGAS)	*Homo sapiens*	1-522	N-terminal His_6-MBP-tag
Flag/HA-hcGAS	Cyclic GMP-AMP synthase (cGAS)	*Homo sapiens*	1-522	N-terminal Flag/HA-tag
eGFP-hcGAS	Cyclic GMP-AMP synthase (cGAS)	*Homo sapiens*	1-522	N-terminal eGFP-tag
mTFAM	Transcription factor A, mitochondrial (TFAM)	*Mus musculus*	43-243	N-terminal His_6-tag
hTFAM	Transcription factor A, mitochondrial (TFAM)	*Homo sapiens*	43-246	N-terminal His_6-tag
HA-TFAM	Transcription factor A, mitochondrial (TFAM)	*Homo sapiens*	43-246	N-terminal HA-tag
mHMGB1	High mobility group protein B1 (HMGB1)	*Mus musculus*	1-215	N-terminal His_6-tag
mHMGB1dCTT	High mobility group protein B1 (HMGB1)	*Mus musculus*	1- 185	N-terminal His_6-tag
lHU	DNA-binding protein HU	*Listeria monocytogenes*	1-121	N-terminal His_6-tag

Supplementary Table 2 Stimulatory DNA sequences

Construct Name	Sequence
20 bp-s	CTACTAGTGATCTATGACTG
20 bp-as	CAGTCATAGATCACTAGTAG
25 bp-s	CTACTAGTGATCTATGACTGATCTG
25 bp-as	CAGATCAGTCATAGATCACTAGTAG
30 bp-s	CTACTAGTGATCTATGACTGATCTGTACAG
30 bp-as	CTGTACAGATCAGTCATAGATCACTAGTAG
35 bp-s	ATCTACTAGTGATCTATGACTGATCTGTACATGAT
35 bp-as	ATCATGTACAGATCAGTCATAGATCACTAGTAGAT
40 bp-s	AGTGTCTACTAGTGATCTATGACTGATCTGTACATGATCT
40 bp-as	AGATCATGTACAGATCAGTCATAGATCACTAGTAGACACT
45 bp-s	TACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTACA
45 bp-as	TGTAGATCATGTACAGATCAGTCATAGATCACTAGTAGATCTGTA
50 bp-s	GATACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTACA ATC
50 bp-as	GATTGTAGATCATGTACAGATCAGTCATAGATCACTAGTAGATCTGT ATC
55 bp-s	TCGATACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTA CAATCACT
55 bp-as	AGTGATTGTAGATCATGTACAGATCAGTCATAGATCACTAGTAGATC TGTATCGA
60 bp-s	AGTCGATACAGATCTACTAGTGATCTATGACTGATCTGTACATGATC TACAATCACTGCA
60 bp-as	TGCAGTGATTGTAGATCATGTACAGATCAGTCATAGATCACTAGTA GATCTGTATCGACT
65 bp-s	CCAAGTCGATACAGATCTACTAGTGATCTATGACTGATCTGTACAT GATCTACAATCACTGCAGT
65 bp-as	ACTGCAGTGATTGTAGATCATGTACAGATCAGTCATAGATCACTAG TAGATCTGTATCGACTTGG
70 bp-s	GACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTGATCTG TACATGATCTACAATCACTGCAGT
70 bp-as	ACTGCAGTGATTGTAGATCATGTACAGATCAGTCATAGATCACTAG TAGATCTGTATCGACTTGGTAGTC
75 bp-s	GACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTGATCTG TACATGATCTACAATCACTGCAGTTACCG
75 bp-as	CGGTAACTGCAGTGATTGTAGATCATGTACAGATCAGTCATAGATC ACTAGTAGATCTGTATCGACTTGGTAGTC
80 bp-s	GACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTGATCTG TACATGATCTACAATCACTGCAGTTACCGTGACC
80 bp-as	GGTCACGGTAACTGCAGTGATTGTAGATCATGTACAGATCAGTCAT AGATCACTAGTAGATCTGTATCGACTTGGTAGTC
85 bp-s	TCCTAGACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTG ATCTGTACATGATCTACAATCACTGCAGTTACCGTGACC
85 bp-as	GGTCACGGTAACTGCAGTGATTGTAGATCATGTACAGATCAGTCAT AGATCACTAGTAGATCTGTATCGACTTGGTAGTCTAGGA
90 bp-s	TCCTAGACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTG ATCTGTACATGATCTACAATCACTGCAGTTACCGTGACCAATGT

90 bp-as	ACATTGGTCACGGTAACTGCAGTGATTGTAGATCATGTACAGATCA GTCATAGATCACTAGTAGATCTGTATCGACTTGGTAGTCTAGGA
95 bp-s	TCCTAGACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTG ATCTGTACATGATCTACAATCACTGCAGTTACCGTGACCAATGTCG ACT
95 bp-as	AGTCGACATTGGTCACGGTAACTGCAGTGATTGTAGATCATGTACA GATCAGTCATAGATCACTAGTAGATCTGTATCGACTTGGTAGTCTA GGA
100 bp-s	TCCTAGACTACCAAGTCGATACAGATCTACTAGTGATCTATGACTG ATCTGTACATGATCTACAATCACTGCAGTTACCGTGACCAATGTCG ACTGGATC
100 bp-as	GATCCAGTCGACATTGGTCACGGTAACTGCAGTGATTGTAGATCAT GTACAGATCAGTCATAGATCACTAGTAGATCTGTATCGACTTGGTA GTCTAGGA
200 bp-s	ATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACT TTAAGAAGGAGATATACATATGTCGTACTACCATCACCATCACCATC ACGATTACATGATCGAAGAAGGTAAACTGGTAATCTGGATTAACGG CGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAG AAAGATACCGGAAT
200 bp-as	ATTCCGGTATCTTTCTCGAATTTCTTACCGACTTCAGCGAGACCGTT ATAGCCTTTATCGCCGTTAATCCAGATTACCAGTTTACCTTCTTCGA TCATGTAATCGTGATGGTGATGGTGATGGTAGTACGACATATGTAT ATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTT ATCCGCTCACAAT

Supplementary Table 3 Data processing and refinement statistics

Space group	**C2 (No. 5)**	**C2 (No. 5)**
Unit cell dimensions a, b, c, α, β, γ	168.5 Å, 122.9 Å, 180.0 Å, 90º, 96.4º, 90º	168.5 Å, 122.9 Å, 180.0 Å, 90º, 96.4º, 90º
	XDS	STARANISO
Resolution range[a]	50 – 3.6 (4.2) Å	50 – 3.6 (4.2) Å
No. of observed reflections	212879	–
No. of unique reflections	41980	28400
Completeness[b]	0.99	0.67 / 0.91
I/σ(I)[c]	9.4	9.3 / 12.6
R_{sym}	22.4% (91.0%)	–
No. of protein atoms		17712
No. of DNA atoms		4370
No. of zinc ions		6
R-factor / Free-R-factor		0.204 / 0.256
Rmsd bond lengths / bond angles		0.010 Å / 1.4º
Ramachandran plot preferred / allowed / outliers		94.9% / 3.1% / 2.0%

[a]Approximate effective resolution in parentheses

[b]Resulting completeness after STARANISO for spherical / elliptical shells

[c]Resulting signal-to-noise ratio after STARANISO for spherical / elliptical shells

4. A fluorescent cyclic dinucleotide and its use in methods of identifying substances having an ability to modulate the cGAS/STING pathway (patent application)

Hopfner, K. P.; Andreeva, L.; Drexler, D. J.; **2019**. A fluorescent cyclic dinucleotide and its use in methods of identifying substances having an ability to modulate the cGAS/STING pathway. European Patent Application EP 17182689.4.

This work provides fluorescence-based assays for analysis of the cGAS-STING pathway using a fluorescent analogue of ATP. The cGAS-STING pathway is an important component of the innate immune system. As it was discovered to have a high potential for immunotherapy, a great interest for drug development has raised. The results of this work include a fluorescence-based assay for quantitative analysis of cGAS activity. This assay uses the fluorescent ATP analogue 2-aminopurine riboside-5′-triphosphate, which is accepted by cGAS. The resulting compound, termed fGAMP, is less fluorescent, allowing a high-throughput real time detection of cGAS activity. In addition, fGAMP was identified as suitable compound for fluorescent STING binding assays. Upon STING binding, the fluorescence of fGAMP is quenched. Subsequently, the flurorescence can be restored by adding STING binding molecules, which correlates with the binding affinity. Major advantages of the provided methods are a simple experimental set-up and readout, high-throughput feasibility and similar properties to the physiological substrates. In particular, fGAMP was shown to bind STING and induce a type I interferon response in HEK293T cells. In conclusion, this work describes methods for drug discovery for the cGAS-STING pathway, which are suitable for industrial applications.

Author contribution

The author of the present thesis analyzed cGAS reaction products. He identified, purified and characterized fGAMP by NMR, mass-spectrometry and ion exchange chromatography. Furthermore, he analyzed the binding affinity of fGAMP to STING by isothermal titration calorimetry. Together with L. Andreeva, he described the fluorescent properties of fGAMP and tested STING binding effects using a fluorescence plate reader. In addition, he performed luciferase reporter assays to analyze STING activation in HEK293T cells, assisted by C. Lässig. He assisted L. Andreeva and K.P. Hopfner in writing the patent application.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 431 484 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
23.01.2019 Bulletin 2019/04

(51) Int Cl.:
C07H 21/02 *(2006.01)* ***C07H 21/04*** *(2006.01)*
G01N 33/52 *(2006.01)*

(21) Application number: **17182689.4**

(22) Date of filing: **21.07.2017**

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
MA MD

(71) Applicant: **Ludwig-Maximilians-Universität München**
80539 München (DE)

(72) Inventors:
- **HOPFNER, Karl-Peter**
  81377 München (DE)
- **ANDREEVA, Liudmila**
  81377 München (DE)
- **DREXLER, David**
  81377 München (DE)

(74) Representative: **Grünecker Patent- und Rechtsanwälte**
PartG mbB
Leopoldstraße 4
80802 München (DE)

(54) A FLUORESCENT CYCLIC DINUCLEOTIDE AND ITS USE IN METHODS OF IDENTIFYING SUBSTANCES HAVING AN ABILITY TO MODULATE THE CGAS/STING PATHWAY

(57) The present invention relates to a non-canonical cyclic di-nucleotide containing an asymmetric 2'-5' and 3'-5' phosphodiester linkage and its use in a method for identifying a substance having an ability to modulate cGAS activity or bind to STING. Furthermore, the cyclic dinucleotide is used in a method of identifying a substance having an ability to modulate the activity of a 2'-5' phosphodiesterase.

EP 3 431 484 A1

Description

Field of the invention

[0001] The present invention relates to a non-canonical cyclic di-nucleotide containing an asymmetric 2'-5' and 3'-5' phosphodiester linkage and its use in a method for identifying a substance having an ability to modulate cGAS activity or bind to STING. Furthermore, the cyclic dinucleotide is used in a method of identifying a substance having an ability to modulate the activity of a 2'-5' phosphodiesterase.

Background of the invention

[0002] The innate immune system relies on germline-encoded recognition receptors (PRRs) to survey the extracellular, endosomal, and cytosolic milieu for signs of pathogen invasion or cellular damage. Following ligand binding, PRRs trigger downstream signaling to elicit innate immune responses and to activate the adaptive immune system. Since nucleic acids are central to the replication and propagation of most if not all pathogens, the detection of aberrant DNA and RNA has evolved as a fundamental mechanism of host defense.

[0003] DNA is usually confined within the nucleus and mitochondria of eukaryotic cells. The presence of cytosolic DNA either through infections or cellular damage triggers robust immune responses including type I interferon induction. The cyclic GMP-AMP synthase (cGAS) is a key sensor required for DNA recognition in the cytosol. Beside endogenous DNA, cGAS senses DNA from cytosolic bacteria, HIV (cDNA) and herpes viruses. In its inactive state cGAS exists in a monomeric apo-form. Upon recognition of cytosolic DNA, cGAS undergoes a conformational change resulting in the formation of an active $cGAS_2$:DNA_2 tetramer. Activated cGAS synthesizes cyclic dinucleotides comprising a conventional 3'-5' phosphodiester linkage and an unconventional 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate (2'3' cyclic dinucleotide CDNs). These CDNs serve as second messengers which then bind to the transmembrane adaptor protein stimulator of interferon genes (STING) associated with the endoplasmic reticulum. Upon binding to STING the adapter protein is activated, interacts with TBK 1 (TANK-binding kinase 1) and traffics through the Golgi to perinuclear endosomal compartments, where TBK 1 phosphorylates IRF 3 and IRF 7 to trigger type I interferon production.

[0004] While under normal conditions DNA sensing by the cGAS/STING axis protects the organism from harmful infections by DNA viruses, retroviruses and bacteria, unsolicited self-DNA sensing can result in autoimmune diseases such as systemic lupus erythematosus (SLE) and Aicardi-Goutières syndrome (AGS). AGS is a rare but generally fatal childhood inflammatory condition with neurological dysfunction which is caused by loss-of-function mutations in nucleic acid metabolizing enzymes resulting in accumulation of endogenous DNA in the cytosol.

[0005] Many genes have been identified as being involved in the development of autoimmune diseases. For instance, TREX1 encodes a 3'-5' DNA exonuclease, and it has been found that a loss of function mutation in this gene results in AGS and SLE (Crow, Hayward et al. "Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus". Nat Genet 38, 917-920, 2006; Lee-Kirsch, Gong et al. "Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 are associated with systemic lupus erythematosus". Nat Genet 39, 1065-1067, 2007). Interestingly, crossing TREX1-deficient mice with mice lacking either STING or the IFNα/β receptor resulted in a rescue of the autoinflammatory phenotypes (Gall, Treuting et al. "Autoimmunity Initiates in Nonhematopoietic Cells and Progresses via Lymphocytes in an Interferon-Dependent Autoimmune Disease". Immunity 36, 120-131, 2012; Ahn, Ruiz et al. "Intrinsic Self-DNA Triggers Inflammatory Disease Dependent on STING". The Journal of Immunology 2014), placing STING signaling central to these interferonopathies. Mutations in RNASEH2A, RNASH2B and RNASEH2C have also been linked to AGS (Crow, Hayward et al. "Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus". Nat Genet 38, 917-920, 2006; Crow, Chase et al. "Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1". American journal of medical genetics. Part A 0, 296-312, 2015). The proteins encoded by these genes form the RNase H2 complex and function to degrade cellular RNA/DNA hybrids. In a mouse model of the disease due to a homozygous A174T knock-in-mutation, the cGAS/STING pathway was central to the autoinflammatory pathology, since the loss of STING in these mice reduced levels of inflammatory cytokines. Taken together, there is increasing evidence that STING plays a central role in autoimmune interferonopathies.

[0006] Several studies indicate that cGAS activation by cytosolic self-DNA is also responsible for the lethal phenotypes associated with SLE and ALS (Gao, Li et al. "Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases". Proceedings of the National Academy of Sciences 112, E5699-E5705, 2015). Overall, these studies provide evidence that the inhibition of the cGAS/STING axis may be an effective therapeutic tool for the treatment of autoimmune diseases such as AGS and SLE.

[0007] Chronic inflammatory signaling may not only initiate autoinflammatory diseases but can also contribute to the

development of cancer, probably through cytokines, chemokines and growth factors that stimulate cellular proliferation and survival, as well as by promoting angiogenesis (Nowarski, Gagliani et al. "Innate Immune Cells in Inflammation and Cancer". Cancer Immunology Research 1, 77-84, 2013). Adaptive T cell responses are important for the control and eradication of tumor cells. Numerous immunotherapeutic strategies involving stimulating the adaptive immune response against cancers through checkpoint blockade are presently under evaluation in clinical trials. It has been shown that type I interferon production is involved in tumor antigen specific T cell activation and that the cGAS/STING pathway plays a key role in stimulating adaptive T cell responses against tumor cells (Woo, Fuertes et al. "STING-Dependent Cytosolic DNA Sensing Mediates Innate Immune Recognition of Immunogenic Tumors". Immunity 41, 830-842, 2014). Overall, the targeted inhibition or activation of the cGAS/STING pathway by small molecule compounds or therapeutic nucleic acids provide a promising tool for the pharmaceutical industry to develop strategies for the treatment of cancer, inflammatory diseases, and autoimmune disorders. In view of the numerous promising therapeutic approaches to modulate the cGAS/STING pathway intensive efforts are currently underway to identify small molecule compounds which are capable to modulate and fine-tune the innate immune system via the cGAS/STING pathway. However, large-scale screening assays for the identification of such modulators are currently hampered by the general lack of sensitive and scalable assays that allow a direct readout of the activity. Measuring the activity of cGAS is usually done via thin-layer chromatography of the produced 2'3'CDNs, using radioactive nucleoside triphosphates as substrates, which is impractical in high-throughput assays.

[0008] Alternatively, 2'3'CDN can be directly detected by mass spectrometry. The latter approach is sensitive and to some extent quantitative, however, it is difficult to measure reactions in parallel at high-throughput or at different time points to analyze steady-state rates. One approach of detecting CDNs is described by Roembke et al. (Roembke, Zhou et al. "A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3',3'-cGAMP". Molecular BioSystems 10, 1568-1575, 2014). In this article the authors suggest using the properties of fluorescent 3'3'-cGAMP analogue - 3'3'-cG(d2AP)MP. The method is based on Mn^{2+}-induced 3'3'-cG(d2AP)MP association with c-diGMP or 3'3'-cGAMP resulting in 3'3'-cG(d2AP)MP quenching. However, the method failed to detect 2'3'-cGAMP and c-di-AMP due to the lack of its association with 3'3'-cG(d2AP)MP. Hence, direct determination of human STING activating CDNs produced by cGAS was not successful with this assay. Furthermore, there is currently no robust and simple-to-implement high-throughput assay that directly measures ligand binding to STING in a quantitative manner, because STING is a receptor protein and not an enzyme. Very elaborate FRET 3'3'-cG(d2AP)MP (Roembke, Zhou et al. "A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3',3'-cGAMP". Molecular BioSystems 10, 1568-1575, 2014) methods have been proposed to measure the conformational change of STING, but these require very specialized equipment and also require modification of STING itself.

[0009] In view of all of the above, there is still a need for a sensitive and scalable biochemical assay that allows a direct readout of cGAS/STING activity, thereby providing a tool for identifying, on industrial scale, small molecule compounds that could modulate the cGAS/STING pathway and which could therefore be promising candidates in anti-inflammatory, anti-autoimmune, and anti-cancer therapy. The object of the present invention is to provide such means.

Summary of the invention

[0010] The present invention relates to a fluorescent cyclic dinucleotide and its preparation. Furthermore, the present invention relates to a method for measuring cGAS activity. The present invention further relates to a method of identifying a substance having an ability to modulate cGAS activity. The present invention also relates to a method of identifying a substance having an ability to bind to STING. The present invention further relates to a method of indentifying a substance having an ability to modulate the activity of a 2'-5' phosphodiesterase. Finally, the present invention relates to a fluorescent cyclic dinucleotide and its use to label cGAS and STING and/or to identify substances having an ability to bind to STING and to complexes of the fluorescent cyclic dinucleotide with cGAS and STING.

[0011] According to one aspect, the invention relates to a cyclic dinucleotide comprising a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate.

[0012] In a further aspect, the invention relates to a method of measuring cGAS activity, wherein the method comprises the steps of: (i) providing an aqueous solution comprising cGAS, cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group; and (ii) measuring the change of the fluorescence signal of the aqueous solution over time.

[0013] In another aspect, the invention provides a method of identifying a substance having an ability to modulate the activity of cyclic GMP-AMP synthase (cGAS), wherein the method comprises the steps of: (i) providing an aqueous solution comprising cGAS, cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a

fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group; (ii) measuring the fluorescence signal of the aqueous solution; (iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence signal is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and (iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance; wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance is a cGAS antagonist, while a measured fluorescence signal which is lower in the presence of the substance than in the absence of the substance indicates that the substance is a cGAS agonist.

[0014] A further aspect of the invention relates to a method of preparing a cyclic dinucleotide, wherein the method comprises the steps of (i) providing an aqueous solution comprising cGAS, cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group, thereby preparing the cyclic dinucleotide; and, optionally, (ii) purifying the cyclic dicnucleotide.

[0015] Another aspect of the invention relates to a method of identifying a substance having an ability to bind to stimulator of interferon genes (STING), wherein the method comprises the steps of (i) providing an aqueous solution comprising a cyclic dinucleotide and STING, wherein the cyclic dinucleotide comprises a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate; (ii) measuring the fluorescence signal of the aqueous solution; (iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and (iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance; wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance has an ability to bind to STING.

[0016] According to a further aspect, the invention provides a method of identifying a substance having an ability to modulate the activity of a 2'-5' phosphodiesterase enzyme, wherein the method comprises the steps of (i) providing an aqueous solution comprising a cyclic dinucleotide and a 2'-5' phosphodiesterase, wherein the cyclic dinucleotide comprises a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate; (ii) measuring the fluorescence signal of the aqueous solution; (iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and (iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance; wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance is a 2'-5' phosphodiesterase agonist, while a measured fluorescence signal which is lower in the presence of the substance than in the absence of the substance indicates that the substance is a 2'-5' phosphodiesterase antagonist.

[0017] Another aspect of the invention relates to a use of a fluorescent nucleoside triphosphate for measuring cGAS activity.

[0018] A further aspect of the invention provides a use of a fluorescent nucleoside triphosphate for identifying a substance with an ability to modulate cGAS activity.

[0019] According to another aspect, the invention also relates to a use of a cyclic dinucleotide for identifying a substance having an ability to bind to STING.

[0020] A further aspect of the invention relates to a use of a cyclic dinucleotide for labeling cGAS or STING.

[0021] In a final aspect, the invention relates to complexes of cGAS or STING with a cyclic dinucleotide.

[0022] Other embodiments of the invention are set out below.

Brief description of the Figures

[0023]

Figure 1: The cGAS/STING axis. Scheme depicting the cGAS/STING axis of cytosolic DNA sensing. Upon DNA recognition, the cGAS protein synthesizes 2'3' cyclic dinucleotides (2'3'CDN) from nucleoside triphosphates (here shown by the example of 2'3'cGAMP synthesized from GTP and ATP). 2'3'CDN binds to the adaptor molecule STING and induces a conformational change, leading to activation of TBK1 and type I interferon (IFN) production.
Figure 2: Domain structure of STING. STING comprises four N-terminal transmembrane domains (TM 40-135aa),

a central c-di-GMP-binding domain (CBD, 150-345aa) and a C-terminal tail (CTT, 346-379). The C-terminal part CBD+CTT - cytosolic domain (CD) - is responsible for CDN binding and signal transduction.

Figure 3: Domain structure of cGAS. cGAS comprises an unstructured and poorly conserved N-terminus spanning amino acid sequences 1-159 and a highly conserved C-terminus (160-513aa, catalytic domain - CD) including Mab-21 domain (213-513 aa). The conserved CD domain comprises both DNA-binding capacity and enzymatic activity.

Figure 4: Structure of 2'3'cGAMP (a) and 2'3'fGAMP (b). 2'3'fGAMP is synthesized by cGAS using 2-aminopurine riboside triphosphate (2-APTP) and GTP substrates.

Figure 5: cGAS activity increases with DNA length. (a) cGAS activation determined by CXCL 10 cytokine expression in BLaER1 cells. Cells were stimulated with 20, 40 and 60 ng DNA of different length (20-100 bp in 5 bp intervals) and herring testes (HT) DNA, and CXCL 10 concentration in the supernatant was measured by ELISA. First two bars of each series represent unstimulated cells and lipofectamine controls. Shown are mean values ± standard deviation, n=3. (b) and (c) Activity of $hcGAS^{cd}$ (truncated human), hcGAS (full length human) (b) and $mcGAS^{cd}$ (truncated mouse) (c) in the presence of DNA of increasing length (20-100 bp) or plasmid DNA. Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) measured by the rate of 2-APTP incorporation into 2'3'fGAMP are plotted against DNA length ± standard deviation, n=3-5. (d) Radiolabeled cGAMP production of cGAS stimulated with DNA of different lengths (20 bp, 35 bp, 55 bp and plasmid). cGAS reactions in presence of [α^{32}P]ATP were stopped at the indicated time points and radiolabeled compounds (shown with black arrows) were visualized.

Figure 6: Fluorescence-based cGAS activity assays. (a) cGAS radiolabeled products formation with different substrate composition: ATP+GTP, 2-APTP +GTP, GTP in presence of [α^{32}P]GTP; or ATP+GTP, 2-APTP+ATP, ATP in presence of [α^{32}P]ATP. Reactions were stopped at the indicated time points and radiolabeled compounds and cGAS reaction products (shown with black arrows) were visualized. (b) and (c) 2'3'fGAMP mobility in anion-exchange chromatography on MonoQ 5/50 GL column. (b) Comparison of 2'3'fGAMP (straight line) arising from cGAS reaction with 2-APTP and GTP, 2-APTP (dashed line) and GTP (dotted line) mobilities. (c) Comparison of 2'3'fGAMP (straight line), 2'3'-cGAMP (dashed line) and 3'3'-cGAMP (dotted line) mobilities. (d) ESI LC/MS spectrum of enzymatically synthesized fGAMP. Arrows indicate ions corresponding to fGAMP-H (m/z 673.09) and fGAMP-2H+Na^+ (m/z 695.07). (e) Tandem mass spectrometry spectrum derived from a parent fGAMP ion (m/z 673.09). Arrows highlight ion products corresponding to depurination of the dinucleotide: guanine (m/z 522.19) or 2-AP (m/z 538.15). The ratio of these peaks is characteristic for the linkage between 2'-OH of GTP and 5'-phosphate of 2-APTP within fGAMP molecule (Ablasser, Goldeck et al. "cGAS produces a 2[prime]-5[prime]-linked cyclic dinucleotide second messenger that activates STING". Nature 498, 380-384, 2013).

Figure 7: The principle of fluorescence-based cGAS activity assay. (a) Fluorescence of 2-APTP is quenched upon its incorporation into cyclic 2'3' fluorescent GMP-AMP (2'3'fGAMP) during the reaction that allows the measurement of the reaction rate by differences in fluorescence intensity. (b) Comparison of 2-APTP fluorescence intensity under physiologically relevant conditions: in ultrapure water (1), 100 mM NaCl+5 mM $MgCl_2$ (2), 40 mM Tris pH 7.5 (3), 500 μM GTP (4), 40 mM Tris pH 7.5 + 100 mM NaCl + 5 mM $MgCl_2$ (5), 40 mM Tris pH 7.5 + 100 mM NaCl + 5 mM $MgCl_2$ + 500 μM GTP (6). 2-APTP fluorescence intensity is not altered by cGAS activity assay buffer in comparison to water. (c-e) General workflow for calculating the initial cGAS reaction rates. Initial fluorescence read-out curves (c) represent the changes in fluorescence during the reaction. Next, for each curve the background fluorescence values were subtracted and inverted to give a positive value (ΔF). The resulting curves were scaled, such that ΔF at time point 0 min equals 0 (ΔF-$\Delta F_{t=0}$) (d). The initial cGAS reaction rates were calculated as a slope of the linear intervals (d, dashed lines) on the resulting curves and are defined as $\Delta F/\Delta t$ [RFU min^{-1}] (e).

Figure 8: Reactions of 2-APTP and GTP with different cGAS constructs. (a) Anion exchange chromatography of *Homo sapiens* (straight line), *Mus musculus* (dashed line) and *Sus scrofa* (dotted line) truncated (Mab21 domain) cGAS reactions with 2-APTP and GTP. 2-APTP is incorporated into 2'3'fGAMP in species-unspecific manner. (b), (c) and (d) - Fluorescence quenching assays with truncated mouse cGAS ($mcGAS^{cd}$) (b) and human cGAS ($hcGAS^{cd}$) (c), as well as with full length human cGAS (hcGAS) (d). Bar charts represent mean values of initial steady state activity rates of reactions (calculated as described in Fig.7) plotted against DNA length ± standard deviation, n=3. The assays are suitable for evaluating activity of different cGAS constructs from different species.

Figure 9: DNA-bending proteins boost cGAS activity. (a-d) cGAS activity measured by the rate of 2-APTP incorporation into 2'3'fGAMP in presence of DNA-bending proteins. Mean values of initial cGAS reaction rates ($\Delta F/\Delta t$) are plotted against increasing concentrations of DNA bending proteins ± standard deviation, n=3-8. (a) mTFAM, (b) mHMGB1 and (c) 1HU robustly enhance $mcGAS^{cd}$ activity in a dose dependent manner until they eventually compete cGAS away. (d) hTFAM activates hcGAS.

Figure 10: STING-fGAMP binding. (a) Luciferase activity assays in HEK293T-mSTING flip-in cells (Ablasser, Schmid-Burgk et al. "Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP". Nature 503, 530-534, 2013) upon the induction with fluorescent cGAS product fGAMP. IFN-β response was measured as a proportion of firefly (FF) luciferase activity to Renilla (Ren) luciferase activity. All ratios were normalized to fGAMP buffer control. Plotted values correspond to mean values, error bars represent ± standard deviation, n=3. IFB-β

response increases with fGAMP concentration meaning fGAMP is capable of STING activation *in vivo*. (b) STING-dependent fGAMP quenching assay. Mean values of fluorescence read-outs ± standard deviation (n=3) are plotted against increasing concentration of hSTING cytoplasmic domain ($STING^{cd}$). fGAMP fluorescence decreases with increasing STING concentration suggesting STING-binding induced fGAMP quenching. (c) and (d) Competition assays based on reversal of fGAMP quenching by STING. 2'5'-cGAMP (c, d) or c-di-GMP (d) were titrated into fGAMP-STING complex. Diagrams represent fluorescence read-outs. Maximal fGAMP fluorescence in the absence of STING for each experiment is shown with horizontal lines. The observed differences in maximal fGAMP fluorescence are due to inaccuracy in experimental setup and can be better estimated by higher number of replicates. STING ligands displace fGAMP from STING binding pocket resulting in an increase of fluorescence. High affinity STING ligands (e.g. cGAMP, c and d) replace fGAMP more efficiently and at lower concentrations than weaker binding ligands (e.g. c-di-GMP, d) leading to a bigger slope of fluorescence intensity increase (dotted lines s_1 and s_2, $s_{1\ (c\text{-}di\text{-}GMP)} < S_{2\ (cGAMP)}$). The method offers a quick and high-throughput approach for searching for potent STING activators or inhibitors using the simple fluorescence readout.

General

[0024] It is to be understood that the forgoing general description as well as the following detailed description are exemplary and explanatory only and are not restrictive of the invention as claimed. In this application, the use of the singular may include the plural unless specifically stated otherwise. Further, the use of the term "including" as well as other grammatical forms such as "includes" and "included", is not limiting.

[0025] As used herein the term "comprising" has the broad standard meaning "including", "encompassing", or "containing". It includes the explicitly recited elements, and also allows the presence of other elements not recited. In addition to this broad meaning, as used herein, the term "comprising" also encompasses the limiting meaning "consisting of'. This means that any aspect or embodiment of the present invention that is defined as comprising features, also includes, in its most limited form, the meaning of consisting (only) of said features and no others, whether this is explicitly stated or not. In addition, the term "comprising" also includes the meaning of "consisting essentially of", which means that other elements may be present which do not alter the technical effect achieved by the explicitly recited elements.

[0026] As used herein the term "about" when referring to a particular value such as a wavelength, sequence homology, and the like, is meant to encompass, in addition to the recited value itself, variations of ± 5.0, ± 4.9%, ± 4.8%, ± 4.7%, ± 4.6%, ± 4.5%, ± 4.4%, ± 4.3%, ± 4.2%, ± 4.1%, ± 4.0%, ± 3.9%, ± 3.8%, ± 3.7%, ± 3.6%, ± 3.5%, ± 3.4%, ± 3.3%, ± 3.2%, ± 3.1%, ± 3.0%, ± 2.9%, ± 2.8%, ± 2.7%, ± 2.6%, ± 2.5%, ± 2.4%, ± 2.3%, ± 2.2%, ± 2.1%, ± 2.0%, ± 1.9%, ± 1.8%, ± 1.7%, ± 1.6%, ± 1.5%, ± 1.4%, ± 1.3%, ± 1.2%, ± 1.1%, ± 1.0%, ± 0.9%, ± 0.8%, ± 0.7%, ± 0.6%, ± 0.5%, ± 0.4%, ± 0.3%, ± 0.2%, or ± 0.1% from the specified value. It is to be understood that the term "about", in reference to the particular value, includes that exact particular value itself, irrespective of any explicit mention that this exact particular value is included. Thus, the absence of an explicit indication that the term "about" includes the particular exact recited values is not to be understood that this particular recited values is excluded from the range of variations created by the term "about". Even in the absence of an explicit indication that the term "about" includes the particular exact recited value, this exact particular value is still included in the range of variation created by the term "about". If the term "about" is recited before a numerical range, the term "about" may refer to the lower end point, the upper end point or both the lower end point and the upper end point.

[0027] The skilled person will understand that "measuring a fluorescent signal" as recited in the methods of the invention requires that the parameters (e.g. excitation and emission wavelength or wavelength range) of the fluorescent signal measured should be adapted to each fluorescent nucleoside either in free form or in a form incorporated into a cyclic dinucleotide in order to obtain optimal quantum yields. Different fluorescent nucleosides either in free form or in a form incorporated into a cyclic dinucleotide will have different excitation and emission wavelength optima. This means for example that depending on the fluorescent nucleoside used in the methods of the invention the wavelengths or range of wavelengths at which a fluorescent nucleoside or a cyclic dinucleotide comprising it is excited (i.e. the excitation maxima) and the wavelengths or range of wavelengths at which the fluorescence signal is measured (i.e. the emission maxima) need to be adapted for each fluorescent nucleoside individually in order to avoid false positives by measuring a fluorescence signal outside the optimal wavelength or range of wavelengths for a given fluorescent nucleoside or a cyclic dinucleotide comprising it. The skilled person will know or can readily determine the wavelength or range of wavelengths at which the fluorescence signal for a given fluorescent nucleoside either in free form or in a form incorporated into a cyclic dinucleotide should be measured in order to obtain optimal quantum yields of the fluorescence signal.

[0028] The methods of the invention include the step of measuring the fluorescence signal "of a solution", e.g. of the aqueous solution. By this is meant that the fluorescence signal which is measured is that which arises from the totality of all fluorescing entities present in the respective solution. This fluorescence may for example arise from fluorescent nucleoside triphosphates in free form or fluorescent nucleoside monophosphates already incorporated into a CDN, which itself may either be a free CDN or a CDN bound by STING, depending on the particular embodiment of the invention,

or may arise from a combination of the above. As such, the intensity of the overall "fluorescence signal of the aqueous solution" measured will depend on how much of a respective fluorescent nucleoside is in a form in which the fluorescence arising from the fluorescent nucleoside is quenched, i.e. reduced. The skilled person understands, for instance, that the measured intensity of the fluorescence of an aqueous solution containing only free (unquenched) fluorescent nucleoside will be greater than that of an aqueous solution in which a fraction of the fluorescent nucleoside is present in free (unquenched) form and another fraction of the fluorescent nucleoside is present in a quenched form. In either case, the readout is the overall fluorescence signal which arises as a summation of the fluorescence intensity from all fluorescing entities in the respective solution, whether free, partially quenched or fully quenched.

[0029] The term "identical conditions" as used herein means that the composition of the aqueous solution in the methods of the invention is the same in the presence and absence of a given substance to be tested. This means that the aqueous solution in the presence and absence of the substance has the same concentrations of e.g. proteins (e.g. cGAS, STING and/or 2'-5' phosphodiesterases), nucleoside triphosphates or cGAS-activating nucleic acids (i.e. for methods relating to cGAS), cyclic dinucleotides (i.e. for methods relating to STING and 2'-5'phosphodiesterases), salts, divalent cations, or any other possible solute. Furthermore, the term "identical conditions" as used herein means that the fluorescence signal of a fluorescent nucleoside or a cyclic dinucleotide comprising it is measured at the same wavelength or range of wavelengths in the presence and absence of a substance to be tested. For example, if it is known that the fluorescent nucleoside either in free form or in a form incorporated into a cyclic dinucleotide fluoresces using a certain excitation wavelength or wavelength range, and that this fluorescence may be measured at a certain emission wavelength or wavelength range, then that same wavelength or wavelength range should be used in measuring the fluorescence of the fluorescent nucleoside or cyclic dinucleotide comprising it in the presence and absence of the substance to be tested. The "identical conditions" as recited in the methods of the invention therefore help ensure comparability of the fluorescence signals measured in the presence and absence of the substance to be tested, so that any change in the relative levels of these signals can confidently be attributed to the effect of that substance itself.

[0030] The term "substantially identical" as used herein has the meaning of identical within normal experimental error, e.g. with regard to the reaction conditions used in the measurement of a fluorescent signal in order to allow valid comparison of the measured fluorescence signals from two or more measuring steps. For example, the term "substantially identical" as used in accordance to the present invention refers to the measurement of two or more fluorescence signals at the same time points in the same aqueous solution (i.e. the aqueous solutions of two or more measuring steps comprises same salt and protein concentrations, same pH and the like) in order to allow a reliable comparison of the measured fluorescence signals in the presence or absence of a substance. The skilled person recognizes that normal variance within experimental error, e.g. due to pipetting solution volumes, weighing solutes and the like, are unavoidable, but still lead to results which are substantially identical.

[0031] The features of any one embodiment disclosed herein are intended as being combinable with those of any other embodiment. Such combinations of one or more features in any one embodiment with one or more features in any other embodiment belong to the disclosure of the present application as filed.

[0032] All documents or portions of documents cited in this application, including but not limited to patents, patent applications, articles and books, are hereby expressly incorporated by reference in their entirety for any purpose.

Detailed description

[0033] As used herein the term "cyclic dinucleotide" (CDN) refers to a molecule in which two nucleoside monophosphates are covalently joined, via two phosphodiester linkages, into a single species. Such "head to toe"-linked CDNs sometimes serve as a second messenger secreted by certain bacteria, such as *Listeria monocytogenes*, or generated by the cyclic GMP-AMP synthase (cGAS) following binding to cytosolic DNA. CDNs produced by the enzyme cGAS comprise an asymmetric 2'-5' and 3'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate. CDNs produced by cGAS are referred to as "2'3' cyclic dinucleotide" or "2'3'CDN" according to the present invention. The first and the second nucleoside monophosphate in the CDN according to the invention can be either a deoxyribonucleoside monophosphate or a ribonucleoside monophosphate provided that one of the first and second nucleoside monophosphate has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group in order to enable the 2'-5' and 3'-5' phosphosdiester linkages. Positions of the sugar backbone of the CDN according to the invention which do not participate in cyclization can be substituted with any moiety such as H-, OH-, methyl-, amino-, methoxy-, fluoro-, methoxyethyl-, -O-propargyl, and O-propylamine, provided that such moiety does not impede cyclization of the two nucleoside monophosphates. Furthermore, the phosphodiester linkages of the 2'3'CDN according to the invention may comprise modifications such as thiophosphate, boranophosphate, methyl phosphonate, N-3'-phosphoramidate (O2P-NH-ribose) e.g. in order to enhance the stability of the phosphodiester linkage.

[0034] As used herein, the terms "nucleoside" and "nucleotide" used in various forms (e.g. "nucleoside monophosphate", "cyclic dinucleotide", etc.) have the commonly accepted meaning in the art. These species contain a 1-β-D-furanose moiety with its stereochemical configuration as in naturally occurring DNA and RNA.

[0035] As used herein the term "cyclic GMP-AMP synthase" or "cGAS" refers to an enzyme belonging to the nucleotidyltransferase (NTase) superfamily which is responsible for cytosolic DNA recognition. Upon recognition of cytosolic dsDNA, cGAS becomes activated by the formation of a cGAS/dsDNA complex. Activated cGAS converts free nucleotide triphosphates into cyclic dinucleotides having an asymmetric 3'-5' and 2'-5' phosphodiester linkage between a first and a second nucleoside monophosphate (i.e. 2'3'CDN). These cyclic dinucleotides can serve as second messengers in downstream signaling. Non-limiting examples for variants of cGAS may be defined by e.g. SEQ ID NO: 10 and SEQ ID NO: 12.

[0036] As used herein the term "stimulator of interferon genes" or "STING" refers to a protein associated with the endoplasmic reticulum that binds to cyclic dinucleotides to trigger cytokine production. STING is expressed in various endothelial and epithelial cell types, as well as in haematopoietic cells, such as T cells, macrophages and dendritic cells. STING is activated upon binding of CDNs and formation of a complex consisting of a V-shaped STING dimer and a CDN molecule. CDNs which are capable of activating STING are either secreted by various bacteria or produced by the enzyme cGAS. 2'3'CDNs produced by cGAS have a much higher binding affinity for STING than CDNs secreted by bacteria which have two conventional 3'5' phosphodiester linkages between the first and the second nucleoside monophosphates (3'3'CDN). Therefore, 2'3'CDNs produced by cGAS play a key role in STING activation. Activated STING then interacts with Tank-binding kinase 1 (TBK1) and traffics through the Golgi to perinuclear endosomal compartments, where TBK1 phosphorylates the transcription factors IRF3 and IRF7 to induce the expression of pro-inflammatory cytokines and chemokines. Non-limiting examples for variants of STING may be defined by e.g. SEQ ID NOs: 2, 4, 6 and 8.

[0037] As used herein the term "2'-5' phosphodiesterase" refers to a pyrophosphate/phosphodiesterase enzyme which degrades the phosphodiester bond in 2'3'CDNs. The enzyme regulates the localization, duration and amplitude of cyclic dinuleotide signaling within subcellular domains and therefore plays an important role in the signal transduction mediated by 2'3'CDN second messenger molecules. Beside the cleavage of 2'3'CDNs, the 2'-5' phosphodiesterase is also capable of cleaving 3'3'CDNs secreted by various bacteria, such as *Listeria monocytogenes.*

[0038] As used herein the term "DNA-bending protein" refers to proteins that upon interaction with DNA induce a structural bend or kink in the DNA. In general, bending or introducing curves or kinks in DNA can lead to DNA compactification or to structural arrangement of DNA binding sites in cis. Examples are bacterial and mitochondrial genome structuring protein (e.g. bacterial HU, mitochondrial TFAM), where DNA bending in conjunction with protein dimerization induces loop structures in DNA, resulting in genome packaging as nucleoids. Other DNA bending proteins such as the transcription factor TBP shape DNA to be recognized by the transcription complexes. Histone proteins also bend or curve DNA, forming nucleosomes. HMG box proteins also bend DNA upon binding. These proteins function in stress response. In one illustrative example of the present invention, a "DNA-bending protein" may represent a "substance" having an ability to enhance cGAS activity. The "DNA-bending proteins" as used in the illustrative examples of the present invention therefore provide evidence that the below described methods of the invention are suitable to identify a substance having an ability to modulate (e.g. enhance) cGAS activity. In any event, the skilled person will understand that a "DNA-bending protein" is nonessential for the herein described methods of the invention and merely serves as an illustrative example representing a substance having the ability to enhance cGAS activity.

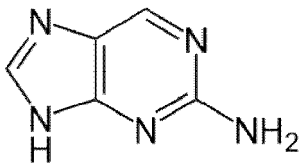
[0039] As used herein the term "affinity tag" refers to a peptide sequence which is genetically linked to the N- or C-terminus of an amino acid sequence of a protein or a polypeptide. Affinity tags according to the invention are used for protein purification, protein detection and enhancing the solubility of a protein. Examples of affinity tags include but are not limited to albumin-binding protein (ABP), alkaline phosphodiesterase (AP), AU1 epitope, AU5 epitope, bacteriophage T7 epitope, bacteriophage V5 epitope, biotin-carboxy carrier protein (BCCP), bluetongue virus tag (B-tag), calmodulin binding peptide (CBP), chloramphenicol acetyl transferase (Crow, Chase et al. "Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1". American journal of medical genetics. Part A 0, 296-312) cellulose binding domain (CBP), chitin binding domain (CBD), choline-binidng domain (CBD), dihydrofolate reductase (DHFR), E2 epitope, FLAG epitope, galactose-binding protein (GBP), green fluorescent protein (GFP), Glu-Glu (EE-tag), glutathione S-transferase (GST), human influenza hemagglutinin (HA), Halo Tag® (Promega), histidine affinity tag (Li, Shu et al. "Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization". Immunity 39, 1019-1031), horseradish peroxidase (HRP), HSV epitope, ketosteorid isomerase (KSI), KT3 epitope, LacZ, luciferase, maltose-binding protein (MBP), Myc epitope, NusA, PDZ domain, PDZ ligand, polyarginine (Arg-tag), polyaspartate (Asp-tag), polycysteine (Cys-tag), polyhistidine (e.g. His_6-tag), polyphenylalanine (Phe-tag), Profinity eXact™ fusion Tag (Bio-Rad), protein C, S1-tag, S-tag, streptavidin-binding peptide (SBP), Staphylococcal protein A (protein A), Staphylococcal protein G (protein G), Strep-tag, Streptavidin, small ubiquitin-like modifier (SUMO), tandem affinity purification (TAP), T7 epitope, thioredoxin (Trx), TrpE, ubiquitin, Universal, VSV-G. Furthermore, the term "affinity tag" as used herein may refer to a combination of one or more of the above recited peptide sequences.

[0040] As used herein the term "nucleoside" refers to a molecule comprising a ribofuranosyl or deoxyribofuranosyl ring linked to a nucleobase via a beta glycosidic linkage. If the 5'oxygen of the (deoxy)ribofuranosyl is linked to one, two or three phosphate groups, the nucleoside is referred to as a nucleoside monophosphate, diphosphate or triphosphate,

respectively, as commonly denoted in the art. The one, two or three phosphate groups can comprise modifications such as thiophosphates, boranophosphate, methyl phosphonate, N-3'-phosphoramidate (O2P-NH-ribose) e.g. in order to increase the stability of phosphodiester bonds. Examples of nucleobases comprised in the nucleoside molecule include but are not limited to adenine, guanine, thymine, uracil, cytosine and fluorescent nucleobase analogues.

[0041] As used herein the term "fluorescent nucleoside" refers to a nucleoside comprising a fluorescent nucleobase analogue having an intrinsic fluorescence, i.e. the fluorescence is a property of the heterocyclic ring of the nucleobase analogue itself and not a property from an extrinsic fluorescent label attached to it. Fluorescent nucleobase analogues resemble the shape of naturally occurring nucleobases, such as cytosine (C), guanine (G), adenine (A), thymine (T) or uracil (U), and also have the ability to form hydrogen bonds with other nucleobases, e.g. in the manner known for nucleic acid base-pairing. The fluorescent nucleobase analogues herein therefore include those having a high structural similarity to the respective natural nucleobases and, when incorporated into cyclic dinucleotides (CDNs) or polynucleotides, result in fluorescent CDNs or fluorescent poynucleotides which resemble the shape of the natural molecules. Importantly, the fluorescent nucleobase analogues according to the present invention have a fluorescence activity which is higher than the fluorescence activity of natural occurring nucleobases. According to the present invention, the intensity of the fluorescence emitted by the fluorescent nucleobase analogue in a "fluorescent nucleoside" is quenched, i.e. reduced, by a change in the chemical environment surrounding the fluorescent nucleobase analogue in the manner well known for fluorescence resonance energy transfer (FRET), as known in the art. This means that if the fluorescent nucleoside changes its location, e.g. when it is incorporated into a larger molecule such as a CDN, the quantum yield of the fluorescent signal emitted by the fluorescent nucleobase analogue will be quenched, i.e. reduced, thereby allowing one to analyze environmental changes of the fluorescent nucleoside comprising the fluorescent nucleobase analogue by measuring its fluorescence. In particular, the fluorescence intensity (i.e. the quantum yield φ) of a fluorescent nucleoside decreases upon incorporation of the fluorescent nucleoside into a CDN molecule, and/or upon binding of the fluorescent nucleoside or a CDN comprising the fluorescent nucleoside to the cGAS/STING signaling pathway, e.g. upon binding to cGAS and/or STING. Due to the dependence of fluorescence intensity of the fluorescent nucleosides on environment, the fluorescent nucleosides of the present invention can therefore be utilized to measure physiological processes including but not limited to enzymatic activity and interaction of molecules as described herein.

[0042] As used herein the term "2-aminopurine" or "2-AP" refers to a highly fluorescent nucleobase analogue of adenine having the lowest energy absorption band centered at 305 nm and a molar absorptivity (ε) of 6000 M^{-1} cm^{-1}. The position of the band is outside the absorption of the natural nucleic acids, so 2-AP can be selectively excited in the presence of natural nucleobases. The high quantum yield of free 2-AP in water at 25°C (φ=0.68) is considerably quenched, i.e. reduced, (~100 fold) when 2-AP is incorporated into a CDN or nucleic acid. It is quenched further still when the 2'3'CDN in which 2-AP exists is bound by the protein STING. This high sensitivity of 2-AP to its microenvironment can be utilized in a variety of assays to study e.g. DNA or nucleotide/protein interactions and enzymatic activity. 2-AP has the following structure:



[0043] As used herein the term "quantum yield" refers to the efficiency of the fluorescence process. The term "quantum yield" as used herein is defined as the ratio of the number of photons emitted to the number of photons absorbed. It is defined by the following equation:

$$\varphi = \frac{\text{Number of photons emitted}}{\text{Number of photons absorbed}}$$

[0044] The maximum fluorescence quantum yield is 1.0 (100%), which means that each photon absorbed results in a photon emitted. The fluorescence signal of a "fluorescent nucleoside" according to the present invention may have a quantum yield (φ) in free form in water at 25°C of about 0.10 to about 1.0 or between about 0.10 and 1.0. In one embodiment of the invention a fluorescent nucleoside according to the invention has a quantum yield in free form in water at 25°C of at least: about 0.10, about 0.11, about 0.12, about 0.13, about 0.14, about 0.15, about 0.16, about 0.17, about 0.18, about 0.19, about 0.20, about 0.21, about 0.22, about 0.23, about 0.24, about 0.25, about 0.26, about 0.27, about 0.28, about 0.29, about 0.30, about 0.31, about 0.32, about 0.33, about 0.34, about 0.35, about 0.36, about 0.37, about 0.38, about 0.39, about 0.40, about 0.41, about 0.42, about 0.43, about 0.44, about 0.45, about 0.46, about 0.47,

about 0.48, about 0.49, about 0.50, about 0.51, about 0.52, about 0.53, about 0.54, about 0.55, about 0.56, about 0.57, about 0.58, about 0.59, about 0.60, about 0.61, about 0.62, about 0.63, about 0.64, about 0.65, about 0.66, about 0.67, about 0.68, about 0.69, about 0.70, about 0.71, about 0.72, about 0.73, about 0.74, about 0.75, about 0.76, about 0.77, about 0.78, about 0.79, about 0.80, about 0.81, about 0.82, about 0.83, about 0.84, about 0.85, about 0.86, about 0.87, about 0.88, about 0.89, about 0.90, about 0.91, about 0.92, about 0.93, about 0.94, about 0.95, about 0.96, about 0.97, about 0.98, about 0.99, or about 1.00. A nucleoside is still considered "intrinsically fluorescent", and therefore a "fluorescent nucleoside" according to the present invention, if it has a quantum yield in free form in water at 25°C of at least 0.10.

[0045] As used herein the term "molar absorptivity" refers to the molar attenuation coefficient or molar extinction coefficient and is a measurement of how strongly a chemical species, e.g. a fluorophore, absorbs light at a given wavelength. It is an intrinsic property of the chemical species, i.e. it depends on the chemical structure of the species.

[0046] As used herein the term "aqueous solution" refers to a solution comprising water, e.g. a solution comprising water and at least one buffer. A buffer comprises a weak acid and its conjugate base or a weak base and its conjugate acid. A buffer keeps the pH of a solution constant by taking up protons that are released during reactions, or by releasing protons when they are consumed by reactions. A buffer therefore stabilizes the H^+ concentration in vitro without affecting the functioning of a system under investigation. As used herein, an "aqueous solution" can further comprise essential cofactors for enzymatic reaction, such as metal ions, critical salts and essential nutrients for cells or tissues. An aqueous solution may also comprise one or more solvents other than water, as long as such solvents are sufficiently miscible in water such that multiple phases do not form.

[0047] As used herein the term "cGAS-activating nucleic acid" refers to any nucleic acid molecule capable of activating cGAS. In particular, a cGAS-activating nucleic acid may be a double-stranded (ds), single-stranded (ss) DNA or a DNA/RNA hybrid. In the case of nucleic acid, e.g. dsDNA, the cGAS-activating nucleic acid has a length of at least 10 base pairs (bp). Furthermore, the double-stranded cGAS-activating nucleic acid may have a minimal length selected from the group consisting of at least 11bp, 12bp, 13bp, 14bp, 15bp, 16bp, 17bp, 18bp, 19bp, 20bp, 21bp, 22bp, 23bp, 24bp, 25bp, 26bp, 27bp, 28bp, 29bp, 30bp, 31bp, 32bp, 33bp, 34bp, 35bp, 36bp, 37bp, 38bp, 39bp, 40bp, 41bp, 42bp, 43bp, 44bp, 45bp, 46bp, 47bp, 48bp, 49bp, 50bp, 51bp, 52bp, 53bp, 54bp, 55bp, 56bp, 57bp, 58bp, 59bp, 60bp, 61bp, 62bp, 63bp, 64bp, 65bp, 66bp, 67bp, 68bp, 69bp, 70bp, 71bp. 72bp, 73bp, 74bp, 75bp, 76bp, 77bp, 78bp, 79bp, 80bp, 81bp, 82bp, 83bp, 84bp, 85bp, 86bp, 87bp, 88bp, 89bp, 90bp, 91bp, 92bp, 93bp, 94bp, 95bp, 96bp, 97bp, 98bp, 99bp, and 100bp. The double-stranded cGAS-activating nucleic acid may also have a minimal length of at least 200bp, 300bp, 400bp, 500bp, 600bp, 700bp, 800bp, 900bp, 1kbp, 2kbp, 3kbp, 4kbp, 5kbp, 6kbp and the like. Corresponding lengths also apply in the event the cGAS-activating nucleic acid, e.g. DNA, is single-stranded. The double-stranded cGAS-activating nucleic acid may be DNA, for example a plasmid, such as a commonly known expression vector. The cGAS-activating nucleic acid may also be in the form of single-stranded DNA or DNA/RNA hybrids. Such DNA/RNA hybrids may also have the lengths as set out above.

[0048] As used herein the term "substance" refers to any molecule which can be tested for its binding to cGAS or STING. The term "substance" as used herein may for example refer to any molecule having an ability to modulate cGAS or 2'-5' phosphodiesterase activity. A "substance" according to the invention may be a small molecule, a protein, a polypeptide, or a single amino acid. The protein may be an enzyme having biological activity, or an antibody. The term "antibody" as used herein comprises the full-length antibody or fragments thereof, wherein the term "antibody fragment" comprises a Fab, F(ab')2, monospecific Fab2, bispecific Fab2, trispecific Fab2, monovalent IgG, scFv, bispecific scFv, bispecific diabody, trispecific triabody, scFv-Fc, minibody and the like. The term "substance" as used herein may also comprise chemical molecules having saturated or unsaturated hydrocarbons, such as alkanes, cycloalkanes, alkenes and the like, i.e. may generally represent any member of the class of "small molecules" which the skilled person recognizes as constituted by substances which may be synthesized by an organic synthetic chemist without the need to implement microorganisms. A "substance" according to the invention may further comprise an alcohol, ester, ether, amide and the like. Furthermore, a "substance" as used herein may be a carbohydrate or a lipid. Nucleic acids or single nucleotides may also be a "substance" according to the present invention. A nucleic acid may be a double-stranded (ds) or single standed (ss) DNA. A nucleic acid may also be an RNA, such as a miRNA or siRNA, or a DNA/RNA hybrid. A nucleic acid may also refer to poly(I:C), poly(dI:dC), poly(dA:dU), or any combination of ribose and deoxyribose backbones. The backbones furthermore can contain phosphothioate esters, instead of phosphates esters between the backbone sugars. In addition, it may contain or consist of peptide nucleic acid (PNA) backbones, locked nucleic acid (LNA) backbones, Glycol nucleic acid (GNA) backbones, threose nucleic acid (TNA) backbones and other Xeno nucleic acid (XNA) alternatives. In one illustrative example, a substance according to the present invention is represented by a "DNA bending protein" which is capable of enhancing cGAS activity as shown by the methods of the present invention.

[0049] As used herein the term "systemic lupus erythematosus" or "SLE" refers to an autoimmune disease in which antibodies specific for DNA, RNA or proteins associated with nucleic acids form immune complexes that damage small blood vessels, especially in the kidneys. Patients with SLE generally have abnormal B and T cell function. SLE can be associated with hyperproduction of type I interferon.

[0050] As used herein the term "Aicardi-Goutières syndrome" or "AGS" refers to a neurodegenerative disorder that

can be caused by STING-dependent cytokine hyperproduction owing to mutations in genes such as TREX-1 (three-prime repair exonuclease 1).

[0051] As used herein the term "idiopathic arterial calcification of infancy" or "IAIC" refers to a rare condition characterized by extensive calcification and stenosis of large and medium sized arteries. IACI is an unusual genetically inherited autosomal recessive condition which may be dystrophic or metastatic.

[0052] As used herein the term "stenosis" refers to an abnormal narrowing in a blood vessel or other tubular organ or structure.

[0053] As used herein the term "pseudoxanthoma elasticum" or "PXE" refers to a progressive disorder that is characterized by the accumulation of deposits of calcium and other minerals (mineralization) in elastic fibers. Elastic fibers are a component of connective tissue, which provides strength and flexibility to structures throughout the body.

[0054] As used herein the term "diabetic kidney disease" refers to a kidney disease caused by or coexistent with diabetes.

[0055] As used herein the terms "engineered", "genetic engineering" or "genetic modification" refer to the manipulation of a nucleotide sequence of a gene encoding a protein or enzyme. A nucleotide sequence may for example be manipulated by introducing mutations into the nucleotide sequence of the gene, thereby altering the amino acid sequence of the encoded protein or enzyme. For example, mutations can be introduced into the active site of an enzyme, thereby altering its enzymatic activity. By choosing the appropriate mutation the enzymatic activity can be increased, decreased or completely silenced compared to the unmodified enzyme. A mutation in the nucleotide sequence of a gene encoding a protein can alter the conformation of the protein thereby influencing its functionality and stability. Furthermore, specific mutations can be introduced into the nucleotide sequence of a gene to investigate polymorphisms, such as single nucleotide polymorphisms, that might be associated with specific diseases. A mutation can be a substitution, deletion or insertion of a single nucleotide or more than one nucleotide in the nucleotide sequence of a gene encoding a protein, e.g. an enzyme. As used herein the terms "engineered", "genetic engineering" or "genetic modification" may also refer to the deletion of nucleotide sequences encoding one or more specific domains of the protein or enzyme, resulting in a truncated version of the molecule. For example, a truncated protein, e.g. a truncated enzyme, might only include the C-terminus, N-terminus or one or more highly conserved catalytic domains compared to the full length molecule. By deleting nucleotide sequences encoding e.g. unstructured regions, the truncated protein, e.g. the enzyme, may gain advantageous properties compared to the full length molecule, including but not limited to increased expression, improved solubility, increased activity or binding affinity. Nucleotide sequences can be further engineered by replacing specific nucleotide sequences of a gene with orthologous nucleotide sequences from different species. Furthermore, nucleotide sequences can be engineered to comprise affinity tags or other labels at their N- or C-terminus allowing for purification or detection of the encoded protein or enzyme. "Genetic engineering" or "genetic modification" is achieved by well known methods in the art including chemical synthesis and PCR technology (Russell. "Molecular Cloning, a laboratory manual". Cold Spring Harbor Laboratory Press 1, 2012);(Hughes, Miklos et al. "Chapter twelve - Gene Synthesis: Methods and Applications". Methods in Enzymology Volume 498, 277-309, 2011; Throop and LaBaer. "Recombinational Cloning Using Gateway and In-Fusion Cloning Schemes". Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.] 110, 3.20.21-23.20.23, 2015).

[0056] As used herein the term "modulate" or "modulating" refers to the capability of a substance to inhibit, activate or enhance the activity or properties of another substance, e.g. cGAS, STING or 2'-5' phosphodiesterase, thereby changing the concentration of CDNs, in particular 2'3'CDNs. A substance that acts as an antagonist of one of the aforementioned enzymes may partially or fully reduce the physiological function of said protein, thereby resulting in a reduction of CDN production (in the case of cGAS inhibition) or CDN cleavage (in the case of 2'-5' phosphodiesterase inhibition). Specifically, a substance which antagonizes the activity of cGAS may partially or fully inhibit the conversion of the nucleoside triphosphates into 2'3'CDNs. A substance which is an antagonist of cGAS might bind to the active center of the enzyme thereby competing with the nucleoside triphosphates for binding to cGAS. A substance which antagonizes cGAS activity (partially or fully) might also bind to the enzyme in an uncompetitive manner, i.e. to an allosteric site of cGAS which results in an alteration of the active binding site thereby preventing nucleoside triphosphate binding. A substance that acts as an antagonist of 2'-5' phosphodiesterase may function in a competitive or uncompetitive manner similar to a substance inhibiting cGAS activity. A substance which acts as an agonist of cGAS may enhance (partially or fully) cGAS activity, e.g. an agonist of cGAS enhances the conversion of nucleoside triphosphates into 2'3'CDNs in comparison to the conversion of nucleoside triphosphates into 2'3'CDNs in the absence of such an agonist. A substance which activates 2'-5' phosphodiesterase (partially or fully) may enhance the cleavage of the 2'-5' phosphodiester linkage of the 2'3' CDN, e.g. an agonist of 2'-5' phosphodiesterase reduce the overall concentration of 2'3' CDN molecules in a biological system.

[0057] As used herein the term "divalent cation" refers to a cation having a valence of 2. A "divalent cation" according to the invention is selected from the group consisting of Mg^{2+}, Ca^{2+}, Zn^{2+}, Fe^{2+}, Cu^{2+}, Ni^{+2} and Mn^{2+}.

[0058] As used herein the term "measuring" refers to determining the intensity of a fluorescence signal measured at a predetermined time point or to determining the change of a fluorescence signal measured continuously or at intervals

over a predetermined time interval.

[0059] The present invention is now described in more detail by reference to certain preferred embodiments.

Cyclic dinucleotides

[0060] As mentioned above, one aspect of the present invention relates to a cyclic dinucleotide comprising a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate. The fluorescent cyclic dinucleotides of the present invention are hereinafter referred to as 2'3'fCDNs.

[0061] The fluorescent nucleoside comprises a fluorescent nucleobase analogue which possesses intrinsic fluorescence. Furthermore, the intrinsic fluorescence of the fluorescent nucleobase analogue is sensitive to environmental changes. This means that if the fluorescent nucleoside comprising the fluorescent nucleobase analogue changes its disposition with respect to other chemical species, e.g. other nucleobases in a CDN, the quantum yield will be quenched, i.e. reduced, thereby allowing the analysis of environmental changes of the fluorescent nucleoside. In particular, the fluorescence intensity, i.e. the quantum yield of the fluorescent nucleoside, is reduced upon incorporation of the fluorescent nucleoside into a CDN, and/or upon binding of the nucleoside or a CDN comprising it to cGAS and/or STING. Due to their environmental sensitivity the fluorescent nucleosides of present invention can therefore be utilized to measure physiological processes including but not limited to enzymatic activity and interaction of molecules as will be discussed in more detail below.

[0062] The fluorescent nucleoside according to the present invention has a fluorescence which is higher than the fluorescence of natural occurring nucleosides, such as adenosine, guanosine, cytosine, thymidine, uracil and the like. In a further embodiment of the present invention, the fluorescence signal of a fluorescent nucleoside according to the present invention has a quantum yield (φ) in free form in water at 25°C of about 0.10 to about 1.0 or between about 0.10 and 1.0. In one embodiment of the invention a fluorescent nucleoside according to the invention has a quantum yield in free form in water at 25°C of at least: about 0.10, about 0.11, about 0.12, about 0.13, about 0.14, about 0.15, about 0.16, about 0.17, about 0.18, about 0.19, about 0.20, about 0.21, about 0.22, about 0.23, about 0.24, about 0.25, about 0.26, about 0.27, about 0.28, about 0.29, about 0.30, about 0.31, about 0.32, about 0.33, about 0.34, about 0.35, about 0.36, about 0.37, about 0.38, about 0.39, about 0.40, about 0.41, about 0.42, about 0.43, about 0.44, about 0.45, about 0.46, about 0.47, about 0.48, about 0.49, about 0.50, about 0.51, about 0.52, about 0.53, about 0.54, about 0.55, about 0.56, about 0.57, about 0.58, about 0.59, about 0.60, about 0.61, about 0.62, about 0.63, about 0.64, about 0.65, about 0.66, about 0.67, about 0.68, about 0.69, about 0.70, about 0.71, about 0.72, about 0.73, about 0.74, about 0.75, about 0.76, about 0.77, about 0.78, about 0.79, about 0.80, about 0.81, about 0.82, about 0.83, about 0.84, about 0.85, about 0.86, about 0.87, about 0.88, about 0.89, about 0.90, about 0.91, about 0.92, about 0.93, about 0.94, about 0.95, about 0.96, about 0.97, about 0.98, about 0.99, or about 1.00. Any of the above recited quantum yield values may be combined with one another to form a quantum yield range. Also contemplated within such quantum yield ranges are corresponding ranges in which the lower quantum yield is included while the upper quantum yield is excluded, as well as ranges in which the lower quantum yield is excluded while the upper quantum yield is included. In a particular preferred embodiment, the quantum yield of a fluorescent nucleoside in free form in water at 25°C is about 0.68, more preferably is 0.68. This is for example the fluorescent yield of 2-AP in free form in water at 25°C.

[0063] In one embodiment of the present invention, the 3'-5' phosphodiester linkage of the cyclic dinucleotide is between the 3' oxygen of the first nucleoside monophosphate and the 5' oxygen of the second nucleoside monophosphate and the 2'-5' phosphodiester linkage of the cyclic dinucleotide is between the 5' oxygen of the first nucleoside monophosphate and the 2' oxygen of the second nucleoside monophosphate. This asymmetric nature of the phosphodieseters is advantageous for high affinity binding to human STING.

[0064] In a further embodiment of the present invention, the fluorescent nucleoside monophosphate is a fluorescent purine nucleoside monophosphate. The fluorescent purine nucleoside monophosphate is advantageous for recognition by cGAS and may also enhance binding to STING.

[0065] In a preferred embodiment of the present invention, the fluorescent nucleoside monophosphate is selected from the group consisting of a 2-aminopurine nucleoside monophosphate (2-APMP), a 3-methyl-isoxanthopterin nucleoside monophosphate (3-MIMP), a 6-methyl isoxanthopterin nucleoside monophosphate (6-MIMP), 4-amino-6-methyl-8-(2-deoxy-beta-d-ribofuranosyl)-7(8H)-pteridone nucleoside monophosphate (6-MAPMP), a 4-amino-2,6-dimethyl-8-(2'-deoxy-beta-d-ribofuranosyl)-7(8H)-pteridone nucleoside monophosphate (DMAPMP), a pyrrolocytosine nucleoside monophosphate (pyrrolo-CMP), a 6-phenylpyrrolocytosine nucleoside monophosphate (PhpCMP), a (aminoethoxy)phenylpyrrolocytosine nucleoside monophosphate (moPhpCMP), a [bis-o-(aminoethoxy)phenyl]pyrrolocytosine nucleoside monophosphate (boPhpCMP), a hydropyrimidopyrimidine nucleoside monophosphate ($C^{hpp}MP$), a pyrrolopyrimidopyrimidine nucleoside monophosphate ($C^{ppp}MP$), a pyrimidopyrimidoindole nucleoside monophosphate ($C^{ppi}MP$), a benzopyridopyrimidine nucleoside monophosphate (BPPMP), a naphthopyridopyrimidine nucleoside monophosphate (NPPMP), a methoxybenzodeazaadenine nucleoside monophosphate (^{MD}AMP), a methoxybenzodeazain-

osine nucleoside monophosphate (^{MD}IMP), a naphthodeazaadenine nucleoside monophosphate (NDAMP), a furan-modified pyrimidine nucleoside mohophosphate, a thieno[3,2]pyrimidine nucleoside monophosphate, a thieno[3,4]pyrimidine nucleoside monophosphate, a 5-methoxy-quinazoline-2,4-(1H,3H) dione nucleoside monophosphate, a 5-methylpyrimidine-2-one nucleoside monophosphate, a 7-deazapurine nucleoside monophosphate, a 5-alkyluridine nucleoside monophosphate, a benzoquinalozines nucleoside monophosphate, a triazoleadenosine nucleoside monophosphate, and a 1,N^6-ethenoadenosine nucleoside monophosphate.

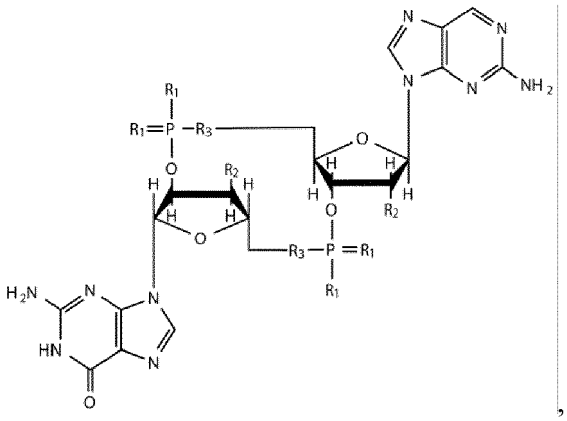
[0066] In a further embodiment of the present invention, a fluorescent nucleoside monophosphate can comprise further chemical modification to the fluorescent nucleobase analogue and the ribofuranosyl backbone provided that such modification does not affect the fluorescence of the fluorescent nucleobase analogue. Furthermore, such modification should not affect the structural and chemical properties of a fluorescent nucleoside monophosphate, i.e. the modification may not affect the fluorescent nucleoside monophosphate's ability to incorporate into nucleotide molecules, such as polynucleotides and CDNs. Possible modifications of the fluorescent nucleoside monophosphate according to the present invention include but are not limited to phosphothioate linker. In addition, it may contain or consist of peptide nucleic acid (PNA) backbones, locked nucleic acid (LNA) backbones, Glycol nucleic acid (GNA) backbones, threose nucleic acid (TNA) backbones and other Xeno nucleic acid (XNA) alternatives.

[0067] In a particularly preferred embodiment, the fluorescent nucleoside monophosphate is a 2-aminopurine nucleoside monophosphate (2-APMP), wherein the fluorescent nucleobase analogue is 2-aminopurine. 2-aminopurine (2-AP) is a highly fluorescent nucleobase analogue of adenine having the lowest energy absorption band centered at 305 nm and a molar absorptivity (ε) of 6000 M^{-1} cm^{-1}. The position of the band is outside the absorption of the natural nucleic acids and, thus, 2-AP can be selectively excited in the presence of natural nucleobases. The high quantum yield of free 2-AP in water (φ=0.68) is considerably quenched, i.e. reduced, (~100 times) when incorporated into biological molecules such as nucleic acids or CDNs. This high sensitivity of 2-AP to its microenvironment can be utilized in a variety of methods of measuring enzyme activity or investigating molecule interactions as will be discussed in more detail below.

[0068] In a further preferred embodiment of the present invention, 2-APMP is the first nucleoside monophosphate. A CDN comprising 2-APMP as the first nucleoside monophosphate resembles the shape of 2'3'CDNs made by cGAS and recognized by STING, thereby making such CDN in particular suitable for methods of e.g. measuring cGAS activity or identifying substances with an ability to modulate the properties of another substance, e.g. cGAS, STING or 2'-5' phosphodiesterases.

[0069] Due to the chemical structure of the 2'3'CDN it is necessary that one of the two nucleoside monophosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group in order to enable the 2'-5' and 3'-5' phosphodiester linkages in the 2'3'CDN. Thus, in a further embodiment of the invention the second nucleoside monophosphate of the cyclic dinucleotide is a purine ribonucleoside monophosphate, preferably a guanosine monophosphate (GMP) having a free 2' hydroxyl group, whereas 2-APMP is the first nucleoside monophosphate having a free 3' hydroxyl group. The chemical structure of a CDN comprising 2-APMP as the first nucleoside monophosphate and guanosine monophosphate as the second nucleoside monophosphate is almost identical to the chemical structure of the 2'3'CDN made by cGAS. Such a CDN is therefore an ideal fluorescent analogue of the naturally occurring 2'3'CDN made by cGAS and recognized by STING, and may thus be used in various methods as set out below in more detail.

[0070] In another preferred embodiment of the present invention, the fluorescent cyclic dinucleotide has the following formula

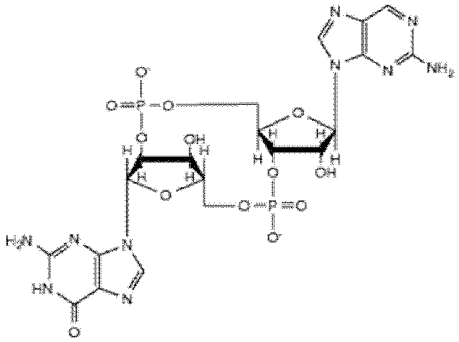


wherein

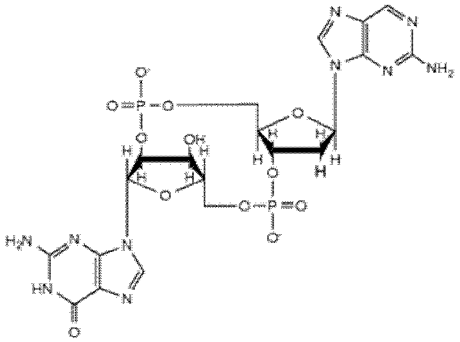
when R_1 is bound to P through a single bond, R_1 is independently selected from the group consisting of O, S, BH_3

and CH_3;
when R_1 is bound to P through a double bond, R_1 is independently selected from the group consisting of O, S and NH;
wherein at least one R_1 bound to each P is O;
R_2 is independently selected from the group consisting of H, OH, methyl, amino-, methoxy-, fluoro-, methoxyethyl-, -O-propargyl and O-propylamine; and
R_3 is O.

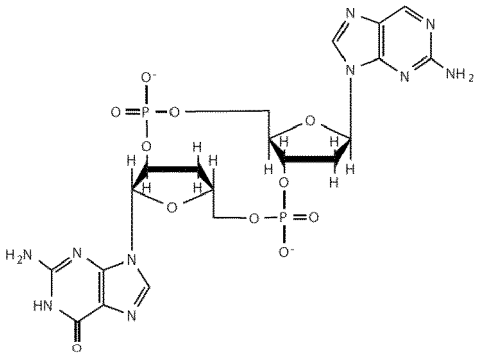
[0071] In a particularly preferred embodiment of the present invention, the cyclic dinucleotide has the following formula:



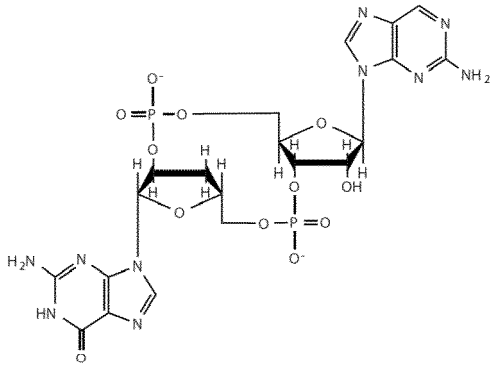
or



or



or



.

This cyclic dinucleotide resembles the structural and chemical properties of cyclic dinucleotide generated by cGAS and recognized by STING and can thus be used in a variety of methods according to the present invention as will be discussed in more detail below.

[0072] For example, the 2'3'fCDNs according to the present invention can be used to identify a substance having an ability to bind to STING. The cyclic dinucleotide of the present invention can be further used to identify a substance having an ability to modulate 2'-5' phosphodiesterase activity. Furthermore, the cyclic dinucleotide of the present invention can be used *in vitro* to label enzymes or proteins, such as cGAS and STING, in order to investigate their cellular localization during different diseases and during bacterial or viral infection. For example, the cyclic dinucleotide according to the present invention may form a complex with cGAS or STING, which can be readily localized in an *in vitro* system, such as a mammalian cell culture, due to the fluorescence signal of the fluorescent nucleoside of the cyclic dinucleotide. The fluorescence signal can be detected by using commonly known techniques such as fluorescence microscopy (Lakowicz. "Principles of Fluorescence Spectroscopy". Springer 3th edition, 2006).

Method of identifying a substance having an ability to bind to STING

[0073] Stimulator of interferon genes (STING) is a protein which is expressed in various endothelial and epithelial cell types, as well as in hematopoietic cells, such as T cells, macrophages and dendritic cells. Homologues of STING have been identified in different eukaryotic species as well as in invertebrates. STING stimulates the transcription of numerous innate immune genes, including type I interferon encoding genes, upon binding of cyclic dinucleotides (CDNs). Beside the bacterial 3'3' CDNs, it was found out that the most potent activator of STING is a 2'3'CDN generated by the enzyme cGAS. This second messenger binds STING with approximately 250-fold higher affinity compared to CDNs secreted by

bacteria. Binding of 2'3'CDN to STING results in dimerization of the signaling molecule and triggers a signaling cascade resulting in the expression of various immune modulatory molecules including proinflammatory cytokines, such as type I interferon, and chemokines. STING therefore plays an essential role in controlling the transcription of numerous immune modulatory genes and its dysfunction is related to the development of several diseases. For example, hyperactivity of STING results in hyperproduction of proinflammatory cytokines and chemokines causing autoimmune disorders and inflammatory diseases such as systemic lupus erythematosus (SLE) and Aicardi-Goutières Syndrome (AGS). On the other hand, impaired activity of STING is associated with the development of cancer, promoting the proliferation and spread of tumor cells (Nowarski, Gagliani et al. "Innate Immune Cells in Inflammation and Cancer". Cancer Immunology Research 1, 77-84, 2013).

[0074] Up to now, methods for identifying substances that are capable of binding to STING do not provide a reliable and simple assay for identifying STING ligands in a quantitative manner. However, due to its central role in the regulation of the innate immune system and its involvement in the development of various diseases, it would be of great importance to have a robust and simple-to-implement high-throughput method that enables for direct identification of substances capable to bind to STING. Such substances could then serve as a promising starting point in the development of therapeutic drugs specifically targeting STING protein and modulating its activity.

[0075] A further aspect of the present invention therefore relates to a method of identifying a substance having an ability to bind to stimulator of interferon genes (STING), wherein the method comprises the steps of (i) providing an aqueous solution comprising a cyclic dinucleotide and STING, wherein the cyclic dinucleotide comprises a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphate is a fluorescent nucleoside monophosphate; (ii) measuring the fluorescence signal of the aqueous solution; (iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence signal is measured under identical or substantially identical conditions following the provision of the respective aqueous solution, and (iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance; wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance has an ability to bind to STING.

[0076] The above method of the present invention is based on the unexpected finding that the fluorescent cyclic dinucleotide (2'3'fCDN) of the present invention, e.g. a CDN comprising 2-APMP, binds to STING, thereby activating the signaling molecule. The inventors surprisingly found out that upon binding of 2'3'fCDN to STING, the fluorescence signal of the cyclic dinucleotide is quenched, thereby allowing the measurement of 2'3'fCDN binding to STING with a simple fluorescence read out technique. While incorporation of 2-AP in the 2-APTP into cyclic 2'3' fluorescent GMP/AMP (2'3'fGAMP) quenches the fluorescence signal of 2-AP relative to its fluorescence as observed in free solution, the binding of 2'3'fGAMP to STING quenches the fluorescence signal of 2-AP further still. Based on this further fluorescence quenching it is thus possible to measure the fluorescence signal in the absence or presence of a substance with a simple-to-implement method and to reliably identify substances that bind to STING. These STING ligands could than serve as promising starting points for the development of therapeutic drugs that specifically modulate STING activity.

[0077] According to the present invention, STING is a protein which may be obtained by conventional cloning and expression systems, including *E. coli* expression strains (e.g. BL21(DE3), Lemo21(DE3), ArcticExpress), baculovirus expression systems (ex. in High Five insect cell line), non-viral insect cells expression system (ex. transient transfection to S2 insect cell line using ExpreS2), yeast strains (ex. BCY123) or mammalian cell line (ex. HEK293T, CHO) expression systems. STING can be expressed as full-length protein or as a truncated version comprising the catalytic domain only. Furthermore, STING can be engineered to contain modifications in its nucleic acid sequence altering the amino acid sequence of the protein. For example, STING can be engineered (Gao, Ascano et al. "Structure-Function Analysis of STING Activation by c[G(2',5')pA(3',5')p] and Targeting by Antiviral DMXAA". Cell 154, 748-762, 2013; Gao, Zillinger et al. "Binding-Pocket and Lid-Region Substitutions Render Human STING Sensitive to the Species-Specific Drug DMXAA". Cell Reports 8, 1668-1676, 2014) in its CDN binding site to result in an increased or decreased binding affinity for the 2'3'CDN compared to unmodified STING. Furthermore, STING can be engineered to be constitutively active. Modifications in its nucleic acid can facilitate STING dimerization and/or TBK1 binding (Tang and Wang. "Single Amino Acid Change in STING Leads to Constitutive Active Signaling". PLOS ONE 10, e0120090, 2015). STING may also be modified by expressing the protein as a fusion construct together with a tag, such as a GFP tag that would allow for determination of its cellular localization.

[0078] In a further embodiment of the present invention, the protein may further comprise affinity tags allowing purification and detection of the expressed protein.

[0079] According to the method of the present invention, STING may be derived from any vertebrate, including mammals such as human, macaque, mouse, sus, bos, and the like, birds, such as chicken and duck, fish, such as zebra fish, other vertebrates such as Xenopus laevis, Choanoflagellate (e.g. Nematostella vectensis), or invertebrates (e.g. insects).

[0080] In a preferred embodiment of the present invention, STING is human STING. The nucleotide sequence of full-length human STING is defined by SEQ ID NO: 1.

SEQ ID NO: 1

[0081]

```
ATGCCCCACTCCAGCCTGCATCCATCCATCCCGTGTCCCAGGGGTCACGGGGCC
CAGAAGGCAGCCTTGGTTCTGCTGAGTGCCTGCCTGGTGACCCTTTGGGGGCTA
GGAGAGCCACCAGAGCACACTCTCCGGTACCTGGTGCTCCACCTAGCCTCCCTG
CAGCTGGGACTGCTGTTAAACGGGGTCTGCAGCCTGGCTGAGGAGCTGCGCCAC
ATCCACTCCAGGTACCGGGGCAGCTACTGGAGGACTGTGCGGGCCTGCCTGGGC
TGCCCCCTCCGCCGTGGGGCCCTGTTGCTGCTGTCCATCTATTTCTACTACTCC
CTCCCAAATGCGGTCGGCCCGCCCTTCACTTGGATGCTTGCCCTCCTGGGCCTC
TCGCAGGCACTGAACATCCTCCTGGGCCTCAAGGGCCTGGCCCCAGCTGAGATC
TCTGCAGTGTGTGAAAAAGGGAATTTCAACGTGGCCCATGGGCTGGCATGGTCA
TATTACATCGGATATCTGCGGCTGATCCTGCCAGAGCTCCAGGCCCGGATTCGA
ACTTACAATCAGCATTACAACAACCTGCTACGGGGTGCAGTGAGCCAGCGGCTG
TATATTCTCCTCCCATTGGACTGTGGGGTGCCTGATAACCTGAGTATGGCTGAC
CCCAACATTCGCTTCCTGGATAAACTGCCCCAGCAGACCGGTGACCATGCTGGC
ATCAAGGATCGGGTTTACAGCAACAGCATCTATGAGCTTCTGGAGAACGGGCAG
CGGGCGGGCACCTGTGTCCTGGAGTACGCCACCCCCTTGCAGACTTTGTTTGCC
ATGTCACAATACAGTCAAGCTGGCTTTAGCCGGGAGGATAGGCTTGAGCAGGCC
AAACTCTTCTGCCGGACACTTGAGGACATCCTGGCAGATGCCCCTGAGTCTCAG
AACAACTGCCGCCTCATTGCCTACCAGGAACCTGCAGATGACAGCAGCTTCTCG
CTGTCCCAGGAGGTTCTCCGGCACCTGCGGCAGGAGGAAAAGGAAGAGGTTACT
GTGGGCAGCTTGAAGACCTCAGCGGTGCCCAGTACCTCCACGATGTCCCAAGAG
CCTGAGCTCCTCATCAGTGGAATGGAAAAGCCCCTCCCTCTCCGCACGGATTTC
TCTTGA
```

[0082] SEQ IN NO: 1 as shown above includes the terminal TGA stop codon. The stop codon may be useful in expressing the protein as a discrete product. However, there might be instances where the protein is to be expressed together with at least one other protein product, e.g. as a protein fusion. In such instances, it is possible to remove the terminal TGA stop codon, replacing it with a nucleic acid sequence encoding the other protein(s) of interest. The skilled person understands under what circumstances the terminal stop codon TGA is to be retained or dispensed with, and is readily able to reproduce the above sequence in the presence or absence of the terminal stop codon TGA.

[0083] In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 1 is at least about 54% to 100% or between about 54% and 100%. In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 1 is at least: about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 1 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING nucleotide sequence having the above recited sequence identities to SEQ ID NO: 1 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 1. The skilled person can incorporate and express sequences comprising the above sequences in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention.

[0084] The amino acid sequence of human full-length STING is defined by SEQ ID NO: 2.

SEQ ID NO: 2

[0085]

MPHSSLHPSIPCPRGHGAQKAALVLLSACLVTLWGLGEPPEHTLRYLVLHLAS
LQLGLLLNGVCSLAEELRHIHSRYRGSYWRTVRACLGCPLRRGALLLLSIYFY

YSLPNAVGPPFTWMLALLGLSQALNILLGLKGLAPAEISAVCEKGNFNVAHGL
AWSYYIGYLRLILPELQARIRTYNQHYNNLLRGAVSQRLYILLPLDCGVPDNL
SMADPNIRFLDKLPQQTGDHAGIKDRVYSNSIYELLENGQRAGTCVLEYATPL
QTLFAMSQYSQAGFSREDRLEQAKLFCRTLEDILADAPESQNNCRLIAYQEPA
DDSSFSLSQEVLRHLRQEEKEEVTVGSLKTSAVPSTSTMSQEPELLISGMEKP
LPLRTDFS

[0086] In some embodiments of the present invention, the identity between STING amino acid sequence and SEQ ID NO: 2 is at least about 36% to 100% or between about 36% and 100%. In some embodiments of the invention, the identity between STING amino acid sequence and SEQ ID NO: 2 is at least: about 36%, about 37%, about 38%, about 39%, about 40%, about 41%, about 42%, about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING amino acid sequence and SEQ ID NO: 2 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING amino acid sequence having the above recited sequence identities to SEQ ID NO: 2 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 2. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 2 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention.

[0087] In another preferred embodiment of the present invention, the nucleotide sequence of full-length STING is as defined by SEQ ID NO: 3. This sequence encodes a human STING variant which is derived from a macrophage cell line having an arginine to histidine substitution at position 220 (R220H) and a histidine to arginine substitution at position 232 (H232R). The STING^{H220R232} variant was shown to have greater affinity to 2'3' cyclic dinucleotides generated by cGAS than wild-type STING encoded by SEQ ID NO: 1, represented by SEQ ID NO: 2.

SEQ ID NO: 3

[0088]

```
ATGCCCCACTCCAGCCTGCATCCATCCATCCCGTGTCCCAGGGGTCACGGGGC
CCAGAAGGCAGCCTTGGTTCTGCTGAGTGCCTGCCTGGTGACCCTTTGGGGGC
TAGGAGAGCCACCAGAGCACACTCTCCGGTACCTGGTGCTCCACCTAGCCTCC
CTGCAGCTGGGACTGCTGTTAAACGGGGTCTGCAGCCTGGCTGAGGAGCTGCG
CCACATCCACTCCAGGTACCGGGGCAGCTACTGGAGGACTGTGCGGGCCTGCC
TGGGCTGCCCCCTCCGCCGTGGGGCCCTGTTGCTGCTGTCCATCTATTTCTAC
TACTCCCTCCCAAATGCGGTCGGCCCGCCCTTCACTTGGATGCTTGCCCTCCT
GGGCCTCTCGCAGGCACTGAACATCCTCCTGGGCCTCAAGGGCCTGGCCCCAG
CTGAGATCTCTGCAGTGTGTGAAAAAGGGAATTTCAACGTGGCCCATGGGCTG
GCATGGTCATATTACATCGGATATCTGCGGCTGATCCTGCCAGAGCTCCAGGC
CCGGATTCGAACTTACAATCAGCATTACAACAACCTGCTACGGGGTGCAGTGA
GCCAGCGGCTGTATATTCTCCTCCCATTGGACTGTGGGGTGCCTGATAACCTG
AGTATGGCTGACCCCAACATTCACTTCCTGGATAAACTGCCCCAGCAGACCGG
TGACCGTGCTGGCATCAAGGATCGGGTTTACAGCAACAGCATCTATGAGCTTC
TGGAGAACGGGCAGCGGGCGGGCACCTGTGTCCTGGAGTACGCCACCCCCTTG
CAGACTTTGTTTGCCATGTCACAATACAGTCAAGCTGGCTTTAGCCGGGAGGA
TAGGCTTGAGCAGGCCAAACTCTTCTGCCGGACACTTGAGGACATCCTGGCAG
ATGCCCCTGAGTCTCAGAACAACTGCCGCCTCATTGCCTACCAGGAACCTGCA
GATGACAGCAGCTTCTCGCTGTCCCAGGAGGTTCTCCGGCACCTGCGGCAGGA
GGAAAAGGAAGAGGTTACTGTGGGCAGCTTGAAGACCTCAGCGGTGCCCAGTA
CCTCCACGATGTCCCAAGAGCCTGAGCTCCTCATCAGTGGAATGGAAAAGCCC
CTCCCTCTCCGCACGGATTTCTCTTGA
```

[0089] SEQ IN NO: 3 as shown above includes the terminal TGA stop codon. The stop codon may be useful in expressing the protein as a discrete product. However, there might be instances where the protein is to be expressed together with at least one other protein product, e.g. as a protein fusion. In such instances, it is possible to remove the terminal TGA stop codon, replacing it with a nucleic acid sequence encoding the other protein(s) of interest. The skilled person understands under what circumstances the terminal stop codon TGA is to be retained or dispensed with, and is readily able to reproduce the above sequence in the presence or absence of the terminal stop codon TGA.

[0090] In some embodiments of the present invention, the identity between STING nucleotide sequence and SEQ ID NO: 3 is at least about 54% to 100% or between about 54% and 100%. In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 3 is at least: about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 3 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING nucleotide sequence having the above recited sequence identities to SEQ ID NO: 3 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 3. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 3 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention, in order to obtain recombinant STING for use in the present invention. The amino acid sequence of human full-length STING ^{H220R232} is as defined by SEQ ID NO: 4.

SEQ ID NO: 4

[0091]

MPHSSLHPSIPCPRGHGAQKAALVLLSACLVTLWGLGEPPEHTLRYLVLHLAS
LQLGLLLNGVCSLAEELRHIHSRYRGSYWRTVRACLGCPLRRGALLLLSIYFY
YSLPNAVGPPFTWMLALLGLSQALNILLGLKGLAPAEISAVCEKGNFNVAHGL
AWSYYIGYLRLILPELQARIRTYNQHYNNLLRGAVSQRLYILLPLDCGVPDNL
SMADPNIHFLDKLPQQTGDRAGIKDRVYSNSIYELLENGQRAGTCVLEYATPL
QTLFAMSQYSQAGFSREDRLEQAKLFCRTLEDILADAPESQNNCRLIAYQEPA
DDSSFSLSQEVLRHLRQEEKEEVTVGSLKTSAVPSTSTMSQEPELLISGMEKP
LPLRTDFS

[0092] In some embodiments of the present invention, the identity between STING amino acid sequence and SEQ ID NO: 4 is at least about 36% to 100% or between about 36% and 100%. In some embodiments of the invention, the identity between STING amino acid sequence and SEQ ID NO: 4 is at least: about 36%, about 37%, about 38%, about 39%, about 40%, about 41%, about 42%, about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING amino acid sequence and SEQ ID NO: 4 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING amino acid sequence having the above recited sequence identities to SEQ ID NO: 4 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 4. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 4 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention. As can be seen from Figure 2 of the description, human STING comprises an endoplasmic reticulum localization signal (1-36 aa) followed by transmembrane domain (TM) at the N-terminus (40-135 aa), a highly conserved cyclic dinucleotide binding domain (CBD) at the C-terminus (139-344 aa) which is sufficient for CDN binding and a C-terminal tail (CTT, 345-379) responsible for signal transduction. The inventors have found that truncated human STING consisting only of the highly conserved CBD domain has improved characteristics when used in the herein described methods of the invention. In particular, truncated human STING consisting only of CBD or of CBD and CTT has improved solubility in the aqueous solution, while retaining its capability to bind to CDN and to form STING dimers upon CDN binding. In addition, the truncated construct can efficiently produced recombinantly in soluble form in *E. coli.*

[0093] Thus, in a further preferred embodiment of the present invention, STING is a truncated version of the human full-length protein comprising the soluble CDN-binding domain CBD (139-344 aa). The nucleotide sequence of truncated human STING is defined by SEQ ID NO: 5.

SEQ ID NO: 5

[0094]

CTGGCCCCAGCTGAGATCTCTGCAGTGTGTGAAAAAGGGAATTTCAACGTGGC
CCATGGGCTGGCATGGTCATATTACATCGGATATCTGCGGCTGATCCTGCCAG
AGCTCCAGGCCCGGATTCGAACTTACAATCAGCATTACAACAACCTGCTACGG
GGTGCAGTGAGCCAGCGGCTGTATATTCTCCTCCCATTGGACTGTGGGGTGCC
TGATAACCTGAGTATGGCTGACCCCAACATTCGCTTCCTGGATAAACTGCCCC
AGCAGACCGGTGACCATGCTGGCATCAAGGATCGGGTTTACAGCAACAGCATC
TATGAGCTTCTGGAGAACGGGCAGCGGGCGGGCACCTGTGTCCTGGAGTACGC
CACCCCCTTGCAGACTTTGTTTGCCATGTCACAATACAGTCAAGCTGGCTTTA
GCCGGGAGGATAGGCTTGAGCAGGCCAAACTCTTCTGCCGGACACTTGAGGAC
ATCCTGGCAGATGCCCCTGAGTCTCAGAACAACTGCCGCCTCATTGCCTACCA
GGAACCTGCAGATGACAGCAGCTTCTCGCTGTCCCAGGAGGTTCTCCGGCACC
TGCGGCAGGAGGAAAAGGAAGAGGTTACTGTGGGC

[0095] In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 5 is at least about 58% to 100% or between about 58% and 100%. In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 5 is at least: about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 5 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING nucleotide sequence having the above recited sequence identities to SEQ ID NO: 5 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 5. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 5 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention.

[0096] The amino acid sequence of truncated human STING is defined by SEQ ID NO: 6.

SEQ ID NO: 6

[0097]

LAPAEISAVCEKGNFNVAHGLAWSYYIGYLRLILPELQARIRTYNQHYNNLLR
GAVSQRLYILLPLDCGVPDNLSMADPNIRFLDKLPQQTGDHAGIKDRVYSNSI
YELLENGQRAGTCVLEYATPLQTLFAMSQYSQAGFSREDRLEQAKLFCRTLED
ILADAPESQNNCRLIAYQEPADDSSFSLSQEVLRHLRQEEKEEVTVG

[0098] In some embodiments of the present invention, the identity between STING amino acid sequence and SEQ ID NO: 6 is at least about 43% to 100% or between about 43% and 100%. In some embodiments of the invention, the identity between STING amino acid sequence and SEQ ID NO: 6 is at least: about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING amino acid sequence and SEQ ID NO: 6 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING amino acid sequence having the above

recited sequence identities to SEQ ID NO: 6 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 6. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 6 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention.

[0099] In a further preferred embodiment of the present invention, the nucleotide sequence of truncated human STING is defined by SEQ ID NO: 7. This sequence represents the truncated STING^{H220R232} variant having a higher binding affinity for 2'3'CDN.

SEQ ID NO: 7

[0100]

CTGGCCCCAGCTGAGATCTCTGCAGTGTGTGAAAAAGGGAATTTCAACGTGGC
CCATGGGCTGGCATGGTCATATTACATCGGATATCTGCGGCTGATCCTGCCAG
AGCTCCAGGCCCGGATTCGAACTTACAATCAGCATTACAACAACCTGCTACGG
GGTGCAGTGAGCCAGCGGCTGTATATTCTCCTCCCATTGGACTGTGGGGTGCC
TGATAACCTGAGTATGGCTGACCCCAACATTCACTTCCTGGATAAACTGCCCC
AGCAGACCGGTGACCGTGCTGGCATCAAGGATCGGGTTTACAGCAACAGCATC
TATGAGCTTCTGGAGAACGGGCAGCGGGCGGGCACCTGTGTCCTGGAGTACGC
CACCCCCTTGCAGACTTTGTTTGCCATGTCACAATACAGTCAAGCTGGCTTTA
GCCGGGAGGATAGGCTTGAGCAGGCCAAACTCTTCTGCCGGACACTTGAGGAC
ATCCTGGCAGATGCCCCTGAGTCTCAGAACAACTGCCGCCTCATTGCCTACCA
GGAACCTGCAGATGACAGCAGCTTCTCGCTGTCCCAGGAGGTTCTCCGGCACC
TGCGGCAGGAGGAAAAGGAAGAGGTTACTGTGGGC

[0101] In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 7 is at least about 60% to 100% or between about 60% and 100%. In some embodiments of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 7 is at least: about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between STING nucleotide sequence and SEQ ID NO: 7 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING nucleotide sequence having the above recited sequence identities to SEQ ID NO: 7 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 7. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 7 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention.

[0102] The amino acid sequence of truncated human STING^{H220R232} is defined by SEQ ID NO: 8.

SEQ ID NO: 8

[0103]

LAPAEISAVCEKGNFNVAHGLAWSYYIGYLRLILPELQARIRTYNQHYNNLLR
GAVSQRLYILLPLDCGVPDNLSMADPNIHFLDKLPQQTGDRAGIKDRVYSNSI
YELLENGQRAGTCVLEYATPLQTLFAMSQYSQAGFSREDRLEQAKLFCRTLED
ILADAPESQNNCRLIAYQEPADDSSFSLSQEVLRHLRQEEKEEVTVG

[0104] In some embodiments of the present invention, the identity between STING amino acid sequence and SEQ ID

NO: 8 is at least about 44% to 100% or between about 44% and 100%. In some embodiments of the invention, the identity between STING amino acid sequence and SEQ ID NO: 8 is at least: about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a range. Also contemplated within such ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the present invention, the identity between STING amino acid sequence and SEQ ID NO: 8 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any STING amino acid sequence having the above recited sequence identities to SEQ ID NO: 8 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative physiological activity as STING defined by SEQ ID NO: 8. The skilled person can incorporate and express sequences comprising the above sequence identities to SEQ ID NO: 8 in recombinant expression systems according to known methods in the art, in order to obtain recombinant STING for use in the present invention.

[0105] In one embodiment of the present invention, the fluorescent signal of the fluorescent nucleoside monophosphate of the 2'3'fCDN is measured at a predetermined time-point. This time-point should be determined in preliminary experiments taking into account the "dead time" of the instrument used as the fluorescence reader, i.e. the shortest time it takes between mixing the reagents and measuring the first data point in a given instrument. Once the correct time point is determined for each experimental setup, measurement of the fluorescent signal at a predetermined time point allows fast and reliable measurement of the fluorescent signal of the fluorescent nucleoside monophosphate in the aqueous solution. This allows fast and reliable identification of a substance having an ability to bind to STING.

[0106] In another embodiment of the present invention, the fluorescence signal is measured continuously or at intervals over a predetermined time interval. In some embodiments of the present invention, the fluorescent signal of the fluorescent nucleoside is measured starting from about 10 seconds after mixing all compounds together up to about 1 hour after mixing all compounds together. Measuring over a time interval includes the measurement over a series of specific time points. This allows one to obtain information about the linearity of the reaction and reduces the possibility that measuring occurs at a time point when the reaction has already long been completed. The measurement over a time interval therefore reduces the possibility that the measured fluorescence signal is misinterpreted.

[0107] According to the method of the present invention, the 2'3'fCDN of the present invention binds to STING in the aqueous solution, thereby resulting in a quenched (i.e. reduced) fluorescent signal of the (at least one) fluorescent nucleoside monophosphate of the 2'3'fCDN of the present invention. A reduction of the fluorescence signal of the aqueous solution therefore directly correlates with the binding of 2'3'fCDN to STING. Thus if, after addition of a substance, the fluorescence signal measured in the presence of the substance increases compared to the fluorescence signal measured in the absence of the substance, this is an indication that previously bound 2'3'fCDN is released from STING, thereby reversing the quenching effect of STING on the fluorescence signal of bound 2'3'fCDN. An increase of the fluorescence signal in the presence of a substance therefore indicates that the 2'3'fCDN of the present invention is released from STING due to an interaction of the substance with STING. Binding of the substance to STING can either occur at the same binding site as the 2'3'fCDN of the present invention, thereby displacing the cyclic dinucleotide, or at a site different from the binding site of 2'3'fCDN, thereby causing a conformational change in STING which then releases the cyclic dinucleotide. Furthermore, the substance might prevent 2'3'fCDN binding to STING to begin with, by blocking the binding site of STING. The method according to the present invention thus provides a robust and reliable tool to identify STING ligands that interfere with 2'3'CDN binding to STING and which could therefore be a promising starting point for the development of therapeutic strategies to specifically target STING with the aim of modulating its activity.

[0108] In one embodiment of the present invention, the increase of the fluorescence signal measured in the presence of the substance compared to the fluorescence signal measured in the absence of the substance may indicate that the substance is a competitive binder of STING. A competitive binder of STING binds at the same site of the signaling molecule than 2'3'fCDN. Due to its higher binding affinity to STING, the competitive binder either displaces already bound 2'3'fCDN or blocks the binding site for 2'3'fCDN. A substance which results in an increase of the fluorescence signal of the aqueous solution may therefore compete with the 2'3'fCDN of the present invention for binding to STING.

[0109] In a further embodiment according to the present invention, the increase in the fluorescence signal measured in the presence of the substance compared to the fluorescence signal measured in the absence of the substance may indicate that the substance is a noncompetitive binder of STING. A noncompetitive binder binds to an allosteric site, other than or in addition to the binding site of 2'3' fCDN of the present invention, thereby inducing a conformational change in STING. This conformational change can affect the complex formation between the 2'3'fCDN and STING by

either preventing binding of 2'3'fCDN to STING, or if the cyclic dinucleotide is already bound to STING, can result in a release of 2'3'fCDN due to a conformational change in the binding site which reduces the binding affinity of 2'3'fCDN to STING. Thus, according to a further embodiment of the present invention, a substance which results in an increase of the fluorescence signal of the aqueous solution may therefore bind in an uncompetitive manner to STING, resulting in a conformational change in the protein that affects the interaction between STING and the 2'3'fCDN of the present invention.

[0110] In one embodiment of the invention, the substance is added to the aqueous solution comprising STING after the addition of the 2'3'fCDN.

[0111] Due to its environmental sensitivity an unchanged fluorescence signal measured in the presence of the substance in comparison with a fluorescence signal measured in the absence of the substance might indicate that the 2'3'fCDN of the present invention did not change its location or that STING did not undergo a conformational change. In either case, an unchanged fluorescence intensity in the presence of the substance compared to the fluorescence signal measured in the absence of the substance might indicate that the 2'3'fCDN is still bound to STING, either because binding of the substance does not affect 2'3'fCDN binding to STING, or because the substance did not bind to STING or binds to STING with such a low affinity that it does not affect 2'3'fCDN binding. However, an unchanged fluorescence signal measured in the presence of the substance compared to the fluorescence signal measured in the absence of the substance might also indicate that the chemical environment of the 2'3'fCDN has changed, but that the old and new environments of 2'3'fCDN are sufficiently similar and therefore give rise to a net unchanged fluorescence signal. So based on an unchanged fluorescence signal measured in the presence of the substance compared to the fluorescence signal measured in the absence of the substance, it may not be possible to positively conclude that the substance did not change the location of the 2'3'fCDN of the present invention and thus did not bind to STING, or indeed that the substance in question bound STING at all. A measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance may be taken as positive proof that the substance has an ability to bind sting, as recited in the present method of the invention. Thus, the method according to the invention, is a method for identifying a substance having an ability to bind to STING, thereby replacing the bound 2'3'fCDN and/or changing its chemical environment in a manner that the quenching effect caused by STING binding is partially or wholly reversed. The present method according to the invention may thus also be characterized as a method for identifying a substance having an ability to bind to STING and modulate, e.g. reduce, binding of the CDN, e.g. the 2'3'fCDN by STING.

[0112] In practice, one would therefore need to tune/choose the system such that a fluorescent change is correlated with a change in localization of the 2'3'fCDN of the present invention (e.g. by choosing appropriate fluorophores, or attaching them to sensitive positions at the 2'3' CDN). In principle, the skilled person could use fluorescence polarization anisotropy to measure CDN binding in the absence of any quenching effects to fine tune the method in order to allow an unambiguous interpretation of the measured fluorescence signal (Zhang H, Wu Q, Berezin MY. Fluorescence anisotropy (polarization): from drug screening to precision medicine. Expert Opin Drug Discov. 10(11):1145-61, 2015).

[0113] According to the present invention, a decrease in the fluorescent signal measured in the presence of the substance in comparison the fluorescent signal measured in the absence of the substance might indicate that the substance either further quenches the fluorescence signal of the fluorescent nucleoside monophosphate of the 2'3'fCDN or stabilizes the interaction between 2'3'fCDN and STING. For example, the substance could lower the dissociation constant resulting in a higher binding affinity of 2'3'fCDN to STING and in increased binding of the molecule to STING, thereby further reducing the measured fluorescence signal in the reaction mixture containing bound and free 2'3'fCDN .

[0114] In one embodiment of the present invention, the measured fluorescence signal of the fluorescent nucleoside monophosphate in the 2'3'fCDN is measured at different wavelengths or ranges of wavelengths depending on the fluorescent nucleoside monophosphate in the 2'3'fCDN used in the method according to the invention. Specifically, each 2'3'fCDN comprising a specific fluorescent nucleoside monophosphate may have different excitation and emission maxima which influence the wavelengths at which the fluorescence signal is measured. Thus, depending on the fluorescent nucleoside monophosphate of the 2'3'fCDN of the present invention, the wavelengths or range of wavelengths at which the fluorescent nucleoside monophosphate is excited (i.e. the excitation maxima) and the wavelengths or range of wavelengths at which the fluorescence signal is measured (i.e. the emission maxima) need to be adapted for each 2'3'fCDN of the present invention individually in order to avoid false positives by measuring a fluorescence signal outside the optimal wavelength or range of wavelengths for a given 2'3'fCDN of the present invention. In some embodiments of the present invention, the excitation wavelength is in the range of about 290nm to about 365nm, or between about 290nm and 365nm. In some embodiments of the invention, the excitation wavelength is about 290nm, about 291nm, about 292nm, about 293nm, about 294nm, about 295nm, about 296nm, about 297nm, about 298nm, about 299nm, about 300nm, about 301nm, about 302nm, about 303nm, about 304nm, about 305nm, about 306nm, about 307nm, about 308nm, about 309nm, about 310nm, about 311nm, about 312nm, about 312nm, about 313nm, about 314nm, about 315nm, about 316nm, about 317nm, about 318nm, about 319nm, about 320nm, about 321nm, about 322nm, about 323nm, is about 324nm, about 325nm, about 326nm, about 327nm, about 329nm, about 330nm, about 331nm, about 332nm, about 333nm, about 334nm, about 335nm, about 336nm, about 337nm, about 338nm, about 339nm, about

340nm, about 341nm, about 342nm, about 343nm, about 344nm, about 345nm, about 346nm, about 347nm, about 348nm, about 349nm, about 350nm, about 351nm, about 352nm, about 353nm, about 354nm, about 355nm, about 356nm, about 357nm, about 358nm, about 359nm, about 360nm, about 361nm, about 362nm, about 363nm, about 364nm, or about 365nm. Any of the above recited excitation wavelength may be combined with one another to form an excitation wavelength range. Also contemplated within such excitation wavelength ranges are corresponding ranges in which the lower excitation wavelength is included while the upper excitation wavelength is excluded, as well as ranges in which the lower excitation wavelength is excluded while the upper excitation wavelength is included. In a particularly preferred embodiment of the invention, the excitation wavelength ranges from about 305nm to about 315nm or between about 305 nm and 315nm. In a most preferred embodiment the excitation wavelength is 305nm.

[0115] In some embodiments of the present invention the emission wavelength ranges from about 345nm to about 500 nm or between about 345nm and 5000nm. In some embodiments the excitation wavelength is about 345nm, about 346nm, about 347nm, about 348nm, about 349nm, about 350nm, about 351nm, about 352nm, about 353nm, about 354nm, about 355nm, about 356nm, about 357nm, about 358nm, about 359nm, about 360nm, about 361nm, about 362nm, about 363nm, about 364nm, about 365nm, about 366nm, about 367nm, about 368nm, about 369, about 370nm, about 371nm, about 372nm, about 373nm, about 374nm, about 375nm, about 376nm, about 377nm, about 378nm, about 379nm, about 380nm, about 381nm, about 382nm, about 383nm, about 384nm, about 385nm, about 386nm, about 387nm, about 388nm, about 389nm, about 390nm, about 391nm, about 392nm, about 393nm, about 394nm, about 395nm, about 396nm, about 397nm, about 398nm, about 399nm, about 400nm, about 401nm, about 402nm, about 403nm, about 404nm, about 405nm, about 406nm, about 407nm, about 408nm, about 409nm, about 410nm, about 411nm, about 412nm, about 413nm, about 414nm, about 415nm, about 416nm, about 417nm, about 418nm, about 419nm, about 420nm, about 421nm, about 422nm, about 423nm, about 424nm, about 425nm, about 426nm, about 427nm, about 428nm, about 429nm, about 430nm, about 431nm, about 432nm, about 433nm, about 434nm, about 435nm, about 436nm, about 437nm, about 438nm, about 439nm, about 440nm, about 441nm, about 442nm, about 443nm, about 444nm, about 445nm, about 446nm, about 447nm, about 448nm, about 449nm, about 450nm, about 451nm, about 452nm, about 453nm, about 454nm, about 455nm, about 456nm, about 457nm, about 458nm, about 459nm, about 460nm, about 461nm, about 462nm, about 463nm, about 464nm, about 465nm, about 466nm, about 467nm, about 468nm, about 469nm, about 470nm, about 471nm, about 472nm, about 473nm, about 474nm, about 475nm, about 476nm, about 477nm, about 478nm, about 479nm, about 480nm, about 481nm, about 482nm, about 483nm, about 484nm, about 485nm, about 486nm, about 487nm, about 488nm, about 489nm, about 490nm, about 491nm, about 492nm, about 493nm, about 494nm, about 495nm, about 496nm, about 497nm, about 498nm, about 499nm or about 500nm. Any of the above recited emission wavelength may be combined with one another to form an emission wavelength range. Also contemplated within such emission wavelength ranges are corresponding ranges in which the lower emission wavelength is included while the upper emission wavelength is excluded, as well as ranges in which the lower emission wavelength is excluded while the upper emission wavelength is included. In a particularly preferred embodiment of the invention, the emission wavelength ranges from about 350nm to about 380nm, or between about 350nm and 380nm. In a most preferred embodiment of the invention the emission wavelength is 363nm.

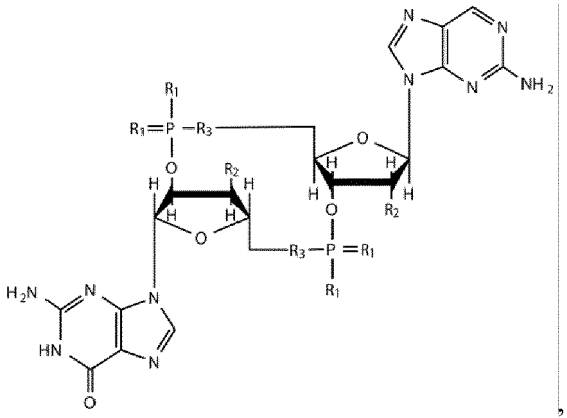
[0116] Furthermore, the quantum yield of a fluorescent nucleoside monophosphate of the 2'3'fCDN the present invention may also differ depending on the experimental conditions. Thus, for each 2'3'fCDN used in the method according to the invention, the experimental conditions may be modified in order to optimize the quantum yield of the fluorescent molecule. Within the biochemical constraints of the reaction under study, tuning of pH, temperature, buffer composition could be optimized. A reference for the quantum yield can be measured e.g. in ultrapure water or a standard reaction buffer. Experimental conditions that might influence the quantum yield of the 2'3'fCDN of the present invention include but are not limited to temperature, pH, ionic strength.

[0117] The change of the fluorescence signal can be measured by any known fluorescence read out technique using a conventional plate reader or a single-cuvette photometer. In a preferred embodiment of the invention, the fluorescence signal is measured with Tecan infinite M1000 using 96-well black non-binding PS plates (Greiner Bio-One), specifically FluoroMax-P spectrofluorometer (Horiba Scientific) or similar instruments can be used.

[0118] The inventors surprisingly found out that, when incorporated into a cyclic dinucleotide of the present invention, 2-aminopurine nucleoside monophosphate (2-APMP) is still fluorescently active and possess a high fluorescence environmental sensitivity which can be exploited to measure 2'3'fCDN binding to STING. In particular, the inventors found out that the fluorescence signal of the 2'3'fCDN comprising 2-APMP is significantly quenched, i.e. reduced, resulting from the environmental change of the 2'3'fCDN of the present invention when bound to STING. The quenching of the fluorescence signal upon binding of 2'3'fCDN comprising 2-APMP to STING thus allows one to measure 2'3'fCDN binding to STING in the presence or absence of a test substance in a continuous assay using conventional fluorescence plate readers, to yield valuable information about that substance's ability to bind to STING.

[0119] Thus, in a preferred embodiment according to the method of the present invention, the 2'3'fCDN comprises 2-APMP.

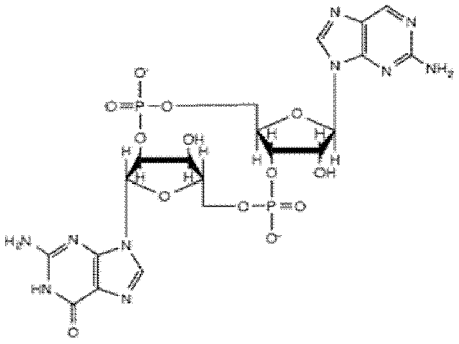
[0120] In another preferred embodiment, the 2'3'fCDN has the structure



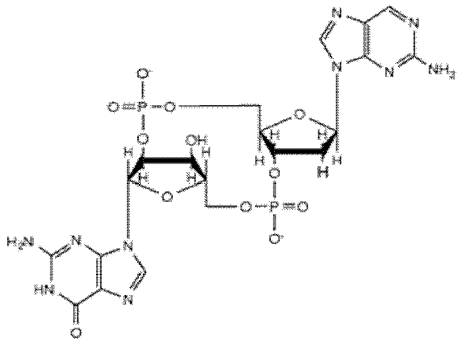
wherein

when R_1 is bound to P through a single bond, R_1 is independently selected from the group consisting of O, S, BH_3 and CH_3;
when R_1 is bound to P through a double bond, R_1 is independently selected from the group consisting of O, S and NH; wherein at least one R_1 bound to each P is O;
R_2 is independently selected from the group consisting of H, OH, methyl, amino-, methoxy-, fluoro-, methoxyethyl-, -O-propargyl and O-propylamine; and
R_3 is O.

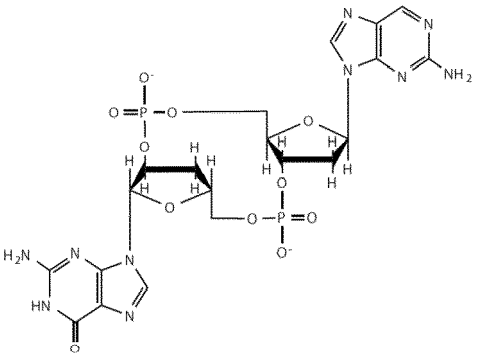
[0121] Most preferred the 2'3'fCDN has the structure:



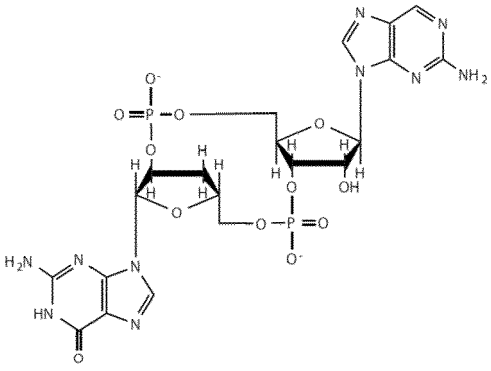
or



or



or



.

[0122] As can be seen from all of the above, the 2'3'fCDN of the present invention can advantageously be used for identifying a substance having an ability to bind to STING according to the inventive method set forth herein. Upon complex formation between STING and the cyclic dinucleotides of the present invention, the cyclic dinucleotides can

further be used to label STING.

Method of identifying a substance having an abitlity to modulate 2'-5' phosphodiesterase activity

[0123] Phosphodiesterases (PDE) are an important group of enzymes with an extensive functional range that are distributed widely and are highly conserved between species. The skilled person can purify any known PDE by commonly known chromatographic fractionation techniques using a cell extract and test the extract on PDE activity by the method of measuring PDE activity according to the invention. Furthermore, the skilled person will be able to identify new PDEs by way of a standard sequence comparison (e.g. BLAST) against a protein and/or nucleic acid sequence data base to identify similarities to already known PDEs.

[0124] In humans the ecto-nucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1, UniProtKB Accession P22413; Human Gene Numberin Committee: 3356; Entrez Gene: 5167) is expressed in a wide range of tissues including cartilage, heart, kidney, parathyroid and collateral model, and in cells such as vascular smooth muscle cells (VSMCs), osteoblasts and chondrocytes. ENPP1 specifically cleaves the 2'-5' phosphodiesterase linkage in cyclic dinucleotides and regulates the localization, duration and amplitude of 2'3'CDNs produced by cGAS. The enzyme therefore plays an important role in the control and mediation of immune responses induced via the cGAS/STING axis. Not surprisingly, ENPP1 dysfunction is associated with a variety of human diseases. For example, mutation in the gene encoding for ENPP1 at the found to be associated with a rare type of idiopathic infantile arterial calcification demonstrating the importance of ENPP 1 in maintaining normal tissue function. Furthermore, ENPP 1 mutation have been reported to result in autosomal recessive conditions causing pre-major onset of arterial calcification resulting in stenosis and pseudoxanthoma elasticum characterized by ectopic calcification of soft connective tissue. Finally, ENPP1 mutations are associated with diabetic kidney diseases (Mackenzie, Huesa et al. "New insights into NPP1 function: Lessons from clinical and animal studies". Bone 51, 961-968, 2012; Li, Yin et al. "Hydrolysis of 2'3'-cGAMP by ENPP1 and design of nonhydrolyzable analogs". Nat Chem Biol 10, 1043-1048, 2014; Sortica, Buffon et al. "Association between the ENPP1 K121Q Polymorphism and Risk of Diabetic Kidney Disease: A Systematic Review and Meta-Analysis". PLOS ONE 10, e0118416, 2015).

[0125] Beside human PDEs, bacteria have developed several immune evasion strategies specifically targeting the CDN-based immune signaling pathways. For example, M. tuberculosis has developed a mechanism to inhibit STING-dependent type I interferon expression via a multifunctional mycobacterial phosphodiesterase CdnP. This PDE mediates hydrolysis of both bacterial-derived and host-derived CDNs thereby directly interfering with the cGAS/STING axis (Dey, Dey et al. "Inhibition of innate immune cytosolic surveillance by an M. tuberculosis phosphodiesterase". Nat Chem Biol 13, 210-217, 2017).

[0126] In view of the essential roles of PDEs in bacterial infections as well as their association with a variety of human diseases, this group of enzymes represents an attractive target for the development of new immune therapeutic approaches as well as in treatment of bacterial infections. The basic idea in the pharmaceutical industry therefore is to modulate the activity of PDE, thereby specifically controlling CDN homeostasis. In particular, it will be of great importance to identify substances having an ability to modulate PDE activity thereby providing potential vaccines against bacterial infections or substances that could be used to treat rare human diseases, such as cancer and arterial calcification.

[0127] Thus, in one aspect the present invention relates to a method for identifying a substance having an ability to modulate 2'-5' phosphodiesterase activity, wherein the method comprises the steps of (i) providing an aqueous solution comprising a cyclic dinucleotide and a 2'-5' phosphodiesterase, wherein the cyclic dinucleotide comprises a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate; (ii) measuring the fluorescence signal of the aqueous solution; (iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and (iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance; wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance is a 2'-5' phosphodiesterase agonist, while a measured fluorescence signal which is lower in the presence of the substance compared to the absence of the substance indicates that the substance is a 2'-5' phosphodiesterase antagonist.

[0128] In one embodiment of the invention, the substance is added to the aqueous solution comprising a 2'-5' phosphodiesterase after the addition of the 2'3'fCDN.

[0129] In one embodiment of the present invention, an increase of the measured fluorescence signal in the presence of the substance compared to the fluorescence signal measured in the absence of the substance, indicates that the substance activates (partially or fully) or enhances 2'-5' phosphodiesterase activity, thereby resulting in increased cleavage of the 2'3' fluorescent dinucleotide (2'3'fCDN) of the present invention. Cleavage of the 2'-5' and/or 3'-5' phosphodiester linkages results in the release of the first and the second nucleoside monophosphates. The release of the fluorescent nucleoside monophosphate of the present invention partially or wholly reduces the quenching effect previously caused

by the incorporation of the free fluorescent nucleoside into the 2'3'fCDN of the present invention by cGAS. Thus, an increase of the fluorescent signal measured in the presence of the substance in comparison to the fluorescence signal measured in the absence of the substance directly correlates with an increased cleavage of 2'3'fCDN of the present invention and indicates that the substance has an ability to activate (partially or fully) or enhance 2'-5' phosphodiesterase activity. In this case, the substance is thus a 2'-5' phosphodiesterase agonist.

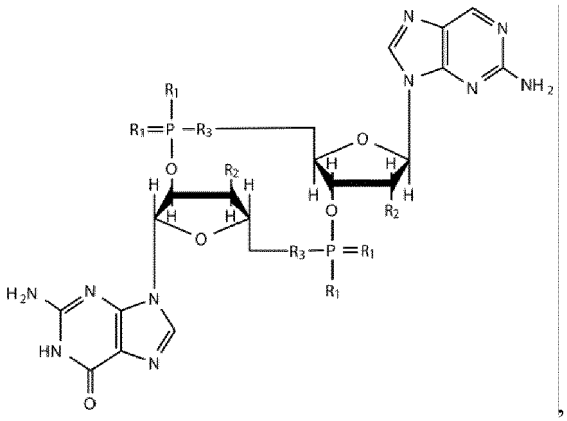
[0130] According to another embodiment of the present invention, a decrease in the fluorescence signal in the presence of the substance compared to the fluorescence signal measured in the absence of the substance indicates that the substance inhibits (partially or fully) 2'-5' phosphodiesterase. A decrease in the measured fluorescent signal in the presence of the substance indicates that less 2'3'fCDN of the present invention is cleaved by the 2'-5' phosphodiesterase, resulting in more fluorescent nucleoside monophosphate kept incorporated in 2'3'fCDN which directly correlates with quenching of the measured fluorescent signal. Thus, a decreased fluorescence signal measured in the presence of the substance compared to the absence of the substance indicates that less fluorescent nucleoside monophosphate is released from 2'3'fCDN due to reduced 2'-5' phosphodiesterase activity. The substance therefore acts as a 2'-5' phosphodiesterase antagonist.

[0131] In a further embodiment of the present invention, the fluorescent signal of the fluorescent nucleoside monophosphate in the 2'3'fCDN is measured at a predetermined time-point. This time-point should be determined in preliminary experiments taking into account the dead time of the instrument used as the fluorescence reader, i.e. the shortest time it takes between mixing the reagents and measuring the first data point in a given instrument. Once the correct time point is determined for each experimental set up, measuring of the fluorescent signal at a predetermined time point allows fast and reliable identification of a substance having an ability to modulate 2'-5' phosphodiesterase activity. Measuring over a time interval includes continuous measurement, and the measurement of a fluorescence signal over a series of discrete time points. This allows one to obtain information about the linearity of the reaction and reduces the possibility that the measurement occurs at a time point when the reaction might already have been completed. The measurement over a time interval may therefore reduce the risk that the measured fluorescence signal is misinterpreted.

[0132] In another embodiment of the present invention, the fluorescence signal is measured over a predetermined time interval. The advantage is a single readout that does not need to be processed further. In some embodiments of the present invention, the fluorescent signal is measured starting from about 10 seconds from mixing all compounds together up to about 1 hour. The advantage is that measuring over a time interval, the slope is a more precise readout than a single time point and also shows the linearity of the reaction. The time interval of the measurement depends strongly on 2'-5' phosphodiesterase activity in the presence of particular fluorescent 2'3'fCDN as substrate and on the experimental setup, such as buffer condition, temperature and substance concentration. In general, for the low 2'-5' phosphodiesterase activity higher protein and 2'3'fCDN concentrations may be used, compared to the case of high 2'-5' phosphodiesterase activity to perform the measurement within 5 min - 1 h time interval.

[0133] In a preferred embodiment according to the method of the present invention, the 2'3'fCDN comprises 2-APMP.

[0134] In another preferred embodiment, the 2'3'fCDN has the structure



wherein

when R_1 is bound to P through a single bond, R_1 is independently selected from the group consisting of O, S, BH_3 and CH_3;

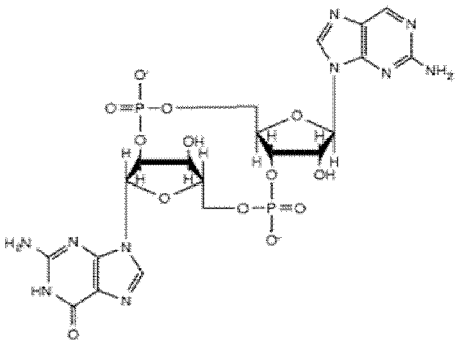
when R_1 is bound to P through a double bond, R_1 is independently selected from the group consisting of O, S and NH;

wherein at least one R_1 bound to each P is O;

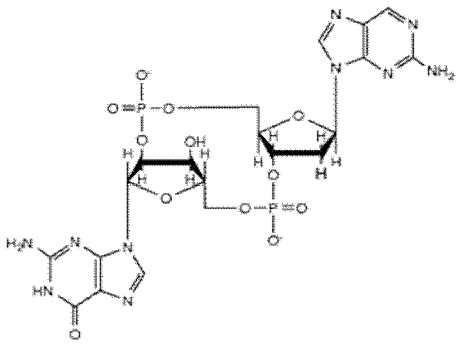
R_2 is independently selected from the group consisting of H, OH, methyl, amino-, methoxy-, fluoro-, methoxyethyl-,

-O-propargyl and O-propylamine; and
R_3 is O.

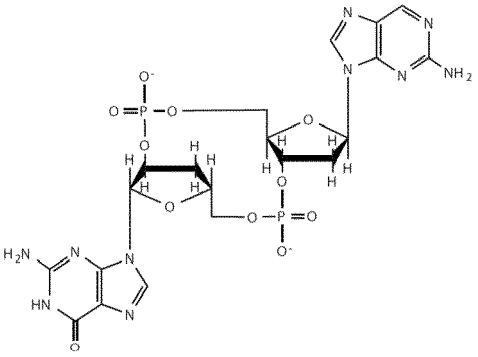
[0135] Most preferred the 2'3'fCDN has the structure:



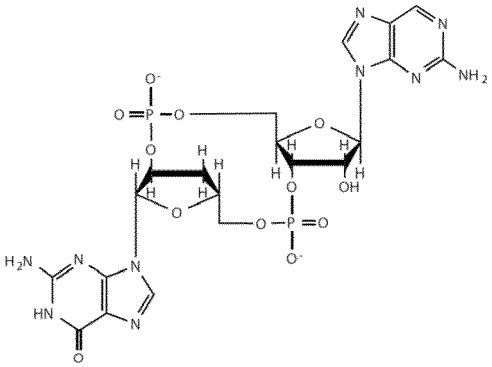
or



or



or



.

[0136] In one embodiment of the invention, the reaction is carried out by incubating 2'3'fCDN with a PDE. The change of the fluorescence signal is than measured in a commonly known fluorescence reader. Suitable cofactors (typical metal ions Mn^{2+}, Mg^{2+}, Zn^{2+} depending on the particular PDE) can be added to the reaction mixture. In case the PDE under investigation is specific for 2'-5' or 3-5' phosphodiesters, an PDE with the other specificity would be added as well, to ensure that the CDN is cleaved into two single nucleoside monophosphates.

[0137] In view of all of the above, the 2'3'fCDN of the present invention can be used in a method of identifying a substance having an ability to modulate 2'-5' phosphodiesterase activity.

Methods of (a) measuring cGAS activity and of (b) identifying a substance having an ability to modulate cGAS activity

(a) Method of measuring cGAS activity

[0138] The present invention is based, in part, on the surprising finding that cGAS can incorporate a fluorescent nucleobase analogue from a fluorescent nucleoside triphosphate into a cyclic dinucleotide which comprises a conventional 3'-5' phosphodiester linkage and an unconventional 2'-5' phosphodiester linkage, resulting in 2'3' fluorescent dinucleotides (2'3'fCDN). The inventors have found that the fluorescence signal of the fluorescent nucleobase analogue of the free fluorescent nucleoside triphosphate in aqueous solution is quenched, i.e. reduced, upon incorporation of the fluorescent nucleoside into a 2'3'fCDN of the present invention. Thus, the change in the measured fluorescence signal in aqueous solution over time can be used to measure cGAS activity.

[0139] Thus, a further aspect of the present invention relates to a method of measuring cGAS activity, wherein the method comprises the steps of (i) providing an aqueous solution comprising cGAS, a cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group and

(ii) measuring the change of the fluorescence signal of the aqueous solution over time. Measurement may be continuous or discontinuous, i.e. at intervals, over a predetermined time interval.

[0140] In one embodiment of the present invention, the method of measuring cGAS activity can be used to study cGAS variants and their efficiency in converting nucleoside triphosphates into 2'3'CDNs of the present invention. For example, the claimed method allows for the identification of engineered cGAS variants having increased or decreased cGAS activity compared to unmodified cGAS. The engineered cGAS variants can be used to study polymorphisms in the gene encoding cGAS that are associated with different diseases. For instance, it may be possible to analyze the enzymatic activity of cGAS encoded by polymorphic genes only found in a particular disease, thereby shedding light on the possible genetic mechanisms underlying said disease. Knowing the genetic mechanism related to cGAS dysfunction in a particular disease could serve as a promising starting point for the development of a therapeutic treatment.

[0141] In a further embodiment of the present invention, the method of measuring cGAS activity can be used e.g. to determine any of (a) the start and end points of measuring the fluorescence signal of the aqueous solution in the presence of different cGAS variants, (b) the effect of different nucleic acids on cGAS activation, (c) the optimal time interval required for measuring the change of the fluorescence signal of the aqueous solution, and (d) the optimal starting concentrations of the reaction partners. Furthermore, the method of measuring cGAS activity can be used to determine the standard fluorescence signal of a given fluorescent nucleoside triphosphate in aqueous solution under predetermined conditions. Thus, the method of measuring cGAS activity according to the present invention can be used to identify optimal reaction parameters which could then be used in the methods of the invention described below. The methods of measuring cGAS activity according to the present invention therefore also serve to calibrate and fine tune the experimental reaction setups for the methods of the invention described below.

[0142] In one embodiment of the present invention, cGAS activity is calculated based on the time-dependent change of the measured fluorescence signal of the aqueous solution. In particular, cGAS activity may be calculated by continuously measuring the change in the fluorescence signal of the aqueous solution in a plate reader or in a single cuvette. The obtained fluorescence curve may then be inverted into a curve representing cGAS activity (see e.g. Figure 7d). The initial slope of the curve representing cGAS activity may then be used to calculate the reaction velocity, which is one way of defining cGAS activity according to the present invention. This particular embodiment has the advantage of determining the linear portion of the curve representing cGAS activity which is not possible when using commonly known techniques such as mass spectrometry or radioactivity. The determination of the initial slope is a more precise measure of activity than measuring end points, few or single-time points. The slope, due to many more individual data points is statistically more significant and allows the determination of the linear range of the reaction, before substrate depletion or product inhibition slows down the reaction. A single point or few points of measurement are typically not sufficient to determine the linear range of a reaction, therefore increasing the possibility to underestimate the activity. In view of the above, the method of measuring cGAS activity can thus be used to optimize and fine-tune cGAS measurement conditions for other cGAS-related methods of the invention, e.g. for optimizing cGAS measurement conditions in a method of identifying a substance having an ability to modulate cGAS activity as described below in more detail.

(b) Method of identifying a substance having an ability to modulate cGAS activity

[0143] As discussed above, cGAS is a key sensor for cytosolic DNA in the innate immune system which elicits a signaling cascade that triggers the host response via production by cyclic dinucleotides comprising a conventional 3'-5' phosphodiester linkage and an unconventional 2'-5' phosphodiester linkage between two nucleoside monophosphates (2'3'CDN). The cyclic dinucleotides generated by cGAS bind to STING, thereby triggering the expression of type I interferon production. While under normal conditions the cGAS/STING axis protects the cell from bacterial and viral DNA infections, impairment of this pathway is associated with a variety of different diseases, such as inflammatory diseases and autoimmune disorders. For example, uncontrolled expression of type-I interferon due to constitutively active cGAS/STING is associated with various autoimmune disorders including systemic lupus erythematosus (SLE) and Aicardi-Goutières Syndrome (AGS). Thus, the identification of substances having the ability to act as cGAS/STING antagonists would be of great importance for the development of therapeutic treatments against such diseases.

[0144] In tumor cells the innate immune response is usually silenced allowing the tumor to proliferate and spread. The general idea in anti-cancer therapy is therefore to specifically trigger the cGAS/STING pathway in cancer cells, thereby eliciting a type I interferon immune response. Such an immune response could then instruct the immune system not only to attack the cancer cells but also to develop longer lasting immunogenic memory T cells against the tumor. Many cancer types developed strategies to become immunologically invisible and have for instance no T cell infiltrates, rendering anti-cancer immunity unsuccessful.

[0145] In view of the above, the identification of substances having an ability to modulate cGAS activity therefore represents an emerging technology for the development of new therapeutic strategies for the treatment of immunological disorders and cancer.

[0146] A further aspect of the present invention therefore relates to a method of identifying a substance having an

ability to modulate the activity of cGAS, wherein the method comprises the steps of (i) providing an aqueous solution comprising cGAS, cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group; (ii) measuring the fluorescence signal of the aqueous solution; (iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence signal is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and (iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance; wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance is a cGAS antagonist, while a measured fluorescence signal which is lower in the presence of the substance than in the absence of the substance indicates that the substance is a cGAS agonist.

[0147] In one embodiment of the invention, the substance is added to the aqueous solution comprising cGAS, cGAS-activating nucleic acid and one or more divalent cation after the addition of the first and second nucleoside triphosphates.

[0148] In one embodiment of the present invention, an increase of the measured fluorescence signal in the presence of the substance compared to the fluorescence signal measured in the absence of the substance, indicates that less fluorescent nucleoside triphosphate is incorporated into 2'3'fCDN resulting in a higher concentration of free fluorescent nucleoside triphosphate in the aqueous solution. Upon incorporation of the fluorescent nucleobase analogue from the fluorescent nucleotide triphosphate into 2'3'fCDN by cGAS, the fluorescent signal of the fluorescent nucleobase analogue is quenched resulting in a decreased fluorescence of the fluorescent nuceobase analogue from the fluorescent nucleoside triphosphate. Thus, a lower fluorescence signal correlates with increased incorporation of the fluorescent nucleobase analogue from the fluorescent nucleotide triphosphate into 2'3'fCDN, whereas a higher fluorescence signal correlates with less incorporation of the fluorescent nucleobase analogue from the fluorescent nucleotide triphosphate into 2'3'fCDN. The higher comparative amount of the measured fluorescence signal is therefore a direct indication for reduced incorporation of the fluorescent nucleoside triphosphate into 2'3'fCDN. Since the reduced incorporation of the fluorescent nucleoside triphosphate into the 2'3'fCDN is directly associated with reduced cGAS activity in the presence of a substance, the increased fluorescence signal allows for identification of a substance having an inhibitory effect on cGAS activity. An increase in the measured fluorescence signal in the presence of the substance compared to the fluorescence signal measured in the absence of the substance therefore indicates that the substance is a cGAS antagonist.

[0149] According to another embodiment of the present invention, a decrease of the measured fluorescence signal in the presence of the substance compared to the fluorescence signal measured in the absence of the substance indicates that more fluorescent nucleoside triphosphate is incorporated into 2'3'fCDN of the present invention which results in a reduced concentration of free fluorescent nucleoside triphosphate in the aqueous solution. Due to its environmental sensitivity, the fluorescence signal of the fluorescent nucleoside triphosphate is quenched upon incorporation of the nucleoside into the 2'3'fCDN of the present invention, which results in a decrease in the measured fluorescence signal of the aqueous solution. Thus, a decrease in the measured fluorescence signal in the presence of the substance compared to the fluorescence signal measured in the absence of the substance indicates that cGAS converts more free fluorescent nucleoside triphosphate into 2'3'fCDN in the presence of the substance than in the absence of the substance. A decrease of the measured fluorescence signal in the presence of the substance therefore indicates that the substance activates (partially or fully) or enhances cGAS enzymatic activity; in this case, the substance is thus a cGAS agonist.

[0150] According to the present invention, cGAS is a recombinant protein obtained by conventional cloning and expression systems. A typical set of cloning and expression systems includes *E. coli,* Hi5 or SF9 insect cells and e.g. baculuvirus based expression plasmids or non-viral insect cells expression system (e.g. transient transfection to S2 insect cell line using ExpreS2), Saccharomyces cerevisiae, Pichia pastoris, mammalian cell lines (e.g. HEK293T, CHO). The protein can also be generated using cell free transcription/translation or translation systems (e.g. based on wheat germ). cGAS can be expressed as full-length protein or as a truncated version comprising the catalytic domain. Furthermore, cGAS can be engineered to have alteration in its amino acid sequence. For example, cGAS can be engineered in its active catalytic domain resulting in increased or decreased enzymatic activity of cGAS compared to the activity of unmodified cGAS. Exemplarily techniques to engineer cGAS include but are not limited to site directed mutagenesis using polymerase chain reaction, cassette ligation, ligation-during-amplification.

[0151] The protein may further contain affinity tags allowing purification and detection of the expressed protein.

[0152] According to the methods of the invention, cGAS is derived from any vertebrate, including mammals such as human, macaque, mouse, sus, bos, and the like, chicken, duck, zebra fish and *Xenopus laevis.*

[0153] In a preferred embodiment of the invention, cGAS is derived from human. The nucleotide sequence of full-length human cGAS is defined by SEQ ID NO: 9.

SEQ ID NO: 9

[0154]

ATGCAGCCTTGGCACGGAAAGGCCATGCAGAGAGCTTCCGAGGCCGGAGCCAC
TGCCCCCAAGGCTTCCGCACGGAATGCCAGGGGCGCCCCGATGGATCCCACCG
AGTCTCCGGCTGCCCCCGAGGCCGCCCTGCCTAAGGCGGGAAAGTTCGGCCCC
GCCAGGAAGTCGGGATCCCGGCAGAAAAAGAGCGCCCCGGACACCCAGGAGAG
GCCGCCCGTCCGCGCAACTGGGGCCCGCGCCAAAAAGGCCCCTCAGCGCGCCC
AGGACACGCAGCCGTCTGACGCCACCAGCGCCCCTGGGGCAGAGGGGCTGGAG
CCTCCTGCGGCTCGGGAGCCGGCTCTTTCCAGGGCTGGTTCTTGCCGCCAGAG
GGGCGCGCGCTGCTCCACGAAGCCAAGACCTCCGCCCGGGCCCTGGGACGTGC
CCAGCCCCGGCCTGCCGGTCTCGGCCCCCATTCTCGTACGGAGGGATGCGGCG
CCTGGGGCCTCGAAGCTCCGGGCGGTTTTGGAGAAGTTGAAGCTCAGCCGCGA
TGATATCTCCACGGCGGCGGGGATGGTGAAAGGGGTTGTGGACCACCTGCTGC
TCAGACTGAAGTGCGACTCCGCGTTCAGAGGCGTCGGGCTGCTGAACACCGGG
AGCTACTATGAGCACGTGAAGATTTCTGCACCTAATGAATTTGATGTCATGTT
TAAACTGGAAGTCCCCAGAATTCAACTAGAAGAATATTCCAACACTCGTGCAT
ATTACTTTGTGAAATTTAAAAGAAATCCGAAAGAAAATCCTCTGAGTCAGTTT
TTAGAAGGTGAAATATTATCAGCTTCTAAGATGCTGTCAAAGTTTAGGAAAAT
CATTAAGGAAGAAATTAACGACATTAAAGATACAGATGTCATCATGAAGAGGA
AAAGAGGAGGGAGCCCTGCTGTAACACTTCTTATTAGTGAAAAAATATCTGTG
GATATAACCCTGGCTTTGGAATCAAAAAGTAGCTGGCCTGCTAGCACCCAAGA
AGGCCTGCGCATTCAAAACTGGCTTTCAGCAAAAGTTAGGAAGCAACTACGAC
TAAAGCCATTTTACCTTGTACCCAAGCATGCAAAGGAAGGAAATGGTTTCCAA
GAAGAAACATGGCGGCTATCCTTCTCTCACATCGAAAAGGAAATTTTGAACAA
TCATGGAAAATCTAAAACGTGCTGTGAAAACAAAGAAGAGAAATGTTGCAGGA
AAGATTGTTTAAAACTAATGAAATACCTTTTAGAACAGCTGAAAGAAAGGTTT
AAAGACAAAAAACATCTGGATAAATTCTCTTCTTATCATGTGAAAACTGCCTT
CTTTCACGTATGTACCCAGAACCCTCAAGACAGTCAGTGGGACCGCAAAGACC
TGGGCCTCTGCTTTGATAACTGCGTGACATACTTTCTTCAGTGCCTCAGGACA
GAAAAACTTGAGAATTATTTTATTCCTGAATTCAATCTATTCTCTAGCAACTT
AATTGACAAAAGAAGTAAAGAATTTCTGACAAAGCAAATTGAATATGAAAGAA
ACAATGAGTTTCCAGTTTTTGATGAATTTTGA

[0155] SEQ IN NO: 9 as shown above includes the terminal TGA stop codon. The stop codon may be useful in expressing the protein as a discrete product. However, there might be instances where the protein is to be expressed together with at least one other protein product, e.g. as a protein fusion. In such instances, it is possible to remove the terminal TGA stop codon, replacing it with a nucleic acid sequence encoding the other protein(s) of interest. The skilled person understands under what circumstances the terminal stop codon TGA is to be retained or dispensed with, and is readily able to reproduce the above sequence in the presence or absence of the terminal stop codon TGA.

[0156] In some embodiments of the invention, the identity between cGAS nucleotide sequence and SEQ ID NO: 9 is at least about 50% to about 100% or between about 50% and about 100%. In some embodiments of the invention, the identity between cGAS nucleotide sequence and SEQ ID NO: 9 is at least: about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a percent identity range. Also contemplated within such percent identity ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between cGAS nucleotide sequence and SEQ ID NO: 9 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled

person will understand that any cGAS nucleotide sequence having the above recited sequence identities to SEQ ID NO: 9 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative enzymatic activity as cGAS defined by SEQ ID NO: 9. The skilled person can incorporate and express sequences comprising the above recited sequence identities to SEQ ID NO: 9 in recombinant expression systems according to known methods in the art, in order to obtain recombinant cGAS for use in the present invention.

[0157] The amino acid sequence of full-length human cGAS is defined by SEQ ID NO: 10.

SEQ ID NO: 10

[0158]

MQPWHGKAMQRASEAGATAPKASARNARGAPMDPTESPAAPEAALPKAGKFGP
ARKSGSRQKKSAPDTQERPPVRATGARAKKAPQRAQDTQPSDATSAPGAEGLE
PPAAREPALSRAGSCRQRGARCSTKPRPPPGPWDVPSPGLPVSAPILVRRDAA
PGASKLRAVLEKLKLSRDDISTAAGMVKGVVDHLLLRLKCDSAFRGVGLLNTG

SYYEHVKISAPNEFDVMFKLEVPRIQLEEYSNTRAYYFVKFKRNPKENPLSQF
LEGEILSASKMLSKFRKIIKEEINDIKDTDVIMKRKRGGSPAVTLLISEKISV
DITLALESKSSWPASTQEGLRIQNWLSAKVRKQLRLKPFYLVPKHAKEGNGFQ
EETWRLSFSHIEKEILNNHGKSKTCCENKEEKCCRKDCLKLMKYLLEQLKERF
KDKKHLDKFSSYHVKTAFFHVCTQNPQDSQWDRKDLGLCFDNCVTYFLQCLRT
EKLENYFIPEFNLFSSNLIDKRSKEFLTKQIEYERNNEFPVFDEF

[0159] In some embodiments of the invention, the identity between cGAS amino acid sequence and SEQ ID NO: 10 is at least about 36% to about 100% or between about 36% and about 100%. In some embodiments of the invention, the identity between cGAS amino acid sequence and SEQ ID NO: 10 is at least: about 36%, about 37%, about 38%, about 39%, about 40%, about 41%, about 42%, about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a percent identity range. Also contemplated within such percent identity ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between cGAS amino acid sequence and SEQ ID NO: 10 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any cGAS amino acid sequence having the above recited sequence identities to SEQ ID NO: 10 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative enzymatic activity as cGAS defined by SEQ ID NO: 10. The skilled person can incorporate and express sequences comprising the above recited sequence identities to SEQ ID NO: 10 in recombinant expression systems according to known methods in the art, in order to obtain recombinant cGAS for use in the present invention.

[0160] As can be seen from Figure 3 human cGAS comprises a highly conserved C-terminal domain (catalytic domain - CD, aa 160-513) containing a Mab21 domain (aa 213 to 513 of human cGAS) and comprising both DNA-binding and enzymatic activity. The N-terminus of human cGAS is unstructured and not well conserved. The inventors found out that truncated human cGAS consisting of the highly conserved C-terminal domain (i.e. aa 155 to 522 of human cGAS) but missing the unstructured N-terminal domain has improved characteristics when used in the herein described methods of the invention. Specifically, truncated cGAS is less prone to proteolytic degradation and aggregation.

[0161] Thus, in a further preferred embodiment of the invention cGAS is a truncated version of the human full length protein, wherein the first 154 amino acid sequence encoding for the unstructured N-terminal domain of human cGAS is missing. The nucleotide sequence of truncated human cGAS is defined by SEQ ID NO: 11.

SEQ ID NO: 11

[0162]

CGGAGGGATGCGGCGCCT
GGGGCCTCGAAGCTCCGGGCGGTTTTGGAGAAGTTGAAGCTCAGCCGCGATGA
TATCTCCACGGCGGCGGGGATGGTGAAAGGGGTTGTGGACCACCTGCTGCTCA
GACTGAAGTGCGACTCCGCGTTCAGAGGCGTCGGGCTGCTGAACACCGGGAGC
TACTATGAGCACGTGAAGATTTCTGCACCTAATGAATTTGATGTCATGTTTAA
ACTGGAAGTCCCCAGAATTCAACTAGAAGAATATTCCAACACTCGTGCATATT
ACTTTGTGAAATTTAAAAGAAATCCGAAAGAAAATCCTCTGAGTCAGTTTTTA
GAAGGTGAAATATTATCAGCTTCTAAGATGCTGTCAAAGTTTAGGAAAATCAT
TAAGGAAGAAATTAACGACATTAAAGATACAGATGTCATCATGAAGAGGAAAA
GAGGAGGGAGCCCTGCTGTAACACTTCTTATTAGTGAAAAAATATCTGTGGAT
ATAACCCTGGCTTTGGAATCAAAAAGTAGCTGGCCTGCTAGCACCCAAGAAGG
CCTGCGCATTCAAAACTGGCTTTCAGCAAAAGTTAGGAAGCAACTACGACTAA
AGCCATTTTACCTTGTACCCAAGCATGCAAAGGAAGGAAATGGTTTCCAAGAA
GAAACATGGCGGCTATCCTTCTCTCACATCGAAAAGGAAATTTTGAACAATCA
TGGAAAATCTAAAACGTGCTGTGAAAACAAAGAAGAGAAATGTTGCAGGAAAG
ATTGTTTAAAACTAATGAAATACCTTTTAGAACAGCTGAAAGAAAGGTTTAAA
GACAAAAAACATCTGGATAAATTCTCTTCTTATCATGTGAAAACTGCCTTCTT
TCACGTATGTACCCAGAACCCTCAAGACAGTCAGTGGGACCGCAAAGACCTGG
GCCTCTGCTTTGATAACTGCGTGACATACTTTCTTCAGTGCCTCAGGACAGAA
AAACTTGAGAATTATTTTATTCCTGAATTCAATCTATTCTCTAGCAACTTAAT
TGACAAAAGAAGTAAAGAATTTCTGACAAAGCAAATTGAATATGAAAGAAACA
ATGAGTTTCCAGTTTTTGATGAATTTTGA

[0163] SEQ IN NO: 11 as shown above includes the terminal TGA stop codon. The stop codon may be useful in expressing the protein as a discrete product. However, there might be instances where the protein is to be expressed together with at least one other protein product, e.g. as a protein fusion. In such instances, it is possible to remove the terminal TGA stop codon, replacing it with a nucleic acid sequence encoding the other protein(s) of interest. The skilled person understands under what circumstances the terminal stop codon TGA is to be retained or dispensed with, and is readily able to reproduce the above sequence in the presence or absence of the terminal stop codon TGA.

[0164] In some embodiments of the invention, the identity between truncated cGAS nucleotide sequence and SEQ ID NO: 11 is about 56% to about 100% or between about 56% and about 100%. In some embodiments of the invention, the identity between truncated cGAS nucleotide sequence and SEQ ID NO: 11 is at least: about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form percent identity a range. Also contemplated within such percent identity ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between truncated cGAS nucleotide sequence and SEQ ID NO: 11 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any cGAS nucleotide sequence having the above recited sequence identities to SEQ ID NO: 11 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative enzymatic activity as cGAS defined by SEQ ID NO: 11. The skilled person can incorporate and express sequences comprising the above recited sequence identities to SEQ ID NO: 11 in recombinant expression systems according to known methods in the art, in order to obtain recombinant cGAS for use in the present invention.

[0165] The amino acid sequence of truncated cGAS is defined by SEQ ID NO: 12.

SEQ ID NO: 12

[0166]

RRDAAPGASKLRAVLEKLKLSRDDISTAAGMVKGVVDHLLLRLKCDSAFRGVG
LLNTGSYYEHVKISAPNEFDVMFKLEVPRIQLEEYSNTRAYYFVKFKRNPKEN
PLSQFLEGEILSASKMLSKFRKIIKEEINDIKDTDVIMKRKRGGSPAVTLLIS
EKISVDITLALESKSSWPASTQEGLRIQNWLSAKVRKQLRLKPFYLVPKHAKE
GNGFQEETWRLSFSHIEKEILNNHGKSKTCCENKEEKCCRKDCLKLMKYLLEQ

LKERFKDKKHLDKFSSYHVKTAFFHVCTQNPQDSQWDRKDLGLCFDNCVTYFL
QCLRTEKLENYFIPEFNLFSSNLIDKRSKEFLTKQIEYERNNEFPVFDEF

[0167] In some embodiments of the invention, the identity between truncated cGAS amino acid sequence and SEQ ID NO: 12 is about 40% to about 100% or between about 40% and about 100%. In some embodiments of the invention, the identity between truncated cGAS amino acid sequence and SEQ ID NO: 12 is at least: about 40%, about 41%, about 42%, about 43%, about 44%, about 45%, about 46%, about 47%, about 48%, about 49%, about 50%, about 51%, about 52%, about 53%, about 54%, about 55%, about 56%, about 57%, about 58%, about 59%, about 60%, about 61%, about 62%, about 63%, about 64%, about 65%, about 66%, about 67%, about 68%, about 69%, about 70%, about 71%, about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, bout 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%; or is 100%. Any of the above recited percent identities may be combined with one another to form a percent identity range. Also contemplated within such percent identity ranges are corresponding ranges in which the lower percent identity is included while the upper percent identity is excluded, as well as ranges in which the lower percent identity is excluded while the upper percent identity is included. In a particularly preferred embodiment of the invention, the identity between truncated cGAS amino acid sequence and SEQ ID NO: 12 is about 100%, more preferably is 100%. In view of the teaching of the present invention, the skilled person will understand that any cGAS amino acid sequence having the above recited sequence identities to SEQ ID NO: 12 may be used according to the present invention, provided that such sequence still has the same qualitative or qualitative and quantitative enzymatic activity as cGAS defined by SEQ ID NO: 12. The skilled person can incorporate and express sequences comprising the above recited sequence identities to SEQ ID NO: 12 in recombinant expression systems according to known methods in the art, in order to obtain recombinant cGAS for use in the present invention.

[0168] In order to convert nucleoside triphosphates into 2'3'CDN cGAS must be activated by a cGAS-activating nucleic acid which induces complex formation of cGAS with the cGAS-activating nucleic acid, thereby activating the enzyme.

[0169] According to the methods of the invention, a cGAS-activating nucleic acid comprises double-stranded (ds) and single-stranded (ss) DNA. Furthermore, the cGAS-activating nucleic acid may be an RNA/DNA hybrid, preferably a poly(A):poly(dT) or poly(dA):poly(T) hybrid.

[0170] In a preferred embodiment of the present invention, the cGAS-activating nucleic acid, e.g double stranded DNA, may have a length at least 10 base pairs (bp). In a particular embodiment of the present invention, the cGAS-activating nucleic acid, e.g. dsDNA, may have a length selected from the group consisting of at least 10bp, 11bp, 12bp, 13bp, 14bp, 15bp, 16bp, 17bp, 18bp, 19bp, 20bp, 21bp, 22bp, 23bp, 24bp, 25bp, 26bp, 27bp, 28bp, 29bp, 30bp, 31bp, 32bp, 33bp, 34bp, 35bp, 36bp, 37bp, 38bp, 39bp, 40bp, 41bp, 42bp, 43bp, 44bp, 45bp, 46bp, 47bp, 48bp, 49bp, 50bp, 51bp, 52bp, 53bp, 54bp, 55bp, 56bp, 57bp, 58bp, 59bp, 60bp, 61bp, 62bp, 63bp, 64bp, 65bp, 66bp, 67bp, 68bp, 69bp, 70bp, 71bp. 72bp, 73bp, 74bp, 75bp, 76bp, 77bp, 78bp, 79bp, 80bp, 81bp, 82bp, 83bp, 84bp, 85bp, 86bp, 87bp, 88bp, 89bp, 90bp, 91bp, 92bp, 93bp, 94bp, 95bp, 96bp, 97bp, 98bp, 99bp, and 100bp. The double-stranded cGAS-activating nucleic acid may also have a minimal length of at least 200bp, 300bp, 400bp, 500bp, 600bp, 700bp, 800bp, 900bp, 1kbp, 2kbp, 3kbp, 4kbp, 5kbp, 6kbp and the like. Corresponding lengths also apply in the event the cGAS-activating nucleic acid, e.g. DNA, is single-stranded. The double-stranded cGAS-activating nucleic acid may be DNA, for example a plasmid, such as a commonly known expression vector. The cGAS-activating nucleic acid may also be in the form of single-stranded DNA or DNA/RNA hybrids. Such DNA/RNA hybrids may also have the lengths as set out above. Any of the above recited length of base pairs may be combined with one another to form a length of base pair range. Also contemplated within such length of base pair ranges are corresponding ranges in which the lower length of base pairs is included while the upper length of base pairs is excluded, as well as ranges in which the upper length of base pairs is excluded while the upper length of base pairs is included.

[0171] In a further embodiment of the invention, the cGAS-activating nucleic acid is a dsDNA molecule in the form of

a plasmid, such as a commonly known expression vector pET, Phil 74, pGEX, pcDNA5, pFBDM, pRS, pETM.

[0172] In a preferred embodiment of the invention, the cGAS-activating nucleic acid is a dsDNA molecule having at least 12bp flanked by guanosine-rich (G-rich) ssDNA overhangs. G-rich overhangs significantly enhance cGAS binding to short DNA.

[0173] In a further preferred embodiment of the invention, the cGAS-activating nucleic acid is a dsDNA molecule having at least 30bp. Below 30bp dsDNA molecules are less efficient in activating cGAS.

[0174] Furthermore, cGAS requires divalent cations as co-factors in order to convert nucleoside triphosphates into 2'3'CDNs. Thus, in a preferred embodiment of the present invention, the one or more divalent cation is selected from the group consisting of Mg^{2+}, Ca^{2+}, Zn^{2+}, Fe^{2+}, Cu^{2+}, Ni^{2+}, and Mn^{2+}. In a most preferred embodiment of the present invention, the aqueous solution further comprises Mg^{2+} and/or Zn^{2+}.

[0175] According to one embodiment, cGAS may be activated by the addition of one or more divalent cations to the aqueous solution already comprising cGAS, a cGAS-activating nucleic acid, first nucleoside triphosphate, and a second nucleoside triphosphate, wherein at least one of the first and second nucleoside triphosphate is a fluorescent nucleoside triphosphate according to the present invention. The conversion of the first and second nucleoside triphosphates into 2'3'fCDN then starts as soon as the first cGAS/cGAS-activating nucleic acid complex is formed. As soon as the first cGAS/cGAS-activating nucleic acid complex is formed, the conversion rate of the first and second nucleoside triphosphates into 2'3'fCDN continuously increases over time as more cGAS molecules become activated. When cGAS activation reaches equilibrium, the conversion rate of the first and second nucleoside triphosphate into 2'3'fCDN reaches a stationary phase. This allows fast and direct measurement of cGAS activity and the identification of a substance having the ability to modulate cGAS activity. Thus, in one embodiment of the present invention, one or more divalent cations are added to the aqueous solution comprising cGAS, cGAS-activating nucleic acid, and a first and second nucleoside triphosphate of the present invention to start the 2'3'fCDN-forming reaction.

[0176] According to another embodiment, cGAS may alternatively be activated prior to the addition of a first and a second nucleoside triphosphates of the present invention. This approach allows cGAS activation to achieve equilibrium prior to the addition of its substrates and therefore allows accurate measurement of the initial conversion rates of the nucleoside triphosphates into 2'3'fCDN. Thus, in a further preferred embodiment of the present invention, cGAS is activated prior to the addition of the first and second nucleoside triphosphates in the aqueous solution comprising cGAS, cGAS-activating nucleic acid and one or more divalent cations. After cGAS activation reaches equilibrium, the first and second nucleoside triphosphate according to the present invention are added to the aqueous solution, wherein at least one of the first and second nucleoside triphosphate is a fluorescent nucleoside triphosphate according to the present invention. The skilled person will be able to measure DNA binding equilibrium by commonly known techniques such as EMSA, ITC or fluorescence anisotropy measurements using a fluorophore attached to DNA.

[0177] In a further embodiment of the present invention, the fluorescent signal of the aqueous solution is measured at a predetermined time-point. This time-point is determined in preliminary experiments taking into account the dead time of the instrument used as the fluorescence reader, i.e. the shortest time it takes between mixing the reagents and measuring the first data point in a given instrument. It is advantageous to determine cGAS activity in a preceding experiment, e.g. by a method of measuring cGAS activity described above, to avoid choosing a time point of measuring the fluorescence signal when either the conversion of the fluorescent nucleoside triphosphate of the present invention into the 2'3'fCDN has not yet started (this would be the case for cGAS having a very low nucleotidyltransferase activity) or all of the free fluorescent nucleoside triphosphate in the aqueous solution has already been converted into 2'3'fCDN due to very high nucleotidyltransferase activity of cGAS molecules. Once the correct time point is determined for each experimental set up, measuring of the fluorescent signal at a predetermined time point allows fast and reliable measurement of the fluorescent signal in aqueous solution. According to one embodiment of the invention, this allows fast and reliable identification of substances having an agonistic or antagonistic effect on cGAS activity.

[0178] In another embodiment of the present invention, the fluorescence signal of the aqueous solution is measured continuously over a predetermined time interval. Measuring over a time interval, the slope can be determined more precisely than using a single time point and the linear interval of the reaction can be determined more accurately. In some embodiments of the present invention, the fluorescent signal is measured starting from about 10 seconds from mixing all compounds together up to about 1 hour. The time interval of the measurement depends strongly on the experimental setup, such as buffer condition, temperature and substance concentration. In general, for the low cGAS activity higher protein and/or activating DNA and/or substance concentrations may be used, compared to the case of high cGAS activity to perform the measurement within 5 min - 1 h time interval.

[0179] In a further embodiment of the present invention, the measured fluorescence signal of the aqueous solution is measured at different wavelengths or ranges of wavelengths depending on the fluorescent nucleoside triphosphate or triphosphates used in the methods according to the invention. Specifically, each fluorescent nucleoside triphosphate may have different excitation and emission maxima which influence the wavelengths at which the fluorescence signal is measured. Thus, depending on the fluorescent nucleoside triphosphate or triphosphates used in the methods according to the invention the wavelengths or range of wavelengths at which the fluorescent nucleoside triphosphate is excited

(i.e. the excitation maxima) and the wavelengths or range of wavelengths at which the fluorescence signal is measured (i.e. the emission maxima) need to be adapted for each fluorescent nucleoside triphosphate or triphosphates individually in order to avoid false positives by measuring a fluorescence signal outside the optimal wavelength or range of wavelengths for a given fluorescent nucleoside triphosphate or triphosphates.

[0180] In some embodiments of the present invention, the excitation wavelength is in the range of about 295nm to about 365nm, or between about 295nm and 365nm. In some embodiments of the invention, the excitation wavelength is about 295nm, about 296nm, about 297nm, about 298nm, about 299nm, about 300nm, about 301nm, about 302nm, about 303nm, about 304nm, about 305nm, about 306nm, about 307 nm, about 308nm, about 309nm, about 310nm, about 311nm, about 312nm, about 312nm, about 313nm, about 314nm, about 315nm, about 316nm, about 317nm, about 318nm, about 319nm, about 320nm, about 321nm, about 322nm, about 323nm, is about 324nm, about 325nm, about 326nm, about 327nm, about 329nm, about 330nm, about 331nm, about 332nm, about 333nm, about 334nm, about 335nm, about 336nm, about 337nm, about 338nm, about 339nm, about 340nm, about 341nm, about 342nm, about 343nm, about 344nm, about 345nm, about 346nm, about 347nm, about 348nm, about 349nm, about 350nm, about 351nm, about 352nm, about 353nm, about 354nm, about 355nm, about 356nm, about 357nm, about 358nm, about 359nm, about 360nm, about 361nm, about 362nm, about 363nm, about 364nm, or about 365nm. Any of the above recited excitation wavelength may be combined with one another to form an excitation wavelength range. Also contemplated within such excitation wavelength ranges are corresponding ranges in which the lower excitation wavelength is included while the upper excitation wavelength is excluded, as well as ranges in which the lower excitation wavelength is excluded while the upper excitation wavelength is included. In a particularly preferred embodiment of the invention, the excitation wavelength ranges from about 295nm to about 315nm or between about 295nm and 315nm. In a most preferred embodiment the excitation wavelength is 305nm.

[0181] In some embodiments of the present invention, the emission wavelength ranges from about 350nm to about 500nm or between about 350nm and 500nm. In some embodiments the excitation wavelength is about 350nm, about 351nm, about 352nm, about 353nm, about 354nm, about 355nm, about 356nm, about 357nm, about 358nm, about 359nm, about 360nm, about 361nm, about 362nm, about 363nm, about 364nm, about 365nm, about 366nm, about 367nm, about 368nm, about 369nm, about 370nm, about 371nm, about 372nm, about 373nm, about 374nm, about 375nm, about 376nm, about 377nm, about 378nm, about 379nm, about 380nm, about 381nm, about 382nm, about 383nm, about 384nm, about 385nm, about 386nm, about 387nm, about 388nm, about 389nm, about 390nm, about 391nm, about 392nm, about 393nm, about 394nm, about 395nm, about 396nm, about 397nm, about 398nm, about 399nm, about 400nm, about 401nm, about 402nm, about 403nm, about 404nm, about 405nm, about 406nm, about 407nm, about 408nm, about 409nm, about 410nm, about 411nm, about 412nm, about 413nm, about 414nm, about 415nm, about 416nm, about 417nm, about 418nm, about 419nm, about 420nm, about 421nm, about 422nm, about 423nm, about 424nm, about 425nm, about 426nm, about 427nm, about 428nm, about 429nm, about 430nm, about 431nm, about 432nm, about 433nm, about 434nm, about 435nm, about 436nm, about 437nm, about 438nm, about 439nm, about 440nm, about 441nm, about 442nm, about 443nm, about 444nm, about 445nm, about 446nm, about 447nm, about 448nm, about 449nm, about 450nm, about 451nm, about 452nm, about 453nm, about 454nm, about 455nm, about 456nm, about 457nm, about 458nm, about 459nm, about 460nm, about 461nm, about 462nm, about 463nm, about 464nm, about 465nm, about 466nm, about 467nm, about 468nm, about 469nm, about 470nm, about 471nm, about 472nm, about 473nm, about 474nm, about 475nm, about 476nm, about 477nm, about 478nm, about 479nm, about 480nm, about 481nm, about 482nm, about 483nm, about 484nm, about 485nm, about 486nm, about 487nm, about 488nm, about 489nm, about 490nm, about 491nm, about 492nm, about 493nm, about 494nm, about 495nm, about 496nm, about 497nm, about 498nm, about 499nm or about 500nm. Any of the above recited emission wavelengths may be combined with one another to form an emission wavelength range. Also contemplated within such emission wavelength ranges are corresponding ranges in which the lower emission wavelength is included while the upper emission wavelength is excluded, as well as ranges in which the lower emission wavelength is excluded while the upper emission wavelength is included. In a particularly preferred embodiment of the invention, the emission wavelength ranges from about 350nm to about 380nm, or between about 350nm and 380nm. In a most preferred embodiment of the invention the emission wavelength is 363nm.

[0182] Furthermore, the quantum yield of a fluorescent nucleoside triphosphate may also differ depending on the experimental conditions. Thus, for each fluorescent nucleoside triphosphate the experimental conditions need to be modified in order to optimize the quantum yield of the fluorescent molecule. Experimental conditions that might influence the quantum yield of a given fluorescent nucleoside triphosphate of include but are not limited to pH, temperature, ionic strength, buffer composition

[0183] The change of the fluorescence signal can be measured by any known fluorescence read out technique using a conventional plate reader or a single-cuvette photometer. In a preferred embodiment of the present invention, the fluorescence signal of the aqueous solution may be measured by a Tecan infinite M1000 using 96-well black non-binding PS plates (Greiner Bio-One), however a FluoroMax-P spectrofluorometer (Horiba Scientific) or similar instrument can be used.

[0184] According to the invention, the fluorescent nucleoside triphosphate is a fluorescent and purine nucleoside triphosphate. In one embodiment of the present invention, the fluorescent nucleoside triphosphate can be selected from the group consisting of a 2-aminopurine nucleoside triphosphate (2-APTP), a 3-methyl-isoxanthopterin nucleoside triphosphate (3-MITP), a 6-methyl isoxanthopterin nucleoside triphosphate (6-MITP), 4-amino-6-methyl-8-(2-deoxy-beta-d-ribofuranosyl)-7(8H)-pteridone nucleoside triphosphate (6-MAPTP), a 4-amino-2,6-dimethyl-8-(2'-deoxy-beta-d-ribofuranosyl)-7(8H)-pteridone nucleoside triphosphate (DMAPTP), a pyrrolocytosine nuceloside triphosphate (pyrrolo-CTP), a 6-phenylpyrrolocytosine nucleoside triphosphate (PhpCTP), a (aminoethoxy)phenylpyrrolocytosine nucleoside triphosphate (moPhpCTP), a [bis-o-(aminoethoxy)phenyl]pyrrolocytosine nucleoside triphosphate (boPhpCTP), a hydropyrimidopyrimidine nucleoside triphosphate ($C^{hpp}TP$), a pyrrolopyrimidopyrimidine nucleoside triphosphate ($C^{ppp}TP$), a pyrimidopyrimidoindole nucleoside triphosphate ($C^{ppi}TP$), a benzopyridopyrimidine nucleoside triphosphate (BPPTP), a naphtopyridopyrimidine nucleoside triphosphate (NPPTP), a methoxybenzodeazaadenine nucleoside triphosphate (^{MD}ATP), a methoxybenzodeazainosine nucleoside triphosphate (^{MD}ITP), a naphtodeazaadenine nucleoside triphosphate (^{ND}ATP), a furan-modified pyrimidine nucleoside states, a thieno[3,2]pyrimidine nucleoside triphosphate, a thieno[3,4]pyrimidine triphosphate, a 5-methoxy-quinazoline-2,4-(1H,3H) dione nucleoside triphosphate, a 5-methylpyrimidine-2-one nucleoside triphosphate, a 7-deazapurine nucleoside triphosphate, a 5-alkyluridine nucleoside triphosphate, a benzoquinalozines nucleoside triphosphate, a triazoleadenosine nucleoside triphosphate, and a $1,N^6$-ethenoadenosine nucleoside triphosphate, wherein the nucleoside triphosphate is a deoxyribonucleoside triphosphate or a ribonucleoside triphosphate. In a most preferred embodiment, the fluorescent nucleoside triphosphate is 2-amino purine ribonucleoside triphosphate (2-APTP) comprising the fluorescent nucleobase analogue 2-aminopurine (2-AP). The inventors surprisingly found out that the high fluorescence quantum yield of 2-APTP in aqueous solution (about 0.68) is considerably reduced when incorporated into 2'3'fCDN. The high sensitivity of 2-AP to the microenvironment can thus be exploited to measure cGAS activity in the presence or absence of a substance. In particular, the inventors found out that cGAs can utilize 2-APTP instead of adenine nucleoside triphosphate (ATP) to efficiently generate 2'3'fCDN in the presence of a second nucleoside triphosphate. The quenching of the fluorescence signal upon incorporation of 2-APTP into 2'3'fCDN allows the measurement of cGAS activity in a continuous assay using a conventional fluorescence plate reader or single cuvette photometer. The finding that 2-APTP is utilized by cGAS for incorporation into 2'3'fCDN in conjunction with a significant quenching of the fluorescence signal allows the determination of quantitative steady state activity rates of cGAS in an easily scalable and high-throughput manner. Based on the unexpected properties of 2-APTP in the methods of the present invention, the inventors were thus able to develop a fast and reliable method of measuring cGAS activity and a method to screen a large range of substances for their ability to modulate cGAS activity in a fast and reliable manner.

[0185] In a preferred embodiment of the present invention, the first nucleoside triphosphate is 2-APTP.

[0186] As evident from all of the above, the fluorescent nucleoside triphosphate can be used for measuring cGAS activity or for identifying a substance having an ability to modulate cGAS activity. The fluorescent nucleoside triphosphate therefore represents a powerful tool to measure cGAS activity and to identify substances having an ability to modulate cGAS activity in a fast and reliable method according to the invention.

[0187] In a further embodiment of the present invention, the second nucleoside triphosphate is a purine ribonucleoside triphosphate. The ribofuranosyl backbone of the second nucleoside triphosphate contains a hydroxyl group at position 2 of the ribose ring, thereby allowing the formation of the 2'5'-phosphodiester linkage between the first and the second nucleoside triphosphate. It is essential that one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group in order to enable the formation of the 2'-5' and 3'-5' phosphodiester linkages in the 2'3'CDN.

[0188] In a most preferred embodiment of the present invention, the second nucleoside triphosphate is guanosine triphosphate (GTP) having a free 2' hydroxyl group, whereas the first nucleoside triphosphate is 2-APTP having a free 3' hydroxyl group.

[0189] According to the methods of the present invention, the skilled person will realize that the fluorescent nucleoside triphosphates set forth herein can be used for measuring cGAS activity and for identifying a substance having an ability to modulate cGAS activity.

Method of preparing a fluorescent cyclic dinucleotide

[0190] In a further aspect, the present invention relates to a method of preparing a fluorescent cyclic dinucleotide (2'3'fCDN) of the present invention. The method comprises the steps of (i) providing an aqueous solution comprising cGAS, a cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphate is a fluorescent nucleoside triphosphate of the present invention, and wherein one of the first and second nucleoside triphosphate has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group, thereby preparing the cyclic dinucleotide of the present invention. The method is based on the finding that activated cGAS generates cyclic dinucleotides by utilizing fluorescent

nucleoside triphosphates of the present invention instead of natural occurring nucleoside triphosphates with only slightly reduced efficiency.

[0191] In a preferred embodiment of the present invention, the fluorescent cyclic dinucleotide of the present invention is further purified. For example, the reaction mixture of the method of preparation can be subjected to commonly known purification techniques such as reversed-phase HPLC, ion exchange chromatography, solid phase extraction, and the like.

[0192] In a further embodiment of the invention, the method of preparing a 2'3'fCDN further comprises additional steps of incubating the aqueous solution for a specified time period and under predetermined conditions sufficient to allow cGAS to convert the first and the second nucleoside triphosphate into a CDN of the invention. In particular, about 50 nM to about 50 μM, most preferably about 10 μM, cGAS is incubated with about 2.6 to 650 ng/μl, most preferably about 195 ng/μl, DNA in the aqueous solution comprising about 40 mM Tris, about 100 mM NaCl, about pH 7.5, about 5 mM $MgCl_2$, about 50 μM to about 5 mM of a first nucleoside triphosphate (i.e. a fluorescent nucleoside triphosphate), and about 50 μM to about 5 mM of a second nucleoside triphosphate. The reaction is performed at about 32°C to about 37°C for about 30 minutes to about 3 hours. Most preferably, the reaction mixture is incubated at about 37°C for about 60 minutes.

[0193] As evident from the overall teaching of the present invention, the 2'3'fCDN of the present invention can advantageously be used to label cGAS. Upon complex formation between cGAS and the 2'3'fCDN of the present invention, the cyclic dinucleotides can further be used to label cGAS due to its (at least one) fluorescent nucleotide monophosphate.

[0194] The following examples, including the experiments conducted and the results achieved, are provided for illustrative purposes only and are not construed as limiting the present invention.

Example 1: Cell line and reagents

[0195] The following cell lines and reagents were used in the herein described experiments. All DNA oligonucleotides were purchased from Metabion.

[0196] The following cell lines were used: HEK293T STING KI (Ablasser, Schmid-Burgk et al. "Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP". Nature 503, 530-534, 2013), BLaER1(Rapino, Robles et al. "C/EBPα Induces Highly Efficient Macrophage Transdifferentiation of B Lymphoma and Leukemia Cell Lines and Impairs Their Tumorigenicity". Cell Reports 3, 1153-1163, 2013) and BLaER1 cGAS-KO (cGAS deficient cell line, a general gift from M. Gaidt, group of V. Hornung, Gene Center, Ludwig-Maximilians-Universitat München, Center for Integrated Protein Science Munich, Germany).

Example 2: Constructs and cloning

[0197] The following constructs were used in the below described experiments:

The plasmid encoding full-length (aa 1-522) and truncated (catalytic domain - ^{cd}) *Homo sapiens* (h) (aa 155-522), *Mus musculus* (m) (aa 141-507) and *Sus scrofa* (s) (aa 135-495) cGAS for N-terminal His_6-MBP fusion protein expression were obtained from Dr. F. Civril (Civril, Deimling et al. "Structural mechanism of cytosolic DNA sensing by cGAS". Nature 498, 332-337, 2013). The plasmid encoding hSTING cytosolic domain (^{cd}, aa 139-379, R220H+H232R variant) for N-terminal His_6-SUMO1 fusion protein expression was obtained from Dr. T. Deimling (Cavlar, Deimling et al. "Species-specific detection of the antiviral small-molecule compound CMA by STING". The EMBO Jourhal 32, 1440-1450, 2013). These plasmids were used to transform *E.coli* Rosetta (DE3) protein expression strain cells (Novagen).

Example 3: Cell lines and cell culture

[0198] HEK293T STING KI were cultured in DMEM (Thermo Fisher Scientific or Sigma-Aldrich, respectively) supplemented with 10% heat inactivated FBS (Thermo Fisher Scientific or Biochrom, respectively) and incubated at 37°C with 5% CO_2.

[0199] BLaER1 and BLaER1 cGAS-KO cells were cultivated in RPMI medium containing heat-inactivated 10% FCS, Penicillin (100 U/ml), Streptomycin (100 μg/ml) (Gibco, Life Technologies) and 1 mM sodium pyruvate (Thermo Fisher Scientific). For trans-differentiation, 5 x 10^4 cells were seeded per well of a flat bottom 96 well plate and cultivated in the presence of ß-estradiol (100 nM, Sigma-Aldrich), hr-IL-3 (10 ng/ml) and M-CSF (10 ng/ml) (both PeproTech) for five days prior to the experiment (Gaidt, Ebert et al. "Human Monocytes Engage an Alternative Inflammasome Pathway". Immunity 44, 833-846, 2016).

Example 4: Protein expression and purification

[0200] All cGAS protein constructs (e.g. hcGAS^{cd}, mcGAS^{cd}, scGAS^{cd}) were overexpressed in *E. coli* Rosetta (DE3) for 16-18h at 18°C after induction with 0.2 mM IPTG. Cells were lysed by sonication in 50 mM Tris, 500 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% Glycerol, pH 7.5, supplemented with 2 mM β-mercaptoethanol and protease inhibitor cocktail (Sigma-Aldrich) and purified with Ni-NTA agarose resin (Qiagen). For truncated cGAS proteins His_6-MBP tag was removed with TEV protease (1:50 mass ratio) during 16 h dialysis against 30 mM Tris, 100 mM NaCl, 2 mM DTT, pH 7.0. cGAS proteins were further purified by cation-exchange chromatography (30 mM Tris, 100 mM/l M NaCl, 2 mM DTT, pH 7.0) on HiTrap SP HP columns (GE Healthcare) followed by size-exclusion chromatography on HiLoad S200 16/60 column (GE Healthcare) equilibrated with 20 mM Tris, 100 mM NaCl, pH 7.5. Full-length hcGAS with N-terminal His_6-MBP tag (hcGAS) and mcGAS proteins were concentrated to 8-12 mg ml^{-1}. Truncated hcGAS was concentrated to 4 mg ml^{-1}. All proteins were flash-frozen in liquid nitrogen and stored at -80°C.

[0201] mTfam and mHMGB1 (both DNA bending proteins) were purified as described for full-length hcGAS, except the cation-exchange chromatography step was skipped and after dialysis against 20 mM Tris, 300mM NaCl, pH 7.5 size exclusion chromatography on HiLoad S75 16/60 (GE Healthcare) column was performed. Proteins were concentrated to 10-13 mg ml^{-1}.

[0202] 1HU (another DNA bending protein) was purified as described for full-length hcGAS, except instead of cation-exchange chromatography affinity chromatography on HiTrap Heparin HP (GE Healthcare) column was performed (20 mM Tris, 100 mM/l M NaCl, 2 mM DTT, pH 7.5). HiLoad S75 16/60 (GE Healthcare) equilibrated with 20 mM Tris, 100 mM NaCl, pH 7.5 was used for size exclusion chromatography. Protein was concentrated to 7 mg ml^{-1}.

[0203] hSTING^{cd} construct was expressed and purified as follows. *E. coli* Rosetta (DE3) cells were grown in 31 of LB media supplemented with Kanamycin (50mg/l) and Chloramphenicol (34mg/l) at 37°C for 3 hours, to OD_{600}=0.8. Expression of N-terminal His_6-SUMO1-tagged STING was induced by adding IPTG (Roth) to a final concentration of 0.2mM. Expression was done overnight at 18°C. Cells were collected by centrifugation and were resuspended in lysis buffer (50mM Tris, 500mM NaCl, 10mM imidazole, 5% glycerol, 2mM β-mercaptoethanol, pH7.5) and lyzed by sonification. The soluble His_6-SUMO1-STING was purified by Ni-affinity chromatography. The His_6-SUMO1-tag was removed by proteolytic cleavage with SenP2 protease during dialysis over night (20mM Tris, 150mM NaCl, 3% glycerol, 2mM β-mercaptoethanol, pH7.5) and a second Ni-affinity chromatography step. To remove additional contaminants, a HiTrap Q FF (GE Healthcare) purification was applied. Finally, the STING containing flow-through was used in a HiLoad 16/60 Superdex 75 prep grade (GE Healthcare) size-exclusion chromatography step (20 mM Tris, 100 mM NaCl, pH 7.5). Purified STING was concentrated (9-15mg/ml) with a 10kDA cut-off centrifugal concentrator device (Millipore) and was flash frozen in liquid nitrogen for storage (-80°C).

Example 5: Characterization of cGAS activating nucleic acids

[0204] cGAS is activated by cytosolic DNA. In order to study the effect of different DNA constructs on cGAS activation, DNA of different length was tested in an enzyme-linked immunoabsorbent assay as follows (ELISA). This assay provided information about efficient cGAS-activating nucleic acids that could be used to induce robust cGAS activity in the herein described methods of the invention.

[0205] $5x10^4$ trans-differentiated BLaER1 cells expressing cGAS were transfected with 20, 40 or 60 ng DNA of different length (20-100 bp in 5 bp intervals) and herring testis (HT) DNA using 0.5 μl Lipofectamine 2000 Transfection Reagent (Thermo Fischer Scientific) in 50 μl Opti-MEM Reduced Serum medium (Thermo Fischer Scientific). CXCL 10 expression in the supernatants was quantified 8 h after transfection using ELISA (BD OptEIA™, human IP-10 ELISA Set). Cells stimulated with transfection reagent only and unstimulated cells served as control.

[0206] As can be seen from Figure 5a, the data indicate a concentration and DNA length-dependent activation of cGAS measuring CXCL10 production as a surrogate parameter of cGAS activity. Long herring testes (HT) DNA robustly activated cGAS at all DNA concentrations, while shorter DNA molecules required increasing concentrations of DNA. The data show a clear preference for long DNA at low DNA concentrations, while short DNA is more effective at higher DNA amounts.

[0207] The same effect was also observed in fluorescence-based cGAS activity assays described in Example 9. The data clearly indicate that at a constant concentration of both human full-length cGAS and human truncated cGAS (hcGAS^{cd}) a DNA length-dependent activation was observed which was independent of the presence or absence of cGAS N-terminal domain (Figure 5b). The data further indicate that robust activation in vitro proceeds most readily when DNA > 40bp (Figure 5b). The length-dependence was furthermore not species-specific, since truncated mouse cGAS (mcGAS^{cd}) showed a comparable length dependent activation, with a gradual increase in activity until about 75bp (Figure 5c). Plasmid DNA was an even more potent activator (Figure 5c).

[0208] Overall, the data indicate that cGAS is activated in a manner dependent on the length of the nucleic acid used, wherein cGAS activity directly correlates with the length of the tested DNA. Furthermore, the data indicate that the

nucleotide sequence does not influence cGAS activation and that the observed length-dependence was not species-specific. In summary, an increase in DNA length directly correlates with an increase in cGAS activity.

Example 6: Radiolabeled cGAS activity assays

[0209] The above results were confirmed by radiolabeled cGAS assays. In this assay 13 ng μl^{-1} DNA was mixed with 2 μM truncated mouse cGAS (mcGAS^{cd}) and reaction was started by adding 50 μM ATP, 500 μM GTP, 5 mM $MgCl_2$ in 40 mM Tris pH 7.5 and 100 mM NaCl containing 1:800 [α^{32}P]ATP (3000 Ci mmol-1, Hartman Analytic). Samples were incubated at 35°C and the reaction was stopped by plotting on PEI-Cellulose F plates (Merck) and analyzed by thin layer chromatography (TLC) with 1M $(NH_4)_2SO_4$/1.5M KH_2PO_4 as running buffer. The radiolabeled products were visualized with Typhoon FLA 9000 phosphor imaging system.

[0210] The results clearly confirmed the length-dependent cGAS activation already described in Example 5 above (see Figure 5d).

[0211] For testing of 2-APTP incorporation into cGAS enzymatic activity product 1 μM mcGAS^{cd} was mixed with 13 ng μl^{-1} ($\approx$1 μM binding sites) 55 bp, 500 μM 2-APTP/ATP, 500 μM GTP, 5 mM $MgCl_2$ and 1:160 [α^{32}P]ATP or [α^{32}P]GTP (3000 Ci mmol-1, Hartman Analytic) in the same condition.

[0212] The data demonstrate that cGAS is capable of incorporating 2-APTP together with GTP into 2'3'fGAMP (see Figure 6a).

Example 7: Analysis of chemical properties of 2'3'fGAMP

[0213] To analyze the physical and chemical properties of 2'3'fGAMP, the fluorescent cyclic dinucleotide was analyzed by anion-exchange chromatography. 10 μM truncated mouse cGAS (mcGAS^{cd}), truncated porcine cGAS (scGAS^{cd}) or truncated human cGAS (hcGAS^{cd}) were incubated with 195 ng μl^{-1} ($\approx$15 μM binding sites) plasmid DNA at 35°C for 2 h in buffer containing 40 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM 2-APTP and 2 mM GTP. Reaction mixtures were centrifuged for 10 min, 16100 rcf, and the supernatant was separated from cGAS by ultrafiltration (30 kDa, Amicon). Resulting flow through was diluted in 50 mM Tris pH 9.0 and loaded on Mono Q 5/50 GL (GE Healthcare) for anion exchange chromatography (50 mM Tris, 0 M/1 M NaCl, pH 9.0). The fractions containing cGAS product fGAMP were collected, lyophilized and stored at -20°C. Control runs with 2-APTP, GTP, 2'3'- and 3'3'cGAMP were made analogously to validate the resulting peaks of cGAS reaction products.

[0214] As can be seen from the chromatograms in Figures 6b and 6c, 2'3'fGAMP generated by truncated mouse (mcGAS^{cd}) or human cGAS (hcGAS^{cd}) was detected in the same elution volume as natural 2'3'cGAMP at approximately 10 mL, indicating that the fluorescent 2'3'fGAMP has similar chemical and physical properties as the natural 2'3'cGAMP.

Example 8: Analysis of enzymatically synthesized fGAMP by mass spectrometry.

[0215] To further determine the exact chemical composition of fGAMP produced by cGAS from 2-APTP and GTP, the resulting fGAMP was analyzed with mass spectrometry as follows. 10 μM truncated mouse cGAS (mcGAS^{cd}) was incubated with 195 ng μl^{-1} ($\approx$15 μM binding sites) plasmid DNA at 35°C for 2 h in buffer containing 40 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 2 mM 2-APTP and 2 mM GTP. Reaction mixtures were centrifuged for 10 min, 16100 rcf, and the supernatant was separated from cGAS by ultrafiltration (30 kDa, Amicon). Resulting flow through was diluted in 50 mM Tris pH 9.0 and loaded on Mono Q 5/50 GL (GE Healthcare) for anion exchange chromatography (50 mM Tris, 0 M/l M NaCl, pH 9.0). The fractions containing cGAS product fGAMP were collected, lyophilized and stored at -20°C.

[0216] For further purification reverse phase high-performance liquid chromatography (HPLC) was used with solvent A (triethylammonium acetate:acetonitrile:H_2O [0.1 M:3%:97%], w:v:v) and solvent B (methanol:acetonitrile:H_2O [45%:45%:10%], v:v:v). Lyophilized fGAMP was diluted in solvent A and subjected to a C18 column. Two step linear gradient 0%-10% B (2 CV), 10%-50% B (2 CV) was used as described in (Gao, Ascano et al. "Cyclic [G(2',5')pA(3',5')p] Is the Metazoan Second Messenger Produced by DNA-Activated Cyclic GMP-AMP Synthase". Cell 153,1094-1107,2013).

[0217] The sample was lyophilized and diluted in 100 μL H_2O for ESI LC/MS analysis.

[0218] Flow injection analysis with 10 μL fGAMP sample was performed with a Surveyor MS pump with a flow rate of 100 μL/min 20-80% (H_2O:acetonitrile) using an injection filter. ESI-LC/MS analysis was performed using a Finnigan LTQ FT Ultra Fourier Transform Ion Cyclotron Resonance mass spectrometer (Thermo Fisher Scientific) with an IonMax source with ESI head. The resolution was set to 100.000 at m/z 400 and the spectra were acquired in a mass range from m/z 150 to 1,000. The ion transfer capillary temperature was set to 250 °C, the spray voltage to 4 kV, liquid nitrogen Sheathgasflow to 20 and the Sweepgasflow to 5 units. The acquired spectrum of enzymatically synthesized fGAMP shows ions corresponding to fGAMP-H (m/z 673.09) and fGAMP-2H+Na^+ (m/z 695.07) (Figure 6d). In tandem mass spectrometry spectrum (Figure 6e) derived from a parent fGAMP ion (m/z 673.09) ion products corresponding to depu-

rination of the dinucleotide were observed. Depurination of guanine resulted in product with m/z 522.19 and depurination of 2-AP - in product with m/z 538.15. The ratio of these peaks (more depurinated guanines than 2-AP) is characteristic for the linkage between 2'-OH of GTP and 5'-phosphate of 2-APTP within the fGAMP molecule corresponding to 2'3'fGAMP as cGAS product (Ablasser, Goldeck et al. "cGAS produces a 2[prime]-5[prime]-linked cyclic dinucleotide second messenger that activates STING". Nature 498, 380-384, 2013).

Example 9: Fluorescence-based cGAS activity assay

[0219] To precisely measure initial rates of cGAS activity in high-throughput, the inventors developed a fluorescence-based cGAS activity assay based on the incorporation of the fluorescent nucleoside analog 2-aminopurine nucleoside triphosphate (2-APTP) along with GTP into cyclic dinucleotide denoted 2'3'fGAMP (Figure 7a).In this assay 2-APTP and 6.5 ng μl^{-1} DNA of different lengths (20 -100 bp in 5 bp intervals, and pET28M-SUMO1-GFP vector (EMBL) (6.2 kbp, plasmid)) that corresponds to roughly 0.5 μM 20 bp (approximate length of cGAS binding site) were premixed with 0.5 μM cGAS in 40 mM Tris pH 7.5 and 100 mM NaCl. Alternatively, 2.6 ng μl^{-1} DNA ($\approx$0.2 μM binding sites) and 1μM cGAS were used. The reaction was started by adding 5 mM $MgCl_2$ together with 500 μM GTP and 50 μM 2-APTP and performed at 32°C. Fluorescence decrease was measured in 96-well black non-binding PS plates (Greiner Bio-One) on Tecan infinite M1000 (λ_{ex}=305 nm, λ_{em}=363 nm, gain 100, settle time 10ms, kinetic interval 2 min).

[0220] Additionally, 2-APTP fluorescence alone was studied in different conditions, relevant for the assay, as follows. 2.5, 10, 50 or 100 μM 2-APTP was mixed in ultrapure water, 100mM NaCl + 5mM $MgCl_2$, 40mM Tris pH 7.5, 500 μM GTP, 40mM Tris pH 7.5 + 100mM NaCl + 5mM $MgCl_2$ or 40mM Tris pH 7.5 + 100mM NaCl + 5mM $MgCl_2$ and fluorescence was measured as described above. The fluorescence intensity of 2-APTP was not influenced by the used components in the aqueous solution and did not differ from that of fluorescence intensity in ultrapure water (Figure 7b).

[0221] The data clearly indicate that the fluorescence signal of free 2-APTP is quenched upon the incorporation of 2-APTP into 2'3'fGAMP during the reaction which allows the measurement of the reaction rate by measuring the change in fluorescence intensity (Figure 7a,c). Initial fluorescence read-out curves (Figure 7c) represent the changes in fluorescence during the reaction. For each curve the background fluorescence values were subtracted and inverted to give the positive value (ΔF). The resulting curves were scaled, such that ΔF at time point 0 min equals 0 ($\Delta F - \Delta F_{t=0}$) (Figure 7d). The initial cGAS reaction rates were calculated as a slope of the linear intervals (dashed lines) on the resulting curves and are defined as $\Delta F/\Delta t$ [RFU min^{-1}] (Figure 7e).

[0222] The data indicate that the new assay allows measurement of cGAS activity in a simple-to-implement high-throughput assay. Furthermore, the data suggest that the assay can be used to identify substances having an ability to modulate cGAS activity as described above.

Example 10: cGAS incorporates 2-APTP into 2'3'fGAMP in a species-independent manner

[0223] In order to investigate whether 2-APTP is incorporated into 2'3'fGAMP by cGAS in a species-dependent manner, three truncated constructs comprising Mab21 from human (hcGAS^{cd}), mouse (mcGAS^{cd}) and *Sus scrofa* (scGAS^{cd}) and human full length cGAS (hcGAS) were generated as described above.

[0224] These constructs were then analyzed using anion exchange chromatography of cGAS products on MonoQ 5/50 GL column (GE Healthcare), as described in Example 7. The data show that the incorporation of 2-APTP into 2'3'fGAMP does not depend on the origin of cGAS as indicated by the calculated cGAS activity rates for the different species (see Figure 8a).

[0225] The constructs were further tested in fluorescence-based cGAS activity assays. 0.5 μM mcGAS^{cd} (Figure 8b), hcGAS^{cd} (Figure 8c) or hcGAS (Figure 8d) were premixed with 6.5 ng μl^{-1} DNA of different lengths (20 -100 bp), and pET28M-SUMO1-GFP vector (EMBL) (6.2 kbp, plasmid)) that corresponds to roughly 0.5 μM 20 bp in 40 mM Tris pH 7.5 and 100 mM NaCl. The reaction was started by adding 5 mM $MgCl_2$ together with 500 μM GTP and 50 μM 2-APTP and performed at 32°C. Fluorescence decrease (quenching) was measured in 96-well black non-binding PS plates (Greiner Bio-One) on Tecan infinite M1000 (λ_{ex}=305 nm, λ_{em}=363 nm, gain 100, settle time 10ms, kinetic interval 2 min). cGAS activity was calculated as described in Example 9 above. All tested constructs showed the same increase of cGAS activity with DNA length.

[0226] The results further indicate that the described assay is suitable for evaluating the activity of both truncated and full length cGAS constructs, as well as cGAS constructs originating from different species.

Example 11: DNA-bending proteins enhance cGAS activity

[0227] In order to investigate the effect of DNA-bending proteins on cGAS activity, a fluorescence-based cGAS activity assay as described above was performed in the presence several DNA stress associated factors. The experiment was performed as follows:

13 ng μl^{-1} DNA was premixed with DNA-bending proteins mTFAM, mHMGB1 or 1HU in concentrations 0-5 μM, after which 100 nM cGAS in 40 mM Tris pH 7.5 and 100 mM NaCl was added. The reaction was started by adding 5 mM $MgCl_2$ together with 500 μM GTP and 50 μM 2-APTP and carried out at 32°C for $mcGAS^{cd}$ or 37°C for hcGAS. Fluorescence was measured in 96-well black non-binding PS plates (Greiner Bio-One) on Tecan infinite M1000 (λ_{ex}=305 nm, λ_{em}=363 nm, gain 100, settle time 10ms, kinetic interval 2 min). Based on the measured fluorescence cGAS activity was calculated as described in Example 9 above.

[0228] The data indicate that with low concentrations of cGAS only background activity is observed in the absence of DNA-bending proteins, whereas adding increasing amounts of mTFAM, mHMGB1 and 1Hu robustly activated truncated mouse cGAS ($mcGAS^{cd}$) in vitro up to ~25 fold (Figure 9a-c) A similar activation was also seen for hTFAM and full-length human cGAS (hcGAS) (Figure 9d). Each DNA-bending protein tested had an optimal concentration at which maximal stimulation of cGAS activity was observed. Above such concentration cGAS activity was sharply abolished.

[0229] The data indicate that the above described assay is suitable for identifying a substance, illustratively represented by one or more DNA bending proteins, which has an ability to modulate, e.g. enhance, cGAS activity.

Example 12: STING binding assays

[0230] In order to investigate whether 2'3'fCDN is able to bind to STING, luciferase activity assays were performed in which HEK293T cells stably expressing STING (HEK293T STING KI) (Ablasser, Schmid-Burgk et al. "Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP". Nature 503, 530-534, 2013) were induced with 2'3'fGAMP.

[0231] Briefly, $1x10^5$ HEK293T STING KI cells were seeded on 96-well plates and transfected with 100 ng p-125Luc, 10 ng pGL4.74 (Promega) plasmids and fGAMP diluted in DPBS (Gibco, Thermo Scientific) per well using Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific) as transfection reagent according to the vendor's protocol. After 14h cells were lysed in 50 μL passive lysis buffer (Promega, Madison, WI, USA). Immunoactivity experiments using the Dual-Glo luciferase assay system (Promega, Madison, WI, USA) were performed according to manufacturer's instructions. The luciferase activity was determined in a 96-well plate reader.

[0232] Interferon-β response was measured as a proportion of firefly (FF) luciferase activity to Renilla (Ren) lucifearse activity. All ratios were normalized to 2'3'fGAMP buffer control. The data clearly show that IFN-β response increases with 2'3'fGAMP concentration (see Figure 10a) indicating that is capable to bind and activate STING in a cellular system.

Example 13: STING-dependent 2'3'fCDN quenching assay

[0233] To investigate the effect of STING binding on the fluorescene signal of 2'3'fCDN, STING-dependent 2'3'fGAMP assays were performed as follows:

Lyophilized was dissolved in water, diluted for the experiment to the final absorption 2.34 at 250 nm (A_{250}=2.34) and mixed with 0-250 μM human truncated STING ($hSTING^{cd}$). The samples were incubated for 15 min at room temperature and measured in 96-well black non-binding PS plates (Greiner Bio-One) on Tecan infinite M1000 (λ_{ex}=305 nm, λ_{em}=363 nm, gain 140, settle time 10ms) at 25°C.

[0234] The data indicate that fluorescence considerably decreases with increasing concentrations of STING, suggesting that STING-binding induces fluorescence quenching in 2'3'fCDN (see Figure 10b). Thus, while incorporation of 2-AP into 2'3'fGAMP quenches the fluorescence signal of 2-APTP as observed in free solution (see e.g. Example 9), the binding of 2'3'fGAMP to STING quenches the fluorescence signal of 2-AP even further. The results shown in Figure 10b therefore indicate that 2'3'fCDN can be used in a method to study 2'3'fCDN binding to STING either in the presence of a substance or in the absence of a substance.

[0235] The quenching effect of STING on 2'3'fGAMP can be further used to search and compare STING-binding compounds. The competition assays with 2'3'-fGAMP-bound STING were performed as follows:

Lyophilized was dissolved in water, diluted for the experiment to the final absorption 2.34 at 250 nm (A_{250}=2.34) and mixed with 200 μM human truncated STING ($hSTING^{cd}$). The samples were incubated for 15 min at room temperature after which 0-150 μM cGAMP 2'3' or c-di-GMP were added. After second round of incubation for 15 min the samples were measured in 96-well black non-binding PS plates (Greiner Bio-One) on Tecan infinite M1000 (λ_{ex}=305 nm, λ_{em}=363 nm, gain 140, settle time 10ms) at 25°C.

[0236] The data show that adding a STING-binding molecule such as 2'3'cGAMP leads to an increase of observed fluorescent signal (Figure 10c). Such molecules compete for STING binding, leading to the release of 2'3'fGAMP from

the STING-2'3'fGAMP complex. So the quenched fluorescence of 2'3'fGAMP can be regained. Comparison of two STING-binding molecules with different affinities - 2'3'cGAMP and c-di-GMP - shows different efficiency of STING binding. In particular, if comparing the increase in fluorescence by increasing concentrations of STING-binding molecules, those with higher affinity towards STING (e.g. 2'3'cGAMP) will show a more rapid fluorescence increase than those with lower affinity (e.g. c-di-GMP) (compare s1 and s2 from Figure 10d).

[0237] Thus, the STING-dependent 2'3'fCDN quenching assay provides a robust tool for the identification of substances having an ability to bind to STING as discussed above.

References

[0238]

Ablasser, A., M. Goldeck, et al. (2013). "cGAS produces a 2[prime]-5[prime]-linked cyclic dinucleotide second messenger that activates STING." Nature 498(7454): 380-384.

Ablasser, A., J. L. Schmid-Burgk, et al. (2013). "Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP." Nature 503(7477): 530-534.

Ahn, J., P. Ruiz, et al. (2014). "Intrinsic Self-DNA Triggers Inflammatory Disease Dependent on STING." The Journal of Immunology.

Cavlar, T., T. Deimling, et al. (2013). "Species-specific detection of the antiviral small□molecule compound CMA by STING." The EMBO Journal 32(10): 1440-1450.

Civril, F., T. Deimling, et al. (2013). "Structural mechanism of cytosolic DNA sensing by cGAS." Nature 498(7454): 332-337.

Crow, Y. J., D. S. Chase, et al. (2015). "Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1." American journal of medical genetics. Part A 0(2): 296-312.

Crow, Y. J., B. E. Hayward, et al. (2006). "Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus." Nat Genet 38(8): 917-920.

Dey, R. J., B. Dey, et al. (2017). "Inhibition of innate immune cytosolic surveillance by an M. tuberculosis phosphodiesterase." Nat Chem Biol 13(2): 210-217.

Gaidt, M. M., T. S. Ebert, et al. (2016). "Human Monocytes Engage an Alternative Inflammasome Pathway." Immunity 44(4): 833-846.

Gall, A., P. Treuting, et al. (2012). "Autoimmunity Initiates in Nonhematopoietic Cells and Progresses via Lymphocytes in an Interferon-Dependent Autoimmune Disease." Immunity 36(1): 120-131.

Gao, D., T. Li, et al. (2015). "Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases." Proceedings of the National Academy of Sciences 112(42): E5699-E5705.

Gao, P., M. Ascano, et al. (2013). "Structure-Function Analysis of STING Activation by c[G(2',5')pA(3',5')p] and Targeting by Antiviral DMXAA." Cell 154(4): 748-762.

Gao, P., T. Zillinger, et al. (2014). "Binding-Pocket and Lid-Region Substitutions Render Human STING Sensitive to the Species-Specific Drug DMXAA." Cell Reports 8(6): 1668-1676.

Hughes, R. A., A. E. Miklos, et al. (2011). Chapter twelve - Gene Synthesis: Methods and Applications. Methods in Enzymology. V. Christopher, Academic Press. Volume 498: 277-309.

Lakowicz, J. R. (2006). "Principles of Fluorescence Spectroscopy." Springer 3th edition.

Lee-Kirsch, M. A., M. Gong, et al. (2007). "Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 are associated with systemic lupus erythematosus." Nat Genet 39(9): 1065-1067.

Li, L., Q. Yin, et al. (2014). "Hydrolysis of 2'3'-cGAMP by ENPP1 and design of nonhydrolyzable analogs." Nat Chem Biol 10(12): 1043-1048.

Li, X., C. Shu, et al. "Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization." Immunity 39(6): 1019-1031.

Mackenzie, N. C. W., C. Huesa, et al. (2012). "New insights into NPP1 function: Lessons from clinical and animal studies." Bone 51(5): 961-968.

Nowarski, R., N. Gagliani, et al. (2013). "Innate Immune Cells in Inflammation and Cancer." Cancer Immunology Research 1(2): 77-84.

Rapino, F., Eloy F. Robles, et al. (2013). "C/EBPα Induces Highly Efficient Macrophage Transdifferentiation of B Lymphoma and Leukemia Cell Lines and Impairs Their Tumorigenicity." Cell Reports 3(4): 1153-1163.

Roembke, B. T., J. Zhou, et al. (2014). "A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3',3'-cGAMP." Molecular BioSystems 10(6): 1568-1575.

Russell, J. F. S. a. D. W. (2012). "Molecular Cloning, a laboratory manual." Cold Spring Harbor Laboratory Press 1(4th edition).

Sortica, D. A., M. P. Buffon, et al. (2015). "Association between the ENPP1 K121Q Polymorphism and Risk of Diabetic Kidney Disease: A Systematic Review and Meta-Analysis." PLOS ONE 10(3): e0118416.
Tang, E. D. and C.-Y. Wang (2015). "Single Amino Acid Change in STING Leads to Constitutive Active Signaling." PLOS ONE 10(3): e0120090.
Throop, A. L. and J. LaBaer (2015). "Recombinational Cloning Using Gateway and In-Fusion Cloning Schemes." Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et a1.] 110: 3.20.21-23.20.23.
Woo, S.-R., Mercedes B. Fuertes, et al. (2014). "STING-Dependent Cytosolic DNA Sensing Mediates Innate Immune Recognition of Immunogenic Tumors." Immunity 41(5): 830-842.

SEQUENCE LISTING

<110> Ludwig-Maximilians-Universität München

<120> A fluorescent cyclic dinucleotide and its use in methods of identifying substances having an ability to modulate the cGAS/STING pathway

<130> EP111308SZpau

<140> not yet assigned
<141> herewith

<160> 12

<170> PatentIn version 3.5

<210> 1
<211> 1140
<212> DNA
<213> Homo sapiens

<400> 1
atgccccact ccagcctgca tccatccatc ccgtgtccca ggggtcacgg ggcccagaag 60

gcagccttgg ttctgctgag tgcctgcctg gtgacccttt gggggctagg agagccacca 120

gagcacactc tccggtacct ggtgctccac ctagcctccc tgcagctggg actgctgtta 180

aacggggtct gcagcctggc tgaggagctg cgccacatcc actccaggta ccggggcagc 240

tactggagga ctgtgcgggc ctgcctgggc tgccccctcc gccgtggggc cctgttgctg 300

ctgtccatct atttctacta ctccctccca aatgcggtcg gcccgccctt cacttggatg 360

cttgccctcc tgggcctctc gcaggcactg aacatcctcc tgggcctcaa gggcctggcc 420

ccagctgaga tctctgcagt gtgtgaaaaa gggaatttca acgtggccca tgggctggca 480

tggtcatatt acatcggata tctgcggctg atcctgccag agctccaggc ccggattcga 540

acttacaatc agcattacaa caacctgcta cggggtgcag tgagccagcg gctgtatatt 600

ctcctcccat tggactgtgg ggtgcctgat aacctgagta tggctgaccc caacattcgc 660

ttcctggata aactgcccca gcagaccggt gaccatgctg gcatcaagga tcgggtttac 720

agcaacagca tctatgagct tctggagaac gggcagcggg cgggcacctg tgtcctggag 780

tacgccaccc ccttgcagac tttgtttgcc atgtcacaat acagtcaagc tggctttagc 840

cgggaggata ggcttgagca ggccaaactc ttctgccgga cacttgagga catcctggca 900

gatgcccctg agtctcagaa caactgccgc ctcattgcct accaggaacc tgcagatgac 960

agcagcttct cgctgtccca ggaggttctc cggcacctgc ggcaggagga aaaggaagag 1020

gttactgtgg gcagcttgaa gacctcagcg gtgcccagta cctccacgat gtcccaagag 1080

cctgagctcc tcatcagtgg aatggaaaag cccctccctc tccgcacgga tttctcttga 1140

<210> 2

<211> 379
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Pro His Ser Ser Leu His Pro Ser Ile Pro Cys Pro Arg Gly His
1               5                  10                  15

Gly Ala Gln Lys Ala Ala Leu Val Leu Leu Ser Ala Cys Leu Val Thr
            20                  25                  30

Leu Trp Gly Leu Gly Glu Pro Pro Glu His Thr Leu Arg Tyr Leu Val
        35                  40                  45

Leu His Leu Ala Ser Leu Gln Leu Gly Leu Leu Leu Asn Gly Val Cys
    50                  55                  60

Ser Leu Ala Glu Glu Leu Arg His Ile His Ser Arg Tyr Arg Gly Ser
65                  70                  75                  80

Tyr Trp Arg Thr Val Arg Ala Cys Leu Gly Cys Pro Leu Arg Arg Gly
                85                  90                  95

Ala Leu Leu Leu Leu Ser Ile Tyr Phe Tyr Tyr Ser Leu Pro Asn Ala
            100                 105                 110

Val Gly Pro Pro Phe Thr Trp Met Leu Ala Leu Leu Gly Leu Ser Gln
        115                 120                 125

Ala Leu Asn Ile Leu Leu Gly Leu Lys Gly Leu Ala Pro Ala Glu Ile
    130                 135                 140

Ser Ala Val Cys Glu Lys Gly Asn Phe Asn Val Ala His Gly Leu Ala
145                 150                 155                 160

Trp Ser Tyr Tyr Ile Gly Tyr Leu Arg Leu Ile Leu Pro Glu Leu Gln
                165                 170                 175

Ala Arg Ile Arg Thr Tyr Asn Gln His Tyr Asn Asn Leu Leu Arg Gly
            180                 185                 190

Ala Val Ser Gln Arg Leu Tyr Ile Leu Leu Pro Leu Asp Cys Gly Val
        195                 200                 205

Pro Asp Asn Leu Ser Met Ala Asp Pro Asn Ile Arg Phe Leu Asp Lys
    210                 215                 220

Leu Pro Gln Gln Thr Gly Asp His Ala Gly Ile Lys Asp Arg Val Tyr
```

225 230 235 240

Ser Asn Ser Ile Tyr Glu Leu Leu Glu Asn Gly Gln Arg Ala Gly Thr
245 250 255

Cys Val Leu Glu Tyr Ala Thr Pro Leu Gln Thr Leu Phe Ala Met Ser
260 265 270

Gln Tyr Ser Gln Ala Gly Phe Ser Arg Glu Asp Arg Leu Glu Gln Ala
275 280 285

Lys Leu Phe Cys Arg Thr Leu Glu Asp Ile Leu Ala Asp Ala Pro Glu
290 295 300

Ser Gln Asn Asn Cys Arg Leu Ile Ala Tyr Gln Glu Pro Ala Asp Asp
305 310 315 320

Ser Ser Phe Ser Leu Ser Gln Glu Val Leu Arg His Leu Arg Gln Glu
325 330 335

Glu Lys Glu Glu Val Thr Val Gly Ser Leu Lys Thr Ser Ala Val Pro
340 345 350

Ser Thr Ser Thr Met Ser Gln Glu Pro Glu Leu Leu Ile Ser Gly Met
355 360 365

Glu Lys Pro Leu Pro Leu Arg Thr Asp Phe Ser
370 375

<210> 3
<211> 1140
<212> DNA
<213> Homo sapiens

<400> 3
atgccccact ccagcctgca tccatccatc ccgtgtccca ggggtcacgg ggcccagaag 60

gcagccttgg ttctgctgag tgcctgcctg gtgacccttt gggggctagg agagccacca 120

gagcacactc tccggtacct ggtgctccac ctagcctccc tgcagctggg actgctgtta 180

aacggggtct gcagcctggc tgaggagctg cgccacatcc actccaggta ccggggcagc 240

tactggagga ctgtgcgggc ctgcctgggc tgccccctcc gccgtggggc cctgttgctg 300

ctgtccatct atttctacta ctccctccca aatgcggtcg gcccgccctt cacttggatg 360

cttgccctcc tgggcctctc gcaggcactg aacatcctcc tgggcctcaa gggcctggcc 420

ccagctgaga tctctgcagt gtgtgaaaaa gggaatttca acgtggccca tgggctggca 480

tggtcatatt acatcggata tctgcggctg atcctgccag agctccaggc ccggattcga 540

acttacaatc agcattacaa caacctgcta cggggtgcag tgagccagcg gctgtatatt 600

ctcctcccat tggactgtgg ggtgcctgat aacctgagta tggctgaccc caacattcac 660

ttcctggata aactgcccca gcagaccggt gaccgtgctg gcatcaagga tcgggtttac 720

agcaacagca tctatgagct tctggagaac gggcagcggg cgggcacctg tgtcctggag 780

tacgccaccc ccttgcagac tttgtttgcc atgtcacaat acagtcaagc tggctttagc 840

cgggaggata ggcttgagca ggccaaactc ttctgccgga cacttgagga catcctggca 900

gatgcccctg agtctcagaa caactgccgc ctcattgcct accaggaacc tgcagatgac 960

agcagcttct cgctgtccca ggaggttctc cggcacctgc ggcaggagga aaaggaagag 1020

gttactgtgg gcagcttgaa gacctcagcg gtgcccagta cctccacgat gtcccaagag 1080

cctgagctcc tcatcagtgg aatggaaaag cccctccctc tccgcacgga tttctcttga 1140

<210> 4
<211> 379
<212> PRT
<213> Homo sapiens

<400> 4

Met Pro His Ser Ser Leu His Pro Ser Ile Pro Cys Pro Arg Gly His
1 5 10 15

Gly Ala Gln Lys Ala Ala Leu Val Leu Leu Ser Ala Cys Leu Val Thr
20 25 30

Leu Trp Gly Leu Gly Glu Pro Pro Glu His Thr Leu Arg Tyr Leu Val
35 40 45

Leu His Leu Ala Ser Leu Gln Leu Gly Leu Leu Leu Asn Gly Val Cys
50 55 60

Ser Leu Ala Glu Glu Leu Arg His Ile His Ser Arg Tyr Arg Gly Ser
65 70 75 80

Tyr Trp Arg Thr Val Arg Ala Cys Leu Gly Cys Pro Leu Arg Arg Gly
85 90 95

Ala Leu Leu Leu Leu Ser Ile Tyr Phe Tyr Tyr Ser Leu Pro Asn Ala
100 105 110

Val Gly Pro Pro Phe Thr Trp Met Leu Ala Leu Leu Gly Leu Ser Gln
115 120 125

Ala Leu Asn Ile Leu Leu Gly Leu Lys Gly Leu Ala Pro Ala Glu Ile
130 135 140

```
Ser Ala Val Cys Glu Lys Gly Asn Phe Asn Val Ala His Gly Leu Ala
145                 150                 155                 160

Trp Ser Tyr Tyr Ile Gly Tyr Leu Arg Leu Ile Leu Pro Glu Leu Gln
                165                 170                 175

Ala Arg Ile Arg Thr Tyr Asn Gln His Tyr Asn Asn Leu Leu Arg Gly
            180                 185                 190

Ala Val Ser Gln Arg Leu Tyr Ile Leu Leu Pro Leu Asp Cys Gly Val
        195                 200                 205

Pro Asp Asn Leu Ser Met Ala Asp Pro Asn Ile His Phe Leu Asp Lys
    210                 215                 220

Leu Pro Gln Gln Thr Gly Asp Arg Ala Gly Ile Lys Asp Arg Val Tyr
225                 230                 235                 240

Ser Asn Ser Ile Tyr Glu Leu Leu Glu Asn Gly Gln Arg Ala Gly Thr
                245                 250                 255

Cys Val Leu Glu Tyr Ala Thr Pro Leu Gln Thr Leu Phe Ala Met Ser
            260                 265                 270

Gln Tyr Ser Gln Ala Gly Phe Ser Arg Glu Asp Arg Leu Glu Gln Ala
        275                 280                 285

Lys Leu Phe Cys Arg Thr Leu Glu Asp Ile Leu Ala Asp Ala Pro Glu
    290                 295                 300

Ser Gln Asn Asn Cys Arg Leu Ile Ala Tyr Gln Glu Pro Ala Asp Asp
305                 310                 315                 320

Ser Ser Phe Ser Leu Ser Gln Glu Val Leu Arg His Leu Arg Gln Glu
                325                 330                 335

Glu Lys Glu Glu Val Thr Val Gly Ser Leu Lys Thr Ser Ala Val Pro
            340                 345                 350

Ser Thr Ser Thr Met Ser Gln Glu Pro Glu Leu Leu Ile Ser Gly Met
        355                 360                 365

Glu Lys Pro Leu Pro Leu Arg Thr Asp Phe Ser
    370                 375
```

<210> 5
<211> 618

<212> DNA
<213> Homo sapiens

<400> 5
ctggccccag ctgagatctc tgcagtgtgt gaaaaaggga atttcaacgt ggcccatggg 60

ctggcatggt catattacat cggatatctg cggctgatcc tgccagagct ccaggcccgg 120

attcgaactt acaatcagca ttacaacaac ctgctacggg gtgcagtgag ccagcggctg 180

tatattctcc tcccattgga ctgtggggtg cctgataacc tgagtatggc tgaccccaac 240

attcgcttcc tggataaact gccccagcag accggtgacc atgctggcat caaggatcgg 300

gtttacagca acagcatcta tgagcttctg gagaacgggc agcgggcggg cacctgtgtc 360

ctggagtacg ccacccccct gcagactttg tttgccatgt cacaatacag tcaagctggc 420

tttagccggg aggataggct tgagcaggcc aaactcttct gccggacact tgaggacatc 480

ctggcagatg cccctgagtc tcagaacaac tgccgcctca ttgcctacca ggaacctgca 540

gatgacagca gcttctcgct gtcccaggag gttctccggc acctgcggca ggaggaaaag 600

gaagaggtta ctgtgggc 618

<210> 6
<211> 206
<212> PRT
<213> Homo sapiens

<400> 6

Leu Ala Pro Ala Glu Ile Ser Ala Val Cys Glu Lys Gly Asn Phe Asn
1 5 10 15

Val Ala His Gly Leu Ala Trp Ser Tyr Tyr Ile Gly Tyr Leu Arg Leu
20 25 30

Ile Leu Pro Glu Leu Gln Ala Arg Ile Arg Thr Tyr Asn Gln His Tyr
35 40 45

Asn Asn Leu Leu Arg Gly Ala Val Ser Gln Arg Leu Tyr Ile Leu Leu
50 55 60

Pro Leu Asp Cys Gly Val Pro Asp Asn Leu Ser Met Ala Asp Pro Asn
65 70 75 80

Ile Arg Phe Leu Asp Lys Leu Pro Gln Gln Thr Gly Asp His Ala Gly
85 90 95

Ile Lys Asp Arg Val Tyr Ser Asn Ser Ile Tyr Glu Leu Leu Glu Asn
100 105 110

Gly Gln Arg Ala Gly Thr Cys Val Leu Glu Tyr Ala Thr Pro Leu Gln

115 120 125

Thr Leu Phe Ala Met Ser Gln Tyr Ser Gln Ala Gly Phe Ser Arg Glu
130 135 140

Asp Arg Leu Glu Gln Ala Lys Leu Phe Cys Arg Thr Leu Glu Asp Ile
145 150 155 160

Leu Ala Asp Ala Pro Glu Ser Gln Asn Asn Cys Arg Leu Ile Ala Tyr
165 170 175

Gln Glu Pro Ala Asp Asp Ser Ser Phe Ser Leu Ser Gln Glu Val Leu
180 185 190

Arg His Leu Arg Gln Glu Glu Lys Glu Glu Val Thr Val Gly
195 200 205

<210> 7
<211> 618
<212> DNA
<213> Homo sapiens

<400> 7
ctggccccag ctgagatctc tgcagtgtgt gaaaaaggga atttcaacgt ggcccatggg 60

ctggcatggt catattacat cggatatctg cggctgatcc tgccagagct ccaggcccgg 120

attcgaactt acaatcagca ttacaacaac ctgctacggg gtgcagtgag ccagcggctg 180

tatattctcc tcccattgga ctgtggggtg cctgataacc tgagtatggc tgaccccaac 240

attcacttcc tggataaact gccccagcag accggtgacc gtgctggcat caaggatcgg 300

gtttacagca acagcatcta tgagcttctg gagaacgggc agcgggcggg cacctgtgtc 360

ctggagtacg ccacccccctt gcagactttg tttgccatgt cacaatacag tcaagctggc 420

tttagccggg aggataggct tgagcaggcc aaactcttct gccggacact tgaggacatc 480

ctggcagatg cccctgagtc tcagaacaac tgccgcctca ttgcctacca ggaacctgca 540

gatgacagca gcttctcgct gtcccaggag gttctccggc acctgcggca ggaggaaaag 600

gaagaggtta ctgtgggc 618

<210> 8
<211> 206
<212> PRT
<213> Homo sapiens

<400> 8

Leu Ala Pro Ala Glu Ile Ser Ala Val Cys Glu Lys Gly Asn Phe Asn
1 5 10 15

```
Val Ala His Gly Leu Ala Trp Ser Tyr Tyr Ile Gly Tyr Leu Arg Leu
            20                  25                  30

Ile Leu Pro Glu Leu Gln Ala Arg Ile Arg Thr Tyr Asn Gln His Tyr
        35                  40                  45

Asn Asn Leu Leu Arg Gly Ala Val Ser Gln Arg Leu Tyr Ile Leu Leu
    50                  55                  60

Pro Leu Asp Cys Gly Val Pro Asp Asn Leu Ser Met Ala Asp Pro Asn
65                  70                  75                  80

Ile His Phe Leu Asp Lys Leu Pro Gln Gln Thr Gly Asp Arg Ala Gly
                85                  90                  95

Ile Lys Asp Arg Val Tyr Ser Asn Ser Ile Tyr Glu Leu Leu Glu Asn
            100                 105                 110

Gly Gln Arg Ala Gly Thr Cys Val Leu Glu Tyr Ala Thr Pro Leu Gln
        115                 120                 125

Thr Leu Phe Ala Met Ser Gln Tyr Ser Gln Ala Gly Phe Ser Arg Glu
    130                 135                 140

Asp Arg Leu Glu Gln Ala Lys Leu Phe Cys Arg Thr Leu Glu Asp Ile
145                 150                 155                 160

Leu Ala Asp Ala Pro Glu Ser Gln Asn Asn Cys Arg Leu Ile Ala Tyr
                165                 170                 175

Gln Glu Pro Ala Asp Asp Ser Ser Phe Ser Leu Ser Gln Glu Val Leu
            180                 185                 190

Arg His Leu Arg Gln Glu Glu Lys Glu Glu Val Thr Val Gly
        195                 200                 205
```

<210> 9
<211> 1569
<212> DNA
<213> Homo sapiens

<400> 9
atgcagcctt ggcacggaaa ggccatgcag agagcttccg aggccggagc cactgccccc 60

aaggcttccg cacggaatgc caggggcgcc ccgatggatc ccaccgagtc tccggctgcc 120

cccgaggccg ccctgcctaa ggcgggaaag ttcggccccg ccaggaagtc gggatcccgg 180

cagaaaaaga gcgccccgga cacccaggag aggccgcccg tccgcgcaac tggggcccgc 240

gccaaaaagg cccctcagcg cgcccaggac acgcagccgt ctgacgccac cagcgccccc 300

ggggcagagg ggctggagcc tcctgcggct cgggagccgg ctctttccag ggctggttct 360

tgccgccaga ggggcgcgcg ctgctccacg aagccaagac ctccgcccgg gccctgggac 420

gtgcccagcc ccggcctgcc ggtctcggcc cccattctcg tacggaggga tgcggcgcct 480

ggggcctcga agctccgggc ggttttggag aagttgaagc tcagccgcga tgatatctcc 540

acggcggcgg ggatggtgaa aggggttgtg gaccacctgc tgctcagact gaagtgcgac 600

tccgcgttca gaggcgtcgg gctgctgaac accgggagct actatgagca cgtgaagatt 660

tctgcaccta atgaatttga tgtcatgttt aaactggaag tccccagaat tcaactagaa 720

gaatattcca acactcgtgc atattacttt gtgaaattta aaagaaatcc gaaagaaaat 780

cctctgagtc agtttttaga aggtgaaata ttatcagctt ctaagatgct gtcaaagttt 840

aggaaaatca ttaaggaaga aattaacgac attaaagata cagatgtcat catgaagagg 900

aaaagaggag ggagccctgc tgtaacactt cttattagtg aaaaaatatc tgtggatata 960

accctggctt tggaatcaaa aagtagctgg cctgctagca cccaagaagg cctgcgcatt 1020

caaaactggc tttcagcaaa agttaggaag caactacgac taaagccatt ttaccttgta 1080

cccaagcatg caaaggaagg aaatggtttc caagaagaaa catggcggct atccttctct 1140

cacatcgaaa aggaaatttt gaacaatcat ggaaaatcta aaacgtgctg tgaaaacaaa 1200

gaagagaaat gttgcaggaa agattgttta aaactaatga aataccttt agaacagctg 1260

aaagaaaggt ttaaagacaa aaaacatctg gataaattct cttcttatca tgtgaaaact 1320

gccttctttc acgtatgtac ccagaaccct caagacagtc agtgggaccg caaagacctg 1380

ggcctctgct ttgataactg cgtgacatac tttcttcagt gcctcaggac agaaaaactt 1440

gagaattatt ttattcctga attcaatcta ttctctagca acttaattga caaaagaagt 1500

aaagaatttc tgacaaagca aattgaatat gaaagaaaca atgagtttcc agtttttgat 1560

gaattttga 1569

<210> 10
<211> 522
<212> PRT
<213> Homo sapiens

<400> 10

Met Gln Pro Trp His Gly Lys Ala Met Gln Arg Ala Ser Glu Ala Gly
1 5 10 15

Ala Thr Ala Pro Lys Ala Ser Ala Arg Asn Ala Arg Gly Ala Pro Met
20 25 30

Asp Pro Thr Glu Ser Pro Ala Ala Pro Glu Ala Ala Leu Pro Lys Ala
35 40 45

```
Gly Lys Phe Gly Pro Ala Arg Lys Ser Gly Ser Arg Gln Lys Lys Ser
    50                  55                  60

Ala Pro Asp Thr Gln Glu Arg Pro Pro Val Arg Ala Thr Gly Ala Arg
65                  70                  75                  80

Ala Lys Lys Ala Pro Gln Arg Ala Gln Asp Thr Gln Pro Ser Asp Ala
                85                  90                  95

Thr Ser Ala Pro Gly Ala Glu Gly Leu Glu Pro Pro Ala Ala Arg Glu
            100                 105                 110

Pro Ala Leu Ser Arg Ala Gly Ser Cys Arg Gln Arg Gly Ala Arg Cys
        115                 120                 125

Ser Thr Lys Pro Arg Pro Pro Pro Gly Pro Trp Asp Val Pro Ser Pro
    130                 135                 140

Gly Leu Pro Val Ser Ala Pro Ile Leu Val Arg Arg Asp Ala Ala Pro
145                 150                 155                 160

Gly Ala Ser Lys Leu Arg Ala Val Leu Glu Lys Leu Lys Leu Ser Arg
                165                 170                 175

Asp Asp Ile Ser Thr Ala Ala Gly Met Val Lys Gly Val Val Asp His
            180                 185                 190

Leu Leu Leu Arg Leu Lys Cys Asp Ser Ala Phe Arg Gly Val Gly Leu
        195                 200                 205

Leu Asn Thr Gly Ser Tyr Tyr Glu His Val Lys Ile Ser Ala Pro Asn
    210                 215                 220

Glu Phe Asp Val Met Phe Lys Leu Glu Val Pro Arg Ile Gln Leu Glu
225                 230                 235                 240

Glu Tyr Ser Asn Thr Arg Ala Tyr Tyr Phe Val Lys Phe Lys Arg Asn
                245                 250                 255

Pro Lys Glu Asn Pro Leu Ser Gln Phe Leu Glu Gly Glu Ile Leu Ser
            260                 265                 270

Ala Ser Lys Met Leu Ser Lys Phe Arg Lys Ile Ile Lys Glu Glu Ile
        275                 280                 285

Asn Asp Ile Lys Asp Thr Asp Val Ile Met Lys Arg Lys Arg Gly Gly
```

290 295 300

Ser Pro Ala Val Thr Leu Leu Ile Ser Glu Lys Ile Ser Val Asp Ile
305 310 315 320

Thr Leu Ala Leu Glu Ser Lys Ser Ser Trp Pro Ala Ser Thr Gln Glu
325 330 335

Gly Leu Arg Ile Gln Asn Trp Leu Ser Ala Lys Val Arg Lys Gln Leu
340 345 350

Arg Leu Lys Pro Phe Tyr Leu Val Pro Lys His Ala Lys Glu Gly Asn
355 360 365

Gly Phe Gln Glu Glu Thr Trp Arg Leu Ser Phe Ser His Ile Glu Lys
370 375 380

Glu Ile Leu Asn Asn His Gly Lys Ser Lys Thr Cys Cys Glu Asn Lys
385 390 395 400

Glu Glu Lys Cys Cys Arg Lys Asp Cys Leu Lys Leu Met Lys Tyr Leu
405 410 415

Leu Glu Gln Leu Lys Glu Arg Phe Lys Asp Lys Lys His Leu Asp Lys
420 425 430

Phe Ser Ser Tyr His Val Lys Thr Ala Phe Phe His Val Cys Thr Gln
435 440 445

Asn Pro Gln Asp Ser Gln Trp Asp Arg Lys Asp Leu Gly Leu Cys Phe
450 455 460

Asp Asn Cys Val Thr Tyr Phe Leu Gln Cys Leu Arg Thr Glu Lys Leu
465 470 475 480

Glu Asn Tyr Phe Ile Pro Glu Phe Asn Leu Phe Ser Ser Asn Leu Ile
485 490 495

Asp Lys Arg Ser Lys Glu Phe Leu Thr Lys Gln Ile Glu Tyr Glu Arg
500 505 510

Asn Asn Glu Phe Pro Val Phe Asp Glu Phe
515 520

<210> 11
<211> 1107
<212> DNA
<213> Homo sapiens

<400> 11
cggagggatg cggcgcctgg ggcctcgaag ctccgggcgg ttttggagaa gttgaagctc 60

agccgcgatg atatctccac ggcggcgggg atggtgaaag gggttgtgga ccacctgctg 120

ctcagactga agtgcgactc cgcgttcaga ggcgtcgggc tgctgaacac cgggagctac 180

tatgagcacg tgaagatttc tgcacctaat gaatttgatg tcatgtttaa actggaagtc 240

cccagaattc aactagaaga atattccaac actcgtgcat attactttgt gaaatttaaa 300

agaaatccga aagaaaatcc tctgagtcag tttttagaag gtgaaatatt atcagcttct 360

aagatgctgt caaagtttag gaaaatcatt aaggaagaaa ttaacgacat taaagataca 420

gatgtcatca tgaagaggaa aagaggaggg agccctgctg taacacttct tattagtgaa 480

aaaatatctg tggatataac cctggctttg gaatcaaaaa gtagctggcc tgctagcacc 540

caagaaggcc tgcgcattca aaactggctt tcagcaaaag ttaggaagca actacgacta 600

aagccatttt accttgtacc caagcatgca aaggaaggaa atggtttcca agaagaaaca 660

tggcggctat ccttctctca catcgaaaag gaaattttga acaatcatgg aaaatctaaa 720

acgtgctgtg aaaacaaaga agagaaatgt tgcaggaaag attgtttaaa actaatgaaa 780

taccttttag aacagctgaa agaaaggttt aaagacaaaa aacatctgga taaattctct 840

tcttatcatg tgaaaactgc cttctttcac gtatgtaccc agaaccctca agacagtcag 900

tgggaccgca aagacctggg cctctgcttt gataactgcg tgacatactt tcttcagtgc 960

ctcaggacag aaaaacttga gaattatttt attcctgaat tcaatctatt ctctagcaac 1020

ttaattgaca aaagaagtaa agaatttctg acaaagcaaa ttgaatatga aagaaacaat 1080

gagtttccag tttttgatga attttga 1107

<210> 12
<211> 368
<212> PRT
<213> Homo sapiens

<400> 12
Arg Arg Asp Ala Ala Pro Gly Ala Ser Lys Leu Arg Ala Val Leu Glu
1 5 10 15

Lys Leu Lys Leu Ser Arg Asp Asp Ile Ser Thr Ala Ala Gly Met Val
20 25 30

Lys Gly Val Val Asp His Leu Leu Leu Arg Leu Lys Cys Asp Ser Ala
35 40 45

Phe Arg Gly Val Gly Leu Leu Asn Thr Gly Ser Tyr Tyr Glu His Val
50 55 60

```
Lys Ile Ser Ala Pro Asn Glu Phe Asp Val Met Phe Lys Leu Glu Val
65                  70                  75                  80

Pro Arg Ile Gln Leu Glu Glu Tyr Ser Asn Thr Arg Ala Tyr Tyr Phe
                85                  90                  95

Val Lys Phe Lys Arg Asn Pro Lys Glu Asn Pro Leu Ser Gln Phe Leu
            100                 105                 110

Glu Gly Glu Ile Leu Ser Ala Ser Lys Met Leu Ser Lys Phe Arg Lys
        115                 120                 125

Ile Ile Lys Glu Glu Ile Asn Asp Ile Lys Asp Thr Asp Val Ile Met
    130                 135                 140

Lys Arg Lys Arg Gly Gly Ser Pro Ala Val Thr Leu Leu Ile Ser Glu
145                 150                 155                 160

Lys Ile Ser Val Asp Ile Thr Leu Ala Leu Glu Ser Lys Ser Ser Trp
                165                 170                 175

Pro Ala Ser Thr Gln Glu Gly Leu Arg Ile Gln Asn Trp Leu Ser Ala
            180                 185                 190

Lys Val Arg Lys Gln Leu Arg Leu Lys Pro Phe Tyr Leu Val Pro Lys
        195                 200                 205

His Ala Lys Glu Gly Asn Gly Phe Gln Glu Glu Thr Trp Arg Leu Ser
    210                 215                 220

Phe Ser His Ile Glu Lys Glu Ile Leu Asn Asn His Gly Lys Ser Lys
225                 230                 235                 240

Thr Cys Cys Glu Asn Lys Glu Glu Lys Cys Cys Arg Lys Asp Cys Leu
                245                 250                 255

Lys Leu Met Lys Tyr Leu Leu Glu Gln Leu Lys Glu Arg Phe Lys Asp
            260                 265                 270

Lys Lys His Leu Asp Lys Phe Ser Ser Tyr His Val Lys Thr Ala Phe
        275                 280                 285

Phe His Val Cys Thr Gln Asn Pro Gln Asp Ser Gln Trp Asp Arg Lys
    290                 295                 300

Asp Leu Gly Leu Cys Phe Asp Asn Cys Val Thr Tyr Phe Leu Gln Cys
305                 310                 315                 320
```

```
Leu Arg Thr Glu Lys Leu Glu Asn Tyr Phe Ile Pro Glu Phe Asn Leu
                325                 330                 335

Phe Ser Ser Asn Leu Ile Asp Lys Arg Ser Lys Glu Phe Leu Thr Lys
            340                 345                 350

Gln Ile Glu Tyr Glu Arg Asn Asn Glu Phe Pro Val Phe Asp Glu Phe
        355                 360                 365
```

Claims

1. A cyclic dinucleotide comprising a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate,
wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate.

2. The cyclic dinucleotide according to claim 1, wherein

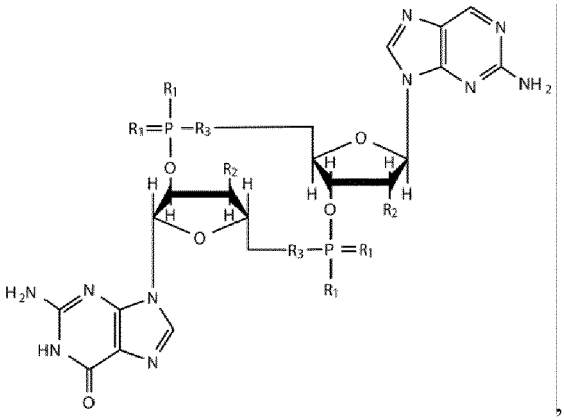
   • the 3'-5' phosphodiester linkage is between the 3'-oxygen of the first nucleoside monophosphate and the 5'-oxygen of the second nucleoside monophosphate, and
   • the 2'-5' phosphodiester linkage is between the 5'-oxygen of the first nucleoside monophosphate and the 2'-oxygen of the second nucleoside monophosphate.

3. The cyclic dinucleotide according to claims 1 or 2, wherein the fluorescent nucleoside monophosphate is a fluorescent purine nucleoside monophosphate.

4. The cyclic dinucleotide according to claims 1 or 2, wherein the fluorescent nucleoside monophosphate is selected from the group consisting of a 2-aminopurine nucleoside monophosphate (2-APMP), a 3-methyl-isoxanthopterin nucleoside monophosphate (3-MIMP), a 6-methyl isoxanthopterin nucleoside monophosphate (6-MIMP), 4-amino-6-methyl-8-(2-deoxy-beta-d-ribofuranosyl)-7(8H)-pteridone nucleoside monophosphate (6-MAPMP), a 4-amino-2,6-dimethyl-8-(2'-deoxy-beta-d-ribofuranosyl)-7(8H)-pteridone nucleoside monophosphate (DMAPMP), a pyrrolocytosine nuceloside monophosphate (pyrrolo-CMP), a 6-phenylpyrrolocytosine nucleoside monophosphate (PhpCMP), a (aminoethoxy)phenylpyrrolocytosine nucleoside monophosphate (moPhpCMP), a [bis-o-(aminoethoxy)phenyl]pyrrolocytosine nucleoside monophosphate (boPhpCMP), a hydropyrimidopyrimidine nucleoside monophosphate ($C^{hpp}MP$), a pyrrolopyrimidopyrimidine nucleoside monophosphate ($C^{ppp}MP$), a pyrimidopyrimidoindole nucleoside monophosphate ($C^{ppi}MP$), a benzopyridopyrimidine nucleoside monophosphate (BPPMP), a naphthopyridopyrimidine nucleoside monophosphate (NPPMP), a methoxybenzodeazaadenine nucleoside monophosphate (^{MD}AMP), a methoxybenzodeazainosine nucleoside monophosphate (^{MD}IMP), a naphthodeazaadenine nucleoside monophosphate (^{ND}AMP), a furan-modified pyrimidine nucleoside mohophosphate, a thieno[3,2]pyrimidine nucleoside monophosphate, a thieno[3,4]pyrimidine nucleoside monophosphate, a 5-methoxy-quinazoline-2,4-(1H,3H) dione nucleoside monophosphate, a 5-methylpyrimidine-2-one nucleoside monophosphate, a 7-deazapurine nucleoside monophosphate, a 5-alkyluridine nucleoside monophosphate, a benzoquinalozines nucleoside monophosphate, a triazoleadenosine nucleoside monophosphate, and a $1,N^6$-ethenoadenosine nucleoside monophosphate.

5. The cyclic dinucleotide according to claim 4, wherein the 2-APMP is the first nucleoside monophosphate.

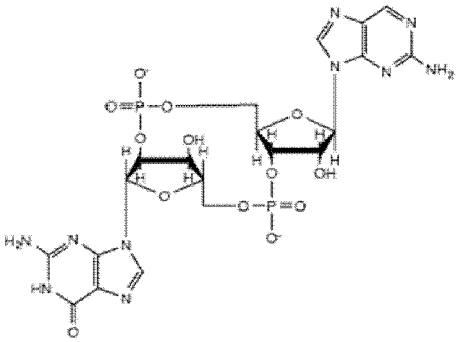
6. The cyclic dinucleotide according to claim any one of claims 1 to 5, having the following formula:



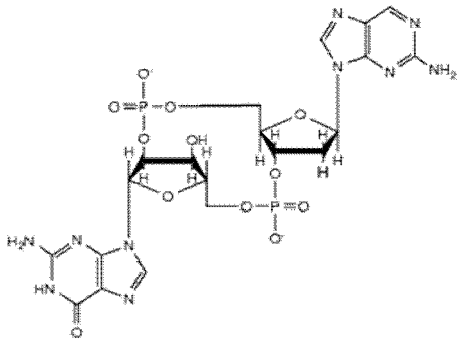
wherein

when R_1 is bound to P through a single bond, R_1 is independently selected from the group consisting of O, S, BH_3 and CH_3;
when R_1 is bound to P through a double bond, R_1 is independently selected from the group consisting of O, S and NH;
wherein at least one R_1 bound to each P is O;
R_2 is independently selected from the group consisting of H, OH, methyl, amino-, methoxy-, fluoro-, methoxyethyl-, -O-propargyl and O-propylamine; and
R_3 is O.

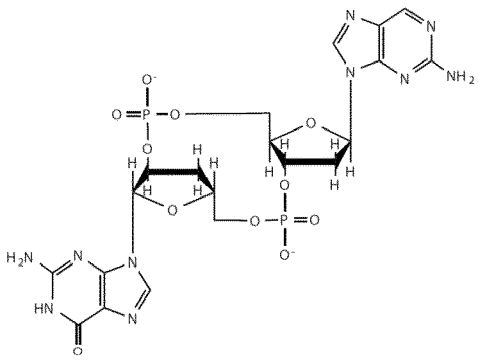
7. The cyclic dinucleotide according to any one of claims 1 to 6 having the following formula:



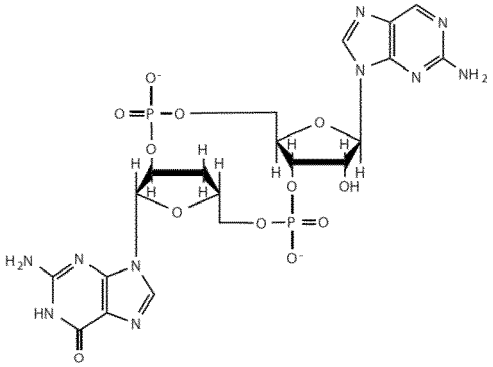
or



or



or



.

8. A method of measuring cGAS activity, the method comprising:

(i) providing an aqueous solution comprising cGAS, cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group; and
(ii) measuring the change of the fluorescence signal of the aqueous solution over time.

9. A method of identifying a substance having an ability to modulate the activity of cyclic GMP-AMP synthase (cGAS), the method comprising:

(i) providing an aqueous solution comprising cGAS, cGAS-activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group;
(ii) measuring the fluorescence signal of the aqueous solution;
(iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence signal is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and
(iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance;

wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance is a cGAS antagonist, while a measured fluorescence signal which is lower in the presence of the substance than in the absence of the substance indicates that the substance is a cGAS agonist.

10. A method of preparing a cyclic dinucleotide according to any one of claims 1 to 7, the method comprising:

(i) providing an aqueous solution comprising cGAS, cGAS activating nucleic acid, a first nucleoside triphosphate, a second nucleoside triphosphate, and one or more divalent cation, wherein at least one of the first and second nucleoside triphosphates is a fluorescent nucleoside triphosphate, and wherein one of the first and second nucleoside triphosphates has a free 2' hydroxyl group and the other one has a free 3' hydroxyl group, thereby preparing the cyclic dinucleotide; and optionally
(ii) purifying the cyclic dinucleotide.

11. The method according to any one of claims 8 to 10, wherein the fluorescent nucleoside triphosphate is 2-aminopurine nucleoside triphosphate (2-APTP).

12. A method of identifying a substance having an ability to bind to stimulator of interferon genes (STING), the method comprising:

(i) providing an aqueous solution comprising a cyclic dinucleotide and STING, wherein the cyclic dinucleotide comprises a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate;
(ii) measuring the fluorescence signal of the aqueous solution;
(iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and
(iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance;

wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance has an ability to bind to STING.

13. A method of identifying a substance having an ability to modulate the activity of a 2'-5' phosphodiesterase, the method comprising

(i) providing an aqueous solution comprising a cyclic dinucleotide and a 2'-5' phosphodiesterase, wherein the

cyclic dinucleotide comprises a 3'-5' phosphodiester linkage and a 2'-5' phosphodiester linkage between a first nucleoside monophosphate and a second nucleoside monophosphate, wherein at least one of the first and second nucleoside monophosphates is a fluorescent nucleoside monophosphate;
(ii) measuring the fluorescence signal of the aqueous solution;
(iii) repeating steps (i) and (ii), wherein upon said repetition the aqueous solution further comprises the substance, and wherein the fluorescence is measured under identical or substantially identical conditions following the provision of the respective aqueous solution; and
(iv) comparing the fluorescence signal measured in the presence of the substance with the fluorescence signal measured in the absence of the substance;

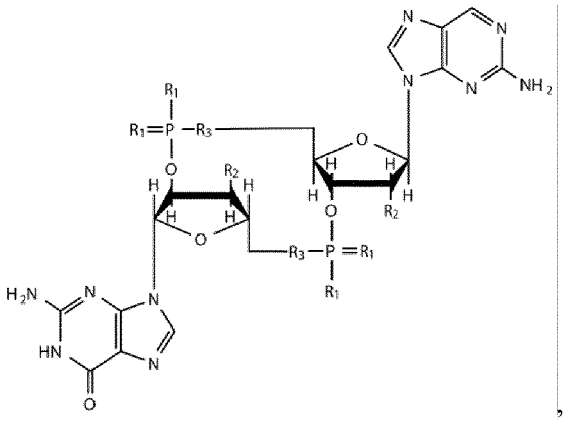
wherein a measured fluorescence signal which is higher in the presence of the substance than in the absence of the substance indicates that the substance is a 2'-5' phosphodiesterase agonist, while a measured fluorescence signal which is lower in the presence of the substance compared to the absence of the substance indicates that the substance is a 2'-5' phosphodiesterase antagonist.

14. The method according to claim 12 or 13, wherein the 3'-5' phosphodiester linkage of the cyclic dinucleotide is between the 3'-oxygen of the first nucleoside monophosphate and the 5'-oxygen of the second nucleoside monophosphate, and
wherein the 2'-5' phosphodiester linkage of the cyclic dinucleotide is between the 5'-oxygen of the first nucleoside monophosphate and the 2'-oxygen of the second nucleoside monophosphate.

15. The method according to any one of claims 12 to 14, wherein the fluorescent nucleoside monophosphate is a fluorescent purine nucleoside monophosphate.

16. The method according to any one of claims 13 to 15, wherein the first nucleoside monophosphate is 2-aminopurine nucleoside monophosphate (2-APMP).

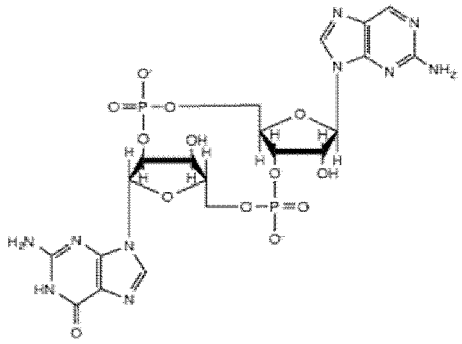
17. The method according to any one of claims 12 to 16, having the following formula:



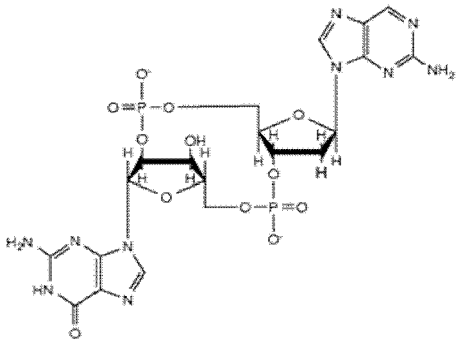
wherein

when R_1 is bound to P through a single bond, R_1 is independently selected from the group consisting of O, S, BH_3 and CH_3;
when R_1 is bound to P through a double bond, R_1 is independently selected from the group consisting of O, S and NH;
wherein at least one R_1 bound to each P is O;
R_2 is independently selected from the group consisting of H, OH, methyl, amino-, methoxy-, fluoro-, methoxyethyl-, -O-propargyl and O-propylamine; and
R_3 is O.

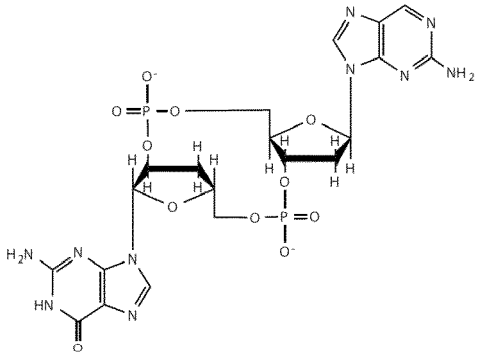
18. The method according to any one of claims 12 to 17, wherein the cyclic dinucleotide has the following formula:



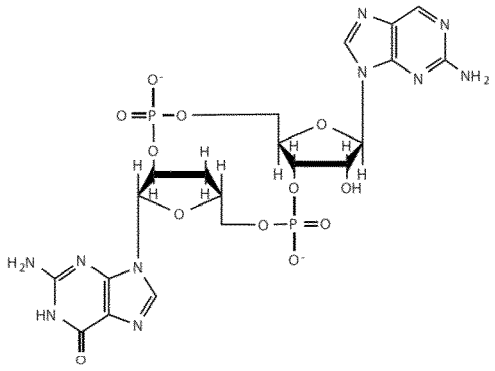
or



or



or



.

19. Use of a fluorescent nucleoside triphosphate for measuring cGAS activity or for identifying a substance having an ability to modulate cGAS activity.

20. Use of a cyclic dinucleotide according to any one of claims 1 to 7 for identifying a substance with the ability to bind to STING.

21. Use of a cyclic dinucleotide according to any one of claims 1 to 7 for labeling cGAS or STING.

22. A complex comprising STING and a cyclic dinucleotide according to any one of claims 1-7.

Figure 1

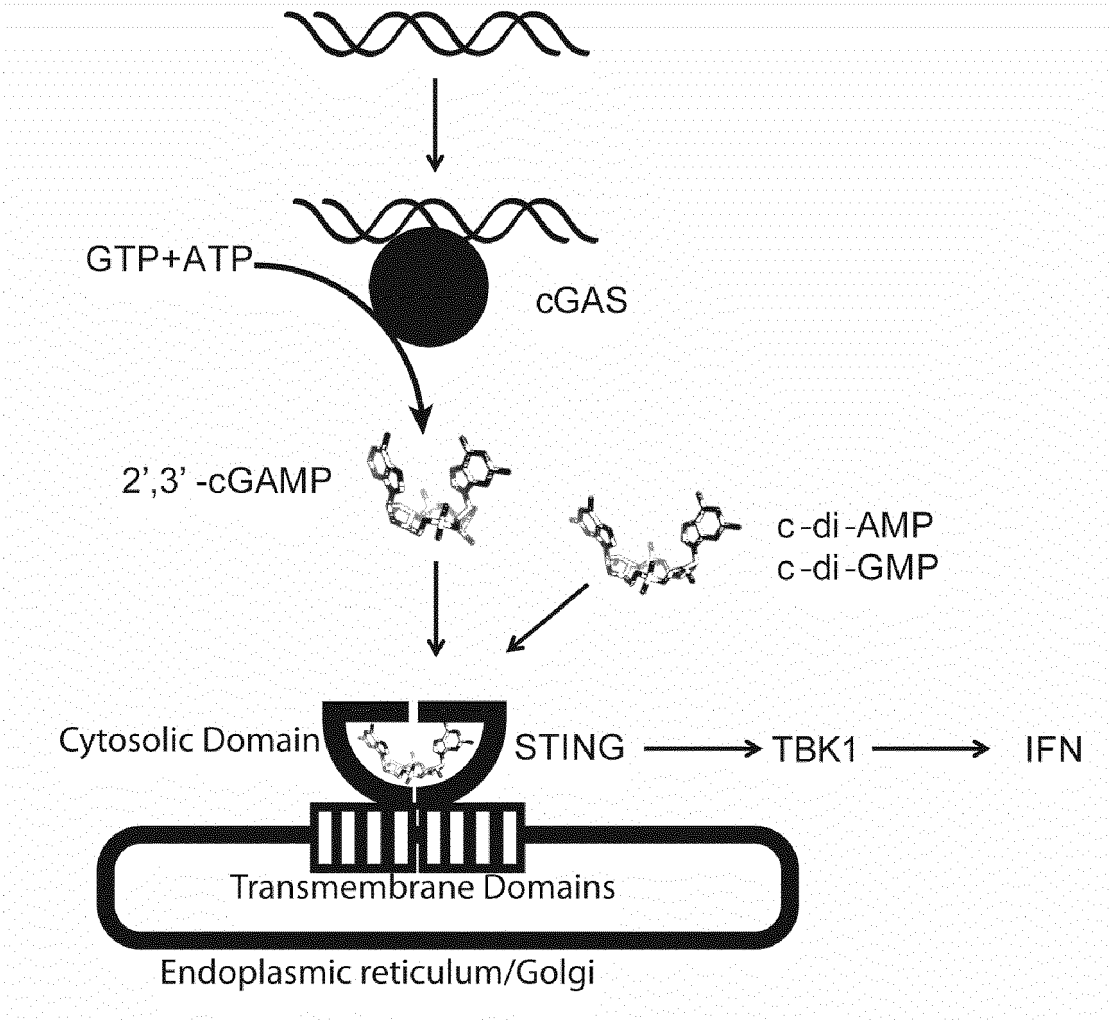

GTP+ATP
cGAS
2',3' -cGAMP
c-di-AMP
c-di-GMP
Cytosolic Domain
STING
TBK1
IFN
Transmembrane Domains
Endoplasmic reticulum/Golgi

Figure 2

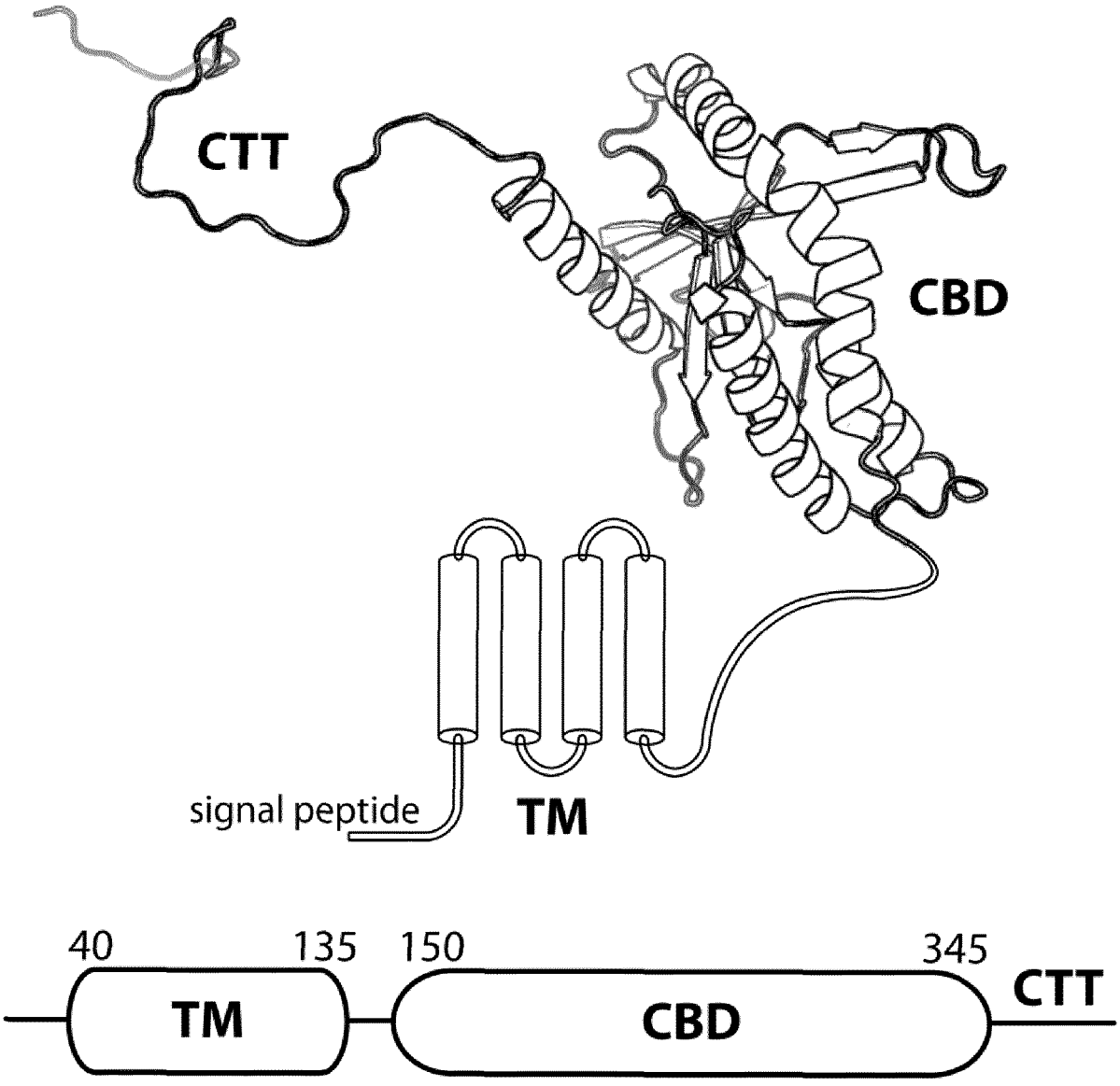


Figure 3

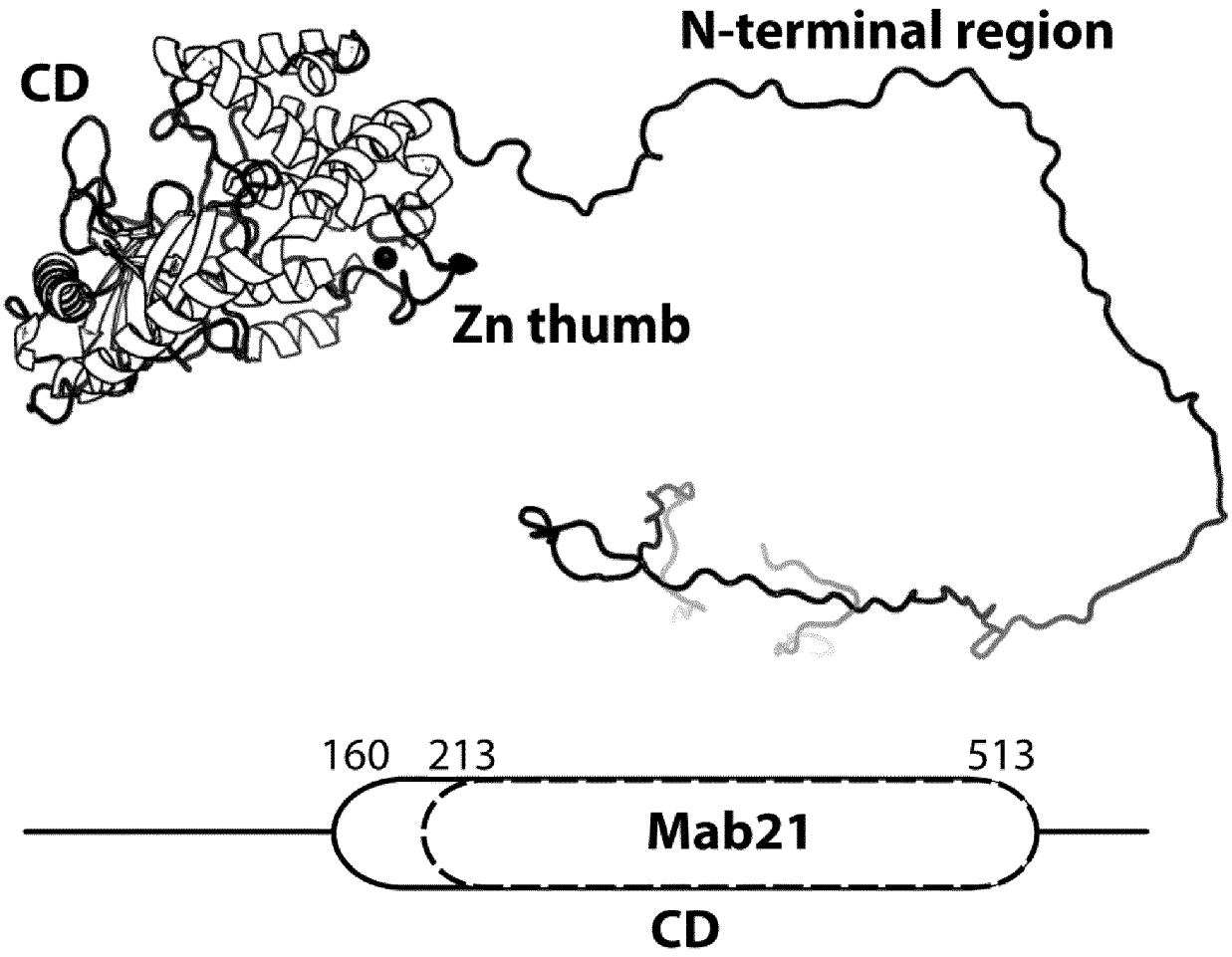
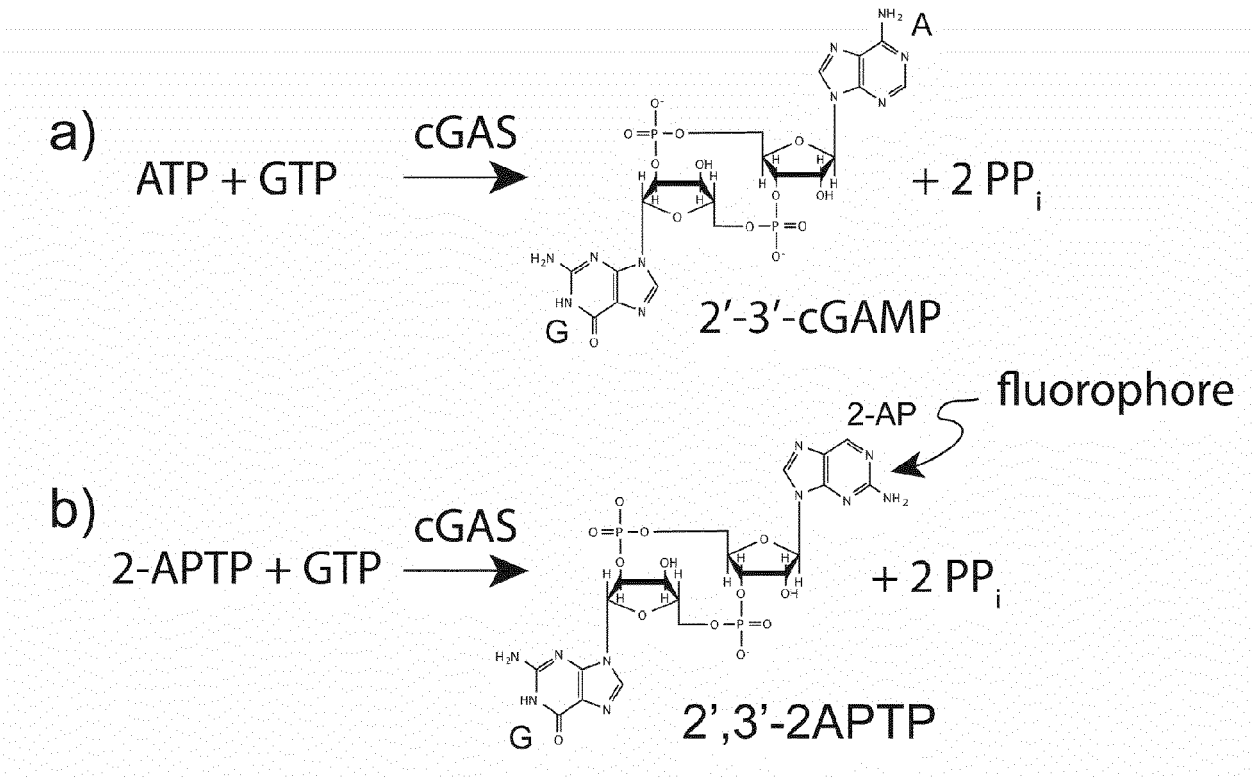


Figure 4


a)
ATP + GTP
cGAS
+ 2 PPi
A
G
2′-3′-cGAMP
fluorophore
2-AP
b)
2-APTP + GTP
cGAS
+ 2 PPi
G
2’,3’-2APTP

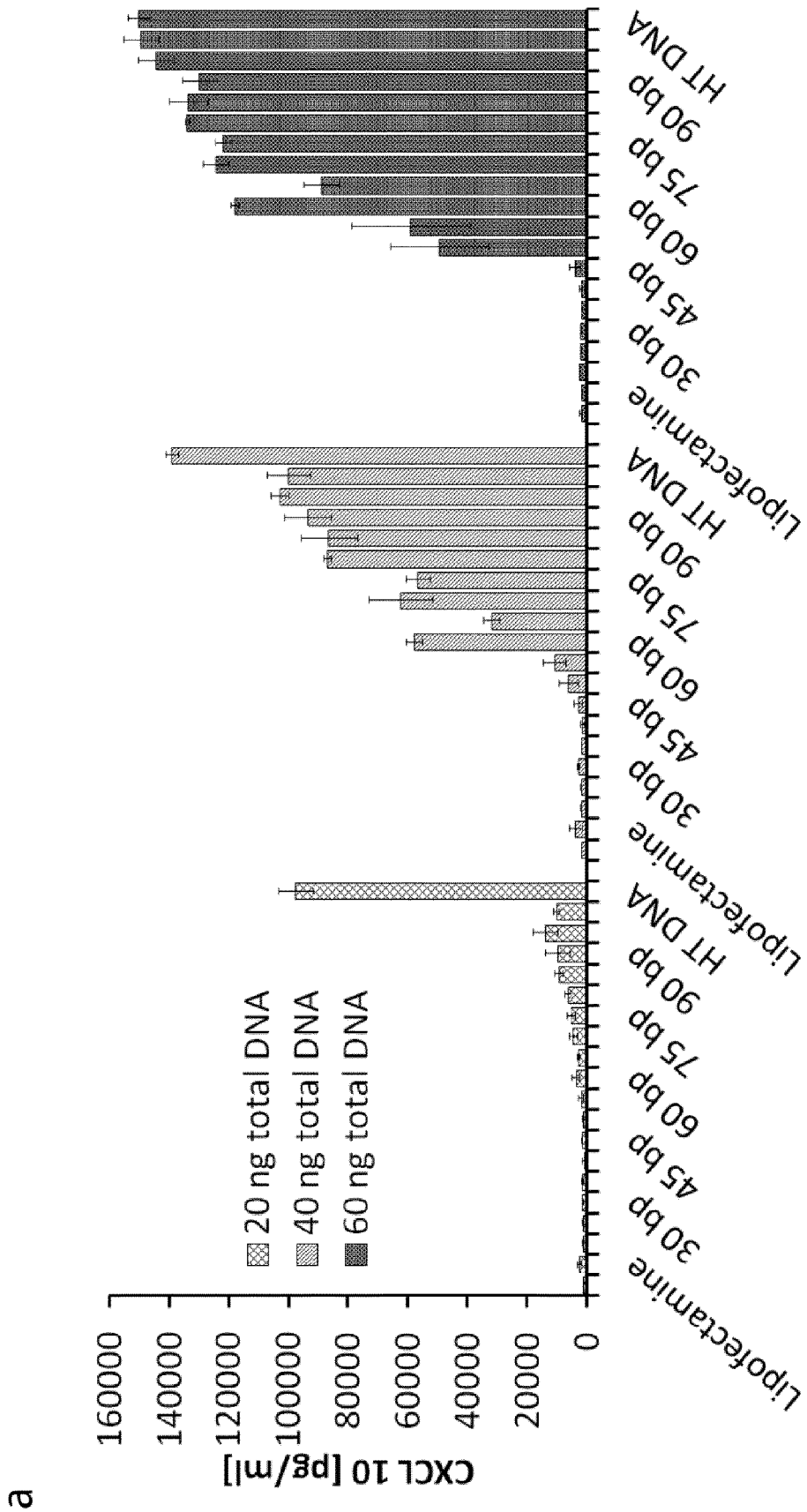

a
CXCL 10 [pg/ml]
160000
140000
120000
100000
80000
60000
40000
20000
0
20 ng total DNA
40 ng total DNA
60 ng total DNA
Lipofectamine
30 bp
45 bp
60 bp
75 bp
90 bp
HT DNA
Lipofectamine
30 bp
45 bp
60 bp
75 bp
90 bp
HT DNA
Lipofectamine
30 bp
45 bp
60 bp
75 bp
90 bp
HT DNA

Figure 5a

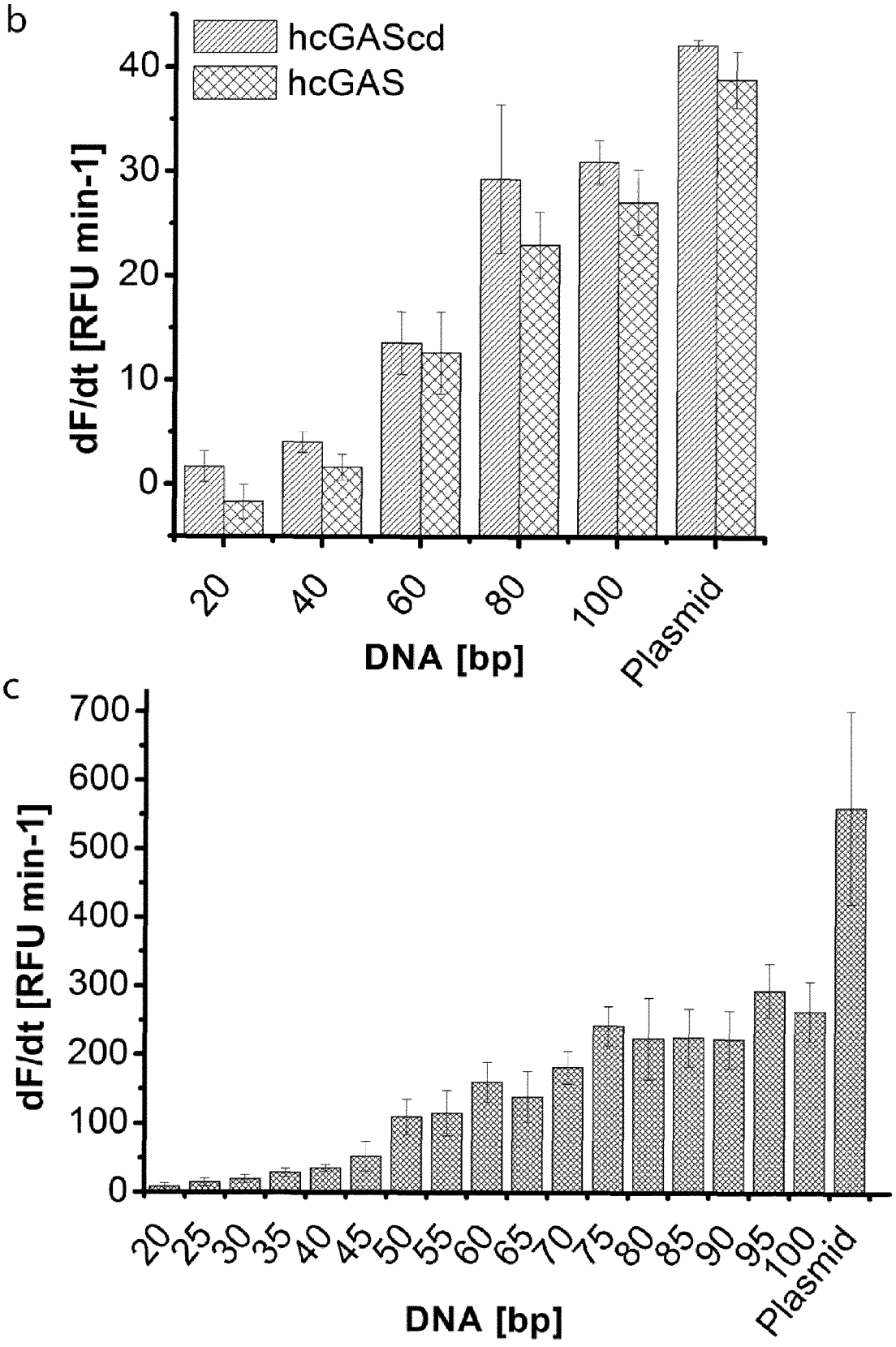

b
hcGAScd
hcGAS
dF/dt [RFU min-1]
0
10
20
30
40
20
40
60
80
100
Plasmid
DNA [bp]
c
dF/dt [RFU min-1]
0
100
200
300
400
500
600
700
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Plasmid
DNA [bp]

Figure 5b_c

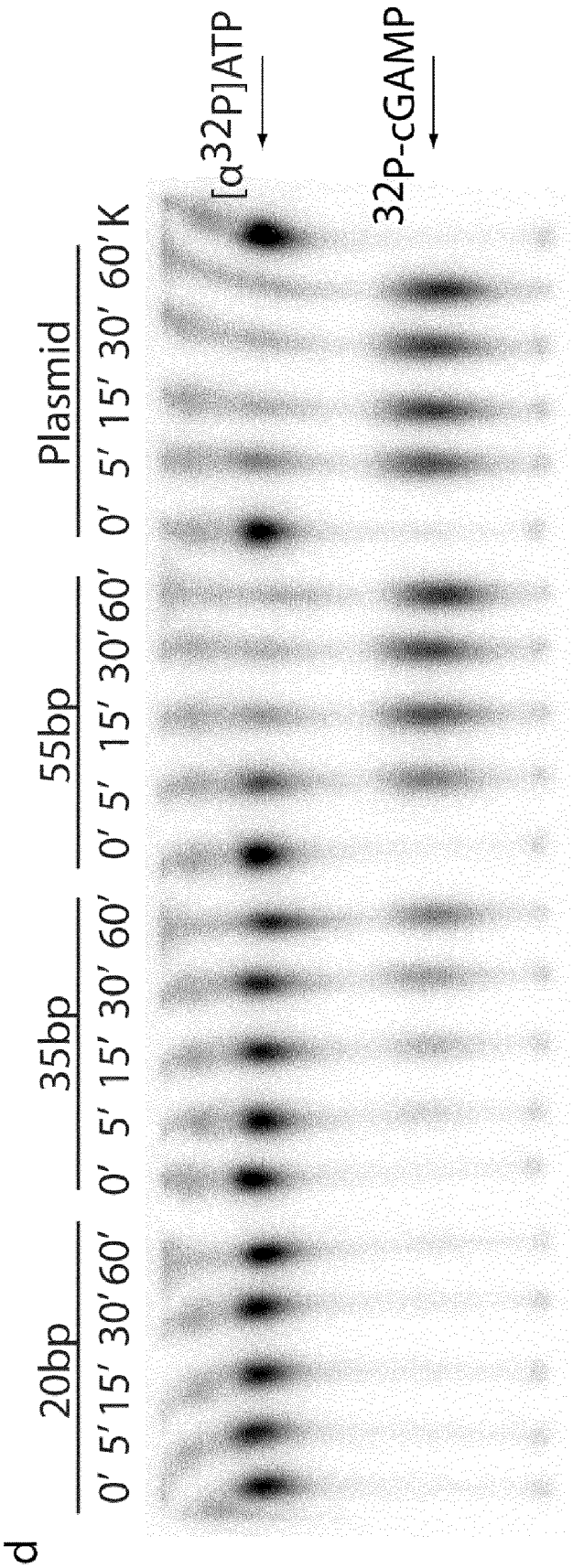

d
20bp
0′ 5′ 15′ 30′ 60′
35bp
0′ 5′ 15′ 30′ 60′
55bp
0′ 5′ 15′ 30′ 60′
Plasmid
0′ 5′ 15′ 30′ 60′ K
[α32P]ATP
32P-cGAMP

Figure 5d

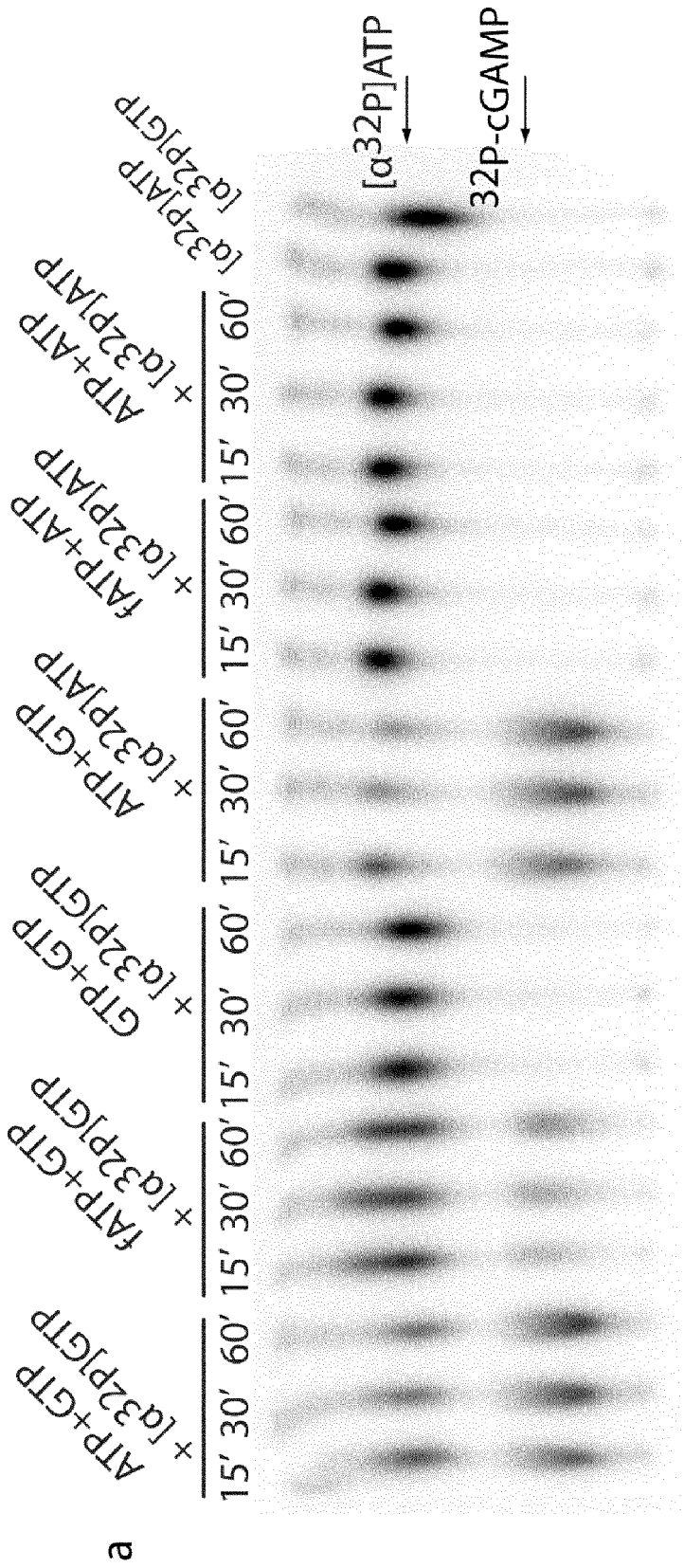
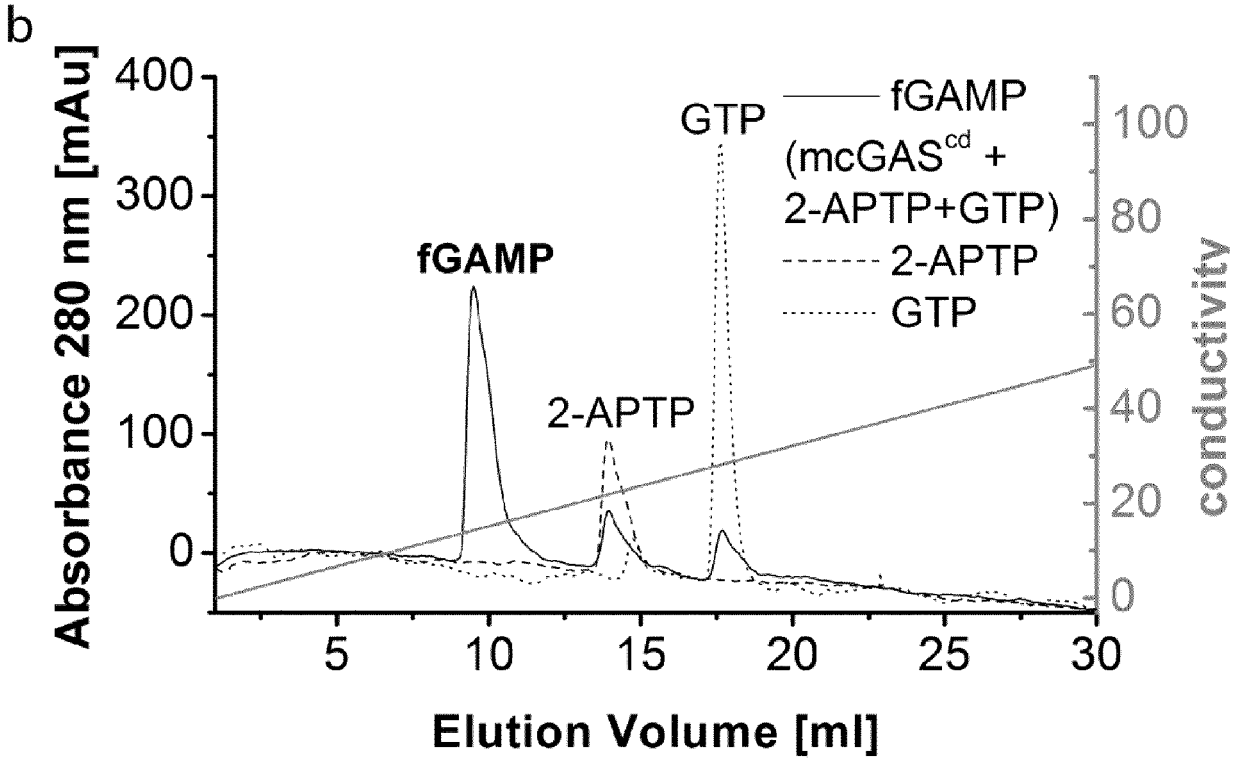

[α32P]ATP
32P-cGAMP
[α32P]ATP
[α32P]GTP
ATP+ATP + [α32P]ATP
fATP+ATP + [α32P]ATP
ATP+GTP + [α32P]ATP
GTP+GTP + [α32P]GTP
fATP+GTP + [α32P]GTP
ATP+GTP + [α32P]GTP
15′ 30′ 60′ 15′ 30′ 60′ 15′ 30′ 60′ 15′ 30′ 60′ 15′ 30′ 60′
a

Figure 6a


b
Absorbance 280 nm [mAu]
400
300
200
100
0
fGAMP
2-APTP
GTP
fGAMP
(mcGAS^cd +
2-APTP+GTP)
2-APTP
GTP
100
80
60
40
20
0
conductivity
5
10
15
20
25
30
Elution Volume [ml]

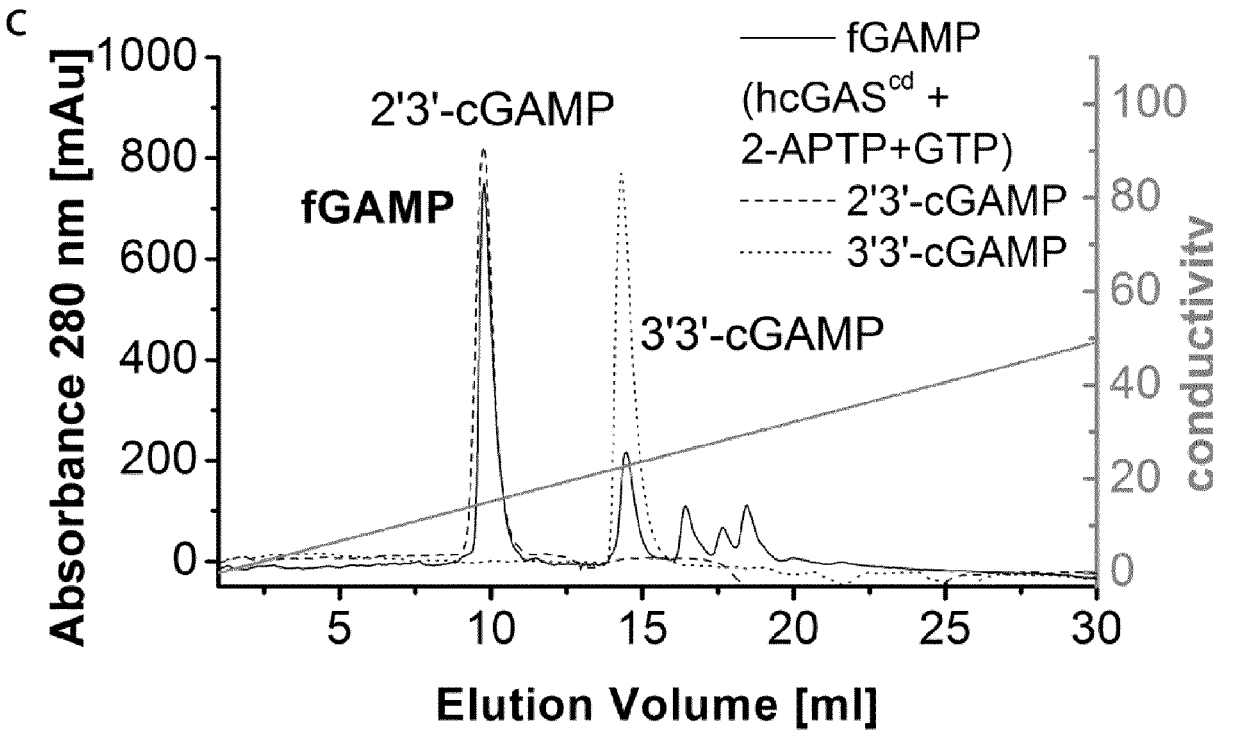

c
Absorbance 280 nm [mAu]
1000
800
600
400
200
0
2'3'-cGAMP
fGAMP
3'3'-cGAMP
fGAMP
(hcGAS^cd +
2-APTP+GTP)
2'3'-cGAMP
3'3'-cGAMP
100
80
60
40
20
0
conductivity
5
10
15
20
25
30
Elution Volume [ml]

Figure 6bc

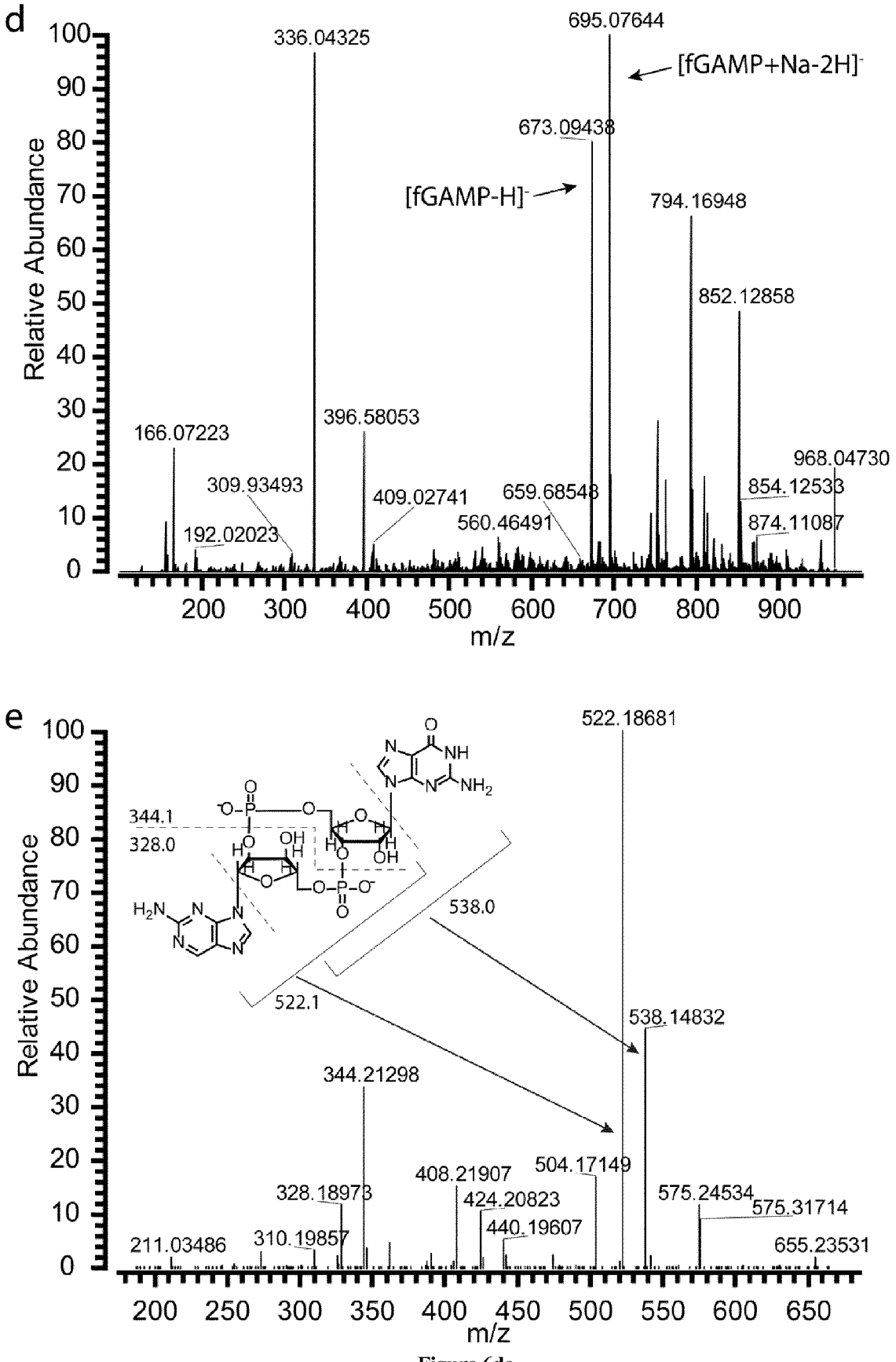
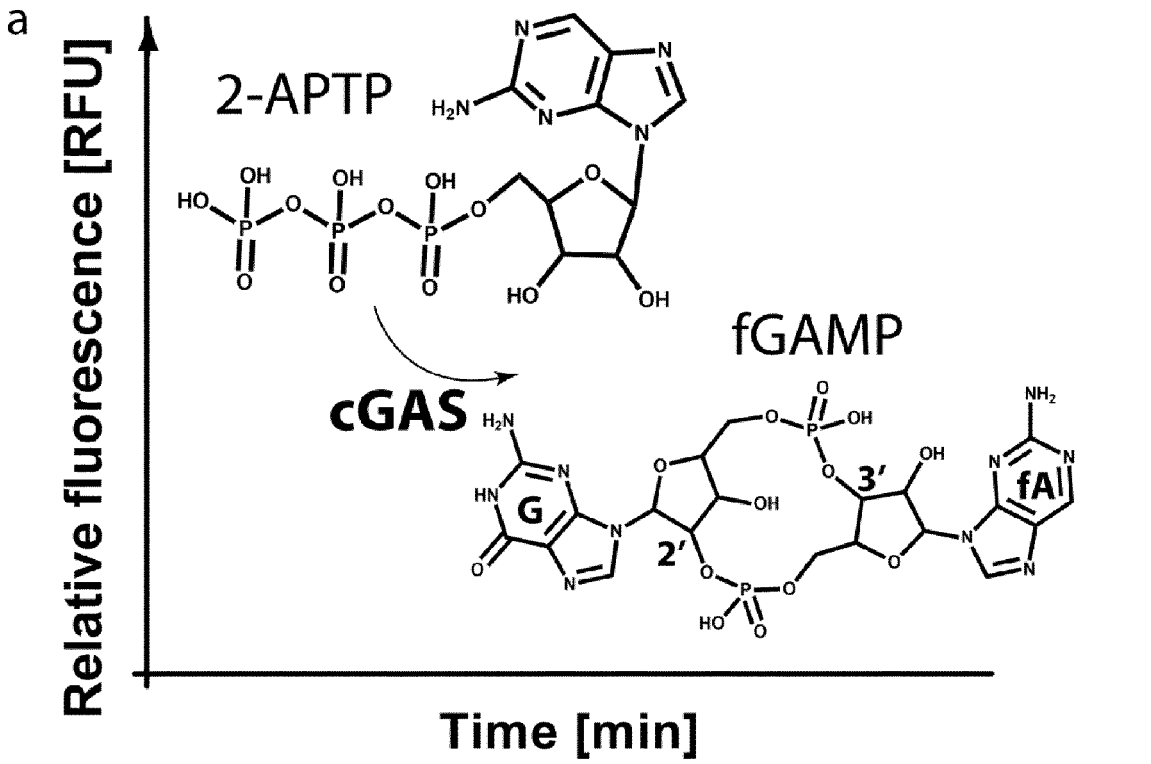

d
Relative Abundance
m/z
336.04325
695.07644
[fGAMP+Na-2H]-
673.09438
[fGAMP-H]-
794.16948
852.12858
396.58053
166.07223
968.04730
309.93493
409.02741
659.68548
854.12533
560.46491
874.11087
192.02023
e
Relative Abundance
m/z
522.18681
344.1
328.0
538.0
522.1
538.14832
344.21298
504.17149
408.21907
328.18973
424.20823
575.24534
575.31714
440.19607
211.03486
310.19857
655.23531

Figure 6de


a
2-APTP
cGAS
fGAMP
Relative fluorescence [RFU]
Time [min]

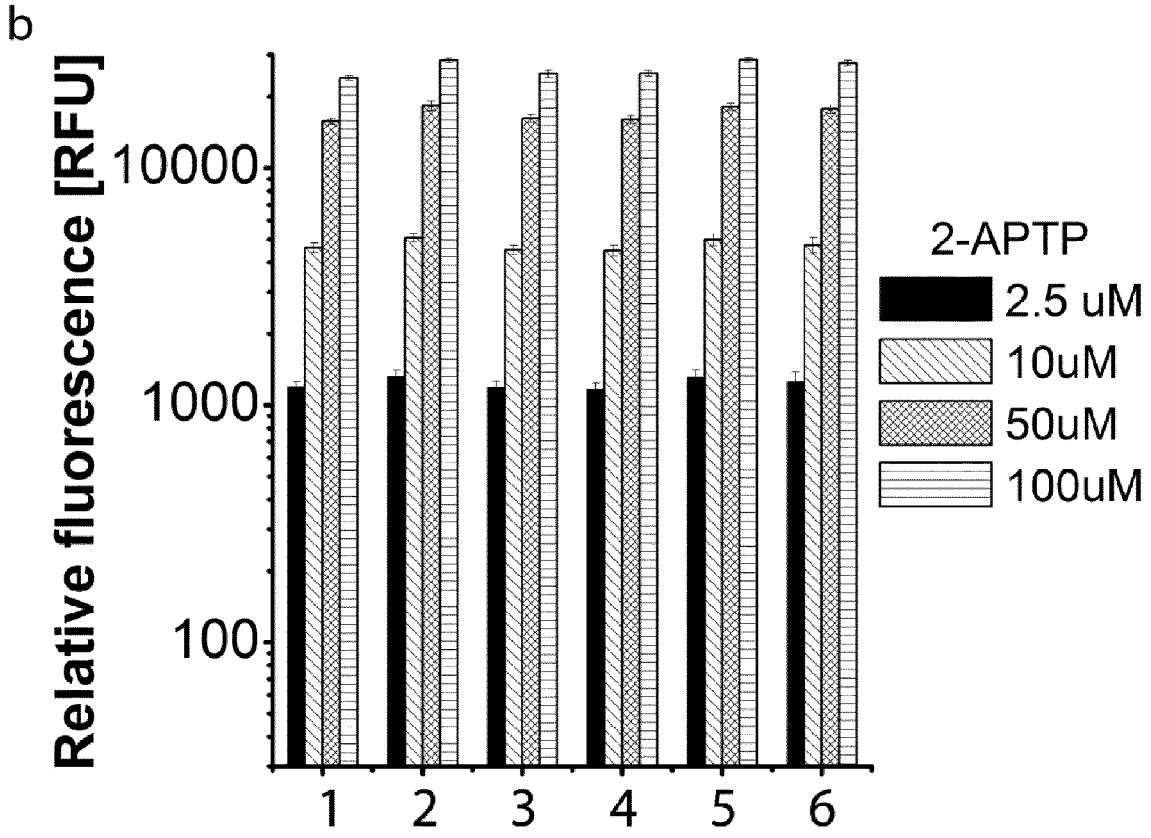

b
Relative fluorescence [RFU]
10000
1000
100
1
2
3
4
5
6
2-APTP
2.5 uM
10uM
50uM
100uM

Figure 7ab

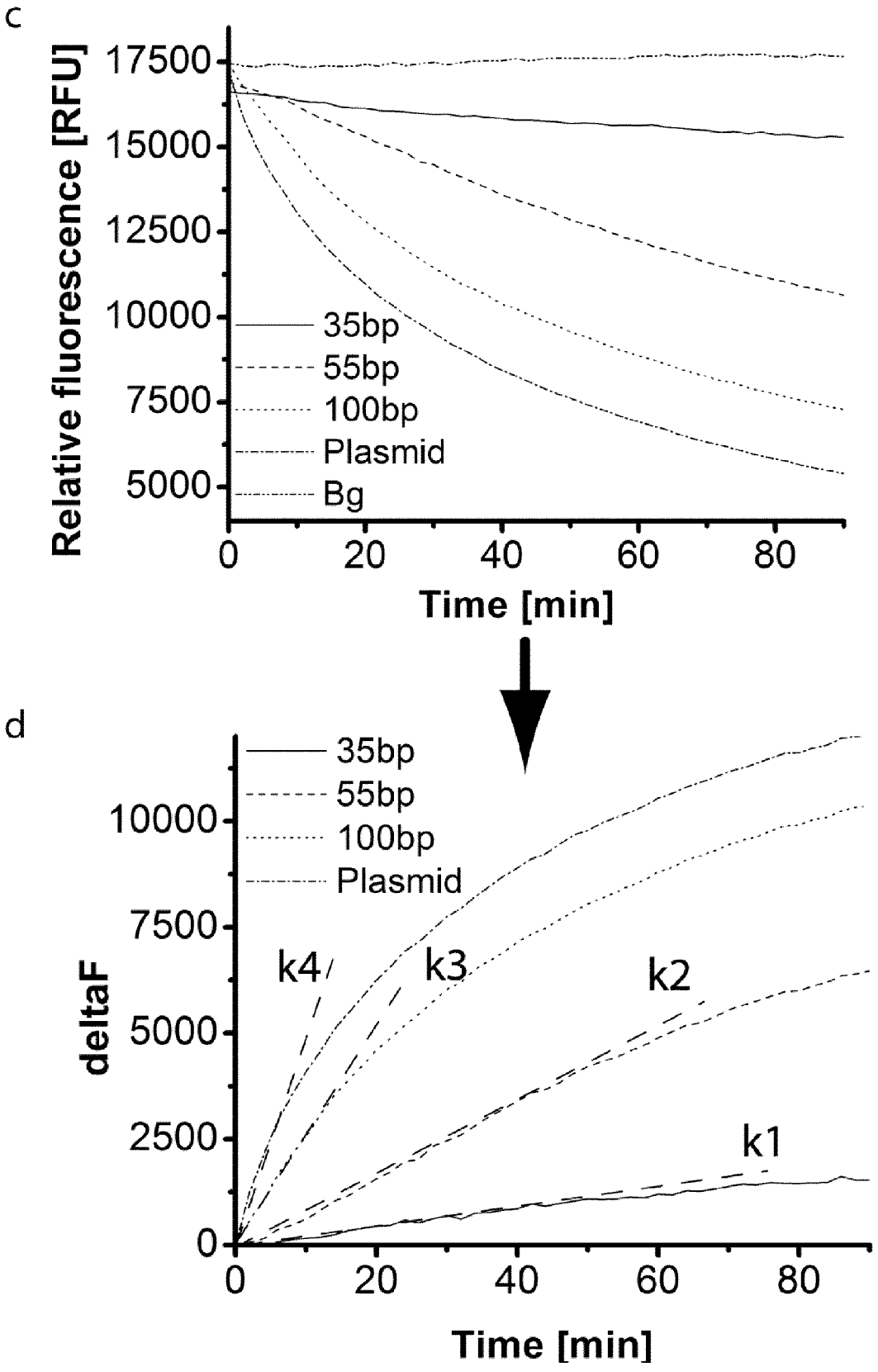

c
Relative fluorescence [RFU]
17500
15000
12500
10000
7500
5000
35bp
55bp
100bp
Plasmid
Bg
0
20
40
60
80
Time [min]
d
35bp
55bp
100bp
Plasmid
10000
7500
5000
2500
0
deltaF
k4
k3
k2
k1
0
20
40
60
80
Time [min]

Figure 7cd

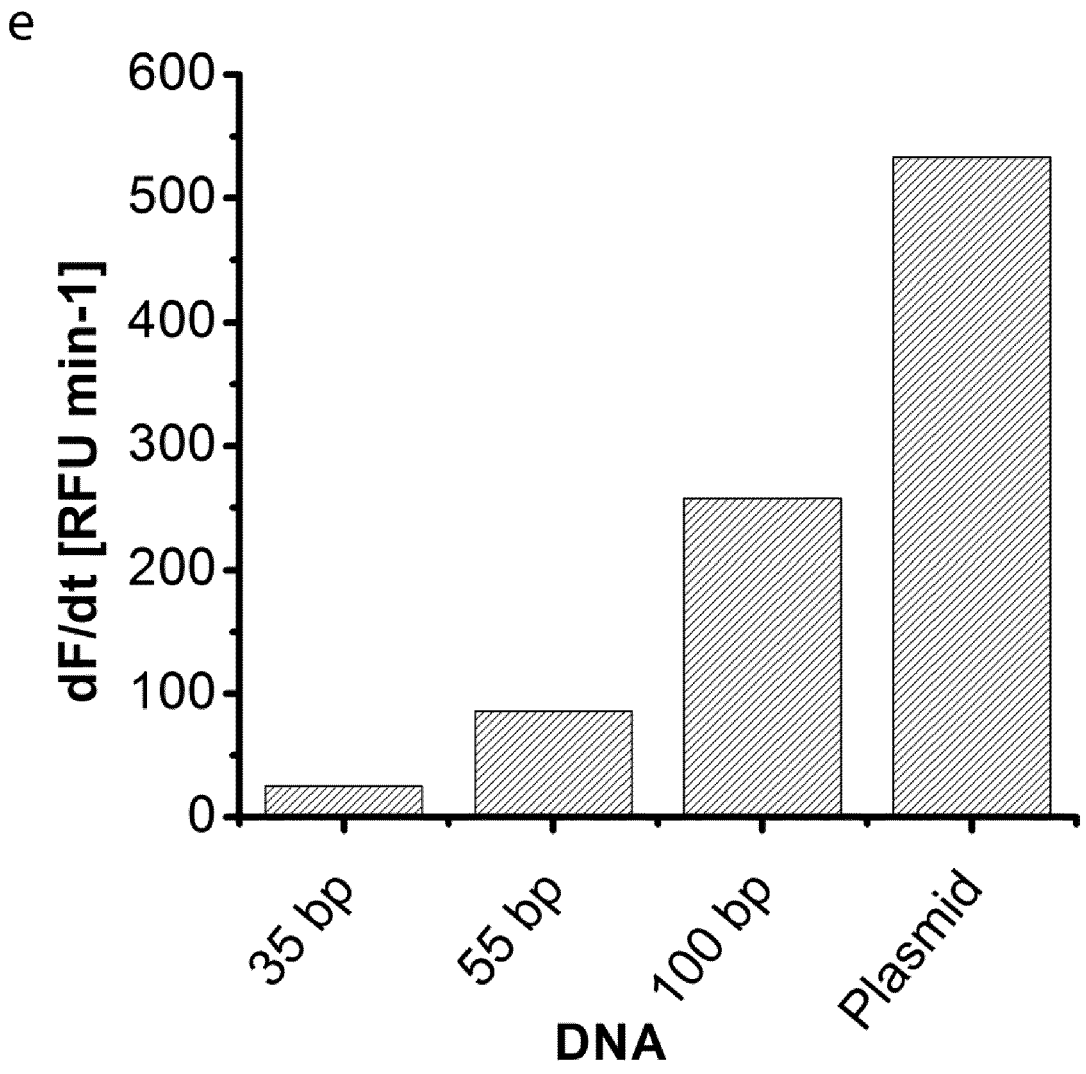
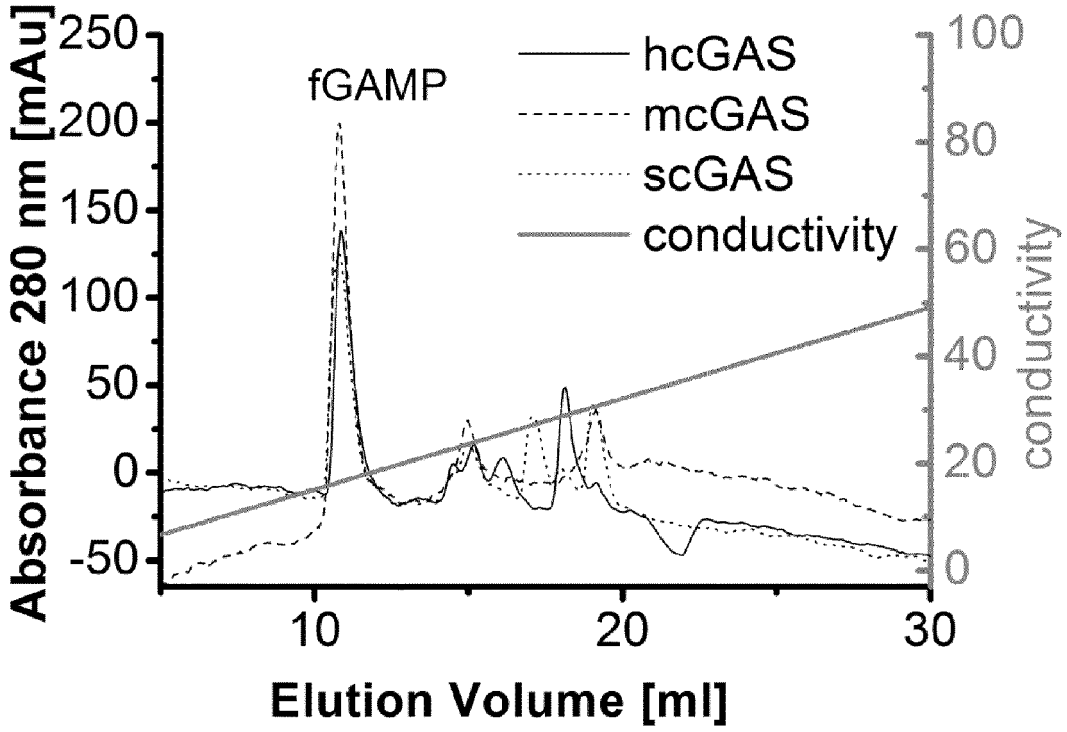

e
600
500
400
300
200
100
0
dF/dt [RFU min-1]
35 bp
55 bp
100 bp
Plasmid
DNA

Figure 7e

a



b

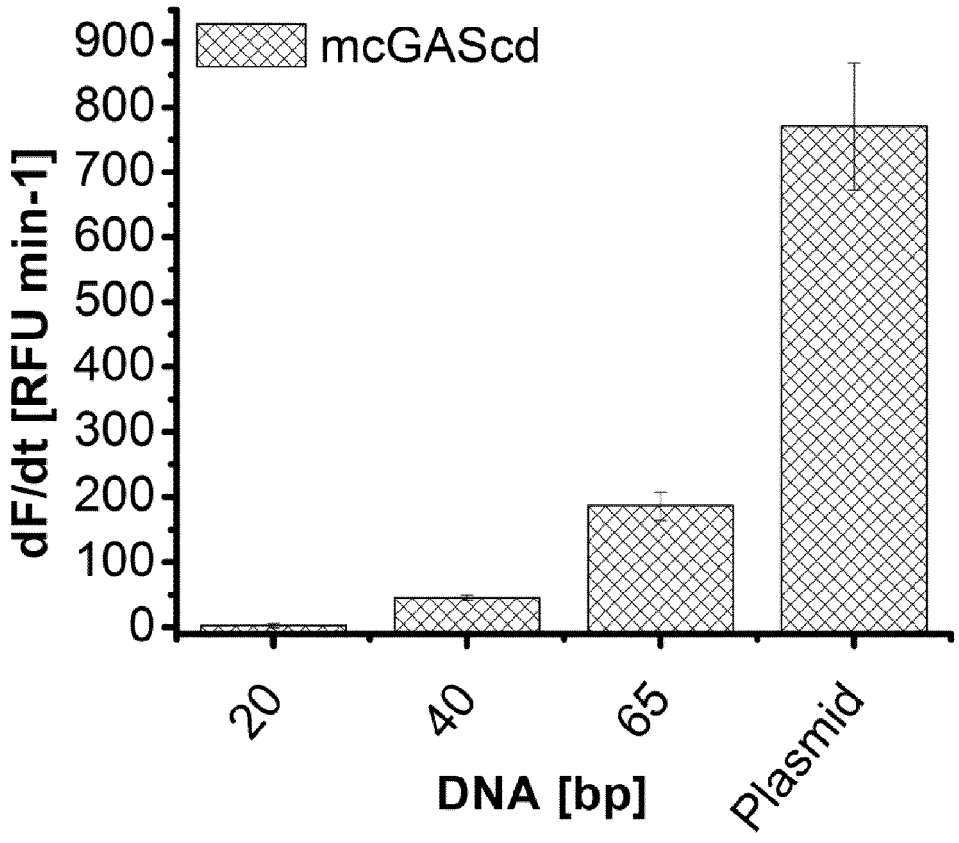


Figure 8_ab

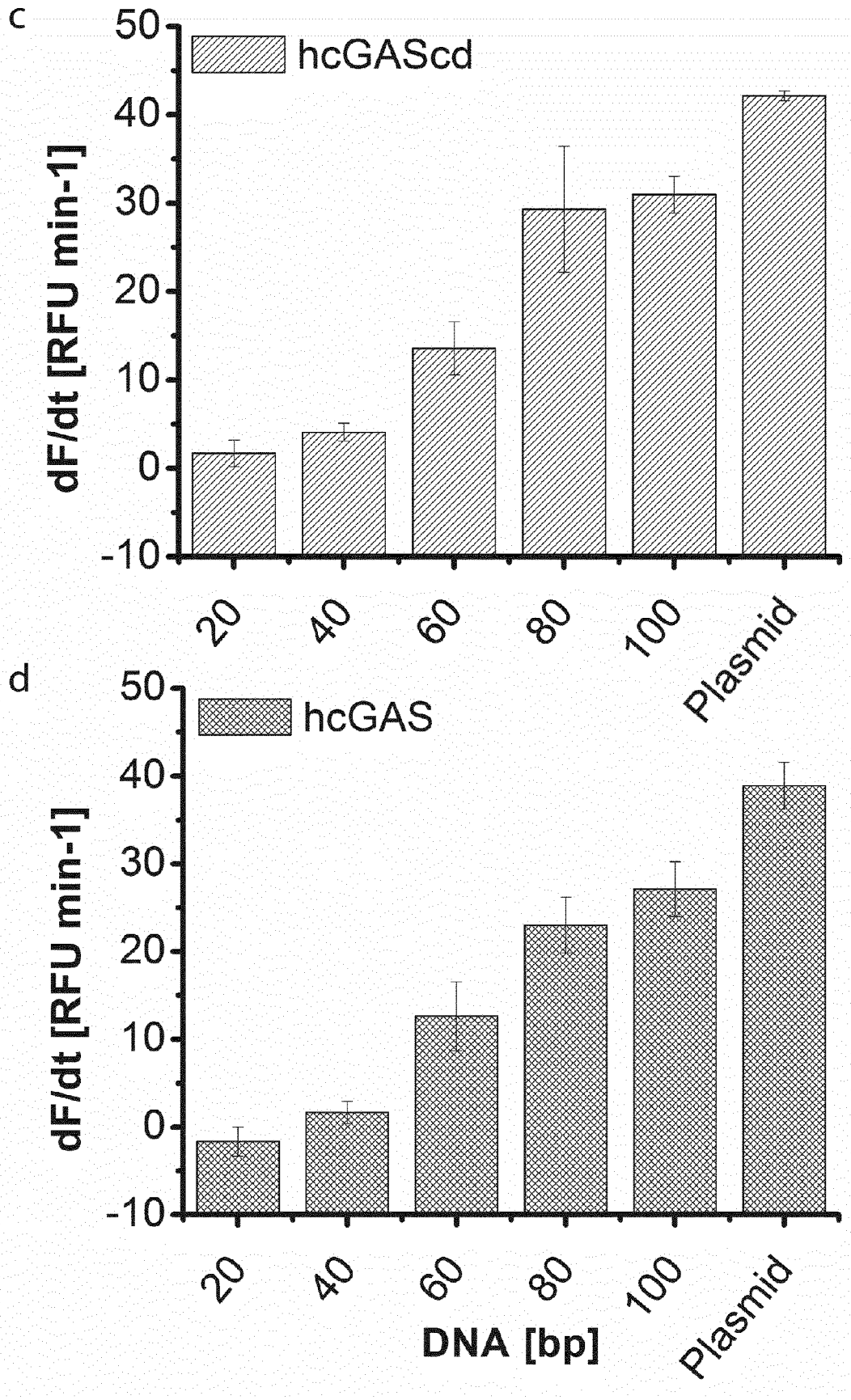

c
hcGAScd
dF/dt [RFU min-1]
50
40
30
20
10
0
-10
20
40
60
80
100
Plasmid
d
hcGAS
dF/dt [RFU min-1]
50
40
30
20
10
0
-10
20
40
60
80
100
Plasmid
DNA [bp]

Figure8_cd

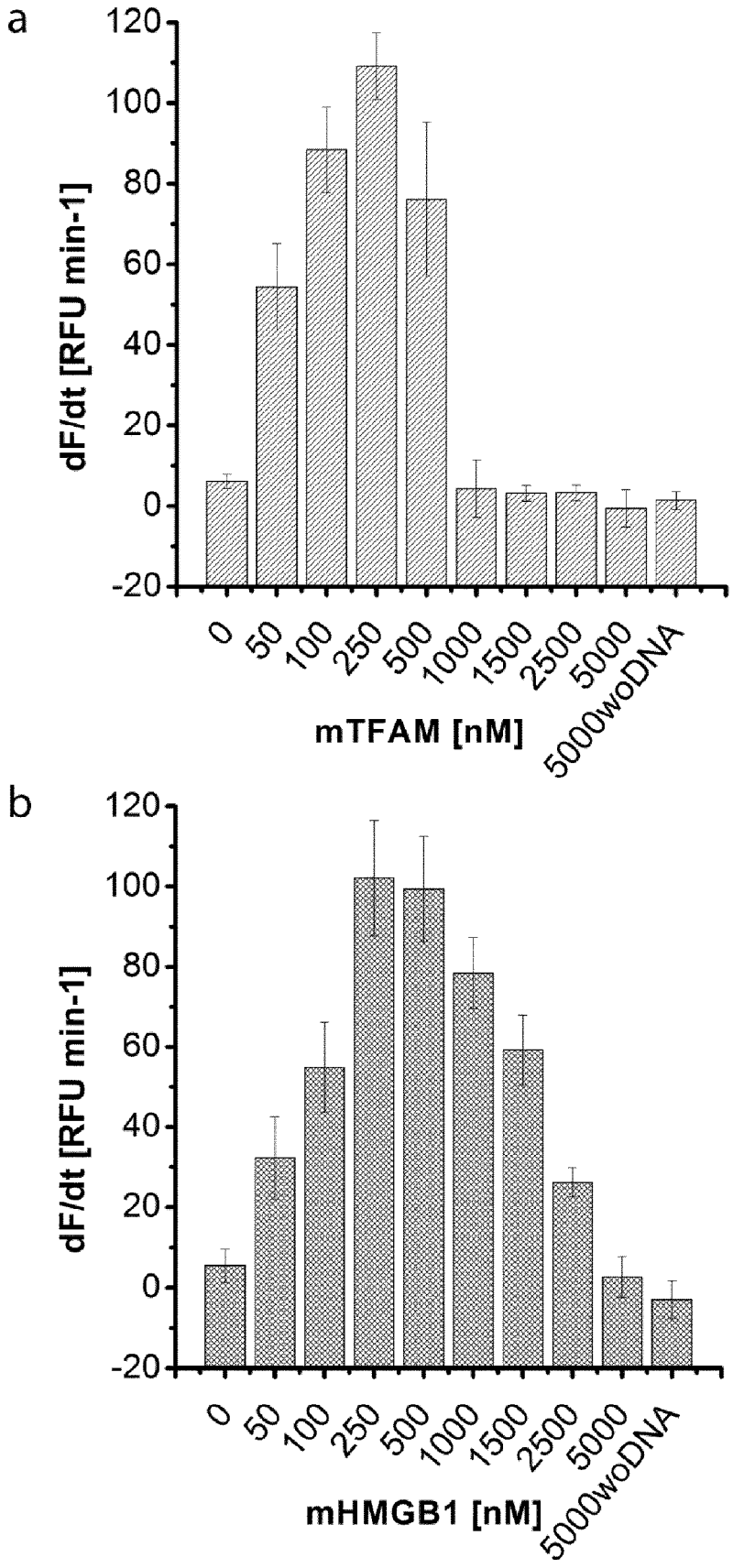

a
120
100
80
60
40
20
0
-20
dF/dt [RFU min-1]
0
50
100
250
500
1000
1500
2500
5000
5000woDNA
mTFAM [nM]
b
120
100
80
60
40
20
0
-20
dF/dt [RFU min-1]
0
50
100
250
500
1000
1500
2500
5000
5000woDNA
mHMGB1 [nM]

Figure 9ab

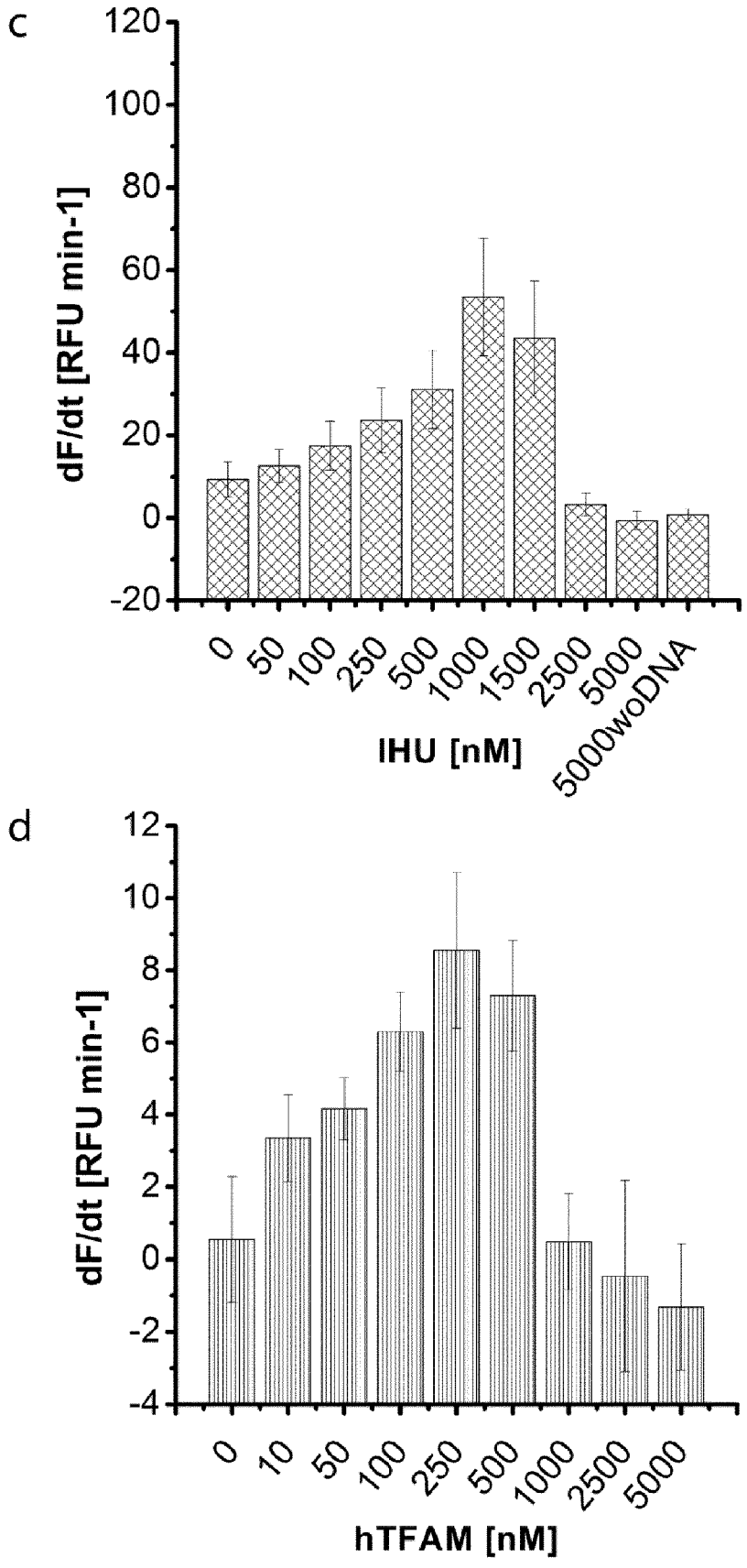

c
dF/dt [RFU min-1]
120
100
80
60
40
20
0
-20
0
50
100
250
500
1000
1500
2500
5000
5000woDNA
IHU [nM]
d
dF/dt [RFU min-1]
12
10
8
6
4
2
0
-2
-4
0
10
50
100
250
500
1000
2500
5000
hTFAM [nM]

Figure9cd

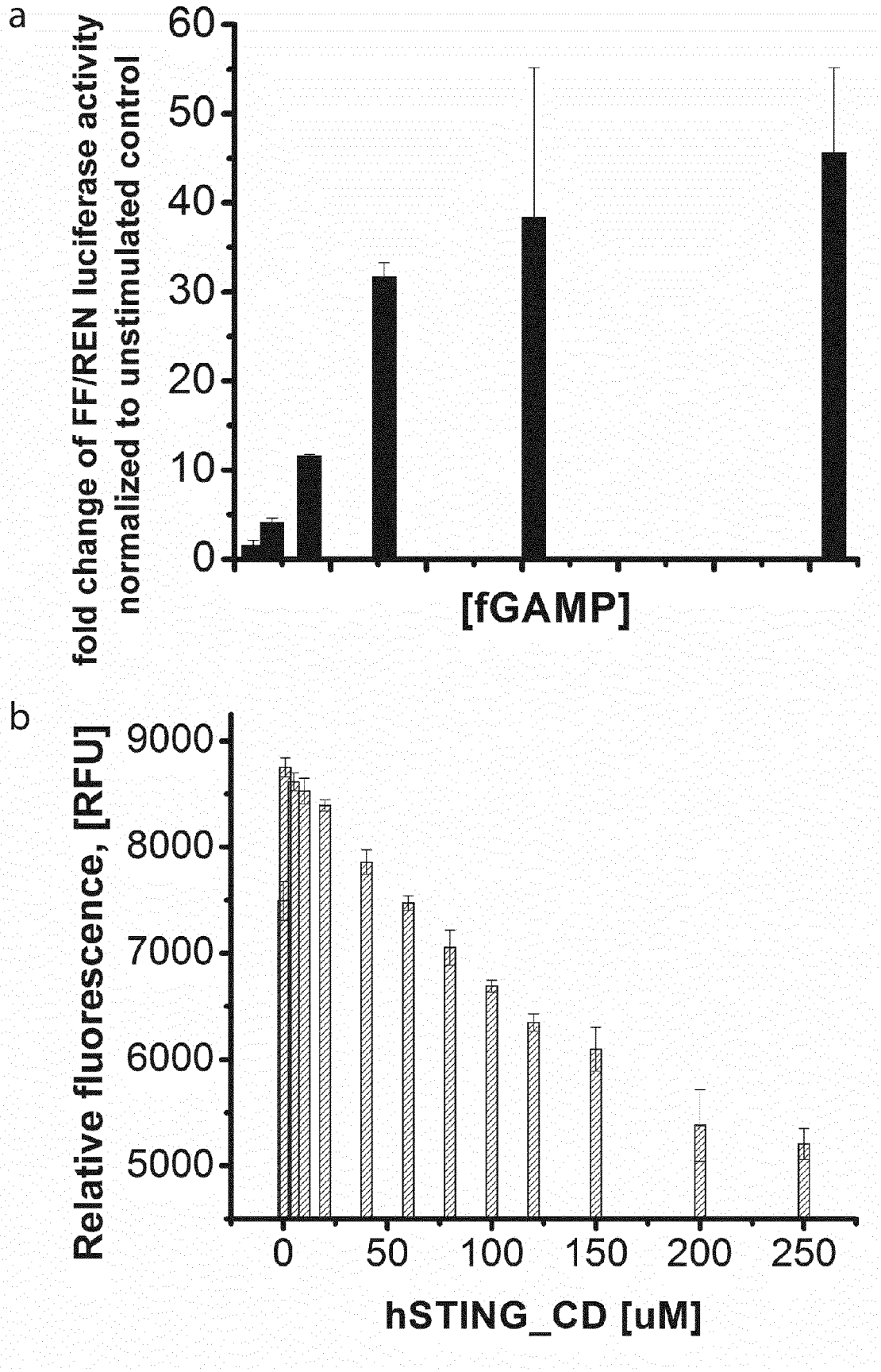

a
fold change of FF/REN luciferase activity
normalized to unstimulated control
60
50
40
30
20
10
0
[fGAMP]
b
Relative fluorescence, [RFU]
9000
8000
7000
6000
5000
0
50
100
150
200
250
hSTING_CD [uM]

Figure 10ab

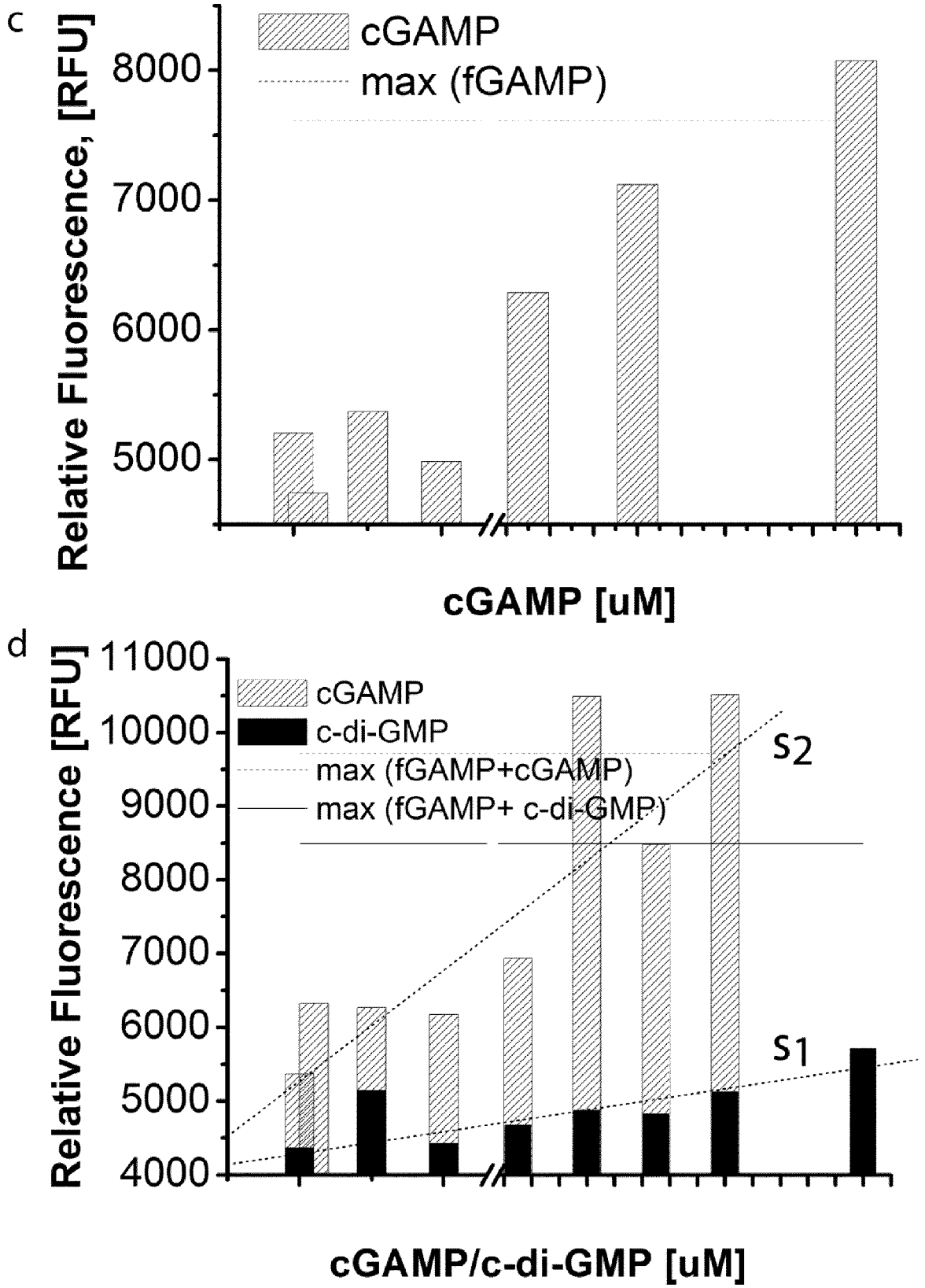

c
cGAMP
max (fGAMP)
Relative Fluorescence, [RFU]
8000
7000
6000
5000
cGAMP [uM]
d
cGAMP
c-di-GMP
max (fGAMP+cGAMP)
max (fGAMP+ c-di-GMP)
S2
S1
Relative Fluorescence [RFU]
11000
10000
9000
8000
7000
6000
5000
4000
cGAMP/c-di-GMP [uM]

Figure10_cd

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 17 18 2689

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	WO 2017/027646 A1 (MERCK SHARP & DOHME CORP., USA) 16 February 2017 (2017-02-16) * the whole document in particular the examples *	1-7,11, 20-22	INV. C07H21/02 C07H21/04 G01N33/52
X	WO 2017/093933 A1 (GLAXOSMITHKLINE IP DEV LTD [GB]) 8 June 2017 (2017-06-08) * the whole document in particular the examples *	1-3, 20-22	
X	WO 2015/077354 A1 (UNIV CHICAGO [US]) 28 May 2015 (2015-05-28) * the whole document in particular claim 23 and the examples *	1-3, 20-22	
X	WO 2014/189805 A1 (AURO BIOTECH, INC., USA; UNIVERSITY OF CALIFORNIA) 27 November 2014 (2014-11-27) * the whole document in particular Figures 4-6 *	1-3, 20-22	
X	WO 2017/027645 A1 (MERCK SHARP & DOHME [US]; ALTMAN MICHAEL D [US]; ANDRESEN BRIAN [US];) 16 February 2017 (2017-02-16) * the whole document in particular the examples *	1-3, 20-22	TECHNICAL FIELDS SEARCHED (IPC) C07H G01N
X	WO 2017/075477 A1 (ADURO BIOTECH INC [US]; NOVARTIS AG [CH]) 4 May 2017 (2017-05-04) * the whole document *	1-3, 20-22	

-/--

~~The present search report has been drawn up for all claims~~

Place of search	Date of completion of the search	Examiner
Munich	11 April 2018	Klein, Didier

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 17 18 2689

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	HIROAKI SAWAI ET AL: "Preparation and properties of oligocytidylates with 2'-5' internucleotide linkage.", BULLETIN OF THE CHEMICAL SOCIETY OF JAPAN, vol. 58, no. 1, 1 January 1985 (1985-01-01), pages 361-366, XP055293995, JP ISSN: 0009-2673, DOI: 10.1246/bcsj.58.361 * table 1 *	1,2	
X	HIROAKI SAWAI ET AL: "Synthesis of 2'-5' Linked Oligouridylates in Aqueous Medium using the Pd2+ Ion", CHEM. PHARM.BULL, vol. 29, no. 8, 1 January 1981 (1981-01-01), pages 2237-2245, XP055322566, * table 1 *	1,2	
T	D. ONIDAS ET AL: "Fluorescence Properties of DNA Nucleosides and Nucleotides: A Refined Steady-State and Femtosecond Investigation", THE JOURNAL OF PHYSICAL CHEMISTRY B, vol. 106, no. 43, 1 October 2002 (2002-10-01), pages 11367-11374, XP55068275, ISSN: 1520-6106, DOI: 10.1021/jp026063g * the whole document *		TECHNICAL FIELDS SEARCHED (IPC)

-/--

~~The present search report has been drawn up for all claims~~

Place of search	Date of completion of the search	Examiner
Munich	11 April 2018	Klein, Didier

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 17 18 2689

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A,D	BENJAMIN T. ROEMBKE ET AL: "A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3',3'-cGAMP", MOLECULAR BIOSYSTEMS, vol. 10, no. 6, 1 January 2014 (2014-01-01), page 1568, XP055172186, ISSN: 1742-206X, DOI: 10.1039/c3mb70518h * the whole document *	12,14-18	
A	US 2013/039933 A1 (BARBER GLEN N [US]) 14 February 2013 (2013-02-14) * the whole document *	12,14-18	

TECHNICAL FIELDS SEARCHED (IPC)

~~The present search report has been drawn up for all claims~~

Place of search	Date of completion of the search	Examiner
Munich	11 April 2018	Klein, Didier

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

Application Number
EP 17 18 2689

CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing claims for which payment was due.

☐ Only part of the claims have been paid within the prescribed time limit. The present European search report has been drawn up for those claims for which no payment was due and for those claims for which claims fees have been paid, namely claim(s):

☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for those claims for which no payment was due.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

see sheet B

☐ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.

☐ As all searchable claims could be searched without effort justifying an additional fee, the Search Division did not invite payment of any additional fee.

☒ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respect of which search fees have been paid, namely claims:

12, 15-17(completely); 14, 18(partially)

☐ None of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims, namely claims:

☐ The present supplementary European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims (Rule 164 (1) EPC).

Europäisches Patentamt
European Patent Office
Office européen des brevets

LACK OF UNITY OF INVENTION
SHEET B

Application Number
EP 17 18 2689

The Search Division considers that the present European patent application does not comply with the requirements of unity of invention and relates to several inventions or groups of inventions, namely:

1. claims: 1-7, 10, 20-22(completely); 11(partially)

   Compounds comprising a 2'-5' and 3'-5'-dinucleotide wherein at least one of the nucleotide is fluorescent and uses thereof.

   ---

2. claims: 8, 9, 19(completely); 11(partially)

   A method of measuring CGAS activity

   ---

3. claims: 12(completely); 14-18(partially)

   A method of identifying a substance having the ability to bind to stimulators of STING

   ---

4. claims: 13(completely); 14-18(partially)

   A method of identifying a substance having the ability to modulate the activity of 2',5'-phosphodiesterase

   ---

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 17 18 2689

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

11-04-2018

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017027646 A1	16-02-2017	AU 2016304899 A1	01-02-2018
		CA 2995365 A1	16-02-2017
		TW 201718619 A	01-06-2017
		US 2017044206 A1	16-02-2017
		WO 2017027645 A1	16-02-2017
		WO 2017027646 A1	16-02-2017
WO 2017093933 A1	08-06-2017	AU 2016362697 A1	04-01-2018
		CN 107849084 A	27-03-2018
		KR 20180009812 A	29-01-2018
		TW 201731862 A	16-09-2017
		US 2017158724 A1	08-06-2017
		US 2017233430 A1	17-08-2017
		UY 37007 A	30-06-2017
		WO 2017093933 A1	08-06-2017
WO 2015077354 A1	28-05-2015	EP 3071209 A1	28-09-2016
		JP 2016538344 A	08-12-2016
		US 2016287623 A1	06-10-2016
		US 2018028553 A1	01-02-2018
		WO 2015077354 A1	28-05-2015
WO 2014189805 A1	27-11-2014	AU 2014268836 A1	24-09-2015
		CA 2904536 A1	27-11-2014
		CL 2015002522 A1	30-12-2016
		CN 105228450 A	06-01-2016
		CR 20150616 A	19-04-2016
		CU 20150158 A7	30-05-2016
		DO P2015000281 A	29-02-2016
		EP 2996473 A1	23-03-2016
		HK 1219024 A1	24-03-2017
		HK 1222512 A1	07-07-2017
		JP 2016520085 A	11-07-2016
		KR 20160009039 A	25-01-2016
		PE 00802016 A1	21-02-2016
		PH 12015502438 A1	28-03-2016
		SG 11201508273R A	30-12-2015
		US 2015056224 A1	26-02-2015
		US 2017333552 A1	23-11-2017
		WO 2014189805 A1	27-11-2014
WO 2017027645 A1	16-02-2017	AU 2016304899 A1	01-02-2018
		CA 2995365 A1	16-02-2017
		TW 201718619 A	01-06-2017
		US 2017044206 A1	16-02-2017
		WO 2017027645 A1	16-02-2017

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 17 18 2689

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

11-04-2018

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		WO 2017027646 A1	16-02-2017
WO 2017075477 A1	04-05-2017	TW 201726700 A	01-08-2017
		UY 36969 A	31-05-2017
		WO 2017075477 A1	04-05-2017
US 2013039933 A1	14-02-2013	US 2013039933 A1	14-02-2013
		US 2018085432 A1	29-03-2018

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **CROW, HAYWARD et al.** Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat Genet,* 2006, vol. 38, 917-920 **[0005]**
- **LEE-KIRSCH, GONG et al.** Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet,* 2007, vol. 39, 1065-1067 **[0005]**
- **GALL, TREUTING et al.** Autoimmunity Initiates in Nonhematopoietic Cells and Progresses via Lymphocytes in an Interferon-Dependent Autoimmune Disease. *Immunity,* 2012, vol. 36, 120-131 **[0005]**
- **AHN, RUIZ et al.** Intrinsic Self-DNA Triggers Inflammatory Disease Dependent on STING. *The Journal of Immunology,* 2014 **[0005]**
- **CROW, CHASE et al.** Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *American journal of medical genetics,* 2015, 296-312 **[0005]**
- **GAO, LI et al.** Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proceedings of the National Academy of Sciences,* 2015, vol. 112, E5699-E5705 **[0006]**
- **NOWARSKI, GAGLIANI et al.** Innate Immune Cells in Inflammation and Cancer. *Cancer Immunology Research,* 2013, vol. 1, 77-84 **[0007] [0073]**
- **WOO, FUERTES et al.** STING-Dependent Cytosolic DNA Sensing Mediates Innate Immune Recognition of Immunogenic Tumors. *Immunity,* 2014, vol. 41, 830-842 **[0007]**
- **ROEMBKE, ZHOU et al.** A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3',3'-cGAMP. *Molecular BioSystems,* 2014, vol. 10, 1568-1575 **[0008]**
- **ABLASSER, GOLDECK et al.** cGAS produces a 2[prime]-5[prime]-linked cyclic dinucleotide second messenger that activates STING. *Nature,* 2013, vol. 498, 380-384 **[0023]**
- **ABLASSER, SCHMID-BURGK et al.** Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature,* 2013, vol. 503, 530-534 **[0023] [0196] [0230]**
- **CROW, CHASE et al.** Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *American journal of medical genetics,* 296-312 **[0039]**
- **LI, SHU et al.** Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization. *Immunity,* vol. 39, 1019-1031 **[0039]**
- **RUSSELL.** Molecular Cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, 2012, vol. 1 **[0055]**
- **HUGHES, MIKLOS et al.** Gene Synthesis: Methods and Applications. *Methods in Enzymology,* 2011, vol. 498, 277-309 **[0055]**
- Recombinational Cloning Using Gateway and In-Fusion Cloning Schemes. **THROOP ; LABAER.** Current protocols in molecular biology. 2015, vol. 110, 3.20.21-23.20.23 **[0055]**
- **LAKOWICZ.** Principles of Fluorescence Spectroscopy. Springer, 2006 **[0072]**
- **GAO, ASCANO et al.** Structure-Function Analysis of STING Activation by c[G(2',5')pA(3',5')p] and Targeting by Antiviral DMXAA. *Cell,* 2013, vol. 154, 748-762 **[0077]**
- **GAO, ZILLINGER et al.** Binding-Pocket and Lid-Region Substitutions Render Human STING Sensitive to the Species-Specific Drug DMXAA. *Cell Reports,* 2014, vol. 8, 1668-1676 **[0077]**
- **TANG ; WANG.** Single Amino Acid Change in STING Leads to Constitutive Active Signaling. *PLOS ONE,* 2015, vol. 10, e0120090 **[0077]**
- *Expert Opin Drug Discov.,* 2015, vol. 10 (11), 1145-61 **[0112]**
- **MACKENZIE, HUESA et al.** New insights into NPP1 function: Lessons from clinical and animal studies. *Bone,* 2012, vol. 51, 961-968 **[0124]**
- **LI, YIN et al.** Hydrolysis of 2'3'-cGAMP by ENPP1 and design of nonhydrolyzable analogs. *Nat Chem Biol,* 2014, vol. 10, 1043-1048 **[0124]**
- **SORTICA, BUFFON et al.** Association between the ENPP1 K121Q Polymorphism and Risk of Diabetic Kidney Disease: A Systematic Review and Meta-Analysis. *PLOS ONE,* 2015, vol. 10, e0118416 **[0124]**
- **DEY, DEY et al.** Inhibition of innate immune cytosolic surveillance by an M. tuberculosis phosphodiesterase. *Nat Chem Biol,* 2017, vol. 13, 210-217 **[0125]**
- **RAPINO, ROBLES et al.** C/EBPα Induces Highly Efficient Macrophage Transdifferentiation of B Lymphoma and Leukemia Cell Lines and Impairs Their Tumorigenicity. *Cell Reports,* 2013, vol. 3, 1153-1163 **[0196]**

- **M. GAIDT.** Gene Center. Ludwig-Maximilians-Universitat München, Center for Integrated Protein Science Munich **[0196]**
- **CIVRIL, DEIMLING et al.** Structural mechanism of cytosolic DNA sensing by cGAS. *Nature,* 2013, vol. 498, 332-337 **[0197]**
- **CAVLAR, DEIMLING et al.** Species-specific detection of the antiviral small-molecule compound CMA by STING. *The EMBO Jourhal,* 2013, vol. 32, 1440-1450 **[0197]**
- **GAIDT, EBERT et al.** Human Monocytes Engage an Alternative Inflammasome Pathway. *Immunity,* 2016, vol. 44, 833-846 **[0199]**
- **GAO, ASCANO et al.** Cyclic [G(2',5')pA(3',5')p] Is the Metazoan Second Messenger Produced by DNA-Activated Cyclic GMP-AMP Synthase. *Cell,* 2013, vol. 153, 1094-1107 **[0216]**
- **GOLDECK et al.** cGAS produces a 2[prime]-5[prime]-linked cyclic dinucleotide second messenger that activates STING. *Nature,* 2013, vol. 498, 380-384 **[0218]**
- **ABLASSER, A., M. GOLDECK et al.** cGAS produces a 2[prime]-5[prime]-linked cyclic dinucleotide second messenger that activates STING. *Nature,* 2013, vol. 498 (7454), 380-384 **[0238]**
- **ABLASSER, A. ; J. L. SCHMID-BURGK et al.** Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature,* 2013, vol. 503 (7477), 530-534 **[0238]**
- **AHN, J. ; P. RUIZ et al.** Intrinsic Self-DNA Triggers Inflammatory Disease Dependent on STING. *The Journal of Immunology,* 2014 **[0238]**
- **CAVLAR, T. ; T. DEIMLING et al.** Species-specific detection of the antiviral small□molecule compound CMA by STING. *The EMBO Journal,* 2013, vol. 32 (10), 1440-1450 **[0238]**
- **CIVRIL, F. ; T. DEIMLING et al.** Structural mechanism of cytosolic DNA sensing by cGAS. *Nature,* 2013, vol. 498 (7454), 332-337 **[0238]**
- **CROW, Y. J. ; D. S. CHASE et al.** Characterization of Human Disease Phenotypes Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *American journal of medical genetics,* 2015, vol. 0 (2), 296-312 **[0238]**
- **CROW, Y. J. ; B. E. HAYWARD et al.** Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat Genet,* 2006, vol. 38 (8), 917-920 **[0238]**
- **DEY, R. J. ; B. DEY et al.** Inhibition of innate immune cytosolic surveillance by an M. tuberculosis phosphodiesterase. *Nat Chem Biol,* 2017, vol. 13 (2), 210-217 **[0238]**
- **GAIDT, M. M. ; T. S. EBERT et al.** Human Monocytes Engage an Alternative Inflammasome Pathway. *Immunity,* 2016, vol. 44 (4), 833-846 **[0238]**
- **GALL, A. ; P. TREUTING et al.** Autoimmunity Initiates in Nonhematopoietic Cells and Progresses via Lymphocytes in an Interferon-Dependent Autoimmune Disease. *Immunity,* 2012, vol. 36 (1), 120-131 **[0238]**
- **GAO, D. ; T. LI et al.** Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proceedings of the National Academy of Sciences,* 2015, vol. 112 (42), E5699-E5705 **[0238]**
- **GAO, P. ; M. ASCANO et al.** Structure-Function Analysis of STING Activation by c[G(2',5')pA(3',5')p] and Targeting by Antiviral DMXAA. *Cell,* 2013, vol. 154 (4), 748-762 **[0238]**
- **GAO, P. ; T. ZILLINGER et al.** Binding-Pocket and Lid-Region Substitutions Render Human STING Sensitive to the Species-Specific Drug DMXAA. *Cell Reports,* 2014, vol. 8 (6), 1668-1676 **[0238]**
- **HUGHES, R. A. ; A. E. MIKLOS et al.** Gene Synthesis: Methods and Applications. Methods in Enzymology. Academic Press, 2011, vol. 498, 277-309 **[0238]**
- **LAKOWICZ, J. R.** Principles of Fluorescence Spectroscopy. Springer, 2006 **[0238]**
- **LEE-KIRSCH, M. A. ; M. GONG et al.** Mutations in the gene encoding the 3[prime]-5[prime] DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat Genet,* 2007, vol. 39 (9), 1065-1067 **[0238]**
- **LI, L. ; Q. YIN et al.** Hydrolysis of 2'3'-cGAMP by ENPP1 and design of nonhydrolyzable analogs. *Nat Chem Biol,* 2014, vol. 10 (12), 1043-1048 **[0238]**
- **LI, X. ; C. SHU et al.** Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization. *Immunity,* vol. 39 (6), 1019-1031 **[0238]**
- **MACKENZIE, N. C. W. ; C. HUESA et al.** New insights into NPP1 function: Lessons from clinical and animal studies. *Bone,* 2012, vol. 51 (5), 961-968 **[0238]**
- **NOWARSKI, R. ; N. GAGLIANI et al.** Innate Immune Cells in Inflammation and Cancer. *Cancer Immunology Research,* 2013, vol. 1 (2), 77-84 **[0238]**
- **RAPINO, F. ; ELOY F. ROBLES et al.** C/EBPα Induces Highly Efficient Macrophage Transdifferentiation of B Lymphoma and Leukemia Cell Lines and Impairs Their Tumorigenicity. *Cell Reports,* 2013, vol. 3 (4), 1153-1163 **[0238]**
- **ROEMBKE, B. T. ; J. ZHOU et al.** A cyclic dinucleotide containing 2-aminopurine is a general fluorescent sensor for c-di-GMP and 3',3'-cGAMP. *Molecular BioSystems,* 2014, vol. 10 (6), 1568-1575 **[0238]**
- **RUSSELL, J. F. S. A. D. W.** Molecular Cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, 2012, vol. 1 **[0238]**
- **SORTICA, D. A. ; M. P. BUFFON et al.** Association between the ENPP1 K121Q Polymorphism and Risk of Diabetic Kidney Disease: A Systematic Review and Meta-Analysis. *PLOS ONE,* 2015, vol. 10 (3), e0118416 **[0238]**

- **TANG, E. D. ; C.-Y. WANG.** Single Amino Acid Change in STING Leads to Constitutive Active Signaling. *PLOS ONE,* 2015, vol. 10 (3), e0120090 **[0238]**
- Recombinational Cloning Using Gateway and In-Fusion Cloning Schemes. **THROOP, A. L. ; J. LABAER.** Current protocols in molecular biology. 2015, vol. 110, 3.20.21-23.20.23 **[0238]**
- **WOO, S.-R. ; MERCEDES B. FUERTES et al.** STING-Dependent Cytosolic DNA Sensing Mediates Innate Immune Recognition of Immunogenic Tumors. *Immunity,* 2014, vol. 41 (5), 830-842 **[0238]**

5. A Click-Chemistry Linked 2′3′-cGAMP Analogue

Dialer, C. R.; Stazzoni, S.; Drexler, D. J.; Müller, F. M.; Veth, S.; Pichler, A.; Okamura, H.; Witte, G.; Hopfner, K. P.; Carell, T. A Click-Chemistry Linked 2′3′-cGAMP Analogue. *Chem. - A Eur. J.* **2019** *25* (8), 2089–2095.

DOI: https://doi.org/10.1002/chem.201805409

URL: https://onlinelibrary.wiley.com/doi/full/10.1002/chem.201805409

This publication describes the synthesis of a 2′3′-cGAMP analogue with modified phosphodiester linkages. 2′3′-cGAMP is a signalling molecule in the cGAS-STING pathway, which is an important component of the innate immune system. As this pathway was discovered to have a high potential for immunotherapy, a great interest in drug discovery evolved. Cyclic dinucleotides represent suitable STING agonists but possess major issues in membrane crossing due to the negative charge. The compound from this publication has modified uncharged linkages to increase membrane permeability. It consists of one amide and one triazole linkage. The synthesis includes a Cu^{I}-catalyzed click reaction for the triazole incorporation and a macrolactamization for cyclization.

Author contribution

The author of the present thesis analyzed STING binding using isothermal titration calorimetry and differential scanning fluorimetry. Furthermore, he assisted in manuscript writing.

Nucleotide Analogues

A Click-Chemistry Linked 2′3′-cGAMP Analogue

Clemens Reto Dialer,[a] Samuele Stazzoni,[a] David Jan Drexler,[b] Felix Moritz Müller,[a] Simon Veth,[a] Alexander Pichler,[a] Hidenori Okamura,[a] Gregor Witte,[b] Karl-Peter Hopfner,[b] and Thomas Carell*[a]

Abstract: 2′3′-cGAMP is an uncanonical cyclic dinucleotide where one A and one G base are connected via a 3′-5′ and a unique 2′-5′ linkage. The molecule is produced by the cyclase cGAS in response to cytosolic DNA binding. cGAMP activates STING and hence one of the most powerful pathways of innate immunity. cGAMP analogues with uncharged linkages that feature better cellular penetrability are currently highly desired. Here, the synthesis of a cGAMP analogue with one amide and one triazole linkage is reported. The molecule is best prepared via a first Cu^{I}-catalyzed click reaction, which establishes the triazole, while the cyclization is achieved by macrolactamization.

Introduction

Cyclic dinucleotides (CDNs) are important cellular messenger molecules in a variety of organisms.[1] The compounds play a crucial role in a wide range of biological processes, such as signal transduction, control of biofilm formation or quorum sensing.[2] Bacteria produce molecules in which two purine bases are linked via two 3′-5′ phosphate linkages to give symmetrical cyclophane structures.[3] One main example for such a molecule is the c-di-GMP compound **1** shown in Figure 1.[4,5] Biochemically, the compound is generated from the corresponding nucleotide-5′-triphosphates. Recently, an unsymmetrical cyclic dipurine molecule (cGAMP, **2**) was discovered in mammalian cells.[6,7] In this molecule, the two purines are connected via one 3′-5′ and another 2′-5′ linkage.[8] The dinucleotide **2** is assembled by the cyclase cGAS (cyclic GMP-AMP synthase). cGAS is a cytosolic DNA sensor and part of the innate immune system.[9,10] 2′3′-cGAMP (**2**) binds to the transmembrane receptor STING (stimulator of interferon genes) with nanomolar affinity (k_d = 4.59 nM),[11] which activates the type 1 interferon (IFN) pathway.[12–14] Subsequent degradation of cGAMP **2** occurs by the specific cleavage of the 2′-5′ phosphodiester bond by ENPP1 highlighting the importance of this unusual connection.[15,16]

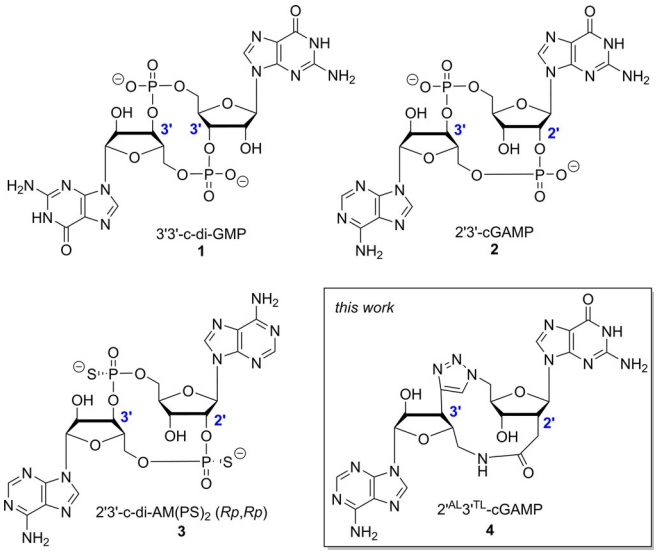


Figure 1. Depiction of the symmetrical microbial c-di-GMP **1**, the unsymmetrical STING activator cGAMP **2**, as well as the bisphosphorothioate analogue **3**, together with the molecule **4** targeted here. AL = amide linked, TL = triazole linked.

There is currently tremendous interest to develop synthetic routes towards analogues of cGAMP **2** as potential agonists or antagonists for cGAS and STING.[17–19] The bisphosphorothioate cGAMP derivative **3**,[20,21] for example, is already in clinical trials.[22,23] Alternative targeting of STING with small molecules is also known.[24–26] Particularly, compounds which lack the negatively charged phosphodiester linkages are discussed as new immune-regulatory pharmaceuticals.[27] While such derivatives are available for symmetric 3′-5′ dinucleotides,[28–32] to the best of our knowledge, uncharged cGAMP **2** analogues do not exist.

[a] C. R. Dialer, S. Stazzoni, F. M. Müller, S. Veth, A. Pichler, Dr. H. Okamura, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@lmu.de
Homepage: www.carellgroup.de

[b] D. J. Drexler, Dr. G. Witte, Prof. Dr. K.-P. Hopfner
Gene Center and Department of Biochemistry
Ludwig-Maximilians-Universität München
Feodor-Lynen-Strasse 25, 81377 Munich (Germany)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/chem.201805409.

In this article, we describe the modular synthesis of a neutral cGAMP analogue **4** that features one triazole and one amide linkage. The triazole was generated by a Cu^{I} catalyzed alkyne-azide click reaction (CuAAC) that was found to be particularly efficient on nucleotides and oligonucleotides.[33–35]

Results and Discussion

We decided to start our synthetic study by synthesizing the cGAMP analogue **4**, in which the 5′-G-3′-A linkage is replaced by a triazole unit and the 2′-G-5′-A linkage is substituted by an amide bond.

Molecular modeling (Figure 2) showed that the analogue **4** is able to adopt a conformation that is similar to the natural ligand bound to STING.[11,36]

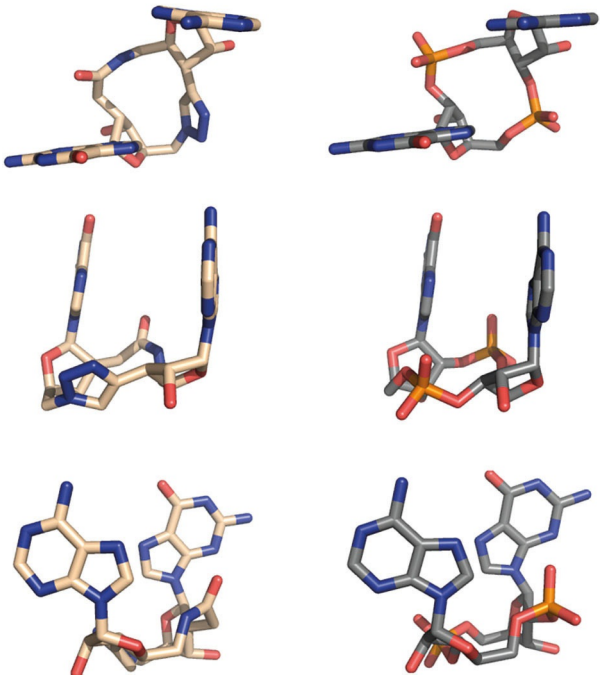


Figure 2. 3D representation showing the potential conformational similarity between compound **4** (left) and natural 2′3′-cGAMP (right, conformation of **2** bound to STING, PDB: 4LOH).

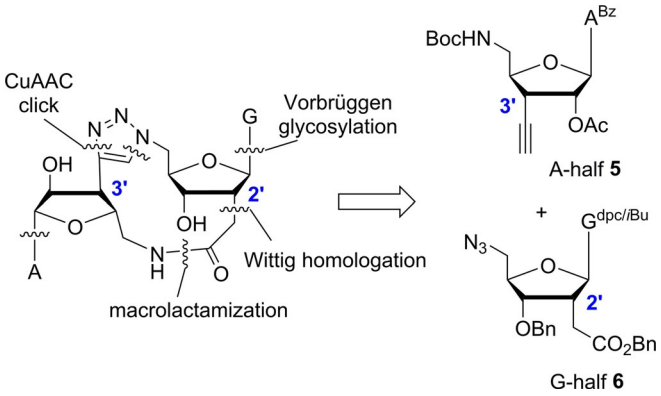
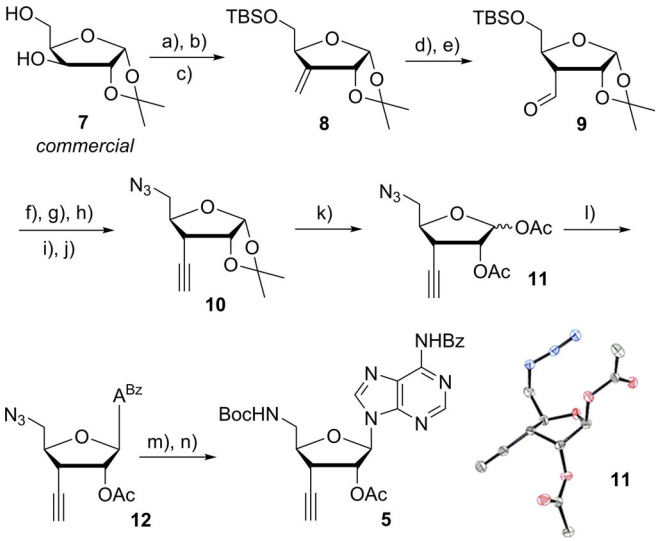


Figure 3. Synthetic strategy towards compound **4**. dpc = diphenylcarbamoyl, *i*Bu = isobutyryl.



Scheme 1. Synthesis of the A-half **5** in 14 steps. a) TBSCl, Py, RT, 2 h, 97%; b) $(COCl)_2$, DMSO, NEt_3, DCM, −60 °C, 3 h; c) CH_3PPh_3Br, BuLi, THF, RT, 6 h, 81% (over two steps); d) BH_3·DMS, THF, RT, 12 h then 30% H_2O_2, 2 N NaOH, RT, 2 h, 76%; e) $(COCl)_2$, DMSO, NEt_3, DCM, −60 °C, 3 h, 93%; f) CBr_4, PPh_3, DCM, 0 °C, 1 h then RT, 12 h, 85%; g) BuLi, THF, −78 °C, 1.5 h, 83%; h) TBAF, THF, RT, 4 h, 95%; i) TsCl, Py, RT, 18 h, 87%; j) NaN_3, DMF, 80 °C, 3 h, 94%; k) HOAc/Ac_2O, H_2SO_4 (cat.), RT, 5 h, 78%; l) 6-*N*-benzoyladenine, BSA, TMSOTf, DCE, 80 °C, 4 h, 61%; m) PMe_3, H_2O, THF, 40 °C, then RT, 12 h, 66%; n) Boc_2O, NEt_3, DCM, RT, 16 h, 64%. Overall yield starting from **7**: 6%.

In both cases, the macrocycle is thought to force the bases into a shifted parallel orientation with the imidazole part of the nucleobases pointing towards each other. This requires *anti*-conformations of both glycosidic bonds. The preferred conformation of compound **4** will be governed by aromatic the triazole unit. For the conformation of the amide we assume a *syn*-conformation due to the small ring size.

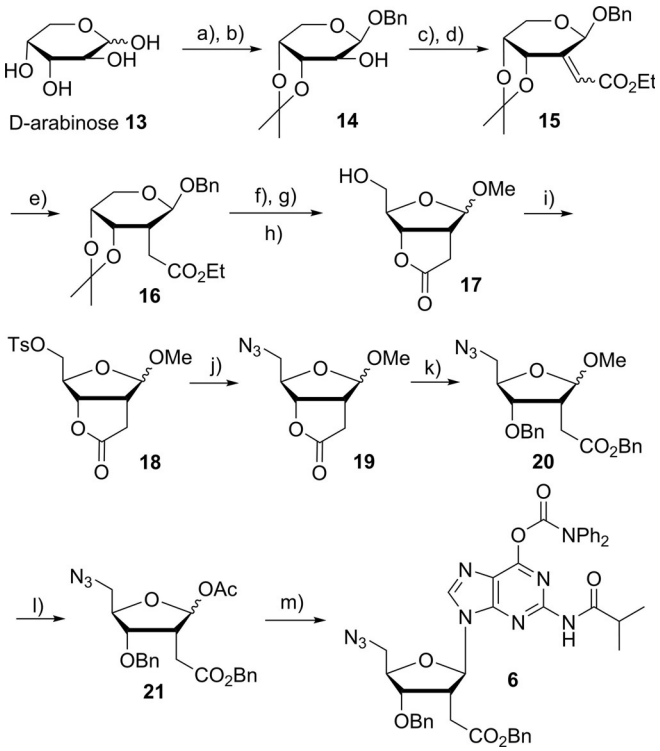
Analysis of potential synthetic accesses of **4** shows that it can be generated by Cu^{I} catalyzed azide alkyne reaction plus a preceding or following lactamization. We developed the synthesis based on the A-half **5** and the corresponding G-half **6** as depicted in Figure 3. For the synthesis of the A-half **5** (Scheme 1), we started with the commercially available 1,2-acetonide protected xylofuranoside **7** (two steps from D-xylose), which we converted in three steps into the 5-TBS-1,2-acetonide protected 3-methylene xylofuranoside **8**.

After stereoselective hydroboration (BH_3·DMS, dr: 9:1) of **8** and Swern oxidation, we obtained the carbonyl compound **9**, which we subjected to a Corey–Fuchs alkinylation (CBr_4, BuLi). TBS deprotection and conversion of the primary hydroxyl group into the azide gave the key intermediate **10**. X-ray analysis of the structure of **10** proved the right configuration of the compound (recrystallisation from isohexanes/ethyl acetate).

Subsequent cleavage of the isopropylidene group and acetyl protection of the hydroxyl groups provided compound **11**, which was the sugar building block for the following glyco-

sylation step. The Vorbrüggen reaction to **12** was found to be most efficient under BSA/TMSOTf conditions with a benzoyl protected A-heterocycle (α/β: 1:12). Finally, we converted the azide via a Staudinger reduction (PMe_3 worked better than PPh_3) into the corresponding amine, which was Boc-protected afterwards to give the A-half **5**.

The desired G-half (Scheme 2) was synthesized starting from D-arabinose (**13**). 1-*O*-Benzyl and 3,4-acetonide protection yielded alcohol **14**.



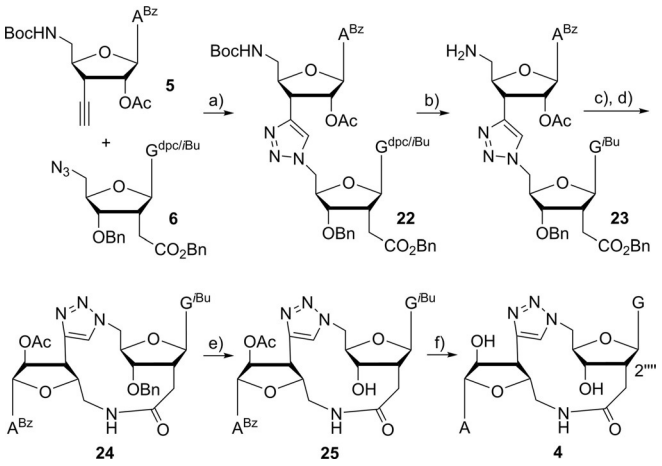
Scheme 2. Synthesis of the G-half **6** in 13 steps. a) AcCl, BnOH, 60 °C, 5 h, 80%; b) $Me_2C(OMe)_2$, Me_2CO, *p*-TsOH (cat.), 60 °C, 2 h, 84%; c) $(COCl)_2$, DMSO, NEt_3, DCM, −60 °C, 3 h; d) Ph_3PCHCO_2Et, DCM, RT, 12 h, 86% (over two steps); e) H_2, Raney-Ni, EtOH, RT, 20 h, 90%; f) H_2, Pd/C, EtOH/THF, 36 h, 88%; g) 80% HOAc, RT, 24 h; h) H_2SO_4 (cat.), MeOH, 4 °C, 3d, 72% (over two steps); i) TsCl, Py, RT, 18 h, 76%; j) NaN_3, DMF, 80 °C, 3 h, 75%; k) BnBr, KOH, THF, reflux, 5 h, 91%; l) HOAc/Ac_2O, H_2SO4 (cat.), RT, 3 h, 85%; m) 6-*O*-(diphenylcarbamoyl)-2-*N*-isobutyrylguanine ($G^{dpc/iBu}$), BSA, TMSOTf, DCE, 80 °C, 2 h, 72%. Overall yield starting from **13**: 10%.

Subsequent Swern oxidation and Wittig homologation provided the intermediate **15** (*E*/*Z*: 4:1). Employing the acetonide protective group as a stereoselective directing group, compound **16** was almost exclusively obtained in *R*-configuration via a Raney-Ni-assisted hydrogenation (dr: 20:1).

Under these reduction conditions the 1-*O*-benzyl group remained unaffected—keeping the sugar in its pyranoside configuration. Removal of the protective groups and treatment with catalytic amounts of acid furnished at 4 °C selectively the ribofuranoside **17**. This was followed by an in situ lactonization. The resulting alcohol **17** was tosylated and reacted with NaN_3 to give azide **19**. The absolute configuration of the compounds was again proven with a crystal structure of **18** (Supporting Information).

We subsequently opened the lactone ring to compound **20** via hydroxide-mediated benzyl protection and converted it into its 1-*O*-acetyl derivative **21**. The glycosylation reaction to the G-half **6** was performed by a so far unreported Vorbrüggen pattern in high β-selectivity (α/β: 1:14) and good yields (79%).

The assembly of nucleoside building blocks A (**5**) and G (**6**) was initiated by a CuAAC reaction. This reaction went smoothly and provided the dinucleotide **22** in fair yield of 80% (Scheme 3). We noticed that click-approaches with the Boc-deprotected amine compound A gave rise of several side products as monitored by thin-layer chromatography (TLC).



Scheme 3. The assembly towards cyclic dinucleotide **4** in 6 steps. a) $CuSO_4$, Na-ascorbate, THF/*t*BuOH/H_2O, RT, 24 h, 80%; b) TFA/DCM (1:1), 0 °C, 1 h, 81%; c) H_2, Pd/C, EtOH, 36 h; d) HATU, DIPEA, DMF (1 mM), RT, 24 h, 52% (over two steps); e) BCl_3, DCM, −40 °C, 3d; f) NH_3, H_2O/MeOH, 50 °C, 20 h, 48% (over two steps). Overall yield starting from **5** and **6**: 16%.

TFA treatment of dinucleotide **22** resulted in the cleavage of both the Boc and the diphenylcarbamoyl (dpc) group. Besides, this was the last step of the consecutive synthesis where purification could be easily conducted by flash column chromatography (DCM/MeOH, 10:1) due to the increasing polarity of the following compounds. A palladium catalyzed hydrogenation reaction deprotected the benzyl ester by leaving the secondary 3′′′′-*O*-benzyl ether intact. Final macrolactamization with HATU furnished the cyclized dinucleotide **24**.

Deprotection of the 3′′′′-*O*-benzyl ether under BCl_3/DCM conditions (−40 °C) proved to be the best option even though solubility in organic solvents decreased with ongoing removal of protective groups. Final ammonolysis revealed our target molecule **4** in 2% overall yield starting from the G-pathway (19 steps) and 1% starting from the A-pathway (20 steps), respectively. Compounds **24**, **25** and **4** were purified by RP-HPLC and subjected to further NMR-studies.

Conformational Analysis and Conclusion

We performed detailed NOESY experiments in order to determine the conformation preferences of target compound **4** in

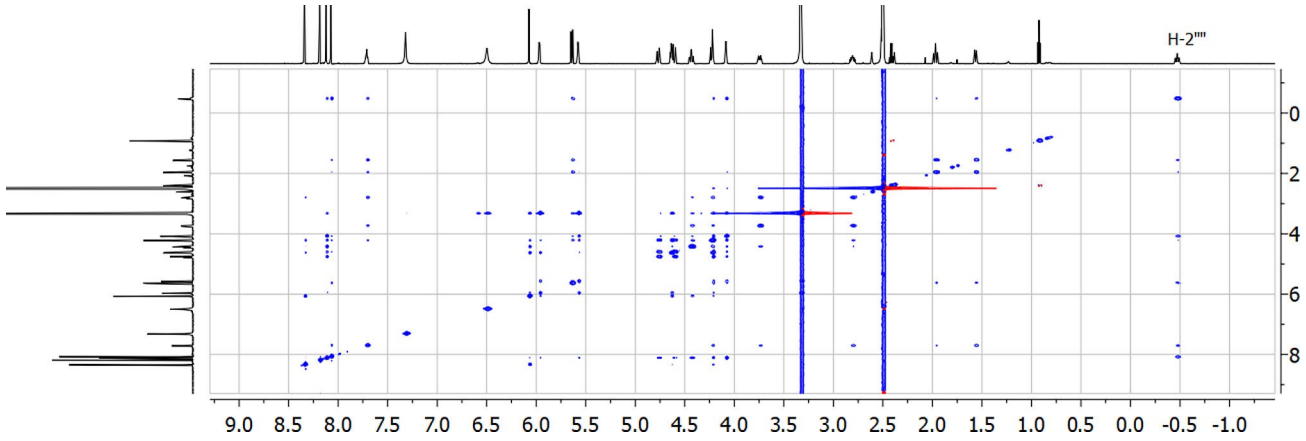


Figure 4. NOESY spectrum of the final compound **4** in [D_6]DMSO.

respect to potential STING binding. The spectrum is shown in Figure 4. The most informative NOE contacts together with a depiction of the modelling results of **4** in solution is shown in Figure 5. The NMR data confirm the overall structure with two

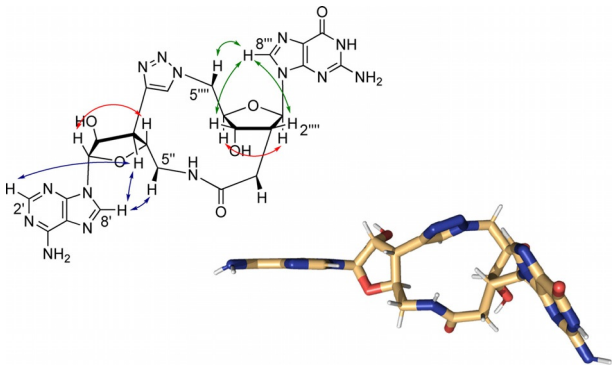


Figure 5. Selected NOE contacts of compound **4** (in [D_6]DMSO) and modelling of the preferred conformation based on the NOE data.

β-configured glycosidic bonds both in anti-conformation. Most interesting, however, is the large shielding of proton H-2′′′′, which shifts from δ(compound **23**) = 4.12 ppm to δ(compound **4**) = −0.47 ppm. This dramatic shift indicates that the proton is positioned just on top of the aromatic triazole ring.

According to this low chemical shift, it is assumed that H-2′′′′ points directly to the triazole ring within the cyclized structures of compounds **24**, **25** and **4**. Unraveling of the conformation just based on the NOE data shows that compound **4** likely adopts a more open conformation in solution ([D_6]DMSO) compared to cGAMP **2**, with the two heterocycles being not parallel to each other.

Potential binding of compound **4** to STING was tested in vitro by nanoDSF assays and analysis of thermal unfolding of the STING constructs hSTING L139 (human STING AA139-379) and mSTING L138 (mouse STING AA138-378). We used the physiological ligand 2′3′-cGAMP and a ligand with lower affinity, 3′3′-cGAMP, as positive controls. As expected after the conformational analysis of compound **4**, binding to hSTING or mSTING could not be detected. This result was confirmed with ITC experiments (see supporting information). Based on the more open structure of the here prepared compound **4**, we believe that interaction studies with cGAS or ENPP1 may be more promising. Investigations in this direction are on the way.

In summary, we report the first synthesis of a 2′3′-cGAMP analogue which features uncharged bridges that should provide membrane crossing properties. The synthetic strategy involved first linking of the two nucleotides by a Cu^{I}-catalyzed click reaction followed by a macrolactamization to close the cycle. The synthesis of medium size ring structures is always difficult. We believe that the here described strategy will open the access to a variety of derivatives of **4**. This allows systematic scanning of the conformational space of the two nucleobases relative to each other regarding the binding to the involved proteins STING, cGAS and ENPP1.

Experimental Section

Unless otherwise specified, all reactions were magnetically stirred under an N_2 atmosphere. Reaction vessels were dried under high vacuum at 550 °C prior to use. Dry solvents and reagents were purchased from commercial suppliers, such as Sigma–Aldrich, Acros Organics, Carbosynth, TCI Europe, ABCR, VWR, stored under septum over molecular sieves and used as received. The reaction progress and fractions during column chromatography were monitored by TLC on silica gel 60-F_{254} plates purchased from Merck and visualized by irradiation with UV-light (254 nm or 366 nm) and *p*-anisaldehyde staining solution (*p*-anisaldehyde (3.7 mL), EtOH (135 mL), conc. H_2SO_4 (5 mL), conc. AcOH (1.5 mL)). Purification was performed using flash column chromatography with silica gel (Merck, particle size 0.063–0.200 mm). The eluents used were determined by TLC. Purification of the crude dinucleotides **24**, **25** and **4** was operated by Waters 2695 reversed phase high performance liquid chromatography (RP-HPLC) using Nucleosil columns (250/4 mm, C18ec, particle size 3 µm for analysis or 250/10 mm, C18ec, 5 µm for purification) from Machery–Nagel with a buffer-free H_2O/MeCN eluent system. Water was purified by a Milli-Q Plus system from Merck Millipore. NMR-spectra were measured on a *Bruker Ascend 400* or *Bruker ARX 600* at room temperature operating at 400 MHz or 600 MHz for ^{1}H-nuclei and at 101 MHz or 151 MHz for ^{13}C-nuclei. The chemical shift (δ) in the NMR-spectra is

reported in parts per million (ppm) and referenced by the residual solvent signal. Measurements were performed in $CDCl_3$ and [D_6]DMSO. The spectra were referenced to the residual protons and carbons of the solvent ($CHCl_3$: $\delta(^1H)=7.26$ ppm, $\delta(^{13}C)=77.16$ ppm; [D_6]DMSO: $\delta(^1H)=2.50$ ppm, $\delta(^{13}C)=39.52$ ppm). Proton-spectra also show the integral intensity, the multiplicity, abbreviated with s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) and the coupling constant (*J* in Hz). Assignments of the signals were performed using 2D-NMR techniques such as homonuclear correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY), heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond coherence (HMBC). All spectra were analyzed with the software MestReNOVA 10.0 from Mestrelab Research, S. L. Atom labelling and nomenclature are not in correspondence with IUPAC. High resolution mass spectra (HRMS) were measured on a *Thermo Finnigan MAT 95* (EI) and a *Thermo Finnigan LTQ FTICR* (ESI). IR-measurements were performed on a PerkinElmer Spectrum BX FT-IR spectrometer with a diamond-ATR (Attenuated Total Reflection) setup. Uncorrected melting points were determined with an automated Stanford Research Systems *EZ-Melt* apparatus (digital image processing technology). Samples were loaded in open capillary tubes. X-ray crystallography of single crystals was performed on an *Oxford XCalibur diffractometer* and further analysis by the software Ortep-3.[37] The structure of the synthesized analogue **4** in Figure 2 was obtained using the geometry optimization tool of the open source software Avogadro and visualized by PyMol.

Nano differential scanning fluorimetry (nanoDSF)

Thermal melting experiments of STING constructs were performed using a Tycho NT.6 instrument (NanoTemper Technologies). In brief, the samples were heated up in a glass capillary and while heating, the internal fluorescence at 330 nm and 350 nm was recorded. Data analysis, data smoothing, and calculation of derivatives was done using the internal evaluation features of the NT.6 instrument. All measurements were repeated to confirm robustness of the assay.

Isothermal titration calorimetry

ITC experiments were performed using a Malvern PEAQ-ITC system with 20 μM protein in ITC-buffer (20 mM HEPES pH 7.5, 150 mM NaCl) in the cell. The positive controls of cGAMP ligands (Biolog) were titrated in a concentration of 200 μM into the cell by 19 injections of 2 μL, spaced 150 s apart, at 25 °C. Compound **4** was used in a concentration of 291 μM for titration. The results were analyzed using the MicroCal PEAQ-ITC analysis software provided with the instrument. All titrations were repeated to confirm robustness of the assay.

Cloning, expression and purification

Human STING AA139-379 and mouse STING AA138-378 constructs were cloned according to previous studies.[38] The plasmids were used to transform *E. coli* Rosetta (DE3) protein expression strain cells (Novagen). The cells were grown in 1 L of Turbo Broth™ media (Molecular Dimensions) supplemented with Kanamycin (50 mg L^{-1}) and Chloramphenicol (34 mg L^{-1}) at 37 °C to an $OD_{600}=1.3$ and expression was induced by adding IPTG to a final concentration of 0.2 mM. Purification of the STING constructs has been performed as described previously.[38]

4-[6′-Benzoylamino-9′-(2″-*O*-acetyl-5″-amino-3″,5″-dideoxy-β-D-ribofuranosyl)-9′*H*-purin-3″-yl]-1-{9‴-[3‴″-*O*-benzyl-2‴″,5‴″-dideoxy-2‴″-*C*-carboxymethyl-β-D-ribofuranosyl]-2‴-*N*-isobutyrylguanin-5‴″-yl}-2‴″,5″-lactame-1,2,3-triazole (24)

To a stirred solution of dinucleotide **22** (2.12 g, 1.61 mmol, 1.00 equiv.) in dry DCM (40 mL) was added TFA (20 mL) at 0 °C under N_2. The mixture was stirred for 1 h at this temperature and then concentrated in vacuo. The brown residue was purified by flash-column chromatography (silica gel, DCM/MeOH, 100:2→100:5→5:1) to give amino compound **23** as a colorless solid (1.33 g, 1.30 mmol, 81%). mp 128 °C (decomp.). $R_f=0.39$ (DCM/MeOH = 5:1). ESI-HRMS calcd for $[C_{51}H_{52}N_{14}O_{10}+H]^+$: 1021.4064, found: 1021.4038. ESI-HRMS calcd for $[C_{51}H_{52}N_{14}O_{10}-H]^-$: 1019.3918, found: 1019.3918.

To a solution of amino compound **23** (1.08 g, 1.06 mmol, 1.00 equiv.) in EtOH (50 mL) was added Pd/C (10 wt.%, 0.30 g) under nitrogen stream at RT. The reaction vessel was evacuated and flushed with hydrogen three times. The mixture was stirred under hydrogen atmosphere for 36 h and then filtered through celite. The solution was concentrated to dryness under reduced pressure. The residue was used in the next step without further purification. ESI-HRMS calcd for $[C_{44}H_{46}N_{14}O_{10}+H]^+$: 931.3594, found: 931.3594. ESI-HRMS calcd for $[C_{44}H_{46}N_{14}O_{10}-H]^-$: 929.3448, found: 929.3450.

Finally, the title compound was prepared according to a modified procedure of Horne et al.[45] and Kinzie et al.[46] To a yellow solution of the hydrogenated compound **23** and HATU (0.60 g, 1.58 mmol, 1.50 equiv.) in dry DMF (1000 mL) was added DIPEA (0.72 mL, 0.54 g, 4.21 mmol, 4.00 equiv.) at RT. The solution turned orange and was stirred at RT for 24 h. After addition of MeOH (5 mL), volatile materials were removed under reduced pressure and the crude product was purified by flash-column chromatography (silica gel, DCM/MeOH, 100:2→100:5→5:1) to yield cyclized compound **24** as a colorless solid (506 mg, 0.55 mmol, 52% over 2 steps). An analytical sample was provided by RP-HPLC. mp 185 °C (decomp.). $R_f=0.57$ (DCM/MeOH = 5:1). $R_t=16.1$ min (RP-HPLC, 15% to 80% MeCN gradient elution). IR (ATR): $\tilde{\nu}=3220, 1682, 1608, 1454, 1403, 1222, 1049, 797, 708\ cm^{-1}$. ^1H NMR, COSY, NOESY (600 MHz, [D_6]DMSO): $\delta=12.06$ (s, 1H, N*H*), 11.59 (s, 1H, N*H*), 11.27 (s, 1H, N*H*), 8.83 (s, 1H, H-2′), 8.68 (s, 1H, H-8′), 8.54 (s, 1H, H-5), 8.35 (s, 1H, H-8‴), 8.08–8.03 (m, 2H, Bz-*o*-C*H*), 8.06–8.03 (s, 1H, $CH_2CONHCH_2$), 7.68–7.63 (m, 1H, Bz-*p*-C*H*), 7.59–7.53 (m, 2H, Bz-*m*-C*H*), 7.52–7.49 (m, 2H, Bn-*o*-C*H*), 7.44–7.40 (m, 2H, Bn-*m*-C*H*), 7.38–7.34 (m, 1H, Bn-*p*-C*H*), 6.46 (d, $^3J=1.1$ Hz, 1H, H-1″), 5.92 (dd, $^3J=5.9$ Hz, $^3J=1.1$ Hz, 1H, H-2″), 5.66 (d, $^3J=10.2$ Hz, 1H, H-1‴″), 4.85 (dd, $^2J=15.0$ Hz, $^3J=3.3$ Hz, 1H, H_a-5‴″), 4.80–4.77 (m, 1H, H_b-5‴″), 4.80 (d, $^2J=10.9$ Hz, 1H, OCH_2Ph), 4.76 (dd, $^3J=10.6$ Hz, $^3J=6.0$ Hz, 1H, H-3″), 4.68 (d, $^2J=10.9$ Hz, 1H, OCH_2Ph), 4.68–4.66 (m, 1H, H-4‴″), 4.57 (td, $^3J=10.5$ Hz, $^3J=4.1$ Hz, 1H, H-4″), 4.15 (d, $^3J=3.4$ Hz, 1H, H3‴″), 3.87–3.81 (m, 1H, H_a-5″), 2.91–2.83 (m, 1H, H_b-5″), 2.67 (hept, $^3J=6.9$ Hz, 1H, C*H*$(CH_3)_2$), 2.11 (s, 3H, OCOCH_3), 2.04 (t, $^3J=12.0$ Hz, 1H, CH_2CONH), 1.76–1.70 (m, 1H, CH_2CONH), 1.07, (d, $^3J=6.9$ Hz, 3H, CH(CH_3)$_2$), 1.06, (d, $^3J=6.9$ Hz, 3H, CH(CH_3)$_2$), −0.19–0.26 (m, 1H, H-2‴″) ppm. ^{13}C NMR, HSQC, HMBC (151 MHz, [D_6]DMSO): $\delta=180.1$ (*i*Bu-*C*ONH), 170.8 ($CH_2CONHCH_2$), 169.1 (OCOCH$_3$), 165.7 (*N*-*C*O-Ph), 154.8 (C-2‴), 151.8 (C-2′), 151.5 (C-4′), 150.7 (C-6′), 149.1 (C-6‴), 148.30 (C-4‴), 144.0 (C-8′), 142.8 (C-4), 138.2 (OCH$_2$*Ph*-C-1), 137.9 (C-8‴), 133.3 (Bz-*C*-CO-N), 132.5 (Bz-*p*-*C*H), 128.53 (Bz-*o*-*C*H), 128.49 (Bz-*m*-*C*H), 128.47 (Bn-*m*-*C*H), 128.3 (Bn-*o*-*C*H), 127.9 (Bn-*p*-*C*H), 126.8 (C-5), 126.2 (C-5′), 119.7 (C-5‴), 89.3 (C-1″), 83.3 (C-1‴″), 81.2 (H-4″), 79.9 (H-3‴″), 79.8 (H-4‴″), 77.5 (C-2″), 72.0 (OCH$_2$Ph), 53.0 (C-5‴″), 46.9 (C-2‴″), 44.5 (C-3″), 42.4 (C-

5′′), 34.7 (*C*H(CH_3)$_2$), 29.7 (*C*H_2CONH), 20.7 (OCO*C*H_3), 18.81 (CH(*C*H_3)$_2$), 18.79 (CH(*C*H_3)$_2$) ppm. ESI-HRMS calcd for $[C_{44}H_{44}N_{14}O_9+H]^+$: 913.3489, found: 913.3495. ESI-HRMS calcd for $[C_{44}H_{44}N_{14}O_9-H]^-$: 911.3343, found: 911.3348.

4-[6′-Benzoylamino-9′-(2′′-*O*-acetyl-5′′-amino-3′′,5′′-dideoxy-β-D-ribofuranosyl)-9′*H*-purin-3′′-yl]-1-{9′′′-[2′′′′,5′′′′-dideoxy-2′′′′-*C*-carboxymethyl-β-D-ribofuranosyl]-2′′′-*N*-isobutyrylguanin-5′′′′-yl}-2′′′′,5′′-lactame-1,2,3-triazole (25)

To a solution of dinucleotide **24** (340 mg, 0.37 mmol, 1.00 equiv.) in dry DCM (300 mL) was added BCl_3 (5.96 mL, 5.96 mmol, 1 M in DCM, 16.0 equiv.) at −40 °C. The mixture was stirred for 3 days at this temperature, quenched by addition of MeOH (5 mL) and extracted with saturated sodium bicarbonate (20 mL) and DCM (4× 50 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated in vacuo. The compound was used in the next step without further purification. An analytical sample was prepared by RP-HPLC to yield a colorless solid. mp 258 °C (decomp.). R_t = 12.5 min (RP-HPLC, 15% to 80% MeCN gradient elution). IR (ATR): $\tilde{\nu}$ = 3234, 1756, 1677, 1613, 1460, 1403, 1220, 1047, 796, 707 cm^{-1}. ^{1}H NMR, COSY, NOESY (600 MHz, [D_6]DMSO): δ = 12.07 (s, 1H, N*H*), 11.69 (s, 1H, N*H*), 11.26 (s, 1H, N*H*), 8.82 (s, 1H, H-2′), 8.66 (s, 1H, H-8′), 8.43 (s, 1H, H-5), 8.33 (s, 1H, H-8′′′), 8.08–8.03 (m, 1H, Bz-*o*-C*H*), 7.93–7.87 (m, 1H, CH_2CON*H*CH_2), 7.68–7.63 (m, 1H, Bz-*p*-C*H*), 7.59–7.53 (m, 2H, Bz-*m*-C*H*), 6.42 (d, ^3J = 1.1 Hz, 1H, H-1′′), 5.88 (dd, ^3J = 5.9 Hz, ^3J = 1.1 Hz, 1H, H-2′′), 5.76 (d, J = 10.2 Hz, 1H, H-1′′′′), 5.68 (d, J = 3.4 Hz, 1H, O*H*-3′′′′), 4.78 (dd, ^2J = 15.0 Hz, ^3J = 3.3 Hz, 1H, H_a-5′′′′), 4.71 (dd, ^3J = 10.6 Hz, ^3J = 6.0 Hz, 1H, H-3′′), 4.68 (dd, ^2J = 15.0 Hz, ^3J = 1.8 Hz, 1H, H_b-5′′′′), 4.50 (td, ^3J = 10.4 Hz, ^3J = 3.9 Hz, 1H, H-4′′), 4.32–4.29 (m, 1H, H-4′′′′), 4.17–4.13 (m, 1H, H3′′′′), 3.82–3.74 (m, 1H, H_a-5′′), 2.89–2.79 (m, 1H, H_b-5′′), 2.73 (hept, ^3J = 6.8 Hz, 1H, C*H*(CH_3)$_2$), 2.09 (s, 3H, OCO*C*H_3), 2.02 (t, ^3J = 12.0 Hz, 1H, *C*H_2CONH), 1.67–1.60 (m, 1H, *C*H_2CONH), 1.11, (d, ^3J = 6.8 Hz, 3H, CH(*C*H_3)$_2$), 1.10, (d, ^3J = 6.8 Hz, 3H, CH(*C*H_3)$_2$), −0.39−0.49 (m, 1H, H-2′′′′) ppm. ^{13}C NMR, HSQC, HMBC (151 MHz, [D_6]DMSO): δ = 180.1 (*i*Bu-*C*ONH), 171.0 (CH_2*C*ONHCH$_2$), 169.2 (O*C*OCH$_3$), 165.7 (*N*-*C*O-Ph), 154.8 (C-2′′′′), 151.8 (C-2′), 151.5 (C-4′), 150.7 (C-6′), 149.1 (C-6′′′), 148.30 (C-4′′′), 144.0 (C-8′), 142.6 (C-4), 138.2 (C-8′′′), 133.3 (Bz-*C*-CO-N), 132.5 (Bz-*p*-C*H*), 128.52 (Bz-*o*-C*H*), 128.49 (Bz-*m*-C*H*), 126.6 (C-5), 126.2 (C-5′), 119.7 (C-5′′′), 89.3 (C-1′′), 83.2 (C-1′′′′),83.0 (C-4′′′′), 81.0 (C-4′′), 77.4 (C-2′′), 70.6 (C-3′′′′), 52.5 (C-5′′′′), 47.1 (C-2′′′′), 44.5 (C-3′′), 42.5 (C-5′′), 34.8 (*C*H(CH_3)$_2$), 29.2 (*C*H_2CONH), 20.6 (OCO*C*H_3), 18.85 (CH(*C*H_3)$_2$), 18.83 (CH(*C*H_3)$_2$) ppm. ESI-HRMS calcd for $[C_{37}H_{38}N_{14}O_9+H]^+$: 823.3019, found: 823.3015. ESI-HRMS calcd for $[C_{37}H_{38}N_{14}O_9-H]^-$: 821.2873, found: 821.2873.

4-[6′-Amino-9′-(5′′-amino-3′′,5′′-dideoxy-β-D-ribofuranosyl)-9′*H*-purin-3′′-yl]-1-{9′′′-[2′′′′,5′′′′-dideoxy-2′′′′-*C*-carboxymethyl-β-D-ribofuranosyl]-guanin-5′′′′-yl}-2′′′′,5′′-lactame-1,2,3-triazole (4)

The crude compound **25** was dissolved in MeOH (15 mL) and aqueous ammonia (25%, 15 mL) in a sealed vessel at RT. The mixture was stirred at 50 °C for 20 h. Volatile components were removed under reduced pressure. The residue was purified by preparative RP-HPLC to provide the final compound 4 as a colorless solid (109 mg, 0.18 mmol, 48% over 2 steps). mp 270 °C (decomp.). R_t = 7.8 min (RP-HPLC, 15% to 80% MeCN gradient elution). IR (ATR): $\tilde{\nu}$ = 3338, 1639, 1599, 1477, 1419, 1209, 1089, 1047, 1005, 730 cm^{-1}. ^{1}H NMR, COSY, NOESY (600 MHz, [D_6]DMSO): δ = 10.60 (s, 1H, Guanine-N*H*), 8.34 (s, 1H, H-8′), 8.19 (s, 1H, H-2′), 8.12 (s, 1H, H-5), 8.07 (s, 1H, H-8′′′), 7.73–7.69 (m, 1H, CH_2CON*H*CH_2), 7.32 (s, 2H, A-NH_2), 6.50 (s, 2H, G-NH_2), 6.07 (d, ^3J = 1.1 Hz, 1H, H-1′′), 5.97 (d, ^3J = 5.1 Hz, 1H, H-2′′), 5.64 (d, J = 10.3 Hz, 1H, H-1′′′′), 5.58 (d, J = 3.6 Hz, 1H, O*H*-3′′′′), 4.77 (dd, ^2J = 14.9 Hz, ^3J = 3.1 Hz, 1H, H_a-5′′′′), 4.65–4.62 (m, 1H, O*H*-2′′), 4.61 (dd, ^2J = 14.9 Hz, ^3J = 1.5 Hz, 1H, H_b-5′′′′), 4.43 (td, ^3J = 10.5 Hz, ^3J = 4.1 Hz, 1H, H-4′′), 4.23–4.21 (m, 1H, H-4′′′′), 4.23 (dd, ^3J = 10.6 Hz, ^3J = 5.6 Hz, 1H, H-3′′), 4.10–4.07 (m, 1H, H-3′′′′), 3.77–3.71 (m, 1H, H_a-5′′), 2.84–2.77 (m, 1H, H_b-5′′), 1.97 (m, 1H, *C*H_2CONH), 1.56 (dd, ^2J = 12.0 Hz, ^3J = 2.2 Hz,1H, *C*H_2CONH), −0.44−0.41 (m, 1H, H-2′′′′) ppm. ^{13}C NMR, HSQC, HMBC (151 MHz, [D_6]DMSO): δ = 170.9 (CH_2*C*ONHCH$_2$), 156.8 (C-2′′′), 156.2 (C-6′), 153.8 (C-6′′′), 152.6 (C-2′), 151.7 (C-4′′′), 148.7 (C-4′), 143.7 (C-4), 139.81 (C-8′), 136.0 (C-8′′′), 126.8 (C-5), 119.5 (C-5′), 116.1 (C-5′′′), 91.9 (C-1′′), 83.7 (C-4′′′′), 82.7 (C-1′′′′), 80.8 (C-4′′), 75.9 (C-2′′), 70.6 (C-3′′′′), 52.2 (C-5′′′′), 46.7 (C-2′′′′), 45.9 (C-3′′), 42.7 (C-5′′), 29.3 (*C*H_2CONH) ppm. ESI-HRMS calcd for $[C_{24}H_{26}N_{14}O_6+H]^+$: 607.2233, found: 607.2231. ESI-HRMS calcd for $[C_{24}H_{26}N_{14}O_6-H]^-$: 605.2087, found: 605.2090.

Acknowledgements

We thank Milda Nainytė, Dr. Tynchtyk Amatov and Dr. Markus Müller for scientific discussions, Dr. David Stephenson for NMR spectroscopy, Dr. Peter Mayer for X-ray analysis and Dr. Werner Spahl for high-resolution mass spectrometry. We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support via the programs SFB1032 (TP-A5), SFB749 (TP-A4) and SPP-1784. Additional funding from the Excellence Cluster EXC117 CiPS^M is acknowledged. This project has received additional funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° EPiR 741912).

Conflict of interest

The authors declare no conflict of interest.

Keywords: cGAMP analogue · click chemistry · cyclophane · macrolactamization · STING

[1] O. Danilchanka, J. J. Mekalanos, *Cell* **2013**, *154*, 962 – 970.
[2] U. Römling, M. Gomelsky, M. Y. Galperin, *Mol. Microbiol.* **2005**, *57*, 629 – 639.
[3] M. Gomelsky, *Mol. Microbiol.* **2011**, *79*, 562 – 565.
[4] R. Hengge, *Nat. Rev. Microbiol.* **2009**, *7*, 263.
[5] C. Shu, G. Yi, T. Watts, C. C. Kao, P. Li, *Nat. Struct. Mol. Biol.* **2012**, *19*, 722.
[6] A. Ablasser, M. Goldeck, T. Cavlar, T. Deimling, G. Witte, I. Röhl, K.-P. Hopfner, J. Ludwig, V. Hornung, *Nature* **2013**, *498*, 380.
[7] J. Wu, L. Sun, X. Chen, F. Du, H. Shi, C. Chen, Z. J. Chen, *Science* **2013**, *339*, 826 – 830.
[8] P. Gao, M. Ascano, Y. Wu, W. Barchet, B. L. Gaffney, L. Barbara, T. Zillinger, Serganov, A. Artem, Y. Liu, Jones, A. Roger, G. Hartmann, T. Tuschl, Patel, J. Dinshaw, *Cell* **2013**, *153*, 1094 – 1107.
[9] L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, *Science* **2013**, *339*, 786 – 791.
[10] Q. Chen, L. Sun, Z. J. Chen, *Nat. Immunol.* **2016**, *17*, 1142
[11] X. Zhang, H. Shi, J. Wu, X. Zhang, L. Sun, C. Chen, Z. J. Chen, *Mol. Cell* **2013**, *51*, 226 – 235.
[12] K. Kato, H. Omura, R. Ishitani, O. Nureki, *Annu. Rev. Biochem.* **2017**, *86*, 541 – 566.
[13] C. Shu, X. Li, P. Li, *Cytokine Growth Factor Rev.* **2014**, *25*, 641 – 648.
[14] X. Cai, Y.-H. Chiu, Z. J. Chen, *Mol. Cell* **2014**, *54*, 289 – 296.

[15] L. Li, Q. Yin, P. Kuss, Z. Maliga, J. L. Millán, H. Wu, T. J. Mitchison, *Nat. Chem. Biol.* **2014**, *10*, 1043.
[16] K. Kato, H. Nishimasu, D. Oikawa, S. Hirano, H. Hirano, G. Kasuya, R. Ishitani, F. Tokunaga, O. Nureki, *Nat. Commun.* **2018**, *9*, 4424.
[17] C. Opoku-Temeng, J. Zhou, Y. Zheng, J. Su, H. O. Sintim, *Chem. Commun.* **2016**, *52*, 9327 – 9342.
[18] F. Schwede, H.-G. Genieser, A. Rentsch in *Non-canonical Cyclic Nucleotides. Handbook of Experimental Pharmacology, Vol. 238* (Ed.: R. Seifert), Springer, Cham, **2017**, pp. 307 – 337.
[19] P. Clivio, S. Coantic-Castex, D. Guillaume, *Chem. Rev.* **2013**, *113*, 7354 – 7401.
[20] K. W. Knouse, J. N. deGruyter, M. A. Schmidt, B. Zheng, J. C. Vantourout, C. Kingston, S. E. Mercer, I. M. Mcdonald, R. E. Olson, Y. Zhu, C. Hang, J. Zhu, C. Yuan, Q. Wang, P. Park, M. D. Eastgate, P. S. Baran, *Science* **2018**, *361*, 1234 – 1238.
[21] T. Lioux, M.-A. Mauny, A. Lamoureux, N. Bascoul, M. Hays, F. Vernejoul, A.-S. Baudru, C. Boularan, J. Lopes-Vicente, G. Qushair, G. Tiraby, *J. Med. Chem.* **2016**, *59*, 10253 – 10267.
[22] J. Fu, D. B. Kanne, M. Leong, L. H. Glickman, S. M. McWhirter, E. Lemmens, K. Mechette, J. J. Leong, P. Lauer, W. Liu, K. E. Sivick, Q. Zeng, K. C. Soares, L. Zheng, D. A. Portnoy, J. J. Woodward, D. M. Pardoll, T. W. Dubensky, Y. Kim, *Science Transl. Med.* **2015**, *7*, 283ra252.
[23] L. Corrales, S. M. McWhirter, T. W. Dubensky Jr., T. F. Gajewski, *J. Clin. Invest.* **2016**, *126*, 2404 – 2411.
[24] S. M. Haag, M. F. Gulen, L. Reymond, A. Gibelin, L. Abrami, A. Decout, M. Heymann, F. G. van der Goot, G. Turcatti, R. Behrendt, A. Ablasser, *Nature* **2018**, *559*, 269 – 273.
[25] S. Khiar, M. Lucas-Hourani, S. Nisole, N. Smith, O. Helynck, M. Bourgine, C. Ruffié, J.-P. Herbeuval, H. Munier-Lehmann, F. Tangy, P.-O. Vidalain, *Sci. Rep.* **2017**, *7*, 2561.
[26] A. D. F. Adli, R. Jahanban-Esfahlan, K. Seidi, S. Samandari-Rad, N. Zarghami, *Chem. Biol. Drug Des.* **2018**, *91*, 996 – 1006.
[27] S. Farkona, E. P. Diamandis, I. M. Blasutig, *BMC Med.* **2016**, *14*, 73.
[28] T. Fujino, K. Okada, H. Isobe, *Tetrahedron Lett.* **2014**, *55*, 2659 – 2661.
[29] C. Pal, T. K. A. Chakraborty, *Asian J. Org. Chem.* **2017**, *6*, 1421 – 1427.
[30] B. L. Gaffney, R. A. Jones, *Org. Lett.* **2014**, *16*, 158 – 161.
[31] T. Kline, S. R. Jackson, W. Deng, C. L. M. J. Verlinde, S. I. Miller, *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 1282 – 1300.
[32] S. Fernicola, I. Torquati, A. Paiardini, G. Giardina, G. Rampioni, M. Messina, L. Leoni, F. Del Bello, R. Petrelli, S. Rinaldo, L. Cappellacci, F. Cutruzzolà, *J. Med. Chem.* **2015**, *58*, 8269 – 8284.
[33] J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T. Carell, *Org. Lett.* **2006**, *8*, 3639 – 3642.
[34] A. H. El-Sagheer, T. Brown, *Chem. Soc. Rev.* **2010**, *39*, 1388 – 1405.
[35] A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli, T. Brown, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11338 – 11343.
[36] P. Gao, M. Ascano, T. Zillinger, W. Wang, P. Dai, A. A. Serganov, B. L. Gaffney, S. Shuman, R. A. Jones, L. Deng, G. Hartmann, W. Barchet, T. Tuschl, D. J. Patel, *Cell* **2013**, *154*, 748 – 762.
[37] L. Farrugia, *J. Appl. Crystallogr.* **2012**, *45*, 849 – 854.
[38] T. Cavlar, T. Deimling, A. Ablasser, K. P. Hopfner, V. Hornung, *EMBO J.* **2013**, *32*, 1440 – 1450.
[39] V. V. Betkekar, S. Panda, K. P. Kaliappan, *Org. Lett.* **2012**, *14*, 198 – 201.
[40] N.-S. Li, J. Lu, J. A. Piccirilli, *Org. Lett.* **2007**, *9*, 3009 – 3012.
[41] M. J. Webber, S. A. Warren, D. M. Grainger, M. Weston, S. Clark, S. J. Woodhead, L. Powell, S. Stokes, A. Alanine, J. P. Stonehouse, C. S. Frampton, A. J. P. White, A. C. Spivey, *Org. Biomol. Chem.* **2013**, *11*, 2514 – 2533.
[42] T. F. Jenny, K. C. Schneider, S. A. Benner, *Nucleosides Nucleotides* **1992**, *11*, 1257 – 1261.
[43] M. J. Robins, R. Zou, Z. Guo, S. F. Wnuk, *J. Org. Chem.* **1996**, *61*, 9207 – 9212.
[44] K. S. Singh, K. V. Sharma, K. Bohra, E. C. Olsen, K. A. Prasad, *Curr. Org. Synth.* **2014**, *11*, 757 – 766.
[45] W. S. Horne, C. A. Olsen, J. M. Beierle, A. Montero, M. R. Ghadiri, *Angew. Chem. Int. Ed.* **2009**, *48*, 4718 – 4724; *Angew. Chem.* **2009**, *121*, 4812 – 4818.
[46] C. R. Kinzie, A. D. Steele, S. M. Pasciolla, W. M. Wuest, *Tetrahedron Lett.* **2014**, *55*, 4966 – 4968.

Manuscript received: October 29, 2018
Accepted manuscript online: December 10, 2018
Version of record online: January 17, 2019

CHEMISTRY
A European Journal

Supporting Information

A Click-Chemistry Linked 2′3′-cGAMP Analogue

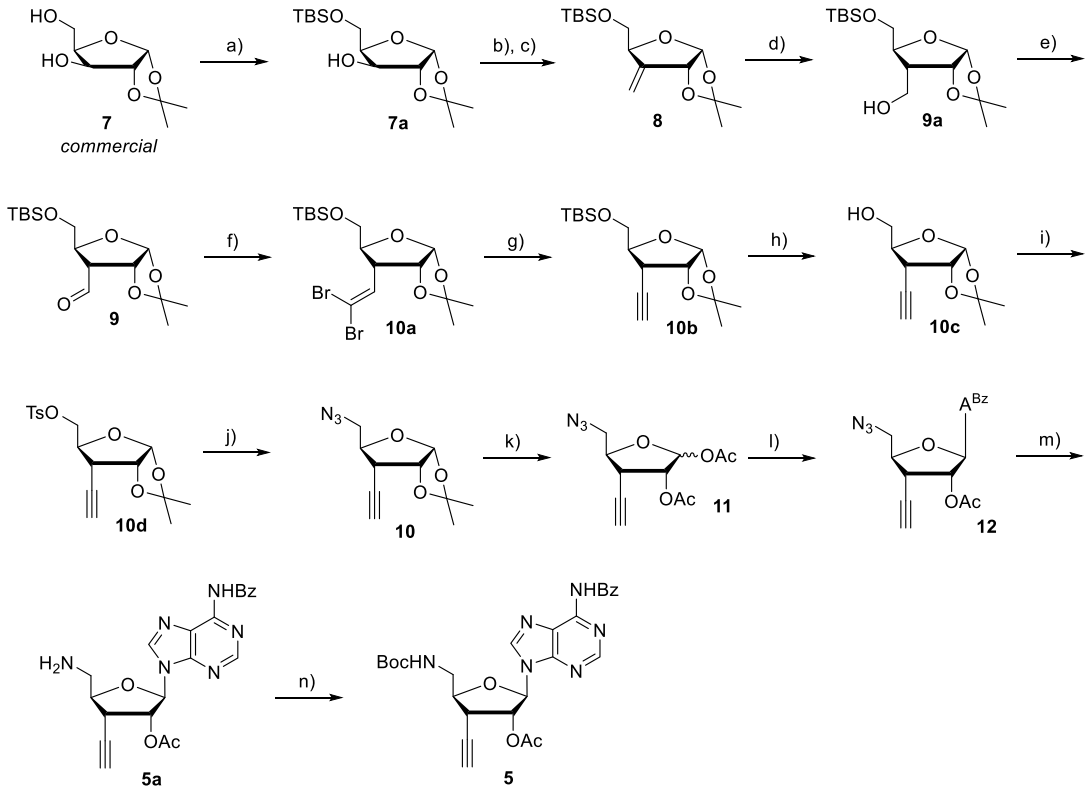
Clemens Reto Dialer,[a] Samuele Stazzoni,[a] David Jan Drexler,[b] Felix Moritz Müller,[a] Simon Veth,[a] Alexander Pichler,[a] Hidenori Okamura,[a] Gregor Witte,[b] Karl-Peter Hopfner,[b] and Thomas Carell*[a]

chem_201805409_sm_miscellaneous_information.pdf

Supporting Information

Table of Contents

1. Experimental procedures



Synthesis of A-half **5**. a) TBSCl, Py, RT, 2h, 97%; b) $(COCl)_2$, DMSO, NEt_3, DCM, -60°C, 3h; c) CH_3PPh_3Br, BuLi, THF, RT, 6h, 81% (over two steps); d) BH_3·DMS, THF, RT, 12h then 30% H_2O_2, 2N NaOH, RT, 2h, 76%; e) $(COCl)_2$, DMSO, NEt_3, DCM, -60°C, 3h, 93%; f) CBr_4, PPh_3, DCM, 0°C, 1h then RT, 12h, 85%; g) BuLi, THF, -78°C, 1.5h, 83%; h) TBAF, THF, RT, 4h, 95%; i) TsCl, Py, RT, 18h, 87%; j) NaN_3, DMF, 80°C, 3h, 94%; k) HOAc/Ac_2O, H_2SO_4 (cat.), RT, 5h, 78%; l) 6-*N*-Benzoyladenine, BSA, TMSOTf, DCE, 80°C, 4h, 61%; m) PMe_3, H_2O, THF, 40 °C, then RT, 12h, 66%; n) Boc_2O, NEt_3, DCM, RT, 16h, 64%. Overall yield starting from **7**: 6%.

5-*O*-(*tert*-Butyldimethylsilyl)-3,3-deoxymethylene-1,2-*O*-isopropylidene-α-D-xylofuranose (8):
The title compound was prepared according to a modified procedure of Betkekar et al.[1] To a solution of oxalyl chloride (12.3 mL, 17.9 g, 141 mmol, 1.10 eq.) in dry DCM (450 mL) was slowly added DMSO (20.0 mL, 22.0 g, 282 mmol, 2.20 eq.) under N_2 at -78 °C. The temperature was maintained below -60 °C and evolving gas was purged. After the mixture was stirred for 1 h at -60 °C, a solution of 5-*O*-(*tert*-butyldimethylsilyl)-1,2-*O*-isopropylidene-α-D-xylofuranose[2] (39.0 g, 128 mmol, 1.00 eq.) in dry DCM (125 mL) was added to the reaction mixture over 5 min and stirred for 2 h. Triethylamine (53.6 mL, 38.9 g, 384 mmol, 3.00 eq.) was added and the suspension was stirred for a further hour at -60 °C. The reaction mixture was warmed to RT, quenched with saturated aqueous $NaHCO_3$ (200 mL) and extracted with DCM (3x200 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo* to obtain 5-*O*-(*tert*-butyldimethylsilyl)-3-oxo-1,2-*O*-isopropylidene-α-D-xylofuranose[3] as a waxy yellow solid. The compound was used in the next step without further purification.
Methyl triphenyl phosphonium bromide (86.2 g, 241 mmol, 2.00 eq.) was suspended in THF (360 mL) and cooled to -78 °C under N_2. *n*-Butyl lithium (96.5 mL, 241 mmol, 2.5M in hexanes, 2.00 eq.) was carefully added dropwise and the resulting red suspension (LiBr precipitates) was stirred for 1 h at 0°C. Subsequent addition of a solution of the crude ketone (36.5 g, 121 mmol, 1.00 eq.) in THF (60 mL) over 10 min gave a slurry which was stirred at RT for 6h. The reaction mixture was quenched with saturated aqueous NH_4Cl (50 mL) and then extracted with ethyl acetate (3 x 200 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. The crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1→4:1) to yield compound **8** as a colorless syrup (31.1 g, 104 mmol, 81% over 2 steps). R_f = 0.68 (silica, isohexanes/EtOAc = 4:1). IR (ATR): $\tilde{\nu}$ = 2930, 1463, 1372, 1252, 1071, 1018, 775 cm^{-1}. ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 5.85 (d, ^3J = 4.1 Hz, 1H, H-1), 5.42 (dd, ^4J = 2.2 Hz, ^4J = 1.1 Hz, 1H, C=CH^a), 5.26 (m, 1H, C=CH^b), 4.88 (dd, ^3J = 4.1 Hz, ^4J = 1.4 Hz, 1H, H-2), 4.75 (ddd, ^3J = 4.2 Hz, ^3J = 3.8 Hz, ^4J = 2.2 Hz, 1H, H-4), 3.75 (dd, ^2J = 10.6 Hz, ^3J = 4.2 Hz, 1H, H_a-5), 3.67 (dd, ^2J = 10.6 Hz, ^3J = 3.8 Hz, 1H, H_b-5), 1.49 (s, 3H, C(CH_3)$_2$), 1.38 (s, 3H, C(CH_3)$_2$), 0.87 (s, 9H, SiC(CH_3)$_3$), 0.043 (s, 3H, Si(CH_3)$_2$), 0.040 (s, 3H, Si(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 147.7 (C-3),112.6 (*C*(CH$_3$)$_2$), 111.7 (C=*C*H$_2$), 105.1 (C-1), 82.1 (C-2), 81.0 (C-4), 65.9 (C-5), 27.7 (C(*C*H$_3$)$_2$), 27.5 (C(*C*H$_3$)$_2$), 26.0 (SiC(*C*H$_3$)$_3$), 18.4 (Si*C*(CH$_3$)$_3$), -5.24 (Si(*C*H$_3$)$_2$), -5.31 (Si(*C*H$_3$)$_2$) ppm. ESI-HRMS calcd. for $[C_{15}H_{28}O_4Si + NH_4]^+$ 318.2095, found: 318.2098.

5-*O*-(*tert*-Butyldimethylsilyl)-3-deoxy-3-(hydroxymethyl)-1,2-*O*-isopropylidene-α-D-ribofuranose (9a):
The title compound was prepared according to a modified procedure of Betkekar et al.[1] To a solution of vinyl compound **8** (29.5 g, 98.2 mmol, 1.00 eq.) in dry THF (300 mL) was added borane dimethyl sulfide complex (73.6 mL, 147 mmol, 2M in THF, 1.50 eq.) at 0 °C. After the solution was stirred for 12 h at RT, aqueous 2 N NaOH (225 mL) was carefully added under strong gas evolution at 0 °C to give a turbid suspension. The reaction mixture was treated slowly with 30% aqueous hydrogen peroxide (98.0 mL) at the same temperature to avoid heat development. The suspension was stirred for further 2 h at RT, quenched with saturated aqueous $Na_2S_2O_3$ solution (200 mL) at 0 °C and finally extracted with EtOAc (3 x 300 mL). The combined organic layers were washed with brine, dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 4:1→2:1→1:1) to afford alcohol **9a** (23.8 g, 74.7 mmol, 76%) as a colorless oil. R_f = 0.76 (isohexanes/EtOAc = 1:1). IR (ATR): $\tilde{\nu}$ = 3456, 2931, 1463, 1381, 1253, 1105, 1019, 778 cm^{-1}. ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 5.79 (d, ^3J = 3.7 Hz, 1H, H-1), 4.75 (dd, ^3J = 4.9 Hz, ^3J = 3.7 Hz, 1H, H-2), 4.08 (ddd, ^3J = 9.9 Hz, ^3J = 6.6 Hz, ^3J = 3.5 Hz, 1H, H-4), 3.88 (dd, ^2J = 10.6 Hz, ^3J = 3.5 Hz, 1H, H_a-5), 3.86 (d, ^3J = 6.2 Hz, 2H, CH_2OH), 3.66 (dd, ^2J = 10.6 Hz, ^3J = 6.6 Hz, 1H, H_b-5), 3.06 (s, 1H, O*H*), 2.19-2.09 (m, 1H, H-3), 1.52 (s, 3H, C(CH_3)$_2$), 1.32 (s, 3H, C(CH_3)$_2$), 0.90 (s, 9H, SiC(CH_3)$_3$), 0.091 (s, 3H, Si(CH_3)$_2$), 0.087 (s, 3H, Si(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 112.2 (*C*(CH$_3$)$_2$), 105.0 (C-1), 82.6 (C-2), 80.7 (C-4), 64.0 (C-5), 59.6 (*C*H$_2$OH), 50.0 (C-3), 26.9 (C(*C*H$_3$)$_2$), 26.4 (C(*C*H$_3$)$_2$), 26.0 (SiC(*C*H$_3$)$_3$), 18.4 (Si*C*(CH$_3$)$_3$), -5.32 (Si(*C*H$_3$)$_2$) ppm. ESI-HRMS calcd. for [$C_{15}H_{30}O_5Si$ + NH_4]$^+$: 336.2201, found: 336.2204.

5-*O*-(*tert*-Butyldimethylsilyl)-3-deoxy-3-formyl-1,2-*O*-isopropylidene-α-D-ribofuranose (9):
The title compound was prepared according to a modified procedure of Parr et al.[3] To a solution of oxalyl chloride (6.91 mL, 10.1 g, 79.6 mmol, 1.10 eq.) in dry DCM (380 mL) was slowly added DMSO (11.3 mL, 12.4 g, 159 mmol, 2.20 eq.) under N_2 at -78 °C. The temperature was maintained below -60 °C and evolving gas was purged. After the mixture was stirred for 1 h at -60 °C, a solution of alcohol **9a** (23.0 g, 72.2 mmol, 1.00 eq.) in dry DCM (70 mL) was added to the reaction mixture over 5 min and stirred for 2 h. Triethylamine (30.2 mL, 21.9 g, 217 mmol, 3.00 eq.) was added and the suspension was stirred for a further hour at -60 °C. The reaction mixture was warmed to RT, quenched with saturated aqueous $NaHCO_3$ (150 mL) and extracted with DCM (3x150 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and volatile components removed under reduced pressure. Purification of the crude product was performed by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1→4:1) to give aldehyde **9** (21.2 g 67.0 mmol, 93%) as a colorless oil. It was also possible to use the crude product in the next step without further purification. R_f = 0.61 (isohexanes/EtOAc = 4:1). IR (ATR): $\tilde{\nu}$ = 2931, 1726, 1472, 1382, 1253, 1100, 1019, 778 cm^{-1}. ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 9.78 (d, ^4J = 1.2 Hz, 1H, C*H*O), 5.87 (d, ^3J = 3.7 Hz, 1H, H-1), 5.03 (dd, ^3J = 5.2 Hz, ^3J = 3.7 Hz, 1H, H-2), 4.55 (ddd, ^3J = 9.5 Hz, ^3J = 3.6 Hz, ^3J = 3.2 Hz, 1H, H-4), 3.86 (dd, ^2J = 11.3 Hz, ^3J = 3.6 Hz, 1H, H_a-5), 3.78 (dd, ^2J = 11.3 Hz, ^3J = 3.2 Hz, 1H, H_b-5), 3.02 (dd, ^3J = 9.5 Hz, ^3J = 5.2 Hz, ^4J = 1.2 Hz, 1H, H-3), 1.48 (s, 3H, C(CH_3)$_2$), 1.33 (s, 3H, C(CH_3)$_2$), 0.86 (s, 9H, SiC(CH_3)$_3$), 0.04 (s, 3H, Si(CH_3)$_2$), 0.03 (s, 3H, Si(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 198.0 (*C*HO), 113.1 (*C*(CH$_3$)$_2$), 105.7 (C-1), 80.8 (C-2), 78.2 (C-4), 62.8 (C-5), 56.6 (C-3), 26.8 (C(*C*H$_3$)$_2$), 26.6 (C(*C*H$_3$)$_2$), 26.0 (SiC(*C*H$_3$)$_3$), 18.5 (Si*C*(CH$_3$)$_3$), 5.31 (Si(*C*H$_3$)$_2$), -5.31 (Si(*C*H$_3$)$_2$) ppm. EI-HRMS calcd. for [$C_{15}H_{28}O_5Si$ - CH_3]$^+$: 301.1466, found: 301.1475.

5-*O*-(*tert*-Butyldimethylsilyl)-3-deoxy-3-(2,2-dibromovinyl)-1,2-*O*-isopropylidene-α-D-ribofuranose (10a):
The title compound was prepared according to a modified procedure of Betkekar et al.[1] A solution of tetrabromomethane (43.0 g, 130 mmol, 2.00 eq.) in DCM (350 mL) was mixed with triphenylphosphine (68.0, 259 mmol, 4.00 eq.) under N_2 at 0 °C and stirred for 1 h at this temperature. The resulting orange solution was treated with a solution of aldehyde **9** (20.5 g, 64.8 mmol, 1.00 eq.) in DCM (120 mL). The dark suspension was stirred at RT for 12 h. After removal of volatile materials, the crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1→4:1) to provide dibromo compound **10a** (26.1 g, 55.3 mmol, 85%) as a slightly yellow oil. R_f = 0.83 (isohexanes/EtOAc = 4:1). IR (ATR): $\tilde{\nu}$ = 2929, 1471, 1382, 1252, 1097, 1020, 777 cm^{-1}. ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 6.49 (d, ^3J = 9.4 Hz, 1H, Br_2CCH), 5.82 (d, ^3J = 3.5 Hz, 1H, H-1), 4.68 (dd, ^3J = 4.5 Hz, ^3J = 3.5 Hz, 1H, H-2), 4.03 (ddd, ^3J = 9.9 Hz, ^3J = 3.6 Hz, ^3J = 3.4 Hz, 1H, H-4), 3.81 (dd, ^2J = 11.5 Hz, ^3J = 3.4 Hz, 1H, H_a-5), 3.66 (dd, ^2J = 11.5 Hz, ^3J = 3.6 Hz, 1H, H_b-5), 3.00 (td, ^3J = 9.9 Hz, ^3J = 4.5 Hz, 1H, H-3), 1.53 (s, 3H, C(CH_3)$_2$), 1.32 (s, 3H, C(CH_3)$_2$), 0.89 (s, 9H, SiC(CH_3)$_3$), 0.06 (s, 6H, Si(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 133.2 (Br_2CCH), 112.2 (*C*(CH$_3$)$_2$), 105.2 (C-1), 91.8 (Br_2CCH), 81.7 (C-2), 80.9 (C-4), 62.3 (C-5), 49.2 (C-3), 26.9 (C(*C*H$_3$)$_2$), 26.4 (C(*C*H$_3$)$_2$), 26.1 (SiC(*C*H$_3$)$_3$), 18.5 (Si*C*(CH$_3$)$_3$), 5.16 (Si(*C*H$_3$)$_2$), -5.22 (Si(*C*H$_3$)$_2$) ppm. EI-HRMS calcd. for [$C_{16}H_{28}Br_2O_4Si$ - CH_3]$^+$: 454.9884, found: 454.9882

5-*O*-(*tert*-Butyldimethylsilyl)-3-deoxy-3-ethynyl-1,2-*O*-isopropylidene-α-D-ribofuranose (10b):
The title compound was prepared according to a modified procedure of Betkekar et al.[1] Dibromo compound **10a** (25.0 g, 52.9 mmol, 1.00 eq.) was dissolved in dry THF (270 mL) and cooled to -78 °C under N_2. *n*-Butyl lithium (48.7 mL, 122 mmol, 2.5M in hexanes, 2.30 eq.) was added dropwise over a period of 10 min until a red solution was formed. After stirring for 1.5 h at this temperature, the reaction mixture was quenched with saturated aqueous NH_4Cl (100 mL) and extracted with EtOAc (3 x 200 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. Purification of the residue was conducted by flash-column chromatography (silica gel, isohexane/EtOAc, 9:1→4:1) to afford ethynyl compound **10b** (13.7 g 43.8 mmol, 83%) as a yellowish oil. R_f = 0.60 (isohexanes/EtOAc, 4:1). IR (ATR): $\tilde{\nu}$ = 3279, 2930, 1472, 1373, 1252, 1215, 1110, 1017, 815, 777 cm^{-1}. ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 5.81 (d, ^3J = 3.6 Hz, 1H, H-1), 4.72 (dd, ^3J = 4.1 Hz, ^3J = 3.6 Hz, 1H, H-2), 4.11 (ddd, ^3J = 10.1 Hz, ^3J = 2.9 Hz, ^3J = 2.0 Hz, 1H, H-4), 3.97 (dd, ^2J = 12.0, ^3J = 2.0 Hz, 1H, H_a-5), 3.79 (dd, ^2J = 12.0, ^3J = 2.9 Hz, 1H, H_b-5), 2.97 (ddd, ^3J = 10.1 Hz, ^3J = 4.1 Hz, ^4J = 2.6 Hz, 1H, H-3), 2.21 (d, ^4J = 2.6 Hz, 1H, CC*H*), 1.57 (s, 3H, C(CH_3)$_2$), 1.36 (s, 3H, C(CH_3)$_2$), 0.89 (s, 9H, SiC(CH_3)$_3$), 0.08 (s, 3H, Si(CH_3)$_2$), 0.07 (s, 3H, Si(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 112.3 (*C*(CH$_3$)$_2$), 105.1 (C-

1), 81.9 (C-4), 81.2 (C-2), 78.2 (*C*CH), 72.3 (C*C*H) , 61.1 (C-5), 36.3 (C-3), 26.8 (C(*C*H$_3$)$_2$), 26.5 (C(*C*H$_3$)$_2$), 26.1 (SiC(*C*H$_3$)$_3$), 18.6 (Si*C*(CH$_3$)$_3$), -5.1 (Si(*C*H$_3$)$_2$), -5.2 (Si(*C*H$_3$)$_2$) ppm. ESI-HRMS calcd. for [$C_{16}H_{28}O_4Si$ + NH_4]$^+$: 330.2095, found: 330.2098.

3-Deoxy-3-ethynyl-1,2-*O*-isopropylidene-α-D-ribofuranose (10c): The title compound was prepared according to a modified procedure of Betkekar et al.[1] To a yellow solution of ethynyl compound **10b** (13.3 g, 42.6 mmol, 1.00 eq.) in THF (200 mL) was added TBAF (55.3 mL, 55.3 mmol, 1M in THF, 1.30 eq.) at RT. The resulting dark solution was stirred for 4 h at this temperature. The reaction mixture was quenched with silica and the solvent was concentrated under reduced pressure. The crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 2:1→1:1→2:3) to yield alcohol **10c** (8.03 g, 40.5 mmol, 95%) as colorless crystals. M.p. = 43 – 45 °C. R_f = 0.38 (isohexanes/EtOAc = 1:1). IR (ATR): ṽ = 3456, 3279, 2936, 1375, 1249, 1215, 1105, 1007, 871 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 5.83 (d, ^3J = 3.5 Hz, 1H, H-1), 4.75 (dd, ^3J = 4.1 Hz, ^3J = 3.5 Hz, 1H, H-2), 4.17 (ddd, ^3J = 10.3 Hz, ^3J = 3.2 Hz, ^3J = 3.0 Hz, 1H, H-4), (4.00 (dd, ^2J = 12.2 Hz, ^3J = 3.2 Hz, 1H, H_a-5), 3.71 (ddd, ^2J = 12.2 Hz, ^3J = 8.8 Hz, ^3J = 3.0 Hz, 1H, H_b-5), 2.96 (ddd, ^3J = 10.3 Hz, ^3J = 4.1 Hz, ^4J = 2.5 Hz, 1H, H-3), 2.23 (d, ^4J = 2.5 Hz, 1H, CC*H*), 1.84 (dd, ^3J = 8.8 Hz, 4.2 Hz, 1H, O*H*), 1.58 (s, 3H, C(C*H*$_3$)$_2$), 1.37 (s, 3H, C(C*H*$_3$)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 112.7 (*C*(CH$_3$)$_2$), 105.1 (C-1), 81.4 (C-4), 81.3 (C-2), 77.6 (*C*CH), 72.6 (C*C*H), 60.6 (C-5), 36.3 (C-3), 26.7 (C(*C*H$_3$)$_2$), 26.5 (C(*C*H$_3$)$_2$) ppm. EI-HRMS calcd. for [$C_{10}H_{14}O_4$ - CH_3]$^+$: 183.0652, found: 183.0652.

3-Deoxy-3-ethynyl-1,2-*O*-isopropylidene-5-*O*-tosyl-α-D-ribofuranose (10d): *p*-Toluenesulfonyl chloride (15.5 g, 56.8 mmol, 1.50 eq.) was dissolved in dry pyridine (50.0 mL) and added to a solution of alcohol **10c** (7.5 g 37.8 mmol, 1.00 eq.) in pyridine (150 mL) at 0 °C. The reaction mixture was stirred for 18 h at RT and finally quenched with MeOH (10 mL). After removal of volatile components, purification of the residue by flash-column chromatography (silica gel isohexanes/EtOAc, 2:1→1:1) gave tosyl compound **10d** (11.7 g, 33.2 mmol, 87%) as colorless crystals. Crystallization from isohexanes/EtOAc (vapor diffusion) provided suitable single crystals for X-ray characterization. M.p. = 113 – 114 °C. R_f = 0.73 (isohexanes/EtOAc = 1:1). IR (ATR): ṽ = 3296, 2990, 1598, 1450, 1360, 1176, 1097, 958, 813 665 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.80 (d, ^3J = 8.2 Hz, 2H, aryl-C*H*-CSO$_3$), 7.34 (d, ^3J = 8.2 Hz, 2H, aryl-C*H*-CCH$_3$), 5.72 (d, ^3J = 3.5 Hz, 1H, H-1), 4.68 (dd, ^3J = 4.0 Hz, ^3J = 3.5 Hz, 1H, H-2),, 4.40 – 4.31 (m, 1H, H_a-5), 4.22 – 4.14 (m, 2H, H-4, H_b-5), 2.84 (ddd, ^3J = 9.7 Hz, ^3J = 4.0 Hz, ^4J = 2.4 Hz, 1H, H-3), 2.44 (s, 3H, aryl-C*H*$_3$), 2.21 (d, ^4J = 2.4 Hz, 1H, CC*H*), 1.52 (s, 3H, C(C*H*$_3$)$_2$), 1.34 (s, 3H, C(C*H*$_3$)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 145.1 (aryl-*C*-SO$_3$), 132.7 (aryl-*C*-CH$_3$), 130.0 (aryl-*C*H-CCH$_3$), 128.2 (aryl-*C*H-CSO$_3$), 112.8 (*C*(CH$_3$)$_2$), 105.0 (C-1), 80.7 (C-2), 78.5 (C-4), 76.5 (*C*CH), 73.3 (C*C*H), 67.4 (C-5), 37.0 (C-3), 26.7 (C(*C*H$_3$)$_2$), 26.4 (C(*C*H$_3$)$_2$), 21.8 (aryl-*C*H$_3$) ppm. ESI-HRMS calcd. for [$C_{17}H_{20}O_6S$ + H]$^-$: 353.1054, found: 353.1057. ESI-HRMS calcd. for [$C_{17}H_{20}O_6S$ + NH_4]$^+$: 370.1319, found: 370.1317.

5-Azido-3,5-dideoxy-3-ethynyl-1,2-*O*-isopropylidene-α-D-ribofuranose (10): A mixture of tosyl compound **10d** (10.5 g, 29.8 mmol, 1.00 eq.) and sodium azide (8.86 g, 95.3 mmol, 3.20 eq.) was suspended in DMF (300 mL) and stirred under N_2 at 80 °C for 3 h. The yellow suspension was diluted with brine (200 mL) and extracted with EtOAc (4 x 300 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo.* Purification of the crude product by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1→4:1) yielded azide compound **10** (6.25 g, 28.0 mmol, 94%) as a colorless oil. R_f = 0.67 (isohexanes/EtOAc = 2:1). IR (ATR): ṽ = 3280, 2989, 2100, 1375, 1216, 1166, 1105, 1012, 871 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 5.86 (d, ^3J = 3.6 Hz, 1H, H-1), 4.75 (dd, ^3J = 4.1 Hz, ^3J = 3.6 Hz, 1H, H-2), 4.24 (ddd, ^3J = 10.2 Hz, ^3J = 3.9 Hz, ^3J = 2.7 Hz, 1H, H-4), 3.76 (dd, ^2J = 13.6 Hz, ^3J = 2.7 Hz, 1H, H_a-5), 3.37 (dd, ^2J = 13.6 Hz, ^3J = 3.9 Hz, 1H, H_b-5), 2.88 (ddd, ^3J = 10.2 Hz, ^3J = 4.1 Hz, ^4J = 2.5 Hz, 1H, H-3), 2.25 (d, ^4J = 2.5 Hz, 1H, CC*H*), 1.57 (s, 3H, C(C*H*$_3$)$_2$), 1.37 (s, 3H, C(C*H*$_3$)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 112.8 (*C*(CH$_3$)$_2$), 105.1 (C-1), 81.0 (C-2), 79.7 (C-4), 76.9 (*C*CH), 73.2 (C*C*H), 50.7 (C-5), 37.9 (C-3), 26.7 (C(*C*H$_3$)$_2$), 26.4 (C(*C*H$_3$)$_2$) ppm. EI-HRMS calcd. for [$C_{10}H_{13}N_3O_3$ - CH_3]$^+$: 208.0717, found: 208.0717.

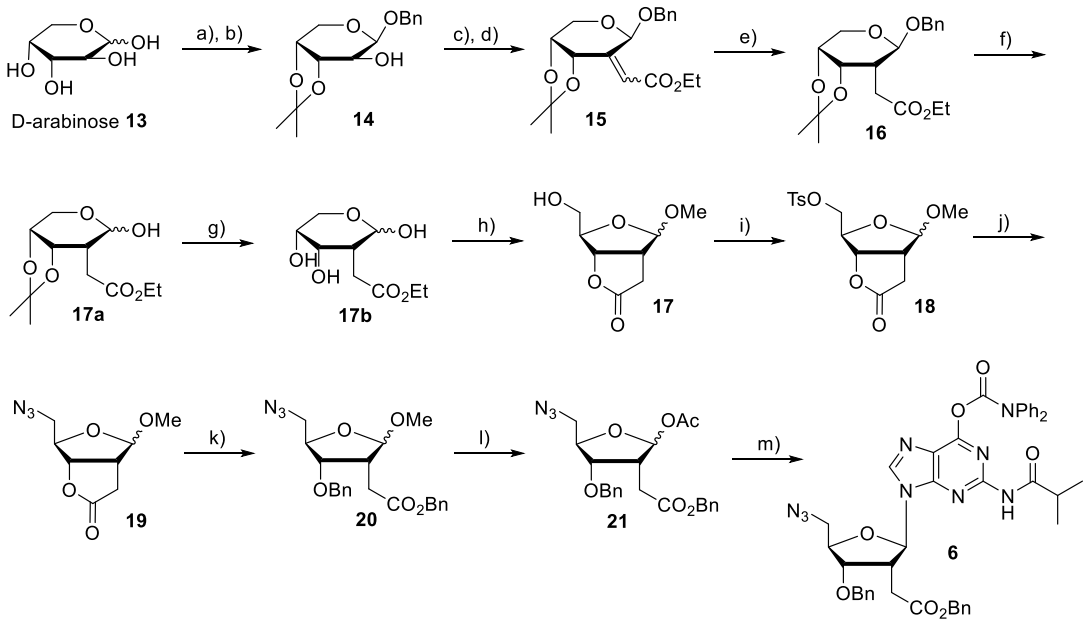
1,2-di-*O*-Acetyl-5-azido-3,5-dideoxy-3-ethynyl-D-ribofuranose (11): A stirred solution of azide compound **10** (6.00 g, 26.9 mmol, 1.00 eq.) in acetic acid (100 mL) and acetic anhydride (50 mL) was treated with concentrated sulfuric acid (96%, 1.30 mL) at 0 °C. The reaction mixture turned dark and was stirred for 5 h at RT. After careful quenching with saturated aqueous $NaHCO_3$ solution (200 mL) and solid $NaHCO_3$ until CO_2 evolution stopped, the reaction was extracted with DCM (4 x 200 mL). The organic phase was washed with brine (200 mL), dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo.* The crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1) to afford diacetate compound **11** (5.61 g, 21.0 mmol, 78%) as colorless crystals. α/β = 1:6. β anomer could be isolated for analysis. Crystallization from isohexanes/EtOAc (vapor diffusion) provided suitable single crystals of the β anomer for X-ray characterization. M.p. = 81 – 82 °C. R_f (α anomer) = 0.55 (isohexanes/EtOAc = 4:1). R_f (β anomer) = 0.46 (isohexanes/EtOAc, 4:1). IR (ATR, β anomer): ṽ = 3281, 2934, 2099, 1743, 1438, 1371, 1205, 1097, 1024, 959, cm^{-1}. Major β anomer: ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 6.12 (s, 1H, H-1), 5.37 (d, ^3J = 4.5 Hz, 1H, H-2), 4.38 (ddd, ^3J = 10.0 Hz, ^3J = 3.2 Hz, ^3J = 3.0 Hz, 1H, H-4), 3.75 (dd, ^2J = 13.7 Hz, ^3J = 3.0 Hz, 1H, H_a-5), 3.39 (ddd, ^3J = 10.0 Hz, ^3J = 4.5 Hz, ^4J = 2.5 Hz, 1H, H-3), 3.24 (dd, ^2J = 13.7 Hz, ^3J = 3.2 Hz, 1H, H_b-5), 2.16 (d, ^4J = 2.5 Hz, 1H, CC*H*), 2.14 (s, 3H, C-2-OCOC*H*$_3$), 2.08 (s, 3H, C-1-OCOC*H*$_3$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 169.5 (C-2-O*C*OCH$_3$), 169.2 (C-1-O*C*OCH$_3$), 98.7 (C-1), 83.7 (C-4), 76.8 (C-2), 76.0 (*C*CH), 73.6 (C*C*H), 51.0 (C-5), 34.5 (C-3), 21.12 (C-1-OCO*C*H$_3$), 20.73 (C-2-OCO*C*H$_3$) ppm. ESI-HRMS calcd. for [$C_{11}H_{13}N_3O_5$ + H]$^+$: 268.0928, found: 268.0930. ESI-HRMS calcd. for [$C_{11}H_{13}N_3O_5$ + NH_4]$^+$: 285.1193, found: 285.1196.

6-Benzoylamino-9-(2-*O*-acetyl-5-azido-3,5-dideoxy-3-ethynyl-β-D-ribofuranosyl)-9*H*-purine (12): *N*,*O*-Bis(trimethylsilyl)acetamid (BSA) (3.66 mL, 3.05 g, 15.0 mmol, 4.00 eq.) was added under N_2 to a stirred suspension of diacetate compound **11** (1.00 g, 3.74 mmol, 1.00 eq.) and 6-*N*-benzoyladenine (1.79 g, 7.48 mmol,

2.00 eq.) in dichloroethane (40 mL) and heated to 80 °C for 1 h until a clear solution was obtained. The reaction mixture was brought to RT and treated with trimethylsilyl triflate (TMSOTf) (1.36 mL, 1.66 g, 7.48 mmol, 2.00 eq.). The dark red solution was stirred at 80 °C for 4 h and additional 8 h at RT. The reaction was quenched with saturated aqueous $NaHCO_3$ (30 mL) and extracted with DCM (4 x 50 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. The residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 2:1→1:1→1:2→EtOAc) to provide nucleoside **12** (1.02 g, 2.29 mmol, 61%) as a colorless foam. The reaction could also be performed on a 4 g scale of the diacetate starting material **11** (yield: 50%). M.p. = 110 °C (decomp.). R_f = 0.20 (isohexanes/EtOAc = 1:2). R_f = 0.24 (DCM/MeOH, 100:5). IR (ATR): $\tilde{\nu}$ = 3296, 3060, 2932, 2103, 1747, 1698, 1581, 1218, 1072, 797 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 9.27 (s, 1H, N*H*), 8.74 (s, 1H, H-2), 8.16 (s, 1H, H-8), 7.98 (d, ^3J = 7.7 Hz, 2H, aryl-o-C*H*), 7.58 – 7.53 (m, 1H, aryl-p-C*H*), 7.46 (t, ^3J = 7.7 Hz, 2H, aryl-m-C*H*), 6.09 (s, 1H, H-1'), 5.87 (d, ^3J = 5.7 Hz, 1H, H-2'), 4.38 (ddd, ^3J = 10.0 Hz, ^3J = 4.8 Hz, ^4J = 2.7 Hz, 1H, H-4'), 4.07 (ddd, ^3J = 10.0 Hz, ^3J = 5.7 Hz, ^4J = 2.4 Hz, 1H, H-3'), 3.79 (dd, ^2J = 13.6 Hz, ^3J = 2.7 Hz, 1H, H_a-5'), 3.57 (dd, ^2J = 13.6 Hz, ^3J = 4.8 Hz, 1H, H_b-5'), 2.26 (d, ^4J = 2.4 Hz, 1H, CC*H*), 2.18 (s, 3H, CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 169.6 (O*C*OCH$_3$), 164.8 (N-*C*O-aryl), 152.8 (C-2), 151.2 (C-4), 149.9 (C-6), 142.0 (C-8), 133.6 (aryl-*C*-CO-N), 132.9 (aryl-p-*C*H), 128.9 (aryl-m-*C*H), 128.0 (aryl-o-*C*H), 123.7 (C-5), 89.7 (C-1'), 83.0 (C-4'), 76.9 (C-2'), 75.7 (*C*CH), 74.4 (C*C*H), 51.2 (C-5'), 36.2 (C-3'), 20.7 (OCO*C*H$_3$) ppm. ESI-HRMS calcd. for $[C_{21}H_{18}N_8O_4 + H]^+$: 447.1524, found: 447.1529.

6-Benzoylamino-9-(2-*O*-acetyl-5-amino-3,5-dideoxy-3-ethynyl-β-D-ribofuranosyl)-9*H*-purine (5a): Trimethylphosphine (4.48 mL, 4.48 mmol, 1.0M in THF, 2.00 eq.) was added to a stirred solution of nucleoside **12** (1.00 g, 2.24 mmol, 1.00 eq.) in THF (25 mL). After 5 min the reaction mixture turned turbid under N_2 evolution and was heated to 40 °C for 1.5 h. The reaction mixture was treated with water (0.44 mL, 24.6 mmol, 11.0 eq.) and stirred for 10 h at RT. Volatile materials were removed under reduced pressure and the residue was purified by flash-column chromatography (silica gel, DCM/MeOH, 100:2→100:5) to give amino compound **5a** (0.62 g, 1.48 mmol, 66%) as a colorless foam. The reaction could also be performed on a 4 g scale of the azide starting material **12** (yield: 56%). R_f = 0.34 (DCM/MeOH = 5:1). IR (ATR): $\tilde{\nu}$ = 3366, 3275, 2918, 1747, 1640, 1422, 1296, 1138, 943, 860 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 8.75 (s, 1H, H-2), 8.27 (s, 1H, H-8), 8.00 (d, ^3J = 7.7 Hz, 2H, aryl-o-C*H*), 7.61 – 7.55 (m, 1H, aryl-p-C*H*), 7.49 (t, ^3J = 7.7 Hz, 2H, aryl-m-C*H*), 6.08 (s, 1H, H-1'), 5.81 (d, ^3J = 5.9 Hz, 1H, H-2'), 4.27 (ddd, ^3J = 9.9 Hz, ^3J = 4.8 Hz, ^4J = 3.0 Hz, 1H, H-4'), 4.02 (ddd, ^3J = 9.9 Hz, ^3J = 5.9 Hz, ^4J = 2.4 Hz, 1H, H-3'), 3.24 (dd, ^2J = 14.0 Hz, ^3J = 3.0 Hz, 1H, H_a-5'), 3.01 (dd, ^2J = 14.0 Hz, ^3J = 4.8 Hz, 1H, H_b-5'), 2.23 (d, ^4J = 2.4 Hz, 1H, CC*H*), 2.19 (s, 3H, CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 169.8 (O*C*OCH$_3$), 164.9 (N-*C*O-aryl), 152.8 (C-2), 151.3 (C-4), 149.9 (C-6), 142.2 (C-8), 133.6 (aryl-*C*-CO-N), 132.9 (aryl-p-*C*H), 128.9 (aryl-m-*C*H), 128.0 (aryl-o-*C*H), 123.7 (C-5), 89.5 (C-1'), 85.4 (C-4'), 77.4 (C-2'), 76.7 (*C*CH), 73.9 (C*C*H), 42.6 (C-5'), 35.8 (C-3'), 20.9 (OCO*C*H$_3$) ppm. ESI-HRMS calcd. for $[C_{21}H_{20}N_6O_4 + H]^+$: 421.1619, found: 421.1623.

6-Benzoylamino-9-(2-*O*-acetyl-5-(*tert*-butoxycarbonyl)amino-3,5-dideoxy-3-ethynyl-β-D-ribofuranosyl)-9*H*-purine (5): A mixture of amine compound **5a** (2.00 g, 4.76 mmol, 1.00 eq.), triethylamine (1.99 mL, 1.44 g, 14.3 mmol, 3.00 eq.) and di-*tert*-butyldicarbonate (1.53 mL, 1.56 g, 7.14 mmol, 1.50 eq.) in dry DCM (40 mL) was stirred at RT for 16 h. MeOH (3 mL) was added and volatile materials were removed *in vacuo*. The residue was purified by flash-column chromatography (silica gel, DCM/MeOH, 100:2→100:3) to give the title compound **5** as a colorless foam (1.59 g, 3.05 mmol, 64%). M.p. = 143 °C (decomp.). R_f = 0.20 (DCM/MeOH = 100:5). IR (ATR): $\tilde{\nu}$ = 3265, 2977, 1749, 1699, 1610, 1516, 1455, 1248, 1227, 1086 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 9.15 (s, 1H, N*H*Bz), 8.79 (s, 1H, H-2), 8.03 (s, 1H, H-8), 8.01 (d, ^3J = 7.7 Hz, 2H, aryl-o-C*H*), 7.63 – 7.56 (m, 1H, aryl-p-C*H*), 7.49 (t, ^3J = 7.7 Hz, 2H, aryl-m-C*H*), 6.26 (dd, ^3J = 7.0 Hz, ^4J = 3.7 Hz, 1H, N*H*Boc), 5.99 (d, ^3J = 2.4 Hz, 1H, H-1'), 5.66 (dd, ^3J = 6.6 Hz, ^3J = 2.4 Hz, 1H, H-2'), 4.43 (ddd, ^3J = 9.4 Hz, ^3J = 3.7 Hz, ^3J = 3.2 Hz, 1H, H-4'), 4.03 (ddd, ^3J = 9.4 Hz, ^3J = 6.6 Hz, ^4J = 2.4 Hz, 1H, H-3'), 3.67 (ddd, ^2J = 14.5 Hz, ^3J = 7.0 Hz, ^3J = 3.7 Hz, 1H, H_a-5'), 3.57 (ddd, ^2J = 14.5 Hz, ^3J = 7.0 Hz, ^3J = 3.2 Hz, 1H, H_b-5'), 2.28 (d, ^4J = 2.4 Hz, 1H, CC*H*), 2.18 (s, 3H, OCOCH_3), 1.46 (s, 9H, C(CH_3)$_3$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 170.0 (O*C*OCH$_3$), 164.7 (N-*C*O-aryl), 156.5 (N-*C*O-O), 152.9 (C-2), 151.1 (C-4), 150.2 (C-6), 142.4 (C-8), 133.5 (aryl-*C*-CO-N), 133.0 (aryl-p-*C*H), 129.0 (aryl-m-*C*H), 128.1 (aryl-o-*C*H), 124.2 (C-5), 90.5 (C-1'), 83.5 (C-4'), 79.6 (*C*(CH$_3$)$_3$), 77.4 (C-2'), 76.4 (*C*CH), 74.4 (C*C*H), 41.6 (C-5'), 35.8 (C-3'), 28.6 (C(*C*H$_3$)$_3$), 20.8 (OCO*C*H$_3$) ppm. ESI-HRMS calcd. for $[C_{26}H_{28}N_6O_6 + H]^+$: 521.2143, found: 521.2150.



Synthesis of G-half **6**. a) AcCl, BnOH, 60°C, 5h, 80%; b) $Me_2C(OMe)_2$, Me_2CO, p-TsOH (cat.), 60°C, 2h, 84%; c) $(COCl)_2$, DMSO, NEt_3, DCM, -60°C, 3h; d) Ph_3PCHCO_2Et, DCM, RT, 12h, 86% (over two steps); e) H_2, Raney-Ni, EtOH, RT, 20h, 90%; f) H_2, Pd/C, EtOH/THF, 36h, 88%; g) 80% HOAc, RT, 24h; h) H_2SO_4 (cat.), MeOH, 4°C, 3d, 72% (over two steps); i) TsCl, Py, RT, 18h, 76%; i) NaN_3, DMF, 80°C, 3h, 75%; k) BnBr, KOH, THF, reflux, 5h, 91%; l) HOAc/Ac_2O, H_2SO4 (cat.), RT, 3h, 85%; m) 6-*O*-(Diphenylcarbamoyl)-2-*N*-isobutyrylguanine ($G^{dpc/iBu}$), BSA, TMSOTf, DCE, 80°C, 2h, 72%. Overall yield starting from **13**: 10%.

Benzyl 3,4-*O*-isopropylidene-β-D-arabinopyranoside (14):
The title compound was prepared according to a modified procedure of Shing et al.[4] Benzyl β-D-arabinopyranoside (48.0 g, 200 mmol, 1.00 eq.) was suspended in acetone (500 mL) and 2,2-dimethoxypropane (49.0 mL, 41.6 g; 400 mmol, 2.00 eq.). After addition of *p*-toluenesulfonic acid monohydrate (1.14 g, 5.99 mmol, 0.03 eq.), the reaction mixture was stirred at 60 °C for 2 h to obtain a clear solution. The reaction was neutralized by treatment with triethylamine (0.84 mL, 0,61 g, 5.99 mmol, 1.00 eq.). Volatile components were removed *in vacuo* and the residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 2:1→3:2→1:1, gradient elution) to furnish the title compound **14** (46.8 g, 167 mmol, 84%) as a colorless oil. R_f = 0.54 (isohexanes/EtOAc = 1:1). IR (ATR): $\tilde{\nu}$ = 3221, 2938, 1499, 1453, 1315, 1252, 1048, 1001, 848, 783, 701 cm^{-1}. ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.42 – 7.28 (m, 5H, aryl-H), 4.94 (d, ^3J = 3.6 Hz, 1H, H-1), 4.79 (d, ^2J = 11.7 Hz, 1H, PhCH_2O), 4.55 (d, ^2J = 11.7 Hz, 1H, PhCH_2O), 4.24 (ddd, ^3J = 6.2 Hz, ^3J = 2.6 Hz, ^3J = 1.2 Hz, 1H, H-4), 4.21 (q, ^3J = 6.2 Hz, 1H, H-3), 4.01 (dd, ^2J = 13.2 Hz, ^3J = 2.6 Hz, 1H, H_a-5), 3.93 (dd, ^2J = 13.2 Hz, ^3J = 1.2 Hz, 1H, H_b-5), 3.80 (dd, ^3J = 6.2 Hz, 3.6 Hz, 1H, H-2), 2.22 (br s, 1H, O*H*), 1.53 (s, 3H, CH_3), 1.36 (s, 3H, CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 137.1 (aryl-*C*-CH_2), 128.7 (aryl-m-*C*H), 128.22 (aryl-p-*C*H), 128.15 (aryl-o-*C*H), 109.4 (*C*$(CH_3)_2$), 97.0 (C-1), 76.1 (C-4), 73.1 (C-3), 70.1 (C-2), 69.9 (Ph*C*H_2O), 59.9 (C-5), 28.0 (*C*H_3), 26.1 (*C*H_3) ppm. ESI-HRMS calcd. for $[C_{15}H_{20}O_5 + NH_4]^+$: 298.1649, found: 298.1651. EI-HRMS calcd. for $[C_{15}H_{20}O_5 - CH_3]^+$: 265.1071, found: 265.1084.

Benzyl 2-deoxy-2-*C*-[(ethoxycarbonyl)methylene]-3,4-*O*-isopropylidene-β-D-arabinofuranoside (15):
The title compound was prepared according to a modified procedure of Kaiya et al.[5] Oxalyl chloride (15.3 mL, 22.4 g, 176 mmol, 1.15 eq.) was dissolved in dry DCM (600 mL). After cooling to −78 °C, dry DMSO (25.1 mL, 27.6 g, 352 mmol, 2.30 eq.) was added dropwise and the mixture was stirred at −60 °C for 1 h until no further gas development was observed. Subsequently, a solution of acetonide compound **14** (43.0 g, 153 mmol, 1.00 eq.) in dry DCM (150 mL) was added slowly over 10 min and the mixture was stirred at -60 °C for 2 h. The reaction mixture was treated with triethylamine (64.1 mL, 46.6 g, 460 mmol, 3.00 eq.), stirred at -60 °C for 1 h, quenched upon addition of saturated aqueous $NaHCO_3$ (300 mL) and extracted with DCM (3 x 300 mL). The combined organic layers were washed with brine, dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure. Ketone **15a** was obtained as a waxy syrup which was used in the next step without further purification. EI-HRMS calcd. for $[C_{15}H_{18}O_5 - CH_3]^+$: 263.0914, found: 263.0914.
A mixture of crude ketone **15a** (42.0 g, 151 mmol, 1.00 eq.) and (carbethoxymethylene)triphenylphosphorane[6] (68.3 g, 196 mmol, 1.30 eq.) in DCM (400 mL) was stirred at RT for 12 h. Volatile components were evaporated and the residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 4:1) to afford the title compound **15** as a colorless oil (46.0 g, 132 mmol, 86% over 2 steps). E/Z = 4:1 (inseparable mixture by fcc). R_f = 0.84 (isohexanes/EtOAc = 2:1). IR (ATR, E/Z-mixture): $\tilde{\nu}$ = 2983, 1718, 1372, 1214, 1150, 1020, 853, 736, 699 cm^{-1}. Major E-Isomer: ^1H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.41 – 7.32 (m, 5H, aryl-H), 6.41 (d, ^4J = 1.8 Hz, 1H, C=C*H*), 6.06 (d, ^3J = 7.5 Hz, 1H, H-3), 5.44 (d, ^4J = 1.8 Hz, 1H, H-1), 4.86 (d, ^2J = 11.9 Hz, 1H, OCH_2Ph), 4.62 (d,

^{2}J = 11.9 Hz, 1H, OCH_2Ph), 4.34 (dd, ^{3}J = 7.5 Hz, ^{3}J = 1.7 Hz, 1H, H-4), 4.20 (q, ^{3}J = 7.1 Hz, 2H, OCH_2CH$_3$), 3.70 (d, ^{3}J = 1.7 Hz, 2H, H-5), 1.53 (s, 3H, C(CH_3)$_2$), 1.40 (s, 3H, C(CH_3)$_2$), 1.30 (t, ^{3}J = 7.1 Hz, 3H, OCH$_2$CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, CDCl$_3$): δ = 165.5 (*C*=O), 147.9 (C-2), 137.6 (aryl-*C*-CH$_2$), 128.6 (aryl-m-*C*H), 128.1 (aryl-p-*C*H), 128.0 (aryl-o-*C*H), 124.4 (C=*C*H), 110.6 (*C*(CH$_3$)$_2$), 95.8 (C-1), 75.2 (C-4), 69.5 (Ph*C*H$_2$O), 68. 6 (C-3), 63.2 (C-5), 60.8 (O*C*H$_2$CH$_3$), 26.4 (C(*C*H$_3$)$_2$), 25.3 (C(*C*H$_3$)$_2$), 14.3 (OCH$_2$*C*H$_3$) ppm. ESI-HRMS calcd. for [C$_{19}$H$_{24}$O$_6$ + NH$_4$]$^+$: 366.1911, found: 366.1911.

Benzyl 2-deoxy-2-*C*-[(ethoxycarbonyl)methyl]-3,4-*O*-isopropylidene-β-D-ribopyranoside (16):
The title compound was prepared according to a modified procedure of Kaiya et al.[5] Vinyl compound **15** (45.0 g, 129 mmol, 1.00 eq.) was dissolved in EtOH (300 mL) and Raney-Ni (ca. 15 mL) was added to the solution at RT. The reaction vessel was evacuated and flushed with hydrogen three times. Subsequently, the mixture was stirred under hydrogen atmosphere for 20 h. Upon completion of the reaction as monitored by TLC, the reaction mixture was filtered through celite. Volatile materials were removed *in vacuo* and the residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 4:1) to yield reduced compound **16** (40.8 g, 116 mmol, 90%) as a colorless oil. dr = 13:1. R$_f$ = 0.33 (isohexanes/EtOAc = 4:1). IR (ATR): ṽ = 2983, 1732, 1455, 1370, 1212, 1071, 1021, 870, 738, 699 cm^{-1}. ^{1}H NMR, COSY (400 MHz, CDCl$_3$): δ = 7.38 – 7.25 (m, 5H, aryl-H), 4.82 (d, ^{3}J = 11.7 Hz, 1H, OCH_2Ph), 4.64 (d, ^{3}J = 8.0 Hz, 1H, H-1), 4.49 (dd, ^{3}J = 7.4 Hz, ^{3}J = 2.8 Hz, 1H, H-3), 4.48 (d, ^{3}J = 11.7 Hz, 1H, OCH_2Ph), 4.24 (ddd, ^{3}J = 7.4 Hz, ^{3}J = 2.6 Hz, ^{3}J = 2.3 Hz, 1H, H-4), 4.10 (q, ^{3}J = 7.2 Hz, 2H, OCH_2CH$_3$), 3.84 (dd, ^{2}J = 12.8 Hz, ^{3}J = 2.6 Hz, 1H, H$_a$-5), 3.63 (dd, ^{2}J = 12.8 Hz, ^{3}J = 2.3 Hz, 1H, H$_b$-5), 2.60 – 2.55 (m, 2H, CH_2COO), 2.28 (ddd, ^{3}J = 8.0 Hz, ^{3}J = 6.5 Hz, ^{3}J = 2.8 Hz, 1H, H-2), 1.47 (s, 3H, C(CH_3)$_2$), 1.31 (s, 3H, C(CH_3)$_2$), 1.22 (t, ^{3}J = 7.2 Hz, 3H, OCH$_2$CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, CDCl$_3$): δ = 172.2 (*C*=O), 138.1 (aryl-*C*-CH$_2$), 128.5 (aryl-m-*C*H), 128.1 (aryl-p-*C*H), 127.8 (aryl-o-CH), 109.1 (*C*(CH$_3$)$_2$), 99.0 (C-1), 73.2 (C-4), 72.6 (C-3), 69.7 (Ph*C*H$_2$O), 62.6 (C-5), 60.7 (O*C*H$_2$CH$_3$), 37.2 (C-2), 33.6 (*C*H$_2$COO), 26.7 (C(*C*H$_3$)$_2$), 25.0 (C(*C*H$_3$)$_2$), 14.3 (OCH$_2$*C*H$_3$) ppm. EI-HRMS calcd. for [C$_{19}$H$_{26}$O$_6$ - CH$_3$]$^+$:335.1489, found: 335.1481.

2-Deoxy-3,4-*O*-isopropylidene-2-*C*-[(ethoxycarbonyl)methyl]-D-ribopyranose (17a): To a stirred solution of ester compound **16** (38.0 g, 108 mmol, 1.00 eq.) in EtOH (250 mL) and THF (100 mL) was added Pd/C (10 wt.%, 1.70 g) under N$_2$ at RT. The reaction vessel was evacuated and flushed with hydrogen three times. The mixture was stirred under hydrogen atmosphere for 24 h and then filtered through celite. The solution was concentrated to dryness under reduced pressure and the residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 4:1→1:1) to furnish anomeric alcohol **17a** as a colorless oil (24.8 g, 95.3 mmol, 88%). α/β = 1:9 (inseparable mixture by fcc). R$_f$ = 0.36 (isohexanes/EtOAc = 1:1). IR (ATR, α/β-mixture): ṽ = 2984, 1731, 1458, 1371, 1213, 1109, 1056, 1020, 868 cm^{-1}. Major β anomer: ^{1}H NMR, COSY (400 MHz, CDCl3): δ = 4.88 (dd, ^{3}J = 8.3 Hz, ^{3}J = 4.8 Hz, 1H, H-1), 4.44 (dd, ^{3}J = 7.1 Hz, ^{3}J = 2.9 Hz, 1H, H-3), 4.23 (ddd, ^{3}J = 7.1 Hz, ^{3}J = 3.7 Hz, ^{3}J = 3.4 Hz, 1H, H-4), 4.16 (q, ^{3}J = 7.1 Hz, 2H, OCH_2CH$_3$), 3.88 (dd, ^{2}J = 12.6 Hz, ^{3}J = 3.4 Hz, 1H, H$_a$-5), 3.60 (dd, ^{2}J = 12.6, ^{3}J = 3.7 Hz, 1H, H$_b$-5), 3.26 (d, ^{3}J = 4.8 Hz, 1H, O*H*), 2.66 (dd, ^{2}J = 16.9 Hz, ^{3}J = 5.9 Hz, 1H, CH_2COO), 2.59 (dd, ^{3}J = 16.9 Hz, ^{2}J = 8.3 Hz, 1H, CH_2COO), 2.20 (ddd, ^{3}J = 8.3 Hz, ^{3}J = 5.9 Hz, ^{3}J = 2.9 Hz, 1H, H-2), 1.45 (s, 3H, (CH_3)$_2$), 1.30 (s, 3H, C(CH_3)$_2$), 1.26 (t, *J* = 7.1 Hz, 3H, OCH$_2$CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, CDCl$_3$): δ = 172.8 (*C*=O), 109.2 (*C*(CH$_3$)$_2$), 94.4 (C-1), 73.4 (C-3), 72.7 (C-4), 63.3 (C-5), 61.0 (O*C*H$_2$CH$_3$), 38.5 (C-2), 33.5 (*C*H$_2$COO), 27.0 (C(*C*H$_3$)$_2$), 25.27 (C(*C*H$_3$)$_2$), 14.3 (OCH$_2$*C*H$_3$) ppm. EI-HRMS calcd. for [C$_{12}$H$_{20}$O$_6$ - CH$_3$]$^+$: 245.1020, found: 245.1023.

Methyl 2-*C*-carboxymethyl-2-deoxy-2,3-lactone-D-ribofuranoside (17): The title compound was prepared according to a modified procedure of Li et al.[7] Anomeric alcohol **17a** (20.7 g, 79.5 mmol, 1.00 eq.) was dissolved in AcOH/H$_2$O (v/v 4:1, 400 mL) and stirred at RT for 24 h. Tthe mixture was heated to 40 °C for additional 2 h. Volatile materials were evaporated and the crude product **17b** was co-evaporated with toluene (3 x 300 mL) and used in the next step without further purification. R$_f$ = 0.05 (isohexanes/EtOAc = 1:1). EI-HRMS calcd. for [C$_{12}$H$_{20}$O$_6$ - 2 x H$_2$O]$^+$: 184.0730, found: 184.0726.
Concentrated sulfuric acid (0.56 mL) was added to a stirred solution of triol compound **17b** in dry methanol (450 mL) at 0 °C. The reaction mixture was stirred at 4 °C for 72 h and then neutralized by addition of solid sodium bicarbonate. The resulting suspension was filtered through celite and the filtrate was concentrated to dryness under reduced pressure. Purification by flash-column chromatography (silica gel, isohexanes/EtOAc, 2:1→1:1→1:3) yielded lactone **17** (10.8 g, 57.4 mmol, 72% over 2 steps) as colorless crystals. α/β = 2:3. β anomer could be isolated for analysis. M.p. = 45 – 47 °C. R$_f$ (α anomer) = 0.44 (DCM/MeOH = 100:5). R$_f$ (β anomer) = 0.30 (DCM/MeOH = 100:5). IR (ATR, β anomer): ṽ = 3442, 2940, 1775, 1172, 1102, 1031, 1003, 932 cm^{-1}. Major β anomer: ^{1}H NMR, COSY (400 MHz, CDCl$_3$): δ = 5.16 (dd, ^{3}J = 7.0 Hz, ^{3}J = 0.9 Hz, 1H, H-3), 4.90 (d, ^{3}J = 1.4 Hz, 1H, H-1), 4.49 – 4.44 (m, 1H, H-4), 3.73 (dd, ^{2}J = 12.8 Hz, ^{3}J = 2.9 Hz, 1H, H$_a$-5), 3.68 (dd, ^{2}J = 12.8, ^{3}J = 3.6 Hz, 1H, H$_b$-5), 3.41 (s, 3H, OCH_3), 3.13 – 3.05 (m, 1H, H-2), 2.87 (dd, ^{2}J = 18.6 Hz, ^{3}J = 11.0 Hz,1H, CH_2COO), 2.55 (dd, ^{2}J = 18.6 Hz, ^{3}J = 3.7 Hz, 1H, CH_2COO) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, CDCl$_3$): δ = 175.6 (C=O), 111.8 (C-1), 86.9 (C-4), 84.8 (C-3), 63.7 (C-5), 55.8 (OCH$_3$), 46.6 (C-2), 32.4 (*C*H$_2$COO) ppm. EI-HRMS calcd. for [C$_8$H$_{12}$O$_5$ - CH$_2$OH]$^+$: 157.0495, found: 157.0494.

Methyl 2-*C*-carboxymethyl-2,5-dideoxy-2,3-lactone-5-tosyl-D-ribofuranoside (18): To a stirred solution of lactone compound **17** (10.0 g, 53.1 mmol, 1.00 eq.) in dry pyridine (300 mL) was added a solution of *p*-toluenesulfonyl chloride (13.2 g, 69.1 mmol, 1.30 eq.) in pyridine (40 mL) at 0 °C. After stirring at RT for 18 h, the reaction was quenched by treatment with MeOH (20 mL). Solvents were removed *in vacuo* and the residue was purified by flash-column chromatography (silica gel, hexane/EtOAc, 3:2→1:1→1:3) to give tosyl compound **18** as colorless crystals (13.9 g, 40.6 mmol, 76%). α/β = 2:3 (inseparable mixture by fcc). Crystallization from isohexanes/EtOAc (vapor diffusion) provided suitable β single crystals for X-ray characterization. M.p. = 78 – 80 °C. R$_f$ = 0.36 (isohexanes/EtOAc = 1:1). IR (ATR, α/β-mixture): ṽ = 2938, 1781, 1598, 1358, 1173, 1111, 979, 815, 665 cm^{-1}. β anomer: ^{1}H NMR, COSY (400 MHz, CDCl$_3$): δ = 7.82 – 7.75 (m, 2H, aryl H-2-2'), 7.40 – 7.32 (m, 2H, aryl

H-3,-3'), 4.98 (dd, ^3J = 7.0 Hz, ^3J = 0.9 Hz, 1H, H-3), 4.86 (d, ^3J = 1.1 Hz, 1H, H-1), 4.41 – 4.34 (m, 1H, H-4), 4.10 (dd, ^2J = 10.3 Hz, ^3J = 7.6 Hz, 1H, H_a-5), 4.06 (dd, ^2J = 10.3 Hz, ^3J = 6.4 Hz, 1H, H_b-5), 3.24 (s, 3H, OCH_3), 3.13 – 3.05 (m, 1H, H-2), 2.82 (dd, ^3J = 18.6 Hz, ^3J = 11.0 Hz, 1H, CH_2COO), 2.56 – 2.47 (m, 1H, CH_2COO), 2.45 (s, 3H, aryl CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 175.3 (C=O), 145.5 (aryl C-4), 132.5 (aryl C-1), 130.13 (aryl C-3,-3'), 128.02 (aryl C-2,-2'), 111.6 (C-1), 83.9 (C-3), 82.2 (C-4), 68.3 (C-5), 55.5 (OCH_3), 45.5 (C-2), 31.6 (CH_2COO), 21.8 (aryl CH_3) ppm. α anomer: ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.82 – 7.75 (m, 2H, aryl H-2-2'), 7.40 – 7.32 (m, 2H, aryl H-3-3'), 4.99 (d, ^3J = 3.8 Hz, 1H, H-1), 4.82 (dd, ^3J = 7.6 Hz, ^3J = 2.7 Hz, 1H, H-3), 4.32 – 4.28 (m, 1H, H-4), 4.23 (dd, ^2J = 11.0 Hz, ^3J = 3.1 Hz, 1H, H_a-5), 4.20 (dd, ^2J = 11.0 Hz, ^3J = 3.4 Hz, 1H, H_b-5), 3.31 (s, 3H, OCH_3), 3.05 – 2.98 (m, 1H, H-2), 2.67 (dd, ^3J = 17.7 Hz, ^3J = 1.5 Hz, 1H, CH_2COO), 2.56 – 2.47 (m, 1H, CH_2COO), 2.45 (s, 3H, aryl CH_3) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 176.2 (C=O), 145.4 (aryl C-4), 132.4 (aryl C-1), 130.11 (aryl C-3,-3'), 128.03 (aryl C-2,-2'), 104.5 (C-1), 83.1 (C-3), 80.2 (C-4), 69.0 (C-5), 55.4 (OCH_3), 44.1 (C-2), 29.0 (CH_2COO), 21.8 (aryl CH_3) ppm. ESI-HRMS calcd. for $[C_{12}H_{20}O_6 + NH_4]^+$: 360.1111, found: 360.1109.

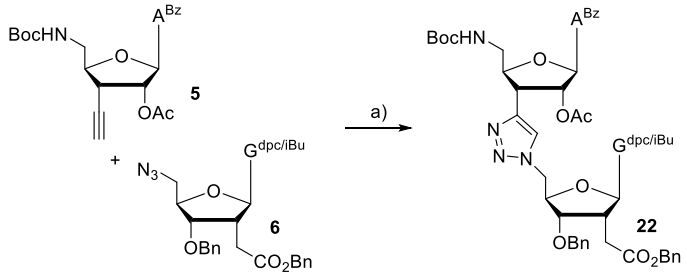
Methyl 5-azido-2-*C*-carboxymethyl-2,5-dideoxy-2,3-lactone-D-ribofuranoside (19): A mixture of tosyl compound **18** (13.0 g, 38.0 mmol, 1.00 eq.) and sodium azide (14.1 g, 152 mmol, 4.00 eq.) was suspended in DMF (300 mL) and stirred under N_2 at 80 °C for 3 h. The yellow suspension was diluted with brine (200 mL) and extracted with EtOAc (4x300 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated under reduced pressure. The residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 4:1→2:1→1:1) to afford azide compound **19** (6.08 g, 28.5 mmol, 75%) as a colorless oil. α/β = 2:3 (inseparable mixture by fcc). R_f = 0.46 (isohexanes/EtOAc = 1:1). IR (ATR, α/β-mixture): $\tilde{\nu}$ = 2936, 2100, 1776, 1444, 1282, 1160, 1110, 1031, 920 cm^{-1}. β anomer: ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 4.95 (dd, ^3J = 7.3 Hz, ^3J = 1.3 Hz, 1H, H-3), 4.92 (d, ^3J = 1.3 Hz, 1H, H-1), 4.41 – 4.36 (m, 1H, H-4), 3.53 (dd, ^2J = 12.7 Hz, ^3J = 7.2 Hz, 1H, H_a-5), 3.39 (s, 3H, OCH_3), 3.38 (m, 1H, H_b-5), 3.18 – 3.11 (m, 1H, H-2), 2.86 (dd, ^2J = 18.6 Hz, ^3J = 10.9 Hz, 1H, CH_2COO), 2.56 (dd, ^2J = 18.6 Hz, ^3J = 4.7 Hz, 1H, CH_2COO) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 175.3 (C=O), 111.9 (C-1), 84.7 (C-3), 84.0 (C-4), 55.7 (OCH_3), 53.1 (C-5), 45.8 (C-2), 31.8 (CH_2COO) ppm. α anomer: ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 5.11 (d, ^3J = 5.2 Hz, 1H, H-1), 4.81 (dd, ^3J = 7.8 Hz, ^3J = 2.9 Hz, 1H, H-3), 4.37 – 4.33 (m, 1H, H-4), 3.68 (dd, ^2J = 13.2 Hz, ^3J = 3.4 Hz, 1H, H_a-5), 3.43 (dd, ^2J = 13.2 Hz, ^3J = 3.8 Hz, 1H, H_b-5), 3.38 (s, 3H, OCH_3), 3.12 – 3.05 (m, 1H, H-2), 2.73 (dd, ^2J = 17.7 Hz, ^3J = 1.6 Hz, 1H, CH_2COO), 2.54 (dd, ^2J = 17.7 Hz, ^3J = 9.1 Hz, 1H, CH_2COO) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 176.3 (C=O), 104.4 (C-1), 83.8 (C-3), 81.6 (C-4), 55.3 (OCH_3), 52.1 (C-5), 44.4 (C-2), 29.0 (CH_2COO) ppm. ESI-HRMS calcd. for $[C_8H_{11}N_3O_4 + NH_4]^+$: 231.1088, found: 231.1088.

Methyl 5-azido-3-*O*-benzyl-2,5-dideoxy-2-*C*-[(benzyloxycarbonyl)methylene]-D-ribofuranoside (20): The title compound was prepared according to a modified procedure of Webber et al.[8] Azide compound **19** (5.05 g, 23.7 mmol, 1.00 eq.) was mixed with KOH (10.6g, 190 mmol, 8.00 eq.) in THF (250 mL). The stirred suspension was treated with benzyl bromide (28.1 mL, 40.5 g, 237 mmol, 10.0 eq.) and refluxed for 5 h. After cooling to 0 °C, the reaction was diluted with water (250 mL) and extracted with EtOAc (3 x 300 mL). The combined organic layers were washed with brine (500 mL), dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. The crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1→4:1→2:1) to furnish benzylated compound **20** as a colorless oil (8.89 g, 21.6 mmol, 91%). α/β = 2:3. β anomer could be isolated for analysis. R_f (α anomer) = 0.57 (isohexanes/EtOAc = 4:1). R_f (β anomer) = 0.48 (isohexanes/EtOAc = 4:1). IR (ATR, β anomer): $\tilde{\nu}$ = 2931, 2100, 1733, 1455, 1282, 1168, 1057, 910, 738, 698 cm^{-1}. β anomer: 1H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.40 – 7.26 (m, 10H, aryl H), 5.11 (d, ^2J = 12.3 Hz, 1H, COOCH_2Ph), 5.06 (d, ^2J = 12.3 Hz, 1H, COOCH_2Ph), 4.82 (d, ^3J = 2.3 Hz, 1H, H-1), 4.43 (s, 2H, OCH_2Ph), 4.16 – 4.11 (m, 1H, H-3), 4.14 – 4.08 (m, 1H, H-4), 3.37 (s, 3H, OCH_3), 3.32 (dd, ^2J = 12.7 Hz, ^3J = 6.2 Hz, 1H, H_a-5), 3.25 (dd, ^2J = 12.7 Hz, ^3J = 4.3 Hz, 1H, H_b-5), 2.86 – 2.79 (m, 1H, H-2), 2.74 (dd, ^2J = 16.5 Hz, ^3J = 7.6 Hz, 1H, CH_2COO), 2.43 (dd, ^2J = 16.5 Hz, ^3J = 7.4 Hz, 1H, CH_2COO) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 172.2 (C=O), 137.5 (OCH_2*Ph*-C-1), 135.9 (COOCH_2*Ph*-C-1), 128.7, 128.6, 128.41, 128.38, 128.08, 127.8 (aryl 10C), 109.4 (C-1), 81.4 (C-4), 80.3 (C-3), 72.6 (OCH_2Ph), 66.6 (COOCH_2Ph), 55.8 (OCH_3), 54.3 (C-5), 44.2 (C-2), 30.5 (CH_2COO) ppm. ESI-HRMS calcd. for $[C_{22}H_{25}N_3O_5 + NH_4]^+$: 429.2132, found: 429.2138.

Acetyl 5-azido-3-*O*-benzyl-2,5-dideoxy-2-*C*-[(benzyloxycarbonyl)methylene]-D-ribofuranoside (21): To a solution of benzylated compound **20** (8.02 g, 19.5 mmol, 1.00 eq.) in AcOH (70 mL) and Ac_2O (70 mL), was added concentrated H_2SO_4 (0.20 mL) at 0 °C. The solution was warmed to RT and stirred for 3 h. After careful quenching with saturated $NaHCO_3$ solution (150 mL) and solid $NaHCO_3$ until CO_2 evolution stopped, the reaction was extracted with DCM (4 x 200 mL), washed with brine (400 mL), dried over anhydrous $MgSO_4$ and filtered. Volatile materials were removed *in vacuo* and the residue was co-evaporated with toluene (2 x 100 mL). The crude product was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 9:1→4:1→2:1) to obtain acetylated compound **21** as a colorless oil (7.27 g, 16.5 mmol, 85%). α/β = 3:2 (inseparable mixture by fcc). R_f = 0.25 (isohexanes/EtOAc = 4:1). IR (ATR): $\tilde{\nu}$ = 2101, 1733, 1455, 1366, 1230, 1170, 1007, 899, 738, 698 cm^{-1}. β anomer: ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.26 (s, 10H, aryl H), 6.09 (d, J = 2.4 Hz, 1H, H-1), 5.09 (d, ^2J = 12.2 Hz, 1H, COOCH_2Ph), 5.04 (d, ^2J = 12.2 Hz, 1H, COOCH_2Ph), 4.45 (s, 2H, OCH_2Ph), 4.28 – 4.24 (m, 1H, H-3), 4.16 (ddd, ^3J = 9.6 Hz, ^3J = 5.0 Hz, ^3J = 4.6 Hz, 1H, H-4), 3.43 (dd, ^2J = 13.2 Hz, ^3J = 4.6 Hz, 1H, H_a-5), 3.23 (dd, ^2J = 13.2 Hz, ^3J = 5.0 Hz, 1H, H_b-5), 3.00 – 2.92 (m, 1H, H-2), 2.78 (dd, ^2J = 16.9 Hz, ^3J = 8.4 Hz, 1H, CH_2COO), 2.61 (dd, ^2J = 16.9 Hz, ^3J = 7.1 Hz, 1H, CH_2COO), 2.07 (s, 3H, CH_3COO) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 171.8 (*C*OOBn), 170.1 (*C*OOCH$_3$), 137.3 (OCH_2*Ph*-C-1), 135.7 (COOCH_2*Ph*-C-1), 128.71, 128.64, 128.51, 128.47, 128.20, 127.84 (aryl 10C), 101.3 (C-1), 82.2 (C-4), 78.9 (C-3), 72.8 (OCH_2Ph), 66.8 (COOCH_2Ph), 52.5 (C-5), 43.7 (C-2), 30.3 (CH_2COO), 21.3 (CH_3COO) ppm. α anomer: ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.41 – 7.22 (m, 10H, aryl H), 6.36 (d, ^3J = 4.8 Hz, 1H, H-1), 5.10 (s, 2H, COOCH_2Ph), 4.46 (d, ^2J = 12.0 Hz, 1H, OCH_2Ph), 4.42 (d, ^2J = 12.0 Hz, 1H, OCH_2Ph), 4.32 – 4.27 (m, 1H, H-4), 4.01 (dd, ^3J = 7.0, ^3J = 2.3 Hz, 1H, H-3),

3.35 (dd, ^2J = 12.9 Hz, ^3J = 5.1 Hz, 1H, H_a-5), 3.15 (dd, ^2J = 12.9 Hz, ^3J = 4.2 Hz, 1H, H_b-5), 2.89 – 2.82 (m, 1H, H-2), 2.77 (dd, ^2J = 16.7 Hz, ^3J = 6.6 Hz, 1H, CH_2COO), 2.61 (dd, ^2J = 16.7 Hz, ^3J = 5.8 Hz, 1H, CH_2COO), 2.05 (s, 3H, CH_3COO) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 172.0 (*C*OOBn), 170.5 (CH$_3$*C*OO), 137.8 (OCH$_2$*Ph*-C-1), 135.8 (COOCH$_2$*Ph*-C-1), 128.74, 128.59, 128.52, 128.51, 128.05, 127.80 (aryl 10C), 98.2 (C-1), 84.4 (C-4), 79.2 (C-3), 72.6 (O*C*H$_2$Ph), 66.7 (COO*C*H$_2$Ph), 52.7 (C-5), 43.1 (C-2), 28.5 (*C*H$_2$COO) 21.3 (*C*H$_3$COO) ppm. ESI-HRMS calcd. for $[C_{23}H_{25}N_3O_6 + NH_4]^+$: 457.2081, found: 457.2082.

9-{5-Azido-3-*O*-benzyl-2,5-dideoxy-2-*C*-[(benzyloxycarbonyl)methylene]-β-D-ribofuranosyl}-6-*O*-(diphenylcarbamoyl)-2-*N*-isobutyrylguanine (6): *N*,*O*-Bis(trimethylsilyl)acetamide (BSA) (2.23 mL, 1.85 g, 9.10 mmol, 4.00 eq.) was added under N_2 to a stirred suspension of compound **21** (1.00 g, 2.28 mmol, 1.00 eq.) and 6-*O*-(diphenylcarbamoyl)-2-*N*-isobutyrylguanine[9,10] (1.90 g, 4.55 mmol, 2.00 eq.) in dichloroethane (30 mL) and heated to 80 °C for 30 min until a clear solution was obtained. The reaction mixture was brought to RT and treated with trimethylsilyl triflate (TMSOTf) (1.07 mL, 1.32 g, 5.92 mmol, 2.60 eq.). The dark red solution was stirred at 80 °C for 2 h. The reaction was quenched with saturated aqueous $NaHCO_3$ (30 mL) at RT and extracted with DCM (4 x 50 mL). The combined organic layers were dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. The residue was purified by flash-column chromatography (silica gel, isohexanes/EtOAc, 4:1→2:1→1:1) to give nucleoside **6** (1.31 g, 1.65 mmol, 72%) as a colorless foam. The reaction could also be performed on a 5 g scale of the starting material **21** (yield: 59%). R_f = 0.68 (isohexanes/EtOAc = 1:1). IR (ATR): ṽ = 3321, 2933, 2102, 1731, 1584, 1492, 1268, 1166, 1047, 694 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 7.99 (s, 1H, H-8), 7.92 (s, 1H, N*H*), 7.47 – 7.18 (m, 20H, aryl-H), 6.00 (d, ^3J = 8.2 Hz, 1H, H-1'), 4.94 (d, ^2J = 12.3 Hz, 1H, COOCH_2Ph), 4.88 (d, ^2J = 12.3 Hz, 1H, COOCH_2Ph), 4.58 (d, ^2J = 11.6 Hz, 1H, OCH_2Ph), 4.443 (d, ^2J = 11.6 Hz, 1H, OCH_2Ph), 4.441 (d, ^3J = 5.8 Hz, ^3J = 2.1 Hz, 1H, H-3'), 4.27 – 4.21 (m, 1H, H-4'), 3.80 (dd, ^2J = 13.0 Hz, ^3J = 6.3 Hz, 1H, H_a-5), 3.74 – 3.63 (m, 1H, H-2'), 3.70 (dd, ^2J = 13.0 Hz, ^3J = 5.3 Hz, 1H, H_b-5), 2.85 (dd, ^2J = 16.7 Hz, ^3J = 8.3 Hz, 1H, CH_2COO), 2.80 – 2.70 (m, 1H, C*H*(CH$_3$)$_2$), 2.48 (dd, ^2J = 16.7 Hz, ^3J = 6.5 Hz, 1H, CH_2COO), 1.26 (s, 3H, CH(CH_3)$_2$), 1.24 (s, 3H, CH(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 174.7 (*C*ONH), 171.3 (*C*OOBn), 156.3 (C-2), 154.6 (C-4), 151.9 (O*C*ONPh$_2$), 150.4 (C-6), 143.4 (C-8), 141.8 (OCON-*Ph*-C1), 137.4 (OCH$_2$*Ph*-C-1), 135.5 (COOCH$_2$*Ph*-C-1), 129.3, 128.70, 128.67, 128.46, 128.41, 128.28, 128.17 (aryl 20C), 122.0 (C-5), 89.5 (C-1'), 82.8 (C-4'), 80.1 (C-3'), 72.4 (O*C*H$_2$Ph), 66.8 (COO*C*H$_2$Ph), 52.6 (C-5'), 42.7 (C-2'), 36.5 (*C*H(CH$_3$)$_2$), 30.2 (*C*H$_2$COO), 19.5 (CH(*C*H$_3$)$_2$), 19.4 (CH(*C*H$_3$)$_2$) ppm. ESI-HRMS calcd. for $[C_{43}H_{41}N_9O_7 + H]^+$: 796.3202, found: 796.3214.



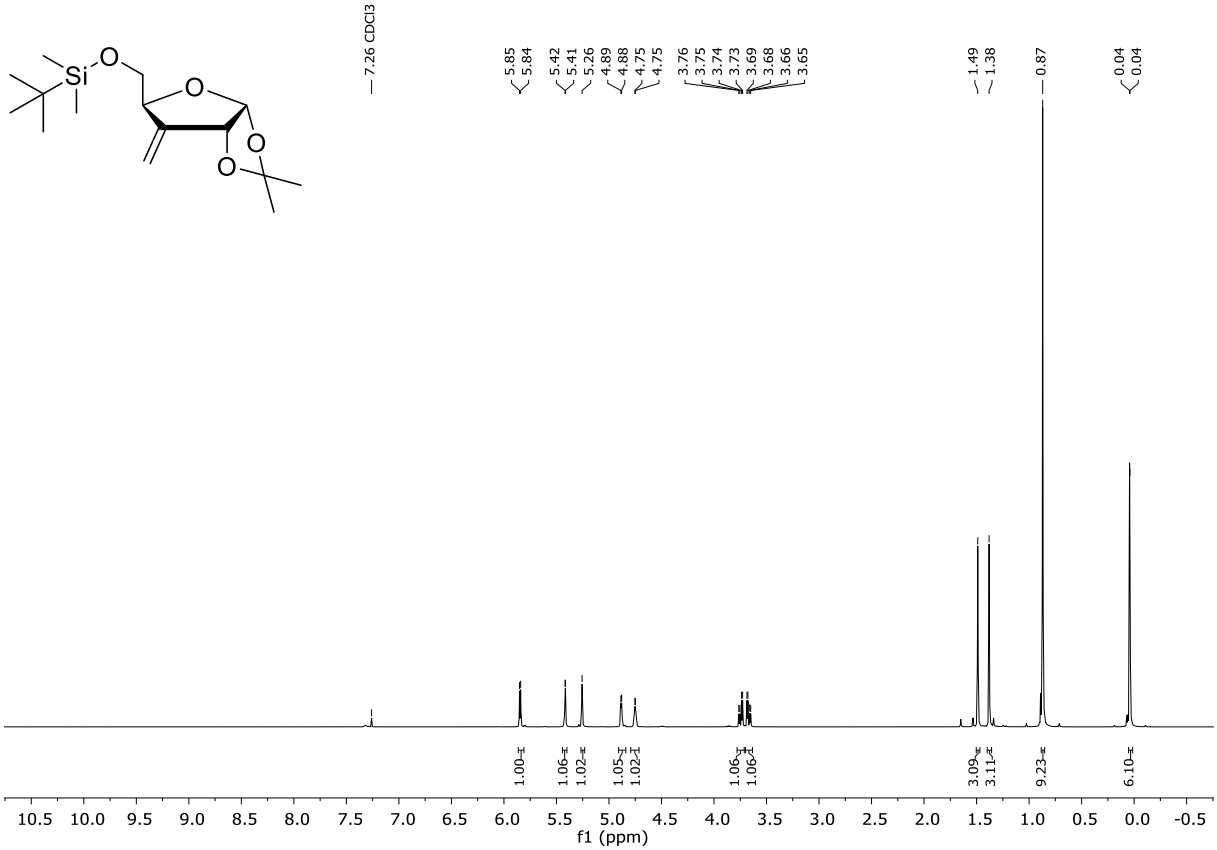
The assembly of dinucleotide **22**. a) $CuSO_4$, Na-Ascorbate, THF/tBuOH/H_2O, RT, 24h, 80%.

4-{6'-Benzoylamino-9'-[2''-*O*-acetyl-5''-(*tert*-butoxycarbonyl)amino-3'',5''-dideoxy-β-D-ribofuranosyl]-9'*H*-purin-3''-yl}-1-{9'''-{3''''-*O*-benzyl-2'''',5''''-dideoxy-2''''-*C*-[(benzyloxycarbonyl)methylene]-β-D-ribofuranosyl}-6'''-*O*-(diphenylcarbamoyl)-2'''-*N*-isobutyrylguanin-5''''-yl}-1,2,3-triazole (22): The title compound was prepared according to a modified procedure of Singh et al.[11] A-half **5** (1.30 g, 2.50 mmol, 1.00 eq.) and G-half **6** (2.39 g, 3.00 mmol, 1.20 eq.) were dissolved in THF/*tert*-BuOH/H_2O (2:2:1, 80 mL) under N_2 at RT. Subsequently, a solution of sodium ascorbate (0.41 g, 2.00 mmol, 0.80 eq.) in water (3 mL) and a solution of copper(II) sulfate (0.16 g, 1.00 mmol, 0.40 eq.) in water (2 mL) was added. The mixture was stirred at RT for 12 h. Volatile components were evaporated and the residue was purified by flash-column chromatography (silica gel, DCM/MeOH, 100:2→100:5→10:1) to provide dinucleotide **22** (2.62 g, 2.00 mmol, 80%) as a colorless foam. M.p. = 183 °C (decomp.). R_f = 0.46 (DCM/MeOH = 10:1). IR (ATR): ṽ = 3268, 2976, 2106, 1738, 1707, 1584, 1452, 1216 1167, 732 cm^{-1}. ^{1}H NMR, COSY (400 MHz, $CDCl_3$): δ = 9.14 (s, 1H, N*H*Bz), 8.86 (s, 1H, H-2'), 8.25 (s, 1H, N*H*^{i}Bu), 8.09 (s, 1H, H-8'), 8.02 (d, ^3J = 7.6 Hz, 1H, Bz-o-C*H*), 7.85 (s, 1H, H-8'''), 7.82 (s, 1H, H-5), 7.62 – 7.56 (m, 1H, Bz-p-C*H*), 7.50 (t, ^3J = 7.6 Hz, 2H, Bz-m-C*H*), 7.48 – 7.12 (m, 20H, aryl-H), 6.81 (dd, ^3J = 7.3 Hz, ^4J = 3.2 Hz, 1H, N*H*Boc), 6.19 (d, ^3J = 3.9 Hz, 1H, H-1''), 5.95 (d, ^3J = 8.2 Hz, 1H, H-1'''''), 5.71 (dd, ^3J = 7.3 Hz, ^3J = 3.9 Hz, 1H, H-2''), 5.15 (dd, ^2J = 14.0 Hz, ^3J = 6.5 Hz, 1H, H_a-5''''), 4.93 (dd, ^2J = 14.0 Hz, ^3J = 8.0 Hz, 1H, H_b-5''''), 4.88 (d, ^2J = 12.3 Hz, 1H, COOCH_2Ph), 4.87 – 4.83 (m, 1H, H-4''''), 4.83 – 4.78 (m, 1H, H-4''), 4.79 (d, ^2J = 12.3 Hz, 1H, COOCH_2Ph), 4.61 (d, ^2J = 11.6 Hz, 1H, OCH_2Ph), 4.40 (d, ^3J = 5.3 Hz, 1H, H-3''''), 4.37 – 4.32 (m, 1H, H-3''), 4.32 (d, ^2J = 11.6 Hz, 1H, OCH_2Ph), 4.15 – 4.04 (m, 1H, H-2''''), 3.61 – 3.51 (m, 1H, H_a-5''), 3.53 – 3.44 (m, 1H, H_b-5''), 2.80 (dd, ^2J = 16.6 Hz, ^3J = 8.3 Hz, 1H, CH_2COO), 2.59 (hept, ^3J = 6.9 Hz, 1H, C*H*(CH$_3$)$_2$), 2.45 (dd, ^2J = 16.7 Hz, ^3J = 7.1 Hz, 1H, CH_2COO), 1.70 (s, 3H, OCOCH_3), 1.45 (s, 9H, C(CH_3)$_3$), 1.25, (d, ^3J = 6.9 Hz, 3H, CH(CH_3)$_2$), 1.21, (d, ^3J = 6.9 Hz, 3H, CH(CH_3)$_2$) ppm. ^{13}C NMR, HSQC, HMBC (101 MHz, $CDCl_3$): δ = 174.3 (iBu-*C*ONH), 171.0 (*C*OOBn), 169.8 (O*C*OCH$_3$), 164.7 (N-*C*O-Ph), 156.7 (N-*C*O-OC(CH$_3$)$_3$), 156.4 (C-2'''), 154.1 (C-4'''), 152.7 (C-2'), 151.6 (O*C*ONPh$_2$), 151.1 (C-4'), 150.4 (C-6'), 150.1 (C-6'''), 144.4 (C-8'''), 142.7 (C-8'), 141.7 (OCON-*Ph*$_2$-C1), 140.7 (C-4), 137.1 (OCH$_2$*Ph*-C-1), 135.4 (COOCH$_2$*Ph*-C-1), 133.5 (Bz-*C*-CO-N), 133.0 (Bz-p-*C*H), 129.3, 129.0,

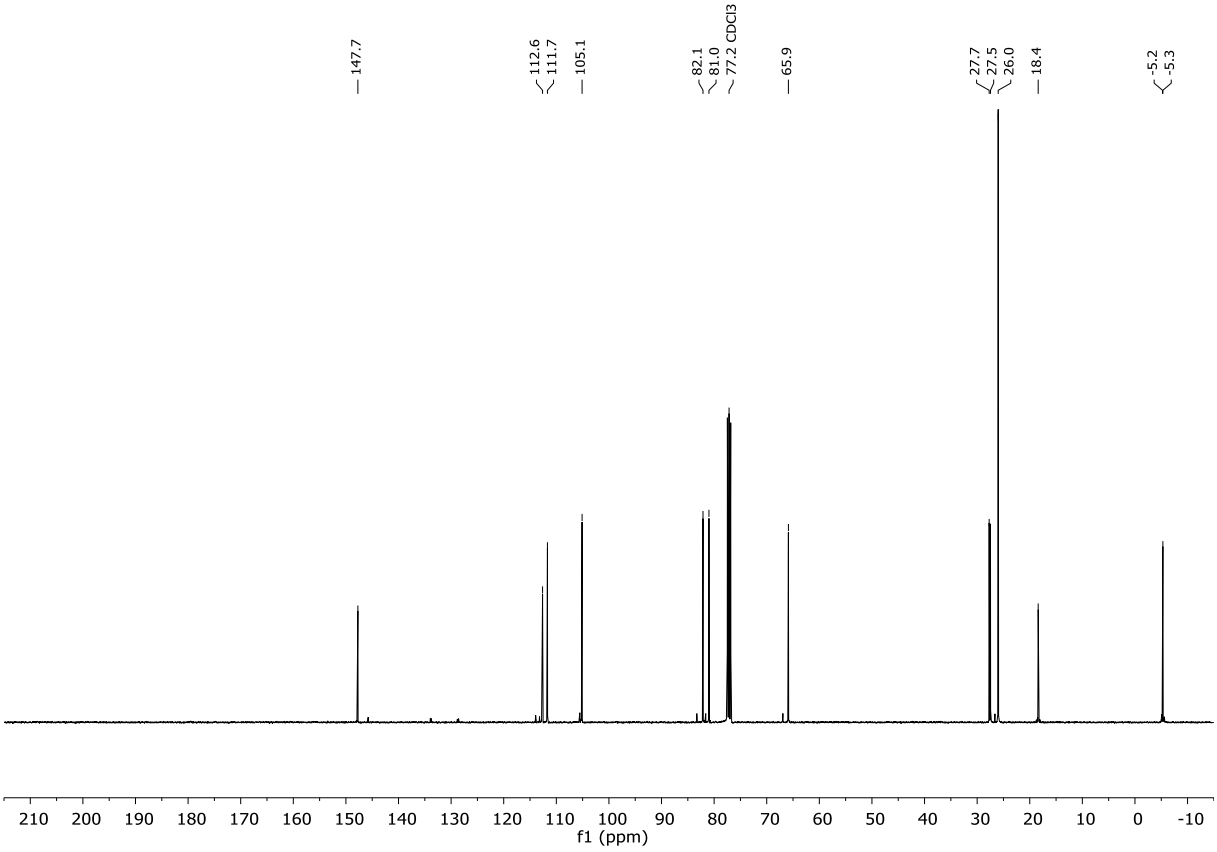
128.61, 128.55, 128.42, 128.25, 128.17, 128.03, 127.98 (aryl 25C), 125.2 (C-5), 124.2 (C-5‘), 122.8 (C-5‘‘‘), 90.9 (C-1‘‘‘‘), 90.2 (C-1‘‘), 83.3 (C-4‘‘), 82.1 (C-4‘‘‘‘), 79.6 (C-3‘‘‘‘), 79.4 ($C(CH_3)_3$), 77.1 (C-2‘‘), 71.8 (OCH_2Ph), 66.7 ($COOCH_2Ph$), 51.2 (C-5‘‘‘‘), 42.4 (C-5‘‘), 40.5 (C-2‘‘‘‘), 39.9 (C-3‘‘), 36.9 ($CH(CH_3)_2$), 30.3 (CH_2COO), 28.6 ($C(CH_3)_3$), 20.3 ($OCOCH_3$), 19.5 ($CH(CH_3)_2$), 19.40 ($CH(CH_3)_2$) ppm. ESI-HRMS calcd. for $[C_{69}H_{69}N_{15}O_{13} + H]^+$: 1316.5272, found: 1316.5330. ESI-HRMS calcd. for $[C_{69}H_{69}N_{15}O_{13} + Na]^+$: 1338.5091, found: 1338.5151.

2. NMR spectra of the synthesized compounds

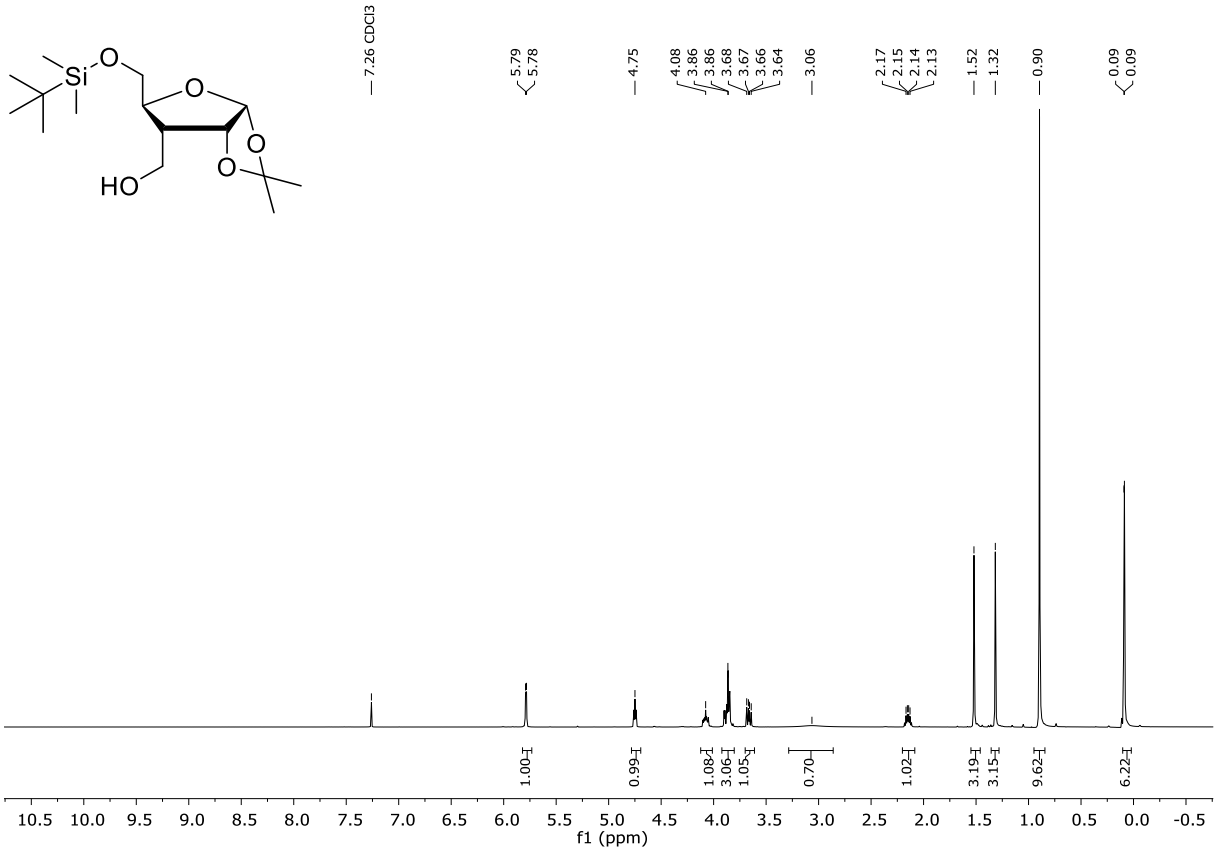
Compound **8** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



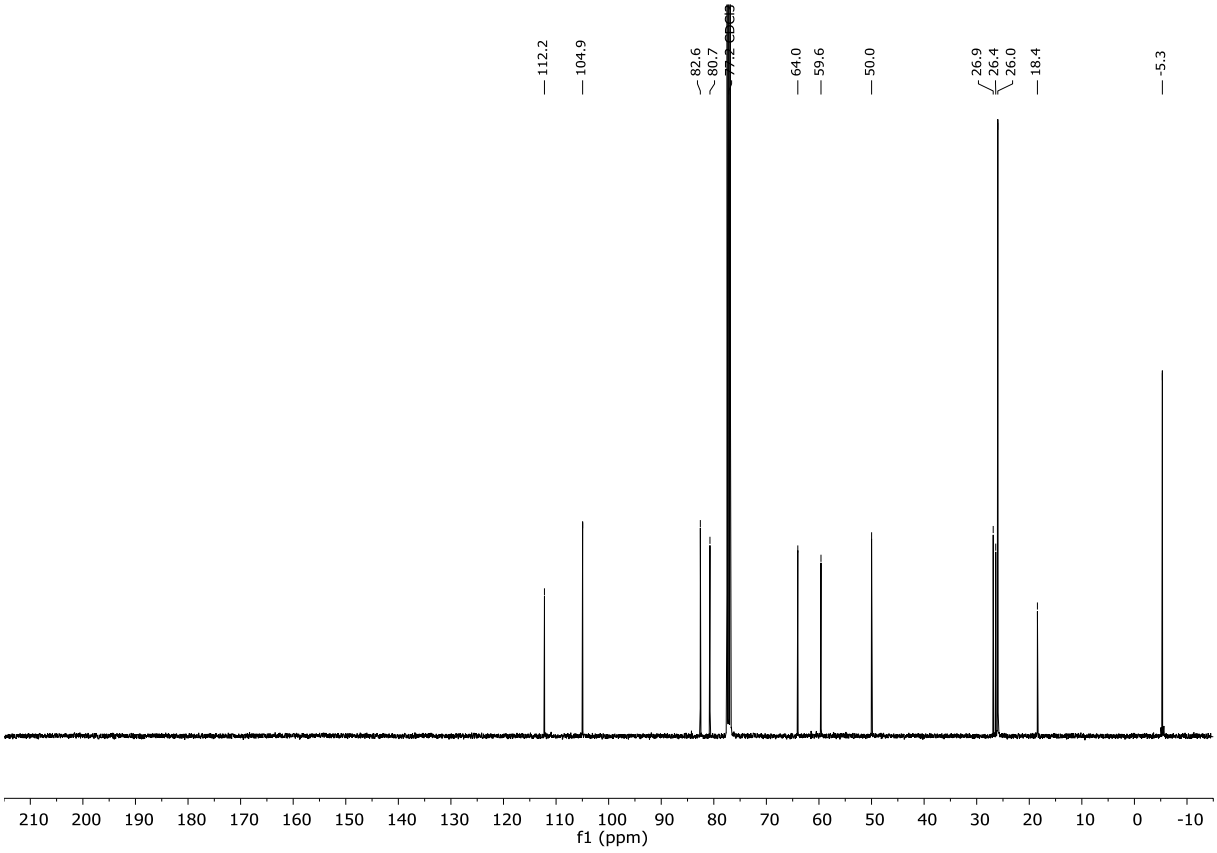
Compound **8** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



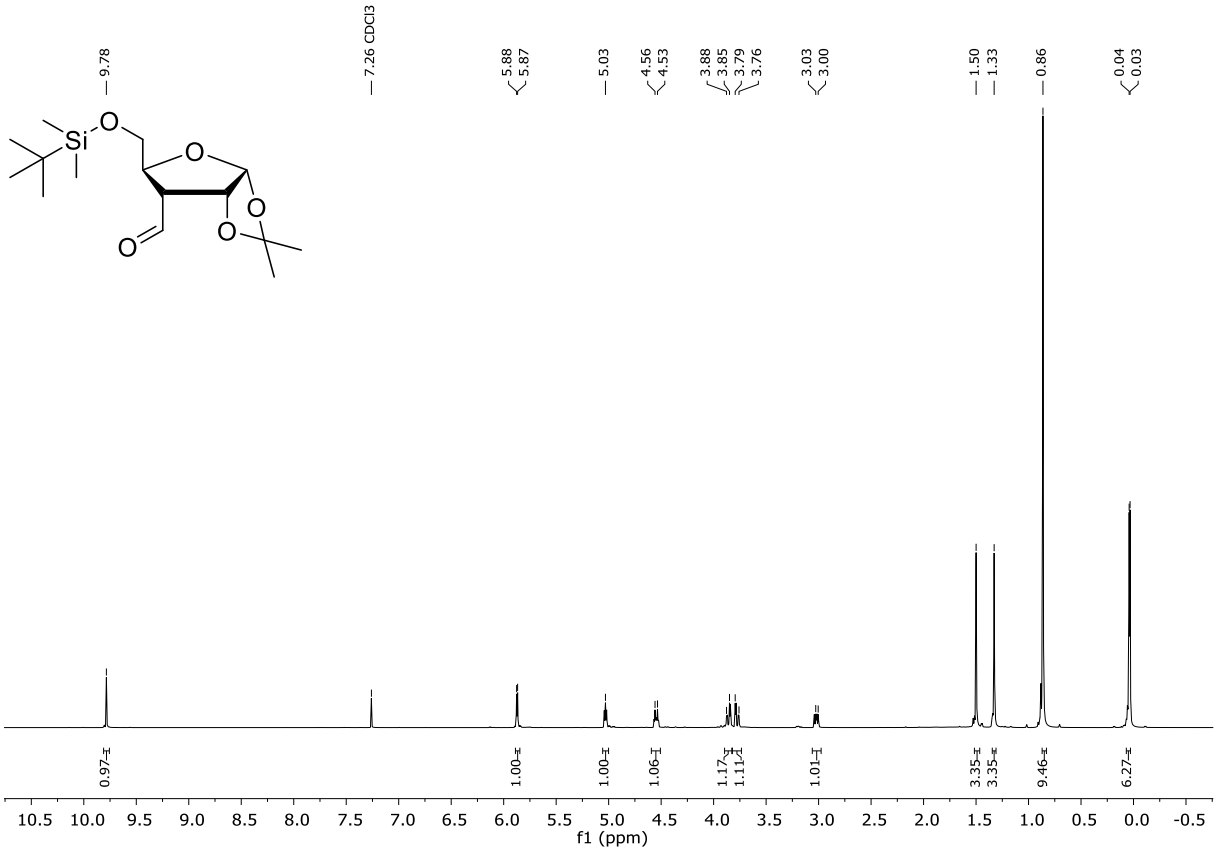
Compound **9a** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



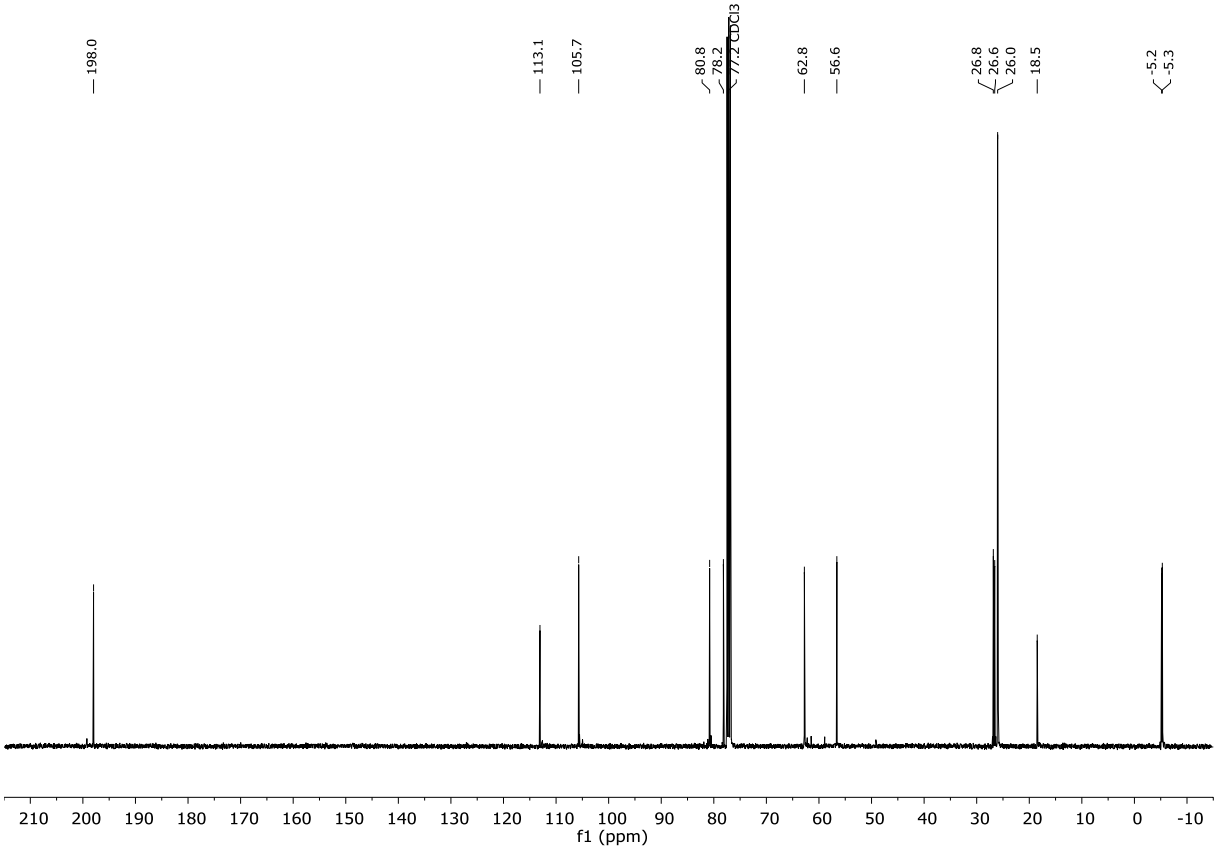
Compound **9a** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



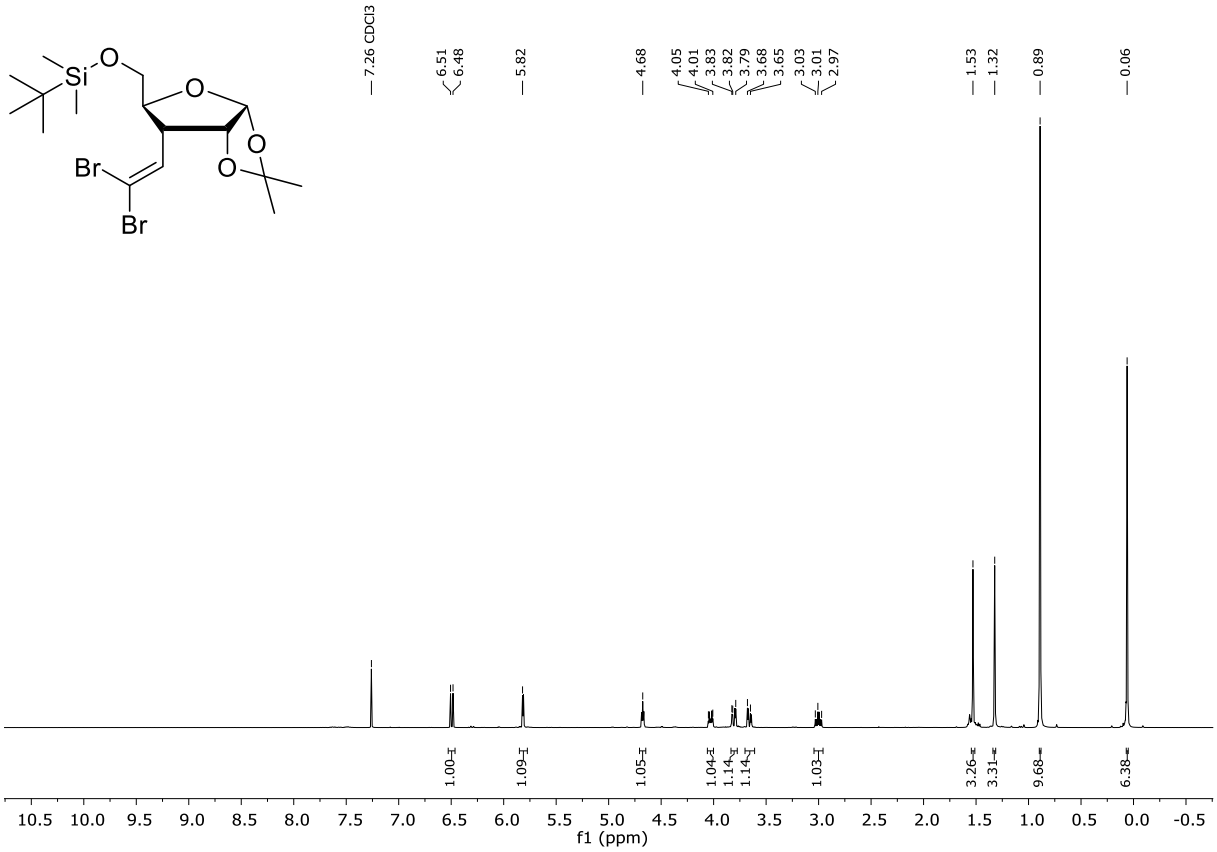
Compound **9** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



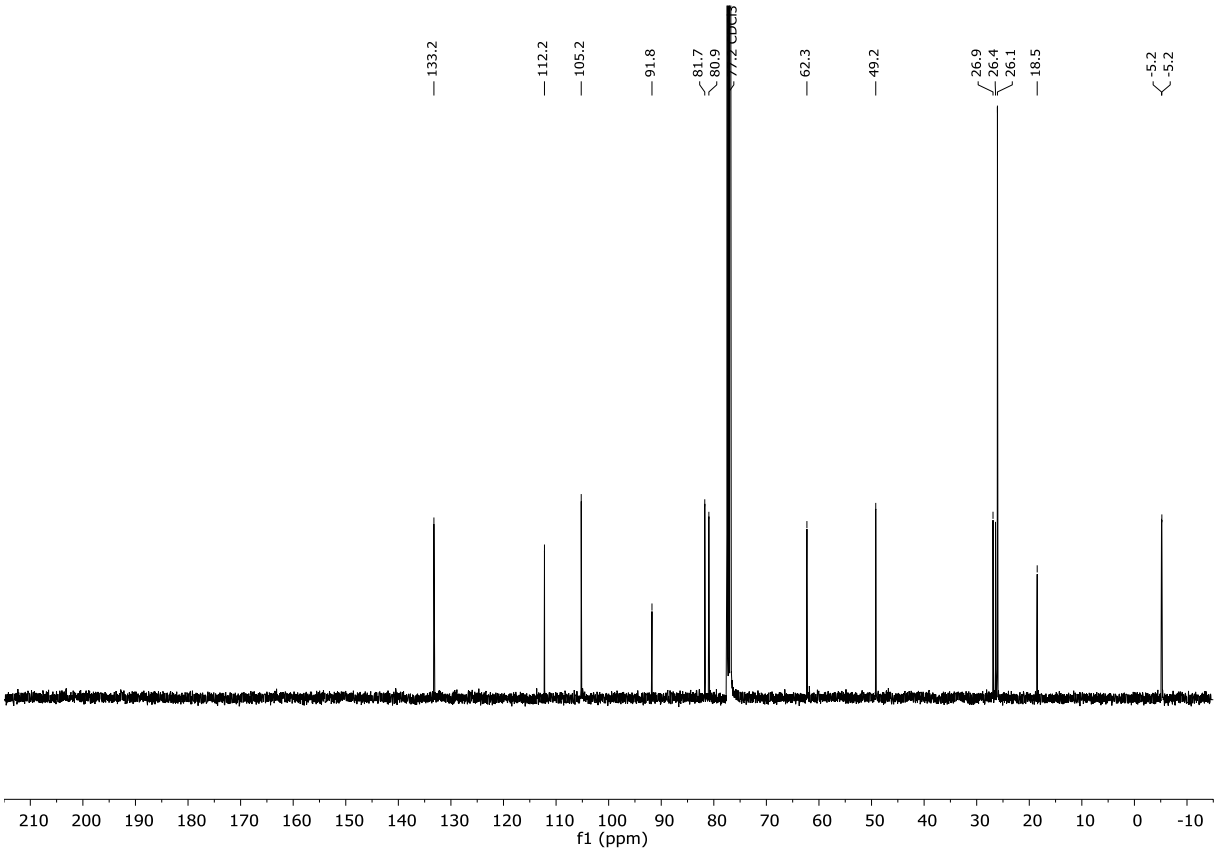
Compound **9** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



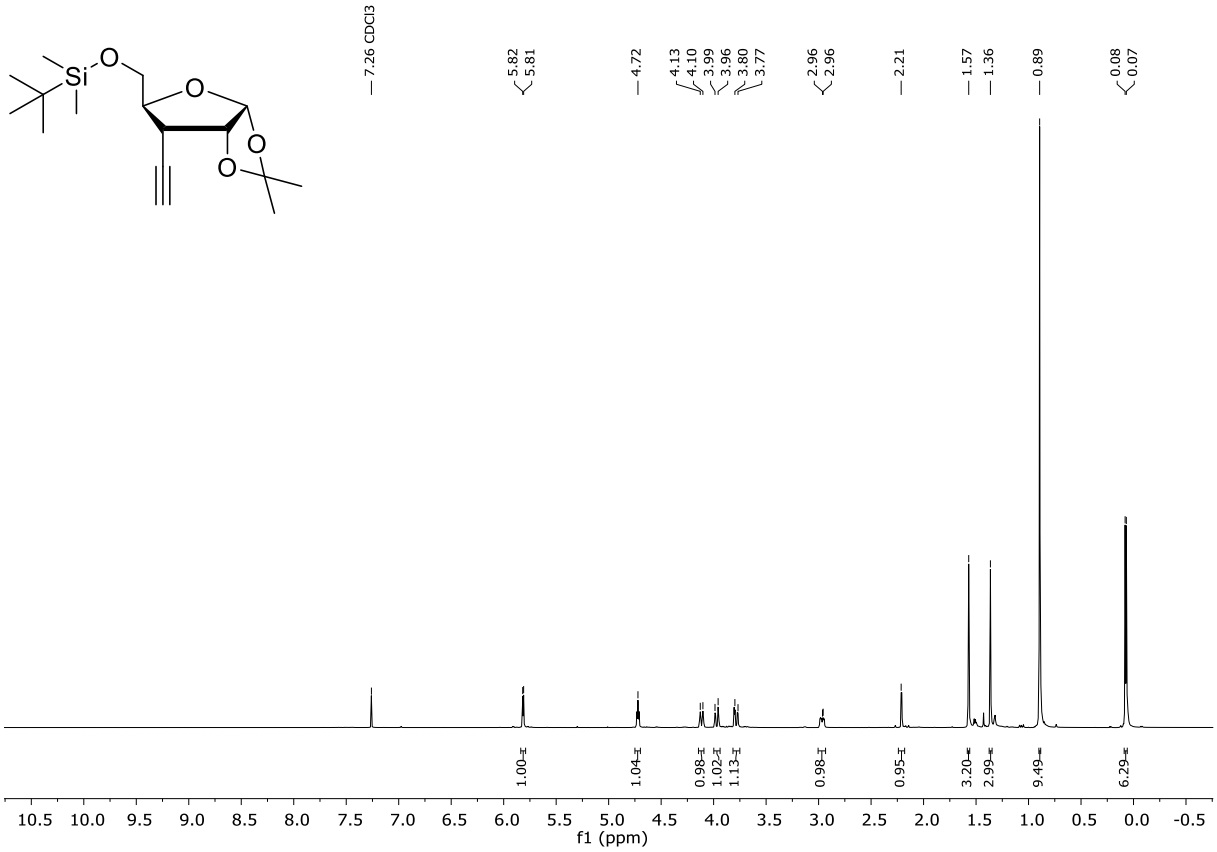
Compound **10a** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



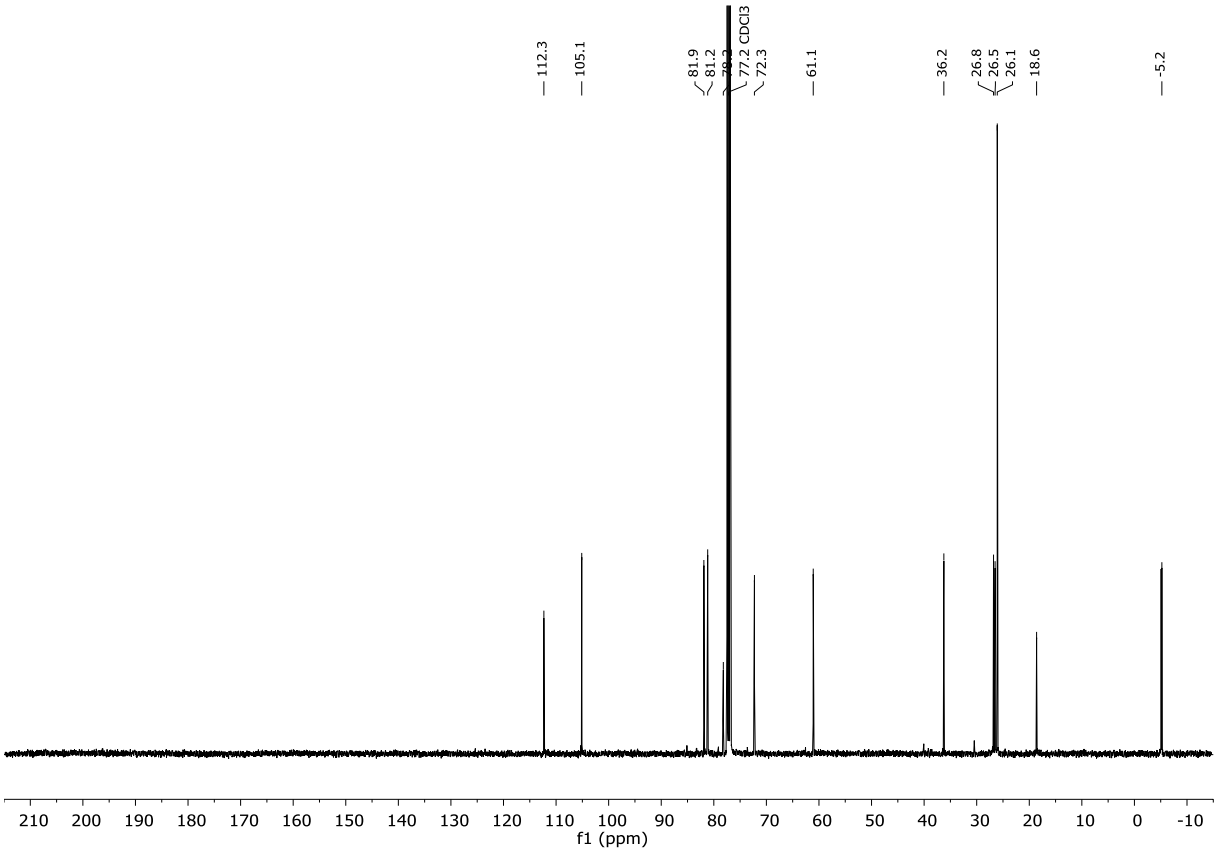
Compound **10a** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



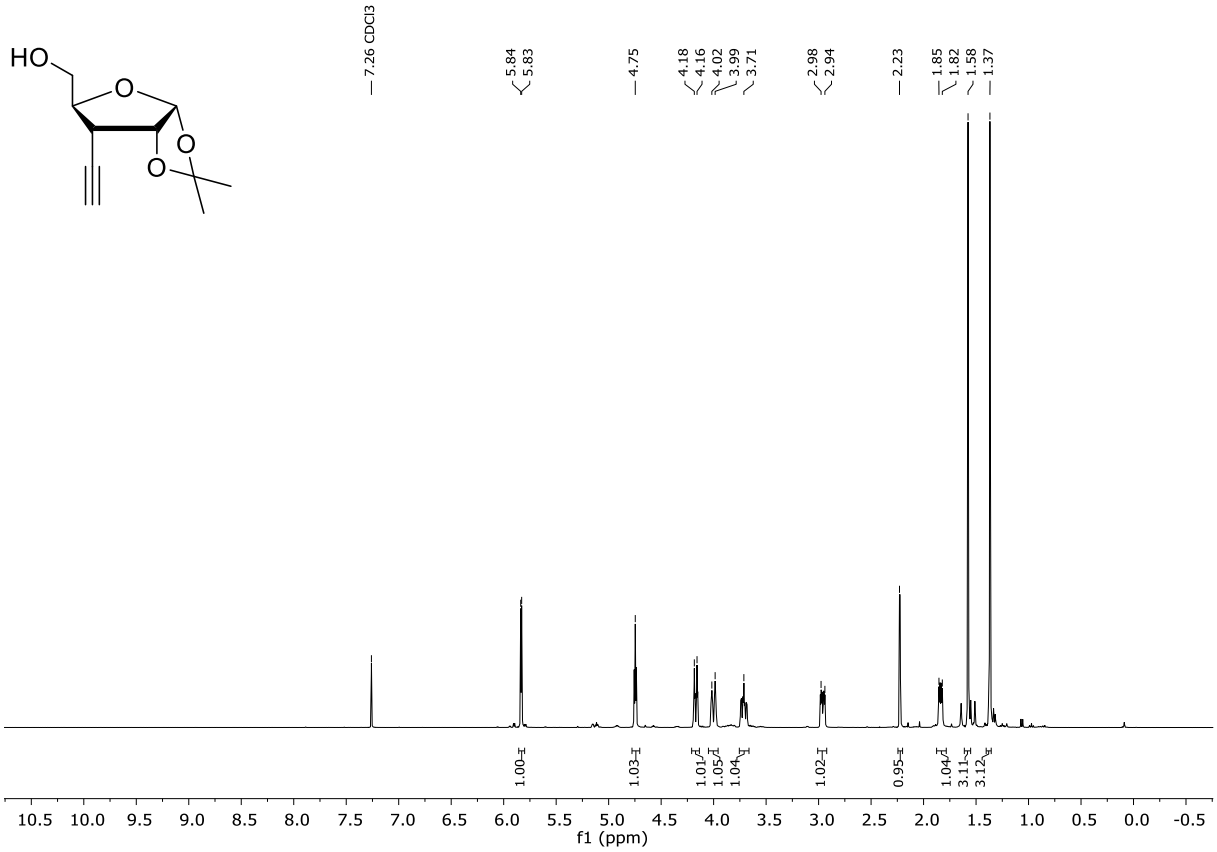
Compound **10b** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



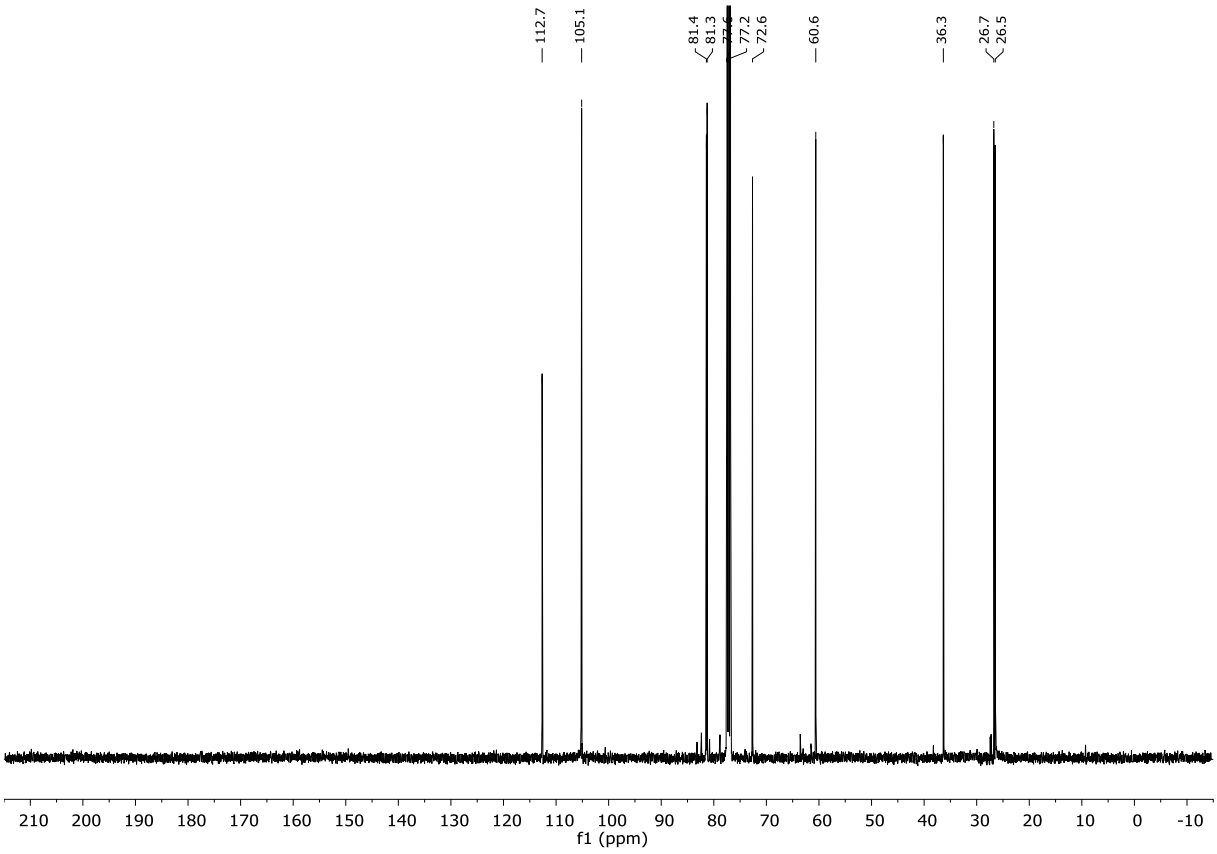
Compound **10b** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



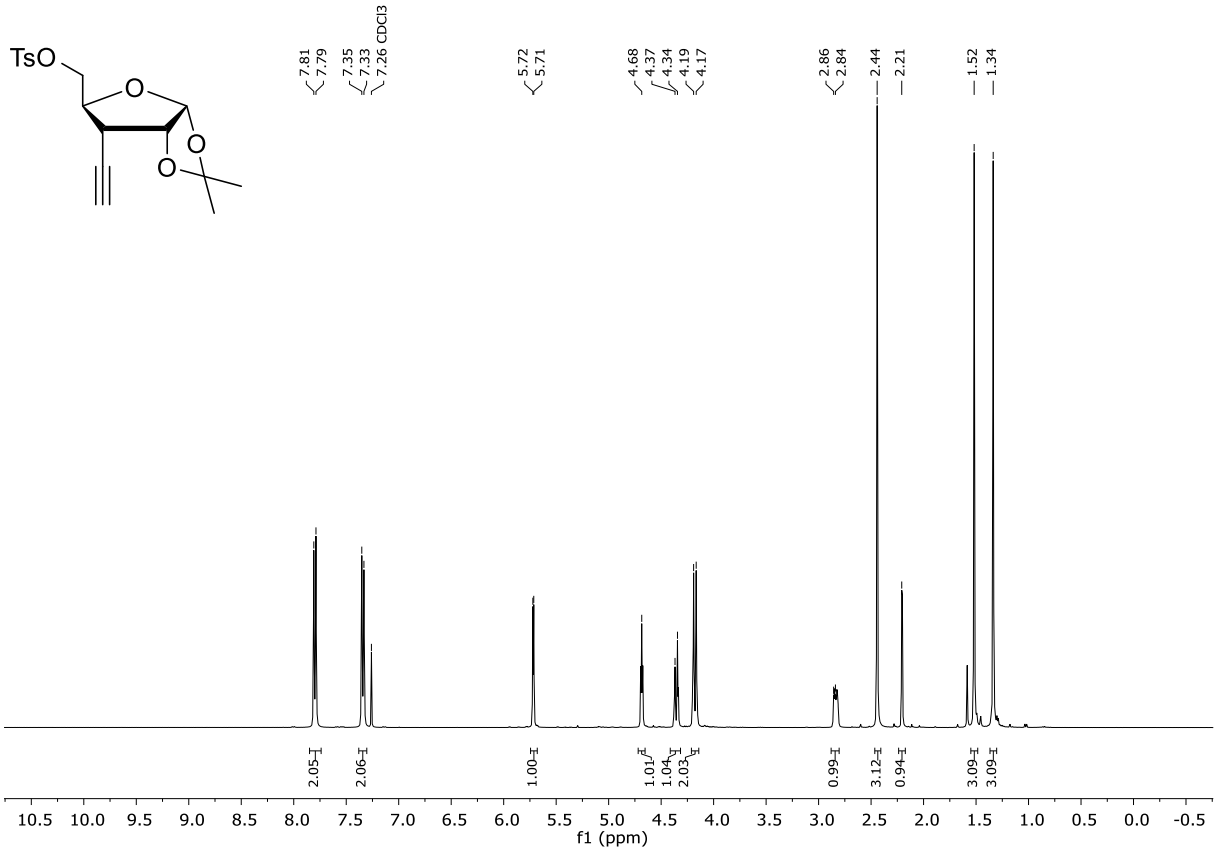
Compound **10c** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



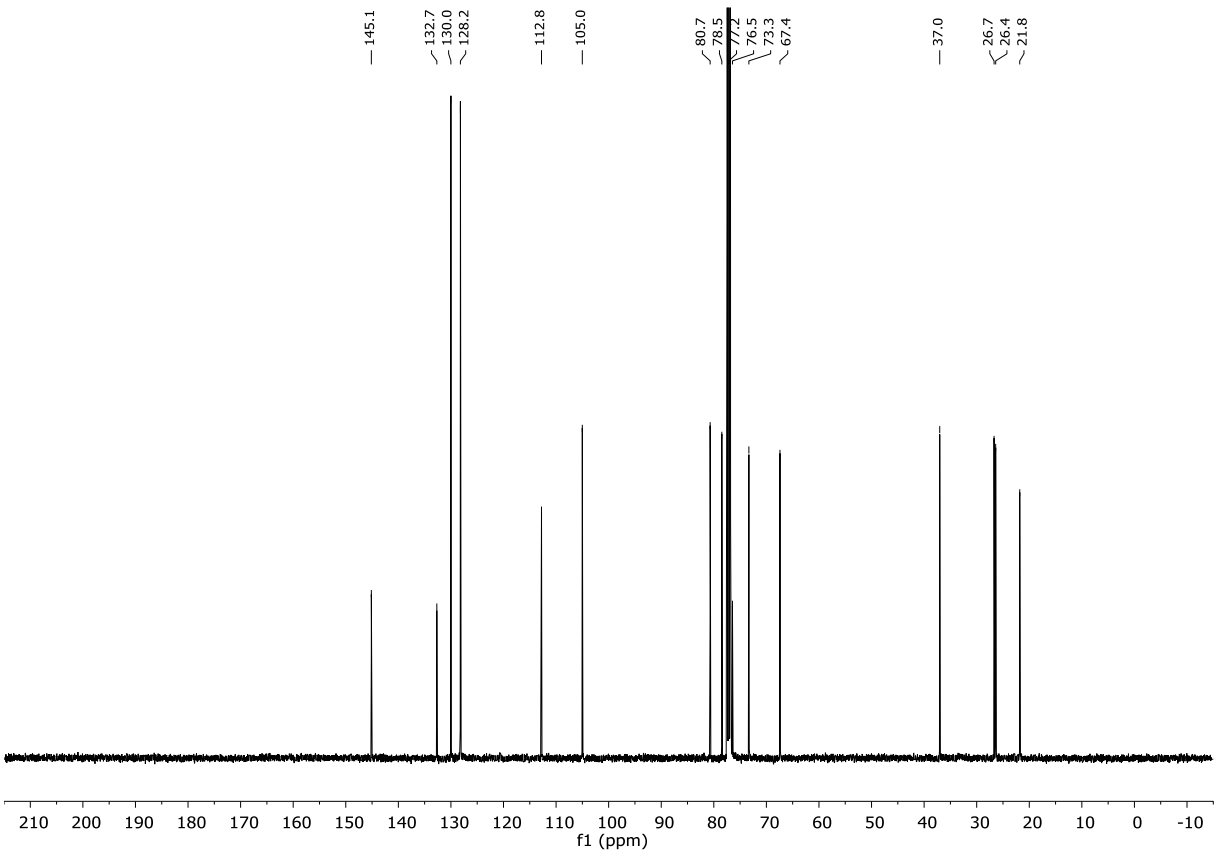
Compound **10c** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



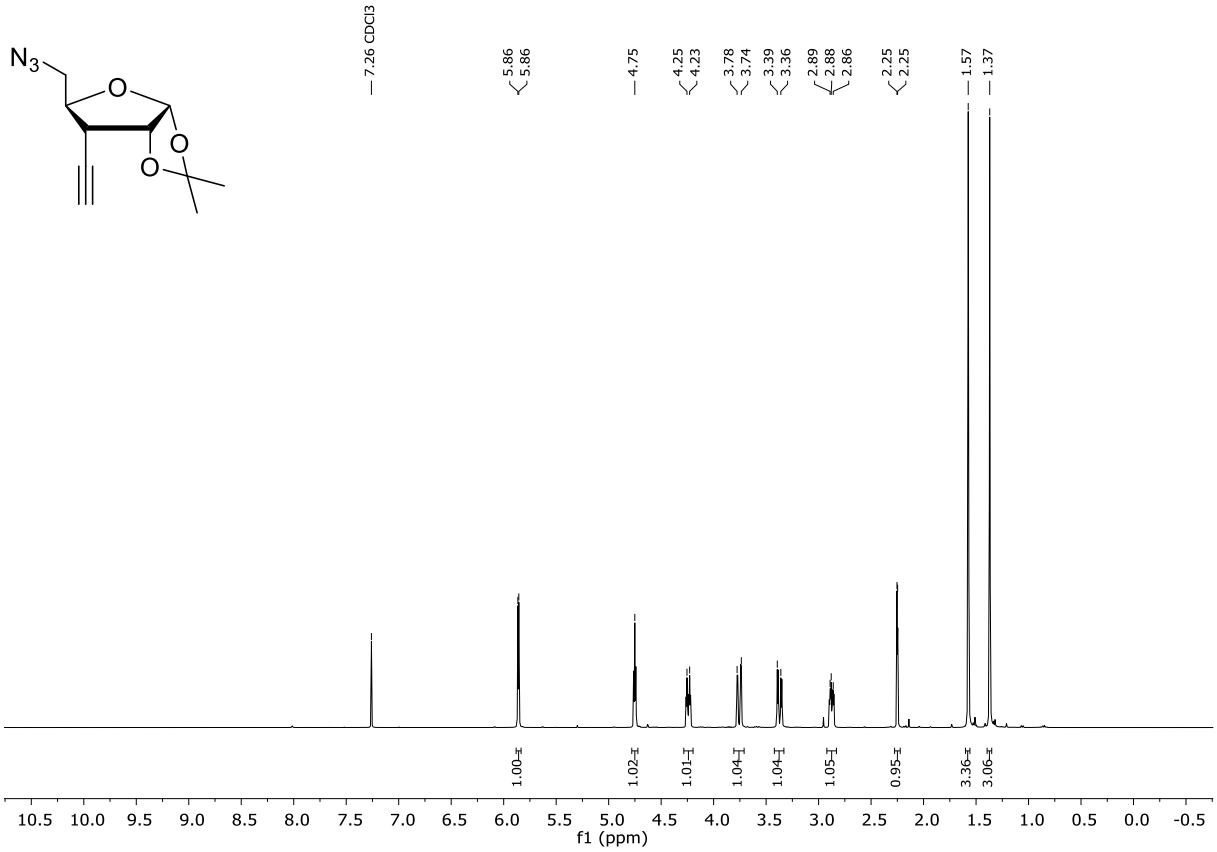
Compound **10d** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



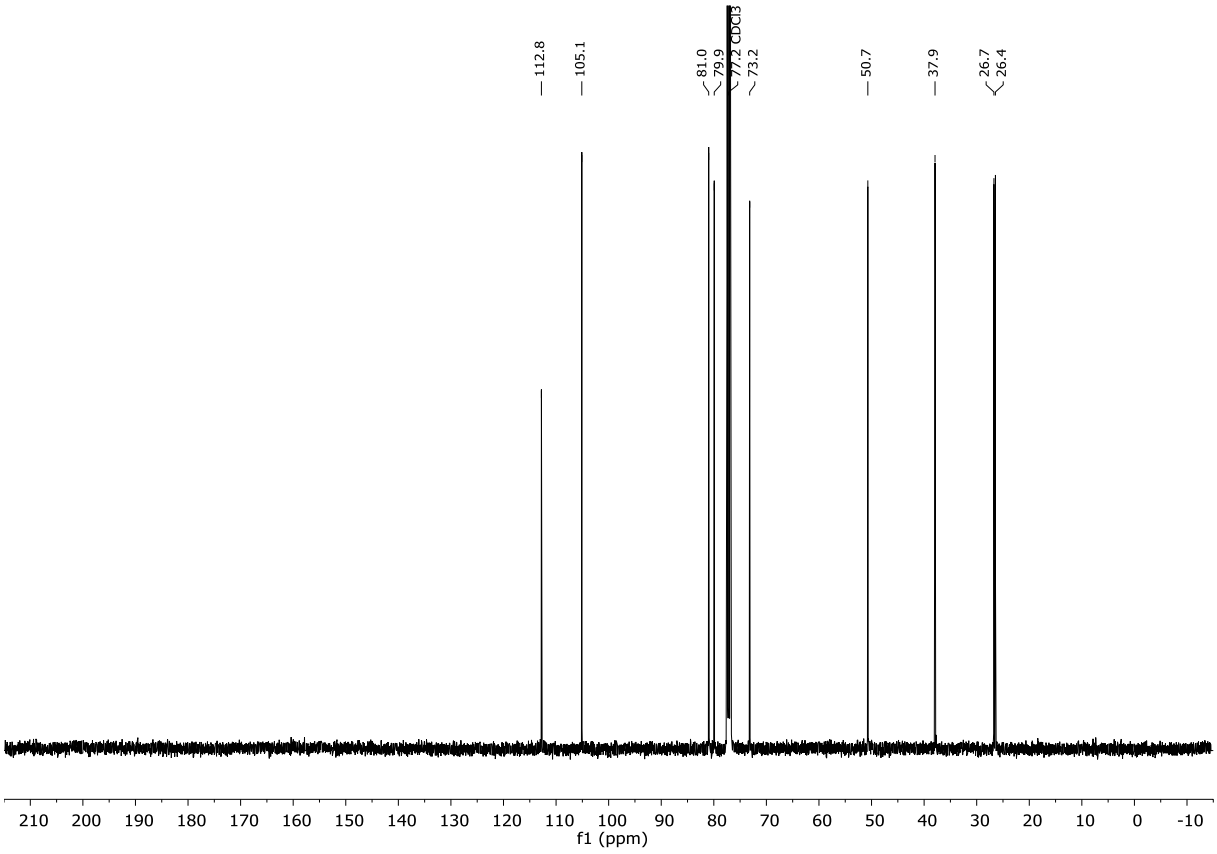
Compound **10d** (^{13}C-NMR, 100 MHz, Chloroform-*d*)



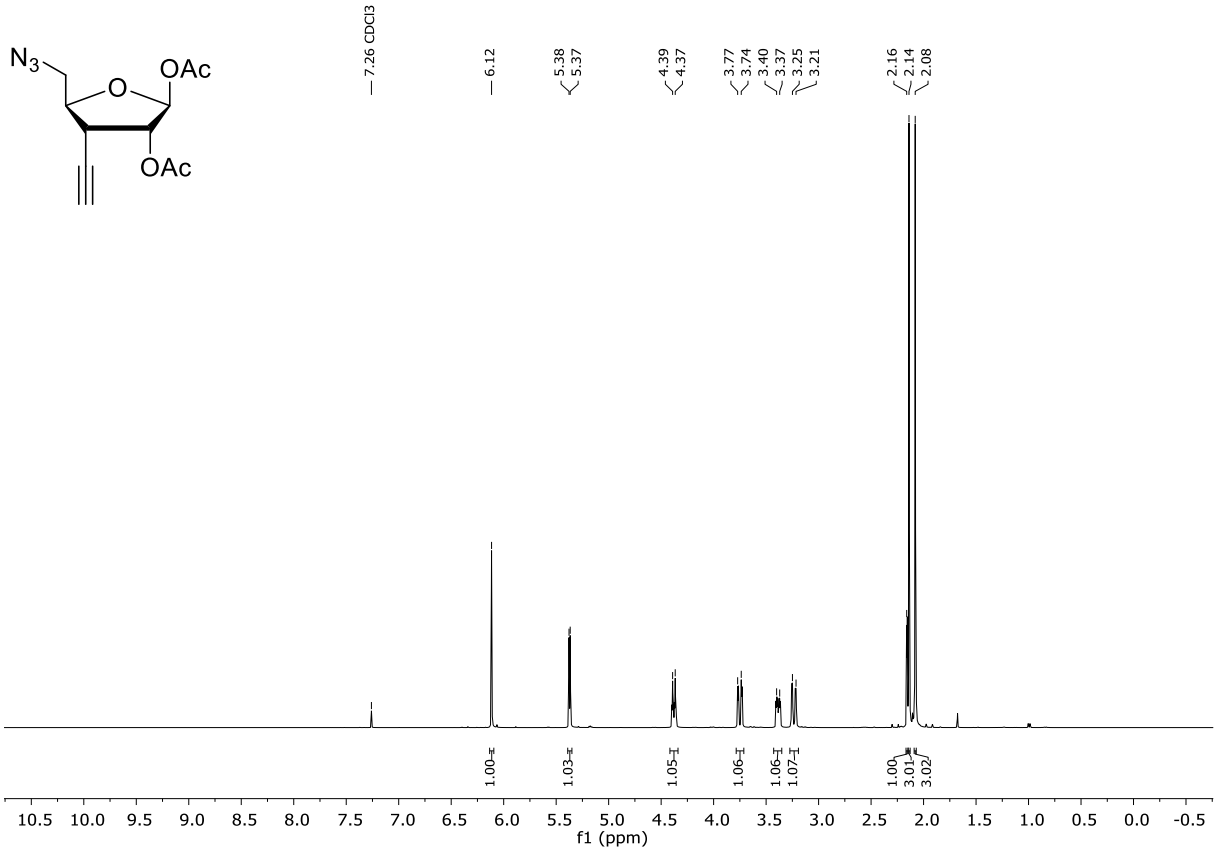
Compound **10** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



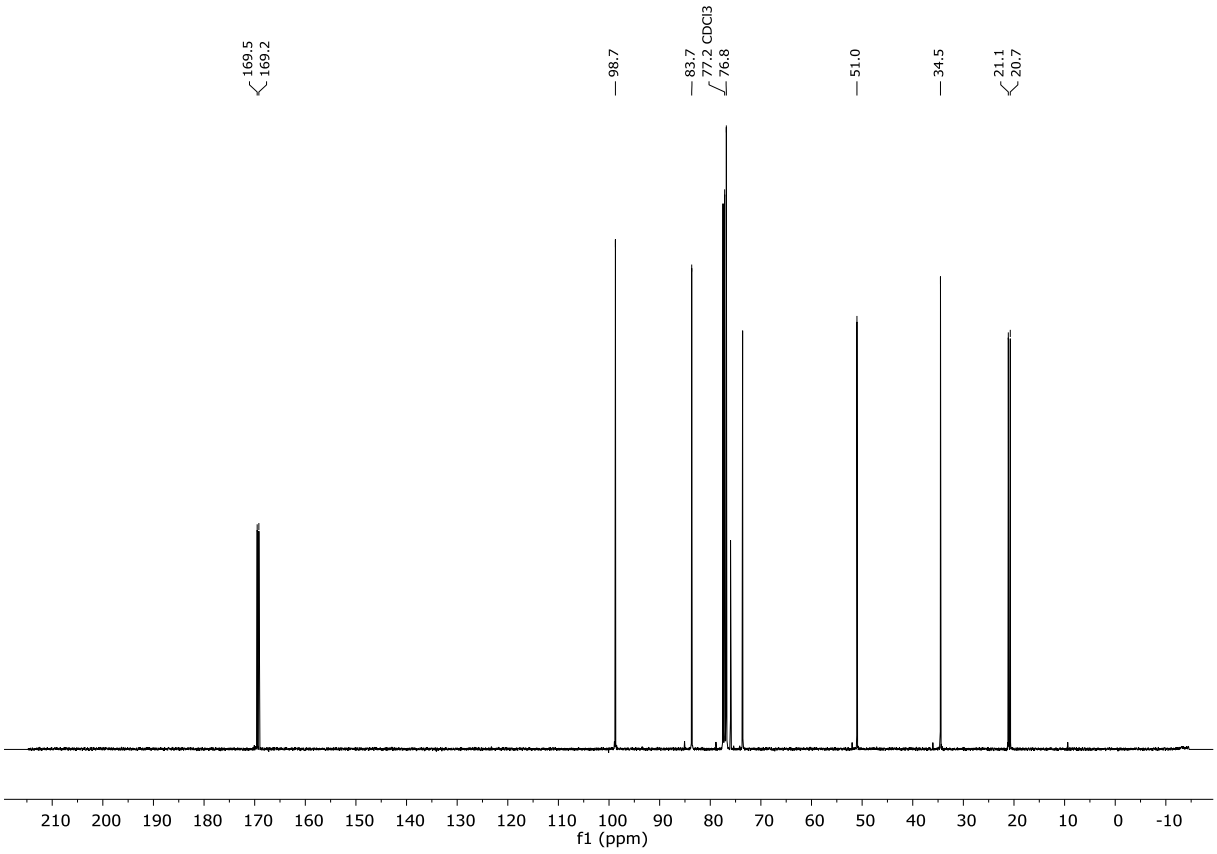
Compound **10** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



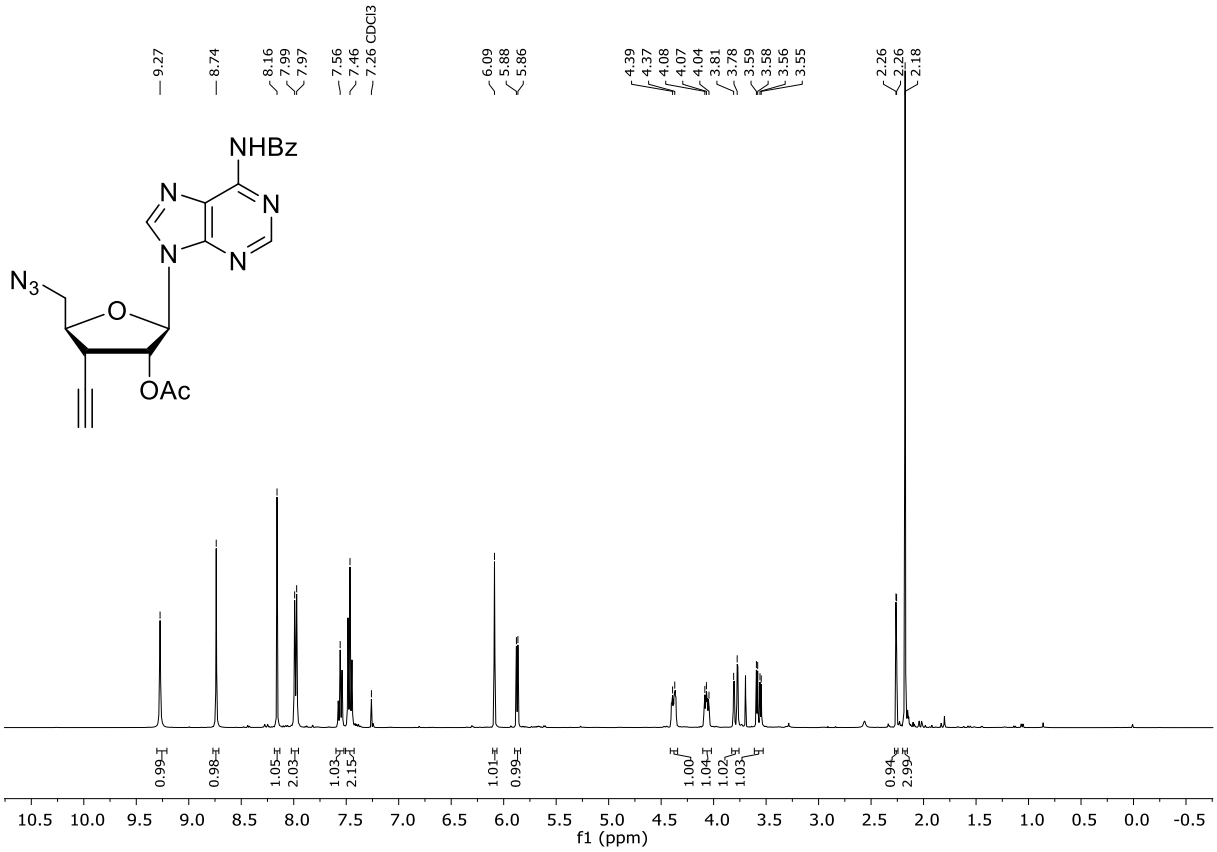
Compound **11** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



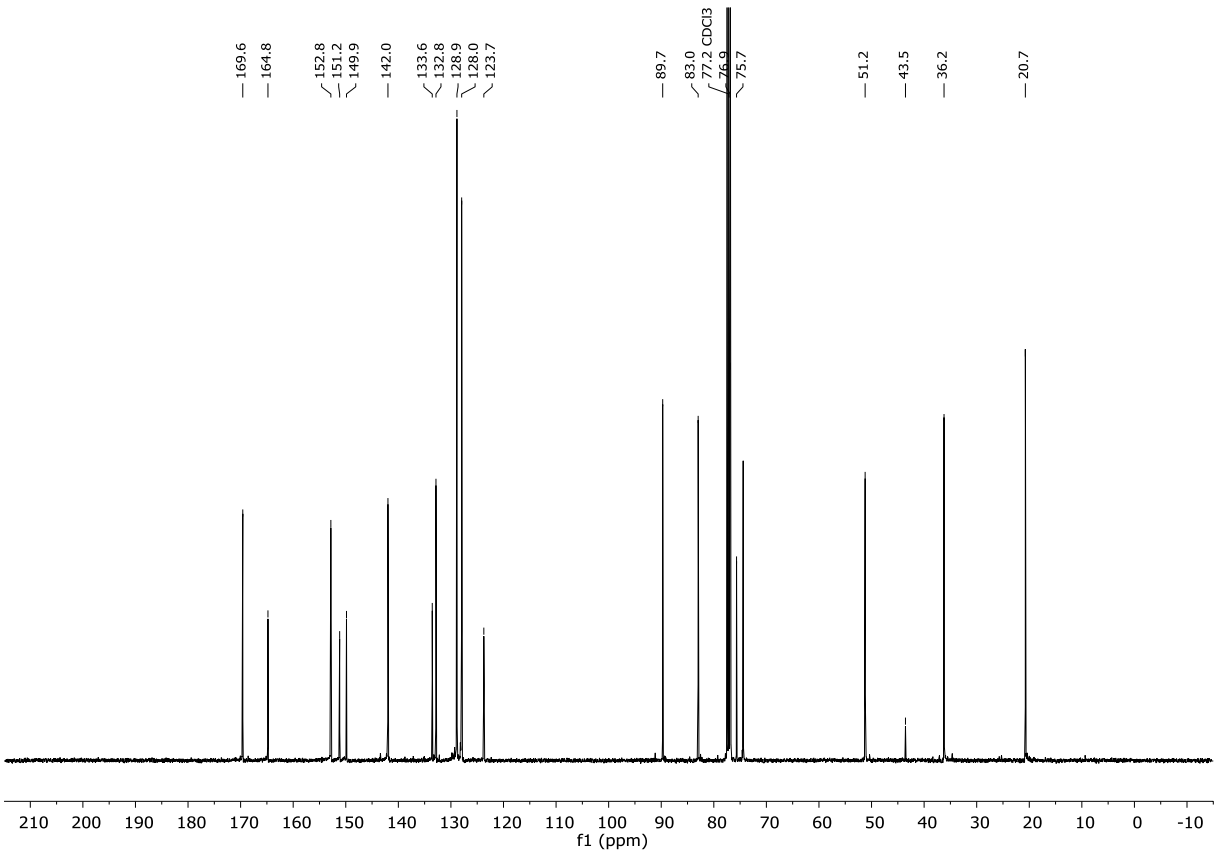
Compound **11** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



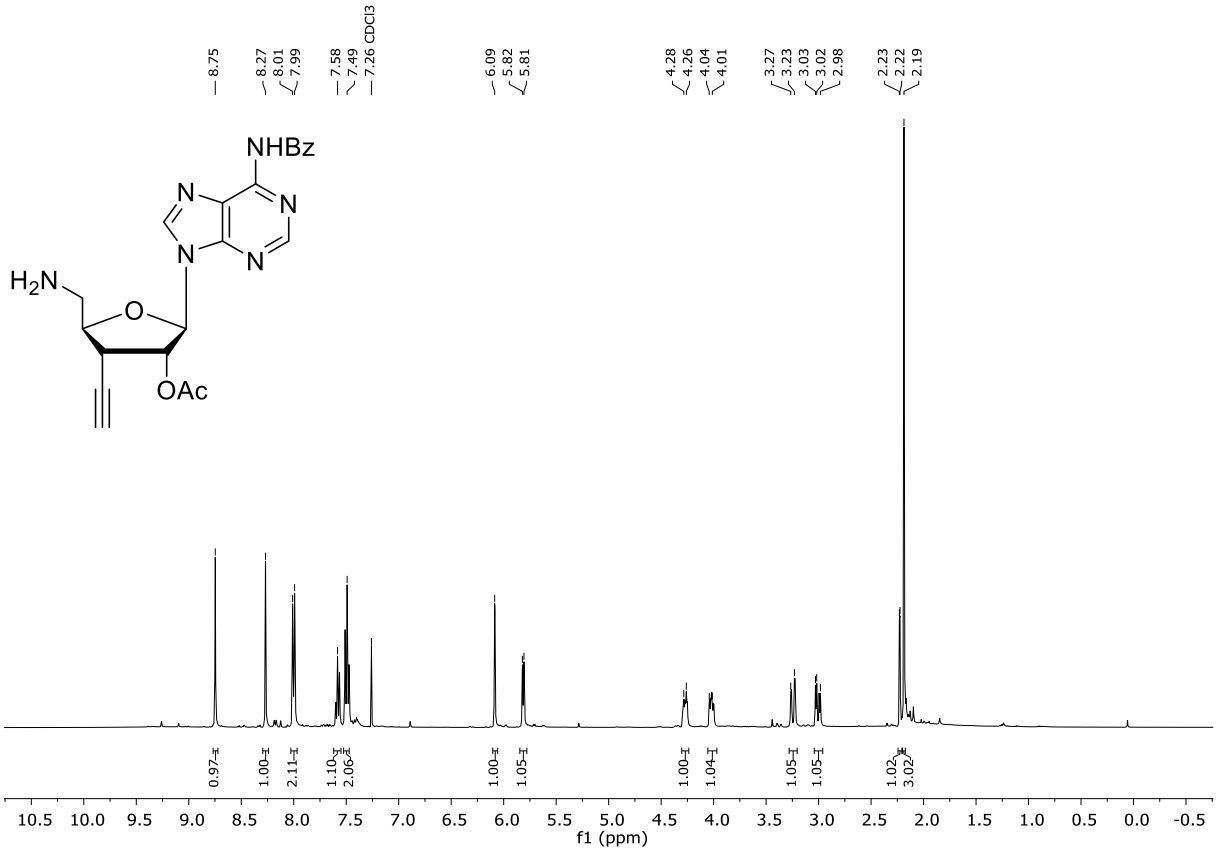
Compound **12** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



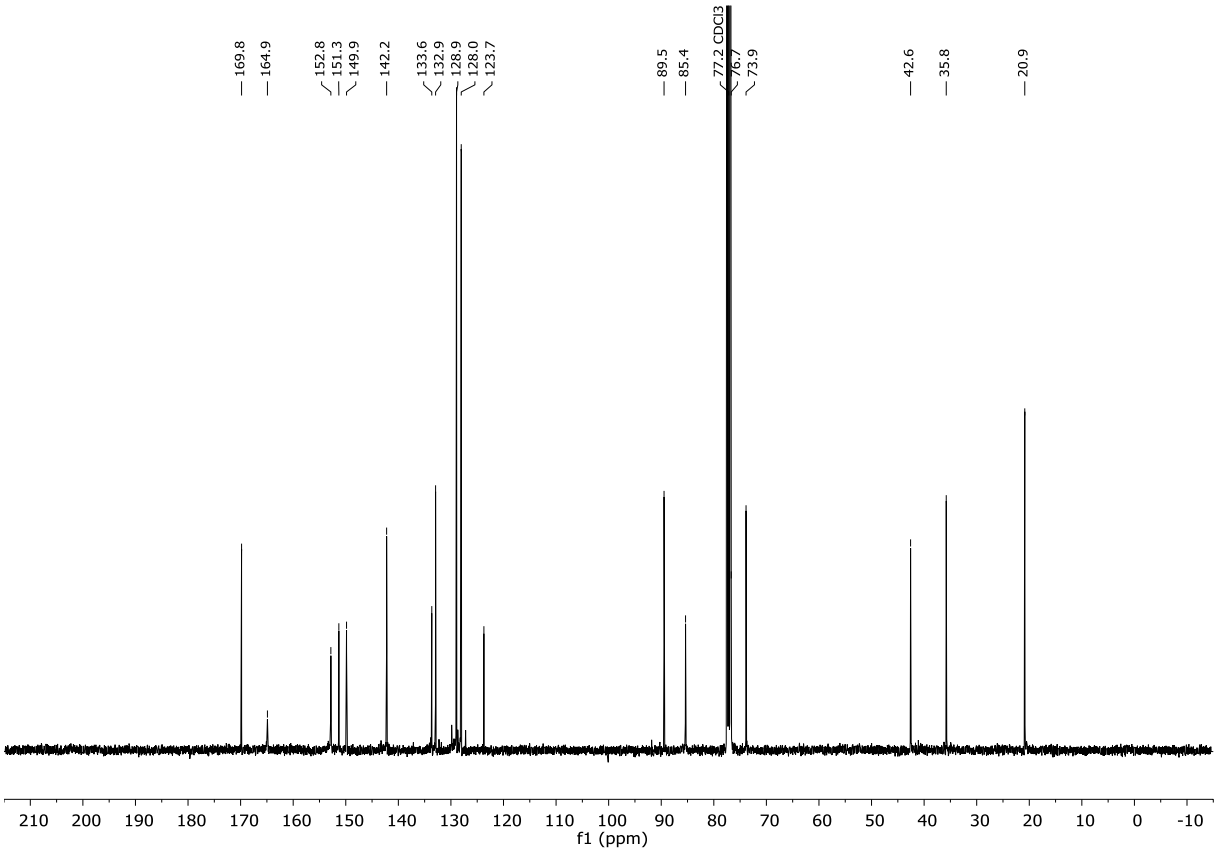
Compound **12** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



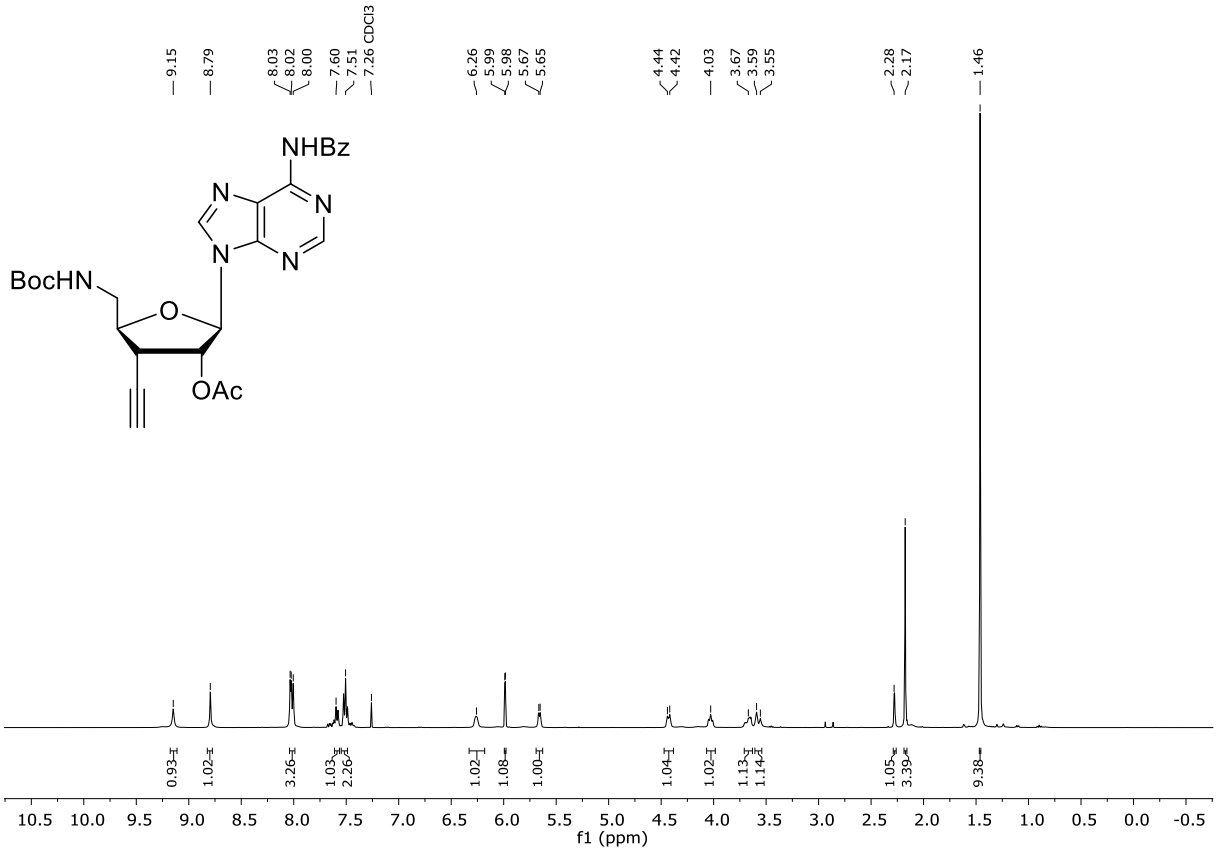
Compound **5a** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



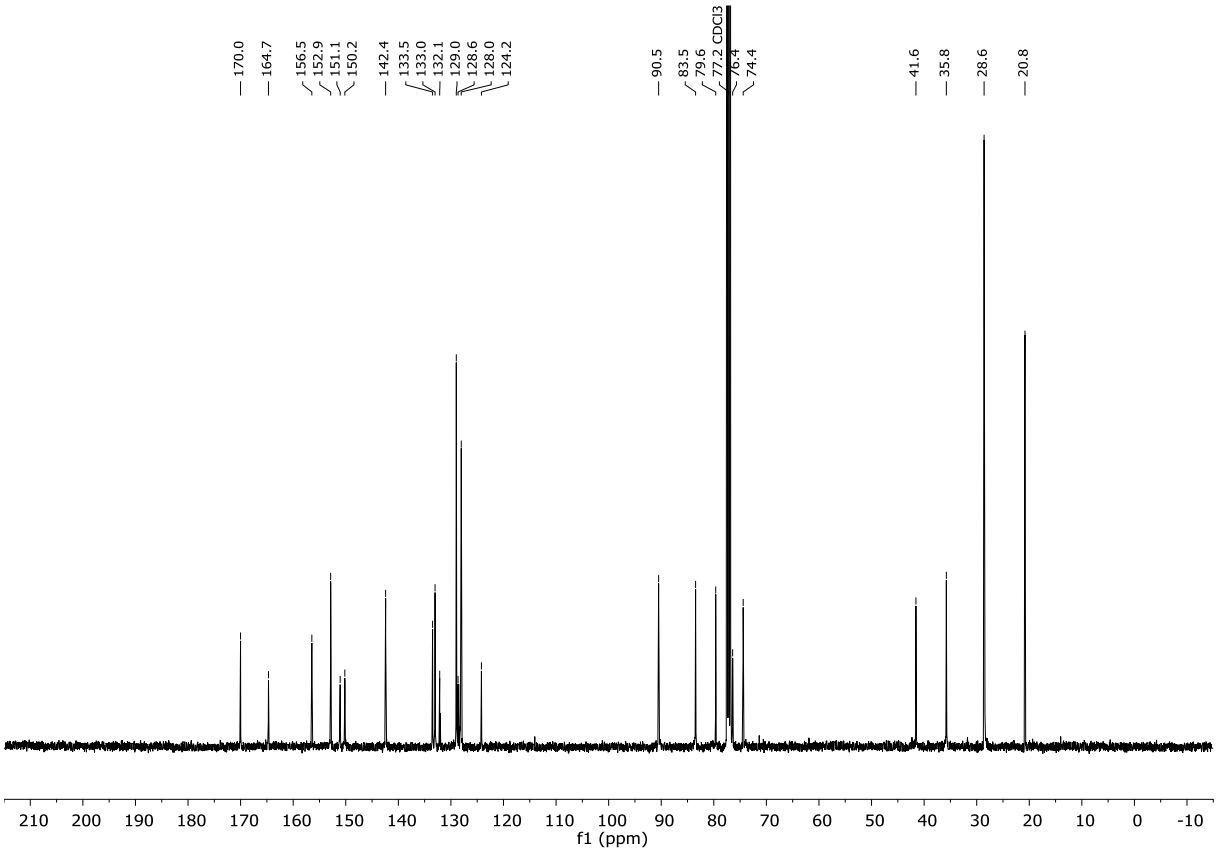
Compound **5a** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



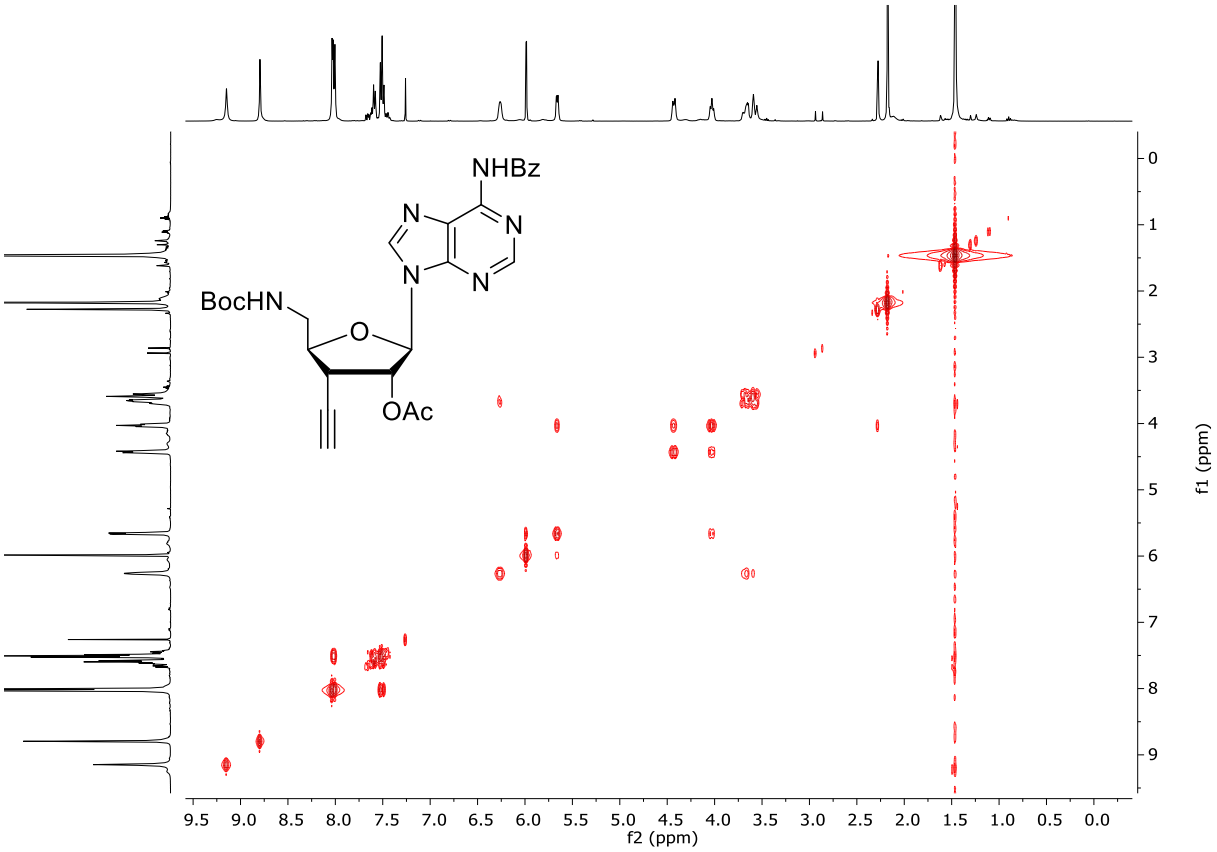
Compound **5** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



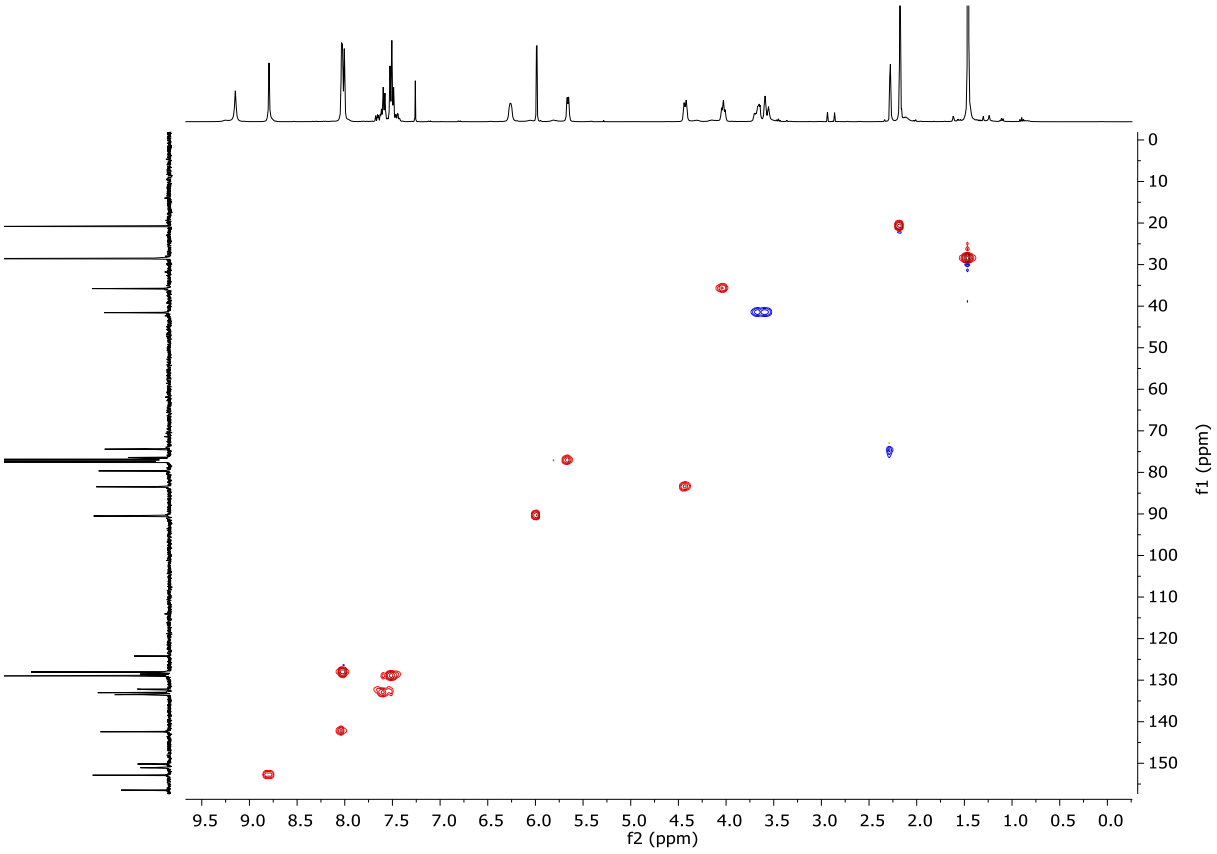
Compound **5** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



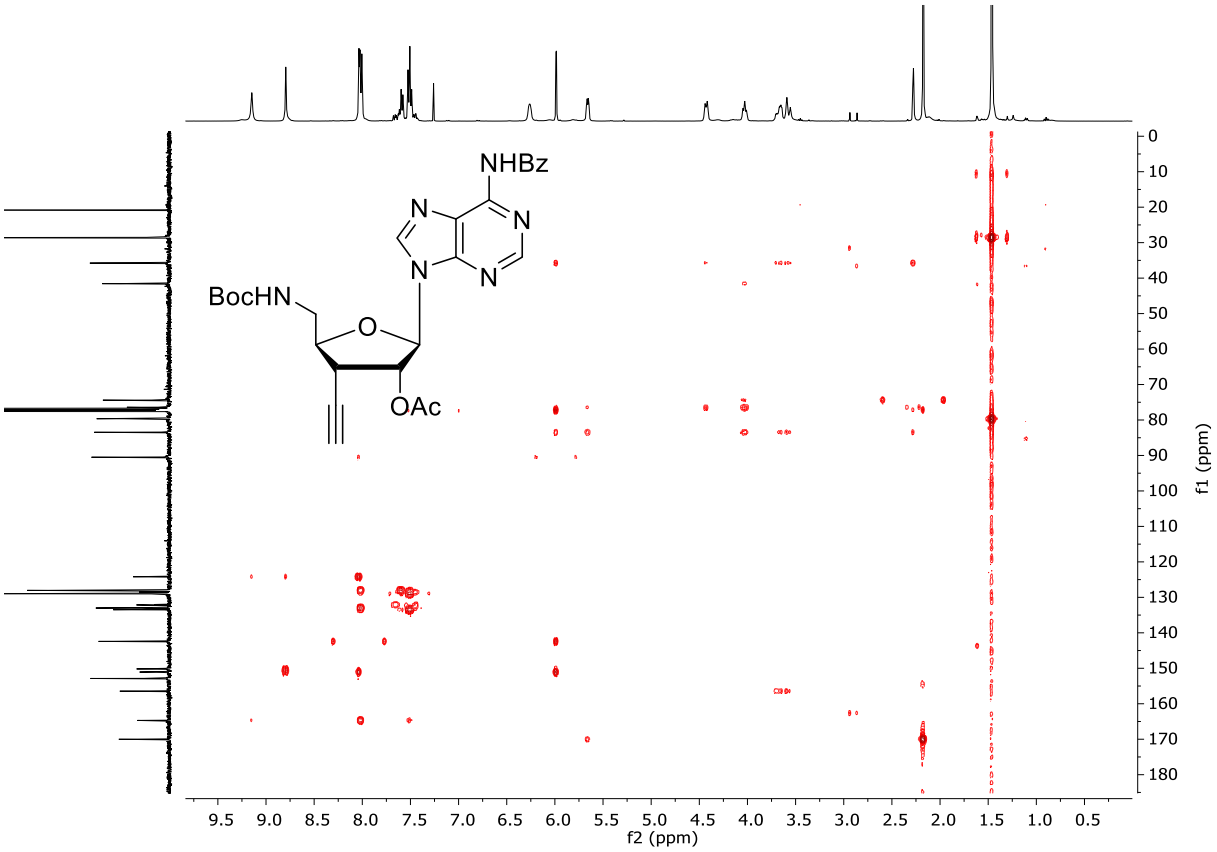
Compound **5** (COSY, 400 MHz, Chloroform-*d*)



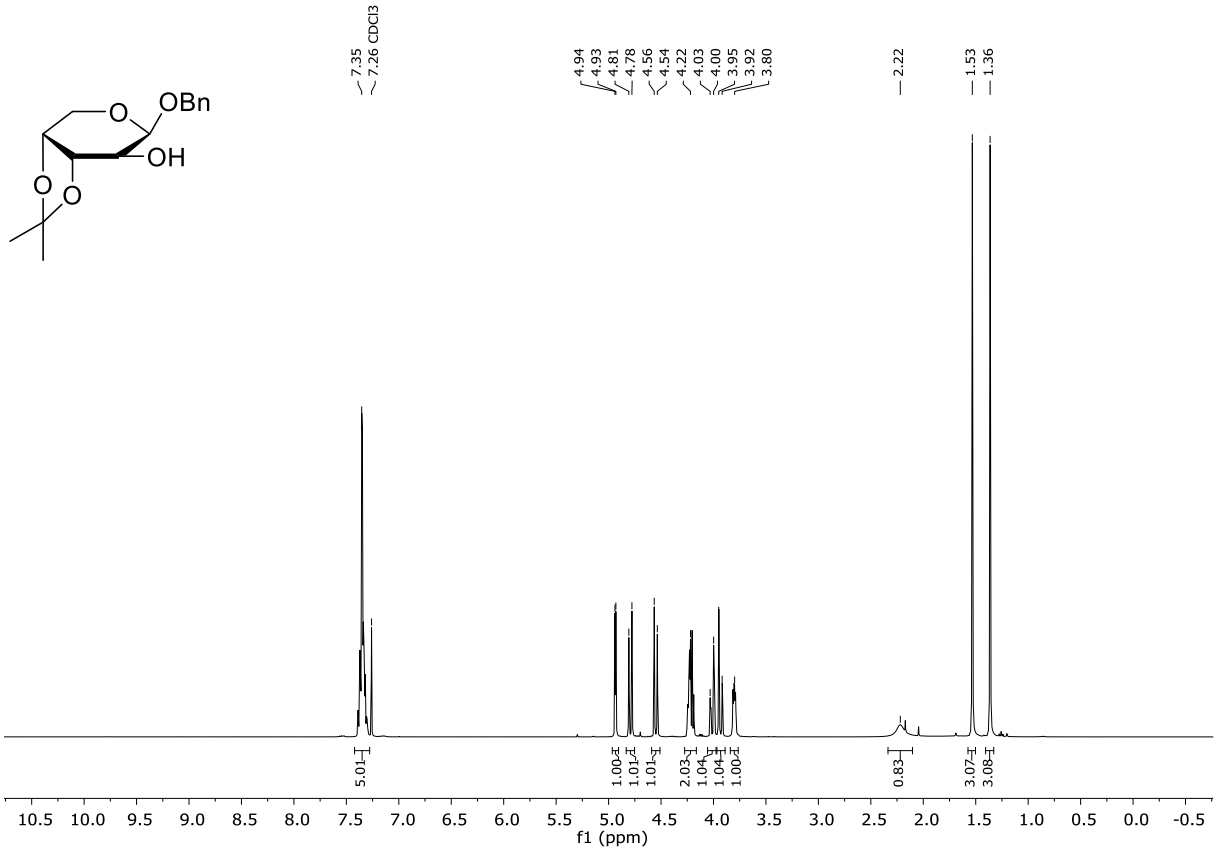
Compound **5** (HSQC, 400 MHz, Chloroform-*d*)



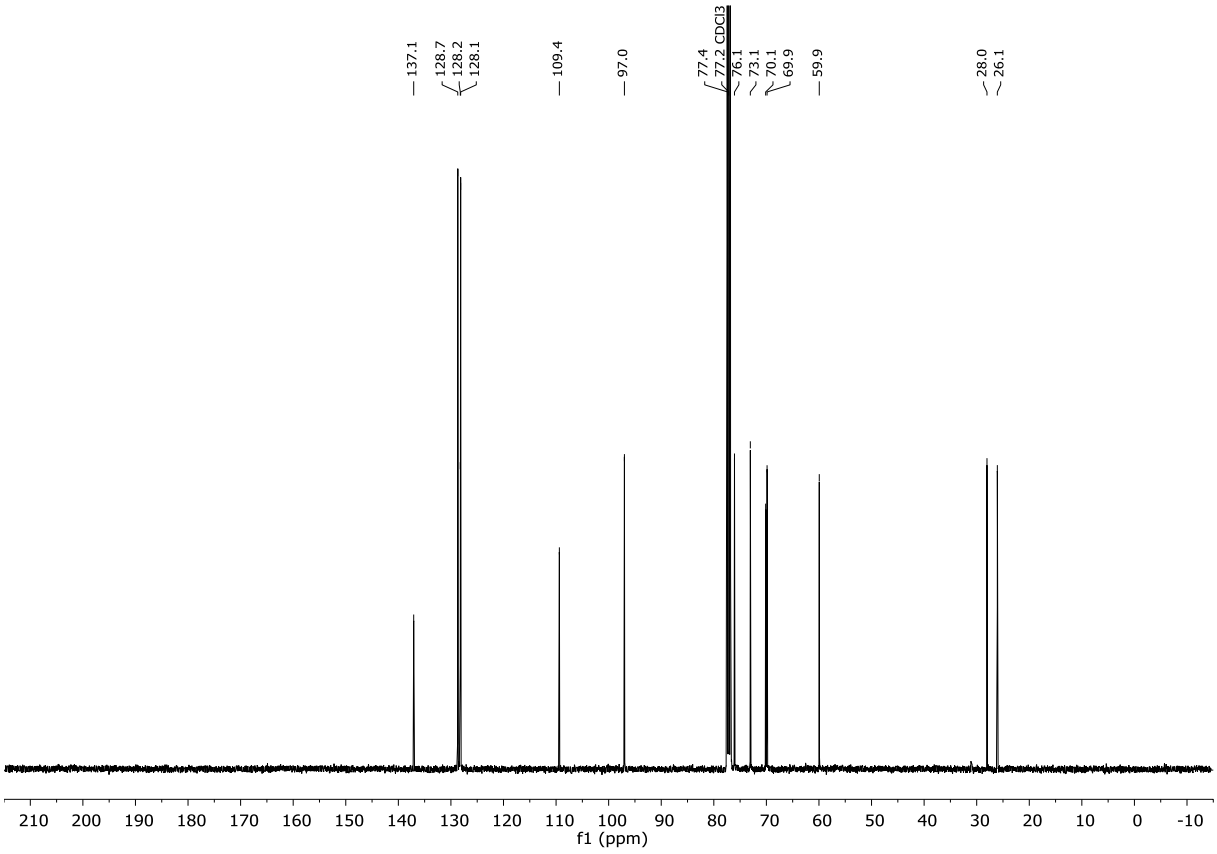
Compound **5** (HMBC, 400 MHz, Chloroform-*d*)



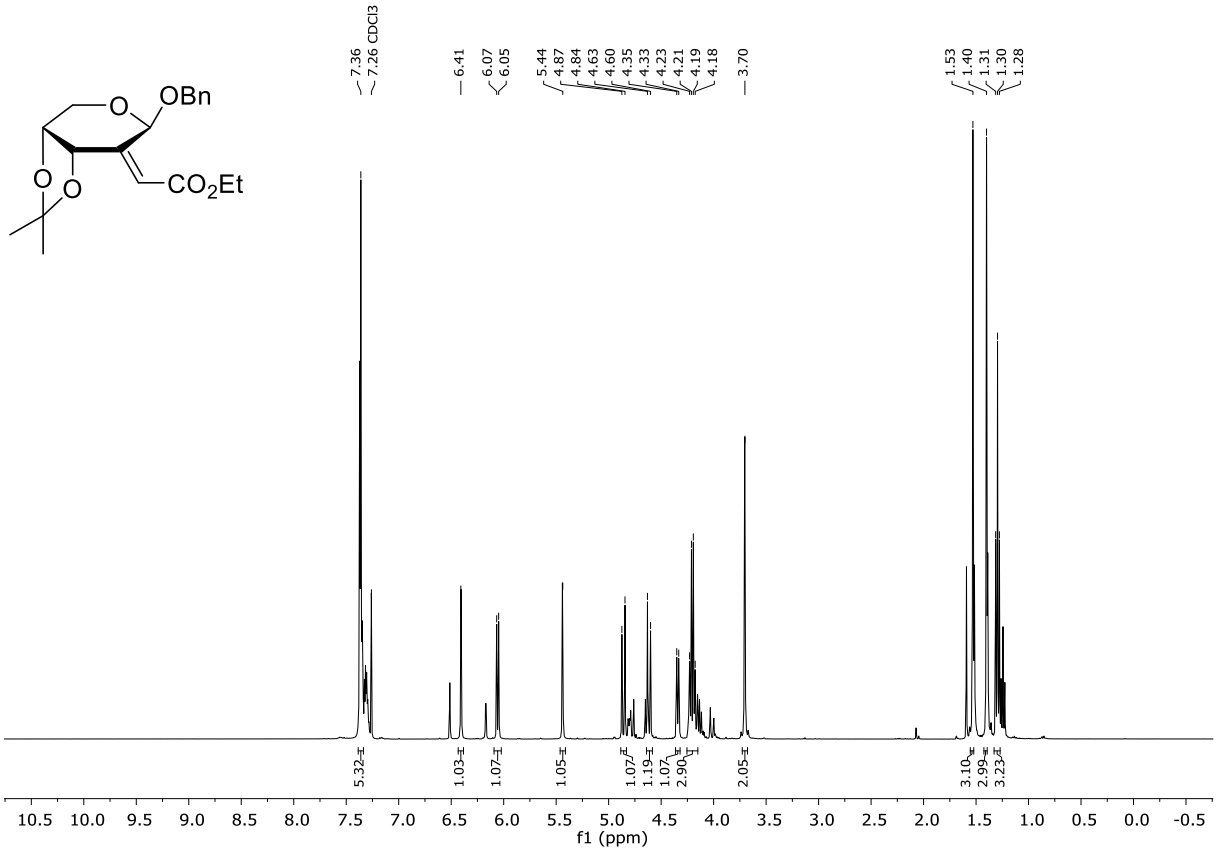
Compound **14** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



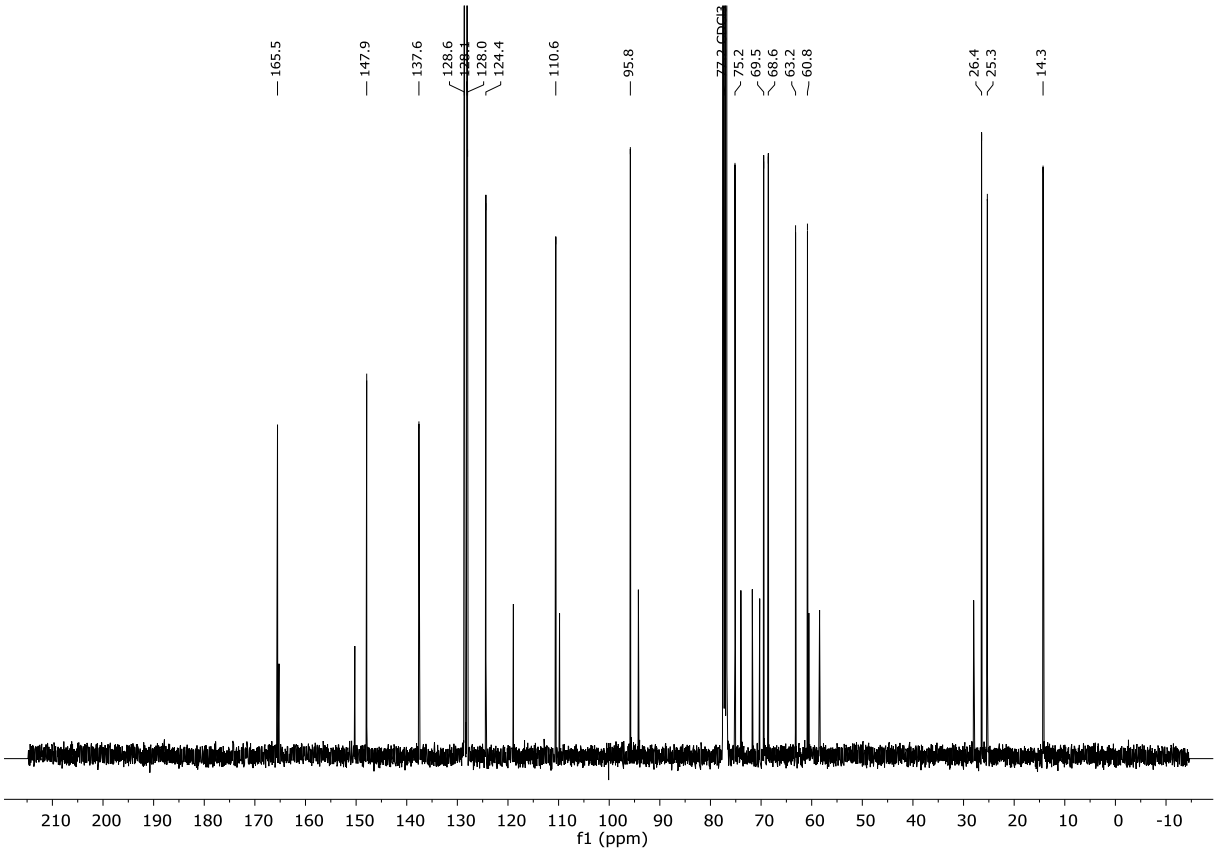
Compound **14** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



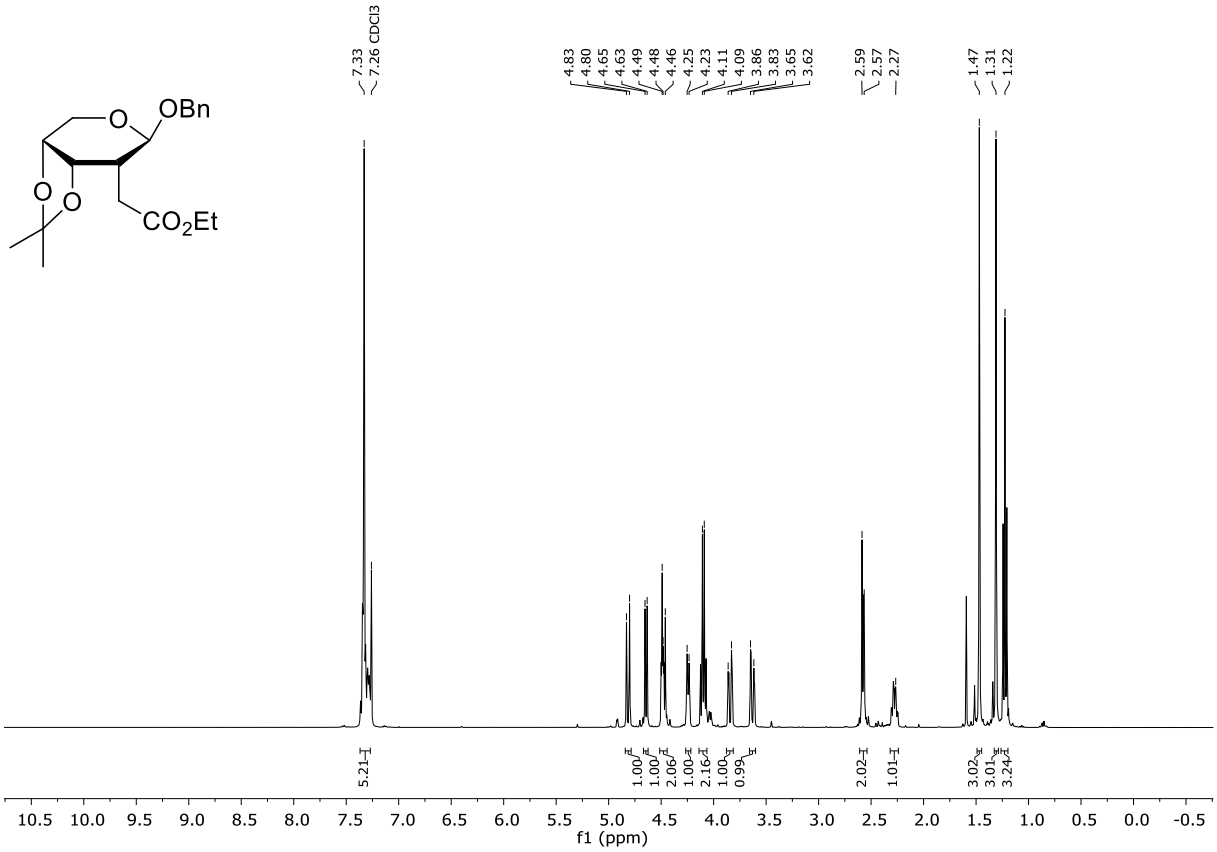
Compound **15** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



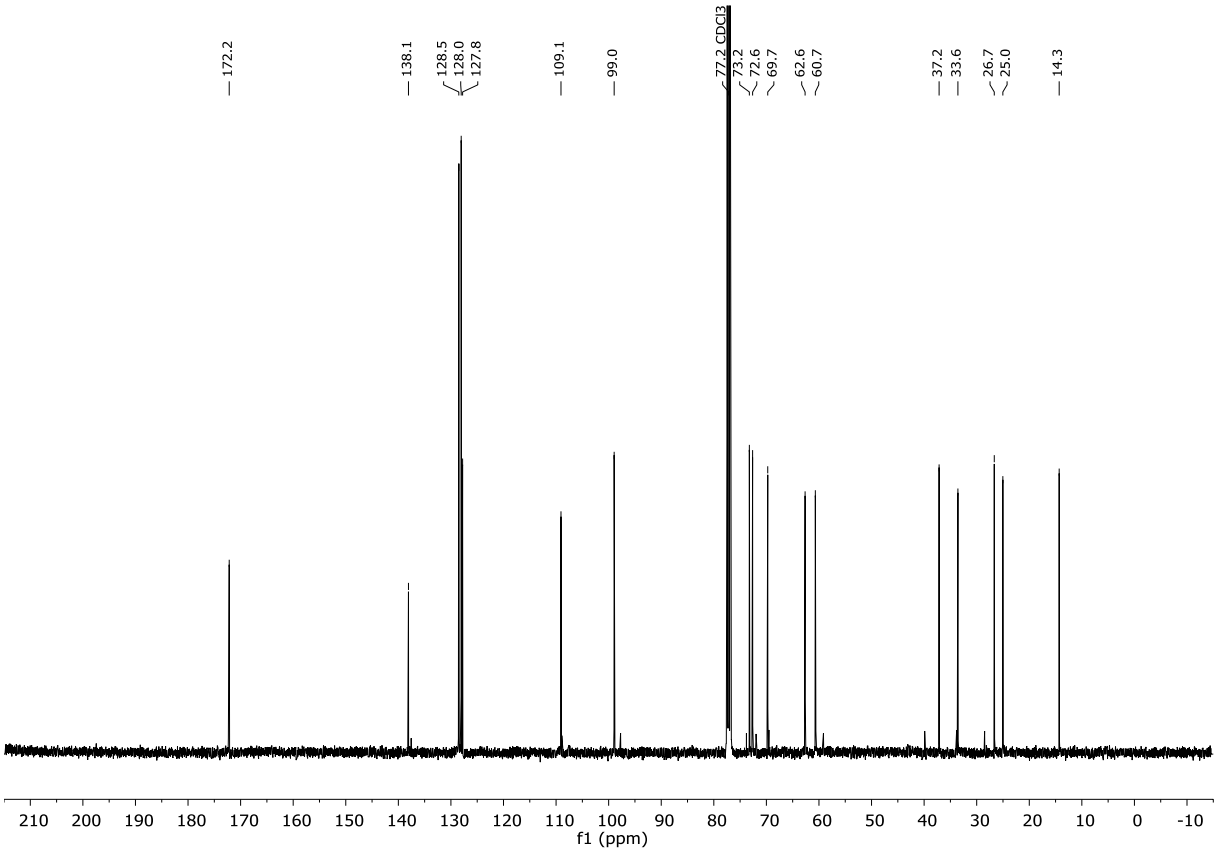
Compound **15** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



Compound **16** (^{1}H-NMR, 400 MHz, Chloroform-*d*)

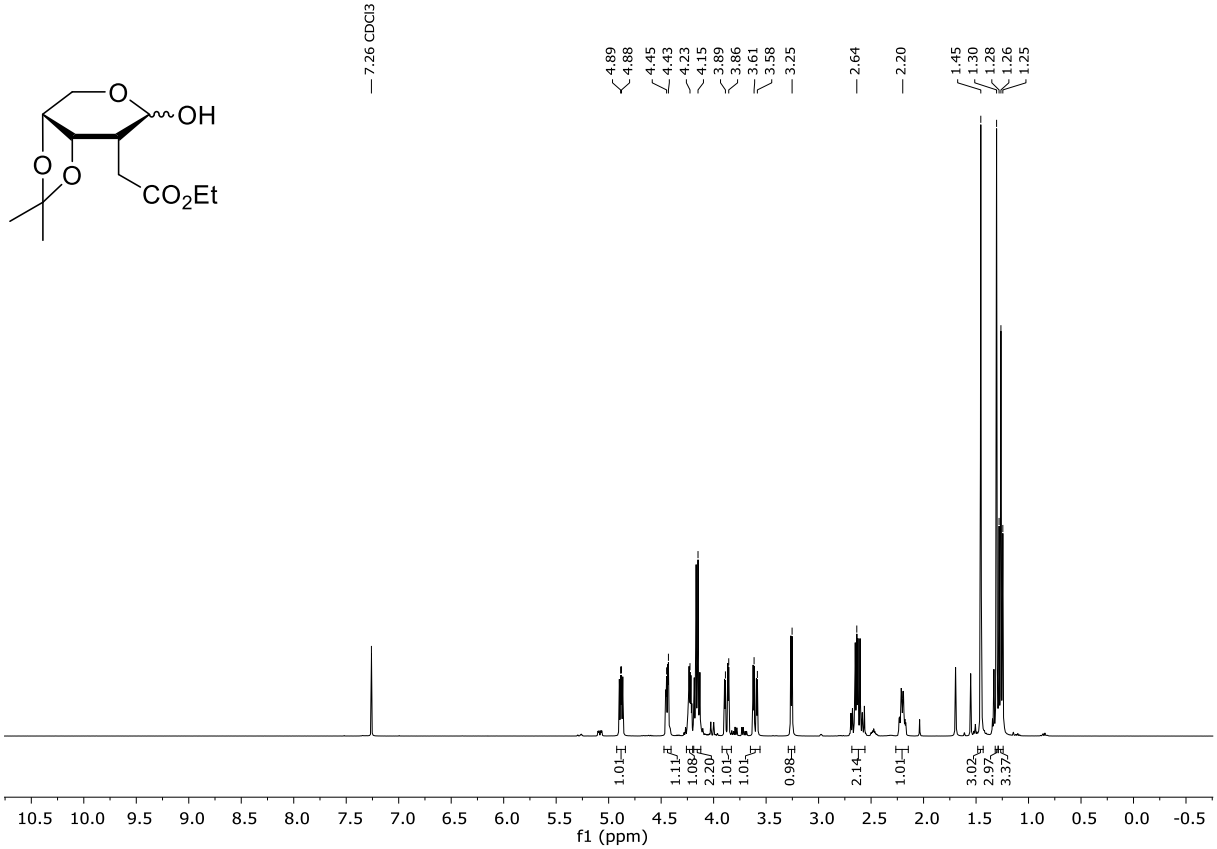


Compound **16** (^{13}C-NMR, 101 MHz, Chloroform-*d*)

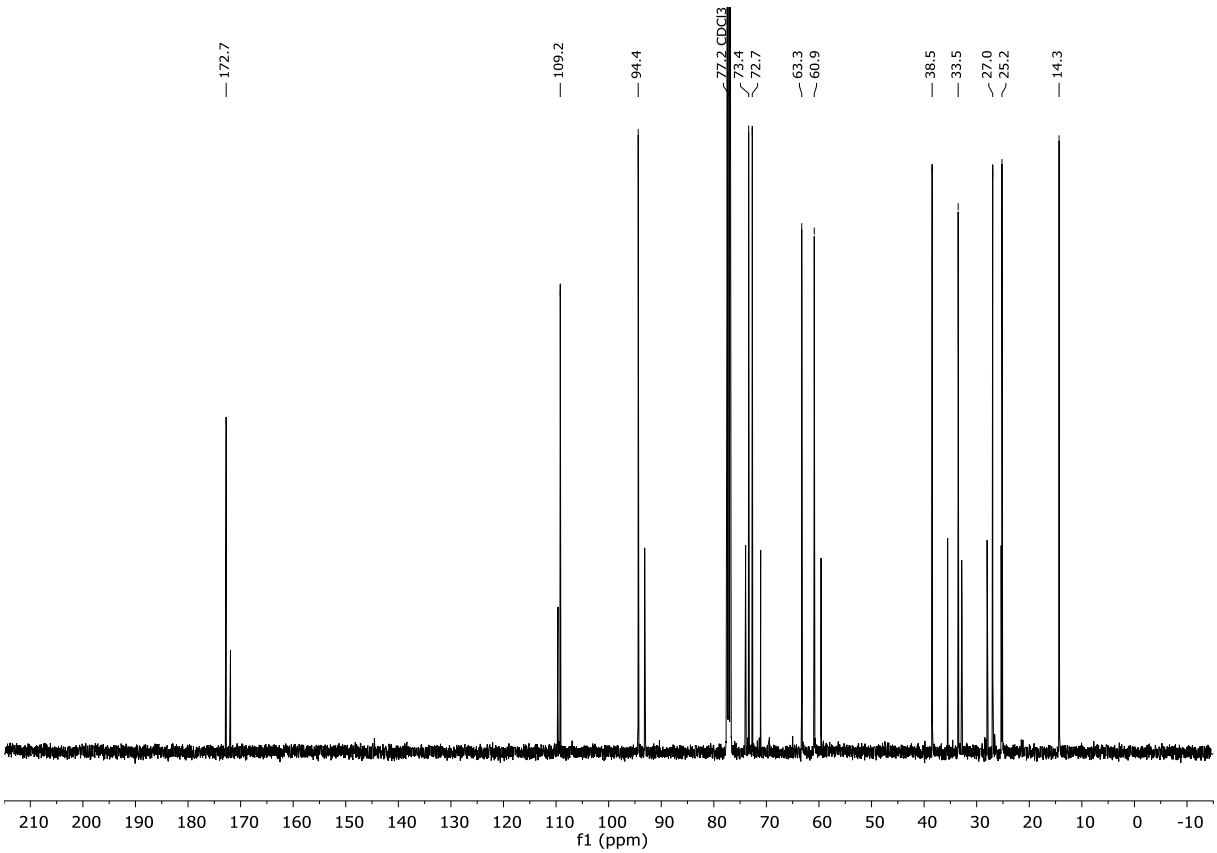


α-/β-anomer:

Compound **17a** (^{1}H-NMR, 400 MHz, Chloroform-*d*)

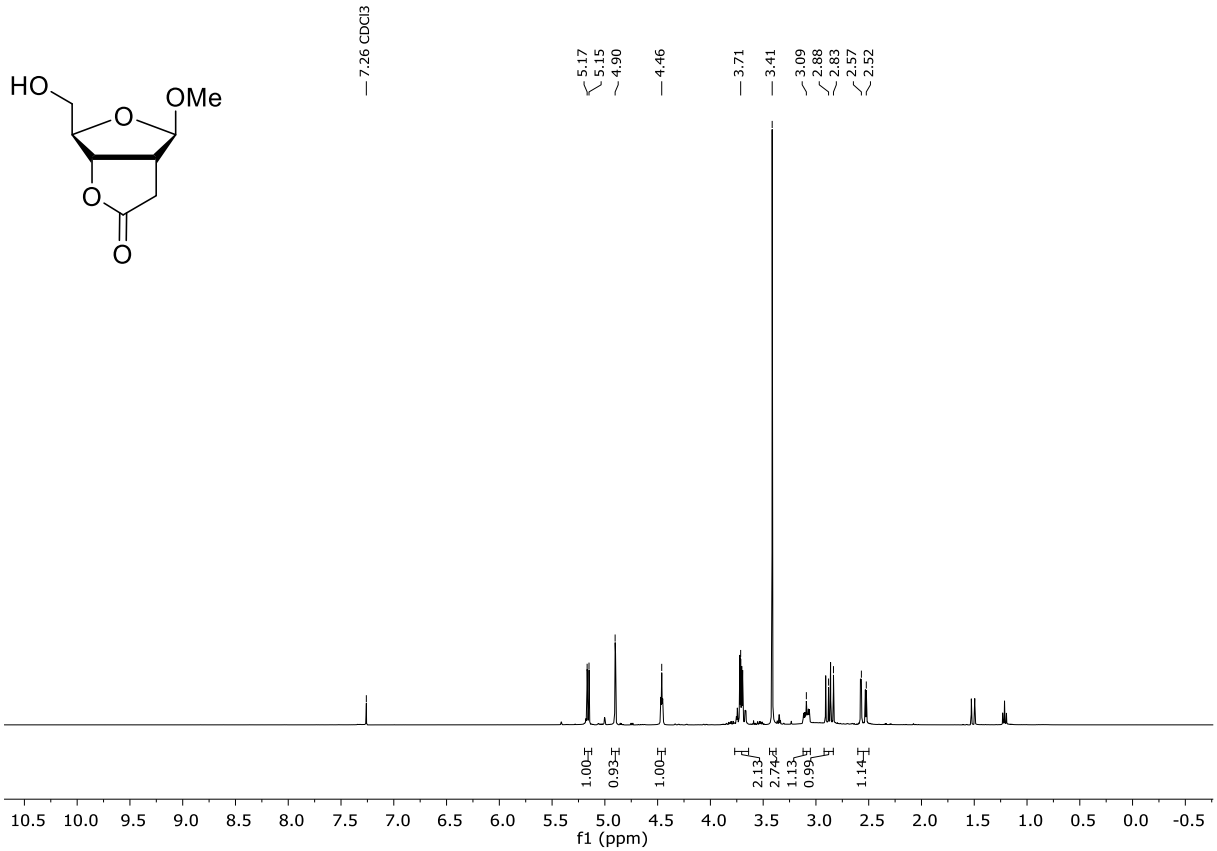


Compound **17a** (^{13}C-NMR, 101 MHz, Chloroform-*d*)

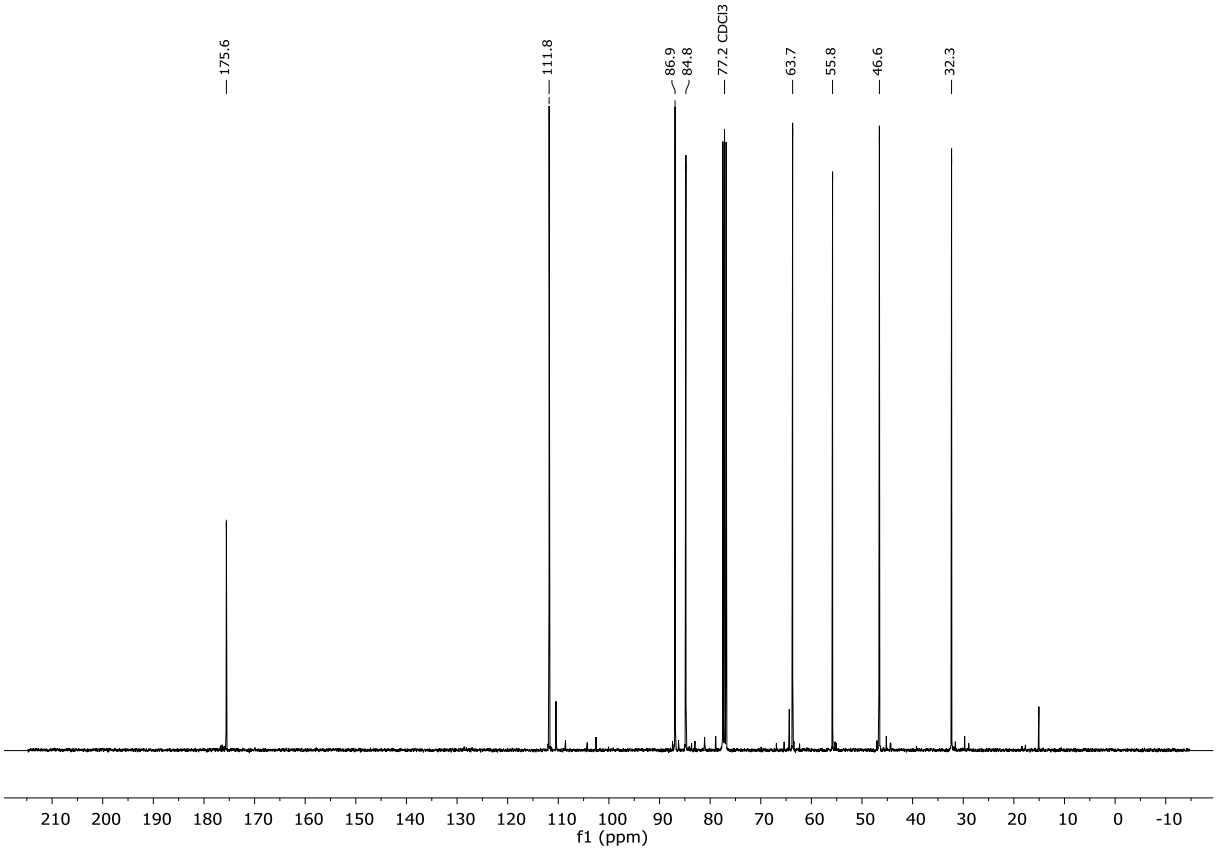


β-anomer:

Compound **17** (^{1}H-NMR, 400 MHz, Chloroform-*d*)

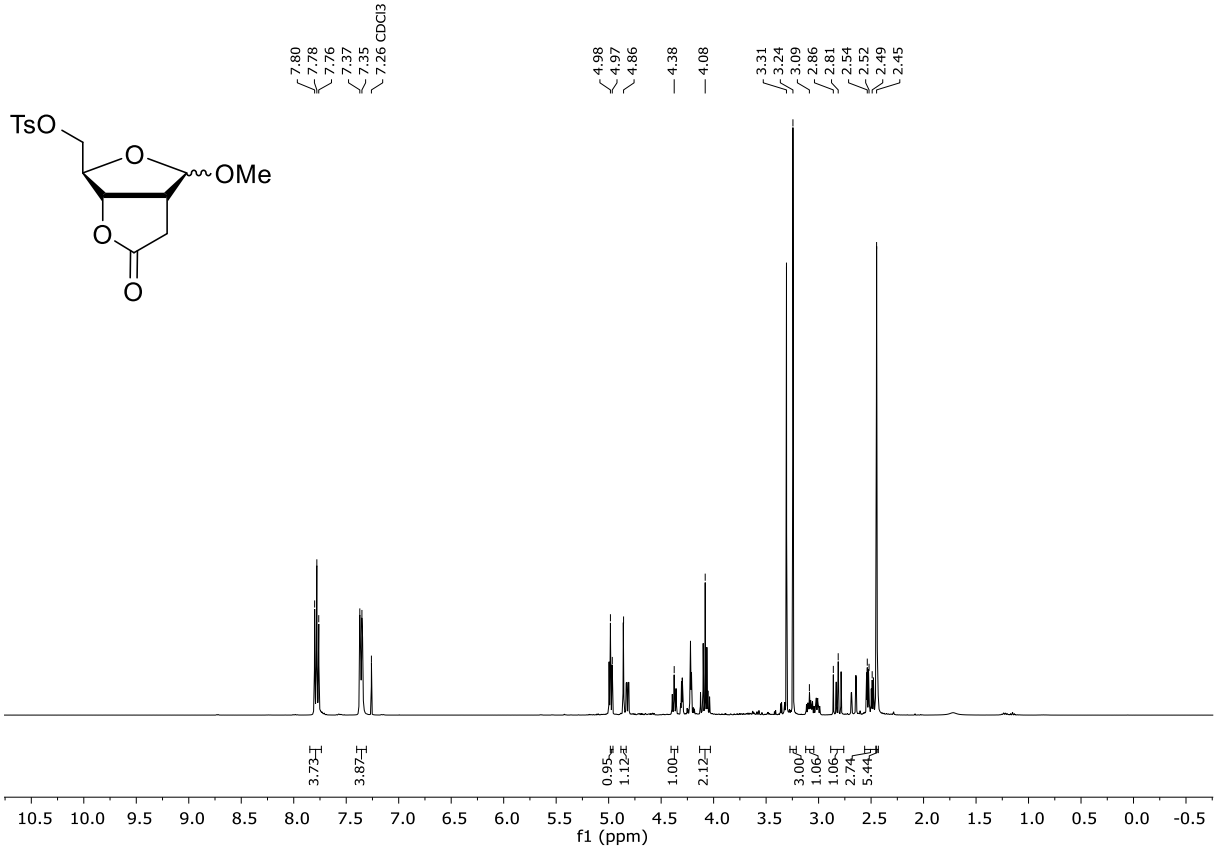


Compound **17** (^{13}C-NMR, 101 MHz, Chloroform-*d*)

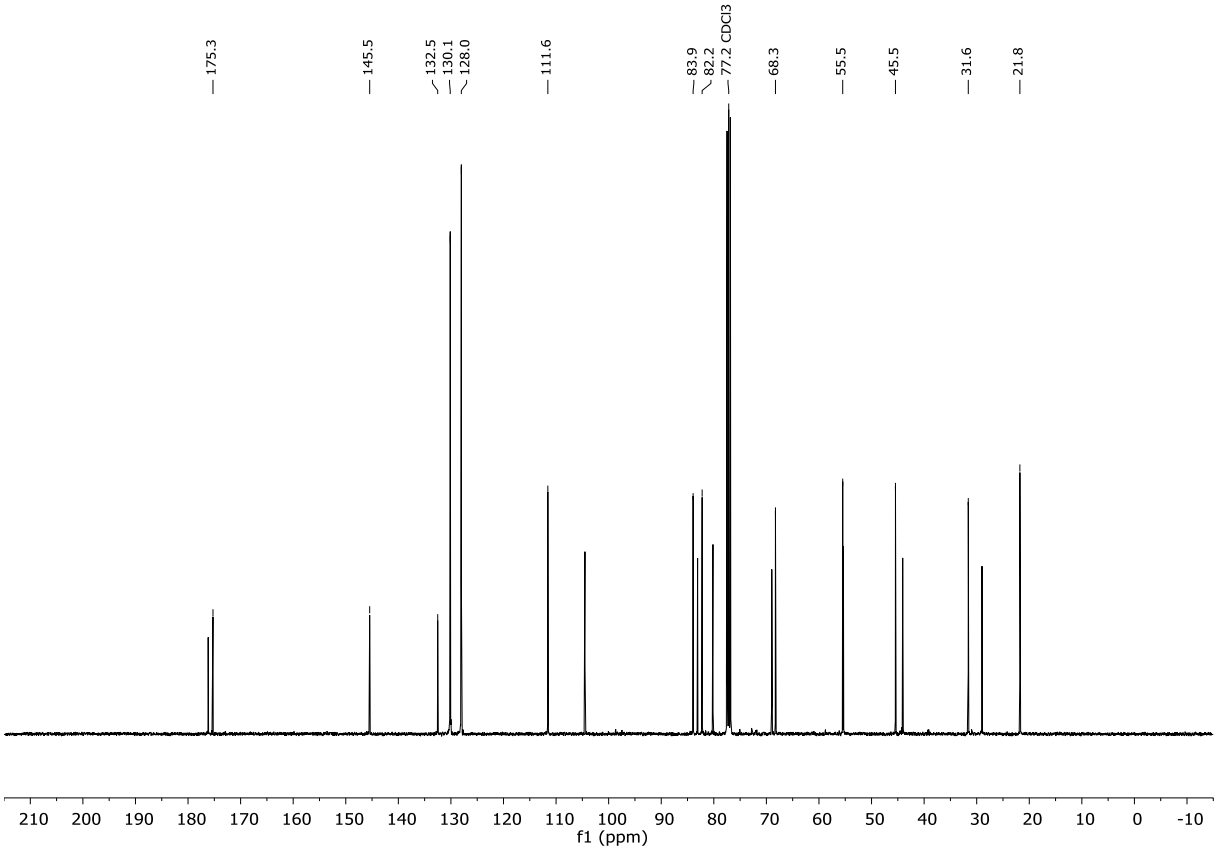


α-/β-anomer:

Compound **18** (^{1}H-NMR, 400 MHz, Chloroform-*d*)

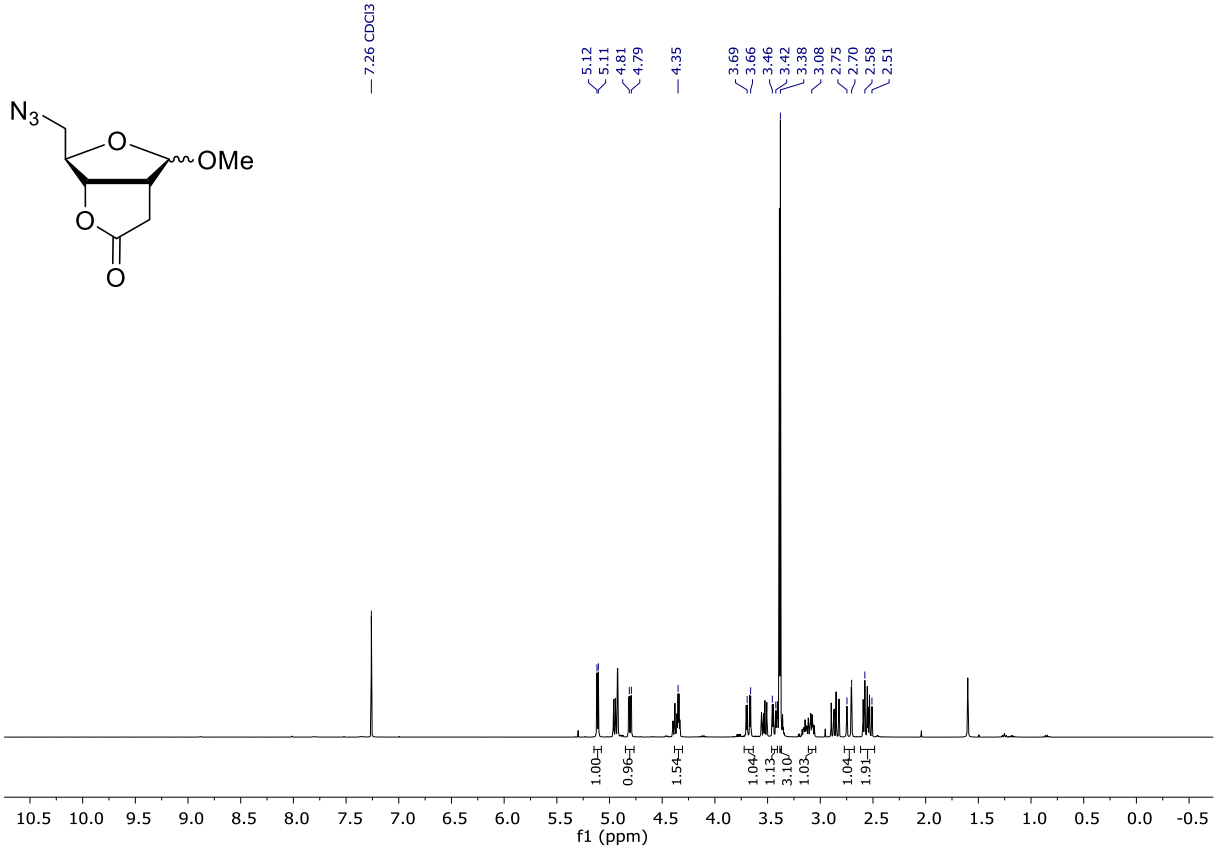


Compound **18** (^{13}C-NMR, 101 MHz, Chloroform-*d*)

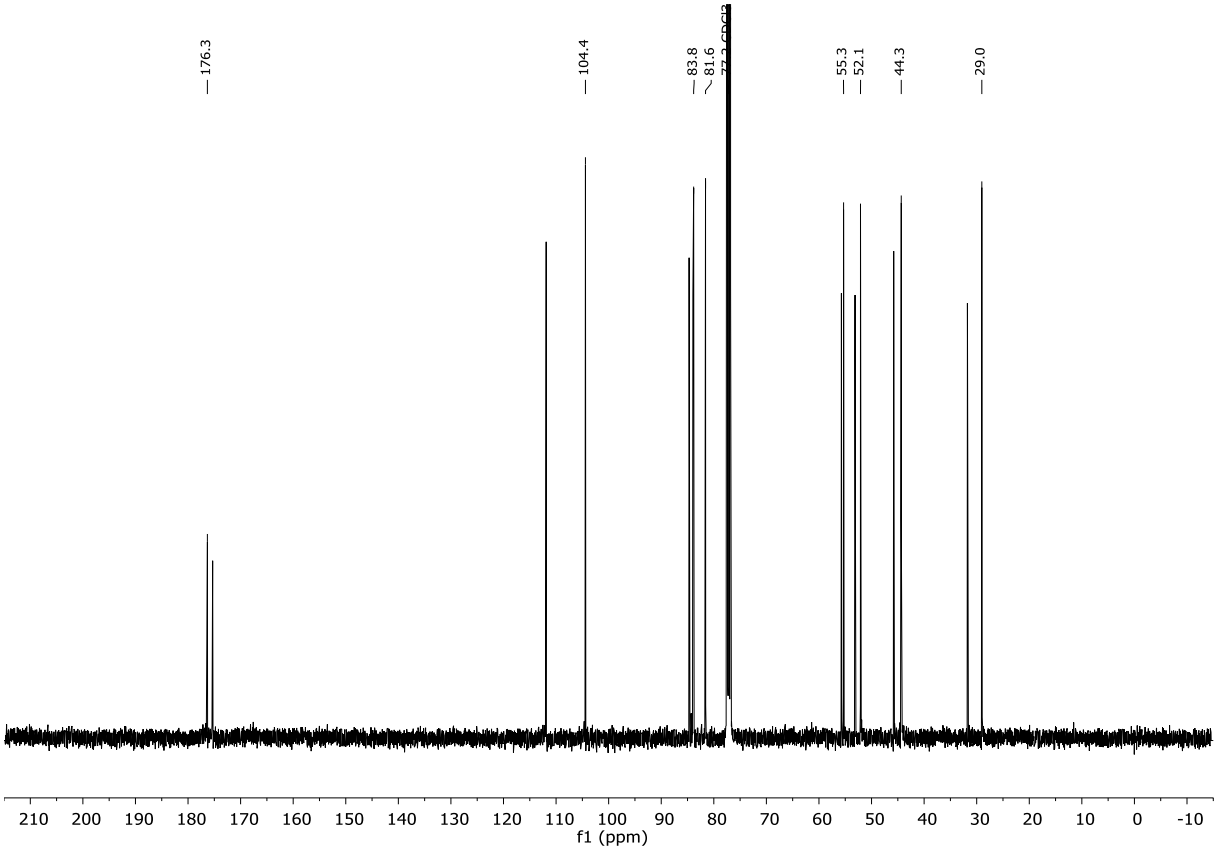


α-/β-anomer:

Compound **19** (^{1}H-NMR, 400 MHz, Chloroform-*d*)

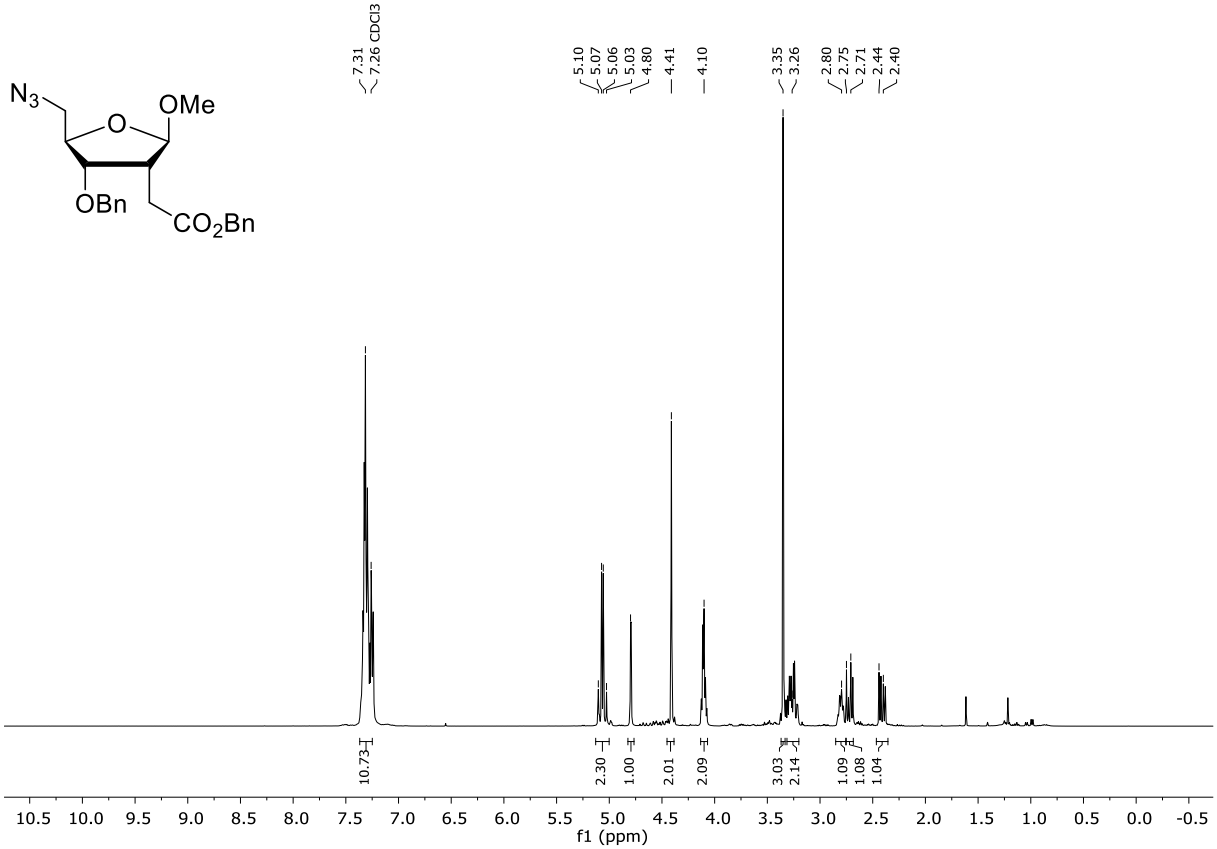


Compound **19** (^{13}C-NMR, 101 MHz, Chloroform-*d*)

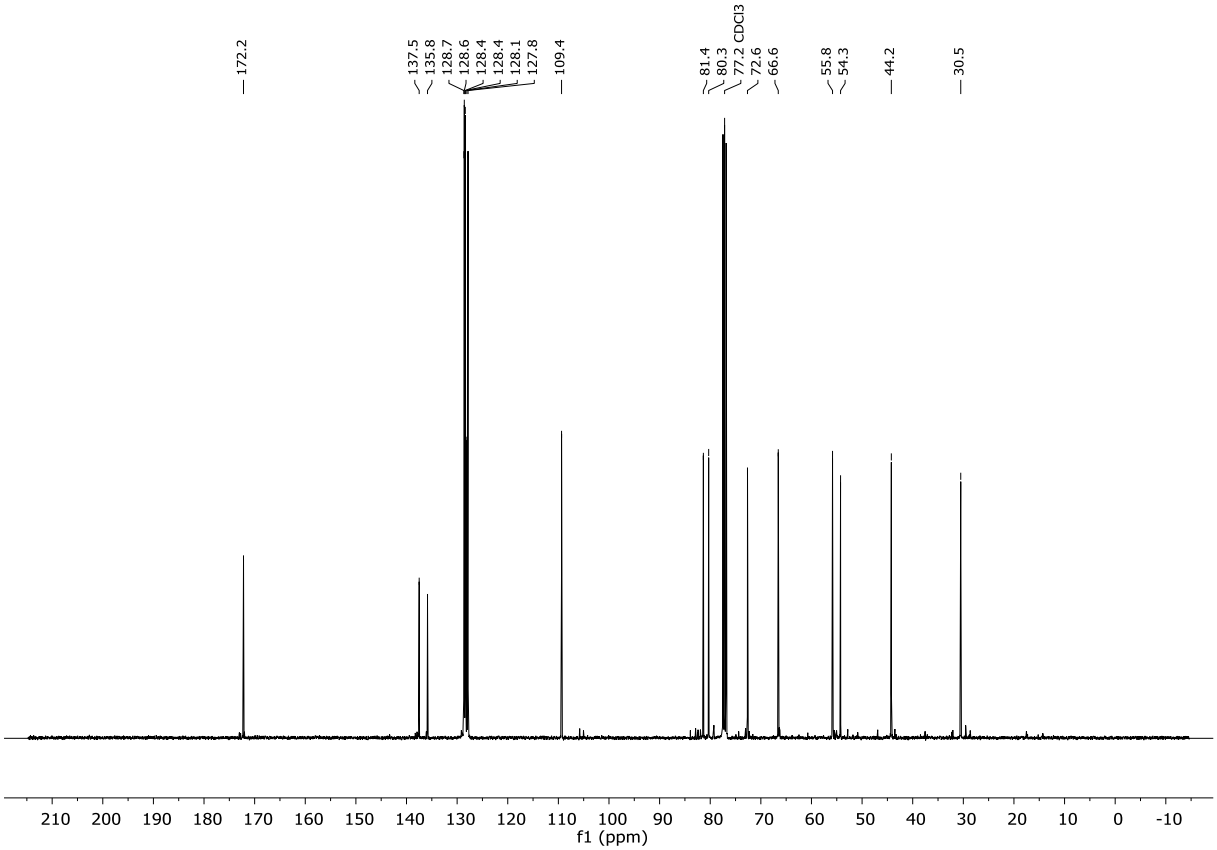


β-anomer:

Compound **20** (^{1}H-NMR, 400 MHz, Chloroform-*d*)

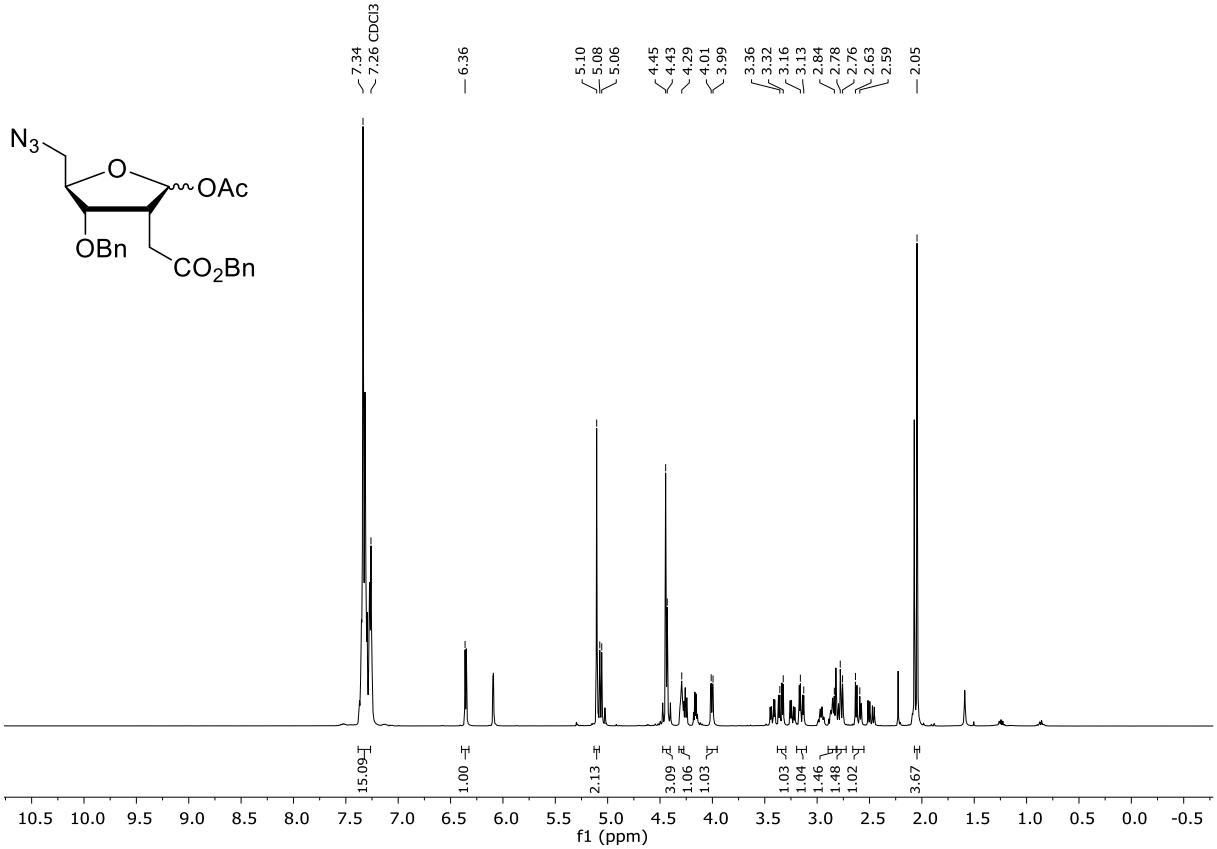


Compound **20** (^{13}C-NMR, 101 MHz, Chloroform-*d*)

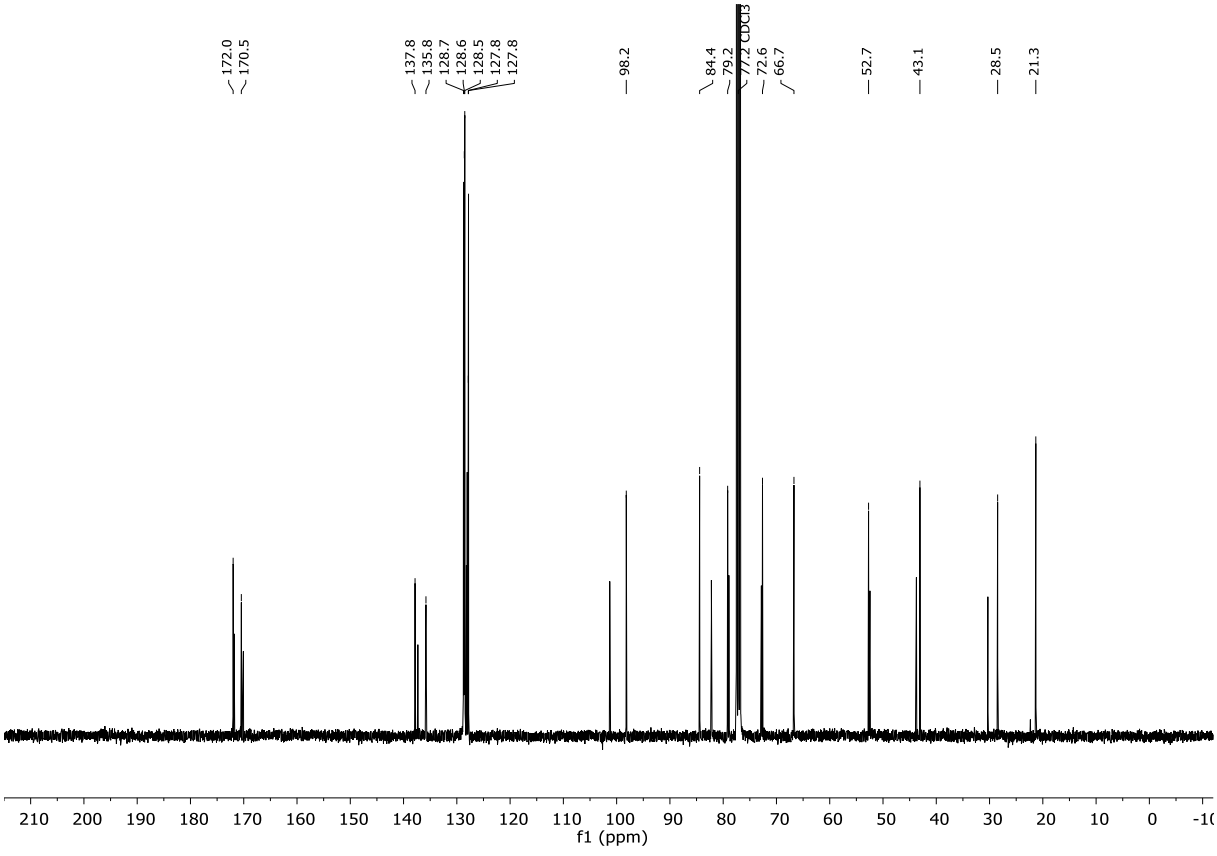


α-/β-anomer:

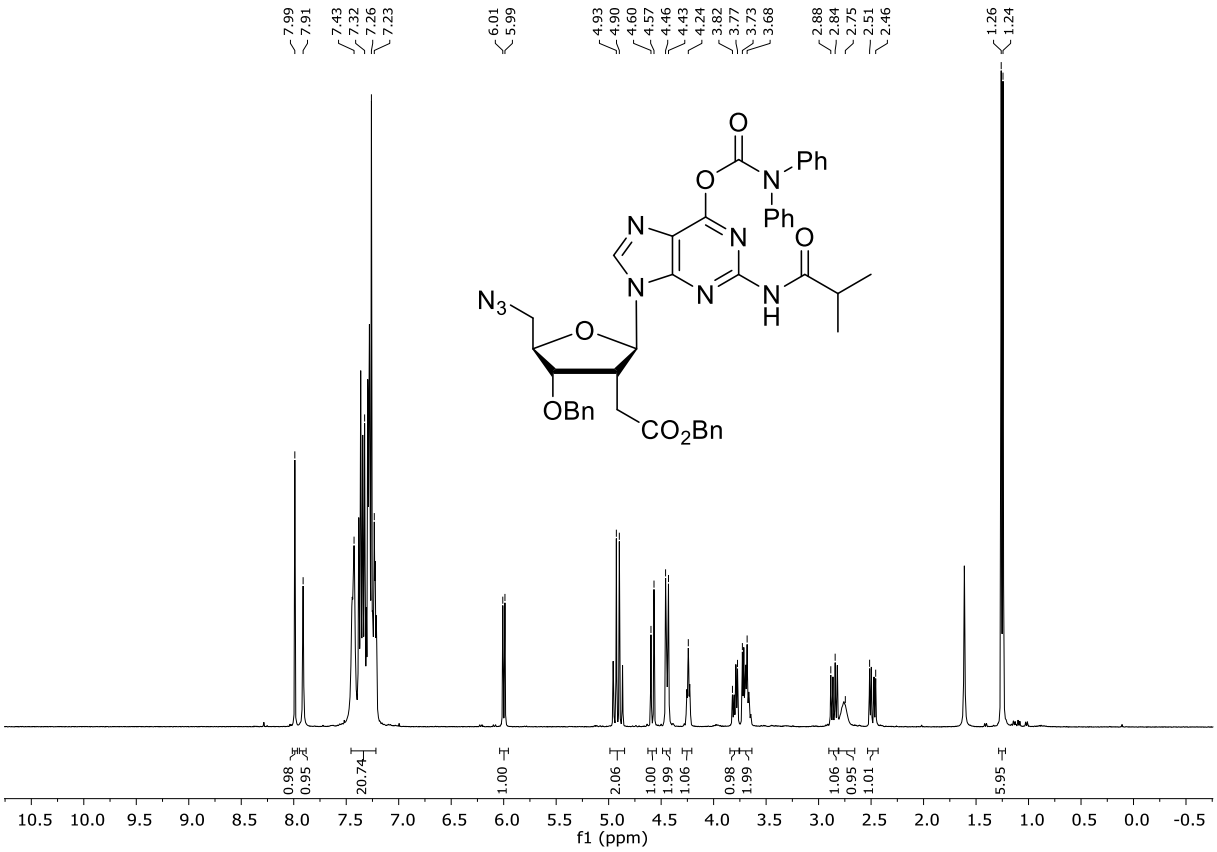
Compound **21** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



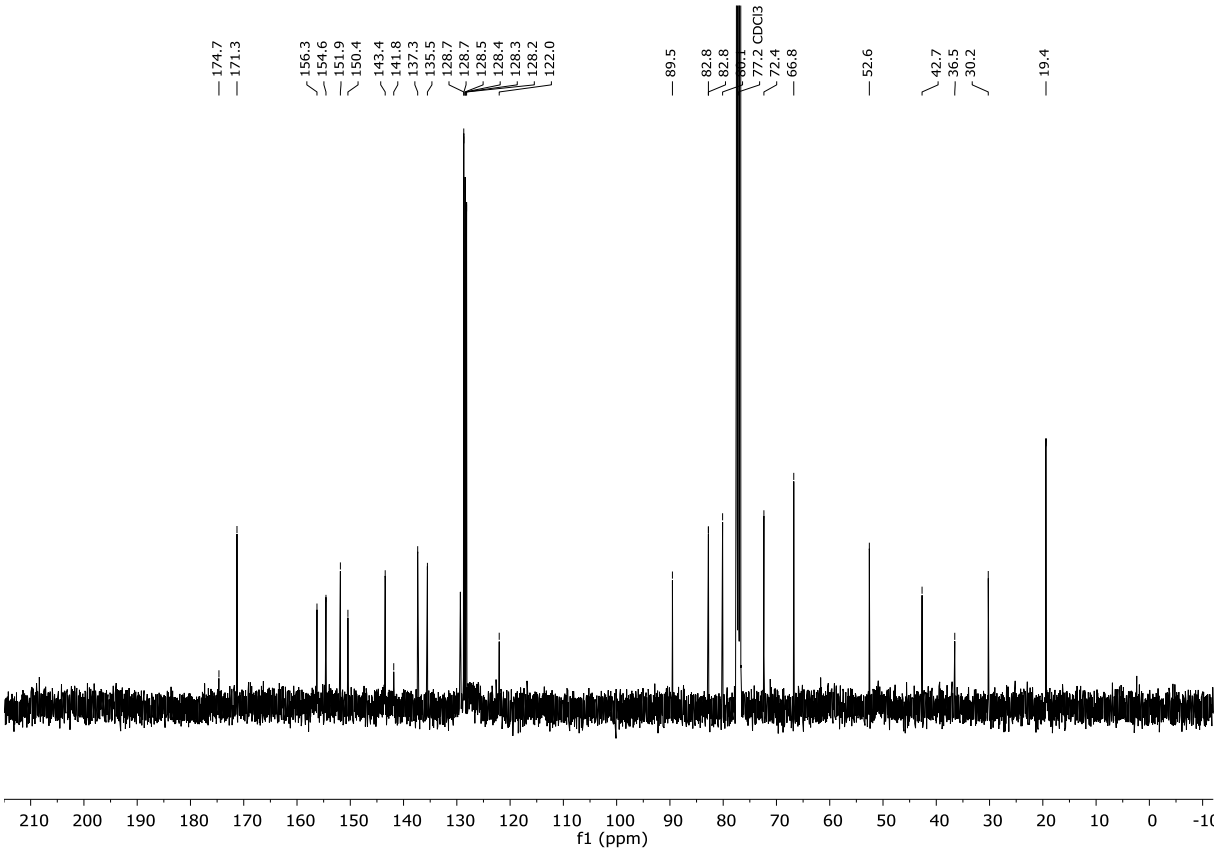
Compound **21** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



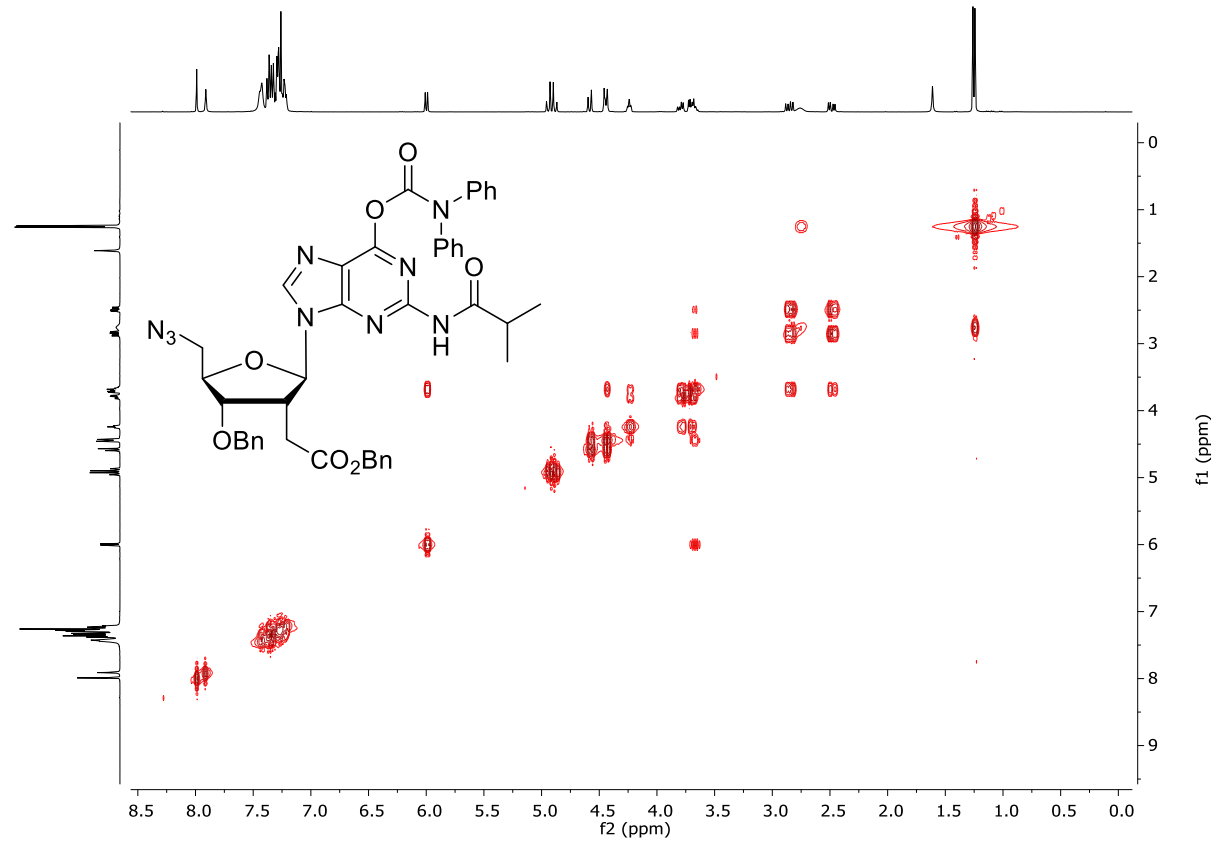
Compound **6** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



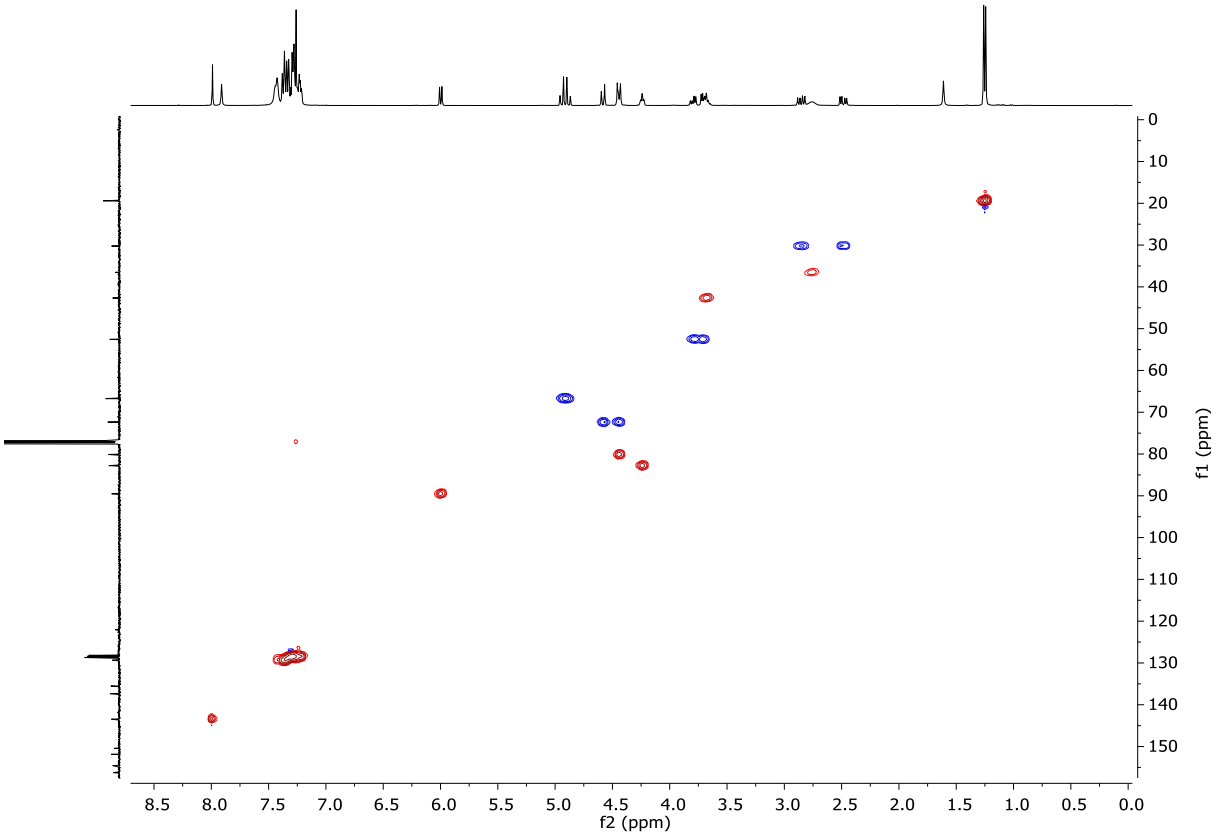
Compound **6** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



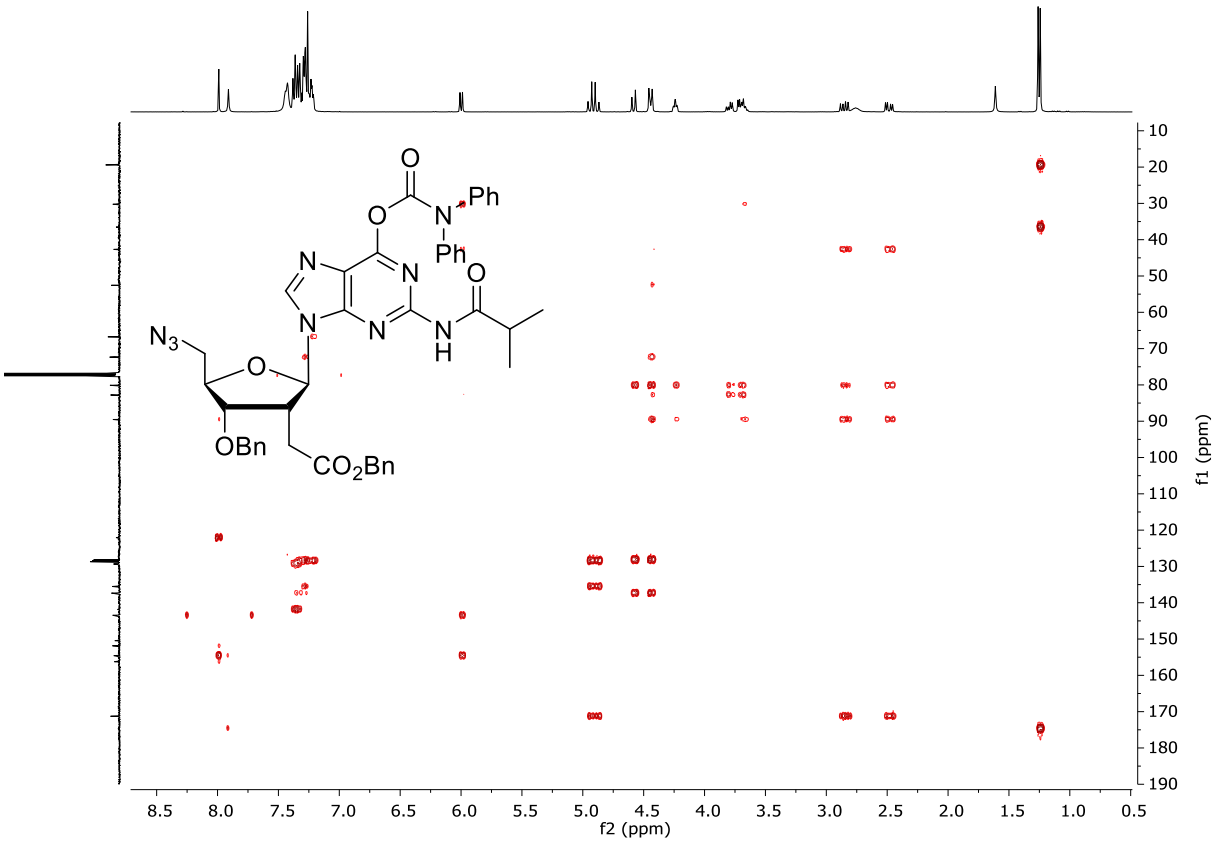
Compound **6** (COSY, 400 MHz, Chloroform-*d*)



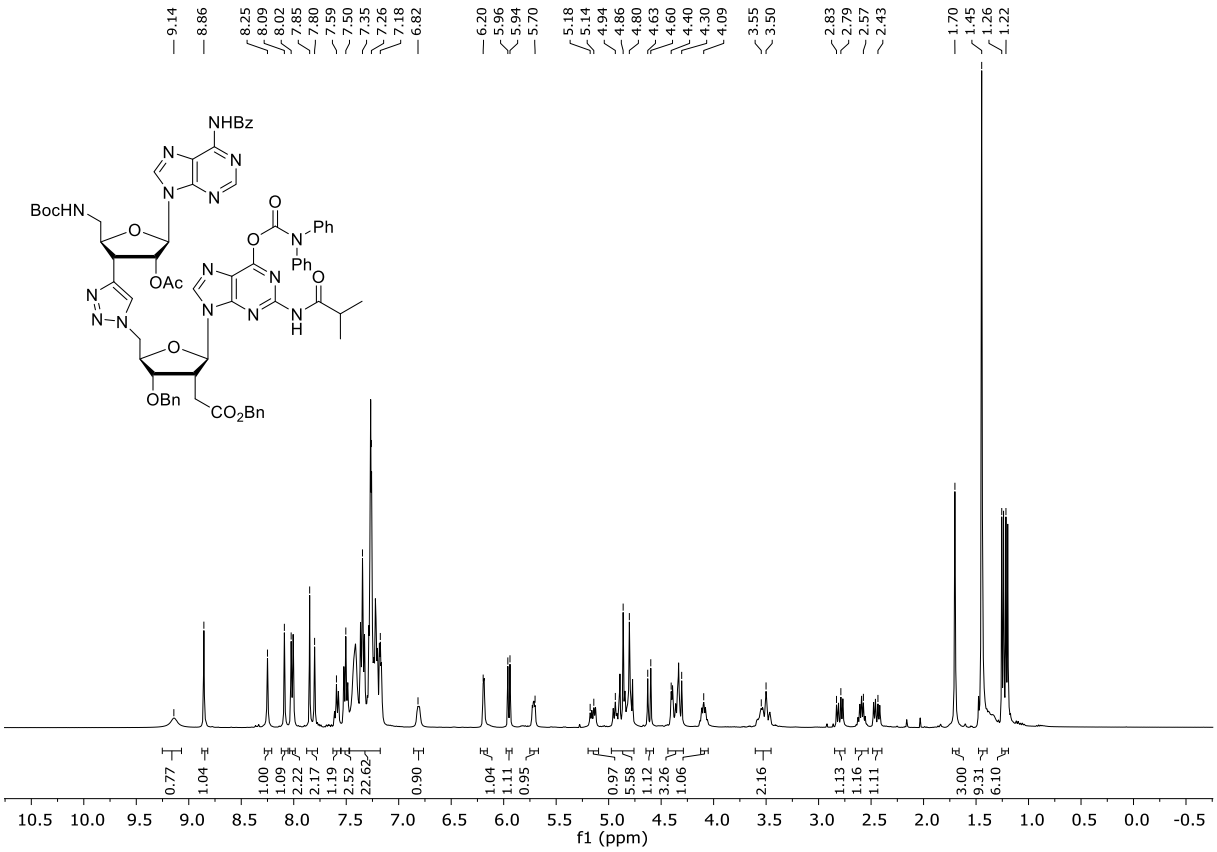
Compound **6** (HSQC, 400 MHz, Chloroform-*d*)



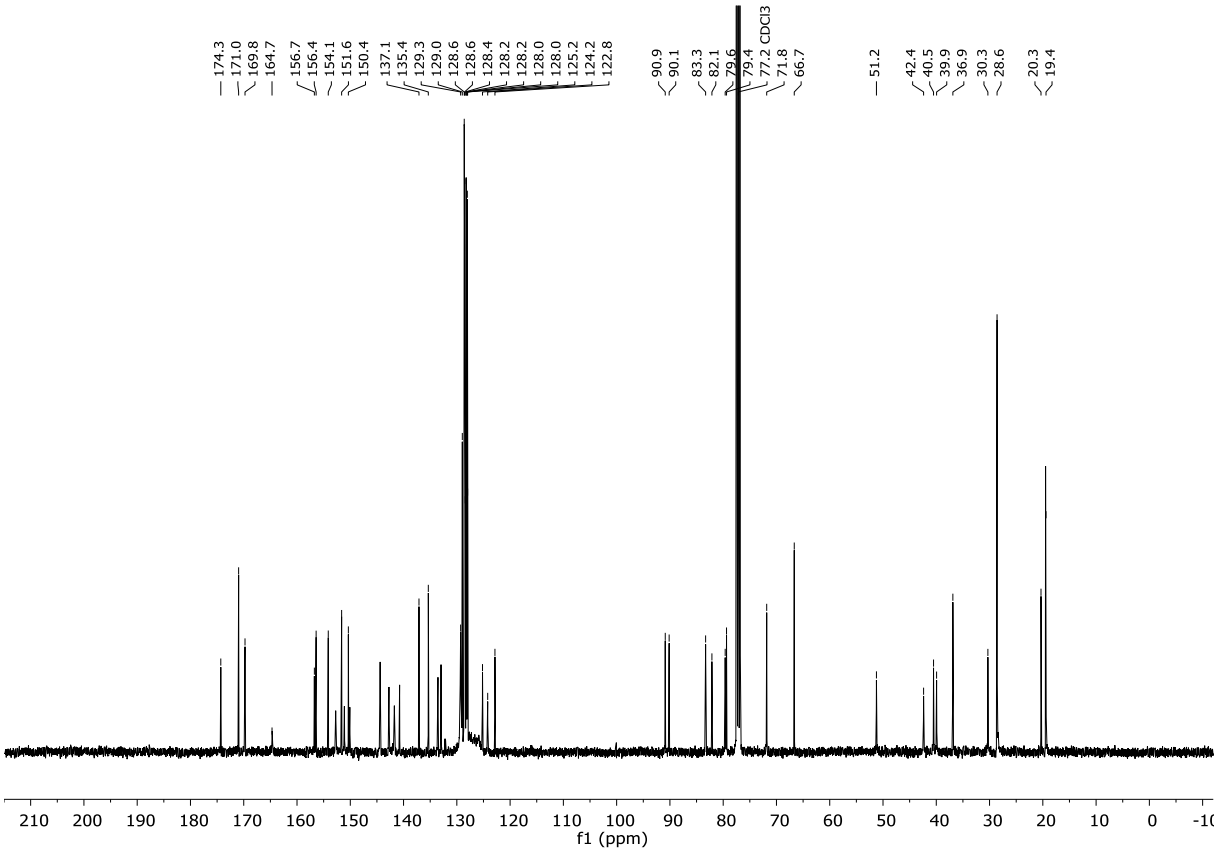
Compound **6** (HMBC, 400 MHz, Chloroform-*d*)



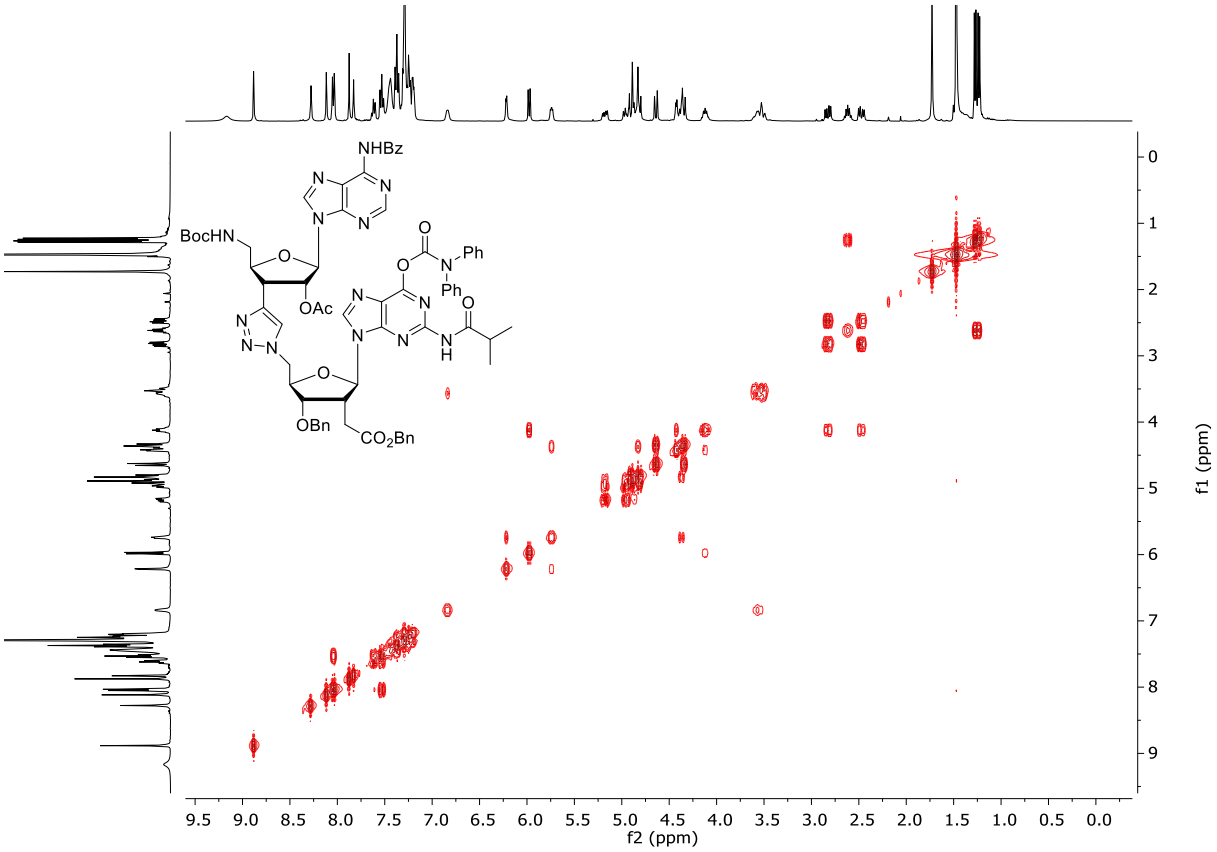
Compound **22** (^{1}H-NMR, 400 MHz, Chloroform-*d*)



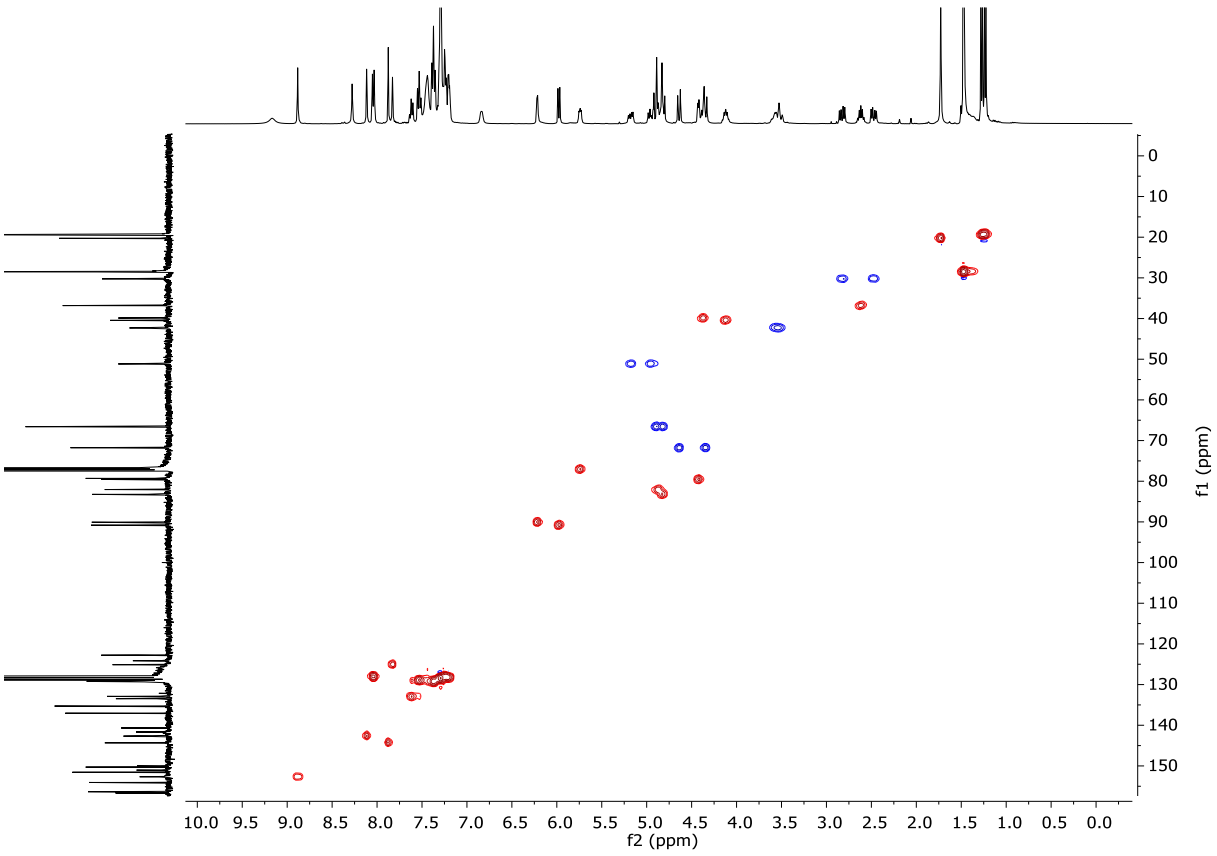
Compound **22** (^{13}C-NMR, 101 MHz, Chloroform-*d*)



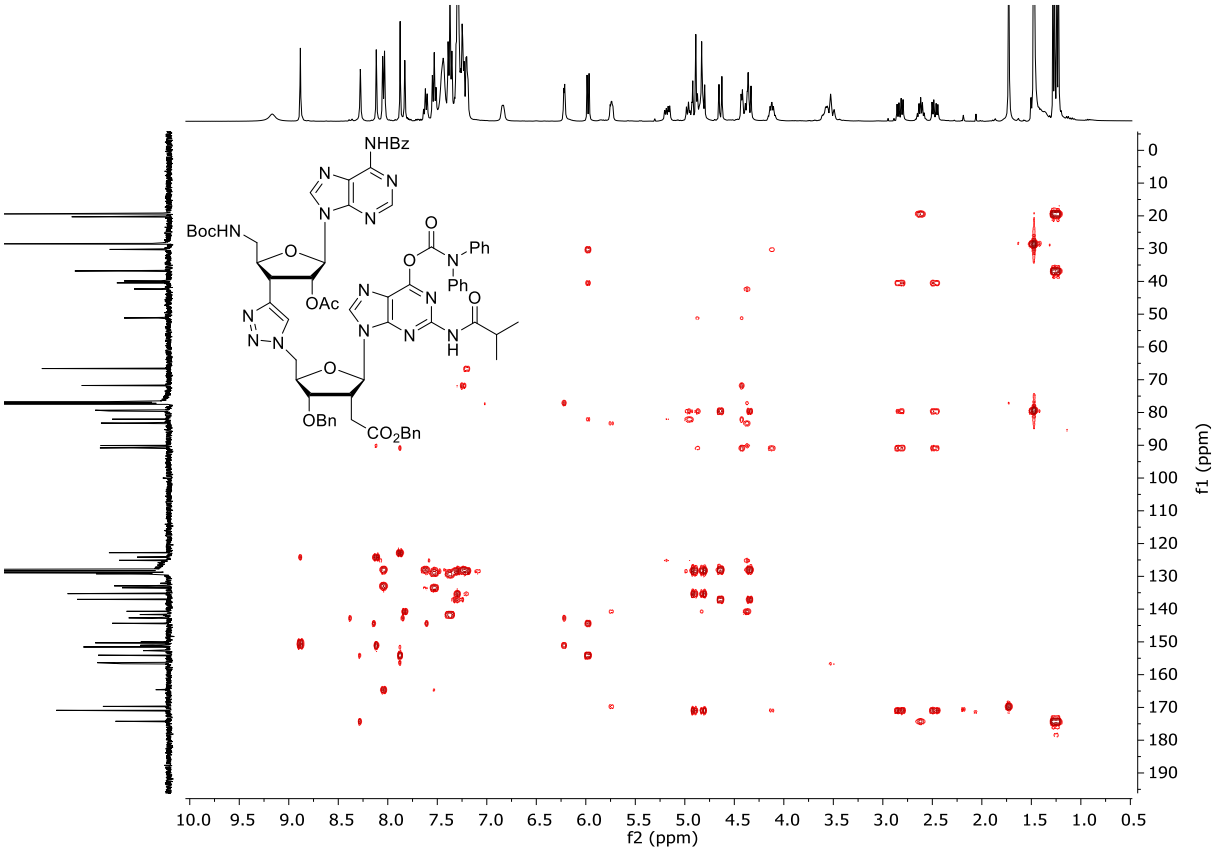
Compound **22** (COSY, 400 MHz, Chloroform-*d*)



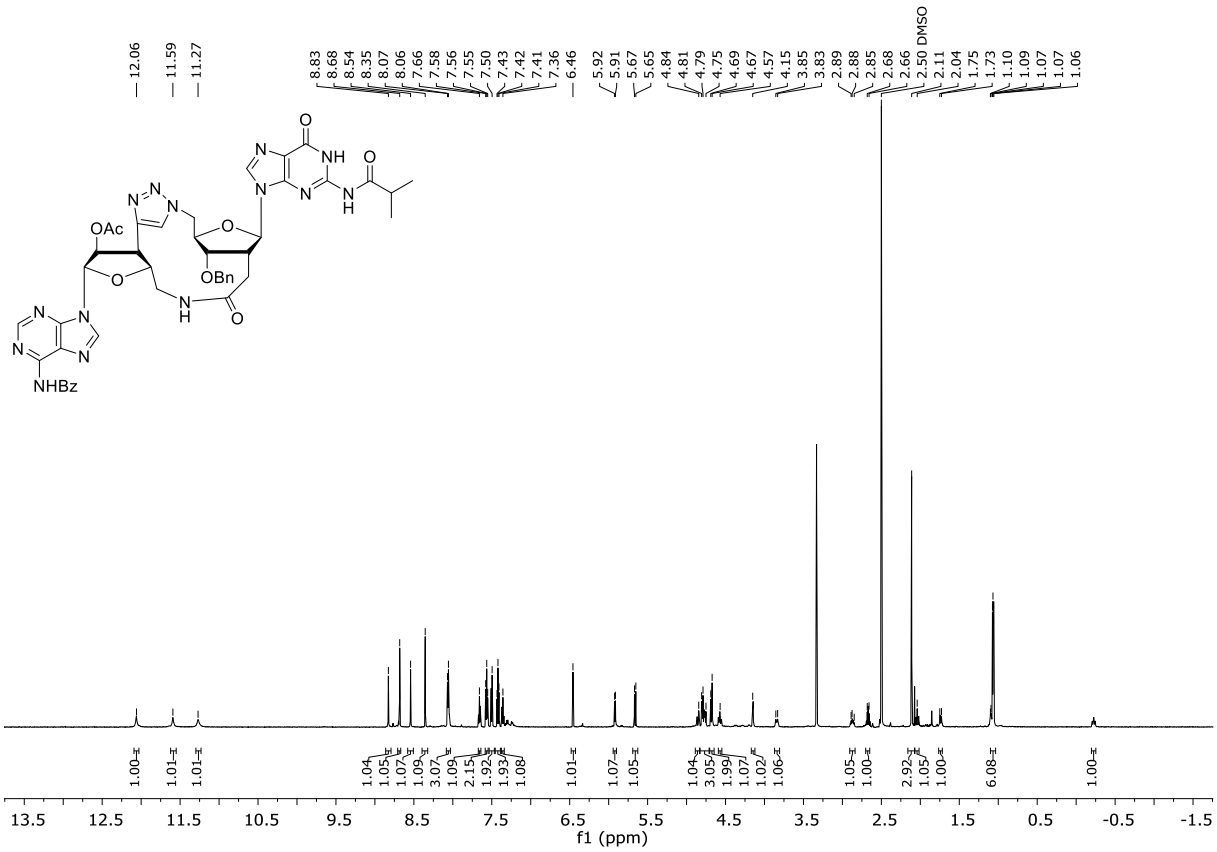
Compound **22** (HSQC, 400 MHz, Chloroform-*d*)



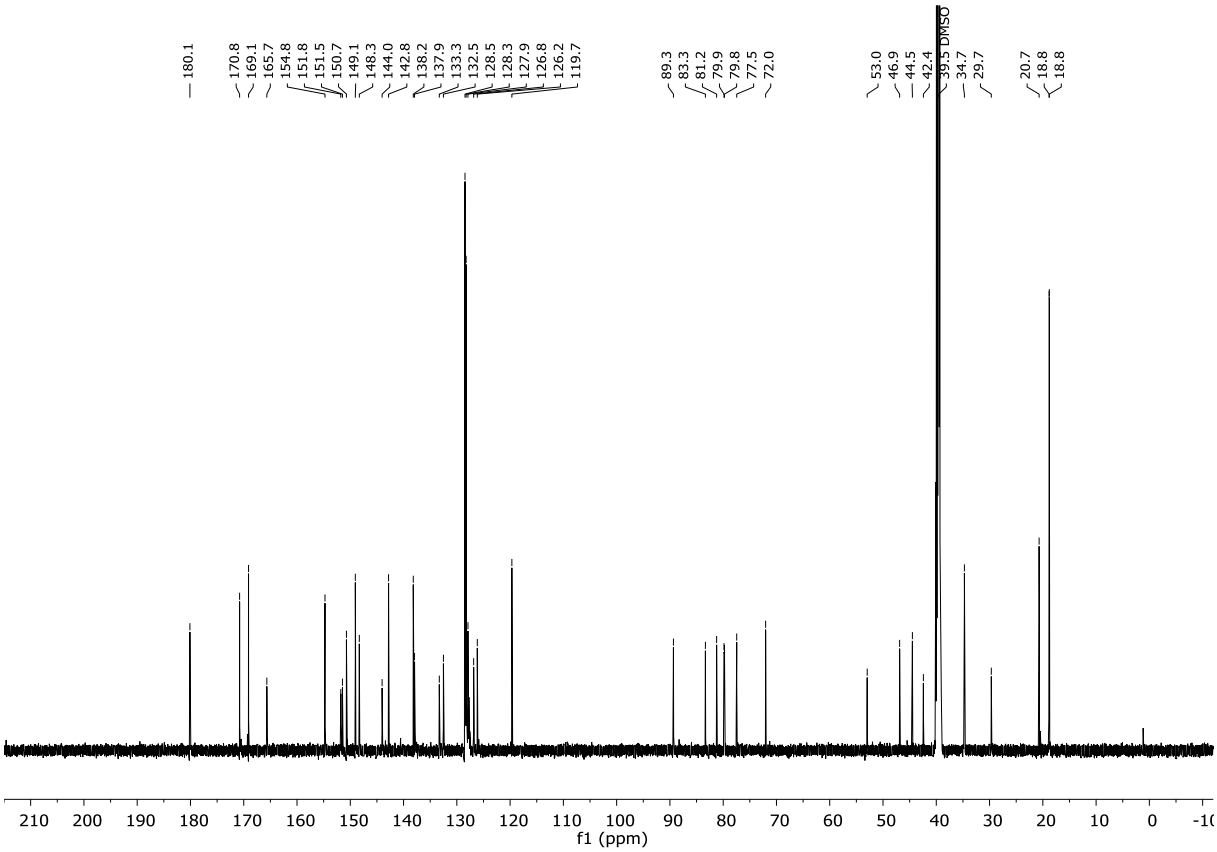
Compound **22** (HMBC, 400 MHz, Chloroform-*d*)



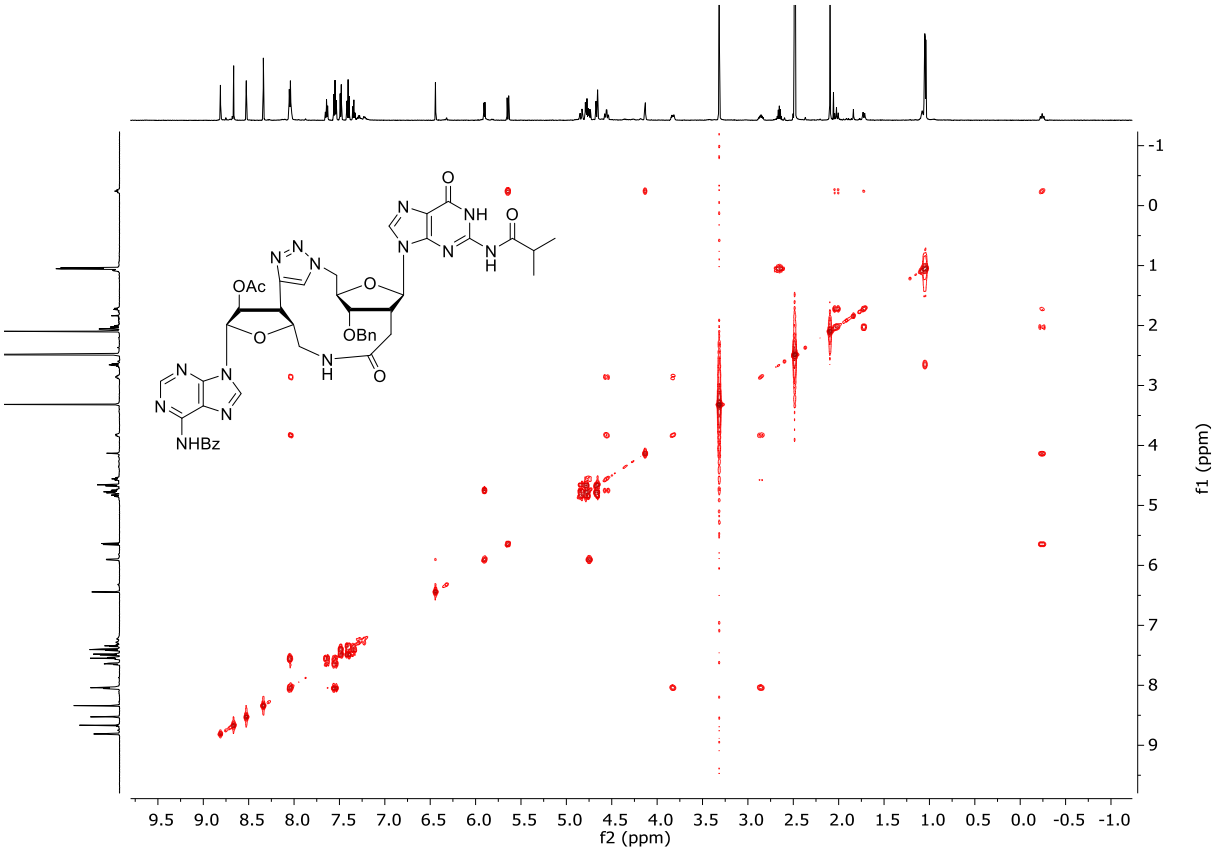
Compound **24** (^{1}H-NMR, 600 MHz, DMSO-d_6)



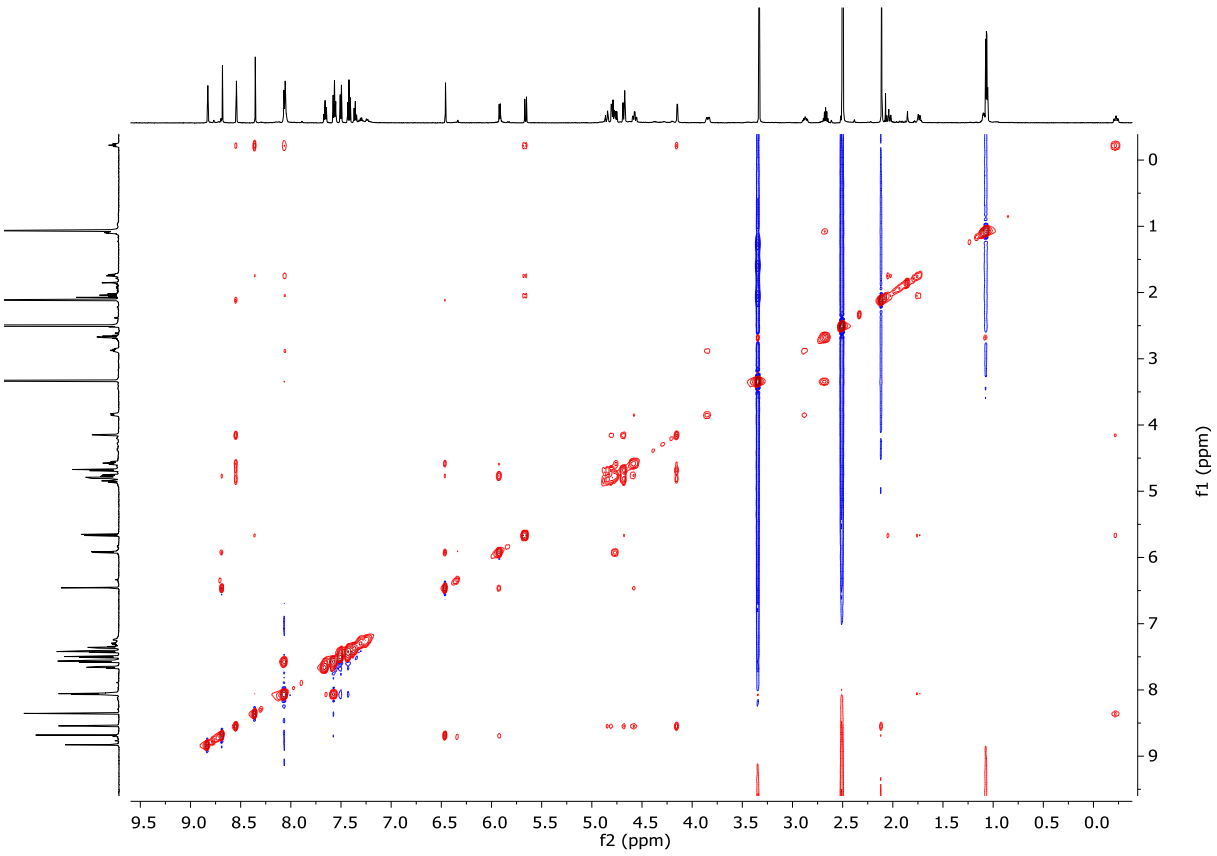
Compound **24** (^{13}C-NMR, 151 MHz, DMSO-d_6)



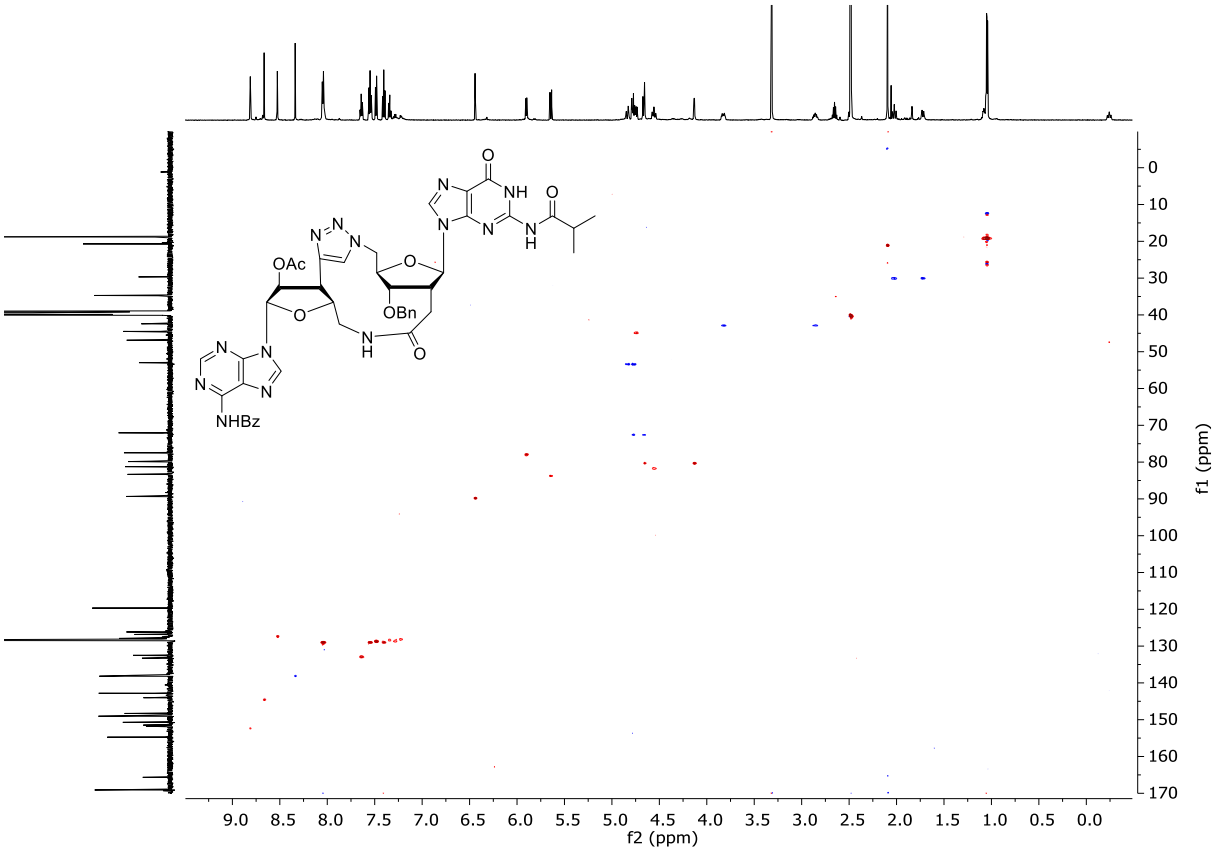
Compound **24** (COSY, 600 MHz, DMSO-d_6)



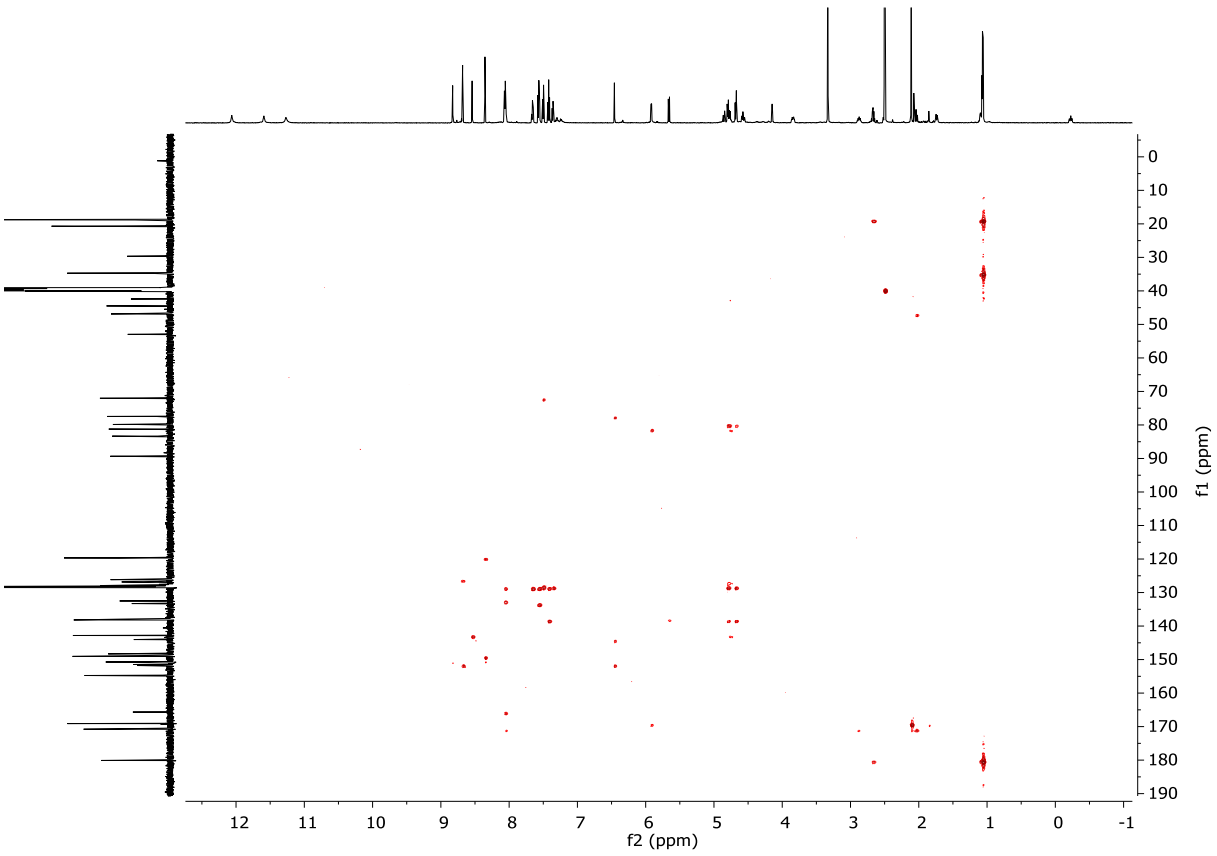
Compound **24** (NOESY, 600 MHz, DMSO-d_6)



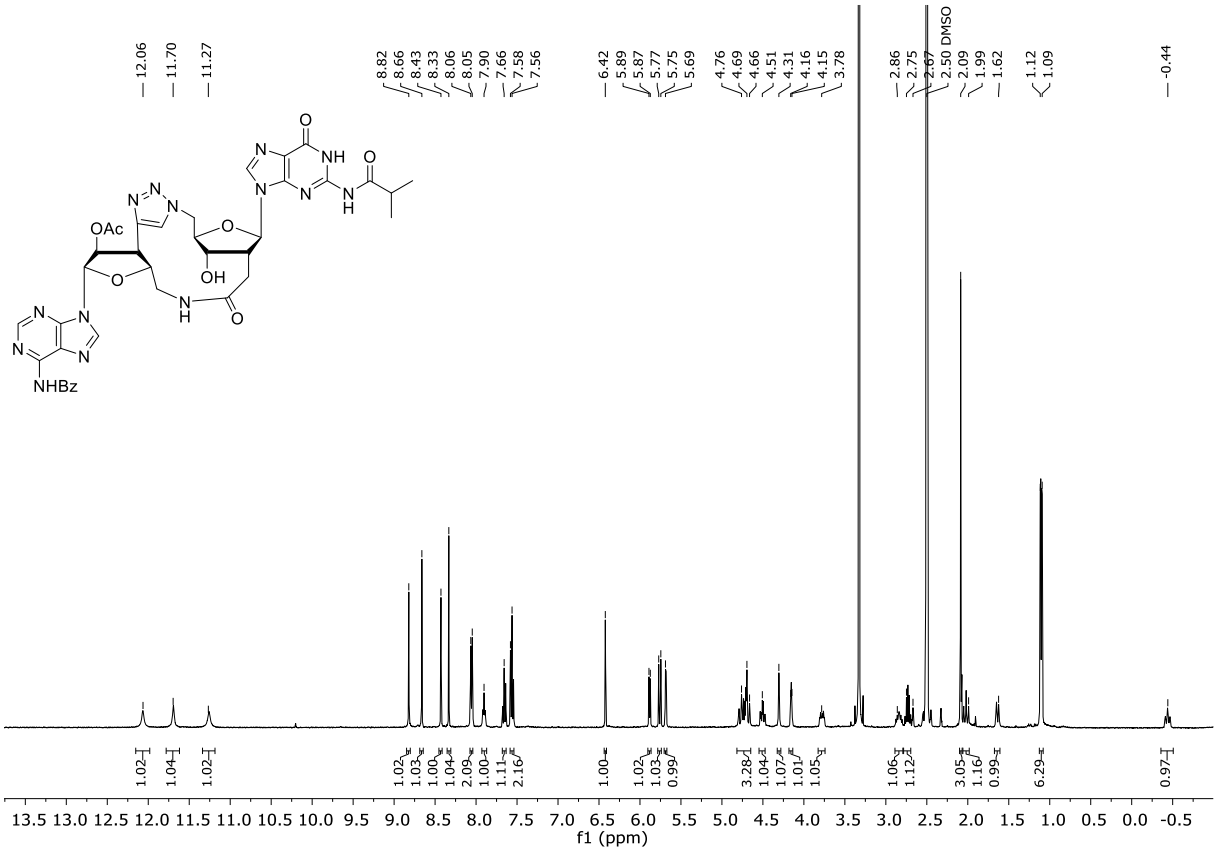
Compound **24** (HSQC, 600 MHz, DMSO-d_6)



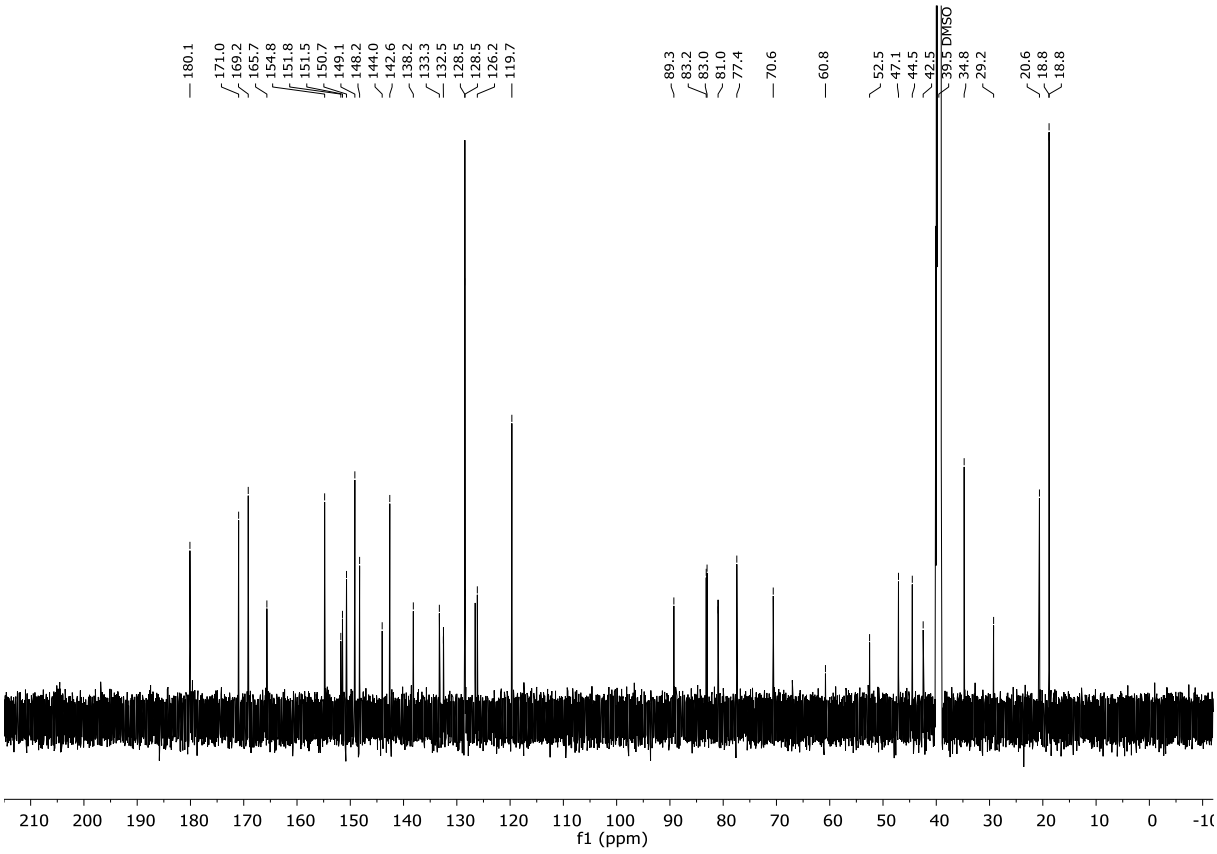
Compound **24** (HMBC, 600 MHz, DMSO-d_6)



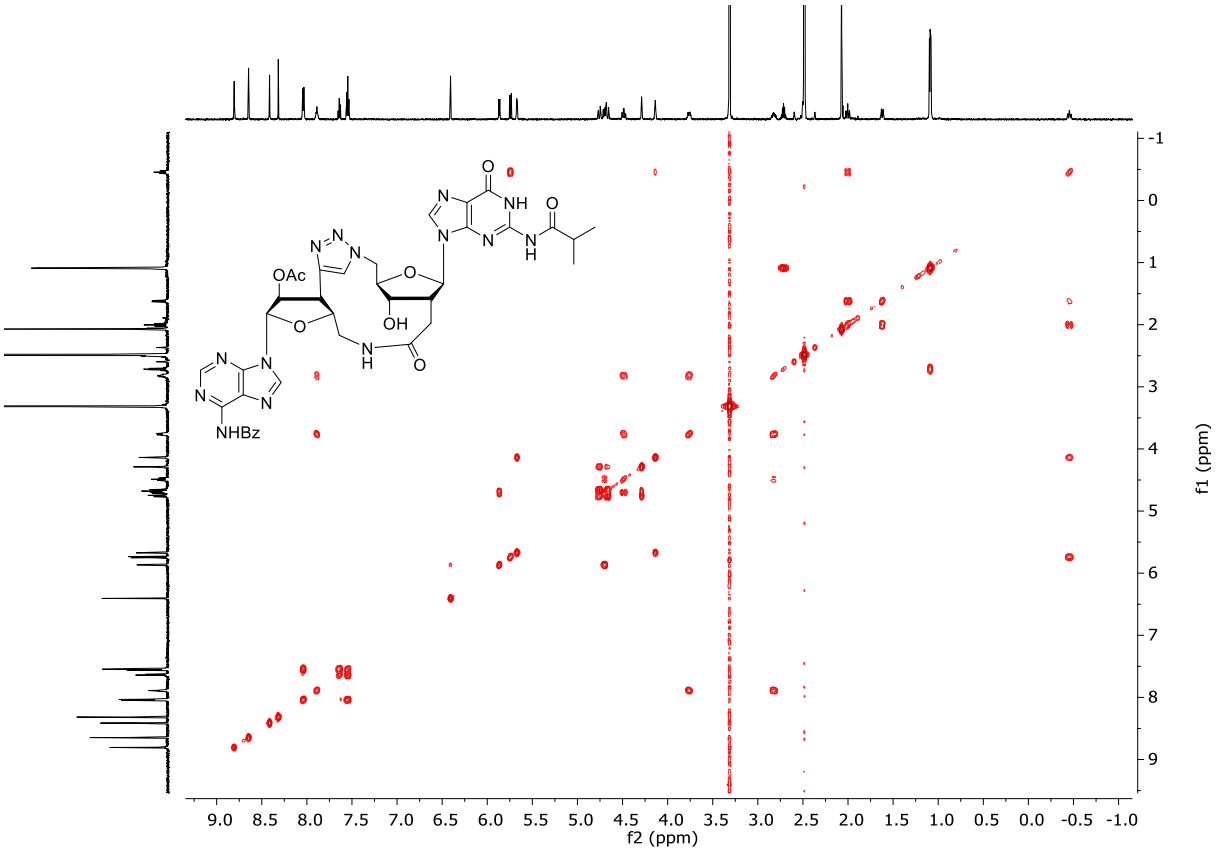
Compound **25** (^{1}H-NMR, 600 MHz, DMSO-d_6)



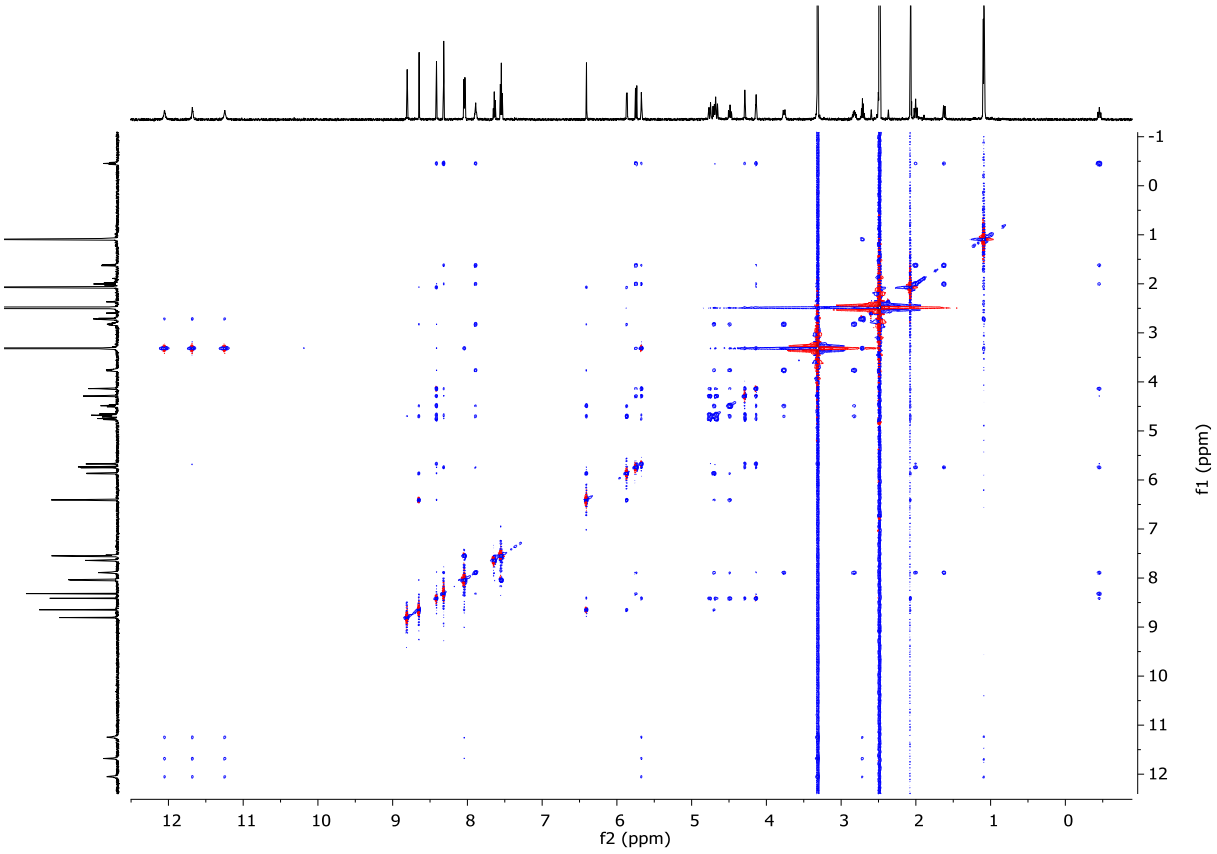
Compound **25** (^{13}C-NMR, 151 MHz, DMSO-d_6)



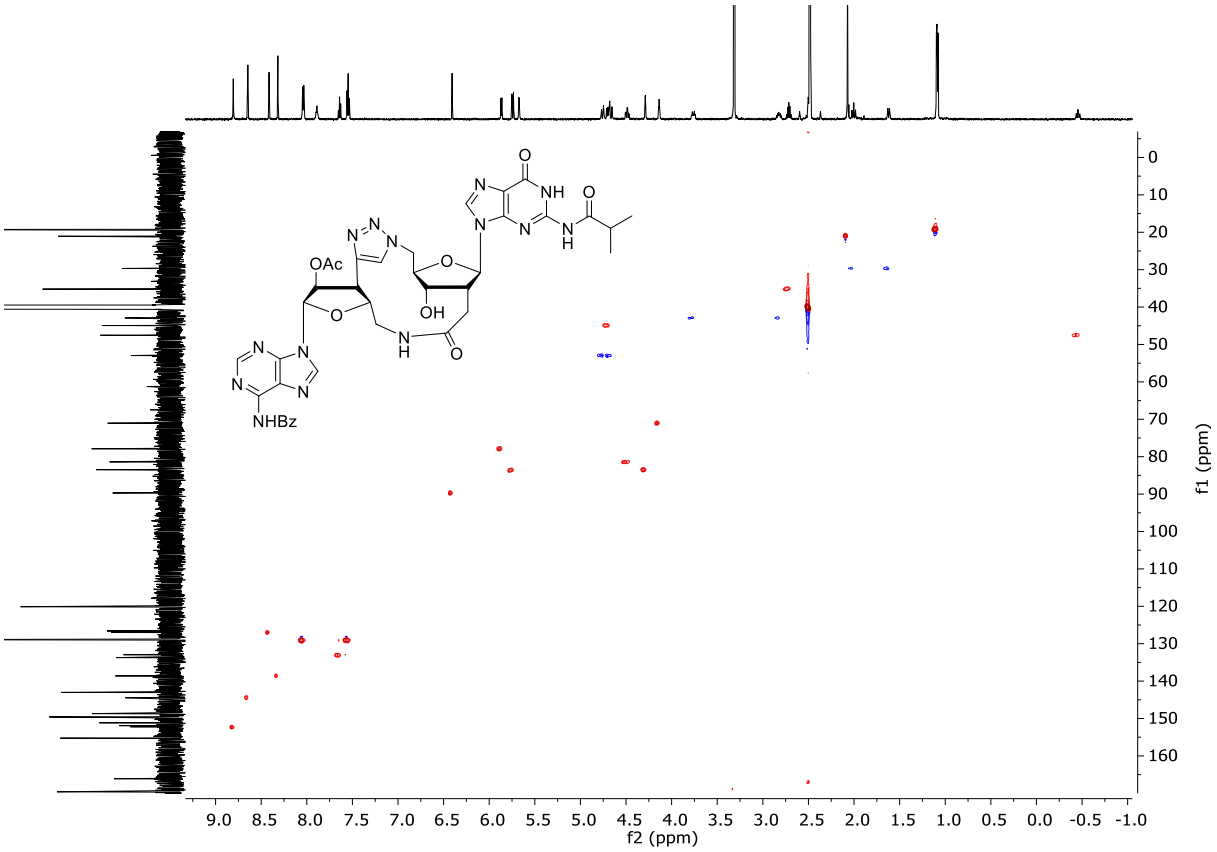
Compound **25** (COSY, 600 MHz, DMSO-d_6)



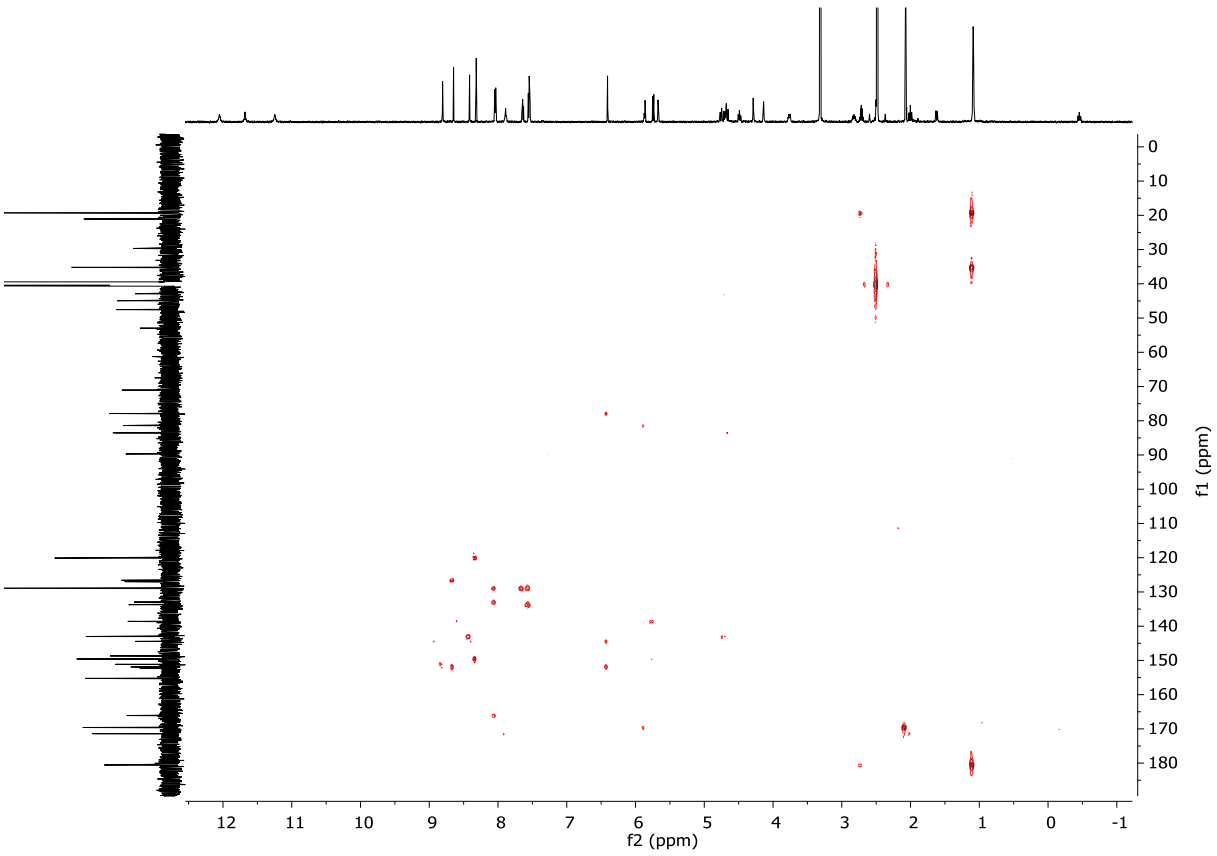
Compound **25** (NOESY, 600 MHz, DMSO-d_6)



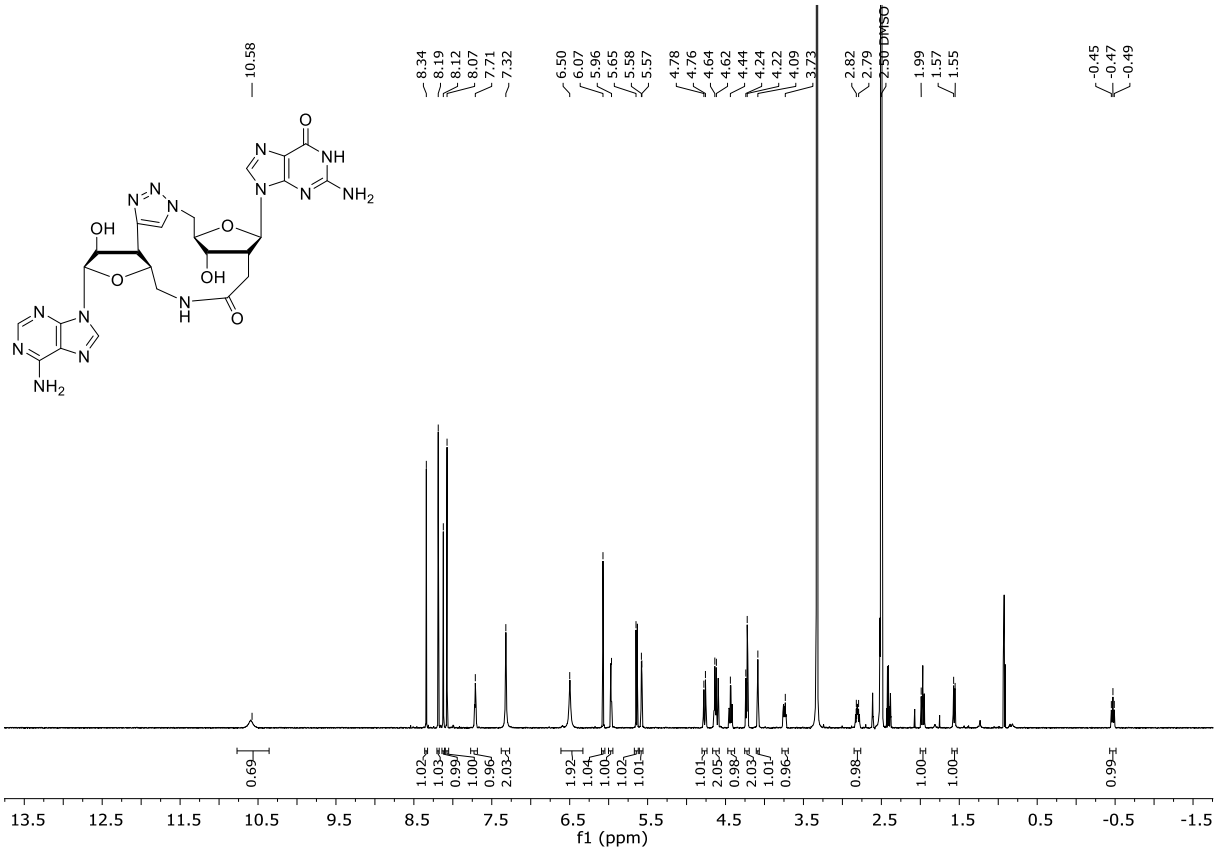
Compound **25** (HSQC, 600 MHz, DMSO-d_6)



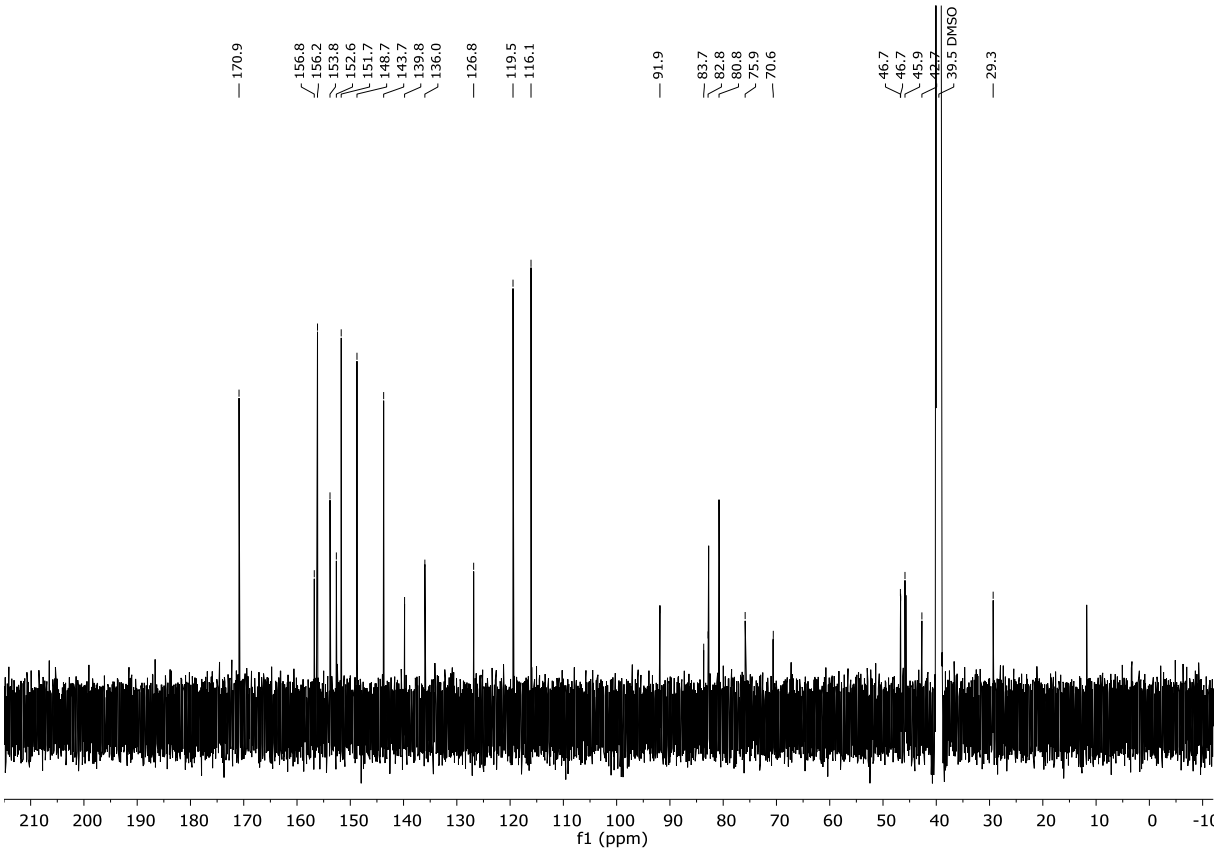
Compound **25** (HMBC, 600 MHz, DMSO-d_6)



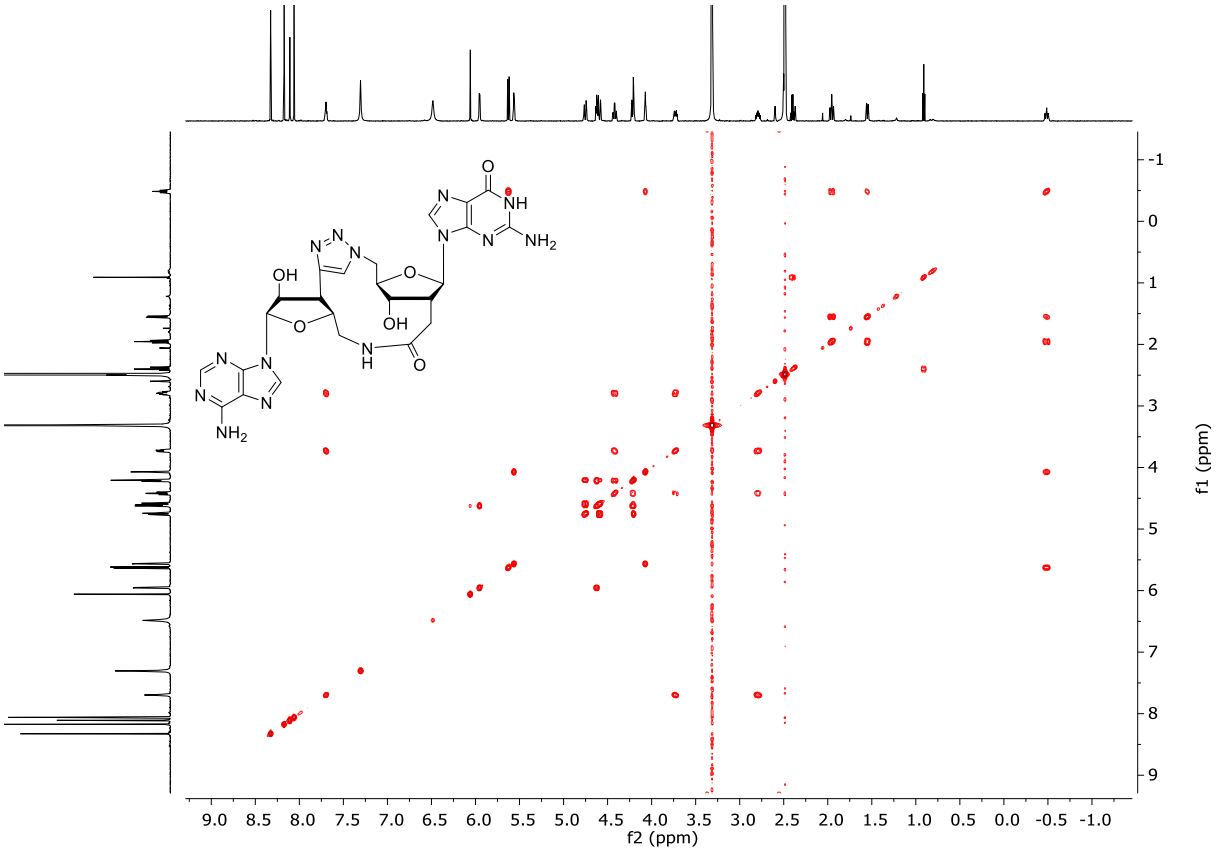
Compound **4** (^{1}H-NMR, 600 MHz, DMSO-d_6)



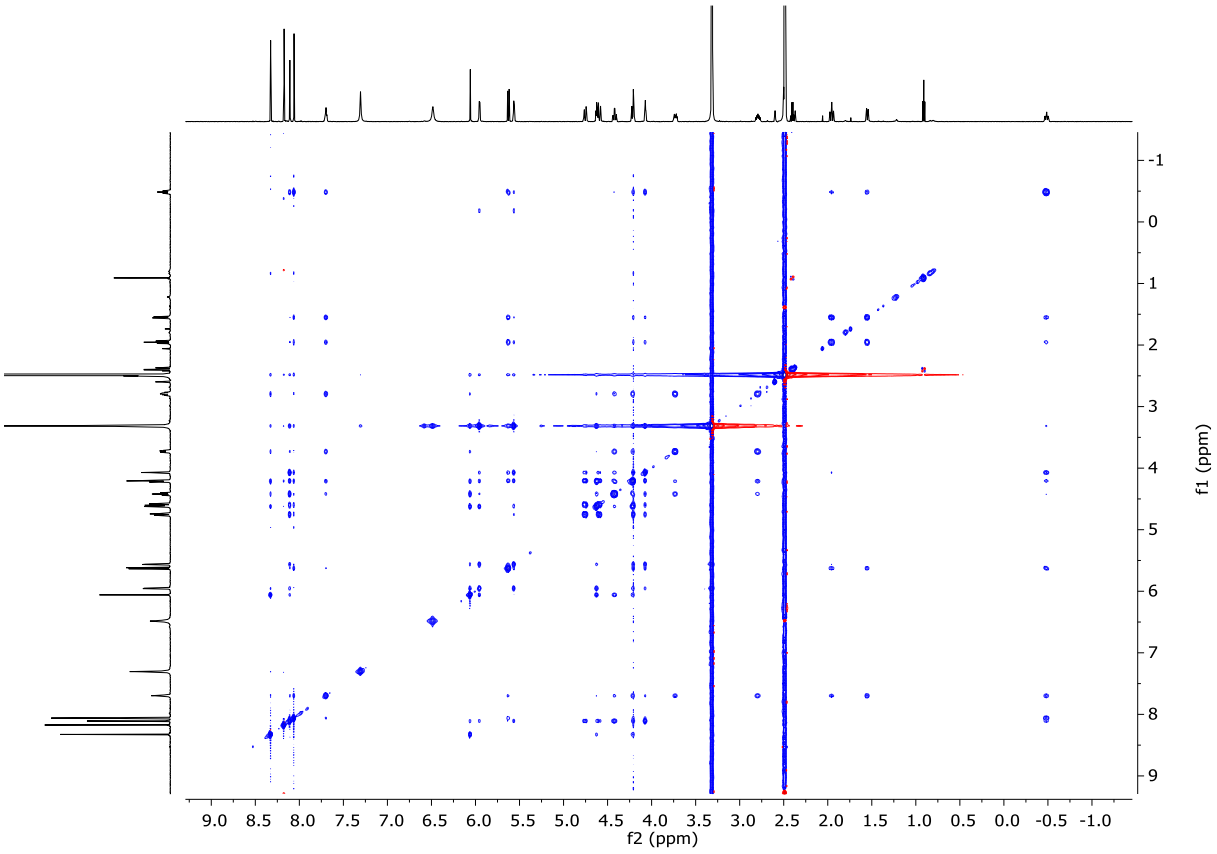
Compound **4** (^{13}C-NMR, 151 MHz, DMSO-d_6)



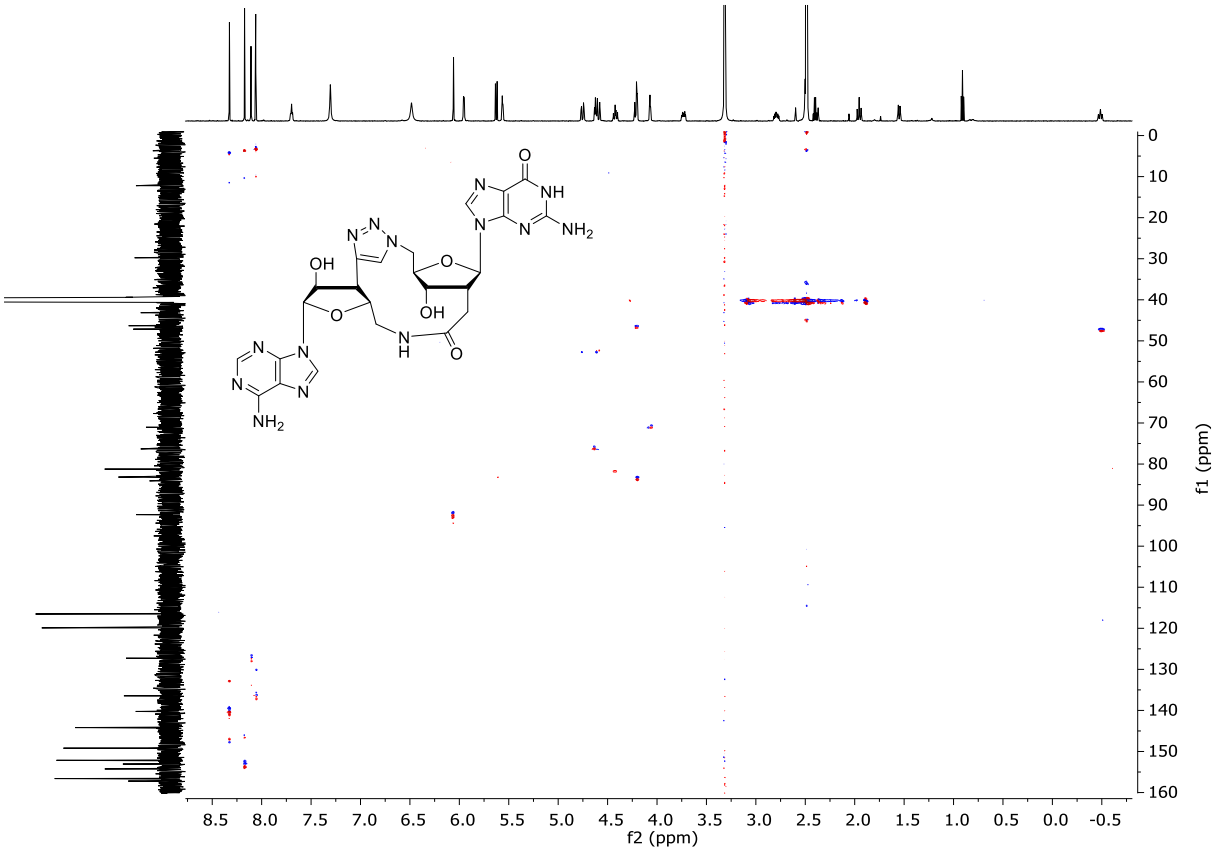
Compound **4** (COSY, 600 MHz, DMSO-d_6)



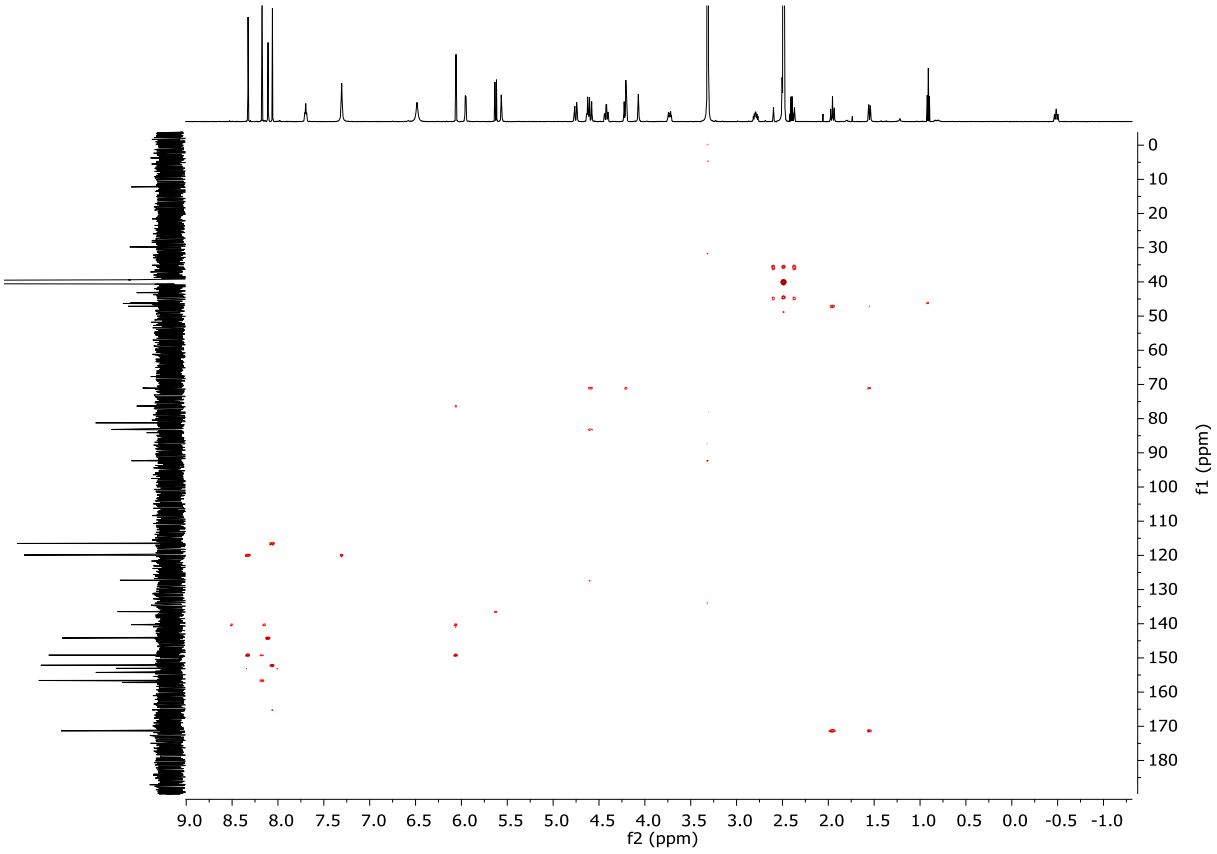
Compound **4** (NOESY, 600 MHz, DMSO-d_6)



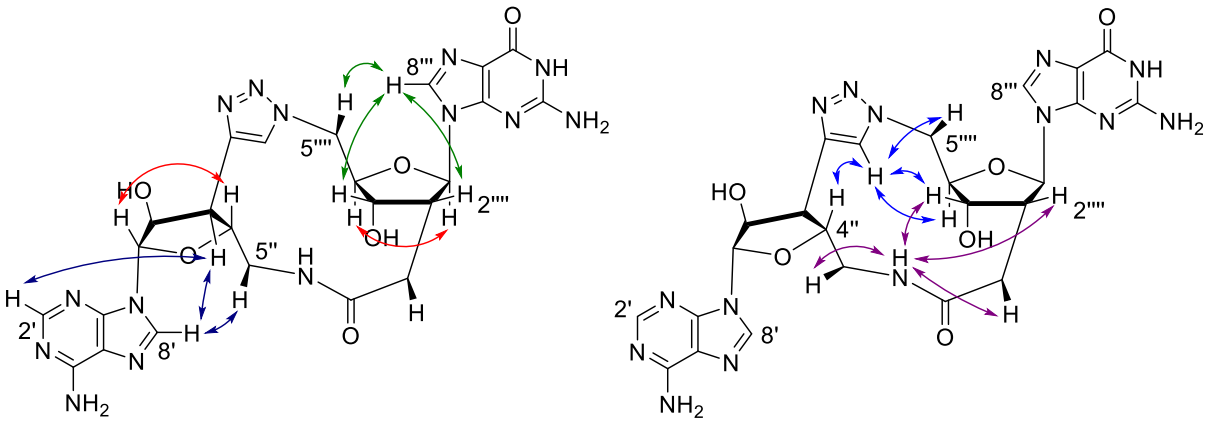
Compound **4** (HSQC, 600 MHz, DMSO-d_6)



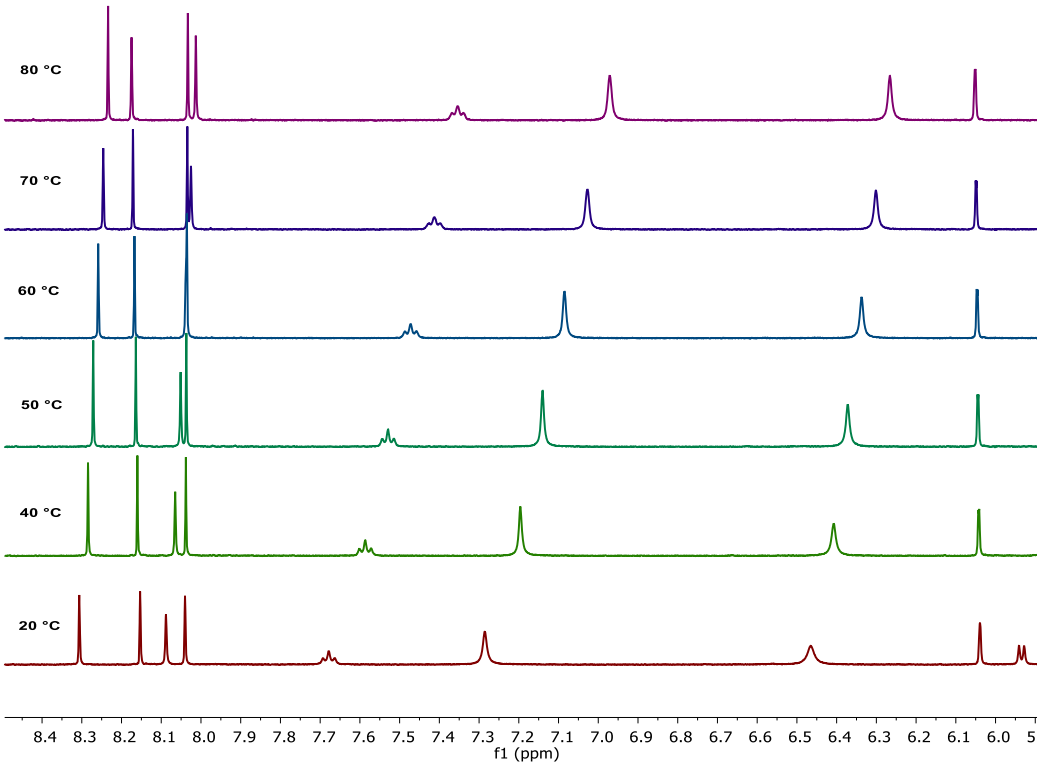
Compound **4** (HMBC, 600 MHz, DMSO-d_6)



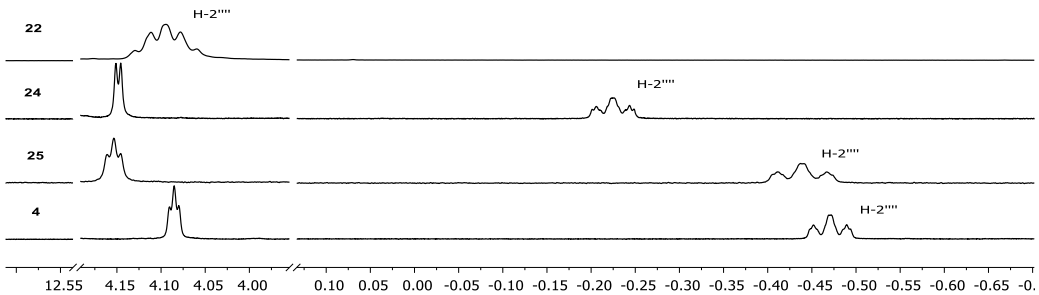
Selected NOE contacts for compound **4**



Temperature dependent ^{1}H-NMR of compound **4** in DMSO-*d*$_6$



Chemical shift of H-2''''



3. RP-HPLC

The purification of compound **24**, **25** and **4** was performed by reversed-phase HPLC. The crude products were dissolved in 30% MeCN, respectively.

Preparative RP-HPLC (flow rate: 5 mL/min)

T / min	0	15	17	22	24	30
A (H_2O) / %	85	80	20	20	85	85
B (MeCN) / %	15	20	80	80	15	15

Product fractions were collected from 23.0 – 25.0 min for **24**, 11.5 – 13.5 min for **25** and 7.5 – 9.5 min for **4**, respectively. Solvents were evaporated and the compounds were lyophilized overnight to give colorless solids.

Analytical RP-HPLC (flow rate: 0.5 mL/min) was conducted with a stronger gradient.

T / min	0	15	17	22	24	30
A (H_2O) / %	85	20	20	20	85	85
B (MeCN) / %	15	80	80	80	15	15

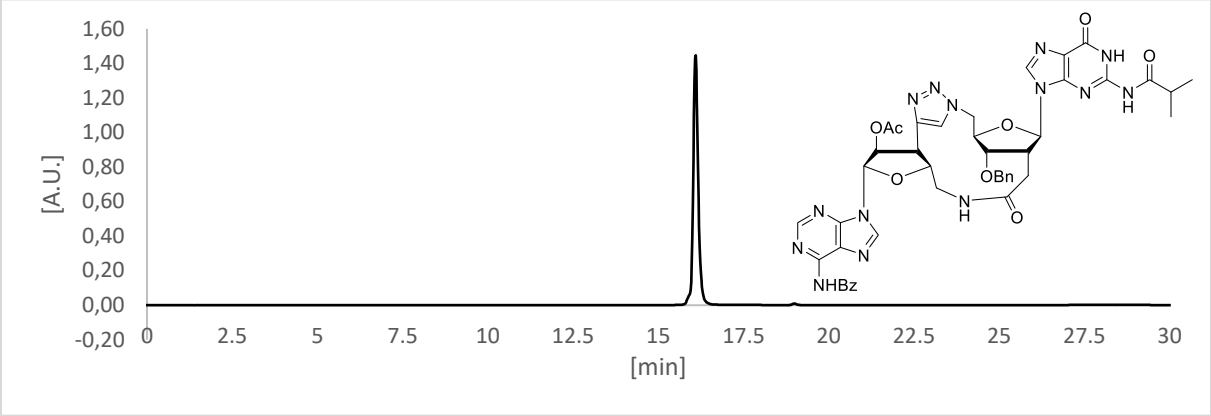
R_t (compound **24**) = 16.1 min

R_t (compound **25**) = 12.5 min

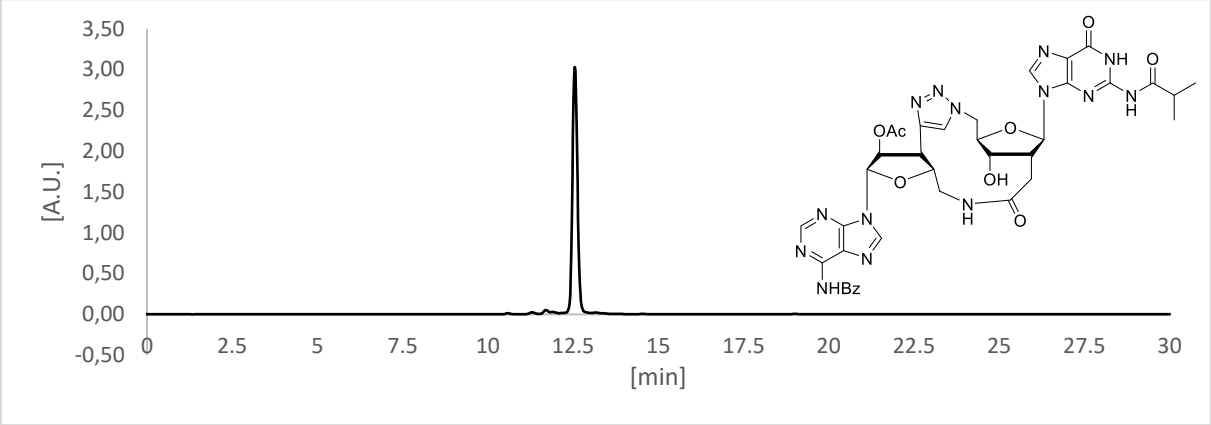
R_t (compound **4**) = 7.8 min

Analytical RP-HPLC (15% to 80% MeCN gradient elution)

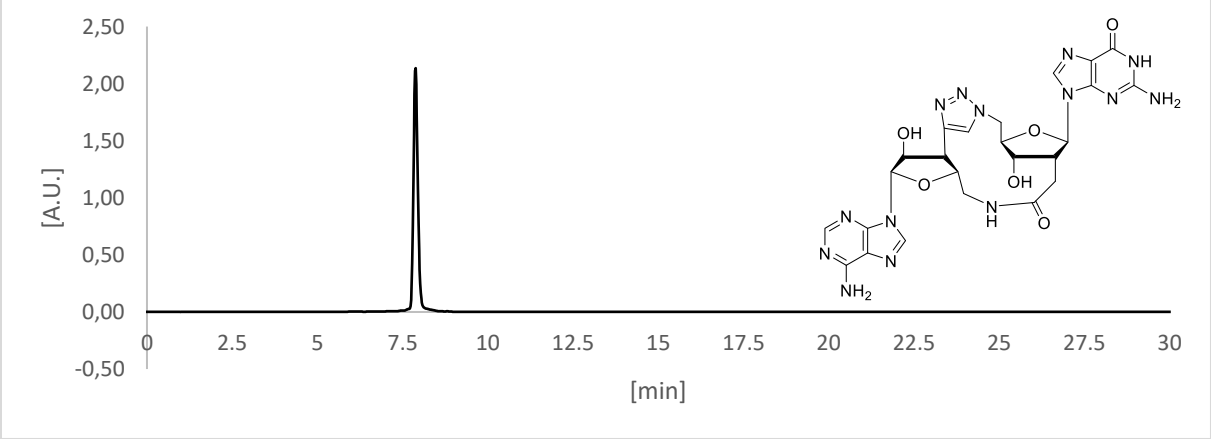
Compound **24**:



Compound **25**:

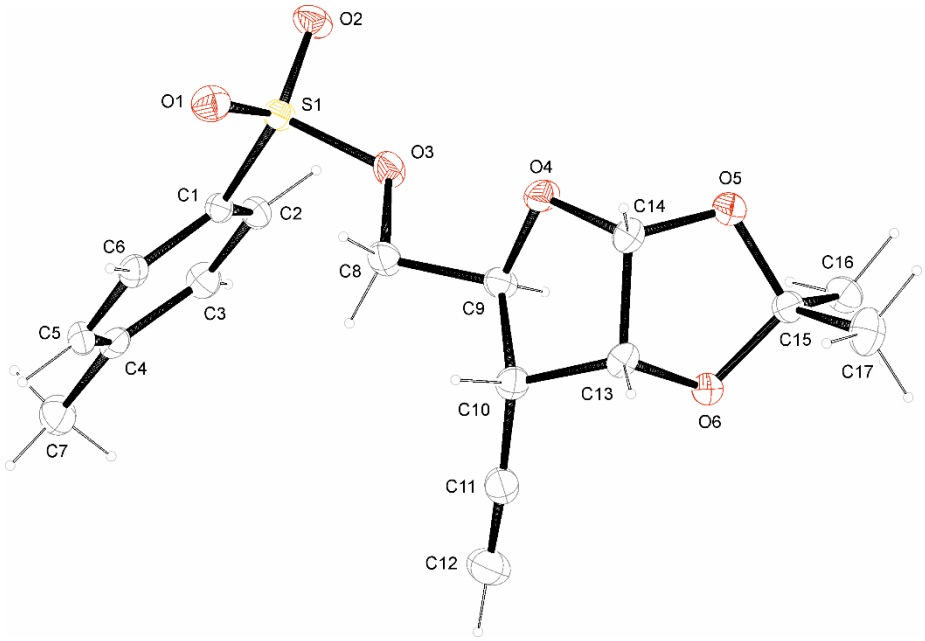


Compound **4**:



4. X-ray crystallography data

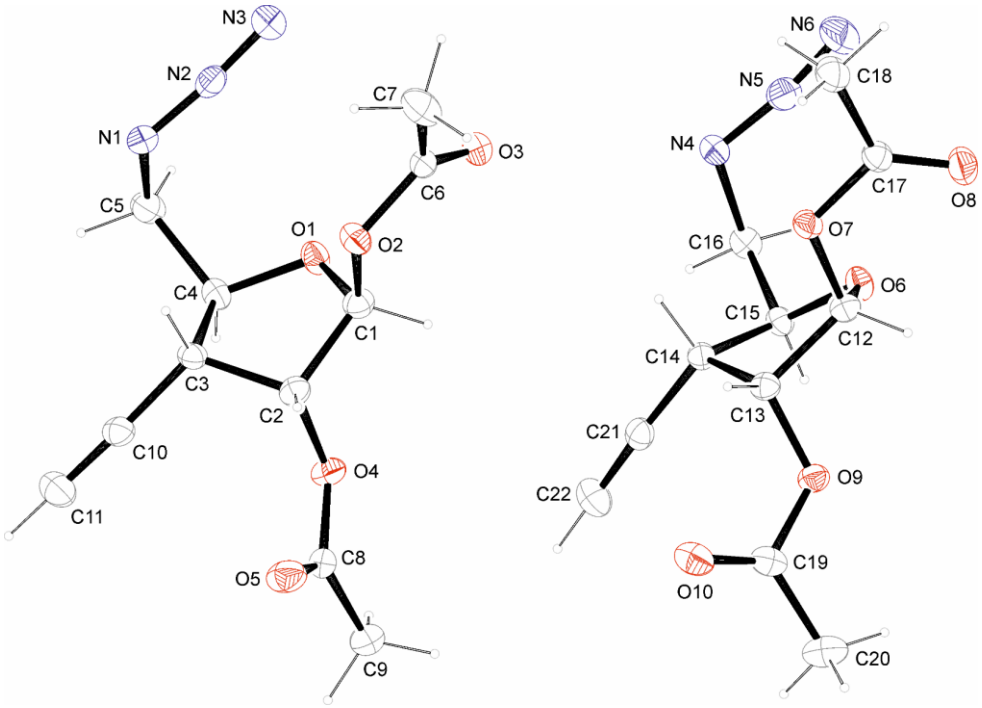
Compound **10d**:



net formula	$C_{17}H_{20}O_6S$
M_r/g mol^{-1}	352.39
crystal size/mm	0.080 × 0.050 × 0.030
T/K	103.(2)
radiation	MoKα
diffractometer	'Bruker D8 Venture TXS'
crystal system	orthorombic
space group	'P 21 21 21'
a/Å	5.5797(3)
b/Å	16.2174(7)
c/Å	18.6761(8)
α/°	90
β/°	90
γ/°	90
V/Å^3	1689.97(14)
Z	4
calc. density/g cm^{-3}	1.385
μ/mm^{-1}	0.221
absorption correction	Multi-Scan
transmission factor range	0.92–0.99
refls. measured	17954
R_{int}	0.0404
mean σ(*I*)/*I*	0.0305
θ range	3.327–27.090
observed refls.	3480

x, y (weighting scheme)	0.0355, 0.4445
hydrogen refinement	constr
Flack parameter	−0.01(3)
refls in refinement	3706
parameters	220
restraints	0
$R(F_{obs})$	0.0290
$R_w(F^2)$	0.0724
S	1.041
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.236
min electron density/e Å^{-3}	−0.366

Compound **11**:

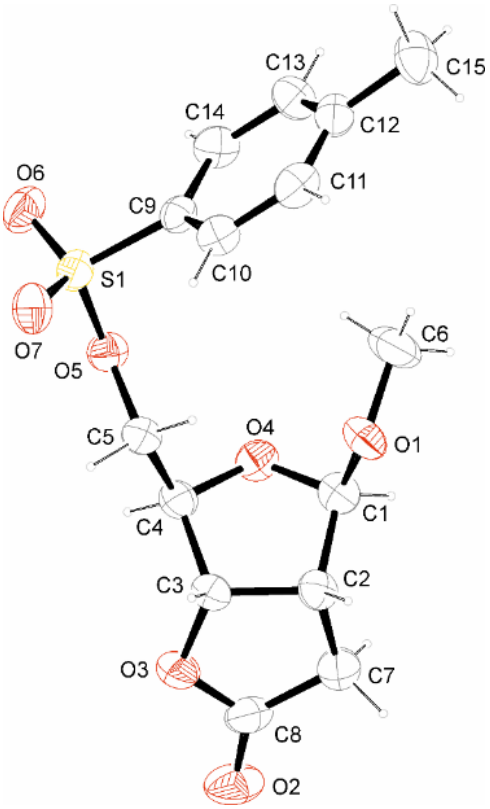


net formula	$C_{11}H_{13}N_3O_5$
M_r/g mol^{-1}	267.24
crystal size/mm	0.080 × 0.050 × 0.030
T/K	103.(2)
radiation	MoKα
diffractometer	'Bruker D8 Venture TXS'
crystal system	monoclinic
space group	'P 1 21 1'
a/Å	10.1530(7)
b/Å	7.0554(5)
c/Å	18.1171(13)

α/°	90
β/°	101.769(2)
γ/°	90
V/Å^3	1270.51(16)
Z	4
calc. density/g cm^{-3}	1.397
μ/mm^{-1}	0.112
absorption correction	Multi-Scan
transmission factor range	0.94–1.00
refls. measured	13325
R_{int}	0.0382
mean σ(I)/I	0.0460
θ range	3.376–26.363
observed refls.	4785
x, y (weighting scheme)	0.0337, 0.2537
hydrogen refinement	constr
Flack parameter	−0.6(5)
refls in refinement	5174
parameters	347
restraints	1
$R(F_{obs})$	0.0344
$R_w(F^2)$	0.0821
S	1.041
shift/error$_{max}$	0.001
max electron density/e Å^{-3}	0.188
min electron density/e Å^{-3}	−0.183

Correct structure derived from synthesis.

Compound **18**:



net formula	$C_{15}H_{18}O_7S$
M_r/g mol^{-1}	342.35
crystal size/mm	0.444 × 0.148 × 0.141
T/K	173(2)
radiation	MoKα
diffractometer	'Oxford XCalibur'
crystal system	monoclinic
space group	'P 21'
a/Å	11.1971(7)
b/Å	5.8612(3)
c/Å	12.8844(8)
α/°	90
β/°	112.211(7)
γ/°	90
V/Å^3	782.84(9)
Z	2
calc. density/g cm^{-3}	1.452
μ/mm^{-1}	0.241
absorption correction	multi-scan
transmission factor range	0.90050–1.00000
refls. measured	4714
R_{int}	0.0268
mean σ(I)/I	0.0502

θ range	4.541–26.367
observed refls.	2507
x, y (weighting scheme)	0.0348, 0.1576
hydrogen refinement	constr
Flack parameter	−0.04(7)
refls in refinement	2844
parameters	210
restraints	1
$R(F_{obs})$	0.0413
$R_w(F^2)$	0.0926
S	1.029
shift/error_{max}	0.001
max electron density/e Å^{-3}	0.249
min electron density/e Å^{-3}	−0.252

5. Binding evaluation of compound 4 to STING *in vitro*

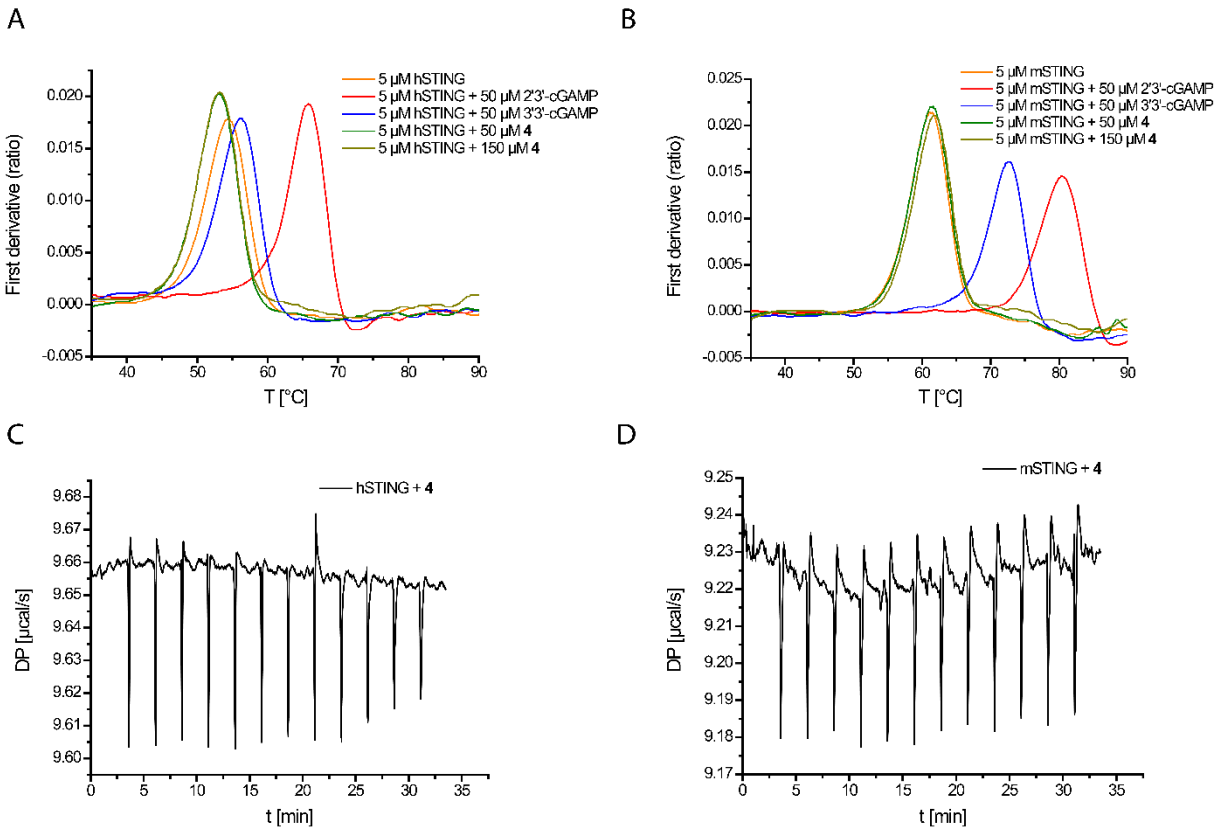


Figure S1: Compound **4** is not binding to STING *in vitro*.
(A) DSF thermal shift first derivative of 5 µM hSTING_L139 (orange), 5 µM hSTING_L139 + 50 µM 2′3′-cGAMP (red), 3′3′-cGAMP (blue), 50 µM **4** (dark green) and 150 µM **4** (light green).
(B) DSF thermal shift first derivative of 5 µM mSTING_L138 (orange), 5 µM mSTING_L139 + 50 µM 2′3′-cGAMP (red), 3′3′-cGAMP (blue), 50 µM **4** (dark green) and 150 µM **4** (light green).
(C) ITC measurement raw data of 20 µM hSTING_L139 titrated with 291 µM **4**.
(D) ITC measurement raw data of 20 µM mSTING_L138 titrated with 291 µM **4**.

6. References

[1] Betkekar, V. V.; Panda, S.; Kaliappan, K. P. *Org. Lett.* **2012**, *14*, 198-201.
[2] Yoshimura, Y.; Sano, T.; Matsuda, A.; Ueda, T. *Chem. Pharm. Bull.* **1988**, *36*, 162-167.
[3] Parr, I. B.; Horenstein, B. A. *J. Org. Chem.* **1997**, *62*, 7489-7494.
[4] Shing, T. K. M.; Leung, Y. C.; Yeung, K. W. *Tetrahedron* **2003**, *59*, 2159-2168.
[5] Kaiya, Y.; Hasegawa, J.-i.; Momose, T.; Sato, T.; Chida, N. *Chem. Asian J.* **2011**, *6*, 209-219.
[6] Mondal, S.; Mohamed, R. K.; Manoharan, M.; Phan, H.; Alabugin, I. V. *Org. Lett.* **2013**, *15*, 5650-5653.
[7] Li, N.-S.; Lu, J.; Piccirilli, J. A. *Org. Lett.* **2007**, *9*, 3009-3012.
[8] Webber, M. J.; Warren, S. A.; Grainger, D. M.; Weston, M.; Clark, S.; Woodhead, S. J.; Powell, L.; Stokes, S.; Alanine, A.; Stonehouse, J. P.; Frampton, C. S.; White, A. J. P.; Spivey, A. C. *Org. Biomol. Chem.* **2013**, *11*, 2514-2533.
[9] Jenny, T. F.; Schneider, K. C.; Benner, S. A. *Nucleosides and Nucleotides* **1992**, *11*, 1257-1261.
[10] Robins, M. J.; Zou, R.; Guo, Z.; Wnuk, S. F. *J. Org. Chem.* **1996**, *61*, 9207-9212.
[11] K. Singh, S.; K. Sharma, V.; Bohra, K.; E. Olsen, C.; K. Prasad, A. *Curr. Org. Synth.* **2014**, *11*, 757-766.

Discussion

The degradation of second messengers is a crucial step to regulate their cellular level. In recent years, different PDEs have been identified to hydrolyze c-di-AMP in bacterial species. However, the precise degradation pathway and its regulation are not fully understood yet and many open questions remain.

Two important enzyme classes in the degradation pathway of c-di-AMP are the two DHH-type PDEs, GdpP and DhhP [113,114]. Although both PDEs possess the same highly conserved catalytic DHH-DHHA1 domain, they exist in parallel in many species. A third PDE class is represented by the HD-domain PDE PgpH, which can be found in several species [94]. The set of these three PDE classes varies among bacteria, which raises several questions. Do all PDEs have the same function in one organism and for which purpose? Are there any differences in activity, substrate specificity or product formation? What are the precise molecular mechanisms to achieve specificity? These questions are discussed in the following chapters.

1. Substrate specificity and product formation of DHH-type PDEs

Studies in different organisms are consistent about substrate specificity and product formation of GdpP PDEs. GdpP hydrolyzes c-di-AMP exclusively to 5′-pApA, which represents an intermediate in the degradation pathway [32,63,65,67,113,123]. However, the substrate specificity of DhhP PDEs demonstrates significant interspecies differences. The DhhP homologues from *M. pneumoniae*, *B. burgdorferi*, *M. tuberculosis* (Rv2837c) and *M. smegmatis* (MsPDE or NrnA) were shown to accept c-di-AMP as a substrate [37,45,107,114]. In contrast, homologues from *S. aureus*, *T. maritima* (TmPDE), *S. pneumoniae* (SpPde2) and *S. mutans* demonstrate a clear preference for linear dinucleotides and are not able to hydrolyze c-di-AMP under physiological conditions [32,55,115,123].

The kinetic values for hydrolysis of 5′-pApA are comparable between *M. tuberculosis* Rv2837c (k_{cat} = 0.87 ± 0.05 s^{-1}, K_M = 129.88 ± 12.95 µM), *T. maritima* TmPDE (k_{cat} = 0.14 s^{-1}, K_M = 204 ± 10 µM), *S. pneumoniae* SpPde2 (K_M = 97 ± 8 µM) and *S. mutans* Pde2 (k_{cat} = 36.5 ± 8.08 s^{-1}, K_M = 199 ± 65.4 µM) [32,107,115,123]. The hydrolysis of c-di-AMP could only be described kinetically for Rv2837c (k_{cat} = 0.23 ± 0.02 s^{-1}, K_M = 30.89 ± 6.71 µM), MsPDE (k_{cat} = 0.52 ± 0.03 s^{-1}, K_M = 6.80 ± 0.84 µM) and SpPde2 (K_M = 16.75 ± 8.70 µM) [32,45,107]. The hydrolysis activity of SpPde2 was also analyzed within this work revealing a contradictory result [115]. According to our studies, SpPde2 and TmPDE have only weak activity for c-di-AMP, whereas 5′-pApA is hydrolyzed at high turnover rates [115]. This observation was also described for *S. mutans* Pde2 by Konno *et al.* [123]. The substrate specificity has been confirmed in the present work by binding analysis of the inactive TmPDE D80N D154N mutant. ITC and

SPR experiments revealed nanomolar affinities for 5′-pApA, whereas the c-di-AMP binding constant was calculated at 4.5 µM with a high error [115].

Another difference in substrate specificity was observed for TmPDE compared to the homologue MsPDE [115]. MsPDE was shown to possess an exonuclease activity and act as a nanoRNase [118]. In contrast, nanoRNase activity was ruled out for TmPDE within this work. Almost no product formation was observed for 5′-pApApA, although a 10-fold higher enzyme concentration was used compared to the highly effective turnover of 5′-pApA [115]. This can also be explained based on structural analysis of TmPDE in complex with dinucleotide substrates. The base residues of the ligand are coordinated at the two access positions of the active site, while the ribose residues are covered in the core. Thus, an elongated RNA chain would not fit in the active site for hydrolysis. In conclusion, an RNase activity can be excluded as physiological role of TmPDE [115].

Apart from these observations, TmPDE was also shown to hydrolyze 5′-pApG and 5′-pGpG effectively [115]. Binding experiments of the TmPDE D80N D154N mutant confirmed this extended substrate specificity and demonstrated affinities in the following order 5′-pApG < 5′-pGpG < 5′-pApA [115]. This is consistent with a previous study that describes a comparable hydrolysis of 5′-pGpG to 5′-pApA by Rv2837c [107]. As observed for c-di-AMP, TmPDE has no hydrolysis activity for c-di-AMP and Rv2837c converts c-di-GMP with much lower activity than c-di-AMP. At this, the conversion of c-di-GMP is catalyzed by EAL domains or HD-GYP domains [20]. Comparably to the first step of c-di-AMP degradation, EAL-domains hydrolyze c-di-GMP exclusively to 5′-pGpG. In contrast, HD-GYP domains catalyse the reaction of c-di-GMP to the final product GMP [20]. Species like *M. tuberculosis* do not possess HD-GYP domains, thus Rv2837c might have an important role for 5′-pGpG hydrolysis [107]. Although HD-GYP domains from *T. maritima* were shown to hydrolyze c-di-GMP to 5′-pGpG and GMP, TmPDE might still contribute to the hydrolysis of 5′-pGpG in the second step. Taken together, these findings indicate that the role of DhhP homologues is not limited to the c-di-AMP degradation pathway and might also be important for the second step in c-di-GMP degradation.

The differences in substrate specificity and product formation have also been analyzed structurally. During this work, the crystal structure of TmPDE has been determined in complex with several ligands. The high-resolution structures of one ligand-free state, two substrate-bound states and three post-reaction states give a detailed insight in the reaction mechanism. The structures in complex with 5′-pApA and 5′-pApG represent the substrate-bound states and display their differences in binding affinity. In comparison to 5′-pApA, the guanosine base of 5′-pApG forms additional interactions from oxygen O6 to Arg 243 and nitrogen N2 to Asn 162. This is in good agreement with the higher binding affinity for 5′-pApG and 5′-pGpG compared to

5′-pApA [115]. Notably, the overall structure of TmPDE in the substrate-bound states is highly similar to the structure of Rv2837c [107,115]. Therefore, it is difficult to depict a difference that explains why Rv2837c accepts c-di-AMP as a substrate, whereas TmPDE only accepts linear dinucleotides as substrates.

In a recent study, the active site of Rv2837c was analyzed by mutational investigations in comparison with a separate DHH domain construct of *S. aureus* GdpP [116]. Here, the active site is divided into the three positions R, C and G. The hydrolysis of c-di-AMP by GdpP occurs in CG position. However, it is unclear if Rv2837c hydrolyzes c-di-AMP in the same position. In this study, a T180R mutant of Rv2837c was designed to reduce space in position G. This resulted in more than 50% diminished c-di-AMP hydrolysis, while the activity for 5′-pApA remained unaffected. It is assumed that the mutation prevents the access of c-di-AMP to position G, which supports the suggestion of c-di-AMP hydrolysis in CG position by Rv2837c [116]. Although position G from TmPDE is highly similar compared to Rv2837c (Figure 18B, C), TmPDE lacks hydrolysis activity for c-di-AMP. Thus, the difference in c-di-AMP specificity of TmPDE and Rv2837c remains unclear and is probably not dependent on the size of position G.

The difference in substrate coordination can also explain exclusive product formation of 5′-pApA by GdpP. In Rv2837c, 5′-pApA can relocate into CR position, which is required for positioning the phosphodiester linkage in close proximity to the metal ions [116]. The bigger R site of Rv2837c is also present in the structure of TmPDE, which is in accordance with the ability to hydrolyze 5′-pApA (Figure 18A). GdpP has a smaller R site that is assumed to prevent access of 5′-pApA.

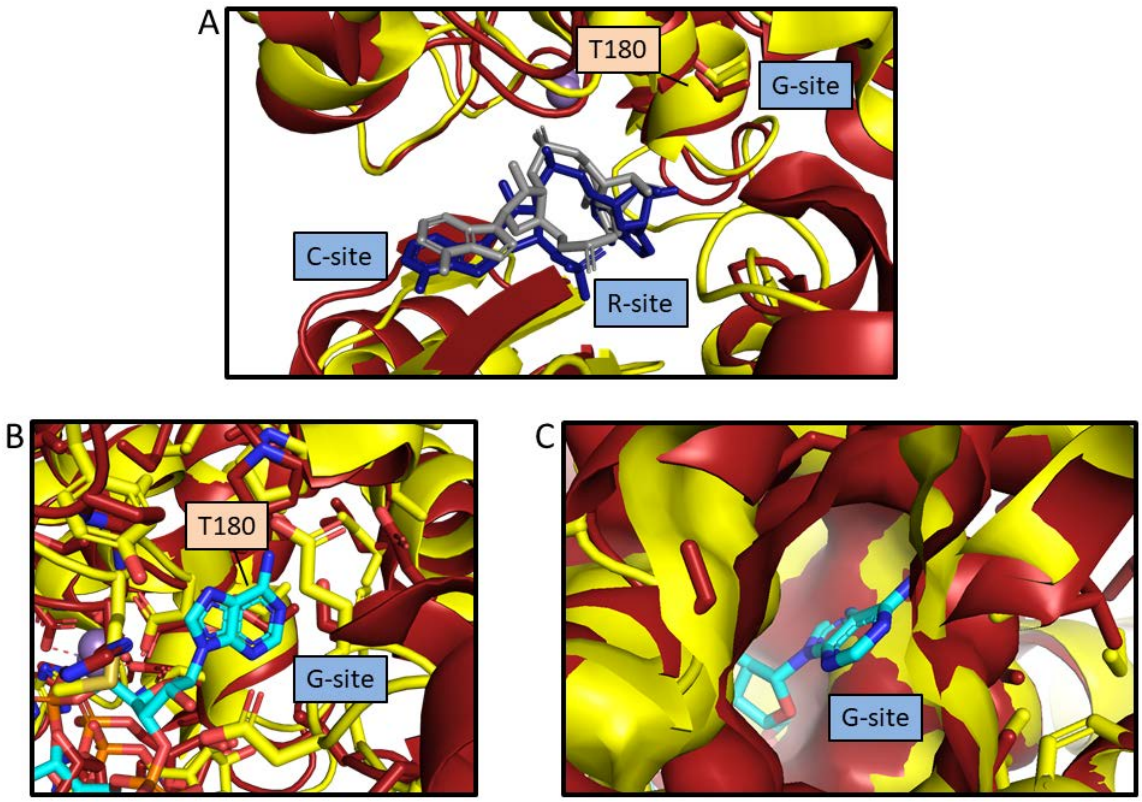


Figure 18: Binding site of TmPDE compared to binding site of Rv2837c. (A) Binding site of TmPDE (yellow) in complex with 5′-pApA (dark blue) (PDB: 5O4Z [115]) and binding site of Rv2837c (red) in complex with 5′-pApA (grey) and Mn^{2+} ions (purple) (PDB: 5JJU [107]). In both active sites 5′-pApA is coordinated in RC position. (B) G-site of TmPDE (yellow, PDB: 5O4Z [115]) and Rv2837c (red, PDB: 5JJU [107]) with modelled c-di-AMP from *S. aureus* GdpP (cyan, PDB: 5XSN [116]). (C) Surface view on G-site of TmPDE (yellow, PDB: 5O4Z [115]) and Rv2837c (red, PDB: 5JJU [107]) with modelled c-di-AMP from *S. aureus* GdpP (cyan, PDB: 5XSN [116]).

The hydrolysis mechanism can be proposed based on the structures of TmPDE and Rv2837c and is consistent with a two metal ion hydrolysis for phosphodiester linkages [107,115]. Particularly, the different reaction states observed in the crystal structures of TmPDE within this work allow a detailed analysis of the reaction sequence. In one structure, 5′-pApA is present with a hydrolyzed phosphodiester linkage. In addition, one crystallization sample has been incubated with 5′-pGpG, which resulted in a structure of TmPDE bound to GMP. This indicates that the two resulting NMP molecules separately leave the active site after cleavage [115]. Another insight is obtained from the observation of a small conformational difference in the apo structure of TmPDE compared to the ligand-bound states. The active site of DHH-type PDEs is located between the DHH and the DHHA1 subunits, which are connected by a flexible linker. In the apo structure of TmPDE, one monomer has a larger distance between the two domains compared to the other monomer. This difference can be described as “open” and “closed” state. In contrast to the apo structure, all ligand-bound states of TmPDE have both active sites closed [115]. The different states can also be observed in structures of other DHH-type PDEs, like the nanoRNase NrnA from *B. subtilis* or the exonuclease RecJ from *Thermus thermophilus* [194–196] (Figure 19). This leads to the suggestion of an opening and closing mechanism of the DHH-DHHA1 domain during the reaction. It could

be possible that an open conformation is required for the substrate access to the active site. In order to bring the substrate in close proximity to the catalytic metal ions, the two subdomains have to close. After hydrolysis, the domains have to open for binding of the next substrate. In addition to the asymmetric open/closed dimer of apo TmPDE in the crystal structure, ITC experiments revealed a binding stoichiometry of n = 0.5 for 5′-pApA to TmPDE. Therefore, it is possible to speculate that the active site of only one monomer can be occupied at a time while the other monomer remains open. In contrast to this theory, all ligand-bound crystal structures have both active sites closed and occupied by ligands. For both observations in the TmPDE structures, an influence by crystallisation conditions and/or packing cannot be excluded for inducing a closed or open state of the active sites [115].

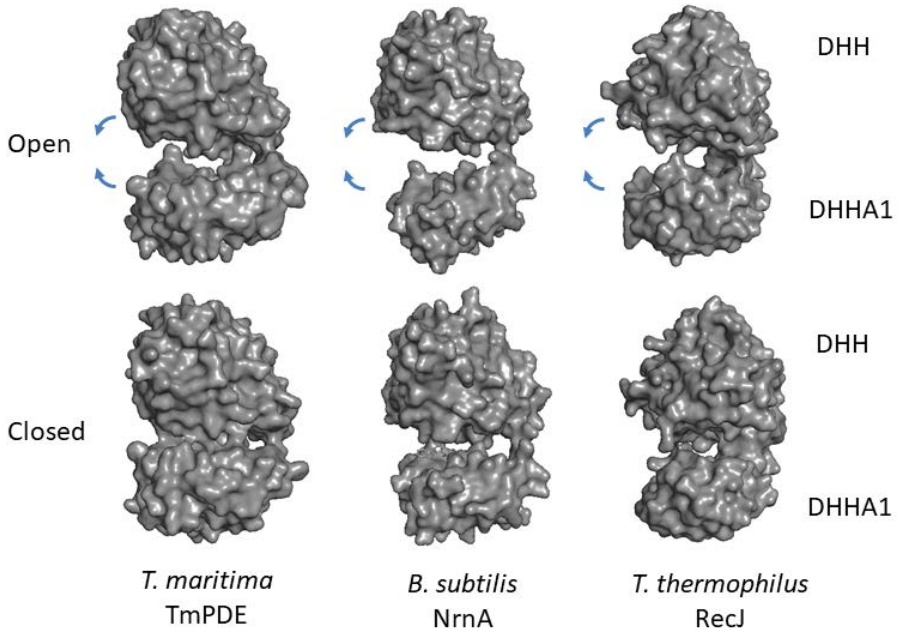


Figure 19: Open and closed conformations of the DHH-DHHA1 domain. Surface representation of the structures of TmPDE in apo state (upper left, PDB: 5O25 [115]) and in complex with 5′-pApA (lower left, PDB: 5O4Z [115]), *B. subtilis* NrnA in complex with pAp (middle upper and lower, PDB: 5IUF [194]) and *T. thermophilus* RecJ (upper right, PDB: 1IR6 [195], lower right, PDB: 2ZXP [196]).

The proposed opening and closing mechanism might also be interesting in regard to the regulation of DHH domains in GdpP PDEs. The hydrolysis activity of GdpP was shown to be inhibited by heme-binding to the PAS domain [119]. Unfortunately, no full-length structure of GdpP could be determined to observe any conformational change induced by the additional domains. The structures of the separate DHH-DHHA1 domain of *S. aureus* GdpP are all present in the same state independent on ligand binding [116]. In comparison with the different conformations of TmPDE or NrnA, the GdpP DHH-DHHA1 is rather similar to the open active site of apo TmPDE. One suggestion for the inhibition of GdpP could be the opening or closing of the active site. Binding of heme could induce a closed conformation, which would reject the substrate access to the active site. However, this mechanism can also be proposed the other way round. Closing of

the active site is required for hydrolysis and heme binding would capture the domains in an open conformation. Although these theories are very speculative, they can be interesting in future studies in order to determine the mechanism of GdpP regulation.

2. Degradation pathways of c-di-AMP

Three different PDE types of the c-di-AMP degradation pathway have been discovered so far. The HD-type PgpH and the DHH-type GdpP are multidomain PDEs located at the membrane. Both are highly specific for the exclusive hydrolysis of c-di-AMP to the intermediate 5′-pApA in many species. The stand-alone DHH domain PDE DhhP has homologues with specificity for c-di-AMP or exclusively for 5′-pApA as discussed in the previous chapter (1. Substrate specificity and product formation of DHH-type PDEs). However, the current research status leads to the suggestion that most analyzed bacteria have different sets of these three PDE types for c-di-AMP degradation. For example, GdpP and DhhP homologues can be found in parallel in *S. aureus*, *S. pneumoniae* and *S. mutans*. Therefore, a two-step degradation pathway for c-di-AMP is proposed for these species with GdpP catalyzing the first step of c-di-AMP hydrolysis into 5′-pApA, which is assumed to be regulated. To avoid accumulation of 5′-pApA, DhhP PDEs catalyze a fast conversion to the final product AMP in the second step. The activity analysis of GdpP and DhhP homologues from *S. aureus*, *S. pneumoniae* and *S. mutans* supports this model [32,55,123]. Other species, such as *T. maritima* do not contain a GdpP homologue but possess a set of PgpH and DhhP homologues. Particularly, the DhhP homologue TmPDE was identified within this work to catalyze the second step of c-di-AMP hydrolysis [115]. Although no information is available yet for the *T. maritima* PgpH homologue, the two-step degradation pathway is presumably performed by these two PDE types. Some species do contain all three PDE homologues, like *L. monocytogenes* or *B. subtilis*. In *L. monocytogenes*, the degradation of c-di-AMP to 5′-pApA is catalyzed predominantly by PgpH or PdeA (GdpP homologue) dependent on the cellular activity of [94]. Both homologues from *B. subtilis* were shown to hydrolyze c-di-AMP, with PgpH having a two-fold higher impact on c-di-AMP degradation [126]. The DhhP homologues from both species were not yet characterized with regard to c-di-AMP hydrolysis. The DHH-type nanoRNase NrnA from *B. subtilis* was shown to degrade 5′-pGpG and is therefore assumed to also accept 5′-pApA as a substrate [197]. Additionally, the DhhP homologue from *L. monocytogenes* was not found in pull down experiments of cell lysate on c-di-AMP beads, whereas PgpH and GdpP could be isolated. This leads to the suggestion that this DhhP homologue has no or minor interactions with c-di-AMP and is possibly responsible for the hydrolysis of 5′-pApA in *L. monocytogenes* [94]. In conclusion, these observations indicate that the two-step degradation pathway of c-di-AMP is also valid in *L. monocytogenes* and *B. subtilis*.

From the three PDEs for c-di-AMP degradation, species, such as *M. pneumoniae*, *B. burgdorferi*, *M. tuberculosis* and *M. smegmatis* only contain one DhhP homologue, which was shown to hydrolyze c-di-AMP in several studies [37,45,107,114]. These studies indicate that the DhhP homologue is a single enzyme for the degradation of c-di-AMP in these species. Interestingly, all these species have an abnormal cell envelope compared to gram-positive bacteria. The cell wall of gram-positive bacteria consists of a thick layer of PG and LTA and is connected to a single membrane by diacylglycerol. In contrast, the cell wall of gram-negative bacteria does not contain LTA and has an additional outer membrane with a periplasmic space between each membrane and the PG layer. Lipopolysaccharides (LPS) are located on the extracellular part of the outer membrane [198]. The cell wall of *B. burgdorferi* has an atypical gram-negative envelope as it contains glycolipids instead of LPS in the outer membrane. In addition, the flagella are present in the periplasmic space [199]. Mycobacteria have a unique cell envelope among prokaryotes, which cannot be assigned by the gram stain. The cell wall consists of thin layers of PG and arabinogalactan and a thick layer of mycolic acids. Additionally, they possess a single inner membrane [198]. Mycoplasmas do not possess a cell wall in general [200]. As GdpP and PgpH PDEs are both membrane-coupled, a different cell envelope might be the reason for the lack of these PDEs. In contrast to this theory, the gram-negative bacterium *T. maritima* has a so far uncharacterized homologue of PgpH. However, the absence of GdpP and PgpH in some species raises the issue of regulation. The degradation of c-di-AMP is assumed to be regulated by the additional non-catalytic domains of GdpP and PgpH, whereas DhhP does not contain any additional domains. Thus, it remains unclear how the degradation of c-di-AMP is regulated in those species, which only contain a DhhP homologue as single c-di-AMP specific PDE.

The proposed two-step degradation pathway for c-di-AMP raises another question. Why should bacteria possess an enzyme which catalyzes an incomplete hydrolysis of a second messenger and would therefore require a second enzyme for the final step? A possible explanation can be derived from the origin of DHH domains. They are also present in nanoRNases like NrnA. As these enzymes hydrolyze one phosphodiester linkage per base, they can catalyze the reaction with a single position in the active site. In contrast, c-di-AMP has two phosphodiester bonds for hydrolysis in opposite positions. The c-di-AMP specific PDEs could have evolved from nanoRNases and therefore contain a single hydrolysis site. Interestingly, all characterized PDEs for c-di-AMP degradation share the common two metal ion hydrolysis mechanism for phosphodiester linkages, which is only possible on one site. Thus, the substrate has to flip or leave the binding pocket and re-associate. Indeed, a flipping mechanism was proposed for the *M. tuberculosis* DhhP homologue Rv2837c by He *et al.* based on molecular docking and quantitative analysis [107]. However, this mechanism might be energetically less favoured compared to possessing a second enzyme for the final degradation step as it is observed in other

species. Another important aspect is the regulation of c-di-AMP hydrolysis. As the activity of GdpP and PgpH is most likely regulated by their additional domains, the substrate specificity of the catalytic domain is crucial. Accepting further substrates, such as 5′-pApA with high affinity in the same active site could interfere with a specific control of c-di-AMP levels. Since the second reaction step of degradation does not need to be regulated, it can be performed very effectively by DhhP PDEs. In addition, due to the phosphodiester linkages c-di-AMP adopts a more rigid conformation than the higher flexible 5′-pApA. Therefore, the active sites for these two molecules might require a different shape, as discussed in the previous chapter.

PgpH, GdpP and DhhP represent the most important enzymes for c-di-AMP hydrolysis in many species. However, some c-di-AMP producing bacteria do not possess any homologues of these enzymes as observed for many species of the typus Actinobacteria. Whereas *M. tuberculosis* and *M. smegmatis* use a DhhP homologue as single c-di-AMP specific PDE [45,107], other Actinobacteria, such as *Streptomyces* do not contain any of the so far characterized c-di-AMP specific PDEs. Therefore, it is unclear how c-di-AMP is hydrolyzed in these species. A recent discovery of a novel and non-canonical PDE by our collaborators (Tschowri group) gives important insights into the degradation of c-di-AMP in *S. venezuelae*. This PDE is mainly present in Actinobacteria and therefore termed AtaC (actinobacterial PDE targeting c-di-AMP hydrolysis). The present work made a substantial contribution to the identification of AtaC as c-di-AMP specific PDE. AtaC was shown to hydrolyze c-di-AMP to AMP with a high activity under physiological conditions. In high substrate concentrations, the formation of the intermediate 5′-pApA was observed. Notably, the turnover for 5′-pApA is ~2-fold higher than for c-di-AMP. Whereas DhhP homologues do not distinguish between adenosine and guanosine in linear dinucleotides [115], AtaC has only weak activity for c-di-GMP, 5′-pGpG or 5′-pApG. This indicates a high specificity for two adenosine residues in the active site. Unfortunately, no protein structure could be determined so far. Nevertheless, structural insights were derived from prediction models, SEC-RALS, SAXS analysis and mutational approaches. Based on an HHpred [201] analysis, the structural homologue with the highest similarity is predicted to be the phosphonoacetate hydrolase PhnA from *Sinorhizobium meliloti* [202]. Similar to PhnA, a monomeric structure was also indicated for AtaC by SEC-RALS experiments. In addition, an overall shape of AtaC obtained from SAXS analysis fits to the monomeric prediction model derived from HHpred [201]. AtaC was shown to bind Mn^{2+} ions, indicating a metal ion-dependent hydrolysis mechanism as common for c-di-AMP specific PDEs. Based on this observation in combination with the prediction model, conserved aspartate residues were suggested to be essential for Mn^{2+}-coordination and hydrolysis activity. This was verified for aspartate D269 as a mutation to asparagine (D269N) results in abolished Mn^{2+}-binding and activity. Similar to the TmPDE D80N D154N mutant, AtaC D269N is inactive but still able to bind c-di-AMP. A

monomeric structure with a single active site is also present in PDEs of the ENPP (Ecto-nucleotide pyrophosphatase phosphodiesterase) subfamily, which are predicted to be structurally similar to AtaC. Interestingly, one member of this subfamily, ENPP1, was shown to be important for degradation of 2′3′-cGAMP in mammalian cells [203]. Therefore, it can be suggested that AtaC and ENPP1 share a similar conserved active site, which is specific for cyclic dinucleotides.

Within the work of our collaboration partner Andreas Latoscha (Tschowri group) the function of AtaC was also analzyed *in vivo*. Significantly increased levels of c-di-AMP in AtaC deletion strains of *S. venezuelae* confirm the important role of this PDE for c-di-AMP degradation. Furthermore, the analysis of the AtaC deletion strain on solid sporulation medium reveals sporulation and growth defects. This indicates a correlation of c-di-AMP levels on sporulation and growth in *S. venezuelae*. Homologues of AtaC can also be found in *M. tuberculosis* and *M. smegmatis*. In both species, only a single c-di-AMP specific PDE has been discovered so far, while most species possess at least two PDEs for c-di-AMP hydrolysis. Thus, it is possible that these AtaC homologues can also hydrolyze c-di-AMP specifically and contribute to the c-di-AMP degradation. In conclusion, AtaC represents a new identified and highly interesting PDE for the degradation of c-di-AMP but requires many more investigations to be fully understood.

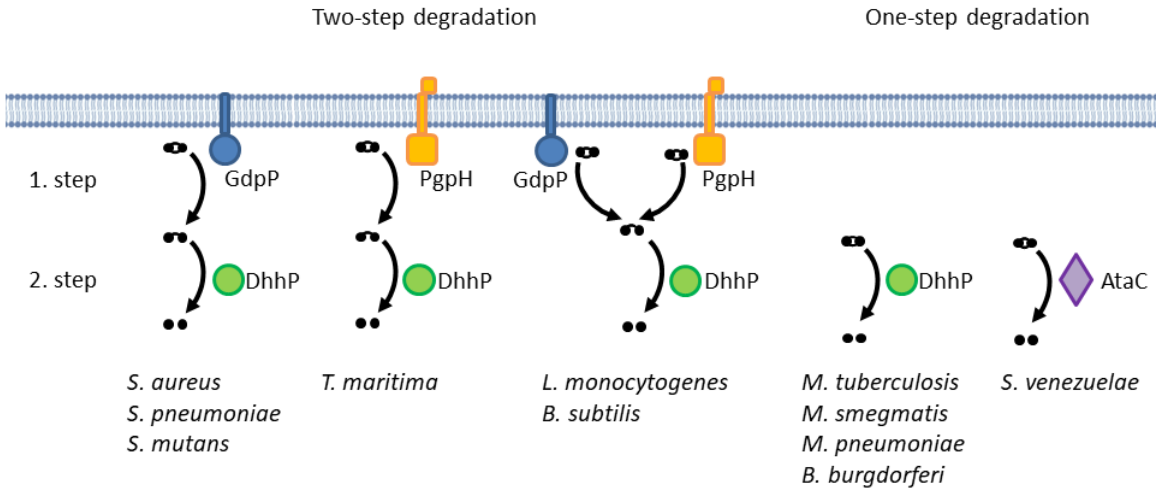


Figure 20: Proposed degradation pathways of c-di-AMP in different species. The membrane-coupled GdpP and PgpH PDEs hydrolyze c-di-AMP exclusively to 5′-pApA. The DhhP PDE has a preference for 5′-pApA hydrolysis, except for species that contain DhhP homologues as single c-di-AMP-degrading enzymes. AtaC is a non-canonical c-di-AMP specific PDE discovered in *S. venezuelae*.

3. Applications of a fluorescent 2′3′-cGAMP analogue

The cGAS-STING pathway is a key component of the mammalian innate immune system. It is crucial for the cytosolic recognition of different threats that derive from bacteria, viruses or defective cellular functions. Foreign DNA in the cytosol is detected by cGAS, which subsequently synthesizes 2′3′-cGAMP. Upon activation by 2′3′-cGAMP or bacterial CDNs,

STING induces a type I interferon response. Since an excessive activation of this pathway stands in correlation with autoimmune diseases, suitable inhibitors are of high interest in drug research. In recent years, the opposite modification, activation of cGAS or STING has become of particular relevance for potential cancer treatment approaches [157]. Therefore, assays to monitor the activity of both proteins are highly required. A very effective possibility to analyze protein interactions are fluorescence-based methods. Andreeva *et al.* developed a fluorescent high throughput screen to analyze the activity of cGAS [164]. This method replaces ATP by the fluorescent analogue 2-aminopurine riboside-5′-triphosphate (fATP). During the reaction, fATP is cyclized with GTP leading to a decrease in fluorescence, which can be measured [164]. However, the product of this reaction retains fluorescent properties. Within the present work, this molecule has been characterized as a fluorescent 2′3′-cGAMP analogue, termed fGAMP (Figure 21). Since fGAMP is highly similar to the physiological molecule, it is interesting for further applications in the cGAS-STING pathway. In particular, the analysis of STING interactions is relevant for drug development and therefore requires a high throughput screen. For an optimal set up of this type of assay, the following properties are important: The respective experiments have to be economical, fast in preparation and suitable for large-scale analysis. In addition, a simple readout without any modifications of the analyzed molecules ensures conditions close to the physiological situation. One commonly used method for STING interactions are thermal shift experiments by differential scanning fluorimetry (DSF) [181,184,190,192,204,205]. Ligand interaction has a significant influence on the stability of STING and thus results in a thermal shift proportional to the binding affinity. However, this method is limited in precision and cannot provide quantitative information. More precise protein interaction methods used for STING binding analysis, such as isothermal titration calorimetry (ITC) [178,180,206,207], are not suitable for large-scale applications.

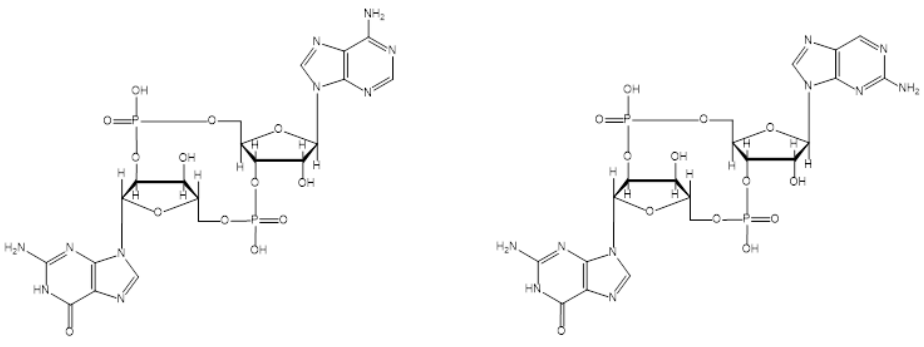


Figure 21: Structure of fGAMP compared to 2′3′-cGAMP.

The results of the present work indicate the potential of fGAMP in a fluorescence-based STING binding assay. In the desired model, STING binding to fGAMP results in fluorescence quenching. By adding a competitive ligand, the fluorescence is restored due to the release of fGAMP. Ligands with lower affinity do not show this effect unless applied in higher concentrations. Indeed, fGAMP was confirmed to be a suitable molecule regarding these properties. STING interacts with fGAMP, which leads to a decrease in fluorescence. The addition of increasing concentrations of 2′3′-cGAMP correlates with the fluorescence intensity. In contrast, c-di-GMP only induces a fluorescence increase when used at higher concentrations. This provides the possibility to monitor the binding event of STING by using a simple readout under different conditions. Moreover, transfection of fGAMP in HEK293T cells induces a type I interferon response, which indicates its similarity to the physiological pendant. Unfortunately, the experiments had to be performed without the possibility to calculate the concentration of fGAMP as the extinction coefficient was unknown. In addition, it has to be mentioned that the fluorescence quantum yield of fGAMP seems to be rather low and the molecule has to be applied in high amounts for detection. Subsequently, high concentrations of STING are necessary for the fluorescence quenching assays. Therefore, further investigations are required to optimize the set-up and establish a more sensitive assay based on the present results. However, a STING binding assay using fGAMP would have several advantages compared to common methods. It is highly similar to the physiological substrate and does not require further modifications of protein or ligand. Furthermore, the described assay has a simple readout suitable for large-scale applications. Collectively, this work provides the characterization of a novel 2′3′-cGAMP analogue with high potential for fluorescence-based STING applications.

References

[1] Kalia, D.; Merey, G.; Nakayama, S.; Zheng, Y.; Zhou, J.; Luo, Y.; Guo, M.; Roembke, B. T.; Sintim, H. O. Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chem. Soc. Rev.* **2013**, *42* (1), 305–341.

[2] Fahmi, T.; Faozia, S.; Port, G.; Cho, K. H. The Second Messenger c-di-AMP Regulates Diverse Cellular Pathways Involved in Stress Response, Biofilm Formation, Cell Wall Homeostasis, SpeB Expression and Virulence in Streptococcus pyogenes. *Infect. Immun.* **2019**, *8* (197).

[3] Aslam, B.; Wang, W.; Arshad, M. I.; Khurshid, M.; Rasool, M. H.; Nisar, M. A.; Aslam, M. A.; Qamar, M. U. Antibiotic resistance : a rundown of a global crisis. *Infect. Drug Resist.* **2018**, *11*, 1645–1658.

[4] Berthet, J.; Rall, T. W.; Sutherland, E. W. The relationship of epinephrine and glucagon to liver phosphorylase. IV. Effect of epinephrine and glucagon on the reactivation of phosphorylase in liver homogenates. *J. Biol. Chem.* **1957**, *224* (1), 463–475.

[5] Ashman, D. F.; Lipton, R.; Melicow, M. M.; Price, T. D. Isolation of adenosine 3′,5′-monophosphate and guanosine 3′,5′-monophosphate from rat urine. *Biochem. Biophys. Res. Commun.* **1963**, *11* (4), 330–334.

[6] Hardman, J. G.; Davis, J. W.; Sutherland, E. W. Measurement of Guanosine 3’,5’-Monophosphate and Other Cyclic Nucleotides. *J. Biol. Chem.* **1966**, *241* (20), 4812–4816.

[7] Cashel, M.; Kalbacher, B. The Control of Ribonucleic Acid Synthesis in Escherichia Coli. *J. Biol. Chem.* **1970**, *245* (9), 2309–2318.

[8] Sun, L.; Wu, J.; Du, F.; Chen, X.; Chen, Z. J. Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. *Science* **2013**, *339* (6121), 786–791.

[9] Wu, J.; Sun, L.; Chen, X.; Du, F.; Shi, H.; Chen, C.; Chen, Z. J. Cyclic GMP-AMP Is an Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. **2013**, No. February, 826–831.

[10] Ablasser, A.; Goldeck, M.; Cavlar, T.; Deimling, T.; Witte, G.; Röhl, I.; Hopfner, K.-P.; Ludwig, J.; Hornung, V. cGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING. *Nature* **2013**, *498*, 380.

[11] Gao, P.; Ascano, M.; Wu, Y.; Barchet, W.; Gaffney, B. L.; Zillinger, T.; Serganov, A. A.; Liu, Y.; Jones, R. A.; Hartmann, G.; Tuschl, T.; Patel, D. J. Metazoan Second Messenger Produced by DNA-Activated Cyclic GMP-AMP Synthase. *Cell* **2013**, *153* (5), 1094–1107.

[12] Diner, E. J.; Burdette, D. L.; Wilson, S. C.; Monroe, K. M.; Kellenberger, C. A.; Hyodo, M.; Hayakawa, Y.; Hammond, M. C.; Vance, R. E. Report The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. *Cell Rep.* **2013**, *3* (5), 1355–1361.

[13] Whiteley, A. T.; Eaglesham, J. B.; de Oliveira Mann, C. C.; Morehouse, B. R.; Lowey, B.; Nieminen, E. A.; Danilchanka, O.; King, D. S.; Lee, A. S. Y.; Mekalanos, J. J.; Kranzusch, P. J. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **2019**, *567* (7747), 194–199.

[14] Ross, P.; Weinhouse, H.; Aloni, Y.; Michaeli, D.; Weinberger-Ohana, P.; Mayer, R.; Braun, S.; de Vroom, E.; van der Marel, G. A.; van Boom, J. H.; Benziman, M. Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* **1987**, *325* (6101), 279–281.

[15] Tischler, A. D.; Camilli, A. Cyclic diguanylate (c-di-GMP) regulates Vibrio cholerae biofilm formation. *Mol. Microbiol.* **2004**, *53* (3), 857–869.

[16] Simm, R.; Morr, M.; Kader, A.; Nimtz, M.; Römling, U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. *Mol. Microbiol.* **2004**, *53* (4), 1123–1134.

[17] Valentini, M.; Filloux, A. Biofilms and Cyclic di-GMP (c-di-GMP) signaling: Lessons from Pseudomonas aeruginosa and other bacteria. *J. Biol. Chem.* **2016**, *291* (24), 12547–12555.

[18] Sarenko, O.; Klauck, G.; Wilke, F. M.; Pfiff, V.; Richter, A. M.; Herbst, S.; Kaever, V.; Hengge, R. More than Enzymes That Make or Break Cyclic Di-GMP — Local Signaling in the. *MBio* **2017**, *8* (5), 1–18.

[19] Wilksch, J. J.; Yang, J.; Clements, A.; Gabbe, J. L.; Short, K. R.; Cao, H.; Cavaliere, R.; James, C. E.; Whitchurch, C. B.; Schembri, M. A.; Chuah, M. L. C.; Liang, Z. X.; Wijburg, O. L.; Jenney, A. W.; Lithgow, T.; Strugnell, R. A. MrKH, a novel c-di-GMP-dependent transcriptional activator, controls klebsiella pneumoniae biofilm formation by regulating type 3 fimbriae expression. *PLoS Pathog.* **2011**, *7* (8), e1002204.

[20] Romling, U.; Galperin, M. Y.; Gomelsky, M. Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiol. Mol. Biol. Rev.* **2013**, *77* (1), 1–52.

[21] Davies, B. W.; Bogard, R. W.; Young, T. S.; Mekalanos, J. J. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. *Cell* **2012**, *149* (2), 358–370.

[22] Severin, G. B.; Ramliden, M. S.; Hawver, L. A.; Wang, K.; Pell, M. E.; Kieninger, A.-K.; Khataokar, A.; O'Hara, B. J.; Behrmann, L. V.; Neiditch, M. B.; Benning, C.; Waters, C. M.; Ng, W.-L. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor Vibrio cholerae. *Proc. Natl. Acad. Sci.* **2018**, *115* (26), E6048–E6055.

[23] Kellenberger, C. A.; Wilson, S. C.; Hickey, S. F.; Gonzalez, T. L.; Su, Y.; Hallberg, Z. F.; Brewer, T. F.; Iavarone, A. T.; Carlson, H. K.; Hsieh, Y.-F.; Hammond, M. C. GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP . *Proc. Natl. Acad. Sci.* **2015**, *112* (17), 5383–5388.

[24] Nelson, J. W.; Sudarsan, N.; Phillips, G. E.; Stav, S.; Lünse, C. E.; McCown, P. J.; Breaker, R. R. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc. Natl. Acad. Sci.* **2015**, *112* (17), 5389–5394.

[25] Li, F.; Cimdins, A.; Rohde, M.; Jänsch, L.; Kaever, V.; Nimtz, M.; Römling, U. DncV Synthesizes Cyclic GMP-AMP and Regulates Biofilm Formation and Motility in Escherichia coli ECOR31 . *MBio* **2019**, *10* (2), 1–21.

[26] Kato, K.; Ishii, R.; Hirano, S.; Ishitani, R.; Nureki, O. Structural basis for the catalytic mechanism of DncV, bacterial homolog of cyclic GMP-AMP synthase. *Structure* **2015**, *23* (5), 843–850.

[27] Hallberg, Z. F.; Wang, X. C.; Wright, T. A.; Nan, B.; Ad, O.; Yeo, J.; Hammond, M. C. Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3′, 3′-cGAMP). *Proc. Natl. Acad. Sci.* **2016**, *113* (7), 1790–1795.

[28] Gao, J.; Tao, J.; Liang, W.; Zhao, M.; Du, X.; Cui, S.; Duan, H.; Kan, B.; Su, X.; Jiang, Z. Identification and characterization of phosphodiesterases that specifically degrade 3′3′-cyclic GMP-AMP. *Cell Res.* **2015**, *25* (5), 539–550.

[29] Witte, G.; Hartung, S.; Büttner, K.; Hopfner, K. P. Structural Biochemistry of a Bacterial Checkpoint Protein Reveals Diadenylate Cyclase Activity Regulated by DNA

Recombination Intermediates. *Mol. Cell* **2008**, *30* (2), 167–178.

[30] Whiteley, A. T.; Pollock, A. J.; Portnoy, D. A. The PAMP c-di-AMP Is Essential for Listeria monocytogenes Growth in Rich but Not Minimal Media due to a Toxic Increase in (p)ppGpp. *Cell Host Microbe* **2015**, *17* (6), 788–798.

[31] Song, J.-H.; Ko, K. S.; Lee, J.-Y.; Baek, J. Y.; Oh, W. S.; Yoon, H. S.; Jeong, J.-Y.; Chun, J. Identification of Essential Genes in Streptococcus pneumoniae by Allelic Replacement Mutagenesis. *Mol. Cells* **2005**, *19* (3), 365–374.

[32] Bai, Y.; Yang, J.; Eisele, L. E.; Underwood, A. J.; Koestler, B. J.; Waters, C. M.; Metzger, D. W.; Bai, G. Two DHH subfamily 1 proteins in Streptococcus pneumoniae possess cyclic Di-AMP phosphodiesterase activity and affect bacterial growth and virulence. *J. Bacteriol.* **2013**, *195* (22), 5123–5132.

[33] Chaudhuri, R. R.; Allen, A. G.; Owen, P. J.; Shalom, G.; Stone, K.; Harrison, M.; Burgis, T. A.; Lockyer, M.; Garcia-Lara, J.; Foster, S. J.; Pleasance, S. J.; Peters, S. E.; Maskell, D. J.; Charles, I. G. Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics* **2009**, *10* (291).

[34] Corrigan, R. M.; Bowman, L.; Willis, A. R.; Kaever, V.; Gründling, A. Cross-talk between two nucleotide-signaling pathways in staphylococcus aureus. *J. Biol. Chem.* **2015**, *290* (9), 5826–5839.

[35] Luo, Y.; Helmann, J. D. Analysis of the role of Bacillus subtilis σM in β-lactam resistance reveals an essential role for c-di-AMP in peptidoglycan homeostasis. *Mol. Microbiol.* **2012**, *83* (3), 623–639.

[36] Mehne, F. M. P.; Gunka, K.; Eilers, H.; Herzberg, C.; Kaever, V.; Stülke, J. Cyclic Di-AMP homeostasis in bacillus subtilis: Both lack and high level accumulation of the nucleotide are detrimental for cell growth. *J. Biol. Chem.* **2013**, *288* (3), 2004–2017.

[37] Blötz, C.; Treffon, K.; Kaever, V.; Schwede, F.; Hammer, E.; Stülke, J. Identification of the components involved in cyclic di-AMP signaling in Mycoplasma pneumoniae. *Front. Microbiol.* **2017**, *8* (JUL), 1–10.

[38] El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C. E.; Finn, R. D. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47* (D1), D427–D432.

[39] Müller, M.; Deimling, T.; Hopfner, K.-P.; Witte, G. Structural analysis of the diadenylate cyclase reaction of DNA-integrity scanning protein A (DisA) and its inhibition by 3′-dATP. *Biochem. J.* **2015**, *469* (3), 367 LP – 374.

[40] Kellenberger, C. A.; Chen, C.; Whiteley, A. T.; Portnoy, D. A.; Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. *J. Am. Chem. Soc.* **2015**, *137* (20), 6432–6435.

[41] Braun, F.; Thomalla, L.; Does, C.; Quax, T. E. F.; Allers, T.; Kaever, V.; Albers, S. Cyclic nucleotides in archaea: Cyclic di-AMP in the archaeon Haloferax volcanii and its putative role. *Microbiologyopen* **2019**, *e829*.

[42] Commichau, F. M.; Heidemann, J. L.; Ficner, R.; Stülke, J. Making and breaking of an essential poison: The cyclases and phosphodiesterases that produce and degrade the essential second messenger cyclic di-AMP in bacteria. *J. Bacteriol.* **2019**, *201* (1), 1–14.

[43] Bejerano-Sagie, M.; Oppenheimer-Shaanan, Y.; Berlatzky, I.; Rouvinski, A.; Meyerovich, M.; Ben-Yehuda, S. A Checkpoint Protein That Scans the Chromosome for Damage at the

Start of Sporulation in Bacillus subtilis. *Cell* **2006**, *125* (4), 679–690.

[44] Manikandan, K.; Sabareesh, V.; Singh, N.; Saigal, K.; Mechold, U.; Sinha, K. M. Two-step synthesis and hydrolysis of cyclic di-AMP in Mycobacterium tuberculosis. *PLoS One* **2014**, *9* (1).

[45] Tang, Q.; Luo, Y.; Zheng, C.; Yin, K.; Ali, M. K.; Li, X.; He, J. Functional analysis of a c-di-AMP-specific phosphodiesterase MsPDE from Mycobacterium smegmatis. *Int. J. Biol. Sci.* **2015**, *11* (7), 813–824.

[46] Oppenheimer-Shaanan, Y.; Wexselblatt, E.; Katzhendler, J.; Yavin, E.; Ben-Yehuda, S. C-di-AMP reports DNA integrity during sporulation in Bacillus subtilis. *EMBO Rep.* **2011**, *12* (6), 594–601.

[47] Zhang, L.; He, Z. G. Radiation-sensitive gene A (RadA) targets DisA, DNA integrity scanning protein a, to negatively affect cyclic Di-AMP synthesis activity in Mycobacterium smegmatis. *J. Biol. Chem.* **2013**, *288* (31), 22426–22436.

[48] Gándara, C.; de Lucena, D. K. C.; Torres, R.; Serrano, E.; Altenburger, S.; Graumann, P. L.; Alonso, J. C. Activity and in vivo dynamics of Bacillus subtilis DisA are affected by RadA/Sms and by Holliday junction-processing proteins. *DNA Repair (Amst).* **2017**, *55* (May), 17–30.

[49] Valenzuela-García, L. I.; Ayala-García, V. M.; Regalado-García, A. G.; Setlow, P.; Pedraza-Reyes, M. Transcriptional coupling (Mfd) and DNA damage scanning (DisA) coordinate excision repair events for efficient Bacillus subtilis spore outgrowth. *Microbiologyopen* **2018**, *7* (5), 1–15.

[50] Tosi, T.; Hoshiga, F.; Millership, C.; Singh, R.; Eldrid, C.; Patin, D.; Mengin-Lecreulx, D.; Thalassinos, K.; Freemont, P.; Gründling, A. Inhibition of the Staphylococcus aureus c-di-AMP cyclase DacA by direct interaction with the phosphoglucosamine mutase GlmM. *PLoS Pathog.* **2019**, *15* (1), 1–28.

[51] Rosenberg, J.; Dickmanns, A.; Neumann, P.; Gunka, K.; Arens, J.; Kaever, V.; Stülke, J.; Ficner, R.; Commichau, F. M. Structural and biochemical analysis of the essential diadenylate cyclase CdaA from Listeria monocytogenes. *J. Biol. Chem.* **2015**, *290* (10), 6596–6606.

[52] Heidemann, J. L.; Neumann, P.; Dickmanns, A.; Ficner, R. Crystal structures of the c-di-AMP synthesizing enzyme CdaA. *J. Biol. Chem.* **2019**, *294* (27), 10463–10470.

[53] Pham, T. H.; Liang, Z. X.; Marcellin, E.; Turner, M. S. Replenishing the cyclic-di-AMP pool: regulation of diadenylate cyclase activity in bacteria. *Curr. Genet.* **2016**, *62* (4), 731–738.

[54] Rismondo, J.; Halbedel, S.; Kaever, V.; Rosenberg, J.; Commichau, F. M.; Gibhardt, J. Phenotypes Associated with the Essential Diadenylate Cyclase CdaA and Its Potential Regulator CdaR in the Human Pathogen Listeria monocytogenes. *J. Bacteriol.* **2015**, *198* (3), 416–426.

[55] Bowman, L.; Zeden, M. S.; Schuster, C. F.; Kaever, V.; Gründling, A. New insights into the cyclic di-adenosine monophosphate (c-di-AMP) degradation pathway and the requirement of the cyclic dinucleotide for acid stress resistance in Staphylococcus aureu. *J. Biol. Chem.* **2016**, *291* (53), 26970–26986.

[56] Mengin-lecreulx, D.; Mengin-lecreulx, D.; Heijenoort, J. Van; Heijenoort, J. Van. Characterization of the Essential Gene. *Mol. Biol.* **1996**, *271* (1), 32–39.

[57] Zhu, Y.; Pham, T. H.; Nhiep, T. H. N.; Vu, N. M. T.; Marcellin, E.; Chakrabortti, A.; Wang, Y.; Waanders, J.; Lo, R.; Huston, W. M.; Bansal, N.; Nielsen, L. K.; Liang, Z. X.;

Turner, M. S. Cyclic-di-AMP synthesis by the diadenylate cyclase CdaA is modulated by the peptidoglycan biosynthesis enzyme GlmM in lactococcus lactis. *Mol. Microbiol.* **2016**, *99* (6), 1015–1027.

[58] Mehne, F. M. P.; Schröder-Tittmann, K.; Eijlander, R. T.; Herzberg, C.; Hewitt, L.; Kaever, V.; Lewis, R. J.; Kuipers, O. P.; Tittmann, K.; Stülke, J. Control of the Diadenylate Cyclase CdaS in Bacillus subtilis . *J. Biol. Chem.* **2014**, *289* (30), 21098–21107.

[59] Mehne, F. M. P.; Gunka, K.; Eilers, H.; Herzberg, C.; Kaever, V.; Stülke, J. Cyclic Di-AMP Homeostasis in Bacillus subtilis . *J. Biol. Chem.* **2012**, *288* (3), 2004–2017.

[60] Soufo, H. J. D. A novel cell type enables B. Subtilis to escape from unsuccessful sporulation in minimal medium. *Front. Microbiol.* **2016**, *7* (1810).

[61] Setlow, P. Spores of Bacillus subtilis: Their resistance to and killing by radiation, heat and chemicals. *J. Appl. Microbiol.* **2006**, *101* (3), 514–525.

[62] Ghosh, J.; Larsson, P.; Singh, B.; Pettersson, B. M. F.; Islam, N. M.; Sarkar, S. N.; Dasgupta, S.; Kirsebom, L. A. Sporulation in mycobacteria. *Proc. Natl. Acad. Sci.* **2009**, *106* (26), 10781–10786.

[63] Corrigan, R. M.; Abbott, J. C.; Burhenne, H.; Kaever, V.; Gründling, A. C-di-amp is a new second messenger in staphylococcus aureus with a role in controlling cell size and envelope stress. *PLoS Pathog.* **2011**, *7* (9).

[64] Dey, B.; Dey, R. J.; Cheung, L. S.; Pokkali, S.; Guo, H.; Lee, J. H.; Bishai, W. R. A bacterial cyclic dinucleotide activates the cytosolic surveillance pathway and mediates innate resistance to tuberculosis. *Nat. Med.* **2015**, *21* (4), 401–408.

[65] Du, B.; Ji, W.; An, H.; Shi, Y.; Huang, Q.; Cheng, Y.; Fu, Q.; Wang, H.; Yan, Y.; Sun, J. Functional analysis of c-di-AMP phosphodiesterase, GdpP, in Streptococcus suis serotype 2. *Microbiol. Res.* **2014**, *169* (9–10), 749–758.

[66] Dengler, V.; McCallum, N.; Kiefer, P.; Christen, P.; Patrignani, A.; Vorholt, J. A.; Berger-Bächi, B.; Senn, M. M. Mutation in the C-Di-AMP Cyclase dacA Affects Fitness and Resistance of Methicillin Resistant Staphylococcus aureus. *PLoS One* **2013**, *8* (8).

[67] Witte, C. E.; Whiteley, A. T.; Burke, T. P.; Sauer, J.-D.; Portnoy, D. A.; Woodward, J. J. Cyclic di-AMP is critical for Listeria monocytogenes growth, cell wall homeostasis, and establishment of infection. *MBio* **2013**, *4* (3), e00282-13.

[68] Gundlach, J.; Commichau, F. M.; Stülke, J. Perspective of ions and messengers: an intricate link between potassium, glutamate, and cyclic di-AMP. *Curr. Genet.* **2018**, *64* (1), 191–195.

[69] Corrigan, R. M.; Campeotto, I.; Jeganathan, T.; Roelofs, K. G.; Lee, V. T.; Grundling, A. Systematic identification of conserved bacterial c-di-AMP receptor proteins. *Proc. Natl. Acad. Sci.* **2013**, *110* (22), 9084–9089.

[70] Moscoso, J. A.; Schramke, H.; Zhang, Y.; Tosi, T.; Dehbi, A.; Jung, K.; Gründling, A. Binding of cyclic di-AMP to the Staphylococcus aureus sensor kinase KdpD occurs via the universal stress protein domain and downregulates the expression of the Kdp potassium transporter. *J. Bacteriol.* **2016**, *198* (1), 98–110.

[71] Bai, Y.; Yang, J.; Zarrella, T. M.; Zhang, Y.; Metzger, D. W.; Bai, G. Cyclic Di-AMP impairs potassium uptake mediated by a cyclic Di-AMP binding protein in streptococcus Pneumoniae. *J. Bacteriol.* **2014**, *196* (3), 614–623.

[72] Quintana, I. M.; Gibhardt, J.; Turdiev, A.; Hammer, E.; Commichau, F. M.; Lee, V. T.;

Magni, C.; Stülke, J. The KupA and KupB Proteins of Lactococcus Lactis IL1403 Are Novel c-di-AMP Receptor Proteins Responsible for Potassium Uptake. *J. Bacteriol.* **2019**, *201* (10), 1–13.

[73] Gundlach, J.; Herzberg, C.; Kaever, V.; Gunka, K.; Hoffmann, T.; Weiß, M.; Gibhardt, J.; Thürmer, A.; Hertel, D.; Daniel, R.; Bremer, E.; Commichau, F. M.; Stülke, J. Control of potassium homeostasis is an essential function of the second messenger cyclic di-AMP in Bacillus subtilis. *Sci. Signal.* **2017**, *10* (475), eaal3011.

[74] Chin, K. H.; Liang, J. M.; Yang, J. G.; Shih, M. S.; Tu, Z. Le; Wang, Y. C.; Sun, X. H.; Hu, N. J.; Liang, Z. X.; Dow, J. M.; Ryan, R. P.; Chou, S. H. Structural Insights into the Distinct Binding Mode of Cyclic Di-AMP with SaCpaA-RCK. *Biochemistry* **2015**, *54* (31), 4936–4951.

[75] Vieira-Pires, R. S.; Szollosi, A.; Morais-Cabral, J. H. The structure of the KtrAB potassium transporter. *Nature* **2013**, *496* (7445), 323–328.

[76] Holtmann, G.; Bakker, E. P.; Uozumi, N.; Bremer, E. KtrAB and KtrCD: Two K+ Uptake Systems in Bacillus subtillis and Their Role in Adaptation to Hypertonicity. *Society* **2003**, *185* (4), 1289–1298.

[77] Kröning, N.; Willenborg, M.; Tholema, N.; Hänelt, I.; Schmid, R.; Bakker, E. P. ATP Binding to the KTN/RCK subunit KtrA from the K+-uptake system KtrAB of Vibrio alginolyticus: Its role in the formation of the KtrAB complex and its requirement in vivo. *J. Biol. Chem.* **2007**, *282* (19), 14018–14027.

[78] Kim, H.; Youn, S. J.; Kim, S. O.; Ko, J.; Lee, J. O.; Choi, B. S. Structural studies of potassium transport protein KtrA regulator of conductance of K^+ (RCK) C domain in complex with cyclic diadenosine monophosphate (c-di-AMP). *J. Biol. Chem.* **2015**, *290* (26), 16393–16402.

[79] Devaux, L.; Sleiman, D.; Mazzuoli, M. V.; Gominet, M.; Lanotte, P.; Trieu-Cuot, P.; Kaminski, P. A.; Firon, A. Cyclic di-AMP regulation of osmotic homeostasis is essential in Group B Streptococcus. *PLoS Genet.* **2018**, *14* (4), 1–23.

[80] Wang, X.; Cai, X.; Ma, H.; Yin, W.; Zhu, L.; Li, X.; Lim, H. M.; Chou, S.-H.; He, J. A c-di-AMP riboswitch controlling kdpFABC operon transcription regulates the potassium transporter system in Bacillus thuringiensis. *Commun. Biol.* **2019**, *2* (1), 151.

[81] Nelson, J. W.; Sudarsan, N.; Furukawa, K.; Weinberg, Z.; Wang, J. X.; Breaker, R. R. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat. Chem. Biol.* **2013**, *9* (12), 834–839.

[82] Commichau, F. M.; Gibhardt, J.; Halbedel, S.; Gundlach, J.; Stülke, J. A Delicate Connection: c-di-AMP Affects Cell Integrity by Controlling Osmolyte Transport. *Trends Microbiol.* **2018**, *26* (3), 175–185.

[83] Schuster, C. F.; Bellows, L. E.; Tosi, T.; Campeotto, I.; Corrigan, R. M.; Freemont, P.; Gründling, A. The second messenger c-di-AMP inhibits the osmolyte uptake system OpuC in Staphylococcus aureus. *Sci. Signal.* **2016**, *9* (441), 1–31.

[84] Huynh, T. A. N.; Choi, P. H.; Sureka, K.; Ledvina, H. E.; Campillo, J.; Tong, L.; Woodward, J. J. Cyclic di-AMP targets the cystathionine beta-synthase domain of the osmolyte transporter OpuC. *Mol. Microbiol.* **2016**, *102* (2), 233–243.

[85] Verheul, A.; Rombouts, F. M.; Beumer, R. R.; Abee, T. An ATP-dependent L-carnitine transporter in Listeria monocytogenes Scott a is involved in osmoprotection. *J. Bacteriol.* **1995**, *177* (11), 3205–3212.

[86] Pham, H. T.; Nhiep, N. T. H.; Vu, T. N. M.; Huynh, T. A. N.; Zhu, Y.; Huynh, A. L. D.;

Chakrabortti, A.; Marcellin, E.; Lo, R.; Howard, C. B.; Bansal, N.; Woodward, J. J.; Liang, Z. X.; Turner, M. S. Enhanced uptake of potassium or glycine betaine or export of cyclic-di-AMP restores osmoresistance in a high cyclic-di-AMP Lactococcus lactis mutant. *PLoS Genet.* **2018**, *14* (8), 1–23.

[87] Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **2009**, *7* (4), 263–273.

[88] Townsley, L.; Yannarell, S. M.; Huynh, T. N.; Woodward, J. J.; Shank, E. A. Cyclic di-AMP Acts as an Extracellular Signal That Impacts Bacillus subtilis Biofilm Formation and Plant Attachment. *MBio* **2018**, *9* (2), e00341-18.

[89] Peng, X.; Zhang, Y.; Bai, G.; Zhou, X.; Wu, H. Cyclic di-AMP mediates biofilm formation. *Mol. Microbiol.* **2016**, *99* (5), 945–959.

[90] Gundlach, J.; Rath, H.; Herzberg, C.; Mäder, U.; Stülke, J. Second messenger signaling in Bacillus Subtilis: Accumulation of cyclic di-AMP inhibits biofilm formation. *Front. Microbiol.* **2016**, *7* (804).

[91] Limoli, D. H.; Jones, C. J.; Wozniak, D. J. Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. *Microbiol. Spectr.* **2015**, *3* (3), 1–19.

[92] Cho, K. H.; Kang, S. O. Streptococcus pyogenes c-di-AMP Phosphodiesterase, GdpP, Influences SpeB Processing and Virulence. *PLoS One* **2013**, *8* (7), 1–12.

[93] Yang, J.; Bai, Y.; Zhang, Y.; Gabrielle, V. D.; Jin, L.; Bai, G. Deletion of the cyclic di-AMP phosphodiesterase gene (cnpB) in Mycobacterium tuberculosis leads to reduced virulence in a mouse model of infection. *Mol. Microbiol.* **2014**, *93* (1), 65–79.

[94] Huynh, T. N.; Luo, S.; Pensinger, D.; Sauer, J.-D.; Tong, L.; Woodward, J. J. An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. *Proc. Natl. Acad. Sci.* **2015**, *112* (7), E747–E756.

[95] Zhang, L.; Li, W.; He, Z. DarR , a TetR-like Transcriptional Factor , Is a Cyclic Di-AMP-responsive Repressor in Mycobacterium smegmatis * □. **2013**, *288* (5), 3085–3096.

[96] Sureka, K.; Choi, P. H.; Precit, M.; Delince, M.; Pensinger, D. A.; Huynh, T. N.; Jurado, A. R.; Goo, Y. A.; Sadilek, M.; Iavarone, A. T.; Sauer, J.-D.; Tong, L.; Woodward, J. J. The Cyclic Dinucleotide c-di-AMP Is an Allosteric Regulator of Metabolic Enzyme Function. *Cell* **2014**, *158* (6), 1389–1401.

[97] Choi, P. H.; Vu, T. M. N.; Pham, H. T.; Woodward, J. J.; Turner, M. S.; Tong, L. Structural and functional studies of pyruvate carboxylase regulation by cyclic di-AMP in lactic acid bacteria. *Proc. Natl. Acad. Sci.* **2017**, *114* (35), E7226–E7235.

[98] Merrick, M. Post-translational modification of PII signal transduction proteins. *Front. Microbiol.* **2015**, *6* (763).

[99] Conroy, M. J.; Durand, A.; Lupo, D.; Li, X.-D.; Bullough, P. A.; Winkler, F. K.; Merrick, M. The crystal structure of the Escherichia coli AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. *Proc. Natl. Acad. Sci.* **2007**, *104* (4), 1213–1218.

[100] Rajendran, C.; Gerhardt, E. C. M.; Bjelic, S.; Gasperina, A.; Scarduelli, M.; Pedrosa, F. O.; Chubatsu, L. S.; Merrick, M.; Souza, E. M.; Winkler, F. K.; Huergo, L. F.; Li, X.-D. Crystal structure of the GlnZ-DraG complex reveals a different form of PII-target interaction. *Proc. Natl. Acad. Sci.* **2011**, *108* (47), 18972–18976.

[101] Campeotto, I.; Zhang, Y.; Mladenov, M. G.; Freemont, P. S.; Gründling, A. Complex structure and biochemical characterization of the Staphylococcus aureus cyclic diadenylate monophosphate (c-di-AMP)-binding protein PstA, the founding member of a new signal

transduction protein family. *J. Biol. Chem.* **2015**, *290* (5), 2888–2901.

[102] Choi, P. H.; Sureka, K.; Woodward, J. J.; Tong, L. Molecular basis for the recognition of cyclic-di-AMP by PstA, a PII-like signal transduction protein. *Microbiologyopen* **2015**, *4* (3), 361–374.

[103] Gundlach, J.; Dickmanns, A.; Schröder-Tittmann, K.; Neumann, P.; Kaesler, J.; Kampf, J.; Herzberg, C.; Hammer, E.; Schwede, F.; Kaever, V.; Tittmann, K.; Stülke, J.; Ficner, R. Identification, characterization, and structure analysis of the cyclic di-AMP-binding PII-like signal transduction protein DarA. *J. Biol. Chem.* **2015**, *290* (5), 3069–3080.

[104] Müller, M.; Hopfner, K. P.; Witte, G. C-di-AMP recognition by Staphylococcus aureus PstA. *FEBS Lett.* **2015**, *589* (1), 45–51.

[105] Gundlach, J.; Krüger, L.; Herzberg, C.; Turdiev, A.; Poehlein, A.; Tascón, I.; Weiß, M.; Hertel, D.; Daniel, R.; Hänelt, I.; Lee, V. T.; Stülke, J. Sustained sensing in potassium homeostasis : Cyclic di-AMP controls potassium uptake by KimA at the levels of expression and activity. **2019**, 1–20.

[106] Mitić, N.; Smith, S. J.; Neves, A.; Guddat, L. W.; Gahan, L. R.; Schenk, G. The Catalytic Mechanisms of Binuclear Metallohydrolases. *Chem. Rev.* **2006**, *106* (8), 3338–3363.

[107] He, Q.; Wang, F.; Liu, S.; Zhu, D.; Cong, H.; Gao, F.; Li, B.; Wang, H.; Lin, Z.; Liao, J.; Gu, L. Structural and biochemical insight into the mechanism of Rv2837c from mycobacterium tuberculosis as a c-di-NMP phosphodiesterase. *J. Biol. Chem.* **2016**, *291* (7), 3668–3681.

[108] Woodward, J. J.; Lavarone, A. T.; Portnoy, D. A. C-di-AMP secreted by intracellular Listeria monocytogenes activates a host type I interferon response. *Science* **2010**, *328* (5986), 1703–1705.

[109] Andrade, W. A.; Firon, A.; Schmidt, T.; Hornung, V.; Fitzgerald, K. A.; Kurt-Jones, E. A.; Trieu-Cuot, P.; Golenbock, D. T.; Kaminski, P. A. Group B Streptococcus Degrades Cyclic-di-AMP to Modulate STING-Dependent Type I Interferon Production. *Cell Host Microbe* **2016**, *20* (1), 49–59.

[110] Huynh, T. A. N.; Woodward, J. J. Too much of a good thing: Regulated depletion of c-di-AMP in the bacterial cytoplasm. *Curr. Opin. Microbiol.* **2016**, *30*, 22–29.

[111] Aravind, L.; Koonin, E. V. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem. Sci.* **1998**, *23* (12), 469–472.

[112] Uemura, Y.; Nakagawa, N.; Wakamatsu, T.; Kim, K.; Montelione, G. T.; Hunt, J. F.; Kuramitsu, S.; Masui, R. Crystal structure of the ligand-binding form of nanoRNase from Bacteroides fragilis, a member of the DHH/DHHA1 phosphoesterase family of proteins. *FEBS Lett.* **2013**, *587* (16), 2669–2674.

[113] Rao, F.; See, R. Y.; Zhang, D.; Toh, D. C.; Ji, Q.; Liang, Z. X. YybT is a signaling protein that contains a cyclic dinucleotide phosphodiesterase domain and a GGDEF domain with ATPase activity. *J. Biol. Chem.* **2010**, *285* (1), 473–482.

[114] Ye, M.; Zhang, J.-J.; Fang, X.; Lawlis, G. B.; Troxell, B.; Zhou, Y.; Gomelsky, M.; Lou, Y.; Yang, X. F. DhhP, a Cyclic di-AMP Phosphodiesterase of Borrelia burgdorferi, Is Essential for Cell Growth and Virulence. *Infect. Immun.* **2014**, *82* (5), 1840–1849.

[115] Drexler, D. J.; Müller, M.; Rojas-Cordova, C. A.; Bandera, A. M.; Witte, G. Structural and Biophysical Analysis of the Soluble DHH/DHHA1-Type Phosphodiesterase TM1595 from Thermotoga maritima. *Structure* **2017**, *25* (12), 1887-1897.e4.

[116] Wang, F.; He, Q.; Su, K.; Wei, T.; Xu, S.; Gu, L. Structural and biochemical

characterization of the catalytic domains of GdpP reveals a unified hydrolysis mechanism for the DHH/DHHA1 phosphodiesterase. *Biochem. J.* **2017**, *475* (1), 191–205.

[117] Wakamatsu, T.; Kim, K.; Uemura, Y.; Nakagawa, N.; Kuramitsu, S.; Masui, R. Role of RecJ-like Protein with 5′-3′ exonuclease activity in oligo(deoxy)nucleotide degradation. *J. Biol. Chem.* **2011**, *286* (4), 2807–2816.

[118] Srivastav, R.; Kumar, D.; Grover, A.; Singh, A.; Manjasetty, B. A.; Sharma, R.; Taneja, B. Unique subunit packing in mycobacterial nanoRNase leads to alternate substrate recognitions in DHH phosphodiesterases. *Nucleic Acids Res.* **2014**, *42* (12), 7894–7910.

[119] Rao, F.; Ji, Q.; Soehano, I.; Liang, Z. X. Unusual heme-binding PAS domain from YybT family proteins. *J. Bacteriol.* **2011**, *193* (7), 1543–1551.

[120] Tan, E.; Rao, F.; Pasunooti, S.; Pham, T. H.; Soehano, I.; Turner, M. S.; Liew, C. W.; Lescar, J.; Pervushin, K.; Liang, Z. X. Solution structure of the PAS domain of a thermophilic YybT protein homolog reveals a potential ligand-binding site. *J. Biol. Chem.* **2013**, *288* (17), 11949–11959.

[121] Holland, L. M.; O'Donnell, S. T.; Ryjenkov, D. A.; Gomelsky, L.; Slater, S. R.; Fey, P. D.; Gomelsky, M.; O'Gara, J. P. A staphylococcal GGDEF domain protein regulates biofilm formation independently of cyclic dimeric GMP. *J. Bacteriol.* **2008**, *190* (15), 5178–5189.

[122] Christen, M.; Christen, B.; Folcher, M.; Schauerte, A.; Jenal, U. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J. Biol. Chem.* **2005**, *280* (35), 30829–30837.

[123] Konno, H.; Yoshida, Y.; Nagano, K.; Takebe, J.; Hasegawa, Y. Biological and Biochemical Roles of Two Distinct Cyclic Dimeric Adenosine 3',5'-Monophosphate-Associated Phosphodiesterases in Streptococcus mutans. *Front. Microbiol.* **2018**, *9* (September), 2347.

[124] Cron, L. E.; Stol, K.; Burghout, P.; van Selm, S.; Simonetti, E. R.; Bootsma, H. J.; Hermans, P. W. M. Two DHH subfamily 1 proteins contribute to pneumococcal virulence and confer protection against pneumococcal disease. *Infect. Immun.* **2011**, *79* (9), 3697–3710.

[125] Anantharaman, V.; Aravind, L. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. *BMC Genomics* **2003**, *4*, 1–20.

[126] Gundlach, J.; Mehne, F. M. P.; Herzberg, C.; Kampf, J.; Valerius, O.; Kaever, V.; Stülke, J. An Essential Poison: Synthesis and Degradation of Cyclic Di-AMP in Bacillus subtilis. *J. Bacteriol.* **2015**, *197* (20), 3265–3274.

[127] Kaplan Zeevi, M.; Friedman, S.; Shaham, S.; Herskovits, A. A.; Shafir, N. S.; Sigal, N.; Nir Paz, R.; Boneca, I. G. Listeria monocytogenes Multidrug Resistance Transporters and Cyclic Di-AMP, Which Contribute to Type I Interferon Induction, Play a Role in Cell Wall Stress. *J. Bacteriol.* **2013**, *195* (23), 5250–5261.

[128] Jin, L.; Hill, K. K.; Filak, H.; Mogan, J.; Knowles, H.; Zhang, B.; Perraud, A.-L.; Cambier, J. C.; Lenz, L. L. MPYS Is Required for IFN Response Factor 3 Activation and Type I IFN Production in the Response of Cultured Phagocytes to Bacterial Second Messengers Cyclic-di-AMP and Cyclic-di-GMP. *J. Immunol.* **2011**, *187* (5), 2595–2601.

[129] Sauer, J. D.; Sotelo-Troha, K.; Von Moltke, J.; Monroe, K. M.; Rae, C. S.; Brubaker, S. W.; Hyodo, M.; Hayakawa, Y.; Woodward, J. J.; Portnoy, D. A.; Vance, R. E. The N-ethyl-N-nitrosourea-induced Goldenticket mouse mutant reveals an essential function of sting in the in vivo interferon response to Listeria monocytogenes and cyclic dinucleotides.

Infect. Immun. **2011**, *79* (2), 688–694.

[130] Blanco, P.; Hernando-Amado, S.; Reales-Calderon, J.; Corona, F.; Lira, F.; Alcalde-Rico, M.; Bernardini, A.; Sanchez, M.; Martinez, J. Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance Determinants. *Microorganisms* **2016**, *4* (1), 14.

[131] Schwartz, K. T.; Carleton, J. D.; Quillin, S. J.; Rollins, S. D.; Portnoy, D. A.; Leber, J. H. Hyperinduction of Host Beta Interferon by a Listeria monocytogenes Strain Naturally Overexpressing the Multidrug Efflux Pump MdrT. *Infect. Immun.* **2012**, *80* (4), 1537–1545.

[132] Tadmor, K.; Pozniak, Y.; Burg Golani, T.; Lobel, L.; Brenner, M.; Sigal, N.; Herskovits, A. A. Listeria monocytogenes MDR transporters are involved in LTA synthesis and triggering of innate immunity during infection. *Front. Cell. Infect. Microbiol.* **2014**, *4* (February), 1–14.

[133] Sutcliffe, I. C.; Shaw, N. Atypical lipoteichoic acids of gram-positive bacteria. *J. Bacteriol.* **1991**, *173* (22), 7065–7069.

[134] Barker, J. R.; Koestler, B. J.; Carpenter, V. K.; Burdette, D. L.; Waters, C. M.; Vance, R. E.; Valdivia, R. H. STING-dependent recognition of cyclic di-AMP mediates type I interferon responses during Chlamydia trachomatis infection. *MBio* **2013**, *4* (3), 1–11.

[135] Burdette, D. L.; Monroe, K. M.; Sotelo-Troha, K.; Iwig, J. S.; Eckert, B.; Hyodo, M.; Hayakawa, Y.; Vance, R. E. STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **2011**, *478* (7370), 515–518.

[136] McFarland, A. P.; Luo, S.; Ahmed-Qadri, F.; Zuck, M.; Thayer, E. F.; Goo, Y. A.; Hybiske, K.; Tong, L.; Woodward, J. J. Sensing of Bacterial Cyclic Dinucleotides by the Oxidoreductase RECON Promotes NF-κB Activation and Shapes a Proinflammatory Antibacterial State. *Immunity* **2017**, *46* (3), 433–445.

[137] Xia, P.; Wang, S.; Xiong, Z.; Zhu, X.; Ye, B.; Du, Y.; Meng, S.; Qu, Y.; Liu, J.; Gao, G.; Tian, Y.; Fan, Z. The ER membrane adaptor ERAdP senses the bacterial second messenger c-di-AMP and initiates anti-bacterial immunity article. *Nat. Immunol.* **2018**, *19* (2), 141–150.

[138] Mouali, Y. El; Kim, H.; Ahmad, I.; Brauner, A.; Liu, Y.; Skurnik, M.; Galperin, M. Y. Stand-Alone EAL Domain Proteins Form a Distinct Subclass of EAL Proteins Involved in Regulation of Cell Motility and Biofilm Formation in Enterobacteria. *J. Bacteriol.* **2017**, *199* (18), 1–17.

[139] Römling, U. Rationalizing the evolution of EAL domain-based cyclic di-GMP-specific phosphodiesterases. *J. Bacteriol.* **2009**, *191* (15), 4697–4700.

[140] Schmidt, A. J.; Ryjenkov, D. A.; Gomelsky, M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J. Bacteriol.* **2005**, *187* (14), 4774–4781.

[141] Tchigvintsev, A.; Xu, X.; Singer, A.; Chang, C.; Brown, G.; Proudfoot, M.; Cui, H.; Flick, R.; Anderson, W. F.; Joachimiak, A.; Galperin, M. Y.; Savchenko, A.; Yakunin, A. F. Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. *J. Mol. Biol.* **2010**, *402* (3), 524–538.

[142] Barends, T. R. M.; Hartmann, E.; Griese, J. J.; Beitlich, T.; Kirienko, N. V.; Ryjenkov, D. A.; Reinstein, J.; Shoeman, R. L.; Gomelsky, M.; Schlichting, I. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* **2009**, *459* (7249), 1015–1018.

[143] Seshasayee, A. S. N.; Fraser, G. M.; Luscombe, N. M. Comparative genomics of cyclic-di-

GMP signalling in bacteria : post-translational regulation and catalytic activity. *Nucleic Acids Res.* **2010**, *38* (18), 5970–5981.

[144] Phippen, C. W.; Mikolajek, H.; Schlaefli, H. G.; Keevil, C. W.; Webb, J. S.; Tews, I. Formation and dimerization of the phosphodiesterase active site of the Pseudomonas aeruginosa MorA , a bi-functional c-di-GMP regulator. *FEBS Lett.* **2014**, *588* (24), 4631–4636.

[145] Ferreira, R. B. R.; Antunes, L. C. M.; Greenberg, E. P.; Mccarter, L. L. Vibrio parahaemolyticus ScrC Modulates Cyclic Dimeric GMP Regulation of Gene Expression Relevant to Growth on Surfaces *J. Bacteriol.* **2008**, *190* (3), 851–860.

[146] An, S.; Wu, J.; Zhang, L. Modulation of Pseudomonas aeruginosa Biofilm Dispersal by a Cyclic-Di-GMP Phosphodiesterase with a Putative Hypoxia-Sensing Domain *Appl. Environ. Microbiol.* **2010**, *76* (24), 8160–8173.

[147] Liu, C.; Liew, C. W.; Wong, Y. H.; Tan, S. T.; Poh, W. H.; Manimekalai, M. S. S.; Rajan, S.; Xin, L.; Liang, Z.; Rice, S. A.; Lescar, J. Insights into Biofilm Dispersal Regulation from the Crystal Structure of the PAS-GGDEF-EAL Region of RbdA from Pseudomonas aeruginosa. *J. Bacteriol.* **2018**, *200* (3), 1–19.

[148] Rinaldo, S.; Paiardini, A.; Stelitano, V.; Brunotti, P.; Cervoni, L.; Fernicola, S.; Protano, C.; Vitali, M.; Cutruzzolà, F. Structural Basis of Functional Diversification of the HD-GYP Domain Revealed by the Pseudomonas aeruginosa PA4781 Protein , Which Displays an Unselective Bimetallic Binding Site. *J. Bacteriol.* **2015**, *197* (8), 1525–1535.

[149] Ryan, R. P.; Fouhy, Y.; Lucey, J. F.; Crossman, L. C.; Spiro, S.; He, Y.-W.; Zhang, L.-H.; Heeb, S.; Cámara, M.; Williams, P.; Dow, J. M. Cell–cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci.* **2006**, *103* (17), 6712 LP – 6717.

[150] Stelitano, V.; Giardina, G.; Paiardini, A.; Castiglione, N.; Rinaldo, S. C-di-GMP Hydrolysis by Pseudomonas aeruginosa HD- GYP Phosphodiesterases : Analysis of the Reaction Mechanism and Novel Roles for pGpG. *PLoS One* **2013**, *8* (9), 1–13.

[151] Ryan, R. P.; Lucey, J.; Donovan, K. O.; Mccarthy, Y.; Yang, L.; Tolker-nielsen, T.; Dow, J. M. HD-GYP domain proteins regulate biofilm formation and virulence in Pseudomonas aeruginosa. *Environ. Microbiol.* **2009**, *11* (5), 1126–1136.

[152] Lovering, A. L.; Capeness, M. J.; Lambert, C.; Hobley, L.; Sockett, R. E. The Structure of an Unconventional HD-GYP Protein from Bdellovibrio Reveals the Roles of Conserved Residues in this Class of Cyclic-di-GMP Phosphodiesterases. *MBio* **2011**, *2* (5), e00163-11.

[153] Bellini, D.; Caly, D. L.; Mccarthy, Y.; Bumann, M.; An, S.; Dow, J. M.; Ryan, R. P.; Walsh, M. A. Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Mol. Microbiol.* **2014**, *91* (1), 26–38.

[154] Miner, K. D.; Kurtz, D. M. Active Site Metal Occupancy and Cyclic Di-GMP Phosphodiesterase Activity of Thermotoga maritima HD-GYP. *Biochemistry* **2016**, *55* (6), 970–979.

[155] Kawai, T.; Akira, S. Review Toll-like Receptors and Their Crosstalk with Other Innate Receptors in Infection and Immunity. *Immunity* **2011**, *34* (5), 637–650.

[156] Kato, K.; Omura, H.; Ishitani, R.; Nureki, O. Cyclic GMP–AMP as an Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Annu. Rev. Biochem.* **2017**, *86* (1), 541–566.

[157] Chen, Q.; Sun, L.; Chen, Z. J. Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. *Nat. Immunol.* **2016**, *17* (10), 1142–1149.

[158] Kuchta, K.; Knizewski, L.; Wyrwicz, L. S.; Rychlewski, L.; Ginalski, K. Comprehensive classification of nucleotidyltransferase fold proteins : identification of novel families and their representatives in human. *Nucleic Acids Res.* **2009**, *37* (22), 7701–7714.

[159] Kranzusch, P. J.; Lee, A. S.; Berger, J. M.; Doudna, J. A. Report Structure of Human cGAS Reveals a Conserved Family of Second-Messenger Enzymes in Innate Immunity. *Cell Rep.* **2013**, *3* (5), 1362–1368.

[160] Lohöfener, J.; Steinke, N.; Manstein, D. J.; Fedorov, R.; Steinke, N.; Kay-fedorov, P.; Baruch, P.; Nikulin, A.; Tishchenko, S. The Activation Mechanism of 2'-5'-Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. *Structure* **2015**, *23* (5), 851–862.

[161] Clemens, M. J.; Williams, B. R. G.; Hill, M. Inhibition of Cell-Free Protein Synthesis by pppA2'p5'A2'p5'A: a Novel Oligonucleotide Synthesized by Interferon-Treated L Cell Extracts. *Cell* **1978**, *13* (3), 565–572.

[162] Chebath, J.; Benech, P.; Revel, M.; Vigneron, M. Constitutive expression of (2 ' -5 ') oligo A synthetase confers resistance to picornavirus infection s s Corrigendum Thermal X-ray emission from supernova. *Nature* **1987**, *330* (10), 587–588.

[163] Dong, B.; Silverman, R. H. A Bipartite Model of 2-5A-dependent RNase L *. *J. Biol. Chem.* **1997**, *272* (35), 22236–22242.

[164] Andreeva, L.; Hiller, B.; Kostrewa, D.; Lässig, C.; De Oliveira Mann, C. C.; Jan Drexler, D.; Maiser, A.; Gaidt, M.; Leonhardt, H.; Hornung, V.; Hopfner, K. P. cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. *Nature* **2017**, *549* (7672), 394–398.

[165] Zhang, X.; Wu, J.; Du, F.; Xu, H.; Sun, L.; Chen, Z.; Brautigam, C. A.; Zhang, X.; Chen, Z. J. The Cytosolic DNA Sensor cGAS Forms an Oligomeric Complex with DNA and Undergoes Switch-like Conformational Changes in the Activation Loop. *Cell Rep.* **2014**, *6* (3), 421–430.

[166] Li, X.; Shu, C.; Yi, G.; Chaton, C. T.; Shelton, C. L.; Diao, J.; Zuo, X.; Kao, C. C.; Herr, A. B.; Li, P. Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization. *Immunity* **2013**, *39* (6), 1019–1031.

[167] Civril, F.; Deimling, T.; Mann, C. C. D. O.; Ablasser, A.; Moldt, M.; Witte, G.; Hornung, V. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **2013**, *498* (7454), 332–337.

[168] Wu, X.; Wu, F. H.; Wang, X.; Wang, L.; Siedow, J. N.; Zhang, W.; Pei, Z. M. Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.* **2014**, *42* (13), 8243–8257.

[169] Tao, J.; Zhang, X.-W.; Jin, J.; Du, X.-X.; Lian, T.; Yang, J.; Zhou, X.; Jiang, Z.; Su, X.-D. Nonspecific DNA Binding of cGAS N Terminus Promotes cGAS Activation. *J. Immunol.* **2017**, *198* (9), 3627–3636.

[170] Gentili, M.; Lahaye, X.; Nadalin, F.; Nader, G. F. P.; Puig Lombardi, E.; Herve, S.; De Silva, N. S.; Rookhuizen, D. C.; Zueva, E.; Goudot, C.; Maurin, M.; Bochnakian, A.; Amigorena, S.; Piel, M.; Fachinetti, D.; Londoño-Vallejo, A.; Manel, N. The N-Terminal Domain of cGAS Determines Preferential Association with Centromeric DNA and Innate Immune Activation in the Nucleus. *Cell Rep.* **2019**, *26* (9), 2377-2393.e13.

[171] Barnett, K. C.; Coronas-Serna, J. M.; Zhou, W.; Ernandes, M. J.; Cao, A.; Kranzusch, P.

J.; Kagan, J. C. Phosphoinositide Interactions Position cGAS at the Plasma Membrane to Ensure Efficient Distinction between Self- and Viral DNA. *Cell* **2019**, *176* (6), 1432-1446.e11.

[172] Cai, X.; Chiu, Y.; Chen, Z. J. Review The cGAS-cGAMP-STING Pathway of Cytosolic DNA Sensing and Signaling. *Mol. Cell* **2014**, *54* (2), 289–296.

[173] Zhang, C.; Shang, G.; Gui, X.; Zhang, X.; Bai, X.; Chen, Z. J. Structural basis of STING binding with and phosphorylation by TBK1. *Nature* **2019**, *567* (7748), 394–398.

[174] Shang, G.; Zhang, C.; Chen, Z. J.; Bai, X.; Zhang, X. Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP–AMP. *Nature* **2019**, *567* (7748), 389–393.

[175] Zhao, B.; Shu, C.; Gao, X.; Sankaran, B.; Du, F.; Shelton, C. L.; Herr, A. B.; Ji, J.-Y.; Li, P. Structural basis for concerted recruitment and activation of IRF-3 by innate immune adaptor proteins. *Proc. Natl. Acad. Sci.* **2016**, *113* (24), E3403 LP-E3412.

[176] Danilchanka, O.; Mekalanos, J. J. Review Cyclic Dinucleotides and the Innate Immune Response. *Cell* **2013**, *154* (5), 962–970.

[177] Gao, P.; Ascano, M.; Zillinger, T.; Wang, W.; Dai, P.; Serganov, A. A.; Gaffney, B. L.; Shuman, S.; Jones, R. A.; Deng, L.; Hartmann, G.; Barchet, W.; Tuschl, T.; Patel, D. J. Structure-function analysis of STING activation by c[G(2′,5′) pA(3′,5′)p] and targeting by antiviral DMXAA. *Cell* **2013**, *154* (4), 748–762.

[178] Zhang, X.; Shi, H.; Wu, J.; Zhang, X.; Sun, L.; Chen, C.; Chen, Z. J.; Adams, P. D.; Afonine, P. V.; Bunkóczi, G. et al. Cyclic GMP-AMP Containing Mixed Phosphodiester Linkages Is An Endogenous High-Affinity Ligand for STING. *Mol. Cell* **2013**, *51* (2), 226–235.

[179] Yin, Q.; Tian, Y.; Kabaleeswaran, V.; Jiang, X.; Tu, D.; Eck, M. J.; Chen, Z. J. Cyclic di-GMP Sensing via the Innate Immune Signaling Protein STING. *Mol. Cell* **2012**, *46* (6), 735–745.

[180] Shu, C.; Yi, G.; Watts, T.; Kao, C. C.; Li, P. Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat. Struct. Mol. Biol.* **2012**, *19* (7), 722–724.

[181] Ouyang, S.; Song, X.; Wang, Y.; Ru, H.; Shaw, N.; Jiang, Y.; Niu, F.; Zhu, Y.; Qiu, W.; Parvatiyar, K.; Li, Y.; Zhang, R.; Cheng, G.; Liu, Z. Structural Analysis of the STING Adaptor Protein Reveals a Hydrophobic Dimer Interface and Mode of Cyclic di-GMP Binding. *Immunity* **2012**, *36* (6), 1073–1086.

[182] Huang, Y. H.; Liu, X. Y.; Du, X. X.; Jiang, Z. F.; Su, X. D. The structural basis for the sensing and binding of cyclic di-GMP by STING. *Nat. Struct. Mol. Biol.* **2012**, *19* (7), 728–730.

[183] Chin, K.; Tu, Z.; Chen, H.; Lay-cheng, M. Novel c-di-GMP recognition modes of the mouse innate immune adaptor protein STING research papers. *Acta Crystallogr. Sect. D Struct. Biol.* **2013**, *69* (3), 352–366.

[184] Cavlar, T.; Deimling, T.; Ablasser, A.; Hopfner, K.; Hornung, V. Species-specific detection of the antiviral small-molecule compound CMA by STING. *EMBO J.* **2013**, *32* (10), 1440–1450.

[185] Jin, L.; Xu, L.; Yang, I. V; Davidson, E. J.; Schwartz, D. A.; Wurfel, M. M.; Cambier, J. C. Identification and characterization of a loss-of-function human MPYS variant. **2011**, *12* (4), 263–269.

[186] Li, A.; Yi, M.; Qin, S.; Song, Y.; Chu, Q.; Wu, K. Activating cGAS-STING pathway for the optimal effect of cancer immunotherapy. *J. Hematol. Oncol.* **2019**, *12* (1), 1–12.

[187] Bose, D. CGAS/STING pathway in cancer: Jekyll and hyde story of cancer immune response. *Int. J. Mol. Sci.* **2017**, *18* (11).

[188] Prantner, D.; Perkins, D. J.; Lai, W.; Williams, M. S.; Sharma, S.; Fitzgerald, K. A.; Vogel, S. N. 5,6-Dimethylxanthenone-4-acetic acid (DMXAA) activates stimulator of interferon gene (STING)-dependent innate immune pathways and is regulated by mitochondrial membrane potential. *J. Biol. Chem.* **2012**, *287* (47), 39776–39788.

[189] Pang, J. H.; Cao, Z.; Joseph, W. R.; Baguley, B. C.; Ching, L. M. Antitumour activity of the novel immune modulator 5,6- dimethylxanthenone-4-acetic acid (DMXAA) in mice lacking the interferon- gamma receptor. *Eur. J. Cancer* **1998**, *34* (8), 1282–1289.

[190] Corrales, L.; Glickman, L. H.; McWhirter, S. M.; Kanne, D. B.; Sivick, K. E.; Katibah, G. E.; Woo, S. R.; Lemmens, E.; Banda, T.; Leong, J. J.; Metchette, K.; Dubensky, T. W.; Gajewski, T. F. Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *Cell Rep.* **2015**, *11* (7), 1018–1030.

[191] Lara, P. N.; Douillard, J. Y.; Nakagawa, K.; Von Pawel, J.; McKeage, M. J.; Albert, I.; Losonczy, G.; Reck, M.; Heo, D. S.; Fan, X.; Fandi, A.; Scagliotti, G. Randomized phase III placebo-controlled trial of carboplatin and paclitaxel with or without the vascular disrupting agent vadimezan (ASA404) in advanced non-small-cell lung cancer. *J. Clin. Oncol.* **2011**, *29* (22), 2965–2971.

[192] Conlon, J.; Burdette, D. L.; Sharma, S.; Bhat, N.; Thompson, M.; Jiang, Z.; Rathinam, V. A. K.; Monks, B.; Jin, T.; Xiao, T. S.; Vogel, S. N.; Vance, R. E.; Fitzgerald, K. A. Mouse, but not Human STING, Binds and Signals in Response to the Vascular Disrupting Agent 5,6-Dimethylxanthenone-4-Acetic Acid. *J. Immunol.* **2013**, *190* (10), 5216–5225.

[193] Kramer, M. J.; Cleeland, R.; Grunberg, E. Antiviral activity of 10 carboxymethyl 9 acridanone. *Antimicrob. Agents Chemother.* **1976**, *9* (2), 233–238.

[194] Schmier, B. J.; Nelersa, C. M.; Malhotra, A. Structural Basis for the Bidirectional Activity of Bacillus nanoRNase NrnA. *Sci. Rep.* **2017**, *7* (1), 1–13.

[195] Yamagata, A.; Kakuta, Y.; Masui, R.; Fukuyama, K. The crystal structure of exonuclease RecJ bound to Mn2+ ion suggests how its characteristic motifs are involved in exonuclease activity. *Proc. Natl. Acad. Sci.* **2002**, *99* (9), 5908–5912.

[196] Wakamatsu, T.; Kitamura, Y.; Kotera, Y.; Nakagawa, N.; Kuramitsu, S.; Masui, R. Structure of RecJ exonuclease defines its specificity for single-stranded DNA. *J. Biol. Chem.* **2010**, *285* (13), 9762–9769.

[197] Orr, M. W.; Weiss, C. A.; Severin, G. B.; Turdiev, H.; Kim, K.; Turdiev, A.; Liu, K.; Tu, B. P.; Christopher, M.; Winkler, W. C.; Lee, V. T. A subset of exoribonucleases serve as degradative enzymes for pGpG in c-di-GMP signaling. *J. Bacteriol.* **2018**, *200* (24), e00300-18.

[198] Brown, L.; Wolf, J. M.; Prados-Rosales, R.; Casadevall, A. Through the wall: Extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat. Rev. Microbiol.* **2015**, *13* (10), 620–630.

[199] Meriläinen, L.; Herranen, A.; Schwarzbach, A.; Gilbert, L. Morphological and biochemical features of Borrelia burgdorferi pleomorphic forms. *Microbiology* **2015**, *161* (3), 516–527.

[200] Kornspan, J. D.; Rottem, S. The Phospholipid Profile of Mycoplasmas. *J. Lipids* **2012**, *2012*, 1–8.

[201] Hildebrand, A.; Remmert, M.; Biegert, A.; Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (SUPPL. 9), 128–132.

[202] Agarwal, V.; Borisova, S. A.; Metcalf, W. W.; Van Der Donk, W. A.; Nair, S. K. Structural and mechanistic insights into C-P bond hydrolysis by phosphonoacetate hydrolase. *Chem. Biol.* **2011**, *18* (10), 1230–1240.

[203] Li, L.; Yin, Q.; Kuss, P.; Maliga, Z.; Millán, J. L.; Wu, H.; Mitchison, T. J. Hydrolysis of 2'3'-cGAMP by ENPP1 and design of nonhydrolyzable analogs. *Nat. Chem. Biol.* **2014**, *10* (12), 1043–1048.

[204] Sali, T. M.; Pryke, K. M.; Abraham, J.; Liu, A.; Archer, I.; Broeckel, R.; Staverosky, J. A.; Smith, J. L.; Al-Shammari, A.; Amsler, L.; Sheridan, K.; Nilsen, A.; Streblow, D. N.; DeFilippis, V. R. Characterization of a Novel Human-Specific STING Agonist that Elicits Antiviral Activity Against Emerging Alphaviruses. *PLoS Pathog.* **2015**, *11* (12), 1–30.

[205] Gall, B.; Pryke, K.; Abraham, J.; Mizuno, N.; Botto, S.; Sali, T. M.; Broeckel, R.; Haese, N.; Nilsen, A.; Placzek, A.; Morrison, T.; Heise, M.; Streblow, D.; DeFilippis, V. Emerging Alphaviruses Are Sensitive to Cellular States Induced by a Novel Small-Molecule Agonist of the STING Pathway. *J. Virol.* **2017**, *92* (6), 1–19.

[206] Kranzusch, P. J.; Wilson, S. C.; Lee, A. S. Y.; Berger, J. M.; Doudna, J. A.; Vance, R. E. Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol. Cell* **2015**, *59* (6), 891–903.

[207] Zhang, H.; Han, M. J.; Tao, J.; Ye, Z. Y.; Du, X. X.; Deng, M. J.; Zhang, X. Y.; Li, L. F.; Jiang, Z. F.; Su, X. D. Rat and human STINGs profile similarly towards anticancer/antiviral compounds. *Sci. Rep.* **2015**, *5* (18035), 1–12.

Acknowledgements

First of all, I would like to thank my supervisors PD Dr. Gregor Witte and Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work at the gene center and all their scientific input and great support. I am especially happy that I could work on projects from both research fields, c-di-AMP and innate immunity. Prof. Dr. Karl-Peter Hopfner included me in the innate immunity projects and supported me with great ideas, project meetings and collaborations - I appreciate this very much! Special thanks go to PD Dr. Gregor Witte for taking a lot of time for the close supervision with exciting projects, fruitful and enjoyable discussions, synchroton trips, conferences and many other situations.

I am very grateful to our collaboration partners Dr. Natalia Tschowri and Andreas Latoscha for the very positive and successful cooperation. Thank you for including me to the AtaC project and all the constructive discussions.

I would also like to thank Prof. Dr. Thomas Carell and his group, in particular Clemens Dialer, Simon Veth and Samuele Stazzoni for the great collaboration on cGAMP analogues and all the support with RP-HPLC purification.

I thank my dear colleagues of the c-di-AMP projects, including Dr. Martina Omland, Adrian Bandera, Dr. Sandra Schuller and Olga Fettscher for all the support, constructive discussions, cooperation and a nice time at work. I would also like to thank Dr. Liudmila Andreeva, Dr. Charlotte Lässig, Dr. Carina Mann and Sebastian Michalski for introducing me to the cGAS projects and to many methods – it was a great time working with you! I also thank Brigitte Kessler for taking every question about chemicals, orders, centrifuges and many other things and keeping the lab organized. Furthermore, I thank the whole AG Hopfner for the positive and constructive working atmosphere. All colleagues have contributed to feel comfortable in the group from the beginning, thank you!

I also thank the GRK1721 for the organization of workshops, retreats, talks and seminars and supporting conferences, altogether enabling a great scientific exchange.

Finally, I would like to thank my family and Moni for supporting me in every possible situation!