Structural studies of Toxins and Toxin-like Proteins

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Hugo Aragão Correia

von Portugal

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Sebastian Hiller
Prof. Dr. Daniel Müller

Basel, 13.12.2016

Prof. Dr. Jörg Schibler
The Dean of the Faculty

Table of contents

If you can’t explain it simply, you don’t understand it well enough.

Albert Einstein

Summary

Toxins are an ancient mechanism of interaction between cohabiting organisms: basal concentrations serve as an informal cue, enough as a warning signal; too much and the dialog is over. As such, the evolutionary race to arms led to the development of a vast trove of molecular unique biochemical mechanisms, from small molecules to protein toxins. The study of these mechanisms is not only essential for the treatment of toxin-related pathologies, but also as the potential source for novel therapeutic drugs.

In this thesis, a series of studies of different toxins and toxin-like proteins are compiled. To further understand the biological function and relevance of each toxin, their detailed study and characterization were pursued. Here are described the advances made using a combination of different complementary biophysical and structural methods, chosen in each case to specifically target each molecule characteristics. In the first chapter, the general biological theme of this thesis is introduced: toxins, particularly protein toxins, their description, and classification, as well as the role of structural biology in the study of proteins in general. To set the theoretical background of the following chapters, are also described the general principles of two of the most prominent methods for the study of proteins in structural biology: nuclear magnetic resonance (NMR) spectroscopy, and X-ray diffraction. In the second chapter, the interaction between human FKBP12 chaperone protein and two similar bacterial small molecule toxins is detailed: rapamycin initially used as an anti-fungal before the discovery of its potent immunosuppressive properties as a mTOR inhibitor; and mycolactone, a bacterial toxin responsible for the disease Buruli ulcers in humans. In the third chapter, the cell-free protein expression system is introduced as a technique best suited for the expression of cytotoxic proteins and otherwise difficult targets, as explored further in the following chapters. In the fourth chapter, advancements towards the structural and conformational characterization of the membrane-inserted state of two similar pore-forming toxins are detailed: the bacterial Colicin Ia protein; and the human Bax protein, an apoptosis effector; using X-ray crystallography, solution NMR and solid-state NMR. Finally, in the fifth chapter, two FIC-domain bacterial toxins are investigated: the bacterial VbhTA toxin-antitoxin protein complex, and the structural determination with its cognate target, DNA GyraseB enzyme; and the auto-activation of the bacterial NmFIC protein; in both cases using a combination of X-ray crystallography and NMR spectroscopy, as well as other biophysical techniques.

Chapter 1

Introduction

1. General introduction

1.1. Toxins

Toxins are the collective term given to a broad family of toxic molecules produced by different organisms, from microorganisms to animals. Their sole purpose is to directly exploit the negative impact on target organisms, towards maximizing the chances of survival of the producer organism. Thus, toxicity occurs upon direct toxin contact or absorption by the target organism. Different organisms produce a structural and functional diverse molecular set of toxins, with widely different mechanisms of action and consequently targets. Toxins exert their function through interaction with membranes, cell-receptors, enzymes, or other macromolecules of the target organism, which lead to the impairment or even disruption of its homeostasis (1, 2).

Prokaryotic toxins are kept in an inactive state due to either the lack of a cognate target, or complex formation with a specific antitoxin. Hence, toxins are released either through excretion (exotoxins) or cell lysis (endotoxins), targeting hosts through different transport systems such as secretory systems, pore-formation, host-receptor binding, and other endocytic pathways (3–5).

Eukaryotic toxins are termed venoms and exhibit two main functions: predation and defense. The complex chemical composition of venoms encompasses a mixture of different toxin types, often targeting several targets. For instance, each of the hundreds of cone snail species (genus *Conus*) has a specific venom profile, containing a mixture of up to thousands of different peptides, each targeting a specific target organism receptor (6, 7).

A constant survival pressure represents the driving force for the continuous pursuit of novel toxins with new biochemical properties, aiming towards new targets. In fact, toxin-encoding genes are evolutionary hot spots within genomes, which lead to highly complex and diverse toxins (8, 9). In general, the increase in organism complexity from prokaryotes to eukaryotes led to the increase of both functional and structural complexity of toxins, from small molecules to peptides, and protein toxins. Moreover, below are described the known distinct toxin classes.

1.1.1. Small molecule toxins

Small molecule toxins are a vast family of organic molecules of low molecular mass produced by different organisms. Due to wild habitat and nutrient competition, bacteria, cyanobacteria, fungi, and other species have developed a diverse arsenal of small molecule toxins, ranging from alkaloids and

macrolides to polyamines, which target host protein receptors as well as enzymes with distinct mechanisms and pathologies (1, 3, 5, 10). As an example, the Ochratoxin, a common food contaminant produced by the *Penicillium* and *Aspergillus* fungi species, is a known nephrotoxic with pleiotropic effects in animals including humans, such as inhibition of macromolecular synthesis, increased lipid peroxidation and inhibition of mitochondrial respiration (11, 12). Furthermore, small molecule toxins are also found in the venom of several animal species (13). In chapter 3, two examples of small molecule toxins and their interaction with a cognate protein target are described and discussed: the known complex of rapamycin with the FK-506 binding protein 12 kilodalton (FKBP12), and the proposed complex of the exotoxin mycolactone with the FKBP12 protein.

1.1.2. Peptide toxins

Peptide toxins are expressed by many organisms with different activities, and are mostly categorized into three classes: receptor-binding peptides, as ion channels activity modulators (6, 13, 14); membrane destabilizing/permeating peptides (13–15); and enzyme inhibitors (16, 17). Accordingly, peptide toxins exhibit a variety of pathologies, though much of their properties, cellular targets, and functional mechanisms remain poorly understood. For instance, the cone snail produces peptides termed conotoxins for both defense and predation. These conotoxins are hypervariable peptides, known to modulate the ion channel activity of the target organism (6, 7).

1.1.3. Protein toxins

Protein toxins are a vast group of proteins with widely diverse targets and mechanisms of action (2). Toxins target the host organism essentially at three levels: extracellular space, including tissues and fluids; cell membrane; and intracellular space. Membrane interacting toxins can be further classified according to their mechanism: receptor modulators, membrane disruptors, and pore-formers.

Extracellular toxins

Certain toxins act on host fluids and tissues by the impairment of host defenses to further promote toxin and pathogen diffusion. Snake venom serine proteases are known to modulate the host homeostasis such as blood coagulation, blood pressure, fibrinolysis, and the complement as well as nervous systems in addition to prompting host tissue-digestion (18–20).

Receptor modulator toxins

This group of toxins binds directly to a host cell membrane cognate receptor, leading to the disruption of cell signaling pathways through an agonist or antagonist effect on the receptor. For instance, the

main component of the Elapidae snake's venom: the three-fingered toxin α-cobratoxin (21), causes paralysis upon prey injection by binding competitively to different nicotinic acetylcholine receptors and thus preventing acetylcholine binding (22).

Membrane disruptors

Certain toxins disrupt host membrane integrity to facilitate pathogen dissemination and promote disease. As such, the first ever described toxin enzyme: *Clostridium perfringens* phospholipase α-toxin, is the major pathogenic factor responsible for the infection *Clostridial myonecrosis*, known as the common gas gangrene (23). The phospholipase cleaves phospholipids into smaller secondary messengers to disrupt the host's signaling, promoting further infection (24).

Pore-forming toxins

Pore-forming protein toxins (PFTs) are one of Nature's most common and effective biochemical weapons. These toxins are found in a wide range of organisms, particularly bacteria, yet share remarkable domain organization, structure, and function (25, 26). PFTs are classified according to the secondary structure of its members, either α-helical (α-PFTs) or β-sheet (β-PFTs) proteins (25). Members of the α-PFTs group, are generally expressed as non-active soluble α-helical bundles, where the core hydrophobic α-helices are surrounded by amphipathic α-helices. The soluble structure of β-PFTs is mainly composed of amphipathic β-sheets, which combine through oligomerization to form integral membrane β-barrels. Due to their tendency to form fixed-stoichiometric oligomeric pores, β-PFTs are the best-characterized class (27–29). In both cases, membrane interaction leads to the formation of an integral membrane pore, as depicted in Figure 1-1. Pore formation enables either cell-permeation of the cytotoxic domain, or directly to membrane disruption and consequent cell-death.

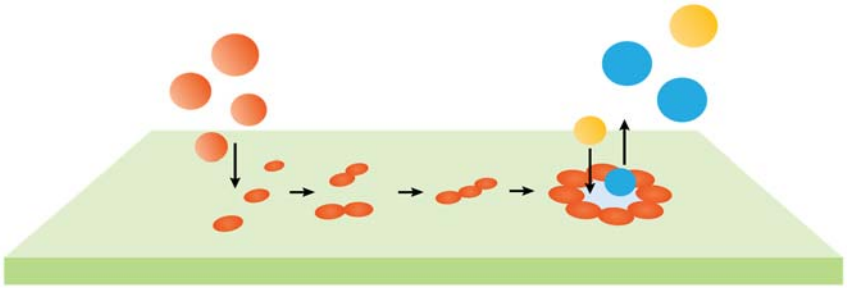


Figure 1-1 – Scheme of pore-formation. Pore-forming toxins, in orange, oligomerize upon membrane insertion leading to the formation of a membrane pore permeable to different molecules.

Several α-PFT's share a common structural topology, in particular members of the Colicin and *Bacillus thuringiensis* Cry family of pore-forming toxins share a similar fold (30, 31). These proteins exhibit a similar helical bundle pore-forming domain, where the central hydrophobic helix or hairpin leads to the initial membrane insertion and consequent pore formation. Similarly, the T-domain of Diphtheria

toxin shows an inner hydrophobic hairpin, surrounded by several amphiphilic helices, as depicted in Figure 1-2 (32).

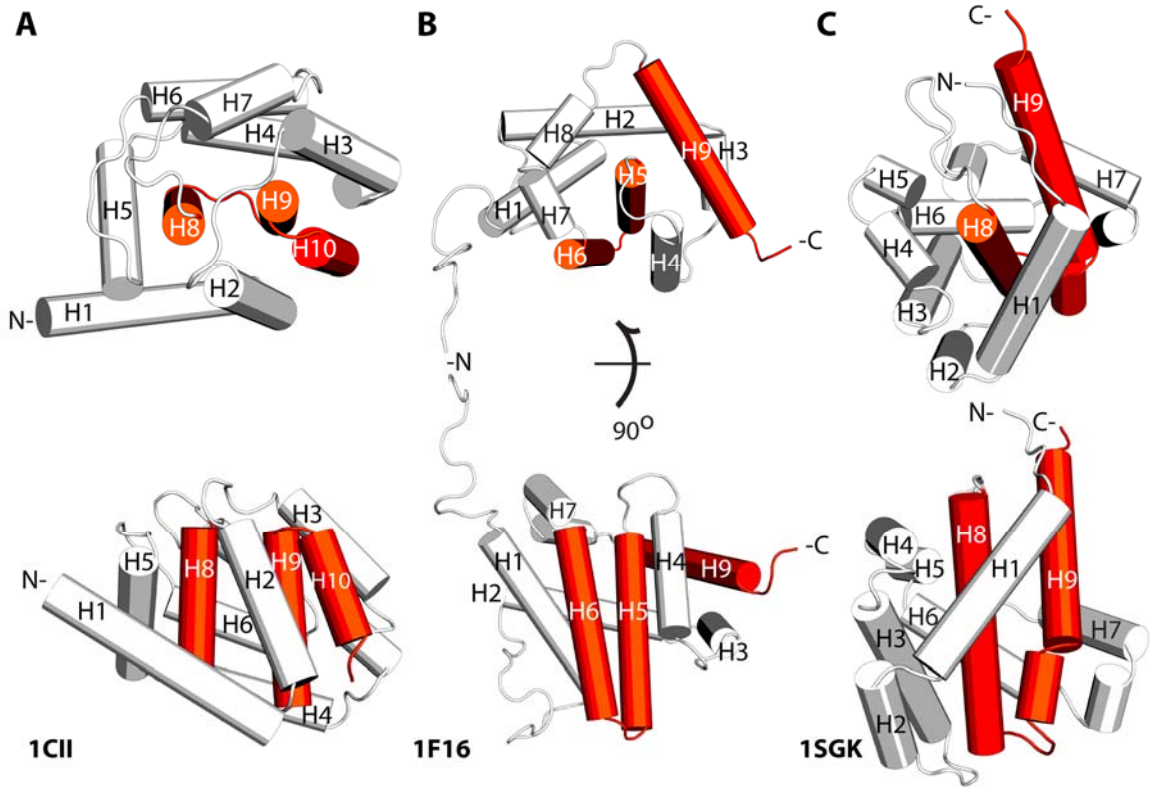


Figure 1-2 – Structures of (A) Colicin IA C-terminal 447-667, (B) human Bcl-2 member Bax and (C) Diphtheria toxin T-domain with annotated helices and the respective PDB ID. Highlighted in color is the pore-forming domain: the hairpin-forming central hydrophobic helices. In all cases, a helical bundle of hydrophilic helices shields the hydrophobic hairpin in its core.

This domain forms a membrane pore responsible for the translocation of its catalytic channel domain (C-domain) into the host (33). Interestingly, eukaryotic organisms evolved PFT-like proteins such as perforins as part of the innate immune response (34), and the Bcl-2 protein family as part of the cellular apoptosis regulation (Figure 1-2B) (35). Structurally, the common denominator of all α-PFT's seems to be the nuclear hydrophobic hairpin surrounded by amphipathic helices highlighted in Figure 1-2, indicating a common membrane insertion mechanism.

Still, much is unknown about the structure and mechanisms of most PFTs (36). In chapter 4, two examples: the *E. coli* Colicin Ia C-domain; and the PFT-like human Bcl-2 member Bax; are described as well as the advances towards their structural characterization.

Intracellular toxins

Intracellular protein toxins modify host-specific targets through a variety of mechanisms such as post-translational modification (PTM) (37–39), cross-linking (40) protein hydrolysis or even non-covalent interactions (37), in order to modulate and manipulate host cell physiology beneficially to the toxic

organism. Due to the complexity of cell-machinery, toxins have evolved to exploit most of the cell macromolecular infrastructures, often in a pleiotropic fashion. For instance, the *E. coli* secreted protein G (EspG), from enterohaemorrhagic *E. coli*, is implicated in tight junction disruption (37, 41). In the host-cell, EspG binds non-covalently to the host guanosine triphosphate (GTP) hydrolases (GTPases) of the ADP-ribosylation factor (ARF) family, locking them in the active ARF-GTP complex state, which in turn inhibits endogenous-ligand binding, effectively blocking Golgi traffic (42).

Enzyme-catalyzed PTM is a common mechanism to alter protein function through covalent attachment or removal of functional groups, such as AMPylation: the catalytic addition of an adenosine monophosphate (AMP) group. The filamentation induced by cyclic AMP (FIC) protein family, targets proteins for PTM altering their endogenous function, usually through AMPylation. This family is found in most organisms (38), whereas in bacteria these are present as toxin-antitoxin (TA) complexes, which use different modification mechanisms to modulate host physiology. Interestingly, FIC proteins are also present in eukaryotic organisms. The human Huntingtin yeast-interaction protein E (HypE) protein acts as an on-off switch through AMPylation of the Binding immunoglobulin protein (BiP), a major chaperone in the unfolded protein response residing in the endoplasmatic reticulum (ER) (122).

In chapter 5, the structural and functional characterization of two bacterial intracellular PTM toxins is described: the *Neisseria meningitidis* FIC (NmFIC); and the *Bartonella schoenbuchensis* octopine Ti plasmid virulence B homologous T and A complex (VbhT/A), in complex with its cognate target, the *Escherichia coli* DNA topoisomerase IIa Gyrase B (Gyr B).

1.1.4. Toxins as drug development leads

The unique properties of toxins such as: receptor targeting and cell type specificity (43), the capacity to evade the innate immune system, cell proteases and degradation systems (15, 44), as well as their distinctive catalytic mechanisms, make them an ideal lead for drug development and drug delivery (45–49). In fact, many toxins were discovered through drug lead screening programs (50). Scientific research and pharmaceutical industry have in the past explored toxins mostly as antibiotics, with very successful results, although its excessive and often neglectful use has led to the rise of multiresistant pathogens (51). Several potential therapeutic toxins have been described in the past (52); in fact, numerous toxin-based pharmaceuticals are currently in clinical trial, and some approved products made it to market launches (53, 54). Likewise, the pertussis toxin produced by *B. pertussis*, responsible for whooping cough, is in clinical research for its therapeutic role in a number of common diseases, such as hypertension (55), autoimmune diseases (56), and human immunodeficiency virus replication

(57). Effective research towards the better understanding of the structure and action mechanisms of toxins provides the basis for new therapies towards pathogenicity and venom antitoxins. Additionally, and more impactful, the extremely ingenious chemical space exploration of toxin molecules, paves the way for the development of new strategies based on toxin derivatives, towards novel therapies for human pathologies.

1.2. Structural Biology for protein structure and conformation characterization

With the advent of high-throughput genome sequencing and the completion of the Human Genome Project (58), as well as other genomes, the available amount of protein sequence data has grown exponentially. Nevertheless, the knowledge of protein sequence is not sufficient for the determination of protein folding and function, representing only the first step towards this goal. The understanding of proteome function requires extensive knowledge on both each protein's interacting-network, as well as their respective chemical environment; moreover, these fundamental characteristics depend on each distinct protein structure. Each protein's exposed surface properties, such as charge, as well as hydrophobicity, minutely detail these interactions and its extent. Besides, protein flexibility and conformational changes further modulate the functional properties of proteins. Structural biology aims to determine and subsequently understand protein structures as well as dynamics at atomic resolution. Further, this knowledge is not limited to understand the functional principles of each organism, but as well to improve the current state of drug research. As a consequence, structural knowledge permits the use of rational design for the improvement of current drugs, as well as the identification of new drug leads for the research and subsequent development of novel drugs and therapies with improved efficacy (59).

Currently, the three foremost techniques for structure determination at atomic or near-atomic resolution in biology are X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy. These research techniques vary significantly in their approach to structure determination. Each technique different advantages and disadvantages can be used to complement each other.

X-ray crystallography uses X-ray diffraction patterns to obtain static structures of molecules of interest. It depends on the obtainment of well-ordered macroscopic diffracting crystals for each molecule, which typically requires serialization testing of crystallization conditions. X-ray crystallography can determine complex high-resolution structures ranging from small molecules to molecules as large as the prokaryotic and eukaryotic ribosomes (60, 61), but it proves ineffectual for

dynamic and disordered systems, as it requires periodicity. The recent advent of X-ray free-electron laser sources and serial femtosecond crystallography have been shown to overcome some of the obstacles of X-ray crystallography, such as crystal size and radiation damage, while concurrently possibly introducing time-resolved information (62–65).

Cryo-electron microscopy (cryo-EM) probes molecules of interest through exposure of frozen sample grids to an electron beam at cryogenic temperatures. It requires minimal sample amounts while allowing the study of proteins with different conformations and long flexible loops. Structure determination relies on computer selection and averaging of thousands of particles, leading to a slow throughput. Since the determination of the first near-atomic resolution structures (66), cryo-EM structure determination has mostly relied on large and relatively stable complexes, such as ribosomes. Due to recent developments such as improved detector hardware, and beam-induced motion correction, cryo-EM is evolving at a rapid pace, currently on the verge of reaching atomic resolution (67, 68). As a consequence, the previous macromolecule experimental lower size limit of 300 kDa is closer to the theoretical size limit of 38 kDa (69), as the recently published structure of the cancer target isocitrate dehydrogenase with a size of 93 kDa at 3.8 Å resolution shows (70).

Solution NMR spectroscopy allows the structure determination of macromolecules at atomic resolution and is ideally suited to dynamical characterization such as ligand binding, conformational changes, and macromolecular dynamics. A significant drawback of biomolecular NMR is its size limitation, as the slower molecular tumbling and shorter signal relaxation times of larger molecules lead to peak broadening and loss of sensitivity. Besides, the increased number of nuclei complicates the spectra and subsequent spectral analysis. In the last few decades, several advances in NMR hardware and methodology such as: increased magnetic fields and improved electronics, water handling, perdeuteration, transverse relaxation optimized spectroscopy (TROSY) based experiments, heteronuclear isotopic labeling and multidimensional experiments (71–74), have increased the current size limit for *de novo* structure determination to about 70 kDa. Furthermore, sophisticated labeling schemes such as methyl group labeling increased the overall limits of molecule size which can be quantitatively and qualitatively studied by NMR spectroscopy to 1 MDa (75, 76).

Magic-angle spinning (MAS) solid-state NMR spectroscopy allows the study of macromolecules in the solid-state, where molecules typically lack motion and spin anisotropic interactions lead to peak broadening. This can be overcome by both fast sample spinning (77), as well as specific radiofrequency (rf) pulses that attenuate undesired couplings, enabling identical approaches used in solution-state. Theoretically, molecular size is not a limitation for solid-state NMR (78). This allows both the structural and dynamical characterization of huge macromolecule assemblies such as fibrils, an ideal application

for solid-state NMR. Currently, solid-state NMR spectroscopy applications are limited due to sample preparation, though advancements such as faster rotor spinning which allow proton detection, are advancing at a fast pace (79–82).

The different advantages and drawbacks of the previously discussed techniques have led to the procurement of integrated structural approaches. The NMR spectroscopy overall size limitation can be overcome with solid-state NMR, cryo-EM and X-ray-derived structures. The dynamic rich information and atomic resolution of NMR solved structures of individual monomers can be complemented with the higher-sized complex structures achievable with both techniques. For instance, the recent structure and assembly determination of the apoptosis-associated speck-like protein containing a CARD (ASC)-filament, was solved with an iterative algorithm using an ensemble of cryo-EM, solution, and solid-state NMR data in an integrated fashion (83).

1.2.1. NMR spectroscopy principles

NMR uses the spin angular momentum and magnetic moment of nuclei to study its local chemical environment. A non-zero nuclear spin (S) is associated with a magnetic moment (μ) *via* the relation

$$\mu = \gamma \, . \, S$$

where γ is the nuclear gyromagnetic ratio (84–86). In the absence of an external magnetic field, the nuclei are oriented randomly. Under an external magnetic field (B_0) the bulk magnetization precesses around the magnetic field (Figure 1-3) with a frequency

$$\omega_0 = \gamma \, . \, B_0$$

known as the Larmor frequency (84, 87).

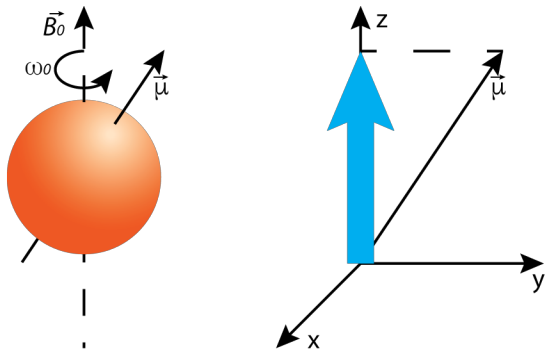


Figure 1-3 – Spin precession. On the left is depicted the spin precession under an external field B_0. On the right is shown the z-axis projection of the magnetic moment, the bulk magnetization.

Since γ is nucleus-type specific, each nucleus has a specific Larmor frequency. The Larmor frequency is independent of the angle between the magnetic field and the magnetic moment direction. A

transversal radiofrequency pulse applied at the Larmor frequency forces the tipping of the bulk magnetization towards the transversal plane.

Chemical shift

Nuclei with a non-zero spin, under a magnetic field, are split into different equally spaced energy levels proportional to B_0, in a phenomenon known as the Zeeman splitting (85, 88). Each nucleus interacts with the surrounding nuclei and electrons, further shifting and splitting these energy levels (89). This specific chemical environment within the bulk magnetization slightly changes its precessing frequency. Hence, the local electron chemical environment perturbation of each nucleus-specific Larmor frequency ω_0 is termed chemical shift (90–92).

1D NMR

An NMR experiment relies on the manipulation of a group of spins through rf-pulses. A high bandwidth pulse produces a broad excitation of different nuclei frequencies. Also, when applied with a precise duration the precession of the excited nuclei around the axis of the pulse corresponds to a well-defined angle. Each experiment applies different pulses in sequence, separated by intervals, and is usually composed of at least two time periods: relaxation and acquisition as shown in Figure 1-4.

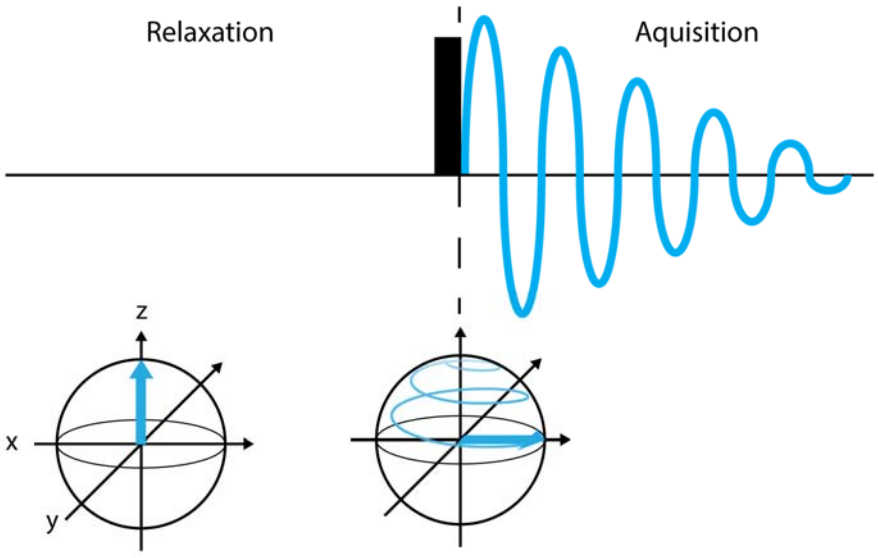


Figure 1-4 – Pulse-sequence with a 90° y pulse followed by an evolution delay t_1, and an acquisition delay t_2. Below is shown the pulse sequence effects on the bulk magnetization excited spins (shown in blue). B_0 aligns with the vertical z axis, x, and y transversal to it on the horizontal axis.

The relaxation delay allows the excited spins to relax back to equilibrium between consecutive pulse sequences. After recovery, one or more pulses excite a group of spins tipping the bulk magnetization to a particular angle. After all pulses and delays take place, the sum of all spins free-induction decay (FID) is recorded during acquisition (100, 101). So, the deconvolution of the recorded data by a Fourier

transformation yields a frequency vs. intensity 1D plot as exemplified in Figure 1-5 (102). Also, to increase the signal to noise ratio, each experiment is repeated multiple times in sequence.

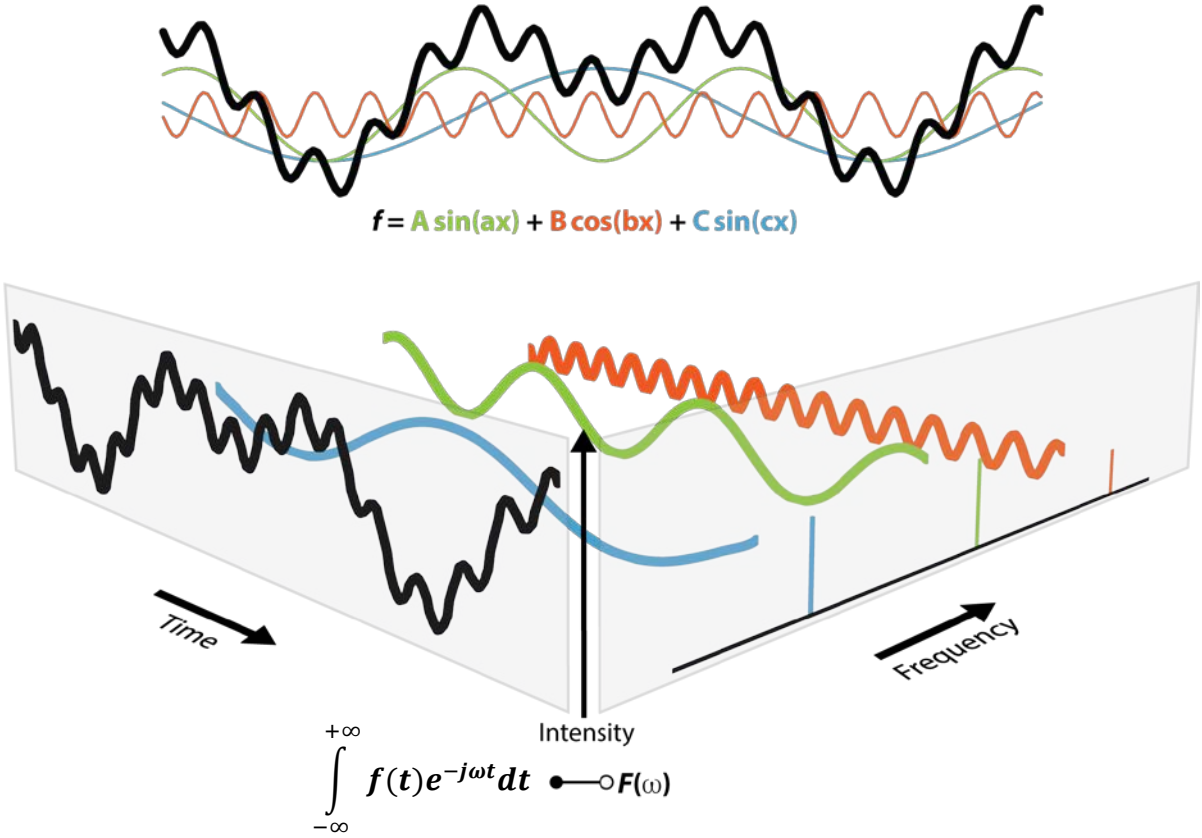


Figure 1-5 – A periodic function *f* (black), is composed of individual sinusoidal components (blue, green and red) with different amplitudes (A, B, C) and frequencies (a,b,c). A Fourier transform deconvolutes the time-domain signal sum, such as an FID, into the frequency domain, the specific frequencies of each sinusoid that constitute it.

2D NMR

Signal overlap due to the high number of homonuclear spins present in macromolecules results in complicated, frequency-overcrowded spectra. Therefore, the complex analysis of 1D spectra led to the extension of NMR experiments from one to two dimensions (103). In two-dimensional NMR the signal is recorded as a function of two time variables, t_1 and t_2, where the evolution time t_1 is stepwise increased, and for each different t_1 an FID is recorded. As a result, two Fourier transformations take place one as a function of t_1 and the second as a function of t_2, yielding a spectrum function of two frequency variables F1 and F2 as depicted in Figure 1-6 (104).

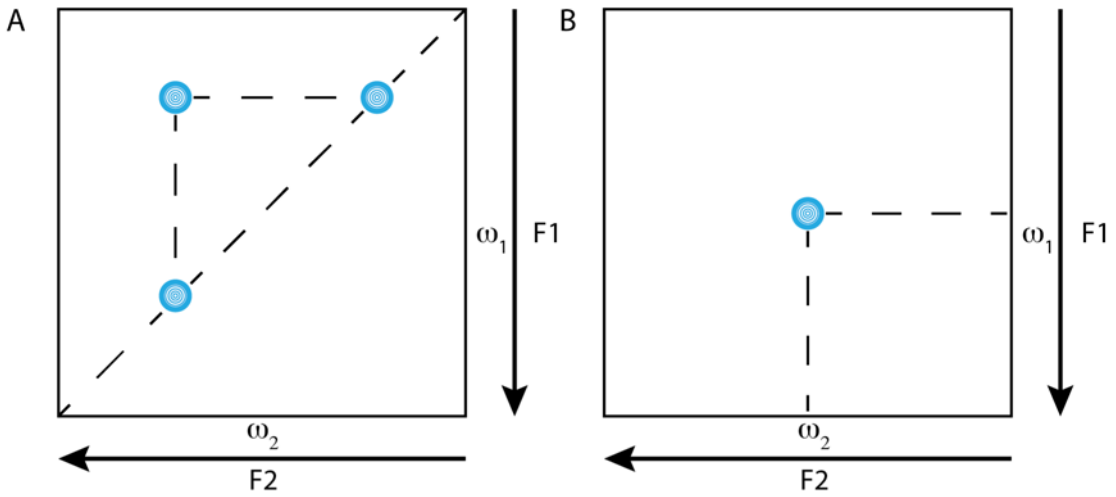


Figure 1-6 – 2D-NMR spectrum following the second Fourier transform. In A is shown a 2D- homonuclear type spectrum, where F1 and F2 frequencies correspond to the same type of nucleus. In B is shown a 2D-heteronuclear type spectrum, where F1 and F2 correspond to the frequencies of distinct nuclei types.

The introduction of a second dimension allows to better resolve overlapping signals in 1D NMR-spectra, it also helps obtain new chemical shift information. Correlation experiments can differ on the type of magnetization transfer. Homonuclear correlation spectroscopy, such as correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) identify same isotope correlated spins, e.g., the H^N-H^α pair (105, 106). In two-dimensional homonuclear correlation spectra, both F1 and F2 frequencies result from the same isotope usually ^1H. Diagonal peaks correspond to the 1D isotope spectrum; off-diagonal peaks have different frequencies in each axis and correspond to coupled resonances (Figure 1-6-A). In a TOCSY experiment, between the evolution period and acquisition, a composite spin-lock pulse is introduced during the mixing time (Figure 1-7). Spin-locking achieves an isotropic mixing of protons, resulting in coherence transfer between coupled spins belonging to the same spin-system. The extension of the mixing time allows the magnetization to spread through farther bonds.

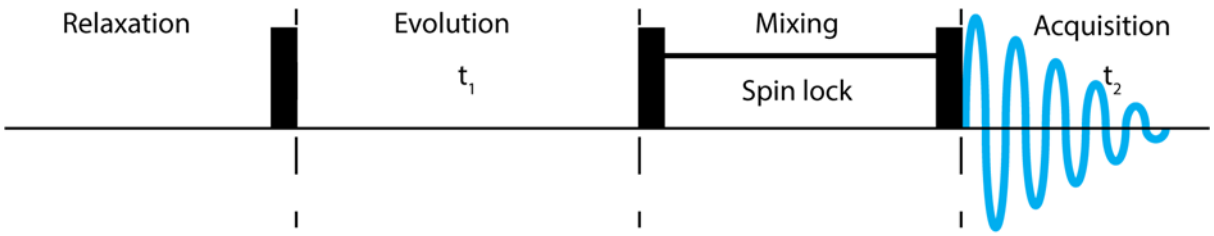


Figure 1-7 – Scheme for a 2D NMR [^{1}H,^{1}H]-TOCSY experiment pulse sequence.

In heteronuclear NMR a two-dimensional spectrum is recorded where the F2 dimension coordinates correspond to the chemical shift of one type of nucleus (e.g., ^{15}N), and the F1 dimension coordinates to the chemical shift of another nucleus (e.g., ^1H) coupled to the first (Figure 1-6-B) (107). In protein NMR, the 2D-[^{15}N,^{1}H]-HSQC experiment correlates in the F2 dimension the backbone –NH ^{15}N, and in

the F1 dimension the -NH ^{1}H (108). The 2D plot shows resonance peaks for each –NH and its distinctness is often termed as the protein's fingerprint spectra. Peak chemical shift dispersion is a measure of the protein folding state, as folded proteins residues have a local specific chemical environment distinct from its respective random coil shifts.

Multidimensional NMR, assignment and structural determination

Specific resonance assignment marks the initial step towards structure determination, conformational and dynamical characterization. For small molecules, sequential assignment is obtained through 2D-homonuclear experiments (109) such as through-bond [^{1}H,^{1}H]–COSY (110), [^{1}H,^{1}H]–TOCSY (106, 111) and through-space [^{1}H,^{1}H]–NOESY (104, 112) spectra. Spectral analysis for larger molecules is complicated due to peak crowding. To solve this, multi-dimensional NMR experiments, which can correlate three types of nuclei, ^{1}H, ^{15}N, and ^{13}C have been developed. As depicted in Figure 1-8, the spread of resonances through a new dimension effectively reduces overlap, allowing sequential assignment and extraction of additional chemical shift information from these nuclei. In protein NMR, these are used for the specific backbone and side-chain resonance assignment (113). Moreover, different residue types have distinct chemical shifts, which can be used in their identification and conformation characterization.

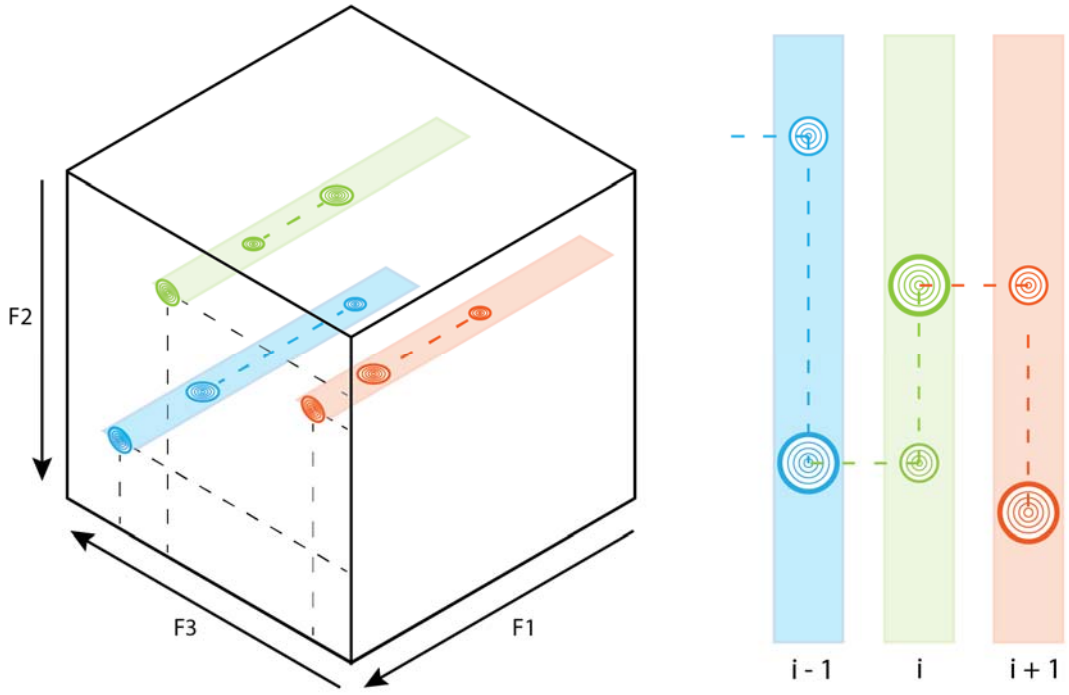


Figure 1-8 – Heteronuclear 3D-NMR spectrum, where F1, F2, and F3 correspond to the frequencies of three distinct nuclei types. On the left is shown a 3D spectrum, with the crosspeaks of the 2D-F3, F2 projection of the "root" spectrum, with each F1-dimension corresponding strip highlighted in corresponding colors. On the right is shown the sequential assignment of residues i-1, i, and i+1, from connected cross-peaks of each F1-dimension strip.

After complete assignment, backbone and side-chain chemical shift data can be used to determine protein secondary structure (114, 115). From scalar coupling data, protein dihedral angles can be derived (116). Other phenomena such as the NOE, and residual dipolar couplings can be used as conformational restraints for structural characterization; and relaxation data can be used to derive protein dynamics information (109, 117, 118). Finally, in the last step towards structure determination, an ensemble of structures is calculated and refined in agreement with the restraint data, which leads to the characterization of the structure and conformation of a specific macromolecule or complex of macromolecules.

1.2.2. X-ray crystallography principles

Structural X-ray crystallography uses X-ray scattering to study molecules of interest. Contrary to visible light (λ=400–700 nm), X-ray radiation (λ≈1 Å) is able to diffract with the electron cloud of molecules. Macroscopic crystals are composed of highly similar structural motifs that form each unit cell, repeated periodically throughout the whole crystal. As each molecule diffracts X-rays identically, a massive enhancement of identical constructive scattering events produces a measurable diffraction pattern. From it, it is possible to obtain an image of the electron clouds that surround the molecules in the crystal, and consequently a model of the molecule's structure (119).

Protein crystallization

The study of protein structure using X-rays requires high-quality protein crystals. Proteins crystals are grown through slow, controlled precipitation from aqueous solution under conditions that do not denature the protein, while increasing the protein and precipitant concentrations through evaporation, commonly by vapor diffusion. Under certain conditions, molecules solidify to form a crystal, with individual proteins adopting one or a few identical orientations resulting in a three-dimensional molecule matrix-bound through non-covalent interactions, hydrogen bonds and salt-bridges (120-121).

After protein purification, crystallization serial trials are set up with common crystallization buffers in search for crystal formation. After a successful condition is found, this is then finely optimized for crystal quality and size, suitable for diffraction measurement. Seeding with crushed pre-formed crystals can also be attempted for further optimization. As proteins often interact with partners or ligands, in order to understand these interactions and their effect on structure, proteins can be co-crystallized with its partner, or even soaked in the crystalline form with it, if small enough to diffuse into the binding locus in the protein (121).

X-ray diffraction

Molecule crystals are a three-dimensional repetition of an individual geometrical defined unit cell, with specific dimensions and angles that define its shape, edges, and vertices that form the crystal lattice (Figure 1-9). Within a unit cell, the smallest unit that can reproduce the whole unit cell through symmetry operations is termed the asymmetric unit, shown in Figure 1-9; and the symmetry operations that characterize a crystal lattice, define its space group (119, 122). In protein crystallography, the asymmetric unit is usually a protein, domain or protein complex.

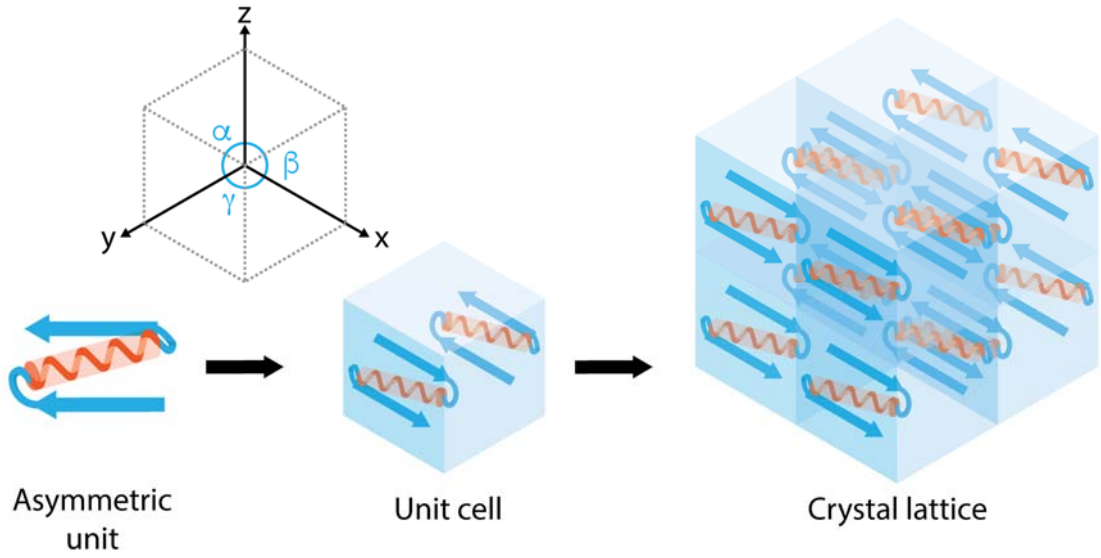


Figure 1-9 – Architecture of crystal lattice.

In the unit cell real space, three-dimensional Cartesian coordinates (x,y,z) define the position of each molecule's atom, with one of the vertices as its origin (x,y,z=[0,0,0]) (Figure 1-9). Each unit cell is composed by an infinite number of atom planes, which diffract an incoming X-ray at a particular angle (120). Thus, the scattered X-ray beam produces an observable pattern of reflections termed diffraction pattern, as depicted in Figure 1-10.

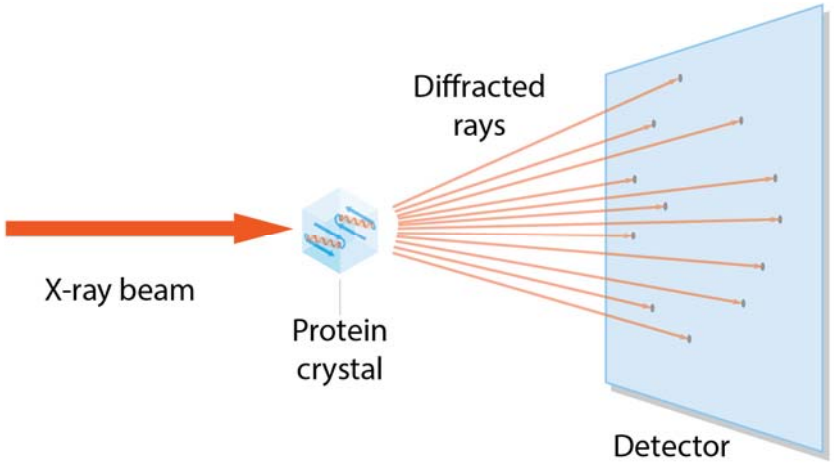


Figure 1-10 – X-ray beam diffraction by a protein crystal, and respective diffraction pattern.

From each reflection both the position, which correlates with the diffraction angle and the intensity of the diffracted spots can be measured. The distances between reflection spots outline the reciprocal lattice, which inversely correlates with the crystal (real) lattice. In reciprocal space, each reflection defined by a set of coordinates (h,k,l), correspond to a family of diffracting atom planes in real space. Counting from the center of reflections h,k,l=[0,0,0], an integer coordinate is given to each reflection during indexing. Moreover, the intensity of each reflection relates to the positions of electrons in real space (122).

As a crystal is a three-dimensional lattice, its rotation (from the X-ray beam) exposes a different section of atom planes capable of diffraction and therefore a different diffraction pattern. Therefore, through crystal rotation, it is possible to obtain the full three-dimensional diffraction pattern that describes the three-dimensional unit cell. In other words, the crystal structure is encoded in the diffracted X-rays, the shape and symmetry of the unit cell define the directions of the diffracted beams, and the locations of all atoms in the unit cell define their intensities.

From diffraction to structure

Each diffracted X-ray that produces a reflection is the sum of the contributions of all scatterers in the unit cell, described as a Fourier sum, the structure factor equation F_{hkl} (122). Each scattered X-ray wave is characterized by an amplitude, frequency, and phase, however, unlike the amplitude and frequency, it is not possible to infer the hitting wave's phase from its diffraction pattern reflection.

In order to characterize the electron distribution in the asymmetric unit of the unit cell, it is necessary to obtain the phases, in what is known as the phase problem. Different methods exist to solve this, being the most popular heavy metal insertion, molecular replacement, in the case where a homologous structure is known and increasingly anomalous diffraction (123).

After solving the phase problem, through the Fourier transform of the structure factors sum, it is possible to obtain an initial approximate electron-density distribution. The known residue sequence enables model building in a chemical and conformational appropriate state that fits the electron density map. This can be improved in an iterative fashion between model building and refinement until an accurate structural model of the target protein is obtained, as depicted in Figure 1-11 (122).

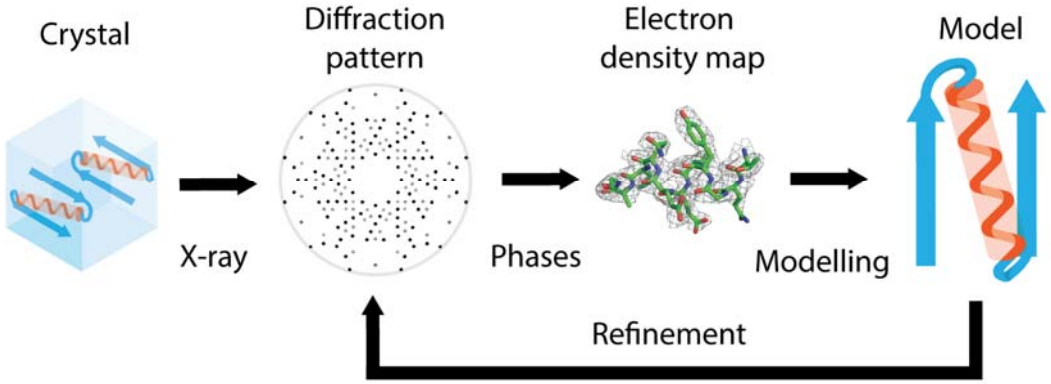


Figure 1-11 – X-ray protein structure determination workflow.

1.3. Thesis outline

The complexity of protein toxin targets often requires an integrated structural approach. Different biophysical methods, such as solution and solid-state NMR, X-ray crystallography, cryo-EM, and AFM can be combined in order to unravel the molecular details of this complex group of proteins. In the following chapters different practical examples are described, where the chosen approach was not the choice of a single method but an integration between different methods.

In the second chapter, the set up, optimization and expression test of a cell-free protein expression system required for expression of cytotoxic targets such as toxins are described, detailed in further chapters. In the third chapter, two small molecule toxins and the interaction with their cognate target are described. In the fourth chapter, two examples of pore-forming toxins are described, and the advances towards their membrane-inserted characterization, using solution NMR spectroscopy, solid-state NMR spectroscopy, and X-ray crystallography are described. Finally, in the fifth chapter, two examples of PTM toxins are described, their structural and functional characterization, using both solution NMR spectroscopy and X-ray crystallography, among other biophysical methods.

1.4. References

1. J. J. P. Genevieve S. Bondy, James J. Pestka, Immunomodulation By fungal toxins. *J. Toxicol. Environ. Heal. Part B*. **3**, 109–143 (2000).
2. D. Ladant, J. E. Alouf, M. R. Popoff, *The comprehensive sourcebook of bacterial protein toxins* (Elsevier, ed. 3, 2006).
3. M. M. Lubran, Bacterial toxins. *Ann. Clin. Lab. Sci.* **18**, 58–71 (1988).
4. T.-T. Tseng, B. M. Tyler, J. C. Setubal, Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol.* **9**, S2 (2009).
5. K. Sandvig, M. L. Torgersen, N. Engedal, T. Skotland, T. G. Iversen, Protein toxins from plants and bacteria: Probes for intracellular transport and tools in medicine. *FEBS Lett.* **584**, 2626–2634 (2010).
6. W. R. Gray, B. M. Olivera, L. J. Cruz, Peptide toxins from venomous conus snails. *Annu. Rev. Biochem.* **57**, 665–700 (1988).
7. H. Terlau, Conus Venoms: A rich source of novel ion channel-targeted peptides. *Physiol. Rev.* **84**, 41–68 (2004).
8. K. Nakashima *et al.*, Accelerated evolution in the protein-coding regions is universal in crotalinae snake venom gland phospholipase A2 isozyme genes. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5605–5609 (1995).
9. L. H. Liow, L. Van Valen, N. C. Stenseth, Red Queen: From populations to taxa and communities. *Trends Ecol. Evol.* **26**, 349–358 (2011).
10. H. M. Liyanage, D. N. M. Arachchi, T. Abeysekara, L. Guneratne, Toxicology of freshwater cyanobacteria. *J. Environ. Sci. Heal. Part C*. **501**, 137–168 (2016).
11. J. L. Richard, G. A. Bray, D. H. Ryan, Mycotoxins as immunomodulators in animal systems. *Mycotoxins, Cancer, Heal.*, 197–220 (1991).
12. T. Kuiper-Goodman, P. M. Scott, Risk assessment of the mycotoxin ochratoxin A. *Biomed. Environ. Sci.* **2**, 179–248 (1989).
13. G. Corzo, P. Escoubas, Pharmacologically active spider peptide toxins. *Cell. Mol. Life Sci.* **60**, 2409–2426 (2003).
14. T. Honma, K. Shiomi, Peptide toxins in sea anemones: Structural and functional aspects. *Mar. Biotechnol.* **8**, 1–10 (2006).
15. D. J. Craik, Plant cyclotides: Circular, knotted peptide toxins. *Toxicon*. **39**, 1809–1813 (2001).
16. S. H. Chiou *et al.*, Inhibition of protein kinase C by snake venom toxins: comparison of enzyme inhibition, lethality and hemolysis among different cardiotoxin isoforms. *Biochem. Mol. Biol. Int.* **35**, 1103–1112 (1995).
17. R. . Dawson, The toxicology of microcystins. *Toxicon*. **36**, 953–962 (1998).
18. T. Matsui, Y. Fujimura, K. Titani, Snake venom proteases affecting hemostasis and thrombosis. *Biochim. Biophys. Acta* **1477**, 146–156 (2000).
19. R. M. Kini, Serine proteases affecting blood coagulation and fibrinolysis from snake venoms. *Pathophysiol. Haemost. Thromb.* **34**, 200–204 (2005).
20. T. Morita, Structures and functions of snake venom CLPs (C-type lectin-like proteins) with anticoagulant-, procoagulant-, and platelet-modulating activities. *Toxicon*. **45**, 1099–1114 (2005).
21. A. V. Osipov *et al.*, Naturally occurring disulfide-bound dimers of three-fingered toxins: A paradigm for biological activity diversification. *J. Biol. Chem.* **283**, 14571–14580 (2008).
22. O. H. Del Brutto, V. J. Del Brutto, Neurological complications of venomous snake bites: A review. *Acta Neurol. Scand.* **125**, 363–372 (2012).
23. M. G. Macfarlane, B. C. J. G. Knight, The biochemistry of bacterial toxins. *Biochem. J.* **42**, 590–595 (1948).
24. J. Sakurai, M. Nagahama, M. Oda, Clostridium perfringens α-toxin: Characterization and mode of action. *J. Biochem.* **136**, 569–574 (2004).

25. M. W. Parker, S. C. Feil, Pore-forming protein toxins: From structure to function. *Prog. Biophys. Mol. Biol.* **88**, 91–142 (2005).
26. G. Anderluh, J. H. Lakey, Disparate proteins use similar architectures to damage membranes. *Trends Biochem. Sci.* **33**, 482–490 (2008).
27. J. Rossjohn, S. C. Feil, W. J. McKinstry, R. K. Tweten, M. W. Parker, Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form. *Cell*. **89**, 685–692 (1997).
28. M. Fivaz, L. Abrami, Y. Tsitrin, F. G. van der Goot, Aerolysin from *Aeromonas hydrophila* and related toxins. *Curr. Top. Microbiol. Immunol.* **257**, 35–52 (2001).
29. R. K. Tweten, M. W. Parker, A. E. Johnson, The cholesterol-dependent cytolysins. *Curr. Top. Microbiol. Immunol.* **257**, 15–33 (2001).
30. P. Boonserm, P. Davis, D. J. Ellar, J. Li, Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications. *J. Mol. Biol.* **348**, 363–382 (2005).
31. N. Galitsky *et al.*, Structure of the insecticidal bacterial δ-endotoxin Cry3Bb1 of *Bacillus thuringiensis*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 1101–1109 (2001).
32. C. E. . E. D. Bell, Crystal structure of nucleotide-free diphtheria toxin. *Biochemistry*. **36**, 481–488 (1997).
33. K. J. Oh *et al.*, Organization of diphtheria toxin T domain in bilayers: A site-directed spin labeling study. *Science* **273**, 810–812 (1996).
34. M. E. Pipkin, J. Lieberman, Delivering the kiss of death: progress on understanding how perforin works. *Curr. Opin. Immunol.* **19**, 301–308 (2007).
35. P. E. Czabotar, G. Lessene, A. Strasser, J. M. Adams, Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* **15**, 49–63 (2014).
36. D. Butzke, A. Luch, High-molecular weight protein toxins of marine invertebrates and their elaborate modes of action. *Molecular, Clinical and Environmental Toxicology* **100**, 213–232 (2010).
37. K. Aktories, Bacterial protein toxins that modify host regulatory GTPases. *Nat. Rev. Microbiol.* **9**, 487–498 (2011).
38. A. Harms, F. V. Stanger, C. Dehio, Biological diversity and molecular plasticity of FIC domain proteins, *Annu Rev Microbiol.* **70**, 341–360 (2016).
39. C. E. Bell, D. Eisenberg, Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry*. **35**, 1137–1149 (1996).
40. K. J. F. Satchell, Actin Crosslinking Toxins of Gram-Negative Bacteria. *Toxins.* **1**, 123–133 (2009).
41. J. A. Guttman, B. B. Finlay, Tight junctions as targets of infectious agents. *Biochim. Biophys. Acta.* **1788**, 832–841 (2009).
42. A. S. Selyunin *et al.*, The assembly of a GTPase-kinase signalling complex by a bacterial catalytic scaffold. *Nature*. **469**, 107–111 (2011).
43. K. Strebhardt, A. Ullrich, Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat. Rev. Cancer*. **8**, 473–480 (2008).
44. B. G. Fry *et al.*, The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. G.* **10**, 483–511 (2009).
45. P. Escoubas, F. Bosmans, Spider peptide toxins as leads for drug development. *Expert Opin. Drug Discov.* **2**, 823–835 (2007).
46. Y. N. Utkin, Animal venom studies: Current benefits and future developments. *World J. Biol. Chem.* **6**, 28–33 (2015).
47. J. Chaisakul, W. C. Hodgson, S. Kuruppu, N. Prasongsook, Effects of animal venoms and toxins on hallmarks of cancer. *J Cancer*. **7**, 1571–1578 (2016).
48. L. González-Mariscal *et al.*, Strategies that target tight junctions for enhanced drug delivery. *Curr. Pharm. Des.* **22**, 1–1 (2016).
49. R. Mir, S. Karim, M. Kamal, C. Wilson, Z. Mirza, Conotoxins: Structure, therapeutic potential and pharmacological applications. *Curr. Pharm. Des.* **22**, 582–589 (2016).
50. W. Bergmann, R. J. Feeney, Contributions to the study of marine products. XXXII. The

nucleosides of sponges. *J. Org. Chem.* **16**, 981–987 (1951).
51. D. M. Cornforth, K. R. Foster, Antibiotics and the art of bacterial war. *Proc. Natl. Acad. Sci.* **112**, 10827–10828 (2015).
52. A. Pickett, K. Perrow, Towards new uses of botulinum toxin as a novel therapeutic tool. *Toxins.* **1,** 63–81 (2011).
53. S. Potala, S. K. Sahoo, R. S. Verma, Targeted therapy of cancer using diphtheria toxin-derived immunotoxins. *Drug Discov. Today*. **13**, 807–815 (2008).
54. Z. Takacs, N. York, Animal venoms in medicine. *Encycl. Toxicol.* **1**, 252–259 (2014).
55. C. Kost, W. A. Herzer, P. J. Li, E. K. Jackson, Pertussis toxin-sensitive G-proteins and regulation of blood pressure in the spontaneously hypertensive rat. *Clin. Exp. Pharmacol. Physiol.* **26**, 449–455 (1999).
56. Z. Tang *et al.*, Pertussis toxin attenuates experimental autoimmune encephalomyelitis by upregulating neuronal vascular endothelial growth factor. *Neuroreport*. **24**, 469–475 (2013).
57. C. Rizzi *et al.*, Pertussis toxin B-oligomer suppresses IL-6 induced HIV-1 and chemokine expression in chronically infected U1 cells *via* inhibition of activator protein 1. *J Immunol*. **176**, 999–1006 (2006).
58. I. Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature*. **431**, 931–945 (2004).
59. A. C. Anderson, The process of structure-sased drug design. *Chem. Biol.* **10**, 787–797 (2003).
60. A. Korostelev, S. Trakhanov, M. Laurberg, H. F. Noller, Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*. **126**, 1065–1077 (2006).
61. A. Ben-Shem *et al.*, The structure of the eukaryotic ribosome at 3.0 A resolution. *Science.* **334**, 1524–1529 (2011).
62. R. Neutze, J. Hajdu, Femtosecond time resolution in X-ray diffraction experiments. *Proc. Natl. Acad. Sci.* **94**, 5651–5655 (1997).
63. E. E. Lattman, Molecular structures from femtosecond X-ray pulses. *Proc. Natl. Acad. Sci.* **98**, 6535–6536 (2001).
64. H. N. Chapman *et al.*, Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nat. Phys.* **2**, 839–843 (2006).
65. H. N. Chapman *et al.*, Femtosecond X-ray protein nanocrystallography. *Nature*. **470**, 73–77 (2011).
66. R. Henderson *et al.*, Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 (1990).
67. A.-C. Milazzo *et al.*, Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *J. Struct. Biol.* **176**, 404–408 (2011).
68. X. Li *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods*. **10**, 584–590 (2013).
69. R. Henderson, The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (2009).
70. A. Merk *et al.*, Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* (2016).
71. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci.* **95**, 13585–13590 (1998).
72. S. Hiller, G. Wider, T. Etezady-Esfarjani, R. Horst, K. Wüthrich, Managing the solvent water polarization to obtain improved NMR spectra of large molecular structures. *J. Biomol. NMR*. **32**, 61–70 (2005).
73. S. W. Fesik, E. R. P. Zuiderweg, Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Q. Rev. Biophys.* **23**, 97–131 (1990).
74. D. M. LeMaster, F. M. Richards, NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. *Biochemistry*. **27**, 142–150 (1988).
75. J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, NMR analysis of a 900K GroEL GroES

complex. *Nature*. **418**, 207–211 (2002).
76. R. Sprangers, L. E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*. **445**, 618–622 (2007).
77. E. R. Andrew, A. Bradbury, R. G. Eades, Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature*. **182**, 1659–1659 (1958).
78. Vaughan; W. Robert, R. W. Vaughan, R. W. Vaughan, High-resolution, solid state NMR. *Annu. Rev. Phys. Chem.* **29**, 397–419 (1978).
79. M. J. Knight *et al.*, Fast resonance assignment and fold determination of human superoxide dismutase by high-resolution proton-detected solid-state MAS NMR spectroscopy. *Angew. Chemie Int. Ed.* **50**, 11697–11701 (2011).
80. D. H. Zhou *et al.*, Solid-state NMR analysis of membrane proteins and protein aggregates by proton detected spectroscopy. *J. Biomol. NMR*. **54**, 291–305 (2012).
81. R. Zhang, K. H. Mroue, A. Ramamoorthy, Proton chemical shift tensors determined by 3D ultrafast MAS double-quantum NMR spectroscopy. *J. Chem. Phys.* **143**, 1442011–1442016 (2015).
82. E. Barbet-Massin *et al.*, Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* **136**, 12489–12497 (2014).
83. L. Sborgi *et al.*, Structure and assembly of the mouse ASC inflammasome by combined NMR spectroscopy and cryo-electron microscopy. *Proc. Natl. Acad. Sci.* **112**, 13237–13242 (2015).
84. J. Larmor, LXIII. On the theory of the magnetic influence on spectra; and on the radiation from moving ions. *Philos. Mag. Ser. 5*. **44**, 503–512 (1897).
85. W. Pauli, Zur Frage der theoretischen Deutung der Satelliten einiger Spektrallinien und ihrer Beeinflussung durch magnetische Felder. *Naturwissenschaften*. **12**, 741–743 (1924).
86. P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, 1930).
87. J. Larmor, *Aether and Matter* (Cambridge University Press, 1900).
88. P. Zeeman, VII. Doublets and triplets in the spectrum produced by external magnetic forces. *Philos. Mag. Ser. 5*. **44**, 55–60 (1897).
89. C. P. Slichter, *Principles of Magnetic Resonance* (Springer Berlin Heidelberg, ed. 2, 1978), vol. 1 of *Springer Series in Solid-State Sciences*.
90. W. G. Proctor, F. C. Yu, The dependence of a nuclear magnetic resonance frequency upon chemical compound. *Phys. Rev.* **77**, 717–717 (1950).
91. N. F. Ramsey, Magnetic shielding of nuclei in molecules. *Phys. Rev.* **78**, 699–703 (1950).
92. G. Lindström, An experimental investigation of the nuclear magnetic moments of D2 and H1. *Physica*. **17**, 412–419 (1951).
93. H. S. Gutowsky, D. W. McCall, C. P. Slichter, Coupling among nuclear magnetic dipoles in molecules. *Phys. Rev.* **84**, 589–590 (1951).
94. E. L. Hahn, D. E. Maxwell, Chemical shift and field independent frequency modulation of the spin echo envelope. *Phys. Rev.* **84**, 1246–1247 (1951).
95. N. F. Ramsey, E. M. Purcell, Interactions between nuclear spins in molecules. *Phys. Rev.* **85**, 143–144 (1952).
96. G. E. Pake, Nuclear resonance absorption in hydrated crystals: fine structure of the proton line. *J. Chem. Phys.* **16**, 327–336 (1948).
97. H. S. Gutowsky, G. B. Kistiakowsky, G. E. Pake, E. M. Purcell, Structural investigations by means of nuclear magnetism. I. Rigid Crystal Lattices. *J. Chem. Phys.* **17**, 972–981 (1949).
98. E. M. Purcell, H. C. Torrey, R. V. Pound, Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.* **69**, 37–38 (1946).
99. J. H. Van Vleck, The dipolar broadening of magnetic resonance lines in crystals. *Phys. Rev.* **74**, 1168–1183 (1948).
100. F. Bloch, Nuclear induction. *Phys. Rev.* **70**, 460–474 (1946).
101. E. L. Hahn, Nuclear induction due to free larmor precession. *Phys. Rev.* **77**, 297–298 (1950).
102. R. R. Ernst, Application of fourier transform spectroscopy to magnetic resonance. *Rev. Sci.*

Instrum. **37**, 93–102 (1966).
103. R. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of nuclear magnetic resonance in one and two dimensions* (Clarendon Press, 1988).
104. J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553 (1979).
105. W. P. Aue, E. Bartholdi, R. R. Ernst, Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229–2246 (1976).
106. A. Bax, D. G. Davis, MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360 (1985).
107. R. H. Griffey, A. G. Redfield, Proton-detected heteronuclear edited and correlated nuclear magnetic resonance and nuclear Overhauser effect in solution. *Q. Rev. Biophys.* **19**, 51–82 (1987).
108. G. Bodenhausen, D. J. Ruben, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189 (1980).
109. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley-Interscience, ed. 1, 1987).
110. W. P. Aue, Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **64**, 2229–2246 (1976).
111. L. Braunschweiler, R. . Ernst, Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521–528 (1983).
112. A. Kumar, R. R. Ernst, K. Wüthrich, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6 (1980).
113. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton, in *Protein NMR Spectroscopy* (Elsevier, ed. 2, 2007).
114. D. S. Wishart, B. D. Sykes, F. M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*. **31**, 1647–1651 (1992).
115. D. Wishart, B. Sykes, The ^{13}C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*. **4** (1994).
116. J. C. Hoch, C. M. Dobson, M. Karplus, Vicinal coupling constants and protein dynamics. *Biochemistry*. **24**, 3831–3841 (1985).
117. N. Tjandra, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science.* **278**, 1111–1114 (1997).
118. J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci.* **92**, 9279–9283 (1995).
119. L. Ooi, Principles of X-ray Crystallography (Oxford, 2010)
120. J. Drenth, Principles of Protein X-Ray Crystallography (Springer, 2007).
121. A. McPherson, J. A. Gavira, Introduction to protein crystallization. *Acta Crystallogr. Sect. F Structural Biol. Commun.* **70**, 2–20 (2014).
122. D. R. Tobergte, S. Curtis, *Crystallography made crystal clear* (Elsevier, ed. 3, 2013).
123. A. Ilari, C. Savino, Protein structure determination by X-ray crystallography. Methods Mol Biol. 452, 63–87 (2008).
124. A. Sanyal, A. Chen, E. S. Nakayasu, C. S. Lazar, E. Zbornik; C. A. Worby, A. Koller, S. Mattoo, A novel link between fic (filamentation induced by cAMP)-mediated adenylylation/AMPylation and the unfolded protein response. J. Biol. Chem. **290**, 8482–8499 (2015).

Chapter 2

The small molecule exotoxin mycolactone and purported cognate target FKBP12

2. The small molecule exotoxin mycolactone and purported cognate target FKBP12

2.1. Introduction

Buruli ulcer is a chronic necrotizing skin disease caused by *Mycobacterium ulcerans* (1, 2). In humans, topic infection leads to the systemic diffusion of the mycolactone small molecule toxin, causing local analgesia, and impaired inflammation (3). Further, this macrolide has been shown to have pleiotropic cellular effects, such as impaired protein translocation, immunosuppressive, and cell-death (4–7). Macrolides are natural compounds produced by different organisms, which belong to the polyketide class of secondary metabolites mainly produced by cyclization and derivatization of metabolites by mega-synthases. Furthermore, these have a wide range of activities, such as antibiotic, antifungal, immunosuppressant, and cytotoxic effects in animals (8-11). Due to their high biological activity and easy derivatization, these compounds are commonly studied leads in drug discovery (*12–14*). Mycolactone A/B is composed of an invariant 12-membered core macrocyclic lactone ring (in Figure 2-1 is shown the PG-119 derivative, which is composed of only the lactone ring) and two variable polyketide-derived highly unsaturated acyl side chains. As such, it exhibits lipid-like characteristics, such as hydrophobicity. Different bacterial strains express distinct congeners of mycolactone, which differ in one of the unsaturated acyl chains (Figure 2-1) leading to different biological activities (15).

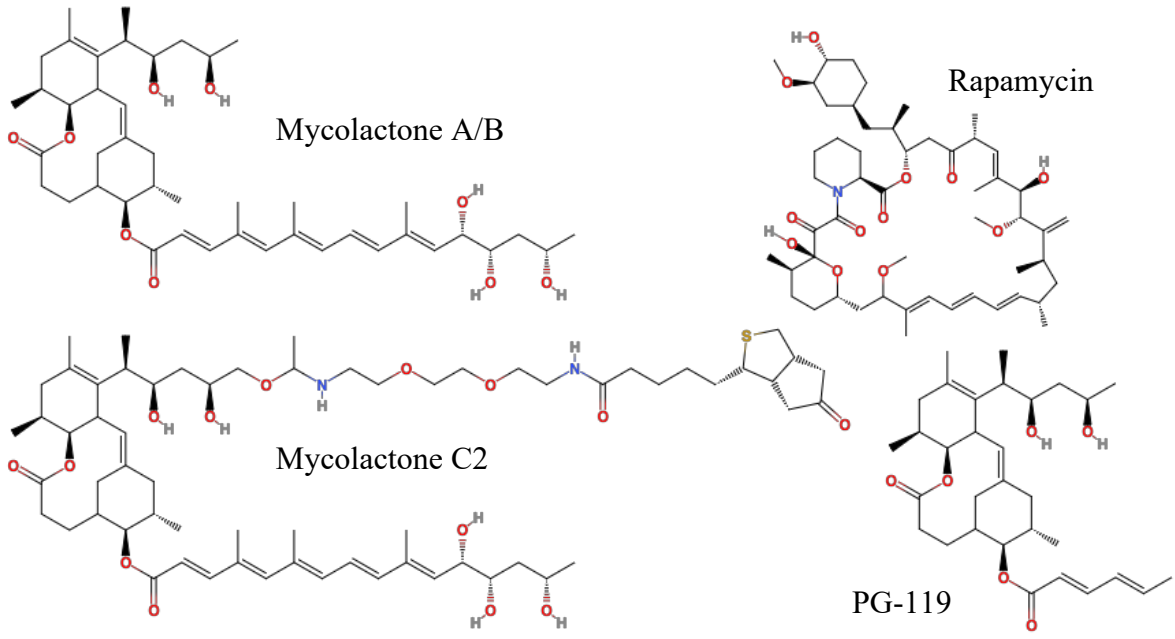


Figure 2-1 – Chemical structures of rapamycin, mycolactone A/B and derivatives, C2 and PG-119.

Previously, it has been suggested that mycolactone shares structural and functional features with the immunosuppressant rapamycin, an antifungal macrolide produced by *Streptomyces hygroscopicus* (Figure 2-1) (7). In mammals, rapamycin acts as an immunosuppressant through inhibition of the mechanistic target of rapamycin complex 1 (mTORC1), a multi-factor sensor protein complex involved in the regulation of cell growth, cell proliferation, cell survival, autophagy, transcription and translation (16). Rapamycin forms a complex with both mTORC1 and cytosolic mTORC1 associated FKBP12 protein (17–19), leading to the direct inhibition of mTORC1 through steric hindrance (20). FKBP12 is a known chaperone foldase of proline-rich proteins, which catalyzes the cis-trans isomerization of peptidyl-prolyl bonds (21). The exact cellular-target and systemic infection molecular mechanisms that underlie the Buruli ulcer disease are currently not completely known. Due to the structural and functional similarities between both macrolides, the potential interaction between different mycolactone congeners with FKBP12, in comparison with the known FKBP12/rapamycin complex, was addressed in this work using solution NMR spectroscopy.

2.2. Material and methods

Protein expression and purification

The FKBP12 construct was kindly provided by Alvar Gossert (Novartis, Basel). A single colony was picked from a fresh transformation in BL21 (λDE3) competent cells and inoculated into 5 ml LB/kanamycin and grown overday at 37°C. At the end of the day, 1 ml was inoculated in 20ml of M9 minimal medium (^{15}N-ammonium chloride) and grown overnight at 37°C. On the next day, the pre-inoculum was transferred to 1 l M9 minimal medium and grown at 37°C. At an OD_{600}=0.8 the culture was induced with 1 mM IPTG. After 4h of expression the cells were harvested (10min, 5,500 rpm, 4°C, SLA3000 rotor), resuspended in 30 ml buffer A (50 mM Tris-HCl, pH=8.0, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole) and frozen at -20°C.

The cells were thawed, DNase I and lysozyme were added, and the solution was stirred for approximately 15 min on ice. Afterward, the cells were lysed with a French press (2runs, 1500 PSI), and the cell debris was cleared by centrifugation (30 min, 13,000 rpm, 4°C, SS34 rotor). The cleared supernatant was applied to a 5 ml pre-equilibrated Ni-NTA gravity column (Genscript, USA) and incubated for 30 min at RT on a shaking device. Afterward, the flow-through was collected, the column was washed with 10 CV of buffer A, 10 CV 5% of buffer B (buffer A with 150 mM NaCl, 300 mM imidazole) and then eluted with 5 CV of 100% buffer B. The elution was left dialyzing in 2 l dialysis

buffer (150 mM NaCl, 300 mM imidazole, 1 mM TCEP and 0.5 mM EDTA) with a 3,000 MWCO (Thermo Scientific, USA) dialysis membrane at RT. After 2 h, 50:1 mg of PreScission protease (protein:protease) was added and left dialyzing overnight in 5 l dialysis buffer.

The next day the dialysate was transferred to a 2 x 2.5 l buffer A dialysis for 1h each and then cleared by centrifugation (10 min, 12,000 rpm). The cleared dialysate was applied to a 5 ml pre-equilibrated Ni-NTA gravity column and incubated for 30 min at RT on a shaking device. Afterward, the cleaved protein was collected both in the flow-through and in the buffer A wash (10 CV). PreScision protease and un-cleaved protein were eluted with 10 CV 100% buffer B. Both cleaved protein containing fractions were pooled, concentrated (Amicon centrifugal filtering device with 5,000 MWCO) and ran on a pre-equilibrated S75 SEC column (GE Healthcare; 25 mM KPO_4, pH=7), using an Äkta system (GE Healthcare, USA). Final yield was 60 mg/l on H_2O-based M9 minimal media, determined by UV/Vis spectroscopy. An SDS-PAGE confirmed protein purity and the protein pooled stock frozen at -20°C.

Complex titration sample preparation

Both rapamycin and mycolactone (A/B, C2, and PG-119 congeners) were kindly provided by Gerd Pluschke's group (Swiss TPH, Basel). Due to the limited solubility of mycolactone in aqueous buffer, all samples were prepared with either 1 or 10% (v/v) final DMSO-d_6. To ensure a constant percentage of DMSO between titration points, each point was prepared separately. A mother solution with FKBP12 in KPi buffer, NaN_3 (0.02% (w/v) final concentration), DSS (1mM final concentration), and D_2O (10% (v/v) final concentration), was prepared and then mixed with different amounts of mycolactone or rapamycin. For each titration point, either mycolactone or rapamycin (1 or 10% (v/v) final DMSO-d_6 concentration) were added. FKBP12 was mixed with each of the macrolides in different ratios up to 1:10 (FKBP12:macrolide) whenever possible. In addition, complex formation was tested in the presence of PPIase activity buffer (50 mM Hepes, pH8, 100 mM NaCl, 5 mM DTT, 0.005% (w/v) NP-40) (21). For this, FKBP12 was buffer exchanged twice (using a 5,000 MWCO Amicon centrifugal filtering device) with PPIase activity buffer, NaN_3 (0.02% (w/v) final concentration), DSS (1 mM final concentration), and D_2O (10% (v/v) final concentration), before adding Mycolatone A/B (1% (v/v) final DMSO-d_6 concentration).

Solution NMR measurement and data analysis

Standard 2D [^{15}N,^{1}H]-TROSY-HSQC (23) NMR spectra were recorded at 25°C on a Bruker Ascend 700 spectrometer equipped with a cryogenic triple-resonance probe. NMR data were processed with TOPSPIN 3.0 (Bruker Biospin) and analyzed with CCPNMR (24).

$$CSP = \sqrt{\left[\left(\Delta\delta(^{1}H)\right)^{2} + \left(0.2 * \Delta\delta(^{15}N)\right)^{2}\right]},$$

where $\Delta\delta(^{1}H)$ and $\Delta\delta(^{15}N)$ are the chemical shift changes of the ^{1}H and ^{15}N nuclei, respectively.

2.3. Results

The previously characterized protein:ligand interaction between rapamycin and the protein FKBP12 part of the mTorC1 was first assessed (25). To establish a titration protocol, the FKBP12 protein was first titrated with increasing amounts of rapamycin (in a buffer containing 10% (v/v) DMSO due to rapamycin's poor solubility) using solution NMR spectroscopy. The titration chemical shift perturbation was analyzed using the known FKBP12 sequence-specific backbone assignment (Figure 2-2 A) (26). A full titration series was performed up to a 1:5 (protein to ligand) molar ratio. The FKBP12 last titration point with rapamycin in 10% (v/v) DMSO-d_6 spectra is shown in Figure 2-2 B.

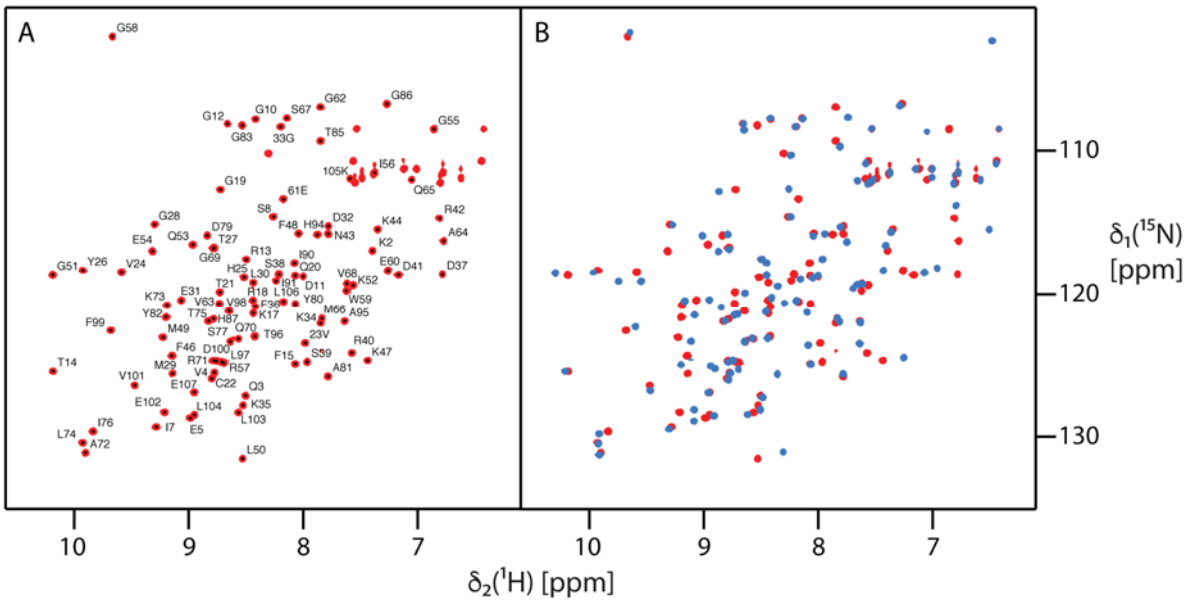


Figure 2-2 – 2D-[^{1}H,^{15}N]-HSQC spectra of apo FKBP12 in comparison with FKBP12/rapamycin complex. In A, is shown the spectra of apo-FKBP12 in red. In B is shown the spectra superposition of FKBP12/rapamycin 1:5 complex in 10% (v/v) DMSO depicted in blue, and the apo-FKBP12 depicted in red.

The titration spectra allowed the identification of substantial chemical shifts corresponding to several backbone resonances. Using the known FKBP12 protein backbone assignment, the measured chemical shift perturbation was plotted *vs*. residue sequence. The analysis of the different chemical shift perturbation showed a localized perturbation around FKPB12 residues 37-64 (Figure 2-3).

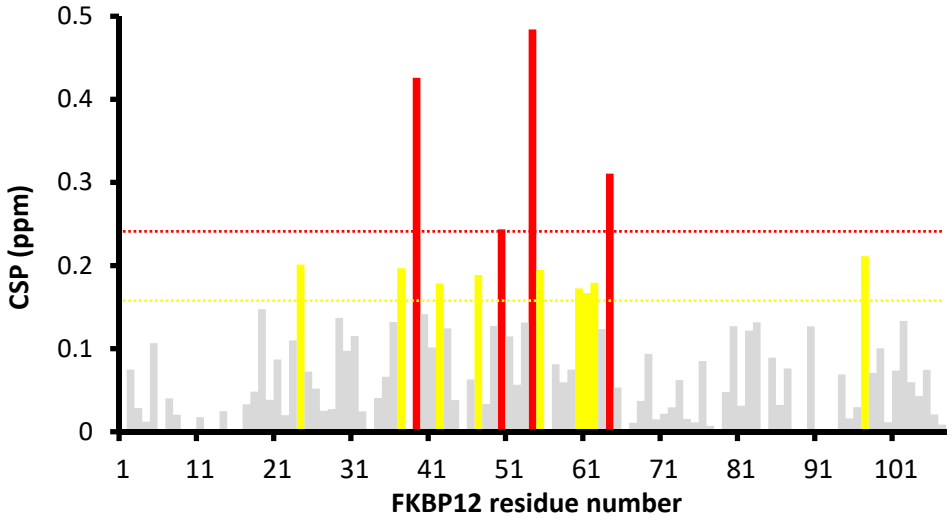


Figure 2-3 – Chemical shift perturbation *vs.* sequence plot of the FKBP12/rapamycin 1:1 complex titration. In yellow are shown the residues with a CSP above the sum of the average and the standard deviation of each residues CSP, and in red the residues above the sum of the average plus twice the standard deviation.

The measured chemical shifts, which showed perturbation above the highlighted thresholds, were mapped onto the known FKBP12/rapamycin complex structure (PDB ID: 1C9H). As shown in Figure 2-4, all significant backbone chemical shifts demonstrate a good correlation between the NMR titration data and the known X-ray crystal structure of the protein:ligand complex (26).

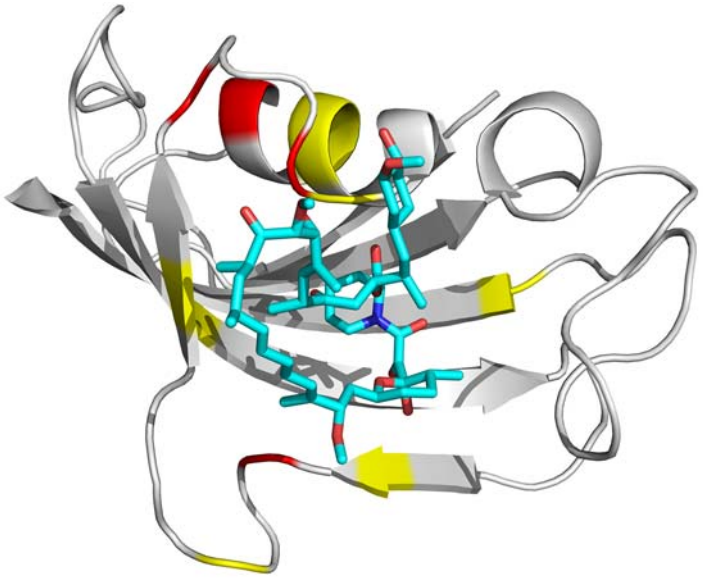


Figure 2-4 – Chemical shift perturbation of the FKBP12/rapamycin 1:1 complex in 10% (v/v) DMSO mapped onto the known complex X-ray structure in cartoon representation, with rapamycin depicted as stick representation (PDB ID: 1C9H).

After demonstrating a suitable titration protocol for the soluble interaction between FKBP12 and an interacting ligand, the titration with the FKBP12 purported ligand mycolactone A/B was tested. Due to the limited solubility of mycolactone molecule, the titration was initially performed in the presence

of 10% (v/v) DMSO. In stark contrast to the above results for rapamycin, mycolactone A/B did not show considerable backbone chemical shift perturbation (Figure 2-5 A). To further rule out DMSO as a solubility factor, the titration was repeated in 1% (v/v) DMSO with similar results (Figure 2-5 B).

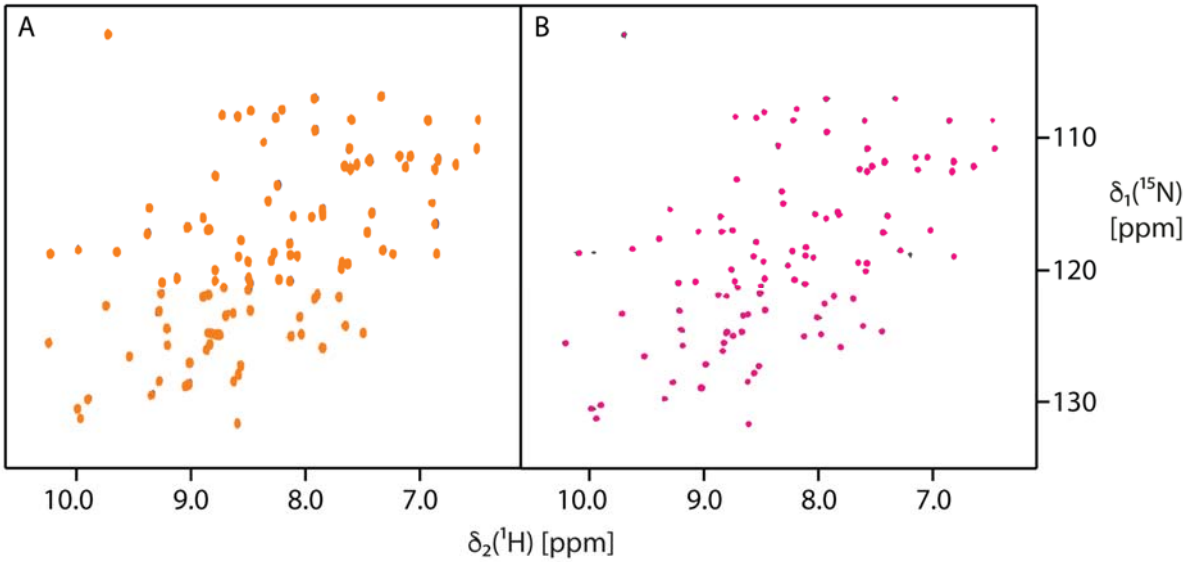


Figure 2-5 – 2D-[^{1}H,^{15}N]-HSQC spectra of the titration end points between FKBP12 and mycolactone A/B. In A, the spectra of FKBP12/mycolactone A/B 1:2 complex in 10% (v/v) DMSO, apo-FKBP12 is shown in dark blue, and the complex in orange. In B, the spectra of FKBP12/mycolactone A/B 1:5 complex in 1% DMSO, apo-FKBP12 is shown in grey, and the complex in pink.

The FKBP12 PPIase catalytic activity was previously studied with *in vitro* functional UV-Vis spectrophotometric essays (21). As such, the FKBP12/mycolactone A/B complex titration was repeated using the FKBP12 PPIase activity functional essay buffer. As before, the FKBP12 2D-[^{1}H,^{15}N]-HSQC spectra showed no significant backbone resonance shifts (Figure 2-6).

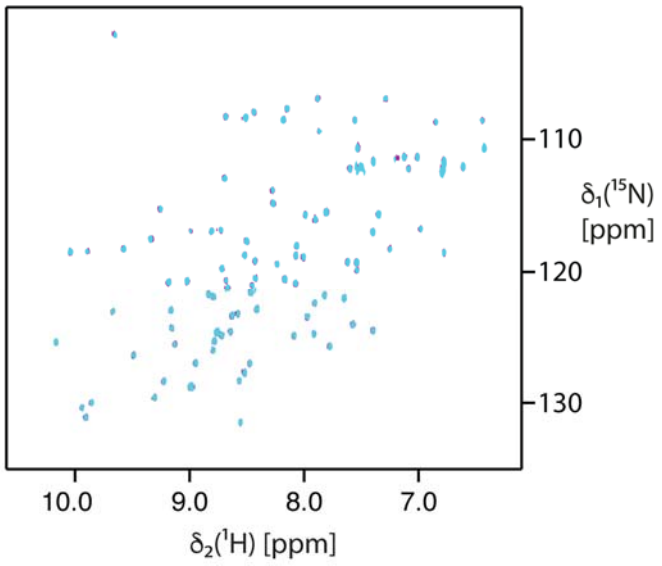


Figure 2-6 – 2D-[^{1}H,^{15}N]-HSQC spectra of the titration end point of FKBP12/mycolactone A/B 1:10 complex in in PPIase activity essay buffer with 1% (v/v) DMSO. Apo-FKBP12 is shown in cyano, and the complex in purple.

Due to mycolactone's lipid-like nature, and consequent limited aqueous solubility, the FKBP12 protein was titrated with different mycolactone congeners with substituted unsaturated acyl chains. Likewise, both the acyl chain double substituted mycolactone C2 (Figure 2-7-A) and the unsaturated acyl chain lacking PG-119, which consists of the macrocyclic lactone ring only (Figure 2-7-B), showed no significant chemical shift perturbation.

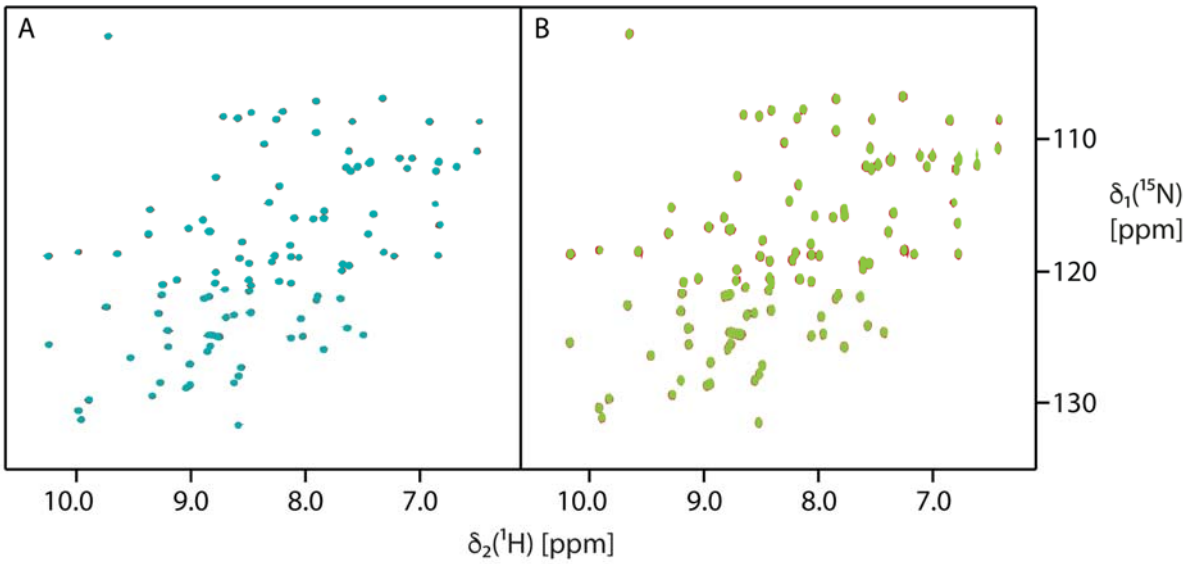


Figure 2-7 – 2D-[^{1}H,^{15}N]-HSQC spectra of the titration end points of FKBP12 with mycolactone C2 or PG-119. In A, the spectra of FKBP12/mycolactone C2 1:5 complex in 10% (v/v) DMSO, apo-FKBP12 is shown in brown, and the complex in teal. In B, the spectra of the FKBP12/PG-119 1:10 complex in 10% DMSO, apo-FKBP12 is shown in dark red, and the complex in light green.

2.4. Discussion

The NMR titration of the FKBP12/rapamycin complex revealed significant backbone resonance shifts, confirming complex formation. As shown in Figure 2-4, mapping of significant backbone resonance shifts onto the X-ray crystal structure of the FKBP12/rapamycin complex shows the chemical shift perturbation data to perfectly match the FKBP12 and rapamycin ligand interaction surface. Due to the insolubility of these macrolides in aqueous buffer, all titrations were performed in the presence of DMSO. Each FKBP12/ligand titration point was prepared separately, in order to maintain a constant DMSO percentage and therefore eliminate the induced DMSO chemical shift change as a factor. The successful complex titration experiment, demonstrated by interaction surface mapping, served both as a positive titration control as well as to eliminate DMSO as a constraint.

In the same conditions, the titration of FKBP12/mycolactone A/B, FKBP12 showed no significant chemical shift perturbations, indicating no complex formation. To further exclude DMSO concentration as a concern, the titration was repeated with 1% instead of 10% (v/v) DMSO. Still, no

significant chemical shift perturbation was observed. To further exclude the tested experimental conditions as a factor, the previously described FKBP12 functional activity assay buffer was tested. The presence of a small concentration of surfactant, NP-40, was thought to enhance mycolactone's solubility by hindering the formation of micelle-like structures in solution, effectively limiting the availability of the compound to interact with FKBP12. Again, mycolactone A/B produced no significant chemical shift perturbations. A different mycolactone congener, mycolactone C2, with an extra upper acyl chain was also tested with similar results. The macrocyclic PG-119 compound, composed of only the lactone ring, was also tested to exclude the possibility both mycolactone A/B and mycolactone C2 congeners might form micelle-like structures due to the presence of the acyl side chains. Likewise, this compound produced no significant chemical shift perturbations. Hence, after testing the FKBP12/mycolactone purported interaction under different buffer conditions, DMSO concentrations, as well as distinct mycolactone congeners, the lack of significant backbone chemical shift perturbation leads to the conclusion that under the described experimental conditions the FKBP12/mycolactone complex is not present.

2.5. References

1. D. P. O'Brien *et al.*, The urgent need for clinical, diagnostic, and operational research for management of Buruli ulcer in Africa. *Lancet Infect. Dis.* **14**, 435–440 (2014).
2. T. Junghanss, R. C. Johnson, G. Pluschke, in *Manson's Tropical Infectious Diseases* (Elsevier, ed. 23, 2014).
3. M. Goto *et al.*, Nerve damage in *Mycobacterium ulcerans*-infected mice. *Am. J. Pathol.* **168**, 805–811 (2006).
4. J. B. Gama *et al.*, Proteomic analysis of the action of the *Mycobacterium ulcerans* toxin mycolactone: Targeting Host Cells Cytoskeleton and Collagen. *PLoS Negl. Trop. Dis.* **8**, 3066 (2014).
5. L. Guenin-Macé *et al.*, Mycolactone activation of Wiskott-Aldrich syndrome proteins underpins Buruli ulcer formation. *J. Clin. Invest.* **123**, 1501–1512 (2013).
6. B. S. Hall *et al.*, The pathogenic mechanism of the *Mycobacterium ulcerans* virulence factor, mycolactone, depends on blockade of protein translocation into the ER. *PLoS Pathog.* **10**, 1004061 (2014).
7. C. Demangel, T. P. Stinear, S. T. Cole, Buruli ulcer: reductive evolution enhances pathogenicity of Mycobacterium ulcerans. *Nat. Rev. Microbiol.* **7**, 50–60 (2009).
8. N. Khan, B. Rawlings, P. Caffrey, A labile point in mutant amphotericin polyketide synthases. *Biotechnol. Lett*. **33**, 1121–1126 (2011).
9. H. G. Floss, T.-W. Yu, Rifamycin-mode of action, resistance, and biosynthesis. *Chem. Rev.* **105**, 621–632 (2005).
10. P. E. Wallemacq, R. Reding, FK506 (tacrolimus), a novel immunosuppressant in organ transplantation: Clinical, biomedical, and analytical aspects. *Clin. Chem.* **39**, 2219–2228 (1993).
11. G. B. Watson, Actions of insecticidal spinosyns on γ-aminobutyric acid responses from small-diameter cockroach neurons. *Pestic. Biochem. Physiol.* **71**, 20–28 (2001).
12. L. Guenin-Mace *et al.*, Shaping mycolactone for therapeutic use against inflammatory disorders. *Sci. Transl. Med.* **7**, 1-10 (2015)
13. M. H. Lacoske, E. A. Theodorakis, Spirotetronate polyketides as leads in drug discovery. *J. Nat. Prod.* **78**, 562–575 (2015).
14. J. I. Levin, *Macrocycles in Drug Discovery* (Royal Society of Chemistry, 2014).
15. A. Mve-Obiang, R. E. Lee, F. Portaels, P. L. C. Small, Heterogeneity of mycolactones produced by clinical isolates of *Mycobacterium ulcerans*: implications for virulence. *Infect. Immun.* **71**, 774–783 (2003).
16. N. Hay, Upstream and downstream of mTOR. *Genes Dev.* **18**, 1926–1945 (2004).
17. J. Choi, J. Chen, S. L. Schreiber, J. Clardy, Structure of the FKBP12-rapamycin complex interacting with binding domain of human FRAP. *Science.* **273**, 239–242 (1996).
18. J. Chen, X. F. Zheng, E. J. Brown, S. L. Schreiber, Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. *Proc. Natl. Acad. Sci. USA* **92**, 4947–4951 (1995).
19. J. Heitman, N. Movva, M. Hall, Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. *Science.* **253**, 905–909 (1991).
20. A. Y. Choo, J. Blenis, Not all substrates are treated equally: Implications for mTOR, rapamycin-resistance, and cancer therapy. *Cell Cycle*. **8**, 567–572 (2009).
21. M. A. Wear, A. Patterson, M. D. Walkinshaw, A kinetically trapped intermediate of FK506 binding protein forms in vitro: Chaperone machinery dominates protein folding in vivo. Protein Expr. Purif. 51, 80–95 (2007).
22. J. J. Siekierka, S. H. Hung, M. Poe, C. S. Lin, N. H. Sigal, A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. Nature. **341**, 755–757 (1989).
23. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T2 relaxation by mutual cancellation

of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).
24. W. F. Vranken *et al.*, The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.* **59**, 687–696 (2005).
25. C. C. S. Deivanayagam, M. Carson, A. Thotakura, S. V. L. Narayana, R. S. Chodavarapu, Structure of FKBP12.6 in complex with rapamycin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 266–271 (2000).
26. M. K. Rosen, S. W. Michnick, M. Karplus, S. L. Schreiber, Proton and nitrogen sequential assignments and secondary structure determination of the human FK506 and rapamycin binding protein. Biochemistry. **30**, 4774–4789 (1991).

Chapter 3

Cell-free protein expression of eukaryotic proteins

3. Cell-free protein expression of eukaryotic proteins

3.1. Introduction

Functional and structural protein studies are often limited by the availability of suitable amounts of protein. In particular, NMR spectroscopy requires several milligrams of isotopically labeled material. Historically, *in vivo* prokaryotic systems have been extensively used for heterologous protein expression. This is mainly due to the system's robustness and associated low cost of the preparation. However, the low achievable yield of important soluble targets, degradation, and aggregation represent the main disadvantages of this approach. Alternatively, Eukaryotic systems have compelling advantages, such as correct folding and PTM's. These are however expensive, difficult to set up and to maintain, and generally have low yields. In the last decades, *in vitro* protein synthesis has become a viable alternative for the production of problematic proteins. Improvements in both, the system and extract preparation, has led to the system's current state with many advantages such as considerable high expressed protein yields, fast expression times, easy screening and robust optimization of expression conditions, supplementation with suitable compounds (e.g., membrane mimetics and co-factors) (1–3), expression of toxic proteins (4, 5) and specifically for NMR the scrambling-free labeled amino acid incorporation (6, 7). The system continues to evolve at a fast pace (8, 9). Recently, microfluidics cell-free reactors have been introduced (10, 11), which have enabled research towards the development of single-dose protein production bioreactors for point-of-care systems (12).

System description

In cell-free systems, a target protein is translated by ribosomes extracted from cell lysates. Coupled transcription-translation systems combine exogenous DNA transcription with mRNA translation in a simultaneous process, avoiding RNA problematic isolation. The required gene of interest is under the control of a specific promoter, enabling specific transcription initiation through the addition of an exogenous purified RNA polymerase. Which leads to expression specificity and yield enhancement, through the increase of copy numbers for the gene of interest. Early systems expressed protein in a single container, the so-called batch-mode (13), which were limited to a few hours of expression before byproducts would inhibit expression. The continuous exchange method was introduced to overcome this issue (14, 15). In this system, the reaction mixture (RM) containing high molecular weight compounds is stored in a dialysis membrane surrounded by the feeding mixture (FM) containing molecular weight essential substrates (Figure 3-1). The continuous exchange between by-

products and substrates allows a near steady-state, allowing protein expression until exhaustion of substrates, allowing expression of milligrams of each protein of interest per reaction milliliter.

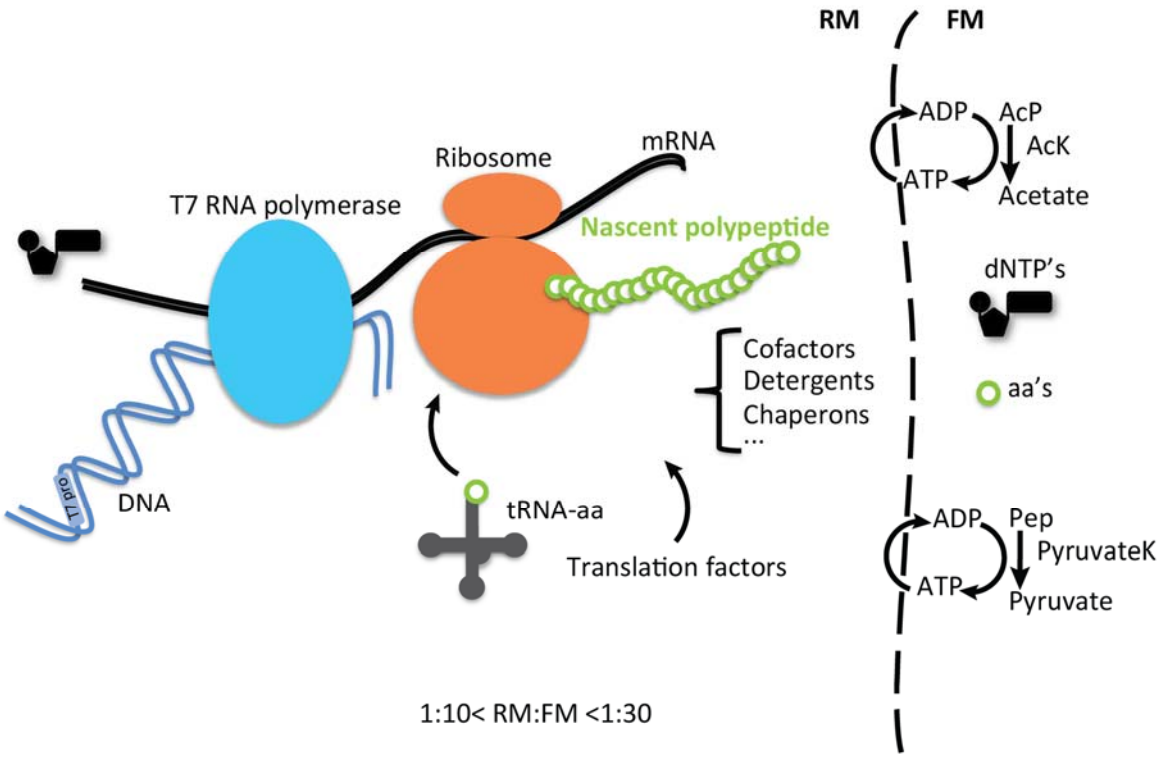


Figure 3-1 – Cell-free continuous exchange system setup, showing the main components of both the RM and FM compartments. In the RM, are present the high molecular weight components essential for transcription and translation, separated by a dialysis membrane from the FM, where the free-flowing small metabolites are in dynamic equilibrium with the RM, replenishing essential metabolites and removing by-products.

Newer methods have been developed through the reconstitution of the complete protein synthesis machinery from purified components (16). PURE is a modular, tailored system for cell-free expression. However, the limited yields, up to a third of conventional cell-free systems, and high cost, have limited its general application (17).

Cell-free system protein production for NMR

Cell-free protein expression systems have been shown in the past to successfully express complex targets such as toxic and transmembrane proteins in sufficient amounts for structural biology (4, 18). Several unique characteristics such as the lack of a functional metabolism and the inherent flexibility allow efficient isotope labeling of proteins, avoiding scrambling. The open nature of cell-free systems allows the manipulation of the reaction conditions and co-factor supplementation, in order to further enhance and stabilize tricky protein targets for high throughput production. Approaches such as the stereo-array isotope labeling method (SAIL), which uses specially synthesized amino acids containing stereospecific and region-specific isotope labeling patterns optimized for NMR, allowing spectra simplification (19).

3.1.1. Cell-free protein expression targets

The cell-free system is an ideal system for cytotoxic, such as toxins, as well as otherwise complicated expression targets, due to its lack of metabolism, and the possibility of supplementation with required ligands incompatible with conventional expression systems. The cell-free protein expression system robustness was tested with two distinct problematic eukaryotic targets, where expression had proven difficult with common expression systems.

Human Abelson tyrosine kinase

Protein kinases are key regulators of several cellular processes and have been implicated in neurological, cancer, infection, and immunological disorders (20, 21), one of the most pursued classes of targets for drug research (22). The high incidence of kinase inhibitor resistance during patient treatment has created a high demand for the development of novel drugs (23). Human Abelson tyrosine kinase (Abl) has been implicated in numerous processes (24), such as cell differentiation (25, 26), cell adhesion (27) and stress response (28). The abnormal, reciprocal translocation between Abl1 gene in chromosome 9 and the breakpoint cluster region (Bcr) gene in chromosome 22 generates the deregulated fusion protein Bcr-Abl associated with chronic myeloid leukemia (29, 30). Imatinib was one of the first-generation kinase inhibitors developed, used in the treatment of multiple cancers. Despite the early therapeutic success of developed inhibitors, the occurrence of different single point mutations in the Bcr-Abl tyrosine kinase domain has led to patient relapse upon treatment due to drug resistance. Of these, the “T315I” gatekeeper (GK) mutation exhibits resistance to most currently available kinase inhibitor drugs (31, 32). The design of next-generation drugs to overcome drug resistance can benefit from structural insights on the gatekeeper mutant and its resistance mechanism.

Human LPS-binding protein

Septic shock is a sepsis-induced hypotension, which leads to abnormalities in cellular metabolism. Patient infection can cause multiple organ failure and death in up to 50% of the cases (33). Bacterial infection leads to the immunity system exposure to lipopolysaccharide (LPS), the main component of the outer leaflet of Gram-negative bacteria. The septic response has been shown to be associated with the LPS-binding protein (LBP) and the opsonic receptor CD14 (34). The interaction complex between LPS and LBP is structurally uncharacterized. Due to the high affinity of LBP to LPS, the lack of LPS in cell-free expression makes it an ideal system for the expression of LBP. The structural characterization of the human LBP and LPS complex should benefit rapid patient diagnosis and treatment (35, 36).

3.2. Materials and methods

Cell-free extract preparation

E. coli S30 cell extract preparation was prepared following published protocols (37). In short, harvested A19 cells were washed three times with S30 buffer A1 (10 mM Tris-acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 6 mM β-ME). After resuspension in S30-buffer B1 (10 mM Tris-acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 1 mM DTT, 0.1 mM PMSF), cells were lysed using a French press and centrifuged twice at 30,000×*g*, 4°C. To remove endogenous mRNA, the extract was incubated 45 min at 42°C and then dialyzed at 4°C overnight against S30-buffer C (10 mM Tris-acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 0.5 mM DTT) with a 12-14 kDa MWCO. To remove the remaining cellular debris, the cell extract was centrifuged at 30,000×*g*, 4°C, aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

DNA amplification

High amounts of purified target DNA are required for cell-free expression. The isolation of plasmid DNA was performed using midiprep or maxiprep kits (Macherey-Nagel) and respective protocols.

T7 RNA polymerase expression and purification

A pT7-911 expression plasmid containing T7-RNA polymerase under the control of a T7 promoter carrying an ampicillin resistance gene with N-terminal 6xHis-tag was obtained from Stephan Grzesiek (Biozentrum, Basel). Using the QuickChange protocol (Agilent), and two consecutive reactions with the forward primers 1st half: 5'-CGCATCACCATCACCATCACGGAGAAAACCTGTCCATGAACACG and 2nd half: 5'-CCATCACGGAGAAAACCTGTACTTCCAGTCCATGAACACGATTAACA and reverse primers 1st half: 5'-CGTGTTCATGGACAGGTTTTCTCCGTGATGGTGATGGTGATGCG and 2nd half: 5'-TGTTAATCGTGTTCA TGGACTGGAAGTACAGGTTTTCTCCGTGATGG, a TEV cleavage site (residues -ENLYFQ-) was introduced between the N-terminal 6xHis-tag and the protein sequence (see Appendix for complete sequence).

For protein expression single colony was picked from a fresh transformation in BL21 (λDE3) competent cells and inoculated into 5 ml Luria Broth (LB) medium with ampicillin (Amp) and grown overday at 37°C. At the end of the day, 1 ml was inoculated in 20 ml of LB/Amp medium and grown overnight at 37°C. On the next day, the pre-inoculum was transferred to 1 l LB/Amp medium and grown at 37°C. At an OD_{600}=0.8 the culture was induced with 0.5 mM IPTG. After 4h of expression the cells were harvested (10 min, 5,500 rpm, 4°C, SLA3000 rotor), resuspended in 30 ml buffer A2 (50 mM Na_2HPO_4, pH=7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT, 20mM imidazole) and frozen at -20°C.

For purification the cells were thawed, DNase I (Roche), cOmplete (Roche), and lysozyme (Carl Roth) were added and stirred for 15 min on ice. Afterward, the cells were lysed with a French press (twice, 1500 PSI), and centrifuged (30 min, 16,000×*g*, 4°C, SS34 rotor). The cleared supernatant was applied to a 5 ml pre-equilibrated Ni-NTA gravity column (Genscript) and incubated for 2 h at 4°C on a shaking device. Afterward, the flow-through was collected, the column was washed with 10 column volume (CV) of buffer A, 5 CV 5% buffer B1 (buffer A1 with 500 mM imidazole) and then eluted with 5CV 75% buffer B1. The elution was left dialyzing in 2 l dialysis buffer (10 mM K_2HPO_4, pH=8, 150 mM NaCl, 1 mM DTT and 0.5 mM EDTA, 5% (v/v) glycerol) with a 10,000 MWCO (Thermo Scientific) dialysis membrane at 4°C. After 2 h, 30:1 mg of TEV protease (protein:protease) was added and left dialyzing overnight in 5 l dialysis buffer at 4°C.

The next day the dialysate was cleared by centrifugation (10 min, 16,000×*g*, 4°C). The cleared dialysate was applied to a 5 ml pre-equilibrated Ni-NTA gravity column and incubated for 1 h at 4°C on a shaking device. Afterward, the cleaved protein was collected both in the flow-through and in the dialysis buffer wash (5 CV). TEV protease and un-cleaved protein was eluted with 5 CV 100% buffer B. Both cleaved protein-containing fractions were pooled, concentrated (Amicon centrifugal filtering device with 30,000 MWCO) to about 15–20 mg/ml, and the glycerol adjusted to 50% (v/v), to a final protein concentration of 7.5–10 mg/ml. Final yield was 70 mg/l in LB medium, determined by UV/Vis spectroscopy. Protein purity was confirmed by an SDS-PAGE, afterward, the protein was aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.

Cell-free protein expression

A pET-based expression plasmid containing enhanced green fluorescent protein (GFP+) with C-terminal 6xHis-tag was used as a positive control of expression (38). Both Bcr-Abl, GK-Bcr-Abl and LBP constructs were obtained from Stephan Grzesiek's group (Biozentrum, Basel). For analytical-scale expression, 55 µl reactions were prepared in in-house built mini-continuous exchange cell-free (CECF) reactors used to optimize expressing conditions, with an optimal FM/RM volume ratio of 14.5. Protein expression was left overnight in a thermoshaker at 120 rpm, 30°C in a VWR Incubating Mini Shaker. The next day, the reaction mixture was centrifuged (10 min at 10,000 rpm, Eppendorf Centrifuge 5810 R), and both the resuspended pellet and the supernatant were diluted 1:7 with SDS-buffer and ran on an SDS-PAGE.

For each new batch of T7 RNA polymerase, an analytical-scale reaction comparing the expression of GFP+ with the previous T7 RNA polymerase optimal concentration and a range of concentrations for the new batch was set up. After overnight expression, the reaction mixture was pipetted from the analytical-scale reactor, and centrifuged (10 min, 12,000 rpm, Eppendorf Centrifuge 5810 R). The

supernatant was discarded, and the precipitate resuspended in an equal volume of buffer. For each reaction (in duplicate), both the precipitate and supernatant were run on an SDS-PAGE. For each new batch of extract, the optimal magnesium concentration was tested similarly.

Purification of cell-free expressed GFP+

The recovered cell-free GFP+ supernatants of 10 analytical-scale reactions were pooled and then incubated with 0.5 ml of Ni-NTA beads (Genescript, USA) for 60 min. After incubation, the beads were washed with 5 CV 2.5% buffer B1 and eluted with 20% buffer B1. The eluate was left dialyzing overnight in 5 l of 50 mM Hepes pH=7.5, 150 mM NaCl, 2mM DTT, with a 10,000 MWCO (Thermo Scientific) dialysis membrane at RT. The following day, the protein was concentrated (Amicon centrifugal filtering device with 10,000 MWCO). The concentration of GFP+ was measured to be 2.2 mg/ml with the BCA method (ThermoFisher Scientific, USA).

3.3. Results

Cell-free system optimization

The T7 RNA polymerase optimal concentration for protein expression was tested by comparing performance in the cell-free system. For each new batch of T7 RNA polymerase,GK abl an analytical-scale reaction was set up, comparing the soluble expression of GFP+ between batches of T7 RNA polymerase. The optimal soluble expression of GFP+ was found to be between 0.04–0.08 mg/ml, so the latter was chosen as the optimal concentration (Figure 3-2).

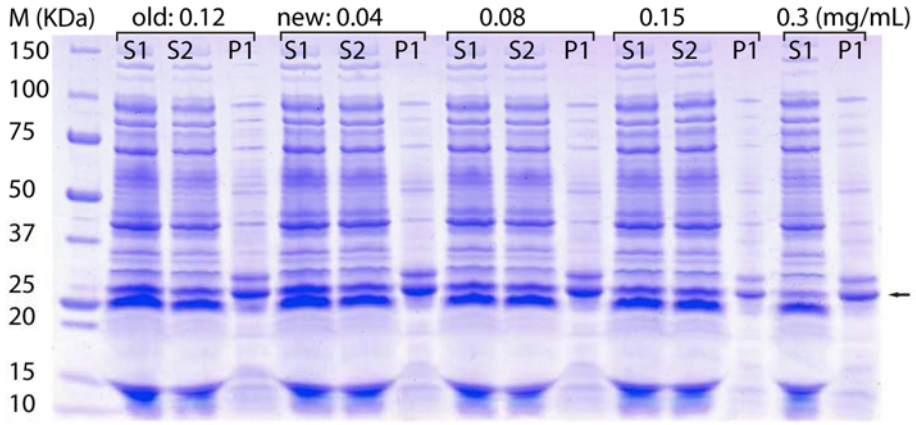


Figure 3-2 – SDS-PAGE of T7 RNA polymerase new batch test expression, comparing expression levels of GFP+ in the supernatant (S) and precipitate (P) between two different (old and new) batches.

Since expression rates are strongly dependent on the concentration of magnesium ions, magnesium concentration was adjusted for each batch of cell extract. In Figure 3-3, is exemplified the expression of GFP+ at different concentrations of magnesium. For this particular cell extract batch, the optimal

magnesium concentration was found to be between 18 mM and 20 mM per reaction mixture, where maximum protein yield was observed.

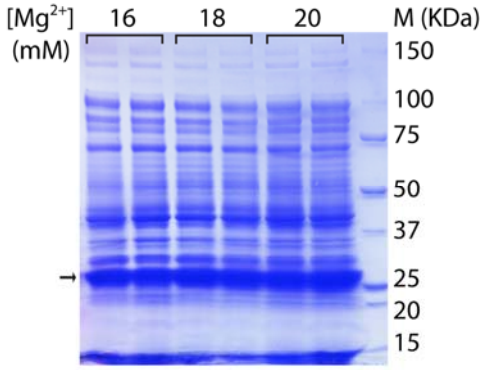


Figure 3-3 – SDS-PAGE of magnesium optimal concentration for cell-free soluble protein expression screen.

After optimization of vital components concentrations in the cell-free medium, the total quantity of soluble protein expression per ml of reaction medium was checked, once again with GFP+ as a control protein. A standard curve of purified GFP+ was prepared and analyzed in a fluorescence plate reader (Synergy H1, Biotek, USA) in triplicate (Figure 3-4).

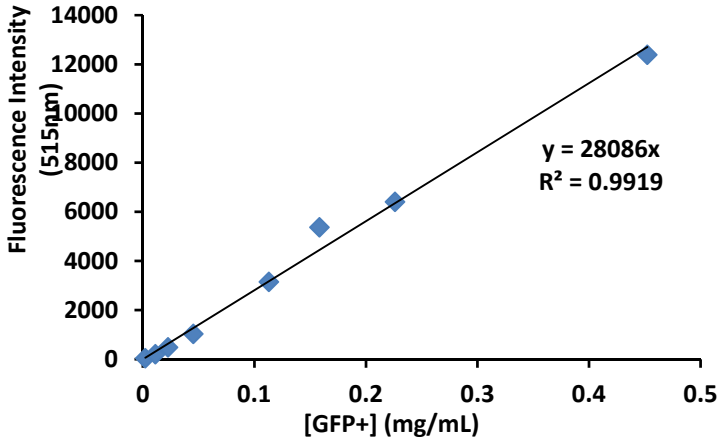


Figure 3-4 – Fluorescence intensity of standard curve prepared with purified GFP+.

The fluorescence values of a fresh magnesium optimization expression were measured and found to vary between ~10-18 mg/ml of GFP+ mg per ml of cell-free reaction volume, depending of magnesium concentration (Table 3-1).

Table 3-1 – Fluorescence intensity values and calculated GFP+ concentration, based on the calculated standard curve, of a GFP+ magnesium optimization reaction series.

[Mg2+] (mM)	12	14	16	18
I (515nm)	5520.5	6153	8356	9575.25
[sGFP] (mg/ml)	9.8	11	14.9	17

Protein expression

After successful setup and testing of *in vitro* protein expression, both eukaryotic proteins LBP and Abl were tested for soluble expression in comparison with GFP+ as positive control, shown in Figure 3-5. LBP protein expressed well, but as a precipitate. Different constructs were tested, with similar results. No further optimization was tried. Abl protein expressed poorly and solely as a precipitate.

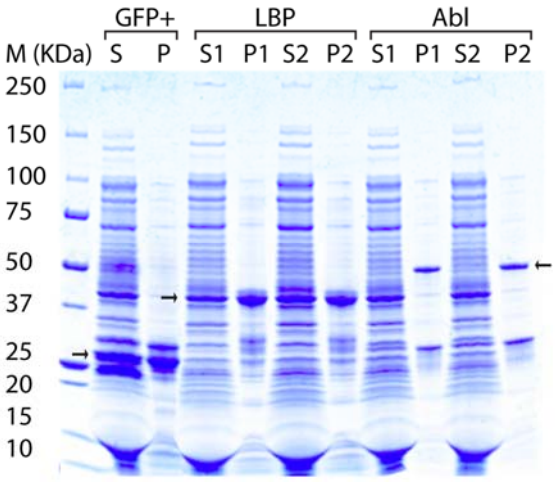


Figure 3-5 – SDS-PAGE of *in vitro* cell-free expressed LBP, Abl and positive control GFP+ supernatant (S) and precipitate (P) numbered when in duplicate.

The GK Abl mutant was also tested for *in vitro* expression. *Yersinia pestis* protein tyrosine phosphatase H (YopH) toxin was tested for expression and co-expression with its cognate target Abl. In addition, two Abl inhibitor drugs (GNF5 and Desatinib) were supplemented in the cell-free medium to test for soluble enhanced expression (Figure 3-6). In all cases, GK Abl expressed as a precipitate.

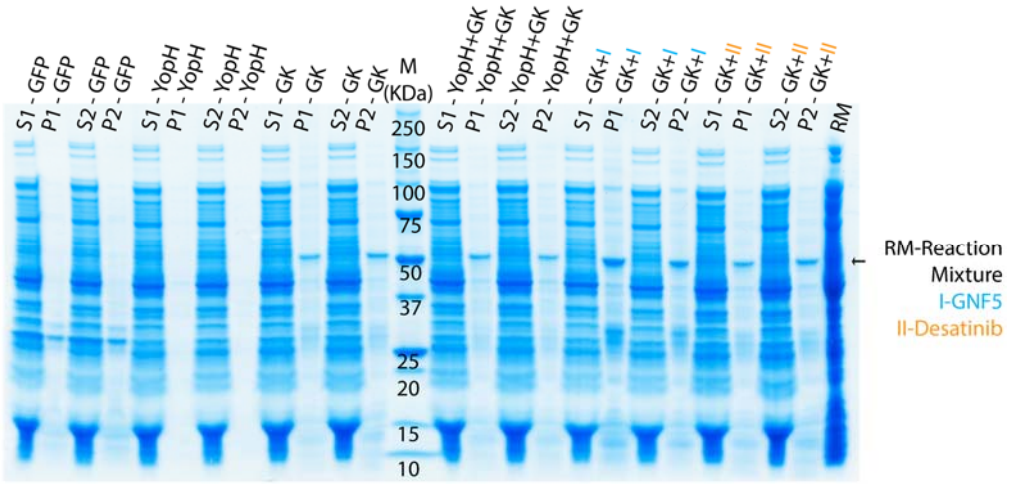


Figure 3-6 – SDS-PAGE of *in vitro* cell-free expressed Abl GK mutant (indicated) co-expressed with YopH phosphatase in the presence of inhibitors GNF5 and Desatinib. Supernatant (S) and precipitate (P) fractions are numbered when performed in duplicate.

3.4. Discussion

The cell-free continuous flow system was set up successfully, as shown with the *in vitro* expression of GFP+. The protein was expressed well-folded with a yield between 10-17 mg/ml of cell-free medium, a very high-level of expression. The cell-free system was set up primarily for the expression of cytotoxic proteins. The *in vitro* expression system robustness was also tested for the successful soluble expression with two significant eukaryotic protein targets of known difficult expression. Human LBP expressed in high amounts but insoluble, so no further optimization was tested. Abl and the gatekeeper mutant are challenging low expressing proteins in both prokaryotic and eukaryotic traditional expressions systems. Cell-free protein expression of the gatekeeper mutant was tested in the presence of different commercial kinase inhibitors as well as in the presence of YopH a bacterial toxin that targets Abl kinase, to test for protein stabilization. In all cases, almost all protein was expressed in the precipitate form. Protein expression optimization was not performed. Co-factors, such as chaperones could be used in the future to enhance correct folding and therefore solubility of the gatekeeper mutant.

These two examples also serve to demonstrate the current limitations of *in vitro* protein expression systems. The optimal protein high yield of prokaryotic systems proves to be insufficient possibly due to lack of post-translation modifications, required for specific eukaryotic targets. In sub-chapters 4.2 and 5.2, the successful expression of two labeled protein toxins is shown, using the described *in vitro* cell-free continuous flow protein expression system, in sufficient amounts for NMR studies.

3.5. References

1. Greiner, P., Hannappel, A., Werner, C. and Ludwig, B. Biogenesis of cytochrome c oxidase - *in vitro* approaches to study cofactor insertion into a bacterial subunit I. Biochimica et Biophysica Acta, **1777**, 904–911 (2008).
2. Lyukmanova, E.N., Shenkarev, Z.O., Khabibullina, N.F., Kopeina, G.S., Shulepko, M.A., Paramonov, A.S., et al. Lipid-protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: Comparison with detergent micelles, bicelles and liposomes. Biochimica et Biophysica Acta, **1818**, 349–358 (2012).
3. Shrestha, P., Smith, M.T. and Bundy, B.C. Cell-free unnatural amino acid incorporation with alternative energy systems and linear expression templates. New Biotechnology, **31**, 28–34 (2014).
4. Chalmeau, J., Monina, N., Shin, J., Vieu, C. and Noireaux, V. α-Hemolysin pore formation into a supported phospholipid bilayer using cell-free expression. Biochimica et Biophysica Acta, **1808**, 271–278 (2011).
5. Xu, Z., Chen, H., Yin, X., Xu, N. and Cen, P. High-level expression of soluble human β-defensin-2 fused with green fluorescent protein in *Escherichia coli* cell-free system. Applied Biochemistry and Biotechnology, **127**, 53–62 (2005).

6. Guignard, L., Ozawa, K., Pursglove, S.E., Otting, G. and Dixon, N.E. NMR analysis of *in vitro*-synthesized proteins without purification: A high-throughput approach. FEBS Letters, **524**, 159–162 (2002).
7. Terada, T. and Yokoyama, S. *Escherichia coli* Cell-Free Protein Synthesis and Isotope Labeling of Mammalian Proteins. 1st ed., Elsevier Inc. (2015)
8. Galinis, R., Stonyte, G., Kiseliovas, V., Zilionis, R., Studer, S., Hilvert, D., et al. DNA nanoparticles for improved protein synthesis *in vitro*. Angewandte Chemie International Edition, **55**, 3120–3123 (2016).
9. Liutkus, M., Fraser, S.A., Caron, K., Stigers, D.J. and Easton, C.J. Peptide synthesis through cell-free expression of fusion proteins incorporating modified amino acids as latent cleavage sites for peptide release. ChemBioChem, **17**, 908-912 (2016).
10. Zubay, G. In vitro synthesis of protein in microbial systems. Annual review of genetics, **7**, 267–87 (1973).
11. Spirin, A.S., V. I. Baranov, L. A. Ryabova, S. Y. Ovodov and Y. B. Alakhov A continuous cell-free translation system capable of producing polypeptides in high yield. Synthesis, **242**, 1162–1164 (1988).
12. Kim, D.M. and Choi, C.Y. A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. Biotechnology Progress, **12**, 645–649 (1996).
13. Kuruma, Y. and Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. Nature protocols, **10**, 1328–44 (2015).
14. Hillebrecht, J.R. and Chong, S. A comparative study of protein synthesis *in vitro* systems: from the prokaryotic reconstituted to the eukaryotic extract-based. BMC.Biotechnol., **8**, 58 (2008).
15. Reckel, S., Gottstein, D., Stehle, J., Löhr, F., Verhoefen, M.-K., Takeda, M., et al. Solution NMR Structure of Proteorhodopsin. Angewandte Chemie International Edition, **50**, 11942–11946 (2011).
16. Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A. and Güntert, P. Optimal isotope labelling for NMR protein structure determinations. Nature, **440**, 52–57 (2006).
17. Manning, G. The protein kinase complement of the human genome. Science, **298**, 1912–1934 (2002).
18. Lahiry, P., Torkamani, A., Schork, N.J. and Hegele, R.A. Kinase mutations in human disease: interpreting genotype–phenotype relationships. Nature Reviews Genetics, **11**, 60–74 (2010).
19. Cohen, P. and Alessi, D.R. Kinase Drug Discovery – What's Next in the Field? ACS Chemical Biology, **8**, 96–104 (2013).
20. Barouch-Bentov, R. and Sauer, K. Mechanisms of drug resistance in kinases. Expert Opinion on Investigational Drugs, **20**, 153–208 (2011).
21. Wang, J.Y.J. The Capable ABL: What Is Its Biological Function? Molecular and Cellular Biology, **34**, 1188–1197 (2014).
22. Kua, H.-Y., Liu, H., Leong, W.F., Li, L., Jia, D., Ma, G., et al. c-Abl promotes osteoblast expansion by differentially regulating canonical and non-canonical BMP pathways and p16INK4a expression. Nature Cell Biology, **14**, 727–737 (2012).
23. Ghosh-Choudhury, N., Mandal, C.C., Das, F., Ganapathy, S., Ahuja, S. and Ghosh Choudhury, G. c-Abl-dependent Molecular Circuitry Involving Smad5 and Phosphatidylinositol 3-Kinase Regulates Bone Morphogenetic Protein-2-induced Osteogenesis. Journal of Biological Chemistry, **288**, 24503–24517 (2013).
24. Woodring, P.J. Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. Journal of Cell Science, **116**, 2613–2626 (2003).
25. Shaul, Y. and Ben-Yehoyada, M. Role of c-Abl in the DNA damage stress response. Cell Research, **15**, 33–35 (2005).
26. Lugo, T., Pendergast, A., Muller, A. and Witte, O. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. Science, **247**, 1079–1082 (1990).
27. Groffen, J., Stephenson, J.R., Heisterkamp, N., de Klein, A., Bartram, C.R. and Grosveld, G.

Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell, **36**, 93–99 (1984).
28. Mian, A.A., Schüll, M., Zhao, Z., Oancea, C., Hundertmark, A., Beissert, T., et al. The gatekeeper mutation T315I confers resistance against small molecules by increasing or restoring the ABL-kinase activity accompanied by aberrant transphosphorylation of endogenous BCR, even in loss-of-function mutants of BCR/ABL. Leukemia, **23**, 1614–1621 (2009).
29. Bose, P., Park, H., Al-Khafaji, J. and Grant, S. Strategies to circumvent the T315I gatekeeper mutation in the Bcr-Abl tyrosine kinase. Leukemia research reports, **2**, 18–20 (2013).
30. Martin, G.S. Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. Expert Review of Anti-infective Therapy, **10**, 701–706 (2012).
31. Wright, S.D., Ramos, R.A., Tobias, P.S., Ulevitch, R.J. and Mathison, J.C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. Science, **249**, 1431–3 (1990).
32. Kipnis, E., Ogé, L., Soudan, B., Leroy, B., Vallet, B. and Lebuffe, G. Time course of IL-6 and LBP, candidate biomarkers of sepsis in surgical critical care. Critical Care, **14**, 38 (2010).
33. Zweigner, J., Gramm, H.J., Singer, O.C., Wegscheider, K. and Schumann, R.R. High concentrations of lipopolysaccharide-binding protein in serum of patients with severe sepsis or septic shock inhibit the lipopolysaccharide response in human monocytes. Blood, **98**, 3800–3808 (2001).
34. Reckel, S., Sobhanifar, S., Durst, F., Löhr, F., Shirokov, V.A., Dötsch, V., et al. Strategies for the cell-free expression of membrane proteins. In Methods in Molecular Biology, **607**, 187–212 (2010).
35. Karunakaran, R. A family of promoter probe vectors incorporating autofluorescent and chromogenic reporter proteins for studying gene expression in Gram-negative bacteria. Microbiology, **151**, 3249–3256 (2005).

Chapter 4

Structural and conformational elucidation of membrane-bound and the membrane-inserted states of pore-forming toxins

4. Structural and conformational elucidation of membrane-bound and the membrane-inserted states of pore-forming toxins

4.1. The bacterial toxin Colicin Ia

4.1.1. Introduction

Colicin family

Colicins are a family of bactericidal proteins produced by *E. coli* to kill competing susceptible bacterial strains. This protein family exhibits distinct cytotoxic activities, and thus far four cytotoxic classes of colicin have been identified: the pore-forming colicins such as ColB, ColIa, and ColN; RNase colicins, such as ColE3; DNase colicins such as ColE9; and inhibitors of cell wall synthesis such as ColM (1, 2). Though they exhibit different activities, all share domain function and structural organization. All members contain three domains each involved in the three parasitic stages: the N-terminal translocator domain (T-domain) responsible for outer membrane (OM) translocation; the middle receptor binding domain (R-domain) responsible for binding to each respective receptor in the OM of susceptible targets (typically OM porins); and the C-domain, which is responsible for the onset of cell death (2, 3) (Figure 4-1). Colicins can further be subdivided depending on which host system is exploited for their translocation. Group A colicins, exploit *E.coli*'s Tol-dependent translocation system, whereas group B exploits the Ton-dependent translocation system. Both groups hijack and exploit the proton motive force of the host cell for colicin insertion into the inner membrane (1, 2).

Colicins are plasmidic proteins, which are co-expressed with their specific immunity protein. The immunity protein inactivates the toxic activity of the colicin, protecting the expressing cell, which is achieved through the formation of a tight 1:1 complex with the C-domain. Disassembly of the complex occurs upon interaction with the OM receptor of the host cell, prior to OM permeation (2).

Colicin Ia

ColIa is a 626 residue (69 kDa) α-helical protein. Structurally, the N-terminal translocator-binding domain is separated from the C-terminal pore-forming domain through a long coiled-coil, with the receptor-binding domain sitting in between as depicted in Figure 4-1. The C-domain is the cytotoxic domain, exhibiting pore-forming activity in the presence of membranes.

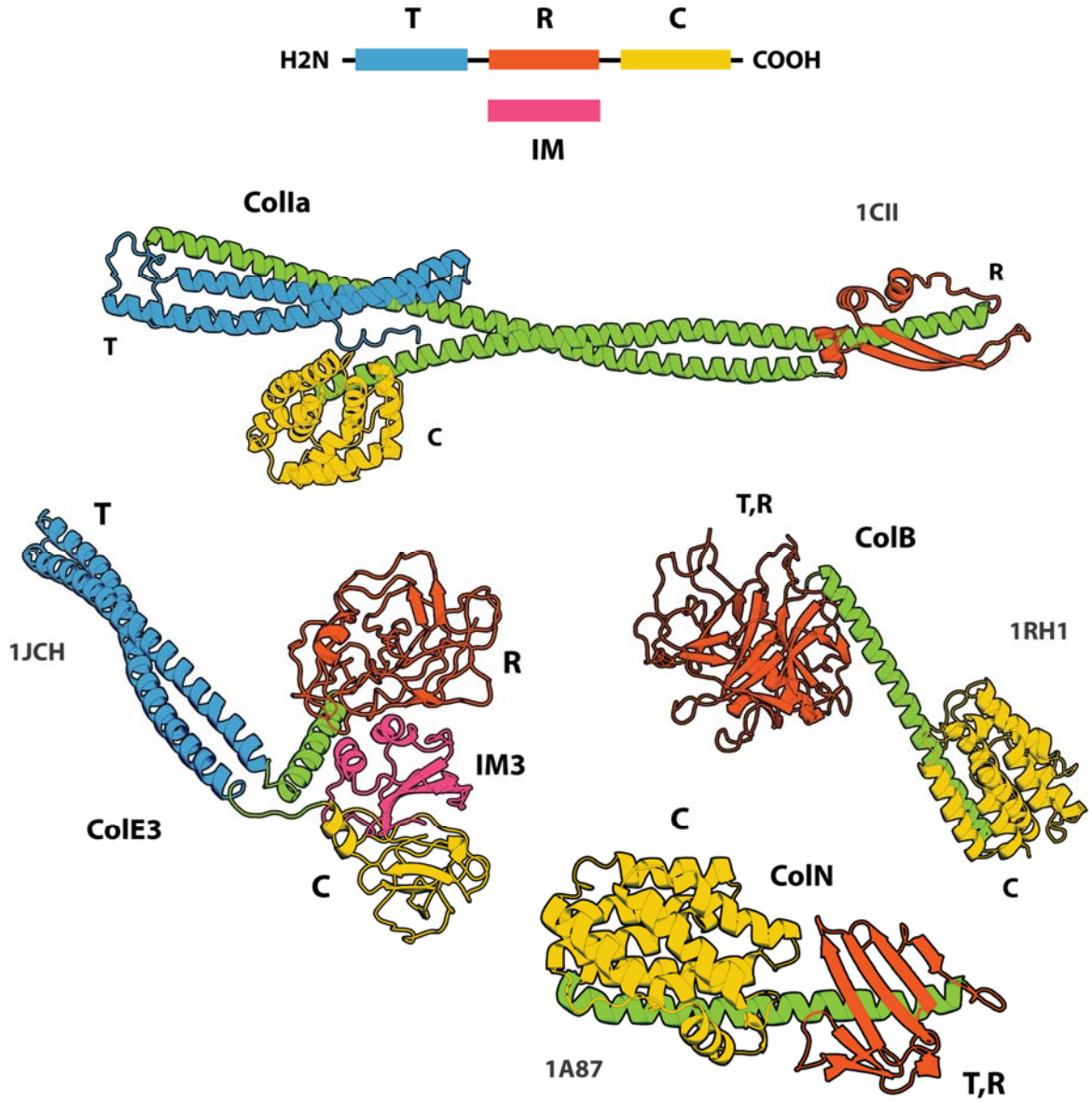


Figure 4-1 – General domain architecture of colicin family members and known full-length structures of different colicin members with respective PDB ID. The three domains involved in receptor binding (R-domain), translocation (T-domain) and cytotoxicity (C-domain) are highlighted in different colors. For colicins ColB and ColN the T as well as the R-domain are comprised of a single domain and are indicated as "T,R" shown in red. For ColE3, the complex with the immunity protein (IM3) is shown.

ColIa is part of the colicin group B since it exploits the host Ton system as energy donor for translocation (4). Further, it uses the colicin I receptor (Cir; 70 kDa β-barrel protein) both as the receptor (4) and translocator (5) (Figure 4-2). During OM translocation the R-domain first binds to Cir with high affinity, then the T-domain acts as a scaffold to subsequently search for a second copy of Cir to which it binds with lower affinity. Subsequently, the inner membrane (IM) host Ton system recognizes a TonB box in the N-terminal part of ColIa, which provides the necessary energy for translocation. This, in turn, leads to a necessary partial or possibly full unfolding of both T and C-domain. It was previously suggested that both the release of the immunity protein and C-domain cleavage occurs at this point. Nevertheless, the order of events and the precise insertion mechanisms

are currently unclear (1, 2). Likewise, other non-pore-forming colicins were shown to require the cleavage of the C-domain for cytotoxicity (6, 7). The mechanism and conformational changes occurring after translocation of the C-domain into the periplasm, leading to the insertion of the C-domain into the IM are still unclear.

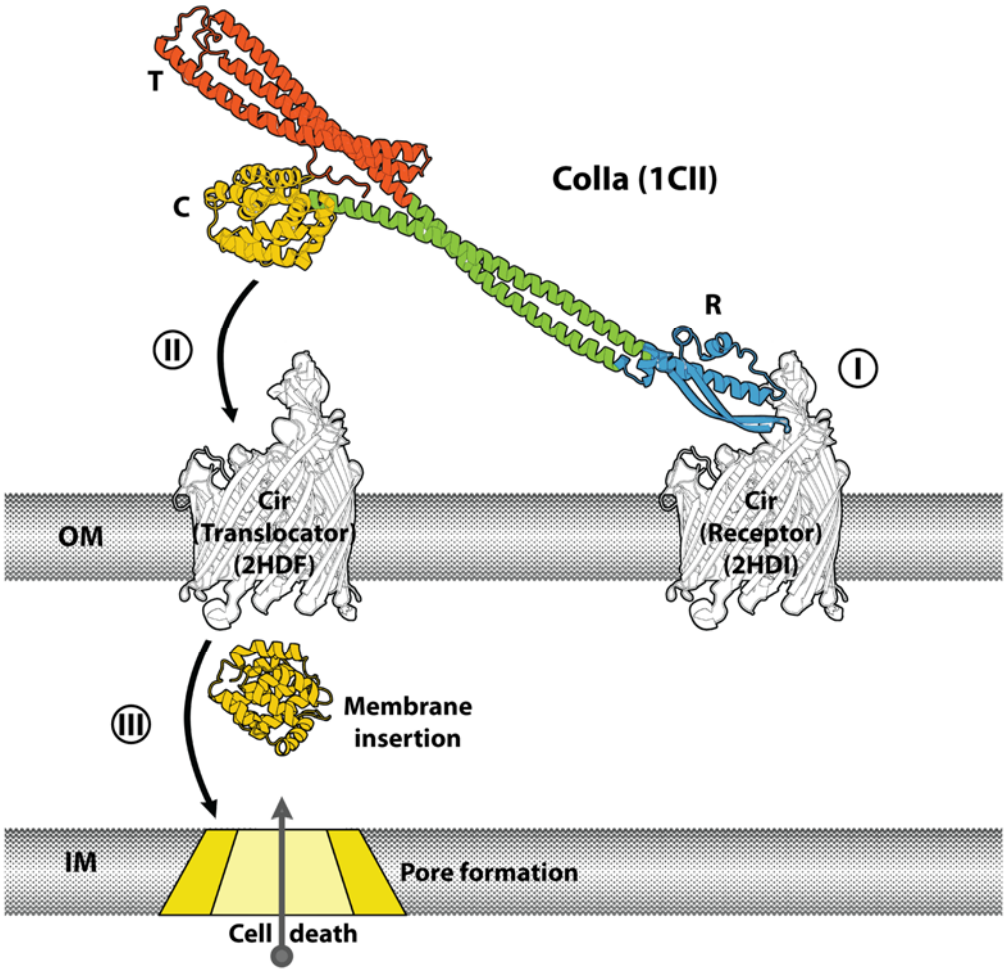


Figure 4-2 – Translocation of Colicin Ia into *E.coli*. In the first step, the R-domain (blue) binds the Cir receptor with high affinity. Meanwhile, the T-domain (red) recognizes the Cir translocator, helping the C-domain (yellow) permeate the outer membrane (OM). In the last step, ColIa inserts itself in the inner membrane (IM) forming a pore through oligomerization.

Colicin Ia C-domain

Due to the limited cavity size of the translocator Cir, ColIa is thought to unfold during translocation, as has been shown for other colicins (4, 8). Very little is known concerning the Cir bound T- and R-domains fate following C-domain translocation and pore formation. It represents a steric barrier for several ColIa's to bind adjacent Cir's in order to span the membrane to form an oligomeric pore within the IM, while the R-domain remains bound to Cir. The most plausible explanations include the release of the R-domain or the cleavage of the C-domain. Conversely, dissociation constant in the nanomolar range (9) poses an impressive energy barrier for the former. While for the latter, it is possible that the

unfolding of the C-domain exposes an unknown cleavage site necessary for translocation. This would energetically and sterically corroborate the multimeric pore complex hypothesis. Interestingly, it could be shown that the C-domain alone is not cytotoxic, whereas a mutant lacking the receptor domain showed about half the killing efficiency of the whole ColIa (5). This lower efficiency can be attributed to the lower binding affinity of T-domain to the Cir translocator. Therefore, pore formation remains possible due to the second binding event. Furthermore, the fact that the R-domain's absence does not preclude pore-formation, indicating that full-length ColIa bound to the Cir receptor copy on the OM is not a prerequisite for the formation of the pore, shows that the long coiled-coil linker between the domains, including helix 1, is not essential for pore formation.

Colicin Ia C-domain morphology

In the full-length ColIa amino acid sequence, helix 1 (residues 359-467; Figure 4-3) is situated between the C- and R-domains. The exact sequence limits between helix 1 and the C-domain are not clearly defined (10). Due to the lack of detailed understanding of the translocation mechanism, the significance of ColIa C-domain's helix 1 in translocation and IM channel formation remains unclear. It is generally agreed that the C-domain comprises residues 438/453-626 of full-length ColIa, including several residues attributed to helix 1 (10) (Figure 4-3).

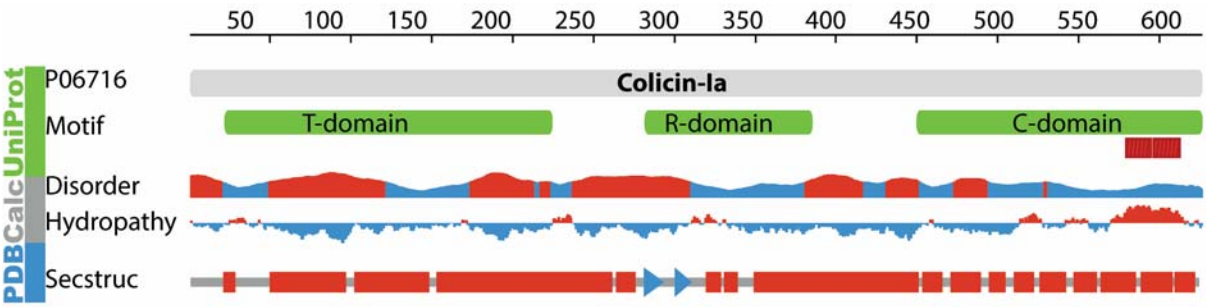


Figure 4-3 – Colicin Ia protein feature view of PDB entries mapped to UniProtKB P06716 sequence.

Helix 1 is a prime target for the location of the previously mentioned cleavage site, responsible for ColIa's C-domain release before IM insertion. In the soluble structure of full-length ColIa, the C-domain exhibits the common PFT compact fold as shown in Figure 4-4 (11). It consists of ten α-helices including part of helix 1. The hydrophobic pore-forming hairpin comprised of helices 8 and 9 is embedded within the core of the protein, surrounded by several amphiphilic helices. The C-terminal transmembrane helix 10 completes the PFT analogous fold.

The early steps of ColIa's translocation are currently well described in the literature, but the C-domain's unfolding and the subsequent insertion mechanisms remain elusive. So far the only structure of a colicin IM pore is based on the imaging of an oligomer of full-length ColIa in a lipid bilayer by

negative staining electron microscopy (12), showing a low-resolution structure of a hexamer assembly protruding from the membrane.

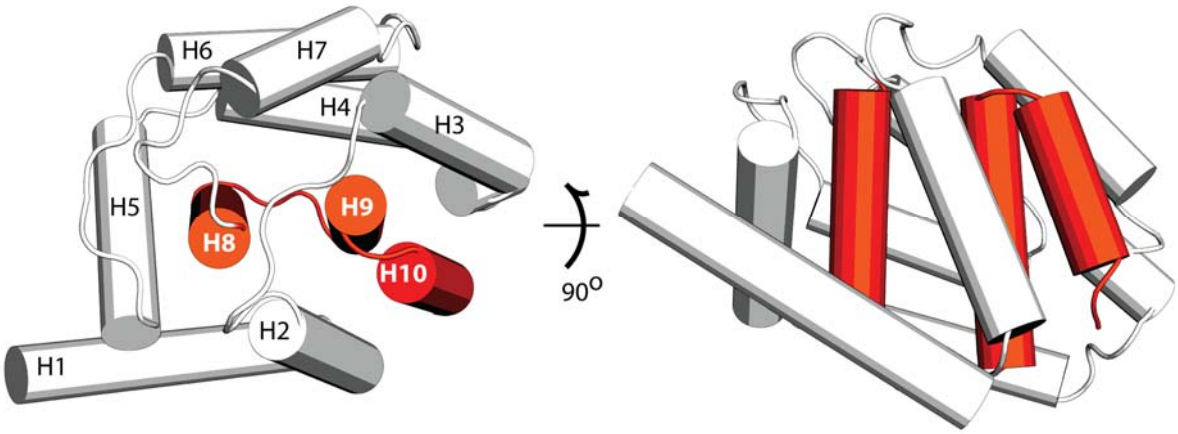


Figure 4-4 – Structure of the C-domain of ColIa (PDB ID: 1CII), showing both top and side views and annotated helices. The pore-forming domain (helix 8 and 9) and the transmembrane domain (helix 10) are highlighted in red.

4.1.2. Material and methods

Construct cloning

The CT-M C-domain construct (10) containing residues 438-626 of FL-ColIa with an N-terminal His_6-His-tag followed by a thrombin cleavage site was changed to a TEV cleavage site using the Quickchange II site-directed mutagenesis (Agilent) protocol and the following primers: 5'-CATCACAG CAGCGGCGAAAACCTGTACTTCCAGCATATGCTCGAGGAG-3' and 5'-CTCCTCGAGCATATGCTGGAAGTAC AGGTTTTCGCCGCTGCTGTGATG-3'. From here forthwith, amino-acid sequence numbering refers to the cleaved product of the used construct (from the full-length sequence 438-626 to 4-192)

Protein expression and purification

A single colony was picked from a fresh transformation in BL21 (λDE3) chemically competent cells, inoculated into 5 ml LB/Amp and grown overday at 37°C. Afterward, a 1:20 inoculation in 20ml of LB (unlabeled protein) or M9 minimal medium (protein labeled with $^{15}NH_4Cl$ only or $^{15}NH_4Cl$ and ^{13}C-glucose in D_2O) was grown overnight at 37°C. The following day, the pre-inoculum was transferred to 1 l of medium and grown at 37°C until OD_{600}=0.8, when the culture was induced with 1 mM IPTG. After 4 h of expression the cells were harvested (10 min, 5,500 rpm, 4°C, SLA3000 rotor), resuspended in 30 ml buffer A (20 mM Hepes, pH=7.0, 100 mM NaCl, 20 mM imidazole) and frozen at -20°C.

For protein purification, the cells were thawed and both DNase I as well as lysozyme were added and the suspension was stirred for approximately 15 min at room temperature. Afterward, the cells were lysed with a French press (2 runs, 1500 PSI), and the cell debris was cleared by centrifugation (30min,

13,000 rpm, 4°C, SS34 rotor). The cleared supernatant was mixed with 5ml of pre-equilibrated Ni-NTA beads (Genscript, USA) in a gravity column and incubated for 30min at RT on a shaking device. Afterward, the flow-through was collected, the column was washed with 10 column volume (CV) of buffer A, 10 CV of 5% buffer B (buffer A with 500 mM imidazole) and then eluted with 5 CV of 75% buffer B. The elution was left dialyzing overnight against 5 l Buffer A (pH=7.2) with a 10,000 MWCO (Thermo Scientific, USA) dialysis membrane at RT.

The resulting dialysate was then cleared by centrifugation (10 min, 12,000 rpm) and was run on a 5ml Hitrap SP HP (GE Healthcare, USA) prepacked column equilibrated with IEX buffer A (50mM Hepes, pH=7.2), using an Äkta (GE Healthcare, USA) system. The protein was eluted with IEX buffer B (IEX buffer A supplemented with 1 M NaCl) over a gradient of 10CV. The elution fractions were collected pooled, and then incubated with 20:1 mg of tobacco etch virus (TEV) protease (protein:protease ratio), and left dialyzing overnight in 5 l TEV dialysis buffer (50 mM Tris-HCl, pH8, 150 mM NaCl, 1 mM DTT and 0.5 mM EDTA) with a 10,000 MWCO (Thermo Scientific, USA) dialysis membrane at RT.

Subsequently, the dialysate was transferred to a 2x 2.5 l buffer A dialysis for 1 h each, and then cleared by centrifugation (10 min, 12,000 rpm). The cleared dialysate was applied to a 5ml pre-equilibrated Ni-NTA gravity column and incubated for 30 min at RT on a shaking device. Afterward, the cleaved protein was collected both in the flow-through and in the buffer A wash (10 CV). TEV protease and un-cleaved protein were eluted with 10 CV 100% buffer B. Both cleaved protein containing fractions were pooled, concentrated with a 5,000 MWCO Amicon centrifugal filtering device (Sartorius, Germany) and ran on a pre-equilibrated S75 SEC column (GE Healthcare, USA), (20mM NaOAc, pH=4.5, or 50 mM Tris-HCl, pH=8) as a polishing step. Protein purity was checked with an SDS-PAGE.

Bicelle preparation

A 40% w/w 250 µl stock with q=2.8 DHPC/DMPC (Avanti Polar Lipids, USA) was prepared by weighting individual DMPC lipid and DHPC into a sample vial, adding milli-Q water, and left shaking on a vortex overnight at 4°C. In the morning the bicelle solution was quickly spun-down and heated at 42°C before storing at -20°C. A 10% and 40% stock of CHAPSO/DMPG stock was prepared similarly.

Crystallization

An 8% bicelle/ColIa (15-20 mg/ml) solution was prepared from a 40% bicelle stock solution (DMPC/DHPC, q=2.8), and incubated for 30 min at 4°C. A series of screening plates were set up using the sitting-drop vapor diffusion method with 1:2 bicelle/protein to reservoir solution, against 80 µl reservoir, set up manually at 4°C and stored at RT.

After successful condition screening, crystal quality was improved through crystal seeding. For that, the mother liquor was prepared by mixing 4:1 solution of protein buffer (50mM Tris pH8, 300mM NaCl) with bicelle stock, and then mixed with 0.1M potassium thiocyanate, 30% w/v PEG monoethyl-ether 2000 at a 1:2 ratio. Initial crystals were transferred to a sample vial with 50-100µl of mother liquor solution, and then crushed with the help of a pipette and subsequently vortexed. The seed stock was diluted 1:10 with mother liquor five times, every time using the previous diluted solution to a final dilution factor of 1:100,000. For the seeding trials, to 9 µl bicelle/protein mixture 1µl of each seeding dilution was added. In each crystallization plates well, 0.5 µl of different bicelle:protein seed dilution and 1µl of reservoir solution were pipetted. In the second well, the same solutions were pipetted, except for the seed dilution. The screening plate was then sealed and stored at RT. Drops with seeding optimized crystals were cryoprotected with a drop of perfluoropolyether, and then the crystals were fished and flash frozen in liquid N_2.

Solution NMR measurement and data analysis

2D [^{15}N,^{1}H]-TROSY-HSQC (13) NMR spectra were recorded at 30°C on a Bruker Ascend 700 spectrometer equipped with a cryogenic triple-resonance probe. NMR data were processed with TOPSPIN 3.2 (Bruker Biospin, USA) and analyzed with CCPNMR (26). For the sequence-specific backbone resonance assignment of [*U*-99%,^{2}H,^{13}C,^{15}N]-ColIa, the following NMR experiments were recorded: 2D [^{15}N-^{1}H]-TROSY-HSQC (13), 3D TROSY-HNCA, 3D TROSY-HNCACB, 3D TROSY-HNCO, 3D TROSY-HN(CO)CA, and 3D TROSY-HN(CA)CO (*14*). For side-chain resonance assignment, the following NMR experiments were recorded: 3D H(CCO)NH and 3D CC(CO)NH (15, 16).

4.1.3. Results

X-ray structure determination of ColIa C-domain

Bicelles were used as a membrane mimetic in an attempt to study the membrane intermediary and inserted states using both NMR spectroscopy and X-ray diffraction, due to the adequate bicelle size for solution NMR, and optimal membrane-mimetic properties due to the combination of the soluble detergent and lipid environment. At first, crystallization was attempted with the common

DHPC/DMPC bicelles. After an initial screening, crystal diffraction patterns showed peak twinning artifacts. A seeding trial was set up to improve the quality of the crystals. After optimization, crystals diffracted to ~1.8 Å (Figure 4-5). Processing of the acquired dataset showed a P1 space group and six molecules per asymmetric unit (unit cell constants: 37.15, 37.17, 209.63, 94.4, 90.44, 119.93) (Figure 4-6 A).

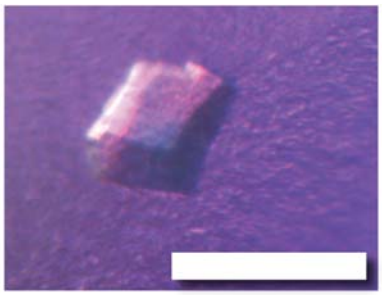


Figure 4-5 – Crystals of ColIa grown in the presence of DMPC/CHAPSO bicelles, which diffracted to a resolution of 1.8 Å. Scale bar: 200 nm.

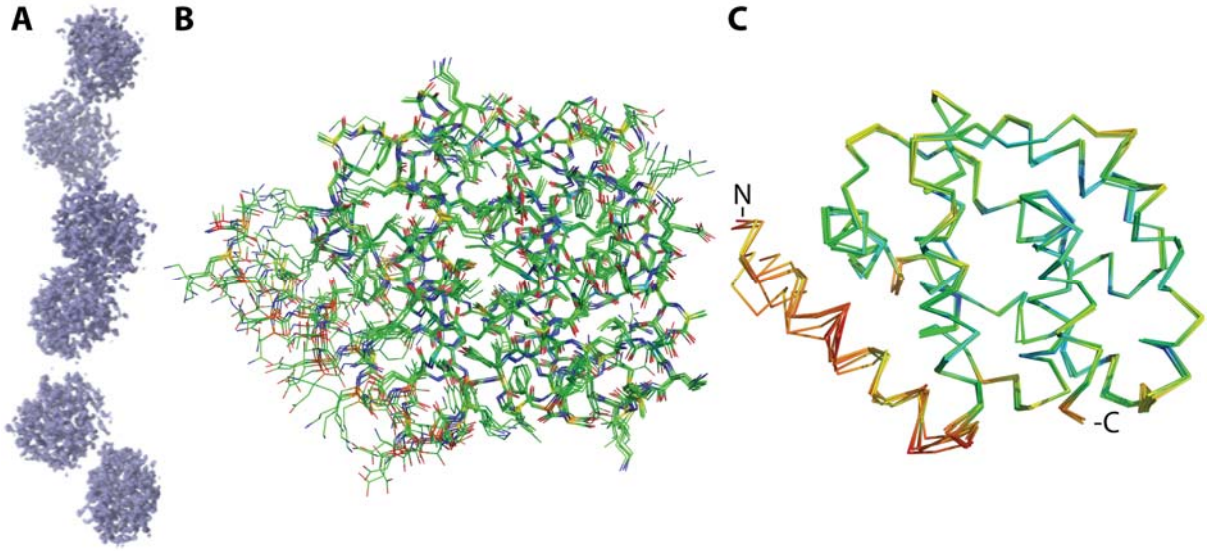


Figure 4-6 – In A is shown the electron density of the asymmetric crystal unit cell; B shows the superimposition of the each ColIa C-domain structure models in stick representation; in C, the backbone is colored by the crystal B-factors where blue is less dynamic and red is more dynamic.

The known structure of full-length ColIa (PDB ID: 1CII) was used as search models for molecular replacement. The resulting model revealed an identical conformation to the known soluble structure of ColIa C-domain. Figure 4-6 B and C, show the alignment of the six individual C-domains models, where the N-terminal helix 1 is shown to be considerably more flexible than other helices as evidenced by the increased B-factors.

Assignment of the soluble form of ColIa

Sequence-specific resonance assignment of ColIa was accomplished, using a triple-labeled (^{2}H,^{15}N,^{13}C) ColIa sample and several triple-resonance experiments (Figure 4-7, Figure 4-8 and Figure 4-9). With

these experiments, 90.5% (172 of 190 assignable residues) assignment of the backbone, and 64% of the proton side-chain assignments were achieved (Figure 4-9).

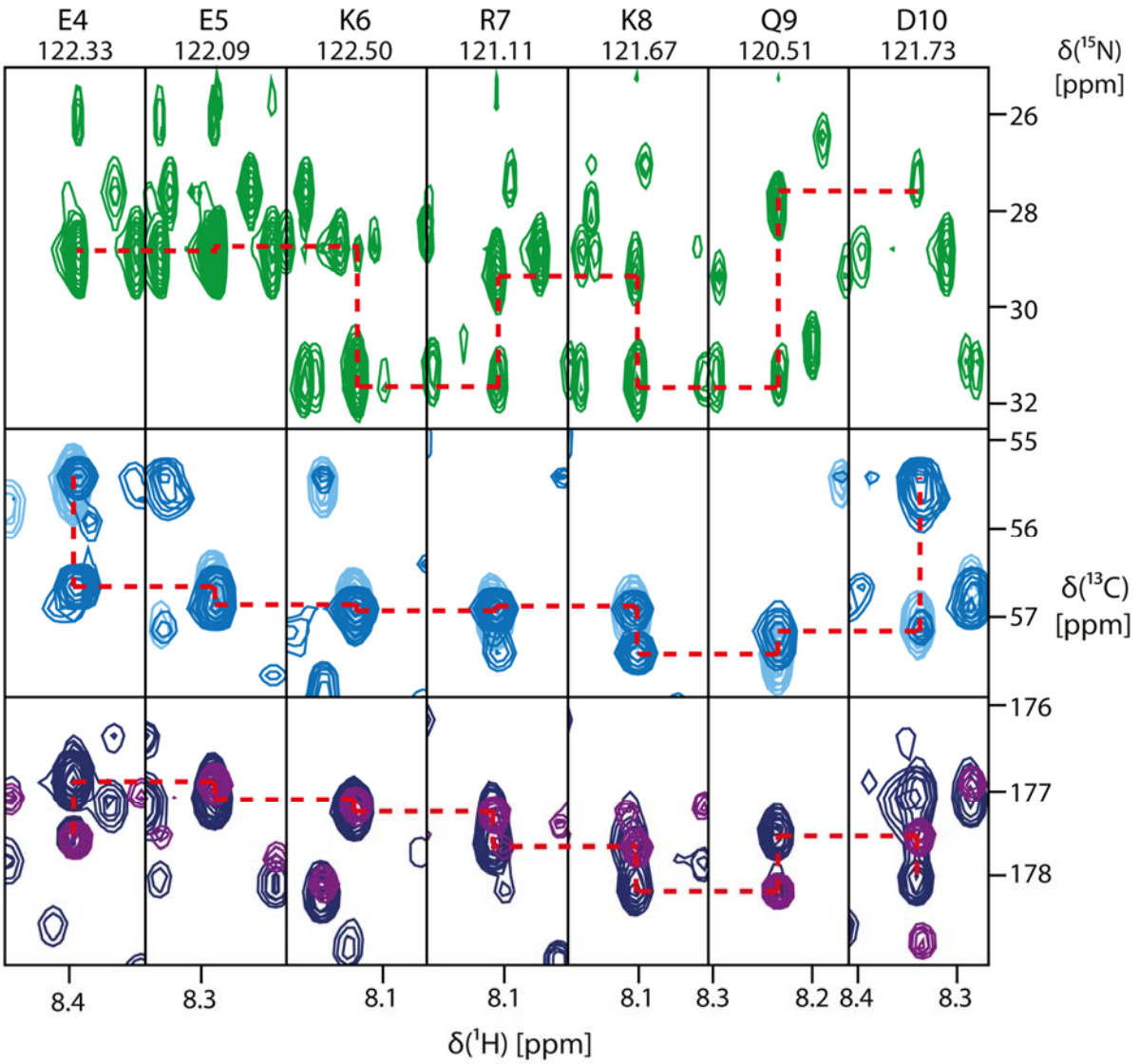


Figure 4-7 – Colla 3D backbone assignment strips corresponding to residues 4-10 of the C-domain. In the upper half in green are shown the HNCACB strips corresponding to the $C\beta_i$ and $C\beta_{i-1}$. In the middle panel, in blue are shown the HNCACB's $C\alpha_i$ and $C\alpha_{i-1}$ peaks, and in light blue the HNcoCA $C\alpha_{i-1}$ peaks, with sequential interresidue connections shown in red. In the bottom panel are shown in dark blue the HNcaCO's CO_i and CO_{i-1}, and in purple the HNCO's CO_{i-1}. The sequential interresidue connections are shown with a red dashed line.

The thirteen missing backbone assignments reside in helices 8 (residues 150-156) and 9 (residues 170-174, and 177-178), which comprise the pore-forming domain. To check for incomplete –NH deuterium/water exchange, a 2D-HSQC was run for a sample incubated for three days at 50°C (Figure 4-12), the superposition with the reference spectrum showed no additional resonances. The complete chemical shift list file was analyzed with the Chemical Shift Index (CSI) 3.0 (17), which utilizes backbone chemical shift and protein sequence data to identify secondary structure (Figure 4-11 C).

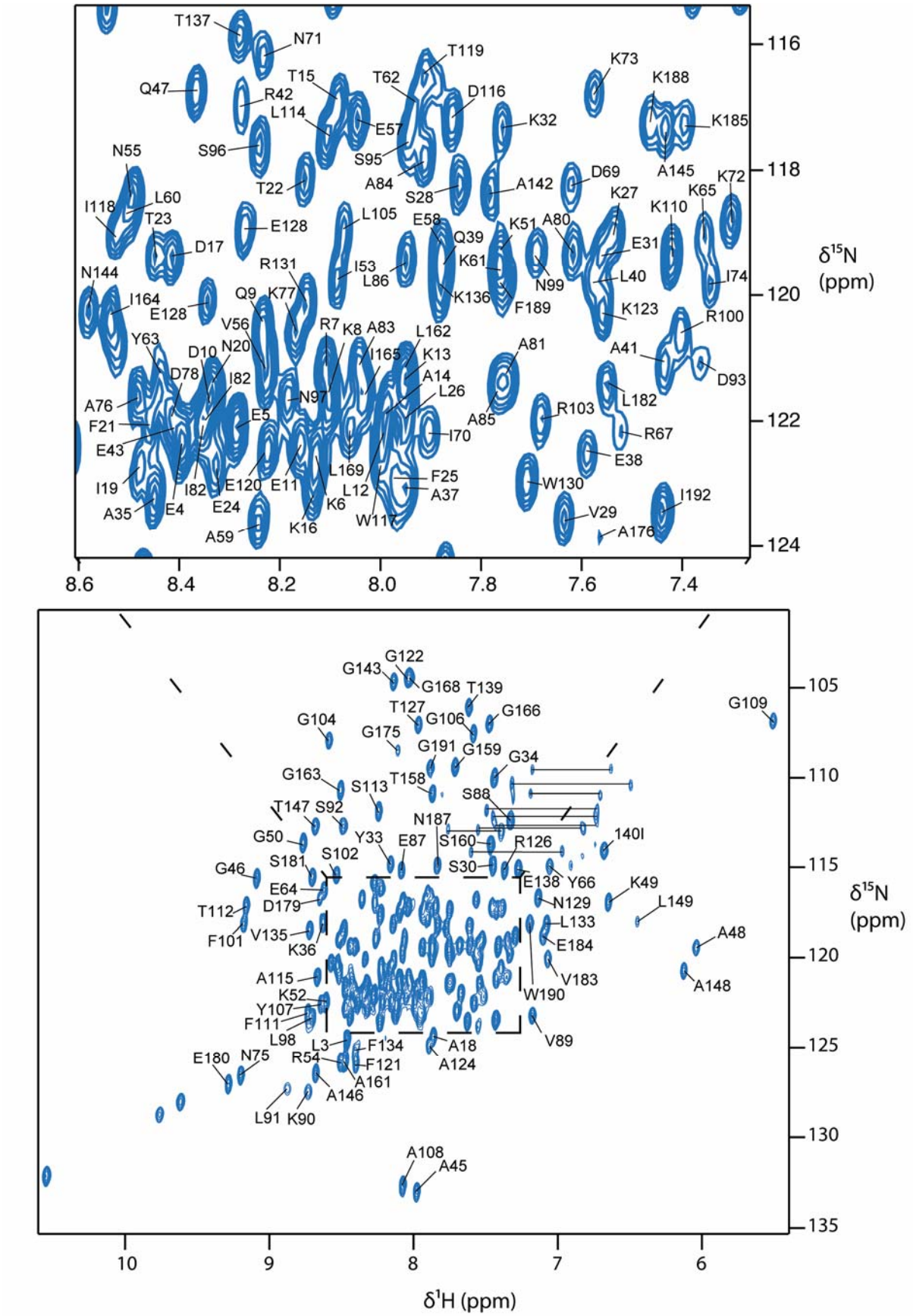


Figure 4-8 – Sequence- specific backbone resonance assignment of the soluble form of ColIa C-domain, shown on a 2D-[^{15}N,^{1}H]-TROSY-HSQC. The crowded region in the inset is shown in detail above.

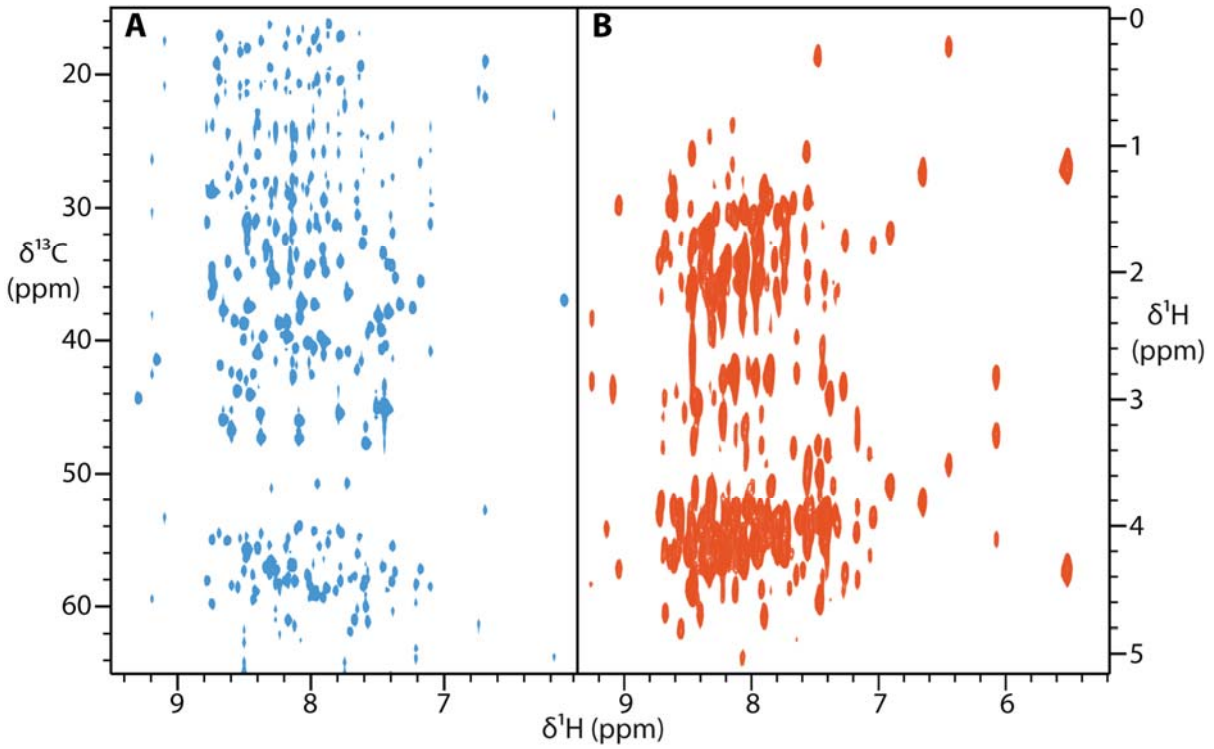


Figure 4-9 – Spectra used for assignment of Colla sidechains. A – 2D [^{1}H-^{13}C] projection of the 3D-CC(CO)NH spectrum; B - 2D [^{1}H-^{13}C] projection of the 3D-H(CCCO)NH spectrum.

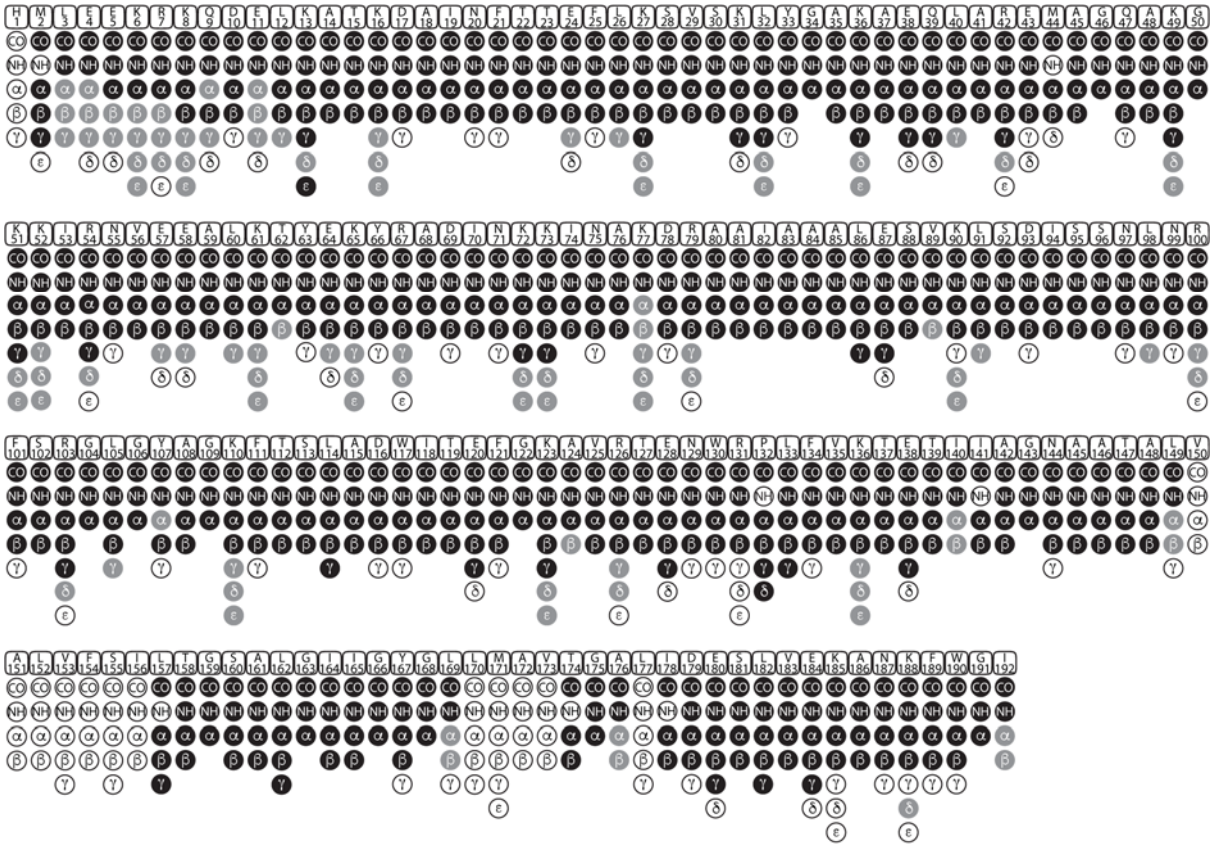


Figure 4-10 – Colla sequence showing backbone and sidechain assignment. For the sidechain full assignment is shown in black, carbon only in grey.

In addition, the difference between the CA and CB chemical shifts to the random coil values was plotted versus the residue sequence (Figure 4-11 D). The resulting secondary structure elements were compared with the here described as well as known crystal structures (Figure 4-11 A and B).

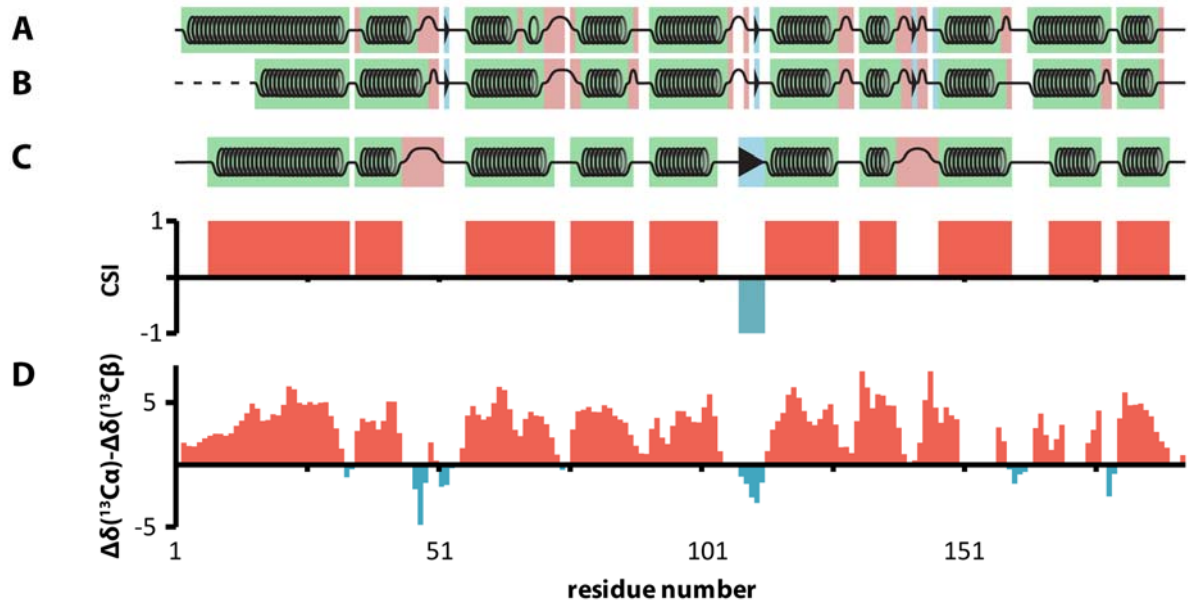


Figure 4-11 – Secondary structure of Colla. A – Secondary structure from Colla PDB ID: 1Cii (residues 448-624), calculated with DSSP; B - Secondary structure from Colla crystal structural model, calculated with DSSP; C – Predicted secondary structure of Colla from assigned chemical shift data, calculated with CSI 3.0; D – Plot of the differences between experimental 13Cα and 13Cβ chemical shifts and their corresponding "random-coil" chemical shifts. A weighting function with weights 1-2-1 for residues (i-1) – I – (i+1) has been applied to the raw data.

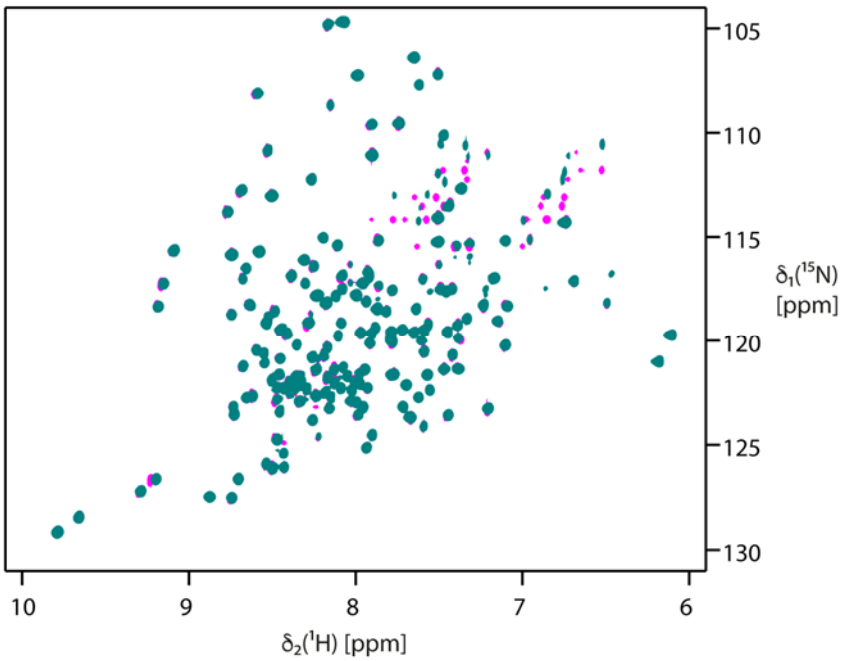


Figure 4-12 – 2D-[^{15}N,^{1}H]-TROSY-HSQC of free Colla (teal) and Colla heated at 50°C (pink).

ColIa Bicelle titration

After the unsuccessful crystallization of the membrane-inserted form of ColIa using DHPC/DMPC bicelles, CHAPSO/DMPG bicelles were tested for membrane interaction. As shown in Figure 4-13, ColIa was titrated with each bicelle type, and the detergents DHPC or CHAPSO as a detergent interaction reference to rule out detergent solubilization. In the case of DHPC, ColIa was shown to strongly interact with both the detergent and the bicelle, as a dramatical resonance disturbance was observed, leading to the conclusion that ColIa strongly interacts with the detergent, being preferentially solubilized by the detergent. As for the CHAPSO bicelles, the chemical shift perturbation of ColIa resonances showed minimal resonance shifts for both cases, leading to the conclusion, that ColIa does not strongly interact with neither.

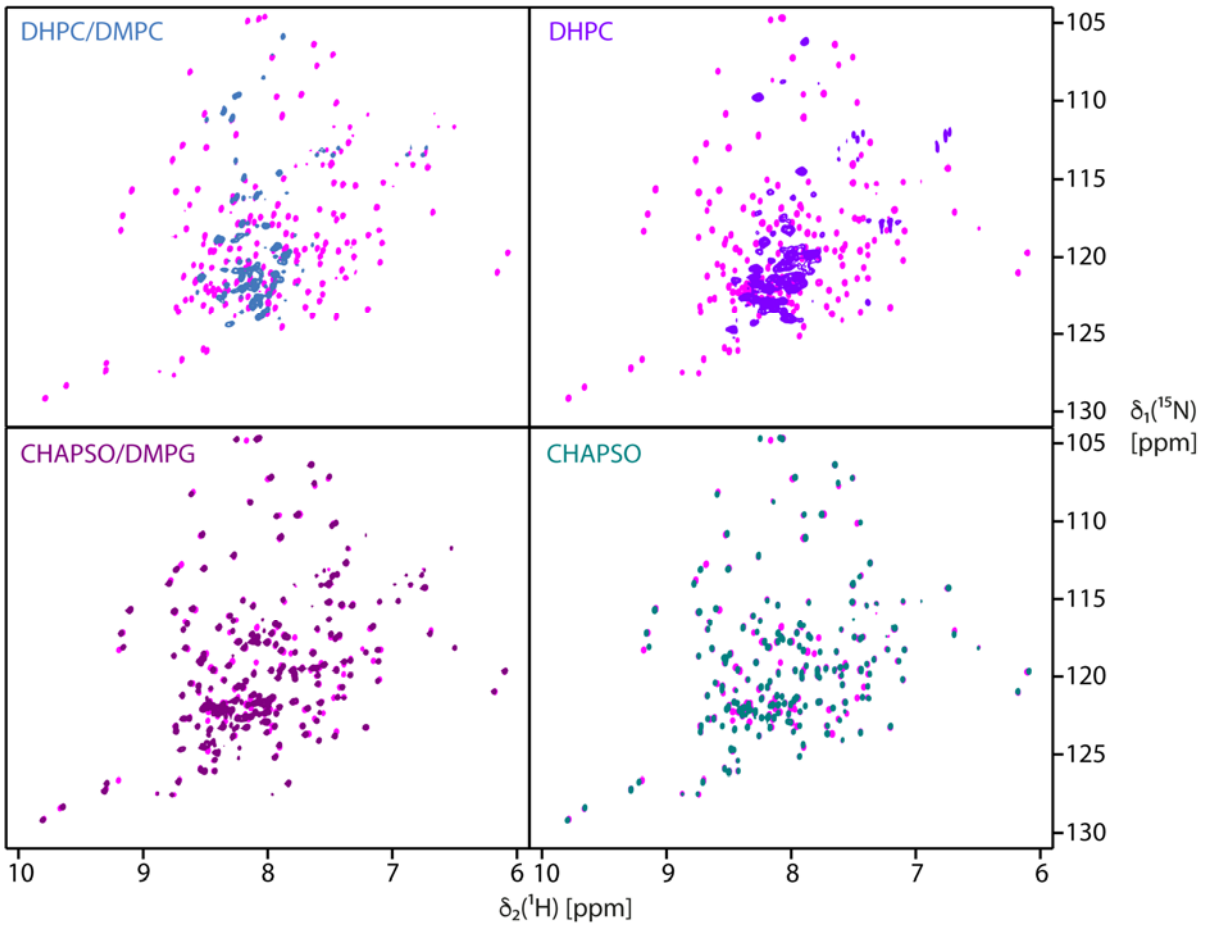


Figure 4-13 – 2D-[15N,1H]-TROSY-HSQC of ColIa (pink) in comparison with bicelles (DHPC/DMPC q=0.3; CHAPSO/DMPG q=0.3) and bicelles detergent only (DHPC and CHAPSO).

4.1.4. Discussion

As bicelles have shown to be a successful membrane mimetic for crystallization (19), crystallization trials using bicelles as membrane mimetic were set up. The ColIa C-domain crystallization trials resulted in the recrystallization of the soluble structure of ColIa, and as such, complete structure refinement was not further pursued.

The promising preliminary crystallization results prompted the procurement of the full backbone and sidechain assignment for the soluble form of ColIa, towards the solution study of bicelle interaction, through titration studies with different bicelle derivatives in search for the membrane-bound intermediate. Residue specific assignment of 90% of backbone residues was achieved for the ColIa C-domain, where the missing assignments were found to lie on the purported pore-forming hairpin helices 8 and 9. Limited H/D-exchange was experimentally ruled out as the cause for peak absence. Therefore, it could be concluded that the missing residues are in an intermediate exchange regime, which leads to peak broadening and consequent peak vanishing. This result indicates the existence of different conformations of the pore-forming hairpin in the soluble form. The secondary structure obtained from the NMR chemical shift assignments was found to match closely with the known crystal structure.

Using solution NMR spectroscopy titration studies, the initial success with the DHPC/DMPC bicelles was experimentally attributed to detergent solubilization. The alternative CHAPSO detergent was shown to only mildly interact with ColIa, and could in principle constitute a starting point for further extensive titration studies using different lipid, and pH combinations. Crystallization trials could be repeated in parallel with solution NMR, to search for a suitable combination of detergent/lipid composition. The lipid screen for a suitable membrane-mimetic could be investigated with other methods such as lipidic cubic phase and specifically AFM, which avoid detergents altogether.

4.2. The proapoptotic Bcl-2 protein Bax

4.2.1. Introduction

Bcl-2 homology family and its role in apoptosis

Apoptosis is a crucial step in the regulation of cellular fate. It is involved in different stages of the life cycle of an organism, such as development, tissue homeostasis, and senescence (20–22). Apoptosis is also involved in the progress of several diseases, like cancer and several autoimmune diseases (23, 24). Due to its ubiquity and importance, apoptosis is tightly regulated at different levels. One important family of proteins involved in apoptosis is the Bcl-2 (B-cell lymphoma 2) homology family. More than 20 members have been described so far, leading to a substantial complex family (20–22, 25, 26). The different members include four highly structurally conserved Bcl-2 homology (BH) domains, outlining both the structural and functional homology between the different members as depicted in Figure 4-14. The BH family can be divided into two main groups: antiapoptotic proteins, which inhibit apoptosis, such as Bcl-2 and B-cell lymphoma-extra large (Bcl-XL); and proapoptotic proteins, which promote it. The proapoptotic BH proteins are further subdivided in death effectors (pore-formers) such as the Bcl-2 homologous antagonist/killer (Bak) and the Bcl-2-associated X protein (Bax); and the BH3-only activators such as the BH3 interacting domain death agonist (Bid). Different group members are regulated by homo- and heterodimerization, which leads to the activation or inhibition of its partner (21, 27, 28). A delicate balance between the expression levels and cytoplasm availability for each group decides between homeostasis or cell death (21).

Bcl-2 homology family relation to protein toxins

The endosymbiotic theory explains the evolution of eukaryotic cells from prokaryote symbiosis. A prime example is the mitochondrium, which is derived from prokaryotic cells capable of oxidative phosphorylation. This is further emphasized by the fact that mitochondrial DNA is closely related to prokaryotic genomes (29). As discussed before, the Bcl-2 homology (BH) family shares a high structural homology between both the putative pore-forming domain of the colicin protein family and the translocation domain of diphtheria toxin (30, 31). The antiapoptotic members of the BH family form inhibiting complexes with the proapoptotic member, which closely resembles the inactivation mechanism of colicins C-Domain through complex formation with their co-expressing immunity conferring proteins. This has lead to the speculation that the BH family, and therefore apoptosis, has its roots in the prokaryote ancestor of mitochondria (32, 33). Although not technically toxins, the structure and function of this family suggest a close link with the α-PFTs.

Bax as a member of the Bcl2 homology family

Bax is a proapoptotic member of the Bcl-2 homology family (Figure 4-14), and it has been shown to trigger cytochrome c release from mitochondria, through pore formation, a crucial step within apoptosis (34). The transition from the monomeric inactive form to the active cytotoxic membrane-inserted form and its precise mechanisms has been named the ''holy grail'' of apoptosis research (26). However, the description of this crucial step has so far been elusive, due to the lack of atomic resolution information on the intermediate steps.

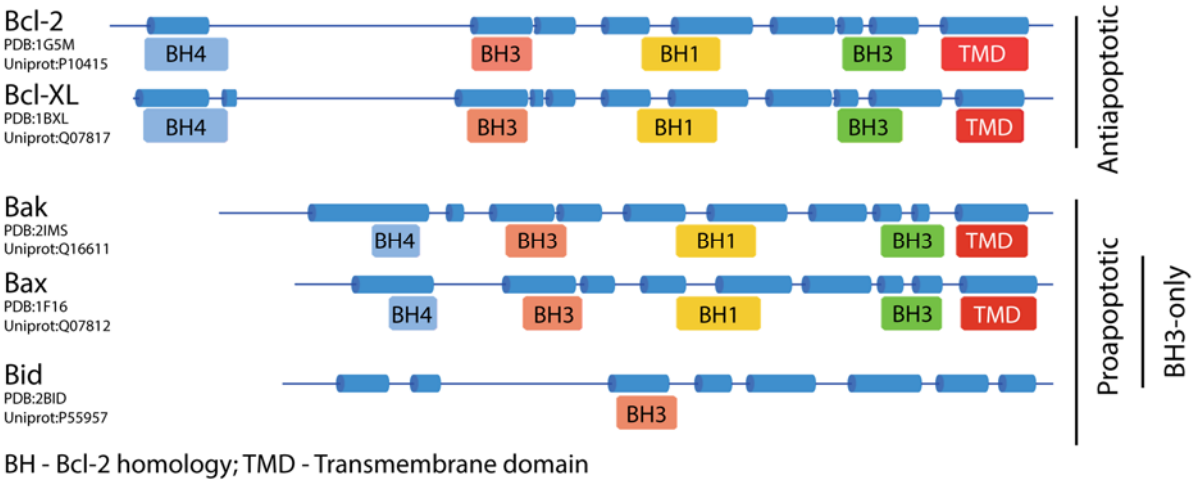


Figure 4-14 – Scheme comparing the secondary structure and domains of some of the different BH family members where structural data is available. Data was obtained both from PDB and Uniprot.

Bax soluble conformation

Bax is a 21 kDa α-helical protein bundle consisting of 9 helices (Figure 1-2 B and Figure 4-15). Occluded from the exterior sits a hydrophobic hairpin formed by helices 5 and 6 in the center of the protein, surrounded by several amphipathic helices (Figure 1-2 B). The C-terminal helix 9 sits on a hydrophobic cleft-pocket formed by helices 2, and 3, termed BH3 domain.

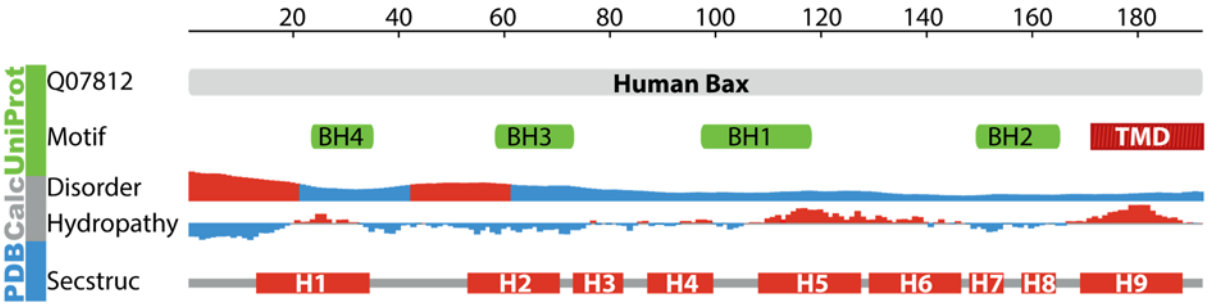


Figure 4-15 – Bax protein feature view of PDB entries mapped to UniProtKB Q07812 sequence.

Physicochemistry of Bax

Many bacterial toxins undergo a pH-dependent conformational change from the soluble to their membrane-inserted form, whereas the translocation domain of diphtheria toxin and the colicin family

of pore-forming toxins are the best-understood examples (35–37). Similar findings were observed for several members of the BH family such as Bcl-XL (*38*), Bcl-2 and Bax (39–41). These observations indicate a pH-dependent activation mechanism, leading to the exposure of the membrane-targeting domain of Bax and subsequent membrane insertion. Nevertheless, recent experimental work by cell-free assays (42), EPR (43) and *in vitro* fluorescence kinetics (44) using liposomes has demonstrated the pH-dependent membrane association and pore formation in a broader pH range.

The exposure of Bax at 43°C to isolated mitochondria leads to the release of cytochrome c, while exposure to either antiapoptotic Bcl-XL or cytosolic extract abrogates this effect (45). In addition to its pH stability, Bax shows remarkable heat stability (46, 47). These observations suggest an alternative activating mechanism for Bax, decoupled from exogenous activators, based on an intermonomeric interaction triggered by heat. Heat-shock proteins are known to stabilize sensible proteins to avoid cellular stress. Conversely, Bax could have an opposing role, leading under extreme conditions to its activation and subsequent cell death. Other indications to the activity of Bax are found in its resistance to chaotrops (46).

The high stability of Bax under extreme conditions suggest the possibility that direct physical or chemical agents *in vivo* can regulate Bax activation. Both temperature and pH-induced conformational changes, and its relation to membrane insertion might be either ancient evolutionary rendered redundant forms of activation/regulation or actual alternative apoptosis regulation pathways, related to cell stress, that further fine tune the precise regulation of apoptosis.

Conformational changes of Bax

The structural similar BH family members (Figure 4-14) can be subdivided into groups according to their activity. These two groups have opposing roles in apoptosis, either pro or anti. The fact that some members have at least two distinct structural forms (the known cytosolic soluble form and the still elusive membrane-inserted form) and probably other intermediate conformations leads to a highly structurally and functionally complex family.

Bax was initially discovered through co-immune-precipitation experiments with Bcl-2, a known antiapoptotic protein at the time (21). The overexpression of Bax was found to counteract the death repressor activity of Bcl-2, leading directly to cell death. Homodimerization and heterodimerization between Bcl-2 members were found to regulate available free monomer, in order to facilitate or hinder membrane interaction (41).

The most significant advancement in the field since the description of the soluble structures of different BH family members (20–22) was the description of the mechanisms that govern Bax

activation and consequent conformation changes mainly by Gavathiotis and coworkers (48, 49). Comparing monomeric and ligand-activated Bax or a detergent mimic thereof, using circular dichroism, thermofluorescence, and protease digestion (46) as well as NMR spectroscopy to analyze ligand based conformational changes (49, 48) a series of events was made clear. First, and in excellent agreement with previous data, where Bax N-terminal region had been shown to be displaced from its original position to a protease available form (46) exposing an essential hallmark of Bax activation by epitope mapping (47, 50, 51); through ligand titration chemical-shift perturbation mapping, the initially buried helix 1, 2 loop, was found to be the displaced region (48). This displacement was concluded to be the first event in the BH3-only direct activation of Bax. The Bcl-2 member conserved BH3 domain (helix 2) implicated in both Bax homodimerization, and interaction with BH3-only activator proteins was thought to be responsible for activation propagation (46, 49, 52). This hypothesis was tested with the titration of ^{15}N-Bax with a BH3-only peptide mimic, where chemical shift mapping showed the binding *locus* to be at the helix 1, 6 interface, in line with helix 1-2 loop displacement (49). The displacement leads to the exposure of Bax BH3 domain, and further activation propagation to other monomers, and constitutes the second step in Bax-ligand activation. Finally, in the third activation step, the C-terminal helix 9 is displaced from a core-protected helix within the monomeric state to a loose helix in the activated form (46, 49, 52). This hydrophobic pocket closely resembles the BH3 antiapoptotic pocket of Bcl-2 and Bcl-XL (44, 52). Its hydrophobic nature allows the stabilization of several hydrophobic residues of amphiphilic helix 9, increasing Bax solubility through the exposure of helix 9 hydrophilic side (52). These three sequential steps outline the fundamental interaction between BH3-only activators and Bax, which lead to its activation. Over decades of extensive research, a clear set of events has become clear: the ligand-activation of Bax sets in motion a chain of events, which results in a conformational reorganization leading to membrane permeation and ultimately cell death.

Pore activity evidence

One of the most significant issues studying Bax conformational changes is the difficulty in acquiring structural information other than the inactive monomer. Different electrophysiological studies showed Bax to cause destabilization of the membrane and to form different sized ion-conducting channels (39, 40, 44, 53–56). The overall trend in this studies is that two types of channels are observed: at first an initial small channel is formed producing simple current recordings, later larger channels appear with higher currents and more complex patterns. A general observation is that low pH enhanced pore formation (38–40). Both observations suggest that Bax might form different oligomeric structures in the membrane.

Two recent publications managed to image Bax in the membrane-embedded form by AFM, showing Bax pores with diameters ranging from 10–20 to 100–300nm (43, 57). Several studies managed to image liposome (43, 58, 59) and nanodisc (60) pores in the presence of Bax, using cryo-EM. In one instance, large solitary pores between 50–90nm in size were shown to increase with time and Bax concentration (59). Fluorescence microscopy studies show the localized release of fluorescein from liposomes upon Bax membrane insertion (44, 61). In both cases, the kinetics of fluorescence loss, indicate a concentration-dependent 10–40nm pore.

The wide range of results from different techniques indicates a heterogeneously shaped pore. Depending on conditions such as pH, protein concentration, and the presence of BH3-only members, Bax seems to form pores with different sizes and properties. The precise clarification of the conformation or conformational steps of Bax in the membrane is vital for the understanding of the role of Bax and the Bcl-2 family in apoptosis.

Bax and the Bcl-2 family have shown to be a remarkable topic of research and discussion within the apoptosis field. Which is in part due to its importance in human biology and therefore pharmacological industry, but also on its intrinsical complexity. Clarifying the role of each Bcl-2 family member in Bax oligomerization, pore formation, and the eventual cascade of effects it leads to, is crucial for the elucidation of the complete picture. This is essential in order to understand the underlying mechanisms of mitochondrial apoptosis and clarify this delicate and intricate puzzle.

Research Article

Hugo Aragão
Joeri Verasdonck
Erich Michel
Beat Meier
and Sebastian Hiller

Sample preparation and initial characterization of the membrane-inserted apoptosis Bax pore by high-resolution solid-state NMR spectroscopy

In preparation

4.2.2. Sample preparation and initial characterization of the membrane-inserted apoptosis Bax pore by high-resolution solid-state NMR spectroscopy

Abstract

The Bcl-2 family of proteins has a central role in the regulation of apoptosis. A mitochondrial membrane-bound form of the Bcl-2 protein Bax leads to pore formation and apoptosis. The structural and conformational changes Bax experiences towards its membrane-inserted form are currently not understood at the atomic level. Here, we show that optimized cell-free expression protocols in combination with solution and solid-state NMR experiments allow monitoring these states at atomic resolution. Using bicelles as a size-constrained membrane mimetic provides access to a trapped, monomeric intermediate step of membrane-associated Bax. Fully formed Bax pores in liposomes, corresponding to the apoptosis-active membrane-inserted form of Bax, are studied by solid-state NMR spectroscopy. In the membrane-inserted state, the C-terminal helices 7–9 contain dynamic residues, and at least one isoleucine residue is shown to be mobile. Our approach provides an experimental route to study the pore-state of Bax at atomic-resolution, as well as intermediate states of the insertion mechanism.

Statement of contribution

In the experimental part, I was responsible for cloning, cell-free system set up, testing and expression of Bax protein, solution NMR data acquisition, and NMR data analysis. As for the manuscript, I was responsible for writing as well as preparing all the figures and tables present in the manuscript.

Introduction

Apoptosis is an essential function for the regulation of individual cellular fate within an organism. It is involved in several stages of life: including development, tissue homeostasis, and senescence (62–64). Also, it is crucially involved in the progress of certain diseases, such as cancer and autoimmune disorders (65, 66). Due to its prominence, apoptotic progress is tightly regulated at several levels, in particular by the Bcl-2 homology (BH) family of proteins, which comprises more than twenty members (20-22, 25, 26). The BH family consists of anti-apoptotic proteins, which inhibit apoptosis, and of pro-apoptotic proteins, which promote its progress. BH proteins can form homo- and heterodimers in different combinations, leading to mutual activation or inhibition (21, 27, 28). The pro-apoptotic BH proteins are further subdivided in death effectors (pore-forming proteins), and activators (BH3-only proteins).

Within the pro-apoptotic group, Bax plays a key role as the final executor of cell death. Within the cell, Bax exists in equilibrium between a cytosolic soluble inactive form and a mitochondrial membrane-inserted active form. Its insertion into the mitochondrial membrane triggers cytochrome c release, one of the hallmarks of apoptosis (34). In its soluble form, Bax forms a compact α-helical bundle, sharing remarkable structural similarities with bacterial pore-forming toxins (Figure 4-16) (30, 31, 34, 52).

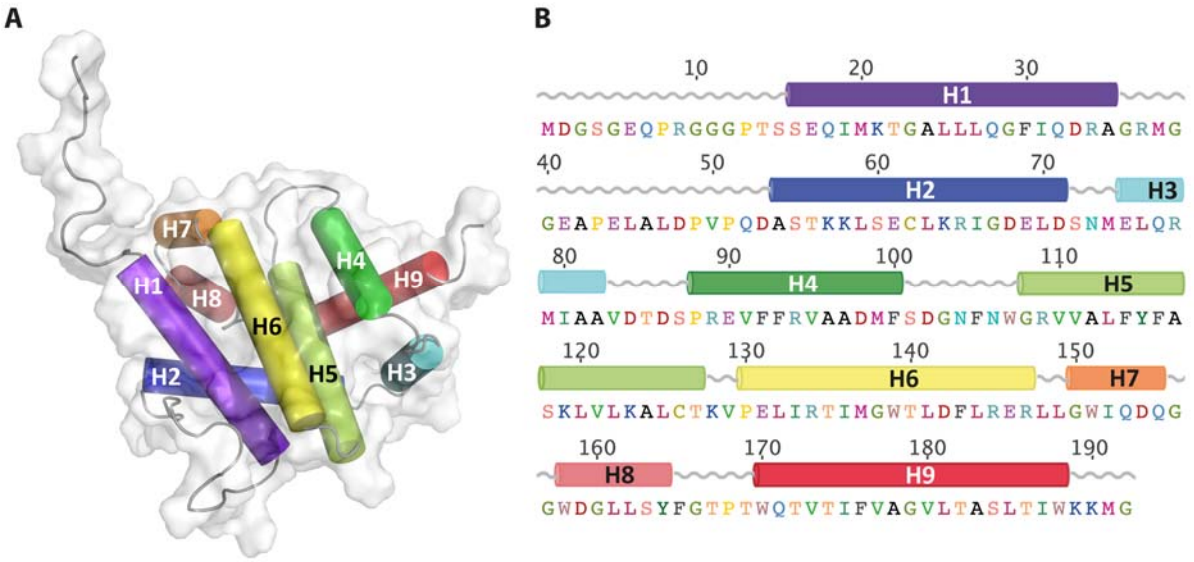


Figure 4-16 – A- Structure of Bax in aqueous solution. The numbered helices are shown in cylinder representation and different colors, the surface is represented in light gray [PDB 1F16 (52)]. B- Amino-acid sequence of full-length Bax, with drawn secondary structure elements (PDB 1F16).

Enclosed in its core are the two hydrophobic helices 5 and 6, the putative pore-forming hairpin (67). Additionally, the C-terminal transmembrane helix 9 is trapped in the hydrophobic BH3 groove formed by helices 2, 3, 4 and 5. Its release and exposure constitute the priming event for oligomerization and membrane interaction. Despite the elucidation of the mechanisms that lead to the exposure of helix 9 (48, 49, 68), structural details of the subsequent events that lead to Bax pore-formation in the mitochondrial outer membrane (MOM) are yet unclear. A structural model based on EPR constraints has provided substantial information on the helix topology (69), but atomic-resolution information, which would be necessary for a full description of the insertion mechanism, is still lacking. Importantly, the membrane-inserted, active Bax pore does not possess a single, well-defined oligomeric structure, but a variable stoichiometry, making its crystallization particularly challenging. Solid-state NMR spectroscopy does not require the formation of repetitive crystal lattices and thus has the potential to provide an atomic resolution structure of the membrane-inserted form of Bax. In addition, solution NMR spectroscopy can provide structural information of trapped intermediate states.

The use of NMR spectroscopy requires protein samples that are labeled with stable isotopes. Bax expression yields in *E.coli* cell culture are low, due to membrane self-insertion and consequent cytotoxicity, and as a result, multiple liters of cell culture are required for a single isotope-labeled protein sample, making extensive studies overly expensive (52). In contrast, cell-free systems often circumvent the toxicity problem due to lack of a functional metabolism and cell-delimiting membranes. Cell-free expression systems have been successfully used to express challenging protein targets such as toxic and transmembrane proteins in milligram amounts for structural biology (70, 71). Also, their inherent flexibility and open nature allow manipulation of the reaction conditions and co-factor supplementation, such as to enhance and stabilize protein targets for high throughput production.

In this work, we describe an optimized cell-free protein expression protocol for the production of an isotopically labeled sample of Bax in milligram quantities, suitable and sufficient for solution and solid-state NMR experiments. Solution NMR spectra of Bax attached to bicelles, and solid-state NMR spectra of Bax proteoliposomes were recorded. The data provide concrete insights into the conformational changes Bax experiences upon membrane binding and insertion in the presence of different membrane mimetics. Our approach opens a new experimental route to study the insertion and pore formation mechanism of Bax at the atomic level.

Materials and Methods

Cell-free extract preparation

E.coli S30 cell extract preparation was prepared following published protocols (72). In short, harvested A19 cells were washed three times with S30 buffer A (10 mM Tris-acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 6 mM β-ME). After resuspension in S30-buffer B (10 mM Tris-acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 1 mM DTT, 0.1mM PMSF), cells were lysed using a French press and centrifuged twice at 30'000×*g*, 4°C. To remove endogenous mRNA, the extract was incubated for 45 min at 42°C and then dialyzed at 4°C overnight against S30-buffer C (10 mM Tris-acetate pH 8.2, 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 0.5 mM DTT) with a 12–14 kDa MWCO. To remove the remaining cellular debris, the cell extract was centrifuged at 30'000×*g* at 4°C, aliquoted, flash frozen in liquid nitrogen and stored at –80°C.

Protein expression and purification

A pET15b-based expression plasmid containing full-length Bax (FL-Bax) with N-terminal His_6-tag and thrombin cleavage site was obtained from Gerhard Wagner (Harvard Medical School). The N-terminal His_6-GB1-tag containing construct was obtained by inserting the Bax sequence into pCFX3 (73) using the XhoI and the NotI restriction sites. The RF cloning protocol (74) was used to replace the N-terminal GB1 tag with the His_{10}-SUMO tag into pCFX3, using the 5'- and 3'-oligonucleotide primers 5'- ATA CCA TGT CTG GTT CTC ATC ATC ATC ATC ATC ATC ATC ATC ATG GAG GTT CTG ACT CCG AAG -3' and 5'-GCT GCT CCC CGG ACC CGT CCA TGC CAC CAA TCT GTT CGA GG -3'.

For analytical scale, 55µl reactions were prepared in home-built mini-continuous exchange cell-free (CECF) reactors used to optimize expressing conditions. For preparative-scale protein production, a home-built maxi-CECF reactor (72) was used with a 0.5–3.0 ml Slide-A-Lyzer dialysis cassette with a 10 kDa MWCO (Pierce, Rockford, IL, USA) and a feeding/reaction mixture volume ratio of 14.5. Final concentrations of critical components such as PEG8000 (2%), Mg^{2+} (16 mM) and T7-RNA polymerase (9.5 mg/ml of reaction mixture) were carefully optimized for maximal protein expression. Both [*U*-^{15}N], and [*U*-^{13}C,^{15}N] amino acid labeling was achieved with a cell-free twenty amino acid mix (Cambridge Isotope Labs) at a final concentration of 0.8 mM in the reaction mixture, and 0.4 mM in the feeding chamber. Protein expression was left overnight in a thermoshaker at 120 rpm, 30°C in a VWR Incubating Mini Shaker.

The reaction mixture was centrifuged (10 min at 10'000 rpm, Eppendorf Centrifuge 5810 R) and the supernatant was diluted 1:2 with buffer A (25 mM HEPES, 500 mM NaCl, 20 mM imidazole, 1 mM

TCEP, pH 7.5) and then incubated with Ni-NTA beads (Thermo Fisher Scientific), pre-equilibrated with buffer A, under gentle agitation for 30 min. Afterward, the flow-through was collected, the beads were washed with 10% buffer B (25 mM HEPES, 500 mM NaCl, 500 mM imidazole, 1 mM TCEP, pH 7.5) and then eluted with buffer B. The elution was buffer exchanged with a PD-10 column (GE Healthcare, Piscataway, NJ). The buffer-exchanged eluate was incubated with His_6-ULP1 SUMO protease at a molar ratio of 50:1 for 30 min, and then with Ni-NTA beads, under gentle agitation for another 30 min. Afterward, the flow-through was collected, the beads were washed with 10% buffer B and then buffer B. The protein was concentrated and buffer exchanged by ultrafiltration (Vivaspin 10-kDa-cutoff concentrator, Vivascience) with NMR buffer (25 mM HEPES, 50 mM NaCl, pH 7.0). The purity of the protein was assessed by SDS-PAGE.

Bicelle preparation and titration

DMPG/CHAPSO bicelles with a molar lipid/detergent ratio of *q*=0.3 were prepared at 10 or 40% (w/w) stock concentration in water. The detergent was solubilized in water with slow agitation for 30min. The detergent solution was then added to the lipid

and incubated for solubilization for 1h. The 40% stock was agitated overnight at 4°C. For the bicelle titration, these were added to FL-Bax in defined steps (0.3, 0.5 and 1% (w/v)), resulting in final molar ratios of 1:0.52, 1:0.86 and 1:1.72, respectively.

Liposome preparation

E. coli polar lipid (ECPL) extract in chloroform solution (Avanti Polar Lipids) was evaporated under a stream of nitrogen gas with gentle agitation to form a lipid film and left overnight exposed to vacuum in a desiccator to completely dry. The resulting film was re-suspended in buffer (50 mM Tris-HCl, 100 mM NaCl, pH 7.5) under agitation, to a final concentration of 12.5 mg/ml. The lipids were subjected to five freeze-thaw cycles and vortexed in-between. Liposomes were prepared by no less than 10 passes in a home-built pressure extruder using 100nm polycarbonate filters (Whatman).

Ni-NTA-Gold negative staining TEM

Bax protein was incubated with liposomes at a 1:6 w/w ratio of protein:lipid (1:160 molar ratio), for 2 h at 30°C with agitation. After proteoliposome assembly, 3 μl of the solution was adsorbed to 200-mesh carbon-coated copper grids rendered hydrophilic by glow-discharging in air for 20s. After blotting the excess solution, the grids were stained with a 1:5 dilution of Ni-NTA-Nanogold, rinsed in double-distilled water and stained with 2% uranyl acetate. Pictures were taken on a Philips CM10 equipped with a LaB_6 filament and operated at an accelerating voltage of 80 kV.

EM and image processing

For cryo-EM, crystal solution on glow-discharged carbon-coated holey carbon grids (Quantifoil R2/2, Quantifoil Micro Tools, Jena, Germany) was blotted and rapidly plunge frozen in liquid-nitrogen-cooled liquid ethane, using a MarkII Vitrobot (FEI, Eindhoven, Netherlands). The frozen grids were transferred to a Gatan-626 cryo-holder and analyzed in a Philips CM200 transmission electron microscope, equipped with a field-emission gun and operated at 200 kV. Pictures were taken at a nominal magnification of 50,000× using low-dose imaging techniques with an electron dose of approximately 5 e−/Å2 and defocus values ranging from 0.5 to 0.9 µm. Images were recorded on Kodak SO-163 film, which was developed for 7 min in full-strength Kodak D19 developer solution. Image quality was assessed by optical diffraction on a home-built laser diffractometer. The best images were digitized using a Heidelberg Primescan D 7100 scanner with a step size of 1 Å/pixel at the specimen level. Digital images were processed using the 2dx software suite (75,76), which is based on the MRC programs (77). Images were corrected for crystal disorders by three rounds of unbending. This was followed by astigmatism and contrast transfer function correction. Symmetry was determined using the allspace program (78) within 2dx.

Proteoliposome preparation and pulldown

Liposomes were incubated with purified protein at different protein to lipid ratios for 1h at 30°C, and ultra-centrifuged at 150'000×*g* for 1 h in PC tubes (TLA 120.1 rotor, in an OPTIMA Max-XP centrifuge, Beckman Coulter, Fullerton, CA). The supernatant was then pipetted out, and the pellet was re-suspended in liposome buffer. Both supernatant and pellet were analyzed by SDS-PAGE.

Solution NMR spectroscopy

2D [^{15}N,^{1}H]-TROSY-HSQC (13) NMR spectra were recorded at 25°C on a Bruker Ascend 700 spectrometer equipped with a cryogenic triple-resonance probe.

Solid-state NMR sample preparation

Proteoliposomes were prepared by mixing ECPL liposomes with 1:3 w/w ratio of protein:lipid (1:80 molar ratio) and agitated at 30°C overnight. The proteoliposomes were ultracentrifuged overnight at 150,000×*g* directly into a 1.9mm rotor (Bruker), using a home-built tool (79) in a Beckman L90-K with a swinging-bucket rotor (SW40 Ti, Beckman Coulter, Fullerton, CA). Solid-state NMR spectra were recorded on a Bruker Avance III spectrometer at a 20 T static field at 20°C. Several 2D correlation experiments were recorded; 2D [^{1}H-^{13}C]-INEPT, 2D [^{1}H-^{15}N]-HSQC, and 2D [^{13}C-^{13}C]-TOBSY (32–37, Table 4-1). All spectra were recorded at 35 kHz MAS, except for the 20 ms 2D DARR spectrum which

was recorded with 18 kHz MAS frequency. NMR data were processed with Topspin (Bruker) and analyzed with CcpNmr (86).

Table 4-1 – Experimental parameters for the solid-state NMR spectra.

Experiment	HSQC	INEPT	INEPT-TOBSY
MAS frequency [kHz]	35	35	25
Transfer 1	HN ref-INEPT	HC ref-INEPT	HC ref-INEPT
Field [kHz]-^{1}H	100	100	100
Field [kHz]-^{13}C	-	62.5	62.5
Field [kHz]-^{15}N	50	-	-
Carrier [ppm]	-	-	-
Time [ms]	10.4	5.0	5.0
Transfer 2	NH ref-INEPT	-	TOBSY C9$\frac{1}{3}$
Field [kHz]-^{1}H	100	-	-
Field [kHz]-^{13}C	50	-	75
Carrier [ppm]	-	-	57
Time [ms]	10.4	-	8.0
t^1 Increments	2048	512	2048
Sweep width (t_1) [kHz]	12	20	67
Max acq. time (t_1) [ms]	85.0	12.8	15.4
t_2 Increments	4096	3072	3072
Sweep width (t_2) [kHz]	20	85	85
Max acq. time (t_2) [ms]	102.4	18.0	18.0
^{1}H Decoupling [kHz]	3.0	90	1.1
Interscan delay [s]	1.33	2.2	2.2
Number of scans	64	64	32
Measurement time [h]	48	20	40

Results

Optimization of Bax expression by cell-free synthesis

Cell-free expression systems lack a cellular membrane, making them a method of choice for pore-forming toxin expression. For solid-state NMR spectroscopy, protein amounts in the mg range are required, and we have thus optimized a continuous-exchange cell-free system to produce isotope-labelled Bax samples. In a first step, to optimize expression and solubility conditions, three different Bax constructs were designed: Bax with N-terminal His_6-tag in a pET15b vector (pET15b-Bax), Bax with N-terminal His_6-GB1 solubility tag in a pCFX3 vector (73) (pCFX3-GB1Bax) and Bax with N-terminal His_{10}-SUMO tag in a pCFX3 vector (pCFX3-SUMOBax). For each of the constructs, the macromolecular crowding concentration was optimized for maximum yield (Fig. 4-17).

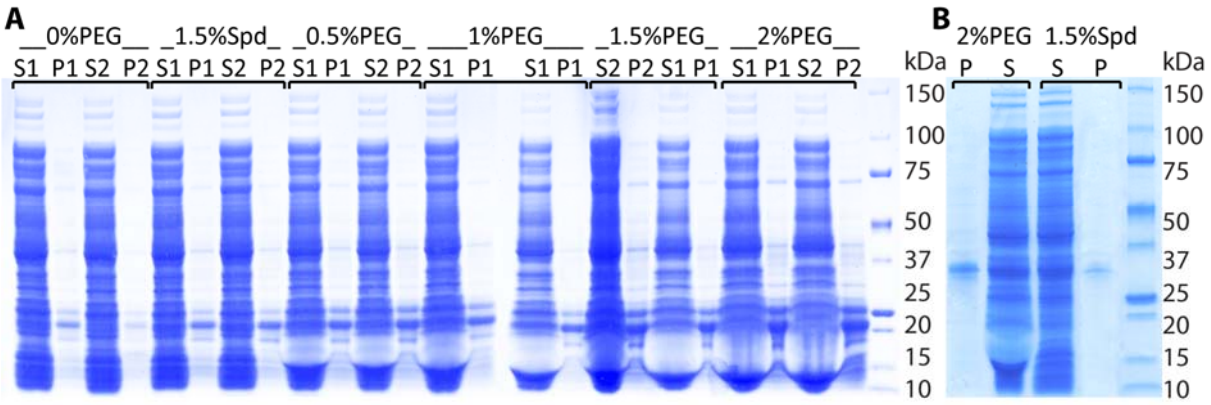


Figure 4-17 – Expression optimization of Bax. A- Comparison of solubility and expression optimization of pET15b-Bax with 1.5% spermidin and increasing percentages of PEG (0-2%) showing both the insoluble (P) and soluble (S) fractions in duplicate; B- Comparison of pCFX3-GB1Bax solubility, while expressed in the presence of different crowding agents (2% PEG and 1.5% Spermidin) of both insoluble and soluble fractions.

Comparison between the different percentages of the two crowding agents Spermidin and PEG8000, led to the conclusion that 1.5–2% PEG8000 gave consistently the highest protein yields. Expression of full-length Bax without solubility tag (pET15b-Bax) resulted in mostly insoluble protein (Figure 4-18 A). When fused to the GB1 solubility tag (pCFX3-GB1Bax), total expression of Bax increased, but around half of the total protein remained insoluble (Figure 4-18 A). The SUMO tag fusion further increased the fraction of soluble Bax (Figure 4-18 A). Expression temperature was varied from 25–30 °C, which showed no noticeable yield improvement (data not shown). After expression optimization, the expression yields of Bax was ~1mg per ml of cell-free reaction volume. During protein purification and concentration, some loss of protein was observed after SUMO tag cleavage (Figure 4-18 B). Overall, a total yield of purified protein of ~0.33mg per ml of reaction volume was obtained. Importantly, these yields are sufficiently high to allow routine expression of isotope-labeled Bax samples.

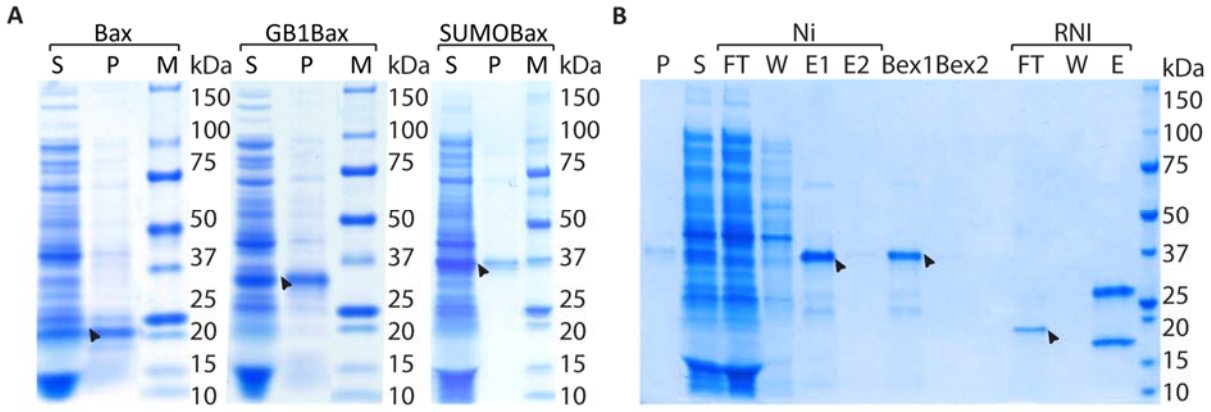


Figure 4-18 – Bax expression and purification. A - Solubility of different Bax constructs (pET15b-Bax; pCFX3-GB1Bax; pCFX3-SUMOBax) comparing fractions in the supernatant (S) and precipitate (P). B- Ni-NTA and reverse Ni-NTA of 10-His-SUMO-Bax; Arrows indicate Bax in the first Ni elution 1 (E1), buffer exchange (Bex1) and reverse Ni (RNI) flow through (R.FT).

Cell-free produced Bax is functional

To assess the functionality of our cell-free expressed Bax preparations, liposomes before and after incubation with Bax were imaged using negative-stain transmission electron microscopy. Empty liposomes appeared flattened upon blotting, but overall their shape seemed intact (Figure 4-19 A).

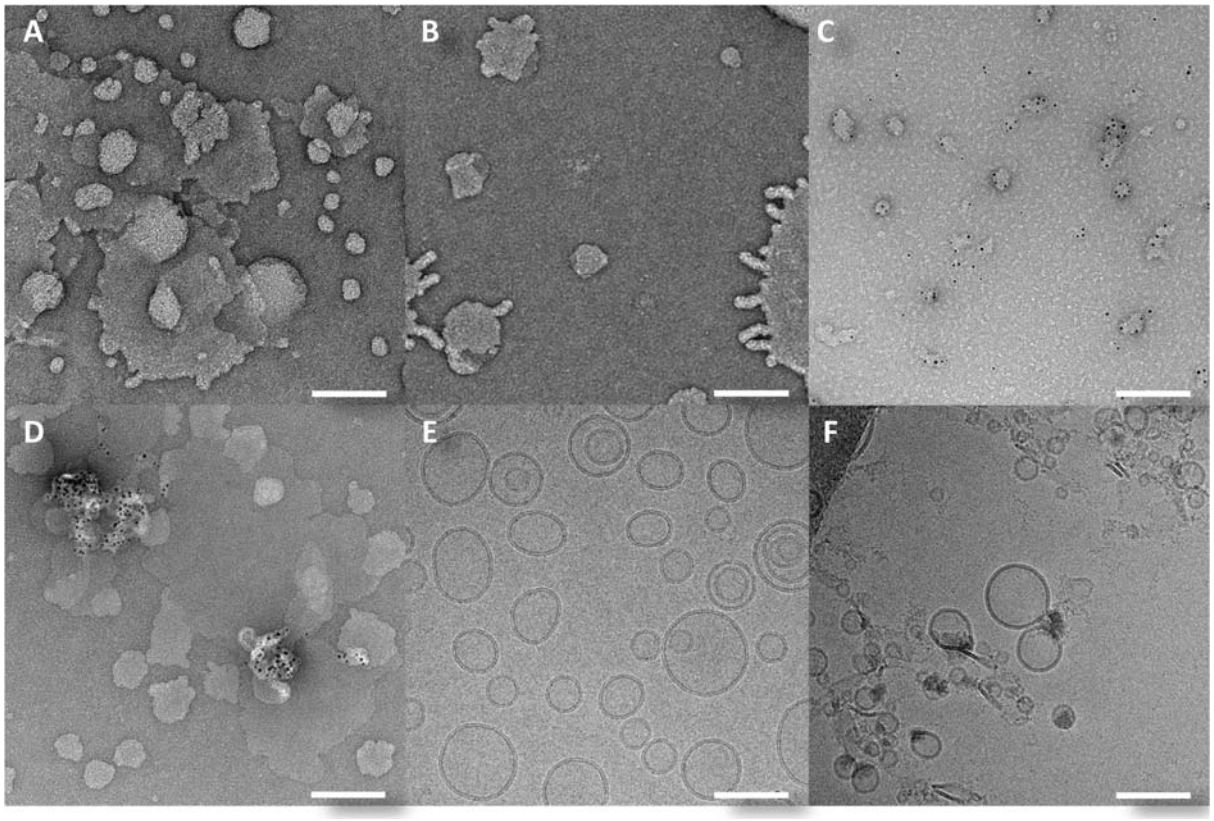


Figure 4-19 – EM images of Bax proteoliposome preparations and empty liposome controls. A- Negative staining EM of empty ECPL liposomes; B- Negative staining EM of empty ECPL liposomes with Ni-NTA-Au staining (negative control); C- Bax in 0.025% DDM Ni-NTA-Au staining (positive control); D- Bax proteoliposome preparation with Ni-NTA-Au staining; E- Cryo-EM image of empty liposomes; F- Cryo-EM image of Bax proteoliposomes. Scale bars: 100 nm.

Bax was localized using Ni-NTA-gold nanoparticles that bind the N-terminal His_6-tag of FL-Bax. The negative control showed no nanoparticle background signal (Figure 4-19 B). Staining of Bax proteoliposomes showed co-localization of multiple Bax molecules at the membrane interface (Figure 4-19 D). When imaged by cryo-EM, empty liposomes showed well-defined unilamellar shape (Figure 4-19 E). In remarkable contrast, Bax proteoliposomes featured mostly broken liposomes and in some instances a Bax pore-like structure spurring a clear rupture of the membrane curvature (Figure 4-19 F). Overall, these data show that our cell-free produced Bax is functional, as it inserts into liposomes forming large pores, in agreement with published data from other preparations (43, 59).

Characterization of a Bax membrane-bound intermediate state by solution NMR spectroscopy

Bicelles were chosen as membrane mimetic to characterize the early stage of Bax association into lipid bilayers. Due to their limited size, bicelle-bound Bax cannot form large oligomers, and this state may thus correspond to an initial stage of membrane insertion. DMPG/CHAPSO bicelles with *q*=0.3, which have a diameter of ~8nm (42) were titrated to Bax and the conformational change monitored by 2D [^{15}N,^{1}H]-TROSY NMR experiments (Figure 4-20).

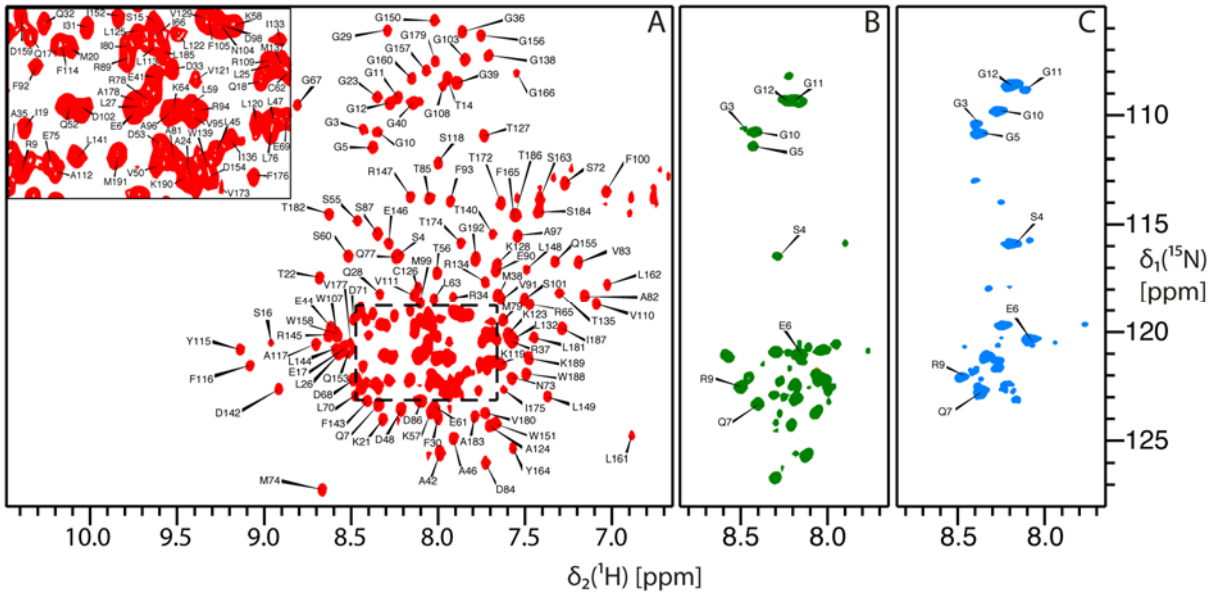


Figure 4-20 – Comparison of free Bax spectra with Bax in bicelles and liposomes calibrated with DSS. A- 2D [^{15}N,^{1}H]-TROSY-HSQC of free Bax in 25mM HEPES pH7, 50mM NaCl. B- 2D [^{15}N,^{1}H]-TROSY-HSQC of 50uM [*U*-^{15}N]-labeled Bax in CHAPSO/DMPG (q=0.3) bicelles at a 1:1.72 molar ratio showing the transferable assigned resonances of Bax in the membrane-mimetic; C- solid-state NMR 2D [^{15}N,^{1}H]-HSQC of [*U*-^{13}C,^{15}N]-labeled Bax in liposomes (Table 2).

Upon bicelle titration most of the resonances broadened beyond detection, remaining around 30 peaks above the noise level (Figure 4-20 B). This reduction of most signals shows that Bax, which was fully protonated in the present samples, associates strongly to the bicelles, leading to slow molecular tumbling and thus reduced signal intensities. A comparison with the known assignment showed that among the remaining signals, the N-terminal segment of residues 1–12 remains observable in the bicelle-attached form. Thus, this segment remains locally flexible relative to the membrane-attached form. It is also further evidenced that Bax can associate to lipid bilayers in the absence of BH3-activators, which is in agreement with earlier observations that showed Bax activation solely upon incubation at high temperature (43).

Preparation of samples for solid-state NMR spectroscopy

The sensitivity in solid-state NMR experiments is proportional to the total amount of labeled protein in the active volume of the rotor. For a given rotor, the amount of total protein should thus be maximized. Hence, the protein/lipid ratio of the proteoliposomes needs to be optimized. To obtain the maximum protein rotor packing, an ultracentrifugation liposome pull-down with decreasing mass ratios of protein:lipid (from 1:1 to 1:3) was tested. At a 1:3 mass ratio, close to all of the protein associated with the liposomes (Figure 4-21).

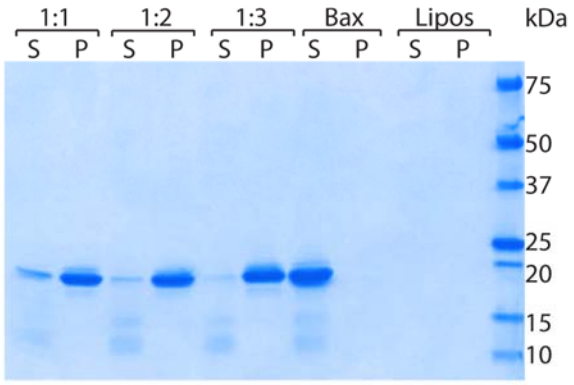


Figure 4-21 – Bax liposome ultracentrifuge pulldown experiment to test the maximum protein to lipid mass ratio; Liposome bound Bax (P), and soluble Bax (S) are shown for different ratios and both free Bax and liposome controls. Optimal w/w ratio is shown to be 1:3, where maximum protein appears in the P form.

Soluble Bax control shows no protein in the insoluble fraction. Note that Bax liposome insertion is Bax specific, as none of the impurities present in the soluble fraction are found on the pellet. Proteoliposomes with a 1:3 ratio were thus used for the solid-state experiments.

Characterization of the Bax membrane-inserted state by solid-state NMR spectroscopy

Uniformly [*U*-^{13}C,^{15}N]-labeled Bax in proteoliposomes was characterized by solid-state experiments that selectively either allow magnetization transfer in locally flexible (INEPT) or in rigid parts (CP) of the protein assembly. Both the 1D ^{13}C solid-state NMR INEPT-based and CP-based spectra of Bax proteoliposomes show sharp signals with narrow line widths (Figure 4-22). Since both spectra contain signals, membrane-inserted Bax features both rigid and flexibly disordered regions with very different relative amino acid composition. The solid-state NMR 2D [^{15}N,^{1}H]-HSQC (Figure 4-20 C) is an INEPT-based experiment, showing the highly flexible residues of liposome-inserted Bax. Around 20–30 resonances are observed, among these several peaks correspond to the N-terminal resonances shown before to remain mobile upon membrane interaction. Overall the spectrum shows a highly similar resonance fingerprint in comparison with the bicelle-bound 2D-TROSY-HSQC (Figure 4-20 B). The

flexible parts are thus similar, suggesting that common structural motives exist in the bicelle- and liposome-bound states and that their comparison might be functionally relevant.

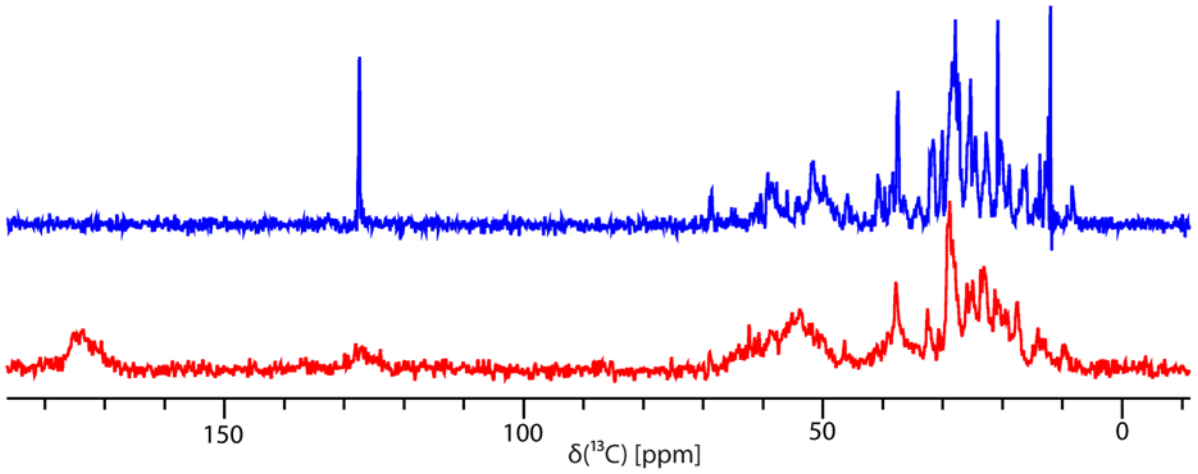


Figure 4-22 –1D ^{13}C ssNMR spectra of Bax in liposomes. Blue is an INEPT-based (1024 scans), red a CP-based spectrum (128 scans), showing the flexible and rigid parts of the protein, respectively. The spectra are scaled to the noise intensity.

Looking at the flexible residues with the help of the 2D [^{13}C,^{13}C]-INEPT-TOBSY and 2D [^{1}H,^{13}C]-INEPT (Figure 4-23 and Figure 4-24), a set of carbon resonance spin systems and cross-peaks are visible.

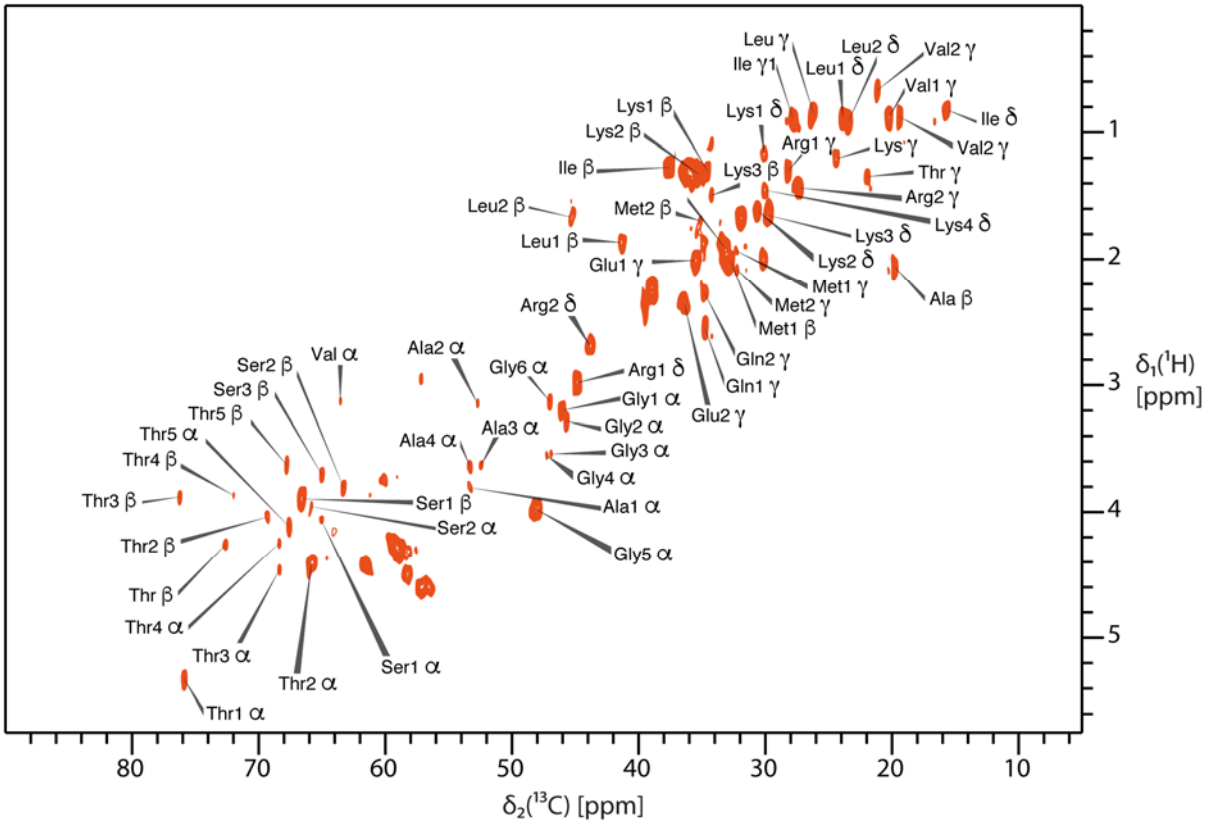


Figure 4-23 – 2D [^{1}H,^{13}C]-INEPT spectrum of [*U*-^{13}C,^{15}N]-labeled Bax in ECPL liposomes at 283 K with a 35 kHz spinning frequency in a 1.9mm rotor (Table 2). Spin-system typing of carbon resonances from mobile residues is indicated.

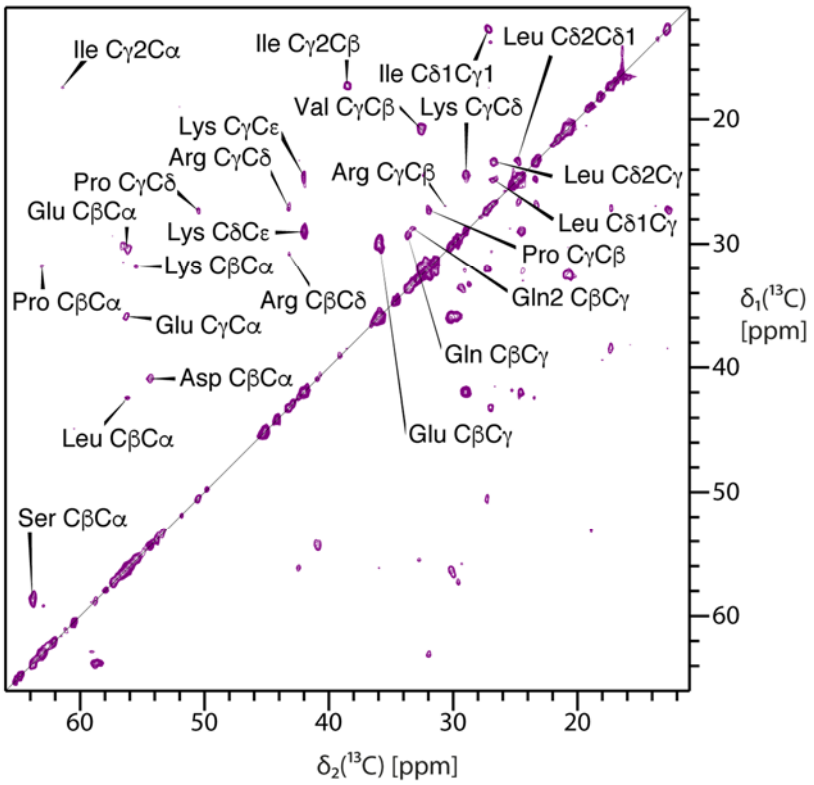


Figure 4-24 – 2D [^{13}C,^{13}C]-INEPT-TOBSY spectrum of [*U*-^{13}C,^{15}N]-labeled Bax in ECPL liposomes at 283 K with a 35 kHz spinning frequency in a 1.9mm rotor (Table 2). Spin systems of cross-peak resonances from mobile residues are indicated.

The identified spin systems are summarized in Table 4-2. An isoleucine residue could be identified on both spectra. As all isoleucines are rigid in the free form, it follows that besides the N-terminal there is at least one other flexible region in the membrane-inserted form.

Table 4-2 – Number of spin systems types identified from the 2D [^{13}C,^{13}C]-INEPT-TOBSY spectrum (with individually identified cross-peaks in parenthesis) and 2D [^{1}H,^{13}C]-INEPT (with individually identified carbon resonances in parenthesis).

Residues	2D [^{13}C,^{13}C]-INEPT-TOBSY	2D [^{1}H,^{13}C]-INEPT	Free Bax flexible	Bax total
Alanine		4 (4 Cα; 1 Cβ)	2	14
Arginine	1 (CγCβ; CγCδ)	2 (2 Cδ; 2 Cγ)	2	11
Aspartate	1 (CβCα)		6	13
Glutamine	2 (CγCβ)	2 (2 Cγ)	2	9
Glutamate	1 (CγCβ)	2 (2 Cγ)	3	10
Glycine	n.a.	6 (6 Cα)	13	21
Isoleucine	1 (Cδ1Cγ1; Cγ2Cβ; Cγ2Cα)	1 (1 Cβ; 1 Cδ; 1 Cγ)	0	9
Leucine	1 (Cδ2Cδ1; Cδ2Cγ; Cδ1Cγ; CβCα)	2 (2 Cβ; 1 Cδ; 1 Cγ)	4	22
Lysine	1 (CγCδ; CγCε; CδCε)	4 (3 Cβ; 4 Cδ; 1 Cγ)	3	9
Methionine		2 (2 Cβ; 2 Cγ)	4	8
Proline	1 (CβCα)		6	8
Serine	1 (CβCα)	3 (2 Cα; 3 Cβ)	5	11
Threonine		5 (5 Cα; 5 Cβ; 1 Cγ2)	4	13
Valine	1 (CγCβ)	3 (1 Cα; 3 Cγ)	3	11

Discussion

In the present work, we have employed cell-free protein expression to improve the total production yield of isotope-labeled samples of the protein Bax. This approach circumvents the toxicity and low expression yields in standard *E.coli* expression systems (52, 88). Furthermore, the tendency of Bax to aggregate and oligomerize was partially resolved through the fusion with different solubility tags. Bax cell-free expression was optimized to ~0.33mg/ml, forming the basis for cost-efficient production of labeled samples (90). The use of solubilizing detergents during Bax purification is described at length among the literature (50, 51, 90, 91). Different detergents have also been shown to activate Bax (50, 92), distinctly than the BH3-only activators (92). After an initial detergent trial, in order to avoid interpretation mistakes, the use of detergents to stabilize Bax was precluded.

Bax 2D [^{15}N,^{1}H]-TROSY-HSQC resonance dispersion demonstrates the expression of a well-folded protein (Figure 4-20 A), which superimposes quite well with the known sequence-specific resonance assignment (52). The membrane-bound form of Bax (Figure 4-20 B) shows that the N-terminal flexible region is not in contact with the bicelle, tumbling at a different regime. It seems the slower tumbling regime due to bicelle immobilization leads to peak broadening concealing non-flexible resonances, this transition from rigid to flexible evidence a C-terminal structural rearrangement upon membrane binding. Additional resonances with lower intensity are also seen. These point to flexible residues in a distinct conformation from the free form. The small diameter of q=0.3 bicelles (~8 nm) (87) could comfortably accommodate a 20 kDa (3-4 nm diameter) globular particle (93), possibly a dimer, but would effectively halt oligomerization. It seems reasonable to conclude that a pore intermediate is bound to the membrane, with several flexible residues including the N-terminal. The bicelle acts as a cage by physically trapping this membrane-bound intermediate limiting further oligomerization. This observation is supported by the accepted mechanism of Bax activation, where a single monomer/dimer is found in the mitochondria outer membrane in equilibrium with the free form. This intermediate is purportedly rescued from pore-formation by Bcl-Xl and other antiapoptotic members or led further down the oligomerization path by t-Bid and other BH3-only activators (28, 94).

High-temperature incubation has been shown before to activate Bax in a similar way to BH3-only members, allowing membrane insertion (43). EM imaging of proteoliposomes confirms Bax to both co-localize and rupture the liposomal membrane after high-temperature incubation (SFig.2). The liposome pull-down shows that after protein/lipid mass ratio optimization, essentially all Bax is inserted into the liposomes. The presence of limited sets of resonances in both flexible and mobile

residues in the solid-state NMR spectra (Fig. 4) and the absence of apparent peak doubling suggests a high degree of conformational homogeneity by the individual Bax molecules in our preparation.

Comparing the liposome-inserted Bax 2D [^{1}H,^{15}N]-HSQC with the bicelle bound (Figure 4-20 B and C), the N-terminal region is found to be mobile in both cases. The flexible residues evidenced in both the 2D [^{13}C,^{13}C]-INEPT-TOBSY and 2D [^{1}H,^{13}C]-INEPT show that at least 33 residues out of the 192 amino acids remain flexible (Table 4-2). Interestingly, one isoleucine is found to be flexible in the membrane-inserted state, as opposed to none within the free form. It seems liposome-inserted Bax is a step further in the oligomerization path, where most residues are buried in the membrane, and the flexible residues experience an entirely different chemical environment, which explains the different chemical shifts for flexible regions, between the liposome inserted and the free form. The presence of a flexible isoleucine unequivocally evidences an overall structure rearrangement upon membrane interaction.

Overall, by combining cell-free expression with solid-state NMR spectroscopy, this work has enabled the first high-resolution spectra of membrane-inserted Bax. Whereas additional work will be required to provide structural information *via* sequence-specific resonance assignments, our approach opens a new experimental route to study the insertion and pore formation mechanism of Bax at the atomic level.

4.3. References

1. C. Kleanthous, Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat. Rev. Microbiol.* **8**, 843–848 (2010).
2. K. S. Jakes, W. A. Cramer, Border crossings: colicins and transporters. *Annu. Rev. Genet*. **46**, 209–31 (2012).
3. E. Gouaux, The long and short of colicin action: the molecular basis for the biological activity of channel-forming colicins. *Structure*. **5**, 313–7 (1997).
4. S. K. Buchanan *et al.*, Structure of colicin I receptor bound to the R-domain of colicin Ia: implications for protein import. *EMBO J.* **26**, 2594–2604 (2007).
5. K. S. Jakes, A. Finkelstein, The colicin Ia receptor, Cir, is also the translocator for colicin Ia. *Mol. Microbiol.* **75**, 567–578 (2010).
6. M. Chauleau, L. Mora, J. Serba, M. De Zamaroczy, FtsH-dependent processing of RNase colicins d and e3 means that only the cytotoxic domains are imported into the cytoplasm. *J. Biol. Chem.* **286**, 29397–29407 (2011).
7. M. De Zamaroczy, L. Mora, A. Lecuyer, V. Géli, R. H. Buckingham, Cleavage of colicin D is necessary for cell killing and requires the inner membrane peptidase LepB. *Mol. Cell*. **8**, 159–168 (2001).
8. G. Kurisu *et al.*, The structure of BtuB with bound colicin E3 R-domain implies a translocon. *Nat. Struct. Biol.* **10**, 948–954 (2003).
9. J. Konisky, B. S. Cowell, Interaction of colicin Ia with bacterial cells. Direct measurement of Ia-receptor interaction. *J. Biol. Chem.* **247**, 6524–6529 (1972).
10. P. K. Kienker, K. S. Jakes, A. Finkelstein, Protein translocation across planar bilayers by the colicin Ia channel-forming domain: where will it end? *J. Gen. Physiol.* **116**, 587–98 (2000).
11. M. Wiener, D. Freymann, P. Ghosh, R. M. Stroud, Crystal structure of colicin Ia. *Nature*. **385**, 461–464 (1997).
12. S. L. Greig, M. Radjainia, A. K. Mitra, Oligomeric structure of colicin ia channel in lipid bilayer membranes. *J. Biol. Chem.* **284**, 16126–16134 (2009).
13. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94**, 12366–12371 (1997).
14. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci.* **95**, 13585–13590 (1998).
15. G. T. Montelione, B. A. Lyons, S. D. Emerson, M. Tashiro, An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* **114**, 10974–10975 (1992).
16. S. Grzesiek, J. Anglister, A. Bax, Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magn. Reson. Ser. B*. **101**, 114–119 (1993).
17. N. E. Hafsa, D. Arndt, D. S. Wishart, CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Res.* **43**, W370-7 (2015).
18. M. W. Parker, J. P. M. Postma, F. Pattus, A. D. Tucker, D. Tsernoglou, Refined structure of the pore-forming domain of colicin A at 2.4 Å resolution. *J. Mol. Biol.* **224**, 639–657 (1992).
19. F. Gruss *et al.*, The structural basis of autotransporter translocation by TamA. *Nat. Struct. Mol. Biol.* **20**, 1318–1320 (2013).
20. D. Hockenbery, G. Nuñez, C. Milliman, R. D. Schreiber, S. J. Korsmeyer, Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature*. **348**, 334–336 (1990).

21. Z. N. Oltval, C. L. Milliman, S. J. Korsmeyer, Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programed cell death. *Cell*. **74**, 609–619 (1993).
22. L. H. Boise *et al.*, Bcl-X, a Bcl-2-Related Gene That functions as a dominant regulator of apoptotic cell death. *Cell*. **74**, 597–608 (1993).
23. K. Eguchi, Apoptosis in autoimmune diseases. *Intern. Med.* **40**, 275–284 (2001).
24. R. Wong, Apoptosis in cancer: from pathogenesis to treatment. *J. Exp. Clin. Cancer Res.* **30**, 1–14. (2011).
25. K. Wang, X. M. Yin, D. T. Chao, C. L. Milliman, S. J. Korsmeyer, BID: A novel BH3 domain-only death agonist. *Genes Dev.* **10**, 2859–2869 (1996).
26. R. J. Youle, A. Strasser, The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* **9**, 47–59 (2008).
27. T. W. Sedlak *et al.*, Multiple Bcl-2 family members demonstrate selective dimerizations with Bax. *Proc. Natl. Acad. Sci.* **92**, 7834–7838 (1995).
28. T. T. Renault, O. Teijido, B. Antonsson, L. M. Dejean, S. Manon, Regulation of Bax mitochondrial localization by Bcl-2 and Bcl-x L: Keep your friends close but your enemies closer. *Int. J. Biochem. Cell Biol.* **45**, 64–67 (2013).
29. C. R. Stocking, E. M. Gifford, Incorporation of thymidine into chloroplasts of *Spirogyra*. *Biochem. Biophys. Res. Commun.* **1**, 159–164 (1959).
30. S. W. Muchmore *et al.*, X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature*. **381**, 335–341 (1996).
31. S. L. Schendel, M. Montal, J. C. Reed, Bcl-2 family proteins as ion-channels. *Cell Death Differ.* **5**, 372–380 (1998).
32. E. Taylor-Brown, H. Hurd, The first suicides: a legacy inherited by parasitic protozoans from prokaryote ancestors. *Parasit. Vectors*. **6**, 108 (2013).
33. E. V Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death Differ.* **9**, 394–404 (2002).
34. S. Manon, B. Chaudhuri, M. Guérin, Release of cytochrome c and decrease of cytochrome c oxidase in Bar-expressing yeast cells, and prevention of these effects by coexpression of Bcl-x(L). *FEBS Lett.* **415**, 29–32 (1997).
35. M. G. Blewitt, L. a Chung, E. London, Effect of pH on the conformation of diphtheria toxin and its implications for membrane penetration. *Biochemistry*. **24**, 5458–5464 (1985).
36. K. Sandvig, S. Olsnes, Diphtheria toxin entry into cells is facilitated by low pH. *J. Cell Biol.* **87**, 828–832 (1980).
37. S. F. Mel, R. M. Stroud, Colicin Ia inserts into negatively charged membranes at low pH with a tertiary but little secondary structural change. *Biochemistry*. **32**, 2082–2089 (1993).
38. a J. Minn *et al.*, Bcl-x(L) forms an ion channel in synthetic lipid membranes. *Nature*. **385** (1997), pp. 353–357.
39. P. H. Schlesinger *et al.*, Comparison of the ion channel characteristics of proapoptotic BAX and antiapoptotic BCL-2. *Proc. Natl. Acad. Sci.* **94**, 11357–11362 (1997).
40. B. Antonsson, Inhibition of Bax Channel-Forming Activity by Bcl-2. *Science.* **277**, 370–372 (1997).
41. Z. Xie, S. Schendel, S. Matsuyama, J. C. Reed, Acidic pH promotes dimerization of Bcl-2 family proteins. *Biochemistry*. **37**, 6410–6418 (1998).
42. P. F. Cartron, L. Oliver, E. Mayat, K. Meflah, F. M. Vallette, Impact of pH on Bax α conformation, oligomerisation and mitochondrial integration. *FEBS Lett.* **578**, 41–46 (2004).
43. S. Bleicken *et al.*, Molecular details of Bax activation, oligomerization, and membrane insertion. *J. Biol. Chem.* **285**, 6636–6647 (2010).
44. P. H. Schlesinger, M. Saito, The Bax pore in liposomes, Biophysics. *Cell Death Differ.* **13**, 1403–8 (2006).
45. L. J. Pagliari *et al.*, The multidomain proapoptotic molecules Bax and Bak are directly activated by heat. *Proc. Natl. Acad. Sci.* **102**, 17975–17980 (2005).

46. S. Bleicken, K. Zeth, Conformational changes and protein stability of the pro-apoptotic protein Bax. *J. Bioenerg. Biomembr.* **41**, 29–40 (2009).
47. J. A. Yethon, R. F. Epand, B. Leber, R. M. Epand, D. W. Andrews, Interaction with a membrane surface triggers a reversible conformational change in Bax normally associated with induction of apoptosis. *J. Biol. Chem.* **278**, 48935–48941 (2003).
48. E. Gavathiotis *et al.*, BAX activation is initiated at a novel interaction site. *Nature*. **455**, 1076–81 (2008).
49. E. Gavathiotis, D. E. Reyna, M. L. Davis, G. H. Bird, L. D. Walensky, BH3-triggered structural reorganization drives the activation of proapoptotic Bax. *Mol. Cell*. **40**, 481–492 (2010).
50. Y. T. Hsu, R. J. Youle, Bax in murine thymus is a soluble monomeric protein that displays differential detergent-induced conformations. *J. Biol. Chem.* **273**, 10777–10783 (1998).
51. Y. Te Hsu, R. J. Youle, Nonionic detergents induce dimerization among members of the Bcl-2 family. *J. Biol. Chem.* **272**, 13829–13834 (1997).
52. M. Suzuki, R. J. Youle, N. Tjandra, Structure of Bax. *Cell*. **103**, 645–654 (2000).
53. G. Basanez *et al.*, Bax, but not Bcl-xL, decreases the lifetime of planar phospholipid bilayer membranes at subnanomolar concentrations. *Proc. Natl. Acad. Sci.* **96**, 5492–5497 (1999).
54. X. Roucou, T. Rostovtseva, S. Montessuit, J.-C. Martinou, B. Antonsson, Bid induces cytochrome c-impermeable Bax channels in liposomes. *Biochem. J.* **363**, 547–552 (2002).
55. T. K. Rostovtseva *et al.*, Bid, but not Bax, regulates VDAC channels. *J. Biol. Chem.* **279**, 13575–13583 (2004).
56. S. H. Lin *et al.*, Bax forms two types of channels, one of which is voltage-gated. *Biophys. J.* **101**, 2163–2169 (2011).
57. R. F. Epand, J. C. Martinou, S. Montessuit, R. M. Epand, C. M. Yip, Direct evidence for membrane pore formation by the apoptotic protein Bax. *Biochem. Biophys. Res. Commun.* **298**, 744–749 (2002).
58. B. Schafer *et al.*, Mitochondrial outer membrane proteins assist Bid in Bax-mediated lipidic pore formation. *Mol. Biol. Cell*. **120**, 3083–3092 (2009).
59. L. A. Gillies *et al.*, Visual and functional demonstration of growing Bax-induced pores in mitochondrial outer membranes. *Mol. Biol. Cell*. **26**, 339–49 (2015).
60. X.-P. Xu *et al.*, Three-dimensional structure of Bax-mediated pores in membrane bilayers. *Cell Death Dis.* **4**, e683 (2013).
61. S. Bleicken, O. Landeta, A. Landajuela, G. Basanez, A. J. García-Sáez, Proapoptotic Bax and Bak proteins form stable protein-permeable pores of tunable size. *J. Biol. Chem.* **288**, 33241–33252 (2013).
62. B. G. Childs, D. J. Baker, J. L. Kirkland, J. Campisi, J. M. Van Deursen, Senescence and apoptosis: dueling or complementary cell fates? *EMBO Rep.* **15**, 1–15 (2014).
63. P. Meier, A. Finch, G. Evan, Apoptosis in development. *Nature*. **407**, 796–801 (2000).
64. M. Pasparakis, Regulation of tissue homeostasis by NF-kappaB signalling: implications for inflammatory diseases. *Nat. Rev. Immunol.* **9**, 778–788 (2009).
65. J. Lopez, S. W. G. Tait, Mitochondrial apoptosis: killing cancer using the enemy within. *Br. J. Cancer*. **112**, 957–62 (2015).
66. P. E. Czabotar, G. Lessene, A. Strasser, J. M. Adams, Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* **15**, 49–63 (2014).
67. S. Nouraini, E. Six, S. Matsuyama, S. Krajewski, J. C. Reed, The putative pore-forming domain of Bax regulates mitochondrial localization and interaction with Bcl-X(L). *Mol. Cell. Biol.* **20**, 1604–15 (2000).
68. E. Gavathiotis, D. E. Reyna, J. a Bellairs, E. S. Leshchiner, L. D. Walensky, Direct and selective small-molecule activation of proapoptotic BAX. *Nat. Chem. Biol.* **8**, 639–645 (2012).
69. S. Bleicken *et al.*, Structural Model of Active Bax at the Membrane. *Mol. Cell*. **56**, 496–505 (2014).
70. S. Reckel *et al.*, Solution NMR Structure of Proteorhodopsin. *Angew. Chemie Int. Ed.* **50**, 11942-

11946 (2011).
71. J. Chalmeau, N. Monina, J. Shin, C. Vieu, V. Noireaux, α-Hemolysin pore formation into a supported phospholipid bilayer using cell-free expression. *Biochim. Biophys. Acta.* **1808**, 271–278 (2011).
72. S. Reckel *et al.*, Cell-Free Protein Production, in *Methods in Molecular Biology* **607**, 187–212 (2010).
73. E. Michel, K. Wüthrich, High-yield *Escherichia coli*-based cell-free expression of human proteins. *J. Biomol. NMR*. **53**, 43–51 (2012).
74. S. R. Bond, C. C. Naus, RF-Cloning.org: An online tool for the design of restriction-free cloning projects. *Nucleic Acids Res.* **40**, 209–213 (2012).
75. B. Gipson, X. Zeng, H. Stahlberg, 2dx - Automated 3D structure reconstruction from 2D crystal data. *Microsc. Microanal.* **14 Suppl 2**, 1290–1291 (2008).
76. B. Gipson, X. Zeng, Z. Y. Zhang, H. Stahlberg, 2dx-User-friendly image processing for 2D crystals. *J. Struct. Biol.* **157**, 64–72 (2007).
77. R. A. Crowther, R. Henderson, J. M. Smith, MRC image processing programs. *J. Struct. Biol.* **116**, 9–16 (1996).
78. J. M. Valpuesta, J. L. Carrascosa, R. Henderson, Analysis of electron microscope images and electron diffraction patterns of thin crystals of Ø29 connectors in ice. *J. Mol. Biol.* **240** (1994), pp. 281–287.
79. A. Böckmann *et al.*, Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR*. **45**, 319–327 (2009).
80. G. a. Morris, R. Freeman, Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc.* **233**, 760–762 (1979).
81. D. P. Burum, R. R. Ernst, Net polarization transfer *via* a J-ordered state for signal enhancement of low-sensitivity nuclei. *J. Magn. Reson.* **39**, 163–168 (1980).
82. O. C. Andronesi *et al.*, Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **127**, 12965–12974 (2005).
83. G. Bodenhausen, D. J. Ruben, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189 (1980).
84. M. Baldus, B. H. Meier, Total correlation spectroscopy in the solid state. The use of scalar couplings to determine the through-bond connectivity. *J. Magn. Reson. Ser. A*. **121**, 65–69 (1996).
85. E. H. Hardy, R. Verel, B. H. Meier, Fast MAS total through-bond correlation spectroscopy. *J. Magn. Reson.* **148**, 459–464 (2001).
86. W. F. Vranken *et al.*, The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.* **59**, 687–696 (2005).
87. K. J. Glover *et al.*, Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys. J.* **81**, 2163–2171 (2001).
88. C. J. Tsai *et al.*, BAX-induced apoptosis can be initiated through a conformational selection mechanism. *Structure*. **23**, 139–148 (2015).
89. T. Terada, S. Yokoyama, *Escherichia coli Cell-Free protein synthesis and isotope labeling of mammalian proteins* (Elsevier Inc., ed. 1, 2015), vol. 565.
90. S. Montessuit, G. Mazzei, E. Magnenat, B. Antonsson, Expression and purification of full-length human Bax alpha. *Protein Expr. Purif.* **15**, 202–206 (1999).
91. B. Antonsson, S. Montessuit, S. Lauper, R. Eskes, J. C. Martinou, Bax oligomerization is required for channel-forming activity in liposomes and to trigger cytochrome c release from mitochondria. *Biochem. J.* **345 Pt 2**, 271–278 (2000).
92. T. Li, T. Brustovetsky, B. Antonsson, N. Brustovetsky, Dissimilar mechanisms of cytochrome c release induced by octyl glucoside-activated BAX and by BAX activated with truncated BID. *Biochim. Biophys. Acta.* **1797**, 52–62 (2010).
93. H. P. Erickson, Size and shape of protein molecules at the nanometer level determined by

sedimentation, gel filtration, and electron microscopy. *Biol. Proced. Online*. **11**, 32–51 (2009).
94. H. Kim *et al.*, Hierarchical regulation of mitochondrion-dependent apoptosis by BCL-2 subfamilies. *Nat. Cell Biol.* **8**, 1348–58 (2006).

Chapter 5

Structural and functional characterization of FIC-domain post-translational modification toxins

5. Structural and functional characterization of FIC-domain post-translational modification protein toxins

5.1. FIC-domain protein family

FIC-domain proteins are conserved in all domains of life, including viruses (1, 2), from bacteria to higher eukaryotes (3). They were first identified in an *E.coli* filamentation growth-arrest phenotype caused by cyclic-AMP, where the first identified FIC protein, the gain-of-function *FIC-1,* was found to be the responsible agent (4). Recently, FIC proteins were identified as PTM proteins (5, 6), and since several different target proteins have been identified, such as GTPases, unstructured protein segments, and bacterial enzymes (7–9), which implicates FIC proteins in distinct roles of cell function, such as pathogenicity (6, 8, 10), regulation of translation and stress response (11), as well as eukaryotic signal transduction (12).

FIC-domain proteins, which contain the FIC motif HPFx(D/E)GN(G/K)RxxR catalyze the transfer of a phosphoryl-containing group, through cleavage of a substrate phosphodiester bond, onto a target protein (13). The transfer of an AMP moiety, from an ATP substrate, is termed adenylylation, or AMPylation, and the respective catalytic enzymes, adenylyl (or AMP) transferases. FIC proteins with a degenerate FIC motif were found to display distinct catalytic activities, such as phosphocholine transferase (8) and kinase (11). In these, the catalytic site binds a nucleotide substrate in an inverse conformation compared to the canonical FIC motif, leading to the transfer of a different moiety. This is exemplified by the Death on curing (Doc) protein toxin, which catalyzes the transfer of a phosphate group from ATP onto the target protein (11), and the Ankyrin repeat-containing protein X (AnkX), which catalyzes the transfer of a phosphocholine from CDP-choline (8). Apart from the FIC-domain proteins PTM modification of target proteins, FIC proteins have also been shown to be able to catalyze automodification reactions such as auto-AMPylation (13–15).

The conserved FIC-domain is structurally defined by an α-helical core enclosing the catalytic motif (Figure 5-1). Further, a flexible loop found in most FIC proteins so far, termed flap, located in close proximity to the active site, aid target docking through the formation of an intermolecular antiparallel β-sheet-like element (15, 16). The invariant catalytic histidine deprotonates an incoming target side-chain hydroxyl group, which subsequently enables the nucleophilic attack on the ligand phosphodiester bond (7, 17, 18).

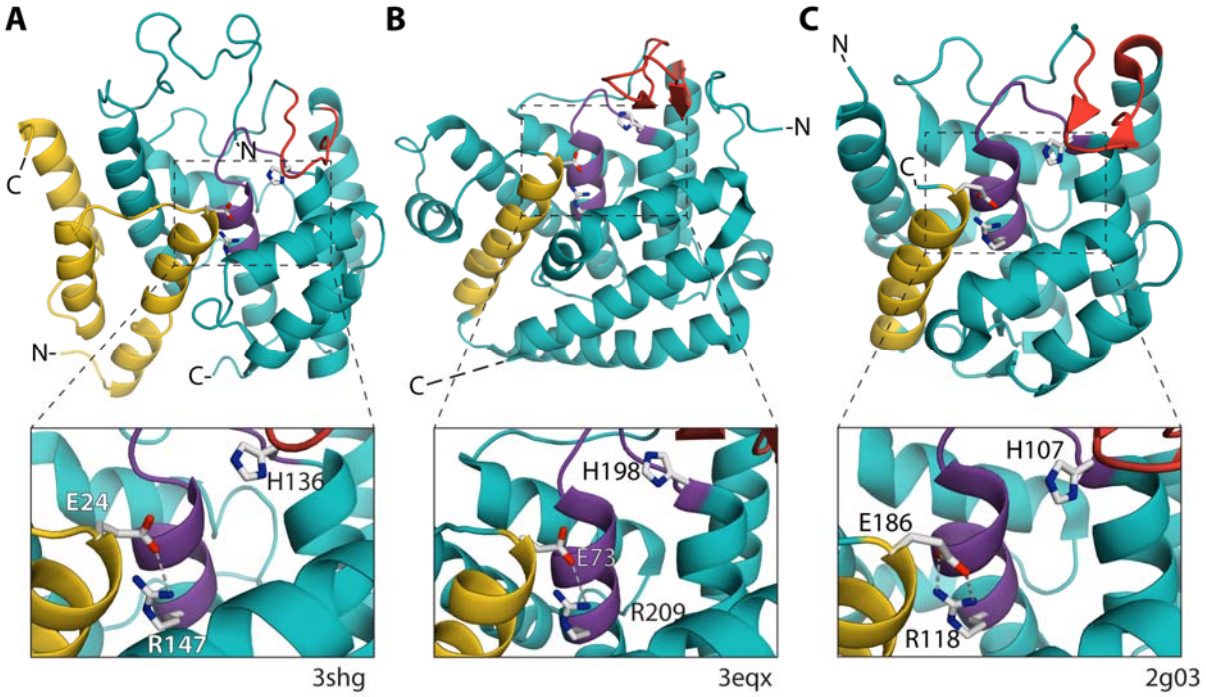


Figure 5-1 – Structures of FIC-domain members from distinct classes in cartoon representation, with the α_{inh} highlighted in yellow, the flap domain in red, the catalytic motif in purple, and their respective PDB IDs. In A is shown the Class I VbhT toxin in complex with its cognate VbhA antitoxin; in B is shown the N-terminal *Shewanella oneidensis* FIC (SoFIC) domain; and in C the *Neisseria meningitidis* FIC (NmFIC). In the insets are shown the catalytic and inhibitory residues in stick representation of each enzyme, as well as the E24-R147 hydrogen bond.

FIC-domain proteins that catalyze the transfer of a nucleotidyl group contain a regulatory element, which is constituted by an inhibitory α-helix (α_{inh}) (depicted in yellow in Figure 5-1). This group contains the conserved (S/T)xxxE(G/N) motif with the invariant glutamate whose carboxyl side-chain binds a conserved arginine in the FIC motif. This interaction effectively prevents the competent binding of an NTP, therefore abolishing the nucleophilic attack by the hydroxyl-containing incoming target protein (2). FIC proteins are classified by the position of the α_{inh} in relation to the FIC-domain, Class I exhibit an intermolecular α_{inh} on a separate polypeptide chain in the form of a toxin/antitoxin complex (Figure 5-1 A), whereas Class II and Class III exhibit an intramolecular N-terminal (Figure 5-1 B) or C-terminal (Figure 5-1 C) α_{inh}, respectively (2). The physiological α_{inh} toxin inhibition relief mechanism remains currently elusive (19).

Class I of FIC-domain proteins

Class I FIC proteins, found exclusively in bacteria, are classic toxin/antitoxin complexes, where the antitoxin bears the α_{inh} motif. Expression of the FIC-domain protein alone is therefore highly toxic to bacteria. Recently, the *Bartonella schoenbuchensis* FIC-domain VbhT/VbhA (Figure 5-1 A) protein toxin/antitoxin complex was found to be the responsible agent for an *E.coli* growth-arrest phenotype, which is abrogated upon co-expressing VbhT with VbhA (2). The formation of the FIC-target complex and nucleotide substrate binding mechanisms were explored further, described below in section 5.3.

Class II of FIC-domain proteins

Approximately 80% of all identified FIC-domain proteins so far belong to Class II of FIC proteins. They are characterized by an N-terminal regulatory domain α_{inh}, relative to the FIC-domain (*2*). The underlying mechanism and function of this large class of FIC proteins, which includes the human FIC homologue HypE, is mostly unknown at this point. Apart from HypE, briefly discussed in the first chapter, another Class II member the *Shewanella oneidensis* FIC (SoFIC) (Figure 5-1 B) AMP transferase was the first described Class II FIC protein (21). Additionally to the N-terminal α_{inh}, SoFIC exhibits a C-terminal winged helix-turn-helix dimerization domain postulated to be a DNA-binding domain (21).

Class III of FIC-domain proteins

This class of FIC proteins is a small group of bacterial single-domain proteins, characterized by a C-terminal regulatory domain α_{inh} relative to the FIC-domain (*2*). Interestingly, *E.coli* expression of an inhibition-relieved mutant of class III Fic-domain protein NmFIC (Figure 5-1 C) was shown to result in the AMPylation of GyrB only, one of the two identified Fic targets (2). The precise role of the AMPylation and consequent toxic inhibition of GyrB within the cellular context is not yet completely understood. NmFIC protein conformation and function were further explored in the framework of this thesis, described below in sections 5.2 and 5.3.

Other FIC-domain proteins

FIC-domain proteins, with different catalytic transferase activities, have different regulatory groups. The Doc/Prevents host death (Phd) protein toxin/antitoxin complex and its homologues, belong to a particular class of FIC proteins, characterized by an incomplete FIC catalytic core, with a distinct catalytic motif (3, 22, 23). The Phd antitoxin replaces the α_{inh} as the substrate-binding regulator, with a non-canonical block of the catalytic site (11). Some FIC-domain proteins even lack a regulatory group altogether like the *Vibrio parahaemolyticus* type III effector (VopS) and the *Histophilus somni* Immunoglobulin binding protein A (IbpA) (7, 10).

Research Article

Frédéric V. Stanger
Björn M. Burmann
Alexander Harms
Hugo Aragão
Adam Mazur
Timothy Sharpe
Cristoph Dehio
Sebastian Hiller
and Tilman Schirmer

Intrinsic regulation of FIC-domain AMP-transferases by oligomerization and auto-modification

PNAS (2016),
113, E529-537

5.2. The *Neisseria meningitidis* FIC protein

5.2.1. Intrinsic regulation of FIC-domain AMP-transferases by oligomerization and auto-modification

In this publication, the complex autoinhibition mechanism of NmFIC is described. NmFIC from *Neisseria meningitidis* undergoes a complex oligomerization dependent control, where autoadenylation *in cis* leads to its activation and subsequent adenylylation of its target, the DNA gyrase subunit B GyrB. The autoadenylated form of NmFIC slows down the growth of ectopically expressing *E.coli* due to its toxicity, leading to very low protein expression. Cell-free expression was key to circumvent this problem and obtain sufficient amounts for the NMR structural characterization and analysis of the dynamic properties of the autoadenylated form of NmFIC. This data was crucial in the comparison of non- and autoadenylated FIC in the identification of the conformational changes induced upon automodification, and relation to its function.

Statement of contributions

I was responsible for the cell-free protein expression of autoadenylated NmFIC, NMR data collection, and analysis of this form of the protein, together with Dr. Björn Burmann.

Intrinsic regulation of FIC-domain AMP-transferases by oligomerization and automodification

Frédéric V. Stanger[a,b], Björn M. Burmann[a], Alexander Harms[b], Hugo Aragão[a], Adam Mazur[c], Timothy Sharpe[d], Christoph Dehio[b,1], Sebastian Hiller[a,1], and Tilman Schirmer[a,1]

[a]Focal Area Structural Biology and Biophysics, Biozentrum, University of Basel, CH-4056 Basel, Switzerland; [b]Focal Area Infection Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland; [c]Research IT, Biozentrum, University of Basel, CH-4056 Basel, Switzerland; and [d]Biophysics Facility, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

Edited by Anthony Maxwell, John Innes Centre, Norwich, United Kingdom, and accepted by the Editorial Board December 15, 2015 (received for review August 27, 2015)

Filamentation induced by cyclic AMP (FIC)-domain enzymes catalyze adenylylation or other posttranslational modifications of target proteins to control their function. Recently, we have shown that Fic enzymes are autoinhibited by an α-helix (α_{inh}) that partly obstructs the active site. For the single-domain class III Fic proteins, the α_{inh} is located at the C terminus and its deletion relieves autoinhibition. However, it has remained unclear how activation occurs naturally. Here, we show by structural, biophysical, and enzymatic analyses combined with in vivo data that the class III Fic protein NmFic from *Neisseria meningitidis* gets autoadenylylated *in cis*, thereby autonomously relieving autoinhibition and thus allowing subsequent adenylylation of its target, the DNA gyrase subunit GyrB. Furthermore, we show that NmFic activation is antagonized by tetramerization. The combination of autoadenylylation and tetramerization results in nonmonotonic concentration dependence of NmFic activity and a pronounced lag phase in the progress of target adenylylation. Bioinformatic analyses indicate that this elaborate dual-control mechanism is conserved throughout class III Fic proteins.

adenylylation | AMPylation | posttranslational modification | enzyme regulation | molecular timer

Fic (filamentation induced by cyclic AMP) proteins containing the FIC domain (pfam 02661) are found in all kingdoms of life. FIC domains encode enzymatic activities that modulate target protein function by diverse posttranslational modifications (1, 2). The vast majority of known Fic proteins are AMP-transferases that use ATP to catalyze the transfer of an AMP moiety onto a target hydroxyl side chain (3, 4). This reaction is akin to the situation in protein kinases, which catalyze γ-phosphate transfer onto target side chains.

Only a few Fic targets have been identified to date (2). IbpA (4) and VopS (3), two bacterial FIC-domain effectors that get translocated into host cells, catalyze the adenylylation of Rho GTPases, resulting in cytoskeleton collapse. Most recently, we have found that a subset of bacterial Fic proteins adenylylates DNA gyrase and topoisomerase IV, which leads to their inactivation and cellular growth arrest (5). The structure of a FIC-domain/target complex (6) and the catalytic mechanism have been determined (6, 7). However, the biological functions, as well as the molecular mechanism, of the vast majority of Fic proteins have remained elusive. Intriguingly, in vitro automodification has been demonstrated for most Fic proteins that have been described so far (6, 8–15), but its physiological relevance has remained unclear.

Recently, we have shown that Fic-mediated adenylylation is tightly regulated (8). In the native state, a helix partly obstructs the ATP binding site with a strictly conserved Glu blocking ATP γ-phosphate binding, thereby preventing productive/competent substrate binding. The inhibitory α-helix (α_{inh}) is either located on a separate protein that forms a tight toxin/antitoxin complex with the Fic enzyme (16) or at the N- or C-terminal position within the same polypeptide chain. These three possibilities lead to a classification of Fic proteins into classes I, II, and III, respectively (8). Mutation of the inhibitory Glu to Gly relieves autoinhibition, thus boosting both target and autoadenylylation (8, 9). However, the identity of the intrinsic or extrinsic factors that in vivo expulse the inhibitory Glu of α_{inh}, and thereby relieve autoinhibition, has not been investigated.

Fic proteins have evolved in bacteria and have spread by horizontal gene transfer into all domains of life (17). Despite structural conservation of the basic FIC-domain fold, there is significant sequence diversity among class I and II Fic proteins. Additionally, these two classes are frequently combined with other protein domains toward multidomain arrangements, demonstrating a high functional plasticity and adaptability (2). In contrast, class III Fic proteins are highly conserved single-domain proteins even though they are found scattered across all classes of Proteobacteria (8). This conservation suggests that class III Fic proteins are stand-alone autoregulated functional entities that act in a plug-and-play manner upon acquisition by horizontal gene transfer.

Here, we dissect the regulatory mechanism of NmFic from *Neisseria meningitidis* as representative for the class III Fic proteins. First, we identify the B-subunit of DNA gyrase as the main

Significance

FIC-domain enzymes are found in all kingdoms of life and catalyze posttranslational modifications of various target proteins to modulate their function. Because the vast majority of Fic proteins are expressed in an inhibited form, their physiological importance has escaped attention for a long time. This article reveals an autonomous mechanism of inhibition relief for class III Fic proteins, which hinges on autoadenylylation of an inhibitory helix. Because the process occurs *in cis*, the Fic enzyme constitutes a molecular timer that operates independent of enzyme concentration. Furthermore, we show that Fic-mediated adenylylation of DNA gyrase leads to bacterial growth arrest. Thus, the time-dependent inactivation of DNA gyrase may serve as a switch to bacterial dormancy under starvation or other stress conditions.

Author contributions: F.V.S., B.M.B., A.H., H.A., A.M., T. Sharpe, C.D., S.H., and T. Schirmer designed research; F.V.S., B.M.B., A.H., H.A., A.M., and T. Sharpe performed research; F.V.S., B.M.B., A.H., T. Sharpe, C.D., S.H., and T. Schirmer analyzed data; and F.V.S., B.M.B., A.H., T. Sharpe, C.D., S.H., and T. Schirmer wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.M. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: Coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5CGL, 5CKL, and 5CMT). Sequence-specific resonance assignments have been submitted to the Biological Magnetic Resonance Data Bank (accession code 26607).

[1]To whom correspondence may be addressed. Email: christoph.dehio@unibas.ch, sebastian.hiller@unibas.ch, or tilman.schirmer@unibas.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516930113/-/DCSupplemental.

bacterial target of NmFic, as for class I Fic proteins (5). We then reveal the crucial role of autoadenylylation in the activation of class III Fic proteins and the opposing role of oligomerization, resulting in a peculiar and intriguing NmFic concentration dependence of target adenylylation. Ultimately, because the oligomerization interfaces are either highly conserved or covaried, and because the modifiable residue Y183 is strictly conserved in class III Fic proteins, we anticipate that the combination of oligomer dissociation and subsequent *cis*-autoadenylylation is the major regulatory mechanism of class III Fic proteins.

Results

NmFic Adenylylates DNA Gyrase. We have previously shown that expression of inhibition-relieved mutants of various bacterial Fic proteins slowed down the growth of ectopically expressing *Escherichia coli* (8). This growth reduction correlated with in vitro adenylylation of an endogenous protein with a mass of about 90 kDa. Subsequently, it was shown that the class I Fic protein VbhT adenylylates the B-subunit of the bacterial topoisomerases DNA gyrase (GyrB) and topoisomerase IV (ParE) (5). The modification affects a conserved Tyr of the ATP binding-site lid, thus interfering with ATPase and topoisomerase activity.

Because expression of (class III) NmFic in its inhibition-relieved form ($NmFic_{E186G}$) also slows down *E. coli* growth (8), we tested whether NmFic modifies the same targets (Fig. 1). Corroborating our previous findings (8), $NmFic_{wt}$ does not adenylylate any of the target proteins. In contrast, $NmFic_{E186G}$ efficiently adenylylates *N. meningitidis* GyrB and the orthologous *E. coli* and *Mycobacterium tuberculosis* proteins, but not ParE of *N. meningitidis* or *E. coli*. Mutation of the *E. coli* GyrB acceptor site (Y109F or Y109A variant) completely abrogates the effect (Fig. 1), confirming the predicted modification. In addition, independent of the presence of the target, strong autoadenylylation of $NmFic_{wt}$ and $NmFic_{E186G}$ is observed (Fig. 1), which is investigated further below.

Conserved Tetrameric Organization of Class III Fic Proteins. NmFic invariably shows a tetrameric arrangement (Fig. 2 and Fig. S1) in different crystal forms and ligation states, even when the C-terminal helix is absent or disordered (Fig. S1 *A–C*). The tetramer is of 222 symmetry (dimer of dimers), and its formation involves two independent interfaces (Fig. 2 *A–C* and Movie S1). Interface 1 is mainly mediated by the apolar interactions of F70 and Y77, with their respective symmetry mates, and by R71 and E102, which form two isologous salt bridges (Fig. 2*B*). The interface 1 residues are strongly conserved among the 197 analyzed sequences of class III Fic proteins. Similarly, interface 2 is formed by apolar residues L155 and F159, as well as the two isologous salt bridges of R149 and E156 (Fig. 2*C*). These residues are not strongly conserved, but the interacting residues show strong covariation with the striking charge reversal of the E-R to K-E salt bridges in a subset of the class III proteins. Taken together, conservation and covariance of the surface residues of NmFic suggest that tetramer formation is of functional importance for class III Fic proteins.

To describe the stability of the tetramer in solution, we determined the dissociation constants by performing size exclusion chromatography coupled to multiangle laser light scattering (SEC-MALLS; Fig. 2*D*, *Top*) and analytical ultracentrifugation sedimentation velocity (AUC-SV), using partially fluorescein-labeled protein (Fig. 2*D*, *Bottom*) measurements at varying protein concentrations. The observed shift of apparent molecular mass and elution volume with protein concentration in SEC-MALLS (Fig. S2*A*) indicated an oligomer equilibrium with fast kinetics relative to the duration of an SEC run. Nonlinear least squares fit of the data with an appropriate thermodynamic model (Fig. S3 *A* and *B*) yielded the dissociation constants $K_{d,1}$ and $K_{d,2}$ for the two independent interfaces. For $NmFic_{wt}$, both values are in the low micromolar range (Table 1). Interestingly, addition of the ATP substrate stabilized the tetramer considerably (Fig. 2*D*, *Bottom*, and Table 1).

To test whether the crystallographically observed tetramer also occurs in solution and whether it is of physiological relevance, we generated NmFic variants with single point mutations in one or both of the interfaces. Based on the crystal structure, charge reversal mutations E102R and E156R were expected to disrupt interfaces 1 and 2, respectively. Indeed, $NmFic_{E102R}$ and $NmFic_{E156R}$ do not form tetramers but exhibit a concentration-dependent monomer/dimer equilibrium (Fig. 2*D*). Furthermore, the dimer dissociation constants of the two variants agree very well with the dimer dissociation constants obtained from the fit of the $NmFic_{wt}$ tetramerization data (Table 1), indicating that the two interfaces are not allosterically coupled. In addition, the crystal structures of $NmFic_{E102R}$ and $NmFic_{E156R}$ (Fig. S2 *B–E* and Table S1) revealed virtually identical interfaces to the interfaces present in the tetrameric WT protein (Fig. S2 *B–E*). Finally, combination of both mutations in $NmFic_{E102R,E156R}$ yielded an oligomerization-deficient monomeric mutant $NmFic_{mono}$, with a concentration-independent mass identical to the theoretical mass of the monomer (Fig. 2*D*).

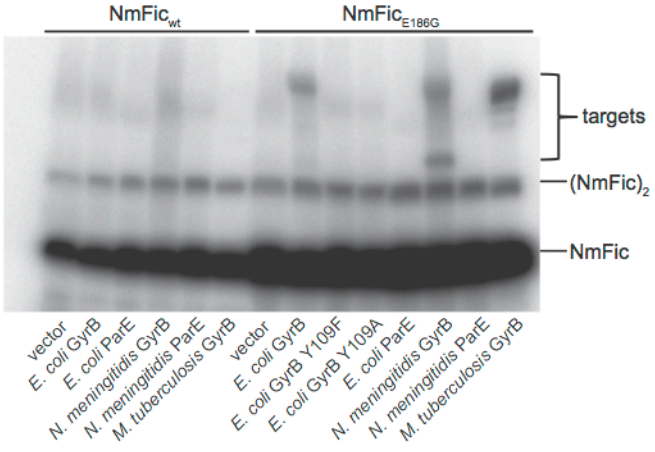


Fig. 1. $NmFic_{E186G}$ adenylylates GyrB. Autoradiographs obtained after incubation of $NmFic_{wt}$ (*Left*) and inhibition-relieved variant $NmFic_{E186G}$ (9) (*Right*) with 40 nM α-[^{32}P]ATP, 15 mM $MgCl_2$, and *E. coli* cell lysates overexpressing potential targets as indicated. Incubation was for 1 h at 30 °C.

Disruption of the Tetramer Activates NmFic. To obtain insight into the physiological role of oligomerization, the tetramerization-deficient mutants were assayed in vivo. Taking a derivative of *E. coli* K12 MG1655 (AB472) as a model organism (8), the bacterial growth inhibition upon expression of NmFic variants was assessed on LB-agar plates supplemented with increasing amounts of isopropyl β-D-thiogalactopyranoside (IPTG; up to 2 mM) as a physiological readout for gyrase inactivation. Fig. 3*A* shows the quantification of bacterial growth by counting colony-forming units (CFU; also Fig. S4 *A* and *B*).

In contrast to WT NmFic, the tetramerization-deficient mutants $NmFic_{E102R}$, $NmFic_{E156R}$, and $NmFic_{E102R,E156R}$ showed a severe, IPTG-dependent, growth defect (Fig. 3*A*), resulting in the loss of almost 1.5, 3.0, and 1.5 $\log_{10}$ CFU/mL viability, respectively. In addition to the strongly reduced numbers, the colonies exhibited a drastically smaller size (Fig. S4*B*). As expected, by additional mutation of the catalytic His (H107A), the observed growth defect was completely abolished (Fig. 3*A*), demonstrating that the phenotype depends on the catalytic activity of NmFic. As expected, the observed *E. coli* growth defect correlates very well with GyrB43 adenylylation, as assayed in vitro by autoradiography (Fig. 3*B*). Again, the combination of the interface mutations

with H107A completely suppressed both autoadenylylation and target adenylylation.

Although all tested interface mutants are active, $NmFic_{E102R}$ and $NmFic_{E102R,E156R}$ show lower activity than $NmFic_{E156R}$. An explanation for this reduced activity may be that the strictly conserved acidic residue 102, which is replaced in both mutants, may contribute to the recognition of the target protein GyrB. Note that the autoadenylylation efficiency is very similar for all oligomerization-deficient mutants (Fig. 3 *B* and *E*), suggesting that the active center, per se, is unperturbed. In summary, the results indicate that NmFic tetramerization renders the enzyme incompetent for efficient autoadenylylation and for target adenylylation.

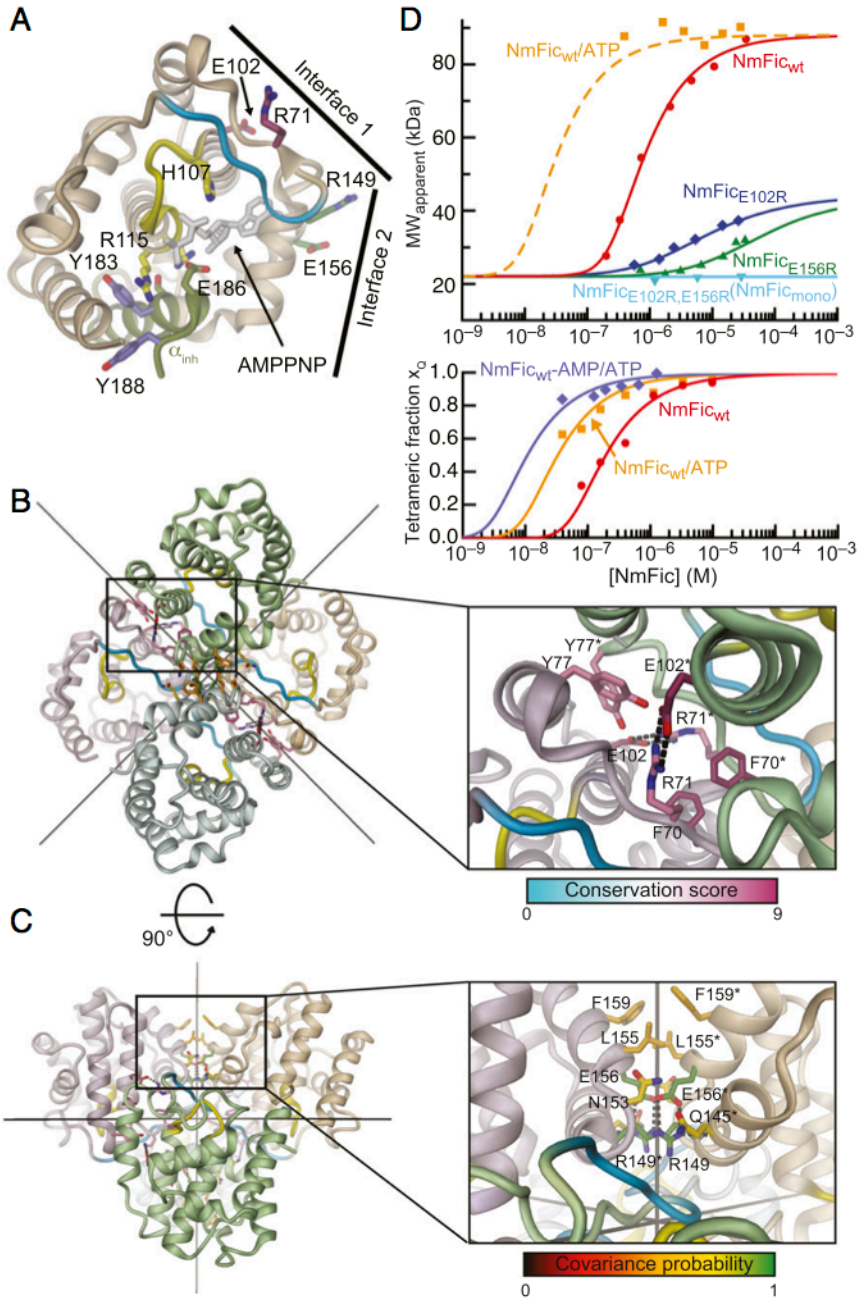


Fig. 2. Structural and oligomeric analysis of NmFic. (*A*) Cartoon of the NmFic structure (PDB ID code 3S6A) with the active site motif highlighted in yellow and the target binding site (flap) in blue. Important residues are shown in full with side-chain carbon atoms of active site residues colored in yellow, modifiable Tyr in purple, and charged residues mediating oligomerization in magenta (interface 1) and green (interface 2). The α_{inh} is shown in olive-green with the inhibitory Glu (E186) shown in full. (*B*) View of the crystal structure of the NmFic tetramer (222 symmetry with the twofold axes indicated) with the subunits distinguished by color (*Left*) and close-up views of the dimerization interface 1, with a contact area of 835 Å^2 (*Right*). Side-chain carbon atoms are colored according to the conservation score [ConSurf (28)]. (*C*) Ninety-degree rotation of the crystal structure shown in *B* (*Left*) and close-up view of the dimerization interface 2, with a contact area of 425 Å^2 (*Right*). Side-chain carbon atoms are colored according to covariance probability [Gremlin (29)]. In *B* and *C*, asterisks denote residues from the neighboring molecule. (*D*) Dynamic oligomerization equilibrium of $NmFic_{wt}$ and variants. (*Top*) Concentration dependence of the apparent molecular weight as determined by SEC-MALLS for $NmFic_{wt}$ and oligomerization interface variants. (*Bottom*) Tetrameric fraction x_Q of $NmFic_{wt}$ in the absence and presence of 5 mM ATP and autoadenylylated $NmFic_{wt}$-AMP in the presence of 5 mM ATP. The lines represent the nonlinear least squares fitting of the monomer/oligomer equilibria according to the model shown in Fig. S3*A*. Resulting dissociation constants are shown in Table 1.

Table 1. Biophysical characterization of NmFic oligomerization

NmFic variants and conditions	Method	$K_{d,1}$, μM	$K_{d,2}$, μM	$K_{d,2'}$, μM
$NmFic_{E102R}$	MALLS	7.4 ± 1.1	—	—
$NmFic_{E156R}$	MALLS	—	60 ± 12	—
$NmFic_{wt}$	MALLS	7.6 ± 1.1	61 ± 9	—
$NmFic_{wt}$*	AUC	2.6 ± 0.1	20	—
$NmFic_{wt}$ (5 mM ATP)*	AUC	0.71 ± 0.04	5.5	—
$NmFic_{wt}$-AMP (5 mM ATP)†	AUC	0.71	—	1.1 ± 0.2

Oligomerization parameters were fitted according to the kinetic scheme (Fig. S3*A*) to the data shown in Fig. 2*D*. $K_{d,1}$ and $K_{d,2}$ correspond to the dissociation constants for NmFic dimerization via interfaces 1 and 2, respectively.

*Due to the limited amount of data points, only one of the dissociation constants ($K_{d,1}$) was fitted and the other ($K_{d,2}$) was computed, keeping $K_{d,2}/K_{d,1} = 7.7$, based on the ratio obtained for $NmFic_{wt}$.

†Based on structural considerations, it can be assumed that the autoadenylylated C-terminal segment does not affect interface 1. Therefore, $K_{d,1}$ was set to the value obtained for the nonmodified protein, and only $K_{d,2'}$ was fitted (Fig. S3*C*).

NmFic Autoadenylylates a Buried Tyr *in Cis*. Automodification has been observed for most Fic proteins described to date (6, 8–15). For NmFic, the sites of autoadenylylation have been mapped to Y183 and Y188 of the α_{inh} by mass spectrometry (MS) (8) (Fig. 2*A*, purple residues). Intriguingly, and in contrast to Y188, the strictly conserved Y183 is completely buried in the hydrophobic core of the protein. Therefore, it can be inferred that the α_{inh} has to detach from the core of the protein for the Y183 adenylylation reaction to occur. Furthermore, for steric reasons, the α_{inh} would no longer be able to repack after modification.

Autoadenylylation of NmFic was monitored in real-time by circular dichroism (CD) spectroscopy. Because $NmFic_{wt}$ shows little autoadenylylation (Fig. 3*B*), we used $NmFic_{mono}$ ($NmFic_{E102R,E156R}$) for this and the following analyses. Upon addition of ATP/$MgCl_2$, the CD spectrum of $NmFic_{mono}$ showed a gradual, time-dependent decrease in amplitude of the negative peaks at 208 and 222 nm (Fig. S5*A*), corresponding to a reduction of α-helical content. This decrease is entirely consistent with (partial) unfolding of the α_{inh} upon autoadenylylation.

As another readout of autoadenylylation, the thermal stability of $NmFic_{mono}$, was monitored by differential scanning fluorimetry (DSF; Fig. S5*B*). The modification resulted in significant destabilization of the protein (decrease in melting temperature of 8 °C), consistent with the partial loss of secondary structure shown above and the concomitant loss of packing interactions. Importantly, the reaction kinetics were found to be independent of the NmFic concentration, indicating that the reaction takes place *in cis*. The data obtained at various $NmFic_{mono}$ concentrations (Fig. 4) could indeed be fitted globally with one apparent first-order rate constant, $k_{cat,1}$ (Table 2). Structural modeling verified that upon α_{inh} unfolding, the disordered C-terminal segment is long enough to allow Y183 to reach the target dock of the same molecule. A model of the structural changes accompanying NmFic autoadenylylation is shown in Movie S2.

Additionally, we tested whether the $NmFic_{mono,H107A}$ mutant, which is catalytically inactive and thus incompetent for *cis* modification, can be modified *in trans*. In mixtures of $NmFic_{mono}$ and $NmFic_{mono,H107A}$, the final concentration of the adenylylated product was found to correspond closely to the final concentration of the active $NmFic_{mono}$ enzyme only (Fig. S5*C*). This observation corroborates the conclusion that autoadenylylation occurs *in cis*.

Ultimately, we subjected our samples to MS analyses (Fig. S6 *A* and *B*). Indeed, $NmFic_{mono,H107A}$ showed the native mass, confirming that $NmFic_{mono}$ is not able to catalyze the modification of the inactive mutant *in trans*. In contrast, the mass of $NmFic_{mono}$ was increased by 1,316 Da, corresponding to the mass of four AMP moieties. Thus, apart from the already identified Tyr residues Y183 and Y188, the presence of two more acceptor sites can be inferred. The MS analysis was also performed on $NmFic_{mono,Y183F}$ (in the presence of $NmFic_{mono}$) and revealed two species corresponding to doubly and triply modified protein. As anticipated, we did not observe modification on four sites because the main acceptor site Y183 had been mutated in this protein variant. Furthermore, the Y183F mutation appears to slow down autoadenylylation (also Fig. 3*G*), because the automodification reaction did not reach completion (triply modified $NmFic_{mono,Y183F}$). Reduced autoadenylylation of the $NmFic_{mono,Y183F}$ mutant compared with $NmFic_{mono}$ was also observed by autoradiography (Fig. S6 *C* and *D*). Finally, residues Y184 and Y185 were identified as the remaining acceptor sites, because the mutant having all four Tyr residues (Y183, Y184, Y185, and Y188) replaced by Phe showed virtually no autoadenylylation. Satisfactorily, the sequential removal of individual acceptor sites in the combinatorial mutants resulted in a linear decrease of the band intensity on the autoradiographs (Fig. S6 *C* and *D*).

α_{inh} of Autoadenylylated NmFic Is Partly Unfolded. We further investigated the structural changes accompanying NmFic automodification by high-resolution nuclear magnetic resonance (NMR) spectroscopy (Fig. 5). We obtained well-dispersed NMR spectra for an unmodified protein (catalytically inactive $NmFic_{mono,H107A}$), as well as for autoadenylylated ($NmFic_{mono}$-AMP) protein. Sequence-specific backbone resonance assignments were obtained for 96% of the residues of $NmFic_{mono,H107A}$ (Fig. S7*A*). Based on these assignments, the chemical shifts of the autoadenylylated form could be assigned to 91% and confirmed by additional triple-resonance experiments. The overlay of the 2D [^{15}N,^{1}H]-HSQC (heteronuclear single quantum coherence) spectra of the native and autoadenylylated proteins shows large chemical shift differences for a subset of resonances (Fig. 5*A*). The residue-specific analysis of chemical shift changes upon autoadenylylation (Fig. 5 *B* and *C*) revealed that helix 8 (α_{inh}) is the most affected region of the protein. Therein, the covalently adenylylated residue Y183 showed the most pronounced chemical shift change [$\Delta\delta$(HN) = 6.7 ppm]. Furthermore, the amide chemical shifts of residues 178–191 populate the random coil region (7.5–8.5 ppm), indicating an unfolded conformation. In addition, the target-binding site (flap) and helix 1 (adjacent to helix 8) show significant chemical shift differences, but to a lesser extent than the α_{inh}, and a few residues in these regions are broadened beyond detection (Fig. 5*B*). Due to the point mutation H107A in the native form, the active site region also shows some chemical shift differences (Fig. 5 *B* and *C*). Despite these changes, the structural scaffold of the protein remains intact upon autoadenylylation. To characterize the secondary structure elements of both NmFic forms in solution, we used the secondary chemical shifts of the backbone ^{13}Cα nuclei (Fig. 5*D*). In aqueous solution, $NmFic_{mono}$ features eight α-helices, which agree in number and positioning with the crystal structure. The picture changes for the autoadenylylated form, in which the first seven helices are maintained, whereas the α_{inh} features helical structure only in its first two helical turns, but is unfolded from the position of residue G178 onward. These secondary structure changes are in full agreement with the chemical shift perturbation in 2D [^{15}N,^{1}H]-HSQC spectra (Fig. 5*D* and Movie S2). Overall, these data confirm the hypothesis that the α_{inh} cannot adopt its original position and conformation upon autoadenylylation of Y183.

NmFic Autoadenylylation Relieves Autoinhibition. We have shown that *cis*-autoadenylylation of NmFic constitutes an in-built mechanism to covalently modify Y183, resulting in partial unfolding of the α_{inh}. Obviously, this modification should be of functional relevance, considering that the α_{inh} in its native form partially

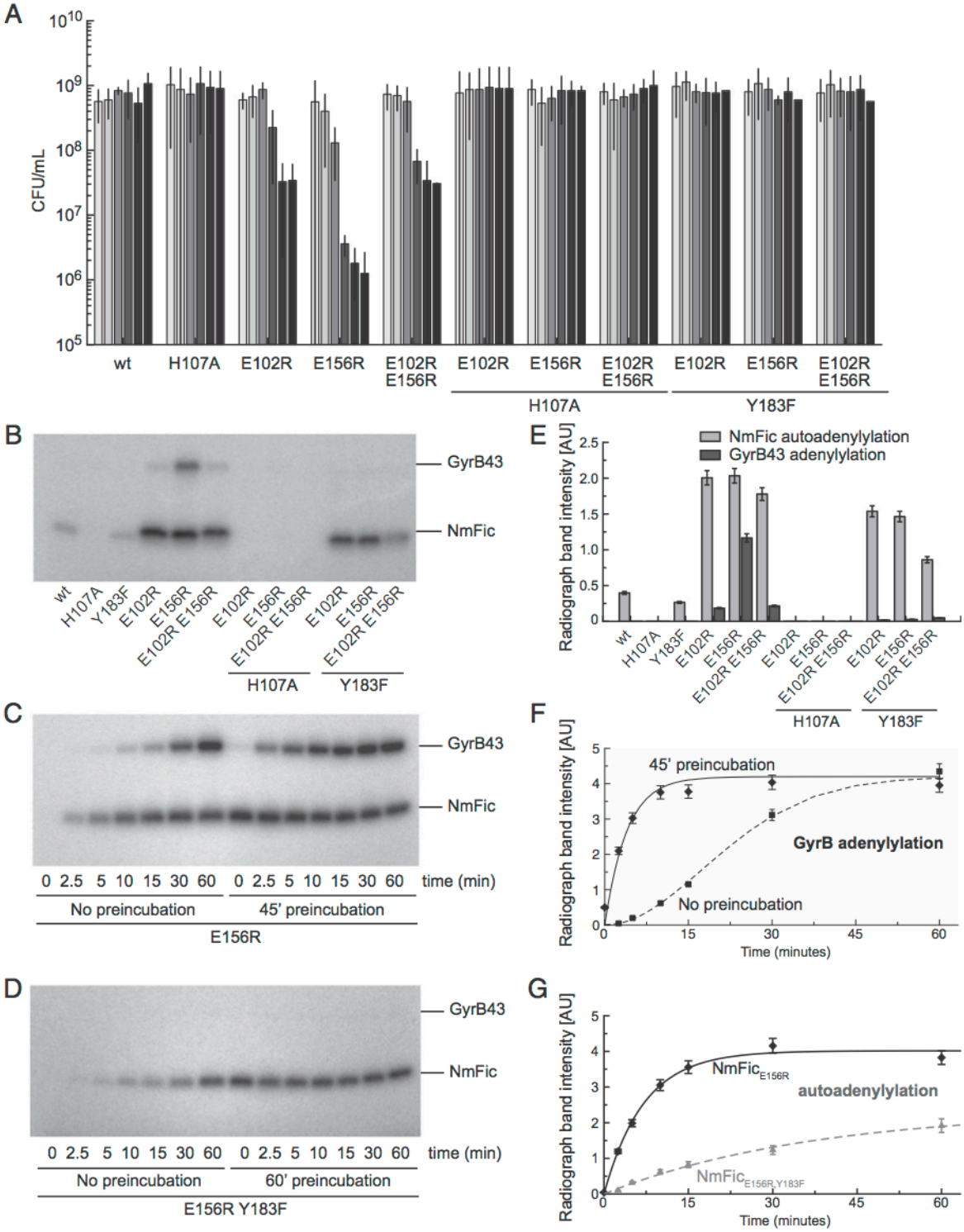


Fig. 3. Activation of NmFic leads to *E. coli* growth defect and in vitro adenylylation of GyrB. (*A*) Quantification of *E. coli* growth upon repression (1% glucose) or induction of the expression of NmFic variants at varying inducer (IPTG) concentrations from a single-copy plasmid (left to right with a color gradient from white to black: 1% glucose, no IPTG, 100 μM IPTG, 500 μM IPTG, 1 mM IPTG, 2 mM IPTG). The bars represent the average of three independent experiments, and error bars represent the SD. Note the remarkable growth defect of mutants $NmFic_{E102R}$, $NmFic_{E156R}$, and $NmFic_{E102R,E156R}$ ($NmFic_{mono}$). Raw data for 1% glucose and 2 mM IPTG are shown in Fig. S4 *A* and *B*. (*B–D*) Autoradiographs obtained after incubation of various NmFic mutants with 25 nM α-[^{32}P]ATP, 5 mM ATP, 10 mM $MgCl_2$, and purified GyrB43 (N-terminal GyrB 43-kDa fragment comprising the ATPase and transducer domains) at 35 °C. (*B*) In vitro adenylylation assay with 1 μM purified NmFic and 2.5 μM purified GyrB43 using an incubation time of 1 h. (*C*) Time course of autoadenylylation and GyrB43 adenylylation by $NmFic_{E156R}$ with or without ATP preincubation using 1 μM $NmFic_{E156R}$ and 5 μM GyrB43. Note that adenylylation of GyrB43 is delayed without preincubation (i.e., activation) of the Fic enzyme. (*D*) Same as in *C*, but for $NmFic_{E156R,Y183F}$. Note that this mutant is unable to adenylylate GyrB43. (*E–G*) Quantification of the radiograph band intensities shown in *B–D*. The assumed error of 5% is depicted as error bars. (*E*) Bar diagram representing NmFic autoadenylylation (light gray) and GyrB43 adenylylation (dark gray) catalyzed by the NmFic variants. (*F*) Time course of GyrB43 adenylylation with or without preincubation of $NmFic_{E156R}$. Also shown are the simulated curves obtained by global fitting of the kinetic model shown in Fig. 6*A* to the data. Resulting parameters are given in Table 2. (*G*) Time course of autoadenylylation of $NmFic_{E156R}$ and $NmFic_{E156R,Y183F}$. Fitting analogously to *F* yielded apparent $k_{cat,1}$ values of 2.4 * 10^{-3}·s^{-1} and 4.5 * 10^{-4}·s^{-1}, respectively. AU, arbitrary units.

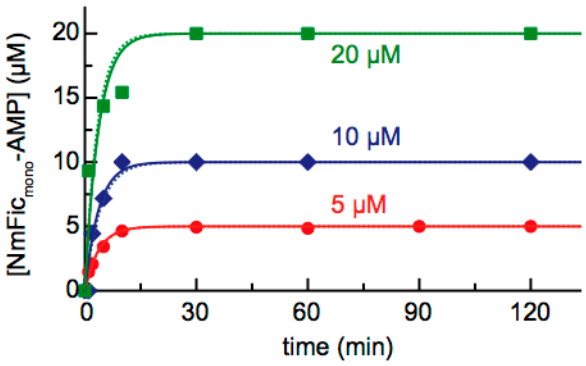


Fig. 4. *Cis*-autoadenylylation of monomeric NmFic. Progress curves of autoadenylylation of NmFic$_{mono}$ acquired at the indicated protein concentrations using 5 mM ATP and 10 mM $MgCl_2$. The data were fitted to a first-order reaction model. Dotted lines are the result of individual fitting, and plain lines are the result of global fitting of the three independent experiments (apparent $k_{cat,1}$ of 4.5 * 10^{-3}·s^{-1}; Table 1). Note that individual and global fitting are virtually identical.

obstructs the nucleotide binding site, and thereby autoinhibits Fic enzymes (8). Thus, to test for any functional role of Y183, we removed its hydroxyl group by introducing an additional Y183F mutation into the alleles of active (i.e., oligomerization-deficient), mutants. Strikingly, the mutation suppressed the growth defect phenotype as efficiently as mutation of the catalytic His (H107A) (Fig. 3*A*, *Right*), suggesting a crucial regulatory role for Y183.

The in vivo observation was faithfully mirrored in the autoradiograph obtained after incubation of the various purified NmFic mutants with radioactive ATP (Fig. 3*B*). In vitro GyrB43 adenylylation, which is efficiently catalyzed by the three interface mutants (Fig. 3*B*, lanes 4–6), is almost completely abolished in their respective Y183F variants (Fig. 3*B*, lanes 10–12).

The remaining autoadenylylation of the Y183F NmFic variants (Fig. 3*B*, *Right*), which is due to the modification of Y184, Y185, and Y188, indicates that enzyme function, per se, is not impaired. As expected, no target adenylylation is observed. Thus, (partial) modification of the additional Tyr residues does not significantly relieve autoinhibition. Still, to corroborate that the Y183F mutation had no unforeseen effect on the enzyme, we determined the high-resolution (0.99 Å) crystal structures of both NmFic$_{E156R}$ and NmFic$_{E156R,Y183F}$ (Fig. S4 *C* and *D*). Indeed, the structures are virtually identical, apart from the absence of the Y183 hydroxyl group in the double mutant.

To test whether the kinetics of autoadenylylation and target adenylylation would be of physiological relevance, we measured time courses of product (GyrB-AMP) formation for NmFic$_{E156R}$ (Fig. 3*C*). The apparent rate of NmFic$_{E156R}$ autoadenylylation derived from Fig. 3*G* agrees well with the one obtained for NmFic$_{mono}$ by DSF analysis (Table 2). Similarly, NmFic$_{E156R,Y183F}$ shows autoadenylylation, although with a considerably slower rate (Fig. 3 *D* and *G*). Such an effect of the Y183F mutation was also observed for NmFic$_{mono,Y183F}$ by MS (Fig. S6*B*). As mentioned above, the modification on residues Y184, Y185, and Y188 is probably not of functional relevance.

Under the used conditions, NmFic$_{E156R}$ fully converts GyrB43 to its adenylylated form within 1 h (Fig. 3 *C* and *F*), but with a pronounced lag phase. This delay is absent in the progress curve of preactivated Fic enzyme, which suggests again that only the adenylylated form is catalytically active. This interpretation was confirmed quantitatively by globally fitting the respective kinetic model (Fig. 6*A*, but without oligomerization) to the data (Fig. 3*F*). Upon setting $K_{m,GyrB}$ to the experimentally determined value of 39 μM, the fit yielded the turnover numbers of autoadenylylation ($k_{cat,1,eff}$) and target adenylylation ($k_{cat,2}$) given in Table 2. Notably, $k_{cat,1,eff}$ turned out to be about one order of magnitude slower that the corresponding rate, $k_{cat,1}$, as measured by DSF. Therefore, it can be concluded that the rate-determining step of Fic activation is the debinding of the modified segment from the active site (A^0→A).

Activity Profile of NmFic$_{wt}$ Is a Consequence of Autoactivation Combined with Oligomerization. We have thus shown that NmFic-catalyzed target adenylylation is controlled by enzyme tetramerization as well as *cis*-autoadenylylation. These results were obtained by analyses of mutants deficient in one or the other function. What, then, would be the combined effect in the WT enzyme? Fig. 6*B* shows the autoradiographs obtained after incubation of NmFic$_{wt}$ (at varying concentrations) with GyrB43 for various durations. At the standard enzyme concentration of 1 μM, weak NmFic$_{wt}$ autoadenylylation and no significant target adenylylation are observed for incubation times up to 8 h, consistent with our earlier result (Fig. 3*B*, first lane). Strikingly, strong GyrB43 adenylylation is observed at lower enzyme concentrations (250 nM and lower), with an abrupt transition from 250 to 500 nM. Because this transition occurs in a very similar range as the monomer-to-tetramer transition (Fig. 2*D*, *Bottom*), it most probably reflects the emergence of catalytically incompetent tetramers at higher concentrations.

Importantly, however, the catalytic incompetence of the tetramer alone is not sufficient to explain the data, because not only the tetramer/monomer ratio but also the absolute concentration of active monomers will increase with total concentration (Fig. S2*H*). To rationalize the observed effects quantitatively, we set up a kinetic scheme including autoactivation and inactivation by oligomerization as shown in Fig. 6*A* (also Fig. S3*C*). In this model, autoinhibited native NmFic (N) is in equilibrium with states N′ (α_{inh} unfolded) and N″ (modifiable Y183 positioned in the active site). The latter state gets autoadenylylated at Y183 in a first-order reaction to yield A^0. Finally, unbinding of the modified C terminus will result in a monomer state A that is predicted to be

Table 2. Enzymatic characterization of NmFic-catalyzed autoadenylylation and target adenylylation

Reaction	Parameter	k_{cat}, s^{-1}	K_m, μM	Method
Autoadenylylation				
(N,N′,N″) → (A^0,A)	$k_{cat,1}$	(4.5 ± 0.53) 10^{-3}	n/a	DSF (Fig. 4)
(N,N′,N″) → (A^0,A)	$k_{cat,1}$	(2.3 ± 0.08) 10^{-3}	n/a	Adenylylation assay (Fig. 3*G*)
(N,N′,N″) → A	$k_{cat,1,eff}$	(2.1 ± 0.25) 10^{-4}	n/a	Adenylylation assay (Fig. 3*F*)
Target adenylylation				
A + G → A + G-AMP	$k_{cat,2}$	0.18 ± 0.01		Adenylylation assay (Fig. 3*F*)
A + G → A + G-AMP	$K_{m,G}$		39.7 ± 3.6	Adenylylation assay (not shown)

All parameters are apparent parameters referring to an ATP concentration of 5 mM. The K_d of ATP binding to native monomeric NmFic has been measured by NMR to 7.8 mM (Fig. S7*C*). The value of $k_{cat,1}$ was determined by DSF and adenylylation assay for NmFic$_{mono}$ and NmFic$_{E156R}$, respectively. The values agree satisfactorily within a factor of 2. n/a, not applicable.

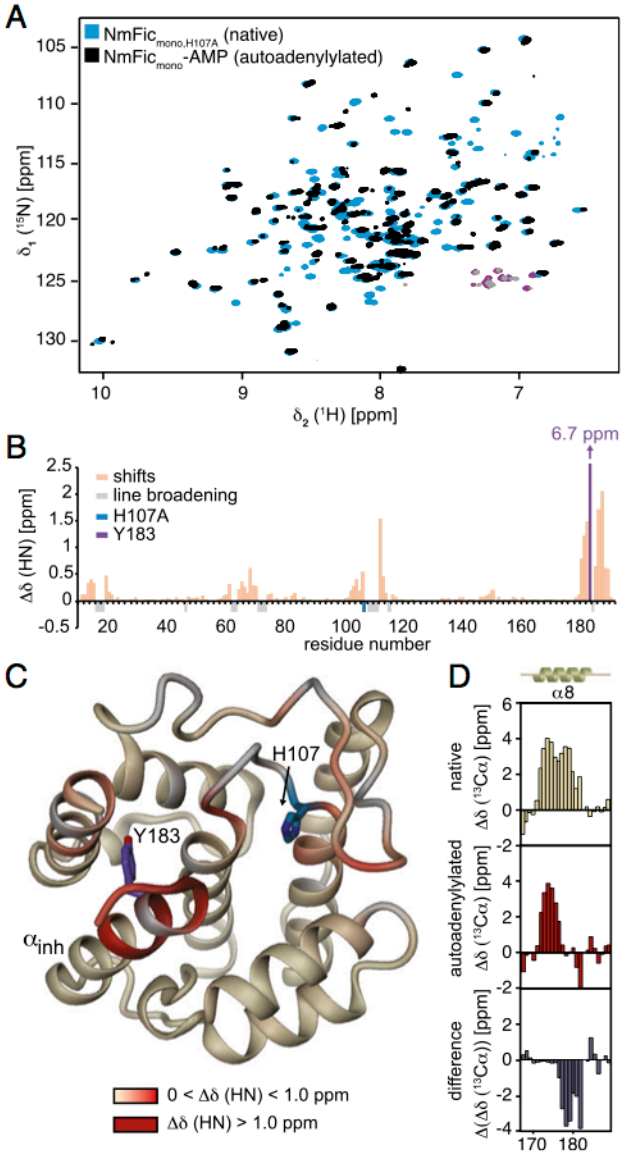


Fig. 5. Structural perturbations of NmFic upon autoadenylylation. (*A*) Overlay of 2D [^{15}N,^{1}H]-HSQC spectra of 500 μM [*U*-^{13}C,^{15}N]-labeled native inactive monomeric NmFic (NmFic$_{mono,H107A}$, cyan) and of 140 μM [*U*-^{13}C,^{15}N]-labeled autoadenylylated monomeric NmFic (NmFic$_{mono}$-AMP, black). (*B*) Amide-group chemical shift difference, Δδ(HN), between NmFic$_{mono,H107A}$ and NmFic$_{mono}$-AMP plotted as a function of NmFic residue number. Residues exhibiting significant line broadening in the adenylylated form are indicated in gray as negative bars. Values at positions 107 (H/A) and 183 (Y) are highlighted in blue and purple, respectively. (*C*) Amide-group chemical shift differences mapped on the structure of NmFic (PDB ID code 3S6A) using the indicated color code. The largest chemical shift differences cluster in helices α1 and α8 (α_{inh}) and in the flap region. (*D*) Secondary ^{13}Cα chemical shifts for residues 167–190 of native NmFic$_{mono,H107A}$ (*Top*), autoadenylylated NmFic$_{mono}$-AMP (*Middle*), and their difference (*Bottom*) relative to the random coil values. A 1-2-1 smoothing function was applied to the raw data. Consecutive stretches with positive and negative values indicate α-helical and β-strand secondary structure, respectively. The extent of α_{inh} as inferred from the crystal structure of the native NmFic is indicated on top. The full residue range is shown in Fig. S7*B*.

competent for target adenylylation. Note, that ATP binding will shift the A^0⇔A equilibrium toward the competent A state. Because monomers are in fast oligomerization equilibrium, active A monomers will repartition into the tetramer that acts as a reservoir, and will therefore be partly sequestered to the inactive oligomeric state.

We reasoned that a contribution of the unfolded autoadenylylated C-terminal segment to the interface might affect the oligomerization affinity. Indeed, in presence of 5 mM ATP, the dissociation constant ($K_{d,2'}$) of A was measured by AUC-SV to be fivefold lower than the corresponding constant of N (Fig. 2*D* and Table 1).

Using the experimentally determined parameters of the kinetic model (Tables 1 and 2), we then simulated GyrB43-AMP production as function of enzyme concentration and incubation time by numeric integration of the respective differential equations (Fig. 6 *C* and *D*). Indeed, as observed experimentally, target adenylylation efficiency drops with enzyme concentration above a certain threshold (Fig. 6*C*). Furthermore, the simulated progress curves (Fig. 6*D*) reproduce the lag phase observed at a high enzyme concentration. Taken together, the salient features of intrinsic NmFic regulation are faithfully captured by the proposed kinetic model.

Discussion

The physiological importance of AMP-transferases with FIC fold has escaped attention for a long time, because their activity is kept in check by intra- or intermolecular active site obstruction (8). In this study, we focused on class III Fic proteins that carry the inhibitory α_{inh} at the C terminus and are composed of a FIC domain only. We reveal that in addition to the previously reported autoinhibition (8), two further mechanisms inversely regulate adenylylation activity, namely, tetramerization and *cis*-autoadenylylation. This combination results in a complex autoregulatory mechanism.

NmFic forms a tetramer with the involved interfaces largely conserved or coevolved among class III members, suggesting conservation of the tetrameric arrangement in this class. Disruption of either interface was achieved by site-directed mutagenesis, thus verifying their role both in the crystalline state and the solubilized state. Due to the cooperative nature of tetramerization, a sharp monomer–tetramer transition occurs in the presence of ATP, at the rather low NmFic concentration of 50 nM (Fig. 2*D* and Fig. S2*G*). Clearly, the tetramer state is of physiological relevance, because expression of interface disruption mutants, but not of NmFic$_{wt}$, resulted in impaired growth of the expressing *E. coli* strain. Thus, it can be inferred that only monomeric NmFic is able to exert the growth retardation effect, which can be attributed to GyrB adenylylation (5). Most probably, the tetramer is catalytically noncompetent, because the binding site for the segment flanking the modifiable side chain [target dock (10)] is partially occluded. Furthermore, the strict conservation of oligomerization interface 1 (as opposed to interface 2, which shows covariation) may point to its involvement in mediating the contact between (monomeric) NmFic and the target. In fact, the surface-exposed area of the tetramer is highly variable, and therefore probably not involved in target recognition (Movie S1).

Fic automodification has been reported repeatedly (6, 8–15), but its role has remained unclear. In fact, nonspecific modification due to the high in vitro Fic concentrations used seemed possible. Here, we have shown that NmFic autoadenylylation crucially controls enzyme activity by relieving the autoinhibitory effect of the α_{inh}, as demonstrated by the Tyr-to-Phe mutation (Y183F) that fully suppressed the activating effect of the interface disruption mutants (Fig. 3*A*) and rendered the enzyme incompetent for target adenylylation (Fig. 3*B*). Biochemical and structural analyses confirmed that the conservative Y183F mutation does not disrupt the FIC fold (Fig. 3 and Fig. S4 *C* and *D*).

Importantly, the automodification occurs *in cis* with a reaction velocity independent of the total enzyme concentration. Indeed, structural modeling shows that upon unfolding of the α_{inh}, the modifiable Tyr can reach the active site of the same molecule (Movie S2). Probably, the segment flanking the Tyr residue would engage in β-strand interaction with the flap, resulting in

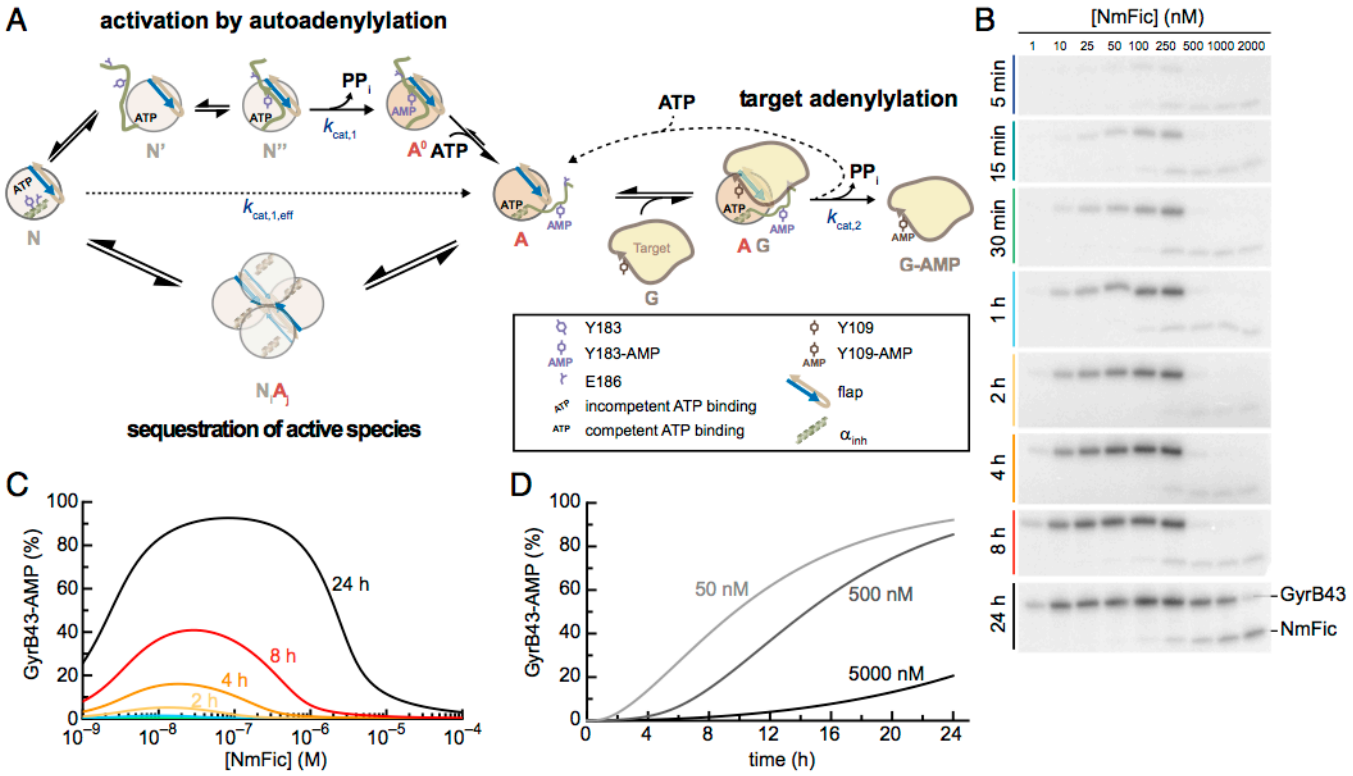


Fig. 6. Concentration and time dependence of $NmFic_{wt}$-catalyzed target adenylylation. (*A*) Simplified model of the regulatory mechanism of class III Fic proteins. (*Top Left*) Native monomeric Fic protein (N) is activated by autoadenylylation to yield A. This reaction involves unfolding of the α_{inh} (N′), binding of the modifiable Tyr Y183 of the unfolded segment to the flap (N″), and subsequent autoadenylylation of Y183 (A^0). (*Bottom*) Both types of monomers (N and A) are in dynamic monomer/tetramer equilibrium to form homotetramers or heterotetramers N_iA_j, with $i + j = 4$. The tetramer acts as a buffer sequestering the active monomers. (*Right*) Only the remaining fraction of free (monomeric) A molecules is competent for target adenylylation. The complete model with all thermodynamic and kinetic parameters is shown in Fig. S3C. (*B*) Autoradiographs showing $NmFic_{wt}$ and GyrB43 adenylylation. The same experimental setup as in Fig. 3C was used but was carried out at various $NmFic_{wt}$ concentrations (1 nM to 2 μM, as indicated). Reactions were stopped after the indicated incubation times. (*C*) Simulation of GyrB adenylylation as a function of total $NmFic_{wt}$ concentration based on the kinetic model shown in *A* and Fig. S3C, with parameters set to their measured values (Tables 1 and 2). (*D*) Replot of the data shown in *C* as a function of incubation time for representative $NmFic_{wt}$ concentrations.

correct registration of the modifiable tyrosyl side-chain within the active site. This binding mode would be analogous to the one seen in the complex structures of $IbpA_{Fic2}$/cdc42 (6) and of noncognate peptide/Fic proteins (9). Because the flap is partially buried at the center of the tetramer, the structural model also explains why the tetramer would be incompetent for automodification. Thus, in the presence of ATP, native monomeric NmFic gets efficiently converted to its active, covalently modified form with a rather fast "in-built" first-order rate constant (Table 2, equivalent to a $t_{1/2}$ of about 150 s). As such, NmFic and, most likely, Fic proteins of class III in general constitute self-contained proteinaceous timers. Whether such an autoactivation mechanism also pertains to Fic proteins of class I and II remains to be seen, but conserved Tyr acceptor candidates can be identified in the N-terminal and central part of the fold, respectively. Noteworthy, *cis*-automodification has also been described for a few kinases as an activation mechanism (18).

The combination of the two coupled and counteracting effects, automodification and tetramerization, results in a quite unusual time and concentration dependence of target adenylylation, as captured in vitro by autoradiography (Fig. 6*B*) and reproduced *in silico* by respective simulations (Fig. 6 *C* and *D*). In essence (Fig. 6*A*), monomeric native Fic molecules (N) get converted with the intrinsic first-order rate to the autoadenylylated, active form (A). However, because Fic monomers are in fast equilibrium with the tetrameric state, they get redistributed such that the A/N ratio will be equal or larger in the tetrameric state. As a consequence, the increase in the absolute number of active monomers (A) will be dampened. Obviously, this buffering effect is most pronounced at high tetramer/monomer ratios (i.e., high enzyme concentrations) and will diminish when the autoadenylylation reaction reaches completion.

The comprehensive characterization of the complex autoregulatory mechanism of class III Fic proteins presented here sets the stage for future investigations of its physiological consequences. It has to be inferred, however, that the intrinsic regulatory mechanism may be further modulated by external factors, such as antagonistic phosphodiesterases [e.g., the de-AMPylase SidD of *Legionella* (19, 20)], specific proteases, or small ligands. Furthermore, Fic biosynthesis and degradation, as well as dilution effects during cell growth and division, will profoundly affect the time profile of the Fic pool activity. Still, solely based on the inherent mechanism, interesting consequences can be envisaged.

The nonmonotonic dependence of AMP-transferase activity on enzyme concentration would render the enzyme exquisitely sensitive to variation of its concentration in the bacterial cell. A twofold drop of NmFic concentration in the right range (e.g., due to cellular growth and division) would drastically increase catalytic activity, and thus GyrB inactivation. In addition, the considerable delay in NmFic-mediated GyrB adenylylation may be important in vivo to ensure that the growth-retarding effect of the target modification kicks in only under certain conditions, such as starvation, when synthesis of fresh NmFic has come to a halt. Molecular timers have been found before, such as the master regulator Spo0A of *Bacillus* that controls, upon phosphorylation, the switch from competence to sporulation via a well-characterized genetic circuit (21, 22) induced by prolonged times of nutriment

starvation. Class III Fic proteins may constitute a new family of molecular timers that are, in contrast to the master regulator Spo0A of *Bacillus*, fully autonomous and not relying on any feedback exerted by other components or a genetic circuit. This hypothesis will have to be tested in future studies.

Materials and Methods

Detailed information on cloning, expression, purification, crystallization, data collection, and structure determination by X-ray crystallography or NMR is provided in *SI Materials and Methods*. Plasmids were constructed as described previously (8, 23, 24) (Tables S2 and S3). Proteins were expressed and purified as described (8, 25). Toxicity tests and adenylylation assays were performed according to the protocols given by Harms et al. (5) and Goepfert et al. (9).

The oligomeric state of NmFic was determined by SEC-MALLS and AUC-SV at varying concentrations. DSF (26) was used to monitor in vitro autoadenylylation of NmFic. For simulations and fitting of functional data to the kinetic model, the programs ProFit 6.2.14 (QuantumSoft) and Complex Pathway Simulator (COPASI) 4.14 (27) were used.

Detailed information on the various experimental procedures is provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We acknowledge Gerd Pluschke (Swiss Tropical and Public Health Institute) for kindly providing the genomic DNA of *N. meningitidis* 2808. We thank Tillmann Heinisch for help with MS analysis. We thank the staff of beamlines X06DA and X06SA of the Swiss Light Source for excellent support. This work was supported by the European Research Council Advanced Investigator Grant (ERC-2013-AdG) FICModFun 340330 (to C.D.) and Swiss National Science Foundation Grants 3100-132979 (to C.D.) and 31003A-138414 (to T. Schirmer).

1. Roy CR, Mukherjee S (2009) Bacterial FIC Proteins AMP Up Infection. *Sci Signal* 2(62): pe14.
2. Garcia-Pino A, Zenkin N, Loris R (2014) The many faces of Fic: Structural and functional aspects of Fic enzymes. *Trends Biochem Sci* 39(3):121–129.
3. Yarbrough ML, et al. (2009) AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* 323(5911):269–272.
4. Worby CA, et al. (2009) The fic domain: Regulation of cell signaling by adenylylation. *Mol Cell* 34(1):93–103.
5. Harms A, et al. (2015) Adenylylation of Gyrase and Topo IV by FicT Toxins Disrupts Bacterial DNA Topology. *Cell Reports* 12(9):1497–1507.
6. Xiao J, Worby CA, Mattoo S, Sankaran B, Dixon JE (2010) Structural basis of Fic-mediated adenylylation. *Nat Struct Mol Biol* 17(8):1004–1010.
7. Luong P, et al. (2010) Kinetic and structural insights into the mechanism of AMPylation by VopS Fic domain. *J Biol Chem* 285(26):20155–20163.
8. Engel P, et al. (2012) Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature* 482(7383):107–110.
9. Goepfert A, Stanger FV, Dehio C, Schirmer T (2013) Conserved inhibitory mechanism and competent ATP binding mode for adenylyltransferases with Fic fold. *PLoS One* 8(5):e64901.
10. Palanivelu DV, et al. (2011) Fic domain-catalyzed adenylylation: insight provided by the structural analysis of the type IV secretion system effector BepA. *Protein Sci* 20(3): 492–499.
11. Kinch LN, Yarbrough ML, Orth K, Grishin NV (2009) Fido, a novel AMPylation domain common to fic, doc, and AvrB. *PLoS One* 4(6):e5818.
12. Pieles K, Glatter T, Harms A, Schmidt A, Dehio C (2014) An experimental strategy for the identification of AMPylation targets from complex protein samples. *Proteomics* 14(9):1048–1052.
13. Feng F, et al. (2012) A *Xanthomonas* uridine 5′-monophosphate transferase inhibits plant immune kinases. *Nature* 485(7396):114–118.
14. Goody PR, et al. (2012) Reversible phosphocholination of Rab proteins by *Legionella pneumophila* effector proteins. *EMBO J* 31(7):1774–1784.
15. Bunney TD, et al. (2014) Crystal structure of the human, FIC-domain containing protein HYPE and implications for its functions. *Structure* 22(12):1831–1843.
16. Goepfert A, Harms A, Schirmer T, Dehio C (2013) Type II toxin-antitoxin loci: The fic family. *Prokaryotic Toxin-Antitoxins* (Springer-Verlag, Berlin Heidelberg), pp 177–187.
17. Khater S, Mohanty D (2015) In silico identification of AMPylating enzymes and study of their divergent evolution. *Sci Rep* 5:10804.
18. Lochhead PA (2009) Protein kinase activation loop autophosphorylation *in cis*: Overcoming a Catch-22 situation. *Sci Signal* 2(54):pe4.
19. Tan Y, Luo Z-Q (2011) *Legionella pneumophila* SidD is a deAMPylase that modifies Rab1. *Nature* 475(7357):506–509.
20. Chen Y, et al. (2013) Structural basis for Rab1 de-AMPylation by the *Legionella pneumophila* effector SidD. *PLoS Pathog* 9(5):e1003382.
21. Levine JH, Elowitz MB (2014) Polyphasic feedback enables tunable cellular timers. *Curr Biol* 24(20):R994–R995.
22. Levine JH, Fontes ME, Dworkin J, Elowitz MB (2012) Pulsed feedback defers cellular differentiation. *PLoS Biol* 10(1):e1001252.
23. Gotfredsen M, Gerdes K (1998) The *Escherichia coli* relBE genes belong to a new toxin-antitoxin gene family. *Mol Microbiol* 29(4):1065–1076.
24. Zheng L, Baumann U, Reymond J-L (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* 32(14):e115.
25. Stanger FV, Dehio C, Schirmer T (2014) Structure of the N-terminal Gyrase B fragment in complex with ADP·Pi reveals rigid-body motion induced by ATP hydrolysis. *PLoS One* 9(9):e107289.
26. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2(9):2212–2221.
27. Hoops S, et al. (2006) COPASI–A COmplex PAthway SImulator. *Bioinformatics* 22(24): 3067–3074.
28. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529–W533.
29. Ovchinnikov S, Kamisetty H, Baker D (2014) Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* 3:e02030.
30. Reckel S, et al. (2010) Strategies for the cell-free expression of membrane proteins. *Methods Mol Biol* 607(Chapter 16):187–212.
31. Spirin AS, Baranov VI, Ryabova LA, Ovodov SY, Alakhov YB (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* 242(4882):1162–1164.
32. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
33. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* 69(Pt 7):1204–1214.
34. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.
35. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
36. Murshudov GN, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):355–367.
37. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
38. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
39. Glaser F, et al. (2003) ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 19(1):163–164.
40. Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94(23):12366–12371.
41. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34(2):93–158.
42. Zhu G, Xia Y, Nicholson LK, Sze KH (2000) Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson* 143(2):423–426.
43. Lee D, Hilty C, Wider G, Wüthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson* 178(1):72–76.
44. Güntert P, Dötsch V, Wider G, Wüthrich K (1992) Processing of multi-dimensional NMR data with the new software PROSA. *J Biomol NMR* 2(6):619–629.
45. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
46. Jaravine V, Ibraghimov I, Orekhov VY (2006) Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. *Nat Methods* 3(8):605–607.
47. Bartels C, Xia T-H, Billeter M, Güntert P, Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 6(1):1–10.
48. Kjaergaard M, Poulsen FM (2011) Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* 50(2):157–165.
49. Sundriyal A, et al. (2014) Inherent regulation of EAL domain-catalyzed hydrolysis of second messenger cyclic di-GMP. *J Biol Chem* 289(10):6978–6990.
50. Kraft D (1994) Algorithm 733: TOMP–Fortran modules for optimal control calculations. *ACM Trans Math Softw* 20(3):262–281.
51. van der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng* 13(2):22–30.
52. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606–1619.
53. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta* 1751(2):119–139.

Supporting Information

Stanger et al. 10.1073/pnas.1516930113

SI Materials and Methods

Plasmid Construction. pRSF-Duet1 derivatives were cloned as described previously (8) and used for overexpression of proteins for purification. For phenotypic analyses, *nmfic* genes were cloned into the single-copy vector pNDM220 (23) under the control of a particularly tight *Plac*. Site-directed mutagenesis was performed following the protocol described by Zheng et al. (24). A list of used plasmids and a list of respective oligonucleotides can be found in Tables S2 and S3.

Expression, Isotope Labeling, and Purification. Plasmids pFVS0015, pFVS0051, pFVS0081, pFVS0109, pFVS0135, pFVS0137, pFVS0138, pFVS0143, pFVS0144, pFVS0167, pFVS0319, pFVS0323, pFVS0326, pFVS0329, and pFVS0332 were transformed into *E. coli* BL21 (λDE3) cells. Plasmids pFVS0059, pFVS0125, pFVS0126, pFVS0134, and pFVS0145 were transformed into *E. coli* BL21 AI cells. NmFic and GyrB43 (residues 1–392) proteins were expressed and purified as described previously by Engel et al. (8) and Stanger et al. (25), respectively. For NMR analysis, [*U*-^{13}C,^{15}N]-labeled $NmFic_{mono,H107A}$ was obtained by growing the expression cells in M9 minimal media supplemented with ($^{15}NH_4$)Cl (Cambridge Isotope Labs) and D-[^{13}C]-glucose (Sigma–Aldrich) using standard NmFic expression and purification conditions.

Cell-Free Expression of $NmFic_{mono}$. The *E. coli* S30 cell extract was prepared following published protocols (30). In short, harvested *E. coli* A19 cells were washed three times with S30-buffer A [10 mM Tris·acetate (pH 8.2), 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 6 mM β-mercaptoethanol]. After resuspension in S30-buffer B [10 mM Tris·acetate (pH 8.2), 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 1 mM DTT, 0.1 mM PMSF], cells were lysed using a French press and centrifuged twice at 30,000 × *g* at 4 °C. To remove endogenous mRNA, the extract was incubated for 45 min at 42 °C and then dialyzed at 4 °C overnight against S30-buffer C [10 mM Tris·acetate (pH 8.2), 14 mM $Mg(OAc)_2$, 0.6 mM KCl, 0.5 mM DTT] with a 12–14 kDa molecular mass cutoff. To remove the remaining cellular debris, the cell extract was centrifuged at 30,000 × *g* at 4 °C, aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C until use. For preparative scale protein production of autoadenylylated NmFic in continuous cell-free mode (31), a maxi-continuous exchange cell free (CECF) reactor built in-house (30) was used with a 0.5- to 3-mL Slide-A-Lyzer dialysis cassette with a 10-kDa molecular mass cutoff (Pierce) and an optimal feeding/reaction mixture volume ratio of 14.5. Final concentrations of critical components, such as PEG 8000 (2%), Mg^{2+} (16 mM per reaction mixture), and T7-RNA polymerase (9.5 mg/mL of reaction mixture), concentrations were optimized for protein expression. Both [*U*-^{15}N], and [*U*-^{13}C,^{15}N] complete amino acid labeling was achieved with a cell-free 20-aa mix (Cambridge Isotope Labs) at a final concentration of 0.8 mM in the reaction mixture and 0.4 mM in the feeding chamber. Protein expression was left overnight in a thermoshaker at 120 rpm at 30 °C in a VWR Incubating Mini Shaker. Expression of the protein yielded between 0.8 and 1.1 mg of protein per milliliter of reaction mixture.

Toxicity Tests. *E. coli* strain AB472 (8), a derivative of *E. coli* K-12 MG1655, was transformed with pNDM220 derivative plasmids using TSS (transformation and storage solution) transformation and handled in LB liquid medium containing appropriate antibiotics, as well as 1% (wt/vol) D-glucose to suppress basal expression before the inoculation of experimental cultures. Bacterial cultures were serially diluted and spotted on LB agar plates containing 0, 50 μM, 100 μM, 250 μM, 500 μM, 1,000 μM, or 2,000 μM IPTG and 30 μg/mL ampicillin, which were then incubated at 37 °C. After 24 h of incubation, the numbers of colony-forming units were determined and plotted against inducer concentrations.

In Vitro Adenylylation Assay. Adenylylation assays were performed using cleared cell lysate of ectopically expressing *E. coli* as described previously (8) [except using BL21 instead of BL21 (λDE3) cells for overexpression] or using purified proteins as described by Goepfert et al. (9). Adenylylation activity of NmFic was assessed by incubating NmFic with 5 mM ATP, 10 μCi [α-^{32}P]-ATP (Hartmann Analytic), 25 mM $MgCl_2$, and GyrB43. In addition, 200 μM novobiocin was added to inhibit the ATPase activity of GyrB43. Adenylylation assays with purified NmFic and GyrB43 were performed at 35 °C, and adenylylation assays using cell lysates were performed at 30 °C. For data evaluation, the intensity of each band on the autoradiographs was quantified using ImageJ (rsbweb.nih.gov/ij/download.html).

Crystallization. $NmFic_{E156R}$ and $NmFic_{E156R,Y183F}$ (10 mg/mL and 8 mg/mL, respectively) crystallized in 500 μL of precipitant solution containing 10 mM Tris (pH 7.8) and 100 mM NaCl using the batch crystallization method at 4 °C. All other crystals were obtained using the sitting-drop vapor-diffusion method by mixing 0.2 μL of protein solution with 0.2 μL of reservoir solution equilibrating against a reservoir of 80 μL at 20 °C. $NmFic_{E102R}$ crystallized after 1 month in 8% (vol/vol) Tacsimate (pH 6.0) (Hampton Research), 20% (wt/vol) PEG 3350, and $NmFic_{E102R,E156R}$ crystallized after 10 days in 8% (vol/vol) Tacsimate (pH 5.0) and 30% (wt/vol) PEG 1500. For data collection, crystals were cryoprotected by soaking into a reservoir solution supplemented with 15–30% glycerol and subsequently flash-frozen in liquid nitrogen.

Structure Determination. X-ray data were collected at the Swiss Light Source (Villigen, Switzerland) on beamline X06SA (PXI) or X06DA (PXIII) at 100 K (Table S1). Diffraction data were indexed and integrated using XDS (32) and subsequently merged and scaled using XSCALE (32) or AIMLESS (33). Data collection and processing statistics are given in Table S1. Structures were determined by molecular replacement [PHASER program (34)] with an $NmFic_{wt}$ monomer [PDB ID code 3S6A (8)] as a search model. Several rounds of model building and refinement were performed using Coot (35) and REFMAC5 (36) or phenix.refine (37). Five percent of the data were excluded from refinement and used for cross-validation. The geometry of the final model was assessed using MolProbity (38) showing >99% of the residues in the core and allowed regions of the Ramachandran plot. Final refinement yielded NmFic models with reasonable R_{work}/R_{free} values for their respective resolution ranges. Refinement statistics are summarized in Table S1. Figures were prepared with DINO (A. Philippsen; www.dino3d.org).

Conservation and Covariance Analyses of Class III Fic Proteins. Conservation scores for each position of the class III Fic protein NmFic were calculated by using the ConSurf server with default parameters (39). The top 200 sequences (46% identity) were selected from a BLAST search against the Uniref90 database and manually curated to select only class III Fic proteins (three sequences that do not contain the C-terminal α_{inh} were removed). For covariance analysis, the NmFic sequence was submitted to the Gremlin server for coevolution analysis (29).

NMR Spectroscopy. NMR experiments were performed on Bruker 600-, 700-, and 900-MHz spectrometers running Topspin3.0 and equipped with cryogenically cooled triple-resonance probes. All experiments were performed in NMR buffer [25 mM MES, 150 mM NaCl (pH 6.5)] at 25 °C. For the sequence-specific backbone resonance assignments of NmFic, the following experiments were recorded: 2D [^{15}N,^{1}H]-TROSY-HSQC (40), 3D HNCA (41), 3D HNCACB (41), 3D HNCO (41), and 3D HN(CA)CO (41). To confirm the chemical shift changes observed upon adenylylation within 2D [^{15}N,^{1}H]-TROSY-HSQC, the following 3D experiments were recorded in a nonuniformly sampled manner with a random sampling schedule generated by Topspin3.0 (Bruker Biospin) to confirm the assignments: 3D HNCA (25% of the full time-domain grid) and 3D HNCO (9%).

For the analysis of the dynamic properties of NmFic, the following experiments were measured: ^{15}N{^{1}H}-NOE (42), T_1(^{15}N) (42), and TROSY for rotational correlation times (TRACT) (43). NMR data were processed with PROSA (44), NMRPipe (45), and mddNMR2.4 (46), and were analyzed with CARA, XEASY (47), and Topspin3.0 (Bruker Biospin). Nonlinear least square fits of relaxation data were done with MATLAB (MathWorks). R_2(^{15}N) values were derived from R_α(^{15}N) and R_β(^{15}N). Error bars for R_1(^{15}N), R_α(^{15}N), and R_β(^{15}N) were calculated by a statistical bootstrapping scheme, and error bars for the ^{15}N{^{1}H}-NOE were calculated from the spectral noise. Secondary chemical shifts were calculated relative to the random coil values of Kjaergaard and Poulsen (48).

The sequence-specific resonance assignment for the apo NmFic has been submitted to the Biological Magnetic Resonance Data Bank under accession code 26607. The chemical shift changes of the amide moiety upon ATP binding were fitted by nonlinear regression analysis to Eq. **S1** by using standard software. The term Δ_{obs} corresponds to the chemical shift difference at a given titration point, and Δ_{max} is the maximal chemical shift difference at the last titration point:

$$\Delta_{obs} = \Delta_{max}\frac{(K_d + [ATP]_0 + [NmFic]_0) - \sqrt{(K_d + [ATP]_0 + [NmFic]_0)^2 - (4[NmFic]_0[ATP]_0)}}{2[NmFic]_0}. \quad [S1]$$

SEC-MALLS Analysis. SEC-MALLS was used for the determination of the NmFic oligomerization dissociation constants by measuring apparent mass values (MW^{app}) at various loading concentrations. UV absorption and differential Refractive Index (dRI) values were used to derive the eluting concentration. A similar experimental setup has been described previously; details are provided in the study by Sundriyal et al. (49). The column was preequilibrated with 10 mM Tris (pH 7.6) and 100 mM NaCl at 4 °C.

Based on the respective mass action law equations (Fig. S3*B*), the dependence of MW^{app} as a function of protein concentration was modeled. Assuming that both dimer interfaces (Fig. 1) are independent, the thermodynamic tetramerization scheme (Fig. 3*A*) contains only two independent parameters, the dissociation constants of the respective interfaces ($K_{d,1}$ and $K_{d,2}$). For data fitting, the reaction extents were used as state variables and inequality constraints were imposed to ensure the correctness of the results. Optimization of reaction extent values was accomplished using the sequential least squares programming method (50) as implemented in the SciPy library (51). For mutants NmFic$_{E102R}$ and NmFic$_{E156R}$, the experimental data were fitted to a simple monomer/dimer equilibrium. All equilibria were modeled as fast-exchange processes with respect to the time scale of chromatographic separation, because all measurements yielded only a single MALLS peak. Theoretical MW^{app} values were calculated from mass concentrations ($c_{m,i}$) and molecular weights (MW_i) of populated species (monomer and dimer for the NmFic$_{E102R}$ and NmFic$_{E156R}$ mutants; monomer, dimer, trimer, and tetramer for NmFic$_{wt}$) as

$$MW^{app} = \frac{\sum_i c_{m,i} MW_i}{\sum_i c_{m,i}}. \quad [S2]$$

All routines were implemented in the Python language using NumPy and SciPy numerical libraries (51).

AUC Analysis. NmFic$_{wt}$ (40 μM) was labeled with Dylight 488 NHS ester using the protocol published by the supplier (Thermo Scientific). A dilution series of unlabeled NmFic$_{wt}$ (10.0 μM, 3.33 μM, 1.11 μM, 0.370 μM, 0.123 μM, 41.0 nM, and 0 nM) supplemented with 40 nM fluorescein-labeled NmFic$_{wt}$ was prepared in the presence and absence of 5 mM ATP. SV experiments were performed on 400-μL samples in double-sector charcoal-Epon centerpieces at 42,000 rpm and 6 °C using a Beckman XL-I analytical ultracentrifuge with the Beckman An-50 Ti rotor. Sedimentation was monitored in 1,000 scans during an overnight run using the Aviv AU-FDS fluorescence detection system. The buffer density (1.0022 g/mL) and viscosity (0.01474 poise) were measured at 6 °C using an Anton Paar DMA 4500M density meter and AMVn viscometer. Similarly, a dilution series of unlabeled autoadenylylated NmFic$_{wt}$-AMP (1.25 μM, 625 nM, 312.5 nM, 156 nM, 78 nM, and 0 nM) supplemented with 40 nM fluorescein-labeled autoadenylylated NmFic$_{wt}$-AMP was prepared in the presence of 5 mM ATP. SV experiments were performed using a Beckman XL-I analytical ultracentrifuge with the An-60 Ti rotor. The SV data were fitted to a diffusion-deconvoluted sedimentation coefficient distribution [$c(s)$] using the software Sedfit (52). A sedimentation coefficient range of 1–15 S was used with a resolution of 200 points. Radially-invariant (RI) noise and time-invariant (TI) noise, baseline, and meniscus position were all fitted. In a first round of fitting, the frictional ratio f/f_0 was fitted for each sample. The data were then refitted using an average value for f/f_0 (1.11) calculated from the free-fitted values. The $c(s)$ distributions were overlaid and normalized by maximum signal intensity in the software GUSSI (Chad Brautigam, University of Texas Southwestern Medical Center, Dallas, TX), and weight-averaged s_w isotherms were calculated from the $c(s)$ distributions over the range 1–5 S. The fraction of monomeric (x_M) and tetrameric (x_O) NmFic species (with sedimentation coefficients of 1.62 S and 3.95 S, respectively) as a function of protein concentration were derived from the relative areas of the respective peaks. Finally, the x_M and x_O distributions were fitted to the kinetic scheme to yield the dimerization constants (as discussed above).

CD Analysis. A time series of CD spectra was recorded at 35 °C in a quartz-suprasil cuvette with a path length of 0.2 cm using the wavelength range 190–270 nm (26.67 nm·min^{-1}) on an Applied Photophysics Chirascan Plus CD spectrophotometer. A measurement was recorded every 5 min. The autoadenylylation

reaction was started in the cuvette (t_0) by addition of 1 mM $MgCl_2$ to a solution containing 20 μM $NmFic_{mono}$, 0.5 mM ATP in 10 mM Tris (pH 7.8), and 125 mM NaCl. The α-helical content was approximated (52) by dividing the mean residue ellipticity at 222 nm by –30,000 deg·cm^2·dmol^{-1}.

DSF. Autoadenylylation was followed in vitro by dye-based thermal denaturation experiments (26) performed on a Qiagen Rotor-gene Q real-time PCR cycler. Experiments were carried out in 200-μL thin-wall PCR tubes (Thermo Scientific) with 50 μL of sample containing the indicated concentration of NmFic and ATP, 10 mM $MgCl_2$, and 5× SyPro Orange dye (Invitrogen) in 10 mM Tris (pH 7.8) and 100 mM NaCl, except when stated differently. For time-course measurements, reactions were stopped at the indicated times by addition of 50 mM EDTA to an aliquot of the reaction. Temperature was increased by a 0.5 °C increment every 30 s between 20 °C and 90 °C. Fluorescence intensity (F) thermal denaturation curves and their first derivatives (dF/dT) were generated using Rotorgene software (Qiagen) and exported as text files. The first derivatives were fitted as the sum of Gaussian peaks using ProFit (QuantumSoft). The fraction of native and autoadenylylated NmFic was derived from the relative areas of the fitted peaks.

MS Analysis. Samples containing 5 μM $NmFic_{mono}$ (or 5 μM $NmFic_{mono,Y183F}$), 5 μM $NmFic_{mono,H107A}$, 1 mM ATP, and 10 mM $MgCl_2$ were incubated for 90 min at 35 °C in a reaction volume of 500 μL. Free ATP was removed by Ni^{2+}-affinity chromatography using a 1-mL HisTrap column (GE Healthcare). Protein samples were concentrated and subsequently desalted using 7-kDa molecular mass cutoff Zeba Spin Desalting Columns (Thermo Scientific). Before direct injection on a Bruker Daltonics micro-TOF mass spectrometer, protein samples were diluted to 5 μM in a buffer containing 50% (vol/vol) acetonitrile and 0.2% (vol/vol) formic acid. Thirty microliters was injected, and the mass spectra were analyzed by maximum entropy deconvolution.

Kinetic Fitting and Simulations. Enzymatic reactions were fitted to standard first-order and second-order reaction kinetics combined with dynamic protein oligomerization (Fig. 6*A*), which yielded the apparent rate constants $k_{cat,1}$ and $k_{cat,2}$ for autoadenylylation and target adenylylation, respectively, at the given substrate (ATP) concentration. The NmFic-GyrB43 dissociation constant ($K_{d,G}$) was set equal to the K_m value obtained from GyrB titration. Oligomerization dissociation constants ($K_{d,1}$, $K_{d,2}$, and $K_{d,2'}$) were set to their experimentally determined values. From the fit of $NmFic_{E156R}$-catalyzed GyrB adenylylation, the effective autoadenylylation rate ($k_{cat,1,eff}$) was obtained. Numerical integration of the differential equations corresponding to each reaction (described in the legend for Fig. S3*C*) was performed using the Complex Pathway Simulator (COPASI) software (27). Fitting and simulations were performed using either ProFit 6.2.14 (QuantumSoft) or COPASI 4.14.

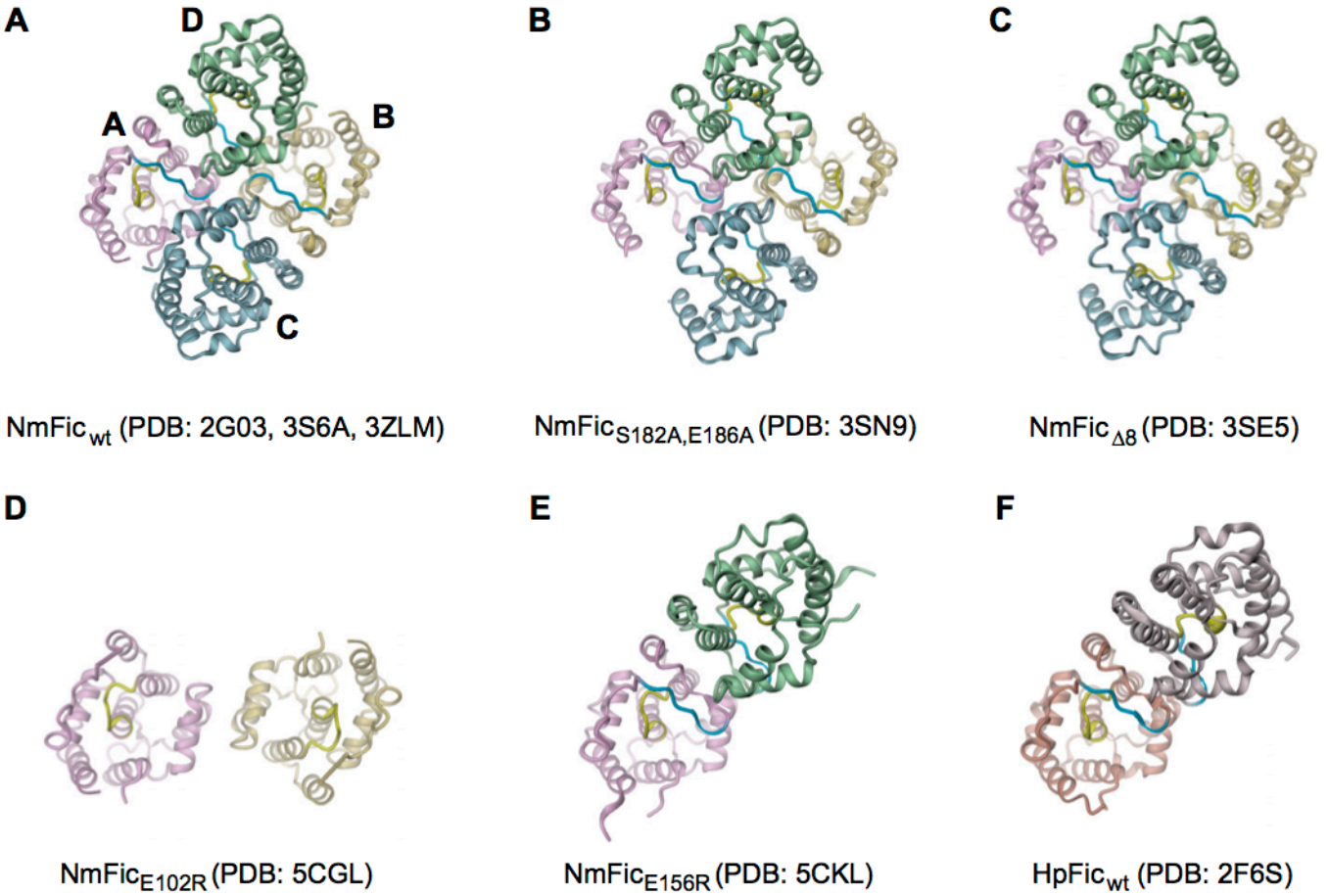


Fig. S1. Quaternary structures of class III Fic proteins. Representations of the structures as in Fig. 2*C* with monomer A (light pink) as a reference. (*A*) Crystallographic tetramer of NmFic observed in the $P6_422$ crystal form of $NmFic_{wt}$ (PDB ID code 2G03), $NmFic_{wt}$/adenosine 5′-(β,γ-imido)triphosphate (AMPPNP) (8) (PDB ID code 3S6A), and $NmFic_{E186G}$/AMPPNP (9) (PDB ID code 3ZLM). The structure of $NmFic_{wt}$/AMPPNP (PDB ID code 3S6A) is depicted. (*B*) $NmFic_{S182A,E186A}$ (8) (space group $P2_1$, four tetramers per asymmetric unit) with unresolved helix 8 and (*C*) $NmFic_{\Delta 8}$ (8) (space group $P2_1$, one tetramer per asymmetric unit) are highly similar to $NmFic_{wt}$ (rmsd of 0.7 Å and 0.9 Å for each 620 Cα-positions of the two structures). (*D*) Dimeric $NmFic_{E102R}$ (PDB ID code 5CGL, local dimer; this work) superimposes with dimer AB from $NmFic_{wt}$ and (*E*) dimeric $NmFic_{E156R}$ (PDB ID code 5CKL, crystallographic dimer; this work) superimposes with dimer AD from $NmFic_{wt}$ with an rmsd of 2.8 Å and 1.0 Å for 323 and 354 Cα-atoms, respectively. (*F*) Crystal structure of homologous protein *Helicobacter pylori* Fic (HpFic) observed as a dimer composed of chain A and symmetry-operated (y, −x + y, z + 5/6) chain B. Note that this dimer superimposes well with the dimer AD from the $NmFic_{wt}$ tetramer (rmsd of 1.6 Å for 354 Cα-atoms). Although the charged residues that would contribute to interface 2 (Fig. 2*A*) are conserved, tetramer formation appears impeded by the presence of the N-terminal tag.

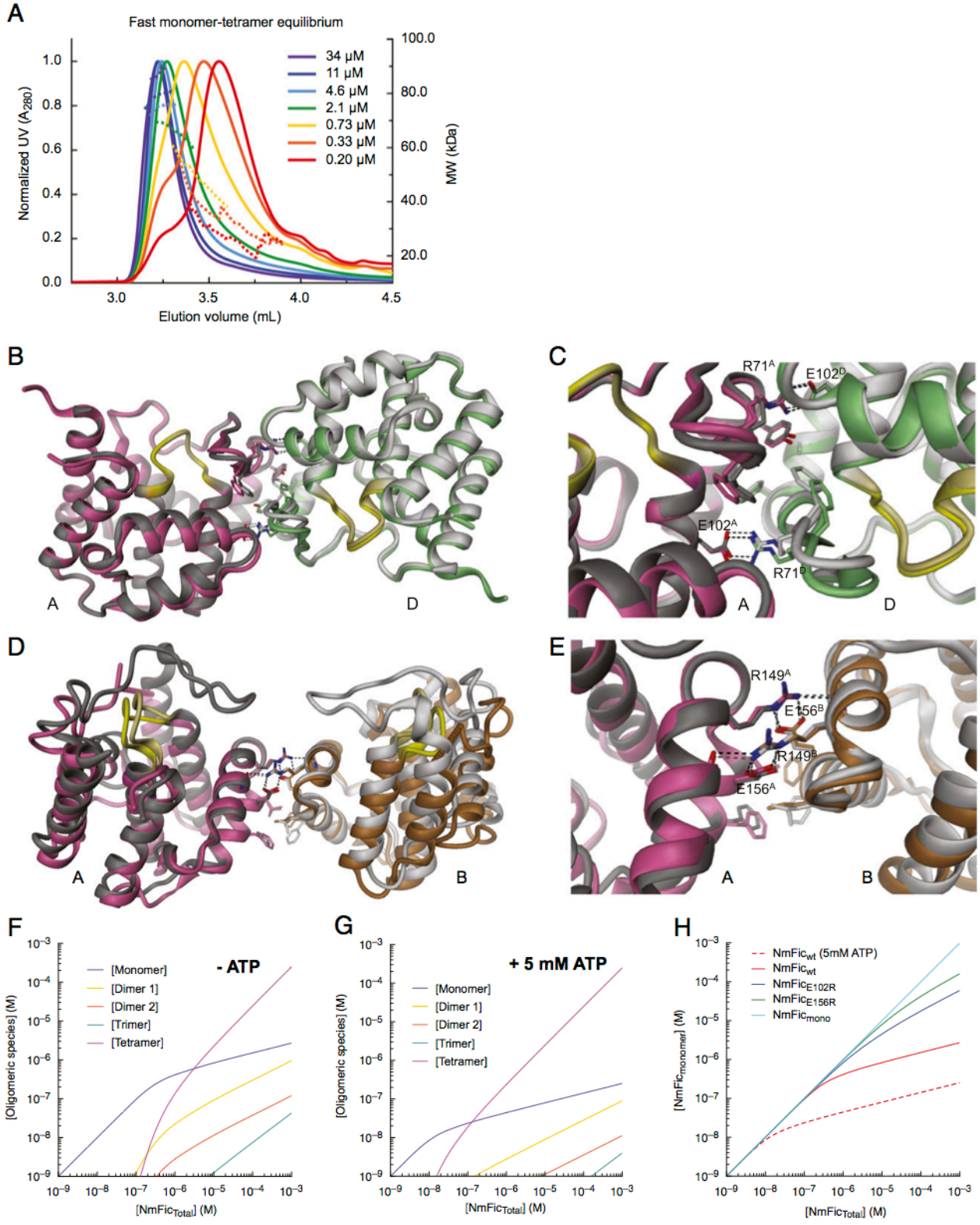
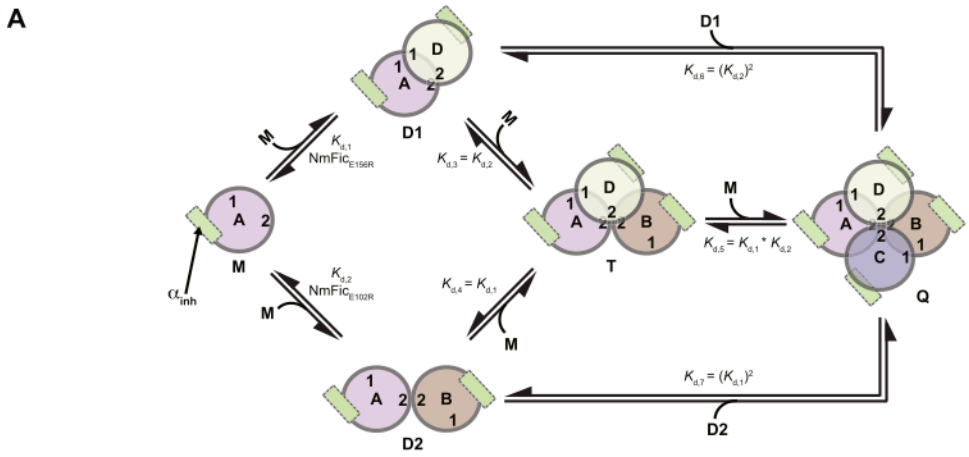


Fig. S2. Detailed analysis of NmFic oligomerization. (*A*) SEC-MALLS elution profile (plain lines) of $NmFic_{wt}$ measured at varying concentrations. The eluting concentration is indicated. The corresponding apparent molecular mass values are shown overlaid (dotted lines). (*B*) Comparison of the structures of $NmFic_{E156R}$ (PDB ID code 5CKL, pink/green) and of the AD dimer of $NmFic_{wt}$ (dark/light gray) as part of the $NmFic_{wt}$ tetramer shown in Fig. 2 *B* and *C* and (*C*) details of the conserved interaction site. (*D*) Comparison of the structures of $NmFic_{E102R}$ (PDB ID code 5CGL, pink/orange) and of the AB dimer of $NmFic_{wt}$ (dark/light gray) as part of the $NmFic_{wt}$ tetramer shown in Fig. 2 *B* and *C* and (*E*) details of the conserved interaction site. (*F–H*) Theoretical distribution of oligomeric $NmFic_{wt}$ species as derived from the kinetic model shown in Fig. S3*A* and the $K_{d,1}$ and $K_{d,2}$ values from Table 1. (*F*) Distribution of the various oligomeric species of $NmFic_{wt}$ as a function of total concentration. All concentrations are given as monomeric molar concentrations. (*G*) Same as in *F*, but in the presence of 5 mM ATP. (*H*) Monomer concentration of the indicated NmFic variants as a function of total concentration.



$$2 * M \leftrightarrows D1 \quad K_{d,1} = \frac{[M]^2}{[D1]} \quad (1)$$

$$2 * M \leftrightarrows D2 \quad K_{d,2} = \frac{[M]^2}{[D2]} \quad (2)$$

$$M + D1 \leftrightarrows T \quad K_{d,3} = K_{d,2} \quad (3)$$

$$M + D2 \leftrightarrows T \quad K_{d,4} = K_{d,1} \quad (4)$$

$$M + T \leftrightarrows Q \quad K_{d,5} = K_{d,1} * K_{d,2} \quad (5)$$

$$2 * D1 \leftrightarrows Q \quad K_{d,6} = (K_{d,2})^2 \quad (6)$$

$$2 * D2 \leftrightarrows Q \quad K_{d,7} = (K_{d,1})^2 \quad (7)$$

$$K_d = constant * e^{\left(\frac{\Delta G}{RT}\right)} \quad with\ constant = 1\ mol/L \quad (8)$$

$$\begin{aligned}(K_d)' &= constant * e^{\left(\frac{\Delta G'}{RT}\right)} \quad with\ \Delta G' = 2 * \Delta G \\ &= constant * e^{2*\left(\frac{\Delta G}{RT}\right)} \\ &= constant * \left(e^{\left(\frac{\Delta G}{RT}\right)}\right)^2 \\ &= constant * \left(\frac{K_d}{constant}\right)^2 \\ &= \frac{(K_d)^2}{constant}\end{aligned} \quad (9)$$

Fig. S3. Thermodynamic model of NmFic oligomerization. (*A*) Schematic view of the thermodynamic equilibrium model with associated parameters and (*B*) complete set of association reactions (1 to 7). D1, dimer 1; D2, dimer 2; M, monomer; Q, tetramer; T, trimer. Eq. **8** in this figure relates the change in Gibbs free energy, ΔG, upon complex formation to the dissociation constant K_d in moles per liter. Assuming no allosteric cooperativity, the ΔG associated with tetramer formation (e.g., dimerization of D2) is twice the ΔG of dimer formation (D1 in this example), because the same interface 1 is used twice. In Eq. **9** in this figure, we show that a twofold increase in ΔG results in a squared numerical value for the corresponding K_d. (*C*) Model of the regulatory mechanism of class III Fic proteins. Monomeric native protein (N) with unfolded α_{inh} (N′), unfolded α_{inh} (Y183) bound to the flap (N″), autoadenylylated protein (A^0), and autoadenylylated NmFic (A) with the modified segment detached are shown as in Fig. 6*A*. It is assumed that both N and A can contribute to oligomer formation, resulting in four distinct dimers (NN, 2*AN, and AA) and 16 distinct tetramers (NNNN, 4*NNNA, 6*NNAA, 4*NAAA, and AAAA), bottom part of the scheme showing the sequestration of active species. Monomeric autoadenylylated protein (A, active) binds the target protein and subsequently transfers an AMP moiety on this target. The respective binding constants ($K_{d,1}$, $K_{d,2}$, $K_{d,2'}$, and $K_{d,G}$) and kinetic parameters ($k_{cat,1}$, $k_{cat,1,eff}$, and $k_{cat,2}$) are indicated. This model has been used to simulate the graphs shown in Fig. 6 *C* and *D*, with the thermodynamic and kinetic parameters indicated in Tables 1 and 2. This scheme remains a simplified model omitting all ATP binding steps [i.e., assuming enzyme saturation with substrate (ATP concentration >> $K_{d,ATP}$)].

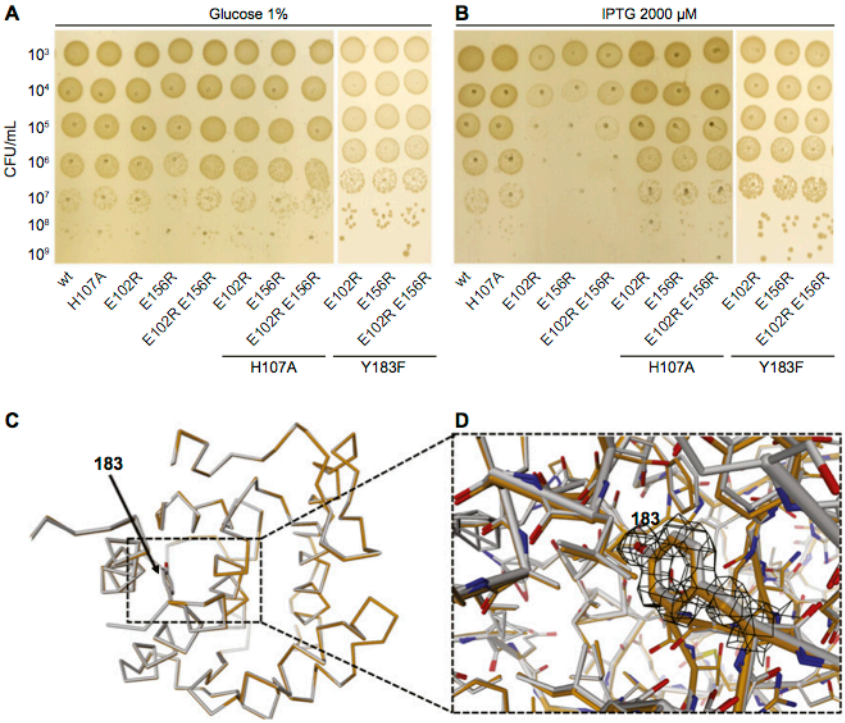


Fig. S4. In vivo and *in crystallo* effects of residue(s) replacement in NmFic. Spotting of *E. coli* strain AB472 expressing NmFic derivatives on plates containing 1% glucose (repressor) (*A*) or 2,000 µM IPTG (inducer) (*B*). Note the remarkable growth defect, both in colony number and colony size of the $NmFic_{E102R}$, $NmFic_{E156R}$, and $NmFic_{E102R,E156R}$ ($NmFic_{mono}$) variants. (*C*) Superimposition of the crystal structures of $NmFic_{E156R}$ (PDB ID code 5CKL, light gray) and $NmFic_{E156R,Y183F}$ (PDB ID code 5CMT, orange) determined at a resolution of 0.99 Å, with an rmsd of 0.23 Å for 176 Cα-positions. (*D*) Close-up view focused on residue 183. The 2Fo-Fc electron density map is shown in black for Y183 ($NmFic_{E156R}$ structure) at a contouring level of 1.2 sigma. For simplicity, only one monomer is depicted in *C* and *D*.

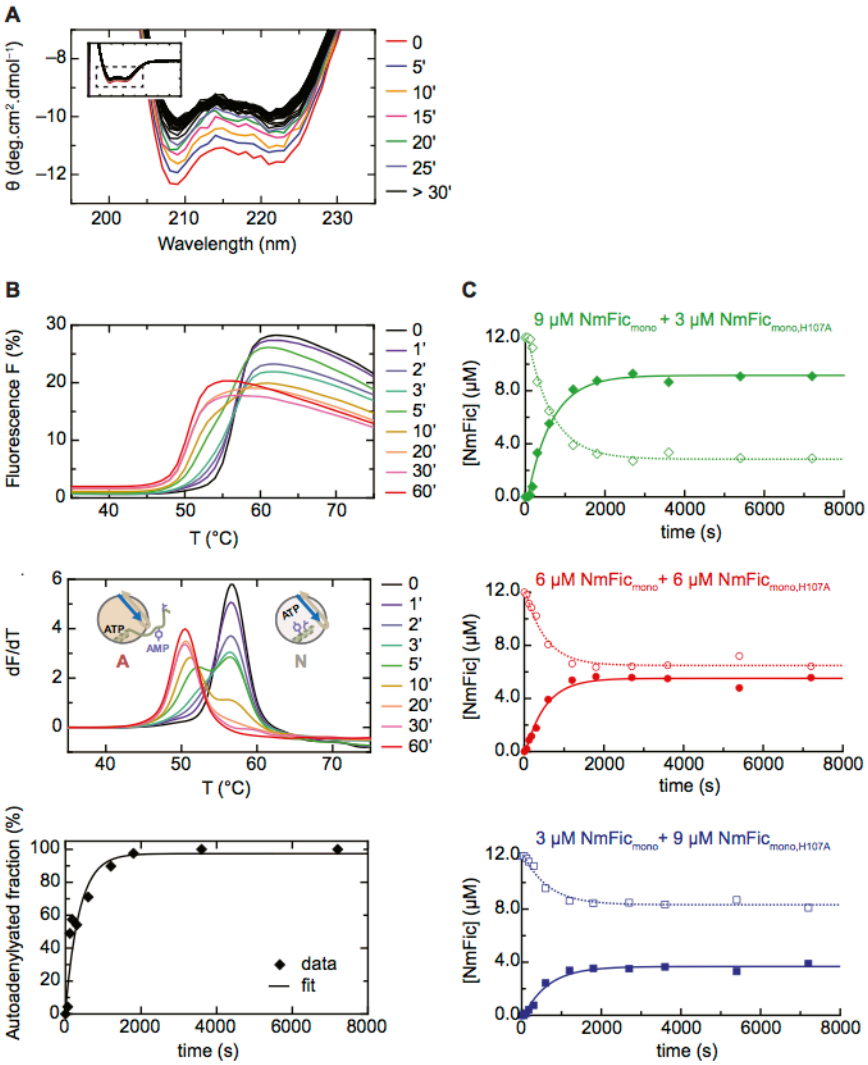


Fig. S5. Analysis of NmFic$_{mono}$ autoadenylylation by real-time CD and DSF. (*A*) CD spectra acquired every 5 min are displayed as indicated. (*Inset*) Full 190- to 270-nm spectra. (*B*) Details of the data evaluation of the melting curves obtained by DSF. Biphasic melting curve [fluorescence (F)] of heterogeneous samples of NmFic (mixture of autoadenylylated and native protein) (*Top*), first derivative dF/dT (*Middle*), and the resulting progress curve of NmFic autoadenylylation (*Bottom*) are shown. Ten micromolar NmFic$_{E156R}$, 2 mM ATP, and 10 mM $MgCl_2$ were used. (*C*) Autoadenylylated NmFic$_{mono}$ (filled) and native NmFic$_{mono}$ (opened) concentrations are plotted as progression curves. The results of three different experiments using the same total amount of monomeric NmFic (12 μM) but distinct amounts of active (NmFic$_{mono}$) and catalytically inactive (NmFic$_{mono,H107A}$) monomeric NmFic are displayed as indicated. Note that only NmFic$_{mono}$ is autoadenylylated, but not NmFic$_{mono,H107A}$, showing that autoadenylylation occurs only *in cis*.

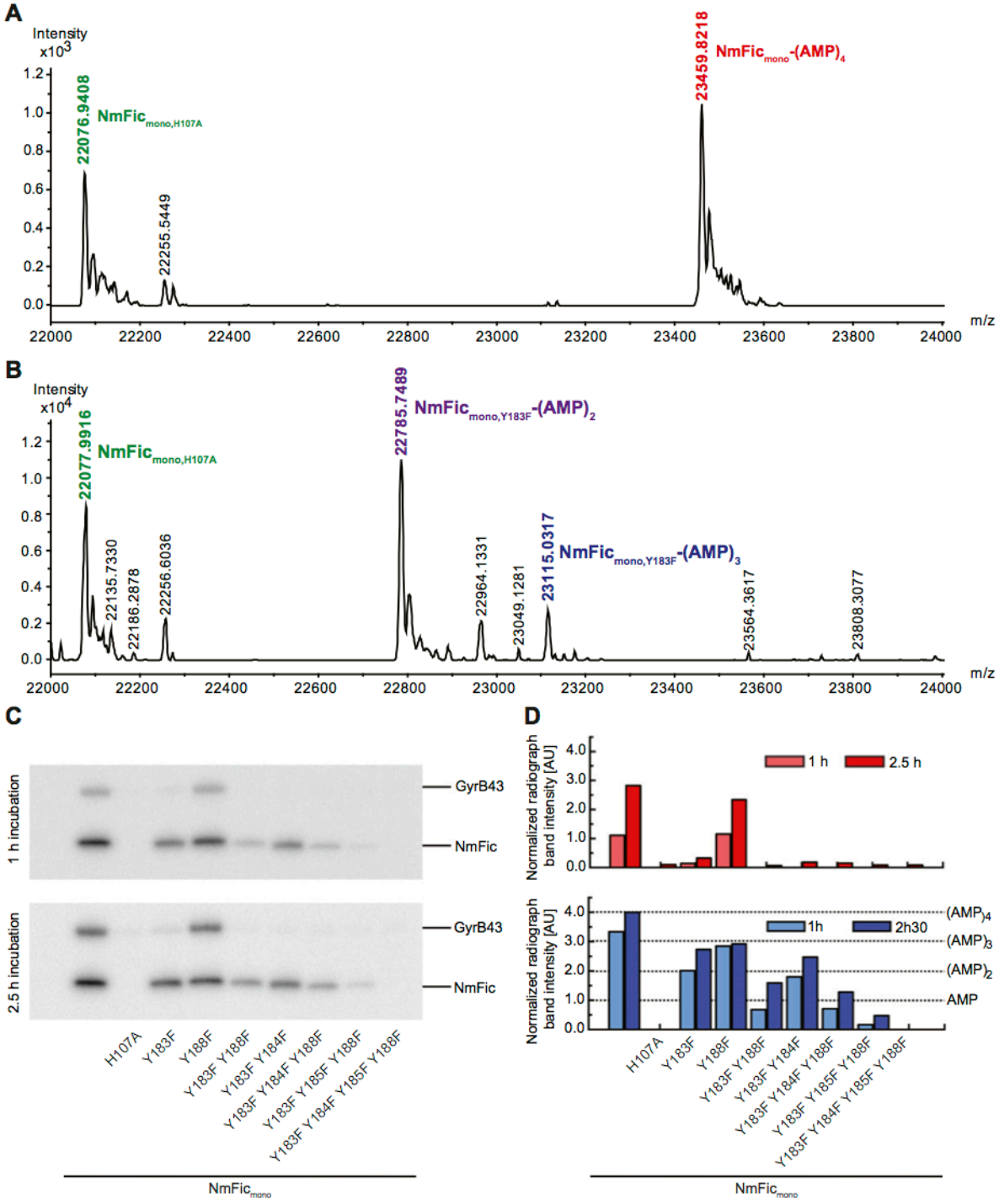


Fig. S6. Analysis of $NmFic_{mono}$ autoadenylylation by MS and autoradiography. (*A*) Mass spectrometric analysis of a 1:1 mixture of $NmFic_{mono}$ and $NmFic_{mono,H107A}$ (5 μM each) incubated for 90 min at 35 °C with 1 mM ATP and 10 mM $MgCl_2$. $NmFic_{mono}$ appears fully tetra-autoadenylylated (MW_{theor} = 22,143 Da + 4 * 329 Da = 23,459 Da). $NmFic_{mono,H107A}$ shows no modification (MW_{theor} = 22,077 Da) confirming that the NmFic automodification does occur only *in cis* and not *in trans*. (*B*) Mass spectrometric analysis of a 1:1 mixture of $NmFic_{mono,Y183F}$ and $NmFic_{mono,H107A}$ (5 μM each) incubated for 90 min at 35 °C with 1 mM ATP and 10 mM $MgCl_2$. Two species corresponding to $NmFic_{mono}$ being modified by two and three AMP moieties (MW_{theor} = 22,785 and 23,114, respectively) are observed. As in *A*, $NmFic_{mono,H107A}$ shows no modification. The experimental values of the identified species agree within 1 Da with the theoretical values. (*C*) In vitro target adenylylation as monitored by autoradiography. Ten micromolar purified GyrB43 was incubated with 1 μM of various purified $NmFic_{mono}$ variants that had one or more of the potential autoadenylylation sites on the α_{inh} (Y183, Y184, Y185, and Y188) replaced by Phe. Incubation times were 1 h (*Top*) or 2.5 h (*Bottom*). All variants with an Y183F mutation were (largely) incompetent for target adenylylation but still showed various degrees of autoadenylylation. Only replacement of all four Tyr abolished autoadenylylation completely. It is noteworthy that all purified protein variants behaved similarly in solution. (*D*) Quantification of the radiograph band intensities shown in *C* displayed as bar diagrams after normalization (1 AU corresponding to 1 AMP moiety). GyrB43 adenylylation (*Top*) and NmFic auto-adenylylation (*Bottom*) are shown. AU, arbitrary units.

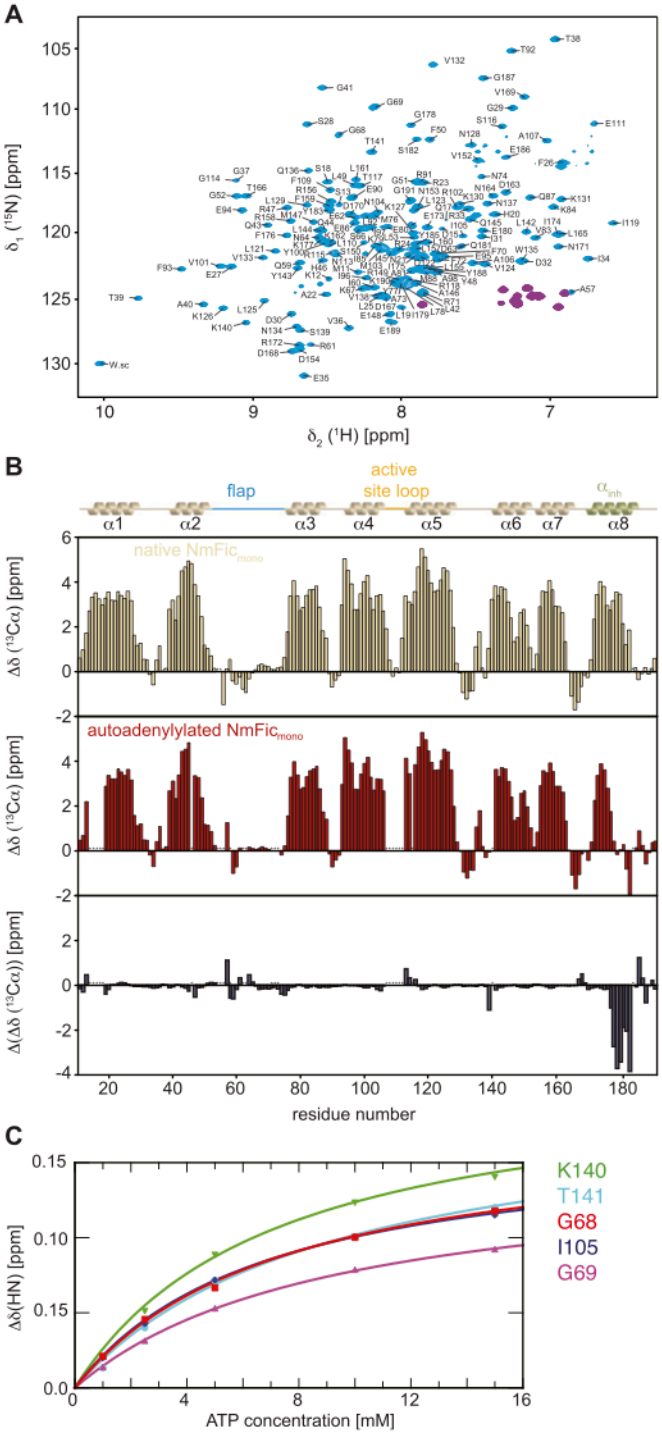


Fig. S7. NMR analysis of $NmFic_{mono}$ upon autoadenylylation. (*A*) Two-dimensional [^{15}N,^{1}H]-HSQC spectrum of 0.5 mM [*U*-^{13}C,^{15}N]-labeled monomeric inactive NmFic ($NmFic_{mono,H107A}$; cyan). Measurements were performed at 700 MHz in 25 mM MES (pH 6.5) and 150 mM NaCl at 25 °C. The sequence-specific resonance assignment is indicated. (*B*) Secondary $^{13}C\alpha$ chemical shifts of native $NmFic_{mono,H107A}$ (*Top*), autoadenylylated $NmFic_{mono}$ (*Middle*), and their difference (*Bottom*) relative to random coil values. A 1-2-1 smoothing function was applied to the raw data. Consecutive stretches with positive and negative values >2 ppm indicate α-helical and β-strand secondary structure, respectively. The secondary structure elements inferred from the crystal structure are indicated above. (*C*) Backbone amide chemical shift perturbations for selected residues upon titration of ATP to [*U*-^{13}C,^{15}N]-labeled $NmFic_{mono,H107A}$ (starting concentration of 500 μM). The lines represent nonlinear least squares best fits of the normalized changes in ^{1}H and ^{15}N chemical shifts using a bimolecular equilibrium binding model. The optimized value of the equilibrium constant is 7.8 ± 1.1 mM.

Table S1. Crystallographic table

	$NmFic_{E102R}$	$NmFic_{E156R}$	$NmFic_{E156R,Y183F}$
PDB ID code	5CGL	5CKL	5CMT
Data collection statistics			
X-ray source	SLS X06SA (PXI)	SLS X06DA (PXIII)	SLS X06DA (PXIII)
X-ray detector	Pilatus 6M	Pilatus 2M	Pilatus 2M
Wavelength, Å	1.0000	0.8000	0.8000
Space group	P 2 2_1 2_1	P 2 2_1 2_1	P 2 2_1 2_1
Cell dimensions a, b, c; Å	48.56, 81.91, 97.53	35.33, 50.54, 130.14	35.38, 50.70, 129.78
Molecules in asymmetrical unit	2	1	1
Matthews coefficient, $Å^3 \cdot Da^{-1}$	2.53	2.73	2.74
Solvent content, %	51.0	54.7	54.8
Resolution limits, Å	48.76–2.35 (2.43–2.35)	65.07–0.99 (1.03–0.99)	64.89–0.99 (1.03–0.99)
R_{merge},* %	7.5 (47.9)	3.2 (54.8)	4.0 (55.1)
R_{meas},† %	8.1	3.4	4.3
CC 1/2, %	99.9 (93.0)	100 (86.9)	99.9 (82.3)
$\langle I/\sigma(I) \rangle$	21.9 (4.6)	29.8 (3.5)	24.8 (3.1)
Total reflections	120,000 (12,268)	839,576 (79,738)	805,834 (74,721)
Unique reflections	16,808 (1,666)	129,866 (12,557)	126,349 (11,682)
Multiplicity	7.1 (7.4)	6.5 (6.4)	6.4 (6.4)
Completeness, %	99.9 (100.0)	99.5 (97.6)	96.6 (90.6)
Mosaicity	0.19	0.10	0.21
Refinement statistics			
R_{work},‡ %	19.9 (25.6)	11.2 (16.3)	13.5 (19.0)
R_{free},§ %	23.2 (26.5)	12.5 (17.1)	15.6 (19.4)
No. of non-H atoms	2,885	2,041	1,911
Macromolecules	2,725	1,670	1,648
Ligands	14	9	15
Solvent	146	355	248
Protein residues	324	181	180
rmsd from ideal			
Bond lengths, Å	0.002	0.029	0.029
Bond angles, °	0.53	2.20	2.43
Ramachandran favored,¶ %	99.4	99.5	99.5
Ramachandran outliers,¶ %	0	0	0
Clashscore¶	4.39	4.80	3.31
Average B values, $Å^2$	38.40	14.30	11.80
Macromolecules	38.30	11.20	10.10
Ligands	48.50	24.10	25.80
Solvent	39.40	39.40	39.40

Numbers in parentheses belong to the outer shell. $\langle I/\sigma(I) \rangle$, mean of intensity over standard deviation.

*$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection and $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

†$R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections, and N is the number of observations of intensity I(hkl).

‡$R_{work} = \sum_{hkl} || F_{obs}| - |F_{calc} || / \sum_{hkl} |F_{obs}|$.

§R_{free} is the R value calculated for 5% of the data set that was not included in the refinement.

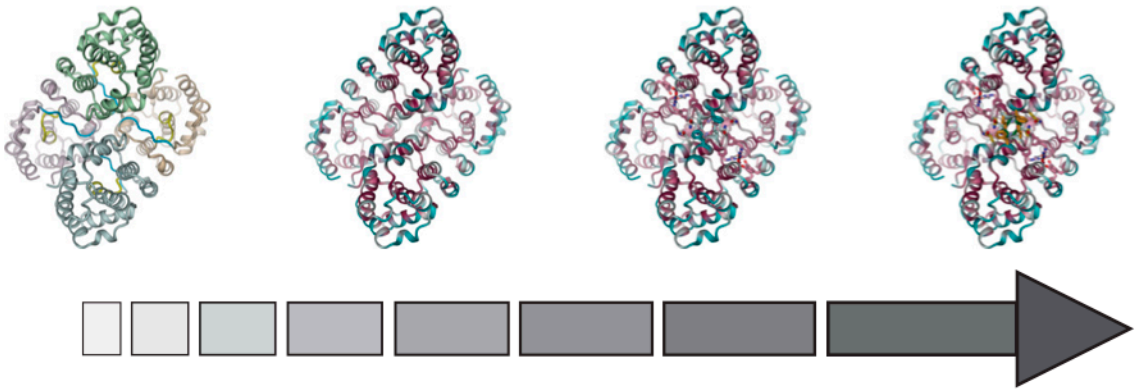
¶Molprobity.

Table S2. List of plasmids used in this study

Plasmid	Backbone	Description	Primer (forward)	Primer (reverse)	Restriction sites	Source
pRSF-Duet1	pRSF-Duet1	Empty vector (RSF1030 *ori*, P*T7*)				Novagen
pFVS0015	pRSF-Duet1	$NmFic_{wt}$				Engel et al. (8)
pFVS0051	pRSF-Duet1	$NmFic_{H107A}$	prFVS149	prFVS150		This study
pFVS0059	pRSF-Duet1	$NmFic_{E186G}$				Goepfert et al. (9)
pFVS0081	pRSF-Duet1	$NmFic_{Y183F}$	prFVS091	prFVS092		This study
pFVS0109	pRSF-Duet1	*E. coli* GyrB 1–392 N_{ter} His_6-tag				Stanger et al. (25)
pFVS0125	pRSF-Duet1	$NmFic_{E102R}$	prFVS119	prFVS120		This study
pFVS0126	pRSF-Duet1	$NmFic_{E156R}$	prFVS121	prFVS122		This study
pFVS0134	pRSF-Duet1	$NmFic_{E102R,E156R}$	prFVS119	prFVS120		This study
pFVS0135	pRSF-Duet1	$NmFic_{E156R,Y183F}$	prFVS091	prFVS092		This study
pFVS0137	pRSF-Duet1	$NmFic_{E102R,H107A}$	prFVS149	prFVS150		This study
pFVS0138	pRSF-Duet1	$NmFic_{H107A,E156R}$	prFVS149	prFVS150		This study
pFVS0143	pRSF-Duet1	$NmFic_{E102R,H107A,E156R}$	prFVS149	prFVS150		This study
pFVS0144	pRSF-Duet1	$NmFic_{E102R,E156R,Y183F}$	prFVS091	prFVS092		This study
pFVS0145	pRSF-Duet1	$NmFic_{E102R,E156R,Y188F}$	prFVS093	prFVS094		This study
pFVS0167	pRSF-Duet1	$NmFic_{E102R,Y183F}$	prFVS091	prFVS092		This study
pFVS0319	pRSF-Duet1	$NmFic_{E102R,E156R,Y183F,Y188F}$	prFVS093	prFVS095		This study
pFVS0323	pRSF-Duet1	$NmFic_{E102R,E156R,Y183F,Y184F}$	prFVS277	prFVS278		This study
pFVS0326	pRSF-Duet1	$NmFic_{E102R,E156R,Y183F,Y184F,Y188F}$	prFVS279	prFVS278		This study
pFVS0329	pRSF-Duet1	$NmFic_{E102R,E156R,Y183F,Y185F,Y188F}$	prFVS280	prFVS281		This study
pFVS0332	pRSF-Duet1	$NmFic_{E102R,E156R,Y183F,Y184F,Y185F,Y188F}$	prFVS282	prFVS283		This study
pNMD220	pNDM220	Empty vector (mini-R1 *ori*, PLac)				Gotfredsen and Gerdes (23)
pAH154_NmFic_wt	pNDM220	$NmFic_{wt}$	prAH525	prAH526	BamHI/EcoRI	This study
pFVS0149	pNDM220	$NmFic_{E102R}$	prFVS119	prFVS120		This study
pFVS0150	pNDM220	$NmFic_{E156R}$	prFVS121	prFVS122		This study
pFVS0151	pNDM220	$NmFic_{H107A}$	prFVS149	prFVS150		This study
pFVS0173	pNDM220	$NmFic_{E102R,E156R}$	prFVS119	prFVS120		This study
pFVS0174	pNDM220	$NmFic_{E102R,Y183F}$	prFVS091	prFVS092		This study
pFVS0175	pNDM220	$NmFic_{E156R,Y183F}$	prFVS091	prFVS092		This study
pFVS0177	pNDM220	$NmFic_{E102R,H107A}$	prFVS149	prFVS150		This study
pFVS0178	pNDM220	$NmFic_{E102R,E156R,Y183F}$	prFVS091	prFVS092		This study
pFVS0179	pNDM220	$NmFic_{H107A,E156R}$	prFVS149	prFVS150		This study
pFVS0197	pNDM220	$NmFic_{E102R,H107A,E156R}$	prFVS149	prFVS150		This study
pGex-6p1	pGex-6p1	Empty vector (ColE1 *ori*, P*lac*)				GE Healthcare
pAH111	pGex-6p1	GyrB of *E. coli* K-12				Harms et al. (5)
pAH119	pGex-6p1	ParE of *E. coli* K-12				Harms et al. (5)
pAH122	pGex-6p1	GyrB of *N. meningitidis* 2808	prAH295	prAH297	BglII/XhoI	This study
pAH141	pGex-6p1	ParE of *N. meningitidis* 2808	prAH416	prAH417	BamHI/XhoI	This study
pAH149	pGex-6p1	GyrB of *M. tuberculosis* H37Rv	prAH453	prAH454	BglII/XhoI	This study

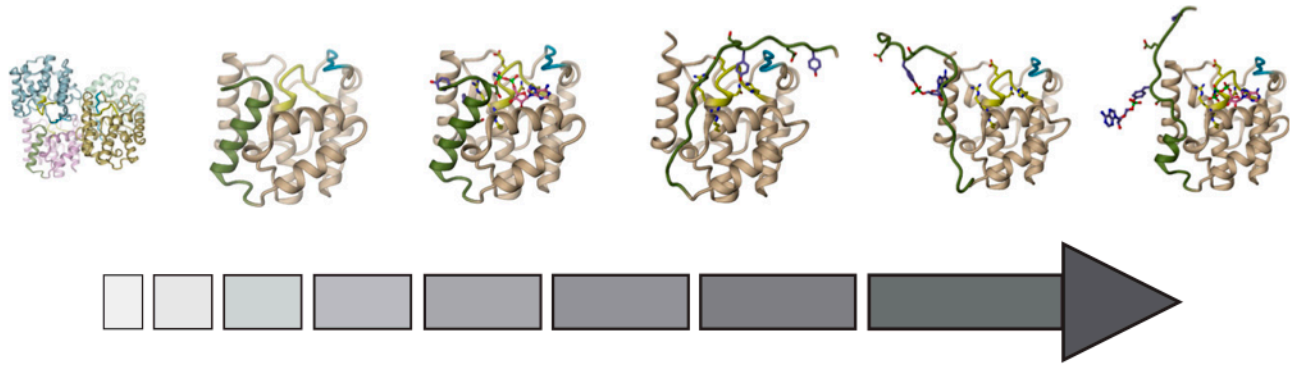
Table S3. List of DNA oligonucleotides used in this study

Primer	Sequence (5′–3′)
prFVS082	GTATTATTACGGCGGGTATGAAAAAGGCTGAC
prFVS083	CATACCCGCCGTAATAATACGACTGCTCG
prFVS091	GAGCAGTCGTTTTATTACGAAGGGTATGAAAAAG
prFVS092	GTAATAAAACGACTGCTCGATACCTTTAAAG
prFVS093	CGAAGGGTTTGAAAAAGGCTGACTCGAGTCTG
prFVS094	CTTTTTCAAACCCTTCGTAATAATACGACTG
prFVS095	CTTTTTCAAACCCTTCGTAATAAAACGACTG
prFVS119	CAAATATGTTCGTATGAACATTGCCCATCC
prFVS120	CAATGTTCATACGAACATATTTGGCGATGATTTC
prFVS121	CAACGATTTACGUCTGCGCTTTTTGTTAAAG
prFVS122	CAAAAAGCGCAGACGTAAATCGTTGACGGGGCTG
prFVS149	GAACATTGCCGCCCCGTTTTTGGAGGGTAATGGCAG
prFVS150	CAAAAACGGGGCGGCAATGTTCATTTCAACATATTTG
prFVS277	GTCGTTTTTTTACGAAGGGTATGAAAAAG
prFVS278	CTTCGTAAAAAAACGACTGCTCGATACC
prFVS279	GTCGTTTTTTTACGAAGGGTTTGAAAAAG
prFVS280	GTTTTATTTCGAAGGGTTTGAAAAAGGCTG
prFVS281	CAAACCCTTCGAAATAAAACGACTGCTCG
prFVS282	GTCGTTTTTTTTCGAAGGGTTTGAAAAAGGCTG
prFVS283	CCCTTCGAAAAAAAACGACTGCTCGATACC
prAH525	GAGCGGGGATCCATAGGAGGAACAATTTTATGAAATCCATAGACGAACAAAG
prAH526	CTCCCGCGAATTCTCAGCCTTTTTCATACCCTTC
prAH537	TTATTACGGCGGGTATGAAAAAGGCTGAGAA
prAH538	TTCATACCCGCCGTAATAATACGACTGCTCG
prAH295	GAGCGAGATCTATGACTGAACAAAAACACGAAG
prAH297	CTCCCGCTCGAGTTATGCGTCGATATTTTGGG
prAH416	ATCCTCCAAAATCGGATCTGGAAGTTCTGTTCCAGGGGCCCCTGGGATCCATGGCTAAAAACAACCAATACA
prAH417	CCGAAACGCGCGAGGCAGATCGTCAGTCAGTCACGATGCGGCCGCTCGAGTCAAATATCCAATTCCGCCG
prAH453	GAGCGAGATCTATGGGTAAAAACGAGGCCAG
prAH454	CTCCCGCTCGAGTTAGACATCCAGGAACCGAA



Movie S1. NmFic tetramer. Rotating view of tetrameric NmFic colored by monomer (same color code as in Fig. 2 and Fig. S1), with the active site and target binding site (flap) highlighted in yellow and blue, respectively (step 1). Rotating view of the same $NmFic_{wt}$ tetramer colored by conservation with a gradient from dark pink to cyan for conserved to variable residues [ConSurf (28)] (step 2). Side chains of residues involved in oligomerization interfaces are depicted as sticks with the carbon atoms colored according to the conservation score (step 3), revealing a strictly conserved interface and a more variable interface. The nonconserved interface is finally colored by covariance [Gremlin (29)] with a gradient from green (covariance probability of 1) to orange (covariance probability of 0.6). Note the high variability at the surface of the tetramer and conservation or covariance at the interfaces.

Movie S1



Movie S2. Model of *cis*-autoadenylylation. NmFic is in monomer/tetramer equilibrium. Once a monomer (N) exits the tetramer, the α_{inh} can unfold and bind to the target binding site, positioning the modifiable Tyr Y183 at a specific position in the active site to allow the nucleophilic attack on the α-phosphorous of ATP (N′′), which results in covalent transfer of AMP onto Y183 (A^0). Note that upon unfolding of the α_{inh}, the inhibitory Glu can no longer occupy its inhibitory position. Upon release of Y183-AMP from the flap, the α_{inh} partially refolds (A). The latter conformation corresponds to the active monomeric NmFic protein.

Movie S2

5.3. FIC-domain protein interaction with cognate target GyrB

5.3.1. Introduction

Most FIC-domain proteins catalyze the transfer of a phosphoryl-containing group onto a target protein, through cleavage of a substrate phosphodiester bond (13). Of these, those with a canonical FIC motif catalyze the transfer of an AMP group, from a bound ATP molecule. Their catalytic activity is under the control of a regulatory element termed α_{inh}, and can be largely classified based on the α_{inh} position in relation to the catalytic motif. The α_{inh} regulatory element contains a conserved motif, where an invariant glutamate carboxyl side-chain binds an invariant arginine part of the FIC motif, effectively preventing competent binding of the nucleotide substrate (2). The physiological α_{inh} toxin inhibition relief mechanism is not yet clear (19).

The first identified FIC protein: the gain-of-function *FIC-1* was found to be the responsible agent In the *E.coli* filamentation growth-arrest phenotype caused by cyclic-AMP (4). Since then, different bacterial FIC-domain proteins have been identified as host-targeted secreted pathogenic factors (6,25); however, the biological function of most FIC-domain proteins remains elusive. VbhTA is a toxin/antitoxin protein complex expressed by the mammalian pathogen *Bartonella schoenbuchensis* (26). The antitoxin VbhA is encoded upstream of the VbhT gene. The full-length VbhT protein contains a C-terminal Bep intracellular delivery (BID) domain, a type IV secretion system signal, responsible for the secretion of the toxin-antitoxin complex onto bacterial or mammalian host-cell targets. Upon host-delivery the inactivating complex is thought to disassemble, leading to the mature active form of the toxin (19). In the inactive form, VbhA forms a tight complex with VbhT, where the conserved residues S20 and specifically, E24 bind R147 of VbhT, obstructing the competent binding of ATP within the catalytic site on the FIC-domain motif. *In vitro*, the $VbhA_{E24G}$ inhibition relief mutant was shown to mimic the antitoxin release mechanism behaving similarly as the active form of VbhT (2). Recently, the VbhT was linked to an *E.coli* growth-arrest phenotype, correlating with the AMPylation of two target proteins: the homologous subunits B of the bacterial type II topoisomerases DNA gyrase GyrB, and topoisomerase IV (topo IV) ParE (20). Topoisomerases control DNA topology through DNA catenation and de-catenation, ensuring DNA and RNA polymerase function, playing an essential role in healthy cell growth and function (24). VbhT catalyzes the post-translational modification of the catalytical site of both ATPases, effectively blocking its intrinsic ATPase activity, causing the abrogation of their biological function and ensuing cell growth arrest (2, 20). The structural characterization of the VbhTA/GyrB apo, as well as the nucleotide-bound complexes is described below.

5.3.2. Material and methods

Protein purification

$NmFIC_{E102R,E156R,H107A}$ monomer mutant, the VbhA/$VbhT_{(1-198)}$ protein complex, the inhibition relieved $VbhA_{E24G}$/VbhT mutant and the $GyrB_{(1-220)}$ ATPase-domain proteins, termed from here on $NmFIC_{mono}$ VbhTA, $VbhTA_{E24G}$, and GyrB, were expressed and purified as previously described in (2, 13, 27). In short, $NmFIC_{mono}$ and $VbhTA_{E24G}$ were expressed in the *E. coli* BL21-AI strain, while VbhTA and GyrB were expressed in the *E. coli* BL21 (λDE3) strain, either in LB or M9 medium (^{15}N-ammonium chloride) for isotopically labeled protein ([*U*-^{15}N]). Subsequently, each protein was purified by Ni-NTA affinity chromatography followed by size exclusion chromatography. An additional anion exchange chromatography step was performed for GyrB.

Solution NMR measurement and data analysis

Standard 2D [^{15}N,^{1}H]-TROSY-HSQC (28) NMR spectra were recorded at 25°C on a Bruker Ascend 700 spectrometer equipped with a cryogenic triple-resonance probe running Topspin 3.0 (Bruker, USA). NMR data were processed with PROSA (29) and analyzed with CARA (30). Chemical shift perturbations were calculated with:

$$CSP = \sqrt{\left[\left(\Delta\delta(^{1}H)\right)^{2} + \left(0.2 * \Delta\delta(^{15}N)\right)^{2}\right]},$$

where Δδ(^{1}H) and Δδ(^{15}N) are the chemical shift changes of the ^{1}H and ^{15}N nuclei, respectively. The earlier determined sequence-specific resonance assignments of $NmFIC_{mono}$, as well as the published assignment of GyrB, was were used for the sequence-specific analysis (31).

VbhTA/GyrB apo and nucleotide complexes co-crystallization

For crystallization, the protein complex was assembled by incubating GyrB 1:1 with the VbhTA complex, and subsequently purified on an S200 increase 10/300 GL gel filtration column (GE Healthcare) equilibrated with buffer G (20 mM phosphate buffer pH 7.2, 40 mM KCL, 40 mM NaCl, 1mM EDTA, 2 mM DTT) using an Äkta (GE Healthcare, USA) system. The complex fractions were analyzed by SDS-PAGE and concentrated to ~10 mg/ml by ultrafiltration with a 10,000 MWCO Amicon centrifugal filtering device (Sartorius, Germany).

Crystals of the protein complex were obtained using the sitting-drop vapor diffusion method after mixing 0.2 µl protein solution with 0.2 µl of reservoir solution, equilibrating against 80 µl of reservoir solution. Crystallization trials were set up using protein-complex screens (ProPlex, and NR-LBD, Molecular Dimensions, USA) as well as the Index-HT screen (Hampton Research, USA) at 20**°**C; the

the VbhTA$_{E24G}$/GyrB complex nucleotide was crystallized in the apo form and subsequently soaked with the nucleotide, as well as co-crystallized with 5 mM and ATP. For cryo-protection, the obtained crystals were soaked in the reservoir solution with 20% glycerol, and flash-frozen in liquid nitrogen for data collection.

Data collection, processing, structure determination, and refinement

Diffraction data were collected at the Swiss Light Source (Villigen, Switzerland) at 100 K, processed using DIALS (32) or XDS (33) and scaled using aimless (34). The complex structures were solved by molecular replacement using the apo structures of VbhTA (PDB ID: 3SHG), and GyrB (PDB ID: 4PRV) as search models using Molrep (35). Several rounds of iterative model building and refinement were performed using Coot (36, 37) and PHENIX (38) or REFMAC5 (39). 5% of the data were excluded from refinement and used for later cross-validation (R-free set).

5.3.3. Results

Solution NMR titration of NmFIC$_{mono}$ with cognate target GyrB

The NmFIC$_{mono}$ sequence-specific backbone assignment as well as the identification of AMPylation protein targets, opened the possibility for the titration of each interacting partner using solution NMR spectroscopy. To determine the possible interaction surfaces of the monomerized mutant NmFIC$_{mono}$ and GyrB, a stepwise titration was performed with increasing amounts of GyrB to [*U*-^{15}N]-NmFIC$_{mono}$ in buffer F (25 mM MES, 105 mM NaCl, pH 6.5), using 2D-[^{15}N,^{1}H]-TROSY-HSQC spectra (Figure 5-2).

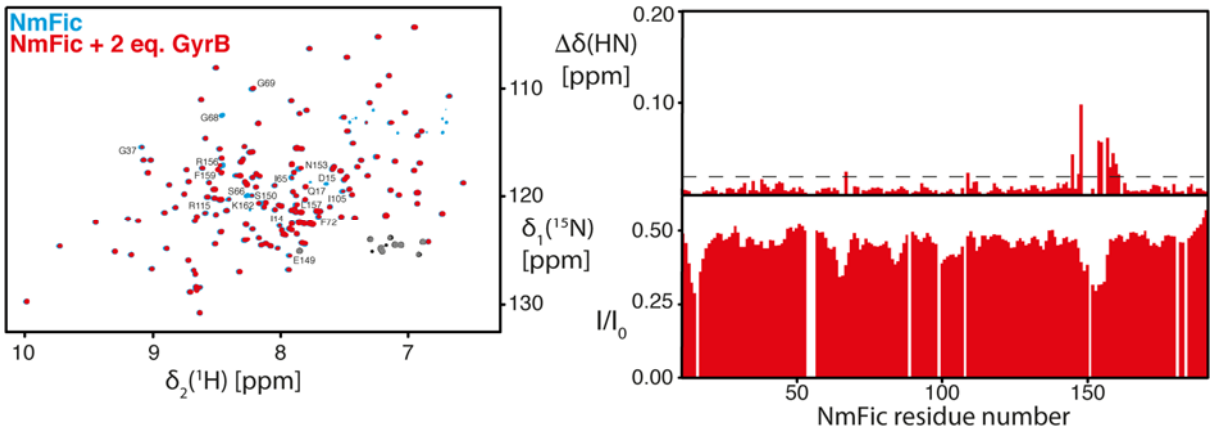


Figure 5-2 – Overlay of 2D-[^{15}N,^{1}H]-TROSY-HSQC spectra of 150 µM [*U*-^{15}N]-NmFIC$_{mono}$ (blue) and after the addition of 300 µM GyrB (red) in buffer F, measured at 25°C. The residue resolved chemical shift perturbations and intensities of backbone amide moieties upon GyrB interaction are shown on the side.

Despite observing localized chemical shift and signal intensity changes, precipitation of GyrB was also observed. As such, a buffer optimization was performed, where buffer G (20 mM phosphate buffer

pH 7.2, 40 mM KCl, 40 mM NaCl, 1mM EDTA, 2mM DTT) was found to stabilize both GyrB and $NmFIC_{mono}$ (31). Subsequently, the NMR-titration was repeated with buffer G, as shown in Figure 5-3.

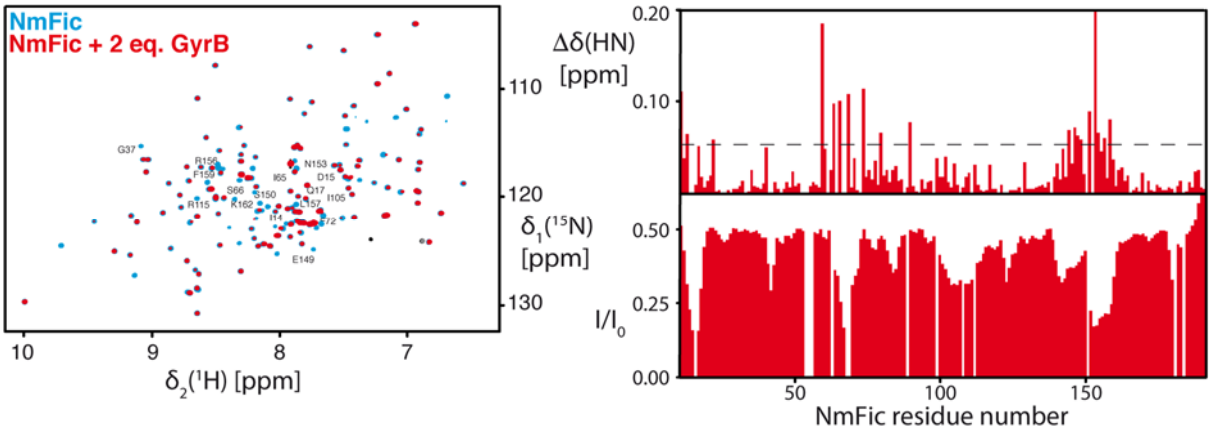


Figure 5-3– Overlay of 2D-[^{15}N,^{1}H]-TROSY-HSQC spectra of 150 µM [*U*-^{15}N]-$NmFIC_{mono}$ (blue) and after the addition of 300 µM GyrB (red) in buffer G, measured at 25°C. The residue resolved chemical shift perturbations and intensities of backbone amide moieties upon GyrB interaction are shown on the side.

The titration under the optimized conditions showed an overall peak intensity decrease, as well as increased chemical shift perturbation. In particular, the measured chemical shift perturbations *vs*. sequence plot show three localized perturbation areas, the N-terminus residues 14-16, residues 65-72, and residues 145-159, as well as intensity drops within these regions (Figure 5-3). The optimized buffer conditions permitted the reverse titration using labeled GyrB for the identification of the interaction surface on the GyrB-side. The [*U*-^{15}N]-GyrB titration spectrum with unlabeled $NmFIC_{mono}$ evidenced several resonance shifts, as well as intensity decrease and peak vanishing (Figure 5-4).

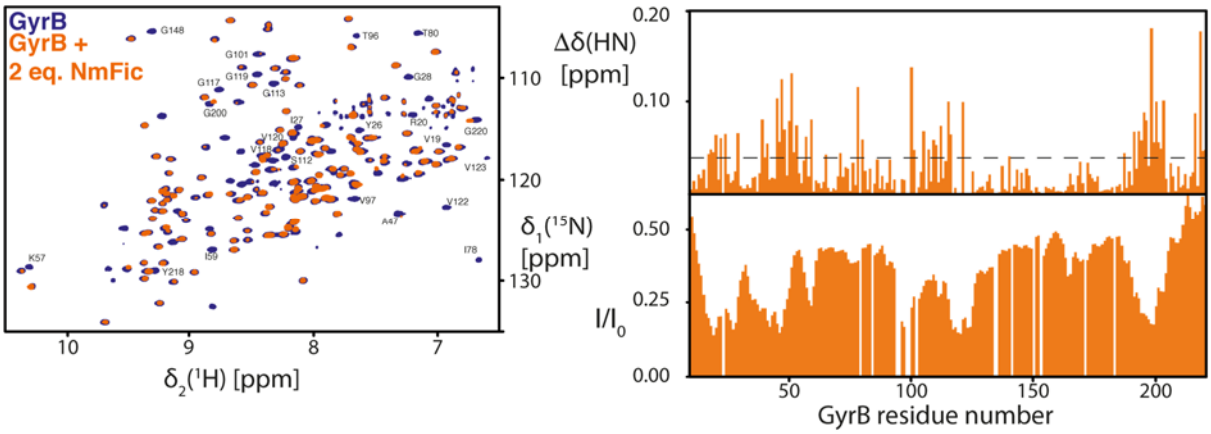


Figure 5-4 – Overlay of 2D-[^{15}N,^{1}H]-TROSY-HSQC spectra of 150 µM [*U*-^{15}N]-GyrB (purple) after the addition of 300 µM $NmFIC_{mono}$ (orange), measured in buffer G and at 25°C. The residue resolved chemical shift perturbations and intensities of backbone amide moieties upon $NmFIC_{mono}$ interaction are shown on the right side.

The chemical shift perturbations *vs*. sequence plot evidenced several chemical shift perturbation areas along the sequence, which correlate well with the intensity *vs*. sequence plot. The overall intensity

decreased is explained due to the decreased molecular tumbling of the protein complex, compared to the protein alone. The measured perturbation data from each side of the titration are shown in Figure 5-5 mapped onto each known protein crystal structure, respectively $NmFIC_{mono}$ (PDB ID: 2g03) and GyrB (PDB ID: 1ei1).

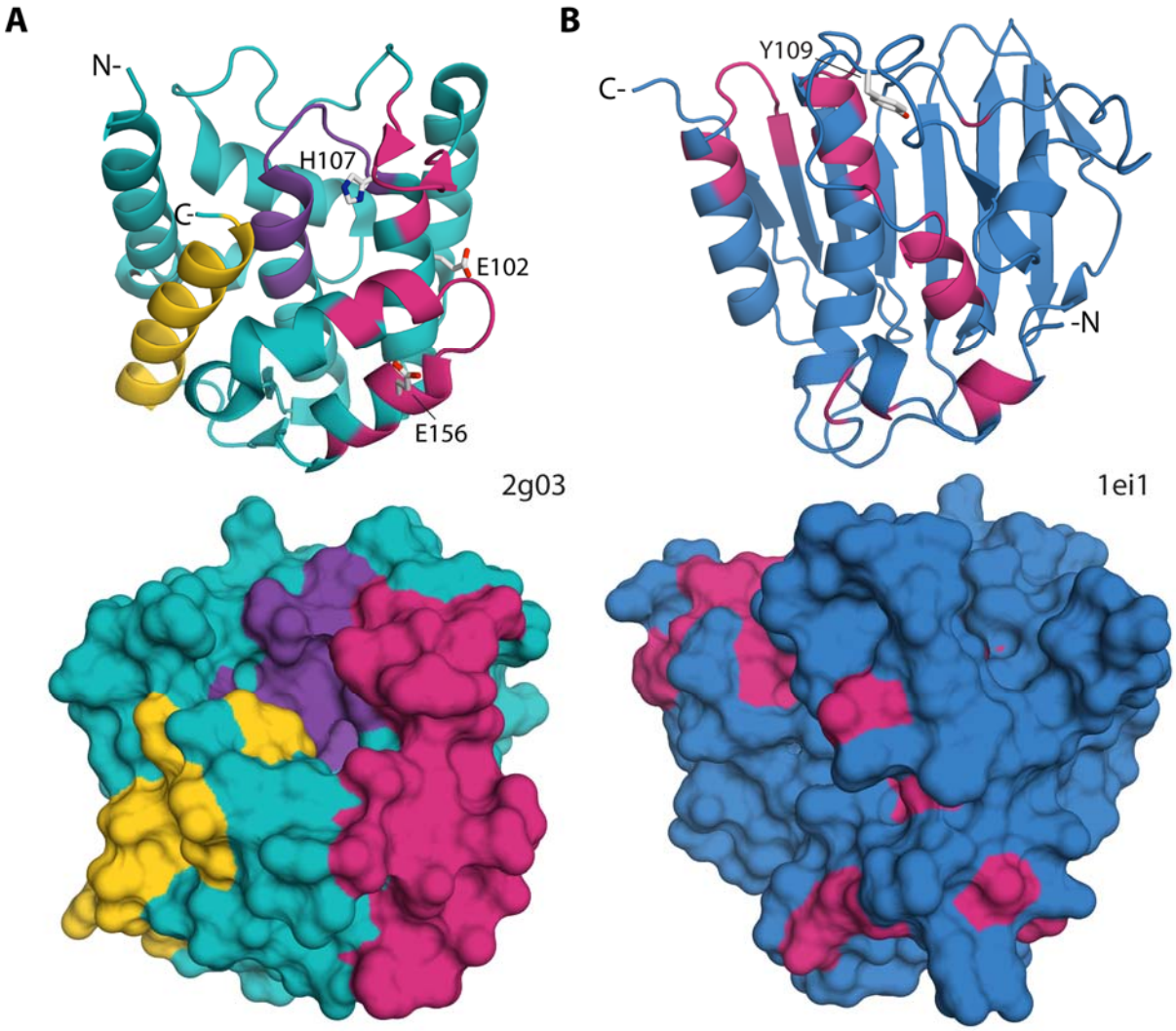


Figure 5-5 – Chemical shift perturbation (shown in dark pink) of the $NmFIC_{mono}$/GyrB complex mapped onto the known complex X-ray structure of each protein, shown in both cartoon and surface representation, with respective PDB IDs. In A, is shown NmFIC colored in teal with α_{inh} highlighted in gold, and the catalytic motif in purple, with the mutated $mutNmFIC_{mono}$ residues highlighted in stick representation. In B is shown GyrB, colored in blue, with Y109 AMPylation target residue depicted as stick representation.

On the $NmFIC_{mono}$ side, the chemical shift perturbation converges onto a binding *locus* close to to the flap region (depicted in red, previously in Figure 5-1 A) and the catalytic domain. The displaced backbone chemical shifts reveal a clear interaction interface, specifically in the surface representation (Figure 5-5 A). Conversely, on the GyrB side, the structure mapped chemical shift perturbation is spread through the N-terminal side of the protein, showing a larger binding interface, including the loop containing the AMPylation target Y109 (Figure 5-5 B). In the surface representation, no clear binding interface is evident. As shown in Figure 5-5 A, the NmFIC monomerizing mutations E102R,

E156R, are both situated in the evidenced binding interface, as such the charge inversion caused by the mutations could affect target binding in addition to the monomerization of the construct. Also, the mutation of the catalytic H107 to an alanine could as well, hinder ligand binding, and consequentially complex formation.

Solution NMR titration VbhT/VbhA complex with cognate target GyrB

As shown for the $NmFIC_{mono}$ and GyrB proteins, the same approach was followed to study the interaction between the VbhTA complex with its cognate target GyrB. Due to the unavailability of VbhTA complex backbone sequence-specific assignment, the titration was performed on the GyrB side only. To determine the possible interaction surfaces of VbhTA and GyrB a stepwise titration of [*U*-^{15}N]-GyrB with increasing amounts of VbhTA in the optimized buffer G was monitored by [^{15}N, ^{1}H]-NMR spectra. The NMR spectra of the titration endpoint overlaid with the reference spectra of GyrB, as well as the chemical shift perturbations *vs*. sequence, and intensity drop *vs*. sequence plots are shown in Figure 5-6.

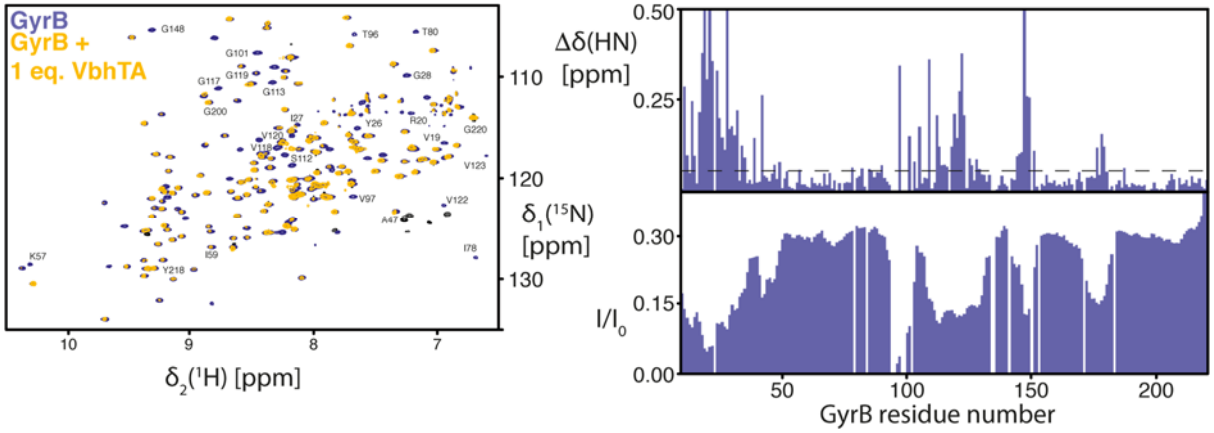


Figure 5-6 – Overlay of 2D-[^{15}N,^{1}H]-TROSY HSQC spectrum of 300 µM [*U*-^{15}N]-GyrB (purple) alone and after the addition of 300 µM VbhTA (yellow), measured in buffer G and at 25°C. The residue resolved chemical shift perturbations and intensities of backbone amide moieties upon $NmFIC_{mono}$ interaction are shown on the right side. The sequence-specific resonance assignment for the most affected resonances is indicated in the spectra.

The [*U*-^{15}N]-GyrB 2D-[^{15}N,^{1}H]-TROSY HSQC spectra evidenced several displaced backbone resonances, as well as several intensity decreased peaks, and a few peaks broadened beyond detection. The chemical shift perturbation plot shows several areas with significant chemical shift perturbation, in particular, the N-terminal region. The intensity drop plot, evidence an overall significant decrease of intensities, caused to the increased rotational correlation time of the higher sized complex when compared to the protein alone. Several localized intensity decreased areas are also observed, that correlate well with the chemical shift perturbation areas, in particular, the N-terminal residues 1-51, residues 90-103, 106-139, 141-153, and 169-183 around the ATP-lid loop. Also, some GyrB resonances belonging to residues situated in the long loop V97-G117 containing the AMPylation target Y109 were

observed to be in a slow exchange regime, for the apo- and the VbhTA-bound species. This observed slow exchange indicate the existence of a stable and distinct form of this region within the GyrB/VbhTA complex. The measured chemical shift perturbation data, mapped onto the GyrB structure, is shown in Figure 5-7 (27).

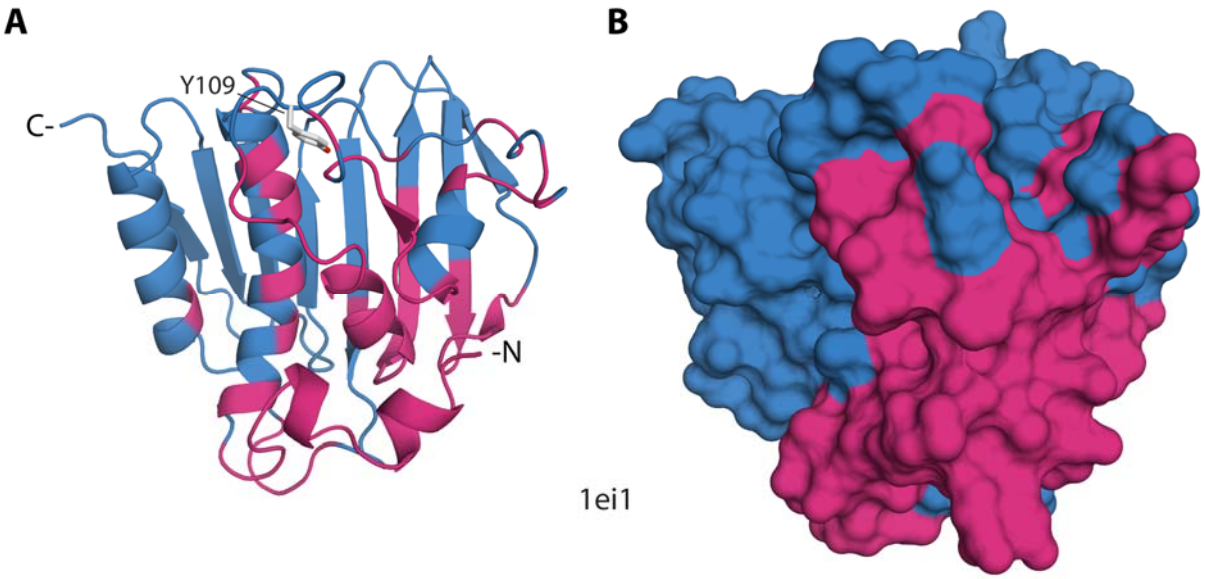


Figure 5-7 – Chemical shift perturbation and intensity drop data (shown in dark pink) for the GyrB/VbhTA complex, mapped onto the known GyrB complex X-ray structure (PDB ID: 3SHG) (27). In A, depicted in cartoon and in B, in surface representation. In A, the GyrB Y109 target residue is highlighted as stick representation.

Unlike the NmFIC titration, the observed chemical shift perturbation and intensity, when mapped onto the known structure of apo GyrB, evidenced a compact binding locus. Several residues comprise the dimerization interface, specifically on the N-terminal side of the GyrB protein, as well as several stretches around the ATP-lid loop containing the AMPylation target residue Y109.

AMP transferase and GyrB catalytic complex formation

Both protein assemblies were applied to a gel filtration column equilibrated with the complex-optimized buffer G. The chromatogram for the assembled $NmFIC_{mono}$/GyrB complex, was compared with each of the individual proteins. As shown in Figure 5-8, both pure $NmFIC_{mono}$ and GyrB elution profiles display individual monomeric peaks, as confirmed by SDS-PAGE. The gel filtration profile for the assembled $NmFIC_{mono}$/GyrB protein complex, matches the chromatogram sum of each of the individual proteins. This was once again confirmed by SDS-PAGE, where each protein is present in fractions from different elution peaks, running separately at a size consistent with the expected monomeric size.

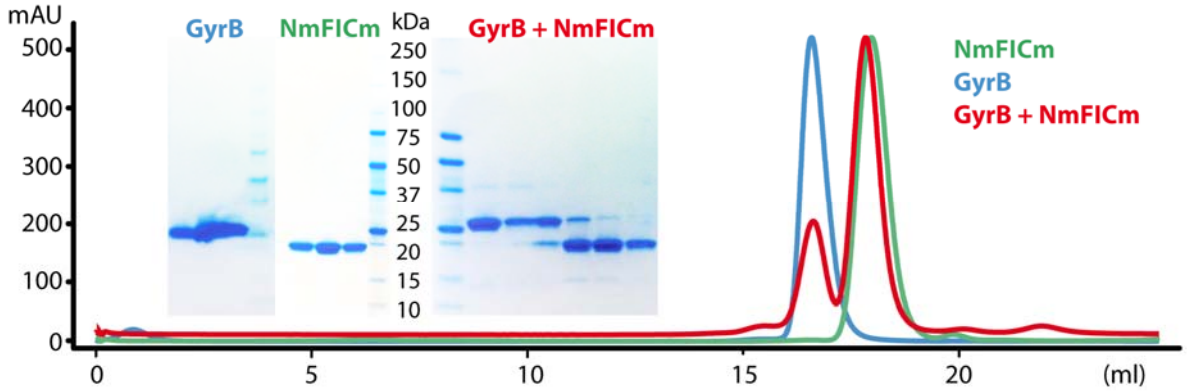


Figure 5-8 – Gel elution profiles as monitored by Abs_{280nm} of $NmFIC_{mono}$ and GyrB-complexes. Recorded at 8°C in buffer G on an S200 (10/300) size exclusion column for the indicated individual proteins and complexes. The inset shows SDS-PAGEs of the peak fractions of the individual size exclusion chromatography runs.

The VbhTA/GyrB complexes were similarly applied to a gel filtration column equilibrated with the optimized buffer G. As shown in Figure 5-9, both pure VbhTA and GyrB elute as individual monodisperse peaks. Conversely, the assembled VbhTA/GyrB complex elutes also as two peaks, but with different retention times when compared to the individual proteins, where the highest molecular weight peak, elutes at a volume consistent with the expected complex molecular weight. The second, lower weight peak, is explained by an excess of VbhTA complex proteins producing the monomeric peak. This behavior was further confirmed through the gel filtration of the reloaded VbhTA/GyrB complex peak, which shows only the highest molecular weight peak. Also, the SDS-PAGE of both complex elution fractions confirmed the co-elution of VbhTA/GyrB complex proteins. Both of these observations strongly indicate a tightly bound complex, under the described experimental conditions.

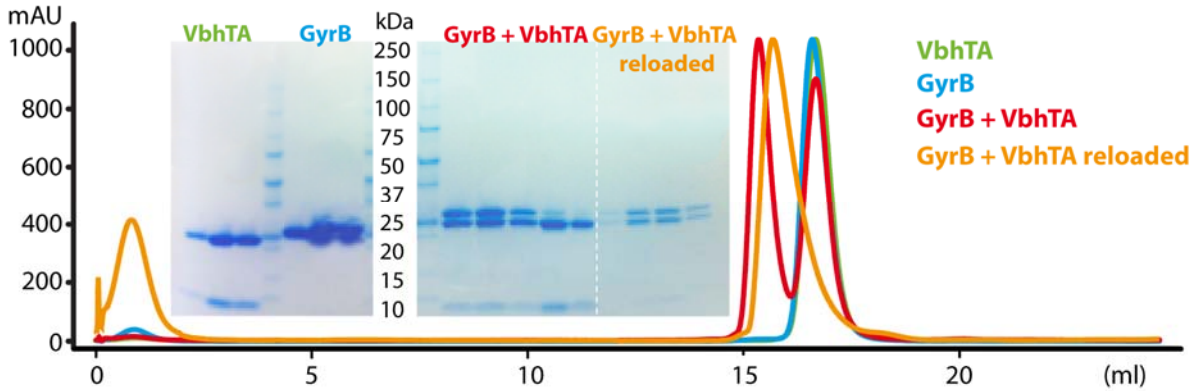


Figure 5-9 – Gel elution profiles as monitored by Abs_{280nm} of VbhTA and GyrB-complexes. Recorded at 8°C in buffer G on an S200 (10/300) size exclusion column for the indicated individual proteins and complexes. The inset shows SDS-PAGEs of the peak fractions of the individual size exclusion chromatography runs.

VbhTA/GyrB apo complex crystallization

Several initial hits were observed in the first week after plate set up, the corresponding buffer, salt, and precipitant conditions are summarized in Table 5-1. These conditions were used for the design of different fine optimization screens, where each buffer pH was varied in 0.5 steps below or above the observed values, while varying the precipitant concentration, maintaining the salt concentration.

Table 5-1 – Summary of initial crystal hit conditions.

Group	Buffer	Salt (M)	Precipitant
1	0.1 M sodium acetate pH 4-5	0.1 M calcium acetate; 0.2 M ammonium acetate;	10-20% (w/v) PEG 4k; 25% PEG 3350.
2	0.1 M trisodium citrate pH 4.5-5	None; 0.1 M magnesium chloride;	15-20% (w/v) PEG 4k; 20% (w/v) PEG 8k.
3	0.1 M HEPES pH 7-8	0.1 M magnesium chloride; 1.3 M lithium sulphate; 0.1-0.2 M NaCl.	None;15-24% (w/v) PEG 4k; 18% PEG 8k.
4	0.1 M MES pH 5.5	0.15-0.2 M ammonium sulfate;	20-25% (w/v) PEG 4k.
5	0.1 M sodium cacodylate pH 5-5.5	0.1 M calcium acetate.	12% PEG 8k.
6	0.1 M Tris-HCl pH 8-9	1-1.6 M lithium sulphate; 0.2 M NaCl	
7	0.1 M MOPS pH 7.5	0.1 M ammonium acetate.	12% PEG 8k.

Of the several tested conditions, mostly crystal needles or plates were observed. Reproducible 3D crystal growth was observed for the screen based on group 3 HEPES buffer conditions, specifically for conditions with $MgCl_2$ and NaCl as salt. Of the several fished crystals, the best diffracting crystals grew after a few days in condition: 0.1 M Hepes pH=7.4, 0.1 M $MgCl_2$ and 18% (w/v) PEG 8k. For crystal fishing and freezing, the cryo-protection was first optimized. Both perfluoropolyether and glycerol were tested, as well as a stepwise increase of glycerol through drop transfer, but ultimately a 20% glycerol/reservoir mixture showed the best results as assayed visually from crystal appearance (cracks or melting of crystals). The fished crystals were flash-frozen in liquid nitrogen for data collection. Crystallization of the complex proved to be successful (Figure 5-10 A and Table 5-2).

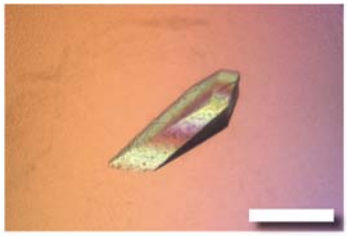


Figure 5-10 – Crystals of VbhTA/GyrB apo complex (0.1 M Hepes pH=7.4, 0.1 M $MgCl_2$ and 18% (w/v) PEG 8K). Scale bar: 200 nm.

Table 5-2 – Data collection and refinement statistics

	VbhTA/GyrB
Wavelength	1
Resolution range	42.19 – 2.01 (2.09 – 2.01)
Space group	P 2_1 2_1 2_1
Unit cell	76.81 84.07 84.38; 90 90 90
Total reflections	73611 (7156)
Unique reflections	36836 (3589)
Multiplicity	2.0 (2.0)
Completeness (%)	1.00 (0.99)
Mean I/sigma(I)	20.45 (1.95)
Wilson B-factor	24.82
R-merge	0.04241 (0.4745)
R-meas	0.05997 (0.469)
CC1/2	0.997 (0.799)
CC*	0.999 (0.969)
Reflections used in refinement	36826 (3586)
Reflections used for R-free	1906 (203)
R-work	0.1755 (0.3123)
R-free	0.2158 (0.3555)
CC(work)	0.962 (0.771)
CC(free)	0.942 (0.695)
Number of non-hydrogen atoms	4187
macromolecules	3667
Protein residues	458
RMS(bonds)	0.004
RMS(angles)	0.57
Ramachandran favored (%)	99
Ramachandran allowed (%)	0.67
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.5
Clashscore	3.97
Average B-factor	32.57
macromolecules	31.39
solvent	40.69
Number of TLS groups	14

Statistics for the highest-resolution shell are shown in parentheses.

VbhTA/GyrB apo complex structure

Of the tested crystals, ~50% showed a unit cell similar to the known VbhTA structure (PDB ID: 3SHG), whereas 50% revealed a space group and unit cell dimensions distinct from each of the known individual crystal structures. From the several collected datasets, for the novel crystal form, the highest resolution dataset was picked for data processing. Due to initial less than optimal crystal cryo-protection, the crystals of the apo-complex diffraction patterns displayed several ice rings, some of

which overlapped with diffraction spots areas. As such, some of the collected reflections were discarded, and the total maximal resolution lowered as a consequence. After successful molecular replacement with each of the known individual crystal structures, the solved structural model of the co-crystallized complex revealed a molecule of VbhTA bound onto GyrB (Figure 5-11) as previously thought (41).

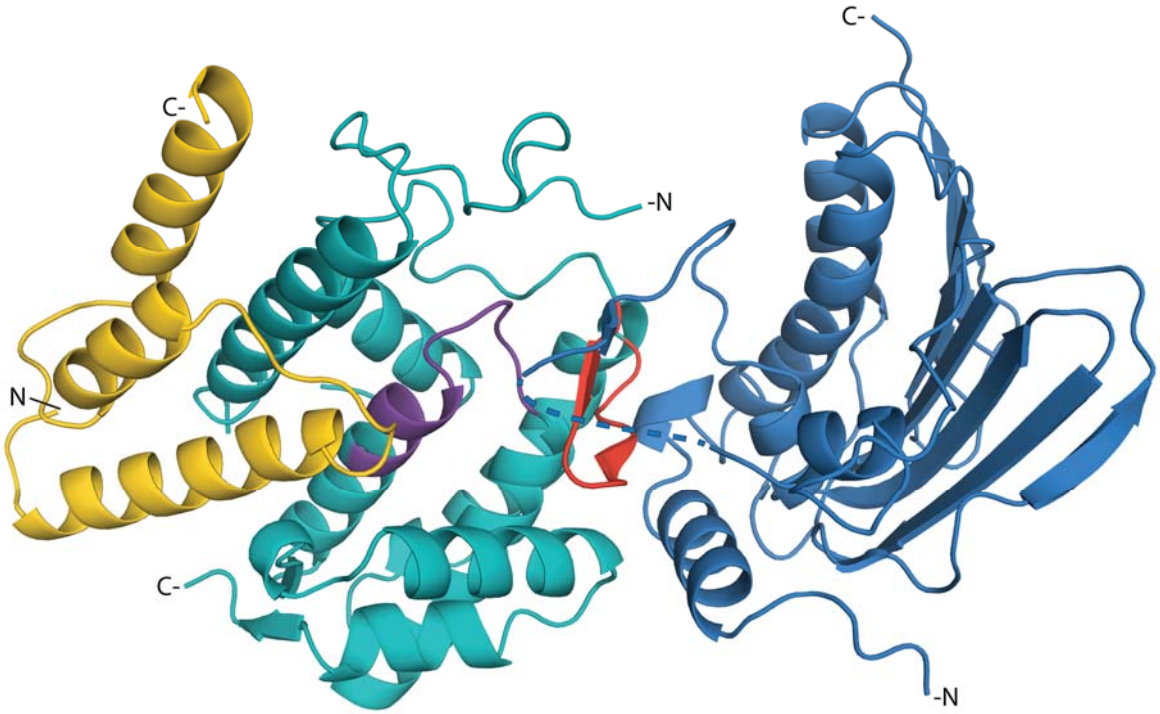


Figure 5-11 – Structural model of the VbhTA/GyrB co-crystallized complex. VbhT is colored in teal, VbhA in gold, and GyrB in blue, depicted in cartoon representation. The VbhT flap domain is highlighted in red, and the catalytic domain in purple.

As expected, from both the NMR titration and size exclusion elution profiles, the complex structure confirmed a 1:1 complex between VbhTA and GyrB. The VbhTA/GyrB complex structure revealed a complex binding interface consistent with the GyrB binding interface obtained from the NMR chemical shift perturbation data. The structural model reveals the overall conformation of each protein to be similar to the known individual crystal structures, with an RMSD of 0.306 for the VbhTA/complex backbone alignment (5 alignment cycles, 438 to 438 atoms out of 524 total atoms) and 0.59 for the GyrB/complex backbone alignment (5 alignment cycles, 356 to 356 atoms out of 454 total atoms) (2, 40). On the VbhTA side, upon GyrB binding, the N-terminal loop is displaced from the catalytic region in order to accommodate GyrB, as had been speculated before (Figure 5-12) (40). The N-terminal loop conformation changes from a position close to the catalytic site, to a solvent accessible conformation.

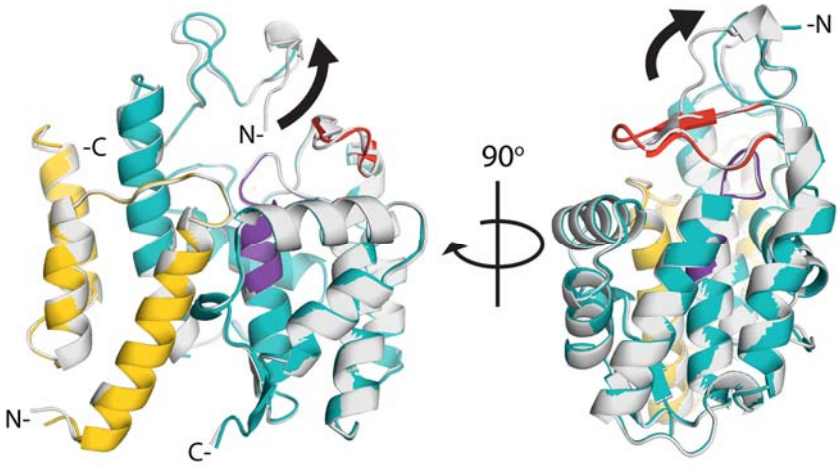


Figure 5-12 – Structure alignment between VbhTA (PDB ID: 3shg) shown in grey and VbhTA from the VbhTA/GyrB complex, in teal (VbhT) and gold (VbhA). The VbhT N-terminal conformational rearrangement induced upon complex binding is highlighted with arrows.

In the apo VbhTA crystal structure, the N-terminal residue is in position to establish two hydrogen bonds with the catalytic residues R139, and E140, as depicted in Figure 5-13. After complex formation, the targeted segment docking forces the N-terminal out of the catalytic pocket, becoming flexible as evidenced by the lack of electron density for residues 1-7.

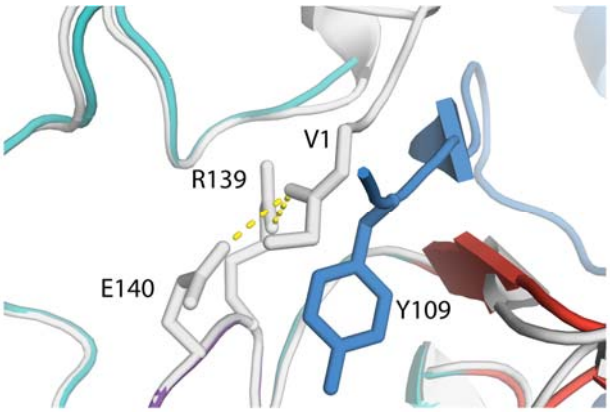


Figure 5-13 – Zoom in of the structure alignment between VbhTA (PDB ID: 3shg) shown in grey and VbhTA from the VbhTA/GyrB complex, in teal (VbhT) and gold (VbhA). The VbhT N-terminal and the two catalytic residues it establishes two intramolecular hydrogen bonds, as well as the target residue Y109, are highlighted in stick representation.

Furthermore, the flap domain maintains its overall unbound conformation, and no other significant structural differences were observed on the VbhTA side. However, on the GyrB side, two major conformation rearrangements take place. First, the N-terminal loop, part of the GyrB ATPase-dimerization interface, adopts a distinct conformation upon complex formation tilting away from the VbhTA/GyrB interaction interface, as evidenced in Figure 5-14.

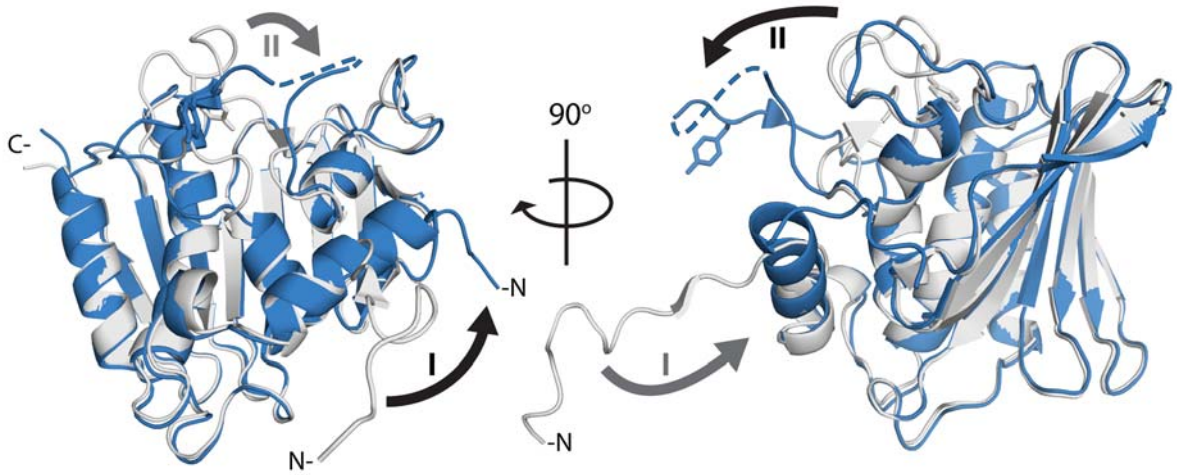


Figure 5-14 – Structure alignment between GyrB (residues 1-220 from PDB ID: 1EI1) shown in grey and GyrB from the VbhTA/GyrB complex, in blue, with the target residue Y109 shown in stick representation. In I, is indicated the GyrB N-terminal tilting; and in II, the ATP-lid loop catalytic approximation conformational rearrangements induced upon complex binding.

Second, the GyrB ATP-lid loop comprising residues 97-120, including the targeted segment with residue Y109, is displaced from the intrinsic ATP binding site towards the VbhTA catalytic pocket. In the catalytic pocket, the catalytic residue H136 establishes a hydrogen bond with the target residue Y109 as seen in Figure 5-15.

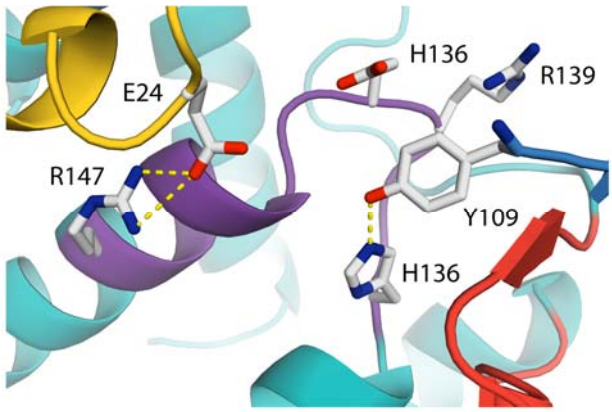


Figure 5-15 – Catalytic site of the VbhTA/GyrB complex. VbhT is colored in teal, VbhA in gold, and GyrB in blue, depicted in cartoon representation, while the catalytic and target residue side-chains are highlighted in stick representation.

Furthermore, in the catalytic pocket the previously described salt-bridge between the canonical FIC α_{inh} motif VbhA's residue E24 and VbhT's catalytic residue R147 responsible for competent ATP binding is also present (Figure 5-15) (2). The mechanism, through which VbhTA holds the GyrB target into place for its AMPylation activity, is revealed to be twofold. As the central binding motif, the VbhTA flap region target dock β-strand establishes an intermolecular antiparallel β-sheet-like structure with the targeted segment of GyrB (residues 108-113) evidenced in Figure 5-16.

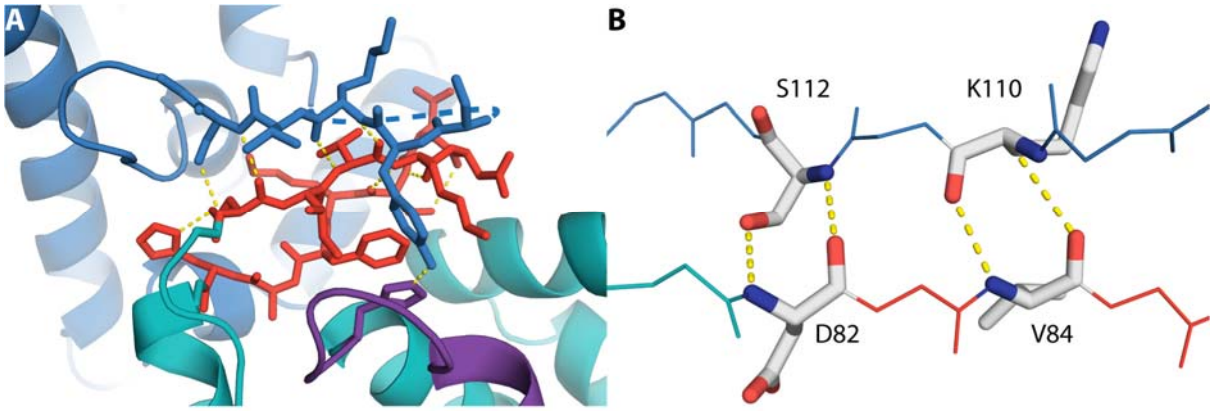


Figure 5-16 – Intermolecular antiparallel β-sheet between the VbhTA flap's target dock (red) and GyrB ATP-lid loop's targeted segment containing target residue Y109 (blue). In A, is shown the flap-lock in stick representation as well as intra- and intermolecular hydrogen bonds; in B, are shown the intermolecular hydrogen bonds between VbhT's target dock, and GyrB's targeted segment, with the interacting residues in stick representation, and the backbone in line representation.

The intermolecular hydrogen bonds established, lead to the β–augmentation of the flap region two β-strands, into a three-stranded antiparallel β-sheet. As shown in Figure 5-16 A, VbhT's target dock is in position to establish four hydrogen bonds between residues V84, D82 and GyrB's K110, S112, which hold Y109 firmly in place for the intermolecular hydrogen bond with residue H136. The electron density of GyrB's ATP lid loop residues 101-106 is poorly defined, evidencing its flexibility. Additionally, VbhTA holds GyrB in place through a second interaction: an additional intermolecular salt-bridge is established between VbhTA residue E131 and GyrB residue R22, shown in Figure 5-17.

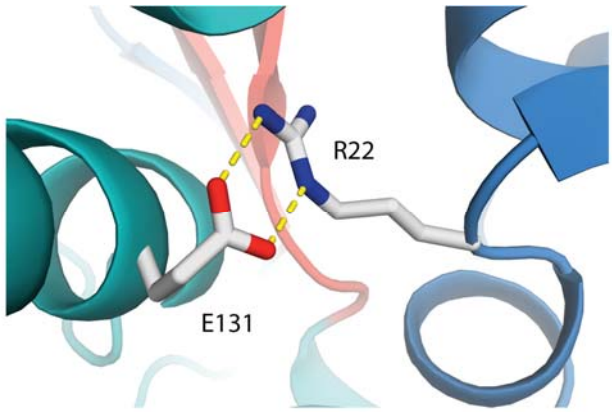


Figure 5-17 – Intermolecular salt-bridge between VbhTA residue E131 and GyrB residue R22.

Both of these interactions help position GyrB's targeted segment into the catalytic site of VbhTA, establishing the H136-Y109 hydrogen bond that initiates the catalytic transfer of the AMP moiety onto the Y109 residue. In fact, the superposition of the VbhTA/GyrB complex with the known structure of the GyrB ATP bound dimer shows that VbhTA binds precisely to the GyrB dimerization interface as depicted in Figure 5-18.

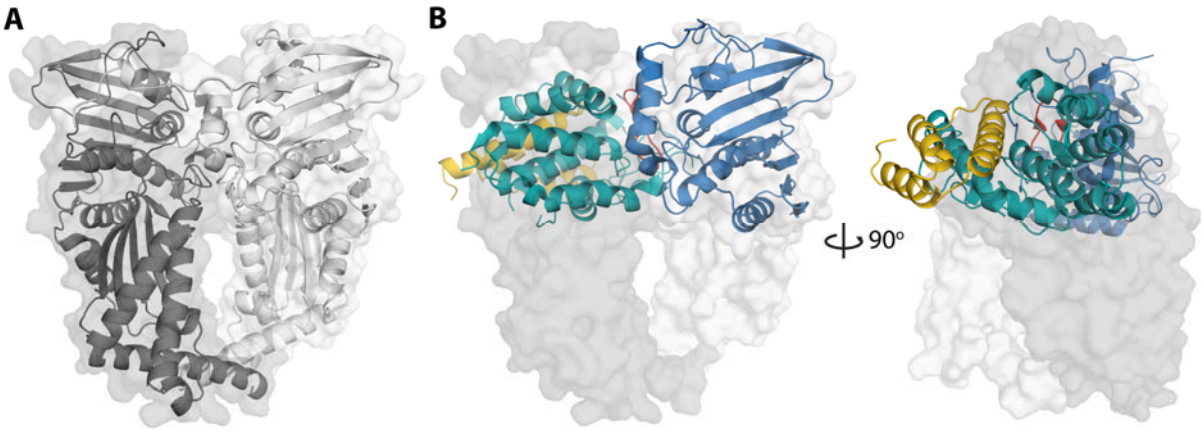


Figure 5-18 – Overlay of the VbhTA/GyrB co-crystallized complex with GyrB dimer (PDB ID: 1ei1). In A, is depicted the GyrB dimer colored dark grey and grey for each monomer, both in cartoon and surface representation. In B, is depicted the VbhTA/GyrB complex, where VbhT is colored in teal, VbhA in gold, and GyrB in blue, in cartoon representation, with GyrB dimer colored in dark grey and grey for each monomer, in surface representation.

The superposition of each complex reveals a steric clash between bound VbhTA and the second copy of GyrB, revealing an exclusive VbhTA complex interaction with the inactive monomeric form of GyrB.

VbhTA/GyrB nucleotide-bound complex crystallization

After the successful crystallization of the apo complex and crystal structure resolution, the description of both the complex interacting residues and the target protein conformational rearrangement into the catalytic site of VbhTA, the description of the AMP moiety transfer onto GyrB, became the next logical step. As such, co-crystallization with different nucleotides was attempted, in search for the nucleotide-bound forms. At first, for the nucleotide-bound complex, the protein complex was co-crystallized in the presence of either 5mM AMP or ATP, using a newly designed fine screen based on the above-described apo-complex crystal buffer conditions, varying both buffer pH (7.4-8), and PEG 8k concentration (15-22% (w/v)). Crystals were observed to grow preferentially at higher concentrations of PEG, and with a few exceptions, at lower pH. VbhTA/GyrB co-crystallized in the presence of AMP nucleotides, but not with ATP, as crystals grew in none of the tested conditions. Of the several AMP co-crystallized crystals, those with higher dimensions and finer edges were selected and cryo-protected with the improved cryo-protection protocol, using the respective reservoir solution with 20% Glycerol, and subsequently flash-frozen. The tested crystal diffraction patterns displayed no ice rings, leading to the elucidation of the AMP-bound structure at resolution of 1.45 Å. (Figure 5-19; Table 5-3).

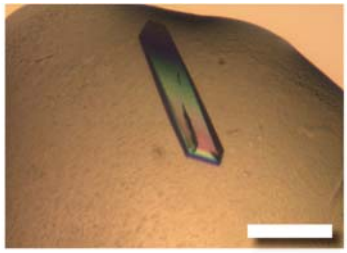


Figure 5-19 – Crystals of VbhTA/GyrB AMP-bound complex (0.1 M Hepes pH=7.8, 0.1 M $MgCl_2$ and 20.1% (w/v) PEG 8K). Scale bar: 200 nm.

Table 5-3 – Data collection and refinement statistics of nucleotide-bound complexes.

	VbhTA/GyrB/AMP
Wavelength	1
Resolution range	42.24 - 1.45 (1.50 - 1.45)
Space group	P 2_1 2_1 2_1
Unit cell	76.94 84.26 84.48; 90 90 90
Total reflections	844927 (78169)
Unique reflections	97299 (9567)
Multiplicity	8.7 (8.2)
Completeness (%)	1.00 (1.00)
Mean I/sigma(I)	26.18 (1.09)
Wilson B-factor	23.31
R-merge	0.03712 (1.673)
R-meas	0.03946 (1.787)
CC1/2	1 (0.56)
CC*	1 (0.848)
Reflections used in refinement	97246 (9553)
Reflections used for R-free	4780 (472)
R-work	0.1806 (0.3455)
R-free	0.2012 (0.3669)
CC(work)	0.962 (0.747)
CC(free)	0.966 (0.716)
Number of non-H atoms	4151
macromolecules	3757
ligands	12
Protein residues	463
RMS(bonds)	0.011
RMS(angles)	1.17
Ramachandran favored (%)	99
Ramachandran allowed (%)	1.1
Ramachandram outliers (%)	1.1
Rotamer outliers (%)	1.3
Clashscore	4.57
Average B-factor	39.51
macromolecules	39.06
ligands	33.77
solvent	44.12
Number of TLS groups	19

Statistics for the highest-resolution shell are shown in parentheses.

VbhTA/GyrB AMP-bound complex structure

The nucleotide-bound solved structure revealed the complex catalytic site to contain an AMP moiety bound in a non-competent state (Figure 5-20). The VbhA α_{inh} E24 salt-bridge with the VbhT catalytic R147 residue, as shown previously, was as well present. In addition, the catalytic R147 was found in position to establish a hydrogen bond with the AMPs adenosine O3, as opposed to the γ-phosphorus in the competent nucleotide binding expected conformation.

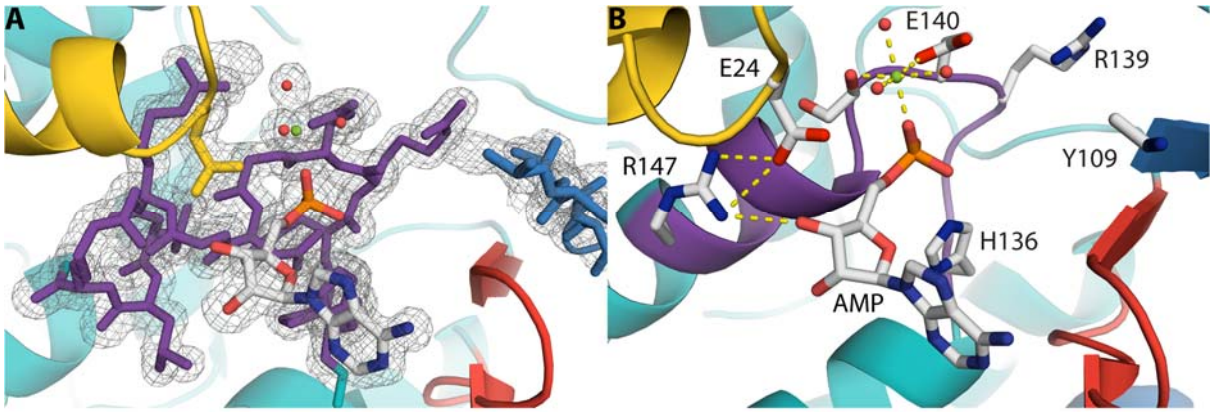


Figure 5-20 – Catalytic site of the VbhTA/GyrB/AMP co-crystallized complex structural model. VbhT is colored in teal, VbhA in gold, and GyrB in blue depicted in cartoon representation. The AMP moiety is depicted in stick representation, with the AMP-bound magnesium colored in green, and respective coordinating waters in red. In A, is additionally shown the Fo-Fc omit map at a contour level of 3.0 σ, while in B, the catalytic residues are highlighted in stick representation, as well as polar interactions between interacting atoms.

This result confirmed the supposed competent substrate binding inhibition role of α_{inh} VbhA E24, through intermolecular binding to VbhT's catalytic R147 residue. In addition, poor electron density for the Y109 residue's side-chain indicates its flexibility upon AMP binding. Furthermore, an additional AMP molecule was also found on GyrB's ATP catalytic site, as shown for the active GyrB dimer (40).

VbhTA$_{E24G}$/GyrB competent nucleotide-bound complex crystallization

The crystallization of the non-competent nucleotide-bound complex led to the pursuit of the crystallization of VbhTA$_{E24G}$ inhibition relieved mutant with GyrB, in the presence of ATP. The crystallization was attempted through both ATP co-crystallization and nucleotide soaking with the VbhTA$_{E24G}$/GyrB apo-complex crystals. Consequently, crystals of the inhibition relieved VbhTA$_{E24G}$/GyrB complex apo- and nucleotide-bound holo-states, were obtained using the previously described fine screen buffer in the absence and presence of 5mM ATP, respectively. For ATP soaking, a nucleotide soaking time course with 5mM ATP was performed with VbhTA$_{E24G}$/GyrB apo crystals. In both cases, leading to high-quality crystals diffracting with a resolution of up to 1.5 Å. (Figure 5-21; Table 5-4).

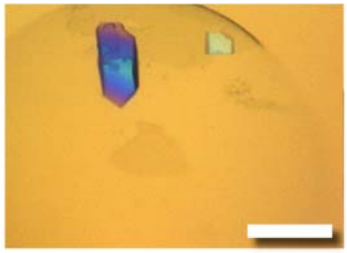


Figure 5-21 – Crystals of VbhTA$_{E24G}$/GyrB ATP co-crystallized complex (0.1 M Hepes pH=7.6, 0.1 M $MgCl_2$ and 21.3% (w/v) PEG 8K). Scale bar: 200 nm.

Table 5-4 – Data collection and refinement statistics for the nucleotide-bound complex.

	VbhTA$_{E24G}$/GyrB ATP
Wavelength	1
Resolution range	45.91 - 1.50 (1.56 - 1.50)
Space group	P 2_1 2_1 2_1
Unit cell	77.10 78.89 82.89; 90 90 90
Total reflections	1056174 (98563)
Unique reflections	81264 (7995)
Multiplicity	13.0 (12.3)
Completeness (%)	1.00 (0.99)
Mean I/sigma(I)	22.53 (1.30)
Wilson B-factor	21.40
R-merge	0.06426 (1.878)
R-meas	0.06694 (1.96)
CC1/2	1 (0.706)
CC*	1 (0.91)
Reflections used in refinement	81168 (7957)
Reflections used for R-free	1662 (165)
R-work	0.1848 (0.3422)
R-free	0.2077 (0.3607)
CC(work)	0.961 (0.833)
CC(free)	0.953 (0.849)
Number of non-H atoms	4220
macromolecules	3699
ligands	64
Protein residues	465
RMS(bonds)	0.008
RMS(angles)	0.99
Ramachandran favored (%)	98
Ramachandran allowed (%)	1.1
Ramachandram outliers (%)	1.1
Rotamer outliers (%)	1.3
Clashscore	5.94
Average B-factor	35.36
macromolecules	34.87
ligands	29.43
solvent	40.18
Number of TLS groups	22

Statistics for the highest-resolution shell are shown in parentheses.

VbhTA$_{E24G}$/GyrB competent nucleotide-bound complex structure

The comparison between the structures of apo and ATP-bound VbhTA$_{E24G}$/GyrB revealed an overall conformational similarity with an RMSD of 0.639 for the VbhTA/complex backbone alignment (5 alignment cycles, 1546 to 1546 atoms out of 1772 total atoms). The provisional structural model of the co-crystallized holo complex revealed an ADP moiety bound to the catalytic site of VbhTA$_{E24G}$, instead of the expected ATP (Figure 5-22).

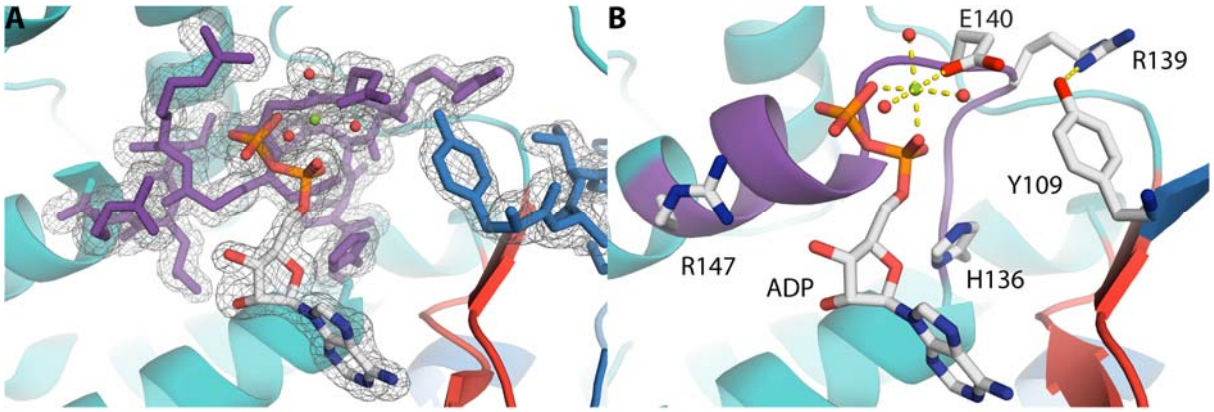


Figure 5-22 – Catalytic site of the VbhTA/GyrB/ADP co-crystallized complex structural model. VbhT is colored in teal, VbhA in gold, and GyrB in blue depicted in cartoon representation. The ADP moiety is depicted in stick representation, with the ADP-bound magnesium colored in green, and respective coordinating waters in red. In A, is additionally shown the Fo-Fc omit map at a contour level of 3.0 σ, while in B, the catalytic residues are highlighted in stick representation, as well as polar interactions between interacting atoms.

Interestingly, upon ADP binding, GyrBs Y109 AMPylation target changes its conformation away from the H136-bound as seen on the apo structure, to the R139 hydrogen bond available position, within the catalytic site. Additionally, one ADP molecule was also found on the ATP binding pocket of GyrB, consistent with the known nucleotide-bound structures (27, 40).

5.3.4. Discussion

The achievement of NmFIC$_{mono}$ backbone resonance specific assignment, discussed in chapter 5.2, paved the way towards the target complex titration studies using solution NMR spectroscopy. Chemical shift perturbation and intensity drop for the NmFIC$_{mono}$/GyrB NMR titration experiment evidenced the binding residues for both proteins, which when mapped onto the known structure helped identify the possible binding interface on each interaction partner, specifically on the NmFIC$_{mono}$ side, where a clear binding interface was observed. On the NmFIC$_{mono}$ side, binding interface expressly demonstrated the suspected role of the flap region on GyrB binding. Conversely, on the GyrB side, the binding interface was less clear. Both NmFIC E102R-E156R mutations introduced previously for the soluble study of the monomeric form of the protein, are situated in the here evidenced NmFIC/GyrB binding interface. The charge inversion of the E2R mutated residues could

induce charge repulsion between interacting residues of each protein, disturbing not only the tetramerization of NmFIC but target-complex formation as well. In addition, the H107A catalytic inactive mutation could further hinder complex interaction, due to the lack of the catalytical hydrogen bond, as shown before for the FIC-target IbpA/Cdc42 complex (42). Gel filtration of the assembled complex, revealed a curve sum of the individual monomeric species, further indicating an unstable complex in these conditions. It is then highly probable the introduced mutations affect target binding, explaining the mixed results.

For the VbhTA/GyrB, the binding interface was only followed on the GyrB side, due to unavailability of the VbhTA backbone assignment. Nevertheless, chemical shift perturbation and residue intensity drop mapping uncovered a compact complex binding interface on the N-terminal side of GyrB. Additionally, NMR titration studies allowed the optimization of GyrB-stabilizing buffer, resulting in the formation of a stable complex, confirmed by size-exclusion chromatography, and SDS-PAGE. After initial crystallization trials and condition optimization, the co-eluting complex was successfully crystallized. The VbhTA/GyrB complex structure, demonstrated the conformational changes each partner experiences upon complex formation. The VbhT N-terminal loop adopts a conformation where it opens up the catalytic site upon GyrB binding. On the GyrB side, as evidenced by the NMR-titrations, GyrB's N-terminus adopts a conformation away from the binding interface. GyrB's N-terminal loop is responsible for holding the active GyrB dimer in place upon ATP substrate binding, essential for GyrB catalytic dimer formation, ATPase activity, and enzyme turnover (40). The GyrB protein is then effectively held through β-augmentation between the VbhT flap loop's target dock and GyrBs ATP-lid targeted segment as an antiparallel β-sheet-like structure; as well as an intermolecular salt-bridge between two residues side-chains, tightly binding GyrB in place for AMPylation. This structural feature matches well with the observed distinct structural state observed with NMR-spectroscopy, showing slow chemical exchange on the NMR-time scale, as well as the known IbpA/Cdc42 complex (42). As a consequence, the antiparallel β-sheet directly forces GyrBs ATP targeted segment Y109 residue, into the VbhT's catalytic site for AMPylation, which abrogates the target's function. The fact that VbhTA binds precisely to the GyrB dimerization interface demonstrates that VbhTA exclusively binds the GyrB monomer, and not the dimer, as initially thought, where the ensuing AMPylation of GyrB's Y109 residue results in concomitant obstruction of GyrB's nucleotide binding, and consequent abrogation of active dimer formation.

The successful description of the apo-complex led to the procurement of the nucleotide-bound and AMPylated complexes. At first, the nucleotide-bound complex was crystallized in the presence of AMP. The resulting structure showed an AMP moiety bound to the catalytic site, albeit in a non-

competent manner, as the alpha-phosphate oxygen sits in the Y109 hydroxyl nucleotide unbound position, resulting in charge repulsion. VbhA antitoxin contains the canonical FIC α_{inh} motif, where residue E24 binds VbhT's R147 catalytic residue responsible for competent ATP binding. The physiological inhibition relief mechanism of the wild-type protein remains currently unknown. It is believed that translocation of the FIC-domain into the host cell, leads to the unfolding of the TA complex, where VbhT translocates unaccompanied, in a similar manner to the proposed colicin translocation unfolding mechanism (discussed in chapter 4.2) (19). *In vitro*, this effect is mimicked by an inhibition-relieved mutant, through the replacement of the glutamate residue to a glycine (2, 13). As such, further crystallization trials were performed using this inhibition-relieved mutant in the presence of ATP, resulting in the solution of a high-quality structure, albeit of the ADP-bound complex. Either residual ADP impurities in the ATP stock or ATP catalyzed by GyrB to ADP, selectively bound the $VbhTA_{E24G}$/GyrB complex, leading to the preferential crystallization of the ADP-bound over the ATP-bound complex. The catalytic site of each structure is shown in Figure 5-23.

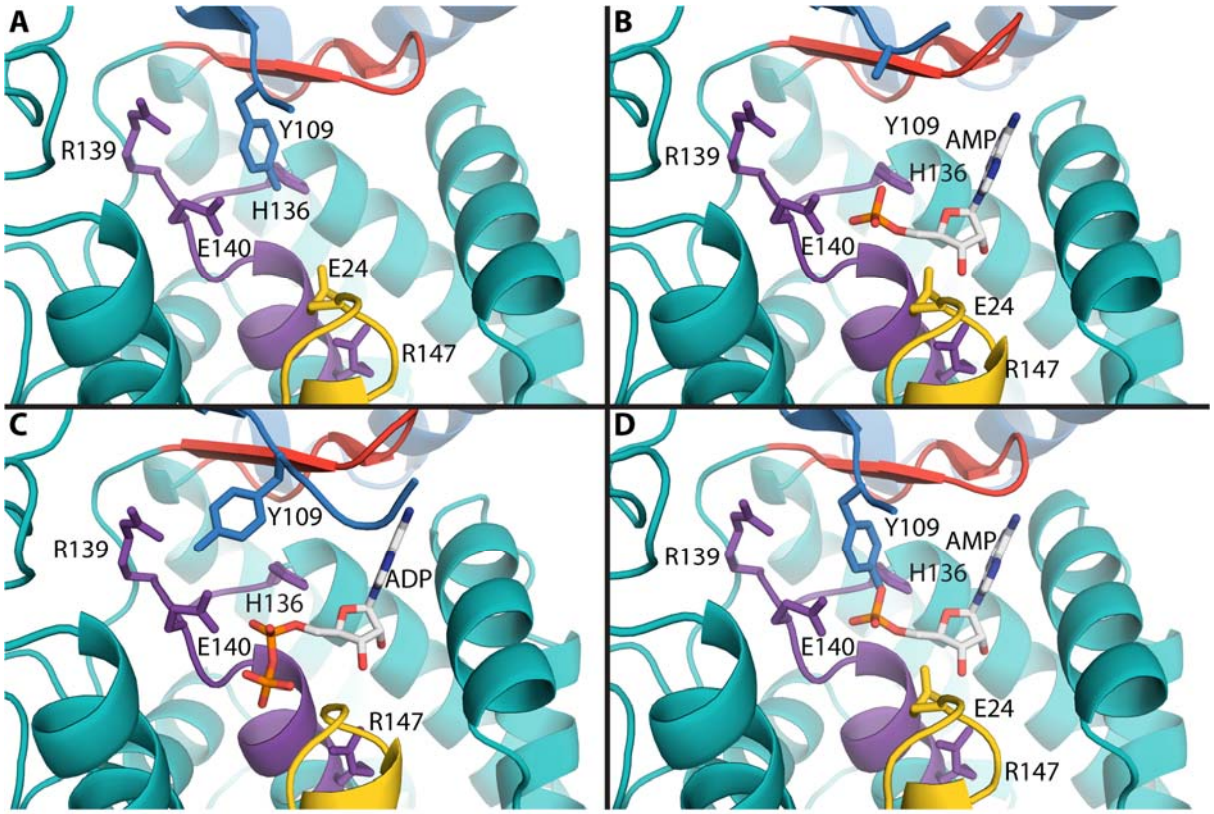


Figure 5-23 – Comparison between the catalytic sites of both apo and nucleotide-bound co-crystallized complexes. In A is shown the apo complex catalytic site, in B, the AMP-bound complex, in C the ADP-bound co-crystallized complex, and in D the superposition between AMP moiety (from the AMP-bound complex) with the apo complex, showing the possible AMPylation product structure. VbhT is depicted in cartoon representation colored in teal, VbhA in gold, and GyrB in blue, and the nucleotide in stick representation.

With the solution of apo-complex and AMP-bound structures (Figure 5-23 A-B), it became clear the role of the α_{inh} VbhA E24 residue on the competent binding of ATP, later demonstrated in the ADP bound structure (Figure 5-23 C). In Figure 5-23 D is depicted the possible structure of the AMPylated product complex, through the overlay of the apo complex's Y109 residue, and the AMP-bound complex AMP moiety, in good agreement with the known FIC-target IbpA/Cdc42 complex (42).

5.4. Conclusion

Solution NMR spectroscopy and X-ray crystallization were explored in an integrated fashion for the conformational and structural characterization of both NmFIC auto-AMPylation and VbhTA/GyrB complex. This approach allowed the biological contextualization of high-resolution crystal structures with NMR dynamical data, to unravel further details of each proteins physiological function. In both cases solution NMR spectroscopy allowed the study of the dynamic properties of each protein and protein complexes, helping uncover aspects such as NmFIC conformational changes induced upon automodification, GyrBs minimal construct and optimal conditions for complex formation; each of these crucial for the conformational characterization of each protein. Whereas X-ray crystallography, allowed the fast resolution of several high-quality atomic resolution structures, aiding the uncovering of each proteins structure, conformational changes, and their catalytic function.

5.5. References

1. V. Anantharaman, L. Aravind, New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.* **4**, 1-15 (2003).
2. P. Engel *et al.*, Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature*. **482**, 107–110 (2012).
3. S. Khater, D. Mohanty, *In silico* identification of AMPylating enzymes and study of their divergent evolution. *Sci Rep*. **5**, 1-17 (2015).
4. R. Utsumi *et al.*, Inhibitory effect of adenosine 3',5'-phosphate on cell division of *Escherichia coli* K-12 mutant derivatives. *J. Bacteriol.* **147**, 1105–1109 (1981).
5. C. A. Worby *et al.*, The Fic domain: Regulation of cell signaling by adenylylation. *Mol. Cell*. **34**, 93–103 (2009).
6. M. L. Yarbrough *et al.*, AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science.* **323**, 269–272 (2009).
7. P. Luong *et al.*, Kinetic and structural insights into the mechanism of AMPylation by VopS Fic domain. *J. Biol. Chem.* **285**, 20155–20163 (2010).
8. S. Mukherjee *et al.*, Modulation of Rab GTPase function by a protein phosphocholine transferase. *Nature*. **477**, 103–106 (2011).
9. A. S. Selyunin *et al.*, The assembly of a GTPase-kinase signalling complex by a bacterial catalytic scaffold. *Nature*. **469**, 107–111 (2011).
10. S. Mattoo *et al.*, Comparative analysis of *Histophilus somni* immunoglobulin-binding protein A (IbpA) with other Fic domain-containing enzymes reveals differences in substrate and nucleotide specificities. *J. Biol. Chem.* **286**, 32834–32842 (2011).
11. D. Castro-Roa *et al.*, The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat. Chem. Biol.* **9**, 811–807 (2013).
12. M. Rahman *et al.*, Visual neurotransmission in Drosophila requires expression of Fic in glial capitate projections. *Nat. Neurosci.* **15**, 871–875 (2012).
13. A. Goepfert, F. V. Stanger, C. Dehio, T. Schirmer, Conserved inhibitory mechanism and competent ATP binding mode for adenylyltransferases with Fic fold. *PLoS One*. **8** 1-9 (2013).
14. L. N. Kinch, M. L. Yarbrough, K. Orth, N. V. Grishin, Fido, a novel ampylation domain common to fic, doc, and AvrB. *PLoS One*. **4**, 1–9 (2009).
15. J. Y. Xiao, C. A. Worby, S. Mattoo, B. Sankaran, J. E. Dixon, Structural basis of Fic-mediated adenylylation. *Nat Struct Mol Biol*. **17**, 1004–1010 (2010).
16. D. Desveaux *et al.*, Type III Effector Activation via nucleotide binding, phosphorylation, and host target interaction. *PLoS Pathog.* **3**, 456-469 (2007).
17. D. V. Palanivelu *et al.*, Fic domain-catalyzed adenylylation: Insight provided by the structural analysis of the type IV secretion system effector BepA. *Protein Sci.* **20**, 492–498 (2011).
18. V. Campanacci, S. Mukherjee, C. R. Roy, J. Cherfils, Structure of the *Legionella* effector AnkX reveals the mechanism of phosphocholine transfer by the FIC domain. *EMBO J.* **32**, 1469–1477 (2013).
19. R. Schulein *et al.*, A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 856–861 (2005).
20. A. Harms *et al.*, Adenylylation of gyrase and topo IV by FicT toxins disrupts bacterial DNA topology. *Cell Rep.* **12**, 1497–1507 (2015).
21. D. Das *et al.*, Crystal structure of the Fic (filamentation induced by cAMP) family protein SO4266 (gi|24375750) from *Shewanella oneidensis* MR-1 at 1.6 Å resolution. *Proteins Struct. Funct. Bioinforma.* **75**, 264–271 (2009).
22. A. Garcia-Pino *et al.*, Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. *J. Biol. Chem.* **283**, 30821–30827 (2008).
23. A. Garcia-Pino, N. Zenkin, R. Loris, The many faces of Fic: structural and functional aspects

of Fic enzymes. *Trends Biochem. Sci.* **39**, 121–129 (2014).
24. R. Menzel, M. Gellert, The biochemistry and biology of DNA gyrase. *Adv.Pharmacol.* **29A**, 39–69 (1994).
25. C. A. Worby et al., The Fic domain: regulation of cell signaling by adenylylation. *Mol. Cell*. **34**, 93–103 (2009).
26. P. Engel et al., Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen bartonella. *PLoS Genet*. **7** 1-16 (2011).
27. F. V. Stanger, C. Dehio, T. Schirmer, Structure of the N-Terminal Gyrase B fragment in complex with ADP*Pi reveals rigid-body motion induced by ATP hydrolysis. *PLoS One*. **9** 1-13 (2014).
28. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94**, 12366–12371 (1997).
29. P. Guntert, V. Dotsch, G. Wider, K. Wüthrich, Processing of multidimensional NMR Data with the new software Prosa. **2**, 619–629 (1992).
30. R. L. J. Keller, *Computer aided resonance assignment tutorial. Cantina* (2004).
31. M. Bellanda *et al.*, Letter to the Editor : Backbone ^{1}H , ^{13}C and ^{15}N resonance assignment of the N-terminal 24 kDa fragment of the gyrase B subunit from *E. coli*. *J. Biom. NMR*. **22**, 369–370 (2002).
32. D. G. Waterman et al., The DIALS framework for integration software. Ccp4 Newsl. Protein Crystallogr. **4**, 16–19 (2013).
33. W. Kabsch, Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
34. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
35. A. Vagin, A. Teplyakov, *MOLREP* : an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
36. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
37. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
38. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D.* **66**, 213–21 (2010).
39. G. N. Murshudov *et al.*, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
40. L. Brino *et al.*, Dimerization of *Escherichia coli* DNA-gyrase B provides a structural mechanism for activating the ATPase catalytic center. *J. Biol. Chem.* **275**, 9468–9475 (2000).
41. F. V Stanger, A. Harms, C. Dehio, T. Schirmer, Crystal Structure of the *Escherichia coli* Fic toxin-like protein in complex with its cognate antitoxin. *PLoS One*. **11**, 1-21 (2016).
42. J. Y. Xiao, C. A. Worby, S. Mattoo, B. Sankaran, J. E. Dixon, Structural basis of Fic-mediated adenylylation. Nat Struct Mol Biol. **17**, 1004–1010 (2010).

Chapter 6

Appendix

6. Appendix

6.1. Protein sequences

GFP+ (28296 kDa; ε280: 21890 $M^{-1}cm^{-1}$):

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDH
MKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITA
DKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGIT
HGMDELYKKLAAALEHHHHHH

T7 RNAp (98942 kDa; ε280: 140260 $M^{-1}cm^{-1}$):

MRGSHHHHHHGENLYFQ/SMNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFE
RQLKAGEVADNAAAKPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFLQEIKPEAVAYITIKTTLACLTSADNTTV
QAVASAIGRAIEDEARFGRIRDLEAKHFKKNVEEQLNKRVGHVYKKAFMQVVEADMLSKGLLGGEAWSSWHKED
SIHVGVRCIEMLIESTGMVSLHRQNAGVVGQDSETIELAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGITGG
GYWANGRRPLALVRTHSKKALMRYEDVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPVEDIPAIEREE
LPMKPEDIDMNPEALTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVYAVSM
FNPQGNDMTKGLLTLAKGKPIGKEGYYWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQ
DSPFCFLAFCFEYAGVQHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEIL
QADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQP
AIDSGKGLMFTQPNQAAGYMAKLIWESVSVTVVAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCAVHWVTP
DGFPVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWAHEKYGIE
SFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFYDQFADQLHESQLDKMPALPAKGNLNLRDILESDFAF
A

FKBP12 (11.974 kDa; ε280: 9970 $M^{-1}cm^{-1}$):

MGSSHHHHHHLEVLFQ/GPGVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEV
IRGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE

Neisseria meningitidis FIC (22.143 kDA; ε280: 24410 $M^{-1}cm^{-1}$):

MHHHHHHMKSIDEQSLHNARRLFESGDIDRIEVGTTAGLQQIHRYLFGGLYDFAGQIREDNISKGGFRFANAMYL
KEALVKIEQMPERTFEEIIAKYVRMNIAHPFLEGNGRSTRIWLDLVLKKNLKKVVNWQNVSKTLYLQAMERSPVND
LRLRFLLKDNLTDDVDNREIIFKGIEQSYYYEGYEKG

human Abl kinase 1 T351I gatekeeper mutant (ε280: ? $M^{-1}cm^{-1}$):

MSSHHHHHHLEVLFQ/GPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCEAQTKNGQGWVPSNYITPV
NSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQRSISLRYEGRVYHYRINTASDGKLYVSSESRFNTLAEL
VHHHSTVADGLITTLHYPAPKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYSLTVAVKT
LKEDTMEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIIIEFMTYGNLLDYLRECNRQEVNAVVLLYMATQISSA
MEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAKFPIKWTAPESLAYNKFSIKSDVWAF
GVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPEGCPEKVYELMRACWQWNPSDRPSFAEIHQAFETMFQE
SSISDEVEKELGKQGV

Colicin Ia C-domain (21.093 kDa; ε280: 23950 $M^{-1}cm^{-1}$):

MGSSHHHHHHSSGENLYFQ/HMLEEKRKQDELKATKDAINFTTEFLKSVSEKYGAKAEQLAREMAGQAKGKKIRN
VEEALKTYEKYRADINKKINAKDRAAIAAALESVKLSDISSNLNRFSRGLGYAGKFTSLADWITEFGKAVRTENWRPL
FVKTETIIAGNAATALVALVFSILTGSALGIIGYGLLMAVTGALIDESLVEKANKFWGI

human Bax (21.322 kDa; ε280: 35980 $M^{-1}cm^{-1}$):

MKGSSHHHHHHSSGLVPRGS/HMDGSGEQPRGGGPTSSEQIMKTGALLLQGFIQDRAGRMGGEAPELALDPVP
QDASTKKLSECLKRIGDELDSNMELQRMIAAVDTDSPREVFFRVAADMFSDGNFNWGRVVALFYFASKLVLKALC
TKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTIFVAGVLTASLTIWKKMG

human GB1-Bax (21.785 kDa; ε280: 35980 $M^{-1}cm^{-1}$):

MSGSHHHHHHSSGIEGR/GRQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVT
ESSGENLYFQ/SGSHMTMDGSGEQPRGGGPTSSEQIMKTGALLLQGFIQDRAGRMGGEAPELALDPVPQDASTK
KLSECLKRIGDELDSNMELQRMIAAVDTDSPREVFFRVAADMFSDGNFNWGRVVALFYFASKLVLKALCTKVPELI
RTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTIFVAGVLTASLTIWKKMG

human SUMO-Bax (21.184 kDa; ε280: 35980 $M^{-1}cm^{-1}$):

MSGSHHHHHHHHHHGGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEM
DSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHLEQIGG/MDGSGEQPRGGGPTSSEQIMKTGALLLQGFIQDRAGR
MGGEAPELALDPVPQDASTKKLSECLKRIGDELDSNMELQRMIAAVDTDSPREVFFRVAADMFSDGNFNWGRV
VALFYFASKLVLKALCTKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDGLLSYFGTPTWQTVTIFVAGVLTASLTI
WKKMG

***Bartonella schoenbuchensis* VbhA (7.270 kDa; ε280: 1490 $M^{-1}cm^{-1}$):**

MLSEEEIEYRRRDARNALASQRLEGLEPDPQVVAQMERVVVGELETSDVIKDLMERIKREEI

Bartonella schoenbuchensis VbhT (24.160 kDa; ε280: 34380 $M^{-1}cm^{-1}$):

MHHHHHHMRKYEGSNDPYTDPETGVMYNLLGIKDQARLERVESAFAYIRSFELGRTSISGKFDLDHMKKIHKKLF
GDVYEWAGKTRLVDIVKDNSKFAHYTQIESYAPQITQQLAREQHLRGLDANEFSQRAGYYMGELNALHPFREGN
GRTLREFIWQLAREAGYHIDWDRVERQEMTRASIESYYGNSDLMSALIRRNLTEFT

Escherichia coli GyrB (25.097 kDa; ε280: 14440 $M^{-1}cm^{-1}$):

MHHHHHHMSNSYDSSSIKVLKGLDAVRKRPGMYIGDTDDGTGLHHMVFEVVDNAIDEALAGHCKEIIVTIHADN

SVSVQDDGRGIPTGIHPEEGVSAAEVIMTVLHAGGKFDDNSYKVSGGLHGVGVSVVNALSQKLELVIQREGKIHRQ

IYEHGVPQAPLAVTGETEKTGTMVRFWPSLETFTNVTEFEYDILAKRLRELSFLNSGVSIRLRDKRDGKEDHFHYEG

H – 6x or 10x His tag

ENLYFQ – TEV protease cleavage sequence

LEVLFQ – PreSc protease cleavage sequence

LVPRGS – Thrombin protease cleavage sequence

IEGR – FactorXa protease cleavage sequence

AA – SUMO tag

AA – bGB1 or GB1 tag

Abbreviations

2D, 2-dimensional

3D, 3-dimensional

2-ME, 2-mercaptoethanol

Abl, Abelson tyrosine kinase

AFM, atomic-force microscopy

ASC, apoptosis-associated speck-like protein containing a CARD

AMP, adenosine triphosphate

ARF, ADP-ribosylation factor

Bcr, breakpoint cluster region

bGB1, basic mutant of B1 domain of protein G from *Streptococcus sp*

BiP, Binding immunoglobulin protein

β-OG, beta-octylglucoside

β-ME, β-mercaptoethanol

BMCF, batch mode cell-free

Brij35, polyoxyethylene-(23)-lauryl-ether

bp, bair pairs

CARD, Caspase activation and recruitment domains

C-domain, channel domain

CECF, continuous-exchange cell-free

Cir, colicin I receptor

Col, colicin

COSY, correlation spectroscopy

cryo-EM, cryo-electron microscopy

DDM, n-dodecyl-D-maltoside

DHPC, 1,2-hexanoyl-sn-glycero-3-phosphocoline

diC7PC, 1,2-diheptanoyl-sn-glycero-3-phosphocoline

DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine

DMPG, 1,2-dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol)

DMSO, dimethylsulphoxide

DNA, deoxyribonucleic acid

Doc, Death on curing

DTT, 1,4-dithiothreitol

EDTA, ethylenediaminetetraacetic acid

ER, endoplasmatic reticulum

FIC, filamentation induced by cyclic AMP

FK-506, Tacrolimus

FKBP12, human peptidyl-prolyl cis-trans isomerase FK506 binding protein 1A, 12kDa

FM, feeding mixture

GB1, B1 domain of protein G from *Streptococcus sp.*

GFP, green fluorescent protein from Aequorea victoria

GK, gatekeeper

GSH, reduced glutathione

GSSG, oxidized glutathione

GTP, guanosine triphosphate

GyrB, *Escherichia coli* DNA topoisomerase IIa Gyrase B

HSQC, heteronuclear single quantum coherence

hAblK1GK, human Abelson kinase 1 T351I gatekeeper mutant

hLBP, human Lipopolysaccharide Binding Protein

HypE, Huntingtin yeast-interaction protein E

IbpA*, Histophilus somni* Immunoglobulin binding protein A

IM, inner membrane

IPTG, isopropyl-β-D-thiogalactopyranoside

kDa, kiloDalton

LB, Luria Bertani

LBP, lipopolysaccharide-binding protein

LPS, lipopolysaccharide

MAS, Magic-angle spinning

mRNA, messenger ribonucleic acid

mTORC1, mammalian target of rapamycin complex 1

MWCO, molecular weight cut-off

NmFIC, Neisseria meningitidis FIC protein

$NmFIC_{mono}$, *Neisseria meningitidis* FIC protein monomeric mutant

NMR, nuclear magnetic resonance

NOE, nuclear Overhauser effect

NOESY, nuclear Overhauser effect spectroscopy

OAc, acetate

OM, outer membrane

PCR, polymerase chain reaction

PFT, pore-forming toxin

Phd, Prevents host death

PMSF, phenylmethylsulfonylfluoride

ppm, parts per million

PreSc, PreScission

PTM, post-translational modification

R-domain, receptor domain

rf, radiofrequency

RM, reaction mixture

rmsd, root-mean-square deviation

RNA, ribonucleic acid

SAIL, stereo-array isotope labeling

SDS-PAGE, sodium dodecyl-sulfate polyacrylamide gel electrophoresis

SoFIC, Shewanella oneidensis FIC

T-domain, translocation domain

T7 RNAp, RNA polymerase from bacteriophage T7

TCEP, tris(2-carboxyethyl)phosphine

TEV, tobacco etch virus

TOCSY, total correlation spectroscopy

TritonX100, polyethylene-glycol P-1,1,3,3-tetramethyl- butylphenyl-ether

tRNA, transfer ribonucleic acid

TROSY, transverse relaxation optimized spectroscopy

VbhTA, *Bartonella schoenbuchensis* octopine Ti plasmid virulence B homologous T and A complex

VopS, *Vibrio parahaemolyticus* type III effector

Acknowledgements

I would like to acknowledge Prof. Dr. Sebastian Hiller, for accepting me as a PhD student. Also, to Prof. Drs. Daniel Müller and Timm Maier, for accepting to examine this work and be part of my PhD defense as co-referee and chair respectively. The great and super talented Postdocs in our lab: Dr. Björn Bürmann, for his infinite availability, general help and true support. Dr. Leonor Morgado, for her immense help, kind hear and specially for always believing in me. Dr. Sina Reckel for introducing me to cell-free protein expression and her supervision during the initial phase of my PhD. Dr. Lorenzo Sborgi, for pushing mostly the right buttons in me. Also Dr. Lichun Hee, Dr. Irena Bürmann, and Dr. Michael Zahn for their help and trust. Fellow PhD student Ricardo Adaixo and Leonildo Delgado for not only their professional help and confidence, but also for helping turn the late “night shift” into not only a bearable experience but also a fun one. And last but nonetheless, all the current and previous PhD students of the Hiller Lab that I had the chance to meet, and for the future students, if you’re reading this: in the end it will definitely be worth it!