Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural studies on autoimmunity regulation by the nucleic acid sensors cGAS and RIG-I

Sebastian Michalski

aus

Bochum, Deutschland

2021

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011

 von Herrn Prof. Dr. Karl-Peter Hopfner betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 17.05.2021

_____

Sebastian Michalski

Dissertation eingereicht am: 17.05.2021

1. Gutachter: Herr Prof. Dr. Karl-Peter Hopfner

2. Gutachter: Herr. Prof. Dr. Veit Hornung

Mündliche Prüfung am: 06.07.2021

This thesis has been prepared from October 2017 to May 2021 in the laboratory of Professor Dr. Karl-Peter Hopfner at the Gene Center of the Ludwig-Maximilians-Universität Munich.

This is a cumulative thesis based on following publications:

Michalski, S.*, de Oliveira Mann*, C.C., Stafford, C.*, Witte, G., Bartho, J., Lammens, K., Hornung, V., Hopfner, K-P. *Nature* **587,** 678–682 (2020). https://doi.org/10.1038/s41586-020-2748-0
*: equal contribution

Lässig, C., Lammens K., Gorenflos López, J-L., Michalski, S., Fettscher, O., Hopfner, K-P. *eLife* (**2018**);7:e38958. https://doi.org/10.7554/eLife.38958

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we fear less.

Marie Curie

Table of contents

1 Summary

Differentiating between self and non-self is the major task of the innate immune system. Therefore, misregulation of the finely orchestrated underlying mechanisms is linked to pathogenic malfunctions of the innate immune system. The major part of this thesis investigates the mode of autoimmunity prevention of the pathogen recognition receptor cyclic GMP-AMP synthase (cGAS). cGAS senses double-stranded DNA (dsDNA) in a sequence-unspecific manner, evoking an innate immune response via its downstream adaptor stimulator of interferon genes (STING). The role of the cGAS-STING axis is well established in sensing pathogenic cytosolic DNA emerging from bacterial or viral sources. Growing evidence also links cGAS to sensing of mislocalized or structurally altered self-DNA. For long time cGAS was thought to be a strictly cytosolic protein, relying on this compartmentalization for differentiation between self and non-self-DNA. This view is nowadays challenged by multiple observations of cGAS colocalizing with nuclear or nuclear-like self-DNA species. Especially, the sequence-unspecific dsDNA sensing of cGAS was intriguing in the case of nuclear envelope breakdown during mitosis, as healthy cells do not elicit an immune response during cell division. This work aims to uncover the mechanism and structural basis for autoimmunity prevention of the cGAS-chromatin interaction. Canonical nucleosomes were found to be a potent inhibitor of cGAS activity, showing magnitudes higher cGAS affinity than the activating ligand double-stranded DNA. The first 3.3 Å cryo-EM structure of the cGAS-Nucleosome complex revealed a conserved unexpected protein-protein interaction of cGAS with the nucleosome core. cGAS was tightly anchored via an invariant `arginine anchor´ to the acidic patch formed at the interface between histone H2A and H2B. This sequestration covered a DNA binding site needed for active cGAS complex formation, thereby keeping cGAS in check at self-DNA encounter. Extensive mutagenic analyses confirmed the specificity of this inhibition mechanism *in vitro* and *in cellulo*. As histone-proteins are among the most conserved proteins found in eukaryotes this detection pattern serves as *bona fide* marker of self. This tight interaction of cGAS with nucleosomes and differences in cGAS activity depending on the compaction level of nucleosome arrays opens a whole new field of nuclear cGAS studies.

Summary

Another part of this thesis aims to understand the constant activity of the retinoic acid-inducible gene I (RIG-I) Singleton-Merten-Sydrome (SMS) variant C268F. RIG-I can be seen as the RNA-sensing counterpart to the cGAS-STING axis, leading to an immune response after sensing of specific pathogenic RNA species. Defects in RIG-I regulation are linked to rare autoimmune diseases like Singleton-Merten- or Aicardi-Goutières-Syndrome (AGS). Similar to how cGAS is tethered to chromatin for inhibition also RIG-I developed mechanisms to prevent activation against self-ligands. As RIG-I structurally belongs to the superfamily 2 (SF2)-helicase family, it harbors an ATPase domain that is not needed for helicase activity, but instead is required for dissociating from self-ligands by ATP-hydrolysis driven translocation. How the RIG-I SMS variant C268F bypasses this regulation leading to constant activation was not known. RIG-I C268F was shown to have higher affinity towards self-like RNA species independent of ATP. Furthermore, the affinity towards ATP was lower compared to the wild-type (WT) protein and the catalytic activity was reduced. These biochemical data suggested a signaling-mode independent of ATP. Indeed, the solved crystal structure of RIG-I C268F in complex with RNA revealed a reorganization of the active site introduced by the bulky phenylalanine residue that mimics an ATP bound state without nucleotide bound. Due to this pseudo-ATP bound state RIG-I C268F lacks the needed regulation and shows constant activity leading to autoimmunity.

Combined, both works in this thesis underline the sophisticated mechanisms that have evolved in innate immunity sensing to efficiently prevent autoimmunity and how small changes can disrupt the tightly regulated differentiation apparatus.

2 Introduction

2.1 Eukaryotic nucleic acid organization

2.1.1 Chromatin – structuring the basis for all self-molecules

Nowadays the central dogma of molecular biology: sequential irreversible information flow by DNA transcribed into RNA that is then translated to proteins is unquestionable (often referred to as `DNA makes RNA makes protein´)[1]. DNA serves as the fundamental information storage found in all cells. In eukaryotic cells DNA is only stored in two places: (1) as dynamic compaction structure called chromatin in the nucleus and as (2) mitochondrial DNA, that can be seen as evolutionary remnant resulting from the switch towards eukaryotic systems[2] (plus chloroplast DNA in plants). The first description of a phosphorus substance within the cell, termed nuclein at this point but later known as DNA, was done by the fundamental work of Miescher in 1871[3]. During his research on cell division, thereby also creating the concept of mitosis, Flemming recognized a readily stainable substance within the nucleus, he therefore named chromatin[4]. The discovery of the protein part of chromatin in 1884, the so-called histones, completed the main ingredients of the eukaryotic genomic organization[5]. Nevertheless, there was no indication of the function of this structure at this point. With the years it became clear, that indeed nuclein or rather DNA, is the carrier of genetic information[6]. The structure of the DNA double-helix obtained by X-ray diffraction might be one of the most well-known scientific achievements in the public perception[7] and many hints towards the functionality of chromatin were made at this time (protein-free regions on chromatin, histone isolation, histone modifications linked to gene expression)[8–10]. The structural role of the protein part became clear, with the first electron microscopic images of chromatin. The DNA is wrapped around the histones in a repetitive manner forming a 10 nm diameter `beads on a string´ like structure[11] (see Fig. 1). Finally, Kornberg published the first structure of the basic repetitive unit of chromatin: around 200 base pairs (bp) of DNA are wrapped around a central octameric unit comprised of four histone dimers building the central repetitive element – the nucleosome[12,13].
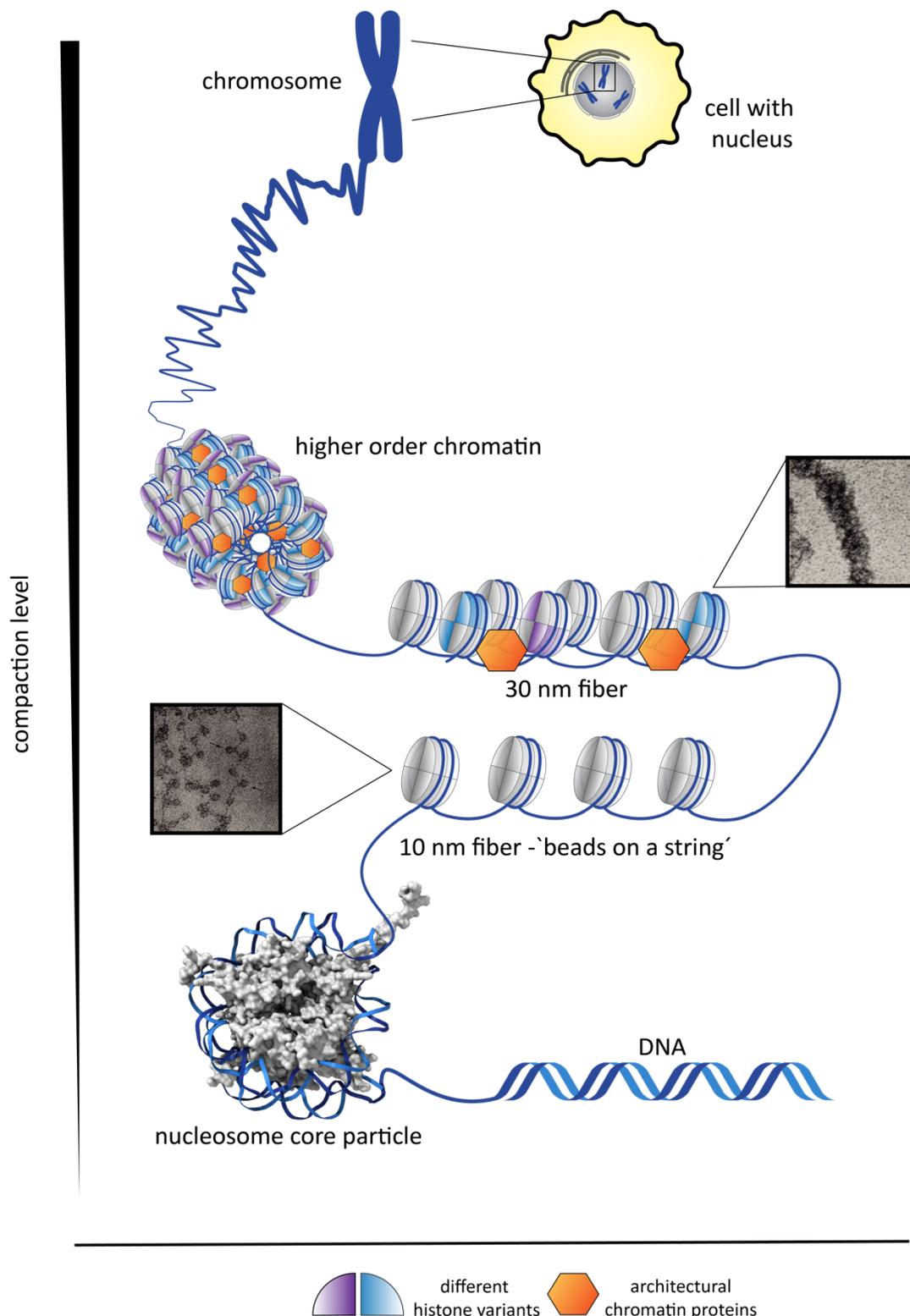
Figure 1: Eukaryotic genome organization. As the first layer of compaction 147 bp of double-stranded DNA is wrapped around the histone octamer, creating the nucleosome core particle. The canonical histone octamer consists of two sets of H2A, H2B, H3 and H4. Next, multiple nucleosomes are linked in a linear fashion creating the 10 nm fiber. The micrograph shows the individual nucleosomes spaced by linker-DNA. In higher order chromatin, the canonical histone subset can be exchanged with histone variants, altering the function and topology of chromatin. *In vitro* the higher order 30 nm fiber is observed (see micrograph on the right). The occurrence of this species *in vivo* is still under debate. Further compaction with the help of architectural chromatin proteins, may end up in the most condensed DNA form – mitotic chromosomes. |micrographs from Olins *et al.* (2003); nucleosome core particle structure from PDB:1AOI

2.1.2 The nucleosome core particle – structural insights

The wrapping of DNA around histone proteins fulfills multiple functions in eukaryotic genome organization. Because the genomic DNA is compacted 7-fold by wrapping around the histones, changes in local compaction levels directly influence gene transcription, DNA replication and repair by altering accessibility of the DNA. Chromatin shows multiple levels of compaction and organization with the nucleosome core particle (NCP) as the primary step (see Fig. 1).

Insight into the exact architecture of the NCP was achieved by a crystal structure of the protein core without DNA, that was later on completed by a first high resolution structure together with DNA by Luger *et al.*[14,15]. Already from the structure without DNA it became clear that the canonical histone octamer core is formed by two copies of each of the so-called core histones (H2A, H2B, H3 and H4). 145 to 147 bp of DNA are wrapped around in 1.65 lefthanded super helical turns, leading to a disc-like structure with a diameter of 11 nm and a height of 5.5 nm[15] (see Fig. 2).

As the compaction of genomic DNA is the basic principle of eukaryotic genome organization the histone core proteins are among the most conserved proteins in eukaryotes[16]. All of them are highly basic proteins with a mostly α-helical C-terminal domain for histone-histone interactions and an N-terminal tail, that is unstructured, when no DNA or other co-factor is bound to the histone octamer[14]. The C-terminal domain harbors the histone-fold – a motif where a central long α-helix (α2) is sandwiched by two shorter α-helices (α1 and α3) connected by short β-sheet containing loops (L1 and L2). Heterotypic interactions between H2A/H2B and H3/H4 are formed by their respective histone-fold motifs known as the `handshake' interaction[17] (see Fig. 2B). A stable (H3/H4)$_2$ tetramer is formed by hydrophobic packing of a four-helix bundle between the two H3 molecules of two H3/H4 dimers. The octamer is completed by two H2A/H2B dimers interacting with a similar four-helix bundle between the H4s of the tetramer with H2B and additional interactions of the docking domain of the H2As with H3.
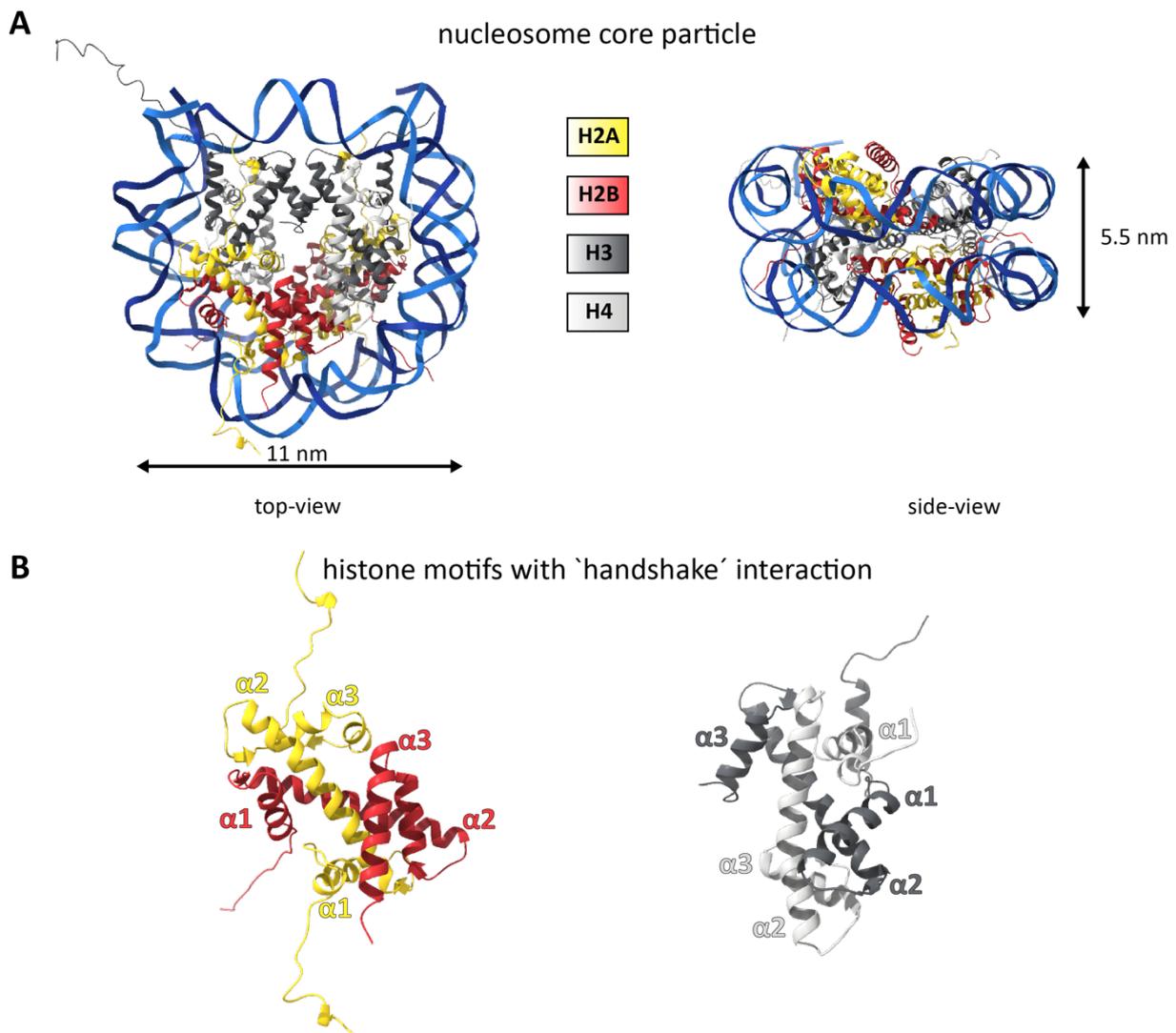
Figure 2: Structural overview of nucleosome core particle and `handshake´ interaction. A The canonical histone octamer subset colored according to the picture. 147 bp of dsDNA (blue) are wrapped around the central octamer. The created disc-like structure is 11 nm in diameter and 5.5 nm in height. **B** specific handshake motif of the individual histone folds, showing the interaction between H2A/H2B and H3/H4. |based on PDB:1AOI

The created disc shows distinct DNA binding sites on its outer vertical surface, created by paired elements (paired β-loops or paired α-helices) from two neighboring dimers, thereby interacting with the DNA backbone where the minor groove faces the octamer[15,17] (see Fig. 3). Many of these interactions are facilitated by highly conserved arginine residues located in the histone-fold domains sticking into the minor grooves and forming hydrogen-bonds with desoxyriboses[18]. The relative position of these interacting residues is fixed in different species, leading to specific DNA localization on the protein core[19]. The combined histone-fold motifs of all histones account for about 120 bp DNA binding surface (of total 147 bp). The residual contacts are made by the non-histone fold α-helix of H3 and the start of its N-terminal tail domain[15].

Figure 3: Surface charge of octamer core. Surface charge of the histone octamer core displayed from positive (blue) to negative (red). The outer surface of the disc is highly positively charged, explaining the binding of DNA around that area, whereas the top and bottom the disc do not show positive charges. |based on PDB:1AOI

This alignment of the DNA leads to a single base pair located at the particle´s pseudo two-fold axis, called the dyad. For non-palindromic sequences this splits the DNA in two halves of 73 and 72 bp. That point defines the relative rotation of the DNA superhelix and is referred to as superhelix location 0 (SHL0). The residual outward facing minor grooves of the superhelix are therefore denoted as SHL ± 1-7 depending on the direction relative to SHL0 (see Fig. 4). Normal B-form DNA has a persistence length of around 150 bp (~50 nm) but is bent into a 1.65 turns lefthanded superhelix with 80 bp/turn when loaded onto a nucleosome. This large distortion is mostly accommodated by base-pair rolling into the minor and major grooves where they face the histone octamer surface[20,21]. The resulting superhelix shows low rise (30 Å rise per 80 bp) with a diameter of around 42 Å and non-uniform bending (highest local curvature at SHL1.5 and SHL 4.5)[15]. Specific local histone-DNA interactions decrease the number of bp per helical turn from 10.5 to 10.2 bp/turn leading to a `twist defect´[20,22]. That was already noticed in the first crystal structure by Luger *et al.*, where the asymmetric splitting of the DNA leads to an overwinding of the DNA by one bp. Nowadays it is known, that this `defect´ is rather a feature needed for chromatin remodeling[23].

Figure 4: Nomenclature of superhelix locations. The pseudo two-fold axis of the nucleosome is called the `dyad´ or superhelix location 0 (SHL0). Depending on the location relative to this axis the outward facing minor grooves of the superhelix are denoted with SHL ± 1-7. From the side view, the visible strands of the superhelix are denoted as gyres. |based on PDB:1AOI

Not all DNA sequences can accommodate these constraints equally well, inducing rotational shifting of the DNA relative to the central axis. The optimal nucleosome positioning sequence was found by systematic evolution of ligands by exponential enrichment (SELEX) and is well-known as the `Widom-601 sequence'[24]. This sequence shows high TA periodicity in one gyre and low periodicity in the other one, separated by the dyad axis, leading to strong positioning[25].

Beyond the formation of the rigid spool-like core structure, all histones have more flexible but yet conserved tail regions. These initially unstructured parts originate from the N-terminal regions of all histones and an additional tail at the C-terminus of H2A. When DNA is bound to the octamer, all tails follow a channel formed by two stacked minor grooves of the DNA superhelix, thereby facing outwards beyond the DNA and making them accessible. Different exit channels are observed for the respective tail-domains, with slight differences found depending on the crystal structure taken as reference. Despite not contributing much to the overall stability of the NCP[26], tail-removal leads to increased DNA release of the nucleosome[27]. Furthermore, it appears that the tail-domains gain more structure in higher order chromatin under physiological conditions[28,29]. Not playing a huge role for the NCP itself, the tail domains are crucial for epigenetic modifications and formation of higher order chromatin structures by facilitating inter-nucleosomal contacts[30,31].

2.1.3 Higher order chromatin – dynamic multilayer architecture

The organization of DNA wrapped around nucleosomes creating the `beads on a string´ structure can be seen as the primary structure of chromatin. Thinking of mitotic chromosomes as the terminal superstructure, multiple organizational sublevels occur depending on histone modifications and interplay, architectural chromatin proteins (ACPs) and ATP-dependent chromatin remodelers[32]. The first addition to the repeating nucleosome array is binding of a linker histone (H1) to extra-nucleosomal DNA at the dyad axis, creating the chromatosome now covering an extra 20 bp of DNA[33]. Metazoan linker histones have a tripartite structure consisting of a flexible N-terminal tail, a central globular domain with winged-helix fold and a highly basic intrinsically unordered C-terminal tail[34]. The globular domain interacts with one or two DNA strands at the dyad, whereas the N-terminal domain is needed for further higher order compaction[31,35].The next architectural layer of chromatin packing is the short-range interaction between neighboring nucleosome arrays forming some kind of higher-order fiber. The oldest concept is the solenoid superhelix with 5-6 nucleosomes/turn creating a 30 nm thick fiber[36]. This view is challenged by the two-point start helix Zig-Zag model[37]. To date, the existence of the so-called `30-nm fiber´ *in vivo* is still controversially discussed in the field. Due to the compactness and electron density of condensed *in vivo* chromatin, there is still no high resolution electron microscopy structure available[38].

All structural data on the three-dimensional arrangement are based on *in vitro* reconstituted systems, leading to different outcomes. The crystal structure of a short-linker length tetranucleosome array without H1 promote the Zig-Zag model[39], whereas EM studies with incorporation of H1 propose a interdigitated solenoid model[38]. Electron microscopy-assisted nucleosome capture (EMANIC) shows a mixture of both types creating a heteromorphic fiber[40], which is also supported by mesoscopic modelling[41]. Most probably there is not a universal model, but different structures can be found depending on DNA sequence, histone variants, and cell type and its cell cycle stage[32,42]. Howsoever the intermediate fiber is formed, these fibers can further fold into larger three-dimensional structures, that end up in condensed chromosomes at metaphase.

This classical view of hierarchal chromosome compaction is increasingly challenged by a liquid-liquid phase separation (LLPS) model, that would favor long distance interactions[43]. LLPS recently has emerged as an important cellular compartmentalization mechanism for non-membranous structures. Especially proteins with intrinsically disordered regions (as found in the histone-tails) are prone to phase separation[44]. Chromatin LLPS was shown to be dependent on the architectural protein heterochromatin protein 1 (HP1)[45] and different properties of the primary structure like array length, histone modifications, and histone H1 (and therefore also salt concentration)[46].

2.1.4 Histone modifications and variants – regulators of the dynamic network

Generally, chromatin can be separated into two subtypes depending on transcriptional activity. Active and therefore less compacted euchromatin and dense but inactive heterochromatin. Modulation of these regions within the three-dimensional structure is achieved by post-translational modifications (PTMs) of the histone cores and exchanging the canonical histone octamer set to alter its interaction with the DNA. All kinds of modifications found on other proteins are also found on histones[47]. The longest known modification is acetylation at lysine residues, mostly in the tail regions (with few exceptions of core acetylation)[8]. Together with phosphorylation at serines, threonines and tyrosines of the tails, these modifications alter the charge of the respective residues, weakening the association with DNA[48]. In contrast, methylation at arginines and lysines does not influence the charge of the histones but adds another layer of regulation by the possibility of multiple modifications on a single amino acid residue. Adding to these most common modifications also ubiquitylation, sumolyation, ADP-ribosylation and β-N-acetylglucosamine can be found[49]. For all modifications specific addition and removal enzymes are available to achieve the dynamic environment found in chromatin. In general histone acetylation is correlated with increased gene expression, whereas the regulation by methylation is more complex. H3K4me and H3K36me are associated with activation whereas H3K9 and H3K27 methylation leads to repression[50]. Opening the chromatin structure by charge repulsion after phosphorylation enables increased accessibility for remodeling and exchange of histone variants[51].

Another perturbation from the canonical nucleosome structure is accomplished by replacement of the core histones with specific histone variants. The expression of the four core histones is strictly coupled to replication, thereby only occurring in S-Phase[52].

In contrast histone variants show a cell type and species specific dynamic expression pattern independent of the cell cycle, thereby acting as a regulatory module[53]. All human core histones have at least one modified counterpart, with H2A and H3 showing the most variations[54]. The best studied and universal histone variants are CenP-A (a H3 variant), H3.3, H2AX and H2AZ. CenP-A is exclusively found at nucleosomes at the centromere region during G1-phase[55]. H3.3 is a hallmark of euchromatin, found in active genomic regions[56]. H2AX shows an additional C-terminal phosphorylation site, implicated in dsDNA repair after activation (γ-H2AX)[57]. H2AZ leads to nucleosome destabilization and is characterized by an increased interaction surface formed between H2A and H2B known as the acidic patch[58].

2.1.5 The acidic patch – most prominent docking site of the nucleosome core

Most interactions with the nucleosome are facilitated via interactions or modifications of the protruding histone-tails, as only 40% of the core surface is solvent accessible. Nevertheless, a specific region formed by the interface between H2A and H2B, called the acidic patch, serves as docking platform for multiple proteins. Six amino acid residues of H2A (human H2A, E56, E61, E64, D90, D91 and E92) and two residues of H2B (human H2B, E105 and E113) create this strongly-negative charged groove, that can coordinate an arginine residue from the interacting protein[59]. Due to the tight binding this specific arginine is referred to as `arginine anchor´[60]. Latency-associated nuclear antigen (LANA)-peptide of Kaposi's sarcoma–associated herpesvirus (KSHV), nucleosome-binding domain of the human cytomegalovirus (CMV) immediate early 1 (IE1) protein and the C-terminal sequence of the chromatin-binding sequence (CBS) motif of the prototype foamy virus (PFV) structural protein GAG, all show binding to the acidic patch via an `arginine anchor´[60–62] (see Fig. 5). Probably these viral proteins interact with the nucleosomes to alter chromatin architecture in their favor regarding self-replication[61]. Other acidic patch binders like regulator of chromosome condensation 1 (RCC1) and silent information regulator (Sir3) bromoassociated homology (BAH) domain show additional binding to other regions of the nucleosome in addition to the `arginine anchor´[63,64]. Beside heterotypic interactions the acidic patch was shown to be relevant for general

chromatin organization. Neutralizing the acidic patch leads to loss of secondary structure formation *in vitro*[65]. The N-terminal tail of H4 interacts *in-trans* with the acidic patch of an adjacent nucleosome (and an additional site in H2B)[66], mediated by H4K16 acetylation[67]. Further *in-cis* interaction of the H4 N-terminal tail with the nucleosome´s own acidic patch indirectly stabilizes the DNA at the dyad[68].
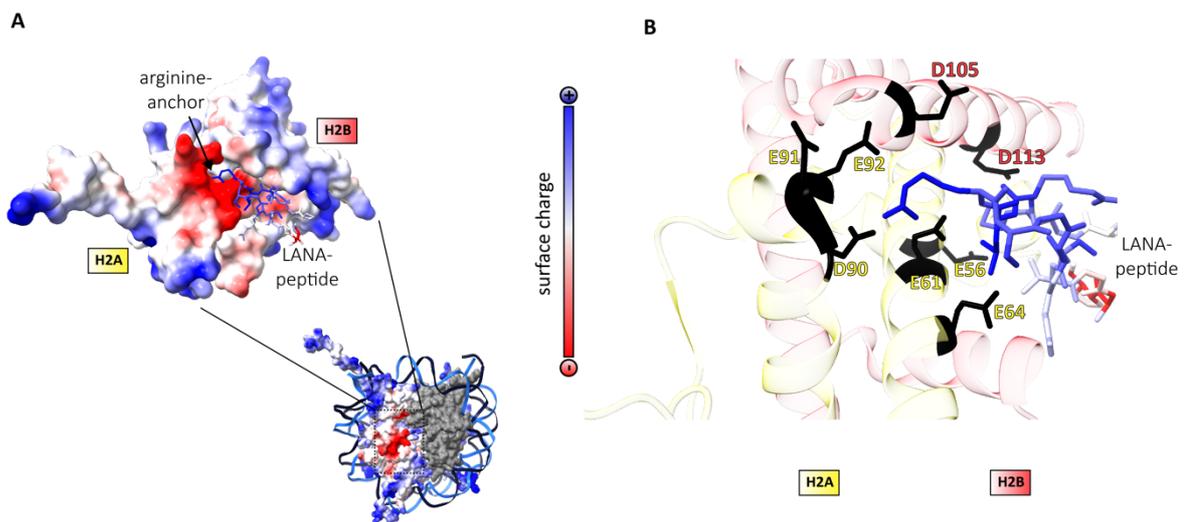
Figure 5: Acidic patch of the nucleosome. A Overview and magnified view of the acidic patch binding pocket of the histone octamer formed between histone H2A and H2B and the LANA peptide specifically binding to this region. Surface potential plot shows the strong negative (red) cavity formed. **B** Close up of the acidic patch region with the relevant residues shown in dark grey. Label color of the respective residues corresponds to the histone´s color code. The LANA peptide is colored according to its charge. The prominent dark blue arginine is shown in the center coordinated by the residues forming the acidic patch. |based on PDB: 1ZLA

2.2 The innate immune system – self vs. non-self

Invasion of pathogens and mislocalization of components also found in the host organism represent a major threat for cellular survival. Therefore, mechanisms against pathogenic invasion can be found even in the most basic forms of life with prokaryotes using restriction enzymes and the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system to protect themselves from pathogenic invasion[69]. The innate immune system displays the first line of evolutionary conserved eukaryotic host-defense mechanisms, found in all branches of eukaryotic life with increasing functionality[70]. Despite its ancient origin it is now well established, that the innate immune system in vertebrates is not simply an evolutionary remnant but a prerequisite for the functionality of the adaptive immune system complementing it at the initial response[71]. As a requirement for innate immunity to work as a defense system, mechanisms had to evolve to distinguish between self and non-self antigens. For this purpose a set of non-clonal germline encoded receptors, the so called pattern recognition receptors (PRRs), emerged[72]. These types of receptors recognize either antigens found in different microbial pathogens, so called Pathogen Associated Molecular Patterns (PAMPs), or antigens produced by the host cell as a result of pathogen invasion, so called Damage Associated Molecular Patterns (DAMPs)[73]. PAMPs are common features of many pathogens, often associated with viability, impeding loss of recognition by the PRRs in an evolutionary time scale[74,75]. In contrast, DAMPs serve distinct functions in the healthy cell but act as an antigen when mislocalized during infection. Binding of any molecular pattern type to PRRs will elicit an innate immune response ultimately leading to inflammation or cell death; directly or via activation of the adaptive immune system[76].

2.3 Pattern recognition receptors

The concept of universal pathogen detection receptors was introduced by Janeway in 1989 and is considered to be the foundation for the understanding of the innate immune system[77]. In contrast to the somatic rearrangement mechanisms used for antigen detection in the adaptive immune system, only found in vertebrates, PRRs are germline encoded. This assumes a limited number of receptor families for common immunogenic targets (see Fig. 6).

The PRRs can be split into different groups by their specific domain architecture, with the major groups being Toll-like receptors (TLRs), C-type lectin receptors (CLRs), nucleotide-binding domain, leucine-rich repeat (LRR)-containing (or NOD-like) receptors (NLRs), RIG-I- like receptors (RLRs), AIM2-like receptors (ALRs) and the cGAS-STING pathway. Further classification can be done based on their cellular localization. TLRs and CLRs are mostly membrane bound proteins detecting extracellular or endocytic pathogens. In contrast NLRs, RLRs, ALRs and cGAS are not membrane bound thereby detecting cytosolic pathogens[78]. Binding to any of these receptors activates specific signaling cascades leading to transcriptional upregulation of pro-inflammatory cytokines and interferons (IFNs) that orchestrate subsequent innate and adaptive immune responses or have direct effects like opsonization, phagocytosis, autophagy or even cell death[79,80]. A common feature of PRRs is the need of downstream adaptor molecules, integrating the signal from different homo- or heterotypic receptors and linking them to enzymatic activity[78].

The most well studied class of PRRs are the TLRs. The Toll receptor from *Drosophila* was the first discovered and eponymous member of this family. Known to be needed for dorsoventral polarity during embryogenesis in *Drosophila*, it was later shown to play a role in antifungal response in insects[81,82]. To date 13 members of the TLR family have been found in mammals, with TLR 1 to 10 found in human and TLR 1-13 (lacking TLR 10) in mice. Members of this large group show a broad ligand diversity against PAMPs originating from bacteria, viruses or fungi. For example, different components of the bacterial cell wall like lipopolysaccharides, lipoproteins or other features like flagellin bind to TLRs[83,84]. Besides, some viral proteins, bacterial and viral nucleic acids or fungal zymosan are also ligands for members of this family[85–87]. Heat-shock-proteins or fibrinogen from the host cell can serve as DAMPs[88,89].

CLRs were originally named after Ca^{2+}-dependent binding to glycan structures often found on microorganisms, but now compose of a large group of soluble and membrane bound proteins with a plethora of cellular functions ranging from development to immunity[90]. The first members of this family were characterized by binding to mannose and galactose often found on the outer surface of bacteria and fungi, leading to an innate immune response. In recent times it became clear, that also other ligands like proteins and lipids may serve as a ligand for the CLR family[91,92].

NLRs can be seen as the intracellular counterpart to TLRs also detecting many structural elements of bacteria like components of the cell wall or flagellin but also toxins. Beside these PAMPs many host-derived DAMPs like uric acid, ATP or damaged membranes can serve as ligand[93]. Many members of the NLR family will oligomerize forming the starting point of inflammasome formation with three out of the four inflammasomes corresponding to members of this family (NLRP1, NLRP3 and NLRC4)[94].

RLRs, ALRs and cGAS serve as cytosolic sensors for non-self or mislocalized host nucleic acids and will be extensively discussed later on[95].
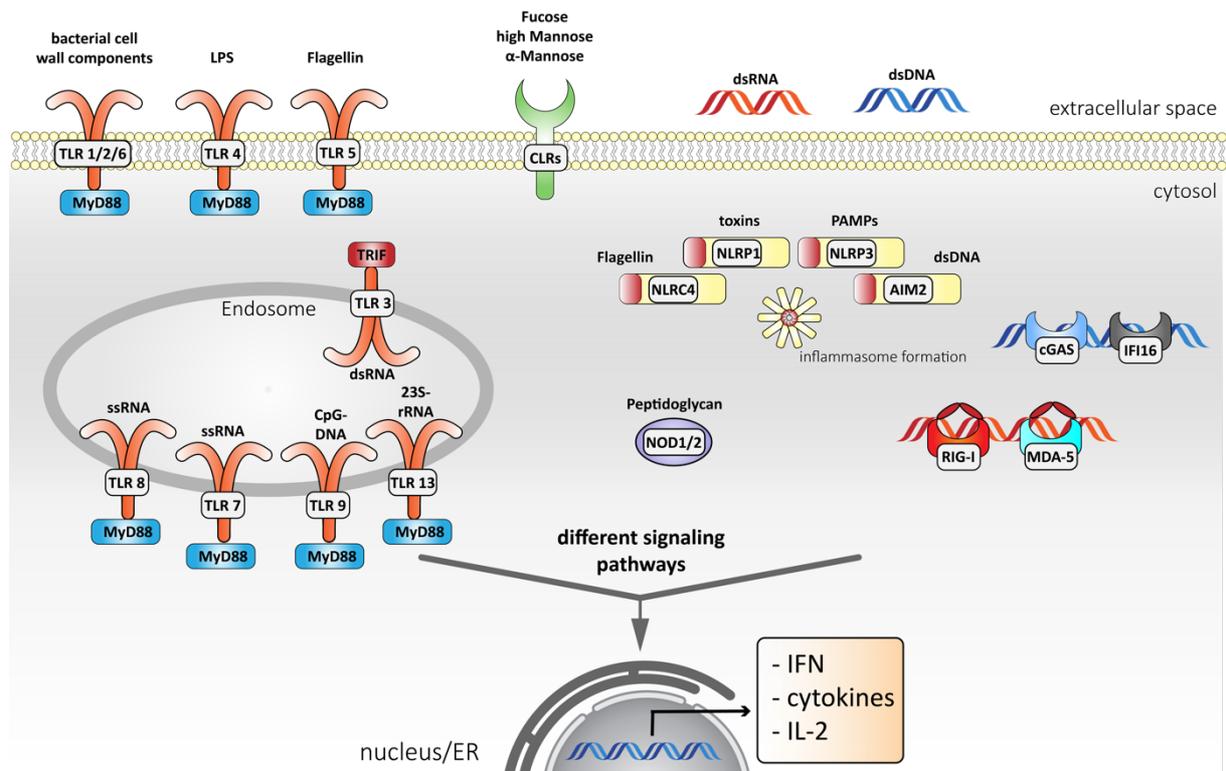
Figure 6: Simplified overview of pattern recognition receptor signaling. Different classes of PRRs are shown regarding their cellular location and their specific ligands (shown in fat letters). The membrane bound subset of TLRs and the CLRs bind to common pathogenic substances or PAMPs available if the pathogen has not entered the cell yet. Examples of these ligands are molecules often found on the cell wall of pathogens like bacterial cell wall components or different sugars. Endosomal membrane bound TLRs can sense many kinds of intracellular nucleic acids that emerge from pathogenic invasion. The intracellular NLRs can sense pathogenic components like flagellin or toxins derived from pathogens, when they already entered the cell. NLR sensing will lead to the formation of the inflammasome singaling complex. NOD1/2 also sense bacterial components like peptidoglycan. cGAS and IFI16 sense intracellular dsDNA, whereas RLRs (RIG-I, MDA-5) are sensors for cytosolic RNAs. All these PRRs will lead to further downstream signaling via adaptor molecules, that will ultimatively end up in an immune response, like the induction of different interferons, cytokines or interleukins.

2.4 Nucleic acid sensing by pattern recognition receptors

2.4.1 Nucleic acid sensing TLRs

All members of the TLR group are type I transmembrane proteins of the Interleukin-I receptor (IL-1R) family. Their shared architecture consists of a horseshoe-shaped N-terminal leucine-rich-repeat (LRR) ligand binding domain, a transmembrane domain and the cytosolic Toll/IL-1 (TIR) domain for signal transduction[96].

As mentioned before there are 13 TLRs known in mammals, with TLR 11 to 13 being specific to mice. Within this group TLR 1, 2, 4, 5 and 6 reside on the cell surface detecting extracellular pathogens, whereas the others are bound to endosomal membranes facing the lumen sensing nucleic acids after pathogen uptake and acidic lysis[97]. TLR 3 is activated by double-stranded RNA (dsRNA); TLR 7, human TLR 8 and murine TLR 13 detect single-stranded RNA (ssRNA) and TLR 9 binds CpG-rich single-stranded DNA (ssDNA)[98–103]. TLR 7 and 8 can also detect nucleobase related compounds[104,105]. Although the structural basis for ligand binding differs for different TLR ectodomains (ECDs), all TLRs will form a m-shaped dimer by pairing the C-terminal ECDs. This induces a conformational change leading to dimerization of the cytosolic TIR domains, priming them for adaptor molecule binding[96]. The linkage to several signaling cascades is modulated by binding of different adaptor molecules. Most nucleic acid sensing TLRs use myeloid differentiation primary-response protein 88 (MyD88) as adaptor molecule. In this pathway MyD88 activates IRAK (IL-1-receptor-associated kinases) and TRAF6 (tumor-necrosis-factor-receptor-associated factor 6), leading to release of nuclear factor-κB (NF-κB) from the inhibitor of nuclear factor-κB (IκB)-kinase complex (IKK-complex), its nuclear translocation and expression of inflammatory cytokines. In contrast TLR 3 uses TIR-domain-containing adaptor protein inducing interferon (IFN)-β (TRIF) to activate interferon-regulatory factor 3 (IRF3) to activate NF-κB[106].

2.4.2 NLRs and ALRs – the inflammasome forming nucleic acid PRRs

Other groups of PRRs able to sense cytosolic DNAs are the NLRs and ALRs, which are known to assemble into the large signaling complex known as the inflammasome. Due to shared domains these two groups can work together.

In contrast to TLRs these receptors are not membrane bound and were found to recognize cytosolic nucleic acids. Proteins of the NLR family share a domain architecture composed of three features: (1) a N-terminal protein-protein interaction domain, which can be a caspase recruitment domain (CARD), a pyrin domain (PYD), or baculovirus inhibitor domain (BIR); (2) a nucleotide-binding and oligomerization domain (NOD/NACHT) for self-oligomerization and (3) a C-terminal leucine-rich repeat (LRR) for ligand binding[107]. While members of the ALR family also have an N-terminal pyrin domain for adapter molecule binding, their dsDNA binding ability is facilitated by hematopoietic expression, interferon-inducible nature, and nuclear localization (HIN) domains[108]. Most of the 23 known human NLRs detect structural components, toxins from microorganism or DAMPs like extracellular ATP or uric acid crystals (detection pattern comparable to intracellular TLRs)[109]. Especially NLRP3 was shown to act as a cytosolic nucleic acid sensor by detecting viral RNA and oxidized mitochondrial DNA (ox-mtDNA)[110,111]. NLRP3 binds its ligand via the LRR-domain, leading to a homotypic oligomerization allowing binding of the adaptor protein Apoptosis-associated speck-like protein containing a CARD (ASC) serving as nucleation point for subsequent helical ASC-filament formation. ASC harbors an N-terminal PYD and a C-terminal CARD domain, linking caspase-1 to the complex via homotypic CARD interactions, thereby cleaving and activating caspase-1[112]. Cleaved caspase-1 activates IL-1β and IL-18 leading to inflammatory cytokine release or directly cleaves gasdermin D inducing pyroptotic cell death[113,114].

The same pathway can be activated by members of the ALR family. Absent in melanoma 2 (AIM2) and γ-interferon inducible gene 16 (IFI16) directly bind dsDNAs via their HIN domains (two in case of IFI16), thereby oligomerizing on the DNA strand to provide the filamentous ASC a binding platform via their PYD domains[115,116]. Additionally, IFI16 can also directly activate STING leading to an IFN-response[117]. Despite first being recognized as cytosolic DNA sensor, a nuclear localization of IFI16 was also shown, facilitating activation upon nuclear non-self-DNA encounter.

Therefore viruses that rely on nuclear replication as herpes simplex virus (HSV), KSHV, CMV, and human immunodeficiency virus (HIV) can be detected by nuclear IFI16[118,119]. Even with the sensing of pathogenic DNA happening in the nucleus activation of the downstream signaling cascade still occurs in the cytoplasm.

2.4.3 RIG-I like receptors – cytosolic RNA sensors

The RIG-I like receptor (RLR) family of cytosolic RNA sensors encompasses three members: (1) Retinoic acid-inducible gene I (RIG-I), (2) melanoma differentiation-associated protein 5 (MDA5) and (3) laboratory of genetics and physiology 2 (LGP2)[120]. RIG-I and MDA5 were found in screenings for type I interferon inducing genes[121,122] and were quickly recognized as the major sensor for viral RNAs[123]. Structurally all members of this family have a central DExD/H-box helicase core (as found for superfamily 2 RNA helicase members). This core is composed of two helicase domains HEL1 and HEL2, with the latter one containing a helicase insertion domain (HEL2i). In addition, all members show a C-terminal RNA binding domain (CTD), also known as repressor domain (RD). The CTD is linked to HEL2 via a helix extension pincer domain. N-terminal double CARD-domains for signal transduction are just found for RIG-I and MDA-5[124]. Apo-RIG-I adopts an inactive conformation with the CARDs being sequestered by interaction with HEL2i. RNA binding releases the CARDs facilitating downstream signaling[125,126]. The released CARDs of RIG-I interact homotypically with mitochondrial antiviral-signaling protein (MAVS) leading to oligomerization[127]. Unlike RIG-I, MDA5 is not kept in check by autoinhibition but oligomerization on long dsRNA is needed for activation[128]. MAVS subsequently activates TANK-binding kinase 1 (TBK1) and IκB kinase-ε (IKKε), which in turn will lead to IRF3/IRF7 and NF-κB activation, driving an IFN and antiviral response (see Fig. 7).

Despite their structural similarity, members of this class detect different specific RNA substrates. RIG-I detects features at the 5´- end of RNAs that are common for viral RNAs but missing in host RNAs. These RNAs need to have a triphosphate or uncapped diphosphate at the 5´-end[129–131], base-pairing at the 5´-end[132] and the 5'-terminal nucleotide needs to be unmethylated at its 2'-O[133]. This specificity enables differentiation between host RNAs and viral RNAs. Negative-sense RNA viruses, contain so called `pan-handles´, 5´-duplex structures that are recognized[132]. The genome of Influenza Virus A and its replication intermediates activate RIG-I[134,135]. Some positive-sense RNA viruses like Dengue and Zika virus are also recognized by RIG-I (uncapped 5´-PPP intermediate)[136]. Less is known about the exact ligand of MDA5. MDA5 needs higher order RNA substrates for sufficient oligomerization[137]. While it is essential for sensing picornaviridae[138], it is dispensable for the recognition of most other virus families[139].
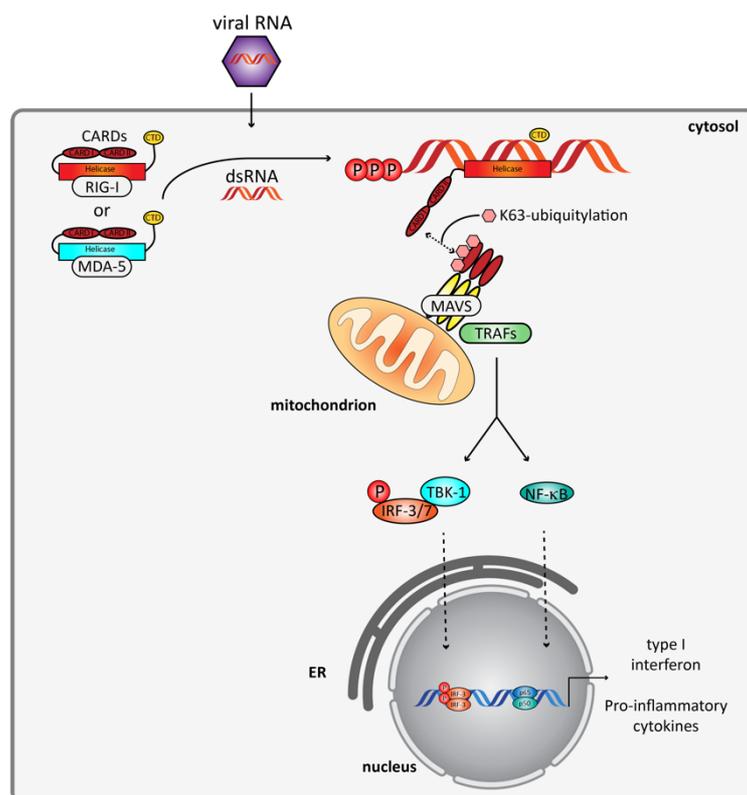
Figure 7: RLR signaling. Exemplary overview of RLR signaling shown for RIG-I. Apo-RIG-I exists in an autoinhibited state with its CARD-domains sequestered. Upon binding of non-self RNAs the CARDs get released and are free for interaction with MAVS. K63-ubiquitylation of this complex further stabilizes it. MAVS oligomerization at the mitochondria serves as a binding platform of TRAFs which will lead to IFN induction or NF-κB activation.

2.4.4 cGAS-STING axis

Many of the introduced nucleic acid sensors like TLRs or RLRs still rely on specific structural or chemical features of the nucleic acids for recognition. Although the ALR family can detect canonical B-form DNA, their presence alone is not sufficient for a full IFN-response[140]. At this stage cGAS, the main cytosolic DNA sensor in most cell types comes into play[141]. cGAS does not only detect viral DNA (HSV-1, vaccinia virus, HIV, CMV)[141–143], and intracellular bacterial DNA (*Chlamydia trachomatis*, *Mycobacterium tuberculosis*, *Francisella novicida*, *Listeria monocytogenes*, *Neisseria gonorrhoea*)[144–147], but also mislocalized self-DNA from different sources.

Prior to the discovery of cGAS as the main cytosolic DNA sensor, its `adaptor´ STING (other names MITA, MPYS or ERIS) was already known to be involved in antiviral DNA sensing. STING was found to be an ER-membrane bound protein with its CTD facing the cytosol, eliciting an IRF3 related IFN response upon viral encounter[148–150]. Initially thought to directly bind DNA, STING turned out to be a receptor for cyclic dinucleotides[151,152]. With the discovery of cGAS´ DNA sensing ability and its catalytic mechanism the cGAS-STING pathway was established. Upon DNA binding cGAS will produce the second messenger cyclic GMP–AMP (cGAMP). cGAMP binds to STING, promoting its translocation to the Golgi apparatus and activation of TBK-1. The latter will phosphorylate IRF-3, leading to its dimerization and nuclear translocation, triggering a type I IFN response and expression of interferon-stimulated genes (ISGs) [153]. Although STING can also be directly activated by bacterial cyclic dinucleotides[152], it shows enhanced binding affinity to cGAMP produced by cGAS. Eukaryotic cGAMP is characterized by an 2´–5´-phosphodiester bond between the 2´O of guanosine and the 5´O of adenosine instead of the bacterial 3´-5´-linkage[154]. Activated STING can also activate NF-κb via IKK and induce signal transducer and activator of transcription 6 (STAT6) signaling[155,156]. Classical induction of ISG expression by cGAS-STING results in an IFN-mediated positive feedback-loop[157] (see Fig. 8).
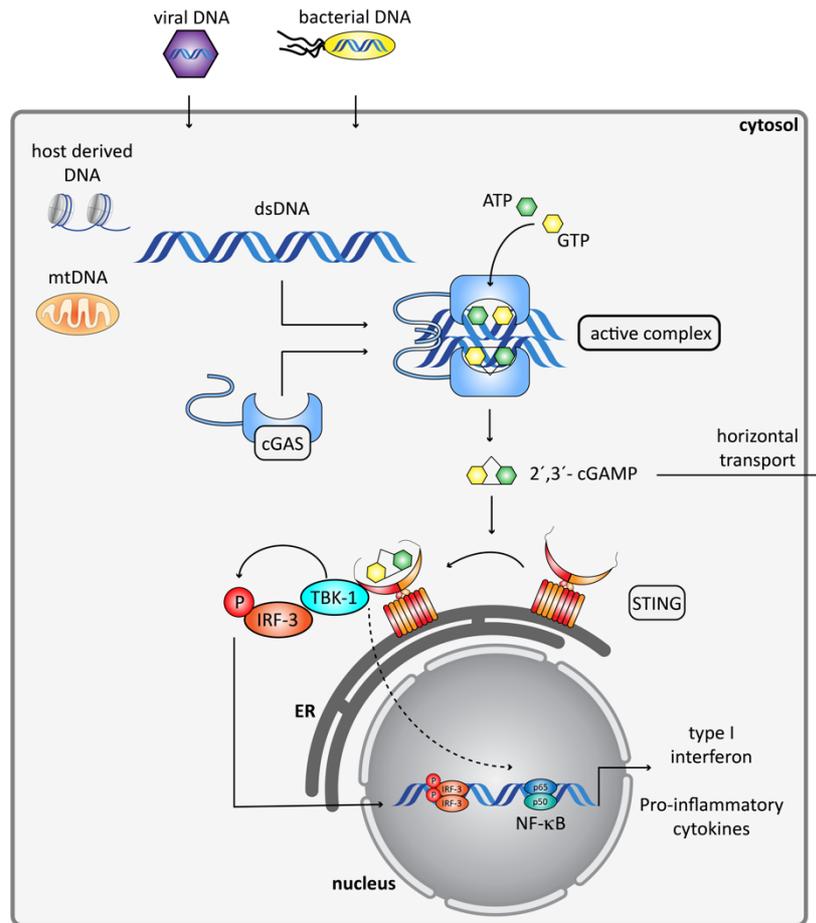
Figure 8: cGAS signaling. cGAS binds sequence unspecifically to dsDNA from different sources occurring in the cytosol. Upon dsDNA binding cGAS will form an active complex. The minimal active complex is a cGAS/DNA 2:2 complex. In this conformation cGAS will produce the second messenger 2´,3´- cGAMP from ATP and GTP. This second messenger can spread to neighboring cells via gap junctional transport. 2´,3´- cGAMP is the ligand for STING, leading to its translocation, oligomerization and activation of TBK-1, that in turn will phosphorylate IRF-3. Phosphorylated IRF-3 can travel to the nucleus and induce the transcription of type I interferons. Another IRF-3 independent mode of STING action is the activation of NF-κB after cGAMP binding.

2.4.5 Other DNA sensors

Prior to the recognition of cGAS as the primary cytosolic DNA sensor other candidates have been discussed. As the name implies DNA-dependent activator of IFN-regulatory factors (DAI) was shown to bind DNA and upregulate IFN-I response via NF-κB and IRF3[158]. While effects on the immune response by this sensor were observed for specific cell types and viral infections[159–161], the underlying sensing was not universal[162,163]. This fact suggested the need of another universal sensor, that was found with cGAS.

RNA Polymerase III was shown to transcribe AT-rich dsDNA, such as poly(dA:dT), into an RNA-containing 5'-triphosphate ligand for RIG-I leading to IFN-β response by indirect DNA sensing[164].

DExD/H box helicase 41 (DDX41), despite having a SF2-helicase domain like RIG-I, was shown to bind to dsDNA but not RNA, acting on the STING-TBK1 pathway. Knockdown of DDX41 leads to compromised response to HSV-1 infection[165,166]. Furthermore, it was described to directly sense bacterial CDNs leading to STING activation[167].

DNA dependent protein kinase (DNA-PK), a heterotrimeric protein complex consisting of Ku70, Ku80 and the catalytic subunit DNA-PKcs known for its function in DNA repair, was shown to sense DNA over 500 bp length leading to IFN-3 response via IRF1 and IRF7[168]. Especially in fibroblasts, it was shown to colocalize with cytosolic DNA leading to IFN-1 induction via STING-TBK-1[169].

Another protein involved in DNA repair, MRE11, was shown to be involved in the cytosolic DNA response, but not to viral DNA[170]. Rad50 as part of the MRE11 complex was shown to translocate to the cytosol upon viral DNA infection, inducing STING-independent NF-κB activation [171].

2.5 Many ways to die – cellular consequences of DNA sensing

The best studied cellular response of nucleic acid sensing is the induction of cytokines, that can be found for all previously described sensors[172]. Regarding DNA-sensing, type I interferons (IFNs) are the most important. IFNs produced after the activation of IFN-genes by DNA sensing pathways will lead to the induction of IFN-stimulated genes (ISGs) via JAK-STAT signaling. ISGs encode for a plethora of immunogenic proteins, can activate positive feedback loops and link DNA sensing to adaptive immunity. Major histocompatibility complex (MHC) and other costimulatory molecules needed for the adaptive immune response are ISGs[173,174]. Additionally, the induction of multiple programmed cell death pathways as *ultima ratio* is linked to nucleic acid sensing. The known death pathways orchestrated by nucleic acid sensing are apoptosis, necroptosis and pyroptosis[175]. Cell death can either be non-lytic and therefore immunogenically silent like apoptosis, or lytic and highly inflammatory like necroptosis and pyroptosis[176]. Hallmarks of apoptosis are cell shrinkage and DNA fragmentation with formation of apoptotic bodies, whereas the other two pathways are characterized by pore formation and therefore loss of membrane integrity, explaining their inflammatory potential[177]. Inducing cell death may have multiple beneficial effects for the host. Additional to the secretion of cytokines found in lytic cell death, intracellular pathogens can directly be cleared, and many DAMPs get released amplifying the immune response[178,179]. Besides, cell death leads to antigen uptake by professional antigen presenting cells linking the innate immune response to adaptive immunity[180]. IFI16 and AIM2 lead to pyroptosis via inflammasome mediated gasdermin D activation. Many studies link the cGAS-STING pathway to the induction of apoptosis, maybe even omitting the need of DNA sensing by cGAS. Apoptosis dependent on STING was observed in malignant B-cells and T-cells[181,182], and upon viral encounter in some other cell types[182,183]. Also necroptosis was observed via the cGAS-STING pathway in bone marrow–derived macrophages after DNA sensing[184]. RNA recognition by RIG-I leads to the `RIG-I-like receptor-induced IRF3 mediated pathway of apoptosis´ (RIPA) needing MAVS and IRF3 activation[185] or also necroptosis via the same pathway as cGAS-STING[184]. Also, nucleic acid sensors of the TLR family are able to induce different types of cell death depending on cell type and stimulus. In summary, any activation of a DNA sensing pathway may ultimately lead to cell death but the exact route will rely on the exact stimulus and cell type, underlining the fine-tuned innate immune response[175].

2.6 cGAS – the most powerful sensor for mislocalized self-DNA

2.6.1 cGAS – the question of cellular location

When cGAS was discovered as the missing link for cytosolic DNA sensing it seemed clear, that it is a predominantly cytosolic protein, relying on this compartmentalization for discrimination between self and non-self. As the sensing is sequence unspecific it would just detect any type of dsDNA not residing in the correct location. Unfortunately, this concept already fails considering healthy cells at the stage of cell division, where the lack of clear distinction between cytosol and nucleus does not result in extensive autoimmunity.

Now growing evidence suggests that cGAS is at least also a nuclear, maybe even predominantly a nuclear protein. cGAS was shown to interact with IFI16 in the nucleus of fibroblasts during HSV-1 sensing[186], entering the nucleus in mitotic-cells during proliferation[187] and being recruited to DNA damage sites[188]. At this stage only the nuclear localization of cGAS was shown, still evoking the question how autoimmunity is prevented. This timepoint represents the knowledge-basis at the start of this work. During this work it became more and more clear, that cGAS also interacts with chromatin directly, especially with nucleosomes. Chromatin was shown as an activator of cGAS, with mononucleosomes being high affinity ligands for cGAS without activating it[189]. Later it was shown, that cGAS is kept in an inactive state by being tethered to chromatin via the conserved acidic patch[190]. The specific interaction between cGAS and nucleosomes is the major finding of this thesis and will be discussed in detail later.

2.6.2 DNA ligands for cGAS

2.6.2.1 Non-self-DNA

The most well established cGAS DNA ligands are of viral or bacterial origin, with many examples already mentioned (1.4.4). The sensing of cytosolic DNAs of these sources is assumed to happen after breakage of viral particles or bacterial cell walls, exposing the DNA to cGAS. With the increasing evidence of nuclear cGAS, DNA sensing of nuclear replicating viruses (as seen for IFI16) may be considered as an additional task of cGAS. This is supported by the fact, that cGAS and IFI16 work together in HSV-1 sensing[117]. In fact, HIV-2 sensing by cGAS happens in the nucleus with the help of non-POU domain-containing octamer binding protein (NONO), linking HIV-capsids to DNA[191].

2.6.2.2 Self-DNA

The major benefit of cGAS´ DNA sensing capabilities compared to other PRRs is the recognition of multiple self-DNA species not related to pathogens but cellular malfunctions like mtDNA leakage, tumor DNA, DNA damage, or occurrence of micronuclei [188,192–194]. Recent work has shown the interaction of cGAS with healthy nuclear DNA suggesting an inactive conformation. In contrast to this, micronuclei or other chromatin intermediates resulting from replication errors, DNA damage or malfunctioning mitosis serve as activators for cGAS. Micronuclei occur after chromosomal mis-segregation due to exogenous DNA damage or spontaneously in human cancer cells, and display chromatin fragments surrounded by components of the nuclear envelope in a composition more prone to membrane rupture[195]. cGAS was shown to colocalize with micronuclear DNA leading to an immune response[196]. The senescence-associated secretory phenotype (SASP), secretion of a specific cytokine pattern as a result of a permanent cell cycle arrest, is associated with cGAS activated by cytosolic chromatin fragments[197–199] (CCFs). Another important self-DNA pattern recognized by cGAS are extrachromosomal telomere repeats (ECTR), a unique feature of cancer cells, produced via the alternative lengthening of telomeres pathway (ALT-pathway). ECTR created in normal fibroblasts lead to a cGAS-STING related type I IFN response and IFN-β production, whereas cancer cells relying on the ALT-pathway are deficient of cGAS-STING signaling[200,201].

Concluding, most alterations of the normal chromosomal structure will lead to detection by cGAS. This is supported by the fact that mutating proteins needed for proper genome integrity, like helicases, topoisomerases or proteins involved in DNA repair, will all lead to activation of the cGAS-STING pathway[202–204]. Still, it is debatable whether the observed effects can be explained just by cGAS´ DNA binding or additional regulatory factors that will recruit cGAS to these specific targets.

Mitochondria are the only other place besides the nucleus in the eukaryotic cell, where DNA should occur under healthy conditions. Mitochondrial stress, leading to aberrant packing of mtDNA by depleting TFAM (transcription factor A, mitochondrial), can lead to cytosolic leakage of mtDNA serving as a ligand for cGAS[192]. Also, some viruses, like HSV or Dengue virus will lead to mtDNA release. This mechanism is especially useful as it enables the detection of the single-stranded RNA Dengue virus via cGAS[205,206]. Interestingly, the well-established mitochondrial DNA and cytochrome C release during apoptosis does not result in cGAS activation, as the involved effector caspases cleave and thereby inactivate cGAS[207].

2.6.3 cGAS structure and active complexes

cGAS is an ~60 kDa sized protein consisting of a ~160 amino acids long, non-structured, positively charged N-terminal tail, and a male abnormal 21 (MAB21)-domain containing, catalytic globular domain (see Fig. 9A). Since cGAS´ discovery many structures of the catalytic domain from multiple species with different ligands have been published, even recently revealing additional features in the human protein. Howsoever the exact structure and function of the N-terminal tail is still puzzling[208–213]. Despite the lack of detailed structural information, the N-terminal part of cGAS can bind DNA on its own, increases the affinity towards DNA, and elevates cGAS activity[214,215]. This increased activity is probably due to the formation of higher order condensation with higher local concentration, mediated by the N-terminus[216]. By its MAB21 domain cGAS is counted as member of the nucleotidyltransferase (NTase) family, all catalyzing phosphodiester bond formations. Despite their low sequence homology (due to diverse cellular functions) members of this family share a common NTase core structure with an α/β-fold composed of β-strands flanked by α-helices and frequent motifs for catalytic activity[217]. The globular domain of cGAS shows a bilobal structure, with the two lobes being connected by a `spine´ α-helix (α1), forming a cleft between them. The N-terminal lobe 1 harbors the NTase-like mixed α/β-fold with an eight-stranded central β-sheet (β1-β8) flanked by two α-helices, and the conserved catalytic triad (hcGAS E225, D227 and D319; mcGAS E211, D213, D307) coordinating two Mg^{2+} ions for subsequent nucleotide alignment (located on β1 and β6). An NTase like hG(GS)-motif (h for hydrophobic) can be found in the `activation loop´ connecting αA and β1 (hcGAS G212/S213; mcGAS G198/S199). The C-terminal lobe 2 consists of a four-helix bundle (αE-αH) stabilized by the spine (α1). The loop connecting αC and αD contains conserved histidine and cysteine residues needed for Zn^{2+} coordination and is therefore referred to as the Zn-Thumb (hcGAS H390, C396, C397 and C404; mcGAS H378, C384, C385 and C392)[208,218,219]. Apo-cGAS shows a rather unstructured activation loop unsuitable for correct nucleotide binding[208,210,212] (see Fig. 9B). It is noteworthy that all structural work on cGAS was performed using only the catalytic domain lacking the N-terminal tail. In this setup all apo-cGAS (no DNA bound) variants appeared as inactive monomers, whereas recently the existence of a dimeric full-length apo-form was identified by size-exclusion chromatography coupled with in-line multiangle light-scattering analysis (SEC-MALS)[211].

The molecular surface opposing the cleft serves as distinctly positively charged platform for DNA binding. Crystal structures of cGAS bound to DNA show sequence independent recognition by relying on extensive electrostatic interactions and hydrogen-bond formation with the phosphate backbone of the DNA. Further, two residues (one arginine finger for the human protein) reach into the minor groove of the DNA (hcGAS R176/H217; mcGAS R161/H203). The Zn-Thumb allows differentiation between A- and B-form DNA, by functioning as molecular ruler (see Fig. 9C).

Binding of B-form dsDNA will introduce a helix break in the spine, closing the cleft between lobe 1 and 2, structuring the activation loop for the catalytically efficient orientation of ATP and GTP coordinated by two Mg^{2+} ions (see Fig. 9D). This first discovered DNA binding site of cGAS is also called site A.
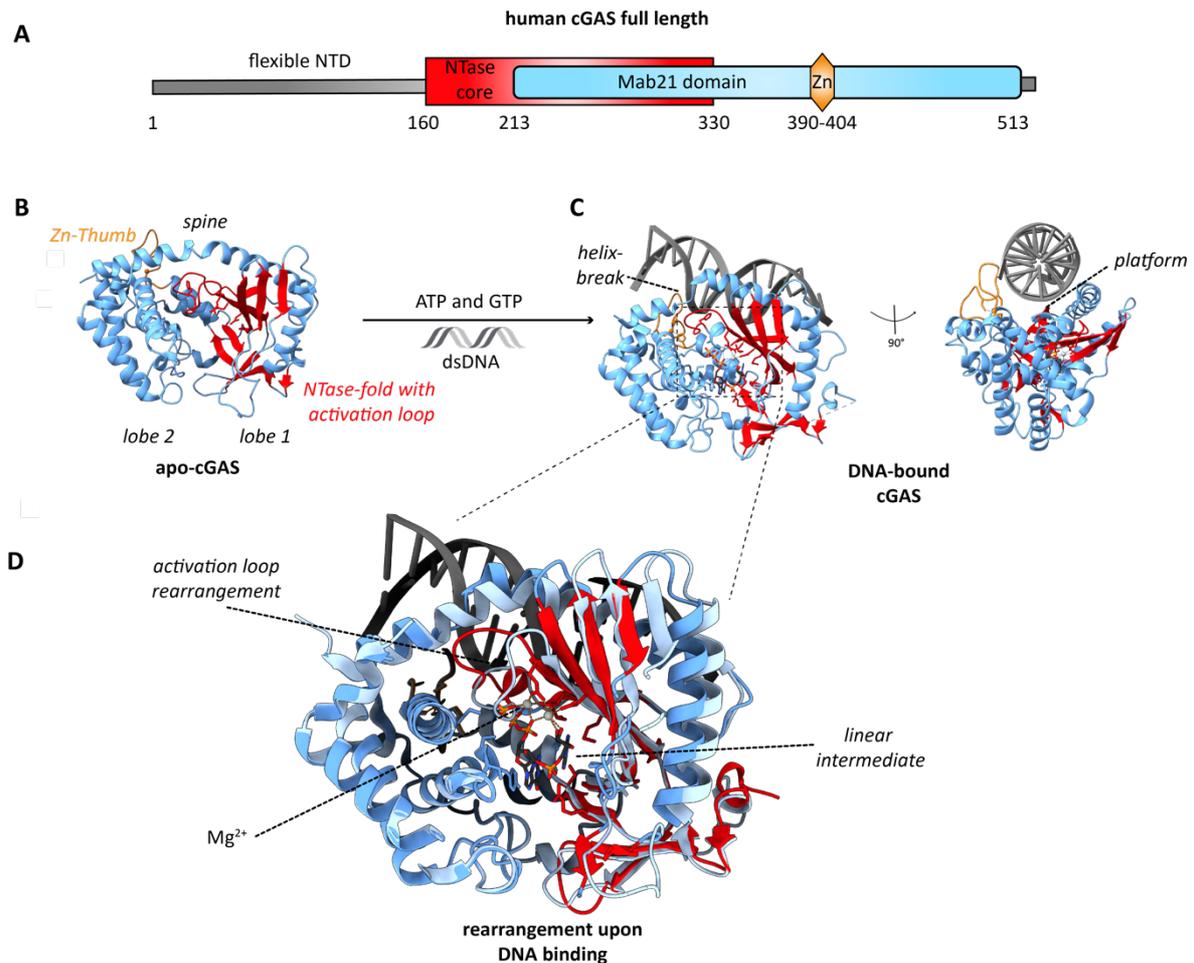
Figure 9: cGAS structure overview and rearrangement upon activation. A Overview of cGAS domain architecture. cGAS consists of a flexible N-terminal tail without structural knowledge to date (grey), that is followed by the Mab21 domain (blue) harboring the NTase core (red). These domains are relevant for ligand binding and formation of cGAMP. Additionally the Zn-Thumb is denoted in orange. **B** Overview of the apo-cGAS structure. Apo-cGAS adopts a bilobal structure, with the NTase fold shown in red. The two lobes are connected by a long spine-helix. In apo-form the activation loop, needed for cGAMP production is unstructured, missing the correct coordinaton of ATP and GTP. **C** DNA binding to cGAS introduces a break in the spine helix, and leads to restructuring of the activation loop. In side view, the platform for DNA binding can be observed. **D** Overlay of apo- and DNA-bound cGAS crystal structure. Upon DNA binding the activation loop in the NTase domain rearranges. The apo-loop is shown in light blue, whereas the rearranged loop is shown in red. The rearranged loop facilitates correct alignment of the catalytic triad leading to proper binding of two Mg^{2+} ions (dark grey) needed for ATP and GTP positioning. The displayed crystal structure shows binding of the first formed linear intermediate pppGp(2′-5′)A. |based on PDB apo cGAS: 4K8V; PDB DNA bound cGAS: 4K98

cGAS bound to DNA does not occur as active 1:1 complex in solution, but needs at least the formation of a 2:2 cGAS/DNA complex for further stabilization, as shown by analytical ultracentrifugation and small angle X-ray scattering [220]. In this complex two dsDNA strands get sandwiched between two cGAS monomers[221]. The dsDNA strand that is bound via site A to one cGAS protomer binds to the second cGAS protomer by an additional cGAS DNA binding site, called site B, and vice-versa. Site B consists of two salt-bridges formed with the DNA backbone (hcGAS R236/K347; mcGAS R222/K335) and a specific residue inserted in the minor groove of the DNA (hcGAS L354; mcGAS R342). Additional protein-protein interaction between residues located in the Zn-Thumbs of the monomers stabilize the complex (hcGAS K394/E398; mcGAS K382/E386)[214,221] (see Fig. 10A and 10B).

The minimal DNA length bound by one cGAS protomer is 16 to 18 bp, which is not sufficient for activation *in vivo*. cGAS shows DNA length depend activation, with dimer- and therefore active complex-formation allosterically favored by longer DNA substrates[214,222] (see Fig. 10C).

As cGAS does not lead to the common adaptor molecule mediated filament formation, local concentration increase is achieved by liquid-phase condensation presumably forming a mesh-like structure.

This super-structure is mediated by Zn^{2+}-depended interaction of the positively charged N-terminal tail with DNA and the involvement of an additional DNA binding site C[211,223]. DNA binding site C is most relevant for the human protein, as the interaction surface area of this site decreases when comparing with less evolved organisms (area of site C interaction: human>pig>mouse). In hcGAS site C has a 663 $Å^2$ interaction surface composed of three main elements, namely the α-region (hcGAS 261–286), the KRKR-loop (hcGAS 299–302), and the KKH-loop (hcGAS 427–432). Q264, K275, K279, K282, and K285 from the α-region; K299, R300, K301 and R302 from the KRKR loop and K427, K428, and K432 from the KKH-loop (numbering for hcGAS) facilitate non-specific electrostatic interactions with the DNA backbone. Additionally, an arginine finger (hcGAS R300) inserts into the minor groove of the DNA[211] (see Fig. 10D).

The main structural features of cGAS can also be found in another phylogenetically unrelated protein family[224]. The oligoadenylatsynthase (OAS) family shows high structural similarity, sharing the bilobal architecture, the conserved catalytic triad as well as the spine helix. In contrast to cGAS, OAS1 is a sensor for dsRNA. While it also forms a 2´–5´-phosphodiester bond, its catalytic product is not cyclic but linear oligoadenylate chains. Due to their structural similarities this group of nucleotidyltransferases is often referred to as the cGAS-OAS family[225]. Oligoadenylates will lead to RNAse L activation and thereby RNA degradation as a clearance mechanism.
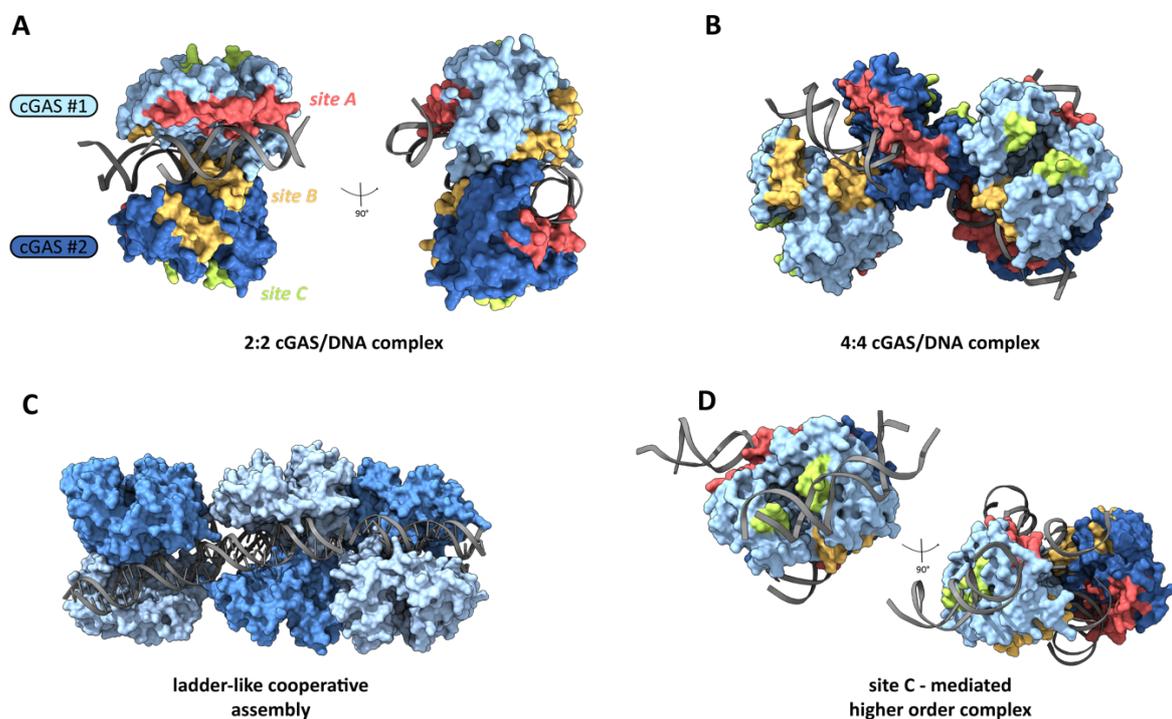
Figure 10: Different DNA binding sites of cGAS and their relevance for complex formation. A Minimal active unit represented by cGAS/DNA 2:2 complex. Individual cGAS monomers are depicted in different shades of blue. DNA binding sites are shown in different colours in their surface representation. cGAS shows three distinct DNA binding sites: A, B and C. Site A (salmon) is the first discovered DNA binding site found on the platform described in Fig. 9. Site B (yellow) is located at the opposing surface and is needed for active complex formation. Site B will bind to the DNA, that is bound by site A of the corresponding cGAS monomer. An additonal DNA binding site C (green) is needed for higher order complexes or phase-seperation. **B** Representation of two times 2:2 complex leading to a 4:4 complex. **C** Ladder-like cooperative assembly of cGAS on longer DNA substrates. When cGAS is binding to longer DNA, than depicted in A, B, and D, longer filament-like structures can be formed. **D** Three DNA strands bound to cGAS with an additonal DNA strand bound at site C. With all three strands bound to the cGAS protomer, formation of a higher order mesh-like active conformation is possible.| PDB A,B: 4LEY; C:5N6I; D: superimposition of 4LEY and 6EDB

2.6.4 cGAS – catalytic mechanism

Prior to their discovery in eukaryotes, cyclic dinucleotides have been known as signaling molecules in bacteria for over 30 years[226], being involved in functions like biofilm formation or different stress responses[227]. However, most bacterial CDNs are defined by two 3´–5´-phosphodiester bonds between the 3´O of one nucleotide and the 5´O of the second one (and vice-versa for cyclization). A hallmark of eukaryotic CDN signaling is the creation of mixed-linkage CDNs, as formed by cGAS and OAS proteins[228]. cGAS produces the mixed-based, and mixed-linked second messenger G(2´-5´)pA(3´-5´)p (also known as 2´-3´cGAMP or just cGAMP) from ATP and GTP, harboring one unusual 2´–5´-phosphodiester bond between the 2´O of guanosine and the 5´O of adenosine[210].

cGAMP production by cGAS is a two-step process with the linear intermediate pppGp(2´-5´)A formed first. After restructuring of the activation loop due to the DNA binding induced conformational switch, two Mg^{2+} ions bind to the catalytic triad (hcGAS E225/D227/D319; mcGAS E211/D213/D307) aligning ATP´s triphosphate. Further ATP triphosphate stabilization is achieved by hydrogen-bonds (hcGAS S213/S452; mcGAS S199/S420). Additional interactions consist of base-stacking of the adenine with tyrosine (hcGAS Y436; mcGAS Y421) and a hydrogen bond with ATP´s ribose (hcGAS E383; mcGAS E371). GTP is sandwiched in a 90° angle base-stacking with adenine on one side and a residue from the protein on the other side (hcGAS T321; mcGAS I309). Hydrogen-bonds specific for GTP guarantee selectivity for the mixed nucleotide linkage (hcGAS S378/S380/T211/R376; mcGAS S366/S368/T197/R364). Correct alignment of ATP and GTP results in efficient localization of ATP´s α-phosphate for nucleophilic attack of GTP´s 2´OH-group. Polarization of the attacking group by aspartate (hcGAS D319; mcGAS D307) enhances the reaction leading to the formation of the linear intermediate pppGp(2´-5´)A[208,210].

For cyclization to occur the intermediate has to bind donor and acceptor pockets in reverse order. GTP still bearing a triphosphate binds to the moiety that was occupied before by ATP using the same residues for triphosphate alignment. The steric constraints introduced by flipping the intermediate in the non-symmetrical binding pockets will lead to the nucleophilic attack of the 3´OH on the α-phosphate of GTP and cyclization[219] (see Fig. 11).

As the recognition site for ATP does not employ the multitude of interactions shown for the GTP binding site, also other linear intermediates like pppG(2'– 5')pG and even pppA(3'–5')pA may occur, but seem not relevant for full cyclization under physiological conditions[229]. Work on the catalytic mechanism revealed that most of the linear intermediate formed is released by cGAS in the first place, to be bound again for full cyclization[230]. Therefore, the increased local concentration of active cGAS complexes in multimeric assemblies may favor the two-step reaction, by providing high local concentrations of intermediate compared to the in general much more frequent ATP and GTP, and omitting linear intermediate accumulation.

Correctly assembled cGAMP will activate the downstream receptor STING in the same cell but can also horizontally spread to neighboring cells via gap-junctional transport of the signaling molecule[231]. Recently, transfer of cGAMP to other cells via the volume-regulated anion channel LRRC8 was shown to be crucial for defense against DNA viruses[232].

Beside producing a CDN by itself cGAS may also bind bacterial CDNs directly, promoting active complex formation and direct association with STING, skipping cGAMP production[233].
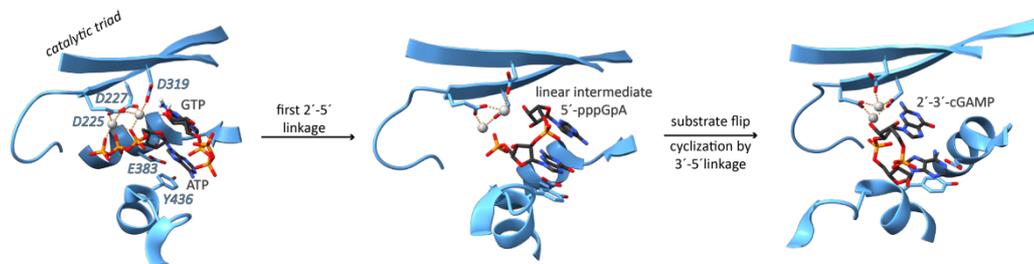
Figure 11: Catalytic mechanism of cGAS. In the first step ATP´s triphosphate is aligned by the catalytic triad D225/D227 and D319 (hcGAS) coordinating two Mg^{2+} ions. At this step Y436 base-stacks with adenine. Polarization of D319 facilitates the nucleophilic attack of GTP´s 2´-OH group leading to the linear intermediate 5´-pppGpA. For full cyclization the linear intermediate has to flip in the binding pocket. Now Y436 does not base-stack with the adenine-base but the guanosine-base. In that coordination the 3´-OH of the adenine base in the linear intermediate can attack the α-phosphate of the guanosine-base forming an unusual 3´-5´linkage and ring-closure. |PDB: 4KB6, 4K98,4K9A, 4K9B

2.7 STING – the downstream adaptor for cGAS

2.7.1 STING structure and IRF3-activation

STING is a ~42 kDa transmembrane protein residing in the membrane of the endoplasmic reticulum but translocating to the Golgi apparatus upon activation. Binding of its ligand cGAMP will ultimately lead to different immune responses with IRF3 activation being the best studied. Furthermore, NF-κB activation, autophagy and lysosomal cell death are possible outcomes of STING activation[234].

STING has four N-terminal transmembrane domains (TM1-TM4), that are connected to the C-terminal cytosolic part via a connection helix. The part facing the cytosol harbors the ligand binding domain (LBD) and the C-terminal tail (CTT) (see Fig. 12A). Whereas multiple crystal structures of the cytosolic part already exist since 2012[235–238], structures of the human and chicken full-length protein were just recently resolved using cryo-EM[239]. Already from the crystal structures one can tell that apo-STING forms a butterfly-shaped dimer creating a V-shaped pocket for binding of a single CDN. Added insights from the cryo-EM structure show packing of TM1 of one monomer with TM3 to TM4 of the other leading to a domain-swapped architecture of the dimer. The two connector helices (one of each monomer) link TM4 to the LBD with the linker loops between connector helices and LBDs crossing and forming a tight interaction (see Fig. 12B). Upon cGAMP binding the LBDs will rotate 180° clockwise (to resolve the right-handed crossing), leading to a parallel orientation of the connector loops[239]. The induced conformational change will lead to coverage of the CDN binding pocket by two β-strands forming a lid preventing cGAMP release (see Fig. 12C). This lid closure is seen in the crystal and also in the cryo-EM structure[238,239]. For subsequent signal transduction STING has conserved TBK1 binding and pLxIS motifs needed for IRF3 activation[240,241]. As found for most innate immunity proteins, STING needs oligomerization for downstream signaling to occur. Oligomerization is needed for *in-trans* TBK-1 phosphorlyation[242] and translocation to the ER-Golgi intermediate compartment (ERGIC) only occurs after correct oligomerization[243]. According to the cryo-EM structures STING oligomerization occurs as lateral stacking of dimers promoted by a platform formed by the conformational change induced by cGAMP binding[239] (see Fig. 12D). Phosphorylated and thereby activated TBK-1 will phosphorylate the conserved serine residue in the pLxIS motif, serving as a binding site for IRF3.

Binding of IRF3 to the pLxIS motif brings it into close proximity with TBK-1, allowing TBK-1 to activate IRF3 by phosphorylation[240]. The activated IRF3 will then dimerize, enter the nucleus, and induce transcription of type I interferons.
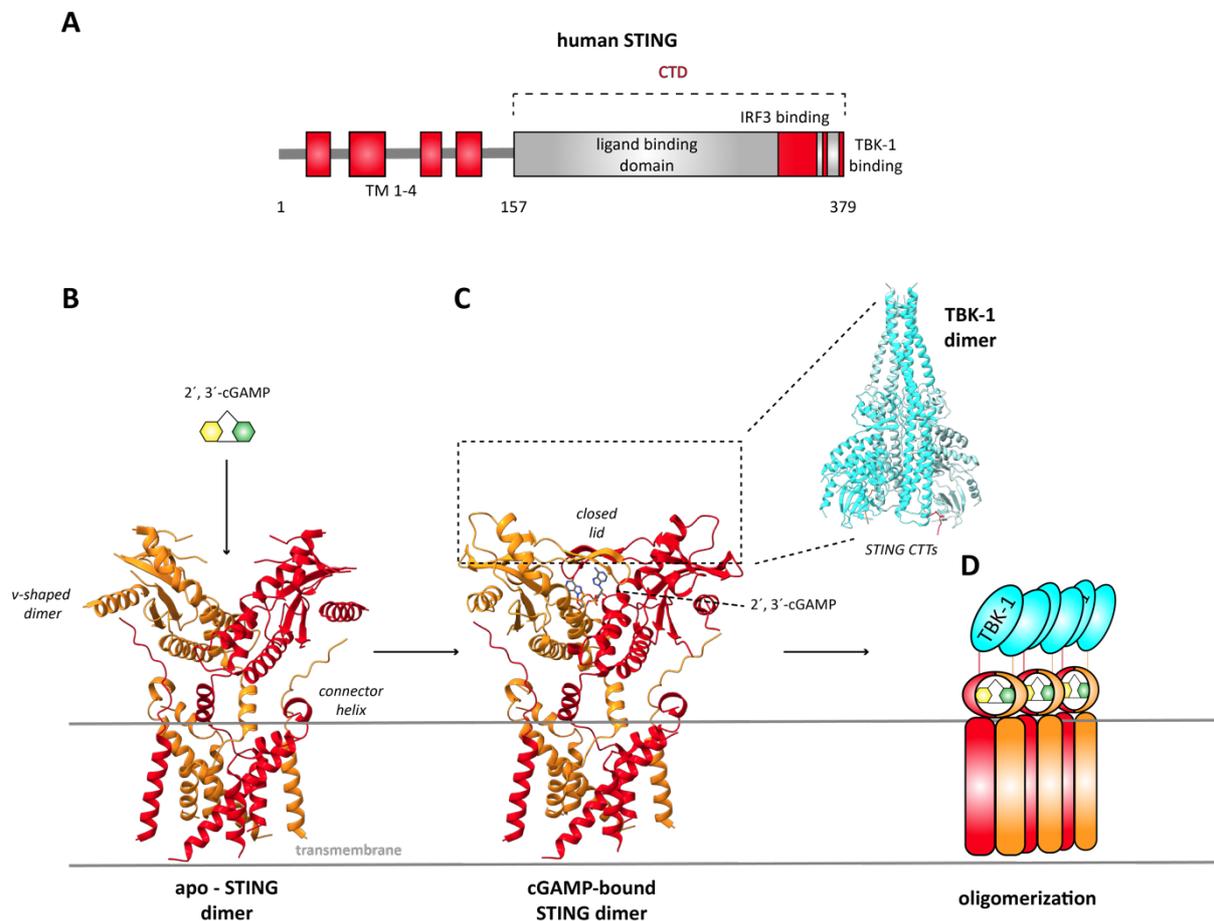
Figure 12: Structural overview of STING activation. A Domain architecture of STING. STING harbors fours transmembrane domains TM 1-4, that are followed by the cytosolic C-terminal domain (CTD) needed for ligand binding and downstream signaling. **B** STING occurs as membrane-bound dimer (red and orange), with the apo-state forming a V-shaped cytosolic pocket for cGAMP binding. The transmembrane domains interact in a domains swapped manner. **C** Upon ligand binding the CTD rearranges and closes a lid over a bound cGAMP molecule. The ligand bound form enables the binding of a TBK-1 dimer (cyan). The residual interaction sites of STINGs CTT are shown in red and orange at the bottom of TBK-1. **D** TBK-1 bound STING oligomerizes facilitating trans-phosphorylation of TBK-1 dimers that in turn will lead to IRF-3 phosphorylation and downstream signaling. |PDB for B: 6NT6; C: overlay of 6NT6 and 4SYK plus 6NT9

2.7.2 NF-κB activation via STING

Supplementing the well-known mechanism of IRF3 activation, NF-κB activation by STING was shown in human and mice to a comparably lower extent. Due to conflicting results the exact mechanism of this activation is still puzzling. Especially as the need for STING´s CTT to elicit a NF-κB activation seems conflicting. While it is shown that the NF-κB response can also be found for invertebrate STING without a CTT, studies in human suggest dependence on TBK-1 which binds at this region[244–246]. Another pathway omitting cGAS activation is STING-mediated NF-κB induction via TRAF6[247].

2.7.3 STING mediated autophagy

STING´s ability to bind CDN is an ancient feature, that can already be found in the homologue of the unicellular eukaryote choanoflagellate *M. brevicollis*, the closest living relative of the animal kingdom (no homologues in other kingdoms of life are found)[248]. Nevertheless, the ability of STING to elicit an IFN response only emerged with introduction of the IFN-system solely found in vertebrates. This fact suggests that STING also has a more ancient function[249].

Data from the sea anemone *Nematostella vectensis* indicate autophagy induction as the primordial function of STING. This pathway being independent from IKK or TBK-1 activation excludes the need of upstream cGAS. ERGIC residing STING serves as a membrane donor for LC3 lipidation, critical for autophagosome formation[250]. This autophagy inducing role of STING was also shown in mammals strengthening the immune response against some Gram-positive bacteria, upon poly(dA:dT) encounter or HSV-1 encounter[251–253].

2.7.4 STING induced lysosomal cell death

Common types of IFN or NF-κB related cell death have been already discussed (2.5) and display the main outcome of STING activation in most cell types. Primarily in human myeloid cells STING can also induce lytic cell death (LCD). In this pathway STING has to translocate to the lysosomal membrane, leading to its rupture and cell death upon the release of lysosomal hydrolases[254]. As STING degradation also mainly happens in the lysosome it has to be clarified how activation of LCD versus shut-down of STING activity is regulated[255].

2.8 Post-translational modifications of the cGAS-STING pathway

Although an interferon regulated feedback loop of cGAS activity exists, cGAS is constitutively expressed in most cell types[256]. Therefore, many post-translational modifications have been shown to mediate cGAS activity. Modifications at the active site directly inhibit its activity by preventing cGAMP production (hcGAS S305-P[257], cleavage at D319[258], L414-Ubi[259]). Active dimer formation is prevented by different modifications at the interaction interface (hcGAS L384-Ac/K394-Ac[259], K231-Sumo/K347-Sumo, K384-Sumo/K394-Sumo[260,261], K173-Ubi/K384-Ubi[262]). The N-terminal part is cleaved off by caspase 1 in apoptosis, rendering cGAS inactive[263]. Caspase 3 cleavage at D319 (for hcGAS) is inactivating in the same context[264]. The sumoylation site K479 (for hcGAS) is not located in any known functional motif[260]. Glutamylation at E302 (for hcGAS) blocks cGAS activity[265]. Just recently multiple phosphorylation sites at the N-terminus have been shown to prevent cGAS autoimmunity on chromatin during mitosis[266] (see Fig. 13 top).

As STING has multiple functions depending on localization and downstream interactors, all proteins involved are targets for modifications to orchestrate the signaling. The most obvious modification is the already described phosphorylation of S366 (for hSTING) by TBK-1 needed for IRF-3 signaling[267]. Another important site for STING regulation is K150 (for hSTING). Depending on the exact type of ubiquitylation status of this residue different effects are observed. In general, K48-linked ubiquitylation is associated with proteasomal degradation, whereas K63-linked ubiquitylation promotes signaling. Both types of modification can be found at K150, either regulating degradation[268,269] or promoting multimerization and therefore activity[270,271]. K63 and K27 ubiquitylation at K224 (for hSTING) promote interaction and phosphorylation by TBK-1[272,273]. Disulfide bridge formation at C148 (for hSTING)[243] and palmitoylation at C88 and C91 (for hSTING) facilitate multimerization [274] (see Fig. 13 bottom).
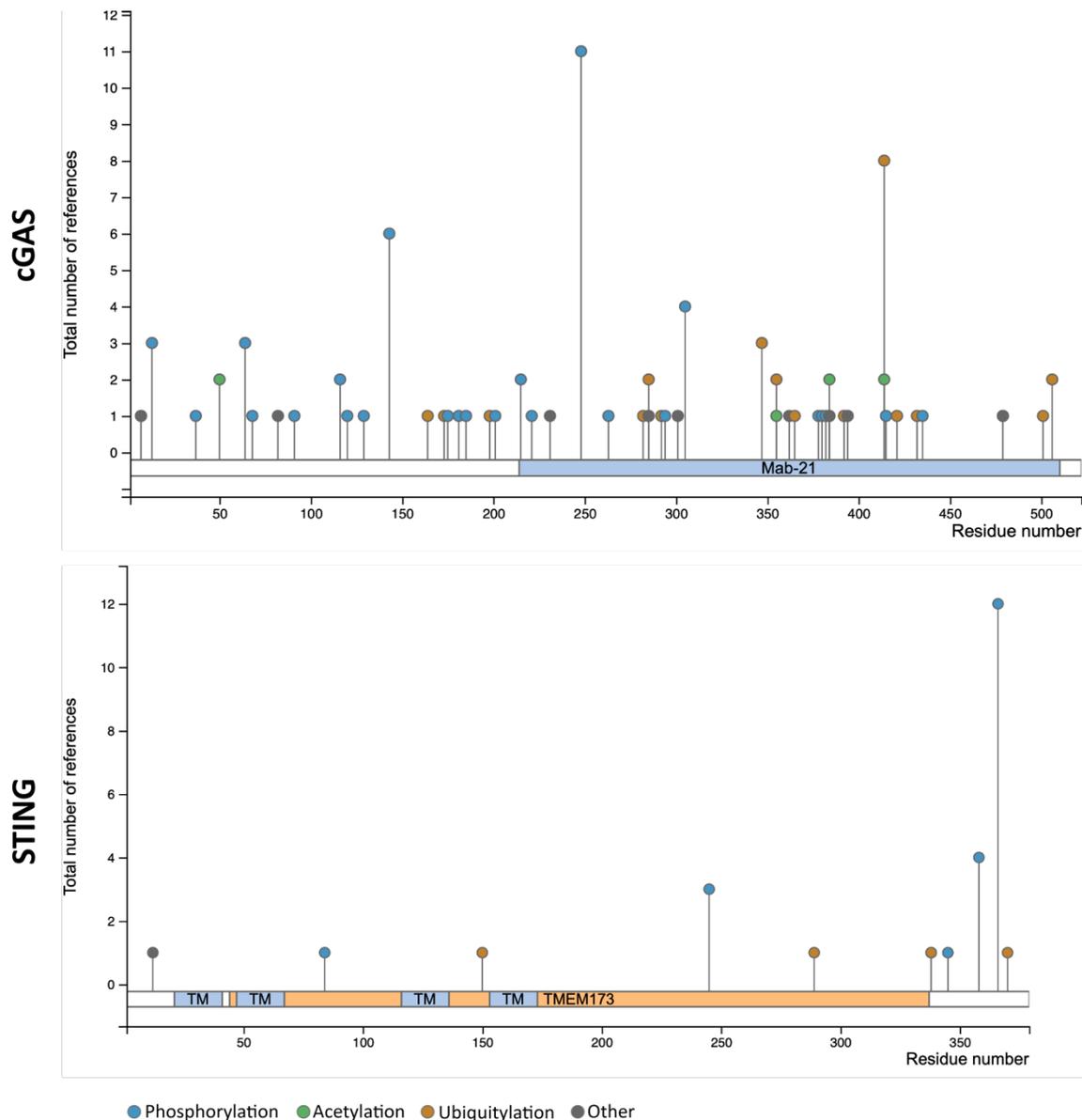
Figure 13: Modifications of cGAS and STING. All modifications of cGAS and STING according to PhosphoSite. Phosphorylation shown with blue dots, acetylation with green dots, ubiquitylation with orange dots and others including cleavage with grey dots. cGAS is heavily modified on all parts of the protein. Especially, the areas needed for cGAMP production (S305-P, cleavage at D319, L414-Ubi) or active dimer formation (L384-Ac/K394-Ac, K231-Sumo/K347-Sumo, K384-Sumo/K394-Sumo, K173-Ubi/K384-Ubi) are modified. Additionally, the NTD can be cleaved off or phosphorylated multiple times during mitosis. In comparison STING does not show that many modifications. Most obviously needed is the phosphorylation at S366 for further downstream signaling. Another noteworthy residue is K150, that can have multiple types of ubiquitylation leading to either impaired or amplified activity. | Phosphosite Plus V.6.5.9.3, www.phosphosite.org, accessed 14.05.2021

2.9 RIG-I the main cytosolic RNA sensor – structure and activation

2.9.1 Inactive open state

As cGAS was introduced as the main sensor for cytosolic dsDNA, RIG-I is the counterpart for dsRNAs. As already discussed before RIG-I does not recognize any kind of RNA nonspecifically (as cGAS does for dsDNA) but needs special recognition motifs. The RNA to be bound by RIG-I leading to activation needs a triphosphate or uncapped diphosphate at the 5´-end[129–131], base-pairing at the 5´-end[132] and the 5'-terminal nucleotide has to be unmethylated at its 2'-O[133].

RIG-I has a highly helical structure with two CARD (CARD1 and CARD2) domains of ~85 amino acids each at the N-terminus followed by a RecA-like helicase domain (Hel1 and Hel2), that is connected via a flexible linker (Hel1i). Hel1 and Hel2 form an ATP binding and hydrolyzing domain at their interface[275]. Another flexible linker (Hel2i) connects the C-terminal RNA binding domain (CTD) to the RecA domain[122]. The CTD harbors a Zn^{2+}-binding domain and forms a positively charged groove for RNA binding[276]. Additional RNA contact is made by the helicase domain. A specific feature of RIG-I compared to other helicases is the so called `pincer-helix´. This long α-helix emerges from Hel2 and bridges back to Hel1, with the interactions between the helix and Hel1 being conserved in different species supporting ATP binding[125] (see Fig. 14A).

Inactive RIG-I occurs in an autoinhibited state with the CARD2 domain sequestered by Hel2 forming a rigid head-to-tail unit[125,277]. The resulting steric constraints prevent MAVSs and polyubiquitin binding, which is needed for downstream signaling. Furthermore, this conformation shields the RNA binding site of Hel2 by CARD2. Due to this, only high affinity ligands bound to the CTD can disrupt this interaction making the full RNA binding site available[278]. Apo-RIG-I adopts an open conformation with Hel1 and Hel2 not in direct contact and Hel2 forming a rigid body with Hel2i (Hel2i-Hel2 module)[125]. As a result, the ATP-binding interface between these domains is not formed correctly disabling ATP hydrolysis when no RNA is bound[279]. The CTD is flexibly linked and available for binding to RNA[125] (see Fig. 14B).

2.9.2 Active closed state

The formation of the active conformation of RIG-I is a sequential process. First, binding of a high affinity RNA will lead to the formation of the functional ATP binding interface between Hel1 and Hel2[277]. Second, ATP binding will introduce a further conformational change compacting the SF2 domain and leading to a ring-like closure of the helicase and CTD domains around the RNA[126]. Hel1 will rotate 60° and Hel2i 23° relative to Hel2, creating a positively charged channel for RNA binding[125]. The pincer-helix adopts a V-shaped conformation protruding in a proline-rich loop, strengthening the closed ring structure. All four domains that participate in RNA binding cover a total area of around 1500 Å^2, binding to about 8-10 bps of RNA[125,126]. Most contacts are established with the sugar-phosphate backbone of the RNA, with a few exceptions in the CTD needed for specific ligand recognition. Nevertheless, most of these interacting residues are the same as in the CTD alone[280]. As the RNAs trajectory is slightly changed in the full-length structure to prevent steric clashes additional contacts between the CTD and RNA can be found. The 3´-terminus is exposed in a highly basic channel, whereas the 5´-end is capped by the CTD. F853 stacks with the terminal base-pair, explaining the preference for dsRNA. S854 binds to the sugar´s hydroxyl group, favoring RNA over DNA as ligand. H830 blocks binding of 2´-OH-Me RNA[280,281]. The only contact of the 3´ RNA strand with the CTD is made at its 5´-end with S906 (numbering for hRIG-I). Multiple motifs within Hel1 and Hel2 facilitate binding to the RNA. Hel1 motif 1a, 1b and 1c, Hel 2 motif IV, IVa, IVb and Hel2i interact with the 3´-RNA strand´s backbone not covered by the CTD. The residual non-capped region of the 5´-strand is recognized by Hel1 motif IIa and Hel2 motif Vc (see Fig. 14C).

The structural change induced by RNA binding separates the interaction of CARD2 with Hel2i, making them accessible for downstream signaling via MAVS[277]. The CARD domains of four RIG-I molecules oligomerize into a tetramer serving as nucleation point for further filament assembly. Addition of ubiquitin chains to the complex stabilizes the formed tetramer. From there on further MAVS can attach via homotypic CARD interactions creating larger filamentous assemblies[282,283].

Figure 14: RIG-I structural overview and rearrangement upon activation. A Domain architecture of RIG-I. RIG-I consits of two N-terminal CARDs (red), that are shielded in the apo-state, followed by the SF2-like helicase and the C-terminal domain for RNA binding (light blue). The SF-helicase consists of the Hel1 (grey), Hel2 (salmon) and the Hel2i insertion domain (yellow). The pincer-helix is denoted in dark-grey. **B** Structure of apo-RIG-I. Colouring of the domain is depicted in **A**. In the apo-state the CARDs (red) are sequestered by binding to Hel2i (yellow). The CTD (light blue) is flexibly attached and available for RNA binding.**C** When bound to RNA, RIG-I adopts a ring-like structure around the RNA molecule. The CARDs are not part of this ring closure and get released. No longer sequestered, the CARDs are availabe for downstream signaling. |PDB for B:4A2W+2QFB; C: 5EH3+2QFB

2.9.3 RIG-I ATPase domain – a closer look and implications for autoimmunity

Structurally RLRs belong to the class of SF2 helicases due to their helicase core. Strikingly the helicase domain of RLRs (and its relative Dicer[284]) have evolved to be activated by double-stranded substrates instead of single-stranded ones as for other members of the SF2-family like SWI/SNF-proteins[285,286]. This subgroup does not even unwind dsRNA robustly and is therefore referred to as `Duplex RNA-activated ATPases´ (DRHs)[284]. The specific ATP binding interface formed between Hel1 and Hel2 after activation is composed of different motifs also found in other SF2-type helicases. Motifs Q, I (a.k.a. `Walker A motif´), II (a.k.a. `Walker B motif´), III, V, and VI at the interface are involved in ATP binding and hydrolysis. Motif I contacts the ATPs phosphates in a typical SF2 helicase manner. Motif II coordinates Mg^{2+} for ATP positioning, with additional interactions of motifs Va and VI contacting the ribose and phosphates of ATP. Finally motif Q enables ATP specificity by selective contacts to adenine[281].

Mutations of the conserved lysine in the Walker A motif (motif I) or the conserved glutamate and aspartate residues in the Walker B motif (motif II) reduce signaling capabilities of RIG-I, whereas mutations in motif III interrupting ATP hydrolysis do not have an effect on signaling [287]. While ATP hydrolysis was implicated to have a role in filament formation[285], it is negligible for signaling[120]. This discrepancy was resolved when ATP was shown to be needed for proper discrimination between self and non-self RNAs[275,288]. ATP hydrolysis leads to faster dissociation of bound self-RNAs compared to the high affinity ligands described before[288,289]. Thereby mutations in the ATP binding and hydrolysis interface show a role in multiple autoimmune diseases, like Aicardi-Goutières syndrome (AGS)[290], Singleton-Merten syndrome (SMS) [291], systemic lupus erythematosus (SLE)[292] or type 1 diabetes[293].

Common residues mutated in the SF2 domain in SMS are C268F and E372A. E372A is located in motif II, where the replacement of glutamate with alanine disables the proper polarization of the water molecule needed for hydrolysis, thereby increasing autoimmune activity[288].This work will cover the influence of a common RIG-I SMS mutation C268F, where the structural basis of constitutive activation was lacking (see Fig. 15).

RIG-I helicase domain
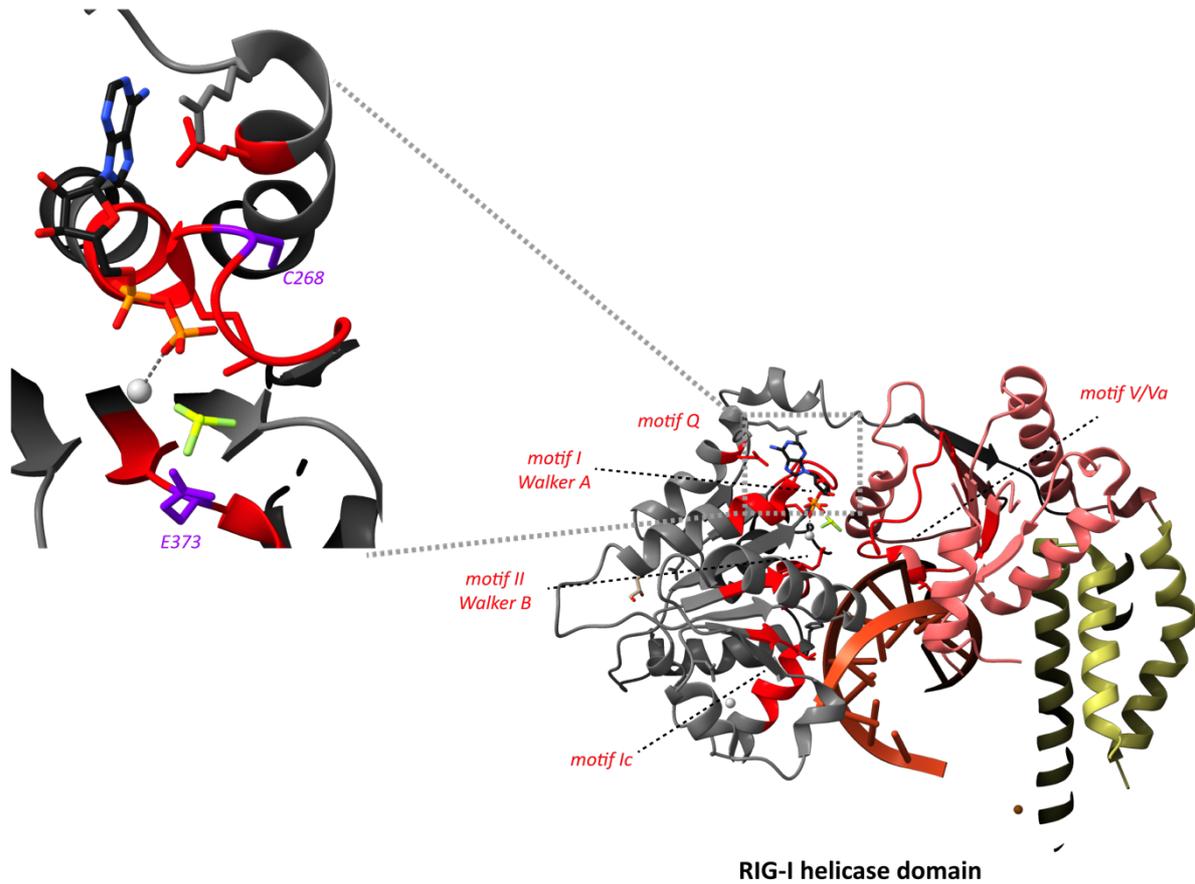
Figure 15: RIG-I SMS variants. RIG-I helicase domain shown in complex with ADP-BeF$_x$ with the relevant ATPase motifs. Magnification shows the pathogenic residues in purple. In case of the substitution E373I the needed polarization of a water molecule for ATP hydrolysis is not possible anymore, thereby shutting down the ATP-hydrolysis driven proof-reading mechanism. The constant activation of the C268F mutation is not explainable from this structure. |PDB:5E3H

3 Publications

3.1 Structural basis for sequestration and autoinhibition of cGAS by chromatin

Michalski, S.*, de Oliveira Mann*, C.C., Stafford, C., Witte, G., Bartho, J., Lammens, K., Hornung, V., Hopfner, K-P. *Nature* **587,** 678–682 (2020). https://doi.org/10.1038/s41586-020-2748-0

*: equal contribution

This work shows the first cryo-EM structure of the innate immune DNA sensor cGAS bound to the canonical nucleosome core particle. The resulting autoinhibition explains the mechanism of autoimmunity prevention against self-DNA in the nucleus. Despite the nuclear localization of cGAS being shown before, the prevention of constant autoimmunity by the sequence-unspecific dsDNA sensing of cGAS was puzzling. Our high-resolution cryo-EM structure revealed that cGAS is tightly anchored via a conserved arginine residue to the acidic patch formed at the interface of histone H2A and H2B. This sequestration prevents active cGAS-dimer formation by covering cGAS´ DNA binding site B, keeping it in an inactive monomeric conformation. Extensive mutagenic analysis of the interface of both interaction partners verified the interaction biochemically using electromobility shift assays and surface-plasmon-resonance. The observed affinity for cGAS towards nucleosomes was magnitudes higher compared to dsDNA. Additionally, the tethering of cGAS to nuclear chromatin was confirmed by nuclear fractionation assays in THP-1 cells. The inhibitory effect of cGAS-nucleosome tethering was shown for different nucleosome species and nucleosome arrays. Detecting the production of radioactive-labelled 2´-3´cGAMP by cGAS, nucleosomes without overhang-DNA (ONO) were identified as potent inhibitors of cGAS activity. This inhibition could not be resolved by the addition of competing dsDNA, but active dsDNA-bound cGAS was inhibited by NCP addition. NCP with extra-nucleosomal DNA can be activating, when the amount of cGAS exceeds the number of inhibitory acidic patches.

For longer nucleosome array constructs, we observed activity depending on the compaction state of the array. Open arrays with accessible acidic patches still inhibited cGAS activity, whereas compact arrays yielded cGAS activation. Again, the observed effects were confirmed by reverse observations using mutated interfaces of cGAS and the nucleosome´s acidic patch vice-versa. The higher activity of tethering defective mutants was confirmed in THP-1 cells measuring cytokine IP-10 levels after HT-DNA infection.

Author contribution:

The author cloned some of the initial cGAS constructs and purified all recombinant cGAS proteins used in the study. He established the production of differently functionalized nucleosomes and nucleosomal arrays to manifest cGAS-nucleosome binding. All nucleosome-related complexes in this study were purified, reconstituted and assembled by the author. He performed the electromobility shifts assays and some of the surface-plasmon-resonance measurements with the help of G.Witte. Besides, the author optimized plunging conditions with the help of C.C de Oliveira Mann and J. Bartho. Furthermore, the author collected the data together with C.C de Oliveira Mann, J. Bartho and K.Lammens. Finally, he participated in manuscript preparation.

Article

Structural basis for sequestration and autoinhibition of cGAS by chromatin

Sebastian Michalski[1,2,3], Carina C. de Oliveira Mann[1,2,3], Che A. Stafford[1,2], Gregor Witte[1,2], Joseph Bartho[1,2], Katja Lammens[1,2], Veit Hornung[1,2] & Karl-Peter Hopfner[1,2 ✉]

Cyclic GMP–AMP synthase (cGAS) is an innate immune sensor for cytosolic microbial DNA[1]. After binding DNA, cGAS synthesizes the messenger 2'3'-cyclic GMP–AMP (cGAMP)[2–4], which triggers cell-autonomous defence and the production of type I interferons and pro-inflammatory cytokines via the activation of STING[5]. In addition to responding to cytosolic microbial DNA, cGAS also recognizes mislocalized cytosolic self-DNA and has been implicated in autoimmunity and sterile inflammation[6,7]. Specificity towards pathogen- or damage-associated DNA was thought to be caused by cytosolic confinement. However, recent findings place cGAS robustly in the nucleus[8–10], where tight tethering of chromatin is important to prevent autoreactivity to self-DNA[8]. Here we show how cGAS is sequestered and inhibited by chromatin. We provide a cryo-electron microscopy structure of the cGAS catalytic domain bound to a nucleosome, which shows that cGAS does not interact with the nucleosomal DNA, but instead interacts with histone 2A–histone 2B, and is tightly anchored to the 'acidic patch'. The interaction buries the cGAS DNA-binding site B, and blocks the formation of active cGAS dimers. The acidic patch robustly outcompetes agonistic DNA for binding to cGAS, which suggests that nucleosome sequestration can efficiently inhibit cGAS, even when accessible DNA is nearby, such as in actively transcribed genomic regions. Our results show how nuclear cGAS is sequestered by chromatin and provides a mechanism for preventing autoreactivity to nuclear self-DNA.

cGAS possesses an approximately 360-amino-acid long cGAS/DncV-like (CD)[11] nucleotidyltransferase domain (cGAS[cat]) that has three DNA-binding sites (A, B and C) and a predicted unstructured 160-amino-acid long N-terminal domain. DNA binding to cGAS[cat] via sites A and B and subsequent dimerization induces a conformational change that structures the active site[3,12–14]. cGAS dimers are unstable but clustering along longer DNA and the formation of liquid–liquid phase-separated condensates involving site C concentrate and cooperatively stabilize active dimers[15–18].

In addition to sensing a wide range of cytosolic non-self DNA[19], cGAS detects self-DNA and chromatin from nuclear and mitochondrial origin that is misplaced in the cytosol or in cytosolic micronuclei, leading to sterile inflammation in cellular senescence or cancer[6,20]. Given this wide range of agonistic ligands, it was surprising that cGAS enters the nucleus during mitosis or after DNA damage[21,22], that it is constitutively localized in the nucleus[8–10] and that tethering to chromatin is even necessary to prevent autoreactivity[8]. cGAS is inhibited by nucleosome core particles (NCPs)[22] but activated by chromatin extracted from cells or synthetic nucleosome arrays[20].

Therefore, to understand how nuclear cGAS as well as cGAS–STING signalling influence inflammation and autoimmunity, we set out to determine how cGAS interacts with chromatin. Both human and mouse cGAS and cGAS[cat] bind NCPs with high affinity (Extended Data Fig. 1a). Surface plasmon resonance (SPR) experiments with mouse cGAS[cat] showed reproducible binding to immobilized nucleosomes at concentrations down to 1 nM, which indicates an affinity in the low nanomolar range (Fig. 1a). Human cGAS[cat] binds nucleosomes in the range of mouse GAS[cat], but owing to high bulk contribution the data did not allow an unbiased determination of kinetic parameters (Fig. 1a). The low nanomolar affinity of cGAS[cat] for nucleosomes is considerably stronger than the affinity for 20-bp double-stranded DNA (dsDNA) (in the micromolar range) and strongly agonistic 80-mer dsDNA (in the 100-nM range)[15].

To test how different nucleosomes can activate or inhibit cGAS, we monitored cGAMP production in vitro. Nucleosomes with 40-bp-long linker dsDNA ('40N40') as well as 147-bp dsDNA (the DNA used to assemble NCPs) activate human cGAS under conditions in which human cGAS is in excess (Fig. 1b). However, NCPs without linker DNA ('0N0') were unable to activate human cGAS (Fig. 1b), although cGAS binds under these conditions (Extended Data Fig. 1a). Here and in published work[20], the amount of cGAS exceeds that of nucleosomes, so it is conceivable that surplus human cGAS is activated by linker DNA. Thus, we performed activity assays with more stoichiometric amounts of human cGAS and NCPs, and challenged activation by the addition of 147-bp dsDNA (Fig. 1c). By contrast, we titrated 0N0 to cGAS pre-incubated with 147-bp dsDNA (Extended Data Fig. 1b). In both cases, 0N0 nucleosomes fully prevented the activation of cGAS by the 147-bp dsDNA. Together, cGAS[cat] binds NCPs with nanomolar affinity and this interaction

[1]Gene Center, Ludwig-Maximilians-Universität, Munich, Germany. [2]Department of Biochemistry, Ludwig-Maximilians-Universität, Munich, Germany. [3]These authors contributed equally: Sebastian Michalski, Carina C. de Oliveira Mann. ✉e-mail: hopfner@genzentrum.lmu.de
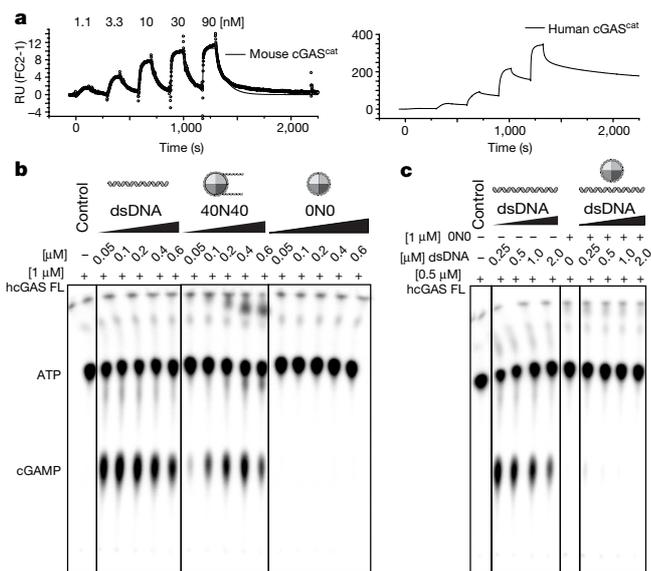
Fig. 1 | Nucleosomes inhibit cGAS activity in the presence of DNA. a, SPR analysis with nucleosomes immobilized via biotinylated DNA and mouse (left) and human (right) cGAS[cat] as analyte. Shown are reference cell (FC1, no nucleosomes) background-corrected data. Data from two subsequent injections (the second is shown in Extended Data Fig. 5a) of 1.1, 3.3, 10, 30 and 90 nM mouse cGAS[cat] were fitted using a 1:1 binding model (global fit, binding constant (K_a) = $2.13 \times 10^6 \, M^{-1} s^{-1}$, dissociation rate constant (k_d) = $0.01 \, s^{-1}$, RU max = 11.5). RU, resonance units. **b**, Human cGAS full-length (hcGAS FL) activity assays with radiolabelled [$\alpha^{32}P$]ATP. cGAS activity in the presence of 147-bp dsDNA, nucleosomes with 40-bp linker DNA (40N40) and nucleosomes with no linker DNA (0N0). Concentrations are indicated. Data are representative of two biological replicates. **c**, Human cGAS activity assay pre-incubated with 0N0 nucleosome followed by titration of dsDNA. Data are representative of two biological replicates.

sequesters cGAS in an inactive state under conditions in which there are more NCPs than cGAS molecules, a situation likely to be present in the nucleus.

To derive a structural basis of cGAS–nucleosome interactions, we used cryo-electron microscopy. Mixing approximately equal amounts of mouse cGAS[cat] with NCPs, we could obtain electron micrographs that showed homogenous, non-aggregated particles. However, human cGAS[cat] was more prone to aggregation. Two-dimensional (2D) classification and three-dimensional (3D) reconstruction resulted in a map with an overall resolution of 3.3 Å, showing a single cGAS bound to the NCP (Fig. 2a, Extended Data Figs. 2, 3). Focused refinement improved the resolution of the NCP and the histone-binding region of cGAS to 3.1 Å (Extended Data Figs. 2, 3), allowing building of a model with good overall stereochemistry (Extended Data Table 1).

Notably, cGAS[cat] binds to the folded parts of the histone 2A–histone 2B (H2A–H2B) dimer via an approximately 880 Å² large, bipartite interface (interfaces I and II), but has almost no contacts with nucleosomal DNA (Fig. 2a, Extended Data Fig. 4a). The interaction is mediated by DNA-binding site B of cGAS, which becomes blocked by the NCP, whereas sites A and C are partially (A) or fully (C) exposed. The interaction places cGAS to the lateral face of the nucleosome, between superhelix locations 5 and 6 of the nucleosome (Fig. 2b). Interface I is formed by 'tethering' loops 1 and 2, and the 'acidic patch' of the nucleosome, a prominent docking site for numerous nucleosome-binding factors[23] (Fig. 2c, d). Tethering loops 1 and 2 contain several evolutionary conserved residues that interact with the acidic patch, which indicates that histone binding is a conserved feature of cGAS (Fig. 2c, e). R241 of mouse cGAS binds to residues E61, D90 and E92 of H2A of the acidic

patch, with additional contacts formed between R222 of mouse cGAS and E61 of H2A, and between K240 of mouse cGAS and E64 of H2A. These residues have been identified as responsible for tight chromatin tethering in living cells, with mutations at R241 and R255 in mouse and human cGAS, respectively, having the most marked effects on gain of autoreactivity[8]. The strong effect of R241 is explained by the unusual interaction with several acidic patch carboxylates in the cavity (Fig. 2d, Extended Data Fig. 4a). R241 adopts the role of the characteristic 'arginine anchor' of many other nucleosome-binding factors[23] (Fig. 2d).

Interface II is formed between histone loops H2A-L2 and H2B-L1 and the core of DNA-binding site B of cGAS (Fig. 2c). In particular, central site B DNA-binding residues R341 and R342 directly bind H2A-L2 and H2B-L1. R341, for example, forms a salt bridge to D51 in H2B. The interactions at interface II lock cGAS into a rigid, homogenous complex that closely matches that of regulator of chromatin condensation 1 (RCC1) (Fig. 2d).

A notable observation is that DNA-binding site B of cGAS recognizes the H2A–H2B dimer via proteinaceous interactions (Fig. 3a). To validate these interactions, we mutated sites A, B and C, and the zinc-thumb cGAS dimerization element, and measured nucleosome interactions. The expression of the mutant proteins did not differ from that of wild-type protein, and, with the notable exception of the zinc-binding site mutants, have wild-type-like thermal stability, indicating correctly folded proteins (Extended Data Fig. 4b, c). Site B mutations in interface I strongly reduce binding to the NCP for arginine anchor mouse cGAS(R241E) and human cGAS(R255E) mutations, or robustly reduce binding to the NCP for mouse cGAS(R222E) and cGAS(K240E) mutations (Fig. 3b, Extended Data Fig. 4d), but leave DNA binding through site A intact (Extended Data Fig. 4e). Similarly, site B mutations in interface II (mouse cGAS(R341E), cGAS(R337E) and to a lesser extent cGAS(R342E)) robustly reduce NCP binding in vitro. Double mutants abolished binding. By contrast, mutations in sites A and the zinc-thumb in human cGAS do not reduce nucleosome binding, which shows that these sites are not involved in NCP binding (Extended Data Fig. 4f). Site C mutants in human cGAS[cat] lead to more defined species in electrophoretic mobility shift assays (EMSAs), rather than a high molecular mass species, which could be caused if site C of cGAS bound to one nucleosome binds to the DNA of a second nucleosome. To rule out that the loss of NCP binding is caused by loss of histone and not (hypothetical) DNA interactions (Extended Data Fig. 4e), we prepared variant NCPs that contain mutations in the acidic patch (0N0[apI]: H2A(E61A), H2A(E64A), H2A(D90A) and H2A(E92A)) and the H2A-L2–H2B-L1 regions (0N0[apII]: H2A(R71A), H2B(H49A) and H2B(D51A)), or both. Both 0N0[apI] and 0N0[apII] had strongly reduced interactions, whereas the double mutant had abolished interactions to mouse cGAS in electrophoretic mobility shift (Fig. 3c, Extended Data Fig. 4g) or SPR experiments (Extended Data Fig. 5a, b). Together, these data validate the structural observations that cGAS DNA-binding site B is a high-affinity binder of the H2A–H2B dimer.

Because NCPs are a dominating competitor to DNA in vitro, we investigated the activation of mouse cGAS by 147-bp DNA or plasmid DNA under conditions in which either cGAS site B or the nucleosome are mutated (Fig. 3d, Extended Data Fig. 5c–k). Mutations in the tethering loop (mouse cGAS mutants R222E, K240E and R241E) led to DNA-dependent activation of cGAS both in the presence and absence of NCPs (Fig. 3d). A similar loss of inhibition by NCPs is observed for human cGAS(R236E) (corresponding to mouse cGAS(R222E)) (Extended Data Fig. 5c, d). Mouse cGAS(G316A/S317A), containing more conservative mutations in interface II, lead to active cGAS that is less inhibited by NCPs (Fig. 3d). By contrast, mutations R341E, R337E and R342E in mouse cGAS resulted in inactive or strongly compromised mouse cGAS irrespective of the presence and absence of NCP, consistent with their important role in dsDNA binding in the active cGAS dimer state[13,14] (Extended Data Fig. 4e, f). None of the tethering mutants displayed activity without DNA (Extended Data Fig. 5g, h), and mutating site A abolishes activity in the tethering mutants as well, showing

Article
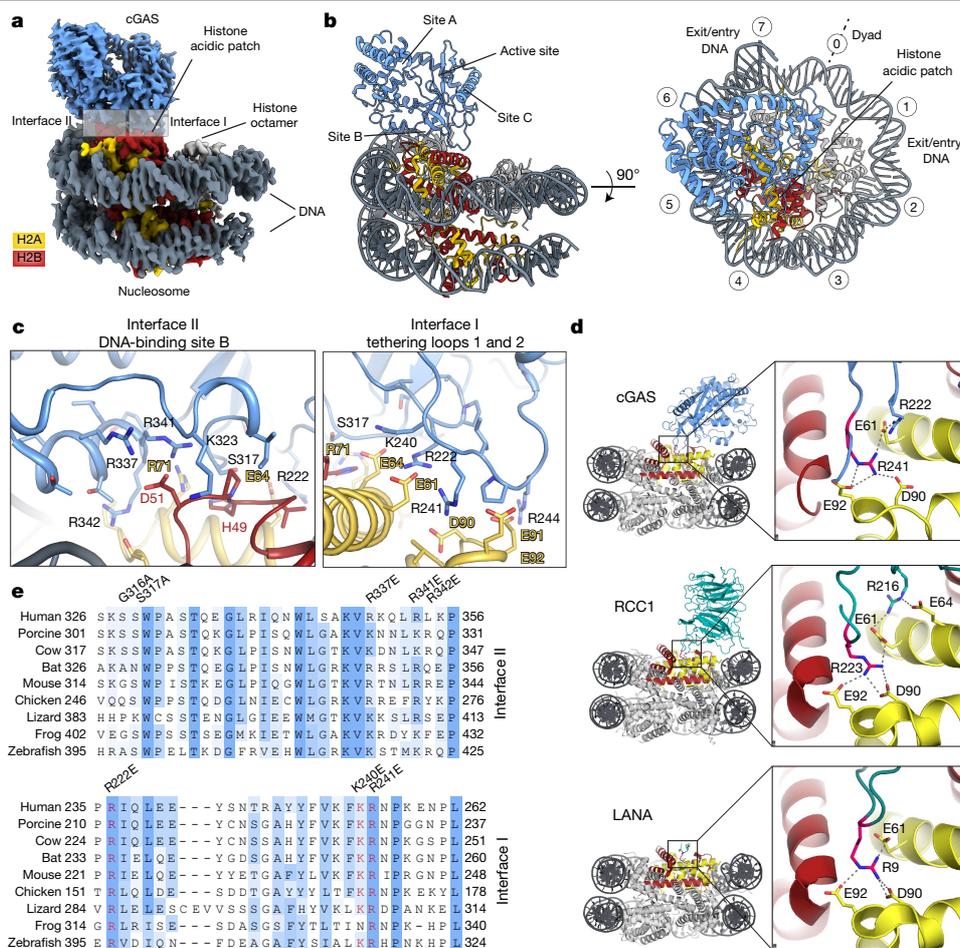
Fig. 2 | Structure of the cGAS catalytic domain–nucleosome complex.
a, A 3.3-Å resolution cryo-EM map showing density for mouse cGAS (light blue) bound to the acidic patch of the nucleosome (dark grey, DNA; yellow, H2A; red, H2B; light grey, H3.3 and H4). **b**, Gyre and disc views of the cGAS–nucleosome structure. cGAS catalytic domain shows no interactions with nucleosomal DNA and interacts with the H2A–H2B acidic patch via a conserved loop. **c**, Close-up view for interface I and interface II. Interacting amino acid residues are depicted (blue, cGAS; yellow, H2A; red, H2B). cGAS conserved loop and key interacting residues R222 and R241 are depicted. **d**, Structures of cGAS–nucleosome, RCC1–nucleosome (PDB code 3MVD) and Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen (LANA)–nucleosome (PDB code 1ZLA) complexes. Close-up views depicting arginine anchors from cGAS R241, RCC1 R223 and LANA R9 and key interacting residues from the nucleosome acidic patch. **e**. Alignment of representative cGAS proteins from different vertebrate species. Darker shadings indicate higher physiochemical conservation according to BLOSUM62 score using Jalview. Amino acids involved in cGAS–nucleosome interactions mutated in this study are depicted. Amino acids marked in red are highly conserved and required for cGAS binding to the nucleosome.

that autoreactivity requires DNA interactions at site A (Extended Data Fig. 5i). Finally, we find that 0N0[ap1] has a markedly reduced ability to inhibit human cGAS in the presence of dsDNA (Extended Data Fig. 5j), and that human cGAS(R236E) is much more robustly activated by 40N40 than wild-type cGAS (Extended Data Fig. 5k). Fluorescence polarization anisotropy studies showed dsDNA binding to human cGAS[cat] in the absence of 0N0 but not in the presence of 0N0 (Extended Data Fig. 5l). Site A mutants K407E and K411E rendered human cGAS inactive in vitro and in living cells[12] but had little effect in these titration assays. Together, these data validate that nucleosome tethering blocks DNA binding to site B (Extended Data Fig. 4d).

Because nucleosomal arrays have been shown to activate cGAS in vitro and chromatin in cytosolic micronuclei is linked to cGAS activation[20], we tested cGAS activation using an array with 12 nucleosomes positioned by Widom-601 sequences spaced by 53-bp linkers (Fig. 3e, Extended Data Fig. 5m). Using the conditions previously described[20], we find robust activation of cGAS by DNA and the 12-mer array. However, these conditions favour highly compact chromatin through interactions of the H4 tail with acidic patches[24], which could prevent cGAS sequestration and allow its activation by linker DNA. Under conditions that reduce array compaction, we do not observe cGAS activation at low cGAS concentrations, whereas free DNA is still activating. Higher cGAS concentrations lead to activation by the arrays, but this activation can be inhibited by adding free NCPs for wild-type but not R241E mutant mouse cGAS. Therefore, compact arrays activate cGAS, at least in the absence of other chromatin factors, but that activation is inhibited by nearby accessible acidic patch.

To complement our biochemistry data, we studied human THP-1 cells, in which cGAS stimulation by long herring testis DNA (HT-DNA) leads to a potent interferon-stimulated gene (ISG) response. Wild-type, but not *CGAS*[−/−], THP-1 cells stimulated with HT-DNA exhibited potent upregulation of the two well-known ISGs viperin and OAS1 (2'-5'-oligoadenylate synthetase 1) (Fig. 3f, Extended Data Fig. 6a). We next complemented *CGAS*[−/−] THP-1 cells to express wild-type or mutated human cGAS variants under a doxycycline-inducible promotor and monitored the expression of viperin and OAS1 in dependence of DNA stimulation.
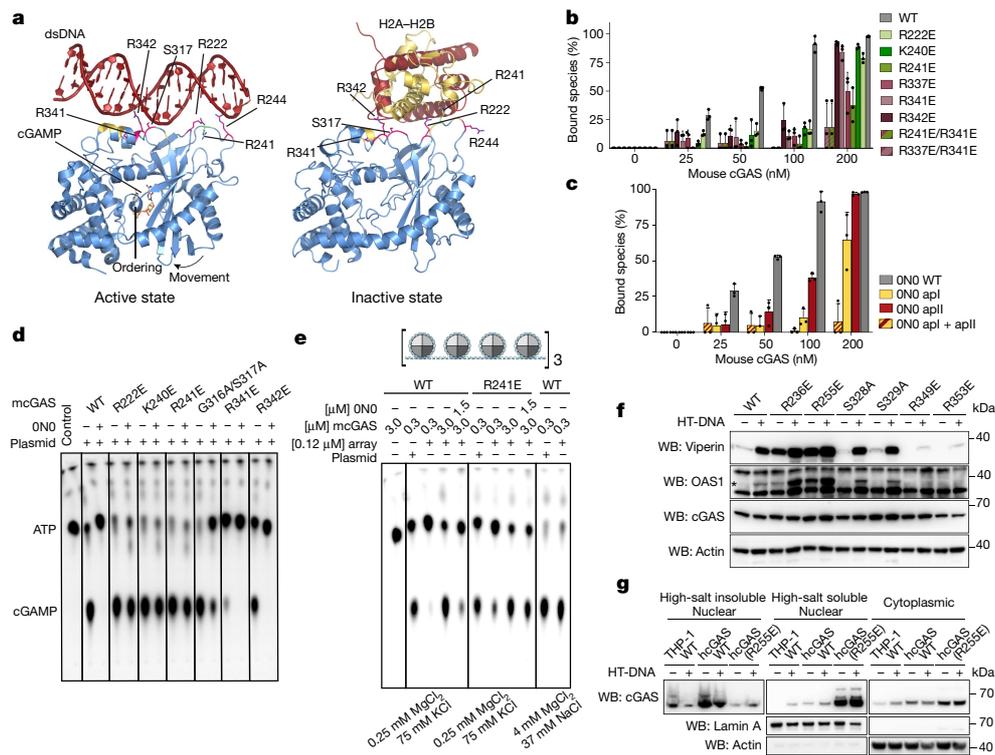
Fig. 3 | cGAS is sequestered by the acidic patch of the nucleosome. a, Left, cGAS bound to DNA and cGAMP (PDB code 4LEZ) in an active state. Site B residues are highlighted (yellow, DNA binding; magenta, DNA and histone binding; green, acidic patch anchor R241). Activating domain movement and activation loop ordering is indicated. Right, cGAS bound to histones H2A–H2B (yellow, red) in an inactive state. Site B residues are highlighted (yellow, histone binding; magenta, DNA and histone binding; green, acidic patch anchor R241). **b**, EMSA of wild-type (WT) and mutant mouse cGAS binding to nucleosomes. Data are mean and s.d., $n = 3$ biological replicates. Samples are derived from the same experiment and gels were processed in parallel. **c**, EMSA of mouse cGAS and nucleosome acidic patch mutants I, II and I + II. Data are mean and s.d. $n = 3$ biological replicates. Samples are derived from the same experiment and gels were processed in parallel. **d**, Activation of mouse cGAS (mcGAS) mutations affecting cGAS–nucleosome interactions by plasmid DNA in the presence and absence of 0N0 nucleosomes. cGAMP production was assayed by thin-layer chromatography. Data are representative of two biological replicates. **e**, Activation of wild-type and R241E mutant mouse cGAS by 12× nucleosomal arrays under low- or high-ionic strength conditions. Data are representative of two biological replicates. **f**, Phorbol 12-myristate 13-acetate (PMA)-differentiated human *CGAS*$^{-/-}$ knockout THP-1 cells reconstituted with wild-type human cGAS or indicated mutants were treated with doxycycline (1 μg ml^{-1}) overnight and either left untreated or treated with HT-DNA (975 ng per 550 μl) for 8 h. Cell lysates were separated on SDS–PAGE gels, analysed by western blot (WB) and probed with the indicated antibodies. Data are representative of three biological replicates. **g**, Nuclear and cytosolic fractionation of wild-type and *CGAS*$^{-/-}$ knockout human THP-1 cells reconstituted with wild-type or R255E mutant human cGAS. High-salt insoluble nuclear (tight chromatin tethering), high-salt soluble nuclear and cytoplasmic fractions were separated on SDS–PAGE gels, analysed by western blot and probed with the indicated antibodies. Data are representative of two biological replicates.

CGAS$^{-/-}$ THP-1 cells reconstituted with wild-type cGAS upregulated viperin and OAS1 after HT-DNA stimulation. Moreover, consistent with previous work[8] and our in vitro analysis, human cGAS mutants R236E and R255E (equivalent to mouse R222E and R241E), led to robust production of viperin and OAS1 even in the absence of HT-DNA, and further increased activity above the wild-type cGAS response in the presence of HT-DNA (Fig. 3f). By contrast, R349E and R353E (equivalent to mouse R337E and R341E) led to inactive human cGAS irrespective of HT-DNA transfection. The more conservative mutations S328A and S329A did not result in a markedly increased ISG response compared to wild-type cGAS; presumably the strength of nucleosome interactions remained sufficient. Analogous data were obtained when we monitored the production of IP-10, an antiviral chemokine (Extended Data Fig. 6b).

We interrogated whether acidic patch interactions regulate human cGAS localization by cell fractionation experiments under different conditions (Fig. 3g). Using a gentle lysis protocol, we separated the cytoplasmic from the nuclear fraction. The nuclear fraction was also separated into a 420-mM KCl soluble and insoluble fraction. In agreement with previous reports[8], we observed endogenous and overexpressed wild-type cGAS in the high-salt insoluble fraction. HT-DNA

treatment reduced endogenous human cGAS in this fraction, coupled with increased levels in the high-salt soluble nuclear fraction and the cytoplasmic fraction. cGAS(R255E) displayed increased cytosolic levels, although most was still nuclear. However, unlike endogenous and overexpressed wild-type human cGAS, the R255E mutant was predominantly found in the high-salt soluble nuclear fraction, rather than in the high-salt insoluble fraction. Thus, nuclear and cytosolic distribution is not, by and large, regulated by acidic patch tethering, but by other mechanisms that need to be addressed in future studies. Once in the nucleus, tethering of cGAS to nucleosomes and prevention of autoreactivity is governed by the acidic patch anchor R255.

Our data reveal how cGAS is tethered to nucleosomes and how this sequestration keeps cGAS inactive (Fig. 4). The interaction is similar to the binding of the mitochondrial dynamics protein of 49 kDa (MID49) to dynamin related protein 1 (DRP1)[25] (Extended Data Fig. 7), suggesting that the tethering loops are more widely used macromolecular anchors of CD nucleotidyltransferases[11]. However, it is possible that other elements besides acidic patch binding affect the subcellular distribution of cGAS because the N-terminal domain has been shown to preferentially localize cGAS to centromeric satellite DNA[10].

Article
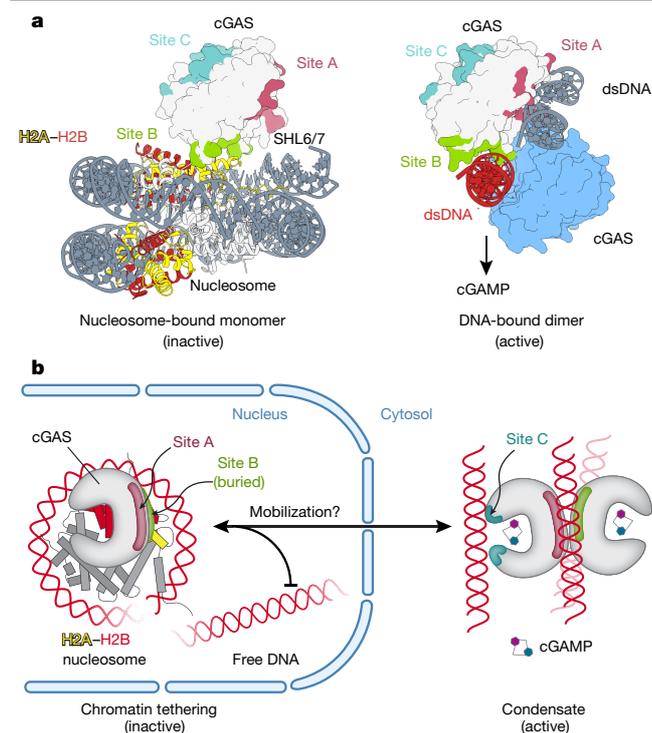
Fig. 4 | Nucleosome sequestration sterically blocks cGAS activation. a, Left, structure of cGAS bound to nucleosome histones H2A–H2B via site B (green) interactions. DNA-binding site A (red) and site C (blue) are depicted. Right, structure of cGAS–DNA complex (PDB code 4LEZ) bound via DNA-binding site A (red) and site B (green). dsDNA (red) binding to site B of cGAS is displaced by the nucleosome. cGAS site B is shared for DNA and nucleosome interactions. **b**, Schematics of cGAS in cytoplasmic active state bound to DNA and in nuclear inactive state bound to histones H2A–H2B of the nucleosome core particle.

Comparison of the structure of monomeric cGAS bound to the nucleosome with that of active cGAS dimer bound to DNA[13,14] reveals two elements of inactivation. First, cGAS binding to the H2A–H2B dimer blocks DNA access to site B. Second, nucleosomal DNA at superhelix location 6/7 occupies the region of the second cGAS molecule in the cGAS dimer. Thus, nucleosome binding also sterically blocks cGAS dimer formation even when DNA is bound to sites A and C. Because DNA binding to both sites A and B and subsequent cGAS dimerization is required to induce a conformation in which active site elements are properly ordered[12–14], our data explain how nucleosomes keep cGAS in an inactive state that is robust enough to prevent activation even in the case of nearby accessible DNA.

The binding of cGAS to the nucleosome acidic patch could be sensitive to the chromatin state and the availability of accessible nucleosomes, and perhaps has evolved to sequester cGAS away from free DNA at sites of transcription or accessible linker DNA. Although nuclear chromatin is inhibitory at physiological cGAS levels[8], chromatin in cytosolic micronuclei, which activate cGAS[20,26,27], might simply not have enough accessible nucleosomes to fully sequester the cytosolic pool of cGAS upon membrane rupture. In future studies, it will be important to investigate the mechanisms by which cytosolic and nuclear distribution of cGAS is regulated, for example, during infections. For instance, activation of THP-1 cells[8] and HT-DNA stimulation (Fig. 3g) lead to partial redistribution of nuclear cGAS to the cytosol. Post-translational modifications of either cGAS or histones, or through cofactors, could be important regulators, and a role of histone tails needs to be clarified. Alternatively, ubiquitination of histones or of human cGAS[28] could

modulate cGAS and nucleosome interactions. Notably, a portion of nuclear cGAS appears to be modified and this modification disappears upon DNA stimulation (Fig. 3g).

In summary, we demonstrate the tethering mechanism and competitive inhibition of cGAS by nucleosomes and provide a framework for the interaction of cGAS with chromatin that explains how nuclear chromatin prevents autoreactivity against self-DNA.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2748-0.

1. 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).
2. Ablasser, A. et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–384 (2013).
3. 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).
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. Hopfner, K. P. & Hornung, V. Molecular mechanisms and cellular functions of cGAS-STING signalling. *Nat. Rev. Mol. Cell Biol.* **21**, 501–521 (2020).
6. Ablasser, A. & Chen, Z. J. cGAS in action: Expanding roles in immunity and inflammation. *Science* **363**, eaat8657 (2019).
7. Motwani, M., Pesiridis, S. & Fitzgerald, K. A. DNA sensing by the cGAS-STING pathway in health and disease. *Nat. Rev. Genet.* **20**, 657–674 (2019).
8. Volkman, H. E., Cambier, S., Gray, E. E. & Stetson, D. B. Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *eLife* **8**, e47491 (2019).
9. Jiang, H. et al. Chromatin-bound cGAS is an inhibitor of DNA repair and hence accelerates genome destabilization and cell death. *EMBO J.* **38**, e102718 (2019).
10. Gentili, M. et al. The N-terminal domain of cGAS determines preferential association with centromeric dna and innate immune activation in the nucleus. *Cell Rep.* **26**, 2377–2393 (2019).
11. Kranzusch, P. J. cGAS and CD-NTase enzymes: structure, mechanism, and evolution. *Curr. Opin. Struct. Biol.* **59**, 178–187 (2019).
12. Civril, F. et al. Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–337 (2013).
13. Li, X. et al. Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019–1031 (2013).
14. 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 Rep.* **6**, 421–430 (2014).
15. Andreeva, L. et al. cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein-DNA ladders. *Nature* **549**, 394–398 (2017).
16. Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* **361**, 704–709 (2018).
17. Xie, W. et al. Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc. Natl Acad. Sci. USA* **116**, 11946–11955 (2019).
18. Hooy, R. M. & Sohn, J. The allosteric activation of cGAS underpins its dynamic signaling landscape. *eLife* **7**, e39984 (2018).
19. 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).
20. Mackenzie, K. J. et al. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).
21. Liu, H. et al. Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* **563**, 131–136 (2018).
22. Zierhut, C. et al. The cytoplasmic DNA sensor cGAS promotes mitotic cell death. *Cell* **178**, 302–315 (2019).
23. McGinty, R. K. & Tan, S. Recognition of the nucleosome by chromatin factors and enzymes. *Curr. Opin. Struct. Biol.* **37**, 54–61 (2016).
24. Arya, G. & Schlick, T. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J. Phys. Chem. A* **113**, 4045–4059 (2009).
25. Kalia, R. et al. Structural basis of mitochondrial receptor binding and constriction by DRP1. *Nature* **558**, 401–405 (2018).
26. Harding, S. M. et al. Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**, 466–470 (2017).
27. Dou, Z. et al. Cytoplasmic chromatin triggers inflammation in senescence and cancer. *Nature* **550**, 402–406 (2017).
28. Seo, G. J. et al. TRIM56-mediated monoubiquitination of cGAS for cytosolic DNA sensing. *Nat. Commun.* **9**, 613 (2018).

Methods

Data reporting
No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Protein production and purification
Human cGAS full-length and cGAS catalytic domain (155–522) genes were purchased from Genescript codon optimized for expression in *Escherichia coli* and cloned into a modified pET28 vector with an N-terminal His-SUMO1-tag. Mouse cGAS full-length and the cGAS catalytic domain (141–507) were cloned in a modified pET21 with an N-terminal His-MBP-tag as previously described[12]. *E. coli* BL21 Rosetta (DE3) cells were cultured in 2× 3-l TB medium until reaching an optical density at 600 nm (OD$_{600}$) of 1.0–2.0, and protein expression was induced at 18 °C with 0.4 mM IPTG for 16 h. Cell pellets were resuspended in lysis buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM β-mercaptoethanol) and disrupted by sonication. Recombinant cGAS proteins and mutants were purified over Ni-NTA affinity chromatography and the His-SUMO1 and His-MBP tags were subsequently removed by SENP2 protease or TEV protease cleavage respectively at 4 °C, overnight dialysis in buffer A (20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM DTT). Proteins were further purified over HiTrap heparin column (GE Healthcare) and eluted with a linear gradient of buffer B (20 mM HEPES pH 7.5, 1 M NaCl, 1 mM DTT). For a final purification step, fractions containing cGAS protein were loaded on a Superdex 16/60 S75 size exclusion chromatography column (GE Healthcare) in 20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM TCEP. Protein samples were concentrated up to 5–10 mg ml^{-1} and 10% glycerol was added before flash freezing in liquid nitrogen and storage at −80 °C.

Purification of histone acidic patch mutants
Canonical human histones H2A (00120), H2B (00101), H3.3 (00102) and H4 (00103) were purchased from The Histone Source–Protein Expression and Purification (PEP) Facility. Human histone H2A (E61A, E64A, D90A and E92A), H2A (R71A) and H2B (H49A, D51A) acidic patch mutants were cloned into pET21 vector and purified as previously described[29]. In brief, histone acidic patch mutants were expressed in *E. coli* BL 21 (DE3) cells (Novagen) for 3 h at 37 °C after induction at OD$_{600}$ 0.6. Pelleted cells were disrupted by resuspension in wash buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT) supplemented with 1 mg ml^{-1} lysozyme (Carl Roth, 8259), 1× protease inhibitor and 250 U benzonase (Merck Millipore, E1014) via three rounds of sonication. Inclusion bodies were washed two times with wash buffer supplemented with 1% Triton X-100 followed by two washes with standard wash buffer. Washed inclusion bodies were incubated with 1 ml DMSO for 30 min at room temperature followed by homogenization in resuspension buffer (7 M GdmCl, 20 mM sodium acetate pH 5.2, 1 mM EDTA) and 1 h incubation at room temperature. The supernatant was dialysed two times for 1.5 h against SAU 50 buffer (8 M urea; 20 mM sodium acetate, pH 5.2, 50 mM NaCl; 1 mM EDTA, 10 mM lysine). As first step of purification histones were separated by cation exchange chromatography (GE Healthcare HiTrap S HP) by applying a salt gradient. Histone-containing fractions were dialysed three times against refolding buffer (15 mM Tris-HCl, pH 8.0) of which one step needs to be over night to ensure proper refolding. Correct folded histones were finally separated by anion exchange chromatography (GE Healthcare HiTrap Q HP) using a salt gradient. Histone containing fractions were shock frozen in liquid nitrogen and lyophilized.

Preparation of mononucleosomes
Histone octamers were assembled by resuspending and mixing lyophilized human H2A (The Histone Source; 00120); H2B (The Histone Source; 00101), H3.3 (The Histone Source; 00102) and H4 (The Histone Source; 00103) or corresponding H2A–H2B acidic patch mutants in 1.2:1.2:1:1 molar ratio in octamer unfolding buffer (25 mM Tris-HCl, pH 7.5; 7 M GdmCl; 1 mM DTT). Refolding into correct octamer was achieved by dialyzing against octamer refolding buffer (25 mM Tris-HCl, pH 7.5, 2 M NaCl, 1 mM DTT). Assembled octamer was purified by size exclusion chromatography. Octamer containing fractions were concentrated and stored at -20 °C in 50% glycerol.

Widom 601 DNA sequence was used for the PCR amplification of nucleosomal DNA. Depending on the needed functionality of the nucleosome to be made, non-fluorescent primers, 5′-FAM-labelled primers or 5′-biotin-labelled primers were used (Metabion, Martinsried). The large-scale PCR amplified product was purified using salt-gradient anion-exchange chromatography (GE Healthcare HiTrap DEAE FF). Nucleosomal DNA containing fractions were dialysed overnight against water to remove salt and concentrated under vacuum.

For reconstitution of functional nucleosomes canonical human octamer or acidic patch mutants were mixed with non-fluorescent, double 6-FAM-labelled or biotin-labelled DNA in a 1:1.1 ratio in high-salt buffer (25 mM Tris-HCl, pH 7.5, 2 M NaCl, 0.25 mM DTT). For nucleosome assembly tenfold volume of low-salt buffer (25 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.25 mM DTT) was added over 24 h at 4 °C. Final assembly took place by dialyzing against fresh low-salt buffer for at least 3 h. Reconstituted nucleosomes were purified by salt-gradient separation via anion-exchange chromatography (GE Healthcare Source Q). Nucleosome containing fractions, were pooled, dialysed against low-salt buffer overnight and stored at 4 °C.

Preparation of nucleosome arrays
For preparation of nucleosome arrays a pUC18-12x601-200 plasmid originating from the Rhodes lab was used as DNA template[30]. Large amounts of plasmid were prepared using a Macherey-Nagel PC-10000 Giga kit (740548), following the manufacturer's instructions. The desired 12× positioning sequence was cut out by an EcoRI (NEB, R3101T) and HindIII (NEB, R3104T) restriction digest. The rest of the vector was fragmented using restriction enzymes DraI (NEB, R0129L) and HaeII (NEB, R0107L). The 12× nucleosomal array DNA was isolated by size-exclusion chromatography using a Sephacryl-S500 16/60 column (GE Healthcare). Reconstitution into nucleosome arrays was performed by mixing array DNA at a final concentration of 0.3 μg μl^{-1} with 1.3 times molar excess histone octamer (in proportion to available binding sites) and 0.1 mg ml^{-1} BSA (Carl Roth, 8076) in the previously described high-salt buffer[31]. Dialysis into low-salt buffer was performed as previously described in this study for mononucleosomes. Correct nucleosomal array reconstitution was assessed by performing a MNase digest. Here, 1 μg of array was digested with 0.005 U MNase (Sigma-Aldrich, 3755) in buffer containing 1.5 μM CaCl$_2$. The reaction was stopped using STOP-buffer (50 mM Tris pH 7.5, 4% SDS, 100 mM EDTA, pH 8.0) at different incubation time points. Afterward samples were incubated for 1 h at 37 °C with Proteinase K (Thermo Fisher Scientific, EO0491) and digested DNA was isolated using the NEB Monarch PCR & DNA Cleanup Kit (NEB, T1030S). The isolated DNA fragments were separated using 1.5% (Biozym, 840004) agarose gel electrophoresis with 1×TAE as running buffer (100 V, 2.5 h) and stained with GelRed (Biotium, 41003). Gels were imaged using a GE Healthcare Typhoon FLA9000 imager.

EMSAs
For EMSAs of cGAS with nucleosomes serial dilutions of cGAS protein (different organisms or mutants) were mixed with double fluorescently labelled canonical or acidic patch mutated nucleosomes. In brief, cGAS protein 0 to 8 times molar excess in twofold dilution, was added to provided 25 nM nucleosome EMSA master mix (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP) and incubated for 30 min on ice. The formed complex was separated on 7% TA-Nu Page gels (Invitrogen, EA0358)

Article

using 1× TG-buffer (25 mM Tris, 192 mM glycine, pH 8.3) under native conditions at 4 °C (100 V const., 105 min). For testing binding to DNA, 147-bp long nucleosomal double fluorescently labelled dsDNA was used. Here, mouse cGAS protein 0 to 16 times molar excess in two-fold dilution, was added to provided 25 nM dsDNA master mix (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP) and incubated for 30 min on ice. Separation of the complex was performed using a 1% agarose gel using 40 mM Tris pH 9.2 as running buffer (4 °C, 100 V const., 90 min). Gels were imaged using a GE Healthcare Typhoon FLA9000 imager. Quantification of the band intensities was performed using ImageJ[32] gel analysis by measuring the decline in intensity of free nucleosomes with increasing cGAS concentrations.

SPR

SPR experiments were performed using a Biacore X100+ (GE Healthcare Biacore) instrument with HSB-P+ and 1 mM TCEP (Cytiva, BR100827) at a flow rate of 10 μl min⁻¹. A CM5 sensor chip was coupled with neutravidin (Thermo Fisher Scientific, 31055) via amino reactive chemistry (EDC/NHS) to both flow cells (FC1, FC2). Nucleosomes with biotinylated DNA were coupled to FC2 (200 RU), FC1 served as a reference cell to correct for unspecific binding to the surface of the chip. We observed some background binding but assume that this is identical for FC2 and FC1. All sensorgrams show reference-subtracted data—that is, response units of FC2−FC1. Data were evaluated using a 1:1 binding model in the Biacore X100 Evaluation Software.

Thermal shift assay

Thermal stability of different protein mutants was examined using Nanotemper Tycho NT.6 by performing thermal shift assays. cGAS proteins (2 mg ml⁻¹) incubated in 20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM TCEP were measured. Change in intrinsic fluorescence of aromatic amino acids was measured at wavelengths of 330 nm and 350 nm from 35 °C to 95 °C (rate 3 °C min⁻¹). The deflection point of the curve and first derivative was calculated by the Tycho NT.6 software (v.1.3.1.868).

Fluorescence anisotropy measurement

Fluorescence anisotropy measurements were carried out using a Tecan Infinite M1000 plate reader. For testing of cGAS binding to small DNA serial dilutions of hcGAS^{cat} or hcGAS^{cat} site A mutant were mixed with 100 nM (final concentration) of 5′-FAM-labelled 20-bp dsDNA in anisotropy buffer (50 mM NaCl, 25 mM HEPES pH 7.5, 1 mM TCEP) in GRE96fb_chimney plates (Greiner BioOne) and incubated for 30 min at room temperature. The maximum amount of cGAS used was limited by the amount of recombinantly produced nucleosome. To test the binding of small DNA to pre-formed cGAS-0N0 complexes the same concentrations of hcGAS^{cat} or hcGAS^{cat} site A mutant as described before were incubated with 0.5 times molar ratio 0N0 for 20 min at room temperature before addition of the 5′-FAM-labelled 20-bp dsDNA. Measurements were performed using the Tecan i-control software in fluorescence polarization mode using the following settings: λ_{ex} = 470 nm (bandwidth = 5 nm), $\lambda_{emission}$ = 520 nm (bandwidth = 10 nm), gain optimal, number of flashes = 10, settle time = 0, z-pos = 20,000, G-factor = 1. Values for anisotropy were calculated by Tecan i-control software and corrected for blank and DNA-only control. Owing to the limited amount of 0N0 the curves were not fitted, as saturation is not reached, but show only qualitative comparison.

In vitro cGAS 2′3′ cGAMP synthesis assay

Radiolabelled cGAS activity assays were performed as previously described[12]. Unless otherwise stated, pre-incubated samples containing different amounts of cGAS, DNA and/or nucleosomes were mixed with 25 μM ATP, 25 μM GTP and trace amounts of [α³²P]ATP (Hartman Analytic, SRP-207) in reaction buffer (50 mM HEPES pH 7.5, 5 mM MgCl₂, 100 mM NaCl, 1 mM TCEP). Reactions were incubated for 45 min at 37 °C and reaction products were spotted on thin-layer chromatography (PEI-Cellulose F plates; Merck, Z122882). 1 M (NH₄)₂SO₄/1.5 M KH₂PO₄ pH 3.8 was used as running buffer and TLC plates were analysed by phosphor imaging (Typhoon FLA 9000, GE Healthcare). In Fig. 1b, 1 μM human cGAS full-length protein was incubated with 0.05, 0.1, 0.2 0.4 and 0.6 μM dsDNA, 40N40 or 0N0 nucleosomes. In Fig. 1c, 0.5 μM human cGAS full-length protein was pre-incubated with 1 μM 0N0 for 25 min followed by the addition of 0.25, 0.5, 1 and 2 μM 147-bp nucleosomal DNA. In Extended Data Fig. 1b, 0.5 μM human cGAS full-length protein was pre-incubated with 0.5 μM 147-bp nucleosomal DNA for 25 min followed by the addition of 0.25, 0.5, 1 and 2 μM 0N0 nucleosome. In Fig. 3d, 0.5 μM mouse cGAS full-length protein wild type and different cGAS mutants were incubated for 2 h at 37 °C with 50 ng plasmid DNA and 1 μM 0N0 nucleosome. In Fig. 3e, 0.3 μM or 3 μM mouse cGAS protein were incubated with 0.12 μM 12× nucleosome array in different buffer conditions. 1.5 μM 0N0 mononucleosomes were added after 5 min pre-incubation with nucleosomal array. In Extended Data Fig. 5j, 0.5 μM of human cGAS full-length protein were pre-incubated with 0, 0.125, 0.25, 0.5 and 1 μM 0N0 wild-type nucleosome or 0N0 acidic patch mutant nucleosome followed by the addition of 0.5 μM 147-bp nucleosomal DNA. In Extended Data Fig. 5k, 0.5 μM 40N40 nucleosome were incubated with 0.125, 0.25, 0.5, 1 and 2 μM human cGAS full-length protein.

Electron microscopy and data collection

Quantifoil Cu 200, R 2/1 grids were plasma cleaned 7 s at 20 mA (Glo-Qube, Quorum). Grids were prepared with a Leica EM GP plunge freezer (Leica), at 10 °C and 95% humidity. Then, 2.8 μM 0N0 nucleosome in low-salt buffer (25 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.25 mM DTT) were mixed in a 1:1 ratio with mouse cGAS catalytic domain, and octyl-β-glucoside was added to a final concentration of 0.05%. Next, 4.5 μl of sample was immediately applied to grids and blotted 2.0 s before vitrification in liquid ethane. Data were collected on a Titan Krios G3 transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV with a K2 Summit direct electron detector (Gatan) operated in counting mode, and GIF energy filter with a slit width of 20 eV (Gatan). Two datasets (1,211 and 3,050 movies) were collected with the EPU software package (TFS), with a defocus range of −2.8 to −1.0 μm, total dose of 44.8 e⁻ Å⁻² over 40 frames, and magnified pixel size of 1.059 Å.

Electron microscopy data processing

Micrograph movies were motion corrected and dose-weighted with MotionCor2[33], and CTF parameters were estimated with CTFFIND4[34]. All further cryo-EM data processing steps were conducted using RELION v.3.0.7[35] unless stated otherwise and resolutions reported are based on the gold standard FSC 0.143 criterion (Extended Data Fig. 2). Five-hundred particles were manually picked from the initial test data set of 1,211 micrographs and extracted using a 200 pixel box size. 2D class averages were calculated and the three best classes were low pass filtered to 20 Å and used for automated template picking of 680,000 particles from the initial data set. These particles were subjected to 2D classification, and 590,000 particles in the selected classes were imported to cryoSPARC v.2.11[36] and used to generate an 3D ab-initio reconstruction. From the second dataset with 3,050 micrographs, 1,650,000 additional particles were picked using the previous 2D classes for automated template picking. The ab initio model from cryoSPARC was used as a reference for the first round of 3D classification containing a total of 2,100,000 particles. Then, 1,380,000 particles were then selected for a second round of 3D classification performing local angular searches. The selected 173,000 particles yielded a cryo-EM map of the cGAS–nucleosome complex at 3.36 Å and subsequent RELION refinement and solvent mask post-processing yielded a cryo-EM map of the complex at 3.1 Å resolution. Directional FSC plot for the final 3D reconstitution

of the cGAS-nucleosome complex was performed using the 3DFSC software[37].

Model building

We performed rigid docking in UCSF Chimera[38] using available crystal structures of mouse cGAS catalytic domain (PDB code 4LEZ) and nucleosome (PDB code 3LZ0). Model building was performed manually in COOT[39] using secondary structure restraints for protein and libg restraints for DNA and final real-space refinement was performed in PHENIX 1.17.

Figure preparation

Figures were prepared using UCSF ChimeraX[40], PyMOL (The PyMOL Molecular Graphics System, version 1.7.4.4 Schrödinger, LLC) and Prism v.8.0 (GraphPad Software).

Generation of cGAS THP-1 knockouts

THP-1 cells were obtained from ATCC and early passages of these cells were used for the experiments. No further authentication was done. CRISRP–Cas9 RNPs were assembled by annealing synthetic, chemically stabilized crRNA:tracrRNA pairs (IDT) at 95 °C for 5 min and incubation at room temperature for 30 min. gRNAs were then mixed with recombinant NLS-Cas9 protein for 20 min at room temperature. Cas9 (40 pmol) was added for each 100 pmol of cGAS gRNA. RNPs were then mixed with the cell suspension and nucleofection was conducted using the SG Cell Line 4D-Nucleofector X Kit S (Lonza) on the X-unit of a 4D-nucleofector (Lonza) (program FF-100). After nucleofection, cells were collected from the nucleofection cuvettes and transferred into a 96-well plate. Cells were allowed to recover for 48 h at 37 °C, 5% CO_2. THP-1 cells were then subject to single cell dilution cloning on in a round-bottom 96-well plate in culture. When colonies became visible after approximately 3 weeks, clones were collected and subjected to MiSeq analysis as previously described[41]. All cell lines were regularly tested for mycoplasma contamination. gRNA1: CAAAACCGCCCGG AGCTTCG; gRNA2: CGCATCCCTCCGTACGAGAA.

Cell culture

THP-1 cells were cultured in RPMI Medium 1640, supplemented with 10% (v/v) FCS, L-glutamine and 100 U ml^{-1} penicillin-streptomycin. For differentiation, cells were incubated with 300 ng ml^{-1} PMA for 3 h. Afterwards, cells were washed twice with PBS and 50,000 cells were seeded into a 96-well tissue culture plate (for ELISA), 500,000 cells into a 24-well tissue culture plate (for ISG immunoblots), or 5 million cells into a 10 cm tissue culture plate (for fractionation) containing culture medium. After 24 h of resting, cells were used for experimentation.

Cell stimulation

THP-1 cells were simulated after a resting phase of one day and an overnight treatment with Doxycycline (1 µg ml^{-1}). For stimulation, HT-DNA (200 ng per 96-well) or (975 ng per 24-well) was incubated for 5 min at room temperature in Opti-MEM (Gibco), complexed with (0.25 µl per 96-well) (1.3 µl per 24 well) of Lipofectamine 2000 per well (Thermo, 11668019) in Opti-MEM, which was also incubated separately for 5 min at room temperature. Afterwards the two reagents were mixed and incubated another 5 min at room temperature. The transfection mix was then applied to each well, totalling 100 µl (96-well) or 550 µl (24-well). Stimulation occurred for 8 h at 37 °C. For ELISA, supernatants were harvested, and IP-10 cytokines were measured at 1:100 dilution using a human IP-10 ELISA (BD Biosciences). For immunoblot, cells were washed twice in PBS, then lysed directly in 1× Laemmli. For THP-1 stimulations in 10 cm plates, 20 µg of HT-DNA was incubated for 5 min room temperature. Polyethylenimine (50 µl) was incubated separately on ice for 5 min. Afterwards, the two reagents were mixed and incubated for 30 min room temperature. The transfection mix was then applied to each plate, totalling 7.5 ml.

Cytoplasmic or nuclear fractionation

After overnight doxycycline treatment and 8 h HT-DNA stimulation, cells were trypsinized for 10 min at 37 °C and lifted with medium. Cells were then washed twice with ice-cold PBS and counted. 2.2 million THP-1 cells were then washed twice with buffer A (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCL, 1 mM DTT supplemented with protease inhibitors), and lysed using 50 µl of buffer A supplemented with 0.01% NP-40 for 15 min on ice, vortexing every 5 min. Lysate was centrifuged at 20,817g for 10 min 4 °C and supernatant was collected (cytoplasmic fraction). Pellet was washed three times in buffer A without detergent. Pellet was then lysed in 20 µl of buffer B (20 mM HEPES pH 7.9, 420 mM $MgCl_2$, 1.5 mM KCl, 1 mM DTT, 25% glycerol, supplemented with protease inhibitors) for 10 min on ice, then subject to 2,000 rpm shaking at 4 °C for 15 min. Nuclear lysate was then centrifuged for 15 min 20,817g and supernatant was collected (420 mM KCl soluble nuclear fraction). Pellet was washed twice in buffer A then lysed directly in 20 µl of 1× Laemmli buffer (Laemmli soluble nuclear fraction). Equal amounts of each fraction were then subject to repeated freeze/boil cycles to destroy DNA and loaded onto SDS–PAGE gels.

Antibodies

Western blots were probed with following primary antibodies: anti-lamin A (Abcam, Ab26300), anti-cGAS (CST, 15102), anti-Viperin (Merck, MABF106), anti-OAS1 (Santa Cruz, SC374656), anti-actin HRP-conjugated (Santa Cruz, SC47778) and secondary antibodies anti-mouse IgG HRP-linked (CST, 7076S) and anti-rabbit IgG HRP-linked (CST, 7074S).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The electron density reconstruction and final model were deposited at the Electron Microscopy Data Bank (EMDB) with accession code EMD-11601, and the Protein Data Bank (PDB) with accession code 7A08. Source data are provided with this paper.

29. Klinker, H., Haas, C., Harrer, N., Becker, P. B. & Mueller-Planitz, F. Rapid purification of recombinant histones. *PLoS One* **9**, e104029 (2014).
30. Huynh, V. A., Robinson, P. J. & Rhodes, D. A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone. *J. Mol. Biol.* **345**, 957–968 (2005).
31. Rogge, R. A. et al. Assembly of nucleosomal arrays from recombinant core histones and nucleosome positioning DNA. *J. Vis. Exp.* **79**, e50354 (2013).
32. Rueden, C. T. et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).
33. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
34. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
35. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
36. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
37. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
38. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
41. 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).

Article

a

b

Extended Data Fig. 1 | cGAS binds nucleosomes even in the presence of free DNA. a, Gel-mobility shift binding analysis of purified cGAS and nucleosomes. Nucleosomes without linker DNA (0N0) were tested for binding to purified human (h) and mouse (m) full-length cGAS and cGAS^cat. Data are representative of two biological replicates. **b**, Human cGAS activity assay preincubated with dsDNA followed by titration of 0N0 nucleosomes. Data are representative of two biological replicates.

Article

a

b

Sphericity = 0.900 out of 1. Global resolution = 3.11 Å

c

0.143 criterion

3.11Å

Histogram of Directional FSC — Global FSC
± 1 S.D. from Mean of Directional FSC

d

test dataset
1211 micrographs (1.059Å/pixel)

523 particles manual picking

2D classification (RELION)

dataset
3050 micrographs (1.059Å/pixel)

681,101 particles

autopicking

1,658,631 particles

autopicking template from test dataset

2D classification (RELION)

2D classification (RELION)

591,603 particles

Homogeneous Refinement

1,523,266 particles

cryoSPARC ab initio model

cryoSPARC reference map

joined datasets
2,114,869 particles

3D classification (RELION)

1.5 %   65 %   15 %   18.5 %

1,379,788 particles

3D classification (RELION) local search

12.5 %   22.5 %   12.5 %   23 %   14 %   15.5 %

172,977 particles

3.36 Å

mask

movie refinement and particle polishing

3.11 Å

model building of cGAS–nucleosome complex

Extended Data Fig. 2 | Cryo-EM data processing for cGAS–nucleosome structure. a, Representative micrograph of the dataset used to determine the cGAS–nucleosome complex structure. **b**, Left, final reconstruction of the cGAS–nucleosome complex coloured by local resolution. Right, representation of angular distribution of particles contributing to the final map. **c**, Histogram and directional Fourier shell correlation (FSC) plot for the final 3D reconstitution of the cGAS–nucleosome complex (3.11 Å). A sphericity of 0.9 was determined indicating very isotropic angular distribution (a value of 1 stands for completely isotropic angular distribution). The global resolution was determined to 3.11 Å (0.143 criterion). Directional FSC determination was performed with the 3DFSC software. **d**, Flow chart for image processing using RELION and cryoSPARC.

a

nucleosome

cGAS
catalytic domain

Histone H2A
residues 39 – 75

Histone H2B
residues 47 – 62

nucleosomal DNA

cGAS
residues 240 – 246

cGAS
residues 314 – 323

cGAS spine
residues 148 – 182

b   interface II
DNA binding site B

interface I
tethering loops 1 and 2
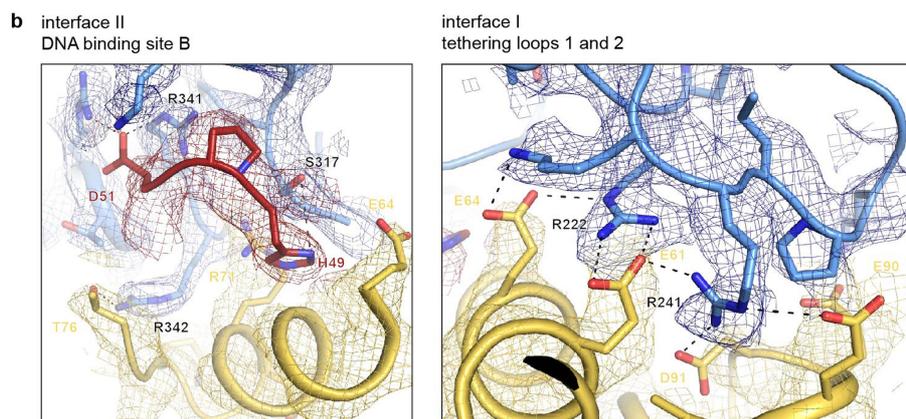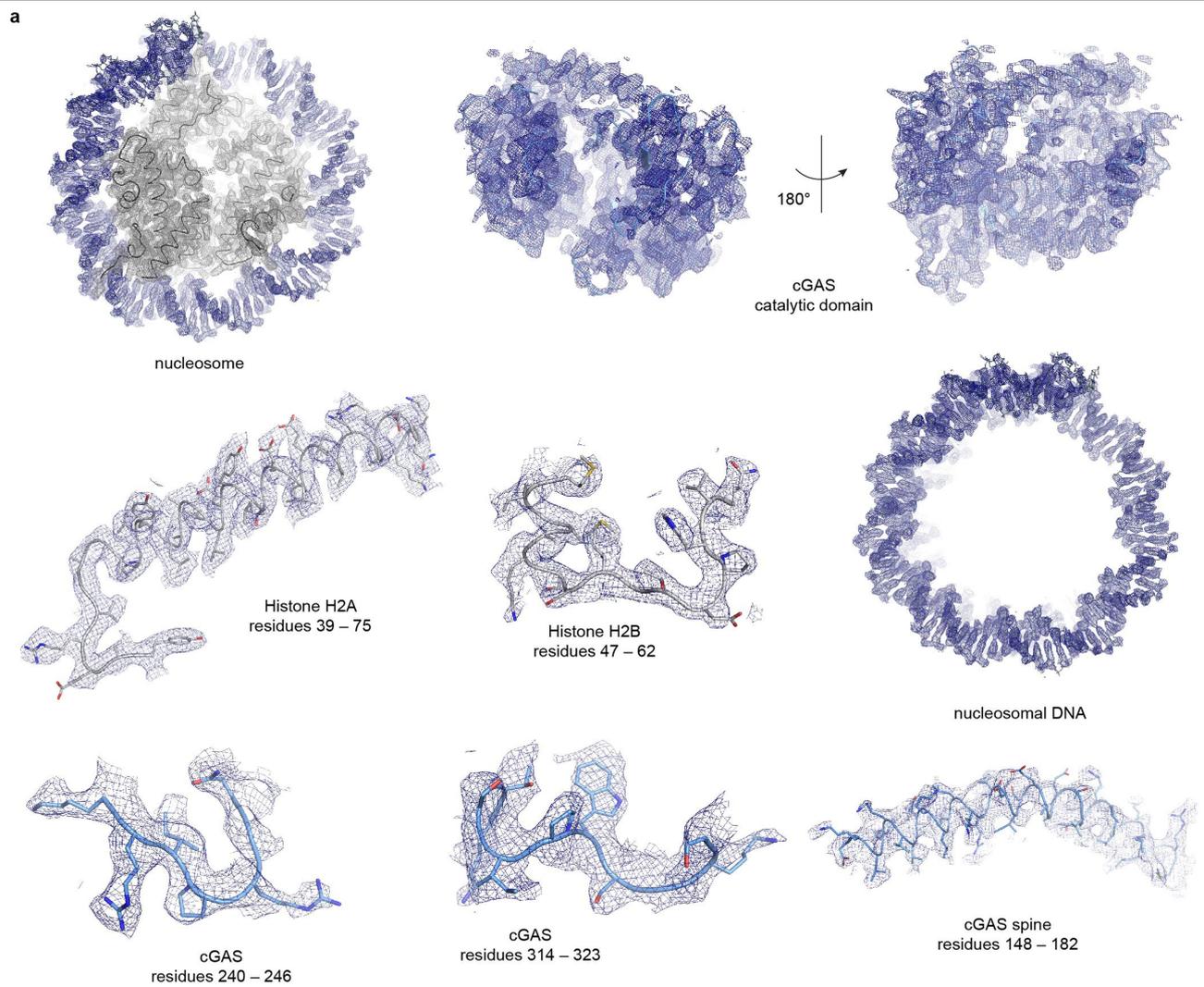
Extended Data Fig. 3 | Sample density maps for cGAS–nucleosome structure. a, Representative examples of cryo-EM map areas of cGAS, nucleosomal DNA and histones used for model building. **b**, Electron density for cGAS (blue) and H2A–H2B (yellow, red) interacting residues in interface I and interface II. cGAS tethering loops 1 and 2 with key interacting residues R222 and R241 are depicted as well as DNA-binding site B residues. Dashed lines represent hydrogen bonds.

Article
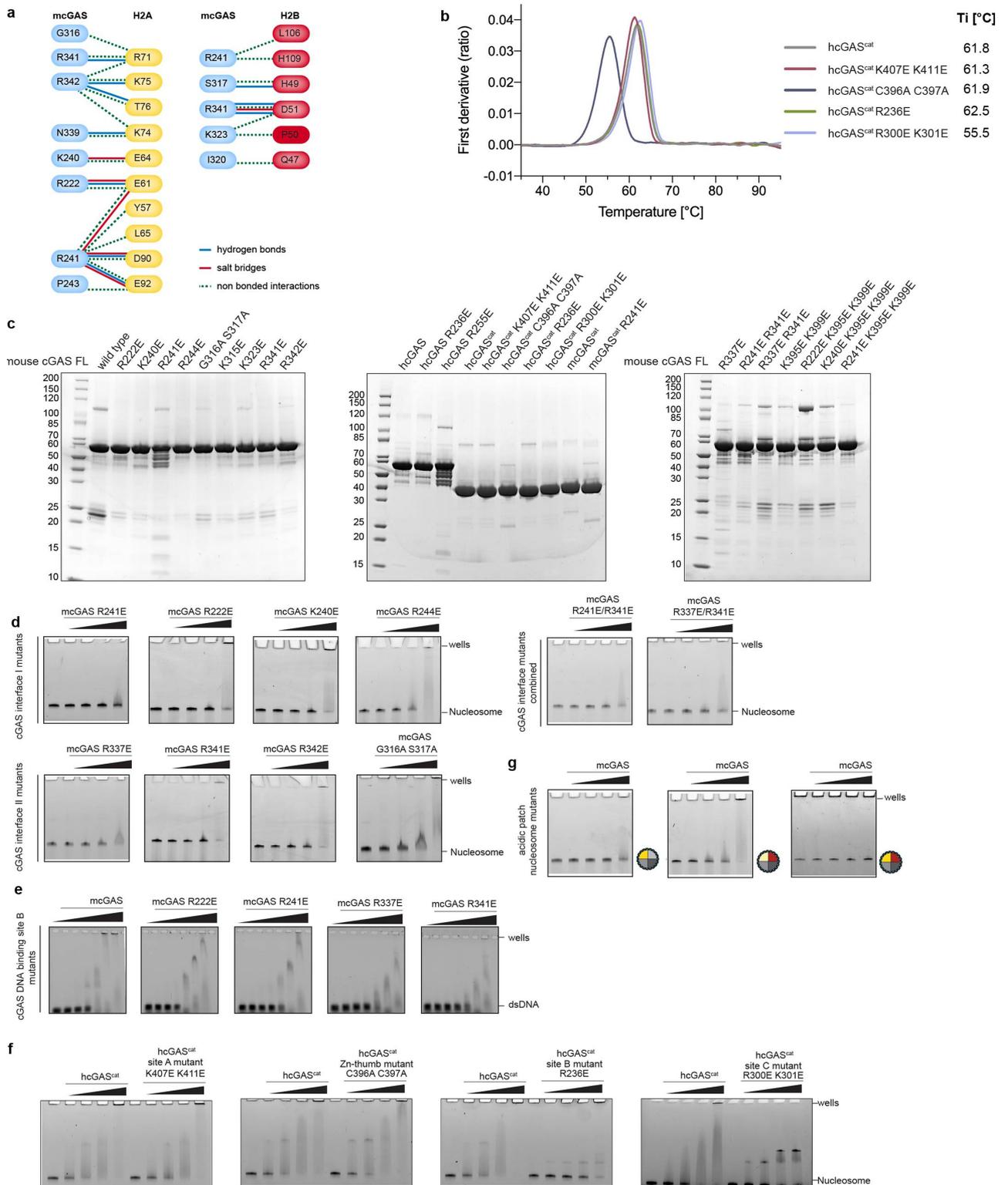
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Mutational analysis of binding interface between cGAS and the nucleosome. a, Protein–protein residue interactions across the interface of cGAS with histone H2A and cGAS with histone H2B. Interacting amino acids are joined by coloured lines, each representing a different type of interaction, as per the key below. Interaction maps for the cGAS–nucleosome complex were generated using PDBsum. **b**, Thermal shift assay derivative melt curve plots of human cGAS^{cat} mutants. Respective inflection temperatures are: cGAS^{cat} 61.8 °C; cGAS^{cat}(K407E/K411E) 61.3 °C; cGAS^{cat}(C396A/C397A (55.5 °C; cGAS^{cat}(R236E) 61.9 °C, cGAS^{cat}(R300E/K301E) 62.5 °C. Data are representative of two biological replicates. **c**, Coomassie stained SDS–PAGE gels of purified recombinant human and mouse cGAS (7 μg each) constructs used in this study. Gels are representative of one replicate. **d**, Representative EMSAs for mouse cGAS–nucleosome interface I and interface II mutants binding to fluorescently labelled nucleosomes. Data are representative of three biological replicates. **e**, EMSAs for mouse cGAS mutants in tethering loops 1 and 2 (R222E, R241E) and DNA-binding site B (R337E, R341E) binding to fluorescently labelled nucleosomal DNA. Data are representative of two biological replicates. **f**, EMSAs for human cGAS full-length and catalytic domain, DNA-binding site A (K407E/K411E), Zn-thumb (C396A/C397A), site B (R236E) and site C (R300E/K301E) mutants binding to fluorescently labelled nucleosomes. Data are representative of two biological replicates. **g**, Representative EMSAs for mouse cGAS full-length binding to fluorescently labelled acidic patch mutant nucleosomes apI (H2A(E61A/E64A/D90A)) and apII (H2A(R71A), H2B(H49A/D51A)) and apI + apII (H2A(E61A/E64A/R71A/D90A/E92A), H2B(H49A/D51A)). Data are representative of three biological replicates.

Article
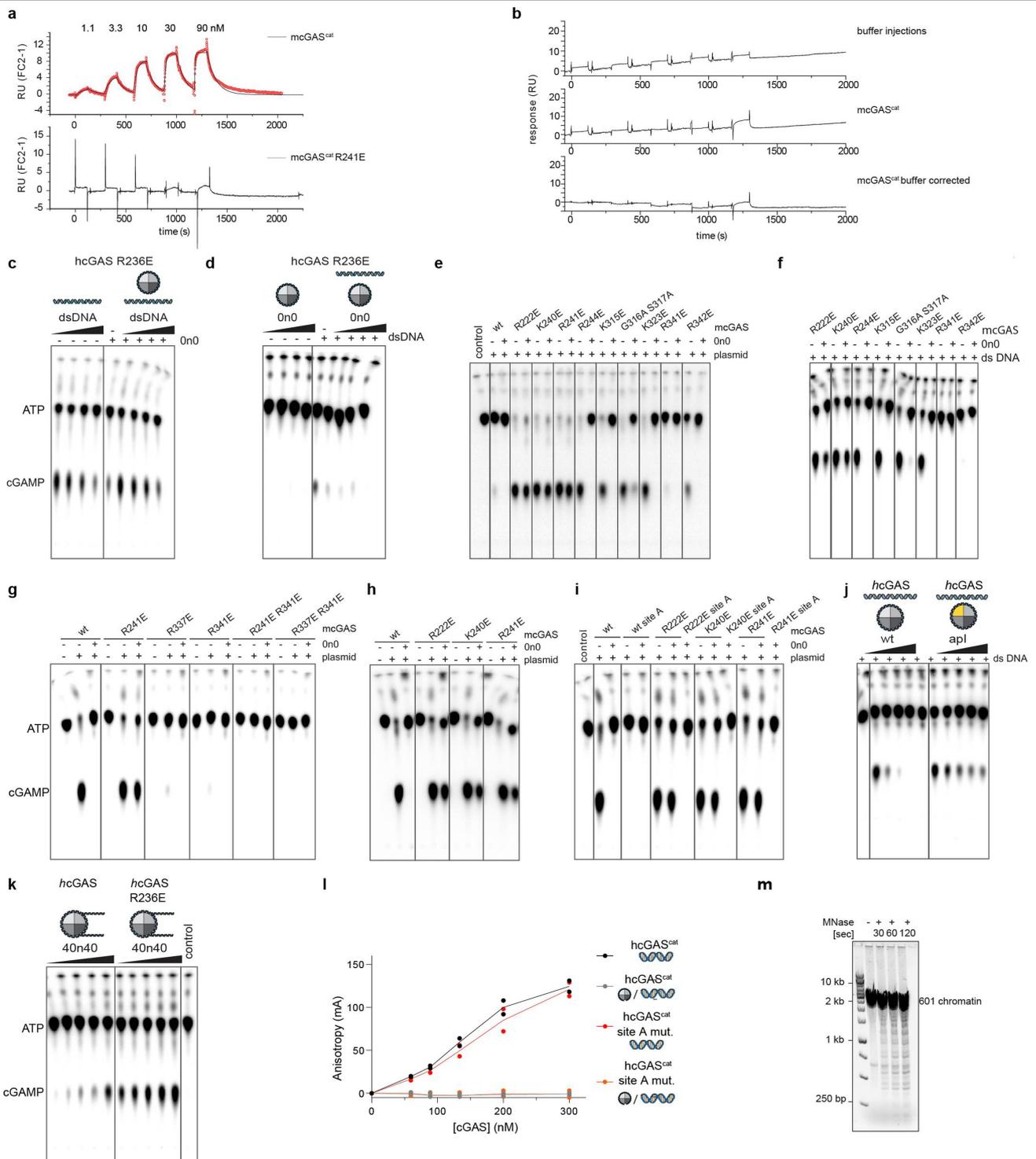
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | cGAS DNA-binding site B is required for cGAS tethering by the nucleosome. a, SPR analysis of single-cycle-kinetics experiment with immobilized nucleosomes via biotinylated DNA and mouse cGAS^{cat} and cGAS^{cat}(R241E) mutant as analytes. Shown are injections of 1.1, 3.3, 9, 10, 30 and 90 nM mouse cGAS^{cat} and cGAS^{cat}(R241E). Data are representative of two biological replicates. **b**, SPR analysis with acidic patch mutant nucleosomes (H2A(E61A/E64A/D90A)) immobilized via biotinylated DNA and cGAS^{cat} as analyte. Shown are buffer injections, injections of 1.1, 3.3, 9, 10, 30 and 90 nM mouse cGAS^{cat} and the cGAS^{cat} background-corrected data. Mouse cGAS^{cat} has orders of magnitude lower affinity to acidic patch mutant nucleosomes than wild-type nucleosomes. Data are representative of two biological replicates. **c**, Human cGAS(R236E) mutant activity assay preincubated with dsDNA followed by titration of 0N0 nucleosome. Data are representative of two biological replicates. **d**, Human cGAS(R236E) mutant activity assay pre-incubated with 0N0 nucleosome followed by titration of dsDNA. Data are representative of two biological replicates. **e**, Mouse cGAS mutations tested affecting cGAS–nucleosome interactions were tested for DNA-dependent activation with plasmid DNA in the presence of 0N0 nucleosomes. cGAMP production was assayed by thin-layer chromatography. Data are representative of two biological replicates. **f**, Mouse cGAS mutations affecting cGAS–nucleosome interactions were tested for DNA-dependent activation with 147-bp nucleosomal DNA in the presence of 0N0 nucleosomes. Data are representative of two biological replicates. **g**, Mouse cGAS single and double mutations of tethering loop and DNA-binding site B were tested for cGAMP production in the presence of plasmid DNA alone or plasmid DNA and 0N0 nucleosome. Data are representative of two biological replicates. **h**, Mouse cGAS mutants R222E, K240E and R241E require plasmid DNA for activation. Data are representative of two biological replicates. **i**, Mouse cGAS mutants R222E, K240E and R241E were tested for DNA-dependent activation with plasmid DNA in the presence of 0N0 nucleosomes. Mutation of DNA-binding site A abolishes activation by plasmid DNA. Data are representative of two biological replicates. **j**, Human cGAS activity assay in the presence of dsDNA, followed by titration of 0N0 and 0N0 acidic patch mutant I (apI; H2A(E61A/E64A/D90A)). Data are representative of two biological replicates. **k**, Wild-type and R236E mutant cGAS activity assays with 40N40 nucleosomes (40-bp linker DNA on each side). Data are representative of two biological replicates. **l**, Fluorescence anisotropy analysis of human cGAS^{cat} human cGAS^{cat} site A mutant (K407E/K411E) binding to fluorescently labelled 20-bp dsDNA and in the presence of 0N0 nucleosomes. Data are representative of two biological replicates. **m**, Agarose gel of micrococcal nuclease (MNase)-digested synthetic 601 chromatin indicating a regular nucleosomal structure. Data are representative of two biological replicates.

Article
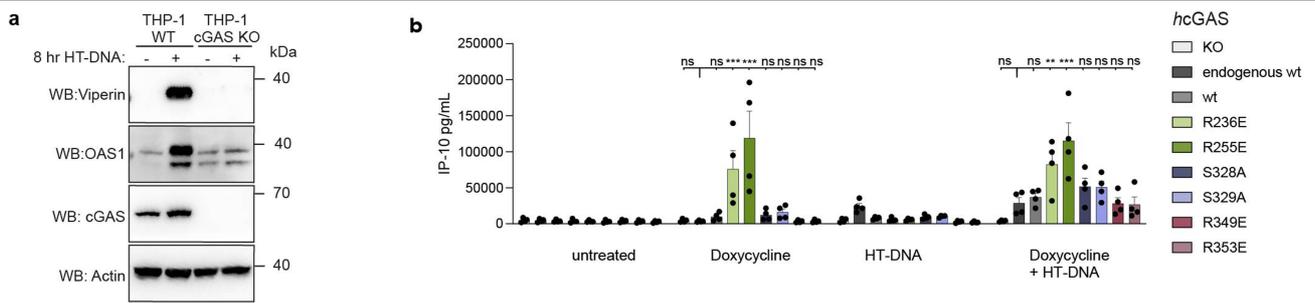
Extended Data Fig. 6 | IP-10 cytokine and ISG production upon self and non-self DNA recognition. a, PMA-differentiated wild-type or knockout human *CGAS*$^{-/-}$ THP-1 cells were treated with doxycycline (1 μg ml^{-1}) overnight and either left untreated or treated with HT-DNA (975 ng per 550 μl) for 8 h. Cell lysates were separated on SDS–PAGE gels, western blotted and probed with the indicated antibodies. Data are representative of three biological replicates. **b**, PMA-differentiated THP-1 cells were left untreated or treated with doxycycline overnight to express the indicated human cGAS mutants, followed by stimulation using HT-DNA (200 ng per well) for 8 h. Supernatant was collected and IP-10 cytokines were measured from supernatant using ELISA. Data are mean and s.e.m. of four biological replicates. ***P < 0.001, **P < 0.01, two-way ANOVA. ns, not significant. Doxycycline R236E *P* < 0.001; doxycycline R255E *P* < 0.001; doxycycline + HT-DNA R236E *P* = 0.002; doxycycline + HT-DNA R255E *P* < 0.001.
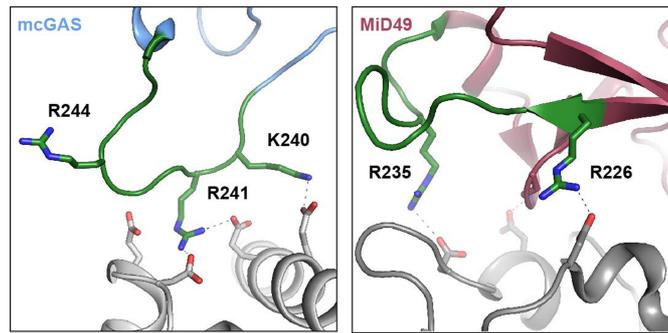
Extended Data Fig. 7 | Structurally conserved loop for protein–protein interactions in MAB21 family nucleotidyltransferases. Structures of cGAS–nucleosome (blue) and MID49–DRP1 (red, PDB code 5WP9). The structurally conserved loop is depicted in green, showing interacting residues. Dashed lines represent hydrogen bonds. Positively charged, basic amino acids on the nucleotidyltransferase loop interact with the acidic patch of the nucleosome or with the protein DRP1, respectively.

Article

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	#1 cGAS–nucleosome (EMDB EMD-11601) (PDB 7A08)
Data collection and processing	
Magnification	130000
Voltage (kV)	300
Electron exposure (e–/Å^2)	44.8
Defocus range (μm)	-1.0 – -2.8
Pixel size (Å)	1.059
Symmetry imposed	C1
Initial particle images (no.)	2114869
Final particle images (no.)	172977
Map resolution (Å)	3.11
0.143 FSC threshold	
Map resolution range (Å)	3.0 – 4.9
Refinement	
Initial model used (PDB code)	4LEZ cGAS 3LZ0 nucleosome
Model resolution (Å)	3.1
0.5 FSC threshold	
Model resolution range (Å)	80 – 3.2
Map sharpening B factor (Å^2)	-84.47
Model composition	
Non-hydrogen atoms	13969
Protein residues	1070
Ligands	1
B factors (Å^2)	
Protein	52.33
DNA	101.85
Ligand	63.23
R.m.s. deviations	
Bond lengths (Å)	0.0056
Bond angles (°)	0.66
Validation	
MolProbity score	1.76
Clashscore	10.0
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	94.01
Allowed (%)	5.909
Disallowed (%)	0

nature research

Corresponding author(s):	Karl-Peter Hopfner
Last updated by author(s):	May 12, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Epu v2.6.1.69REL , TEM User interface Titan v6.15.3.5341REL , Digital Micrograph v3.22.1461
Data analysis	MotionCor2, CTFFIND4, RELION 3.0.7, cryoSPARC 2.11, Phenix 1.17, Coot 0.9-pre, PyMOL 1.7.4.4., UCSF Chimera 1.14, UCSF ChimeraX 0.93, Prism 8.0, Image J 1.52p (JAVA 1.8.0_172), Tycho NT.6 software (v 1.3.1.868), Biacore X100 evalution software (V2.0.2 Plus package)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Electron density reconstruction and final coordinates are deposited at the Protein Data Bank (PDB) and The Electron Microscopy Data Bank (EMDB) under accession codes 7A08 and EMD-11601.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were selected based on previous experience to obtain statistical significance and reproducibility
Data exclusions	no data was excluded
Replication	All experiments were performed with independent replicates as described in the figure legends.
Randomization	no randomization was performed
Blinding	no blinding was perfomed

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	cGAS D1D3G rabbit maB #15102; OAS1 F-3 mouse, sc-374656; Lamin A, lot 686364, ab26300-100; Viperin- clone MaP.VIP, mouse monoclonal, lot 2683305, MABF106; Actin-HRP, sc-47778, Lot K1418
Validation	Validation statements from manufacturer's website: cGAS validated for WB; OAS1 validated for WB; Lamin A KO validated; Viperin validated for WB; Actin-HRP validated for WB

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	THP-1 were originally obtained from ATCC.
Authentication	No methods were used for authentication.
Mycoplasma contamination	Cells have been tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

3.2 Unified mechanisms for self-RNA recognition by RIG-I Singleton-Merten-syndrome variants

Lässig, C., Lammens K., Gorenflos López, J-L., <u>Michalski, S.</u>, Fettscher, O., Hopfner, K-P. *eLife* (**2018**);7:e38958. https://doi.org/10.7554/eLife.38958

This work explains the constant activation of the pathogenic RIG-I Singleton-Merten-Syndrome (SMS) variant C268F using a combination of biochemical assays and X-ray crystallography. Two common mutations of RIG-I in SMS are E373Q and C268F. Whereas, the mechanism of increased autoimmunity of the E373Q variant was investigated before in this group, the mechanism of C268F was still elusive. Especially, the constant activation of this variant compared to the spatially close but catalytically dead K270I variant was intriguing.

Using fluorescence anisotropy measurements, we showed that the affinity towards self-like RNA ligands is higher for RIG-I C268F compared to RIG-I WT and even more interesting independent of ATP. Additionally, reduced catalytic activity was observed for this variant in radioactive ATP-hydrolysis assays. Also, the affinity towards ATP was lower for RIG-I C268F than WT in tryptophan-fluorescence based MANT-ATP FRET assays. This ATP-insensitivity suggested a pseudo-ATP bound state for RIG-I C268F. We confirmed the biochemical observations by solving the structure of RIG-I C268F using X-ray crystallography. The RIG-I C268F mutant in complex with RNA lacked incorporation of any nucleotide. To compensate, the bulky phenylalanine displaces the conserved K270 residue, resulting in salt-bridge formation with E702 creating a pseudo-ATP bound state without actual nucleotide bound.

Concluding, this work shows the constant activation of RIG-I C268F by mimicking an ATP-bound state, thereby bypassing the needed regulation to prevent autoimmunity.

<u>Author contribution:</u>
The author did purify recombinant RIG-I C268F protein and screened a multitude of different novel crystallization conditions. He grew crystals, that were measured with the help of K. Lammens and participated in manuscript preparation.

Unified mechanisms for self-RNA recognition by RIG-I Singleton-Merten syndrome variants

Charlotte Lässig[1,2], Katja Lammens[1,2], Jacob Lucián Gorenflos López[1,2], Sebastian Michalski[1,2], Olga Fettscher[1,2], Karl-Peter Hopfner[1,2,3]*

[1]Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany; [2]Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany; [3]Center for Integrated Protein Science Munich, Munich, Germany

Abstract The innate immune sensor retinoic acid-inducible gene I (RIG-I) detects cytosolic viral RNA and requires a conformational change caused by both ATP and RNA binding to induce an active signaling state and to trigger an immune response. Previously, we showed that ATP hydrolysis removes RIG-I from lower-affinity self-RNAs (*Lässig et al., 2015*), revealing how ATP turnover helps RIG-I distinguish viral from self-RNA and explaining why a mutation in a motif that slows down ATP hydrolysis causes the autoimmune disease Singleton-Merten syndrome (SMS). Here we show that a different, mechanistically unexplained SMS variant, C268F, which is localized in the ATP-binding P-loop, can signal independently of ATP but is still dependent on RNA. The structure of RIG-I C268F in complex with double-stranded RNA reveals that C268F helps induce a structural conformation in RIG-I that is similar to that induced by ATP. Our results uncover an unexpected mechanism to explain how a mutation in a P-loop ATPase can induce a gain-of-function ATP state in the absence of ATP.
DOI: https://doi.org/10.7554/eLife.38958.001

Introduction

Retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs) are cytosolic innate immune sensors that recognize viral double-stranded (ds)RNAs. RLRs (RIG-I, MDA5 and LGP2) are members of the so-called Superfamily II (SF2) helicases/translocases. They share a multi-domain architecture that consists of a central SF2 ATPase domain accompanied by two N-terminal tandem caspase activation and recruitment domains (2CARD, only in RIG-I and MDA5) and a C-terminal regulatory domain (CTD or RD) (*Rawling and Pyle, 2014*). The SF2 domain itself is built out of an N-terminal RecA-like domain (1A) and a C-terminal RecA-like domain (2A) that together form an ATP-binding pocket, as well as an insertion domain (domain 2B). SF2 and RD are crucial for RNA-recognition of RLRs, whereas 2CARD communicates successful RNA-binding events to downstream signaling factors (*Jiang et al., 2011*; *Kowalinski et al., 2011*; *Luo et al., 2011*). Specifically, RIG-I detects dsRNA ends harbouring 5' tri- or diphosphates (*Goubau et al., 2014*; *Schlee et al., 2009*; *Schmidt et al., 2009*). Simultaneous binding of an RNA ligand and ATP to RIG-I switches the protein into an active state in which the otherwise shielded 2CARD is released (*Zheng et al., 2015*). Activated RIG-I homo-tetramerizes via 2CARD (*Jiang et al., 2012*) and nucleates the polymerisation of its adapter protein, mitochondrial antiviral-signaling (MAVS), to elicit the innate immune signaling cascade (*Peisley et al., 2014*; *Wu et al., 2014*).

The similarity of epitopes of viral RNA recognized by RLRs, in particular dsRNA stems, to some endogenous ribonucleic acids has required the immune system to develop mechanisms besides merely recognizing 5'-di/triphosphate-containing RNA ends to discriminate self from non-self. Self-

ribonucleic acids for instance are shielded from RLRs by introducing 2'O-methylations or by destabilizing double-stranded parts through A-to-I editing (*Chung et al., 2018*; *Devarkar et al., 2016*; *Liddicoat et al., 2015*; *Schuberth-Wagner et al., 2015*). In addition, we and others have been able to show that the SF2 domain of RLRs itself confers a proof-reading activity by removing RIG-I from self-RNA so as to avoid autoimmunity (*Anchisi et al., 2015*; *Lässig et al., 2015*; *Louber et al., 2015*; *Rawling et al., 2015*). In particular, ATP turnover can lead to translocation on dsRNA stems that could remove the protein from the RNA and reinstall the inactivated state (*Myong et al., 2009*; *Yao et al., 2015*).

Deficiencies in any of these mechanisms can lead to immune recognition of self-RNAs (*Ahmad et al., 2018*; *Chiang et al., 2018*). For instance, single-nucleotide polymorphisms (SNPs) in RLR genes are known to cause system-wide autoimmune diseases such as Aicardi-Goutière syndrome (AGS) (*Oda et al., 2014*; *Rice et al., 2014*), Singleton-Merten syndrome (SMS) (*Jang et al., 2015*; *Rutsch et al., 2015*), systemic lupus erythematosus (SLE) (*Cunninghame Graham et al., 2011*; *Pettersson et al., 2017*; *Van Eyck et al., 2015*) or type 1 diabetes (*Liu et al., 2009*; *Smyth et al., 2006*). The molecular basis for the development of these diseases is in many cases not understood, because structural data on these RLR variants have been missing.

In this work, we present the biochemical and structural analysis of the RIG-I SMS variant C268F and unveil an ATP-independent signaling mechanism. We show that active site rearrangements of several amino acid side chains in RIG-I C268F mimic an ATP-bound state and activate the protein for signaling upon recognition of RNA ligands in the absence of ATP.

Results and discussion

The RIG-I SMS variants C268F and E373A are located within SF2 ATP binding and hydrolysis motifs I and II, respectively, (*Fairman-Williams et al., 2010*) (*Figure 1—figure supplement 1*) and were previously shown to be constitutively active in reporter cells without any external dsRNA trigger (*Jang et al., 2015*; *Lässig et al., 2015*) (*Figure 1—figure supplement 2A*). RIG-I E373A's enhanced immune signaling ability can be explained by proficient ATP binding but reduced ATP hydrolysis, since the motif II glutamate is implicated in positioning and polarizing the attacking water molecule in the hydrolysis reaction. The stabilized ATP state leads to increased binding of and activation by endogenous dsRNA (*Lässig et al., 2015*; *Louber et al., 2015*). By contrast, the molecular basis for activation of RIG-I C268F is still unknown as this mutation is located in motif I, which is normally associated with ATP binding and which is critical for RIG-I activation. For instance, a designed and widely used mutant with a defect in the invariant ATP phosphate-binding motif I lysine, K270A, possesses reduced ATP binding properties (*Rawling et al., 2015*) and the RIG-I K270A/I mutant is deficient in inducing an immune response (*Lässig et al., 2015*; *Yoneyama et al., 2005*). By contrast, the RIG-I SMS variant C268F shows the opposite effect and is constitutively active, although this mutation is only two amino acids away from K270A in the same ATP phosphate-binding 'P-loop' in motif I (*Figure 1—figure supplement 2A*).

To analyze whether the autoimmune activity of the SMS variant RIG-I C268F is RNA-dependent, we designed a double-mutant that is defective in RNA-binding and carried out interferon (IFN)-β-promoter-driven luciferase reporter assays with overexpressed proteins in HEK293T RIG-I KO cells (*Figure 1A*). In particular, we used a previously described T347A point mutation in the RNA-binding interface of the N-terminal RecA-like domain (1A) (*Figure 1—figure supplement 1*), that abrogates the signaling of wild type (wt) RIG-I in infected cells and decreases the affinity for dsRNA in vitro (*Lässig et al., 2015*). Our assays show, that RIG-I C268F, T347A fails to induce the IFN-β promoter in uninfected as well as in 19mer 5'-triphosphate (ppp)-dsRNA-stimulated cells, indicating that the RIG-I SMS single-mutant C268F induces immune signaling only if bound to endogenous or transfected RNA. In addition, competition assays of RIG-I C268F titrated with signaling-deficient RIG-I lacking the 2CARD domain (RIG-I Δ2CARD) or with RIG-I Δ2CARD E373Q also gradually decreased the IFN-β promoter-driven luciferase activity (*Figure 1—figure supplement 2B*). Thus, RIG-I C268F as well as RIG-I Δ2CARD (E373Q) seem to recognize identical RNA substrates which are, however, saturated with signaling-deficient RIG-I Δ2CARD (E373Q) at higher transfected DNA concentrations thus preventing an immune response. Furthermore, we can exclude the possibility that the RIG-I

eLIFE Research advance   Immunology and Inflammation | Structural Biology and Molecular Biophysics
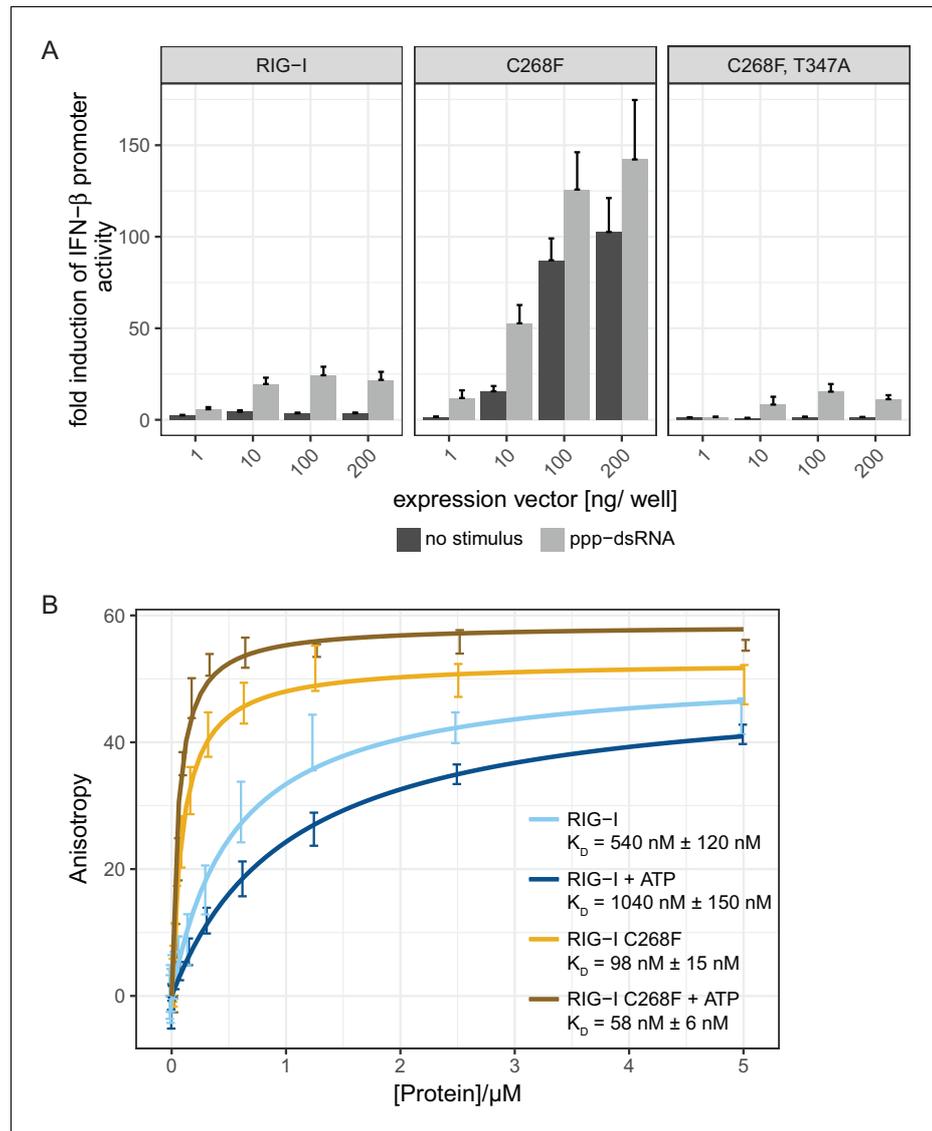
Figure 1. The RIG-I Singleton-Merten syndrome variant C268F signals in response to endogenous dsRNA. (**A**) Fold change of interferon (IFN)-β promoter-driven luciferase activity in uninfected HEK293T RIG-I KO cells or in cells stimulated with a 19mer 5′-triphosphate (ppp)-dsRNA upon overexpression of different RIG-I mutants. Cells were co-transfected with RIG-I expression vectors and p-125luc/pGL4.74 reporter plasmids, and stimulated with ppp-dsRNA 6 hr post transfection. Firefly luciferase activities were determined in respect to Renilla luciferase activities 16 hr after RNA stimulation. All ratios were normalized to an empty vector control. n = 4–12, error bars represent mean values + standard error of the mean (SEM). (**B**) Fluorescence anisotropy changes measured after titrating RIG-I or RIG-I C268F in the presence or absence of ATP into solutions containing a fluorescently labeled 14mer dsRNA. All binding curves were fit to a one-site binding equation using R. n = 4, error bars represent mean values ± standard deviation (SD).

DOI: https://doi.org/10.7554/eLife.38958.002

The following figure supplements are available for figure 1:

Figure supplement 1. Location of RIG-I amino acid substitutions used in *Figure 1*.
DOI: https://doi.org/10.7554/eLife.38958.003

Figure supplement 2. Comparison of the autoimmune signaling activity of RIG-I Singleton-Merten syndrome variants.
DOI: https://doi.org/10.7554/eLife.38958.004

C268F SMS mutation leads to a liberation of the 2CARD signaling module in the absence of RNA, for example by unfolding SF2. Instead, signaling by the RIG-I C268F SMS variant is dependent on 2CARD release triggered by binding to RNA molecules, as it is in wild type RIG-I.

To validate the intact RNA-binding properties of RIG-I C268F, we performed in vitro fluorescence anisotropy assays of purified proteins with a labeled hairpin (hp) RNA containing non-base-paired RNA ends (*Figure 1B*). We chose to use a double-stranded RNA ligand without blunt ends to suppress the dominant binding of RIG-I's RD to terminal RNA base pairs and to simulate recognition of endogenous-like RNA species that could be present within the cytosol. As expected, wild type RIG-I shows a moderate binding affinity to this ligand that is further decreased in the presence of ATP. By contrast, RIG-I C268F displays an already increased affinity to the hpRNA that is even enhanced in the presence of ATP. This confirms the previous results for RIG-I C268F showing an increased co-purification with endogenous RNA molecules (*Lässig et al., 2015*) and indicates that RIG-I C268F has two conformations (apo and ATP-bound) that have increased dsRNA affinity.

Since an intact SF2 ATPase domain of RIG-I is needed so that signaling only occurs when foreign RNA molecules are recognized (*Louber et al., 2015*; *Rawling et al., 2015*), we further analyzed the ATP binding and hydrolysis properties of RIG-I C268F in vitro. The ATP-binding-deficient and hydrolysis-deficient motif I mutant RIG-I K270I and the hydrolysis-deficient motif II mutant RIG-I E373Q both served as references. ATP hydrolysis assays with ^{32}P-labeled ATP confirmed a loss of catalytic activity of RIG-I C268F and of both motif I and II mutants in the presence of dsRNA (*Figure 2A*). In accordance with our RNA-binding experiments, this evidence supports a model in which RIG-I C268F has defects in dissociating from endogenous RNA, because it lacks the capability for ATP turnover and hence translocation. To further analyze the ATP-binding properties of the RIG-I SMS variant, we conducted a tryptophan fluorescence-based FRET assay with MANT-ATP (*Rawling et al., 2015*) (*Figure 2B*). In the absence of any RNA ligand, both wild type RIG-I and ATP-hydrolysis-deficient RIG-I E373Q show comparable affinities for MANT-ATP and MANT-ATPγS in the low μM range (*Table 1*). As expected, ATP-binding-deficient RIG-I K270I has a reduced affinity for ATP (in the medium μM range). In accordance with previous data (*Kohlway et al., 2013*), the presence of RNA increases the affinity of wtRIG-I and RIG-I E373Q, but not of RIG-I K270I for MANT-ATP or MANT-ATPγS. Interestingly, like RIG-I K270I, the RIG-I SMS variant C268F displays reduced ATP-binding affinities compared to that of wtRIG-I or RIG-I E373Q independently of the availability of an RNA ligand (*Figure 2B*, *Table 1*). This is puzzling as both RNA and ATP binding are normally needed to induce a molecular switch within RIG-I in order to release the 2CARD module and to activate an immune response (*Shah et al., 2018*; *Zheng et al., 2015*). RIG-I C268F might thus be able to signal even in the absence of a bound ATP molecule. However, even though it is possible that under cellular conditions of 1 mM ATP or even higher, ATP is still bound by both motif I mutants, the molecular switch to release 2CARD is only triggered in RIG-I C268F but not in RIG-I K270I (*Figure 1—figure supplement 2A*). These peculiarities of the ATP-bound states of the motif I mutants need to be addressed in future studies.

To further investigate the influence of ATP binding to RIG-I C268F, we mutated two residues that are implicated in ATP adenine recognition, R244A and Q247A (*Figure 2—figure supplement 1*), thereby further lowering the ATP-binding affinity of the SMS mutant, and tested the ability of the resulting triple-mutant to induce an immune response (*Figure 2C*). We expected to see no decrease in the autoimmune signaling activity of the triple-mutant compared to that of the RIG-I SMS single-mutant if the single mutant was not able to bind ATP under cellular conditions. However, compared to the SMS variant itself, the IFN-β promoter activity of the ATP-binding mutant RIG-I R244A, Q247A, C268F was reduced, albeit still significantly higher than that of wild type RIG-I. Hence, these data support a model in which RIG-I C268F is able to release its 2CARD domain and start a signaling cascade by a process that is at least partly independent of ATP.

In order to elucidate the molecular basis of autoimmune signaling by RIG-I C268F, we went on to crystalize RIG-I Δ2CARD C268F in the presence of a 14mer dsRNA and ADP·BeF$_x$ or ATP and determined the corresponding structures (*Figure 3—figure supplement 1A*). Intriguingly, despite an overall high structural similarity to wtRIG-I Δ2CARD bound to dsRNA (*Jiang et al., 2011*), the SMS mutant shows crucial amino acid side chain rearrangements in the active site that stabilize the protein in an activated state but prevent binding and co-crystalization of a nucleotide (*Figure 3A*, *Figure 3—figure supplement 1B*). In particular, the bulky F268 side chain displaces the evolutionary invariant motif I K270 from its central ATP phosphate-binding position in the P-loop, into a position

eLIFE Research advance
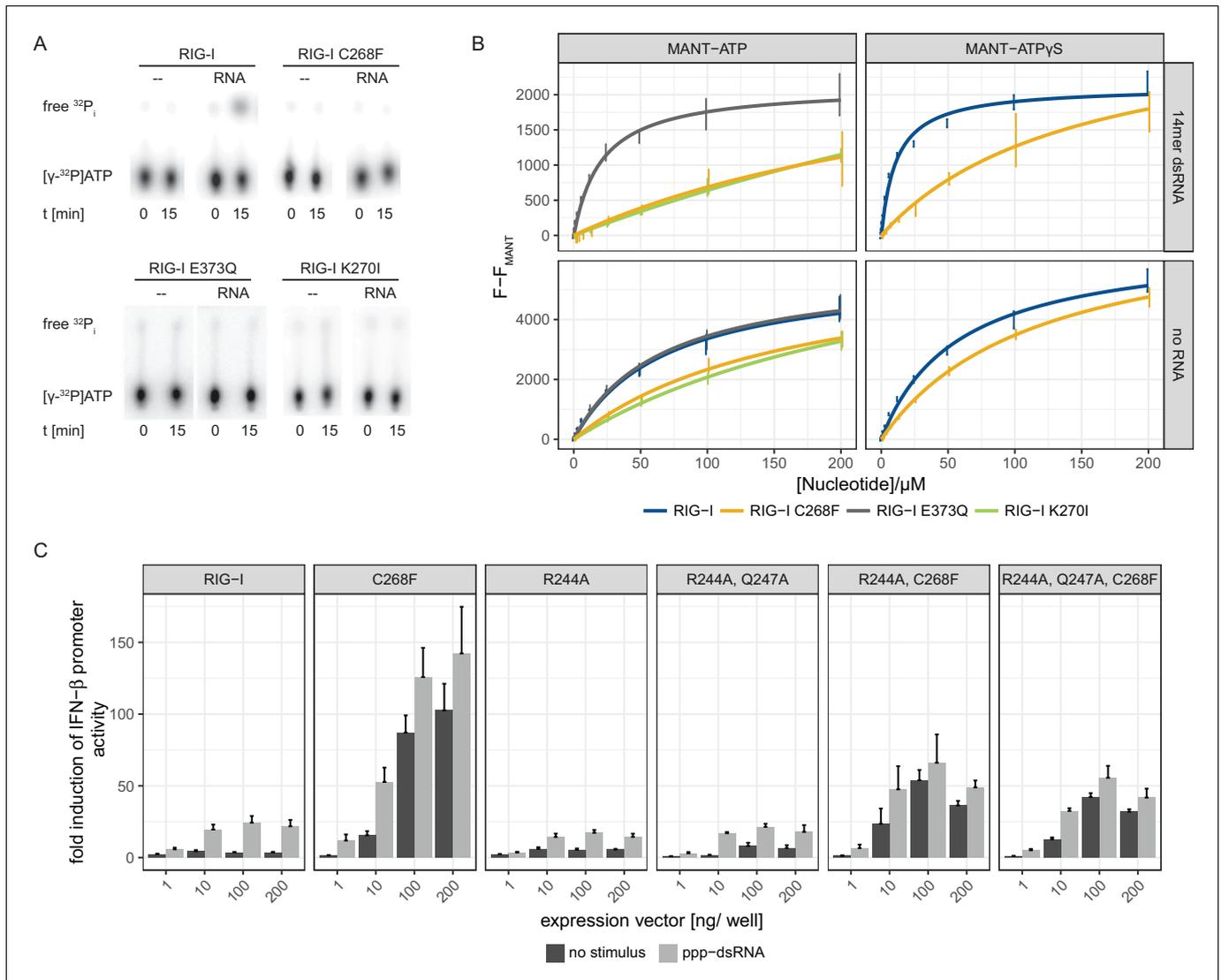
Figure 2. The RIG-I Singleton-Merten syndrome variant C268F is catalytically dead and has reduced ATP-binding-properties. (**A**) ATP hydrolysis activity of RIG-I, the RIG-I Singleton-Merten syndrome (SMS) variant C268F and the RIG-I motif I and II mutants K270I and E373Q. RIG-I proteins were incubated with [γ-^{32}P]-ATP in the presence or absence of a 12mer dsRNA for 15 min at room temperature and free phosphate was separated from ATP by thin layer chromatography. (**B**) Affinity of RIG-I, RIG-I C268F and the RIG-I motif I and II mutants to MANT-ATP or MANT-ATPγS measured by tryptophan fluorescence Förster resonance energy transfer to the MANT-nucleotide. Proteins were incubated with increasing amounts of nucleotides in the presence or absence of a 14mer dsRNA. MANT fluorescence was recorded minus a MANT-nucleotide-only control. n = 4, error bars represent mean values ± SD. (**C**) Fold change of interferon (IFN)-β promoter-driven luciferase activity in uninfected HEK293T RIG-I KO cells or in cells stimulated with a 19mer 5'-triphosphate (ppp)-dsRNA upon overexpression of different RIG-I mutants. Cells were co-transfected with RIG-I expression vectors and p-125luc/pGL4.74 reporter plasmids, and stimulated with ppp-dsRNA 6 hr post transfection. Firefly luciferase activities were determined in respect to Renilla luciferase activities 16 hr after RNA stimulation. All ratios were normalized to an empty vector control. n = 4–12, error bars represent mean values + SEM.

The following figure supplement is available for figure 2:

Figure supplement 1. Location of RIG-I amino-acid substitutions used in *Figure 2*.

Immunology and Inflammation | Structural Biology and Molecular Biophysics

Table 1. Affinities of different RIG-I mutants to MANT-ATP or MANT-ATPγS in the presence or absence of a 14mer dsRNA. n.d., not determined, n.f., no fit possible as no saturation was reached.

Protein	MANT-ATP	MANT-ATPγS
RIG-I	72 ± 13 µM	58 ± 7 µM
RIG-I + RNA	n.d.	11 ± 1 µM
RIG-I E373Q	72 ± 13 µM	n.d
RIG-I E373Q + RNA	28 ± 5 µM	n.d
RIG-I K270I	298 ± 81 µM	n.d
RIG-I K270I + RNA	n.f.	n.d
RIG-I C268F	166 ± 34 µM	116 ± 13 µM
RIG-I C268F + RNA	n.f.	147 ± 55 µM

DOI: https://doi.org/10.7554/eLife.38958.007

where the Nε of K270 is situated at a site normally occupied by the Mg^{2+} ion (*Video 1*). As a result, K270 now forms a salt bridge with E702 from the C-terminal RecA-like domain 2A, which in turn occupies the ATP γ-phosphate binding site. The resulting overall conformation resembles the ATP-bound state of RIG-I, but without ATP, and could thus explain how C268F is able to signal independently of any nucleotide. In order to clarify the impact of the salt bridge, we mutated both side chains in RIG-I C268F and analyzed the resulting double-mutants in our IFN-β promoter activity assay (*Figure 3B*). We expected that a disturbance of the salt-bridge formation in RIG-I C268F would lead to a loss of autoimmune signaling. Indeed, mutation of K270 in RecA-like domain A1 renders RIG-I C268F inactive, probably as the result of a mixture of (i) prevention of the activation of RIG-I C268F in the absence of ATP by disrupting the salt bridge, and/or (ii) the failure to bind ATP altogether through impaired ATP phosphate coordination and thus an impaired 2CARD release. In contrast to our ATP-binding triple mutant R244A, Q247A, C268F, which still allows formation of the salt bridge in the absence of ATP, mutation of motif I K270 in RIG-I C268F thus renders the protein inactive. Mutation of E702 in the RecA-like domain 2A of RIG-I C268F, by contrast, does not disrupt constitutive signaling of the protein, although this activity is at a reduced level compared to that of RIG-I C268F. An explanation for this might be E702's localization within the SF2 helicase motif V, which couples RNA-binding-induced ATP hydrolysis with movement on dsRNA. Similar to our previously described V699A mutant in the same motif (*Lässig et al., 2015*), E702A alone already induces an autoimmune phenotype similar to that induced by RIG-I C268F, E702A. The most plausible explanation for these data is that RIG-I C268F stabilizes an ATP-like state in the absence of ATP, but still allows formation of a proper ATP-bound state. Although E702A could reduce the former (by disrupting the salt bridge), it might stabilize the latter by sterically allowing ATP binding or by increasing the interaction with RNA.

In summary, our data suggest that the RIG-I C268F SMS mutation stabilizes the signal-on state of RIG-I in the presence of RNA but absence of ATP through a salt bridge between K270 from domain 1A and E702 from 2A. Unlike wild type RIG-I, which requires both RNA and ATP bound in order to be activated for downstream signaling, RIG-I C268F can signal independently of ATP (*Figure 2C*) in the presence of

Video 1. Crystal structure of RIG-I Δ2CARD C268F and close-up of the active site. The Singleton-Merten syndrome (SMS) mutation F268, as well as K270 and E702, are represented by a stick model. Theoretic locations of ADP·BeF_3 and Mg^{2+} are indicated in faint sticks and spheres, respectively, according to a superposition with RIG-I Δ2CARD in complex with RNA and nucleotide analogue (PDB 5E3H). K270 is located at the Mg^{2+}-binding site, whereas E702 occupies the BeF_3 (ATP γ-phosphate) position.
DOI: https://doi.org/10.7554/eLife.38958.010

|74|

Figure 3. The RIG-I Singleton-Merten syndrome variant C268F induces amino acid side chain rearrangements within the active site that interfere with nucleotide binding. (**A**) ATP-binding pockets of the RIG-I Singleton-Merten syndrome (SMS) variant C268F (left and middle panels) and the RIG-I wild type (right panel) bound to a 14mer dsRNA. The RIG-I SF2 sub-domains are colored in light gray or light blue (1A and 2A) and dark blue (2B). The RD is depicted in cyan and 2CARD is indicated in yellow. (**B**) Fold change of interferon (IFN)-β promoter-driven luciferase activity in uninfected HEK293T RIG-I KO cells or in cells stimulated with a 19mer 5′-triphosphate (ppp)-dsRNA upon overexpression of different RIG-I mutants. Cells were co-transfected with RIG-I expression vectors and p-125luc/pGL4.74 reporter plasmids, and stimulated with ppp-dsRNA 6 hr post transfection. Firefly luciferase activities were determined in respect to Renilla luciferase activities 16 hr after RNA stimulation. All ratios were normalized to an empty vector control. n = 4–12, error bars represent mean values + SEM.

DOI: https://doi.org/10.7554/eLife.38958.008

The following figure supplement is available for figure 3:

Figure supplement 1. Structural comparison of the RIG-I Singleton-Merten syndrome variant C268F with wild type RIG-I.

DOI: https://doi.org/10.7554/eLife.38958.009
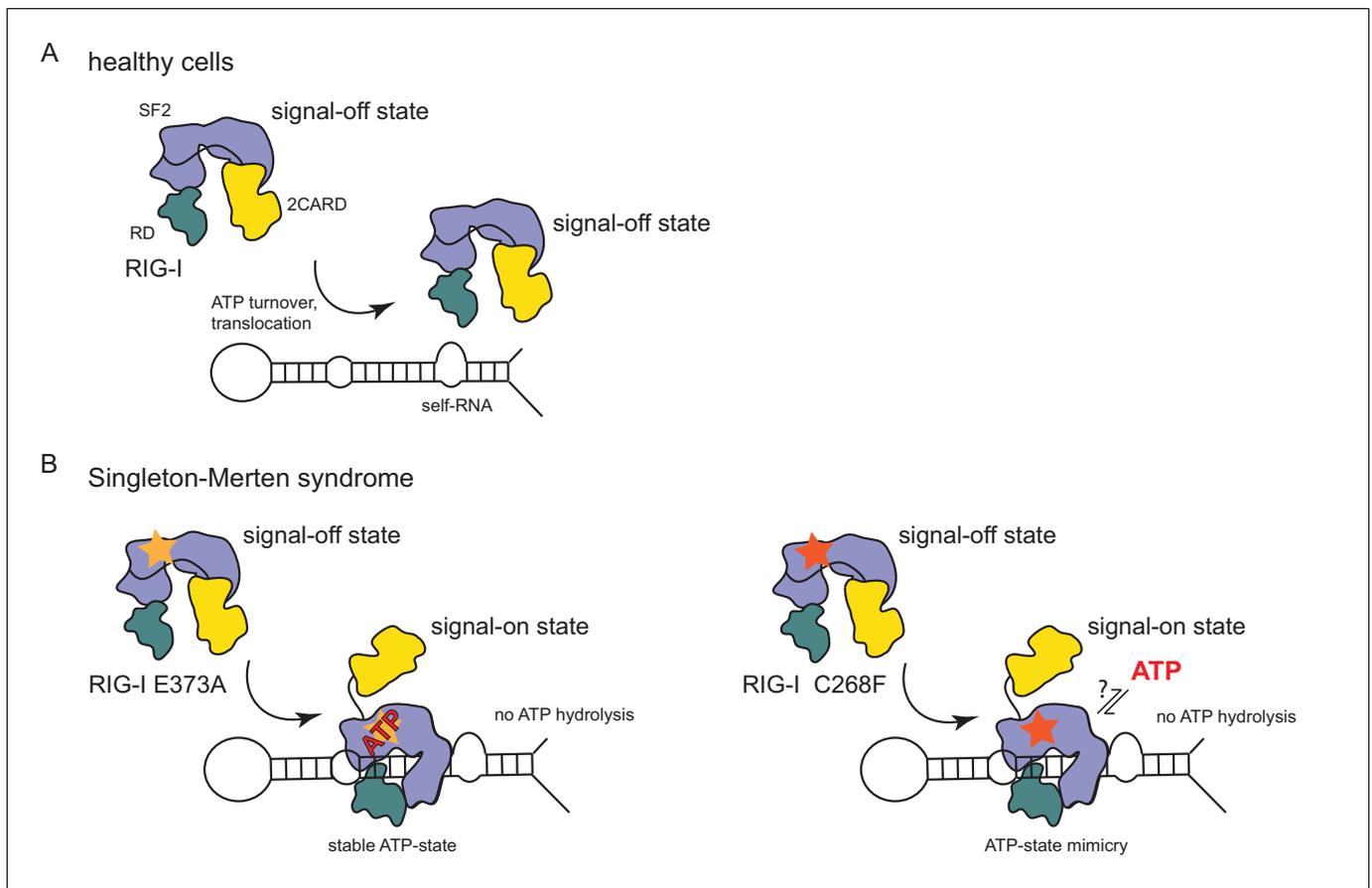
Figure 4. Model for the impact of Singleton-Merten syndrome mutations on self-RNA-induced RIG-I signaling. (**A**) In healthy cells, wild type RIG-I occurs in a signal-off state in which 2CARD is shielded by binding to the insertion domain of SF2. Binding of RIG-I to self-RNAs is efficiently prevented through ATP-turnover-induced dissociation (for a detailed model on self- vs non-self RNA discrimination see also *Lässig et al. (2015)*. (**B**) RIG-I Singleton-Merten syndrome (SMS) mutations either slow down ATP hydrolysis and stabilize the ATP-state (E373A, left side) or mimic the ATP-bound state (C268F, right side), and thus allow formation of the RIG-I signal-on state. In both cases, loss of ATP hydrolysis enhances the interaction with self-RNA and therefore results in pathogenic signaling. SMS mutations are indicated with a yellow or orange star.
DOI: https://doi.org/10.7554/eLife.38958.011

RNA (*Figure 1A*). However, our data also show that ATP further increases the binding of RIG-I C268F to internal dsRNA stems (*Figure 1B*) and that the engineering of ATP-binding mutations R244A, Q247A into RIG-I C268F at least partly reduces its signaling activity in response to endogenous or ppp-dsRNA RNA (*Figure 2C*). Even though we were not able to co-crystalize RIG-I C268F in the presence of a nucleotide, it is very possible that binding of ATP at high molecular concentrations, such as those present in a cellular context, further contributes to the pathogenic signaling activity of the protein. In principle, ATP binding would be sterically allowed if the salt-bridge-forming residue E702 occupies a site as in wtRIG-I. It is not yet clear what happens at the Mg^{2+}-binding site, as sterically F268 would not allow a canonical positioning of motif I K270 and might thus prevent binding of Mg^{2+}. Perhaps K270 remains at the displaced site even in the presence of ATP and simply 'mimics' Mg^{2+}. Such a scenario could still lead to reduced binding of ATP but would prevent ATP hydrolysis, as observed, because of a substantially altered charge distribution. This mechanism unifies the molecular basis for the development of SMS by both RIG-I variants (*Figure 4*). Unlike wtRIG-I, which consumes ATP and uses ATP turnover to decrease its affinity to self-RNA (*Figure 4A*), both motif I and II SMS mutations either mimic or freeze RIG-I in an ATP-bound state (*Figure 4B*). As a consequence, the inability to hydrolyze ATP and thus to dissociate actively from

eLIFE Research advance

RNA (*Figure 1B*, *Figure 2A*) traps the protein in an activated state and thus explains autoimmune signaling in response to self-RNA.

In conclusion, we provide for the first time a detailed biochemical and structural analysis of an RLR autoimmune disease variant. Slight intramolecular rearrangements within the RIG-I C268F ATP-binding pocket appear to compensate for ATP binding and render the protein active in the presence of dsRNA only. At the same time, loss of proof-reading activity leads to increased activation by endogenous RNA. This unusual gain-of-function mutation reveals, for the first time to our knowledge, that a P-loop mutation can mimic the effects of ATP binding.

Materials and methods

Key resources table

Reagent type (species) or source	Designation	Source or reference	Identifiers	Additional information
Cell line (human)	HEK293T RIG-I KO	*Zhu et al. (2014)*		Growth in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) as monolayer
Strain, strain background (*Escherichia coli*)	BL21 (DE3) Rosetta	Novagene		
Strain, strain background (*Escherichia coli*)	DH10multiBac	GenevaBiotech		
Strain, strain background (*Spodoptera frugipeda*)	Sf21 insect cells	Thermo Fisher Scientific	11497013	Growth in SF-900 III serum-free medium
Strain, strain background (*Trichoplusia ni*)	High Five insect cells	Thermo Fisher Scientific	B85502	Growth in Express Five serum-free medium supplemented with 10 mM L-glutamine
Recombinant DNA reagent	pcDNA5/FRT/TO	Thermo Fisher Scientific	V652020	
Recombinant DNA reagent	pcDNA5/FRT/TO-FLAG/HA-RIG-I and various mutants of the same construct	*Lässig et al. (2015)* and this paper		Progenitors: PCR, DDX58 (cDNA) and pcDNA5/FRT/TO
Recombinant DNA reagent	p-125luc	*Yoneyama et al. (1996)*		Firefly luciferase controlled by an interferon-β promoter
Recombinant DNA reagent	pGL4.74	Promega	E6921	Constitutive expression of a Renilla luciferase
Recombinant DNA reagent	pFBDM	*Berger et al. (2004)*		
Recombinant DNA reagent	pFBDM-His-RIG-I and various mutants of the same construct	*Lässig et al. (2015)* and this paper		Progenitors: PCR, DDX58 (cDNA) and pFBDM
Recombinant DNA reagent	pETM11-SUMO3GFP	EMBL Heidelberg, H. Besir	https://www.embl.de/pepcore/pepcore_services/cloning/sumo/	
Recombinant DNA reagent	pETM11-SUMO3-RIG-I-Δ2CARD-C268F	This paper		Progenitors: PCR, DDX58 (cDNA) and pETM11-SUMO3GFP
Sequence-based reagent	19mer 5' triphosphate dsRNA	InvivoGen	tlrl-3prna	1 µg/mL, 5'-pppGCAUGCGACCUCUGUUUGA-3
Sequence-based reagent	14mer dsRNA	Dharmacon		5'-CGACGCUAGCGUCG-3'
Sequence-based reagent	Cy3-hpRNA	Biomers		5'-Cy3-CCACCCGCCCCCCUAGUGAGGGGGGCGGGCC-3'
Chemical compound, drug	Lipofectamine 2000	Thermo Fisher Scientific	11668019	Used at 2.5x excess compared to RNA/DNA mass

Continued on next page

Continued

Reagent type (species) or source	Designation	Source or reference	Identifiers	Additional information
Chemical compound, drug	MANT-ATP	Jena Bioscience	NU-202	
Chemical compound, drug	MANT-ATPγS	Jena Bioscience	NU-232	
Chemical compound, drug	[γ-^{32}P]ATP	Hartmann Analytic	SRP-301	10 nM spiked with 3 mM unlabeled ATP
Commercial assay or kit	Dual-Luciferase Reporter Assay System	Promega	E1910	
Software, algorithm	XDS, XSCALE	*Kabsch (2010)*	http://xds.mpimf-heidelberg.mpg.de/	
Software, algorithm	PHASER	*McCoy et al. (2007)*; *Winn et al. (2011)*	http://www.ccp4.ac.uk/	
Software, algorithm	Coot	*Emsley et al. (2010)*	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/	
Software, algorithm	PHENIX	*Afonine et al. (2012)*	https://www.phenix-online.org/	
Software, algorithm	Pymol	Schrödinger	https://pymol.org/2/	
Software, algorithm	R	*R Development Core Team (2013)*	https://www.r-project.org/	

Vectors and cell lines

Sequences encoding full-length (1-925) or N-terminal truncated (230–925 or 232–925) human RIG-I were cloned into either pcDNA5/FRT/TO (purchased from Thermo Fisher Scientific, Waltham, MA; for expression in human cells), pFBDM (for expression in insect cells) (*Berger et al., 2004*) or pETM11-SUMO3 (EMBL, Heidelberg, Germany; for expression in *E. coli*). All proteins that were overexpressed in human cells contained an N-terminal FLAG/HA-tag, whereas proteins purified from insect cells contained an N-terminal His-tag.

Mutants were generated by site-directed mutagenesis using the QuikChange protocol and PfuUltra polymerase (Agilent, Santa Clara, CA).

HEK293T RIG-I KO cells (*Zhu et al., 2014*) were maintained in high glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with GlutaMAX, pyruvate and 10% fetal bovine serum (FBS) (all purchased from Thermo Fisher Scientific, Waltham, MA) at 37°C/ 5% CO_2 and were regularly tested by PCR for potential mycoplasma contaminations. *Spodoptera frugipeda* Sf21 and *Trichoplusia ni* High Five insect cells were maintained at 27°C/ 150 rpm in SF-900 III serum-free medium and High Five serum-free medium supplemented with 10 mM L-glutamine, respectively (both purchased from Thermo Fisher Scientific, Waltham, MA).

Luciferase reporter-gene assays

Transfection-based reporter gene assays in HEK293T RIG-I KO cells were carried out in 96-well tissue culture plates seeded with 0.2×10^5 cells one day prior to transfection. Cells were transfected with 25 ng p-125Luc (inducible-expression of a Firefly luciferase controlled by an interferon-β promoter) (*Yoneyama et al., 1996*), 5 ng pGL4.74 (constitutive-expression of Renilla luciferase, Promega, Madison, WI) and varying concentrations of FLAG/HA-RIG-I plasmids (1–200 ng) in a total amount of 300 ng DNA per well (filled up with empty pcDNA5 FRT/TO) using OptiMEM and Lipofectamine 2000 (both Thermo Fisher Scientific, Waltham, MA) according to the vendor's protocol. After 6 hr, cells were either stimulated by transfection of 1 µg/mL 19mer 5'triphosphate-dsRNA (InvivoGen, San Diego, CA) in OptiMEM using Lipofectamine 2000 or the respective amount of OptiMEM alone was added. Cells were harvested 16 hr after RNA stimulation in 50 µL passive lysis buffer (Promega, Madison, WI) and frozen at −20°C until they were used. Luciferase activities were determined with a Berthold Luminometer in black 96-well plates using 20 µL cell lysate and the Dual-Luciferase Reporter Assay System (Promega, Madison, WI).

For competition assays of FLAG/HA-RIG-I C268F with N-terminally shortened FLAG/HA-RIG-I Δ2CARD (230-925), cells were seeded as described above. Cells were transfected with 25 ng p-125Luc, 5 ng pGL4.74, 75 ng FLAG/HA-RIG-I C268F plasmid and varying concentrations (1–195 ng) of competitor plasmids containing N-terminal truncated RIG-I in a total of 300 ng DNA per well (filled up with empty pcDNA5 FRT/TO) using OptiMEM and Lipofectamine 2000 as transfection reagent according to the vendor's protocol. Cells were harvested 16 hr after transfection in 50 μL passive lysis buffer and frozen at −20°C. Luciferase activities were determined as described above.

All cell-based assays were performed at least four times in independent experiments and are represented as mean values + SEM.

Protein expression and purification

Recombinant full-length human RIG-I or RIG-I single amino acid mutants (1–925, N-terminal His-tag) were produced in and purified from High Five insect cells as described before (*Cui et al., 2008*; *Lässig et al., 2015*; *Rawling et al., 2015*). Briefly, the open reading frame for human RIG-I was cloned into the pFBDM vector, transformed into *E. coli* DH10MultiBac, and extracted Baculovirus DNA was then transfected into SF21 insect cells. Baculoviruses were propagated twice in SF21 insect cells and subsequently used for infection of High Five insect cells. Expression was carried out for 3 days at 27°C. Harvested cells were shock frozen in liquid nitrogen and stored at −20°C until they were used. For purification, cells were resuspended in lysis buffer (25 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol, pH 7) and lyzed by sonication. Cleared lysate was loaded onto Ni-NTA agarose resin (Qiagan, Hilden, Germany), washed with lysis buffer containing 300 mM NaCl and eluted in elution buffer (25 mM HEPES, 100 mM NaCl, 200 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol, pH 7). Proteins were further purified on a HiTrap Heparin HP column (GE Healthcare, Little Chalfont, UK) in 25 mM HEPES, 10% glycerol, 5 mM β-mercaptoethanol, pH 7 using a linear salt gradient ranging from 100 mM to 1 M NaCl. Finally, fractions containing RIG-I were pooled and loaded onto a HiLoad Superdex 200 16/60 size exclusion column (GE Healthcare, Little Chalfont, UK) using gel filtration buffer (25 mM HEPES, 150 mM NaCl, 5 mM MgCl$_2$, 5% glycerol, 5 mM β-mercaptoethanol, pH 7). Monomeric RIG-I was concentrated to ~6 mg/mL, flash frozen in liquid nitrogen and stored at −80°C. N-terminally truncated RIG-I Δ2CARD C268F (232–925, N-terminal His-tag) expressed in High Five cells was purified as described above.

In addition, N-terminally truncated RIG-I Δ2CARD C268F (232–925, N-terminal His-Sumo3-tag) was produced in and purified from *E. coli* Rosetta (DE3). Cells were induced with 0.2 mM DTT at an OD$_{600}$ of 0.6–0.8 and protein was expressed at 18°C overnight. Harvested cells were shock frozen in liquid nitrogen and stored at −20°C. Protein was purified as described above, except that after metal affinity chromatography, the His-SUMO-tag was cleaved off by adding SenP2 protease (mass ratio 1:500) to pooled eluate fractions during a dialysis step against elution buffer without imidazole (25 mM HEPES, 100 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, pH 7) that was carried out overnight at 4°C. Cleaved protein was separated from the tag during a second metal-affinity chromatography by step-wise elution with elution buffers containing 20 mM and 40 mM imidazole, and subjected to heparin-affinity and gel-filtration chromatography as described above.

For crystallization, N-terminally truncated RIG-I Δ2CARD C268F (purified either from insect cells or from *E. coli*) was concentrated to ~25 mg/mL.

Protein crystalization

Co-crystallization of RIG-I Δ2CARD C268F (170 μM) with equimolar concentrations of 14mer dsRNA (5'-CGACGCUAGCGUCG-3', palindromic RNA, purchased from Dharmacon, Lafayette, CO) was done either in the presence of 2 mM ADP, 2 mM BeCl$_2$ and 10 mM NaF to reconstitute ADP·BeF$_x$ (crystal 1) or with 2 mM ATP (crystal 2) by hanging-drop vapor diffusion at 20°C. In each case, 2.5 μL protein/RNA/nucleotide mix was added to 2.5 μL reservoir solution from a total reservoir volume of 400 μL per well. Crystal 1 was raised in wells containing 0.1 M MOPS pH 7.5, 15% (w/v) PEG 3350, 0.125 M NaSCN and 3% (v/v) 2,2,2-trifluoroethanol as reservoir solution. The reservoir of crystal 2 contained 0.1 M MOPS pH 7.5, 17.5% (w/v) PEG 3350, 0.25 M NaSCN and 3% (v/v) 2,2,2-trifluoroethanol. Crystals appeared after 1–2 days and were transferred into the respective reservoir solutions containing 10% (v/v) 2,3-butanediol as cryoprotectant, flash-frozen and stored in liquid nitrogen.

eLIFE Research advance

Data collection and structure determination

X-ray diffraction data were collected at the SLS X06SA beamline (Swiss Light Source, Villigen, Switzerland). Diffraction datasets from both crystals were indexed and integrated using XDS and scaled with XSCALE (*Kabsch, 2010*). Crystal 1 had space group $P2_12_12_1$ and diffracted to 3.3 Å, crystal 2 had space group $P6_522$ and diffracted until 2.9 Å (*Table 2*). Diffraction data from crystal 1 were used to determine an initial structure of RIG-I Δ2CARD C268F by molecular replacement using PHASER (*McCoy et al., 2007*; *Winn et al., 2011*) and a search model based on a published structure of RIG-I Δ2CARD (PDB entry 5E3H) (*Jiang et al., 2011*). The initial model was created in two iterative rounds of manual model building and refinement using Coot and PHENIX (*Afonine et al., 2012*; *Emsley et al., 2010*). This model was used to phase the second, better-diffracting dataset from crystal 2 using PHASER. The final structure was built and refined in several iterative rounds using Coot and PHENIX. The statistics describing both structures are shown in (*Table 2*). We did not detect any density for a bound nucleotide in either structure.

Table 2. Data collection and refinement statistics.
Values in parentheses are for the highest resolution shell.

	Crystal 1	Crystal 2
PDB code		6GPG
Data collection		
Space group	$P2_12_12_1$	$P6_522$
Wavelength (Å)	1.00	1.00
Cell dimensions		
a, b, c (Å)	112.1, 177.1, 314.8	175.6, 175.6, 109.5
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution range (Å)	47.2–3.3 (3.42–3.30)	46.4–2.9 (3.00–2.89)
R_{merge} (%)	14.3 (112)	7.6 (206)
I/σI	8.45 (1.28)	19.72 (1.15)
$CC_{1/2}$	99.8 (67.6)	99.9 (99.7)
Completeness (%)	95.3 (79.7)	99.7 (97.4)
Redundancy	3.38 (2.91)	13.09 (13.21)
Refinement		
Resolution (Å)	3.3	2.9
No. reflections	90,121	22,649
R_{work}/ R_{free}	22.8/28.3	21.4/25.9
No. atoms		
Macromolecules	35,730	5,810
Ions	10	2
Ramachandran statistics		
Favoured (%)	92.78	92.71
Allowed (%)	6.39	6.98
Outliers (%)	0.83	0.31
R.M.S deviations		
Bond lengths (Å)	0.011	0.009
Angles (°)	1.48	1.43
B-factors		
Macromolecules	109.98	139.89
Ions	105.23	121.74

Figures and movies were created with PyMOL (*Schrödinger, 2015*).

ATP hydrolysis assays

ATP hydrolysis activities of different full-length His-RIG-I constructs were determined using [γ-^{32}P] ATP (Hartmann Analytic, Braunschweig, Germany). 100 nM protein was pre-incubated with 100 nM 12mer 5'triphosphate-dsRNA for 10 min at room temperature in hydrolysis buffer (25 mM HEPES, 50 mM KCl, 5 mM MgCl$_2$, 5 mM TCEP, pH 7.5). The reaction was initiated by the addition of 3 mM unlabeled and 10 nM [γ-^{32}P]ATP and incubated for 15 min at 37°C. Free phosphate was separated from ATP by thin layer chromatography (TLC) in TLC running buffer (1 M formic acid, 0.5 M LiCl) on polyethyleneimine cellulose TLC plates (Sigma-Aldrich, St. Louis, MO). [γ-^{32}P]P$_i$ and [γ-^{32}P]ATP were detected using the Typhoon^{TM} FLA 9500 phosphor-imaging system (GE Healthcare, Little Chalfont, UK).

Fluorescence anisotropy

Affinities of full-length RIG-I or RIG-I C268F (1–925, N-terminal His-tag) to a Cy3-labeled hpRNA (5'-Cy3-CCACCCGCCCCCCUAGUGAGGGGGGCGGGCC-3', purchased from Biomers, Ulm, Germany) were determined using fluorescence anisotropy. All samples were prepared in 96-well black chimney microplates (Greiner Bio-One, Kremsmünster, Austria). 10 nM RNA was pre-incubated with different protein concentrations (2.4 nM – 5 μM) for 15 min at room temperature in assay buffer (25 mM HEPES, 50 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, pH 7). Fluorescence anisotropy was measured after the addition of 5 mM ATP using a TECAN M1000 microplate reader (Tecan, Männedorf, Switzerland) at λ$_{ex}$/$_{em}$ = 530/570 nm and gain = 130 during a time course of 25 min in 1 min intervals. Assays were performed four times in independent experiments using the same protein purification batch. For determination of affinities, anisotropy values between 15 and 20 min measuring time were averaged and fit to the following single-site binding model using R (*R Development Core Team, 2013*):

$$y = B_{max} \frac{[P]}{[P] + K_D} \tag{1}$$

where y is the observed anisotropy, [P] is the protein concentration, B$_{max}$ is the maximal anisotropy and K$_D$ is the dissociation constant. Fitting was performed globally on all available datasets. Representative values in figures are mean values ± SD.

MANT-ATP binding

Binding of MANT-ATP and MANT-ATPγS to different full-length RIG-I mutants (1–925, N-terminal His-tag) was determined via Förster resonance energy transfer from RIG-I to MANT-ATP (Jena Bioscience, Jena, Germany). All samples were prepared in 96-well black chimney microplates. 2.2 μM protein and equimolar concentrations of a 14mer dsRNA (5'-CGACGCUAGCGUCG-3', see 'Protein crystalization') were pre-incubated with different MANT-ATP concentrations (0.2 μM – 200 μM) for 15 min at room temperature in assay buffer (25 mM HEPES, 50 mM NaCl, 5 mM MgCl$_2$, 1 mM DTT, pH 7). Fluorescence of MANT-ATP was measured in a TECAN M1000 microplate reader at λ$_{ex}$/$_{em}$ = 290/448 nm, gain = 170 using an average of five reads per well. Assays were performed four times in independent experiments using the same protein purification batch. For determination of affinities fluorescence values were fitted to *Equation (1)* using R, where y is the fluorescence, [P] is the protein concentration, B$_{max}$ is the maximal fluorescence and K$_D$ is the dissociation constant. Fitting was performed globally on all available datasets. Representative values in figures are mean values ± SD.

eLIFE Research advance

Additional information

Funding

Funder	Grant reference number	Author
Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst	BioSysNet	Karl-Peter Hopfner
Deutsche Forschungsgemeinschaft	CIPSM	Karl-Peter Hopfner
Deutsche Forschungsgemeinschaft	HO2489/8	Karl-Peter Hopfner
Deutsche Forschungsgemeinschaft	CRC1054 project B02	Katja Lammens
Deutsche Forschungsgemeinschaft	CRC/TRR 237	Karl-Peter Hopfner

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Charlotte Lässig, Conceptualization, Formal analysis, Supervision, Validation, Investigation, Visualization, Methodology, Writing—original draft, Writing—review and editing; Katja Lammens, Formal analysis, Supervision, Investigation, Methodology; Jacob Lucián Gorenflos López, Sebastian Michalski, Investigation, Writing—review and editing; Olga Fettscher, Investigation, Cloned and tested mutant proteins for cell-based assays; Karl-Peter Hopfner, Conceptualization, Supervision, Funding acquisition, Writing—original draft, Project administration, Writing—review and editing

Author ORCIDs

Charlotte Lässig (iD) http://orcid.org/0000-0001-6253-7880
Katja Lammens (iD) https://orcid.org/0000-0002-4438-1381
Karl-Peter Hopfner (iD) http://orcid.org/0000-0002-4528-8357

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.38958.020
Author response https://doi.org/10.7554/eLife.38958.021

Additional files

Supplementary files

• Transparent reporting form
DOI: https://doi.org/10.7554/eLife.38958.013

Data availability

Diffraction data have been deposited in PDB under the accession code 6GPG.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Lässig C, Lammens K, Hopfner KP	2018	Structure of the RIG-I Singleton-Merten syndrome variant C268F	https://www.ebi.ac.uk/pdbe/entry/pdb/6gpg	Publicly available at the RCSB Protein Data Bank (accession no. 6GPG)

The following previously published dataset was used:

eLIFE Research advance

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Jiang F, Miller MT, Marcotrigiano J	2015	Structural Basis for RNA Recognition and Activation of RIG-I	https://www.rcsb.org/structure/5E3H	Publicly available at the RCSB Protein Data Bank (accession no. 5E3H)

References

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D Biological Crystallography* **68**:352–367. DOI: https://doi.org/10.1107/S0907444912001308, PMID: 22505256

Ahmad S, Mu X, Yang F, Greenwald E, Park JW, Jacob E, Zhang CZ, Hur S. 2018. Breaching Self-Tolerance to alu duplex RNA underlies MDA5-Mediated inflammation. *Cell* **172**:797–810. DOI: https://doi.org/10.1016/j.cell.2017.12.016, PMID: 29395326

Anchisi S, Guerra J, Garcin D. 2015. RIG-I ATPase activity and discrimination of self-RNA versus non-self-RNA. *mBio* **6**:e02349–02314. DOI: https://doi.org/10.1128/mBio.02349-14, PMID: 25736886

Berger I, Fitzgerald DJ, Richmond TJ. 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnology* **22**:1583–1587. DOI: https://doi.org/10.1038/nbt1036, PMID: 15568020

Chiang JJ, Sparrer KMJ, van Gent M, Lässig C, Huang T, Osterrieder N, Hopfner KP, Gack MU. 2018. Viral unmasking of cellular 5S rRNA pseudogene transcripts induces RIG-I-mediated immunity. *Nature Immunology* **19**:53–62. DOI: https://doi.org/10.1038/s41590-017-0005-y, PMID: 29180807

Chung H, Calis JJA, Wu X, Sun T, Yu Y, Sarbanes SL, Dao Thi VL, Shilvock AR, Hoffmann HH, Rosenberg BR, Rice CM. 2018. Human ADAR1 prevents endogenous RNA from triggering translational shutdown. *Cell* **172**:811–824. DOI: https://doi.org/10.1016/j.cell.2017.12.038, PMID: 29395325

Cui S, Eisenächer K, Kirchhofer A, Brzózka K, Lammens A, Lammens K, Fujita T, Conzelmann KK, Krug A, Hopfner KP. 2008. The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Molecular Cell* **29**:169–179. DOI: https://doi.org/10.1016/j.molcel.2007.10.032, PMID: 18243112

Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvänen AC, Rönnblom L, Behrens TW, Graham RR, Vyse TJ. 2011. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genetics* **7**:e1002341. DOI: https://doi.org/10.1371/journal.pgen.1002341, PMID: 22046141

Devarkar SC, Wang C, Miller MT, Ramanathan A, Jiang F, Khan AG, Patel SS, Marcotrigiano J. 2016. Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *PNAS* **113**:596–601. DOI: https://doi.org/10.1073/pnas.1515152113

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66**:486–501. DOI: https://doi.org/10.1107/S0907444910007493, PMID: 20383002

Fairman-Williams ME, Guenther UP, Jankowsky E. 2010. SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* **20**:313–324. DOI: https://doi.org/10.1016/j.sbi.2010.03.011, PMID: 20456941

Goubau D, Schlee M, Deddouche S, Pruijssers AJ, Zillinger T, Goldeck M, Schuberth C, Van der Veen AG, Fujimura T, Rehwinkel J, Iskarpatyoti JA, Barchet W, Ludwig J, Dermody TS, Hartmann G, Reis e Sousa C. 2014. Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5'-diphosphates. *Nature* **514**:372–375. DOI: https://doi.org/10.1038/nature13590, PMID: 25119032

Jang MA, Kim EK, Now H, Nguyen NT, Kim WJ, Yoo JY, Lee J, Jeong YM, Kim CH, Kim OH, Sohn S, Nam SH, Hong Y, Lee YS, Chang SA, Jang SY, Kim JW, Lee MS, Lim SY, Sung KS, et al. 2015. Mutations in DDX58, which encodes RIG-I, cause atypical Singleton-Merten syndrome. *The American Journal of Human Genetics* **96**:266–274. DOI: https://doi.org/10.1016/j.ajhg.2014.11.019, PMID: 25620203

Jiang F, Ramanathan A, Miller MT, Tang GQ, Gale M, Patel SS, Marcotrigiano J. 2011. Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479**:423–427. DOI: https://doi.org/10.1038/nature10537, PMID: 21947008

Jiang X, Kinch LN, Brautigam CA, Chen X, Du F, Grishin NV, Chen ZJ. 2012. Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* **36**:959–973. DOI: https://doi.org/10.1016/j.immuni.2012.03.022, PMID: 22705106

Kabsch W. 2010. XDS. *Acta Crystallographica. Section D, Biological Crystallography* **66**:125–132. DOI: https://doi.org/10.1107/S0907444909047337, PMID: 20124692

Kohlway A, Luo D, Rawling DC, Ding SC, Pyle AM. 2013. Defining the functional determinants for RNA surveillance by RIG-I. *EMBO reports* **14**:772–779. DOI: https://doi.org/10.1038/embor.2013.108, PMID: 23897087

Kowalinski E, Lunardi T, McCarthy AA, Louber J, Brunel J, Grigorov B, Gerlier D, Cusack S. 2011. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**:423–435. DOI: https://doi.org/10.1016/j.cell.2011.09.039, PMID: 22000019

Lässig C, Matheisl S, Sparrer KM, de Oliveira Mann CC, Moldt M, Patel JR, Goldeck M, Hartmann G, García-Sastre A, Hornung V, Conzelmann KK, Beckmann R, Hopfner KP. 2015. ATP hydrolysis by the viral RNA sensor

|83|

RIG-I prevents unintentional recognition of self-RNA. *eLife* **4**:e10859. DOI: https://doi.org/10.7554/eLife.10859, PMID: 26609812

Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, Li JB, Seeburg PH, Walkley CR. 2015. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**:1115–1120. DOI: https://doi.org/10.1126/science.aac7049, PMID: 26275108

Liu S, Wang H, Jin Y, Podolsky R, Reddy MV, Pedersen J, Bode B, Reed J, Steed D, Anderson S, Yang P, Muir A, Steed L, Hopkins D, Huang Y, Purohit S, Wang CY, Steck AK, Montemari A, Eisenbarth G, et al. 2009. IFIH1 polymorphisms are significantly associated with type 1 diabetes and IFIH1 gene expression in peripheral blood mononuclear cells. *Human Molecular Genetics* **18**:358–365. DOI: https://doi.org/10.1093/hmg/ddn342, PMID: 18927125

Louber J, Brunel J, Uchikawa E, Cusack S, Gerlier D. 2015. Kinetic discrimination of self/non-self RNA by the ATPase activity of RIG-I and MDA5. *BMC Biology* **13**:54. DOI: https://doi.org/10.1186/s12915-015-0166-9, PMID: 26215161

Luo D, Ding SC, Vela A, Kohlway A, Lindenbach BD, Pyle AM. 2011. Structural insights into RNA recognition by RIG-I. *Cell* **147**:409–422. DOI: https://doi.org/10.1016/j.cell.2011.09.023, PMID: 22000018

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658–674. DOI: https://doi.org/10.1107/S0021889807021206, PMID: 19461840

Myong S, Cui S, Cornish PV, Kirchhofer A, Gack MU, Jung JU, Hopfner KP, Ha T. 2009. Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science* **323**:1070–1074. DOI: https://doi.org/10.1126/science.1168352, PMID: 19119185

Oda H, Nakagawa K, Abe J, Awaya T, Funabiki M, Hijikata A, Nishikomori R, Funatsuka M, Ohshima Y, Sugawara Y, Yasumi T, Kato H, Shirai T, Ohara O, Fujita T, Heike T. 2014. Aicardi-Goutières syndrome is caused by IFIH1 mutations. *The American Journal of Human Genetics* **95**:121–125. DOI: https://doi.org/10.1016/j.ajhg.2014.06.007, PMID: 24995871

Peisley A, Wu B, Xu H, Chen ZJ, Hur S. 2014. Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I. *Nature* **509**:110–114. DOI: https://doi.org/10.1038/nature13140, PMID: 24590070

Pettersson M, Bergendal B, Norderyd J, Nilsson D, Anderlid BM, Nordgren A, Lindstrand A. 2017. Further evidence for specific IFIH1 mutation as a cause of Singleton-Merten syndrome with phenotypic heterogeneity. *American Journal of Medical Genetics Part A* **173**:1396–1399. DOI: https://doi.org/10.1002/ajmg.a.38214, PMID: 28319323

R Development Core Team. 2013. R: A language and environment for statistical computing.

Rawling DC, Fitzgerald ME, Pyle AM. 2015. Establishing the role of ATP for the function of the RIG-I innate immune sensor. *eLife* **4**:e09391. DOI: https://doi.org/10.7554/eLife.09391, PMID: 26371557

Rawling DC, Pyle AM. 2014. Parts, assembly and operation of the RIG-I family of motors. *Current Opinion in Structural Biology* **25**:25–33. DOI: https://doi.org/10.1016/j.sbi.2013.11.011, PMID: 24878341

Rice GI, Del Toro Duany Y, Jenkinson EM, Forte GM, Anderson BH, Ariaudo G, Bader-Meunier B, Baildam EM, Battini R, Beresford MW, Casarano M, Chouchane M, Cimaz R, Collins AE, Cordeiro NJ, Dale RC, Davidson JE, De Waele L, Desguerre I, Faivre L, et al. 2014. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nature Genetics* **46**:503–509. DOI: https://doi.org/10.1038/ng.2933, PMID: 24686847

Rutsch F, MacDougall M, Lu C, Buers I, Mamaeva O, Nitschke Y, Rice GI, Erlandsen H, Kehl HG, Thiele H, Nürnberg P, Höhne W, Crow YJ, Feigenbaum A, Hennekam RC. 2015. A specific IFIH1 gain-of-function mutation causes Singleton-Merten syndrome. *The American Journal of Human Genetics* **96**:275–282. DOI: https://doi.org/10.1016/j.ajhg.2014.12.014, PMID: 25620204

Schlee M, Roth A, Hornung V, Hagmann CA, Wimmenauer V, Barchet W, Coch C, Janke M, Mihailovic A, Wardle G, Juranek S, Kato H, Kawai T, Poeck H, Fitzgerald KA, Takeuchi O, Akira S, Tuschl T, Latz E, Ludwig J, et al. 2009. Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity* **31**:25–34. DOI: https://doi.org/10.1016/j.immuni.2009.05.008, PMID: 19576794

Schmidt A, Schwerd T, Hamm W, Hellmuth JC, Cui S, Wenzel M, Hoffmann FS, Michallet M-C, Besch R, Hopfner K-P, Endres S, Rothenfusser S. 2009. 5'-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. *PNAS* **106**:12067–12072. DOI: https://doi.org/10.1073/pnas.0900971106

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**:671–675. DOI: https://doi.org/10.1038/nmeth.2089, PMID: 22930834

Schrödinger L. 2015. *The PyMOL Molecular Graphics System*. 1.8.

Schuberth-Wagner C, Ludwig J, Bruder AK, Herzner AM, Zillinger T, Goldeck M, Schmidt T, Schmid-Burgk JL, Kerber R, Wolter S, Stümpel JP, Roth A, Bartok E, Drosten C, Coch C, Hornung V, Barchet W, Kümmerer BM, Hartmann G, Schlee M. 2015. A conserved histidine in the RNA sensor RIG-I controls immune tolerance to N1-2'O-Methylated Self RNA. *Immunity* **43**:41–51. DOI: https://doi.org/10.1016/j.immuni.2015.06.015, PMID: 26187414

Shah N, Beckham SA, Wilce JA, Wilce MCJ. 2018. Combined roles of ATP and small hairpin RNA in the activation of RIG-I revealed by solution-based analysis. *Nucleic Acids Research* **46**:3169–3186. DOI: https://doi.org/10.1093/nar/gkx1307, PMID: 29346611

Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C, Ionescu-Tirgoviste C, Widmer B, Dunger DB, Savage DA, Walker NM, Clayton DG, Todd JA. 2006. A genome-wide association study of nonsynonymous

eLIFE Research advance

SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nature Genetics* **38**: 617–619. DOI: https://doi.org/10.1038/ng1800, PMID: 16699517

Van Eyck L, De Somer L, Pombal D, Bornschein S, Frans G, Humblet-Baron S, Moens L, de Zegher F, Bossuyt X, Wouters C, Liston A. 2015. Brief report: ifih1 mutation causes systemic lupus erythematosus with selective IgA deficiency. *Arthritis & Rheumatology* **67**:1592–1597. DOI: https://doi.org/10.1002/art.39110, PMID: 25777993

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67**:235–242. DOI: https://doi.org/10.1107/S0907444910045749, PMID: 21460441

Wu B, Peisley A, Tetrault D, Li Z, Egelman EH, Magor KE, Walz T, Penczek PA, Hur S. 2014. Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Molecular Cell* **55**:511–523. DOI: https://doi.org/10.1016/j.molcel.2014.06.010, PMID: 25018021

Yao H, Dittmann M, Peisley A, Hoffmann HH, Gilmore RH, Schmidt T, Schmidt-Burgk J, Hornung V, Rice CM, Hur S. 2015. ATP-dependent effector-like functions of RIG-I-like receptors. *Molecular Cell* **58**:541–548. DOI: https://doi.org/10.1016/j.molcel.2015.03.014, PMID: 25891073

Yoneyama M, Suhara W, Fukuhara Y, Sato M, Ozato K, Fujita T. 1996. Autocrine amplification of type I interferon gene expression mediated by interferon stimulated gene factor 3 (ISGF3). *Journal of Biochemistry* **120**:160–169. DOI: https://doi.org/10.1093/oxfordjournals.jbchem.a021379, PMID: 8864859

Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, Foy E, Loo YM, Gale M, Akira S, Yonehara S, Kato A, Fujita T. 2005. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *The Journal of Immunology* **175**:2851–2858. DOI: https://doi.org/10.4049/jimmunol.175.5.2851, PMID: 16116171

Zheng J, Yong HY, Panutdaporn N, Liu C, Tang K, Luo D. 2015. High-resolution HDX-MS reveals distinct mechanisms of RNA recognition and activation by RIG-I and MDA5. *Nucleic Acids Research* **43**:1216–1230. DOI: https://doi.org/10.1093/nar/gku1329, PMID: 25539915

Zhu J, Zhang Y, Ghosh A, Cuevas RA, Forero A, Dhar J, Ibsen MS, Schmid-Burgk JL, Schmidt T, Ganapathiraju MK, Fujita T, Hartmann R, Barik S, Hornung V, Coyne CB, Sarkar SN. 2014. Antiviral activity of human OASL protein is mediated by enhancing signaling of the RIG-I RNA sensor. *Immunity* **40**:936–948. DOI: https://doi.org/10.1016/j.immuni.2014.05.007, PMID: 24931123

4 Discussion

4.1 cGAS as nuclear protein

Since cGAS´ discovery as a cytosolic DNA sensor, its localization was proposed to be exclusively cytosolic, with this strict compartmentalization leading to the prevention of autoimmunity[294]. In recent times a growing number of studies, describe cGAS as a predominantly nuclear protein that shows its main catalytic function when located in the cytosol[190,191,194,198,266,295–297]. A nuclear localization must already be assumed, based on the need of cGAS for sensing internuclear replicating viruses such as herpesviruses[298] or retroviruses (like HIV), that shield their DNA during reverse transcription in the cytosol[299,300] or simply during mitosis in healthy cells.

4.1.1 Nuclear cGAS localization – tethered to nucleosomes

Although the exact interaction of cGAS with nuclear DNA was unclear prior to the findings described in this work, several interactions with chromatin-related structures have been described. Overexpressed cGAS was shown to colocalize with chromatin in mouse embryonic fibroblasts (MEFs) during mitosis[187] and to bind to heterochromatin marker-positive cytoplasmic chromatin fragments in senescent cells[199]. The same association with chromatin fragments, upon nuclear membrane rupture can be seen with the colocalization of cGAS with micronuclei[301]. All of these data already indicated that an association with nuclear self-DNA can at least occur in a pathogenic setting. While most of these studies used overexpressed cGAS for their experiments, also endogenous levels of cGAS have now been shown to colocalize with chromatin. In fact, most endogenous cGAS was found in the nucleus and will only enter the cytosol upon cytosolic DNA infection. Additionally, its nuclear localization was cell cycle unspecific. Strikingly, in this study cGAS was tethered very tightly to chromatin, suggesting that due to the relatively low affinity to DNA, the DNA itself cannot be the site of interaction[190]. It was shown, that cGAS is binding to nucleosomes rendering it inactive[189]. The interaction site, was identified to be within the catalytic domain of cGAS where arginine 241 (for mcGAS) was especially needed for chromatin tethering[190] and nucleosome binding was dependent on the existence of the acidic patch of the nucleosome[189]. Our findings combine the observed effects and structurally explain the exact mechanism of cGAS tethering to chromatin and thereby its inhibition preventing autoimmunity.

We could show that cGAS is tightly bound to the nucleosome´s acidic patch formed between histone H2A and H2B. The affinity of this interaction is manifolds higher (low nM range) than the described affinity towards DNA (1-2 µM[208]), explaining the previously described salt-resistance of the interaction. We confirmed the nuclear localization of cGAS and salt-resistant chromatin tethering in human cells (THP-1), mediated by its `arginine anchor´ (hcGAS K255; mcGAS R241). We could show that indeed nucleosomes without overhangs (ONO) are potent inhibitors of cGAS activity. This fact is explained by the binding-mode of cGAS to the nucleosome core. When bound to the nucleosome core, cGAS´ DNA-binding site B (see Fig. 16) is covered, preventing formation of an active complex. We tested the influence of mutations in the other DNA binding sites (site A and site C) but did not observe an effect for nucleosome binding. The acidic patches of the nucleosomes need to be accessible for cGAS to bind and act in an inhibitory manner. For a monomeric nucleosome core particle, two acidic patch interfaces are easily accessible. As we used equimolar amounts of cGAS in our structural studies we could observe a 1:1 cGAS/Nucleosome complex. Nevertheless, we observed a 2:1 complex in electromobility shift assays[302]. Several other groups that published structures related to ours also observed higher order complexes[303–307]. For all structures the tight binding of cGAS via its `arginine anchor´ to the acidic patch is the same.
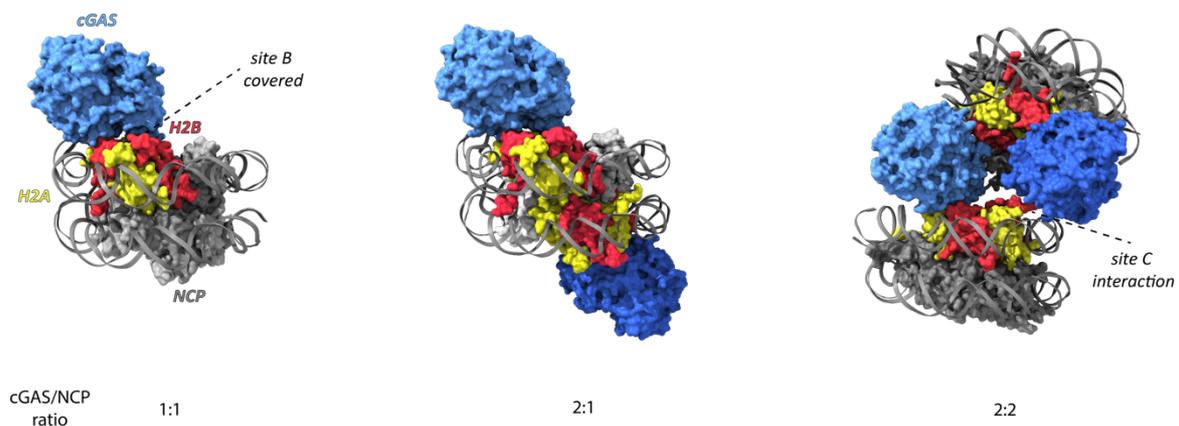
Figure 16: Overview of different cGAS-NCP complexes. Regardless of the exact stoichiometry cGAS binds to the acidic patch formed between histone H2A and H2B via its conserved `arginine anchor´. By this tethering DNA binding site B is blocked, preventing association of the active complex. As we used equimolar ratios of cGAS and NCP in our study, we observed only the 1:1 complex. As the nucleosome is symmetrical also a 2:1 complex can be obtained. For the human protein, harboring the largest site C surface area, a 2:2 complex can be observed in cryo-EM. Site C of cGAS binds to the nucleosomal DNA of a second nucleosome in this scenario. |PDB cGAS/Nucleosome 1:1: 7A08; 2:1: 6xjd; 2:2: 7C0M

Structures emerging from the human cGAS construct, showed some additional distal nucleosome linkage via cGAS´ site C[189,308]. This interaction is likely not seen in our data, as the area of site C is smaller for the mouse protein commonly used. This distal nucleosome interaction is mainly mediated via protein-DNA interactions (hcGAS K285/K299-302), that show a lower binding strength compared to the anchoring at the acidic patch. Additional contacts are made by a β-hairpin loop of cGAS with histone H2A´s C-terminal tail and the N-terminal tails of histone H4 of each nucleosome interacting with one another[308]. As no group used the full-length protein for their structural studies, the N-terminal flexible domain was not resolved in any of these structures. Therefore, it cannot be ruled out, that the NTD, which has DNA binding capabilities on its own can bind to linker DNA or *in-trans* to nucleosomal DNA in a nucleosome array setup. The existence of the 2:2 complex has to be verified *in vivo*, as this configuration would lead to changes in chromatin structure introduced by cGAS. In theory this clustering of nucleosomes would result in a less open, more heterochromatin-like structure. Whether cGAS has a role in modulating the dynamic chromatin topology is questionable, as this intrinsic autoimmunity preventing mechanism might not influence global gene regulation.

4.1.2 Chromatin topology and implications for cGAS activity

Interestingly, we could show that nucleosome arrays (multiple nucleosome core particles linked by spacer DNA) can inhibit cGAS activity when they adopt a linear and open form, whereas more compact arrays still activate cGAS. This behavior has implications for cGAS activity on chromatin *in vivo*, as open euchromatin likely efficiently blocks cGAS activity whereas more compact chromatin regions will not. Whether the higher order cGAS-nucleosome complexes still occur on higher order chromatin substrates remains elusive and needs to be addressed in the future. The activity of cGAS on compact chromatin prerequisites either lower concentrations of nuclear cGAS than accessible nucleosomes available or additional layers of regulation. Just recently hyperphosphorylation of the N-terminal part of cGAS during mitosis was described. Mitotic kinase Aurora B (AurKB) phosphorylates the N-terminal tail disrupting the interaction with mitotic chromatin and preventing active complex formation[266]. Additional transient phosphorylation is performed by the mitotic kinase cyclin-dependent kinase 1-cyclin B (CDK1) complex during mitosis. Another factor for inhibition of nuclear cGAS is barrier-to-autointegration factor 1 (BAF). BAF competes with cGAS for binding to exposed DNA, thereby preventing active complex formation. A combination of all these inhibiting mechanisms keeps cGAS efficiently in check, regardless of the chromatin state. If the number of accessible acidic patches is lower than the amount of cGAS, all cGAS is sequestered by them due to the higher affinity for these sites. In the case of free DNA regions and an excess of cGAS, BAF will compete with cGAS´ DNA binding and prevent autoimmunity. Additionally, phosphorylation of cGAS prevents binding and activation by chromatin during mitosis (see Fig. 17).

Contrary to these inhibition mechanisms, the first examples of cGAS/chromatin colocalization were observed by activating effects with chromatin-like but pathogenic substrates. Thereby the shown inhibition mechanism probably represents the most basic discrimination between self vs. non-self or healthy vs. non-healthy respectively. The canonical histone subset represents one of the most conserved, constitutively expressed proteins in eukaryotes. Thereby detection of the canonical NCP can be seen as the ultimate pattern of self.
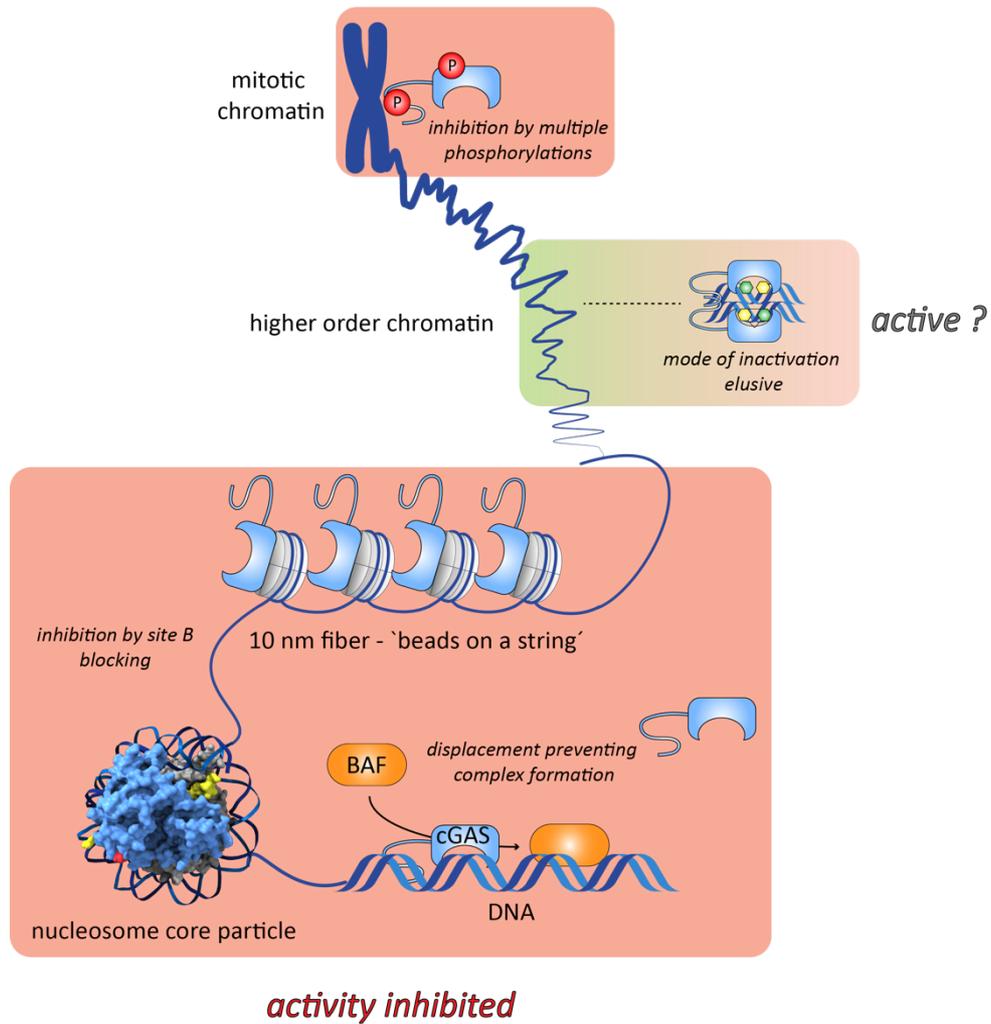
Figure 17: Influence of chromatin topology on cGAS activity. At the level of the NCP cGAS activity is inhibited by binding to the nucleosome´s acidic patch. Due to this tethering cGAS´ DNA binding site B is blocked, preventing the formation of the active complex. At free DNA regions, BAF competes with cGAS for DNA binding, again preventing active complex formation. As long as inhibitory acidic patches are accessible, nucleosomes can still act in an inhibitory manner, even in array configuration. Higher order chromatin, with different histone subsets and higher orders of compaction seems to be activating *in vitro*. The exact mechanism of cGAS inhibition has to be shown. Multiple phoshorylations of cGAS have been shown to prevent activity against mitotic chromatin.

The importance of correct assembled nucleosomes for efficient cGAS inhibition is shown by mutations in the replication-dependent histone (RDH) pre-mRNA-processing complex. Patients suffering from the autoimmune disease AGS can show biallelic mutations in *LSM11* and *RNU7-1*. Mutations in these genes lead to failures in histone RNA processing and thereby change in histone protein composition. Due to this miscomposition they cannot efficiently inhibit cGAS activity anymore, leading to the induction of autoimmunity. Interestingly, lower levels of Histone H1.4 are observed in these patients. In theory, less H1.4 would lead to a more open chromatin form, that would still act as inhibitory in our setup. Possibly a lack of H1.4 in combination with a histone octamer without a correct inhibitory acidic patch lacks the mode of inhibition and adds more areas of free DNA to bind, leading to the observed phenotype.

In conclusion, alterations to the canonical genome organization pattern may still lead to cGAS activity even in a chromatin setup. Many studies show cGAS colocalization with micronuclei [194,301,309]. These small additional nuclei emerge during cancer or cellular mis-segregation and show a different nuclear membrane and histone subset composition[310]. First of all, it must be assumed that the total number of inhibitory nucleosomes is smaller in micronuclei than in the main nucleus. γ-H2AX is a hallmark of DNA damage and is associated with micronuclei´s chromatin fragments. Most studies show colocalization of cGAS with this specific histone. Perhaps this variant shows a different interaction with cGAS as compared to the canonical one, despite there being no relevant structural differences at cGAS´ interaction site (see Fig. 18).
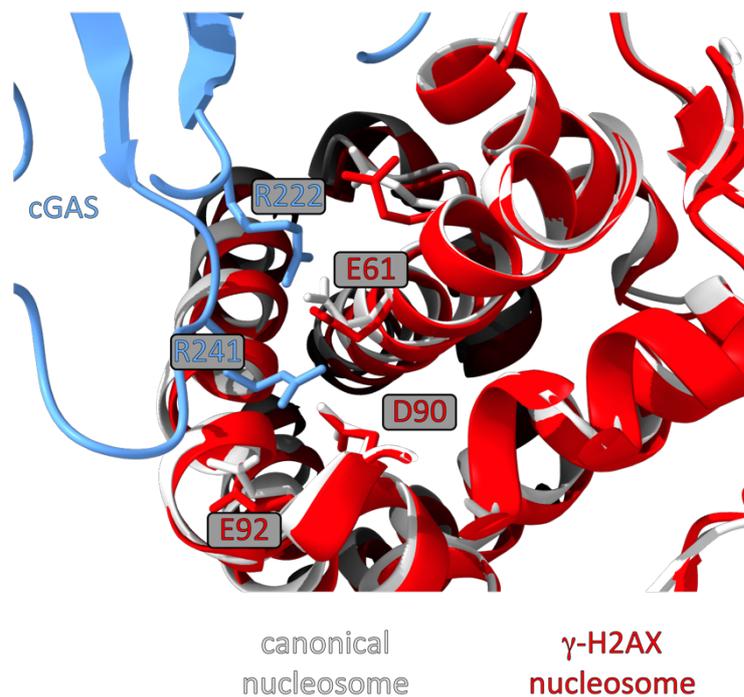
Figure 18: Structural comparison of main cGAS-nucleosome interaction site for the canonical histone octamer and γ-H2AX. Overlay of the γ-H2AX structure (red) with the canonical octamer (grey) bound to cGAS (blue). The most prominent acidic patch interacting residues of cGAS R222 and R241 are shown with the most relevant interacting residues of the acidic patch. No obvious structural differences can be observed that would account for an activating role of cGAS when bound to γ-H2AX. |PDB cGAS-Nucleosome: 7A08, γ-H2AX nucleosome: 6KLI

Additionally, CCFs and micronuclei show distinct markers of heterochromatin like H3K9(me3) and lack euchromatin markers like H3K9ac. These data would support the activity observed with nucleosome arrays. Open euchromatin, with accessible acidic patches, is inhibitory, whereas closed heterochromatin can still activate cGAS. Recently, it was proposed, that it is not the micronuclear chromatin fragments themselves, but rather chromatin bridges formed in this setup that are the actual activator of cGAS[311]. As these bridges have no occupancy with histones no inhibition can take place, supporting our shown data. Chromatin bridges occur more frequently when AurKB is knocked-out. Taken together, the need of AurKB for cGAS-NTD phosphorylation and inhibition, plus its role in chromatin bridge formation, explains the activity of cGAS towards this self-substrate. cGAS colocalization with γ-H2AX was often shown by fluorescent microscopy, but ultimately lacked the resolution to show the direct site of interaction. Therefore, the colocalization seen in other studies might also be rather at histone free regions. Still, cGAS was shown to directly interact with a H2AX peptide in a label-free biomolecular interaction assay with a high affinity of 176.8 nM (compared to dsDNA).

Even if this affinity is lower than the affinity we observed for binding to canonical nucleosomes, it is still higher than the affinity to free DNA[188]. This study also showed interaction with H2AX being dependent on the phosphorylation state of S139 (the residue phosphorylated in γ-H2AX), suggesting a role of this modification for cGAS binding.

Chromatin immunoprecipitation–Sequencing (ChIP-Seq) experiments found nuclear cGAS mostly localized on satellite centromeric DNA and LINE-elements[295]. Again, the same histone pattern can be observed in this case. Pericentromeric chromatin also shows higher levels of H3K9(me3) and lower levels of H3K9ac, as seen for CCFs and micronuclei[312]. The functional role of cGAS preferentially binding to centromere regions is unclear and might just rely solely on the altered chromatin topology in this region. An explanation of cGAS binding to centromeric regions during mitosis is the spatial accumulation of mitotic kinases. As compact chromatin seems to be a good activator of cGAS activity, bringing cGAS in close proximity to the mitotic kinases that phosphorylate and ultimately inactive it might further prevent autoimmunity. This mechanism would facilitate the different outcome of normal mitotic cGAS and the immunogenic sensing of e.g. CCFs.

This influence of chromatin topology might also be used by viruses to prevent recognition by cGAS. Papilloma viruses pack their genome into minichromosomes using cellular histones, that in turn are epigenetically modified similarly to the host-chromatin[313,314]. Especially minichromosomes of papilloma viruses show distinct markers of open chromatin and a lack of histone H1, suggesting a conformation comparable to the linear arrays shown in our study. This packing of viral DNA might inhibit cGAS sensing, by promoting the occurrence of accessible acidic patches[315]. Particularly the existence of higher linked cGAS-nucleosome complexes must be taken with precaution. During homologous recombination (HR) DNA repair cGAS was shown to bind to the damaged genomic region and self-oligomerize. These already described higher order complexes of cGAS, facilitated by its NTD and site C, lead to a DNA compaction that prevents strand invasion by RAD51 needed for HR[316]. Therefore, DNA damage does not get repaired and will lead the cell into the formation of micronuclei, CCFs or senescence. Clustering of nucleosomes via cGAS would lead to a compaction even if no DNA damage has occurred.

4.1.3 Pathogenic effects of altered chromatin structure sensing via cGAS

This work and other studies demonstrate multiple inhibition mechanisms used to keep cGAS in check against non-pathogenic self-DNA. Still, changes in the healthy genomic state are linked to cGAS activity. Thereby the tight regulation of cGAS activity is underlined, as multiple contrary effects can be observed. In general, it can be manifested that cGAS might sense substrates that differ from the canonical one. Accumulation of DNA damage is a major reason for the development of cancer[317]. As a consequence defects in the cGAS-STING axis are linked to cancer development[318,319]. Usually, a set of DNA damage repair mechanisms correct the error before it can multiply in cell division. Multiple studies connect the occurrence of DNA damage with the induction of an interferon-response[320–322]. Damage repair by homologous recombination is inhibited by overexpression of cGAS via its interaction with PARP1[323]. Non-homologous end joining is inhibited by cGAS condensation which prevents strand-invasion needed for Rad51 activity[316]. Another possible interference of cGAS with DNA damage repair (DDR) might be by its occupancy of the acidic patch of nucleosomes. Ubiquitylation of H2A/H2AX serves as signal for activation of the DNA damage response. Blocking the acid patch by the LANA peptide, which occupies the same binding pocket as cGAS (see Fig. 5), impedes ubiquitylation and subsequent recruitment of proteins involved in DDR[324]. As a consequence, cells with damaged genomic DNA will form altered, previously described chromosomal structures and may end up in senescence[198], undergoing autophagic clearance[325] or even apoptosis[326], thereby preventing tumor progression. Yet, cGAS should only inhibit the DNA repair if successful repair cannot be achieved anymore. The exact mechanisms governing the cell´s fate need to be investigated.

4.1.4 Nuclear tethering vs. cytosolic sensing of cGAS

Even if multiple studies show that cGAS is also a nuclear protein, the cytosolic DNA sensing is still essential. In order to tie the tight nuclear tethering with the availability in the cytosol multiple prerequisites are needed. First and foremost, cGAS needs to detach from the nucleosome. Our surface-plasmon-resonance studies showed higher affinity to nucleosomes as compared to free DNA. Still the observed half-time of the cGAS-Nucleosome complex suggests that dissociation without additional factors is at least possible. Nevertheless, most likely, post-translational modifications of cGAS and/or the histones is needed for sufficient release. The acidic patch is a prominent docking site for a plethora of other nuclear proteins, so it cannot be excluded that a specific cofactor might compete with cGAS for nucleosome binding thereby leading to its release. In our nuclear fractionating studies, where we observed the tight nuclear tethering for endogenous and overexpressed cGAS, a modification of cGAS was observed prior to transfection with DNA. This modification disappears after transfection with DNA and might play a role in cytosolic relocalization of cGAS. The exact nature of this modification has to be addressed in further studies. Interestingly, the known modifications of cGAS (see Fig. 13) are well distributed over the whole protein sequence but show no modifications where the interaction with the acidic patch takes place. As most modifications of the histone proteins take place on the histone-tails, their influence on cGAS binding is also unclear to date. At least the formation of higher order cGAS complexes that also show involvement of the histone-tails, might be affected by histone-tail PTMs. Most cancer cell lines have no functional cGAS-STING axis to evade cGAS mediated immunity. Treating colorectal cancer cell lines with histone-deacetylase or histone-demethylase inhibitors restores cGAS expression partially, suggesting an influence of these modifications on cGAS availability[318]. Even for the sensing of nuclear replicating DNA viruses cGAS has to detach from the nucleosomes. Protein arginine methyltransferase 5 (PRMT5) was shown to interact with cGAS in the nucleus thus facilitating non-canonical anti-viral immunity[327]. PRMT5 promotes transcriptional activation by dimethylation of histone H3R2[328]. In the cGAS setting this leads to activation of IFN-genes not related to IRF3 signaling. The site of direct interaction of cGAS with PMRT5 was shown to be located from residue 241 to 380[327]. As this contains the main region of the acidic patch interaction, binding to PMRT5 might prevent nucleosome tethering.

cGAS is also able to sense HIV infection in the nucleus[142]. Immunoprecipitation assays show a direct protein-protein interaction of cGAS with the protein NONO. NONO usually has multiple nuclear functions in all steps of gene regulation. In the case of HIV infection (especially HIV-2) NONO binds to the viral capsid and activates cGAS by direct binding. The exact mode of activation is still unclear, and nothing is known about the exact site of interaction of cGAS with NONO.

Regardless of the actual mechanism of cGAS detachment from the nucleosome, which must be addressed in further studies, cGAS needs to be exported into the cytosol upon cytosolic DNA infection. Like all proteins cGAS is synthesized in the cytosol but will end up in the nucleus at the latest after the first nuclear envelope break down during cell division. Additionally, two nuclear localization signals were discovered at the N-terminal tail of cGAS (NLS1: hcGAS 21-51) and in its CTD (NLS2: hcGAS 295-305). As cGAS interacts with proteins of the importin-α-family, a classical importin-mediated import was proposed[323]. Conversely, a conserved nuclear export signal (hcGAS 169-174) may lead to relocalization to the cytosol, when cGAS is no longer bound to chromatin. The shuttling of cGAS is proposed to be mediated by the main nuclear export receptor chromosomal maintenance 1 (CRM1). Blocking CRM1 by the fungicide LMB disrupts nuclear cGAS translocation upon cytosolic DNA infection[329]. Once located in the cytosol phosphorylation at Y215 (for hcGAS) by B-lymphocyte kinase (BLK) acts as signal for cytosolic cGAS retention. The corresponding phosphatase is not known to date[323]. Disturbing this phosphorylation by knocking-down BLK leads to higher nuclear localization of cGAS. Whether cGAS that has exited the nucleus is recycled by dephosphorylation and nuclear re-entry or is degraded remains to be clarified (see Fig. 19).

In general, the nuclear localization of cGAS assumes a role beyond storage and recognizing a pattern of self for autoinhibition. Most PRRs are interferon-stimulated genes and only occur in low numbers prior to infection. Even if an IFN-mediated positive feedback loop is observed for cGAS in some cell types, nuclear cGAS seems to be present in most cell types. Taking into account the role of cGAS in DDR, senescence and tumorigenesis, a non-canonical role for nuclear cGAS beyond DNA-sensing seems to be likely.
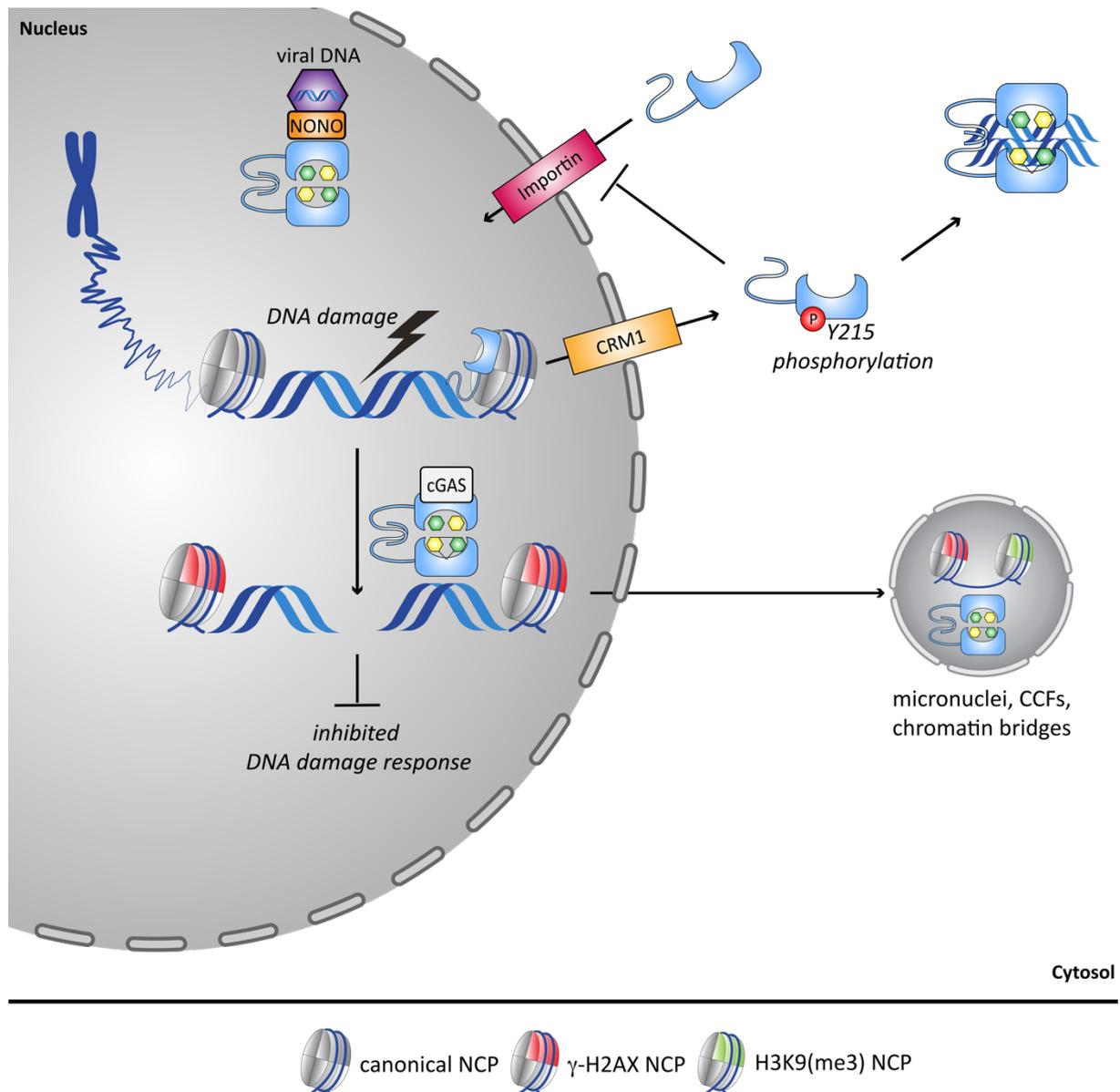
Figure 19: Proposed mechanism of cGAS shuttling. cGAS (blue) is tethered to canonical nucleosomes therby inhibiting its activity (all grey NCP). In the case of DNA damage, often associated with the enrichment of γ-H2AX nucleosomes next to the damage, cGAS seems to colocalize with the site of DNA damage and is active, even inhibiting DNA damage repair. The non-repaired lesions can lead to the formation of abberant chromatin species, like micronuclei or cytosolic chromatin fragments that again can active cGAS. The exact mechanism of cGAS detachement from the nucleosome is not understood, but cGAS seems to be exported via its nuclear export signal by the transporter CRM1. In the cytosol cGAS can also sense free DNA from different sources. Phosphorylation of cGAS at Y215 was shown to inhibit relocalization into the nucleus, relying on a classical nuclear localization signal-dependent importin mediated pathway.

4.2 RIG-I ATPase activity and autoimmunity

The tethering of cGAS to nucleosomes displays a strong discrimination between self and non-self-DNA in innate immune sensing. The functional counterpart of dsDNA sensing by cGAS is RNA sensing by RIG-I. As seen for cGAS, mechanisms have evolved in RNA sensing to differentiate between self and non-self. Failures in this regulation are linked to multiple interferonopathies like SLS and AGS. First of all, the CTD of RIG-I screens the 5´-end of the RNA for specific features solely found in pathogens. A suitable RNA-ligand for RIG-I has three features: (1) a triphosphate at the 5´[330], with base pairing at the 5´(blunt end representing the most potent form)[331] or (2) 5´-diphosphate RNA with uncapped ends[131] and (3) the 5´-terminal nucleotide needs to be unmethylated at its 2´-O[133]. A second layer of regulation is achieved by ATP hydrolysis. As RIG-I structurally belongs to the family of SF2-ATPases it also shows ATP binding and hydrolysis, but the role of the actual helicase activity of this protein family is questionable for RIG-I[281,332]. Many studies show that at least ATP binding is needed in addition to RNA binding for sufficient activation, as the full ring-like active conformation is only achieved after ATP binding to the binding pocket created by initial RNA binding[288,333]. Former work of this group showed that ATP hydrolysis leads to dissociation of RIG-I from low-affinity ligands, like self-RNA, thus preventing autoimmunity[288]. ATP hydrolysis can lead to translocation on dsRNA stems, omitting RNA binding and restoring the inactive CARD-sequestered conformation[334]. Two common mutations in the ATPase domain are found in the context of SLS. E373Q, located in motif II, prevents proper polarization of the water molecule needed for ATP hydrolysis and thereby impairs the connected proof-reading mechanism, leading to increased autoimmunity. The other common mutation C268F is located in motif I of the ATPase motif needed for ATP binding. Strikingly, this mutation leads to an increased RIG-I activity, whereas K270A/I renders RIG-I inactive by omitting ATP binding[288]. We could show how these different outcomes are achieved despite the close spatial localization of these mutations.

Fluorescence anisotropy measurements revealed that RIG-I C268F shows increased affinity to hairpin-RNA (hp-RNA) with unpaired ends (a substrate to simulate an endogenous ligand) compared to the WT, and that is even increased in the presence of ATP. Contrary, the affinity of RIG-I WT to hp-RNA is decreased in the presence of ATP underlining the proof-reading activity and preventing association with suboptimal ligands. The lack of ATP-dependent discrimination already suggests a higher auto-activity. This model is supported by the reduced catalytic activity found for the C268F mutant compared to the WT protein observed in radioactive ATP hydrolysis assays. The affinity of RIG-I C268F towards ATP (measured by a tryptophan-fluorescence based MANT-ATP FRET assay) was lower compared to WT-RIG-I or RIG-I E373Q and similar to the catalytically dead mutant K270I. Interestingly, this affinity did not change for RIG-I C268F after addition of RNA, in contrast to the other described mutants. This is particularly interesting as both ATP and RNA binding are thought to be needed for the formation of the ring-link active conformation. Further prevention of ATP recognition by creating a triple mutant not able to recognize ATP´s adenine (R244A/Q247A/C268F) still led to a significantly higher IFN-β promotor activity compared to the WT. Taken together these data suggest a CARD release of RIG-I C268F, that is at least partially independent of ATP binding.

Indeed, our crystallization approaches did not yield a nucleotide bound structure regardless of the crystallization conditions used. Adversely, the bulky F268 residue shifts the location of the conserved residue K270 from its expected phosphate coordination position towards the space that is usually occupied by the coordinated Mg^{2+}. The displaced K270 forms a salt-bridge with E702 leading to an ATP-bound-like conformation without actual nucleotide bound. As disrupting the salt-bridge by mutating E702 did not completely prevent constant activity, ATP binding might still be possible *in vivo* at higher ATP concentrations. Most likely, RIG-I C268F already allows signaling without ATP, but can show ATP-dependent signaling lacking a proof-reading mechanism at higher ATP concentrations, leading to the observed constantly active phenotype. How Mg^{2+} coordination in the latter setting can be achieved, as its binding site is occupied by K270, is still puzzling. Perhaps K270 fulfills the coordinating function usually achieved by Mg^{2+}.

Our structural and biochemical data is supported by hydrogen deuterium exchange mass spectrometry data (HDX-MS). RIG-I C268F was shown to have a more ring-like structure not dependent on the 5´-modification of the RNA ligand, supporting the activation by endogenous and exogenous RNA ligands. The kinetics for CARD release were shown not to be influenced by RNA modification or external ATP addition (unlike the observed effects for RIG-I WT and RIG-I E373A). These data support our model of RIG-I C268F mimicking the ATP-bound state that disrupts regulation[335].

Still, there is some controversy in the field, whether ATP binding is needed for CARDs release *in vivo*. Already the data of our triple mutant that still shows residual constant activity implicates that the situation *in vivo* might be more complex. Additionally, some RNAs that bind to RIG-I *in vitro* and stimulate ATP hydrolysis do not lead to signaling in cells[336,337]. A newly developed FRET-reporter assay to measure CARDs release directly and not the production of IFN-β (a downstream event) indicates that RNA binding alone is sufficient for CARD presentation and thereby downstream signaling[338]. This work sees the role of ATP in displacing competing proteins from the RNA to facilitate RIG-I activation. The HDX-MS spectra of RIG-I C268F also show increased exposure of the CARDs after RNA binding, but the authors credit the pseudo-ATP-bound state for the drastic change[335]. Even if ATP is not needed to release the CARDs, the pseudo-ATP-bound state will still lead to constant activation of the C268F mutant if the ATP-bound state has a different function.

Transient-state stopped-flow kinetics have been used to analyze the role of ATP binding and hydrolysis on RIG-I activity. This study suggests a kinetic proofreading mechanism by RIG-I using faster off-rates of non-optimal ligands for discrimination[339]. In this model ATP binding alone does not have a proof-reading function, but rather leads to less selective RNA binding, that we also observed in our fluorescence anisotropy measurement with hp-RNA. ATP hydrolysis on the other hand dislocates RIG-I from RNA by translocation. Kinetic measurements of RIG-I C268F showed long-living complexes even in the absence of ATP, representing the same pseudo-ATP-bound state observed in our crystal structure.

In conclusion the ATP-bound-like state of the pathogenic RIG-I C268F variant explains the constant activation and cause of autoimmunity. Even if the exact role of ATP in further downstream signaling is under debate, mimicking ATP binding without hydrolysis will lead to higher activity in any scenario (see Fig. 20).
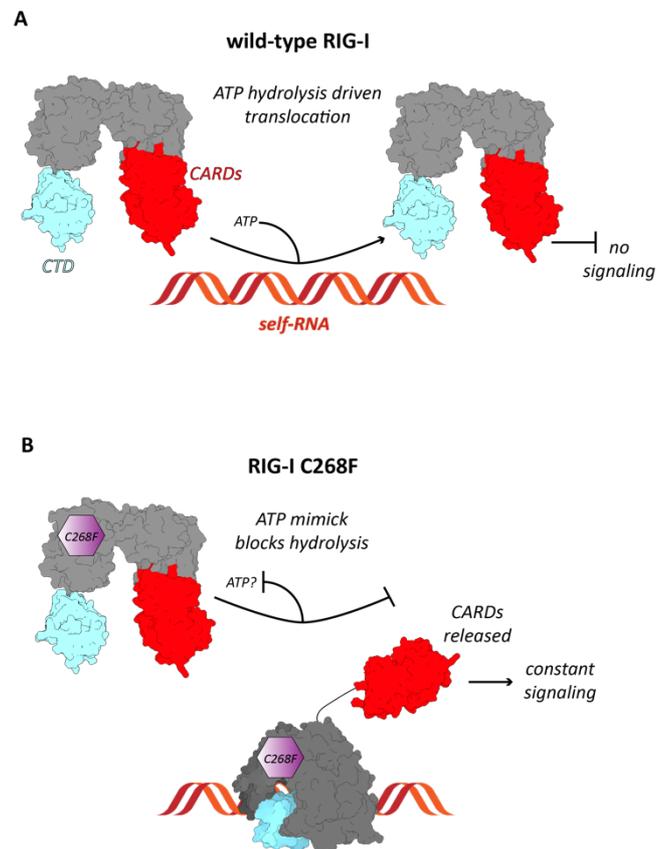
Figure 20: Constant activation by RIG-I C268F. A When RIG-I encounters a suboptimal ligand like self-RNA ATP hydrolysis driven translocation dissociates RIG-I from the RNA, thereby failing to activate signaling. **B** In the case of the RIG-I SMS variant C268F an ATP-bound like state is found. This conformation without actual nucleotide bound bypasses the regulation mechanism and leads to constant activity. Whether any ATP binding is possible *in vivo* is not clear. |PDB for A: 4A2W+2QFB; B: 5EH3+2QFB

5 Literature

1.   Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).

2.   Gray, M. W., Burger, G. & Franz Lang, B. The origin and early evolution of mitochondria. *Genome Biology* vol. 2 reviews1018.1 (2001).

3.   Miescher Friedrich. Ueber die chemische Zusammensetzung der Eiterzellen. *Hoppe-Seyler,Medicinisch-chemische Untersuchungen* **4**, 441–460 (1871).

4.   Flemming, W. *Zellsubstanz, Kern und Zelltheilung*. (F.C.W. Vogel, 1882).

5.   Kossel, A. Ueber die chemische Beschaffenheit des Zellkerns. *Muenchener Medizinische Wochenschrift* 65–69 (1911).

6.   Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *J. Exp. Med.* **79**, 137–158 (1944).

7.   Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. *Nature* vol. 171 737–738 (1953).

8.   Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. United States* **51**, 786–794 (1964).

9.   Clark, R. J. & Felsenfeld, G. Structure of chromatin. *Nat. New Biol.* **229**, 101–106 (1971).

10.   Olins, D. E. & Olins, A. L. Physical studies of isolated eucaryotic nuclei. *J. Cell Biol.* **53**, 715–736 (1972).

11.   Olins, A. L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science (80-.).* **183**, 330–332 (1974).

12.   Kornberg, R. D. Chromatin structure: A repeating unit of histones and DNA. *Science (80-.).* **184**, 868–871 (1974).

13.   Oudet, P., Gross-Bellard, M. & Chambon, P. *Electron Microscopic and Biochemical Evidence that Chromatin Structure Is a Repeating Unit*. *Cell* vol. 4 (1975).

14.   Arents, G., Burlingame, R. W., Wang, B. I. C., Love, W. E. & Moudrianakis, E. N. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10148–10152 (1991).

15. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).

16. White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* **20**, 5207–5218 (2001).

17. Arents, G. & Moudrianakis, E. N. Topography of the histone octamer surface: Repeating structural motifs utilized in the docking of nucleosomal DNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 10489–10493 (1993).

18. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).

19. Wang, D., Ulyanov, N. B. & Zhurkin, V. B. Sequence-dependent kink-and-slide deformations of nucleosomal dna facilitated by histone arginines bound in the minor groove. *J. Biomol. Struct. Dyn.* **27**, 843–859 (2010).

20. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).

21. Tolstorukov, M. Y., Colasanti, A. V., McCandlish, D. M., Olson, W. K. & Zhurkin, V. B. A Novel Roll-and-Slide Mechanism of DNA Folding in Chromatin: Implications for Nucleosome Positioning. *J. Mol. Biol.* **371**, 725–738 (2007).

22. Hayes, J. J., Clark, D. J. & Wolffe, A. P. Histone contributions to the structure of DNA in the nucleosome. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 6829–6833 (1991).

23. Yan, L. & Chen, Z. A Unifying Mechanism of DNA Translocation Underlying Chromatin Remodeling. *Trends in Biochemical Sciences* vol. 45 217–227 (2020).

24. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).

25. Ngo, T. T. M., Zhang, Q., Zhou, R., Yodh, J. G. & Ha, T. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* **160**, 1135–1144 (2015).

26. Ausio, J., Dong, F. & van Holde, K. E. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone 'tails' in the stabilization of the nucleosome. *J. Mol. Biol.* **206**, 451–463 (1989).

27. Polach, K. J., Lowary, P. T. & Widom, J. Effects of core histone tail domains the equilibrium constants for dynamic DNA site accessibility in nucleosomes. *J. Mol. Biol.* **298**, 211–223 (2000).

28. Allan, J., Harborne, N., Rau, D. C. & Gould, H. Participation of core histone 'tails' in the stabilization of the chromatin solenoid. *J. Cell Biol.* **93**, 285–297 (1982).

29. Lee, K. M. & Hayes, J. J. Linker DNA and H1-dependent reorganization of histone-DNA interactions within the nucleosome. *Biochemistry* **37**, 8622–8628 (1998).

30. Schwarz, P. M. & Hansen, J. C. Formation and stability of higher order chromatin structures. Contributions of the histone octamer. *J. Biol. Chem.* **269**, 16284–16289 (1994).

31. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin fiber folding: Requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* **327**, 85–96 (2003).

32. Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: An ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology* vol. 13 436–447 (2012).

33. Simpson, R. T. Structure of the Chromatosome, a Chromatin Particle Containing 160 Base Pairs of DNA and All the Histones. *Biochemistry* **17**, 5524–5531 (1978).

34. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219–223 (1993).

35. Srinivas Bharath, M. M., Chandra, N. R. & Rao, M. R. S. Molecular modeling of the chromatosome particle. *Nucleic Acids Res.* **31**, 4264–4274 (2003).

36. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 1897–1901 (1976).

37. Woodcock, C. L. F., Frado, L. L. Y. & Rattner, J. B. The higher-order structure of chromatin: Evidence for a helical ribbon arrangement. *J. Cell Biol.* **99**, 42–52 (1984).

38. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the '30-nm' chromatin fiber: Evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6506–6511 (2006).

39. Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138–141 (2005).

40. Grigoryev, S. A., Arya, G., Correll, S., Woodcock, C. L. & Schlick, T. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13317–13322 (2009).

41. Arya, G. & Schlick, T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16236–16241 (2006).

42. Grigoryev, S. A. Chromatin Higher-Order Folding: A Perspective with Linker DNA Angles. *Biophysical Journal* vol. 114 2290–2297 (2018).

43. Fussner, E., Ching, R. W. & Bazett-Jones, D. P. Living without 30nm chromatin fibers. *Trends Biochem. Sci.* **36**, 1–6 (2011).

44. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology* vol. 30 39–58 (2014).

45. Sanulli, S. *et al.* HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* **575**, 390–394 (2019).

46. Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e21 (2019).

47. Zhao, Y. & Garcia, B. A. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).

48. Xhemalce, B., Dawson, M. A. & Bannister, A. J. Histone Modifications. in *Encyclopedia of Molecular Cell Biology and Molecular Medicine* (Wiley-VCH Verlag GmbH & Co. KGaA, 2011). doi:10.1002/3527600906.mcb.201100004.

49. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Research* vol. 21 381–395 (2011).

50. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nature Structural and Molecular Biology* vol. 20 259–266 (2013).

51. North, J. A. *et al.* Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling. *Nucleic Acids Res.* **39**, 6465–6474 (2011).

52. Panne, D. *et al.* Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Res.* **46**, 9907–9917 (2018).

53. Talbert, P. B. & Henikoff, S. Histone variants ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology* vol. 11 264–275 (2010).

54. Long, M. *et al.* A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic Acids Res.* **47**, 8399–8409 (2019).

55. Foltz, D. R. *et al.* Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* **137**, 472–484 (2009).

56. Goldberg, A. D. *et al.* Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* **140**, 678–691 (2010).

57. Piquet, S. *et al.* The Histone Chaperone FACT Coordinates H2A.X-Dependent Signaling and Repair of DNA Damage. *Mol. Cell* **72**, 888-901.e7 (2018).

58. Bönisch, C. & Hake, S. B. Histone H2A variants in nucleosomes and chromatin: More or less stable? *Nucleic Acids Res.* **40**, 10719–10741 (2012).

59. Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. *Nature Structural and Molecular Biology* vol. 26 3–13 (2019).

60. Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science (80-.).* **311**, 856–861 (2006).

61. Fang, Q. *et al.* Human cytomegalovirus IE1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome. *Elife* **5**, (2016).

62. Lesbats, P. *et al.* Structural basis for spumavirus GAG tethering to chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 509–5514 (2017).

63. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562–566 (2010).

64. Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science (80-.).* **334**, 977–982 (2011).

65. Zhou, J., Fan, J. Y., Rangasamy, D. & Tremethick, D. J. The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat. Struct. Mol. Biol.* **14**, 1070–1076 (2007).

66. Allahverdi, A. *et al.* The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res.* **39**, 1680–1691 (2011).

67. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (80-.).* **311**, 844–847 (2006).

68. Kan, P.-Y., Lu, X., Hansen, J. C. & Hayes, J. J. The H3 Tail Domain Participates in Multiple Interactions during Folding and Self-Association of Nucleosome Arrays. *Mol. Cell. Biol.* **27**, 2084–2091 (2007).

69. Dunin-Horkawicz, S., Kopec, K. O. & Lupas, A. N. Prokaryotic ancestry of eukaryotic protein networks mediating innate immunity and apoptosis. *J. Mol. Biol.* **426**, 1568–1582 (2014).

70. Desjardins, M., Houde, M. & Gagnon, E. Phagocytosis: The convoluted way from nutrition to adaptive immunity. *Immunological Reviews* vol. 207 158–165 (2005).

71. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nature Immunology* vol. 16 343–353 (2015).

72. Zhang, Q., Zmasek, C. M. & Godzik, A. Domain architecture evolution of pattern-recognition receptors. *Immunogenetics* **62**, 263–272 (2010).

73. Janeway, C. Immunogenecity signals 1,2,3... and 0. *Immunol. Today* **10**, 283–286 (1989).

74. Matzinger, P. Tolerance, Danger, and the Extended Family. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).

75. Tang, D., Kang, R., Coyne, C. B., Zeh, H. J. & Lotze, M. T. PAMPs and DAMPs: Signal 0s that spur autophagy and immunity. *Immunol. Rev.* **249**, 158–175 (2012).

76. Rubartelli, A. & Lotze, M. T. Inside, outside, upside down: damage-associated molecular-pattern molecules (DAMPs) and redox. *Trends Immunol.* **28**, 429–436 (2007).

77. Janeway, C. A. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54 Pt 1**, 1–13 (1989).

78. Brubaker, S. W., Bonham, K. S., Zanoni, I. & Kagan, J. C. Innate immune pattern recognition: a cell biological perspective. *Annu. Rev. Immunol.* **33**, 257–90 (2015).

79. Deretic, V., Saitoh, T. & Akira, S. Autophagy in infection, inflammation and immunity. *Nature Reviews Immunology* vol. 13 722–737 (2013).

80. Lamkanfi, M. & Dixit, V. M. Mechanisms and functions of inflammasomes. *Cell* vol. 157 1013–1022 (2014).

81. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette spatzle/Toll/Cactus controls the potent antifungal response in Drosophila adults. *Cell* **86**, 973–983 (1996).

82. Hashimoto, C., Hudson, K. L. & Anderson, K. V. The Toll gene of drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell* **52**, 269–279 (1988).

83. Schwandner, R., Dziarski, R., Wesche, H., Rothe, M. & Kirschning, C. J. Peptidoglycan- and lipoteichoic acid-induced cell activation is mediated by Toll-like receptor 2. *J. Biol. Chem.* **274**, 17406–17409 (1999).

84. Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).

85. Rassa, J. C., Meyers, J. L., Zhang, Y., Kudaravalli, R. & Ross, S. R. Murine retroviruses activate B cells via interaction with toll-like receptor 4. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2281–2286 (2002).

86. Underhill, D. M. *et al.* The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).

87. Kurt-Jones, E. A. *et al.* Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. *Nat. Immunol.* **1**, 398–401 (2000).

88. Smiley, S. T., King, J. A. & Hancock, W. W. Fibrinogen Stimulates Macrophage Chemokine Secretion Through Toll-Like Receptor 4. *J. Immunol.* **167**, 2887–2894 (2001).

89. Asea, A. *et al.* Novel signal transduction pathway utilized by extracellular HSP70. Role of toll-like receptor (TLR) 2 and TLR4. *J. Biol. Chem.* **277**, 15028–15034 (2002).

90. Brown, G. D., Willment, J. A. & Whitehead, L. C-type lectins in immunity and homeostasis. *Nature Reviews Immunology* vol. 18 374–389 (2018).

91. Weis, W. I., Taylor, M. E. & Drickamer, K. The C-type lectin superfamily in the immune system. *Immunological Reviews* vol. 163 19–34 (1998).

92. Zelensky, A. N. & Gready, J. E. The C-type lectin-like domain superfamily. *FEBS Journal* vol. 272 6179–6217 (2005).

93. Motta, V., Soares, F., Sun, T. & Philpott, D. J. Nod-like receptors: Versatile cytosolic sentinels. *Physiol. Rev.* **95**, 149–178 (2015).

94. Kufer, T. A. & Sansonetti, P. J. NLR functions beyond pathogen recognition. *Nature Immunology* vol. 12 121–128 (2011).

95. Wu, J. & Chen, Z. J. Innate Immune Sensing and Signaling of Cytosolic Nucleic Acids. *Annu. Rev. Immunol.* **32**, 461–488 (2014).

96. Botos, I., Segal, D. M. & Davies, D. R. The structural biology of Toll-like receptors. *Structure* vol. 19 447–459 (2011).

97. Pelka, K., Shibata, T., Miyake, K. & Latz, E. Nucleic acid-sensing TLRs and autoimmunity: novel insights from structural and cell biology. *Immunol. Rev.* **269**, 60–75 (2016).

98. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF-κB by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).

99. Heil, F. *et al.* Species-Specific Recognition of Single-Stranded RNA via Till-like Receptor 7 and 8. *Science (80-.).* **303**, 1526–1529 (2004).

100. Oldenburg, M. *et al.* TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance - Forming modification. *Science (80-.).* **337**, 1111–1115 (2012).

101. Hidmark, A., von Saint Paul, A. & Dalpke, A. H. Cutting Edge: TLR13 Is a Receptor for Bacterial RNA. *J. Immunol.* **189**, 2717–2721 (2012).

102. Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).

103. Bauer, S. *et al.* Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9237–9242 (2001).

104. Hemmi, H. *et al.* Small-antiviral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nat. Immunol.* **3**, 196–200 (2002).

105. Jurk, M. *et al.* Human TLR7 or TLR8 independently confer responsiveness to the antiviral compound R-848 [1]. *Nature Immunology* vol. 3 499 (2002).

106. Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Reviews Immunology* vol. 4 499–511 (2004).

107. Inohara, N. *et al.* An induced proximity model for NF-κB activation in the Nod1/RICK and RIP signaling pathways. *J. Biol. Chem.* **275**, 27823–27831 (2000).

108. Shaw, N. & Liu, Z.-J. Role of the HIN Domain in Regulation of Innate Immune Responses. *Mol. Cell. Biol.* **34**, 2–15 (2014).

109. Shaw, M. H., Reimer, T., Kim, Y. G. & Nuñez, G. NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Current Opinion in Immunology* vol. 20 377–382 (2008).

110. Kanneganti, T. D. *et al.* Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* **440**, 233–236 (2006).

111. Zhong, Z. *et al.* New mitochondrial DNA synthesis enables NLRP3 inflammasome activation. *Nature* **560**, 198–203 (2018).

112. Swanson, K. V., Deng, M. & Ting, J. P. Y. The NLRP3 inflammasome: molecular activation and regulation to therapeutics. *Nature Reviews Immunology* vol. 19 477–489 (2019).

113. Franchi, L., Eigenbrod, T., Muñoz-Planillo, R. & Nuñez, G. The inflammasome: A caspase-1-activation platform that regulates immune responses and disease pathogenesis. *Nature Immunology* vol. 10 241–247 (2009).

114. Shi, J. *et al.* Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **526**, 660–665 (2015).

115. Hornung, V. *et al.* AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458**, 514–518 (2009).

116. Unterholzner, L. *et al.* IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* **11**, 997–1004 (2010).

117. Almine, J. F. *et al.* IFI16 and cGAS cooperate in the activation of STING during DNA sensing in human keratinocytes. *Nat. Commun.* **8**, 14392 (2017).

118. Gariano, G. R. *et al.* The intracellular DNA sensor IFI16 gene acts as restriction factor for human Cytomegalovirus replication. *PLoS Pathog.* **8**, (2012).

119. Jakobsen, M. R. *et al.* IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

120. Goubau, D., Deddouche, S. & Reis e Sousa, C. Cytosolic Sensing of Viruses. *Immunity* vol. 38 855–869 (2013).

121. Andrejeva, J. *et al.* The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-β promoter. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17264–17269 (2004).

122. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.* **5**, 730–737 (2004).

123. Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441**, 101–105 (2006).

124. Leung, D. W. & Amarasinghe, G. K. Structural insights into RNA recognition and activation of RIG-I-like receptors. *Current Opinion in Structural Biology* vol. 22 297–303 (2012).

125. Kowalinski, E. *et al.* Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* **147**, 423–435 (2011).

126. Luo, D. *et al.* Structural insights into RNA recognition by RIG-I. *Cell* **147**, 409–422 (2011).

127. Binder, M. *et al.* Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid-inducible gene-i (RIG-I). *J. Biol. Chem.* **286**, 27278–27287 (2011).

128. Yu, Q., Qu, K. & Modis, Y. Cryo-EM Structures of MDA5-dsRNA Filaments at Different Stages of ATP Hydrolysis. *Mol. Cell* **72**, 999-1012.e6 (2018).

129. Hornung, V. *et al.* 5'-Triphosphate RNA is the ligand for RIG-I. *Science (80-.).* **314**, 994–997 (2006).

130. Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science (80-.).* **314**, 997–1001 (2006).

131. Goubau, D. *et al.* Antiviral immunity via RIG-I-mediated recognition of RNA bearing 59-diphosphates. *Nature* **514**, 372–375 (2014).

132. Schlee, M. *et al.* Recognition of 5' Triphosphate by RIG-I Helicase Requires Short Blunt Double-Stranded RNA as Contained in Panhandle of Negative-Strand Virus. *Immunity* **31**, 25–34 (2009).

133. Schuberth-Wagner, C. *et al.* A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2'O-Methylated Self RNA. *Immunity* **43**, 41–51 (2015).

134. Rehwinkel, J. *et al.* RIG-I Detects Viral Genomic RNA during Negative-Strand RNA Virus Infection. *Cell* **140**, 397–408 (2010).

135. te Velthuis, A. J. W. *et al.* Mini viral RNAs act as innate immune agonists during influenza virus infection. *Nature Microbiology* vol. 3 1234–1242 (2018).

136. Chazal, M. *et al.* RIG-I Recognizes the 5' Region of Dengue and Zika Virus Genomes. *Cell Rep.* **24**, 320–328 (2018).

137. Pichlmair, A. *et al.* Activation of MDA5 Requires Higher-Order RNA Structures Generated during Virus Infection. *J. Virol.* **83**, 10761–10769 (2009).

138. Gitlin, L. *et al.* Essential role of mda-5 in type I IFN responses to polyriboinosinic: polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8459–8464 (2006).

139. Dias Junior, A. G., Sampaio, N. G. & Rehwinkel, J. A Balancing Act: MDA5 in Antiviral Immunity and Autoinflammation. *Trends in Microbiology* vol. 27 75–85 (2019).

140. Gray, E. E. *et al.* The AIM2-like Receptors Are Dispensable for the Interferon Response to Intracellular DNA. *Immunity* **45**, 255–266 (2016).

141. Li, X. D. *et al.* Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science (80-.).* **341**, 1390–1394 (2013).

142. Gao, D. *et al.* Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science (80-.).* **341**, 903–906 (2013).

143. Paijo, J. *et al.* cGAS Senses Human Cytomegalovirus and Induces Type I Interferon Responses in Human Monocyte-Derived Cells. *PLOS Pathog.* **12**, e1005546 (2016).

144. Zhang, Y. *et al.* The DNA Sensor, Cyclic GMP–AMP Synthase, Is Essential for Induction of IFN-β during Chlamydia trachomatis Infection . *J. Immunol.* **193**, 2394–2404 (2014).

145. 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).

146. Storek, K. M., Gertsvolf, N. A., Ohlson, M. B. & Monack, D. M. cGAS and Ifi204 cooperate to produce type I IFNs in response to Francisella infection. *J. Immunol.* **194**, 3236–45 (2015).

147. Hansen, K. *et al.* Listeria monocytogenes induces IFNβ expression through an IFI16-, cGAS- and STING-dependent pathway . *EMBO J.* **33**, 1654–1666 (2014).

148. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674–8 (2008).

149. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type i interferon-dependent innate immunity. *Nature* **461**, 788–792 (2009).

150. Zhong, B. *et al.* The Adaptor Protein MITA Links Virus-Sensing Receptors to IRF3 Transcription Factor Activation. *Immunity* **29**, 538–550 (2008).

151. Abe, T. *et al.* STING Recognition of Cytoplasmic DNA Instigates Cellular Defense. *Mol. Cell* **50**, 5–15 (2013).

152. Burdette, D. L. *et al.* STING is a direct innate immune sensor of cyclic di-GMP. *Nature* **478**, 515–518 (2011).

153. Hopfner, K. P. & Hornung, V. Molecular mechanisms and cellular functions of cGAS–STING signalling. *Nature Reviews Molecular Cell Biology* vol. 21 501–521 (2020).

154. Zhang, X. *et al.* Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51**, 226–35 (2013).

155. Abe, T. & Barber, G. N. Cytosolic-DNA-Mediated, STING-Dependent Proinflammatory Gene Induction Necessitates Canonical NF- B Activation through TBK1. *J. Virol.* **88**, 5328–5341 (2014).

156. Chen, H. *et al.* Activation of STAT6 by STING is critical for antiviral innate immunity. *Cell* **147**, 436–446 (2011).

157. Ma, F. *et al.* Positive Feedback Regulation of Type I IFN Production by the IFN-Inducible DNA Sensor cGAS. *J. Immunol.* **194**, 1545–1554 (2015).

158. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).

159. Upton, J. W., Kaiser, W. J. & Mocarski, E. S. Virus inhibition of RIP3-dependent necrosis. *Cell Host Microbe* **7**, 302–313 (2010).

160. Upton, J. W., Kaiser, W. J. & Mocarski, E. S. DAI/ZBP1/DLM-1 Complexes with RIP3 to Mediate Virus-Induced Programmed Necrosis that Is Targeted by Murine Cytomegalovirus vIRA. *Cell host & microbe* vol. 26 564 (2019).

161. DeFilippis, V. R., Alvarado, D., Sali, T., Rothenburg, S. & Früh, K. Human Cytomegalovirus Induces the Interferon Response via the DNA Sensor ZBP1. *J. Virol.* **84**, 585–598 (2010).

162. Ishii, K. J. *et al.* TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451**, 725–729 (2008).

163. Unterholzner, L. *et al.* IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* **11**, 997–1004 (2010).

164. Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat. Immunol.* **10**, 1065–1072 (2009).

165. Zhang, Z. *et al.* The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat. Immunol.* **12**, 959–965 (2011).

166. Zhang, Z. *et al.* DDX1, DDX21, and DHX36 Helicases Form a Complex with the Adaptor Molecule TRIF to Sense dsRNA in Dendritic Cells. *Immunity* **34**, 866–878 (2011).

167. Parvatiyar, K. *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).

168. Zhang, X. *et al.* Cutting Edge: Ku70 Is a Novel Cytosolic DNA Sensor That Induces Type III Rather Than Type I IFN. *J. Immunol.* **186**, 4541–4545 (2011).

169. Ferguson, B. J., Mansur, D. S., Peters, N. E., Ren, H. & Smith, G. L. DNA-PK is a DNA sensor for IRF-3-dependent innate immunity. *Elife* **2012**, (2012).

170. Kondo, T. *et al.* DNA damage sensor MRE11 recognizes cytosolic double-stranded DNA and induces type i interferon by regulating STING trafficking. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2969–2974 (2013).

171. Roth, S. *et al.* Rad50-CARD9 interactions link cytosolic DNA sensing to IL-1β production. *Nat. Immunol.* **15**, 538–545 (2014).

172. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nature Reviews Immunology* vol. 15 87–103 (2015).

173. Mesev, E. V., LeDesma, R. A. & Ploss, A. Decoding type I and III interferon signalling during viral infection. *Nature Microbiology* vol. 4 914–924 (2019).

174. Schneider, W. M., Chevillotte, D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. (2014) doi:10.1146/annurev-immunol-032713-120231.

175. Maelfait, J., Liverpool, L. & Rehwinkel, J. Nucleic Acid Sensors and Programmed Cell Death. *Journal of Molecular Biology* vol. 432 552–568 (2020).

176. Jorgensen, I., Rayamajhi, M. & Miao, E. A. Programmed cell death as a defence against infection. *Nature Reviews Immunology* vol. 17 151–164 (2017).

177. Paludan, S. R., Reinert, L. S. & Hornung, V. DNA-stimulated cell death: implications for host defence, inflammatory diseases and cancer. *Nature Reviews Immunology* vol. 19 141–153 (2019).

178. Orzalli, M. H. & Kagan, J. C. Apoptosis and Necroptosis as Host Defense Strategies to Prevent Viral Infection. *Trends in Cell Biology* vol. 27 800–809 (2017).

179. Frank, D. & Vince, J. E. Pyroptosis versus necroptosis: similarities, differences, and crosstalk. *Cell Death and Differentiation* vol. 26 99–114 (2019).

180. Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L. & Kroemer, G. Immunogenic cell death in cancer and infectious disease. *Nature Reviews Immunology* vol. 17 97–111 (2017).

181. Tang, C. H. A. *et al.* Agonist-mediated activation of STING induces apoptosis in malignant B cells. *Cancer Res.* **76**, 2137–2152 (2016).

182. Larkin, B. *et al.* Cutting Edge: Activation of STING in T Cells Induces Type I IFN Responses and Cell Death. *J. Immunol.* **199**, 397–402 (2017).

183. Diner, B. A., Lum, K. K., Toettcher, J. E. & Cristea, I. M. Viral DNA sensors IFI16 and cyclic GMP-AMP synthase possess distinct functions in regulating viral gene expression, immune defenses, and apoptotic responses during herpesvirus infection. *MBio* **7**, (2016).

184. Brault, M., Olsen, T. M., Martinez, J., Stetson, D. B. & Oberst, A. Intracellular Nucleic Acid Sensing Triggers Necroptosis through Synergistic Type I IFN and TNF Signaling. *J. Immunol.* **200**, 2748–2756 (2018).

185. Chattopadhyay, S. & Sen, G. C. RIG-I-like receptor-induced IRF3 mediated pathway of apoptosis (RIPA): a new antiviral pathway. *Protein and Cell* vol. 8 165–168 (2017).

186. Orzalli, M. H. *et al.* cGAS-mediated stabilization of IFI16 promotes innate signaling during herpes simplex virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1773-81 (2015).

187. Yang, H. *et al.* cGAS is essential for cellular senescence. doi:10.1073/pnas.1705499114.

188. Liu, H. *et al.* Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* **563**, 131–136 (2018).

189. Zierhut, C. & Funabiki, H. The cytoplasmic DNA sensor cGAS promotes mitotic cell death. *bioRxiv* 168070 (2017) doi:10.1101/168070.

190. Volkman, H. E., Cambier, S., Gray, E. E. & Stetson, D. B. Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *Elife* **8**, (2019).

191. Lahaye, X. *et al.* NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. *Cell* **175**, 488-501.e22 (2018).

192. West, A. P. *et al.* Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553–557 (2015).

193. Won, J. K. & Bakhoum, S. F. The cytosolic DNA-sensing cGAS–sting pathway in cancer. *Cancer Discovery* vol. 10 26–39 (2020).

194. Harding, S. M. *et al.* Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**, 466–470 (2017).

195. Spektor, A., Umbreit, N. T. & Pellman, D. Cell Biology: When Your Own Chromosomes Act like Foreign DNA. *Current Biology* vol. 27 R1228–R1231 (2017).

196. Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).

197. Glück, S. *et al.* Innate immune sensing of cytosolic chromatin fragments through cGAS promotes senescence. *Nat. Cell Biol.* **19**, 1061–1070 (2017).

198. Yang, H., Wang, H., Ren, U., Chen, Q. & Chena, Z. J. CGAS is essential for cellular senescence. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4612–E4620 (2017).

199. Dou, Z. *et al.* Cytoplasmic chromatin triggers inflammation in senescence and cancer. *Nature* **550**, 402–406 (2017).

200. Chen, Y. A. *et al.* Extrachromosomal telomere repeat DNA is linked to ALT development via cGAS-STING DNA sensing pathway. *Nat. Struct. Mol. Biol.* **24**, 1124–1131 (2017).

201. Barroso-González, J. *et al.* RAD51AP1 Is an Essential Mediator of Alternative Lengthening of Telomeres. *Mol. Cell* **76**, 217 (2019).

202. Wang, Z. *et al.* CGAS/STING axis mediates a topoisomerase II inhibitor–induced tumor immunogenicity. *J. Clin. Invest.* **129**, 4850–4862 (2019).

203. Reisländer, T. *et al.* BRCA2 abrogation triggers innate immune responses potentiated by treatment with PARP inhibitors. *Nat. Commun.* **10**, (2019).

204. Heijink, A. M. *et al.* BRCA2 deficiency instigates cGAS-mediated inflammatory signaling and confers sensitivity to tumor necrosis factor-alpha-mediated cytotoxicity. *Nat. Commun.* **10**, 1–14 (2019).

205. Sun, B. *et al.* Dengue virus activates cGAS through the release of mitochondrial DNA. *Sci. Rep.* **7**, 1–8 (2017).

206. Ni, G., Ma, Z. & Damania, B. cGAS and STING: At the intersection of DNA and RNA virus-sensing networks. *PLOS Pathog.* **14**, e1007148 (2018).

207. Rongvaux, A. *et al.* Apoptotic Caspases Prevent the Induction of Type I Interferons by Mitochondrial DNA. *Cell* **159**, 1563–1577 (2014).

208. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–337 (2013).

209. Kato, K. *et al.* Structural and Functional Analyses of DNA-Sensing and Immune Activation by Human cGAS. *PLoS One* **8**, e76983 (2013).

210. 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**, 1094–107 (2013).

211. Xie, W. *et al.* Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 11946–11955 (2019).

212. Kranzusch, P. J., Lee, A. S.-Y., Berger, J. M. & Doudna, J. A. Structure of Human cGAS Reveals a Conserved Family of Second-Messenger Enzymes in Innate Immunity. *Cell Rep.* **3**, 1362–1368 (2013).

213. Zhou, W. *et al.* Structure of the Human cGAS–DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell* **174**, 300-311.e11 (2018).

214. Andreeva, L. *et al.* cGAS senses long and HMGB/TFAM-bound U-turn DNA by forming protein–DNA ladders. *Nature* (2017) doi:10.1038/nature23890.

215. Tao, J. *et al.* Nonspecific DNA Binding of cGAS N Terminus Promotes cGAS Activation. *J. Immunol.* **198**, 1601909 (2017).

216. Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science (80-.).* **361**, 704–709 (2018).

217. 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.* **37**, 7701–7714 (2009).

218. Kranzusch, P. J., Lee, A. S. Y., Berger, J. M. & Doudna, J. A. Structure of Human cGAS Reveals a Conserved Family of Second-Messenger Enzymes in Innate Immunity. *Cell Rep.* **3**, 1362–1368 (2013).

219. Li, X. *et al.* Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019–31 (2013).

220. 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 Rep.* **6**, 421–430 (2014).

221. 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 Rep.* **6**, 421–430 (2014).

222. Hooy, R. M. & Sohn, J. The allosteric activation of cGAS underpins its dynamic signaling landscape. *Elife* **7**, (2018).

223. Du, M. & Chen, Z. J. DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science (80-.).* **361**, 704–709 (2018).

224. Mozzi, A. *et al.* OASes and STING: Adaptive evolution in concert. *Genome Biol. Evol.* **7**, 1016–1032 (2015).

225. Donovan, J., Dufner, M. & Korennykh, A. Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1652–1657 (2013).

226. Ross, P. *et al.* Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* **325**, 279–281 (1987).

227. Danilchanka, O. & Mekalanos, J. J. Cyclic dinucleotides and the innate immune response. *Cell* **154**, 962–70 (2013).

228. Kerr, I. M. & Brown, R. E. pppA2′p5′A2′p5′A: An inhibitor of protein synthesis synthesized with an enzyme fraction from interferon-treated cells. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 256–260 (1978).

229. Kranzusch, P. J. *et al.* Structure-guided reprogramming of human cgas dinucleotide linkage specificity. *Cell* **158**, 1011–1021 (2014).

230. Hall, J. *et al.* The catalytic mechanism of cyclic GMP-AMP synthase (cGAS) and implications for innate immunity and inhibition. *Protein Sci.* **26**, 2367–2380 (2017).

231. Ablasser, A. *et al.* Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP. *Nature* **503**, 530–4 (2013).

232. Zhou, C. *et al.* Transfer of cGAMP into Bystander Cells via LRRC8 Volume-Regulated Anion Channels Augments STING-Mediated Interferon Responses and Anti-viral Immunity. *Immunity* **52**, 767-781.e6 (2020).

233. Liu, H. *et al.* cGAS facilitates sensing of extracellular cyclic dinucleotides to activate innate immunity . *EMBO Rep.* **20**, (2019).

234. Cheng, Z. *et al.* The interactions between cGAS-STING pathway and pathogens. *Signal Transduction and Targeted Therapy* vol. 5 1–15 (2020).

235. Shang, G. *et al.* Crystal structures of STING protein reveal basis for recognition of cyclic di-GMP. *Nat. Struct. Mol. Biol.* **19**, 725–727 (2012).

236. 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.* **19**, 722–724 (2012).

237. Ouyang, S. *et al.* Structural Analysis of the STING Adaptor Protein Reveals a Hydrophobic Dimer Interface and Mode of Cyclic di-GMP Binding. *Immunity* **36**, 1073–1086 (2012).

238. 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.* **19**, 728–730 (2012).

239. Shang, G., Zhang, C., Chen, Z. J., Bai, X. chen & Zhang, X. Cryo-EM structures of STING reveal its mechanism of activation by cyclic GMP–AMP. *Nature* **567**, 389–393 (2019).

240. Zhao, B. *et al.* A conserved PLPLRT/SD motif of STING mediates the recruitment and activation of TBK1. *Nature* **569**, 718–722 (2019).

241. Zhao, B. *et al.* Structural basis for concerted recruitment and activation of IRF-3 by innate immune adaptor proteins. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3403–E3412 (2016).

242. Zhang, C. *et al.* Structural basis of STING binding with and phosphorylation by TBK1. *Nature* **567**, 394–398 (2019).

243. Ergun, S. L., Fernandez, D., Weiss, T. M. & Li, L. STING Polymer Structure Reveals Mechanisms for Activation, Hyperactivation, and Inhibition. *Cell* **178**, 290-301.e10 (2019).

244. de Oliveira Mann, C. C. *et al.* Modular Architecture of the STING C-Terminal Tail Allows Interferon and NF-κB Signaling Adaptation. *Cell Rep.* **27**, 1165-1175.e5 (2019).

245. Konno, H., Konno, K. & Barber, G. N. Cyclic dinucleotides trigger ULK1 (ATG1) phosphorylation of STING to prevent sustained innate immune signaling. *Cell* **155**, 688 (2013).

246. Goto, A. *et al.* The Kinase IKKβ Regulates a STING- and NF-κB-Dependent Antiviral Response Pathway in Drosophila. *Immunity* **49**, 225-234.e4 (2018).

247. Dunphy, G. *et al.* Non-canonical Activation of the DNA Sensing Adaptor STING by ATM and IFI16 Mediates NF-κB Signaling after Nuclear DNA Damage. *Mol. Cell* **71**, 745-760.e5 (2018).

248. Wu, X. *et al.* Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING. *Nucleic Acids Res.* **42**, 8243–8257 (2014).

249. Kranzusch, P. J. *et al.* Ancient Origin of cGAS-STING Reveals Mechanism of Universal 2',3' cGAMP Signaling. *Mol. Cell* **59**, 891–903 (2015).

250. Gui, X. *et al.* Autophagy induction via STING trafficking is a primordial function of the cGAS pathway. *Nature* **567**, 262–266 (2019).

251. Yamashiro, L. H. *et al.* Interferon-independent STING signaling promotes resistance to HSV-1 in vivo. *Nat. Commun.* **11**, 1–11 (2020).

252.   Liu, D. *et al.* STING directly activates autophagy to tune the innate immune response. *Cell Death Differ.* **26**, 1735–1749 (2019).

253.   Moretti, J. *et al.* STING Senses Microbial Viability to Orchestrate Stress-Mediated Autophagy of the Endoplasmic Reticulum. *Cell* **171**, 809-823.e13 (2017).

254.   Gaidt, M. M. *et al.* The DNA Inflammasome in Human Myeloid Cells Is Initiated by a STING-Cell Death Program Upstream of NLRP3. *Cell* **171**, 1110-1124.e18 (2017).

255.   Gonugunta, V. K. *et al.* Trafficking-Mediated STING Degradation Requires Sorting to Acidified Endolysosomes and Can Be Targeted to Enhance Anti-tumor Response. *Cell Rep.* **21**, 3234–3242 (2017).

256.   Ma, F. *et al.* Positive Feedback Regulation of Type I IFN Production by the IFN-Inducible DNA Sensor cGAS. *J. Immunol.* **194**, 1545–1554 (2015).

257.   Seo, G. J. *et al.* Akt Kinase-Mediated Checkpoint of cGAS DNA Sensing Pathway. *Cell Rep.* **13**, 440–449 (2015).

258.   Ning, S. & Wang, L. The Multifunctional Protein p62 and Its Mechanistic Roles in Cancers. *Curr. Cancer Drug Targets* **19**, 468–478 (2018).

259.   Dai, J. *et al.* Acetylation Blocks cGAS Activity and Inhibits Self-DNA-Induced Autoimmunity. *Cell* **176**, 1447-1460.e14 (2019).

260.   Hu, M. M. *et al.* Sumoylation Promotes the Stability of the DNA Sensor cGAS and the Adaptor STING to Regulate the Kinetics of Response to DNA Virus. *Immunity* **45**, 555–569 (2016).

261.   Cui, Y. *et al.* SENP7 Potentiates cGAS Activation by Relieving SUMO-Mediated Inhibition of Cytosolic DNA Sensing. *PLOS Pathog.* **13**, e1006156 (2017).

262.   Wang, Q. *et al.* The E3 ubiquitin ligase RNF185 facilitates the cGAS-mediated innate immune response. *PLOS Pathog.* **13**, e1006264 (2017).

263.   Wang, Y. *et al.* Inflammasome Activation Triggers Caspase-1-Mediated Cleavage of cGAS to Regulate Responses to DNA Virus Infection. *Immunity* **46**, 393–404 (2017).

264.   Ning, X. *et al.* Apoptotic Caspases Suppress Type I Interferon Production via the Cleavage of cGAS, MAVS, and IRF3. *Mol. Cell* **74**, 19-31.e7 (2019).

265.   Xia, P. *et al.* Glutamylation of the DNA sensor cGAS regulates its binding and synthase activity in antiviral immunity. *Nat. Immunol.* **17**, 369–378 (2016).

266.   Li, T. *et al.* Phosphorylation and chromatin tethering prevent cGAS activation during mitosis. *Science (80-.).* eabc5386 (2021) doi:10.1126/science.abc5386.

267. Tanaka, Y. & Chen, Z. J. STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci. Signal.* **5**, ra20 (2012).

268. Tsuchida, T. *et al.* The ubiquitin ligase TRIM56 regulates innate immune responses to intracellular double-stranded DNA. *Immunity* **33**, 765–776 (2010).

269. Zhang, L. *et al.* The deubiquitinase CYLD is a specific checkpoint of the STING antiviral signaling pathway. *PLOS Pathog.* **14**, e1007435 (2018).

270. Zhong, B. *et al.* The Ubiquitin Ligase RNF5 Regulates Antiviral Responses by Mediating Degradation of the Adaptor Protein MITA. *Immunity* **30**, 397–407 (2009).

271. Zhang, J., Hu, M. M., Wang, Y. Y. & Shu, H. B. TRIM32 protein modulates type I interferon induction and cellular antiviral response by targeting MITA/STING protein for K63-linked ubiquitination. *J. Biol. Chem.* **287**, 28646–28655 (2012).

272. Wang, Q. *et al.* The E3 Ubiquitin ligase AMFR and INSIG1 bridge the activation of TBK1 kinase by modifying the adaptor STING. *Immunity* **41**, 919–933 (2014).

273. Ni, G., Konno, H. & Barber, G. N. Ubiquitination of STING at lysine 224 controls IRF3 activation. *Sci. Immunol.* **2**, 7119 (2017).

274. Mukai, K. *et al.* Activation of STING requires palmitoylation at the Golgi. *Nat. Commun.* **7**, 1–10 (2016).

275. Rawling, D. C. & Pyle, A. M. Parts, assembly and operation of the RIG-I family of motors. *Current Opinion in Structural Biology* vol. 25 25–33 (2014).

276. Cui, S. *et al.* The C-Terminal Regulatory Domain Is the RNA 5′-Triphosphate Sensor of RIG-I. *Mol. Cell* **29**, 169–179 (2008).

277. Zheng, J. *et al.* High-resolution HDX-MS reveals distinct mechanisms of RNA recognition and activation by RIG-I and MDA5. *Nucleic Acids Res.* **43**, 1216–1230 (2015).

278. Ramanathan, A. *et al.* The autoinhibitory CARD2-Hel2i Interface of RIG-I governs RNA selection. *Nucleic Acids Res.* **44**, 896–909 (2016).

279. Civril, F. *et al.* The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO Rep.* **12**, 1127–1134 (2011).

280. Ludwig, J. *et al.* Structural and functional insights into 5′-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nat. Struct. Mol. Biol.* **17**, 781–787 (2010).

281. Jiang, F. *et al.* Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature* **479**, 423–427 (2011).

282. Jiang, X. *et al.* Ubiquitin-Induced Oligomerization of the RNA Sensors RIG-I and MDA5 Activates Antiviral Innate Immune Response. *Immunity* **36**, 959–973 (2012).

283. Liu, H. M. *et al.* The mitochondrial targeting chaperone 14-3-3ε regulates a RIG-I translocon that mediates membrane association and innate antiviral immunity. *Cell Host Microbe* **11**, 528–537 (2012).

284. Luo, D., Kohlway, A. & Pyle, A. M. Duplex RNA activated ATPases (DRAs): Platforms for RNA sensing, signaling and processing. *RNA Biology* vol. 10 111–120 (2013).

285. Myong, S. *et al.* Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science (80-.).* **323**, 1070–1074 (2009).

286. Lewis, R., Dürr, H., Hopfner, K. P. & Michaelis, J. Conformational changes of a Swi2/Snf2 ATPase during its mechano-chemical cycle. *Nucleic Acids Res.* **36**, 1881–1890 (2008).

287. Bamming, D. & Horvath, C. M. Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *J. Biol. Chem.* **284**, 9700–9712 (2009).

288. Lässig, C. *et al.* ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *Elife* **4**, (2015).

289. Rawling, D. C., Fitzgerald, M. E. & Pyle, A. M. Establishing the role of ATP for the function of the RIG-I innate immune sensor. *Elife* **4**, (2015).

290. Oda, H. *et al.* Aicardi-goutières syndrome is caused by IFIH1 mutations. *Am. J. Hum. Genet.* **95**, 121–125 (2014).

291. Jang, M. A. *et al.* Mutations in DDX58, which encodes RIG-I, Cause atypical singleton-merten syndrome. *Am. J. Hum. Genet.* **96**, 266–274 (2015).

292. Cunninghame Graham, D. S. *et al.* Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with Systemic Lupus Erythematosus. *PLoS Genet.* **7**, e1002341 (2011).

293. Liu, S. *et al.* IFIH1 polymorphisms are significantly associated with type 1 diabetes and IFIH1 gene expression in peripheral blood mononuclear cells. *Hum. Mol. Genet.* **18**, 358–365 (2009).

294. 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 (80-.).* **339**, 786–791 (2013).

295. Gentili, M. *et al.* The N-Terminal Domain of cGAS Determines Preferential Association with Centromeric DNA and Innate Immune Activation in the Nucleus. *Cell Rep.* **26**, 2377-2393.e13 (2019).

296. Xia, P. *et al.* A Circular RNA Protects Dormant Hematopoietic Stem Cells from DNA Sensor cGAS-Mediated Exhaustion. *Immunity* **48**, 688–701 (2018).

297. Orzalli, M. H. *et al.* CGAS-mediated stabilization of IFI16 promotes innate signaling during herpes simplex virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1773–E1781 (2015).

298. Wu, J. J. *et al.* Inhibition of cGAS DNA Sensing by a Herpesvirus Virion Protein. *Cell Host Microbe* **18**, 333–344 (2015).

299. Lahaye, X. *et al.* The Capsids of HIV-1 and HIV-2 Determine Immune Detection of the Viral cDNA by the Innate Sensor cGAS in Dendritic Cells. *Immunity* **39**, 1132–1142 (2013).

300. Rasaiyaah, J. *et al.* HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* **503**, 402–405 (2013).

301. Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).

302. Michalski, S. *et al.* Structural basis for sequestration and autoinhibition of cGAS by chromatin. *Nature* **587**, 678–682 (2020).

303. Kujirai, T. *et al.* Structural basis for the inhibition of cGAS by nucleosomes. *Science (80-.).* **370**, 455–458 (2020).

304. Boyer, J. A. *et al.* Structural basis of nucleosome-dependent cGAS inhibition. *Science (80-.).* **370**, 450–454 (2020).

305. Cao, D., Han, X., Fan, X., Xu, R. M. & Zhang, X. Structural basis for nucleosome-mediated inhibition of cGAS activity. *Cell Res.* **30**, 1088–1097 (2020).

306. Zhao, B. *et al.* The molecular basis of tight nuclear tethering and inactivation of cGAS. *Nature* **587**, 673–677 (2020).

307. Zierhut, C. *et al.* The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death. *Cell* **178**, 302-315.e23 (2019).

308. Pathare, G. R. *et al.* Structural mechanism of cGAS inhibition by the nucleosome. *Nature* **587**, 668–672 (2020).

309. Bartsch, K. *et al.* Absence of RNase H2 triggers generation of immunogenic micronuclei removed by autophagy. *Hum. Mol. Genet.* **26**, 3960–3972 (2017).

310. Kwon, M., Leibowitz, M. L. & Lee, J. H. Small but mighty: the causes and consequences of micronucleus rupture. *Experimental and Molecular Medicine* vol. 52 1777–1786 (2020).

311. Flynn, P. J., Koch, P. D. & Mitchison, T. J. Chromatin Bridges, not Micronuclei, Activate cGAS after Drug-induced Mitotic Errors in Human Cells. *bioRxiv* 2021.02.02.429360 (2021) doi:10.1101/2021.02.02.429360.

312. Müller, S. & Almouzni, G. Chromatin dynamics during the cell cycle at centromeres. *Nature Reviews Genetics* vol. 18 192–208 (2017).

313. Burley, M., Roberts, S. & Parish, J. L. Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle. *Seminars in Immunopathology* vol. 42 159–171 (2020).

314. Favre, M., Breitburd, F., Croissant, O. & Orth, G. Chromatin-like structures obtained after alkaline disruption of bovine and human papillomaviruses. *J. Virol.* **21**, (1977).

315. Porter, S. S. *et al.* Histone Modifications in Papillomavirus Virion Minichromosomes. *MBio* **12**, (2021).

316. Jiang, H. *et al.* Chromatin-bound cGAS is an inhibitor of DNA repair and hence accelerates genome destabilization and cell death. *EMBO J.* **38**, e102718 (2019).

317. Basu, A. K. DNA damage, mutagenesis and cancer. *International Journal of Molecular Sciences* vol. 19 (2018).

318. Xia, T., Konno, H., Ahn, J. & Barber Correspondence, G. N. Deregulation of STING Signaling in Colorectal Carcinoma Constrains DNA Damage Responses and Correlates With Tumorigenesis Accession Numbers GSE75205. (2016) doi:10.1016/j.celrep.2015.12.029.

319. Konno, H. *et al.* Suppression of STING signaling through epigenetic silencing and missense mutation impedes DNA damage mediated cytokine production. *Oncogene* **37**, 2037–2051 (2018).

320. Rodier, F. *et al.* Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nat. Cell Biol.* **11**, 973–979 (2009).

321. Brzostek-Racine, S., Gordon, C., Van Scoy, S. & Reich, N. C. The DNA Damage Response Induces IFN. *J. Immunol.* **187**, 5336–5345 (2011).

322. Ahn, J., Konno, H. & Barber, G. N. Diverse roles of STING-dependent signaling on the development of cancer. *Oncogene* **34**, 5302–5308 (2015).

323. Liu, H. *et al.* Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* **563**, 131–136 (2018).

324. Leung, J. W. *et al.* Nucleosome Acidic Patch Promotes RNF168- and RING1B/BMI1-Dependent H2AX and H2A Ubiquitination and DNA Damage Signaling. *PLoS Genet.* **10**, e1004178 (2014).

325. Nassour, J. *et al.* Autophagic cell death restricts chromosomal instability during replicative crisis. *Nature* **565**, 659–663 (2019).

326. Zierhut, C. *et al.* The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death. *Cell* **178**, 302-315.e23 (2019).

327. Cui, S. *et al.* Nuclear cGAS Functions Non-canonically to Enhance Antiviral Immunity via Recruiting Methyltransferase Prmt5. *Cell Rep.* **33**, 108490 (2020).

328. Migliori, V. *et al.* Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat. Struct. Mol. Biol.* **19**, 136–145 (2012).

329. Sun, H. *et al.* A Nuclear Export Signal Is Required for cGAS to Sense Cytosolic DNA. *Cell Rep.* **34**, 108586 (2021).

330. Hornung, V. *et al.* 5′-Triphosphate RNA is the ligand for RIG-I. *Science (80-.).* **314**, 994–997 (2006).

331. Schlee, M. *et al.* Recognition of 5′ Triphosphate by RIG-I Helicase Requires Short Blunt Double-Stranded RNA as Contained in Panhandle of Negative-Strand Virus. *Immunity* **31**, 25–34 (2009).

332. Takahasi, K. *et al.* Nonself RNA-Sensing Mechanism of RIG-I Helicase and Activation of Antiviral Immune Responses. *Mol. Cell* **29**, 428–440 (2008).

333. Louber, J., Brunel, J., Uchikawa, E., Cusack, S. & Gerlier, D. Kinetic discrimination of self/non-self RNA by the ATPase activity of RIG-I and MDA5. *BMC Biol.* **13**, 54 (2015).

334. Yao, H. *et al.* ATP-Dependent effector-like functions of RIG-I-like receptors. *Mol. Cell* **58**, 541–548 (2015).

335. Zheng, J. *et al.* HDX-MS reveals dysregulated checkpoints that compromise discrimination against self RNA during RIG-I mediated autoimmunity. *Nat. Commun.* **9**, 1–12 (2018).

336. Kohlway, A., Luo, D., Rawling, D. C., Ding, S. C. & Pyle, A. M. Defining the functional determinants for RNA surveillance by RIG-I. *EMBO Rep.* **14**, 772–779 (2013).

337. Rawling, D. C., Fitzgerald, M. E. & Pyle, A. M. Establishing the role of ATP for the function of the RIG-I innate immune sensor. *Elife* **4**, (2015).

338. Dickey, T. H., Song, B. & Pyle, A. M. RNA binding activates RIG-I by releasing an autorepressed signaling domain. *Sci. Adv.* **5**, eaax3641 (2019).

339. Devarkar, S. C., Schweibenz, B., Wang, C., Marcotrigiano, J. & Patel, S. S. RIG-I Uses an ATPase-Powered Translocation-Throttling Mechanism for Kinetic Proofreading of RNAs and Oligomerization. *Mol. Cell* **72**, 355-368.e4 (2018).

6 List of abbreviations

Å	Ångström
ac	acetylation
ACP	architectural chromatin protein
ADP	adenosine-diphosphate
AGS	Aicardi-Goutières-Syndrome
AIM2	absent in melanoma 2
ALR	AIM2-like-receptors
AMP	adenosine-monophosphate
ASC	associated speck-like protein containing a CARD
ATP	adenosine-triphosphate
AurKB	Aurora Kinase B
BAF	barrier-to-autointegration factor
BAH	bromoassociated homology
BLK	B-lymphocyte kinase
bp	base pairs
CARD	caspase activation and recruitment domain
CCFs	cytosolic chromatin fragments
CDN	cyclic dinucleotides
cGAMP	cyclic GMP–AMP
cGAS	cyclic GMP-AMP synthase
ChIP-seq	chromatin immunoprecipitation coupled with sequencing
CLR	c-type lectin receptors
CRM1	chromosomal region maintenance 1
cryo-EM	cryo-electron microscopy
CTD	C-terminal domain
CTT	C-terminal tail
DAMP	damage associated molecular pattern
DDR	DNA damage repair
DNA	desoxyribonucleic acid
DNA-PK	DNA dependent protein kinase
dsDNA	double-stranded DNA
ECD	ectodomain
ECTR	extrachromosomal telomere repeats
EMANIC	Electron microscopy-assisted nucleosome capture
ERGIC	ER-Golgi intermediate compartment
GMP	guanosine-monophosphate
H2A	histone variant 2A (different variants with same nomenclature)
hcGAS	human cGAS
Hel	helicase domain
HIV	human immunodeficiency virus
hp-RNA	hairpin RNA
HR	homologous recombination

HSV	herpes simplex virus
IE1	immediate early 1
IFI16	interferon inducible protein 16
IFN	interferon
IKK	IκB kinase
IL	interleukin (different variants)
IRAK	IL-1-receptor-associated kinases
IRF	interferon-regulatory factor
ISG	interferon-stimulated gene
kDa	kilo Dalton
KSHV	Kaposi´s sarcoma-associated herpesvirus
LANA	Latency-associated nuclear antigen
LBD	ligand binding domain
LGP2	laboratory of genetics and physiology 2
LLPS	liquid-liquid phase separation
LRR	leucin rich repeat
MAB21	male abnormal 21
MAVS	mitochondrial antiviral-signaling protein
mcGAS	mouse cGAS
MDA5	melanoma differentiation-associated protein 5
Me	methylation
MEFs	mouse embryonic fibroblasts
MHC	major histocompatibility complex
mtDNA	mitochondrial DNA
MyD88	myeloid differentiation primary-response protein 88
NCP	nucleosome core particle
NF-κB	nuclear factor κB
nM	nanomolar
NOD	nucleotide binding domain
NONO	non-POU domain-containing octamer binding protein
NTD	N-terminal domain
OAS	oligoadenylatsynthase
ox	oxidized
PAMP	pathogen associated molecular pattern
PFV	prototype foamy virus
PMRT5	Protein Arginine Methyltransferase 5
PRR	pattern recognition receptor
PTM	post-translational modification
PYD	pyrin-domain
RCC1	regulator of chromatin condensation 1
RD	regulatory domain
RDH	replication-dependent histone
RIG-I	retinoic acid inducible gene I
RIPA	RIG-I-like receptor-induced IRF3 mediated pathway of apoptosis

RNA	ribonucleic acid
SASP	senescence-associated secretory phenotype
SEC-MALS	size exclusion chromatography coupled with in-line multiangle light scattering
SELEX	systematic evolution of ligands by exponential enrichment
SF2	superfamily 2
SHL	superhelix location
Sir3	silent information regulator 3
SMS	Singleton-Merten-Syndrome
STAT	signal transducer and activator of transcription
WT	wild-type

RNA   ribonucleic acid

7 Acknowledgement

First and foremost, I want to thank my supervisor Prof. Dr. Karl-Peter Hopfner for giving me the opportunity to work on this exciting innate immunity project. The unique combination of different biological research areas and the related expertise collaborated and made the findings in this work possible. I want to thank him for giving me the freedom to develop and follow my own ideas and still believing in the project in times of setbacks. Thank you for always seeing the bigger picture. I am extremely grateful to Dr. Carina Mann for joining this project. Her impressive experience in the field of innate immunity enabled the timely success of this highly competitive project. Thank you for your guidance and the work you put into this project. I wish you all the best for your well-deserved own group. Of course, I also want to thank the other people involved in this thesis. Appreciation to Prof. Dr. Veit Hornung and Dr. Che Stafford for complementing our knowledge on the cellular side of innate immunity. Huge thanks to Dr. Joseph Bartho and Dr. Katja Lammens for numerous unsuccessful screening sessions, your patience and introducing me to cryo-EM in general. Also, I am very thankful for Dr.Gregor Witte always helping me in case of biophysical questions and keeping the lab running on the technical side. Thank you for your support and always having an open door. Furthermore, I want to thank Kevin Schall and Dr. Vanessa Niebauer for sharing their nucleosome biology expertise facilitating my foray into the field of nuclear cGAS.

I am very grateful to Dr. Liudmila Andreeva, Dr. Charlotte Lässig and Dr. David Drexler for introducing me to the Hopfner lab and the field of innate immunity during my Master´s thesis. Generally, I want to thank all members of the Hopfner lab for the great working atmosphere, support and scientific discussions. Special thanks to the permanent members of the Hopfner lab Brigitte, Alex, Manu and Olga for keeping the lab running efficiently.

Von ganzem Herzen möchte ich dem erweiterten Kreis meiner Familie danken. Vielen Dank an meine Eltern und Großeltern für eure Liebe, euer Verständnis und eure Unterstützung über all die Jahre hinweg. Ebenso danke ich Wilhelm Behrens, ohne dessen bedingungsloser Unterstützung ich wahrscheinlich nicht in der Lage wäre diese Zeilen zu verfassen. Meiner Partnerin, besten Freundin und Lieblingswissenschaftlerin Vicky möchte Ich von tiefstem Herzen danken. Dank dir war selbst der schlimmste Tag im Labor im nu vergessen. Auch möchte ich deiner Familie danken, mich wie ein Teil von Ihrer zu behandeln.