A Structural View on Mechanisms of Bacterial Communal Life and Toxicity

---

Inaugural–Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Florian Tobias Lindemann

from Cottbus

2023

The presented work is based on experiments performed at the Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP), Department of NMR-supported structural biology in Berlin, Germany between April 2018 and June 2023 under supervision of Prof. Dr. Hartmut Oschkinat.

First reviewer:	Prof. Dr. Hartmut Oschkinat
Second reviewer:	Prof. Dr. Christian Freund
Date of defense	December 15th, 2023

Statutory Declaration / Eidesstattliche Erklärung

I hereby declare that I have written this thesis on my own, marked the sources of any quotations or content obtained otherwise and mentioned any other personal help by name. I also declare that I have not submitted the dissertation in this or any other form to any other institution as a dissertation.

Hierdurch versichere ich, dass ich meine Dissertation selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Die Dissertation ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Place and Date	Signature

Parts of the results presented in this thesis are published as

[1] Yvette Roske[#], **Florian Lindemann**[#], Anne Diehl[#], Nils Cremer, Victoria A. Higman, Brigitte Schlegel, Martina Leidert, Kristina Driller, Kürşad Turgay, Peter Schmieder, Udo Heinemann, and Hartmut Oschkinat, "TapA acts as specific chaperone in TasA filament formation by strand complementation", *Proc. Natl. Acad. Sci. USA* **17**, 120 (2023). DOI: 10.1073/pnas.2217070120

[2] Alexander Belyy[#], **Florian Lindemann**[#], Daniel Roderer, Johanna Funk, Benjamin Bardiaux, Jonas Protze, Peter Bieling, Hartmut Oschkinat, and Stefan Raunser, "Mechanism of threonine ADP-ribosylation of F-actin by a Tc toxin", *Nature Communications* **13**, 4202 (2022). DOI: 10.1038/s41467-022-31836-w

[#] these authors contributed equally

Summary

Nuclear magnetic resonance spectroscopy (NMR) is a versatile tool that can be applied to study the structure of biological systems both in solution and solids. These two approaches complement each other, particularly regarding proteins that can assemble into complexes of different sizes. The focus of this work is set on two bacterial proteins: (i) TasA, which is expressed by *Bacillus subtilis* and a vital component of its biofilm, and (ii) TcHVR, a toxic enzyme produced by *Photorhabdus luminescens* as part of its toxin complex (Tc).

Various bacteria form biofilms with proteinaceous elements, which are often of amyloid nature, and for *B. subtilis*, this main protein component is TasA. Extensive studies have been conducted on this protein, with different forms being described, and a monomeric structure of the protein was recently solved using X-ray crystallography. However, NMR, Electron microscopy (EM), and other biophysical methods have shown that TasA converts into a filamentous form and not an amyloid-like state within the biofilm environment.

Here, the structure of the filamentous form of TasA was analyzed using solid–state magic angle spinning (MAS) NMR. Through ^{1}H-detected experiments, an assignment of 87 % of all backbone NMR signals could be reached. The obtained chemical shifts facilitated the determination of dihedral angles and local secondary structure. Notably, the majority of structural elements of the monomeric structure is retained after the transition to filaments, but significant rearrangements happen at the termini. Most conformational changes occur in the N-terminal region, as confirmed by NMR experiments yielding through-space contacts. In this region, the observed signals pointed towards hydrogen bonding patterns typically found in anti-parallel β-sheets. By preparing filaments by mixing two differently labeled samples, it could be confirmed that these anti-parallel contacts result from inter-molecular interactions. In detail, the N terminus constitutes a newly formed strand β0 that is present exclusively in the filamentous form of TasA. β0 runs anti-parallel to β9 of an adjacent TasA molecule, creating the observed inter-molecular contact. The N-terminal rearrangement also explains the mechanism by which TapA, an accessory protein of TasA, is able to accelerate filament formation. The N terminus of TapA is homologous to that of TasA and can form a similar β0-strand which stimulates filament formation of adjacent TasA molecules.

The observed inter-molecular interaction pattern is characteristic of a polymerization mechanism called ‘donor-strand exchange’, prominently known for its role in the generation of type I pili by gram-negative bacteria. The diverse range of organisms (*e.g.*, *Escherichia coli*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa*) employing donor-strand exchange for biofilm formation indicates the potency of this mechanism. Understanding its impact on bacterial assemblies holds promise for selected biofilm control, either to combat bacterial infection or in plant protection.

Toxin complexes (Tc) are utilized by various bacteria to kill targeted cells. Their toxicity mechanism involves multiple steps: the initial translocation of a toxic enzyme into the target cell by a large assembly (the Tc) and the subsequent covalent modification of its target molecule, often actin, by the small toxic enzyme itself. In the context of *P. luminescens*, the proteins TcdA1, TcdB2, and TccC3 assemble to form a full Tc. This work focuses on a subcomplex formed by TcdB2 and TccC3 (TcB-TcC), which contains the toxic enzyme (referred to as TcHVR). The TcHVR sequence is located at the C-terminal end of TccC3 and, after assembly of the TcB-TcC complex, gets autoproteolytically cleaved and rests inside a shell formed by TcB-TcC until it is injected.

The full Tc has been extensively characterized using various methods, particularly cryo-EM, which focused on the conformational rearrangements needed for proper translocation. The structures and functions of the Tc components are well-known, but the precise conformation of the small TcHVR remained elusive. To address this, we employed solid–state and solution NMR techniques to investigate TcHVR, aiming at a model of the enzyme inside TcB-TcC and after translocation. Specifically, we aimed at determining the folding state of TcHVR inside the TcB-TcC complex and the molecular basis of its toxicity when within the targeted cell.

Investigating TcHVR by solution NMR, a resonance assignment rate of over 91 % of all NMR-active nuclei was achieved. The determined chemical shifts allowed for an analysis of the secondary structure patterns and, based on NOESY data of molecular contacts, the calculation of a structure ensemble. Certain regions of the enzyme, particularly the section adjacent to the binding pocket, exhibited characteristics typical of an ADP-ribosyltransferase. Chemical shift perturbations demonstrated that TcHVR binds to its cofactor, NAD^+, in a manner similar to known ADP-ribosyltransferases. Moreover, cryo-EM analysis of the F-actin bound state, combined with relaxation measurements performed by NMR, indicated significant changes in the NAD^+ pocket during the binding process. A flexible loop comprised of TcHVR residues 198-208 rearranges to accommodate NAD^+, and an ionic interaction between K185 and E265 controls the access to the cofactor pocket. By docking NAD^+ into the cryo-EM structure, we successfully modeled the intermediate ligand-bound state. Consequently, the findings obtained by NMR, cryo-EM, and modeling elucidate the entire process of ligand binding, conformational changes, and catalysis

mechanism of the ribosylation reaction inside the target cell.

The structure ensemble obtained for free TcHVR allowed us to analyze its state within the TcB-TcC subcomplex. We acquired a spectrum of TcHVR inside the TcB-TcC complex by employing ultra-fast MAS at frequencies exceeding 100 kHz. The estimation of signal completeness by analyzing the arginine region of the spectrum showed that signals of TcHVR are observable. By comparing the spectra of TcHVR inside TcB-TcC and the enzyme alone, we found that peaks from residues in the hydrophobic core of folded TcHVR are absent in the TcB-TcC spectrum. From their absence, we deduce that the hydrophobic core of TcHVR does not form when it is inside the shell. The lack of this distinct environment indicates that TcHVR remains unfolded and only adopts its functional state after translocation into the targeted cell.

Zusammenfassung

Die Kernspinresonanzspektroskopie (engl.: ‚Nuclear Magnetic Resonance', NMR) ist eine vielseitige Methode, die zur Untersuchung der Struktur biologischer Systeme genutzt werden kann. Dabei ist es möglich, NMR sowohl in Lösung als auch im festen Zustand (Festkörper-NMR) anzuwenden, wobei beide Ansätze sich ergänzende Erkenntnisse liefern. Dies ist insbesondere bei Proteinen hilfreich, welche sich zu Komplexen unterschiedlicher Größe zusammenschließen können. Der Schwerpunkt dieser Arbeit liegt auf zwei bakteriellen Proteinen, die genau das tun: (i) TasA, welches von *Bacillus subtilis* exprimiert wird und ein wichtiger Bestandteil seines Biofilms ist, und (ii) TcHVR, ein toxisches Enzym, das von *Photorhabdus luminescens* als Teil seines Toxinkomplexes (engl.: ‚toxin complex', Tc) produziert wird.

Eine große Anzahl an Bakterien bilden Biofilme, in die Proteine als wichtige Bestandteile eingebaut werden. Oft bilden diese dort strukturgebende Amyloide aus. Bei dem Bakterium *B. subtilis* übernimmt diese Rolle das Protein TasA. Bezüglich TasA wurden umfangreiche Studien durchgeführt, wobei verschiedene Formen beschrieben wurden, darunter auch eine amyloide. Außerdem konnte vor kurzem eine monomere Struktur des Proteins mit Hilfe von Röntgenkristallographie gelöst werden. Jedoch haben NMR, Elektronenmikroskopie (EM) und andere biophysikalische Methoden gezeigt, dass TasA in der Biofilmumgebung weder einen monomeren noch einen amyloidartigen Zustand annimmt, sondern vielmehr eine filamentöse Form.

In dieser Arbeit wurde die Struktur der filamentösen Form von TasA mit Hilfe der Festkörper-NMR analysiert. Hierbei wird die Probe um den magischen Winkel von 54,7° gedreht, weswegen man von MAS (engl.: ‚Magic Angle Spinning') NMR spricht. Mithilfe von ^{1}H-detektierten Experimenten konnten 87 % der Atome des Proteinrückgrats ein NMR-Signal zugeordnet werden. Die dadurch erhaltenen chemischen Verschiebungen ermöglichten die Bestimmung von Torsionswinkeln des Rückgrats und der lokalen Sekundärstruktur. Dabei konnte festgestellt werden, dass die meisten Sekundärelemente der monomeren Struktur nach dem Übergang zu Filamenten unverändert bleiben.

Allerdings deuteten die Daten auf erhebliche Umlagerungen an den beiden Enden der Proteinsequenz hin. Dabei finden die meisten konformationellen Änderungen in der

Region des N-Terminus statt, was durch Raumkontakte in dafür aufgenommenen NMR-Experimenten bestätigt werden konnte. In diesem Bereich wiesen die beobachteten Signale auf Wasserstoffbrücken hin, die typischerweise in antiparallelen β-Faltblättern zu finden sind. Durch die Nutzung von unterschiedlichen Markierungsmustern in der Herstellung der Filamente konnte gezeigt werden, dass diese antiparallelen Kontakte aus intermolekularen Wechselwirkungen resultieren. Im Detail betrachtet bildet der N-Terminus einen neuen Strang β0 aus, der ausschließlich in der filamentösen Form von TasA vorhanden ist. Der Strang β0 verläuft antiparallel zu dem Strang β9 eines benachbarten TasA-Moleküls, wodurch der beobachtete intermolekulare Kontakt entsteht. Die N-terminale Umstrukturierung erklärt auch den Mechanismus, durch welchen TapA, ein Hilfsprotein von TasA, die Filamentbildung beschleunigen kann. Der N-Terminus von TapA ist homolog zu jenem von TasA und kann einen ähnlichen β0-Strang bilden, der in gleichem Maße die Bildung von Filamenten von TasA-Molekülen induziert.

Das beobachtete intermolekulare Interaktionsmuster ist charakteristisch für einen Polymerisationsmechanismus, der als ‚Donor-Strang-Austausch' bezeichnet wird und vor allem für seine Rolle bei der Bildung von Typ-I-Pili durch gram-negative Bakterien bekannt ist. Die große Vielfalt der Organismen (z.B. *Escherichia coli*, *Acinetobacter baumannii* und *Pseudomonas aeruginosa*), die den Donor-Strang-Austausch für die Biofilmbildung nutzen, zeigt, wie effizient dieser Mechanismus sein muss. Ein besseres Verständnis des Donor-Strang-Austauschs sowie seiner Rolle in bakteriellen Kolonien ermöglicht eine gezielte Manipulation von Biofilmen, entweder zur Bekämpfung von Infektionen oder im Pflanzenschutz.

Toxinkomplexe (Tc) werden von vielen Bakterien genutzt, um ausgewählte Zielzellen zu töten. Der Mechanismus, mit dem diese Toxine agieren, besitzt zwei entscheidende Schritte: die anfängliche Injektion eines toxischen Enzyms in die Zielzelle durch ein großes Protein (den Tc) und die anschließende kovalente Veränderung des Zielmoleküls in der Zelle. Die Modifikation des Ziels wird durch das toxische Enzym herbeigeführt, welches vorher ein Teil des Tc war. Bei den Tc von *P. luminescens* formen die Untereinheiten TcdA1, TcdB2 und TccC3 einen vollständigen Komplex. In dieser Arbeit liegt der Fokus auf einem Unterkomplex, der lediglich von TcdB2 und TccC3 (TcB-TcC) gebildet wird und das toxische Enzym TcHVR enthält. TcHVR befindet sich am C-terminalen Ende der TccC3-Sequenz, wird nach der Bildung des TcB-TcC Komplexes autoproteolytisch abgespalten und ruht danach in einer von TcB-TcC gebildeten Hülle, bis es injiziert wird.

Der vollständige Tc wurde zuvor mit verschiedenen Methoden, insbesondere der Kryo-EM, umfassend charakterisiert. Dabei lag der Schwerpunkt auf lokalen Konformationsänderungen, welche für eine ordnungsgemäße Injektion des TcHVRs erforderlich sind. Die Strukturen und Funktionen der Tc-Komponenten sind daher bekannt, der genaue Zustand des kleinen Enzyms TcHVR ist jedoch noch unklar. Wir haben Festkörper- und Lösungs-

NMR-Techniken eingesetzt, mit dem Ziel, ein Modell des Enzyms innerhalb des TcB-TcC Komplexes, sowie nach der Injektion zu erstellen. Insbesondere der Faltungszustand von TcHVR innerhalb der TcB-TcC-Hülle und die molekulare Grundlage seiner Toxizität in der Zielzelle sollten bestimmt werden.

Bei der Untersuchung des TcHVRs mittels Lösungs-NMR wurden über 91 % aller NMR-aktiven Kerne ihren chemischen Verschiebungen zugeordnet. Die Zuordnung ermöglichte eine Analyse der lokalen Sekundärstrukturmuster und, basierend auf Informationen aus NOESY-Spektren über Molekülkontakte, die Berechnung einer Strukturschar. Bestimmte Bereiche des Enzyms wiesen typische Faltungsmuster einer ADP-Ribosyltransferase auf. Diese Charakteristik war insbesondere in der Region, welche der Bindungstasche angrenzt, stark ausgeprägt. Gemessene Änderungen der chemischen Verschiebung nach Titration mit dem Kofakor NAD^+ zeigten, dass TcHVR diesen auf ähnliche Weise wie bekannte ADP-Ribosyltransferasen bindet. Darüber hinaus wies eine Kryo-EM-Untersuchung des an F-Aktin gebundenen TcHVRs, in Kombination mit NMR-Daten, auf signifikante Veränderungen in der NAD^+-Bindungstasche während des Assoziationsprozesses hin. Ein flexibler Loop, der aus den Resten 198-208 des TcHVRs besteht, ändert seine Orientierung, um NAD^+ binden zu können. Außerdem liegt eine Ladungswechselwirkung zwischen K185 und E265 vor, welche den Zugang zur Bindungstasche kontrolliert und vor dem Andocken des Liganden gelöst werden muss. Durch eine Simulation der Position von NAD^+ in der Kryo-EM-Struktur konnten wir den Zustand von TcHVR mit gebundenem Liganden erfolgreich modellieren. Die durch NMR, Kryo-EM und computergestützte Simulation gewonnenen Erkenntnisse zeigen gemeinsam den Prozess der Ligandenbindung, der Konformationsänderungen und des Katalysemechanismus der Ribosylierung innerhalb der Zielzelle.

Die für das freie TcHVR erhaltene Strukturschar ermöglichte es uns, seinen Zustand innerhalb des TcB-TcC-Unterkomplexes zu untersuchen. Dafür haben wir ein Festkörper-NMR Spektrum des TcB-TcC-Komplexes mit dem darin enthaltenen TcHVR aufgenommen. Dabei wurde die Probe bei hohen MAS-Frequenzen von über 100 kHz vermessen, um ^{1}H-detektierte Experimente aufnehmen zu können. Die Vollständigkeit der Signale im Spektrum wurde durch eine Analyse des Argininbereichs des Spektrums abgeschätzt. Dadurch zeigte sich, dass die Signale des TcHVRs enthalten sein müssen. Mithilfe eines Vergleiches der Spektren von TcHVR im TcB-TcC-Komplex und des Enzyms allein konnten wir feststellen, dass die Signale der Aminosäuren aus dem hydrophoben Kern des gefalteten TcHVR im Komplex fehlen. Daraus schließen wir, dass sich der hydrophobe Kern des TcHVR innerhalb des Komplexes nicht bildet. Das Fehlen dieser besonderen Umgebung innerhalb des Proteins deutet darauf hin, dass das TcHVR ungefaltet bleibt und erst nach der Injektion in die Zielzelle seinen funktionellen Zustand annimmt.

Contents

List of Abbreviations

2D, 3D	two-dimensional, three-dimensional
ART	ADP-ribosyltransferase
BMRB	Biological Magnetic Resonance Data Bank
CA	amino acid backbone C_α
CB	amino acid sidechain C_β
CO, C’	amino acid backbone carbonyl
CP	cross polarization
cryo-EM	cryogenic electron microscopy
CSA	chemical shift anisotropy
CSI	chemical shift index
CSP	chemical shift perturbation
DARR	dipolar-assisted rotational resonance
DREAM	dipolar recoupling enhanced by amplitude modulation
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
EM	electron microsycopy
EPS	exopolysaccharides
HVR	hypervariable region
HSQC	heteronucler single quantum coherence
INEPT	insensitive nuclei enhanced by polarization transfer
ITC	isothermal titration calorimetry
MAS	magic angle spinning
MD	molecular dynamics
MOLP	medium optimal for lipopeptide production
NAD^+	nicotinamide adenine dinucleotide oxidized form
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	nuclear overhauser effect spectroscopy
MSA	multiple sequence alignment
PDB	Protein Data Bank
RF	radiofrequency

RFDR	radio frequency-driven dipolar recoupling
RMSD	root-mean-square deviation
TapA	TasA anchoring/assembly protein
TasA	translocation-dependant antimicrobial spore component
Tc	toxin complex
TcART	toxin complex ADP-ribosyltransferase domain
TcB-TcC	toxin complex TcdB2-TccC3 subcomplex
TcHVR	toxin complex hypervariable region
TMS	tetramethylsilane
WHO	World Health Organization

1. Introduction

1.1. Bacterial infections are increasing worldwide

After the discovery of penicillin in the 1920s [3], mortality rates due to infectious diseases had consistently decreased [4]. However, in recent years the evolutionary pressure on bacteria has led to many strains resistant to one or even multiple antibiotic agents [5]. As a result, infections and deaths caused by bacteria are rising again and are projected to pose a severe public health risk in developing and developed countries [6]. To counter the ever-growing armada of resistant strains, the World Health Organization (WHO) has published a list of priority pathogens [7] requiring urgent research. In recent years, approaches for the discovery of antibiotic agents have diversified, but the process has become increasingly difficult [8]. Therefore the overall approach to treatment and drug target design needs to be put under scrutiny. Non-protein targets have become attractive [9] and methods previously thought to be outdated (*e.g.*, the use of bacteriophages [10]) are again under serious consideration.

To comprehend bacterial virulence, it is essential to investigate the formation of bacterial assemblies that have a role in promoting pathogenicity. The objective here is to explore these mechanisms at a molecular level. This fundamental understanding will not only facilitate the development of novel therapeutic strategies but also enable the customization of benign biotechnological methods. In this thesis, we will focus on two model systems to analyze these aspects:

- TasA, the main component of *Bacillus subtilis* biofilms used by the bacterium to defend against external factors such as antibiotics [11], and
- the Tc toxin complex from *Photorhabdus luminescens*, employed to attack insect cells and whose potential for target delivery was recently shown [12]

To conduct a thorough analysis of these two systems, it is important to take into account both historical information and recent research findings pertaining to them.

1.2. Bacteria gain advantages by forming biofilms

Many bacteria routinely form multicellular communities stabilized by a self-produced matrix [13]. These assemblies, termed biofilms, comprise one or multiple species and form in numerous settings, either by sticking to surfaces or floating in liquids. They can adapt to harsh conditions and naturally occur on rocks, in the soil (*e.g.*, on plant roots), in hydrothermal deep-sea vents, and in hot springs [14], to name a few examples. Biofilms are not malicious *per se* but can cause problems when they arise in industrial (pipes, ship hulls) [15] and medical settings (teeth, catheters, implants) [13, 14]. There, treatment is hindered through reduced penetration of external agents [11] (*e.g.*, antibiotics), enhanced growth, and metabolic coordination [16]. Indeed, many WHO-designated priority pathogens [7] are notorious for biofilm formation, especially in hospital (nosocomial) settings. A few noteworthy examples are gram-negative *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and gram-positive *Staphylococcus aureus* (often abbreviated as MRSA for 'methicillin-resistant *S. aureus*') [17].

The formation of biofilms is also a characteristic of the bacteria whose proteins are examined in this thesis. The non-pathogenic soil bacterium *Bacillus subtilis* is an established biofilm model system, and the components needed for its proper formation have been a subject of intense study [18]. *Bacillus subtilis* is known to form biofilms on solid surfaces (Fig. 1.1A) and at the air-liquid interface of non-perturbed liquids where they are called 'pellicles' (Fig. 1.1B). Next to TasA, other minor protein components are TapA [19], acting as an accessory protein for TasA, and BslA [20, 21], which contributes to the hydrophobicity of the biofilm.

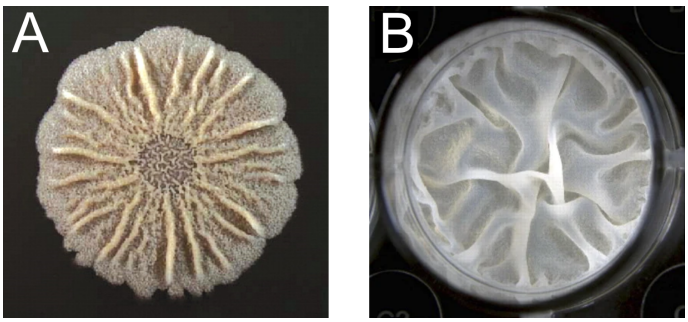

Fig. 1.1.: *Bacillus subtilis* biofilms. (**A**) Biofilm formed by wildtype *Bacillus subtilis* after growth for 3 days on an agar plate. Adapted from [13]. (**B**) A floating biofilm (pellicle) constituted by wildtype *Bacillus subtilis* after 48 hours of growth on MOLP medium at 30°C. Adapted from [22].

As for *Photorhabdus luminescens*, this bacterium is also known to form biofilms [23], but these are not as thoroughly studied yet. All strains of the genus *Photorhabdus* have a dual life cycle where they alternate between the gut of *Heterorhabditis* nematodes (where they live mutualistic) and tissues of other hosts, primarily insects, where they act

pathogenic [24]. These bacteria have never been found outside their biotic environment, and therefore their biofilms are highly specialized. The central role of the *Photorhabdus luminescens* biofilm is likely adherence in the gut of the nematodes and the midgut epithelium in the insect host. Interestingly, the dual life cycle of *Photorhabdus luminescens* has led to multiple biotechnological applications. Its toxicity towards insects is used in plant protection [25], and the antibiotics it produces to shield the nematode host and itself are toxic to otherwise resistant bacteria [26, 27].

1.3. TasA, the main protein in *Bacillus subtilis* biofilms

1.3.1. The genus *Bacillus*

Bacillus subtilis is a well-characterized, gram-positive, and rod-shaped bacterium that occurs in many natural contexts. It has been known since the 19th century and belongs to the genus *Bacillus*, one of the most diverse genera in the class *Bacilli* with more than 250 named species. Its bacteria typically have a low guanine/cytosine content in their DNA [28] and have been the source of many discoveries directly relevant for humans.

Bacillus cereus produces cereulide, a cyclic polypeptide responsible for food poisoning [29], and *Bacillus anthracis* is the cause of anthrax, a severe infection transmissible to humans. *B. anthracis* was discovered by Robert Koch in 1876, and he described its endospore formation in detail [30]. Another endospore-forming example is *Bacillus thuringiensis*, whose insecticidal proteins are used in plant protection [31]. With the advent of genetic modification, crops with inserted *B. thuringiensis* genes have shown significant potential for yield increase, especially in developing countries [32]. Of course, the bacteria can also be manually administered to surface of the plants, where they either proliferate or remain dormant as endospores until the external conditions improve.

Sporulation is a process commonly encountered in *Bacillus* species. Due to its benevolent nature, *B. subtilis* has become the model system for the formation of endospores [33]. In addition to its role in the study of biofilms and endospore-formation, *B. subtilis* is nowadays an essential expression system for medical and industrial enzymes as well as small molecules (*e.g.*, vitamins, hyaluronic acid, or antibiotics) [34]. The produced enzymes play, among other applications, an important role in the detergent, textile, leather, and paper industries. A noteworthy example is subtilisin, which has found broad applications as a detergent additive in cleaning products and in the pharmaceutical and food industry [35]. Due to the presence of *B. subtilis* being allowed in food products in the USA, its proteases can be used for various applications (*e.g.*, meat tenderization and milk coagulation). Furthermore, *B. subtilis* is also applied as a probiotic additive to animal food. It is often administered in the form of resistant endospores that reactivate in the intestinal tract and can help with the degradation of complex carbohydrates [34]. In addition, active *B. subtilis* bacteria help to counter intestinal pathogens through the secretion of antimicrobial peptides.

1.3.2. Characteristics and functions of TasA

The protein TasA was first described as a component of *B. subtilis* endospores in 1999 [36, 37]. It was found that, upon recombinant expression in *Escherichia coli*, TasA presented antimicrobial effects towards a wide variety of other bacteria [37]. This activity and its pres-

ence in endospores led to its characterization as a 'translocation-dependent antimicrobial spore component', or TasA in short. Since then, it was found that the antimicrobial activity was likely an *in vitro* artifact caused by other molecules. The absence of bactericidal properties appears reasonable as TasA these days is routinely produced through recombinant expression in *Escherichia coli* and is known to promote cross-talk and bacterial coordination in its native biofilm environment.

In 2006, TasA was shown to be a key constituent of the *B. subtilis* biofilm. Deletion strains lacking the *tasA* gene do not form a proper biofilm but instead an unstable layer [38]. The observed phenotype was similar to *B. subtilis* lacking the genes responsible for sugar (exopolysaccharides, EPS) production in the biofilm context (Δ*eps*). Remarkably, the deletion of the *tasA* gene can be rescued by supplying Δ*tasA* bacteria with recombinantly produced TasA protein. Upon supplementation, the wildtype phenotype is partially restored [19, 22]. To examine the influence of TasA, we study biofilms of *B. subtilis* that form floating layers at the air-liquid interface of non-perturbed liquids (so-called 'pellicles').

Location and regulation of the *tasA* gene

The gene for TasA is located in the same operon as the ones for TapA, previously called YqxM, and SipW [37]. TapA was initially described in 2011 [19] and is the abbreviation of 'TasA anchoring/assembly protein'. It is known to interact with TasA, which is further detailed below. SipW is a signal peptidase that is situated in the membrane and acts on both TasA and TapA, cleaving their signal peptides [39]. After maturation, the 27 N-terminal amino acids of TasA [36, 37] and the initial 43 of TapA [1] are removed. Interestingly, the regulation of the *tapA-sipW-tasA* operon is under positive control of the master regulator Spo0A [40] which is in turn under control by SinI and SinR [41]. Indeed, the Spo0A, SinI, and SinR are also responsible for the expression of the EPS genes *epsA* to *epsO* [42], which places the protein and sugar components of the biofilm under equal regulation. Interestingly, *B. cereus* appears to possess a similar genetic organization for biofilm proteins as *B. subtilis*. When the homologue genetic region in *B. cereus*, which encodes SipW, two biofilm proteins CalY1 and CalY2, and not characterized hypothetical protein HypP (*sipW-calY2-hypP-calY1* [43]), is replaced with *B. subtilis tapA-sipW-tasA*, then biofilm formation is restored in a manner indistinguishable from the wildtype [44].

In nature, the genetic regulation can adapt to external stimuli [45, 46]. Thiazolyl peptides secreted by *B. cereus* can cause increased expression of biofilm proteins in *B. subtilis* [47]. In a similar manner, *B. subtilis* reacts to the presence of the plant pathogen *Fusarium culmorum* by increasing TasA production [48]. Both these examples illustrate the cross-talk in the native context and the crucial role of TasA.

Structure of the TasA protein and interplay with TapA

Monomeric structures of TasA and TapA are publicly available (Fig. 1.2A and B), but these do not represent the functional state under biofilm conditions. Biofilms of other bacteria often contain proteins in an amyloid-like state (*e.g.*, CsgA curli of *Escherichia coli* [49]) and, accordingly, a similar role was long assumed for TasA. The designation of TasA assemblies ranges from 'amyloid-like' [50] over 'functional amyloid' [51] to 'bacterial amyloid' [52].

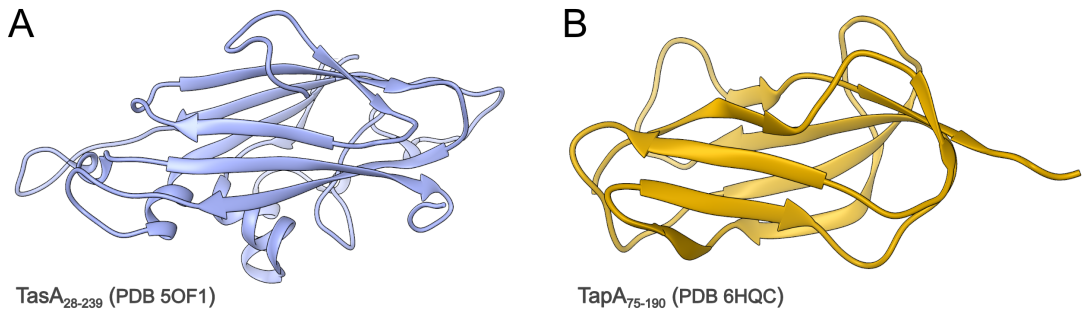

Fig. 1.2.: Structures of TasA and TapA from *Bacillus subtilis*. (**A**) Structure of TasA solved by X-ray crystallography (PDB 5OF1). The C-terminal residues after K239 had to be removed to facilitate crystal formation. (**B**) X-ray structure of the folded domain from TapA (PDB 6HQC). Here, the functionally relevant N-terminal section 44-74 and all C-terminal residues after E190 are not contained in the structure.

The most extensive work in this direction was presented in a number of papers published by Romero et al. [19, 22, 50, 52]. In these, TasA was unfolded through the usage of low pH by the addition of trifluoroacetic or formic acid and subsequently left for refolding at neutral pH. Using this treatment, TasA forms fibers that are visible in electron microscopy (EM) images and are stainable by Thioflavin-T [22]. This dye is commonly used as an indicator of amyloid presence [53] and consequently, amyloid characteristics were attributed to TasA. Under these conditions, the addition of TapA to unfolded TapA accelerated the formation of amyloid fibers, detected in a faster increase of Thioflavin-T fluorescence intensity [50].

However, some studies found that TasA can also occupy a non-amyloid polymeric state. Research conducted by Chai et al. [54] and Erskine et al. [55] showed that TasA can polymerize into a different state not stainable by Thioflavin-T. These 'filaments' form when TasA is kept at neutral pH and at room temperature, conditions which are resembling the biofilm environment. In parallel to the work conducted for this thesis, Böhning et al. [56] examined TasA filaments by atomic force microscopy. The results reinforced the notion that filaments formed by TasA contribute to *Bacillus subtilis* biofilms. Furthermore, our research of TasA filaments by analytical ultracentrifugation [1] showed that their

formation can be stimulated through the addition of TapA as well. For this stimulation, the N terminus (44-55) of TapA is essential.

1.4. Origins of toxicity of *Photorhabdus luminescens*

Next to insecticidal toxins produced by the bacterium *Bacillus thuringiensis* [57], proteins expressed by bacteria from the genera *Photorhabdus*, *Xenorhabdus*, and *Serratia* have long shown potential for insecticidal applications [25]. Bacteria from these three genera are known to live mutualistic with entomopathogenic nematodes and directly contribute to insect lethality. Promising areas of applications include the use in agriculture [58] and mosquito control [59].

Among other virulence factors with potency against insects, these bacteria produce a family of toxins called toxin complexes (Tc in short), which were first described in the year 1998 to be present in *Photorhabdus luminescens* [60]. Genes encoding these toxin complexes are also found in the genomes of other bacteria, including *Bacillus thuringiensis* [61], *Penibacillaceae* [62], and even some human pathogens (*e.g.*, *Yersinia pseudotuberculosis*, *Yersinia pestis*) showing a potential involvement of toxin complexes in human diseases [63]. Furthermore, parts of the toxin complex possess structural homology to teneurins occurring in vertebrate neurons [64, 65], highlighting the broad impact of understanding the structure and function of these assemblies.

1.4.1. Characteristics of bacterial toxin complexes

Toxin complexes are composed of three subunits, termed TcA, TcB, and TcC. The corresponding genes are located in four loci and encode a high number of different variants [66, 67]. Currently, there are seven versions of TcA-type and a further seven TcC-type genes known for the strain TT01 of *Photorhabdus luminescens* alone, whereas TcB genes show less variability than TcA and TcC [68]. Upon expression of the genes, it was found that the TcA subunit alone already shows a certain degree of toxicity. However, the observed effect is potentiated when all three (ABC) subunits are present [69]. These properties are due to the different structures and functions realized by each subunit.

TcA forms a large, homo-pentameric (> 1 MDa) bell-like structure readily observed in cryo-EM micrographs. The superstructure is composed of a central channel surrounded by receptor-binding domains at the periphery [70]. TcB and TcC together form a complex of about 300 kDa that encapsulates the actual cytotoxic enzyme (Fig. 1.3). The toxic moiety is located at the C-terminal end of the TcC sequence. Well-studied variants of TcC with toxic activity are TccC3 and TccC5 [71]. The C-terminal end gets autoproteolytically cleaved from the remainder of the TcC subunit, and afterwards the toxic moiety rests within the TcB-TcC complex [70, 72]. As the toxin section shows a low conservation between different TcC genes, it is often referred to as the ‘hypervariable region’, or TcHVR in short [73]. Upon docking of TcB-TcC to TcA and formation of the holotoxin, a high-

affinity interaction between TcA and TcB triggers a conformational change. The TcB-TcC complex opens and initiates the translocation of the TcHVR into the channel formed by TcA (Fig. 1.3). With this, the ABC holotoxin is fully assembled.

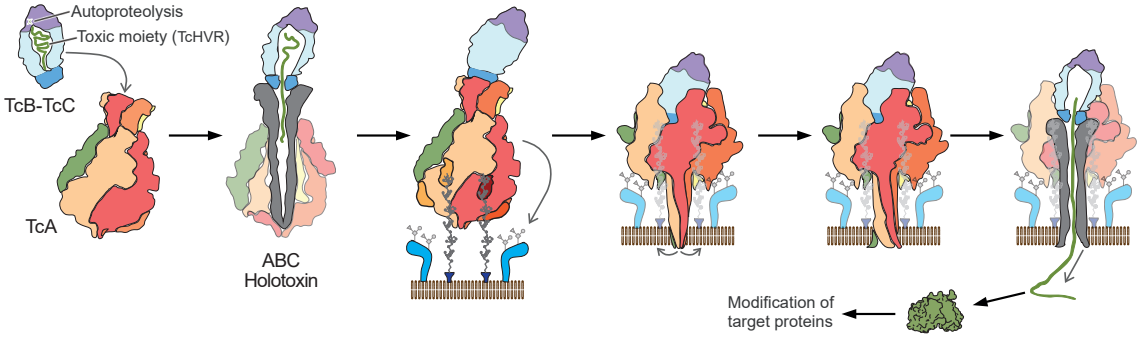

Fig. 1.3.: Mechanism of toxin complex assembly. In order to confer toxicity, toxin complexes need to undergo a series of conformational changes.The homo-pentameric TcA subunit initially assembles with the TcB-TcC complex that carries the actual toxin (TcHVR). The toxic enzyme is part of the TcC subunit before getting autoproteolytically cleaved. Subsequently, TcHVR stays encapsulated inside the TcB-TcC complex. The fully assembled ABC holotoxin docks onto the target cell and injects toxic moiety through conformational changes. During this process, TcA acts like a syringe by first penetrating the membrane of the target cell and injecting the toxin afterwards. Upon arrival in the target, TcHVR adopts its functional structure and harms the cell by modifying target proteins.

After secretion by the bacterium, the ABC toxin complex interacts with glycans and glycosylated receptors on the target cell surface [74]. Afterwards, the TcHVR is translocated into the cytoplasm of the target cell by a syringe-like motion of TcA [75]. It penetrates the membrane of the target and releases the toxin into the cytoplasm. Afterwards, the toxin is expected to either fold or conduct its function as a disordered protein in the target cell. Finally, the toxic activity interferes with essential cellular processes, ultimately leading to cell death. For TccC3 and TccC5, this process occurs through targeted ribosylation.

1.4.2. Mechanism of ADP-ribosylation

The family of ADP-ribosyltransferases (ARTs) is diverse and includes a wide range of toxins from human pathogens, including *Bordetella pertussis*, *Mycoplasma pneumoniae*, *Corynebacterium diphtheriae*, *Clostridium botulinum*, *Streptomyces coelicolor*, and *Vibrio cholerae* [76]. All enzymes classified as ARTs use the cofactor Nicotinamide Adenine Dinucleotide (NAD^+) (Fig. 1.4A) to apply a post-translational modification. During that process, NAD^+ is hydrolyzed into nicotinamide and ADP-ribose, with the latter being covalently attached to the target [77]. For the modification of proteins, the resulting covalent bonds are diverse, from the ADP-ribose being attached to the sidechain nitrogen of arginine, glutamine, asparagine, the sulfur of cysteine, and even the carboxyl carbon of

glutamate [78]. Based on the conservation of enzyme residues critical for their function, ARTs are typically classified into H-Y-E or R-S-E clades. However, some toxins do not belong to either clade. A few example structures of ARTs from the R-S-E clade are shown in Fig. 1.4B. They are characterized by a perpendicular set of β-sheets that border the binding pocket together with two α-helices and the three crucial amino acids arginine, serine, and glutamate in the active center.

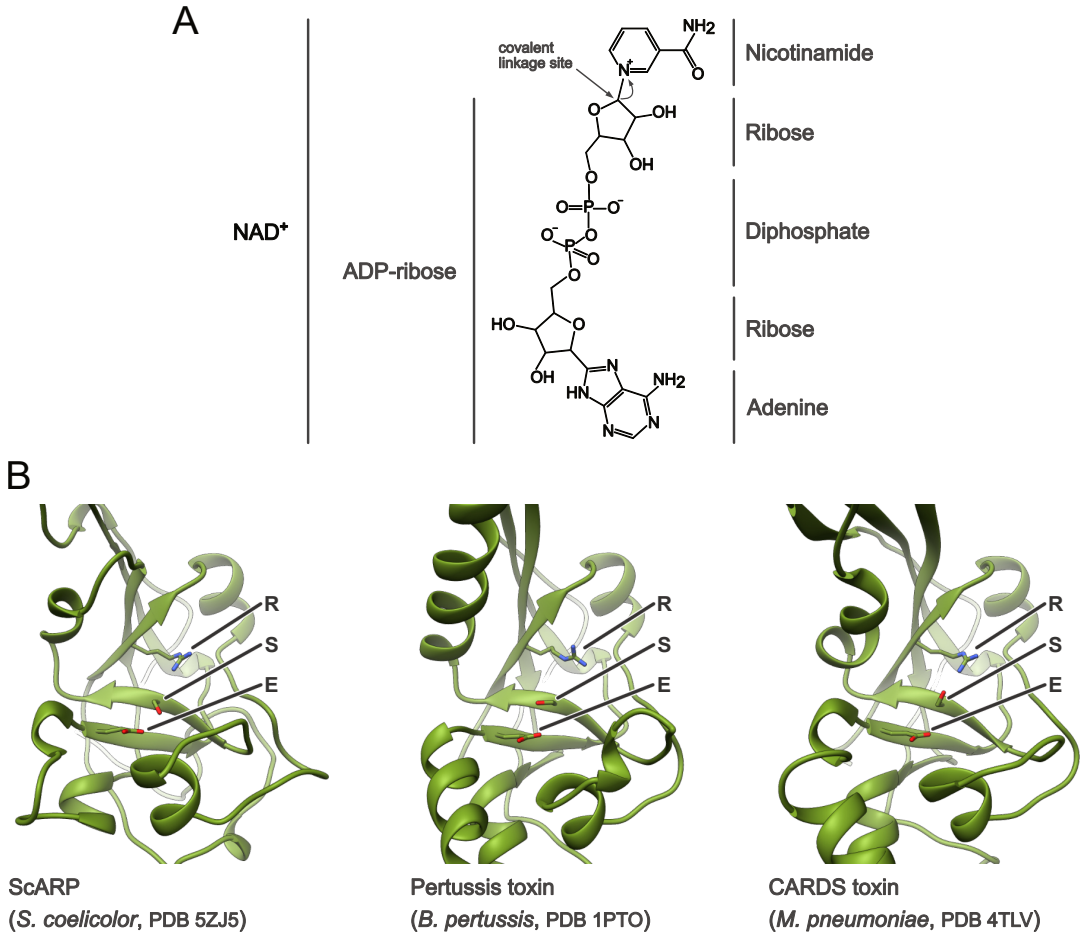

Fig. 1.4.: Characteristics of NAD⁺ and ADP-ribosyltransferases. (**A**) Lewis structure of NAD⁺ with its different chemical groups indicated. Additionally, the site where the covalent linkage to the target molecule occurs and the successive electron pair reorientation are highlighted by arrows. (**B**) View into the active site of selected ADP-ribosyltransferases from the R-S-E clade with the conserved amino acids labeled. Shown are ScARP from *Streptomyces coelicolor* (PDB 5ZJ5) [79], the pertussis toxin from *Bordetella pertussis* (PDB 1PTO) [80], and the CARDS toxin from *Mycoplasma pneumoniae* (PDB 4TLV) [81].

1.4.3. The TccC3 toxic enzyme modifies actin by ADP-ribosylation

For the toxic enzyme of TccC3, TcHVR, the toxicity is achieved by ADP-ribosylation of actin. Recent findings indicate that TcHVR selectively targets F-actin and covalently attaches the ADP-ribose at T148 [71, 82], likely through O-linkage to the hydroxy group. The modification promotes aberrant actin polymerization and causes actin clustering and aggregation, leading ultimately to cell death [83]. However, the molecular mechanism and structural basis of toxin action need to be better understood. No structure of free TcHVR is available, and during the initial stages of the translocation (when TcHVR is inside the ABC holotoxin), only badly resolved density was detected [70, 72, 75, 84]. As a consequence, TcHVR has remained elusive so far.

1.5. Nuclear Magnetic Resonance

Upon exposure to an external magnetic field B_0, a number of chemical isotopes respond to stimulation by electromagnetic waves in the radio frequency range (10^6–10^9 Hz), an observation known as nuclear magnetic resonance (NMR).

The following pages will introduce relevant principles, formulas, and applications for biological purposes that were harnessed for the work presented in this thesis. Detailed examinations of the subjects presented can be found in "Nuclear Magnetic Resonance" by Peter Hore [85] and "Spin dynamics: Basics of nuclear magnetic resonance" by Malcolm Levitt [86]. The basics of solid–state NMR are nicely presented in "Solid–state NMR Spectroscopic Methods in Chemistry" by David Laws, Hans-Marcus Bitter, and Alex Jerschow [87]. A good primer on the current state of solid–state NMR is the recent review "Solid–state NMR spectroscopy" by Bernd Reif et al. [88].

1.5.1. Basic principles

The ability of a nucleus to react to a magnetic field is defined by its spin quantum number I, whereas all nuclei with $I \neq 0$ interact. The magnitude of I determines the number of energy levels ($2 \cdot I + 1$) that arise upon exposure to an external magnetic field B_0. Nuclei with $I = \frac{1}{2}$ are most commonly used ('spin-$\frac{1}{2}$ nuclei'). With $2 \cdot \frac{1}{2} + 1 = 2$, two energy levels that manifest when the nucleus is exposed to B_0 (presented in Figure 1.5). Typically the two states are denoted α and β. The idea of two energy levels is comparable to the nucleus being a bar magnet that occupies minimal energy by aligning itself either parallel or anti-parallel to the axis of an external magnetic field. In general, NMR-active nuclei that have been brought into B_0 are referred to as possessing magnetization.

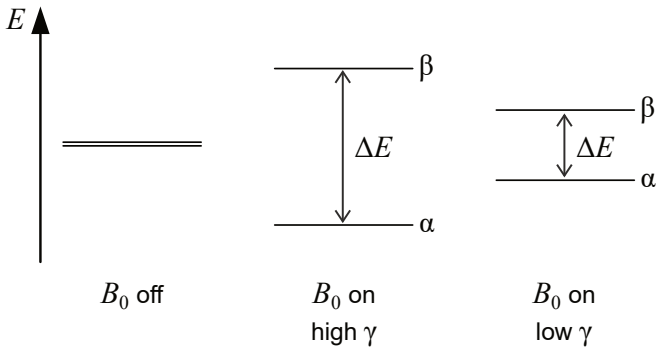

Fig. 1.5.: Energy levels of spin-$\frac{1}{2}$ nuclei. When brought into an external magnetic field B_0, nuclei with spin quantum number $I = \frac{1}{2}$ can occupy two energy levels (Zeeman splitting). The arising difference ΔE is exploited in NMR measurements.

The difference between the two energy levels is proportional to the gyromagnetic ratio γ, an empirical, nucleus-specific parameter originating from the internal structure of

the nucleus. Parameters of isotopes prominently used in routine NMR experiments are summarized in Table 1.1. ^{1}H nuclei are most prominently used for biological purposes and are often simply called 'protons'. Sensible detection of other, less prevalent isotopes (e.g.^{13}C, ^{15}N) usually requires them to be incorporated through chemical or biological production procedures, which also allows for site-specific labeling. As noted in the last row of Table 1.1, the B_0 used for biological NMR measurements (typically 10-20 T) are orders of magnitudes stronger than the earth's magnetic field of 30-60 μT. The strength is often not referred to in Tesla but instead in ^{1}H Larmor frequency (*e.g.*, a '400 MHz spectrometer').

Tab. 1.1.: Common isotopes in biological NMR. Spin quantum number I, gyromagnetic ratio γ, and natural abundance of selected nuclei used for biological NMR. Values from [85].

Isotope	Spin Quantum Number I	Gyromagnetic ratio γ in 10^7 T^{-1} s^{-1}	Natural Abundance in %	Larmor frequency ν at 9.4 T in MHz
^{1}H	$\frac{1}{2}$	26.75	99.985	400.0
^{2}H / D	1	4.11	0.015	61.4
^{13}C	$\frac{1}{2}$	6.73	1.108	100.6
^{15}N	$\frac{1}{2}$	-2.71	0.370	40.5
^{19}F	$\frac{1}{2}$	25.18	100.0	376.5
^{31}P	$\frac{1}{2}$	10.84	100.0	162.1

From the gyromagnetic ratio γ and the magnitude of B_0, the absolute transition energy ΔE between different levels can be calculated by

$$\Delta E = \hbar|\gamma|B_0 \tag{1.1}$$

$\hbar$ being the reduced Planck constant $\frac{h}{2\pi}$. Using the Planck-Einstein-relation ($\Delta E = h\nu$) the transition can be characterized in terms of frequency units:

$$h\nu = \hbar|\gamma|B_0 \Rightarrow \nu = \frac{|\gamma|B_0}{2\pi} \tag{1.2}$$

Equation (1.2) establishes the Larmor frequency ν as a value to quantify ΔE, as shown in Figure 1.5. Waves with defined lengths and intensities ('pulses') that possess a frequency matching ν can stimulate a transition between energy levels. In other words, they cause a change from a low-energy to a high-energy state. These transitions are detected as a signal in the NMR experiment.

However, compared to other spectroscopic techniques, NMR is often considered insen-

sitive. This property arises from the comparable, low-occupancy difference between the two energy levels (denoted as N_α and N_β). The relative distribution of the spins between the two states is described by the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = \exp\left(-\frac{\Delta E}{k \cdot T}\right) = \exp\left(-\frac{\hbar \gamma B_0}{k \cdot T}\right) \tag{1.3}$$

Here, k denotes the Boltzmann constant, and T the absolute temperature. As can be seen from the equation (1.3), the distribution of spins at equilibrium is proportional to the strength of the external magnetic field B_0 and the gyromagnetic ratio γ and inversely proportional to T. As the net absorption that gives rise to the NMR signal is proportional to the occupancy difference of the two levels, a larger population difference leads to a larger NMR signal. However, under typical NMR conditions, the difference is rather low (<0.01 %) [85], which causes NMR to be classified as an insensitive method. In other methods like infrared spectroscopy, higher energy levels are almost unoccupied, which leads to a much higher net absorption and larger signal intensities. As such, it is crucial to optimize NMR conditions by employing strong B_0 fields and using nuclei with high γ. In addition, pulse frequencies and intensities need to be optimized for the intended measurement.

To actively induce a transition between the two energy states, spin-$\frac{1}{2}$ nuclei each need a specific irradiation frequency. ^{1}H nuclei that experience a B_0 of 9.4 T require a pulse of 400 MHz (see Table 1.1) to be stimulated. The exact process for ^{13}C nuclei necessitates a frequency of approximately 100 MHz. A similarly significant difference in Larmor frequencies is present for ^{15}N as well and allows independent excitation and detection of these three nuclei. By appropriate pulse stimulation, a nucleus can transfer its magnetization to another, enabling the characterization of multiple atoms through a sequence of pulses. This mechanism ('magnetization transfer') is the basis for multi-dimensional NMR experiments.

However, the difference in Larmor frequencies of ^{1}H, ^{13}C, and ^{15}N is not the main property exploited in NMR. Each atom also experiences a slightly different local magnetic field and thus presents a unique ν. A typical reference compound is tetramethylsilane (TMS) whose Larmor frequency is used as a reference. For example, this reference frequency could be $\nu_{\text{ref}} = 400$ MHz at a B_0 field of 9.4 T. Conversely, the ^{1}H nuclei of another compound of interest (*e.g.*, acetone) might vary slightly to $\nu = 400.0008$ MHz [85]. The handling of such minor differences can be rather complicated, and to overcome this issue, the chemical shift δ was introduced:

$$\delta = 10^6\,\text{ppm} \cdot \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}} \tag{1.4}$$

In most cases, NMR signals are rather expressed as a function of the chemical shift δ instead of ν. For example, the difference of 800 Hz described above corresponds to a chemical shift difference of 2 ppm. By including the Larmor frequency of the reference compound in the calculation of δ, the data becomes transferable between different spectrometers as the value is independent of B_0. Aside from TMS for organic solvents, 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) is used for experiments in water or the solid–state [89].

1.5.2. Scalar and dipolar coupling in solutions and solids

When interpreting the magnetic properties of nuclei as small bar magnets, it follows that their magnetic fields interact when in close proximity. This effect is known as coupling and can be mediated either by a covalent bond (scalar- or *J*-coupling) or through space by local magnetic fields (dipolar coupling). The different origins of these two lead to profound differences in their nature (Fig. 1.6).

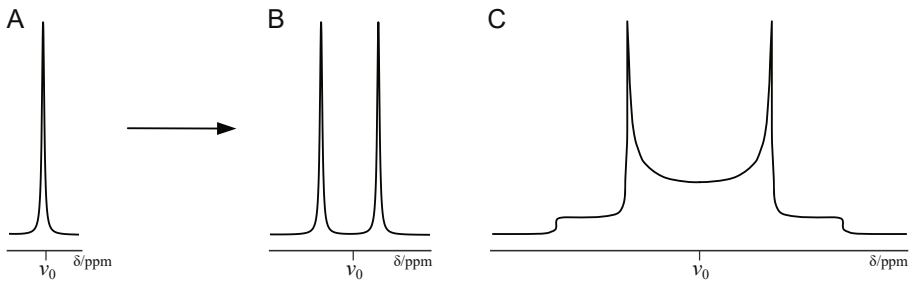

Fig. 1.6.: Coupling leads to the splitting of NMR signals. (**A**) Schematic representation of an NMR signal arising from an uncoupled nucleus. Consequences of scalar (**B**) and dipolar (**C**) coupling with contributions of chemical shift anisotropy (CSA) are shown on the right. In (**C**), a characteristic sloped 'Pake pattern' arises from the angular dependence of the interactions. Shown sizes are not to scale.

J-coupling, or indirect dipole-dipole coupling as termed by Malcolm Levitt [86], is due to nuclei weakly magnetizing the surrounding electron cloud. The effect starts at very small nuclear radii ($<10^{-14}$ m), where nuclei isotropically interact with electrons. Those electrons then propagate it through bonds [85]. This entire coupling process is independent of the spectrometer frequency, has no orientation dependence, and is thus unaffected by molecular tumbling. Typical values are 100-250 Hz for one-bond carbon-proton couplings ($^1J_{CH}$) and up to 20 Hz for homonuclear proton couplings mediated by three bonds ($^3J_{HH}$). The latter are sensitive to conformational differences. Therefore, the extent of $^3J_{HH}$ is a valuable tool in determining the structure of small molecules, carbohydrates, and proteins.

Direct dipole-dipole (or dipolar) coupling arises straight from the magnetic field generated by the nuclear spin. The active radius of this interaction is larger as it, in contrast to *J*-coupling, propagates through space. The magnitude of the dipolar coupling between

two nuclei is termed the dipole-dipole coupling constant $b_{1,2}$ and can be calculated as

$$|b_{1,2}| = -\frac{\hbar\mu_0}{4\pi} \cdot \frac{\gamma_1\gamma_2}{(r_{1,2})^3} \tag{1.5}$$

where μ_0 denotes the vacuum permeability, γ_1, γ_2 the gyromagnetic ratio of each nucleus and $r_{1,2}$ their distance. As such, the strength increases linearly with higher γ (*e.g.*, ^{1}H manifests stronger couplings than ^{13}C) and decreases to the power of three with higher distance. According to this calculation, a heteronuclear ^{1}H / ^{13}C system situated 1.5 Å apart has an absolute coupling $|b_{H,C}|$ of $\sim$9,000 Hz whereas it extents to only $\sim$30 Hz when at 10 Å. A homonuclear ^{1}H / ^{1}H system however creates a $\sim$40,000 Hz strong interaction when at 1.5 Å.

In addition to the parameters mentioned above, the absolute magnitude $|b_{1,2}|$ is further modulated by the relative orientation of the interaction to the external magnetic field B_0. Put simply, the distance $\vec{r}_{1,2}$ is a vector variable. This orientation dependence explains why dipolar coupling is absent in solution NMR but abundant in solids: due to fast molecular tumbling in solution (Brownian motion), all orientation-dependent effects are averaged to 0, whereas they persist for rigid or slowly moving systems (especially in solids).

1.5.3. Magic angle spinning in solid–state NMR

In solid–state NMR, the slow molecular motion leads to a substantial influence of dipolar coupling and chemical shift anisotropy (CSA) on the NMR signal. CSA arises from an asymmetric electron distribution around the nucleus and, like the dipolar coupling, depends on the angle relative to the axis of external magnetic field B_0. The molecules in solid samples are usually distributed over a number of orientations. Hence, the resulting diverse dipolar couplings and the CSA lead to a diversification of chemical shifts, *i.e.*, broadening of the signals.

It can be shown by quantum mechanical calculations that the angular dependence is scaled by a factor of $(3 \cdot \cos^2(\theta) - 1)$, where θ is the angle between the vector of nuclei interacting ($\vec{r}_{1,2}$) and B_0. By tilting the sample to $\theta = 54.7°$ relative to B_0, the term becomes zero for vectors aligned along this axis. Filling the sample in a so-called NMR rotor and spinning it around θ (Figure 1.7A, top) extends this effect to the entire sample [87, 90] as every vector orients along this angle on average (Figure 1.7A, bottom). The approach of rotating the sample around $\theta = 54.7°$ is referred to as magic angle spinning (MAS). Depending on the rotation frequency, line-broadening effects from dipolar coupling and CSA can be increasingly neglected (Figure 1.7B).

The strength of dipolar coupling depends on the nuclei that participate (see previous section), and accordingly, so do the MAS speeds required to remove the broadening effect.

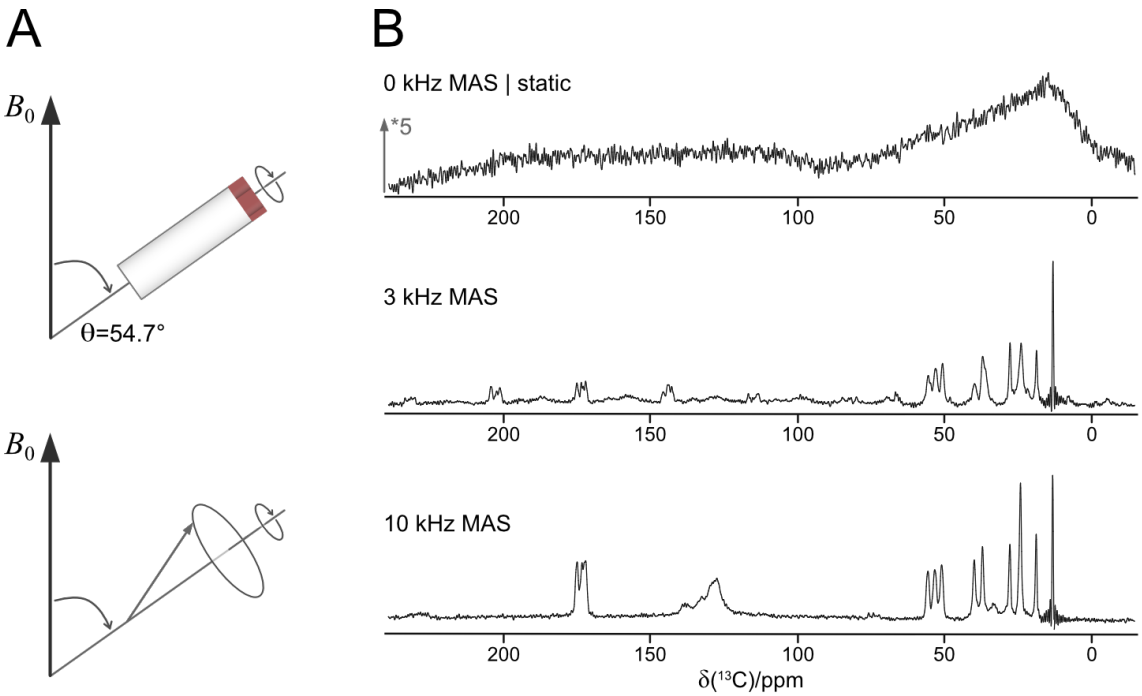

Fig. 1.7.: Magic angle spinning in solid–state NMR. (**A**) Schematic representation of a rotor aligned at the magic angle (54.7°) relative to the external magnetic field B_0 (top). On the bottom, a vector is shown that experiences the spinning. Although its orientation is not along the magic angle, the average position is aligned 54.7° relative to B_0. (**B**) 1D ^{13}C spectra of the tripeptide methionine-leucine-phenylalanine on a 400 MHz spectrometer at different MAS spinning frequencies. The static sample (top) is shown with a five-fold magnification. No discrete signals are visible, and the spectrum is strongly broadened. Spinning at 3 kHz (middle) or 10 kHz (bottom) increases the quality of the spectrum significantly.

As ^{13}C and ^{15}N have a relatively low gyromagnetic ratio γ, their dipolar couplings are comparably weak and can be eliminated by moderate spinning frequencies (8-15 kHz) [91]. Conversely, ^{1}H-^{1}H dipolar couplings are more persistent and require significantly higher MAS rates (100 kHz or more) to be efficiently removed from the spectrum. As a result, initially developed rotors for slower spinning only allowed sensible detection of ^{13}C and ^{15}N.

In order to detect protons at low spinning frequencies, their dense dipolar coupling network has to be disrupted by other means. Replacing ^{1}H with ^{2}H (also designated with D for deuterium or deuterons) yields the desired effect. Deuterium has a much lower γ than ^{1}H (compare Tab. 1.1 on page 13) and therefore develops much weaker dipolar couplings. This 'isotope dilution' is achieved by recombinant expression in heavy water containing a controlled level of deuterons. After the expression, the back-exchange of hydrogen atoms is typically conducted through a protein buffer with a specific $H_2O : D_2O$ ratio. A schematic representation of a fully deuterated (perdeuterated) and 100 % back-exchanged protein backbone is shown in Figure 1.8. This labeling scheme allows ^{1}H-detection at 20-24 kHz MAS yielding well-resolved spectra when appropriate back-exchange levels are used [92–94]. However, roughly 15 years ago, the development of NMR rotors with diameters of 1.9 mm and 1.3 mm enabled higher spinning frequencies (40-60 kHz) and thus allowed back-exchange rates of up to 100 % [95–97].

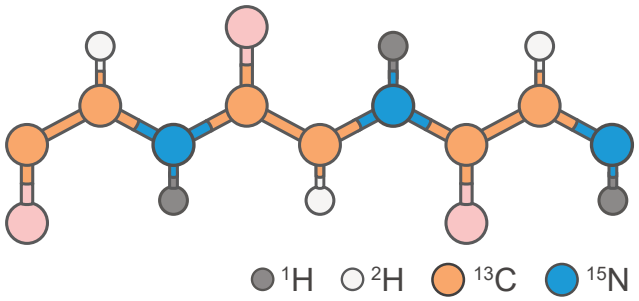

Fig. 1.8.: Labeling pattern after perdeuteration and back-exchange. Schematic representation of a protein backbone after expression in D_2O with ^{13}C and ^{15}N isotopes and subsequent back-exchange in H_2O buffer. Non-exchangeable sites (*e.g.*, H_α) keep the deuterons, while exchangeable sites (*e.g.*, H^N) acquire a proton. The different isotopes are highlighted with multiple colors as indicated, sidechains are omitted for clarity.

Even more recently, commercially available NMR rotors with diameters of 0.7 mm made MAS frequencies of more than 100 kHz possible and enabled the collection of ^{1}H-detected data without perdeuteration [98, 99]. This presented a major leap forward which eased sample preparation for solid–state NMR and enabled experiments that are directly comparable to solution NMR. Now, and likely in the near future, the trend towards smaller rotor sizes and higher spinning frequencies appears to continue. In 2019, MAS rates of up to 126 kHz were achieved [100], and in 2021 data recorded at 150 kHz MAS was

published [101]. At these high speeds, no further simplification can be made with regard to the sample preparation process, but rather only on the NMR side. At 150 kHz MAS, dipolar couplings and other orientation-dependent effects are removed more efficiently, which leads to smaller linewidths and more resolved spectra. Recently in 2022, Bruker has announced to make 0.4 mm diameter rotors, which can spin up to 160 kHz, commercially available, further enhancing general applicability. Interestingly, the high g-forces applied on the sample due to the MAS rotation do not seem to cause problems for the structural integrity of the protein. An overview of rotor diameters and their specifications is given in Table 1.2.

Tab. 1.2.: Selected NMR rotor sizes and their specifications. Diameters, volume, maximum spinning frequencies ω_r , and relative centripetal force F_{cp} at max. speed for rotors used in biological NMR. The F_{cp} calculation for the 0.4 mm rotor is based on the outer diameter.

Outer diameter	Inner diameter	Volume	Max. ω_r	F_{cp}
3.2 mm	2.2 mm	30 µL	24 kHz	$2.5 \cdot 10^6$ g
1.9 mm	1.5 mm	10 µL	42 kHz	$5.2 \cdot 10^6$ g
1.3 mm	0.9 mm	2.5 µL	67 kHz	$7.9 \cdot 10^6$ g
0.7 mm	0.5 mm	0.5 µL	111 kHz	$12 \cdot 10^6$ g
0.4 mm	n.a.	<0.4 µL	160 kHz	$20 \cdot 10^6$ g

1.5.4. Selected NMR experiments and sequential assignment

As detailed above, NMR-active nuclei can be interpreted as small bar magnets that align when brought into an external magnetic field B_0. When in equilibrium, the ensemble of different orientations leads to net magnetization M_0 (also called bulk magnetization), which orients itself along B_0 (Fig. 1.9A). When examining the influence of radiofrequency (RF) pulses on the bulk magnetization, the axis of B_0 is defined as the z-axis in a coordinate frame. This 'vector model', which considers the behavior of the magnetization vector, is often used to illustrate simple NMR experiments [102].

One pulse experiment and relaxation

The simplest possible NMR experiment consists of one pulse that moves M_0 away from its equilibrium position. Typically, one RF pulse is applied that turns the vector of magnetization by 90° (a 90° or π-pulse) into the xy-plane (Fig. 1.9B). There, it starts to precess at the Larmor frequency, which leads to a detectable NMR signal (Fig. 1.9C).

After perturbation, the bulk magnetization gradually returns to its equilibrium. This process is often characterized through the relaxation parameters T_1 and T_2. The T_1 time describes the duration it takes for the magnetization to return to its initial position along the z-axis. Sometimes the inverse of T_1, the relaxation rate R_1, is used. After a perfect 90° pulse, magnetization along z is 0 and, over time, returns to its full extent M_0. This process is typically described by

$$M_z(t) = M_0 \cdot \left(1 - e^{-\frac{t}{T_1}}\right) = M_0 \cdot \left(1 - e^{-R_1 \cdot t}\right) \tag{1.6}$$

with $M_z(0) = 0$ and $M_z(t) = M_0$ when $t \to \infty$.

Conversely, the T_2 time describes how long it takes for magnetization to decay in the xy-plane. The inverse of T_2, the relaxation rate R_2 is also commonly used. After a perfect 90° pulse, the magnetization in the xy-plane is maximal (M_0) and returns to 0. The corresponding exponential function to describe this process is

$$M_{xy}(t) = M_0 \cdot e^{-\frac{t}{T_2}} = M_0 \cdot e^{-R_2 \cdot t} \tag{1.7}$$

with $M_{xy}(0) = M_0$ and $M_{xy}(t) = 0$ when $t \to \infty$.

Due to their nature, T_1 always has to be larger than T_2. For proteins, ranges that are typically encountered are one to multiple seconds for T_1 and tens to hundreds of milliseconds for T_2. Schematic relaxation processes and graphs of (1.6) and (1.7) can be found in Figure 1.9D and E.

These relaxation phenomena occur due to the interactions between spins as well as with spins and their environment. As these effects are intertwined with the molecular motions (since local motions define which other spins and atoms come close), T_1 and T_2 can be used as measures of flexibility. It is possible to determine these rates for each nucleus, making the obtained information precise down to the atom.

Other relaxation measures exist as well. In solution, heteronuclear NOE between ^{1}H and ^{15}N is often used to complement T_1 and T_2. In solid–state NMR, a variation of T_1 called $T_{1\rho}$ is often used [103, 104]. When using multiple relaxation parameters together, a comprehensive picture of the overall and local motions of a protein can be obtained [105].

^{1}H-^{15}N correlations

In large systems, one-dimensional NMR data is often insufficient to resolve the peaks of interest. Increasing the dimensionality of a spectrum spreads the signals out and often has a filtering effect. Multi-dimensional NMR spectra are achieved by applying a series of RF pulses and waiting times in a precise manner which manipulates the magnetization in the intended way. In theory, as many magnetization transfers between nuclei as desired

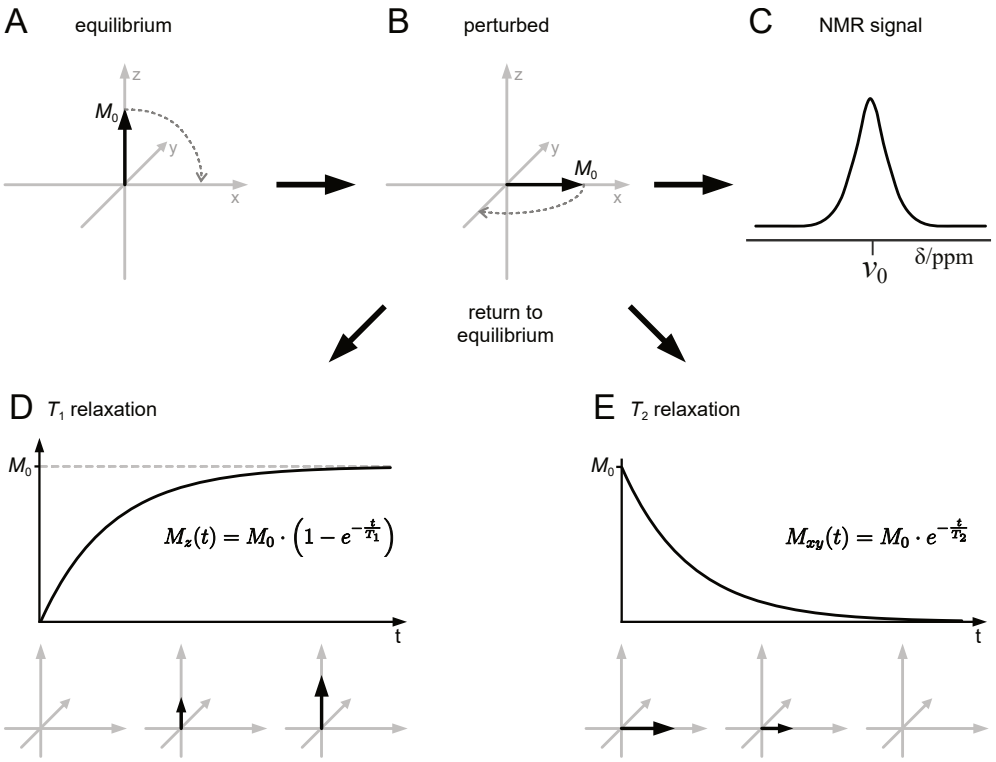

Fig. 1.9.: Vector model description of a simplified NMR experiment. (**A**) When at equilibrium, the bulk magnetization M_0 is aligned along the eternal magnetic field B_0 (= z-axis by definition). Applying a 90° pulse perturbs the magnetization vector and turns it into the xy-plane. (**B**) Afterwards, in the xy-plane, the magnetization vector starts to precess with the Larmor frequency of the nucleus. (**C**) This precession can be detected and, after Fourier transformation, gives rise to an NMR signal in the frequency domain. After being perturbed, the magnetization starts to relax and returns to its equilibrium position. This process is typically split into two orthogonal contributions. (**D**) The return of magnetization along the z-axis (T_1 relaxation) and (**E**) decay of magnetization in the xy-plane (T_2 relaxation).

can be conducted, which gives rises to basically infinitely complex and specific spectra. In practice, however, available pulse sequences are limited by the detected signal (each transfer causes a signal loss) and measurement time at the spectrometer. Here, the focus is set on a few commonly applied examples.

One of, if not the, most abundantly used spectrum in NMR of proteins is the ^{1}H-^{15}N correlation. As the name suggests, ^{1}H and ^{15}N nuclei are correlated through magnetization transfer between them. In solution, *J*-based transfers (*e.g.*, the INEPT scheme [106]) are usually applied, whereas in solid–state NMR dipolar coupling based transfer methods (like cross polarization (CP) [107]) are more commonly used. The respective parameters are chosen in a way that the magnetization is only transferred between adjacent nuclei.

An ^{1}H-^{15}N correlation typically consists of two dimensions (^{1}H and ^{15}N), corresponding to two directly bounded nuclei. Therefore, the majority of signals arise from the backbone amide bond (H^{N}, N), and only a few signals stem from sidechains (*e.g.*, arginine Nε). ^{1}H-^{15}N correlations contain substantial information about the protein backbone and are very sensitive to the overall state of the protein (*e.g.*, folding, pH influence, degradation). Due to these characteristics, an ^{1}H-^{15}N correlation is usually the first spectrum recorded on a protein sample. Further ^{1}H-^{15}N spectra are then routinely measured to check for protein integrity in between more complex experiments. A schematic display of the magnetization transfer and an ^{1}H-^{15}N correlation spectrum is shown in Fig. 1.10A.

Dipolar-assisted rotational resonance

A complementary method to an ^{1}H-^{15}N correlation, which is commonly used in solid–state NMR, is the dipolar-assisted rotational resonance (DARR) scheme [108]. In the applied pulse sequence, correlations between neighboring or more remote ^{13}C nuclei are typically detected through variation of an experimental parameter called mixing time. DARR is one of many pulse sequences that can be used for this type of mixing. For example, magnetization transfer through the nuclear overhauser effect (NOE) has similar properties and is often used in solution NMR. In this example, the resulting symmetric spectrum contains two ^{13}C axes, and displays the ^{13}C chemical shift for every nucleus of the sample (schematically shown in Fig. 1.10B). These ^{13}C-^{13}C spectra contain ample information but can be hard to interpret for large proteins due to signal overlap.

In order to reduce the number of signals, the acquisition of a DARR spectrum can be coupled with selective ^{13}C-labeling of the protein. One approach, which is often applied for structure determination by solid–state NMR, is the usage of [1,3-^{13}C]- or [2-^{13}C]-labeled glycerol as the sole carbon source during the expression [109]. The partial incorporation of ^{13}C leads to characteristic isotope patterns (Appendix Fig A.1 on page 109) which carry information about the type of nucleus. Over all, this approach helps to increase the spectral

resolution by disruption of ^{13}C-^{13}C homonuclear couplings and facilitates peak separation through detection of a signal subset.

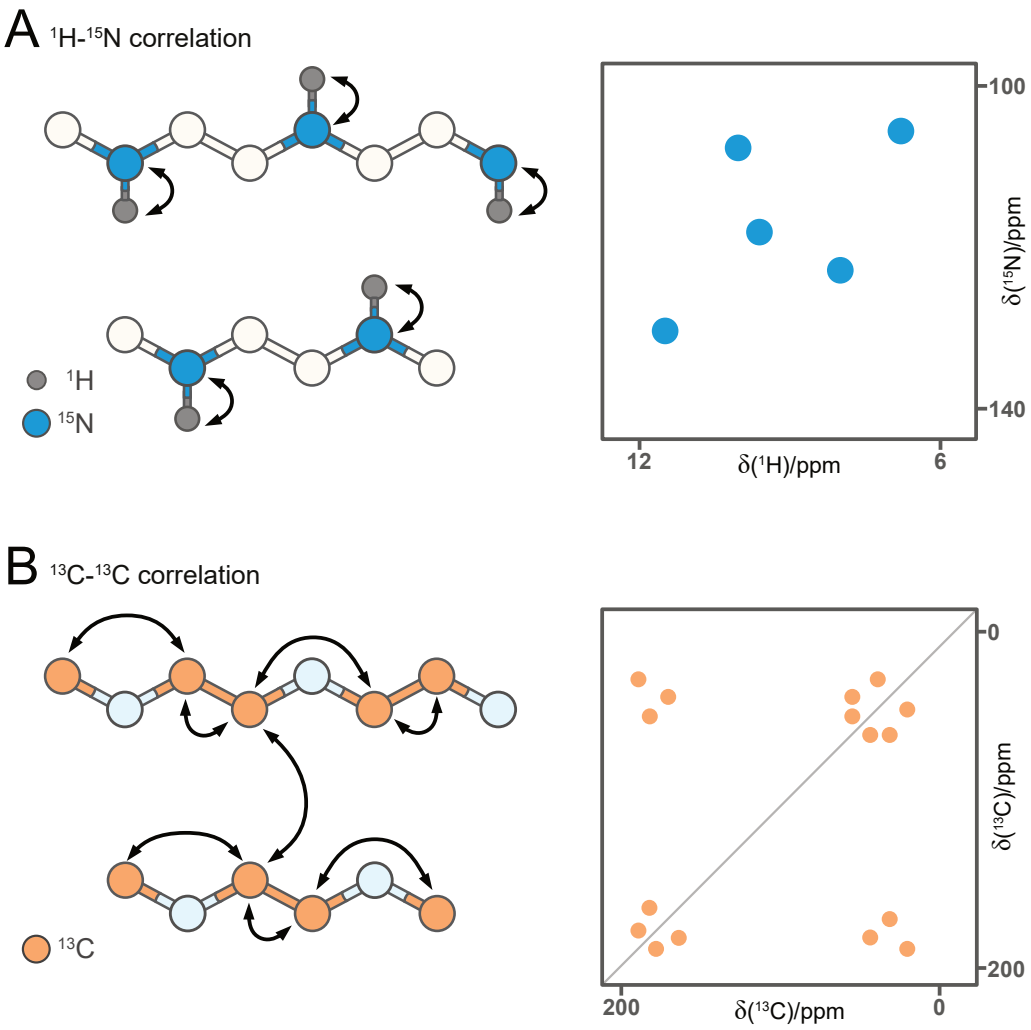

Fig. 1.10.: Schematic representations of ^{1}H-^{15}N and ^{13}C-^{13}C correlations. (**A**) The magnetization transfer in ^{1}H-^{15}N correlation spectra is typically conducted between adjacent nuclei. For proteins, these are predominantly the backbone H^N and N atoms. The resulting spectrum has one ^{1}H- and one ^{15}N-axis and shows approximately one signal per amino acid. (**B**) When acquiring a ^{13}C-^{13}C correlation in solid–state NMR, the magnetization is mostly transferred through space by dipolar coupling (*e.g.*, DARR mixing). The distance over which magnetization is transferred can be varied by changing the mixing time. The obtained spectrum has two ^{13}C-axes, has a diagonal and is symmetric.

Sequential assignment

The spectra obtained by NMR contain atomically resolved data. However, this trait can only be harnessed if it is known which signal belongs to which nucleus (= atom). The process in which NMR signals are mapped to the corresponding nucleus of the protein is called 'assignment' and can be conducted in a number of different ways. The exact nature of the assignment procedure depends on the properties of the protein. However, some traits can be generalized. Most of the time, it is beneficial to know the primary structure of the protein and record complementary spectra. The most commonly used assignment procedure is the 'backbone walk' outlined below.

First, multiple NMR spectra are recorded that share one or more dimensions. With their common dimension, the spectra be related to one another. In a next step, the related chemical shifts are linked. This logical linkage is typically based upon chemical bonds and *J*-based transfers, but spatial proximity can also be used if a structure of the protein is known. Then, the NMR properties are linked until a point is reached where they can be unambiguously assigned to a specific part of the sequence. For a classical 'backbone walk', the procedure is the following:

1. Identifying the chemical shifts of any residue
2. Determining the residue type through characteristic chemical shifts (*e.g.*, glycines)
3. Finding the sequential neighbor and determining their chemical shift
4. Repeating the steps above until the stretch can be mapped to the primary structure

A straightforward set of spectra to illustrate this approach is the combination of hCANH and hCAcoNH (Fig. 1.11). In accordance with the published terminology [110], nuclei written in upper case are measured, and those in lower case are only used for magnetization transfer.

The hCANH correlates three directly bonded nuclei with each other: H^N(i), N(i), and the bonded C_α(i). All three originate from the same residue, i. Conversely, the hCAcoNH takes an additional step through the carbonyl and correlates the C_α of the previous (i-1) residue with H^N(i) and N(i). Therefore, it conducts a sequential transfer between the residues i-1 and i. Using the frequencies of H^N(i) and N(i) as a pivot, a sequential connection between C_α(i) and C_α(i-1) can be made. This sequence can be repeated with H^N(i-1) and N(i-1) to correlate C_α(i-1) and C_α(i-2). Continuous application of this approach allows for determining long stretches of the backbone until the procedure fails due to missing signals or low signal-to-noise.

After linking a sufficient number of residues, they need to be mapped to the sequence. Glycine residues can be easily identified because they have a characteristic C_α chemical

shift of around 45 ppm. A stretch of five residues with two glycines in the manner X-G-X-X-G might allow for unambiguous assignment to the primary sequence. If the motif occurs multiple times or no residue types can be determined, the length of the stretch needs to be increased, or the chemical of further backbone atoms must be determined.

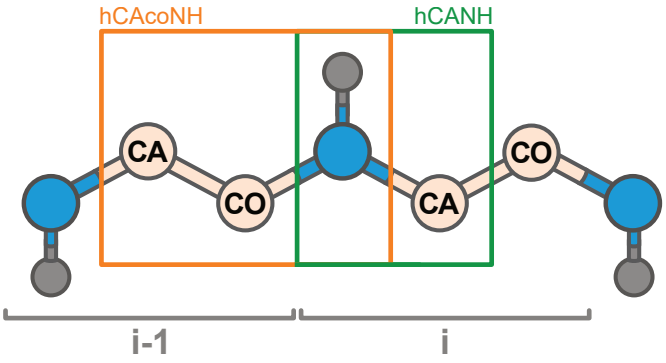

Fig. 1.11.: Correlating backbone atoms by hCANH and hCAcoNH spectra. Schematic representation of how the hCANH and the hCAcoNH spectra can be used for backbone assignment. When aligning both in a way that the ^{1}H and ^{15}N signals are identical, they depict the chemical ^{13}C shift of two adjacent C_α (C_α(i) for hCANH and C_α(i-1) for hCAcoNH).

The assignment approach can be expanded with the chemical shifts of other backbone atoms (C', C_β, H_α) and even by going into the sidechain; however, these complex procedures are outside the scope of this thesis. A comprehensive review of different assignment methods is given in [111].

1.6. Protein structure determination and prediction

The determination of three-dimensional structures for proteins is a crucial step in understanding their function, mechanism of action, and the intermediate steps involved. One practical application of these models is in the design of new binding partners (*e.g.*, drug candidates [112]). By identifying pockets on the protein surface or specific interaction interfaces, potential ligands can be found and known ones refined. In the following section, current methods used for the determination of protein structures will be introduced, and their respective advantages and disadvantages will be highlighted. Protein structures determined by any method are collectively stored in the Protein Data Bank (PDB) [113].

1.6.1. Methods to derive protein structures

X-ray crystallography

X-ray crystallography is a technique that takes advantage of the unique property of X-rays to diffract when they interact with a crystal. This diffraction phenomenon is similar to the scattering of light waves when they pass through a sufficiently narrow grid. In the case of X-rays, they interact with electron densities in the crystal which causes them to scatter. As a result, the relative arrangement of these electron densities is encoded in the diffraction pattern that is obtained after the crystal is irradiated. The diffraction pattern can be converted back to an electron density map of the original sample through mathematical and computational methods. X-ray crystallography can be employed to study various types of molecules [114], but the focus here is on its application to proteins.

The first protein structure derived by X-ray crystallography, one of whale myoglobin, was determined in the late 1950s by John Kendrew together with Max Perutz. For their discovery, both received the Nobel Prize in Chemistry in 1962 [115]. The determination of structures by X-ray crystallography is streamlined nowadays (with more than 175,000 available in 2023, development in Fig. 1.12A on page 29) and yields well-resolved structures rapidly when a high-quality diffraction map is obtained. Typically, hydrogen atoms are not resolved due to their minor electron density, but those can be modeled onto the protein chain through geometric considerations.

The range of accessible proteins is vast, as X-ray crystallography works independent of the protein size as long as well-diffracting crystals can be obtained. However, crystal formation poses the main bottleneck of this method. Some molecules (amyloids, membrane proteins, and more) do not or only hardly form crystals and are not accessible by X-ray crystallography. Even for suitable proteins, finding a fitting crystallization condition can be time-consuming as the process of crystal formation is not well understood [116]. Furthermore, the protein structure can locally diverge from the native state as interactions

enforced by the crystal packing can alter the local structure. The impact of the crystal packing is also reflected in the flexibility of the structures, as the dense packing typically renders the protein more rigid in the crystal than under native conditions.

Cryogenic electron microscopy

Cryogenic electron microscopy, or cryo-EM in short, is another technique that has recently found widespread application in structural biology. The following considerations concern single-particle cryo-EM, which is routinely applied to proteins.

In single-particle cryo-EM, molecules of choice (*e.g.*, proteins) are frozen in amorphous (vitrified) ice and irradiated with an electron beam [117]. The frozen state has multiple advantages: it weakens the damage inflicted on the sample by the electron beam, reduces local motion, freezes the protein conformation, and allows the measurement to be conducted in vacuum. The results of a cryo-EM measurement are images (micrographs) representing two-dimensional projections of the molecule from different angles. It was already shown in 1968 that a three-dimensional structure can be reconstituted from these projection images [118]. However, decades passed until cryo-EM became widely used (Fig. 1.12B).

A broad application was long hampered by tedious data collection and a low signal-to-noise ratio. Only with the advent of ever-increasing computational power and the development of microchip technology in the early 2000s, this changed. Direct electron detectors were introduced, which greatly enhanced the detected signal [119], and algorithms for automated data collection and processing enabled a further increase in quality through image averaging [120]. Today (June 2023), more than 15,000 protein structures that were determined by cryo-EM are available in the PDB. In recognition of their work, the 2017 Nobel Prize in Chemistry was given to Jacques Dubochet, Joachim Frank, and Richard Henderson “for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution” [121].

The main advantages of cryo-EM are the low requirements for accessible samples and the diversity of data that is obtained (*i.e.*, static pictures of interactions and conformation due to freezing the sample). No crystal formation is necessary, and arbitrarily large complexes can be imaged. Notable examples are integral membrane proteins and complexes like the spliceosome [117]. Small proteins (< 100 kDa) are the lower size boundary, and the resolution is currently worse than in X-ray crystallography [120]. In addition, cryo-EM typically does not image the exact location of hydrogen atoms (like X-ray crystallography) due to the nuclei having only a minor interaction with the electron beam.

Nuclear Magnetic Resonance

This section focuses on the significance of NMR in structure determination, building upon the previously discussed basic principles. The main difference to X-ray crystallography and cryo-EM lies in the physical properties that are harnessed to extract the information. Instead of electrons, NMR utilizes the spin properties of nuclei. Here, the nuclei of ^{1}H (protons) are of particular importance as they are abundant and possess a spin of $\frac{1}{2}$.

In a typical workflow, the first step for structure calculation by NMR is to derive the chemical shifts for each nucleus through an assignment procedure. Then, distance restraints between nuclei, mostly ^{1}H, are collected using Nuclear Overhauser Effect Spectroscopy (NOESY). For the exploitation of NOEs for protein structure determination, Kurt Wüthrich was awarded the Nobel Prize in Chemistry in 2002 [122]. If two nuclei are within a distance of 8 Å, NOESY spectra show cross-peaks between them that have a distance-dependent intensity. By gathering a sufficient number of these contact restraints, an overdetermined distance geometry problem can be solved. The NMR structure calculation algorithm attempts to position the nuclei responsible for the contact sites in close proximity, satisfying as many restraints as possible. The output of an NMR structure calculation is not a single structure but rather an ensemble of structures, of which each single one fulfills the distance restraints in a similar manner.

Advancements in computational power have broadened the capabilities of NMR structure determination. Today, programs are available that can calculate NMR structure ensembles fully automated without prior assignment [123, 124]. However, the overall procedure remains time-intensive and can be challenging. Currently (June 2023), ∼14,000 NMR structures are deposited in the PDB, with the yearly increase slowly stagnating (Fig 1.12C).

One significant advantage of NMR over other methods is the ability to study proteins in their native environment (*e.g.*, a buffer mimicking cellular conditions or within membranes for solid–state NMR [125]). Furthermore, NMR data is always obtained at atomic resolution allowing for detailed monitoring of conformational changes. This includes the usage of spectra for the study of ligand binding and target interactions. Additionally, if the assignment is known, other structural parameters such as residual dipolar couplings, dihedral angles, and flexibility can directly be derived from NMR data [126].

However, there are some drawbacks to NMR structure determination. For example, not all proteins are amenable to this technique. Solution NMR is limited by the size of the protein as large proteins (> 50 kDa) produce spectra with significant line-broadening. Conversely, solid–state NMR is not directly limited by size, as the spectral quality remains good even for larger proteins. However, the abundance of NMR signals becomes a challenge with larger molecules, leading to signal overlap that may be difficult to resolve.

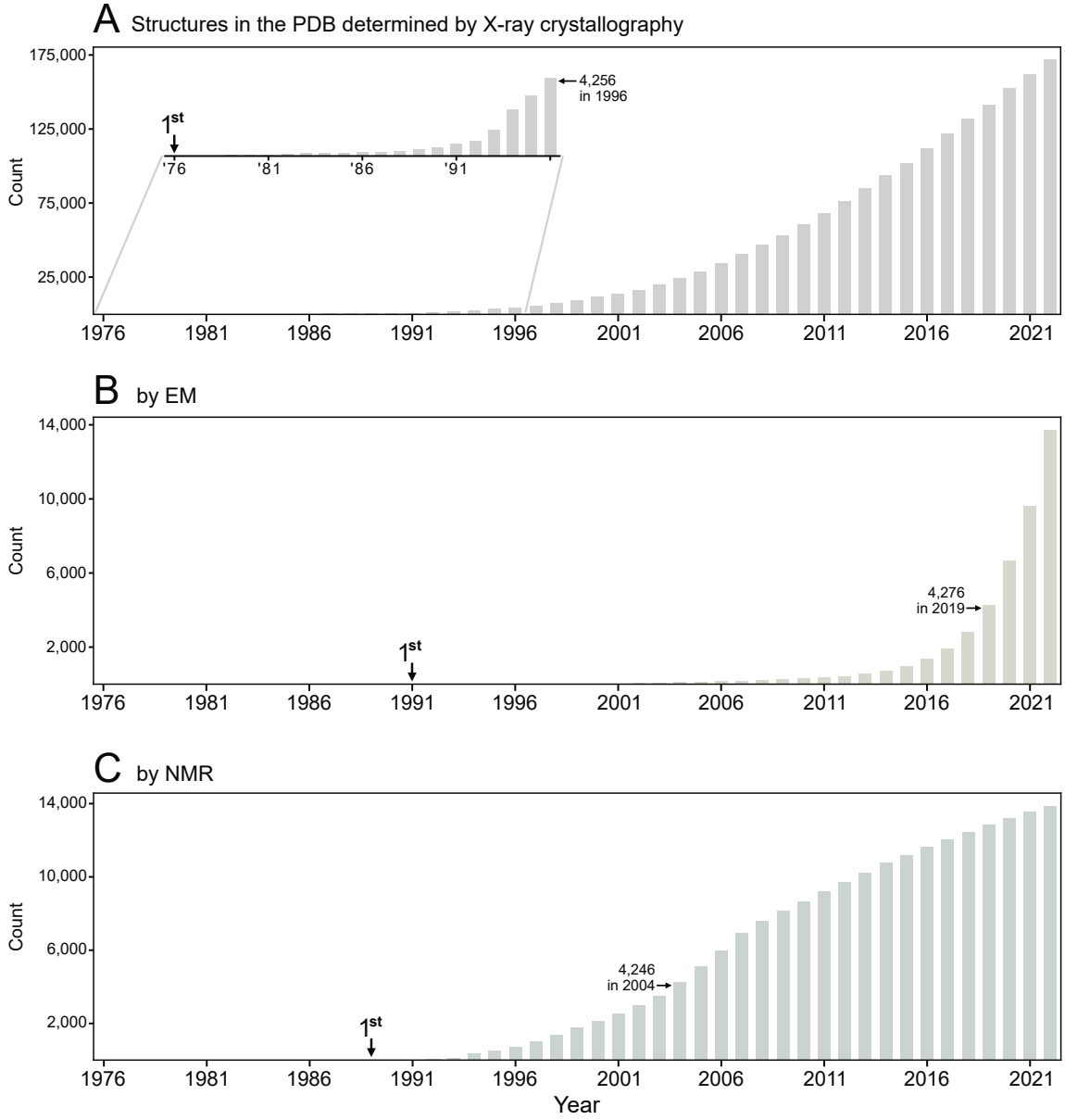

Fig. 1.12.: Total number of structures in the PDB by determination method. The growth of available structures in PDB determined by (**A**) X-ray crystallography, (**B**) EM methods, and (**C**) NMR is shown yearly from 1976 to 2022. For X-ray crystallography, the first 20 years are shown enlarged with the same scaling used for the other methods. In each graph, the years of the first released structure and the one before 5,000 depositions were reached are indicated. The presented data were downloaded from from the PDB in June 2023.

AlphaFold

Although it is not a method for protein structure determination, AlphaFold warrants consideration in this section. It has gained considerable traction since its public release in 2021 and its applications intersect which those of conventional structure determination methods. AlphaFold is an algorithm based on a neural network capable of predicting protein structures solely from their primary sequence [127].

The question of how proteins arrive at their final structure has long puzzled scientists [128]. Ultimately, the structure has to be encoded in the amino acid sequence, as the cell does not offer a specific support (*e.g.*, chaperones) for every protein during folding. Therefore, the decisive difference that determines the protein structure has to be the primary sequence. AlphaFold has, at least partially, solved this question by considering evolutionary, physical, and geometric constraints of protein structures [127]. It uses multiple sequence alignments (MSAs) of the protein of interest with its evolutionary homologues to find conserved and co-evolving residues. Those that show an evolutionary correlation are likely to be spatially close as well - a trait crucial for the algorithm to work. In addition to the MSAs, the AlphaFold neural network is trained with available PDB data (*i.e.*, protein structures determined by the methods mentioned above) and high-confidence structures predicted by a previous iteration of the network itself. During this process, the training data is manipulated by cropping, MSA subsampling, and partial substitution of sequence blocks in a controlled manner which makes the prediction challenging. The results of these difficult predictions are compared to corresponding real structures, and assist the algorithm in adapting to imperfect data. The results are impressive. The refined version of AlphaFold (AlphaFold2) outperformed all competitors in the CASP14 structure prediction competition in 2020 [129].

Soon after the public release of the AlphaFold source code [130], its scope was expanded further. Different MSA algorithms were coupled to AlphaFold, and its performance was boosted, which accelerated the prediction process. An online server was set up that allowed protein structure prediction independent of local computing power [131]. Furthermore, AlphaFold was found to predict complexes and multimers well, although they were not explicitly contained in the training data [132]. Lastly, a public database was established with bulk-predicted monomeric protein structures of available genomes [133]. It currently (June 2023) contains more than 214 million (214,000,000) structures (**!**) which exceeds the extent of the PDB by more than three orders of magnitude (10^3).

As a consequence, a predicted structure for virtually any protein can now be obtained within minutes. These structures can be used as a basis for experiment design, thus enabling a detailed study of the protein function. Furthermore, there are no prerequisites for sample preparation (*e.g.*, as for crystallization) to obtain a structure. While the widespread

availability of protein structures revolutionizes structural biology by itself, it must still be considered that these are only predictions [134]. The influence of point mutations on the entire protein structure is typically not judged correctly by AlphaFold, and complexes and their interfaces are not always correct. Moreover, ligand binding is not at all included in the algorithm, and AlphaFold only excels in the prediction of folded regions, making predicted models for intrinsically disordered proteins unusable.

1.6.2. Integrative structural biology

Integrative structural biology, the combination of multiple techniques to arrive at an improved protein structure, is not a new approach [135] but has changed quite a lot over the years. For example, prior to the 'resolution revolution' in cryo-EM, multiple highly resolved structures determined by X-ray crystallography (2-3 Å) were often fitted into less detailed cryo-EM densities of the complex (4-6 Å) to determine their relative spatial orientation [117]. These days, cryo-EM alone can often determine such structures without any need for integrative modeling.

Nowadays, the boundaries of single techniques are elsewhere. A typical area where integrative approaches are used is the determination of flexibility within a protein. Often X-ray and cryo-EM structures characterize a protein to be more rigid than it is in a cellular context - a subject where NMR excels due to the measurements happening under native-like buffer conditions. In addition, the corresponding parameters are easily accessible through relaxation measurements. Therefore, NMR is often used to refine these structures with regard to motions and conformational changes (*e.g.*, [136]). Of course, the range of methods for an integrative approach is not limited to the ones already presented. Rather, any method that yields additional data can be considered as, *e.g.*, cross-linking mass spectrometry [137], electron paramagnetic resonance [138], or fluorescence [139] methods.

In the future, the integration of AlphaFold predictions with experimental methods appears very promising. The predicted models can assist in conventional structure determination methods, *e.g.*, as a template for molecular replacement in X-ray crystallography or as an initial structure for NMR ensemble calculation. Furthermore, a detailed validation or falsification of the predicted model can be performed. So far, it was found that AlphaFold models rival medium-resolution X-ray structures and are often as, or sometimes more, accurate than solution NMR ensembles [140]. An easy way for cross-validation by NMR is to compare contacts between AlphaFold and NOESY cross-peaks that encode spatial proximity. Alternatively, a comparison to other NMR parameters like residual dipolar couplings or rigidity estimated from secondary chemical shifts [141] can be performed. As structure calculation by NMR is typically very time-consuming, bypassing this step through cross-checking an AlphaFold prediction promises to be a significant progress.

2. Aim and scope of the thesis

Bacteria engage in various interactions with their environment, which can involve plants, soil, or other organisms. These interactions can be associative, as within a biofilm, or aggressive when toxins come into play. They may occur within the same species or between different organisms. The work presented here aims to understand the mechanisms of biofilm formation and toxin activity with two bacterial examples, *Bacillus subtilis* and *Photorhabdus luminescens*, respectively.

The first investigated protein, TasA, is produced by the soil-dwelling bacterium *B. subtilis*. It is the primary protein component of the *B. subtilis* biofilm and has been extensively studied, showing that TasA exists in an amyloid and a non-amyloid state [22, 50, 54–56]. Other bacteria often employ amyloid fibers to support the integrity of their biofilm (*e.g.*, *Escherichia coli* secreting CsgA [49]), and accordingly, TasA was long presumed to behave similarly. However, recent research has collected substantial evidence that a non-amyloid, filamentous state plays a crucial role [54–56] which will be the focus of this work.

Here, the main goal is to understand how TasA filaments assemble and how their formation is catalyzed by the accessory protein TapA. TapA is known to induce the transition of TasA into polymeric states and plays a crucial role in biofilm constitution [19]. Proper biofilm formation with TasA and TapA present leads to more potent adherence to surfaces and increased antibiotic resistance of bacteria. Understanding the molecular basis of this process is the first step toward a controlled application of biofilms in biotechnology. Furthermore, their induced dispersal can be helpful in countering disease-related bacteria in human health.

We plan to employ solid–state NMR to examine TasA filaments. This approach includes the assignment of the protein backbone by ^{1}H-detected experiments and the collection of ^{1}H-^{1}H through-space contacts by NMR. The atomic resolution of the extracted data will lead to a detailed understanding of the TasA-TasA and TapA-TasA interactions. Furthermore, we aim to identify evolutionarily conserved patterns by examining homologies between TasA, TapA, and proteins found in other bacteria. The knowledge of whether the

crucial interaction sites are encountered in distant organisms will indicate how widespread the assembly mechanism of TasA filaments is.

The second mechanistic study relates to a structural analysis of the toxin complexes (Tc) deployed by *P. luminescens* and their pathogenic effect. These Tc function by injecting a toxic enzyme into the target cell, which covalently modifies essential cell components. Bacteria and their Tc have shown potential for pest control in agriculture [58] and the reduction of mosquito populations [59]. Furthermore, the injection mechanism employed by these Tc has recently been indicated as a promising delivery vector [12]. This work focuses on the subcomplex TcdB2-TccC3, called TcB-TcC in short, and the contained toxic enzyme TcHVR. We want to unravel the molecular basis for the enzymatic activity of TcHVR and close the knowledge gaps concerning the injection mechanism.

In order to address these questions, our study utilizes NMR spectroscopy, cryo-EM, and molecular modeling together to elucidate the molecular features of the protein throughout its various states. We want to characterize the molecular basis of TcHVR toxicity by using NMR to derive atomic information about the protein. In detail, this includes the assignment of NMR signals, the examination of water-accessible sites by H-D exchange, the calculation of a structure ensemble, and the characterization of the binding process between TcHVR and its cofactor NAD^+. Cryo-EM and molecular modeling, conducted by our collaborators, will be employed to examine the features of TcHVR bound to its target, the cytoskeletal protein actin.

Given the importance of a conserved R-S-E motif in TcHVR for its enzymatic activity on actin, the active center containing the motif is anticipated to show structural homology to other ADP-ribosyltransferases. While these homologous regions might be responsible for a conserved ligand-binding mode, divergent regions of TcHVR will likely impart target specificity. Determining these characteristics will help to deepen the understanding of ADP-ribosyltransferase functionality and indicate how the actin-specificity is conferred.

Lastly, we aim to transfer the structural knowledge obtained for TcHVR to its state when inside the TcB-TcC subcomplex. While the function and structure of the TcB-TcC subcomplex are known, the folding state of TcHVR when inside has remained unclear in several studies [70, 72, 75, 84]. Understanding the folding state within the complex is crucial, as the injection procedure is the main contributor to host specificity towards insects [64]. By measuring ^{1}H-^{15}N correlations using solid–state NMR at ultra-fast magic angle spinning speeds, we will gather data that is directly comparable to the solution NMR data acquired previously. This comparison will help identify similarities and differences between the free state of TcHVR and its state within the TcB-TcC complex. The results will provide further insights into the criteria for using toxin complexes as a universal protein translocation system.

3. Material and Methods

3.1. Characterization of TasA from *Bacillus subtilis*

3.1.1. Protein expression and purification

Genetic information was kept as a His-Sumo-TasA fusion construct on a pCA 528 plasmid encoding 339 amino acids of which 234 belong to mature $TasA_{28\text{-}261}$ (UniProt P54507) with its signal sequence cleaved. The genetic information was transformed by heat shock protocol into a T7 Express BL21 (DE3) Rosetta 2 *Escherichia coli* strain. In short, 0.5-1 µl of plasmid (typically up to 100 ng) and 50 µL of compatible cells were thawed on ice, mixed, and incubated for 30 min on ice together. The mixture was then exposed to 42 °C for 90 s, subsequently mixed with 400 µL of antibiotic-free medium, and put on ice again for a maximum of 5 min. Afterwards, bacterial growth was stimulated by shaking for 1 h at 37 °C before applying the medium to an agar plate carrying the necessary antibiotics (50 µg/mL kanamycin for pCA 528 and 34 µg/mL Chloramphenicol for Rosetta 2 bacteria). The agar plate carrying transformed bacteria was incubated over night. Afterwards, colonies were picked for protein expression.

First, a pre-culture, 5 mL of lysogeny broth (LB) medium [142] with 50 µg/ml kanamycin and 34 µg/ml Chloramphenicol, was inoculated with the picked colony and agitated for 6-8 h at 37 °C. Afterwards, 200 µl were transferred into 50 mL of M9 minimal medium with the necessary antibiotics (see Tables 3.1 and 3.2 on the following page). The exact composition needs to be adapted for the desired isotope labeling pattern. For ^{13}C and ^{15}N, 3 g/L ^{13}C-glucose and 0.5 g/L ^{15}N-NH_4Cl are used. Perdeuteration of the protein is achieved by preparing the medium, and all containing substances, in heavy water (D_2O, ^2H_2O). If necessary, water-containing substances are lyophilized and dissolved in heavy water. For introduction of sparse ^{13}C labeling, glucose is substituted by 2.7 g/L of either [1,3-^{13}C]- or [2-^{13}C]-glycerol and 3 g/L ^{13}C sodium bicarbonate ($NaHCO_3$) to stabilize the pH. The 50 mL of overnight culture were grown shaking at 180 rpm for 18-20 h at

30 °C. The next day it was diluted with additional, equivalent M9 minimal medium and antibiotics to an absorption value (OD_{600}) of 0.1-0.2 and grown shaking at 37 °C until an OD_{600} of 0.6-0.7 was reached. Subsequently, Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM in culture medium to induce expression. The IPTG-containing culture was left shaking at 22 °C overnight (16-18 h) for protein expression. Afterwards the bacteria were pelleted at 5000 rpm (4400 g) and 4 °C for 10-12 min in a Beckman Avanti J-25 centrifuge with a FIBERLite F10BCI-6x500y rotor (similar to JA-10 from Beckman). The pellets were washed with isotonic (150 mM) NaCl solution and stored at -80 °C if not immediately purified.

Tab. 3.1.: M9 minimal medium. For a volume of 1 L, the following components need to be solubilized in 800 mL of sterile deionized water (resistance >17 MΩ/cm) and adjusted to the complete volume. For labeled expression, ^{13}C-glucose (3 g/L) and ^{15}N-NH_4Cl are used. Sparse labeling is achieved by substituting glucose with either [1,3-^{13}C]- or [2-^{13}C]-glycerol to a final concentration of 2,7 g/l and adding ^{13}C-$NaHCO_3$ (final conc. 3 g/L).

Volume	Compound
10 mL	trace elements
1 mL	$MgSO_4$ (1 M)
0.3 mL	$CaCl_2$ (1 M)
100 mL	M9 salt (10x)
20 mL	20 % glucose
1.5 mL	Thiamin-HCl (1 mg/mL)
15 mL	Biotin (0.1 mg/mL)
2 mL	NH_4Cl (250 mg/mL)
	+ antibiotics

Tab. 3.2.: Components of M9 minimal medium.

(a) Trace elements solution. For a volume of 500 mL, the following components need to be solubilized in sterile deionized water (resistance >17 MΩ/cm). The pH needs to be kept between 7.5 and 7.7 to prevent precipitation, the final pH is 7.6. Additionally, it is recommended to add the compounds in the order they appear.

Amount	Compound
2.5 g	EDTA
457 mg	$FeSO_4$ x 7 H_2O
25 mg	$ZnCl_2$
5 mg	$CuSo_4$

(b) M9 salts (10x). The following components need to be solubilized in 1 L of sterile deionized water (resistance >17 MΩ/cm).

Amount	Compound
80 g	Na_2HPO_4 x 2 H_2O
(or 64 g	Na_2HPO_4)
20 g	KH_2PO_4
5 g	NaCl

For purification, the pellets were resuspended in His Bond buffer A (20 mM Tris pH 7.5, 500 mM NaCl, 5 mM Imidazole) with 6 µl of Benzonase (modified Nuklease A from *Serratia marcescens*) and lyzed by passing the solution through a Microfluidizer LM10 6-7 times until it was homogeneous. The insoluble fraction was removed by centrifugation at 22,000 rpm (48,000 g), 4 °C for 1 h in a Beckman Avanti J-25 centrifuge equipped with a JA-25.50 rotor. The supernatant was decanted and further purified, the pellet disposed. If visible impurities persisted in the supernatant it was passed through a filter with a 0.45 µm cutoff. The protein of interest was separated by metal chelate affinity chromatography using a 5 mL HisTrap HP Ni^{2+} column equilibrated with His Bond buffer A. Subsequently, the bound protein was washed by passing 50 mL of buffer through the column. Elution was achieved by linearly increasing the amount of His Bond buffer B (20 mM Tris pH 7.5, 500 mM NaCl, 500 mM Imidazole) with a slope of 4 %/mL. Afterwards, the protein concentration was determined by measuring absorption at 280 nm (A_{280}) using a NanoDrop 200c and converting the result with the expected molar extinction coefficient of ε_{280} = 16000 $M^{-1}cm^{-1}$.

The protein was then dialyzed to at least 10 times the elution volume of 20 mM Tris pH 7.5, 300 mM NaCl over night (16-20 h) using a 6000 Da cutoff dialysis tube. Simultaneously the N-terminal tag was cleaved off by adding 1 mL of Sumo protease per 100 mg of protein. The dialyzed solution was then passed again through the Ni^{2+} column to remove the tag, this time using the dialysis buffer. Protein-containing fractions were pooled and concentrated with a Pall Macrosep® advance centrifugal device (10 kDa cutoff) equipped with a polyethersulfone (PU) membrane. The concentrated solution was lastly passed through a Superdex 75 gel filtration column with 20 mM phosphate pH 7.0, 150 mM NaCl as buffer. Fractions with protein were again pooled and the concentration was determined by A_{280} measurements (NanoDrop 200c) using the expected ε_{280} = 14400 $M^{-1}cm^{-1}$. If necessary, TasA was again concentrated using a PU membrane prior to filament formation (see below).

3.1.2. Filament formation

For the preparation of pure TasA filaments, monomeric TasA, as purified by size-exclusion chromatography in 20 mM phosphate buffer pH 7.0 with 150 mM NaCl, was concentrated over 2 to 3 days in an Amicon stirring device equipped with a 10 kDa regenerated cellulose membrane at 10 °C. A gel-like, turbid solution was obtained. The filaments were sedimented by ultracentrifugation for 1 h at 8 °C and 75,000 rpm (130,000 g), using a TLA110 fixed angle rotor in Beckman Optima Max ultracentrifuge yielding a translucent pellet with a whitish center.

Alternatively, TasA filaments were grown by mixing 1.5 mL of 100 µM TasA and its

accessory protein TapA (100 µM as well). The mixture was subsequently ultra-centrifuged (129,000 g) overnight at 20 °C in 20 mM phosphate buffer, pH 7.0 with 150 mM NaCl.

3.1.3. Electron Microscopy

TasA was diluted in reduced phosphate buffer (2 mM phosphate pH 7, 5 mM NaCl) to achieve sufficient filament separation. Afterwards, negative stain was achieved by application of 2 % aqueous uranyl acetate twice, dry grids were imaged with a CM100 transmission electron microscope from Philips.

3.1.4. Biofilm preparation, purification, and NMR measurement

Different biofilm strains (strain DK1042 wt and strain DK1042 Δ*tasA*) were grown in MOLP medium for 68 h at 30 °C with 300 µg $^{2}H,^{13}C,^{15}N$-TasA supplemented to the deletion strain 2.5 h after inoculation. In parallel to the Δ*tasA* + $^{2}H,^{13}C,^{15}N$-TasA preparation, cavities with medium and no bacteria were prepared as controls. After maturation, the floating biofilms (pellicles) were removed from the cavities. Pellicles from wt *Bacillus subtilis* typically form a dense phase and can be extracted using an inoculation loop. For the supplemented deletion strain, however, the integrity was not high enough and all material from the cavity had to be removed. Biofilm and medium were transferred into a 15 mL Falcon and the cavity was washed with 20 mM phosphate buffer pH 7, 50 mM NaCl. The mixture of biofilm, buffer, and medium was vortexed and subsequently centrifuged for 20 min at 8 °C and 5,000 g in an Allegra X-22R centrifuge from Beckman Coulter. While wildtype biofilm typically formed a dense pellet, the supplemented deletion strain separated into two phases: a bacterial cell pellet and a slimy supernatant phase. The cell pellet was discarded and the slimy phase (likely consisting of protein and EPS) washed further in 1.5 mL Eppendorf tubes. It was pelleted for 20 min at 8 °C with 20,000 g in an Eppendorf 5430R centrifuge with an FA-45-30-11 rotor. Afterwards, the buffer supernatant was removed, the pellet resuspended in a sufficient amount of phosphate buffer, and again centrifuged as before. In a final compression step, the pellet was resuspended in phosphate buffer, this time with an H_2O/D_2O ratio of 70/30, and transferred into 1.5 mL microfuge tubes (polypropylene) from Beckman Coulter. It was then pelleted at 45,000 rpm (90,000 g) for 1 h at 8 C° in a Beckman Optima Max Ultracentrifuge with a TLA-55 rotor. The dense pellet was subsequently filled into a 1.9 mm ZrO_2 NMR rotor.

Solid–state NMR measurements were conducted on a Bruker spectrometer operating at 700 MHz ^{1}H Larmor frequency equipped with an Avance Neo console and running TopSpin 4. Using a MAS III unit, the rotor was spun at 40 kHz in a four-channel probehead ($^{1}H,^{13}C,^{15}N,^{2}H$) with temperature control achieved by a BCU II unit (260 K at 1400 l/h

flow). A ^{1}H-^{15}N correlation was measured using the fmp.hNH pulse sequence employing two cross polarization (CP) [107] steps (^{1}H to ^{15}N and back with spin-lock frequencies of 76 kHz on ^{1}H and 30 kHz on ^{15}N) with linear 80-100 % and 100-80 % ramps on ^{1}H and a MISSISSIPPI [143] water suppression scheme without homospoil gradients prior to the second CP while magnetization was stored on ^{15}N. 90° pulses were 2.5 µs (100 kHz) for ^{1}H and 7 µs (35.7 kHz) for ^{15}N. During indirect and direct acquisition, ^{1}H and ^{15}N were decoupled with a WALTZ-16 scheme [144], frequency discrimination achieved by STATES–TPPI [145]. Further details are listed in Tab. 3.3 on page 43 for acquisition and in Tab. 3.5 on page 45 for processing parameters (row hNH$^{\#}$).

3.1.5. ^{1}H-detected solid–state NMR experiments

Measurements of TasA filaments formed by addition of TapA

Assignment spectra for TasA filaments formed through the addition of TapA and ultracentrifugation (see Section 3.1.2) were recorded on a Bruker Avance III spectrometer operating at 600 MHz ^{1}H Larmor frequency. Fully ^{2}H, ^{13}C, ^{15}N-labeled and 100 % ^{1}H back-exchanged protein was purified in 20 mM phosphate buffer pH 7.0, 150 mM NaCl and subsequently washed with 20 mM phosphate buffer pH 7.0, 50 mM NaCl. The pellet received after ultracentrifugation was transferred into a 1.3 mm ZrO_2 rotor. A MAS II unit was used to adjust the spinning rate to 60 kHz, temperature control during measurements was achieved with a BCU II, a gas flow of 1200 l/h, and a cooling gas temperature of 252 K, resulting in a sample temperature of ∼297 K. The sample temperature was determined by monitoring the water peak relative to DSS and applying the approximation $T = 485.577 - \delta(H_2O) \cdot 96.9$ where $\delta(H_2O)$ denotes the chemical shift of the water peak in ppm and T the sample temperature in °C. For the assignment spectra, manufacturer-provided pulse programs hNH2D.dcp, hCaNH3D.tcp, hCONH3D.tcp, hcaCbcaNH3D.tcp, hCOcaNH3D.tcp, hcaCbcacoNH3D.tcp and hcoCacoNH3D.tcp with CPs employing tangential ramps for hetero-nuclear magnetization transfer were used, and homo-nuclear (^{13}C-^{13}C) transfer was mediated by J-coupling as described in [110]. Parameter recommendations provided in the pulse programs were closely followed and subsequently optimized for existing conditions. 90° hard pulses were set to 2.5 µs (^{1}H), 5 µs (^{13}C) and 7 µs (^{15}N).

Spatial information was extracted from in-house programmed hNHH and hNhhNH pulse sequences. The hNHH consists of an hNH CP step yielding the two indirect dimensions followed by a 1.5 ms ^{1}H-^{1}H radio frequency-driven dipolar recoupling (RFDR) [146] step prior to acquisition (Fig. 3.1A on page 42). The hNhhNH pulse includes an additional subsequent hNH CP step (Fig. 3.1B). Acquisition and processing parameters can be found in Table 3.3 on page 43 and 3.5 on page 45, respectively.

Measurements of concentration-formed, pure TasA filaments

For a mixed labeling pattern, TasA monomer solutions were prepared containing either [^{2}H, ^{15}N]-, or [2-^{13}C]-isotopes expressed as described in section 3.1.1. These solutions were mixed 50/50 and filaments created by concentration. The resulting pellet was washed with 20 mM phosphate pH 7.0, 50 mM NaCl, and transferred into a 1.3 mm ZrO_2 rotor. For the measurements, NMR spectrometer and settings were identical to the ones used on TapA-formed filaments (previous section). A ^{1}H-^{15}N spectrum (hNH2D.dcp) for comparison was acquired optimizing the parameters previously determined for TapA-formed filaments as well. Inter-molecular contacts were monitored by an hchhNH pulse sequence (schematically shown in Fig. 3.1C on page 42). It consisted of two CP blocks (hch and hNH respectively) in between which a 3 ms RFDR mixing period [146] was introduced (180 rotor synchronized 180° pulses at 60 kHz MAS). Decoupling in the indirect dimensions was achieved by SPINAL-64 [147] (magnetization on ^{13}C) and slTPPM [148] (magnetization on ^{15}N) and by WALTZ-16 [144] during acquisition.

3.1.6. ^{13}C-detected solid–state NMR experiments

Two-dimensional ^{13}C-^{13}C dipolar-assisted rotational resonance (DARR) [108] spectra of different ^{13}C-labeling schemes (uniform-^{13}C, [1,3-^{13}C] or [2-^{13}C]) and mixing times were recorded on either 800 or 900 MHz Bruker spectrometers equipped with Avance III consoles as indicated (Tab. 3.4). Samples were filled into 3.2 mm ZrO_2 rotors and spun at MAS speeds (ω_r) of 12-13 kHz (see Tab. 3.4 as well) in dedicated three-channel probeheads (^{1}H, ^{13}C, ^{15}N, Efree on 800 MHz) with the sample temperature adjusted to 285 K by monitoring the water peak in reference to DSS (which was used for indirect referencing [89]). Typical 90° pulses were 3.125 µs (80 kHz) for ^{1}H and 5 µs (50 kHz) for ^{13}C. The cross polarization [107] transfer was optimized around the $\frac{9}{2}\omega_r$ (^{1}H) and $\frac{7}{2}\omega_r$ (^{13}C) matching condition with a linearly ramped pulse [149] on ^{13}C. DARR mixing was achieved by constant wave irradiation at ω_r for the extent of the transfer period, ^{1}H decoupling through the application of 80 kHz SPINAL-64 [147] during direct and indirect chemical shift evolution. The 2D spectra were acquired in a phase-sensitive manner using STATES-TPPI [145]. Further acquisition details can be found in Tab. 3.4 and processing parameters in Tab. 3.5.

3.1.7. Structure prediction using AlphaFold-Multimer

Structure predictions of tetrameric TasA filaments were generated based on the AlphaFold2 algorithm [127] and run via the publicly available ColabFold (https://github.com/sokrypton/ColabFold) [131] infrastructure through Google Colaboratory. The multiple sequence alignment was set to MMseqs2 (UniRef+Environmental; pair_mode: unpaired+paired),

model_type as AlphaFold2-multimer-v2 [132] and the resulting structure was not relaxed. The template_mode was set to none and num_recycles to 3.

3.1.8. Sequence alignment

Proteins homologous to TasA were located by supplying diverse sequences annotated as homologue (camelysines or TasA-like) to the program 'blast'. Evolutionary relevant sequences were chosen from the results. Alignment was then conducted employing MEGA11 [150] and the ClustalW algorithm [151] with default settings. The resulting alignment was then exported to FASTA format and visualized using the strap [152] web application (https://www.bioinformatics.org/strap/aa/).

3.1.9. Data availability

NMR raw data, chemical shift and peak lists of TasA filaments formed by addition of TapA are deposited in the BMRB as entry 51785. Processed TasA filament spectra, the CCPN 2.4.2 assignment project, TALOS+ predictions and the AlphaFold model are uploaded to zenodo (10.5281/zenodo.7534571).

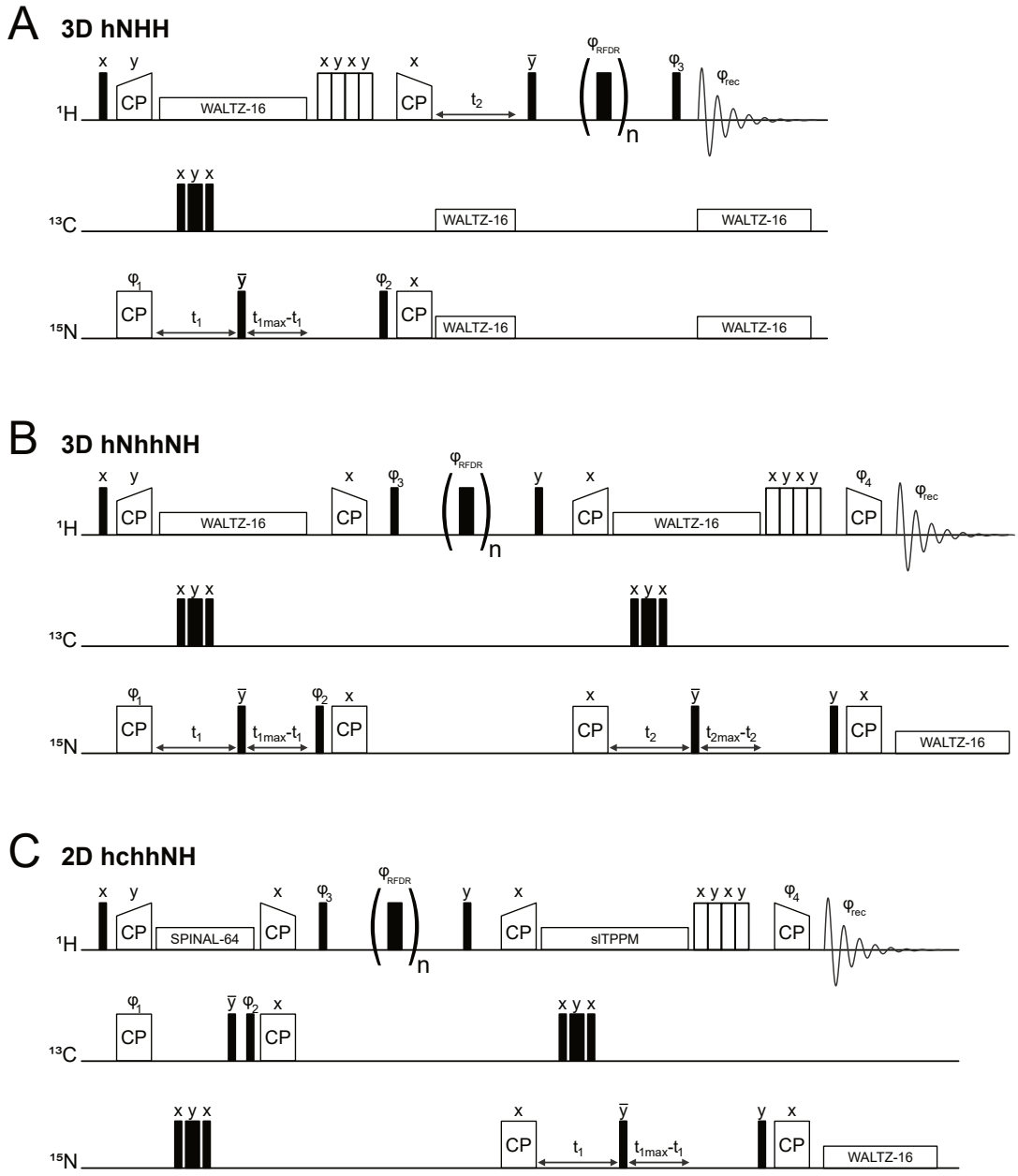

Fig. 3.1.: Pulse sequences employing through-space RFDR mixing. Schematic representations of the three pulse sequences (**A**) hNHH, (**B**) hNhhNH, and (**C**) hchhNH used to obtain through-space contacts by ^{1}H-^{1}H mixing. Black rectangles depict hard pulses (narrow: 90°; broad: 180°), empty ones indicate decoupling sequences (WALTZ-16, SPINAL-64 or slTPPM as specified), cross-polarization (CP) pulses and the MISSISSIPPI solvent suppression scheme (xyxy pulse train). Phase cycling was (**A**): φ_1=00002222, φ_2=13, φ_3=1133, φ_{RFDR}=01011010 (cycled within the RFDR sequence) and φ_{rec}=31131331; (**B**) and (**C**): φ_1=00002222, φ_2=13, φ_3=3311, φ_4=00001111, φ_{RFDR}=01011010 (cycled within the RFDR sequence) and φ_{rec}=20021331.

Tab. 3.3.: Acquisition Parameters of ^{1}H-detected TasA spectra. Parameters used for data acquisition on Bruker spectrometers. Field strength B_0 in MHz, mixing times, scan numbers (ns), the delay between scans (d1) in s, experiment duration (time), and further characteristics for each spectral dimension (F3 to F1, highest is the direct one) are indicated. All experiments of pure filaments were measured on fully ^{2}H,^{13}C,^{15}N-labeled, and 100 % back-exchanged protein at 60 kHz MAS rate. The measurement of washed biofilm (hNH$^{\#}$) was performed at 40 kHz MAS rate.

Experiment	B_0	mixing	(F3x)F2xF1	ns	aq. data points	d1	time	aq. time in ms	SW in ppm
hNH*	600	-	H x N	48	1024 x 256	1.82	7h 1min	21.3 x 55.5	40.1 x 37.9
hNH	600	-	H x N	64	2048 x 256	1.5	7h 45min	42.6 x 51.2	40.1 x 41.1
hNH$^{\#}$	700	-	H x N	512	2048 x 200	1	1d 11h	36.9 x 20	39.7 x 70.5
hCaNH	600	-	H x N x C	32	1024 x 96 x 64	1.03	2d 18h	21.3 x 20 x 6.93	40.1 x 39.5 x 30.6
hCONH	600	-	H x N x C	24	1024 x 80 x 64	1	1d 17h	21.3 x 16.7 x 16	40.1 x 39.5 x 13.3
hCOcaNH	600	-	H x N x C	2*32	1024 x 52 x 40	1	2*(22h)	21.3 x 10.8 x 10	40.1 x 39.5 x 13.3
hcoCacoNH	600	-	H x N x C	2*16	1024 x 84 x 60	1	2*(1d 2h)	21.3 x 16.8 x 7.5	40.1 x 41.1 x 26.5
hcaCbcaNH	600	-	H x N x C	2*32	1024 x 54 x 80	1	2*(1d 22h)	21.3 x 10.8 x 4.0	40.1 x 41.1 x 66.3
hcaCbcacoNH	600	-	H x N x C	3*32	1024 x 54 x 78	1	3*(1d 22h)	21.3 x 10.8 x 3.9	40.1 x 41.1 x 66.3
hNHH	600	1.5 ms	H x N x H	2*32	1024 x 80 x 80	1	2*3d	21.3 x 17.3 x 8.6	40.1 x 37.9 x 8.3
hNhhNH	600	1.5 ms	H x N x N	3*32	2048 x 80 x 80	1	3*(2d 21h)	42.6 x 17.3 x 17.3	40.1 x 37.9 x 37.9
hchhNH	600	3 ms	H x C x N	3*1024	1536 x 1 x 96	1.82	3*(2d 7h)	31.9 x 0 x 20.8	40.1 x 0 x 37.9

*filaments formation induced by concentration

$^{\#}$^{2}H,^{13}C,^{15}N-TasA in Δ*tasA* biofilm

Tab. 3.4.: Acquisition Parameters of 2D ^{13}C-^{13}C DARR TasA spectra. Parameters used for data acquisition on protein samples with different labeling schemes on Bruker spectrometers. Field strength B_0 in MHz, mixing times, magic angle spinning rates (MAS), scan numbers (ns), the delay between scans (d1) in s, experiment duration (time), and further characteristics for each spectral dimension are indicated. The ^{13}C carrier frequency was set to 100 ppm. Filaments listed were formed by concentration.

Labeling	B_0	mixing	MAS	ns	aq. data points	d1	time	aq. time in ms	SW in ppm
[uni-^{13}C]	800	20 ms	12.5 kHz	64	2374 x 960	3	2h 4min	19.9 x 9.6	295.8 x 248.5
[2-^{13}C]	900	50 ms	13 kHz	128	2048 x 800	3	3d 15h	15.0 x 7.7	302.7 x 229.8
[2-^{13}C]	800	400 ms	12.5 kHz	3*128	1536 x 768	3.03	3*(3d 22h)	15.4 x 9.6	248.5 x 198.8
[1,3-^{13}C]	900	50 ms	13 kHz	64	2048 x 800	3	1d 19h	15.0 x 7.7	302.7 x 229.8
[1,3-^{13}C]	800	400 ms	12.5 kHz	3*128	1536 x 800	2.98	3*(4d 0h)	15.4 x 10.7	248.5 x 186.4

Tab. 3.5.: Processing parameters of TasA spectra. Characteristics for each spectral dimension (F3 to F1, highest is the direct one) and applied window functions are indicated.

Experiment	(F3x)F2xF1	proc. data points	window functions
hNH*	H x N	4096 x 1024	qs 3 x qs 3
hNH	H x N	8192 x 1024	qs 3 x qs 3
hNH$^{\#}$	H x N	4096 x 1024	gm (-20, 0.1) x qs 3
hCaNH	H x N x C	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hCONH	H x N x C	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hCOcaNH	H x N x C	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hcoCacoNH	H x N x C	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hcaCbcaNH	H x N x C	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hcaCbcacoNH	H x N x C	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hNHH	H x N x H	2048 x 256 x 256	qs 3 x qs 3 x qs 3
hNhhNH	H x N x N	4096 x 256 x 256	qs 3 x qs 3 x qs 3
hchhNH*	H x N	4096 x 512	em (100) x qs 2
20 ms, [uni-^{13}C]*	C x C	4096 x 4096	qs 3 x qs 3
50 ms, [2-^{13}C]*	C x C	8192 x 4096	gm (-40, 0.06) x qs 3
400 ms, [2-^{13}C]*	C x C	4096 x 4096	qs 3 x qs 3
50 ms, [1,3-^{13}C]*	C x C	8192 x 4096	gm (-40, 0.06) x qs 3
400 ms, [1,3-^{13}C]*	C x C	4096 x 4096	qs 3 x qs 3

*filaments formation induced by concentration
$^{\#2}$H,^{13}C,^{15}N-TasA in Δ*tasA* biofilm

qs = qsine
gm = gaussian (lb,gb)
em = exponential (lb)

3.2. TccC3 from *Photorhabdus luminescens*

3.2.1. Protein expression

Protein expression and purification were performed according to standard protocols in the group of Stefan Raunser (Max Planck Institute of Molecular Physiology, Dortmund) and are detailed in [2]. Samples were frozen and shipped on dry ice to Berlin.

3.2.2. Solution NMR experiments

Comparison of TcHVR and TcART

500 µM of TcHVR, encompassing residues 679-960 of TccC3 (UniProt Q8GF97), were prepared in 500 µL of 20 mM Hepes pH 8, 200 mM NaCl. In addition, TcART (residues 776-960 of TccC3) was as well diluted to a final concentration of 500 µM in 20 mM Hepes pH 7.8, 150 mM NaCl. Both constructs contained the same C-terminal tag.

NMR measurements took place at 750 MHz (TcHVR) and 600 MHz (TcART) Bruker spectrometers equipped with 5 mm cryoprobes possessing four (^{1}H, ^{2}H, ^{13}C, ^{15}N) and five (^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{31}P) channels, respectively, together with Avance III consoles. 50 µL of D_2O were added to he samples for locking. The temperature was set to 280 K and ^{1}H-^{15}N HSQC spectra were measured. Acquisition parameters are detailed in Tab. 3.6 on page 54 in the top two rows and processing parameters similarly in Tab. 3.7 on page 55.

H-D exchange

250 µL of 1.07 mM ^{13}C, ^{15}N-labeled TcART in 20 mM Hepes pH 7.8, 150 mM NaCl were dialyzed to 75 mL of 20 mM d11-Tris pH 7.8, 150 mM NaCl in 99.9 % D_2O at 4 °C. The target buffer was exchanged two times every 12 h resulting in three steps yielding together a $125 \cdot 10^6$ dilution of the starting buffer (1:500 each step) after 36 h. The resulting protein solution was diluted to 522 µL for subsequent NMR experiments.

The sample was then directly used to acquire a water-free ^{1}H-^{15}N SOFAST-HMQC [153] that showed backbone amide signals resistant to bulk water exchange after 2 days. Afterwards, ^{1}H-^{1}H NOE correlations with different mixing times, an HCCH TOCSY (having two indirect ^{13}C and one direct ^{1}H dimensions), and a ^{13}C-edited NOESY HMQC were measured (more details concerning spectra and their interpretation below). Detailed parameters of the spectra can be found in Tab. 3.6 on page 54 for acquisition and in Tab. 3.7 on page 55 for processing.

Signal assignment and peak collection

For backbone assignment purposes 3D HNCO, HNCACB, HNCOCACB spectra [154] of ^{2}H, ^{13}C, ^{15}N-labeled TcART in 20 mM Hepes pH 7.8, 150 mM NaCl with 10 % D_2O were acquired. H^N, N, C', C_α, and C_β chemical shifts were determined and partially transferred to data of fully protonated ^{13}C, ^{15}N-labeled protein by recording a 3D HNCA in the same buffer.

To extend the assignment into the side chain an HCCH TOCSY [155] in D_2O (detailed above under H-D Exchange) with 10 ms FLOPSY-8 mixing [156] and water suppression as published in [157] was used. C_α chemical shifts were propagated from the 3D HNCA spectrum, C_β peaks were derived from deuterated data by applying an isotope correction [158] and both then utilized to connect further proton and carbon signals of the side chain. Ambiguous assignments were cross-checked with the ^{13}C-edited NOESY HMQC [159] (40 ms mixing time) measured on the same sample.

Further assignments of aromatic carbon and proton signals as well as exchangeable sites were obtained through analysis of two ^{13}C-edited NOESY HSQC spectra (centering either aliphatic or aromatic signals) and one ^{15}N-edited NOESY HSQC spectrum [154], all with 80 ms mixing time. The sample used for these experiments was ^{13}C, ^{15}N-labeled, and kept in 20 mM Hepes pH 7.8, 150 mM NaCl with 10 % D_2O.

All measurements were conducted at 280 K sample temperature as TcART degraded quickly at higher temperatures. Acquisition parameters can be found in Table 3.6 and processing parameters in Table 3.7. Peaks used as input for the structure calculation were obtained from all three NOESY spectra measured in aqueous buffer and two acquired in D_2O (a ^{1}H-^{1}H NOESY with 40 ms mixing time and the ^{13}C-edited HMQC specified above).

^{1}H-^{13}C spectra of non-dialyzed and dialyzed TcART

To remove any non-TcART components that could potentially cause NMR signals, a comparison was conducted between the TcART spectra obtained solely through gelfiltration and those obtained after an additional dialysis step. The experiments on non-dialyzed TcART were directly conducted on material received from Daniel Roderer at the Max Planck Institute of Molecular Physiology in Dortmund. A stock solution of 1.72 mM ^{13}C, ^{15}N-TcART in 20 mM Hepes pH 7.8, 150 mM NaCl was diluted with buffer to a final concentration of 300 µM. Afterwards, 20 µL D_2O were added. This sample was also used for the determination of dynamics parameter (see below).

The measurement of non-dialyzed TcART was conducted on a Bruker spectrometer operating at 600 MHz ^{1}H Larmor frequency and with a four-channel (^{1}H, ^{2}H, ^{13}C, ^{15}N), 5 mm cryoprobe at 280 K sample temperarure. An HMQC experiment was acquired with

an inter-scan delay (d1) of 1.3 s, 16 scans (ns), and 4 dummy scans (ds). In the direct (^{1}H) dimension, the time domain (td) was set 1024 points with a spectral width (sw) of 25 ppm and a corresponding acquisition time (aq) of 34 ms. In the indirect (^{13}C) dimension, td was 256, sw was 165.6 with the carrier frequency set to 82.165 ppm, and an aq of 5.12 ms. For data processing with TopSpin 3.5pl6, the data was zero-filled to 4096 (^{1}H) and 1024 (^{13}C) with a quadratic sine (qsine) window function with a sine bell shift (ssb) of 2 being applied for both dimensions.

For the dialysis of the other preparation, the 1.72 M ^{13}C, ^{15}N-TcART in 20 mM Hepes pH 7.8, 150 mM NaCl was dialyzed to 20 mM d11-Tris pH 7.8, 150 mM NaCl at 4 °C. To achieve this, a two step dialysis was conducted using a 'Slide-a-Lyzer' dialysis cassette with 3.5 kDa molecular weight cutoff from ThermoFisher. 74 µL of protein stock solution were diluted with 76 µL of Hepes buffer and, in a first step, dialyzed to 75 mL of d11-Tris buffer (∼1:1600) for 24 h. In a second step, the d11-Tris buffer was renewed and again left for dialysis (∼1:1600) with the sample, this time for 12 h. Afterwards, ∼100 µL of protein solution could be regained from the dialysis cassette to which 360 µL of d11-Tris buffer and 40 µL of D_2O were added. The final protein concentration was determined by NanoDrop (ε_{280} = 22460 $M^{-1}cm^{-1}$) to be 120 µM.

Solution NMR measurements were conducted on the same spectromer, probehead, and with the same settings used for the non-dialyzed sample. An HSQC experiment was acquired with a d1 of 1.3 s, 24 ns, and 4 ds. In the direct (^{1}H) dimension, td was set 1024 points with a sw of 25 ppm and a corresponding aq of 34 ms. In the indirect (^{13}C) dimension, td was 192, sw was 77.96 with the carrier frequency set to 39 ppm, and an aq of 8.16 ms. The small sw window lead to the ^{13}C chemical shifts of the aromatic region being aliased into the acquired range (Nyquist frequency). For data processing with TopSpin 3.5pl6, the data was zero-filled to 4096 (^{1}H) and 1024 (^{13}C) with a quadratic sine (qsine) window function with a sine bell shift (ssb) of 2 being applied for both dimensions. Spectra were visualized using CCPN Analysis Assign 3.1.1 [160].

Dynamics determination

^{15}N relaxation rates (R_1, R_2) were measured as pseudo-3D experiments, guaranteeing matching NMR conditions, at 600 MHz ^{1}H Larmor frequency, 280 K, and with 20 mM Hepes pH 7.8, 150 mM NaCl as buffer. Individual experiments with R_1 (0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.9 and 1.5 s) or R_2 (0.015, 0.035, 0.04, 0.06, 0.08, 0.13, 0.17 and 0.22 s) delays were recorded in a scrambled manner and the FIDs acquired in an interlaced fashion.

Peak heights from resulting ^{1}H-^{15}N HSQC spectra were extracted using CcpNMR version 2.4.2 [161], normalized to the highest intensity, and fit to a mono-exponential decay ($f(x) = a \cdot e^{-b \cdot x}$) using the curve_fit function of the python module SciPy. The peak

height error was estimated by determining twice the standard deviation of the spectra in a signal-free rectangular region (^{1}H: 12-9 ppm, ^{15}N 100-110 ppm). For the decay rates, standard deviations were calculated using the diagonal of the returned covariance matrix (pcov) from curve_fit yielding an estimate of the fit quality. NMR parameters are specified in Tab. 3.6 on page 54 (acquisition) and Tab. 3.7 on page 55 (processing). The pulse programs are published under [162]. For T_1 determination, a difference experiment was performed by alternating the phase of the 90° pulse prior to the incremented relaxation time [163, 164]. As a result, the intensity followed a mono-exponential decay instead of a recovery curve back to M_0.

Chemical Shift Perturbations and ^{31}P spectra

Two dimensional ^{1}H-^{15}N HSQC correlation spectra of 300 µM TcART in 20 mM Hepes pH 7.8, 150 mM NaCl, were recorded at 280 K with two different NAD^+ concentrations (1 and 3 mM) and compared to known data of ligand-free protein. Chemical shift perturbations were determined as a weighted shift distance $d = \sqrt{\Delta\delta_H{}^2 + (0.14 \cdot \Delta\delta_N)^2}$ [165]. NMR parameters are detailed in Tabs. 3.6 and 3.7. Peak tracking, signal assignment, and chemical shift extraction were conducted with CcpNMR 2.4.2 [161].

The TcART/NAD^+ mixtures specified above were as well used to acquire ^{31}P spectra, complemented by 10 mM pure NAD^+ prepared in 20 mM Hepes pH 7.8, 150 mM NaCl to determine chemical shifts of the free molecule. Experiments were run on a Bruker spectrometer operating at 600 MHz ^{1}H Larmor frequency equipped with an Avance III console and a five-channel (^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{31}P) 5 mm cryoprobe. One ^{31}P excitation pulse (10 µs, 25 kHz) was applied with an inter-scan delay (d1) of 2 s to collect 2 s (131072 points) of time domain data with a spectral width of 128.6 ppm (31.25 kHz). During acquisition and d1, 3.125 kHz WALTZ-16 [144] was applied to decouple ^{1}H and enhance the ^{31}P signal through NOE, respectively. The number of scans (dummy scans) was 8 (4) for pure NAD^+ as well as 342 (4) for 1 mM NAD^+ and 8192 (8) for 3 mM NAD^+ with TcART. Exponential modulation (em) with a line broadening (lb) of 1.0 was applied as window function and 65536 FID data points zero-filled to 131072. The baseline between 20 and -26 ppm was subsequently corrected using a 5th-degree polynomial function. Processing was performed with TopSpin 3.5pl6 and spectra were visualized using CCPN Analysis Assign 3.1.1 [160].

3.2.3. Structure calculation

Structure calculation was performed with an iterative NOE assignment procedure using the software ARIA version 2.3.2 [166, 167] coupled with CNS version 1.21 [168].

Cross-peaks from NOESY-type spectra were submitted to ARIA directly from CcpNMR

2.4.2 [161] for several cycles of automated NOE assignment and structure calculation. Peaks in proximity to the water resonance (4.72 - 4.80 ppm) were manually removed. To ensure a better representation of the conformational space allowed from the NOE-derived distance restraints and to prevent over-convergence to a possibly artefactual single state conformation, we implemented a consensus procedure as published by Buchner and Güntert in [169]. A script to perform consensus calculations with ARIA is available at http://aria.pasteur.fr.

The consensus procedure consists of 20 ARIA runs that were performed independently. They used the same input data but different random number seeds, resulting in varying starting conformations and initial velocities for the molecular dynamics-simulated annealing protocol. Then, cross-peaks that remained active (*i.e.*, for which at least one assignment possibility was kept) at the end of at least 12 out of the 20 ARIA runs were collected. For each of the active cross-peaks, the assignment possibilities from individual ARIA runs were combined to yield a new list of consensus (ambiguous) distance restraints. Finally, a new ARIA run is performed with a single iteration and using the consensus distance restraints as input to produce the final consensus structure ensemble.

For all ARIA runs, the NOE data were supplemented with backbone dihedral angle restraints derived from TALOS+ [170] predictions based on H_α, H^N, N^H, C_α, C_β secondary chemical shifts with a possible error of $\pm 20°$ and hydrogen bond restraints extracted from H-D exchange data, secondary structure pattern, and NOE evaluation. In the individual ARIA runs, nine iterations were performed with an adaptive tolerance procedure to discard unsatisfied distance restraints [171] and restraint combination ($4 \rightarrow 4$) [172] was employed for the first 4 iterations. At each iteration, 50 conformers were calculated (except for the last one where 100 were generated). The consensus calculation was run twice and in each case, a single iteration was performed generating 200 conformations of which the 15 lowest-energy ones were refined in a shell of water molecules [173]. Of both water refinements, the 5 lowest energy structures were selected to represent the final coordinates. A log-harmonic energy potential with optimal weighting of distance restraints was always applied during the simulated annealing [174]. To improve convergence, the number of molecular dynamics steps at the two cooling stages of the simulated annealing runs was increased to 40,000 for the individual runs and 100,000 for the consensus runs. Violated restraints were analyzed and assignments manually corrected. Structural and restraints statistics are reported in Table A.2 on page 125, together with excerpts of the PSVS report. The full ensemble is shown in Fig. 4.14A on page 79.

3.2.4. Cryo-EM of ADP-ribosylated F-actin

The cryo-EM analysis of ADP-ribosylated-F-actin was performed by Alexander Belyy at the Max Planck Institute of Molecular Physiology in Dortmund as published under [2].

3.2.5. Docking of NAD^+ into the cryo-EM structure

Docking of NAD^+ into the cryo-EM structure was performed by Jonas Protze as published in [2]. For the molecular docking, Glide [175] as included in the Maestro 12v7 software package (https://www.schrodinger.com/maestro) was used. This algorithm applies a series of hierarchical filters in the search for ligand positions. The *receptor grid* for the binding site of the TcART-F-actin complex was set up with default parameters. The grid was centered on the cryo-EM electron density in the binding pocket. Flexible docking of NAD^+ was carried out with *XP settings* (extra precision) utilizing standard *core pattern comparison* with the ADP-ribose position in the post-reaction state as reference. A tolerance (RMSD) of 2.5 Å was applied. The docking yielded five very similar poses for NAD^+ and the best pose was chosen according to the docking score and fit with the NMR chemical shift experiments.

3.2.6. Comparison of TcHVR and the TcdB2-TccC3 complex

For proper comparison of TcHVR with solid–state NMR experiments conducted on TcdB2-TccC3, sample conditions were chosen to be as similar as possible. The corresponding UniProt entries are Q8GF99 (TcdB2) and Q8GF97 (TccC3).

Solution NMR of TcHVR

For the measurement to be conducted under the same conditions, the TcHVR buffer was exchanged to 20 mM phosphate pH 7, 50 mM NaCl by multiple steps of concentration and dilution in an Amicon® 15 Ultra (Millipore) centrifugal device with a molecular weight cutoff of 10 kDa (protein size >30 kDa).

550 µL of 600 µM protein together 50 µL D_2O were measured at 300 K over night on a 600 MHz Bruker spectrometer equipped with a 5 mm cryoprobe possessing five channels (^1H, ^2H, ^{13}C, ^{15}N, ^{31}P). An ^1H-^{15}N HSQC spectrum was acquired, details can be found in Tabs. 3.6 on page 54 (acquisition) and 3.7 on page 55 (processing) marked with $^{\#\#}$. A small amount of precipitated protein was visible the next day but did not interfere with the measurement of soluble protein.

Solid–state NMR of TcdB2-TccC3

Prior to data acquisition the sample buffer was changed as well to 20 mM phosphate pH 7, 50 mM NaCl by multiple cycles of concentration and dilution in a Amicon® 15 Ultra (Millipore) concentrator with 50 kDa molecular weight cutoff (protein size: >270 kDa). Subsequently, the rotor was filled by ultracentrifugation at 71,100 g, 8 °C for 65 h using a tool as described in [176].

Measurements of the TcdB2-TccC3 complex, further only referred to as TcB-TcC, were conducted at 100 kHz MAS using a narrow bore four-channel (^{1}H, ^{13}C, ^{15}N, ^{2}H) 0.7 mm probe on a 900 MHz Bruker spectrometer equipped with an Avance III console and a MAS III unit. TcB-TcC with and without TcHVR (^{1}H,^{15}N-labeled) were measured. Temperature calibration was performed on a water/DSS sample by monitoring the water peak in reference to DSS, the water peak of the samples was used for indirect referencing [89]. Parameters of the cross polarization-based hNH spectra (analogous to a ^{1}H-^{15}N HSQC in solution) can be found in the bottom rows of Tab. 3.6 and Tab. 3.7.

3.2.7. Co-sedimentation assay

Co-sedimentation experiments were conducted by Alexander Belyy at the Max Planck Institute of Molecular Physiology in Dortmund, as published under [2]. For the assay, an aliquot of freshly thawed G-actin was centrifuged at 150,000 g using a TLA-55 rotor for 20 min at 4 °C to remove possible aggregates. Afterwards, actin was polymerized by incubation in F-buffer (120 mM KCl, 20 mM Tris pH 8, 2 mM $MgCl_2$, 1 mM DTT, and 1 mM ATP) for 2 h at room temperature. To stabilize the filaments, phalloidin was added in 1.5 excess over actin after polymerization. Co-sedimentation assays were performed in 20 µL volumes by first incubating F-actin with the specified amount of protein (in the presence of 1 mM of NAD^{+}, if indicated) for 5 min at room temperature, then centrifuging the mixture at 120,000 g using a TLA120.1 rotor for 20 min at 4 °C. After centrifugation, aliquots of the supernatant and pellet fractions were separated by SDS-PAGE and analyzed by densitometry using Image Lab software version 5.2.1 (Bio-Rad) and Prism version 9 (GraphPad Software). For this thesis, the published data was taken and fit to a function typical for a saturated binding reaction ($f(x) = (a \cdot x)/(b + x)$) with the curve_fit function of the python module SciPy.

3.2.8. Data availability

The molecular coordinates of the NMR structure ensemble are deposited in the PDB under accession code 7ZBQ. NMR raw data and peak lists are available in the BMRB as 34717 (protonated protein assignment and structure calculation data), 51438 (deuterated protein

assignment data), and 51478 (relaxation data). Raw and processed spectral data, the CCPN 2.4.2 assignment project, and structure calculation files are uploaded to zenodo with the doi: 10.5281/zenodo.7998392.

The coordinates determined by our collaborators for the cryo-EM structures of the TcART-F-actin complex and ADPR-F-actin have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-14532 and 14533. Corresponding molecular models for ADPR-F-actin, and the TcART-F-actin complex have been deposited in the PDB with accession codes 7Z7H and 7Z7I.

3.3. Thesis preparation

This thesis was written in LaTeX and compiled using TeXstudio 4.5.1 running on Windows 10. The bibliography was prepared with JabRef 5.9 portable and compiled with BibLaTeX using BibTeX as backend. Figures were prepared with Chimera 1.15 for protein structures and Affinity Designer 2.1 licensed by the Free University Berlin for graphics.

Tab. 3.6.: Acquisition Parameters of TcHVR, TcART and TcB-TcC spectra. Data acquisition parameters on Bruker spectrometers. Field strength B_0 in MHz, mixing times, scan numbers (ns), delay between scans (d1) in s, experiment duration (time) and further characteristics for each spectral dimension (F3 to F1, highest is the direct one) are indicated. Top rows specify values for solution NMR experiments of TcHVR[#] and TcART acquired with TopSpin 3.5pl6, the two bottom for solid–state measurements of the complete TcB-TcC complex conducted with TopSpin 4.1.0.

Experiment	B_0	mixing	(F3x)F2xF1	ns	aq. data points	d1	time	aq. time in ms	SW in ppm
^{1}H-^{15}N HSQC[#]	750	-	H x N	16	1024 x 256	1.3	1h 40min	41 x 33.8	16.7 x 49.8
^{1}H-^{15}N HSQC	600	-	H x N	4	2048 x 512	1.3	50 min	102.4 x 85	16.7 x 49.5
^{1}H-^{15}N HMQC*	600	-	H x N	32	1024 x 256	0.1	25 min	51.2 x 42.5	16.7 x 49.5
HNCO	600	-	H x N x C	8	1024 x 80 x 128	1.3	1d 9h	51.2 x 16 x 25.6	16.7 x 41.1 x 16.6
HNCACB	600	-	H x N x C	16	1024 x 80 x 140	1.3	3d 7h	51.2 x 16 x 7	16.7 x 41.1 x 66.3
HNCOCACB	600	-	H x N x C	16	1024 x 80 x 140	1.3	3d 8h	51.2 x 16 x 7	16.7 x 41.1 x 66.3
HNCA	600	-	H x N x C	32	1024 x 128 x 100	1.3	3d 10h	51.2 x 21.2 x 7.1	16.7 x 49.5 x 46.7
HCCH TOCSY*	600	10 ms	H x C x C	16	1024 x 128 x 128	1	3d 8h	51.2 x 5.2 x 5.2	16.7 x 80.8 x 80.8
^{13}C NOESY HMQC*	600	40 ms	H x C x H	8	1024 x 128 x 256	1.3	3d 18h	51.2 x 5.2 x 15.4	16.7 x 80.8 x 13.9
^{13}C NOESY HSQC aliph.	750	80 ms	H x C x H	8	1024 x 164 x 300	1.3	6d 14h	41 x 5.6 x 15.0	16.7 x 78 x 13.3
^{13}C NOESY HSQC arom.	750	80 ms	H x C x H	8	1024 x 150 x 320	1.3	6d 10h	41 x 5.1 x 16	16.7 x 78 x 13.3
^{15}N NOESY HSQC	750	80 ms	H x N x H	8	1024 x 128 x 300	1.3	5d 4h	41 x 20.5 x 15	16.7 x 41.1 x 13.3
^{1}H-^{1}H NOESY*	600	40 ms	H x H	80	2048 x 1024	1.3	1d 10h	102.4 x 6.1	16.7 x 13.9
^{1}H-^{15}N HSQC 1mM NAD^{+}	600	-	H x N	4	1024 x 256	1.3	24 min	51.2 x 42.5	16.7 x 49.5
^{1}H-^{15}N HSQC 3mM NAD^{+}	600	-	H x N	4	1024 x 256	1.3	24 min	51.2 x 42.5	16.7 x 49.5
T_1 relaxation	600	-	H x N x t1	32	1024 x 256 x 8	1.3	2d 5h	51.2 x 42.5 x 1500	16.7 x 49.5 x 0
T_2 relaxation	600	-	H x N x t2	32	1024 x 256 x 8	1.3	1d 6h	51.2 x 42.5 x 220	16.7 x 49.5 x 0
^{1}H-^{15}N HSQC[##]	600	-	H x N	16	1024 x 256	1.3	1h 40min	51.2 x 42.5	16.7 x 49.5
hNH CP TcB-TcC +hvr	900	-	H x N	32	2048 x 1200	0.98	10h	28.7 x 48	39.7 x 137
hNH CP TcB-TcC -hvr	900	-	H x N	32	2048 x 1200	1	11h	28.7 x 48	39.7 x 137

[#] TcHVR measured [##] TcHVR measured with solid–state NMR buffer *sample measured in 99.9 % D_2O buffer

Tab. 3.7.: Processing parameters of TcHVR, TcART and TcB-TcC spectra. Parameters used for data processing using either TopSpin 3.5pl6 (solution NMR data of TcHVR[#] and TcART, top rows) or TopSpin 4.1.0 (solid–state data of TcB-TcC, three bottom rows). Characteristics for each spectral dimension (F3 to F1, highest is the direct one) and applied window functions are indicated.

Experiment	(F3x)F2xF1	proc. data points	window functions
^{1}H-^{15}N HSQC[#]	H x N	4096 x 1024	qs 3 x qs 3
^{1}H-^{15}N HSQC	H x N	4096 x 1024	qs 3 x qs 3
^{1}H-^{15}N HMQC*	H x N	4096 x 1024	qs 3 x qs 2
HNCO	H x N x C	1024 x 512 x 512	qs 2.5 x qs 2 x qs 2
HNCACB	H x N x C	1024 x 512 x 512	qs 2.5 x qs 2 x qs 2
HNCOCACB	H x N x C	1024 x 512 x 512	qs 2.5 x qs 2 x qs 2
HNCA	H x N x C	4096 x 256 x 256	qs 3 x qs 2 x qs 2
HCCH TOCSY*	H x C x C	2048 x 512 x 256	qs 2.5 x qs 2 x qs 2
^{13}C NOESY HMQC*	H x C x H	2048 x 512 x 256	qs 3 x qs 2 x qs 2
^{13}C NOESY HSQC aliph.	H x C x H	1024 x 512 x 512	qs 2 x qs 2 x qs 2
^{13}C NOESY HSQC arom.	H x C x H	2048 x 512 x 1024	qs 3 x qs 2 x qs 3
^{15}N NOESY HSQC	H x N x H	2048 x 256 x 512	qs 3 x qs 2 x qs 3
^{1}H-^{1}H NOESY*	H x H	4096 x 4096	gm (-40.0, 0.1) x qs 3
^{1}H-^{15}N HSQC 1 mM NAD^{+}	H x N	4096 x 2048	qs 3 x qs 3
^{1}H-^{15}N HSQC 3 mM NAD^{+}	H x N	4096 x 2048	qs 3 x qs 3
T_1 relaxation	H x N x t1	4096 x 1024	qs 2 x qs 2
T_2 relaxation	H x N x t2	4096 x 1024	qs 2 x qs 2
^{1}H-^{15}N HSQC[##]	H x N	4096 x 2048	qs 3 x qs 3
hNH CP TcB-TcC +hvr[§]	H x N	4096 x 4096	gm (-30.0, 0.05) x qs 2
hNH CP TcB-TcC +hvr	H x N	4096 x 4096	gm (-20.0, 0.1) x qs 3
hNH CP TcB-TcC -hvr	H x N	4096 x 4096	gm (-20.0, 0.1) x qs 3

[#]TcHVR measured
qs = qsine
[##]TcHVR measured with solid–state NMR buffer
gm = gaussian (lb,gb)
*sample measured in 99.9 % D_2O buffer
[§]used to determine arginine Nε integrals, 644 time domain points processed in F1

4. Results

4.1. TasA stabilizes *Bacillus subtilis* biofilms

4.1.1. TasA filaments converge to a single state

TasA is the main protein component of *Bacillus subtilis* biofilms and has been an object of intense studies due to its important stabilizing function of biofilm assemblies [22]. However, there is significant divergence in the literature concerning its predominant state under native conditions. While some studies focused on amyloid fibers prepared from unfolded protein as the most relevant state [22, 52], others reported assemblies formed from folded monomers to be of highest importance [54, 56, 177]. To focus on a defined state of TasA during this work, it was necessary to ensure that preparations were homogeneous throughout all experiments. The formation of TasA filaments can be either initiated by high concentrations (> 1 M) or by addition of the accessory protein TapA (details in Material and Methods 3.1.2 on page 37). The resulting pellets were analyzed and showed identical characteristics in transmission EM (negative stain) and solid–state NMR (^{1}H-^{15}N spectra) (Fig. 4.1A-D). In EM micrographs, separated and non-connected filaments are visible that present a beads-on-a-string-like appearance with a periodic oscillation and overall bending. The recorded ^{1}H-^{15}N spectra of pure and TapA-induced TasA filaments match very well, indicating structural similarity on the atomic level. Interestingly, at higher concentrations, bundles are formed that can be disassembled into single filaments through simple dilution. Taken together, it is evident that both ways of polymerizing TasA yield the same filamentous state which we and others recently determined not to be amyloid as no Thioflavin T binding can be detected [1, 54, 55].

The next step was to establish whether the observed filamentous state, prepared from folded, monomeric TasA, resembles the one under native conditions. For this, biofilm cultures were grown by inoculating a Δ*tasA Bacillus subtilis* strain into MOLP medium and allowing it to migrate to the air / liquid interface. Subsequently, purified and folded TasA

was carefully added on top and the biofilm was grown to maturation, forming a floating phase ('pellicle'). The TasA-supplemented pellicle shows increased wrinkle formation and bears a closer resemblance to the wildtype than Δ*tasA Bacillus subtilis* alone (Fig. 4.1E). This indicates that externally provided TasA protein can be incorporated into the biofilm structure.

To characterize TasA in its native environment, ^{2}H, ^{13}C, ^{15}N-labeled TasA was supplied to the pellicle of a Δ*tasA* strain and, after maturation and washing, measured by solid–state NMR. Interestingly, the observed ^{1}H-^{15}N fingerprint (Fig. 4.1F) is close to the one of filaments generated from pure protein (Fig. 4.1C and D). Only minor differences occur compared to to the filaments prepared from recombinant TasA (*e.g.*, in the ^{15}N range of 100-105 ppm), hinting towards a slight change of protein conformation likely at the interface to other biofilm components (polysaccharides, membranes, etc.). Furthermore, incorporating different fractions received from the size-exclusion chromatography to Δ*tasA Bacillus subtilis* yielded the same ^{1}H-^{15}N spectrum, signifying that previously different TasA species converge to the same state in the biofilm (Appendix Fig. A.2 on page 110).

The collected data clearly shows that TasA filaments present a favorable state that is readily adopted when starting from folded protein under both *in vitro* and *in vivo* conditions.

4.1.2. Assignment of NMR signals exposes peak doubling

Having validated the employed preparation method, assignment of the NMR signals of TasA was the next step. For this purpose, different approaches were evaluated. We found that ^{1}H-detected NMR experiments on ^{2}H, ^{13}C, ^{15}N-TasA together with complete back-exchange in H_2O at high speed MAS (60 kHz) provided the most promising results. Assignments were obtained through a suite of NMR spectra recorded with CP during heteronuclear and *J*-based schemes for homonuclear transfers [110]. Especially for experiments with many magnetization transfer steps, we found that for our TasA samples *J*-based transfers out-perform sequences that are exclusively based on dipolar couplings for magnetization exchange between carbon nuclei.

The final assignment extends to 87 % of all backbone atoms of TasA. These atoms distribute over 213 out of 234 residues (91 %) as sometimes not all atoms of each residue were assigned. The quality of the employed spectra may be appreciated from the exemplary strip plots in Appendix Figures A.3 and A.4 (pp. 111 and 112). These strips also show the sequential linkages used in the assignment procedure. A selection of assignments are indicated in Fig. 4.2A. Although at most 227 signal sets are expected in the hCaNH experiment (with exclusion of the seven proline residues that do not possess an amide proton), 245 peaks can be unambiguously mapped to C_α nuclei. This is due to a number

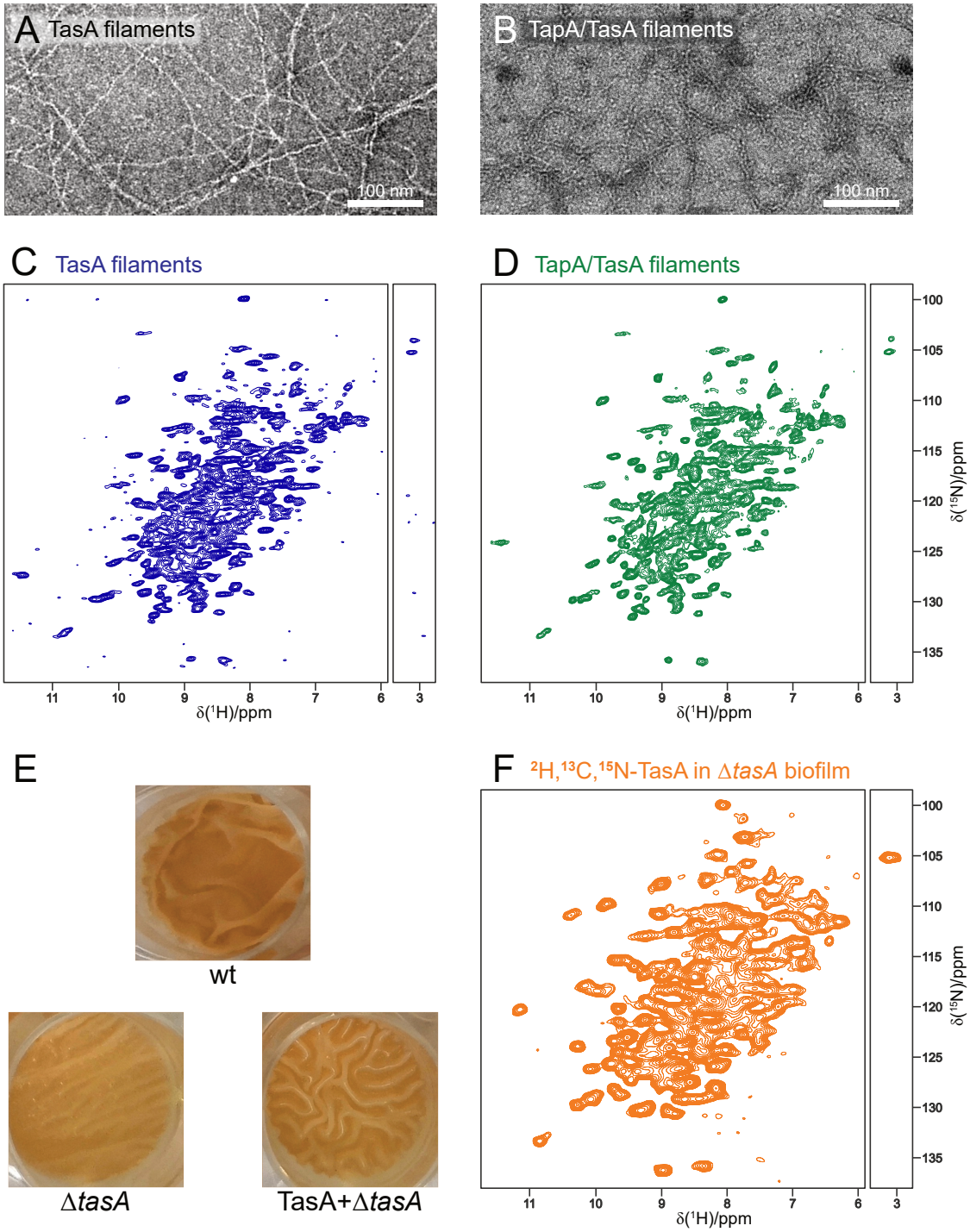

Fig. 4.1.: TasA preparations examined by EM and NMR. Transmission EM micrographs of pure TasA (**A**) and TapA-induced TasA filaments (**B**) show separated filaments. ^{1}H-^{15}N spectra of those preparations are presented in (**C**) for pure TasA and (**D**) for TapA-induced filaments. (**E**) Top view of *Bacillus subtilis* pellicle biofilms observed after 68 h of the wildtype (wt), of a deletion strain unable to produce TasA (Δ*tasA*), and of the deletion strain supplemented with recombinantly expressed TasA protein (TasA+Δ*tasA*). (**F**) ^{1}H-^{15}N spectrum of ^{2}H,^{13}C,^{15}N-labeled TasA added to a Δ*tasA* strain with subsequent isolation of the biofilm phase. The peak in the NMR spectra at $\delta(^{1}\mathrm{H})$ >11 ppm belongs to a histidine sidechain and folds into the spectral range.

of residues that split into two signal sets. Such doubling was identified for 27 amino acids (indicated with red labels in Fig. 4.2A) throughout the sequence, typically manifesting as splitting into two side-by-side peaks with similar but distinct chemical shift. A typical strip is shown in Fig. 4.2B. To ascertain which conclusions can safely be drawn from the spectra, their quality and significance needs to be examined in detail.

The signal-to-noise ratio and linewidth in all acquired spectra are notably good. However, the observed signal doubling is possibly more prevalent than currently estimated. For proper identification of this phenomenon, a sufficiently narrow linewidth is necessary. Even at optimal conditions, the splitting of peaks can only be identified if the difference in chemical shift is large enough relative to the linewidth. The spectral resolution itself highly depends on the acquisition method and overall parameters. Peaks that clearly show a doubling in the two-dimensional ^{1}H-^{15}N correlation at 60 kHz have broadened lines at lower MAS rates (40 kHz) making the splitting hard or impossible to observe. Additionally, signal interpretation is more difficult for spectra with lower signal-to-noise ratio. For instance, this is shown for the peak of K239 (Fig. 4.2C) recorded with different pulse sequences employing either two (hNH) or three (hcaNH) magnetization transfer steps. While the doubling can be easily identified in the hNH spectrum, the lower noise threshold makes this distinction more difficult in the hcaNH. Furthermore, another issue that can hamper the assignment is the overlap of peaks. It is especially prevalent in the central section of two-dimensional spectra (Fig. 4.2A, $\delta(^{1}\mathrm{H})$: 8-9 ppm) and even partially persists when advancing to three dimensions. When taking all these factors into account, it seems reasonable that further peak doubling might be present in the recorded spectra, but is obfuscated by the linewidth, the signal-to-noise ratio, and peak overlap.

However, the overarching goal was not to find singular clues of local structural traits but rather a characterization of the entire TasA protein and its filamentous structure. Such information is readily contained in the NMR assignment. However, a detailed interpretation of the NMR data must based upon an initial model, for which the well-characterized, monomeric state observed by X-ray can be used.

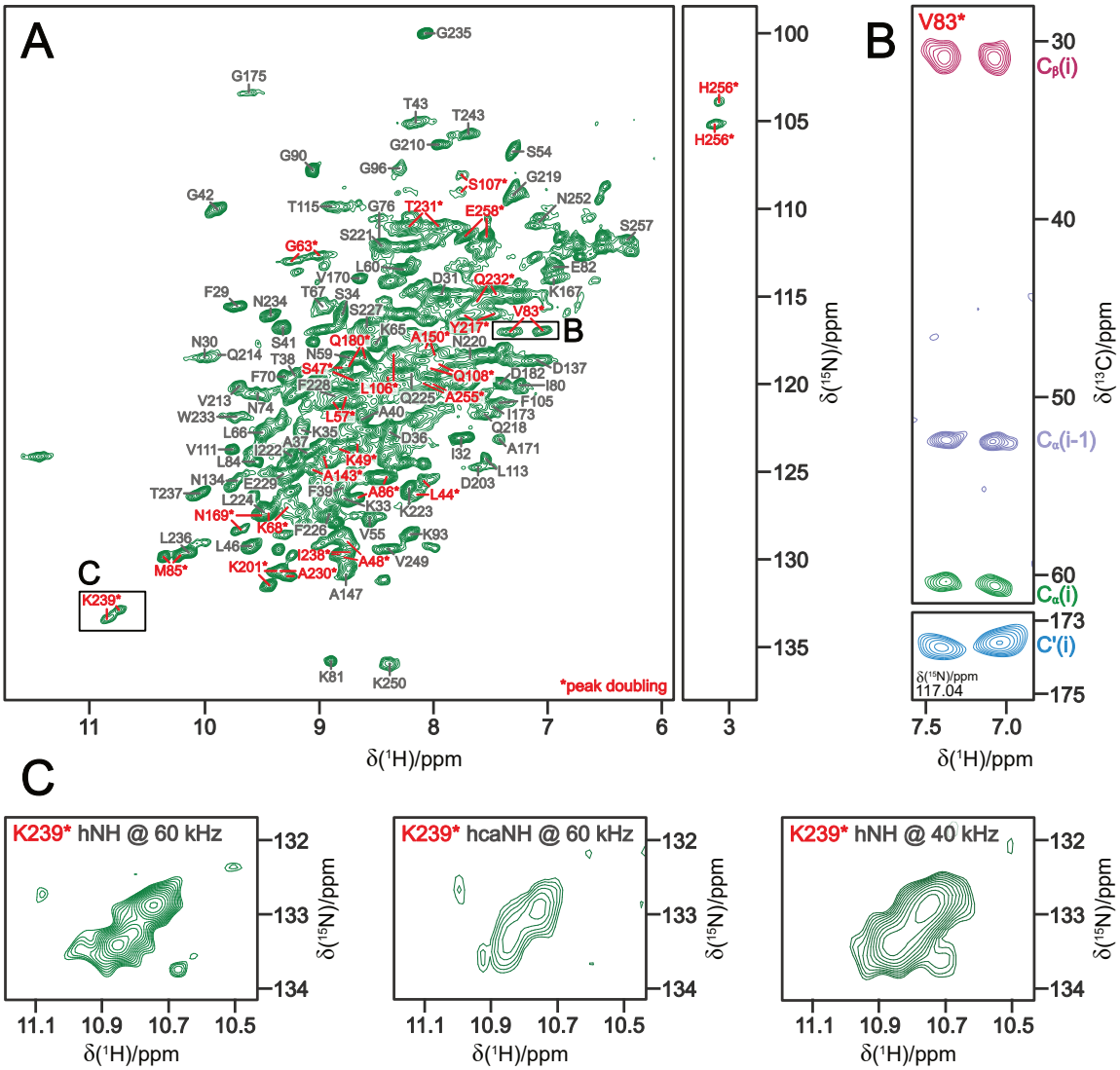

Fig. 4.2.: Peak doubling in TasA spectra. (**A**) ^{1}H-^{15}N spectrum of TasA filaments with selected assignments indicated. A full list of the assignments is uploaded to the BMRB (accession code 51785). Peaks of residues that show a splitting are indicated in red and marked with an asterisk. Black boxes show regions that are examined further in (**B**) and (**C**). (**B**) Strip plot displaying doubling of C_α (green), C_β (rose) and C' (blue) for V83 and the C_α of the previous amino acid (indigo). (**C**) Comparison of the resolution of the K239 peak pair, at 60 kHz MAS for the hNH experiment (left), a projection of the hcaNH under identical conditions (middle) and a hNH at 40 kHz. All three spectra are processed without window functions and visualized with the same contour settings.

4.1.3. TasA retains monomer structure elements when in filaments

As a first step for data interpretation, the NMR results were evaluated in terms of structural properties. The dihedral angles of the protein backbone were predicted by TALOS+ [170] using the set of backbone chemical shifts as input.

Interestingly, the secondary structure measured for filamentous TasA appears to be highly conserved compared to the X-ray structure of the monomer. In particular, helices and most β-strands are predicted at the same position in the sequence, which is apparent when color-coding the monomer structure according to the TALOS+ output (Fig. 4.3). The main differences arise in the N-terminal region where residues preceding β1 (28-41), disordered in the crystal structure, now exhibit typical β-strand chemical shifts. Conversely, the previously ordered strand β2 now displays characteristics of a random coil arrangement. At the other end of the sequence, the C-terminal residues are not part of the X-ray structure since truncation was necessary for crystal formation [177]. However, their signals could be assigned in the solid–state NMR spectra. Although TALOS+ predicted them to be predominantly in random coil conformations, the mere observation of their peaks by solid–state NMR speaks against a high degree of flexibility. This finding is corroborated by H256 having an anomalous H^N chemical shift of 3.2 ppm (see Fig 4.2A), indicative of a well-defined and rigid interaction with an aromatic ring [178, 179]. To create a more detailed picture of retained as well as new structural elements, we recorded NMR spectra from which ^{1}H-^{1}H contacts can be deduced.

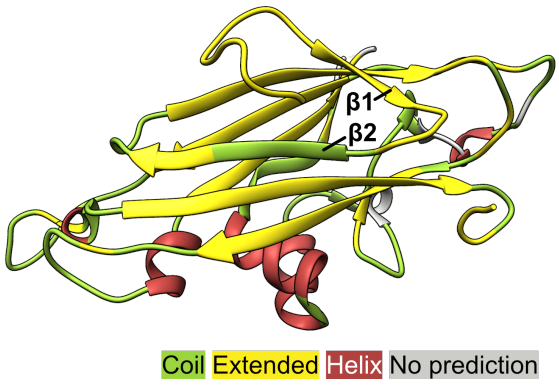

Fig. 4.3.: NMR-derived dihedral angles indicate retained secondary structure. X-ray model of monomeric TasA (PDB 5OF1) color-coded according to the TALOS+ secondary structure prediction based on backbone chemical shifts obtained from filaments. Predicted sheets are colored in yellow, helices in red, coiled regions in green and non-predicted stretches in gray. The initial β-strands β1 and β2 are labeled.

4.1.4. ^{1}H-^{1}H contacts pinpoint structural rearrangements

Backbone contacts confirm overall fold retention

In order to determine ^{1}H-^{1}H contacts by NMR, hNHH and hNhhNH spectra with ^{1}H-^{1}H RFDR mixing to transfer between spatially adjacent sites (Fig. 3.1A and B on page 42) were recorded. The experiments were conducted on filaments prepared from ^{2}H,^{13}C,^{15}N-labeled protein with a 100 % back-exchange rate of the exchangeable protons. As such, the backbone H^N atoms account for the majority of signals. From their symmetric cross-peak pattern, the local secondary structure can be deduced [180]. Loops and α-helices typically show strong peaks for sequential H^N contacts (< 3 Å), with α-helices manifesting additional, weaker signals for the contacts of the types $H^N(i) \leftrightarrow H^N(i+2)$ (∼ 4.2 Å) and $H^N(i) \leftrightarrow H^N(i+3)$ (∼ 4.8 Å). In contrast, β-sheets have weak sequential but rather strong cross-strand contacts. This characteristic is especially pronounced in anti-parallel β-sheets where the H^N cross-strand distance is only ∼3.3 Å (Fig. 4.4A). Due to the nature of anti-parallel β-sheets, the observed $H^N(i) \leftrightarrow H^N(j)$ cross-strand contact corresponds to the formation of two hydrogen bonds for each pair of facing residues (Fig. 4.4A). As TasA consists predominantly of anti-parallel β-sheets (Fig. 4.4B), this bonding pattern is expected to occur rather frequently.

Indeed, most detected contacts (Appendix Tab. A.1 on page 124) are in line with those expected from the secondary structure of the X-ray model. The pattern of through-space cross-peaks provides additional support to the TALOS+ prediction of the conserved positions of α-helices and loops in the sequence. Furthermore, the presence and equivalent pair-wise connectivity of strands in the anti-parallel β-sheets, β3↔β8, β4↔β9, and β5↔β6, is proven. Notably, hydrogen bonds between β5↔β8 and β4↔β7 cannot be validated and β9 does not bind β2 anymore. Moreover, the stretch from G42 to D45, creating β1 in the X-ray model, shows no interaction with β3.

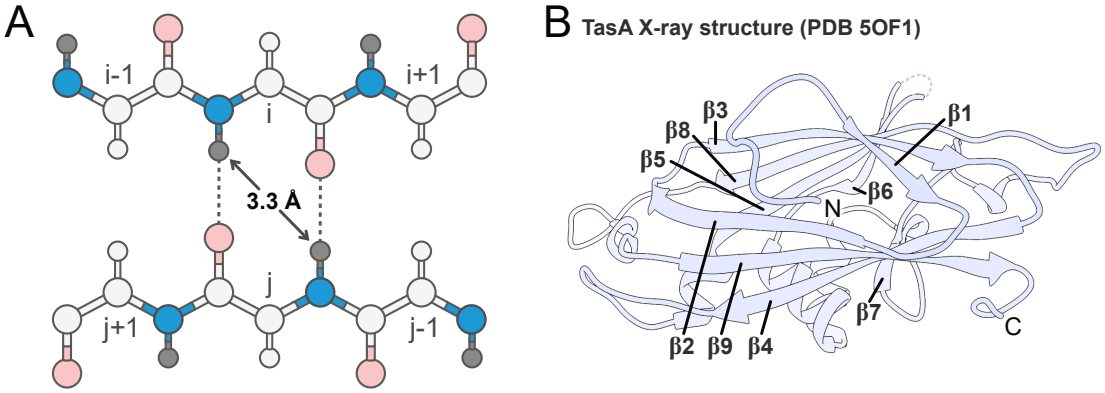

Fig. 4.4.: β-sheets in monomeric TasA. (**A**) Schematic pattern of an anti-parallel β-sheet with the cross-strand hydrogen bonds and the characteristic H^N-H^N distance of 3.3 Å indicated. (**B**) X-ray structure of monomeric TasA (PDB 5OF1) with β-sheet designation shown.

Hydrogen bonds delineate β-sheet contacts

A detailed analysis of ^{1}H-^{1}H contacts, that are not present in the monomeric X-ray model, can help to identify newly formed structural elements in TasA filaments. The corresponding strip plots of through-space spectra are discussed later in a broader context and can be found in Fig. 4.8 on page 71.

The most prominent example of a newly formed structure within the filaments is located at the N terminus. There, residues 31-41 form a new strand (β0), which aligns anti-parallel with the residues 230 to 220 of β9. The formation of this β-sheet (β0↔β9) is apparent from the cross-peak pattern and the $\delta(^{1}\mathrm{H})$ range of the measured chemical shifts. As typical for hydrogen-bonded protons in a β-sheet conformation, the H^{N} signals of β0 display values larger than 8.5 ppm. The only exceptions are S41 and D31 that are located at the outer boundaries of β0. Because 31-41 (new strand β0) and 230-220 (β9) are not aligned in the X-ray model (Fig. 4.5A), and the contacts between β9 and β4 are present in the NMR data, the N-terminal section is expected to rearrange when TasA transitions to filaments.

The conformational change at the N terminus is further corroborated by the location of G42, which was previously part of β1 but now directly follows the newly formed strand β0. It has a distinguished position in the ^{1}H-^{15}N correlation (Fig. 4.2A) of TasA filaments and does therefore create highly distinctive cross-peak signals. The pattern of G42 with N59 and L60 indicates that these residues are close in space, a conformational trait likely associated with N59 and L60 being located in a loop region. The distance between G42 and N59 / L60 is $\sim$ 24 Å in the crystal structure (Fig. 4.5B). In addition, N59 has a less-pronounced cross-peak in the hNhhNH spectrum with A40, which is bordering β0, close to G42, and similarly remote. As these sites must be in close contact to create NMR signals, it is likely that G42 rearranges with β0.

In addition to these new interactions located close to the N terminus, C-terminal contact sites are also pinpointed by the NMR data. K65 shows cross-peaks to L236 and I238. The latter two do not have a cross-peak with each other, indicating that they are not located in a loop region as would be expected from the X-ray model. Together with the fact that K65 and L236 / I238 are remote ($\sim$ 33 Å, Fig. 4.5C), it is apparent that these residues must rearrange to accommodate these contacts in the filament. The significance of the C terminus is underpinned by S58 showing an H^{N}-H^{N} contact with V249. V249 was truncated in the X-ray construct to facilitate crystal formation but seems to play a crucial role in filament formation. Its distance to K65 in the X-ray model cannot be estimated, due to its absence from the structure, but it remains evident that K65 locates far from the C-terminal end (34 Å, Fig. 4.5D).

Taken together, the NMR contacts show that rearrangements in the N- and C-terminal regions happen during filament formation. The next step is to examine whether the nature

of the observed interactions is intra- or inter-molecular.

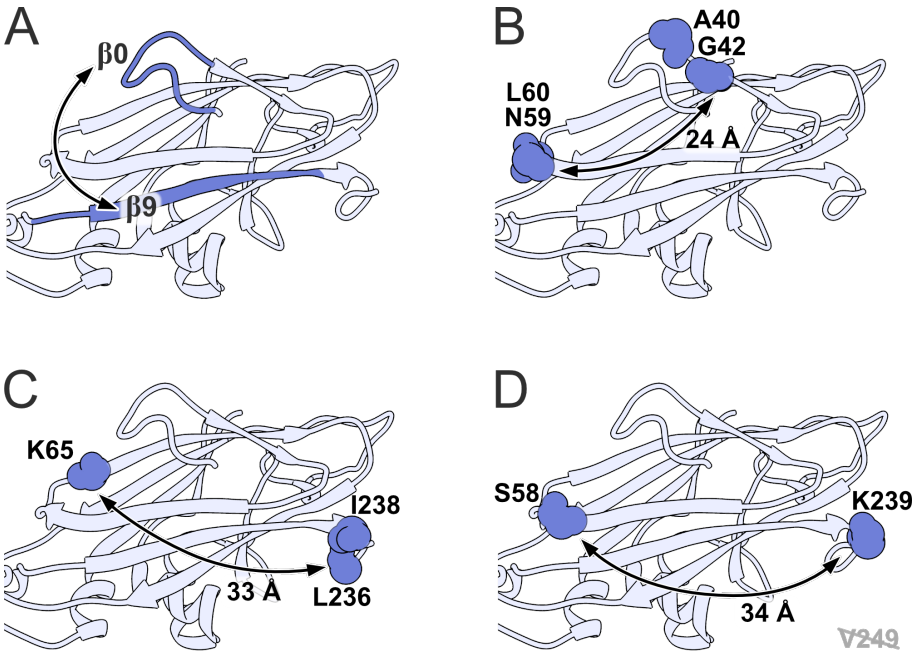

Fig. 4.5.: NMR contacts not consistent with the structure of monomeric TasA. Residues located in close proximity according to H^N-H^N contacts detected by NMR are highlighted on the X-ray structure of monomeric TasA (PDB 5OF1). These include (**A**) the β-strands β9 (230-220) and β0 (31-41), (**B**) A40 / G42 and N59 / L60, (**C**) K65 and L236 / I238, and (**D**) S58 and V249. V249 is not present in the crystallized construct, the C-terminal end K239 is indicated instead.

Validation of the inter-molecular nature of interaction

In order to determine the nature of the interaction, filaments with mixed labeling were prepared. TasA monomers were expressed either with uniform ^{2}H,^{15}N or [2-^{13}C]-glycerol labeling patterns and subsequently mixed in a 50:50 ratio, using a buffer containing 100 % H_2O prior to filament formation. This procedure leads to assemblies of proteins with different NMR-active isotopes, which can be exploited to selectively transfer magnetization between contact interfaces (Fig. 4.6A). In addition to the complementary ^{13}C and ^{15}N labeling, sidechain perdeuteration in the ^{2}H,^{15}N sample avoids the excitation of natural abundance ^{13}C atoms by ^{1}H-^{13}C CP. ^{1}H are only present at the exchangeable sites, which decreases relaxation of H^N nuclei due to the disruption of dense proton networks. Taken together, this labeling pattern predominantly allows the detection of inter-molecular contacts.

The inter-molecular transfer was achieved by running an hchhNH pulse sequence to record a ^{1}H-^{15}N spectrum (details in section 3.1.5 and Fig. 3.1C on page 42). In the chosen pulse sequence, initial ^{1}H polarization is produced by a 90° pulse and subsequently

transferred to ^{13}C nuclei and back using two CP blocks. This section functions as a filter and removes any signals from non-^{13}C labeled material. After this filter, the magnetization resides at ^{1}H and is subsequently transferred through space to adjacent ^{1}H nuclei using a 3 ms RFDR pulse train. Finally, a second CP block is applied, this time from ^{1}H to ^{15}N and back, to filter out everything lacking ^{15}N isotopes. At the end of the pulse sequence, only magnetization that was transferred through space, via ^{13}C and ^{15}N on different molecules, persists. A schematic of the transfer pathway is shown in Fig. 4.6A by black arrows.

The resulting spectrum represents a subset of signals for residues located at inter-molecular interfaces (Fig. 4.6B). The strongest peaks of the selective spectrum do not occur at the same chemical shift values as the most intense ones of an unfiltered spectrum. This shows that previously intense peaks do not contribute to the interface regions. A large number of the detected inter-molecular signals have H^N chemical shifts typical for hydrogen bonded protons, which is expected for those involved in a β-sheet conformation ($\delta(^1\mathrm{H}) > 8.5$ ppm). The peak positions of residues discussed in the previous section are highlighted in Fig. 4.6C. Labels for signals of β0 and β9 are colored in orange (hydrogen bonded) and blue (H^N oriented opposite). Amino acids that are neither part of β0 nor β9 but form other contacts are marked black.

The coverage of β0/β9 signals (labeled in orange and blue in Fig. 4.6B according to their orientation) is quite remarkable and shows that these two strands build a long and consecutive interaction. Due to the complementary labeling and the presence of the peaks of these two β-strands, the interaction between them has to be inter-molecular. Almost all of the expected signals are detected, among which I32, A230, S41 (although slightly shifted), and adjacent F29 (shown in black) are especially significant due to their distinct location in the spectrum. Of the peaks arising from other contact interfaces, I238 and V249 are noteworthy as these are in a separated region of the spectrum as well and can be unambiguously assigned. In contrast, G42 and L236 do not manifest a cross-peak despite being situated at the interface. This absence might be due to an insufficient signal-to-noise ratio or ^{13}C labels being too remote for transfer.

In conclusion, some expected peaks were not detected but the overall abundant coverage of orange- and blue-labeled peaks indicates that strand β0 indeed aligns with β9 in an inter-molecular manner. As such, a model of filamentous TasA which accounts for this characteristic is needed.

4.1.5. Donor-strand complementation supports TasA filaments

Two structures representing filamentous TasA are currently available: a trimeric cryo-EM structure [56] published in parallel to the work presented here and a tetrameric one

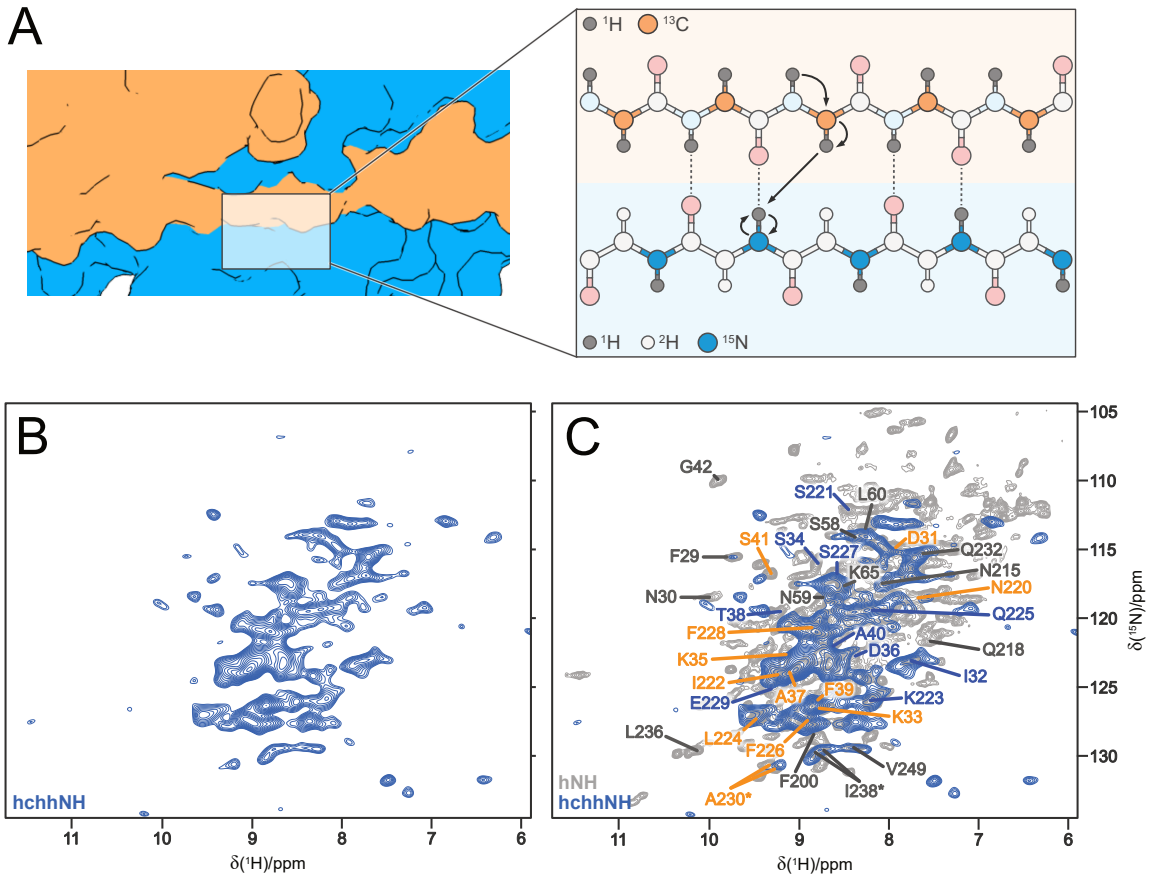

Fig. 4.6.: Measurement of exclusively inter-molecular contacts. (**A**) Schematic representation of an interface with a mixed labeling pattern ([2-^{13}C] versus ^{2}H, ^{15}N, mixed at a ratio of 50:50, and 100 % back-exchanged) at two interacting β-strands. An exemplary magnetization transfer pathway during the hchhNH experiment is indicated by arrows. (**B**) ^{1}H-^{15}N plane of a hchhNH experiment acquired on a sample with the shown labeling scheme. (**C**) Superposition of an hNH spectrum of a uniformly labeled (^{2}H, ^{15}N, ^{13}C and 100 % back-exchanged) preparation, shown in gray, with **B**. Residues directly connecting β0 and β9 by hydrogen bonds are labeled in orange, those in between with their H^N oriented opposite in blue. Other amino acids that form inter-molecular contacts have black as font color.

generated for this thesis by AlphaFold-Multimer [132]. When comparing these structures, PyMOL determines their backbone RMSD as 0.9 Å, indicating very high conformity, which is also apparent from their superposition (Appendix Figure A.5 on page 113). Both models assemble filaments by a mechanism called 'donor-strand exchange' whose characteristics are detailed below. Aside from their common traits, they diverge merely in the region 117-126, which is neither observable in the solid–state NMR spectra nor in the X-ray structure of the monomer. Taking into account that the AlphaFold prediction has a high uncertainty in this region as well (quantified as pLDDT, which correlates with flexibility in MD simulations [181]), it seems plausible that the stretch is flexible and not confined to any precise conformation. With other differences between the cryo-EM and AlphaFold structure being only minor, it is the AlphaFold model that will be referred to for the remainder of this chapter.

Extracting a single TasA molecule from the AlphaFold filament model and color-coding it according to the TALOS+ prediction shows that the earlier observed discrepancies are resolved (Fig. 4.7A and B). The rearrangement during filament formation appears to occur exactly at the previously diverging regions around the old β1 and β2, which are in turn now crucial for the mode of association. The N-terminal rearrangement is a known characteristic of the 'donor-strand exchange' polymerization mechanism, which is also encountered in proteins from other bacteria [182, 183]. Each molecule extends its N terminus, which is then bound to the next, intercalating seamlessly into the β-sheet architecture (Fig. 4.7C). This method of assembly is stabilized predominantly by hydrophobic interactions in the protein core and a regular hydrogen bonding pattern of the anti-parallel β-sheet.

When examining the rearrangements in detail, it is confirmed that the residues 28-41 constitute the new sheet β0 in the AlphaFold model. Additionally, the β0 of the previous protein, $\beta0^{n-1}$, is now aligned with β9 of the next protein in the filament, as indicated by the H^N-H^N contacts observed by NMR. Hence, β0 is involved in an inter-molecular interaction, as expected from the data obtained by the mixed labeling scheme, which is emphasized here by the superscript n-1. After β0, a short connecting stretch (42-46) extends, followed by β1 (47-51) and β2 (56-58), their position in the primary structure being shifted by a few residues compared to the monomeric X-ray structure (Fig. 4.7A). This new architecture, particularly the coiled region between β1 and β2, nicely aligns with the NMR-predicted dihedral angles. As a further validation measure, the H^N contacts that are incompatible with the X-ray structure are well in line with the filament model, as will be discussed in the next section.

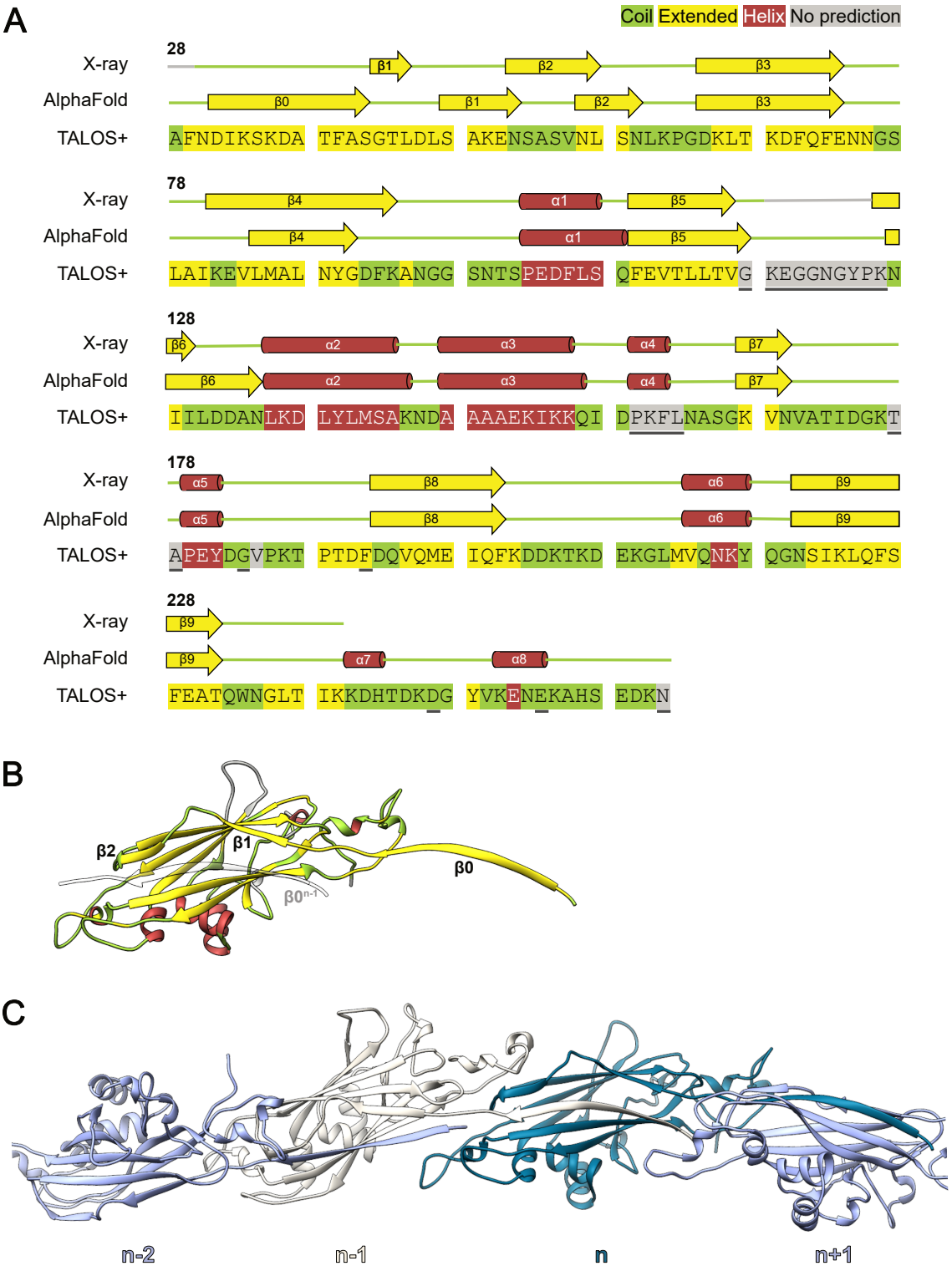

Fig. 4.7.: Rearrangement of TasA in filaments. (**A**) Secondary structure obtained from the X-ray study of monomers (PDB 5OF1), the AlphaFold model of filaments and the TALOS+ prediction based on filament chemical shifts. Sheets are colored in yellow, helices in red, coiled regions in green and not predicted (NMR) or missing regions (X-ray) in gray. Residues without NMR assignments are underlined. (**B**) TALOS+ output color coded onto one molecule of the AlphaFold filament. β1, β2 and β0, which undergo major rearrangement, are labeled. β0 of the previous filament subunit (n-1) is outlined as well. (**C**) TasA filament consisting of four molecules as predicted by AlphaFold-Multimer.

The filament model reflects H^N contacts measured by NMR

The AlphaFold prediction of the TasA filament places all previously characterized (section 4.1.4 on page 64) H^N contacts in proximity (Fig. 4.8A). The longest, continuously interacting stretch with the most contacts is the donor-strand $\beta0^{n-1}$ (residues 31-41) running anti-parallel to $\beta9^n$ (230-220), as shown enlarged in the top of Fig. 4.8B. A total number of 11 backbone hydrogen bonds arise (gray lines), leading to six H^N-H^N contacts indicated by arrows. Indeed, all 12 contributing residues show symmetric cross-peaks in the hNHH spectrum (Fig. 4.8B, bottom). Interestingly, the interaction between S41 and N220 is validated by a small, symmetric cross-peak in accordance with N220 being tilted away (H^N-H^N of 4 Å) and not forming a hydrogen bond. The overall conformation of the sheets is slightly twisted and locates the hydrophobic portions of sidechains in $\beta0^{n-1}$ (K33, K35, A37, and F39) into the core of the succeeding molecule. As such, the interplay of $\beta0^{n-1}$ and $\beta9^n$ forms favorable hydrogen bond interactions and helps hydrophobic complementation.

The backbone atoms of K65 (β3) form a short parallel stretch by binding C' of L236 and H^N of I238 (Fig. 4.8C, top). The orientation of the I238 sidechain aligns with the sidechains of L57 and L60 of the succeeding filament subunit. Hence, these sidechains form an inter-molecular hydrophobic cluster that stabilizes the interaction. The observed backbone distances match the NMR data in multiple respects. K65 displays symmetric cross-peaks with L236 and I238 (Fig. 4.8C, bottom) of which the one I238 contributes to are slightly stronger due to its H^N being in closer proximity. As mentioned before, no direct cross-peak between L236 and I238 is detectable, which is in accordance with their two H^N being distant (> 5.5 Å) in the filament model. An additional parallel-running interaction is formed by the backbone atoms of S58 (β2) and G247 (C') / V249 (H^N) (Fig. 4.8D). Again, this interaction is complemented by hydrophobic inter-molecular contacts between the V249 sidechain and those of L57 / L60 belonging to the succeeding molecule.

The last set of contacts observed in the NMR spectra (G42 / A40 with L60 / N59) arises at the end of β0. The H^N of G42 shows a unique interaction pattern with N59 and L60 (Fig. 4.8E). As visible by the peak pattern in the NMR data, N59 and L60 manifest a kink in the filament structure whose dihedral restrictions bring their H^N very close (∼2.4 Å), resulting in a strong cross-peak between them. The H^N of G42 is relatively close to both (G42 / N59: ∼4.1 Å, G42 / L60: ∼3.2 Å) in line with the observed cross-peak pattern. In addition, intensities match the distances with N59 being further away and therefore having a slightly weaker signal. Furthermore, N59 has a cross-peak with A40 which is also in line with the AlphaFold model (A40 / N59: ∼3.7 Å). This signal is only distinguishable in the hNhhNH spectrum due to both nuclei having similar H^N chemical shift and a comparably weak intensity.

The predicted filament structure places some additional H^N in close proximity, that are

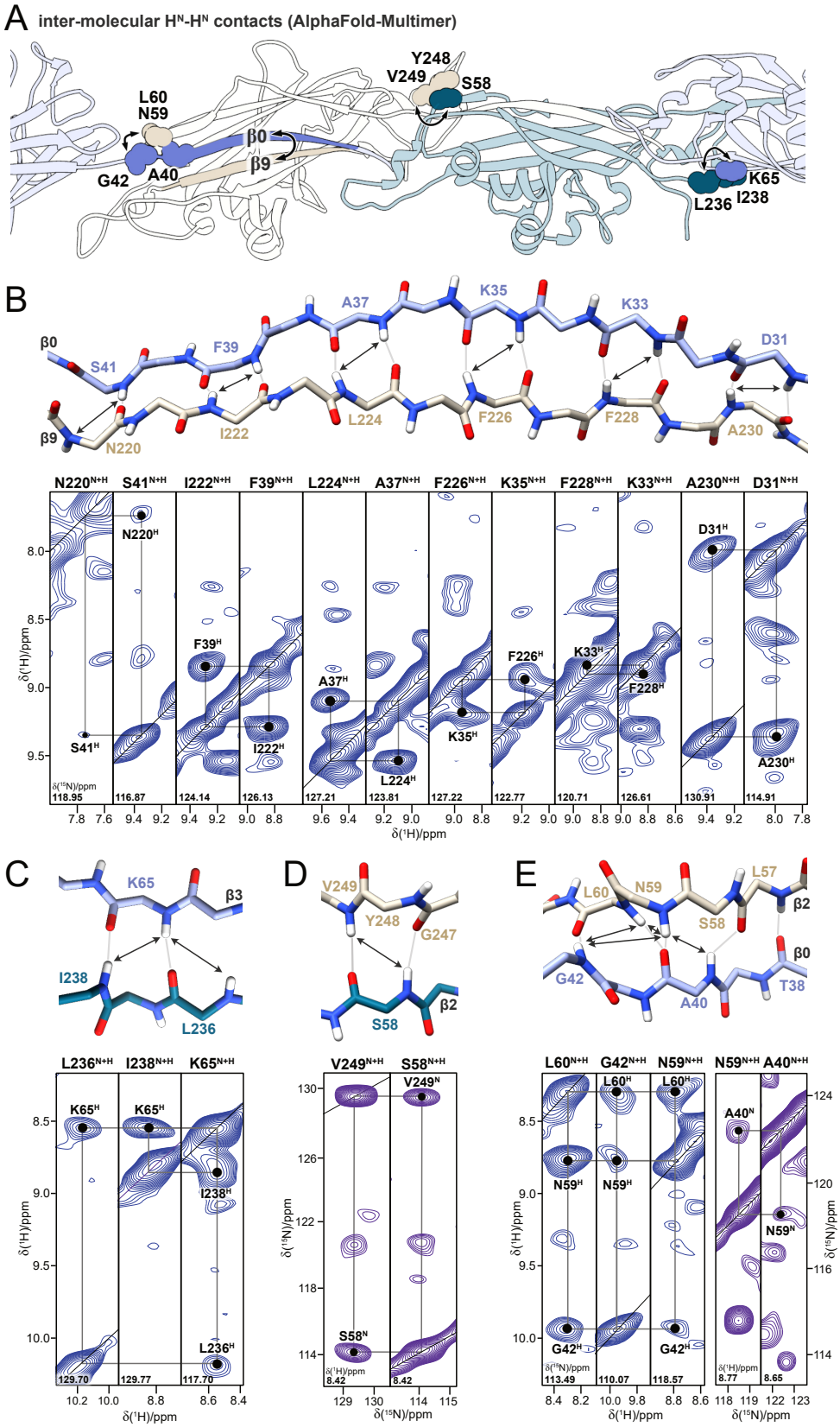

Fig. 4.8.: H^N-H^N contacts in TasA filaments. (**A**) Filament model predicted by AlphaFold-Multimer. (**B**)-(**E**) Detailed examination of H^N-H^N contacts at the interface. Hydrogen bonds are denoted by gray bars, observable interactions by arrows. Contact validation via NMR was performed by either an hNHH (blue) or an hNhhNH spectrum (violet).

not reflected in the NMR data. For instance, hydrogen bonds between L57 and A40 / T38 are predicted (Fig. 4.8E, top right), which should lead to an observable H^N contact between S58 and A40 (∼4.3 Å). However, this one is absent in the NMR data, which therefore deviates from the model in this region. The absence of these NMR signals could be explained by a high backbone flexibility or by the contact site being more distant than predicted.

In addition to these contact sites, the filament model has a more extensive interface where three TasA molecules meet. This region, which we term the ternary interface, only forms in the filamentous state and can be cross-validated with the NMR data as well. As the underlying hydrophobic interactions do not form backbone H^N contacts, we employed ^{13}C-detection for their analysis.

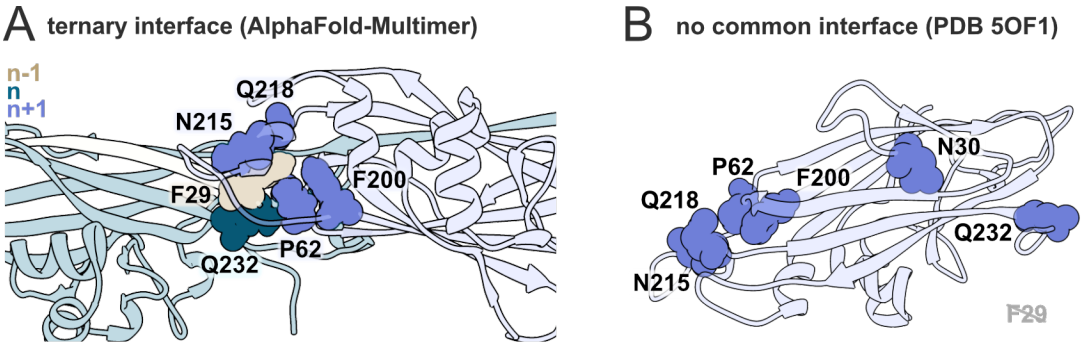

Fig. 4.9.: Three TasA molecules form an interface in the filament model. (**A**) The backbone and sidechain atoms of F29, Q232, P62, F200, N215, and Q218, stemming from three consecutive TasA molecules, form an interface in the filament model. (**B**) These residues are mostly remote in the monomer (PDB 5OF1), with F29 not being part of the structure.

The ternary interface is traceable by hydrophobic contacts

The detection of ^{13}C nuclei has since long been established in solid–state NMR as a valuable method for mapping signals to the polypeptide chain and can complement data obtained with ^{1}H-detected experiments [111, 184]. Here, carbon detection was used to extend the assignment into the sidechain. ^{13}C-^{13}C correlation spectra were recorded on TasA with different mixing times on uniformly ^{13}C-labeled material as well as protein expressed by supplying exclusively [1,3-^{13}C]- or [2-^{13}C]-glycerol to the medium. The latter, glycerol-labeled samples, have a sparse distribution of ^{13}C isotopes and provide less crowded spectra. For instance, the [2-^{13}C]-glycerol pattern is characterized by isotopes being abundant at C_α sites but less prominent on C_β and C' (see Appendix Figure A.1 on page 109). Together with a reduced dipolar coupling network (less NMR active nuclei overall), this leads to more distant contacts being observable in the spectra.

The assignments obtained from ^{1}H-detected experiments on perdeuterated and back-exchanged protein samples were first corrected for deuterium-induced changes of the

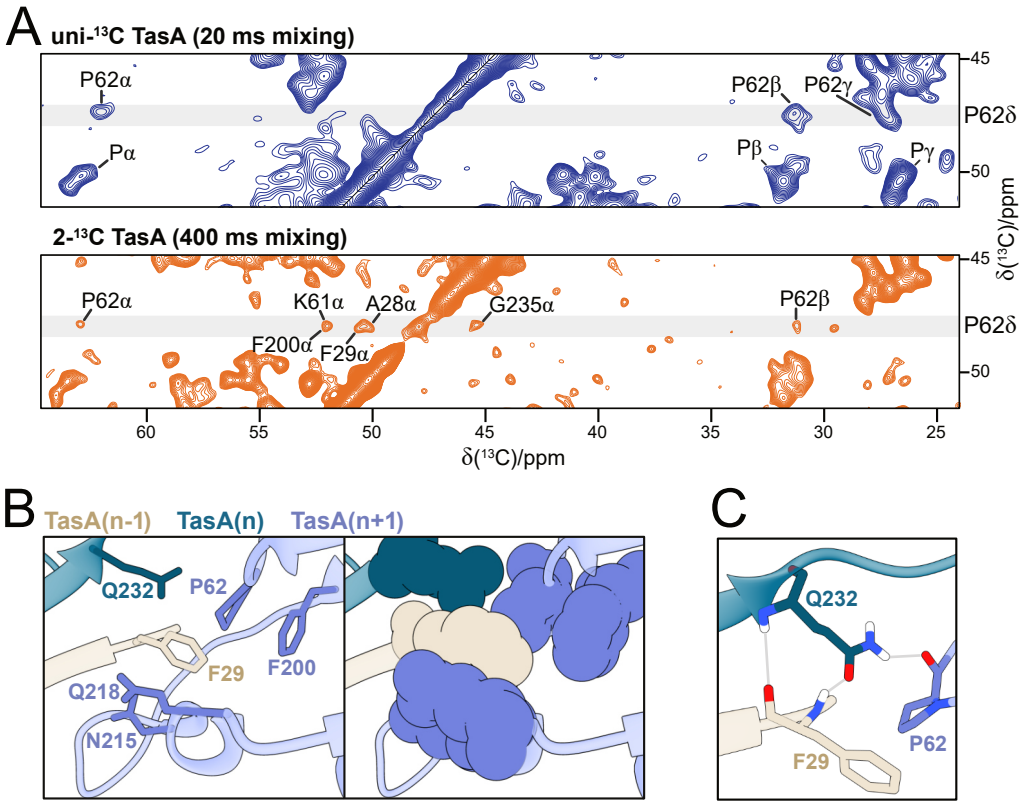

Fig. 4.10.: ^{13}C-detected spectra validate a ternary interface. (**A**) ^{13}C-^{13}C correlations of two samples with labeling schemes and DARR mixing times as indicated. (**B**) Detailed view of the ternary interface with residues either as sticks (left) or spheres (right). Hydrogens are omitted for clarity. (**C**) Q232 stabilizes the site with three hydrogen bonds to F29 and P62.

chemical shift [158]. With the help of the ^{13}C-detected data, the assignment was subsequently extended to the sidechain. Combined with the information encoded in the [1,3-^{13}C]- or [2-^{13}C]-labeling schemes, this made resonance assignment straightforward for serines, threonines, the C_α and C_β of valines, and isoleucines (Appendix Fig. A.6 on page 114). Additionally, P62, which has signals separated from the other prolines (Fig. 4.10A, top), was readily assignable. The unique shift pattern of P62 arises from its characteristic chemical environment, created by two flanking aromatic rings in the filamentous state (Fig. 4.10B, left). P62^{n+1}, together with F200^{n+1}, and F29^{n-1}, forms a so-called 'proline box', a configuration that is known from studies of domains that bind proline-rich peptides [185]. Q232^{n} complements the interface by packing tightly (Fig. 4.10B, right) and being involved in three hydrogen bonds, with the carbonyls of P62^{n+1} and F29^{n-1}, as well as the amide H^N of F29^{n-1} (Fig. 4.10C).

This configuration is corroborated by cross-peaks observed in ^{13}C-^{13}C spectra of [2-^{13}C]-TasA at a long mixing time of 400 ms, as shown in Fig. 4.10A on the bottom. The distinct NMR frequency of P62-C_δ allows the assignment of intra-residue correlations with

C_α and C_β as well as signals with A28-C_α, F29-C_α, F200-C_α, and G235-C_α in line with the labeling pattern and the filament model portrayed above. Taken together, this gives further credit to the available structures of TasA filaments. Strikingly, amino acids at the interface formed by three consecutive filament molecules are conserved in TasA homologues in other bacteria, even when the overall sequence conservation is low (Appendix Fig. A.7 on page 115). This observation and its implications are discussed further in section 5.1.2 on page 98.

4.1.6. Conclusion

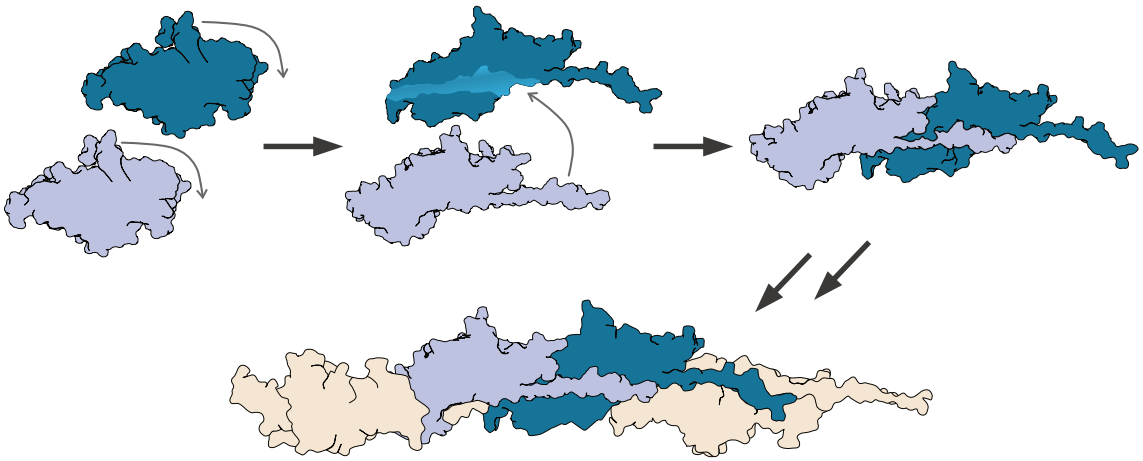

Fig. 4.11.: Schematic representation of TasA filament assembly. In order to transition to a filamentous state, TasA molecules undergo at least two major conformational rearrangements. First, the N-terminal section extends which exposes up an adjacent hydrophobic cleft. Second, two of these molecules need to assemble by donor-strand exchange, where the N-terminal strand of one molecule closes the hydrophobic cleft of the second. These two steps can occur repeatedly to assemble long filaments.

This work aimed to characterize the structure of filaments formed by TasA at the atomic level. Through supplementation of TasA-deficient biofilms by labeled protein, it could be shown that TasA readily forms identical assemblies *in vitro* as it does in its native environment. The excellent resolution achieved in ^{1}H-detected experiments of ^{2}H, ^{13}C, ^{15}N-labeled material and ^{13}C-detected experiments of [2-^{13}C]-labeled filaments allowed for an almost complete backbone assignment exposing additional signal sets for certain residues. By interpretation of the data, supplemented with through-space restraints obtained from mixed samples with complementary labeling, it was determined that TasA filaments employ a conserved linkage pattern called ‘donor-strand complementation’ (Fig. 4.11). This interaction mechanism was previously shown to be present in many bacteria and archaea.

4.2. Toxicity mechanism of TccC3 from *Photorhabdus luminescens*

Tc toxins from different bacteria have been the focus of intense study due to their potential in pest control [59] and as a specific translocation system [12]. Additionally, their homology to proteins pathogenic for humans [186] makes them an important model system. The research presented here focuses on TccC3 expressed by *Photorhabdus luminescens*, which is known to be lethal for insects.

4.2.1. The functional TcART domain is located C-terminal

The toxic moiety of TccC3 resides in the encapsulated C-terminal region [72] termed hypervariable (TcHVR) [64]. After entering the host cell, it ribosylates F-actin at position T148 [71] which causes cell death. Despite this knowledge of how the toxin functions, the size of the functional domain and its overall nature (whether folded or disordered) remained elusive [70, 72, 75, 84]. Using solution NMR, it was possible to determine the minimal domain necessary for modification of F-actin. It acts as an ADP-ribosyltransferase (ART) and we consequently termed it TcART.

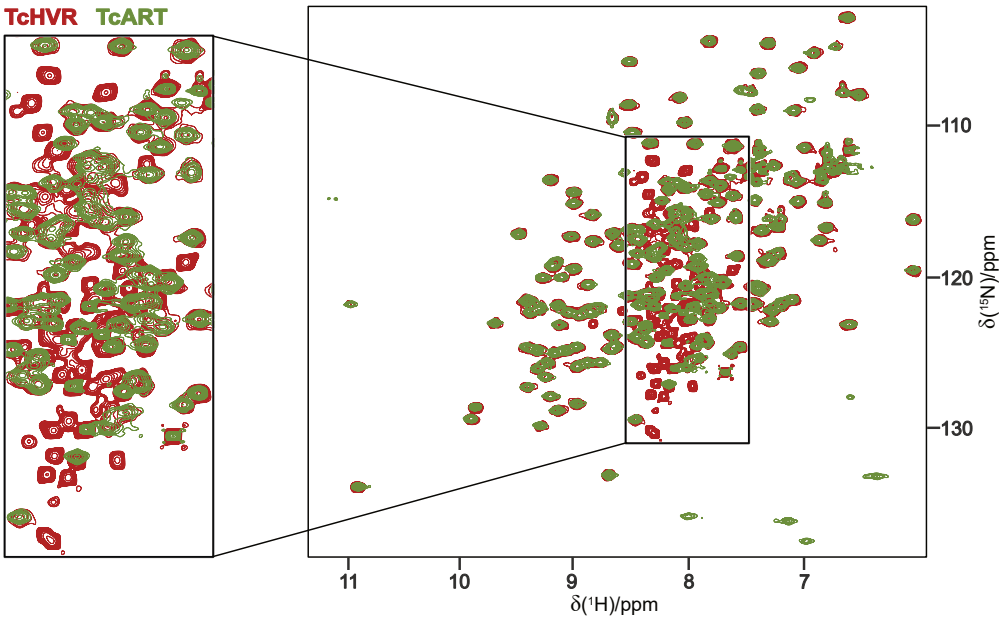

Fig. 4.12.: The N terminus of TcHVR is disordered in solution. Superimposed ^{1}H-^{15}N spectra of the complete hypervariable region of TccC3 (TcHVR, red) and a truncated version lacking 97 N-terminal residues (TcART, green). The central part of the spectrum, characteristic for disordered regions, is shown enlarged on the left.

TccC3 (UniProt Q8GF97) consists of 960 amino acids with the region 679-960 belonging to TcHVR [70]. Jpred4 [187] predicts the first ∼100 residues of TcHVR (679-779) to

be predominantly disordered, which was validated through the ^{1}H-^{15}N spectra depicted in Fig. 4.12. TcART (shown in green), a truncated construct starting at residue 98 of TcHVR (residue 776 of TccC3), comprises the entire folded domain as evident from the perfect overlap observed in spectral regions characteristic for sheets and helices. Except for one peak with a ^{1}H chemical shift of 8.9 ppm, all additional signals observed for TcHVR (red) are located in the ^{1}H range between 7.5 to 8.5 ppm, which is typical for disordered parts of a protein. Deletion of the first 100 amino acids of TcHVR did not affect ligand affinity as confirmed by ITC (Appendix Fig. A.8 on page 116) or enzymatic activity, measured as toxicity in yeast [2]. The main function of the disordered region is likely to act as a linker to the complex and a vehicle to fulfill charge and size requirements needed for proper injection by the complete Tc toxin machinery [12]. Hence, all functional studies were conducted with the shorter TcART construct instead of full-length TcHVR.

4.2.2. The hydrophobic core is rich in β-sheets

The TcART construct (including a C-terminal tag) used for the NMR evaluation had a total length of 194 amino acids and a molecular weight of 21.7 kDa. The size of TcART is comparably large, and in order to achieve an (almost) complete NMR assignment, perdeuteration of sidechains was required for recording HNCACB, HNCOCACB, and HNCO spectra. With this dataset, more than 90 % of backbone atoms could be assigned. The five N-terminal residues were not observable, likely due to unfavorable dynamics. Further unassigned amino acids are scattered throughout the sequence and do not form an extended gap. A few carbonyl signals were not unambiguously assignable due to overlap in the HNCO data and for a small number of residues no signal was observed. It is noteworthy that backbone signals of V190 (^{1}H: 5.2 ppm,^{15}N: 110 ppm) and Nδ276 (^{1}H: 5.3 ppm,^{15}N: 104 ppm) display extreme H^{N} shifts close to the water line. The visibility of their signals strongly depended on sample as well as buffer conditions and was often reduced by the water suppression. With the assignment, information at atomic resolution could be extracted from the data.

In order to probe for sites that do not exchange with the bulk water, TcART was dialyzed to a buffer containing 99.9 % D_2O for a total time of 36 h. Immediately after this, NMR measurements were done to monitor the result of continuous H-D exchange over a period of two days. Strikingly, the ^{1}H-^{15}N correlation showed 28 signals that persisted. These were predominantly located in the spectral region characteristic for β-sheets (> 9 ppm, Fig. 4.13A-C). Visualizing them on the primary structure (Fig. 4.13D) implies that not only both termini contribute to the hydrophobic core, but also that some β-strands do not form tight hydrogen bonds on one side. This is apparent from the alternating pattern as observed for the stretch 265-269 (**E** I **S** F **L**) where I266 and F268, who face opposite from

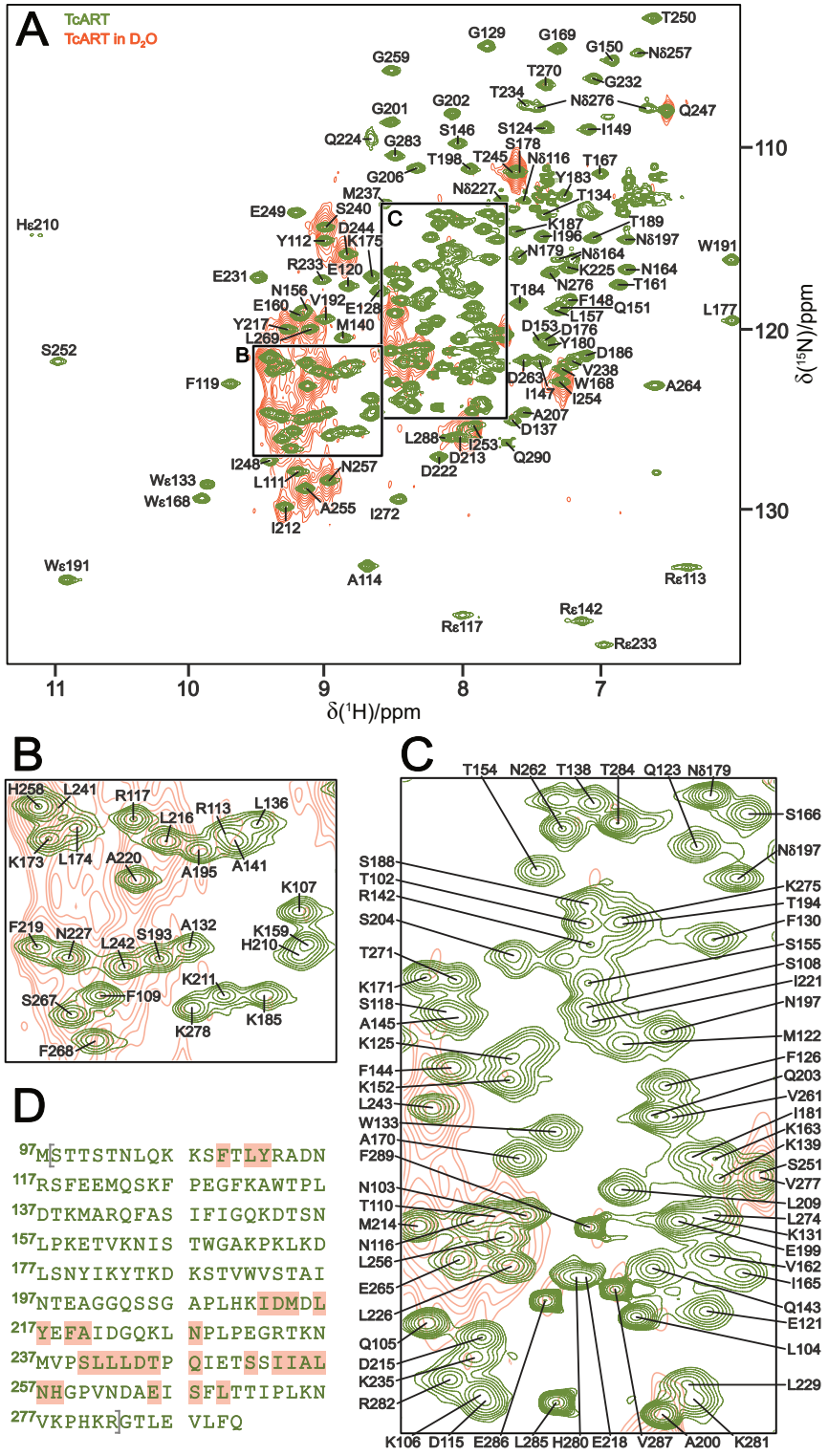

Fig. 4.13.: H-D exchange highlights strongly bonded hydrogens. (**A**) Spectra of TcART in aqueous (green) and D_2O (orange) buffer after 2 days of dialysis. Boxed sections are enlarged in (**B**) and (**C**). Backbone assignments are indicated, a number of side chain labels are omitted for clarity. Signals of V190 (^{1}H: 5.2 ppm,^{15}N: 110 ppm) and Nδ276 (^{1}H: 5.3 ppm,^{15}N: 104 ppm) are outside the displayed range. (**D**) Sequence of the TcART construct with the section originating from TcHVR indicated (98-282, square brackets). Residues whose backbone ^{1}H is not in exchange with the bulk water are marked orange.

E265, S267, and L269, exchange with the bulk water. Using these insights as a basis, further NMR data was collected to calculate a protein structure.

4.2.3. The TcART structure bears homologies to other ARTs

To complement the backbone assignment, further experiments were performed on non-deuterated and ^{13}C,^{15}N-labeled TcART. The existing data could be easily transferred through acquisition of an HNCA spectrum and application of a correction factor for the isotope-induced chemical shift variations [158]. C_α-C_β pairs were then correlated to more remote atoms using TOCSY and NOESY spectra. This extension approach proved difficult for long, flexible side chains (*e.g.*, lysines) due to overlap in disordered spectral regions. Interestingly, the aromatic ring of Y183 displayed increased characteristics of disorder as visible from sharp lines in a TOCSY spectrum focused on the aromatic region. In total, 89 % of ^{1}H and 74 % of the heavy atoms of the sidechains could be assigned which enabled for a systematic analysis of through-space contacts.

The ^{15}N-filtered ^{1}H-^{1}H NOESY spectrum, or ^{15}N NOESY in short, is similar to the hNHH used for TasA in solid–state NMR. It allowed for a detailed examination of the tightly bound β-sheets whose hydrogen bonds do not exchange with the bulk water. As detailed in the TasA section on page 63, their anti-parallel nature is apparent from the combination of weak sequential and strong cross-strand intensities in the through-space ^{1}H-^{1}H pattern (Appendix Fig. A.9A on page 117). In total, 37 hydrogen bonds that restrain the rigid hydrophobic core could be manually derived. Furthermore, ^{13}C NOESY experiments recorded in H_2O and D_2O helped to pinpoint side chain interactions. Prominent examples are the tip of the sidechain of K187 ($H_{\gamma a}$: 0.0 ppm) that packs against the W133 indole ring, or I181 (H_α: 2.9 ppm, $H_{\gamma 1a}$: -2.4 ppm, $H_{\gamma 2}$: 0.1 ppm, H_δ: -0.5 ppm) and I165 (H_α: 1.4 ppm) that have anomalous shifts due to their interaction with the aromatic sidechains of W191/H258 and W168, respectively. Strip plots of these interactions are presented in Appendix Fig. A.9B-D. In order to translate this rather scattered information into a complete fold, ARIA was used for automated NOE assignment and NMR structure calculation.

The iterative scheme employed by ARIA was complemented by a consensus approach (details in Materials and Methods 3.2.3 on page 49) to prevent convergence to an artefactual state. In total, 1,833 chemical shifts and 12,153 peaks were used as input whereas no association between them was transferred. Hence, the iterative process of assignment and disambiguation was performed entirely unbiased. In addition, hydrogen bonds obtained from the NOESY pattern and dihedral angles determined by TALOS+ were supplied as constraints. During the calculation procedure, restraints arising from within one residue were discarded as they gave rise to a considerable number of violations and do not contain

information about the protein tertiary structure. A complete summary of statistics can be found in Appendix Table A.2 on page 125, the NMR ensemble is presented in Fig. 4.14A.

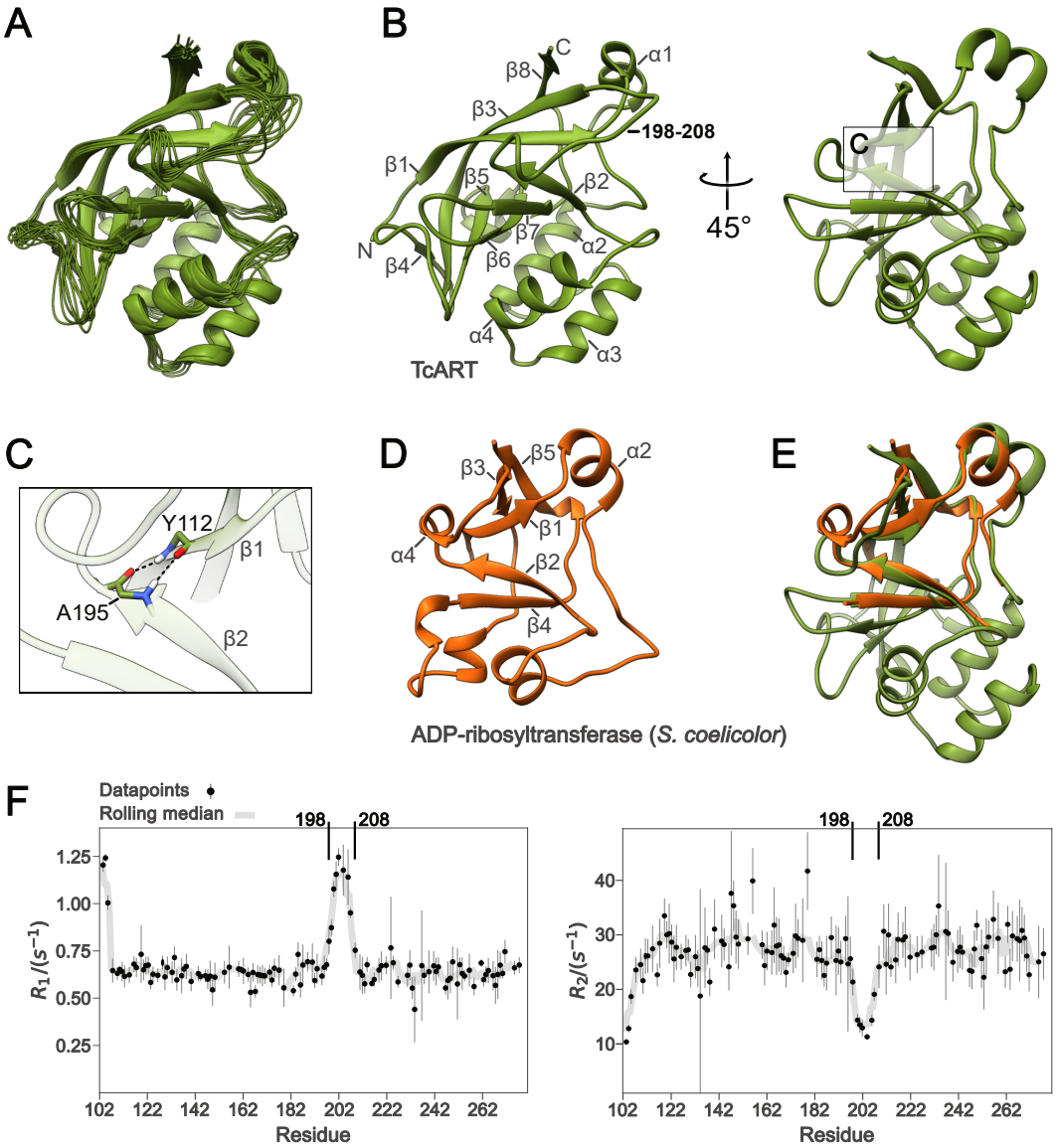

Fig. 4.14.: Characteristics of TcART. (**A**) Structure ensemble determined by NMR. The 10 lowest energy structures are shown, statistics are given in Appendix Tab. A.2 on page 125. Only the folded domain (102-282) is depicted. (**B**) Model closest to the average, shown from two different angles, with secondary structure designation and the flexible loop (198-208) indicated. (**C**) Closed up view of the boxed region indicated in **B**. Two hydrogen bonds between perpendicular β-sheets, characteristic for ribosyltransferases, are present in TcART. (**D**) Core domain of the ScARP ADP-ribosyltransferase from *S. coelicolor* determined by X-ray crystallography (PDB 5ZJ5). Selected secondary structure elements are labeled. (**E**) Superposition of the homologous sections of **D** with TcART. (**F**) R_1 and R_2 relaxation rates determined for TcART with the loop region 198-208 highlighted. Exponential fits used to derive the presented decay rates are shown in Appendix Figs. A.10 on page 118 and A.11 on page 119.

The TcART structure comprises eight anti-parallel β-strands organized in two perpen-

dicular β-sheets (Fig. 4.14B). These strands form a central hydrophobic core that is not accessible to water molecules, which is in line with the results from H-D exchange experiments (Fig. 4.13 on page 77). Interestingly, the perpendicularly oriented strands β1 and β2 are connected by two hydrogen bonds between Y112 (β1) and A195 (β2) as anchoring points (Fig. 4.14C). The overall organization of this fold, with the Y112-A195 connection in particular, is conserved among domains responsible for ADP-ribosylation [78] as shown for ScARP from *Streptomyces coelicolor* (PDB 5ZJ5) in Fig. 4.14D and E. Other homologous structures are, *e.g.*, the pertussis toxin from *Bordetella pertussis* (PDB 1PTO) and the CARDS toxin from *Mycoplasma pneumoniae* (PDB 4TLV).

Of the four α-helices in TcART, only α1 is observed at a canonical position. In place of a helix that usually follows β2 (α4 in *S. coelicolor*, Fig. 4.14D), the corresponding segment in TcART 198-208 manifests as a loop. The R_1 and R_2 rates obtained by NMR characterize this loop as flexible with similar decay rates as the N terminus (Fig. 4.14F). A detailed display of each exponential fit can be found in Appendix Figs. A.10 (p. 118, R_1) and A.11 (p. 119, R_2). Additional unique features of the TcART structure are an exposed triple-helix bundle, formed by α2-4, as well as elongated and strongly twisted strands β1 and β3. Together with β8, they form an extended arch-shaped β-sheet that reaches over the second, less twisted one formed by β2,5,6,7 , thus manifesting a roll-like structure.

The resolved molecular organization of TcART will be instrumental to better understand the ligand binding mode and mechanism of target modification in a next step.

4.2.4. TcART binds NAD^+ in a conserved pocket

The process of NAD^+ binding can be examined from the protein as well as the ligand side. Changes of TcART were monitored by quantifying chemical shift perturbations (CSPs) derived from ^{1}H-^{15}N HSQC spectra. For NAD^+ the interaction was traced by measuring its two ^{31}P nuclei whose NMR resonances offer a simple way of acquiring a sparse spectrum yielding direct information about their local chemical environment.

^{31}P spectra of NAD^+ indicate a complex binding mechanism

NAD^+ was dissolved in protein buffer at 10 mM and measured without TcART as a reference (Fig. 4.15, top). The spectrum shows two coupled doublets with a pronounced roof effect belonging to the respective ^{31}P nuclei of the P-O-P bond. This establishes that the observed splitting is indeed due to ^{31}P-^{31}P coupling as this deviation from the expected 1:1 intensity pattern only occurs when the coupling constant J is close to the difference in Lamor frequencies $\Delta\nu$ (strong coupling regime), which is not possible for a heteronuclear coupling in NAD^+. The observed signal pattern becomes increasingly difficult to interpret with the addition of TcART.

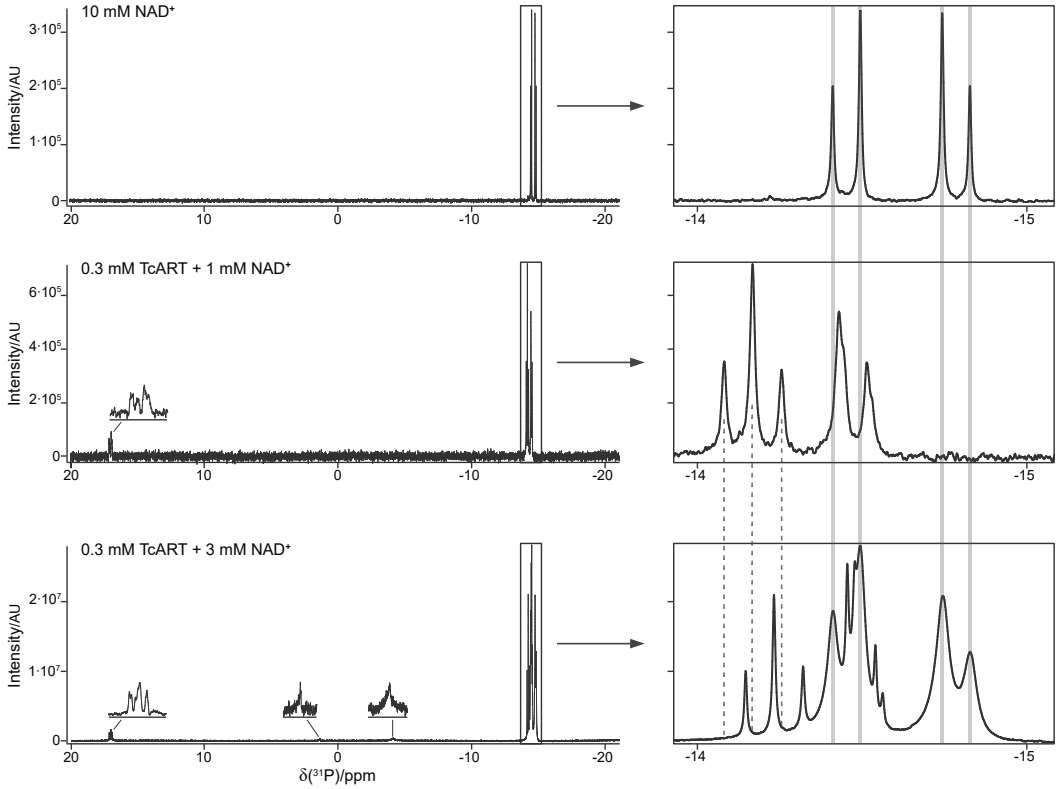

Fig. 4.15.: NAD$^+$ binding to TcART monitored by ^{31}P signals. ^{31}P spectra for free NAD$^+$ and two mixing ratios with TcART. The boxed region is shown enlarged on the right with peak positions belonging to unbound NAD$^+$ indicated by gray bars and those of a bound species by dotted lines. TcART alone does not yield any ^{31}P signals.

Variations of the protein-to-ligand ratio had a significant impact on the observed signals. For instance, all signals that correspond to unbound ligand are absent for a 0.3:1 mixture of TcART to NAD$^+$ (Fig. 4.15 middle, gray columns). This is quite striking considering the excess of NAD$^+$. Only when increasing its relative presence further to 10 times TcART (0.3:3, Fig. 4.15, bottom), these signals appear again, albeit with an increased linewidth. This speaks towards an exchange process with TcART as its slower tumbling could lead to the observed peak broadening. A number of additional peaks arise when NAD$^+$ is titrated to TcART. Minor ones are observed around 17, 1.4, and -4 ppm whereas at the lower ligand concentration, only the signal around 17 ppm is observable. The absence of the two signals at 1.4 and -4 ppm could in part be due to the worse signal-to-noise ratio preventing any definitive conclusion. Major additional peaks are present close to the chemical shift observed for the unbound species. At the 0.3:1 ratio, they consist of two broadened peaks around -14.5 ppm and a sharp, apparent triplet close to -14 ppm. The shoulders observed for the broadened signal could arise from further peaks that are buried beneath. At 0.3:3 ratio, the overall pattern gains complexity with the triplet shifting (dotted lines in Fig. 4.15) and the region around -14.5 ppm becoming difficult to interpret with a multitude of peaks arising. Among them are broad signals from the unbound species (gray columns, described

above) and additional sharp ones not present at 0.3:1.

With the high complexity of the obtained spectra, a detailed interaction mechanism cannot be derived. It appears that slow as well as fast exchange processes are present, as indicated by the observed peak movement characteristics (shifting vs. stationary but broadened). Nevertheless, an interaction between NAD^+ and TcART is evident, which was further quantified through the determination of chemical shift perturbations from the protein side.

^1H-^{15}N correlations indicate multiple regimes of dynamics at the binding site

^1H-^{15}N HSQC spectra of TcART were acquired to monitor ligand-induced chemical shift perturbations (shown in Fig. 4.16). Several signals respond to the presence of NAD^+, whereas different patterns of chemical shift changes arise. Some peaks change their location in a continuous manner (*e.g.*, Y112, G129) while others appear to jump when increasing the relative NAD^+ content from 0.3:1 to 0.3:3 (*e.g.*, G201, G202, and A114). A continuous change of the chemical shift is characteristic for residues that undergo fast exchange processes with a ligand, where the observed chemical shift is an intermediate between the one of bound and unbound state. Peaks gradually shift as the overall population moves towards 100 % binding site occupation. In contrast, a slow exchange mechanism is characterized by the detection of a signal that either reflects the bound or the unbound state. It manifests as a shift of intensity (not of position), where the peak intensity of the initial state decreases and that of the final state increases, possibly giving rise to an apparent 'jump'. A114 has a partial contribution of both these characteristics hinting at a potential intermediate exchange mechanism. This is further illustrated by the fact that cross-peaks between protein and ligand were not observed in NOESY-type spectra, presumably due to unfavorable exchange conditions. As these observations of having multiple dynamic regimes within one protein are hard to interpret without further context, additional insights can be provided by incorporating structural traits and knowledge derived from proteins with matching enzymatic activity.

The NAD^+ binding mode is homologous to other ribosyltransferases

We quantified the chemical shift changes upon addition of 1 mM NAD^+ and plotted them on the protein sequence (Fig. 4.17A). This showed that strong perturbations do not occur at a single site but rather in clusters distributed over the entire protein. The largest effects were observed for Y112, A114, R117, the stretches E128 to W133 and T184 to S188 as well as H210, K211, F268, and the region N276-K281. Additionally, the ^1H-^{15}N signals of Rε113 and Rε117 (Appendix Fig. A.12A and B on page 120) show major changes indicative of their sidechains interacting with the ligand. When adding 3 mM NAD^+, the

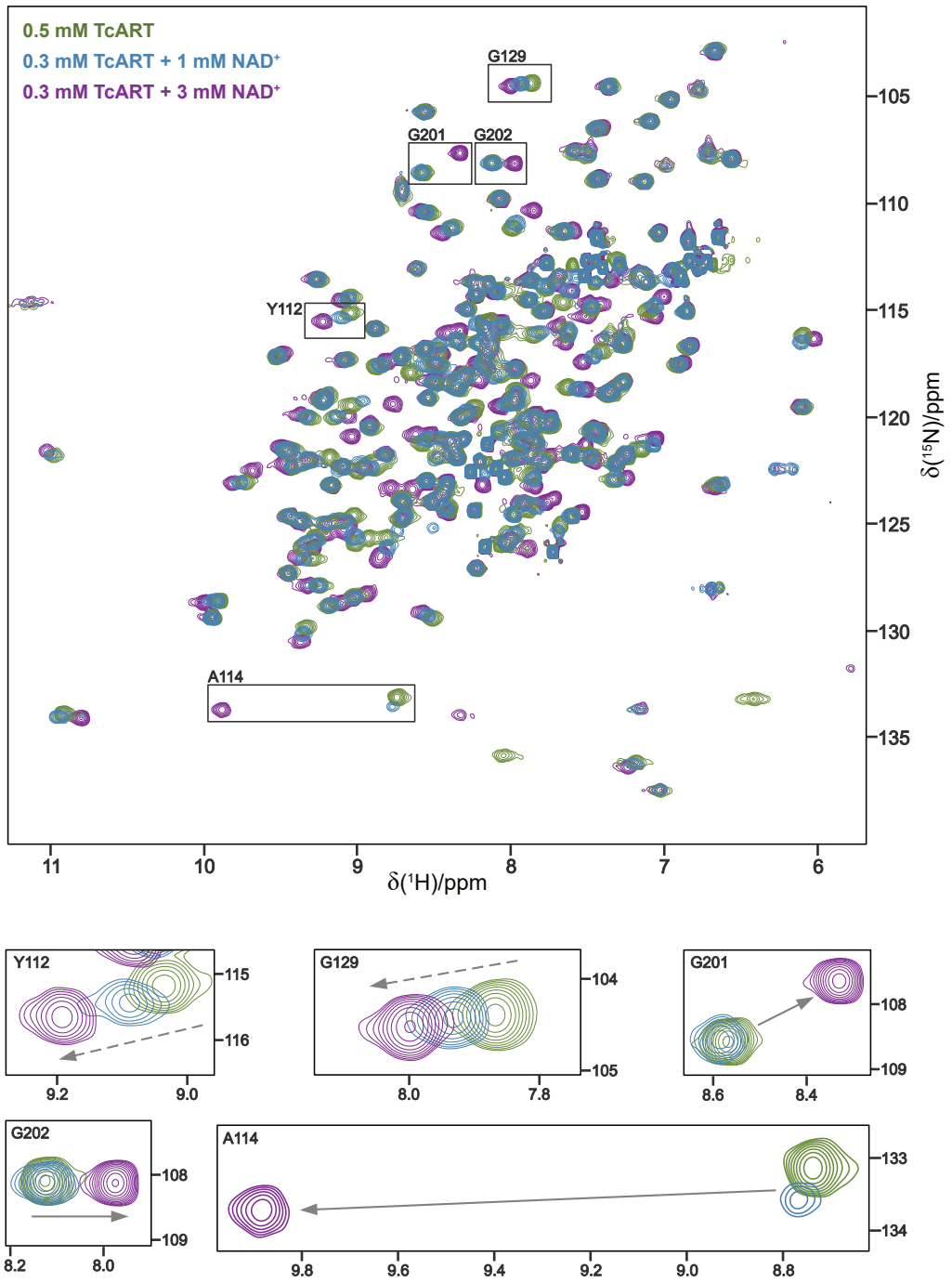

Fig. 4.16.: Changes in TcART ^{1}H-^{15}N spectra upon NAD^{+} addition. Observed ^{1}H-^{15}N correlations of TcART acquired without (green) as well as either 1 mM (blue) or 3 mM (violet) NAD^{+}. Boxed regions are shown magnified on the bottom, gradual peak movements are indicated with dashed arrows, 'jumps' with a solid arrows.

same regions are affected, although their quantitative change is overshadowed by that of A114 (Appendix Fig. A.12C, A114 is also magnified in Fig. 4.16). This observation, together with the contribution of the R113 sidechain binding to NAD^+, speaks towards a homologous coordination mode as in other ADP-ribosyltransferases. This can be derived from the comparison with other proteins.

The homologous enzyme from *S. coelicolor* (PDB 5ZJ5, Fig. 4.14D and E on page 79) is an example of a ribosyltransferase from the so-called 'R-S-E clade' [78]. Characteristic for these enzymes are an arginine (R), serine (S), and glutamate (E) as essential amino acids in the catalytic center (highlighted Fig. 4.17B). Mutation of any of these residues decreases or even fully inhibits catalytic activity [188]. Strikingly, TcART has a similar set of amino acids (R113 in β1, S193 in β2 and E265 at the beginning of β7) arranged in an identical manner (Fig. 4.17C). Significant chemical shift changes uopn NAD^+ addition were observed for the homologue arginine Nε and backbone proton of the adjacent A114. From a closer look at the binding pose of NAD^+ in the X-ray structure of 5ZJ5 (Fig. 4.17D), it is evident that the chemical shift perturbations determined for TcART do indeed cluster in proximity of a similar pocket (Fig. 4.17E). However, the predominant cluster manifests at the lower part of the cleft encompassing the loop T184 to K188. There, K185 interacts with E265, which potentially prevents binding. This suggests that the changes in the ^1H-^{15}N spectrum can be interpreted as an indicator of which protein regions rearrange in presence of the ligand. Hence, the shift changes reflect the residues involved in local induced fit rearrangements to facilitate the binding process.

This line of reasoning is further substantiated by the slow exchange characteristics observed for the flexible loop 198-208 and residue A114. They are situated in a region that binds the nicotinamide moiety of NAD^+ in the homologous structure of *S. coelicolor*. In this ADP-ribosyltransferase, the binding pocket is formed by conformationally similar sheets β1,2,3,5 and the helix α2 (Fig. 4.14D and E). However, a crucial difference arises at the side where helix α4 borders the binding region as its conformational constraints keep the pocket open. In the NMR structure of TcART, the loop 198-208 is situated at the identical position as α4 and prevents access to A114. Therefore, TcART is unable to properly bind NAD^+ in the presented state unless rearrangements occur. Evidence that binding nonetheless can take place is provided by the titration experiment with the higher ligand concentration (Fig. 4.16 and Appendix Fig. A.12). Similar to the formation of two β-strand-like hydrogen bonds between the backbone of S82 and the amide group of nicotinamide in PDB 5ZJ5 (Fig. 4.17F), such bonds could also form for TcART with the backbone of homologous A114. The formation of such two bonds would also explain the strong shift of the A114 signal towards a spectral region characteristic for β-sheet residues upon ligand addition (Fig 4.16). The backbone of A114 in TcART is situated in a very similar configuration as the one of S82 in the homologue (Fig. 4.17G) and could likely be

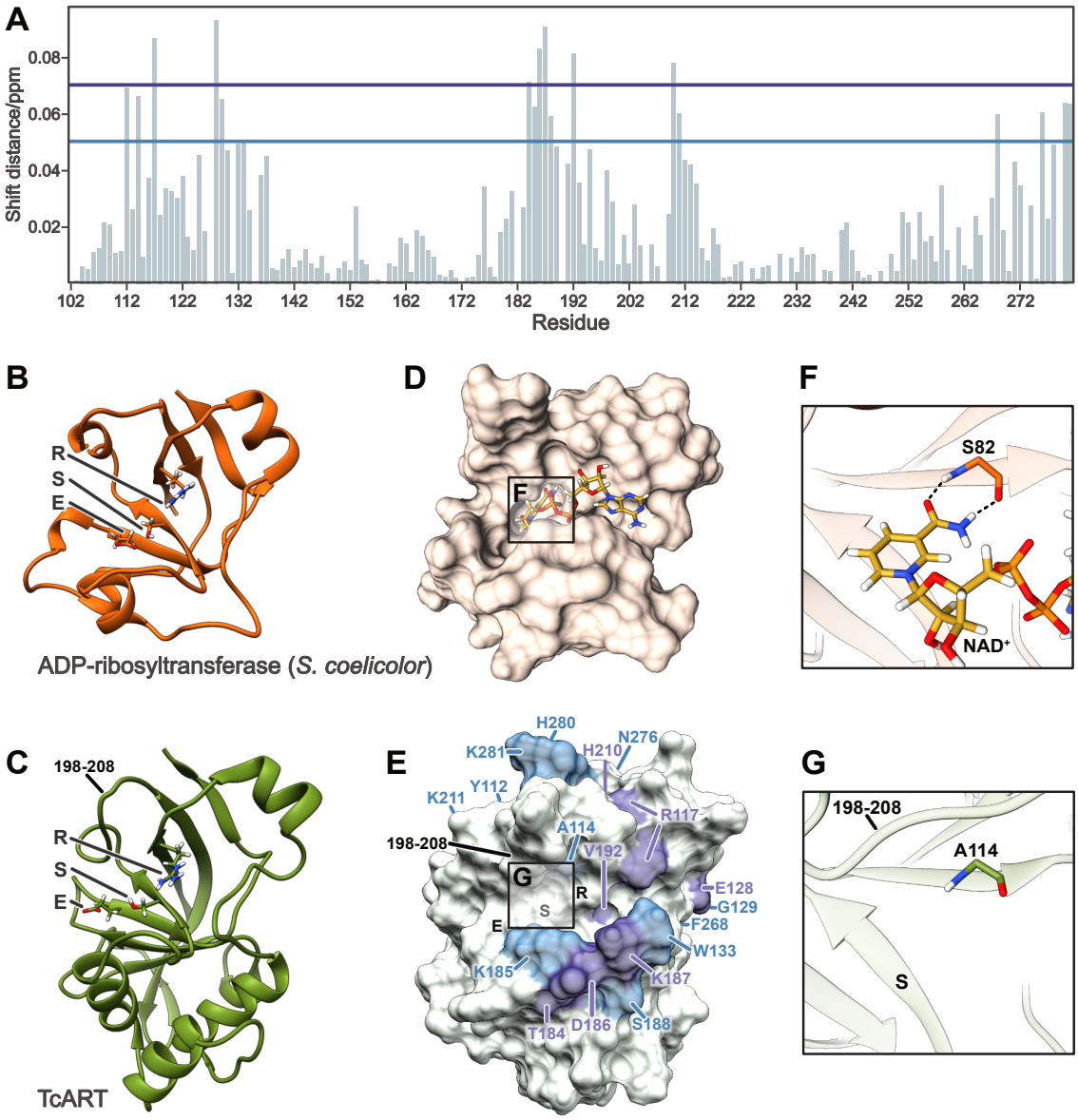

Fig. 4.17.: Chemical shift changes indicate a conserved pocket in TcART. (**A**) Chemical shift perturbations determined for the backbone of TcART (0.3 mM) upon addition of 1 mM NAD^+ (spectra in Fig 4.16). Cutoffs for medium (light blue) and strongly (dark lilac) shifting residues are indicated by horizontal lines. (**B**) ScARP ADP-ribosyltransferase from *S. coelicolor* (PDB 5ZJ5) with the catalytic sidechains R81, S121, and E164 shown. (**C**) TcART with homologous residues R113, S193, and E265. (**D**) Surface of **B** with co-crystallized NAD^+ inside the binding pocket. (**E**) Surface representation of TcART with medium and strongly shifting backbone signals indicated based on the values presented in **A**. The conserved R-S-E motif and the loop 198-208 are indicated. K211, Y112, N276, G129, and F268 are hidden from view. (**F**) The nicotinamide moiety of NAD^+ forms two β-sheet-like hydrogen bonds with the backbone of S82 in the *S. coelicolor* ADP-ribosyltransferase. (**G**) A114 is located at the homologous site in TcART as S82 shown in **F**. Residues 198-208 cover the site, making it inaccessible.

involved in a comparable interaction during binding. As such, in addition to induced fit rearrangements happening fast at the bottom side of the binding cleft (T184 to K188), a second set of slow adaptions seems necessary for complete binding. The loop 198-208 has to move from its position in the NMR structure, exposing the backbone of A114 where nicotinamide binds tightly.

In conclusion, the TcART NMR structure presented here is not competent for NAD^+ binding and additional ways for deriving a model of the TcART-NAD^+ complex need to be taken in consideration.

4.2.5. Snapshots show rearrangements during enzymatic activity

While this work focused on a model of the free TcART state present in solution, cryo-EM studies conducted by collaborators (at the MPI of Molecular Physiology, Dortmund) yielded a model bound of TcART to its target F-actin (PDB 7Z7H) [2].

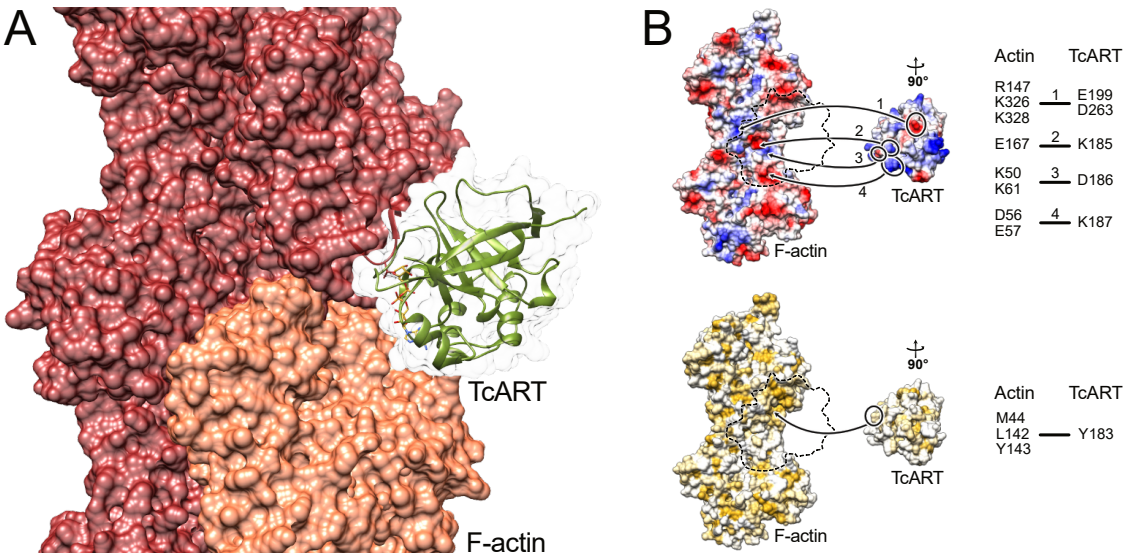

Fig. 4.18.: TcART binds F-actin mainly through electrostatic interactions. (**A**) TcART binds at the interface of two subunits of F-actin (PDB 7Z7H). (**B**) Detailed view of electrostatic (top) and hydrophobic (bottom) interactions stabilizing the complex. Adapted from [2].

TcART binds with two helices of the bundle α2-α4, a binding mode that is not present in ribosyltransferases acting on other proteins. The interface, consisting of two consecutive subunits on the actin side (Fig. 4.18A), is mostly stabilized by electrostatic interactions and salt bridges (Fig. 4.18B). In addition, the sidechain of Y183 of TcART inserts into a hydrophobic pocket and is essential for tight binding. This mechanism is in line with the finding that the Y183 aromatic ring demonstrates increased flexibility in solution NMR measurements, which likely facilitates fast and efficient binding to F-actin.

The cryo-EM structure not only allowed for a characterization of the overall mode of interaction but also provides insight into the ligand binding and reaction mechanisms. In

the cryo-EM maps, two additional electron densities are present in the nucleotide-binding pocket: one can be attributed to the modification product (ADP-ribose) bound to T148 of actin and the other one to nicotinamide in proximity to A114 of TcART (Appendix Fig. A.13A on page 121). While the adenosine moiety of the ADP-ribose interacts with the sidechains of R117 and W133 of TcART, the phosphates are coordinated by those of R113 and K185 of TcART (Appendix Fig. A.13B). The free nicotinamide is kept in place by hydrogen bonds with the backbone of A114 (Appendix Fig. A.13C). These findings are in agreement with the extraordinary chemical shift changes observed in the NMR titration studies (Appendix Fig. A.12 on page 120) and illustrate interactions stabilizing the complex.

In order to obtain a model of the 'pre-reaction' state where NAD^+ is intact, it was docked into the cryo-EM structure. As detailed in the previous section, the NMR structure was not viable for this approach because parts of the binding pocket were blocked by the loops 184-188 and 198-208. In the structure obtained by the docking approach, the overall coordination of NAD^+ is very similar to the one observed in cryo-EM, with the difference that E265 of TcART now forms a hydrogen bond with the ribose 2'-OH and a salt bridge with the N^+ of the nicotinamide. A comparison of all three different models allows for a description of the reaction in detail (Fig. 4.19).

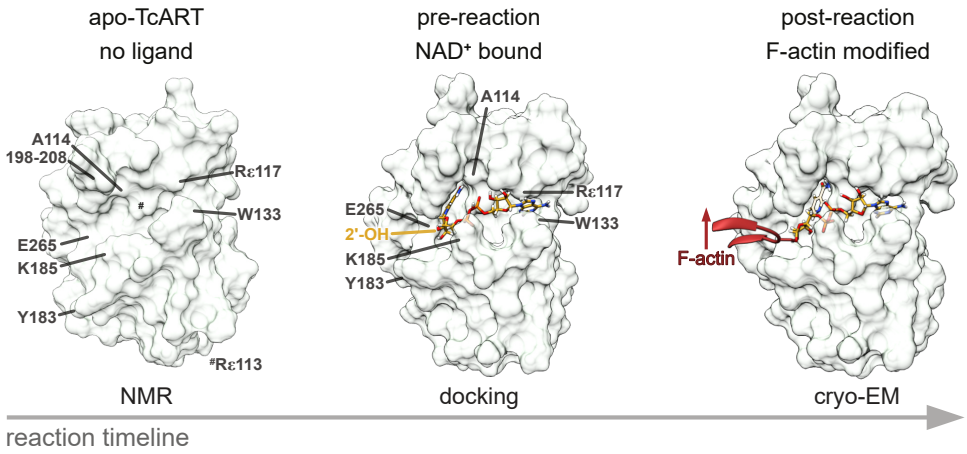

Fig. 4.19.: Timeline of TcART activity. Surface representations of three structures derived by different methods illustrate the conformational states of TcART during its enzymatic activity. Apo-TcART, as determined by NMR, has an empty pocket with the loop 198-208 and the E265 to K185 ionic interaction partially obfuscating the binding site. When NAD^+, shown as sticks, is bound (docking model), these sections move and enable efficient binding together with Rε117, W133, A114, and Rε113. The 2'-OH of one ribose is exposed making it accessible for a nucleophilic attack. After the reaction, as seen in the cryo-EM structure, the ADP-ribose is covalently linked to F-actin through the ribose 2'-OH. The leaving group nicotinamide rests in its prior position.

Apo-TcART, as present in solution, partially masks the binding site through movement of the loop 198-208 and a transient charged interaction between the K185 and E265

sidechains. After initial rearrangements and formation of the pre-reaction state, E265 and A114 fixate the nicotinamide and its ribose bond. At the other end of NAD^+, the phosphates are coordinated by positive charges of R113 and K185 and the adenosine moiety through π-stacking with the R117 and W133 sidechains. Together, all interaction arrange the NAD^+ molecule in an extended conformation where the 2'-OH of the ribose, located next to the nicotinamide, becomes susceptible to an attack of the sidechain oxygen of actin T148. Ultimately, a productive nucleophilic attack occurs, after which the ADP-ribose of NAD^+ is covalently bound to F-actin. The leaving group nicotinamide remains at its original coordination side until TcART dissociates from ADP-ribosylated F-actin. Finally, the interaction between Kζ185 and E265, together with the movement of 198-208, closes the gate of the nucleotide-binding pocket and TcART returns to its apo-state. It can now catalyze another ribosylation reaction.

The molecular organization and functioning mechanism of TcART, the functional toxic subunit of TccC3, does not yet explain its behavior in earlier stages of the transfer cascade. To gain insight in this process, its state inside the TcB-TcC complex, after autoproteolysis but prior to injection into the target cell, was focused on.

4.2.6. TcHVR is unfolded inside the TcB-TcC complex

The complex formed by TcdB2 (UniProt Q8GF99) and TccC3 (UniProt Q8GF97) examined here includes not only the N-terminally truncated TcART but full-length TcHVR (comparison in Fig. 4.12 on page 75). TcdB2-TccC3 (TcB-TcC) has a weight of more than 270 kDa in its entirety and is thus not readily accessible by solution NMR due to slow tumbling of the molecule. In order to acquire directly comparable data, solid–state NMR measurements at 100 kHz MAS were conducted as these yield high resolution spectra without need of perdeuteration of the protein sidechains. To this end, TcB-TcC was expressed ^{15}N-labeled, ultra-centrifuged into a 0.7 mm diameter rotor, and subsequently a ^{1}H-^{15}N correlation was recorded. The foremost issue was to estimate how complete, and therefore representative, this data is.

Arginine Nε signals indicate spectral completeness

Having a total number of 2474 amino acids, the full TcB-TcC complex yields a highly congested ^{1}H-^{15}N correlation in the region characteristic for backbone signals. However, as only 150 of these are arginine, the spectral region in which their sidechain Nε arise (80-90 ppm) is less crowded (Fig. 4.20). Therefore, although a bulk overlap is present in the ^{1}H range between 7 and 8 ppm, specific arginine signals are more dispersed and yield peaks that are countable. By integrating these regions, the number of signals can be estimated to provide a measure of the spectral completeness.

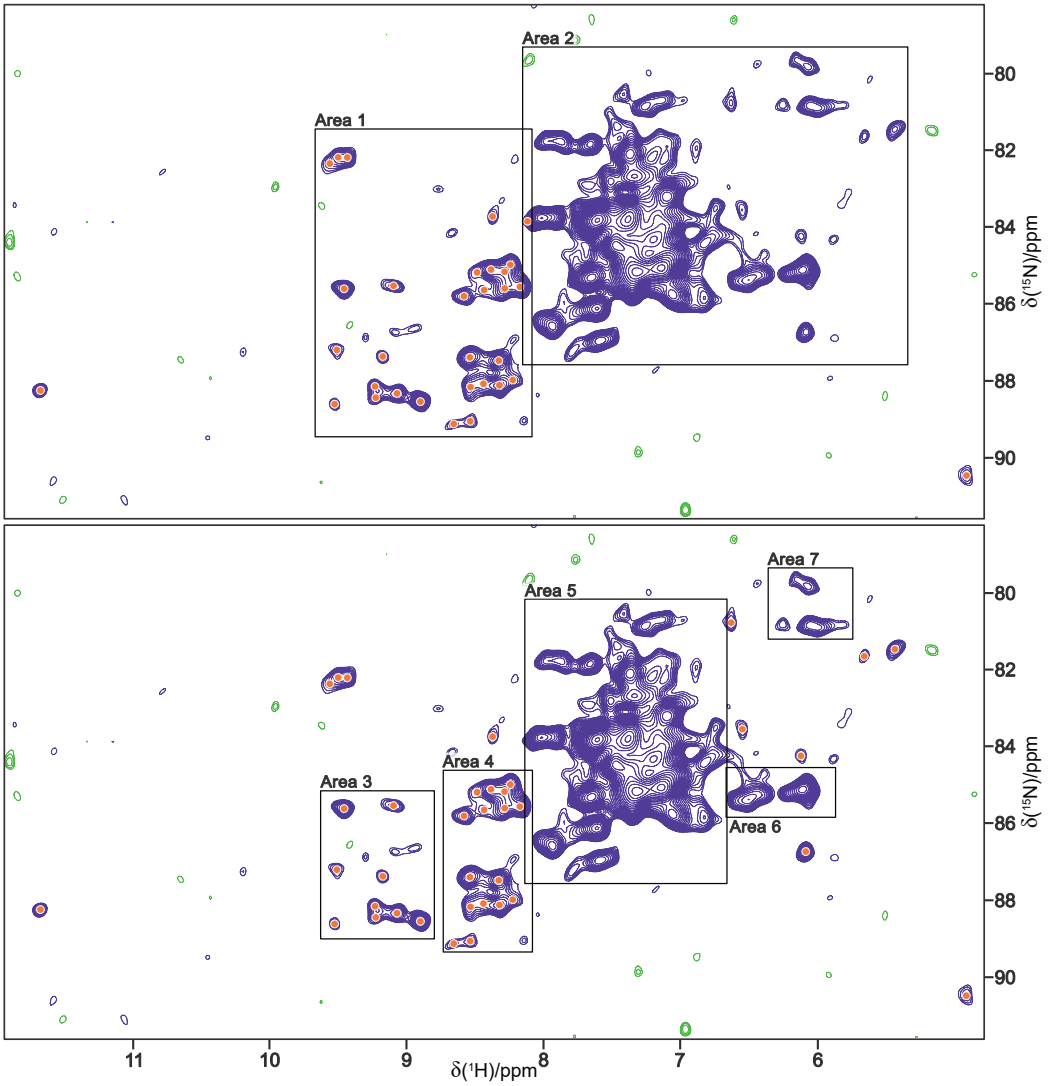

Fig. 4.20.: Arginine NεH region of a ^{1}H-^{15}N correlation of the TcB-TcC complex. Two different integration approaches (rectangular regions) and peak counting schemes (orange) are indicated. Signals with more than two contour lines are considered as peaks. Integrals of the indicated areas can be found in Tabs. 4.1 (below) and A.3 on page 126.

Tab. 4.1.: Selected peak estimates of Arg ^{1}H-^{15}Nε integral regions of TcB-TcC.

ref. area	peak count	area	abs. integral	est. peaks	total
1	**30**	**1**	**$5.50 \cdot 10^{12}$**	**30.00**	
		2	$2.16 \cdot 10^{13}$	117.55	
		-	exterior peaks	2	Σ 149.55
		3	$1.62 \cdot 10^{12}$	8.84	
		4	$2.87 \cdot 10^{12}$	15.64	
		5	$1.80 \cdot 10^{13}$	98.13	
		6	$1.19 \cdot 10^{12}$	6.49	
		7	$6.46 \cdot 10^{11}$	3.52	
		-	exterior peaks	12	Σ 144.62

Different areas in the arginine Nε region were designated for integration and peak counting as presented in Fig. 4.20 by black rectangles and orange dots, respectively. Peaks with more than two contour lines are considered as sufficiently above the noise and therefore valid. As such, the exact number of signals in areas 1, 3, and 4 can be determined and extrapolated to the ones remaining. Starting from 30 signals in area 1 with an absolute integral of $5.50 \cdot 10^{12}$, area 2 can be estimated to contain 117 to 118 signals (Tab. 4.1). Together with the two peaks remote from integrated regions, a total of 150 signals is approximated. Applying this peak-to-integral ratio on areas 3-7 and adding peaks which are not yet included in these areas, an estimate of 145 is obtained (Tab. 4.1). This lower count is not surprising as parts of areas 1 and 2 are considered twice due to overlap. Complementary starting estimates of either 9 peaks in area 3 or 16 peaks in area 4 (Appendix Tab. A.3 on page 126) result in a composite approximation of 149±4 arginine Nε signals. This matches the number of the 150 arginines in TcB-TcC and ensures that TcHVR (possessing 7 arginines) is contained in the signal set. Having validated spectral completeness, a representative comparison of these ^{1}H-^{15}N correlations with results obtained from solution NMR can be made.

Characteristic peaks of TcHVR are absent in TcB-TcC

Upon assembly of TcB-TcC, TcHVR is shielded inside the complex and gets autoproteolytically cleaved through the action of two aspartates [72]. It then resides buried within the complex until the injection process occurs (Fig. 4.21A). This state eluded detailed examination in previous X-ray and cryo-EM studies that both only revealed unresolved density [70, 72, 84]. Cross-linking mass spectrometry showed that while the N-terminal cleavage remains at a fixed position, the rest of the protein assumes random orientations, indicating a certain degree of structural variability [75]. However, the question remains whether TcHVR is folded or unfolded. This was addressed here by combining the presented solid–state NMR data of the TcB-TcC complex and structural insights from solution NMR of TcHVR alone.

In order to extract information from the crowded backbone region, we focused on a set of signals with distinguished chemical shifts in the ^{1}H-^{15}N correlation of TcHVR. These are the peaks arising from the backbone of V190, W191, T250, and S252 as well as the sidechain of Wε191 and Nδ257. All these residues are located in TcART (shown as spheres in Fig. 4.21B) and form either strong hydrogen bonds (T250 to Q247 and S252 to Dγ244) or are embedded inside the hydrophobic core (V190, W191, Wε191, and Nδ257). As such, they are sensitive monitors of the state of the protein, which we also referred to as ‘indicator peaks’. If their local environment is rearranged after un- or misfolding, a change in their chemical shift is expected. The locations of these ‘indicator peaks’ are marked by

circles in Fig 4.21C. If the protein occupies the same folding state inside the complex (dark blue spectrum) as in solution (green spectrum), peaks with very similar chemical shifts are expected for the dark blue spectrum at these positions. As these signals are absent in the complex, it can be concluded that TcHVR must be in a different state in the complex than in solution. Furthermore, expression of TcB-TcC without the inlaying toxin yielded a nearly identical spectrum as the dark blue one in Fig. 4.21C (see Appendix Fig. A.14 on page 122 for a direct comparison). The fact that all other distinguished peaks remain at their position demonstrates that the designated 'indicator peaks' are not slightly shifted but have drastically altered peak positions as can be expected after a complete unfolding.

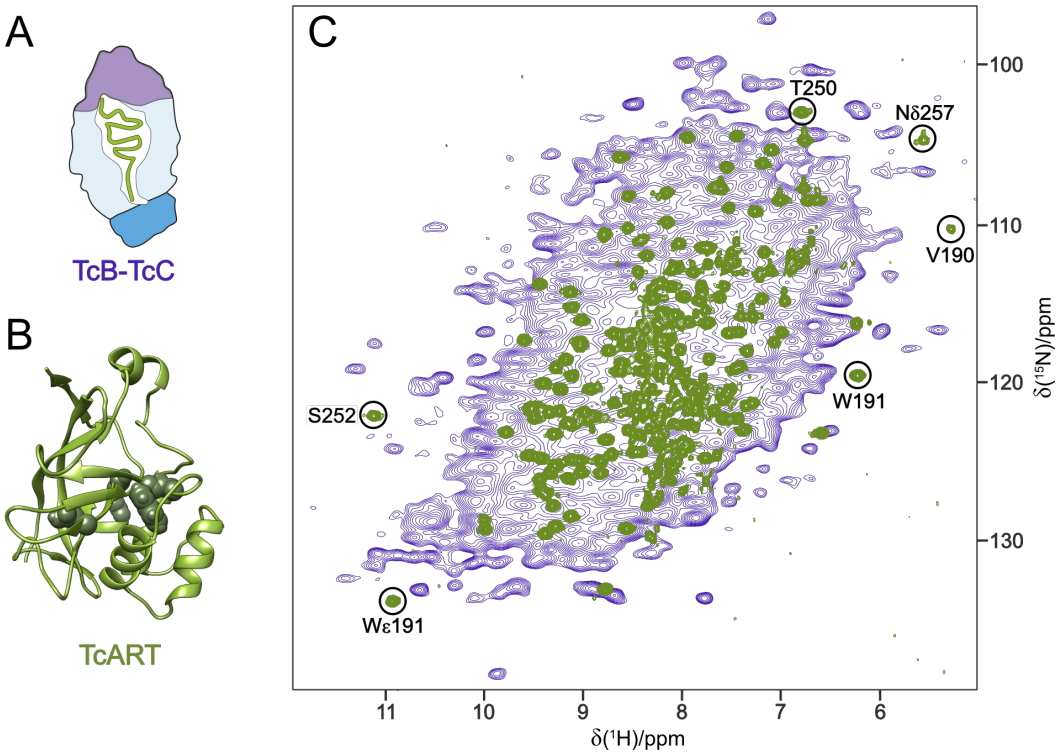

Fig. 4.21.: Comparison of ^{1}H-^{15}N correlations of free and encased TcHVR. (**A**) Schematic representation of the TcB-TcC complex with TcHVR (green) located inside. (**B**) Lowest energy solution NMR structure of the folded and truncated domain of TcHVR, TcART, with the backbone atoms of V190, T250, and S252 visualized as spheres. For Wε191 and Nδ257, backbone and sidechain atoms are shown. (**C**) Superposition of ^{1}H-^{15}N NMR spectra of the full TcB-TcC complex (solid–state, blue) and full-length TcHVR alone (solution, green). Signals diagnostic for the folded structure in solution, as presented in **B**, are indicated by circles.

4.2.7. Conclusion

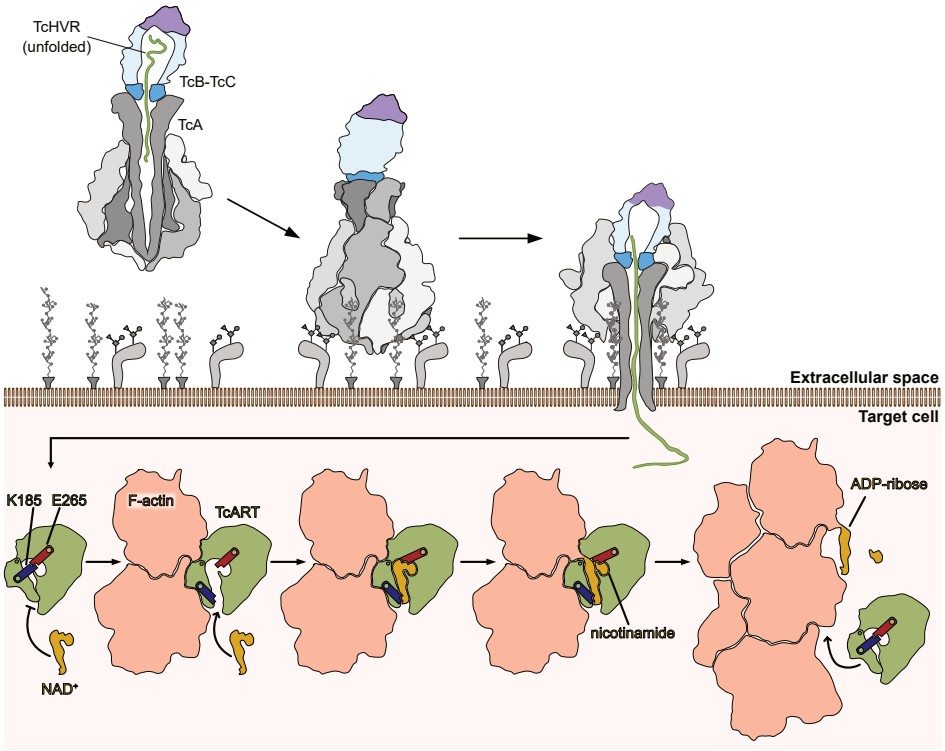

Fig. 4.22.: Schematic representation of ADP-ribosylation by TcART. TcHVR gets injected by action of the TcA, TcB, and TcC complex before adopting its globular structure. It then binds NAD^+ as a cofactor with K185 and E265 playing a crucial role for access to the binding site and alignment of the ligand. Finally, the ADP-ribose moiety gets covalently attached to F-actin.

By a combined study using NMR, cryo-EM, and molecular modeling, the understanding of TccC3 from *Photorhabdus luminescens* could be deepened (Fig. 4.22). Solid–state NMR showed that its C terminus, the actual toxic moiety TcHVR, is unfolded during initial stages of the injection process when TccC3 is forming a complex with TcdB2. Even after insertion into the target cell, its N terminus remains unfolded and is not required for enzymatic activity. Instead, amino acids 102 to 282 form a globular structure, termed TcART, with a stable hydrophobic core that does not exchange with the bulk water. In proximity of the binding pocket, key residues as well as the overall tertiary structure are homologous to other ribosyltransferases. This leads to a conserved binding mode for its ligand NAD^+ whose coordination involves major rearrangements of backbone and sidechains including those of E265, K185, and a flexible loop in the region 198-208. After ligand and target binding, TcART is located at a two-subunit interface of F-actin which it covalently modifies by addition of an ADP-ribose. After dissociation, the toxin returns to its initial state.

5. Discussion and Outlook

5.1. Significance of the TasA assembly mechanism

The results presented in this thesis show that TasA can assemble into highly homogeneous filaments under native conditions. This newly described state has crucial implications for the composition of biofilm assemblies as well as the coordination and cross-talk happening within.

5.1.1. The native state of TasA in biofilms

To be, or not to be, amyloid

An amyloid, or more cautiously termed amyloid-like, state has long been considered certain for TasA. This premise was mainly based on the observation that *Bacillus subtilis* colonies can be stained by Congo Red [19, 22] and that TasA monomers denatured by acid can be reassembled into Thioflavin T-stainable fibers [22, 52, 189]. These dyes are established indicators of amyloids and are routinely used [53], but sometimes Congo Red shows some off-target specificity [55]. However, contrasting results showed that TasA can also assemble into higher-order structures from folded monomers without acid treatment [54], which has recently been validated by a number of publications [55, 56, 177], including the findings presented here. These assemblies, which we now term filaments, are not stainable by Thioflavin T and, thus, non-amyloid [1]. All studies agree that externally supplemented TasA can complement a deficient Δ*tasA* strain. However, which state it then occupies remains elusive. By combining the supplementation approach with the detection of isotope-labeled TasA, the results in section 4.1.1 show that non-amyloid filaments are present within the *Bacillus subtilis* biofilm. Still, the TasA-supplemented Δ*tasA* biofilm is not identical to the wildtype, suggesting that some differences must exist. Possibly, wildtype biofilm contains a fraction of TasA that occupies a different state stainable by Congo Red, giving the biofilm its previously determined characteristic.

This duality of TasA could explain the contrasting results published and would fit the picture of a diverse biofilm where multiple local conditions coexist, and single cells are highly specialized [190]. Indeed, the few differences between the ^{1}H-^{15}N correlations of *in vitro* assembled filaments, and complemented biofilm (Fig. 4.1 on page 59) indicate that TasA filaments undergo minor conformational changes when brought into their native environment.

The TasA interaction interface

Within the biofilm, TasA is embedded into a dense phase consisting of bacteria, exopolysaccharides (EPS), ions, other proteins (*e.g.*, TapA [50]), and possibly other substances [16]. The observed differences between free and biofilm-contained TasA are likely due to interaction with these components. A study that cultivated a Δ*eps Bacillus subtilis* strain for 250 generations to allow evolutionary selection found that mutations in TasA can compensate for EPS absence [191]. Interestingly, the observed substitutions Y124C and G183C, happening independently in different populations, introduce a single cysteine into the TasA sequence. Selectively introducing either mutation into Δ*eps Bacillus subtilis* leads to a restoration of the dense biofilm [191]. For both mutants, negative stain transmission EM micrographs examining the filaments in more detail found an altered morphology with the more abundant formation of TasA filament bundles. These exhibit a step-wise increase of 2 to 6 times the diameter of single filaments, indicating an unchanged nature of the filament as a subunit of the bundles.

It appears reasonable that introducing a cysteine at either location (Y124C / G183C) leads to the formation of disulfide bridges between TasA filaments, stabilizing inter-filament connections. This substitution is especially significant as native TasA contains no cysteines and must rely on less rigid interaction for its association. Bundles of pure, non-mutated TasA have been observed [56], which raises the question of how these interactions are mediated without cysteines and possibly within the biofilm. As Y124 and G183 only mutate in the absence of EPS, the area around these residues is likely part of an interaction interface with this biofilm compound. This raises the question of how the interaction is mediated.

Based on the data presented in this thesis (section 4.1.5 on page 68 and Fig. 4.7 on page 69), the unassigned stretch around Y124, 117-126, is presumed to be mobile. Seeing as these signals are not assigned, it is possible that the binding of other biofilm components renders the entire region 117-126 less mobile, making it observable by solid–state NMR. Therefore, this region could give rise to the newly appearing peaks under *in vivo* conditions (Fig. 4.1 on page 59). Transient, non-covalent interactions could mediate the binding for which other biofilm constituents like EPS or other TasA filaments (bundling) are possible

interaction partners. As the introduction of a single cysteine into the TasA protein occurs only upon removal of EPS, it can be assumed that this region plays a crucial role in wildtype *Bacillus subtilis* where EPS are present. This argument is also valid for G183, which is similarly not assigned in the solid–state NMR data. These observations align with the recently published cryo-EM study of TasA filaments. There, the authors focused on the residues 174-177 which are located close to 117-126 and G183 in the three-dimensional structure (Fig. 5.1).

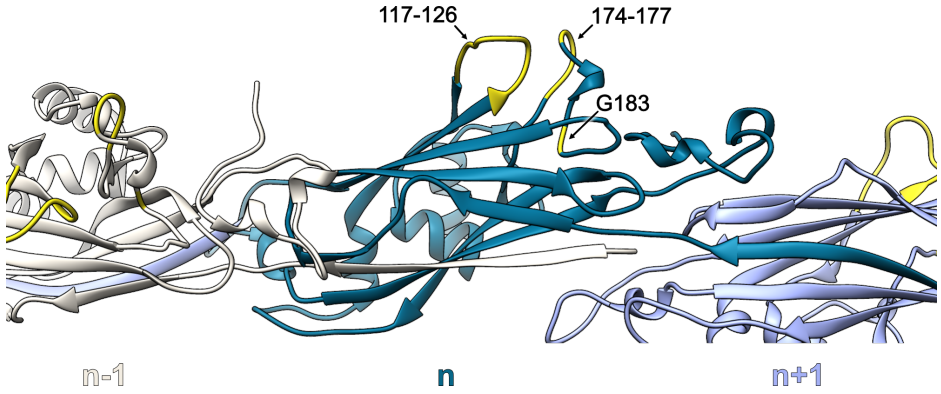

Fig. 5.1.: Regions of TasA filaments relevant for heteromolecular interaction. Prediction of TasA filaments by AlphaFold-Multimer with the regions 117-126, 174-177, and the residue G183 colored yellow.

In their 2022 paper, Böhning et al. [56] propose that the region 174-177 is essential for TasA filament bundling. According to their findings, this stretch is the most remote from the helical center, and mutation of all four residues to alanine ($174\text{-}177_{AAAA}$) abolishes biofilm formation in a manner similar to the Δ*tasA* deletion. This finding partially clashes with results obtained by NMR for residue G175 which is located the stretch 174-177. In both solution NMR measurements of monomeric TasA and solid–state NMR spectra of pure TasA filaments, G175 has the same H^N chemical shift of 9.6 ppm indicative of a highly ordered environment. However, an identical environment in the monomeric and filamentous state is incompatible with the assumption of G175 participating in filament bundling. In contrast to these *in vitro* results, the 9.6 ppm signal is absent in the ^{1}H-^{15}N correlation recorded on unlabeled biofilm supplemented with labeled TasA (Figs. 4.1F and 4.2A). Thus, the NMR data suggests that although G175 appears to have a similar chemical environment in the *in vitro* filamentous and monomeric states, it might rearrange in the native biofilm.

The relevance of the surface created by the stretch 174-177 together with 117-126 and G183, is additionally reinforced by the results obtained by MD simulations presented in [56]. In these calculations performed for filament doublets running anti-parallel, the regions 118-121, 173-179, and 185-190 all participate in the association process. Again, this evidence indicates that the surface area formed by these regions is essential for filament

bundling. The interplay of TasA with itself and other matrix components also becomes a significant issue for its transition from a monomeric to a filamentous state.

Pathway of TasA filament assembly

The partial or complete unfolding of TasA has played a role in previous studies examining its amyloid properties. In addition, the disassembly of its structure is also essential in the biofilm context. TasA is known to exist folded inside bacteria due to its role in the sporulation process [36], and its transport to the extracellular matrix is conducted unfolded via the SecYEG complex [192]. As a result, TasA does not occupy any specific structure when it initially arrives in the biofilm.

Starting from an unfolded state, TasA can either directly transition into filaments or indirectly by first forming globular fold and subsequently assembling into filaments (illustrated in Fig. 5.2A). The assembly of filaments from folded TasA was used in the work presented here and was previously published by Chai et al. [54]. In contrast, the formation of non-amyloid filaments starting from unfolded material has not yet been explicitly shown. However, our NMR measurements indicate that folded TasA has the potential to unfold partially. This behavior is the most reasonable explanation for the almost complete observation of H^N signals from the hydrophobic core of TasA. Typically these residues carry deuterons after recombinant expression in D_2O and only exchange with the bulk water upon unfolding of the protein. For TasA, these signals are present in the spectra without intentional unfolding and thus indicate a loose globular fold of monomeric TasA. This conformational flexibility of TasA could promote structural changes necessary for the filament assembly process.

The transition of folded, monomeric TasA to the filamentous state might happen by a concerted mechanism. β2 covers a hydrophobic cleft of significant size in the X-ray structure which will be exposed upon rearrangement of β1 and β2 (Fig. 4.11 on page 74). This is not advantageous when in an aqueous environment and consequently the cleft is covered differently in the filament model, namely by $\beta0^{n-1}$. The insertion of $\beta0^{n-1}$ between $\beta2^n$ and $\beta9^n$ creates an inter-molecular connection. The sequential proximity of $\beta0^n$ to $\beta2^n$ might indicate an interdependence during this process. Extension of $\beta0^{n-1}$, with or without binding, could induce a pulling force on $\beta1^{n-1}$ and $\beta2^{n-1}$ leading to the observed rearrangement in the adjacent molecule. Vice versa, occupation of the hydrophobic cleft of molecule n by $\beta0^{n-1}$ might cause a $\beta0^n$ extension to become favorable.

Within the native biofilm environment, other components likely assist TasA during its transition to filaments. Possible partners are the contained EPS, membrane components, or proteins. As discussed above, the interplay with EPS as well as with membrane components still needs to be characterized. Of the proteins contained in the biofilm, the interaction

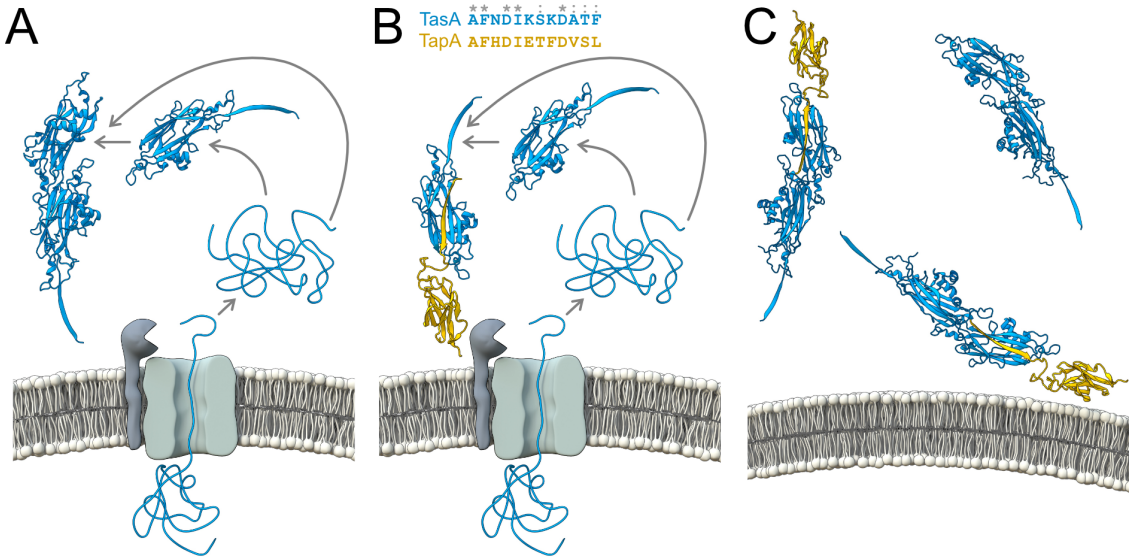

Fig. 5.2.: Possible ways of filament formation in the biofilm. (**A**) TasA (blue) is secreted through the SecYEG channel unfolded. Afterwards, it can either directly assemble into filaments or transition through an intermediate folded state. (**B**) TapA (yellow) can assist TasA filament formation by donating an initiator strand, which is made possible by the N termini of TapA (44-55) and TasA (28-39) being highly similar (shown on top). (**C**) Within the biofilm, all processes might happen simultaneously, with TapA associating with the membrane.

of TasA with its accessory protein TapA has been the focus of several studies. TapA is known to accelerate the formation of amyloid [50] and non-amyloid [1] polymers of TasA starting from folded and unfolded material. Its gene is located in the same operon as TasA, together with their common signal peptidase SipW (*tapA-sipW-tasA*) [19]. The expression of all three proteins is coordinated, and TapA generation precedes that of TasA. Both proteins are exported in the same manner, and TapA was shown to associate with the outer cell wall of *Bacillus subtilis* [19]. Therefore, when TasA molecules arrive unfolded through the SecYEG channel, TapA is already present to assist in filament formation. This process is facilitated by the TapA N terminus, whose initial 12 amino acids are homologous to those of TasA [1, 56]. It can complement TasA by donating an initiator strand (Fig. 5.2B) into the hydrophobic cleft. TapA and TasA form an initiator complex, which then assembles with further TasA molecules to become a filament. We recently validated the TapA-TasA interaction by analytical ultracentrifugation [1], and AlphaFold-Multimer predicts the interaction similar to the one happening in pure TasA filaments (more detailed view in Appendix Fig A.15 on page 123). It is noteworthy that TapA can only be located at the beginning of a filament as its structure allows only donating but not accepting a β-strand. Naturally, the entire process can also occur independently of membrane association (Fig. 5.2C). Whether TapA acts on folded, unfolded, or any mixed state of TasA needs to be examined in further detail.

5.1.2. Potential for TasA in multi-species biofilms

The assembly of TasA filaments and their interaction with TapA has striking similarities to mechanisms observed in other bacteria. For example, pili in the chaperone-usher pathway [193] of many gram-negative bacteria also assemble by donor-strand complementation. This interaction principle is, among others, relevant for uropathogenic *Escherichia coli* [182], the main contributor to periodontal disease *Porphyromonas gingivalis* [194], and the infamous nosocomial pathogens *Acinetobacter baumannii* [183] and *Pseudomonas aeruginosa* [195]. Three of these species are on the WHO list of priority pathogens [7] and require urgent research to establish new treatment options. Indeed, disassembling their biofilms by disrupting the donor-strand complementation mechanism might be a promising attack vector. Due to the ubiquitous occurrence of this assembly mechanism, such an approach could yield wide-ranging anti-microbial effects.

There is reason to assume that filaments similar to those of TasA are also produced by other *Bacillaceae*. TasA-homologue proteins in *Bacillus cereus* are called CalY1 and CalY2 and have previously been proposed to share common traits with TasA [52]. In addition, replacing the *sipW-calY2-calY1* genes of *B. cereus* with the *tapA-sipW-tasA* operon of *B. subtilis* does not influence the appearance of the biofilm pellicle. This interchangeability gives further credence to the idea of a homologous filament formation mechanism.

In a broader context, a characteristic conservation pattern can be found when comparing TasA with CalY1, CalY2, and proteins with similar sequences expressed by organisms with varying evolutionary distances (Appendix Fig. A.7 on page 116). Even when the overall identity of the compared sequences is as low as 26 %, a number of residues situated at the interface remain conserved (spheres in Fig. 5.3). Most noteworthy are G42 and almost all residues (F29, P62, N215, Q218, and Q232) of the ternary contact site where three TasA molecules meet (previously also depicted in Fig. 4.10 on page 73). F200 is not as strongly conserved as it is substituted with other hydrophobic amino acids in some organisms.

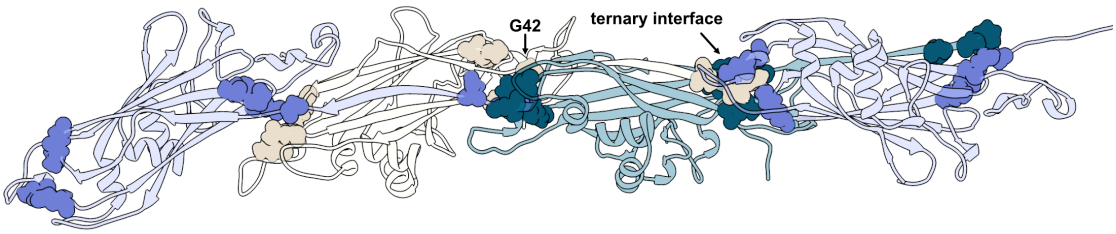

Fig. 5.3.: Conserved residues between TasA and homologues from other organisms. TasA filament model with residues that possess a high conservation, as determined by the alignment shown in Appendix Fig. A.7 on page 116, indicated as spheres. F29, D31, G42, P62, G63, N74, G76, N215, Q218, and Q232 are highlighted.

While the conservation of the ternary interface site is high, the strand β0 is not strictly conserved except for F29 at its beginning. This variability is unsurprising as β-sheet structures typically present functional conservation and do not show exact amino acid identity. However, the high conservation at the TasA-TasA contact area potentially enables the formation of joint assemblies, possibly through a shared 'proline box' mediated by a hydrophobic cluster at the interface. As such, *Bacillus subtilis* might be able to form heterogeneous filaments with homologue proteins originating from other species. Filaments composed of different proteins would be in line with mixed biofilms being commonly encountered in nature and industrial settings [15, 196]. Furthermore, *Bacillus subtilis* biofilm formation was found to be stimulated by other organisms such as *Bacillus cereus* [47], that expresses the homologue proteins CalY1 and CalY2. A fundamental role for the donor-strand complementation mechanism, as shown for TasA and assumed for its homologues, can be envisioned in the interplay of different species within the biofilm.

5.1.3. Further characterization of peak doubling and flexibility

The work presented in this thesis enabled a more detailed understanding of how TasA transitions from its monomeric state into filaments and offered perspectives for interactions of TasA with other biofilm components. Nevertheless, further work on TasA could provide more information about its intrinsic properties.

The exact nature of the observed peak doubling (section 4.1.2 on page 58) is a point that could be investigated further. The observed effect could arise from different conformers fixed within the filament assembly or local flexibility from exchanging conformations. Such information about molecular dynamics can be accessed through the measurement of relaxation times, *e.g.*, $R_{1\varrho}$, which is typically used in solid–state NMR. This data might not only enable a detailed understanding of how atoms and bonds move within the molecule but also explain the bending observed for TasA filaments (Fig. 4.1A and B). This macroscopic effect needs to be intertwined with flexibility on the molecular level. In a different system with similar properties, residues in the SPP1 bacteriophage tail tube that contributed to bending were identified by solid–state NMR [136]. For TasA, G42 (Fig. 5.3) might be highly relevant to facilitate bending. It is situated between strands β0 and β1, directly in the middle of two adjacent TasA molecules. The high conservation of a glycine at this position points towards an inherent need for flexibility.

5.2. Considerations concerning the action of TcART

Actin, the target of TcART, is a highly conserved cytoskeletal protein and an essential platform for many cellular processes [197]. Disruption of its homeostasis has fatal consequences for the cell. As a result, both forms that are present, filamentous F-actin and globular G-actin, are the target of a number of ADP-ribosyltransferase toxins. These proteins are produced by a wide range of bacteria, including some human pathogens [198]. All previously characterized toxins ADP-ribosylate G-actin at position R117, whereas the findings presented here show that TcART exploits T148 with a novel modification mechanism. The area around T148, which is composed of two consecutive F-actin subunits, and the exact reaction timeline employed by TcART warrant a detailed analysis.

5.2.1. Binding mechanism of TcART to F-actin

The TcART-F-actin interface revealed by our cryo-EM structure has previously been described as relevant for other binding partners (*e.g.*, myosin-V [199], ExoY from *Pseudomonas aeruginosa* [200], and the Lifeact peptide [201]). These three molecules target the same hydrophobic pocket as Y183 in TcART, making recognition of this site a crucial property. Interestingly, this area is located directly between two subunits of F-actin and, therefore, does not exist in G-actin. Having a Y183 homologue is thus an essential property for targeting filamentous F-actin specifically.

The contribution of Y183 for binding of TcART to F-actin becomes increasingly important when viewing it in light of the conformational changes that occur. While the tyrosine sidechain is flexible in the apo-state of free TcART (Fig. 4.19 on page 87), multiple other interaction sites are not available. The salt bridge between the sidechains of K185 and E265 in the apo-state (Fig. 4.19 as well) causes the adjacent residues K185, D186, K187, and D263 (Fig 4.18 on page 86) to be unavailable for interaction, which impedes binding. Therefore, the disruption of the K185-E265 salt bridge is necessary for the formation of the final TcART-F-actin complex. As these two residues are also crucial for NAD^+ binding (section 4.2.5), the question arises which process happens first. Is the association with F-actin necessary for a formation of a suitable NAD^+ pocket? Or does the coordination of NAD^+ induce conformational changes that facilitate F-actin binding?

Our co-sedimentation experiments (Appendix Fig. A.16A on page 123) showed that the presence of NAD^+ does not affect the affinity of TcART for non-modifiable F-actin. This means that F-actin does not prefer NAD^+-TcART over apo-TcART. It can be concluded that a NAD^+-induced collapse of the K185-E265 salt bridge and further preparatory rearrangements are not needed for efficient binding of TcART to F-actin. Therefore, the results hint towards F-actin binding being the initial step and NAD^+ coming into play afterwards.

When considering these results together, a possible interaction timeline can be deduced. As a first step, apo-TcART binds to F-actin through the hydrophobic interaction facilitated by Y183, which is always in a binding-competent state. The initial binding induces conformational changes that lead to favorable orientations of the charged residues, including K185, which then constitute electrostatic interactions between TcART and F-actin. In turn, the TcART pocket (*e.g.*, E265) becomes available for tight coordination of NAD^+, and the modeled pre-reaction state (Fig. 4.19) is formed.

5.2.2. Examining the reaction mechanism in further detail

An exact model of the ensuing ADP-ribosylation mechanism is currently unavailable. However, based on the titration experiments, intricate binding and modification processes can be assumed.

The ^{31}P spectra (Fig. 4.15 on page 81) indicate that a simple model of two NAD^+ states (bound and unbound) is insufficient.The shifting of signals alone cannot explain the peak pattern arising upon binding. Additionally, it is remarkable that $\sim$3 times excess of NAD^+ over TcART is insufficient to manifest a signal for the unbound state. An increase to 10 times the concentration of TcART yields peaks at the expected position of unbound ligand. The ligand, however, still appears to interact with the larger, slower-tumbling protein, as indicated by the observed signal broadening. Strikingly, the triplet signals closer to -14 ppm do not show a corresponding increase in linewidth. A possible explanation for these narrow lines arises from the enzymatic activity of TcART. Although NMR data acquisition happened soon after the preparation of the sample, some NAD^+ might already have been hydrolyzed, giving rise to the observed ^{31}P signals. However, the NAD^+ concentration directly influences the triplet signals between -14 and -15 ppm, which makes it unlikely that they are caused by hydrolysis products. The signals close to the noise at $\sim$17, 1.4, and -4 ppm appear more probable to arise from reaction products. The two peaks at 14 and -15 ppm can then be assumed to arise from a flexible, loosely coordinated section of NAD^+.

In addition to the effects observed on the ligand side, the ^{1}H-^{15}N spectra of TcART upon NAD^+ presence also propose a multi-step mechanism. The different patterns of peak movements are directly indicative of fast (G129) and slow exchange (T198, G201, G202, G206), possibly within the same molecule (Fig. 4.16 on page 83). It appears likely that the flexible loop (198-208) undergoes slow motions when alternating between a ligand-bound and unbound state. The bound state would only be observable when sufficiently populated at higher ligand concentrations. The occurrence of the fully bound state is directly evident from A114, which forms two hydrogen bonds with NAD^+ (Appendix Fig. A.13C on page 121), causing a drastic chemical shift perturbation only observable at the 10-fold

ligand excess. Therefore, it is plausible to assume that the ADP-ribose moiety of NAD^+ has a fast exchange not binding tightly. In contrast, the conserved A114 to nicotinamide coordination forms a strong interaction and undergoes slow exchange. In addition, the tight binding of the nicotinamide moiety is in line with it being the electrophilic center of reaction, whose spatial position is essential and needs to be coordinated tightly. For a more detailed examination of peak position changes, it would be beneficial to repeat the titration experiments with a greater variety of protein-to-ligand ratios and non-hydrolyzable analogs of NAD^+.

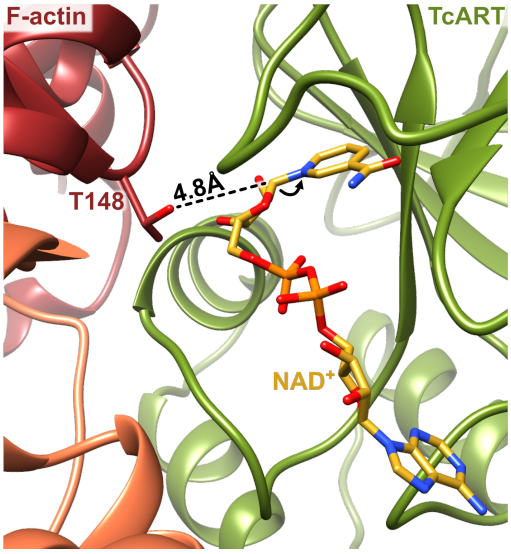

Fig. 5.4.: Location of T148 relative to NAD^+. Model of NAD^+ docked into the cryo-EM structure of the F-actin (red and orange) and TcART (green) complex. The nucleophil T148-OH is located 4.8 Å away from the electrophilic center.

Lastly, the exact nature of the hydrolysis reaction itself remains to be determined. For the nucleophilic attack, the established S_N1 (substitution through an intermediate oxocarbenium ion) and S_N2 (direct substitution) reaction mechanisms [202] must be considered. From the circumstantial evidence, an S_N2-like mechanism appears likely. The strained conformation of bound NAD^+ inside the catalytic center reduces the activation energy [203] and correctly orients the pyridinium N-glycosidic bond for a nucleophilic S_N2 attack from the F-actin side (indicated in Fig. 5.4). In contrast, the F-actin modification site T148 is only a weak base and not suited for such an attack over distance of 4.8 Å in the model. Moreover, the ribosylation mechanism of the iota toxin, possessing an R-S-E motif in the active center as TcART does, was previously proposed to perform an S_N1 reaction [204]. A clear determinant of the mechanism would be the absence (S_N1) or presence (S_N2) of perfect stereochemical inversion at the electrophilic ligand center through the reaction. However, this information was not attainable with the applied experiments. The exact reaction mode could be addressed by introducing a chiral center at the reaction site. Char-

acterizing the change of chirality (racemate formation or inversion) would be an easy way to monitor the nature of the nucleophilic substitution. Further, a systematic analysis of ^{31}P lineshape and *J*-splitting pattern, as previously done for ATP [205], could give insights to the interactions of NAD^+.

5.2.3. Downstream effects of F-actin ADP-ribosylation

After the ADP-ribosylation of F-actin, TcART dissociates from the filament. It can be assumed that the K185-E265 ionic interaction forms again and closes the NAD^+ pocket. The occurrence of this process is indicated by wildtype TcART preferentially co-sedimenting with unmodified F-actin over ADP-ribosylated F-actin (Appendix Fig. A.16B on page 123), showing that the covalently attached ADP-ribose (Fig. 5.5) interferes with the binding process. Furthermore, disruption of the K185 to E265 salt bridge by introducing a mutation of the glutamate to serine causes E265S-TcART to retain its ability to bind F-actin independent of its ribosylation state (Appendix Fig. A.16B as well). Therefore, it can be concluded that K185-E265 employs a gating mechanism that prevents re-binding to modified F-actin and, consequently, futile substrate encounters.

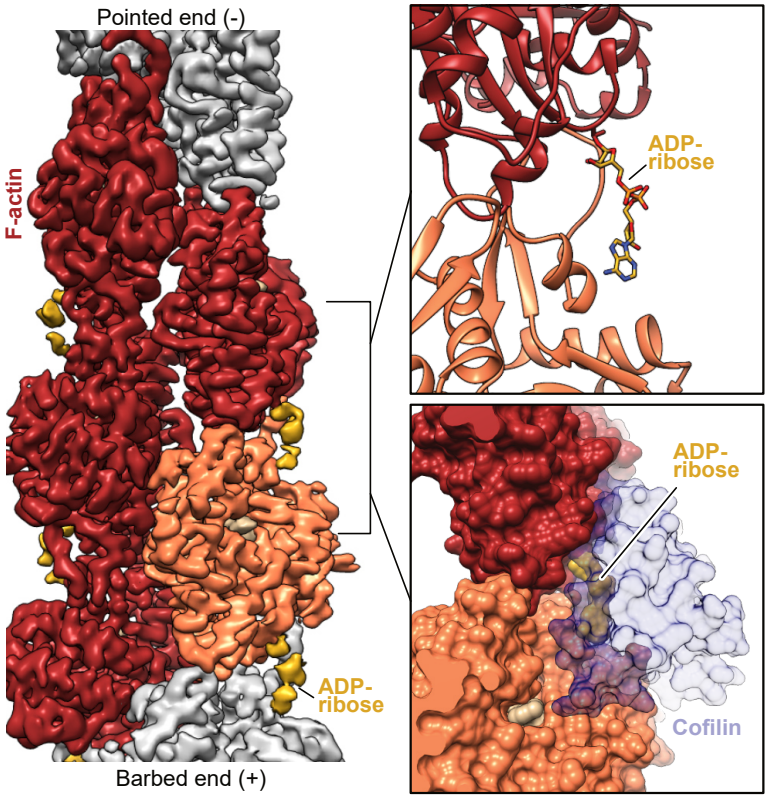

Fig. 5.5.: The ADP-ribose obfuscates the Cofilin binding site. Cryo-EM density (left) and atomic model (right) of ADP-ribosylated F-actin (red and orange). Cofilin is shown as surface representation (blue) at its expected binding site. Cryo-EM density for ADP-ribose (yellow) is shown at a lower threshold than that of F-actin. Figure adapted from [2].

On the side of F-actin, covalent modification through the addition of an ADP-ribose is likely not enough to stabilize the filament in a manner that causes cell death by uncontrolled polymerization. The fatal nature of this modification rather comes from the impairment of interaction with actin depolymerizing factors. The lack of regulating interactions disrupts the F- and G-actin balance and causes filaments to accumulate. For example, the presence of the ADP-ribose directly obfuscates the cofilin binding site [2, 83], as shown in Fig. 5.5. Thus, ADP-ribosylation of T148 stabilizes actin filaments indirectly by hindering cofactor-mediated depolymerization.

5.2.4. Sugars and aromatics interact with the TcART pocket

After examining the canonical pathway of TcART function, it is worth considering off-target specificity. After recombinant expression and protein purification, one or multiple molecules co-purified with the toxin. Their presence is evident from the intensity and chemical shifts observed for non-TcART signals in the NMR experiments. Their signal intensity is comparable to those of the toxin, and ^{13}C-^{13}C magnetization transfer along their bonds was possible. Both properties speak towards incorporated ^{13}C isotopes. At natural abundance ($\sim$1 % ^{13}C), the intensity of these non-TcART signals would be much weaker. Additionally, carbon-to-carbon magnetization transfer would be virtually impossible as two adjacent ^{13}C nuclei are needed. Consequently, these molecules must result from bacterial metabolism during recombinant expression with a ^{13}C-enriched medium.

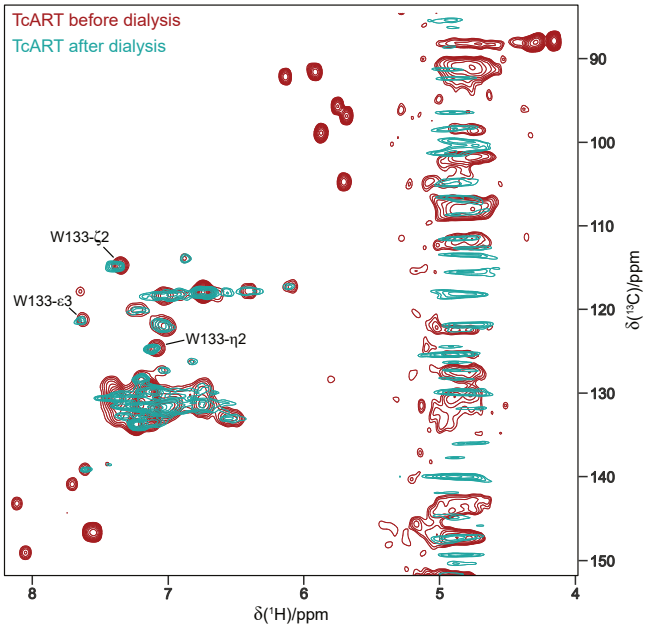

Fig. 5.6.: Non-TcART signals can be removed by dialysis. Aromatic regions from ^{1}H-^{13}C correlations of TcART before (dark red) and after dialysis (green) to an excess of buffer. After dialysis, signals in the $\delta(^{13}C)$ ranges 85-105 and 140-150 ppm are absent. The W133 sidechain signals show minor chemical shift perturbations.

The non-TcART chemical shifts occur in separated spectral regions characteristic of carbohydrates and aromatic rings. In the $\delta(^{13}C)$ range of the acquired ^{1}H-^{13}C spectrum (Fig. 5.6), these are 85-105 ppm (anomeric carbons) and >140 ppm (aromatic rings of DNA bases). Notably, the molecules creating these signals can be removed upon dialysis and their absence did not influence the majority of TcART signals. Both properties speak towards a minor affinity of the co-purified molecules to TcART. Thus, TcART likely binds some molecules, *e.g.*, nucleosides, during expression and carries them through the purification process. From the binding mode of NAD^+, it can be hypothesized that nucleosides can bind in the pocket designated for the adenosine moiety of NAD^+. There, the binding is facilitated by the R117 and W133 sidechains, as shown in Appendix Fig. A.13B on page 121. Indeed, a minor peak shift can be observed for the W133 signals in the ^{1}H-^{13}C spectrum after the removal of the unknown molecules by dialysis (Fig. 5.6). Nevertheless, no significant structural impact of the unknown molecules could be determined.

5.2.5. Implications of the unfolded state of TcHVR within TcB-TcC

We were able to examine the state of TcHVR within the TcB-TcC complex prior to injection by solid–state NMR. The absence of characteristic peaks in ^{1}H-^{15}N spectra indicated the insufficient formation of the hydrophobic core, which we attribute to a completely unfolded TcHVR. Therefore, the protein can be directly translocated without prior unfolding into the TcA channel and subsequently into the target cell. This process differs from that in homologue proteins like the anthrax toxin, which need to be actively unfolded for proper translocation [206, 207]. The undefined structure of TcHVR in the complex also explains why the toxic enzyme was neither resolved in cryo-EM nor X-ray structures of *P. luminescens* and *Y. entomophaga*. Interestingly, a similar trait could be identified for the homologous Rhs proteins from *Pseudomonas protegens* [208] and *Photorhabdus laumondii* [209]. Similarly to TccC3, which carries TcHVR C-terminal, the N-terminal part of Rhs encapsulates the C-terminal region carrying the toxin. Based on the overall homology, it seems likely that Rhs toxins are similarly unfolded when carried within, suggesting that this might be a more general property of these complex-forming toxins.

The Tc toxin complex from *Photorhabdus luminescens* was recently shown to be adjustable as a universal protein translocation system [12]. Knowing that the toxin is kept unfolded inside expands this knowledge further. The scope of proteins for injection by this mechanism is broadened to all unstructured proteins fitting the size requirements. In sync, the scope of possible treatment options by host delivery through this syringe-like system is expanded. Additionally, Rhs toxin complexes can now be considered additional tools for targeted injection.

5.3. The future of structure prediction and determination

The work presented in this thesis relied on different approaches for the determination of protein structures. In addition to established techniques such as NMR spectroscopy and cryo-EM, computational methods have been utilized. Among these, a novel structure prediction tool called AlphaFold [127] has been employed. Given the high degree of agreement between AlphaFold and experimental methods, it is likely that its predictions will significantly impact the future of structure determination.

Traditional approaches to protein structure determination require substantial effort for data collection and the successful conversion of data into a structure model. Many steps needed rely on manual analysis and can only be partially automated. The dominant method used in this thesis was NMR, for which proteins need to be isotopically enriched to reasonably detect nitrogen and carbon nuclei. Moreover, NMR experiments, particularly in the solid–state, necessitate manual optimization to obtain well-resolved data. Even after sample preparation and data acquisition, the assignment procedure, where the data is matched to the atoms of the protein, remains time-consuming. Here, the thorough analysis of NMR data was employed twice: for TasA, where intra- and inter-molecular contacts were directly compared to a structure model, and for TcART, where an ensemble of NMR structures was calculated. Structure calculation requires time for data evaluation and computational resources. In contrast, AlphaFold operates solely on the primary structure and provides a model within minutes. Given the significant differences in time and resources invested, the generation of AlphaFold models and their subsequent validation or falsification will become a more common approach in the future.

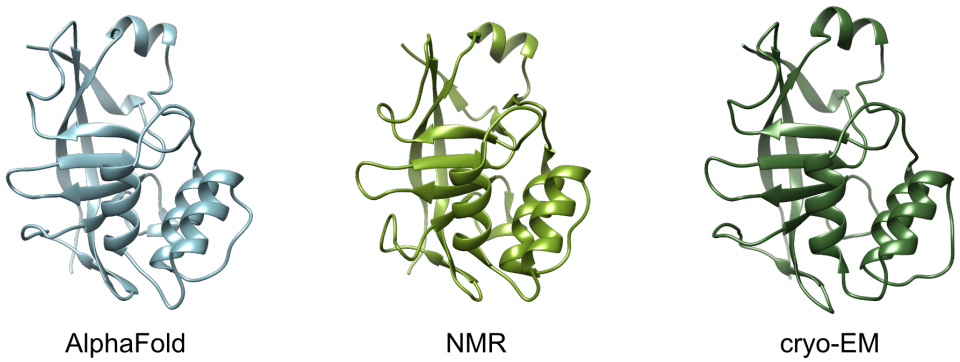

Fig. 5.7.: Structure of TcART as determined by different methods. TcART as predicted by AlphaFold (left) and as determined by NMR (middle, PDB: 7ZBQ) and cryo-EM in for the molecule in complex with F-actin (right, PDB: 7Z7H).

In this thesis, the validity of the AlphaFold model for filamentous TasA was demonstrated through the analysis of contacts obtained by NMR. Furthermore, the cryo-EM structure of TasA filaments [56] was also consistent with the prediction (Appendix Fig. A.5 on page 113). For TcART, AlphaFold similarly predicted a model whose overall fold agrees

with our calculated structures (Fig. 5.7). Analyzing consistent and diverging traits highlights the strengths and limitations of AlphaFold. While the overall composition of folded proteins is usually well predicted by AlphaFold, it sometimes fails to capture details essential for the protein function. In the case of TcART, this is evident in the region bordering the binding pocket, including the loop 198-208. Our NMR data revealed the presence of flexibility in this loop region. Conversely, in the AlphaFold models, the loop consistently appeared in the same position, although predicted with lower confidence. This output indicates that while AlphaFold acknowledges certain areas where the prediction might be inaccurate, it does not reflect alternative conformations. This issue is intertwined with the limiting factor that AlphaFold cannot (yet) deduce conformational changes that enable protein functionality. For TcART, these are, *e.g.*, movement of the loop upon binding and rearrangement of the K185-E265 salt bridge. Moreover, AlphaFold, in its current state, is not able to predict intrinsically disordered proteins and the impact of point mutations.

All things considered, AlphaFold will still assume a significant role in the future of protein structure determination by supplying an initial model that can be refined further. Still, it is crucial to recognize that the most critical details (*e.g.*, conformational changes upon binding, protein flexibility) will still likely be derived through traditional methods. Integrating AlphaFold predictions with experimental approaches will be vital to obtaining quick, comprehensive, and accurate structural insights. For the interplay of NMR and AlphaFold, multiple approaches based on easily attainable NMR data are already available [140, 141], illustrating the power of this comprehensive approach.

5.4. A perspective for the future of NMR

In recent years, the emergence of ultra-fast spinning probes (>100 kHz) has made ^{1}H-detection in solid–state NMR a valuable approach. As a result, numerous experiments previously limited to solution NMR can now be routinely employed for insoluble molecules.

In this work, a variety of solid–state experiments have been conducted. They encompass more traditional approaches like ^{13}C-detection with varying labeling schemes, ^{1}H-detection on perdeuterated and back-exchanged proteins as well as the detection of ^{1}H in fully protonated samples. Notably, the TccC3 toxin project demonstrated that ^{1}H-^{15}N correlations recorded by solid–state and solution NMR of these fully protonated proteins can be used for comparison. Overall, NMR now holds the potential to be applied to a much wider range of samples.

In the forthcoming years and decades, solid–state NMR methodology will continue to advance. Current efforts focus on developing even smaller diameter rotors that can spin at faster speeds [100, 210]. As a result, narrower NMR linewidths are acquired, but the transfer of the sample into the rotor becomes more complicated. Additionally, the implementation of optimal control CP techniques presents an opportunity to enhance existing NMR experiments by selectively manipulating magnetization subsets, thereby improving the efficiency of transfer methods through precise adjustments of pulse phase and intensity [211]. Of course, the general development of NMR hardware advances as well, making stronger magnetic fields with higher spin polarization and resolution available [212].

Furthermore, sedimentation NMR has emerged as a potential application for investigating diverse molecules in the solid–state [213]. This technique involves concentrating soluble molecules by centrifuging them into an NMR rotor. As a result, high concentrations in the rotor are achieved, which further increase along the rotor wall during measurement due to the centrifugal forces caused by MAS (Tab. 1.2 on page 19). This approach enables efficient measurement of soluble proteins by solid–state NMR. Considering these advancements, it can be expected that solid-state NMR methods will undergo further refinement in the future, expanding their range of applications even more and enabling the measurement of previously inaccessible samples.

A. Appendix

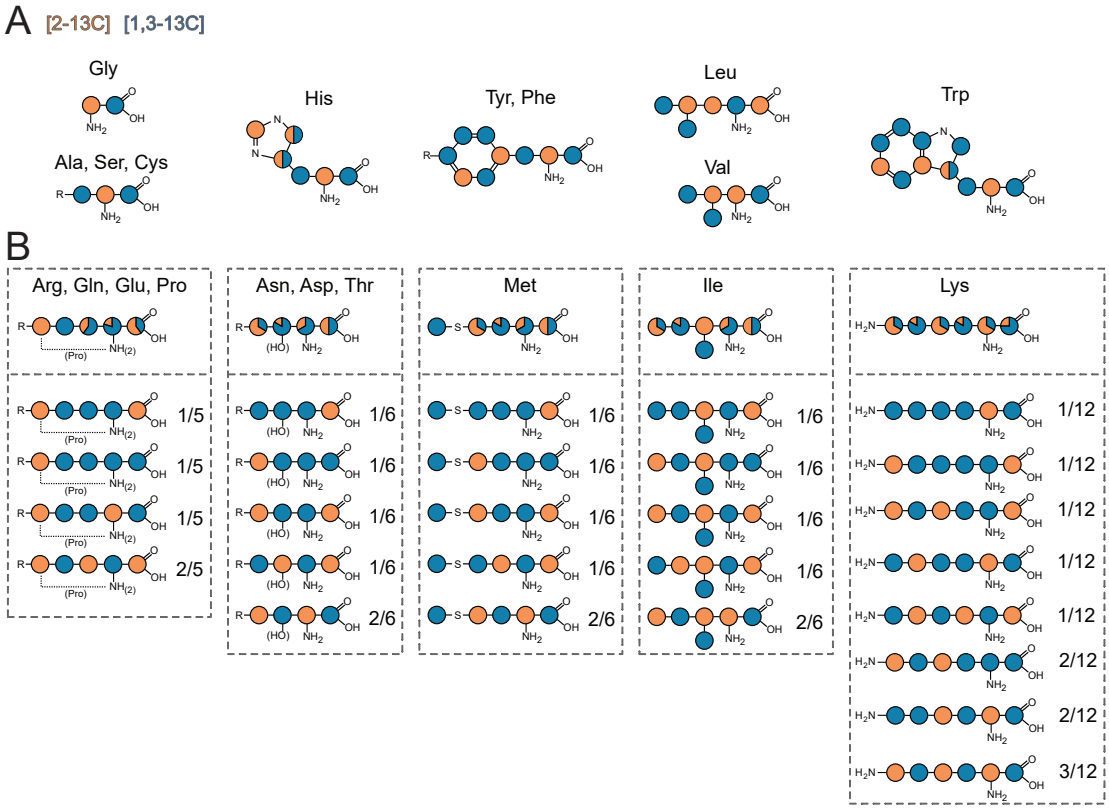

Fig. A.1.: Isotopomers after sparse ^{13}C-labeling. ^{13}C enrichment for all amino acids as observed for the α-spectrin SH3 domain when expressed in *E. coli* BL21(DE3). Sites that incorporate ^{13}C when the bacteria are supplied during growth with [2-^{13}C]-glycerol are marked blue, those enriched upon [1,3-^{13}C]-glycerol addition are orange. Depending on their metabolic pathways, amino acids can either have distinct or distributed isotope enrichment at a given site. Amino acids with a defined pattern in the aliphatic sidechain are shown in (**A**), those with multiple schemes are presented in (**B**). Modified from [214].

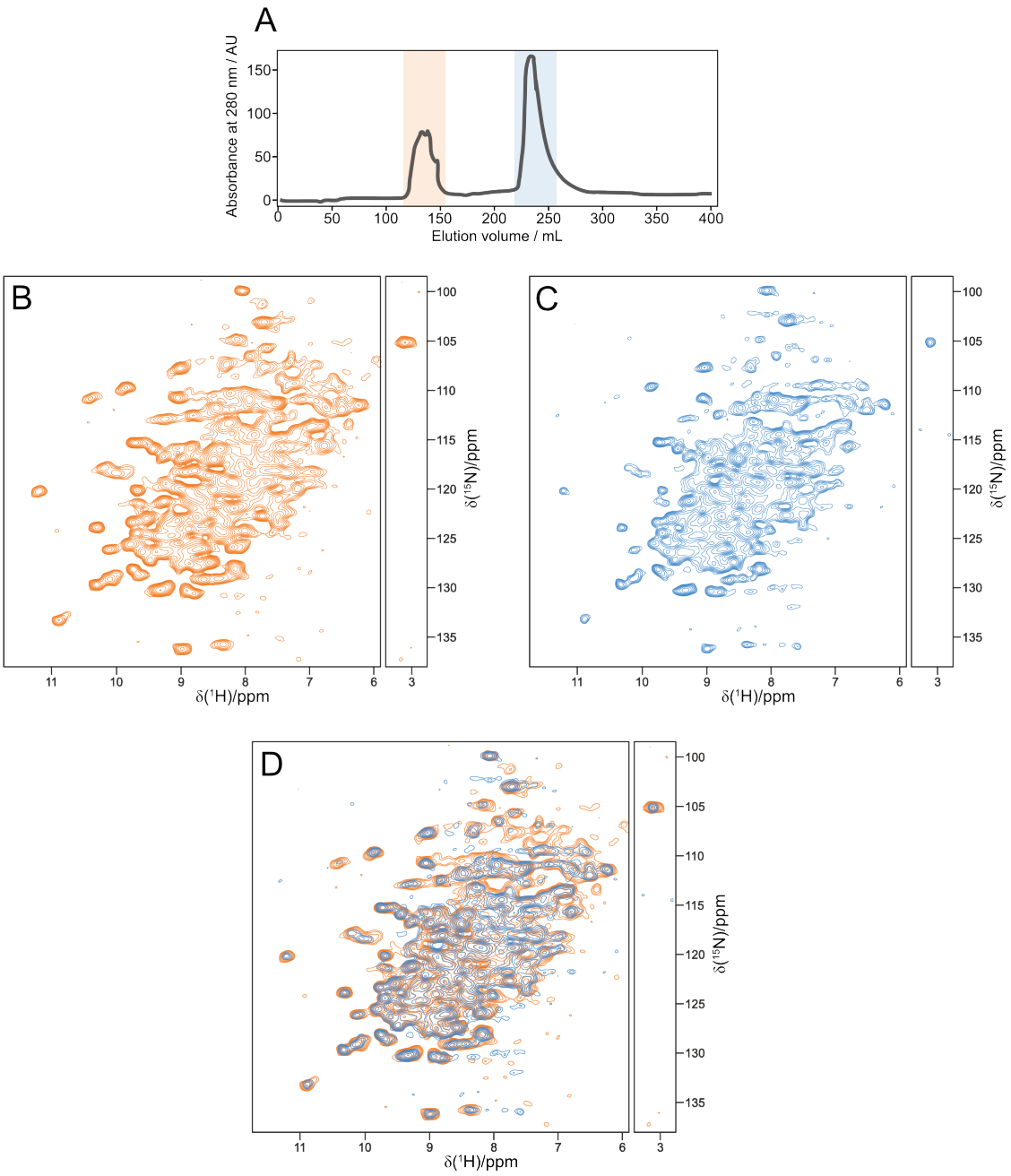

Fig. A.2.: TasA converges to the same state when added to *ΔtasA B. subtilis*. (**A**) Purification of recombinant TasA by size exclusion chromatography yields two elution maxima (data as published in [177]). The plot shows the amount of protein eluted (measured as absorbance at 280 nm) plotted against the elution volume in mL. Elution at a low volume of running buffer typically indicates a high-molecular species whereas late elution at a high volume of running buffer usually indicates low-molecular species (*i.e.*, monomers). Addition of either species to *ΔtasA B. subtilis* biofilm, and acquisition of an ^{1}H-^{15}N correlation of the pellicle after washing, yields the same spectrum. ^{1}H-^{15}N correlations are shown for supplementation of *ΔtasA B. subtilis* with ^{2}H,^{13}C,^{15}N-TasA from an early (**B**) and late elution volume (**C**). (**D**) Superposition of **A** and **B**.

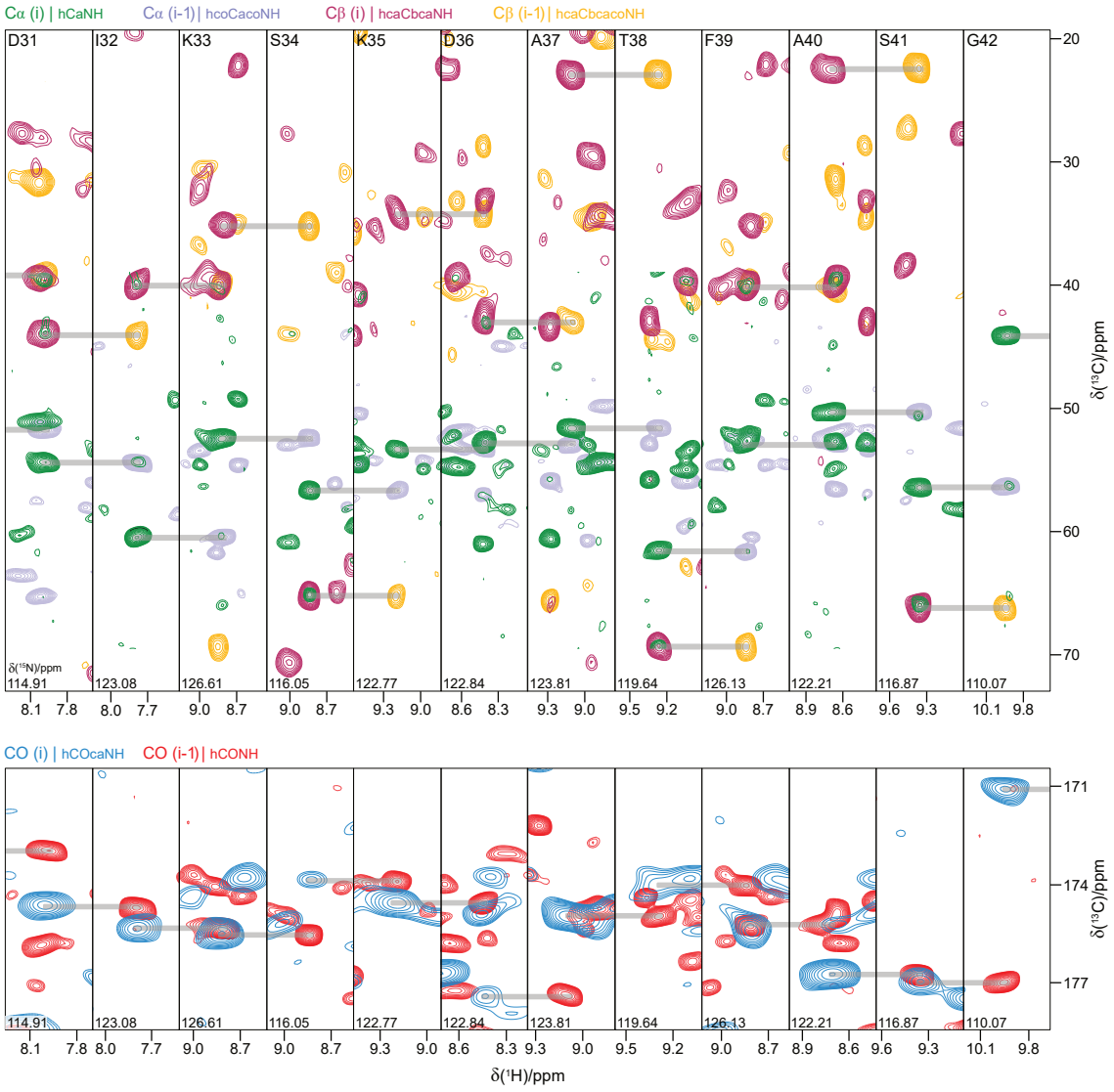

Fig. A.3.: Strip plots of NMR assignment spectra for residues 31-42. Contour lines are shown in green (hCaNH), light blue (hcoCAcoNH), raspberry rose (hcaCbcaNH) and yellow (hcaCbcacoNH) for aliphatic carbons on the top and in blue (hCOcaNH) and red (hCONH) for carbonyls on the bottom. The annotation at the top of each strip denotes whether atoms from the amino acid itself (i) or the preceding one (i-1) are measured. Connecting sequential links are highlighted with gray bars.

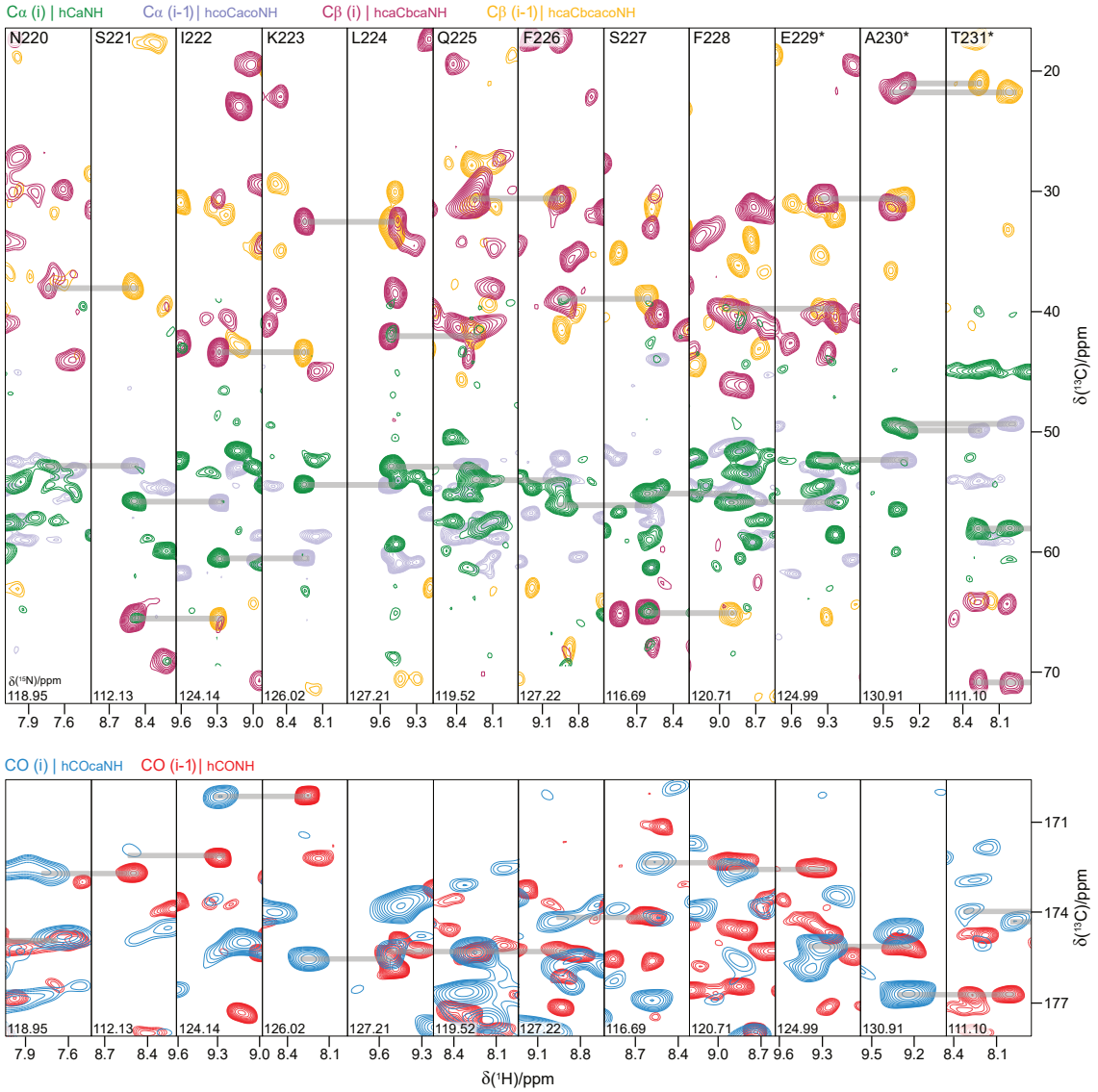

Fig. A.4.: Strip plots of NMR assignment spectra for residues 220-231. Contour lines are shown in green (hCaNH), light blue (hcoCAcoNH), raspberry rose (hcaCbcaNH) and yellow (hcaCbcacoNH) for aliphatic carbons on the top and in blue (hCOcaNH) and red (hCONH) for carbonyls on the bottom. The annotation at the top of each strip denotes whether atoms from the amino acid itself (i) or the preceding one (i-1) are measured. Connecting sequential links are highlighted with gray bars. Doubling of peaks is observed for residues 229, 230, and 231 which are marked with an asterisk. 231 shows clearly separated peaks in the ^{1}H and ^{13}C dimensions. For the other signals, the effect is less pronounced and manifests as line broadening in ^{1}H and ^{13}C (for 230) and in ^{1}H (for 229).

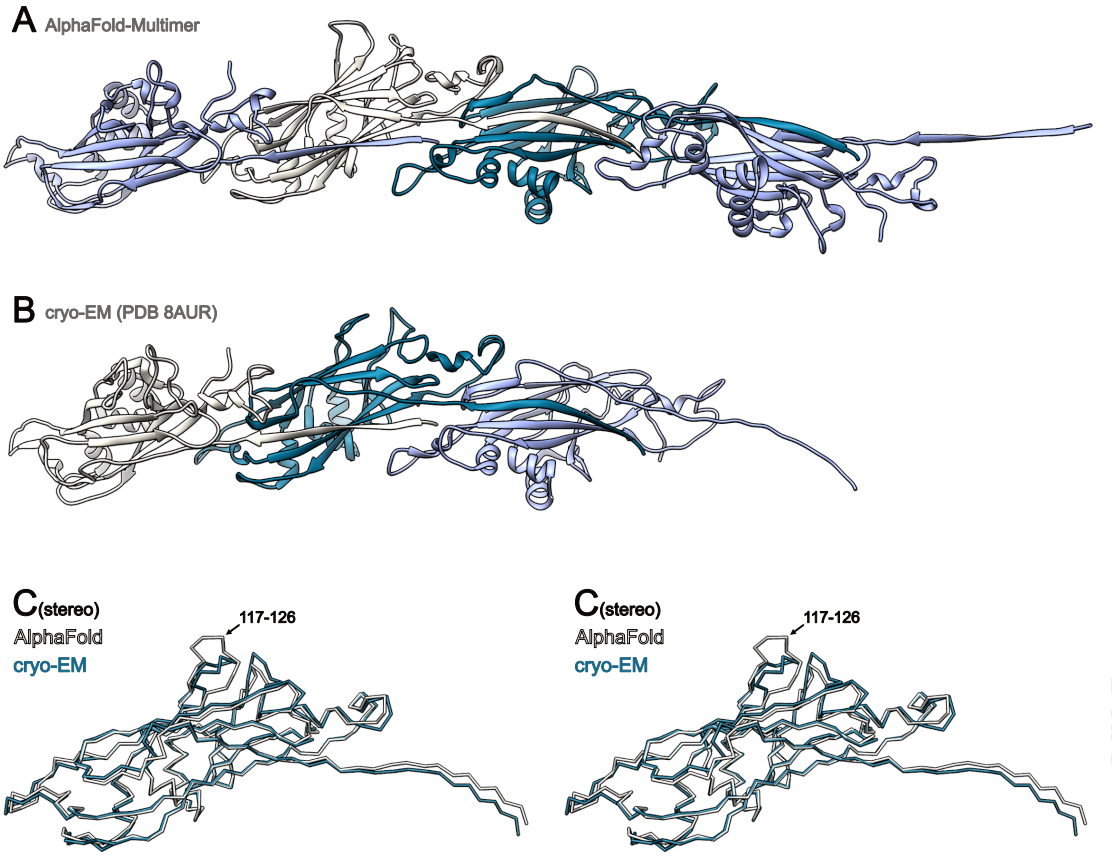

Fig. A.5.: Comparison of TasA filament models by AlphaFold and cryo-EM. (**A**) Tetrameric TasA filament as predicted by AlphaFold. (**B**) Trimeric TasA filament solved by cryo-EM (PDB 8AUR) at an average resolution of 3.5 Å. (**C**) Superposition of the backbone traces from molecules extracted from AlphaFold (white) and cryo-EM (blue) filaments shown in stereoscopic view. The diverging loop 117-126 is indicated.

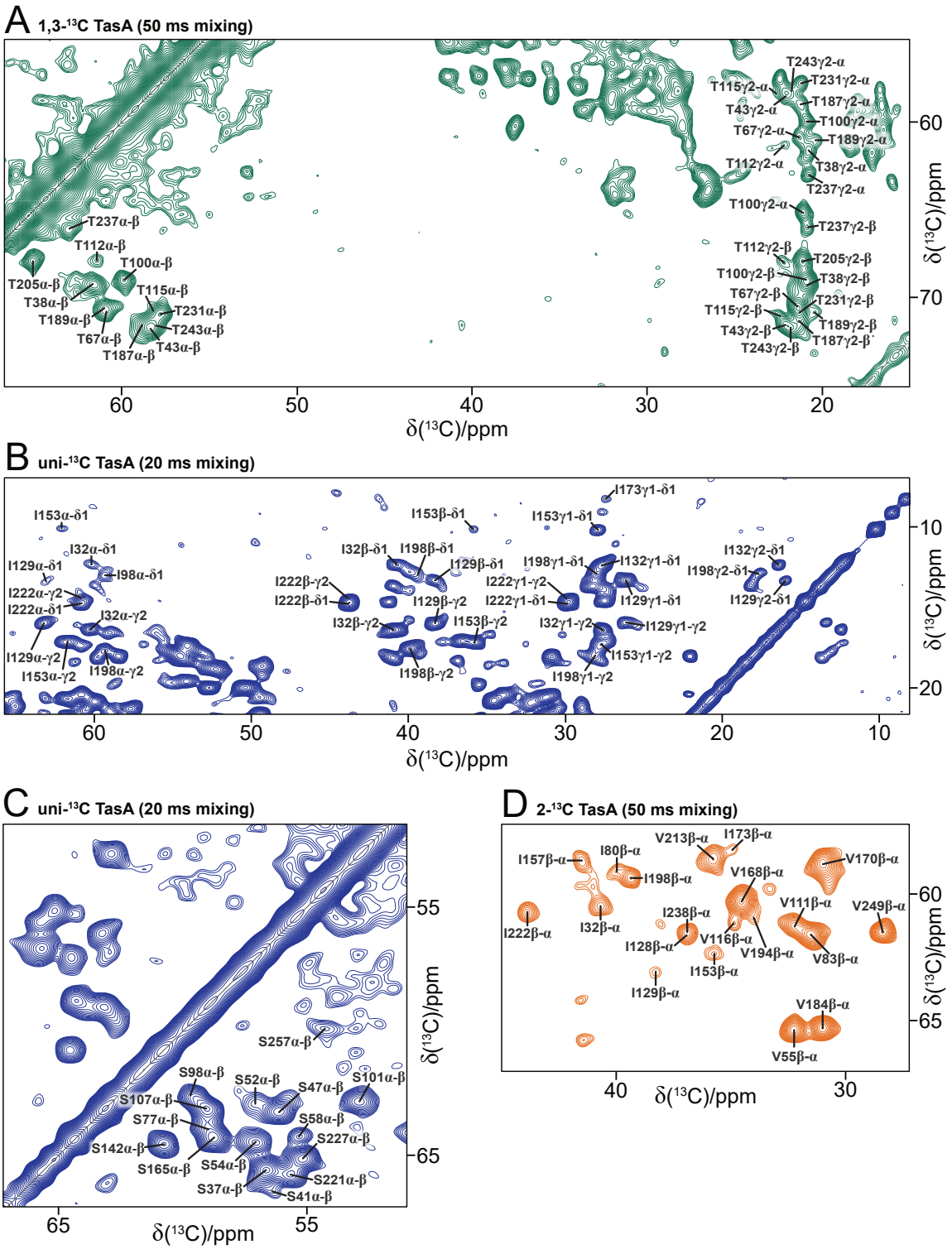

Fig. A.6.: Sidechain assignment in ^{13}C-^{13}C spectra. (**A**) Threonine assignments plotted on a ^{13}C-^{13}C spectrum acquired on a sample with a [1,3-^{13}C]-glycerol labeling pattern at a mixing time of 50 ms. Isoleucine (**B**) and serine (**C**) sidechain assignments are shown on a spectrum recorded on uniformly ^{13}C-labeled protein at a mixing time of 20 ms. (**D**) Isoleucine and valine sidechain assignments plotted on a spectrum recorded on a sample prepared with a [2-^{13}C]-glycerol labeling pattern at a mixing time of 50 ms.


29 42 62
1 TasA B. subtilis (strain 168)|P54507 MGMKKKLSLG.VASAALGLALVGGGTWAAFNDIKS.KDATFASGTLDLS.....AKENSASVN..LSNLKPGD.....
2 CalY1 B. cereus (clone B905)|A0A0A3VYJ9 MSLKKKLGMG.VASAALGLALIGGGTFAFFSDKEV.SNNTFAAGTLDLTLN.....PKTLVDIKDLKPGD.....
3 CalY2 B. cereus (clone B905)|A0A0E3SV09 MTLKKKLGMG.ITSAVLGAALVGGGTFAFFSDKEV.SNNTFAAGTLDLALN.....PSTVVNVSNLKPGD.....
4 Thermococcus gorgonarius|A0A2Z2M7U5 MKSIAVA.LLLIGLVFAGAGYGTWAIYQDTETSEGNSIELGTLDLVLTDG.....VPVANGQWFIPNAVPGS.....
5 Bacillus mycoides|C2Q5K3 MIWEEQKMSLKKKLGVG.VVSAALGLSLIGGGTYAYFSDQVV.TNNSFAAGTLDLAMQ.....PTTSLNLENLKPGD.....
6 Bacillus thuringiensis|K0G0I4 MSLKQKVGMG.VVTATLGLSLTCGGVFAYFSDSET.SSNSFQAGTLDLSIN.....PSVIVDVKDLKPGD.....
7 Haloferax gibbonsii|A0A871BKP0 MPEKVFKLTRRRALGGLLTIGAGSAAAGVGTFALFNDTESSTENTLQAGTLSLADG.....ASSASFSATGLAPGDE.....
8 Paenibacillus graminis|A0A089M0B1 MGIKKTLGFG.IASAALGLSLIGGGTFAYFSDTAT.STATFAAGTLDLNSD.....PSVIVDIQNLKPGD.....
9 Aquisalibacillus elongatus|A0A3N5B9I6 MDIKKKLGIG.VLTGALGLSLVGGGTWAAFNDVEE.TTNVFAAGVLDLE.....IGEESTMDFSLDNLKPGD.....
10 Bacillus oleivorans|A0A285CS81 MSLKKKMSMG.ILSGALGVSLIGGGTWAAFNDVEV.VNNTLAAGTLDLV.....VGENTTMNFTLSNLKPGD.....
11 Aureibacillus halotolerans|A0A4R6UFZ6 MGLKRKLATS.VMSASLGLSLVGGGTWAAFNDVEN.IENTFAAGTLDLAVI.....TKDKSNSVFK.LLNMKPGD.....
12 Bacillus sp. 1NLA3E|N0ATI4 MSIKRKLGLG.AASAALGLSLIGGGTFAYFSDAAT.LHNGFAAGTLNVELQK.....AWDFPLNFDLSNIQPGD.....
13 Blautia luti|A0A564VQ52 MMKNRKLVKGI.ALAAVIT.ALAAGGTAAYLTDFET.ATNSFTVGKVDIDLD.....EPNWKPEDNTDLVPTQ.....
14 Bacillus oleivorans|A0A285CTH3 MSLKKKLTMG.ALSATLGMSLVGAGTWAAFNDIEE.VNASLAAGSLDLV.....VDEYNGPVNFNISNLKPGD.....
15 Paenibacillus taihuensis|A0A3D9SB88 MSIKAKLGMG.IITASLGLSLIGGGTYAYFNDAGSSIQNSFASGTINLNATK.....AWDFPANFDLVNMKPGD.....
16 Neobacillus mesonae|A0A3T0I131 MGIKKKLGLG.MASAALGLSLIGGGTFAYFSDEAT.IHNGFAAGTLNLEVGK....YPGTQWPVNFDLSNLRPGD.....
17 Methanosarcina barkeri|A0A0G3CB53 MANKRVLVS.ILALGIAVTLAGAGTWAAFQDTETSTGNTFSSGTLDMKLSNDGSTYTDGVTATWTSPANFKPGD.....
18 Lachnospiraceae bacterium|A0A7G9GBU6 MKRG..TKTI.MLAAALAGVLIVGGVSAYFTDADT.ATNTFTVGKISLDLQ.....EPSWNPDQVKDLTPNQ.....
19 Virgibacillus salinus|A0A1H1GBR8 MNIKKKMGTV.ILAGALGLSLVGSGTFAAFNDVEK.IDNTFTAGTLDLV.....LTGQEKTSFEFSNLEPGE.....
20 Sporosarcina sp. resist|A0A7G8X4J2 MGIKQKLAMG.IATGALAVSMIGGGTYAYFNDVET.STNTFAAGTLDLSLK.....PETIIDVKNIKPGD.....
21 Neobac. massiliamazoniensis|A0A0U1NYG4 MKIKTKLGMV.AASGALGLALISGGTYAFFNSKVE.ATSTFEMGTLDLSVN.....PSTIVSVKNLLPGD.....
22 Piscibacillus halophilus|A0A1H9E238 MSIKKKLGMG.VLSAALGISLIGGGTFAYFSDTEE.TNNTFAAGTLDLSVD.....PTTVIDVDNIKPGD.....
23 Hungatella hathewayi|A0A174ECC6 MRRNYRKASQIVSLGALAIVLIIGGTLAYFNQDLS.ASNVFLTGKYDTDLH.....EEFKPPGDWQPGV.....
24 Alteribac. persepolensis|A0A1G8JRA6 MKKTKKALIATTLAGAVAVGAG.FGTFSWFTDSQS.TDGEITAGTLTLNDLS.....EYTVFSEGNMQPIDPEITPDLI
25 Bacillus velezensis|A0A8F9SKQ0 MGMKKKLSLG.VASAALGLALVGGGTWAAFNDVKS.TDATFASGTLDLS.....AKEQSANVN..LSNLKPGD.....
26 Psychrobacillus sp. OK028|A0A1G9SR36 MSIKQKLAMG.ILTGGLAISMIGGGTYAYFNDVEV.NNSSFAAGTLDINVAG.....NDADNAIINVTNIKPGD.....
27 Bacillus pumilus|A0A8G1J983 MAKKRSIRLG.VLSGALGLALIGGGTWAAFNDVEK.ANAVYSTGELDLS.....AKENSGAIN..LANLKPGD.....
28 Haloechinothrix alba|A0A238Y1I1 MRIEKKTAVG.VAAAAAAALAITGGTFATFSDTEEAPNPSITAGTLDLTTG.....GEEQSESFAIEDVAPGYSSEE...A
29 Clostridium cf. saccharolyticum|D6DLJ6 MKYYKQLG.IILAGGVLAVGISATWAYFSDRND.IANPFHTSNSAIDMI.....ENYNPAETVLPGE.....
30 Salimicrobium halophilum|A0A1G8S890 MSVKKKLAVG.VMTGALGLSLVGGGTLAAYNDVAK.VNNHFASGELNLEV.....GDSHNKAISFDLTNMKPGD.....
31 Bacterium BMS3Abin01|A0A2H6K5U5 MNKRILIS.VLAVSLLAMLAGAGTYAYFSDSETSSGNSFAAGTLDLKVADNDE.GYGDGVSQTWVIGNMVPGVS.....
74 76
1 KLTKDFQFENNGSLA.IKEVLMALNYGD..FKA.....NGGSNTSPEDFLSQFEVTLLTVGKEGGNGYPKNIILDDANLKDLYLMSAKN..DAAAAEKIKKQIDPKFLNASGKVNVATIDGKTAP
2 SVKKEFLLQNSGSLT.IKDVKLATK.YTVKDAKGD.....NAGEDFGKHVKVKFLWNWDKQSE.....PVYETTLADLQKVDPDLLAK.DIFAPEWG.....
3 TIEKEFKLENKGSLD.IKKVLLKTD.YNVEDVKKD.....NK.DDFGKHIKVTFLKNVDKHET.....IVKQTTLDKLKGDTLTAVDN.DLSAWFWD.....
4 LYNGDIKIINNGTVD.ADHVEIAIELVCWEDDDGNIDNGRAWGPEPDEVNGVGDLPQEIKVNMQYREEWIG.....PSATLQLENGAALGDVVYNVYNLPAPPRVNG
5 KILKKFNLTNSGTLD.IKDIMMKID.YTVNDLKQN.....NTTEDFGKHIKVQFLLDWDPAKSP.....VYETTLAELKSQSPEIASKKVFYSKWTET.....
6 FIEKNFTLENKGSLD.IAKVALETS.YEVTDAKLN.....NDGEDLGDHIIVKFLVNDGKPSNPNDDHE.....ILWETKLSKLKTMKPEDVSTSLERHNLID.....
7 TDTFEVDLENDGTLDGSLDIDLALTENDSDATAGDD.....GDADASVADLAGVLDVTTFEVG.....SLDVLADRFGGSATLADVVANNAGED..DDS
8 TVTKTFKLKNDGTLD.IGSILLKTT.TEVTDTLGD.....NNGADLSKFIKVKFLKNNDKGTIISP.....EVVVYETTLDALKGQTPDLVNN.TGWDQILT.....
9 HWTETLVLRNNGTLD.INQILT.SIEVTG.WEDVDALDLNSKVYDGAGNNKLSDFLKQFEVTVT.....DGDGNV.FFDGTLNQLHGQSNID.ITGT
10 YFNKSLVLTNGGSLD.INQILVHANKLAG.WTDKDVLDLNTEIGANSGNNSEDDFLSQFKVTITKL.....AVDPANNVDVFDGTLADLVAGISVDELTDT
11 SKTAEFELKNDGSLA.IKEVLLDITVPEDGFKNGEN...EYVNEHGGEENEAIDFLEQFQVEILRIDEETG.....YDFTLIEEKD..GVTLADFAKKDLPAEIETVDGAINLAPIH.TERP
12 SFERQFVLKNSGSLA.IEDVFMGVT.ANVTNPLGTG.....GTSDQFLDALNVKYFVDATPPYIDPDAPDA.....GYLLINSQDISLREAIAGDFAG.KIKADYIKNG.KLNL
13 VIRKDPYVANKGVNE.AFVYLEVSVPVRNVITAAK.....DGTRNALAKTELFSFTKNKDWTQL.....ERTEVGQNMVYTYAYN.....HILKPGTKTTT
14 TMTRYIKLDNAGTLA.IKDVLLSIDSVN..FTDYLP...VGADSDIYGANTADEYLSQFKVTLLKTGIEGADEEIISS.ADNITLADIYLATNPDQLDPVAIAKLSAAIG.NGHWVDGHVNLASAA...QN
15 VTERQFVLQNIGSLD.IANTLMSFT.....SDA.....GSDAGLLNTLTISYFVSVGNN.....N.....EYLLMNSQDITLSDAIAGNYAG.KIKAEYLTADGKLNL
16 TVERTFDLKNVGSLA.IEDTYLDFSKVSVENPLNTG.....ATDDDFLSALSVTYFYETVVN..GQYVPES.....MLVNSQKISLKDAIAGNYVG.KIQPRFLKGT.KLNL
17 TFGNELRFTNVGSVN.INHLYMYPEDLTNSGGTGG.....IKLSDKIIITKVTGHITRDGSNLVWG.....NRADFIGSVVGNRDGTLTLTEFCSAKFVAYCLSPIIFG
18 VVQKDPQIKNDGIND.EFVFLEVTVPYKNVVTVND.....AGVKQTAADTELVKYTVNAGWTELG.....SGEKDTVAETVLHRYVYGTDQACTALAKDQVTPA
19 SFTKTITLKNNGSLA.IKEILMSSMAIEGSWIDKT....NQKLPDN.GMNTVDDFLEQFNVIVN.....QGETEVFNGTLADLNSG...KVIMGS
20 WMNRDFTLKNDGTLD.ISKVLLKTE.YTVTDKNGD.....NVENFGSHIRVNFMWNHDKSITGPWTGL.....DNVVYQTTLAELQGMTPDAVAN.KVFVPYLE.....
21 YMTRSFNLKNTGTVD.IAKVLLKTE.YTVTNKDGAP.....ANTDDLGKHIRVNFLFNQDKAALGTWLPE.....DQVVFQTTLYDLQNMTPDAVEN.RIFHDHLE.....
22 VMVREFELINGGTLD.ISSIDLSTA.YEVIDAEGN.....NTDDFGKHIKVLFLENADKTGDGWIIGDY.....NDVIAETTLYDLQNQTPDAVENVQGFFTHLLGL...NG
23 EIPKKVDIKNKGNVDVVAVARMTESCVRKEDVFITTYETVD.GRKTERQEKVASKGDVLPLQFEDSDGTQQEFALKNFG.....SDVVPYAEAKSPEEYRNKWVVYTYDENSKAYYFIYMGLIEGGNTSPG
24 TEGENITVENTGSLEMVPRLQLDVEIKDDKGNDAS.....ETLSDEYMIKMTFDR.....NDMNWSEEISADKVIDNRGFVDPDWFPSD
25 KLTKDFEFRNNGSLA.IKEVLMALNFTD..FKG.....AKKGNESAEDFLNQFEITVLTVGKEGGNGYPKNIILKAASLKDLYLMSTKQ..DKAAAEAISKHIDPKFLSESGKVNVATINGKTAP
26 TMLRTFKLNNTGSLD.VSKVLLTSK.YTVVDAKGSD.....NGGADFGEHIKVKFLINKDKRTE.....VVHETTLKDLS..KTDVTDR.DILGWILDG.....
27 RIKKEFHFENKGSLA.INQVLMSLDYSQ..FQDG.....ASAKNGGKNTAEEFLSQFQVSVLTVGAEGGNGYPKNIILDHANLLDLHQLTSKQ..DQTAFEKLRHAVDEKFLHESGKINVATVDGTVAP
28 GEGKTIQYENEGSIDGDLTLTVDYSTDEVGCNDPE.....IDAEKAFYDDESIEDCGN.....DTELDEQLYAEIIGPDGTVAGEDLVGHLKG
29 TVEKEPYFVNQGSTDLVLRLKMDTYWIDQNGNRKD.....DLGGAYIFGGLQPEEDD.....YVTLYLPDGWDDDWVQIEDYYYYKHILPG
30 NVQRIFTLNNIGSLA.IKDVLMEGSAGAF.....SDLDGDSNKNGFLKQFEIDFMQVDGESNQWEP.....REDVVMSGKT...LTLHDVVHGNTTSKVNSSHLVGNKINLASVAAG
31 SVTNSMSLQNAGSVN.ADHIELGFSYSIDETSNPVES.....DTNPASAPGDMARMIEITAMTYDGVN.....FATSFVDANGNGWFDLEDLSLPPYSSSGG.YLDN
215 218 232
1 EYDGVPKTPT.DFDQVQMEIQFKDDKTKDE...KGLMVQNKYQGNSIKLQFSFEATQWNGLTIK.KDHTDKDGYVKENEKAHSEDKN
2 EKGGLAAGT...EDYLWVQFEFVDD.....GKDQNIFQGDTLNLEWTFNANQEAGEEK
3 EKG.ISAGK...SDKFKVKFEFVDN.....GKDQNQFQGDKLQLNWTFDAQQTAGEER
4 DVNDALWNKK..YGQFDMYLWIEDT.....GQPQNYWQGDVCEIIVHVALAQDSSQQILSPYGFTEYPVPVLPE
5 ..GGLKPGK...MDWFWIKFVFEDN.....GTDQNVFQGDSIALKMEFQANQTEGQER
6 ...GITSGE...TDYLHVLFSFEDN.....DQDQNKFQGDSLQLNWTFHAEQTKGVEK
7 NVRDIALASG.ETKTFSFALEMQES.....AGNEFQADGLDIQFDFELNQRDSQ
8 EADGIKAQS...SDSFTVKFEFVDT.....GEAQNYFQKDKLKLTWNFEAKQGPKKDY
9 DTNTVGLSVN.DNMTYQVDFKFKNVDETFDG..SRFQVQNKYQDEQAEINVKFEATQMPGEEQS.....NGWAN
10 DATTVGLASG.AQQIYNVYVEFVEDSTTFPG..SRLHVQNKYQDEMSNLEFVFEATQMPGEDRS.....ND
11 EYTGLPVNPL.DDEVVKIKIVFVEDDTIDDS.VPGEYLQNRYQGDEIELNFGLEATQWGGLDIQ.DGDVAENGTIERNKNANSETPKE
12 TPLGINSSE...TNRYRIIIEFKDS.....GQPQNNLQGMVAKVAFNFDARQVMAPKYHDQNAPNGNLTGNGIIG..DRATTDSTSNVANQPAVVDAPAWQD
13 LFDTVTFANIIEGQ.LDTQQIDMPVRAYAIQ.ATNTGDDKTTVLEQATAAYQKYVNQ..NKGQAGGVTK
14 QWTGLPVNPA.DFDTVEMSIEFVEYSIRDA...RGVEEQNKYQGDTADITFTLEARQWEGQDVS.....DEEGYVESNEQARNGQ
15 TPDGIKAGAGVDDHRFRMMVKFPET.....NVPQNDLQGKQAHLTFNFDARQGAG.TVNNQN.....IGN.....ANPATIANPITNAN
16 SPDGIDAGS...YGRYRIMISFPET.....NVAQNKLQGMVAKVDFKLDARQVMGNKSHNAQGPNGTITGNGKQGLTDKTVTPNGTGDLVDPDTTDKAAWEDTNQ
17 IFPPILKAGDHKDWGLIIEGKFADD.....ADNHYQGSSCTFNMRCRGSQNSPTEGFEEITGNGAD
18 LFDSITVANVVEDQTLEQTTQNIQVEAFGIQ.ANDIGADDKT...APSDVWKVLVNQGPSKDKTGEDTKTDIVGG
19 EGG..ALESG.ANMSIDIEISFVNNVKKYPN..SLFMEQNKFQGESINLDLIFEATQMSGEDRS
20 EKGGLKAGN...SDTLSVQFEFVEN.....DKDQNEFQGDSLELKWTFDAQQTKG.EKK
21 EIGGLKKGD...KDKLFVQFEFVDN.....GEDQNQFQGDSLALKWTFEGTQPVGKERKN
22 EGDGLEAGT...SDQMYVGFEFVDN.....GEDQNEFQGDALELTWTFEAHQTEG.ELR
23 LLESVTMNPRAQATVTHTKLVSDYDKENKKDRYTFSYDTNSFGYDSAEYRLNITAKTVQATKDAVDEVLTGNREMVPENFDKLAASLKEMCR
24 AGFKYPLGPD...ESFDVQFDVYLN.....KDAGNEYQGSSLTGTLNVEGGQSVDGAEFEDGTETDGSEG
25 EYDGVPKTPA.DYDQVRMEIQFKNDTAKTA...DGLSVQNKFQGNAISLQFSFEATQWNGLTIT.KDHTDKDGYVKENEKAHSEDKN
26 EPDGLKAGT...SDTLSVMFEFVEN.....KKDQNVFQGDSLKLEWTFDAKQTSGSAR
27 EYDGIPKNPF.DYDKMEMIIEFVNDQTKDK...AGHYIQNKYQGDAIQIDLSFEATQWNGLTINPKKHTDEKGYVKENEKAHSEDKK
28 ETVDAGELGGGESRSYTVHFSVPDS.....AGNEIQSDKVTLEITADLEQNETQAD
29 KEEGVQNTTSNFLDHLTLSPEVSND.....EHAADYSGCKFIVDFHAEAVPADELAIEVADWDLIQTDAGGLLDWAKLLPES
30 PDRGIPLDPQ.DSDDVFIQITFKDDQTKLG...NGEYAQNKYMNASVDFYFDLEATQWDGVHVD...SDDKNGAINNEQANTPDPVTDGSGGTHSNDVTDND
31 LLAPLPVQVG..SQSLNMSLFFRPE.....AGNDIQGDILTMSVTSTLNQDASQ

Fig. A.7.: (from previous page) **Alignment of TasA with homologue proteins.** Sequence alignment of *Bacillus subtilis* TasA, *Bacillus cereus* CalY1/CalY2, and close primary sequences from organisms with varying evolutionary distance generated by ClustalW. *Bacillicae*, *Chlostridia*, *Haloferacaceae*, and *Archea* are covered. Residues conserved in more than 90 % of the sequences are colored according to the Taylor coloring scheme provided by strap [152]. A description of the workflow can be found in section 3.1.8 on page 41. The conserved last alanine of the signal peptide (residue 27 in TasA) is shown in bold.

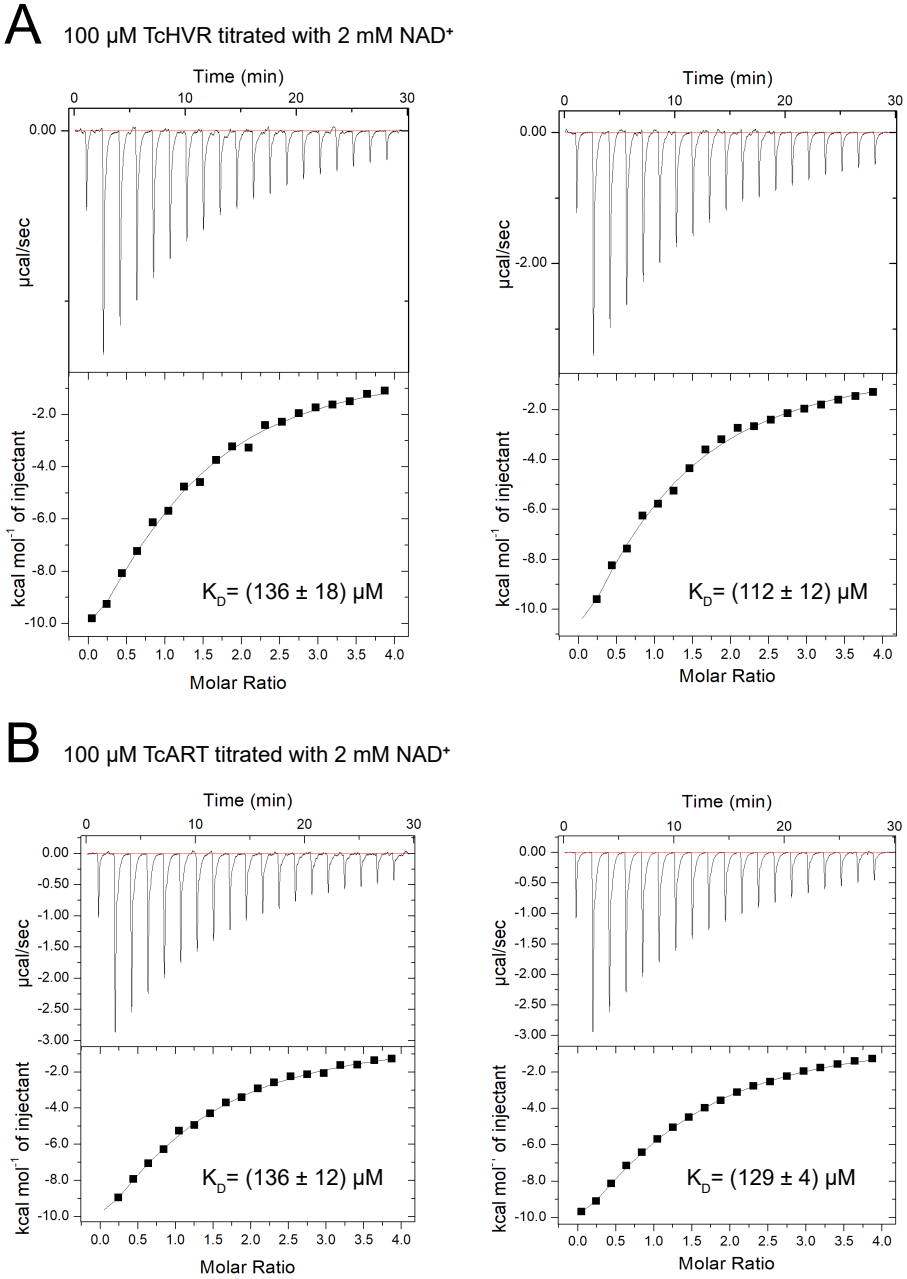

Fig. A.8.: Ligand binding affinities of TcHVR and TcART. ITC measurements in duplicates for titrating 2 mM NAD^+ to 100 µM of either TcHVR (**A**) or TcART (**B**). Experiments were conducted by Daniel Roderer at the Max Planck Institute of Molecular Physiology in Dortmund.

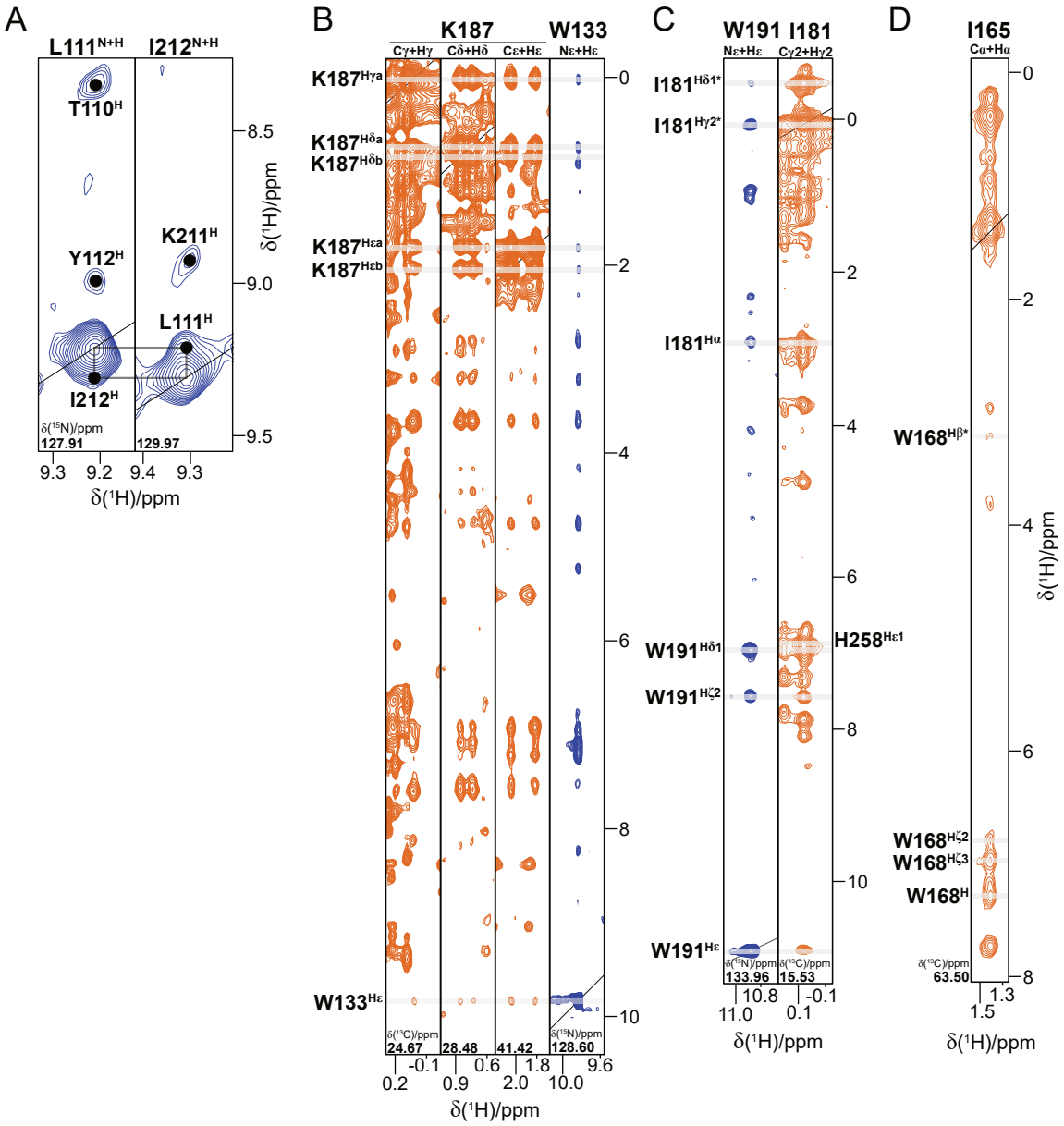

Fig. A.9.: Exemplary through-space NOESY cross-peaks in TcART spectra. (**A**) Strong peaks observed for cross-strand contacts of Y112 and I212 as characteristic for anti-parallel β-sheets. Further contacts, displaying distinguished chemical shifts due to interaction with aromatic sidechains, are present between (**B**) K187 and W133, (**C**) W191 and I181 as well as (**D**) I165 and W168. Full assignments are available in the BMRB under accession code 34717.

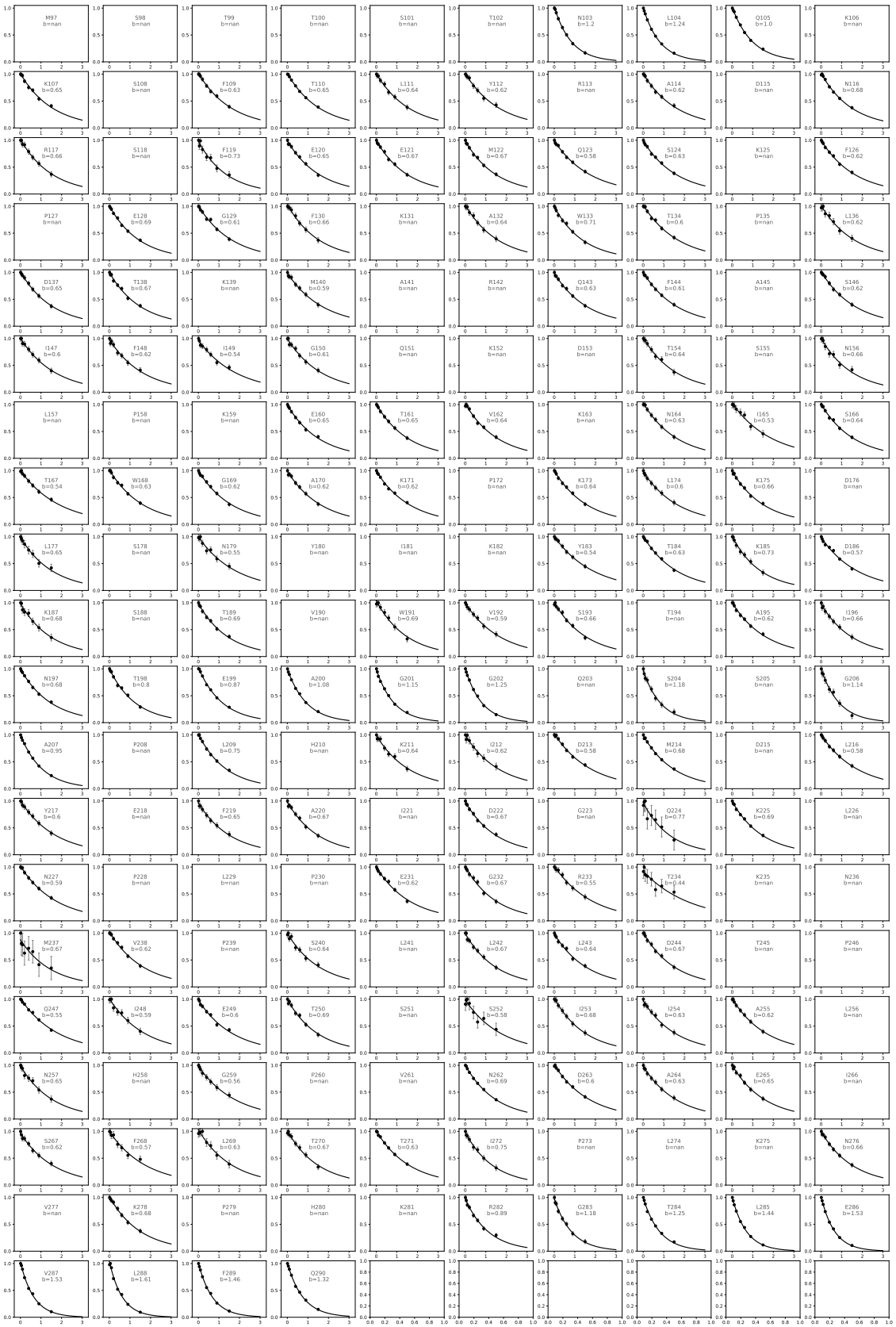

Fig. A.10.: Exponential fit functions for R_1 determination. For each non-overlapped residue, R_1 relaxation rates were fitted to a mono-exponential decay $f(x) = a \cdot e^{-b \cdot x}$ as described in Methods section 3.2.2 on page 48. The extracted rate $R_1 = b$ is indicated on each plot. Errors correspond to twice the standard deviation of a signal-free spectral region.

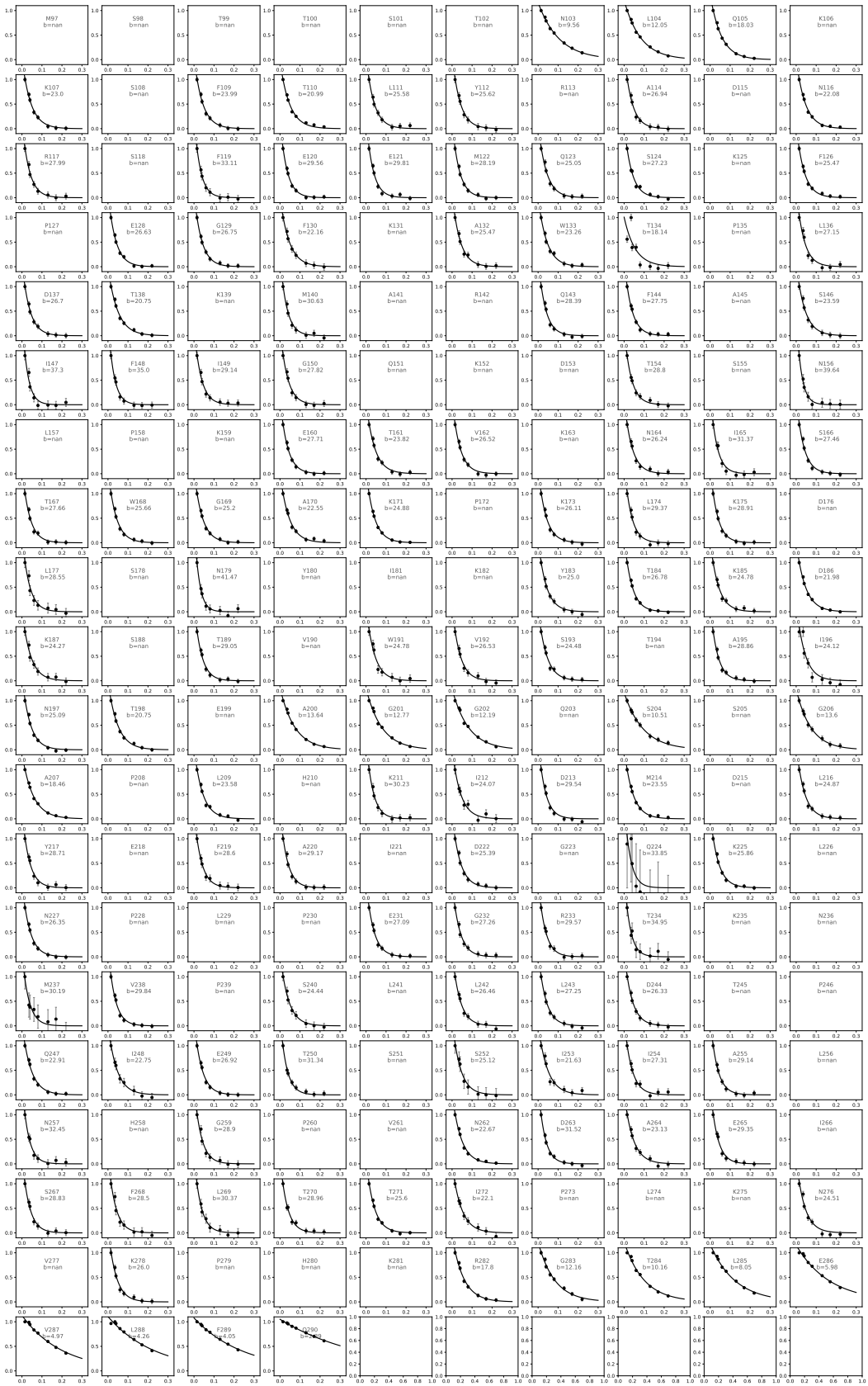

Fig. A.11.: Exponential fit functions for R_2 determination. For each non-overlapped residue, R_2 relaxation rates were fitted to a mono-exponential decay $f(x) = a \cdot e^{-b \cdot x}$ as described in Methods section 3.2.2 on page 48. The extracted rate $R_2 = b$ is indicated on each plot. Errors correspond to twice the standard deviation of a signal-free spectral region.

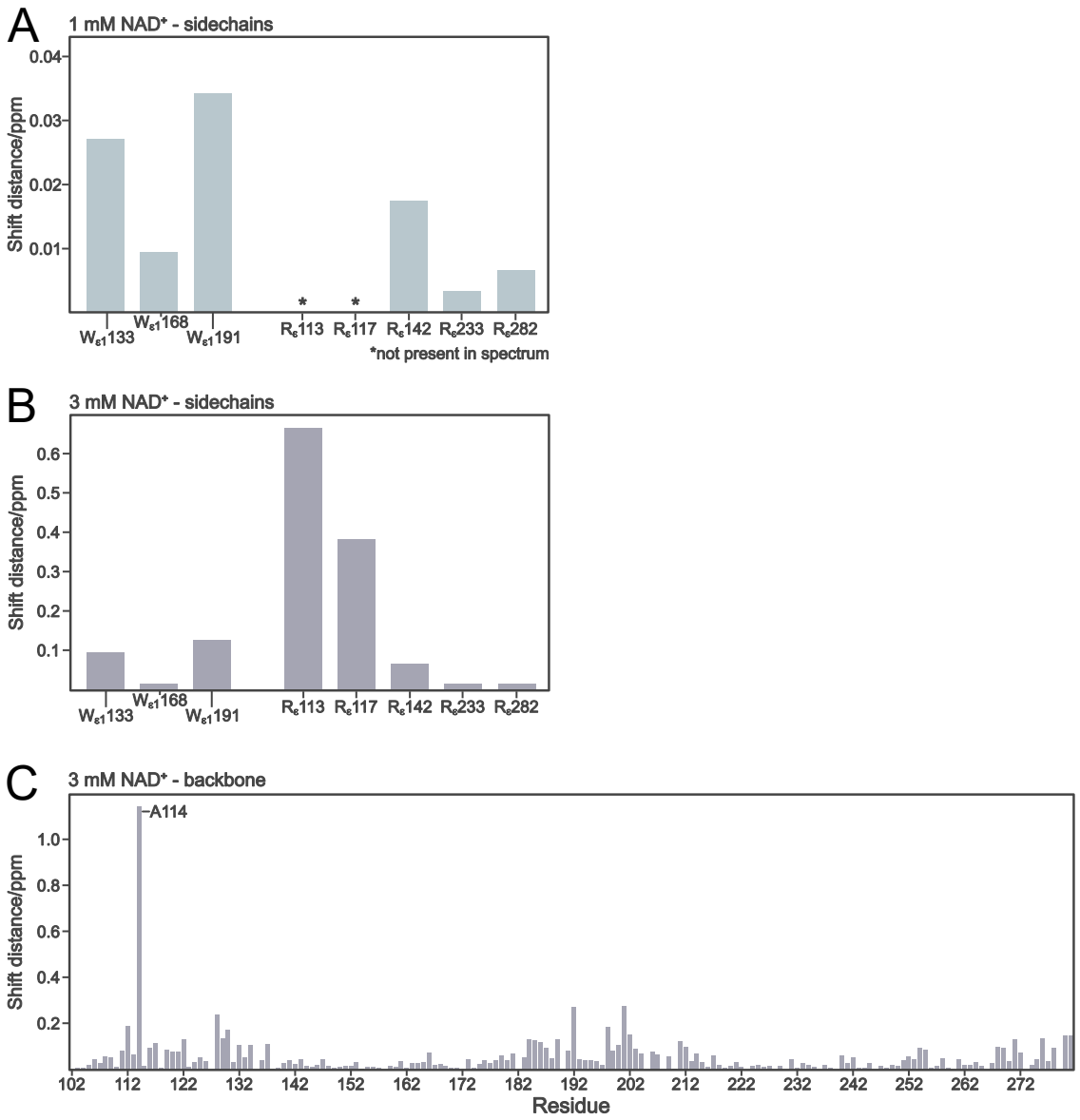

Fig. A.12.: Chemical shift perturbations of sidechains and at 3 mM NAD$^+$. Chemical shift change observed in the ^{1}H-^{15}N correlation per sidechain peak upon addition of 1 mM (**A**) or 3 mM (**B**) NAD$^+$ to 0.3 mM TcART. (**C**) Backbone chemical shift changes after addition of 3 mM NAD$^+$ to 0.3 mM TcART. Values were derived from the NMR data as described in Materials and Methods section 3.2.2 on page 49.

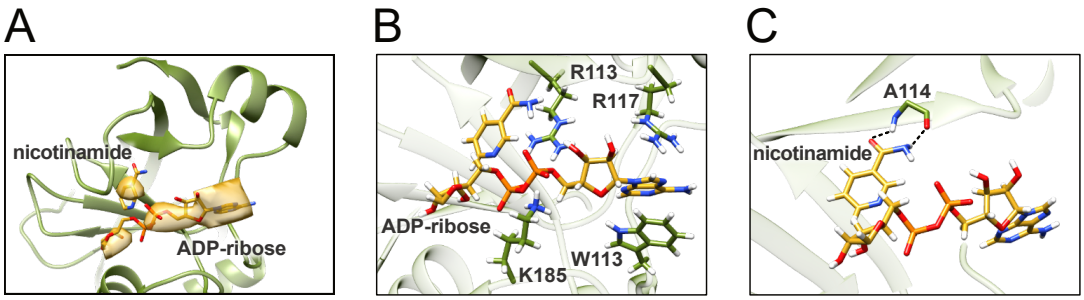

Fig. A.13.: Interactions of post-reaction TcART modeled by cryo-EM. (**A**) Surplus densities not belonging to TcART allowed modeling of nicotinamide and ADP-ribose (attached to F-actin) into the structure. (**B**) ADP-ribose is stabilized by charged interactions between the R113/K185 sidechains and the phosphorus groups as well as π-stacking with the conjugated double bonds of R117 and W113. (**C**) Nicotinamide is held in place by two β-sheet-like hydrogen bonds with the backbone of A114.

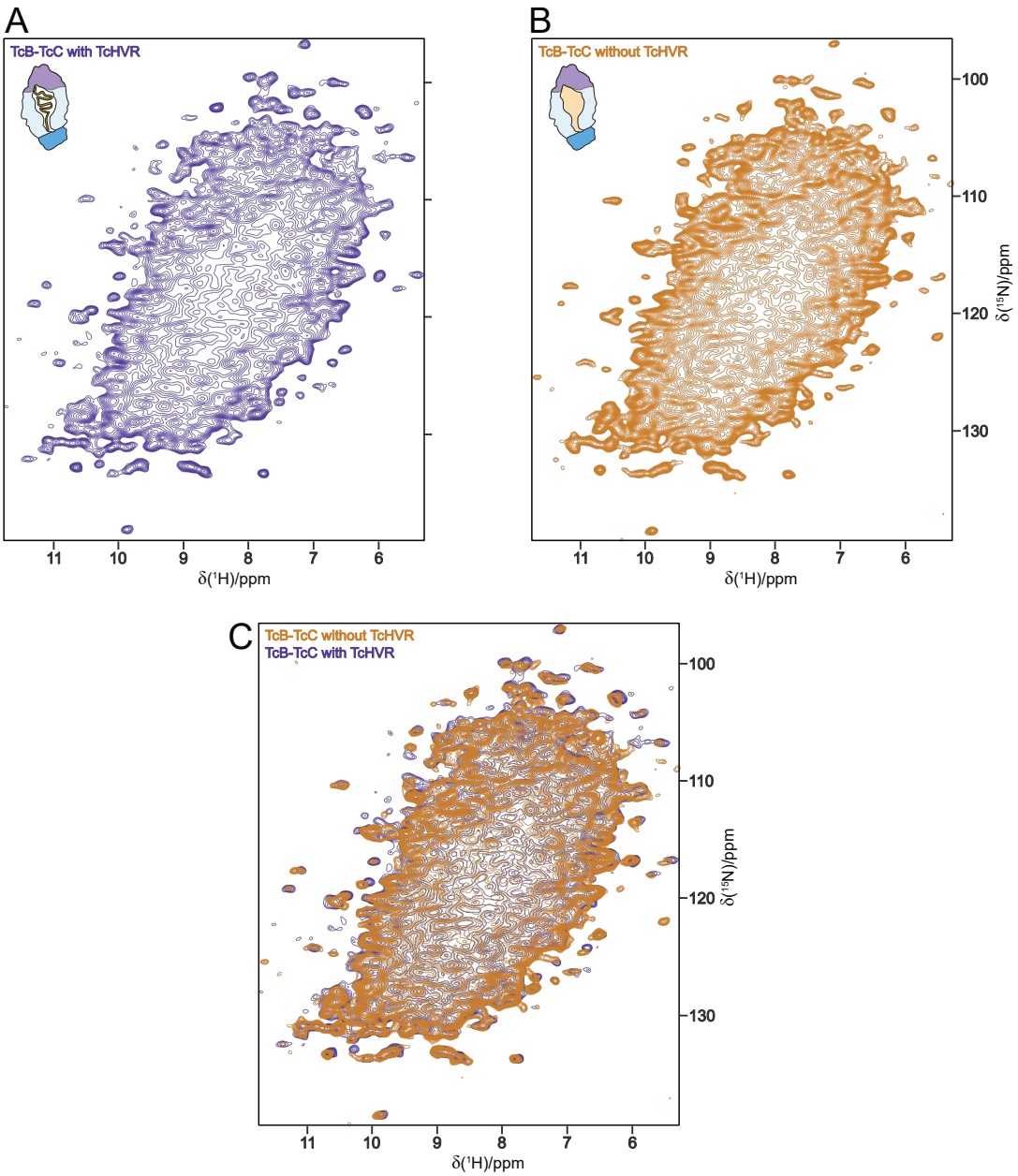

Fig. A.14.: ^{1}H-^{15}N spectra of the TcB-TcC complex with and without TcHVR. Backbone region of a ^{1}H-^{15}N correlation acquired at 100 kHz MAS by solid–state NMR for the full-length TcB-TcC complex (**A**) and the assembly without TcHVR (**B**). The superposition of both spectra is shown in (**C**).

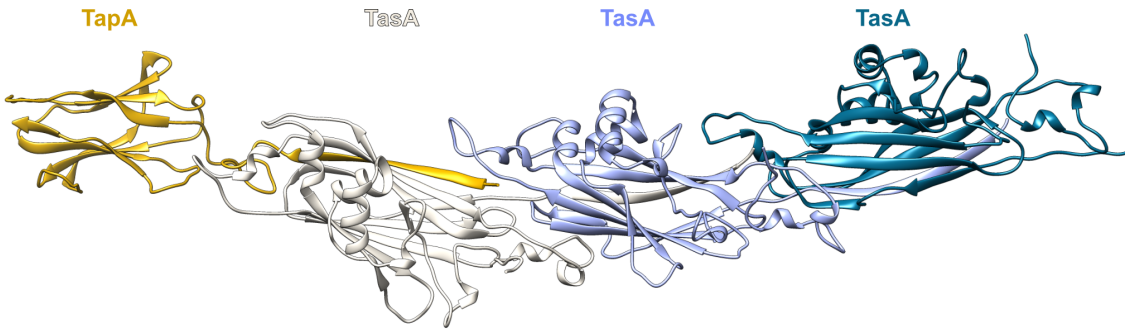

Fig. A.15.: Model of a TapA-induced TasA filament. AlphaFold-Multimer prediction received when providing one TapA and three TasA sequences. TapA (orange) donates its N-terminal β-strand as an initiator of TasA (white, light blue, and blue) filaments. Due to its slightly different fold, TapA itself cannot accept a donor-strand and must be located at the start of a filament [1].

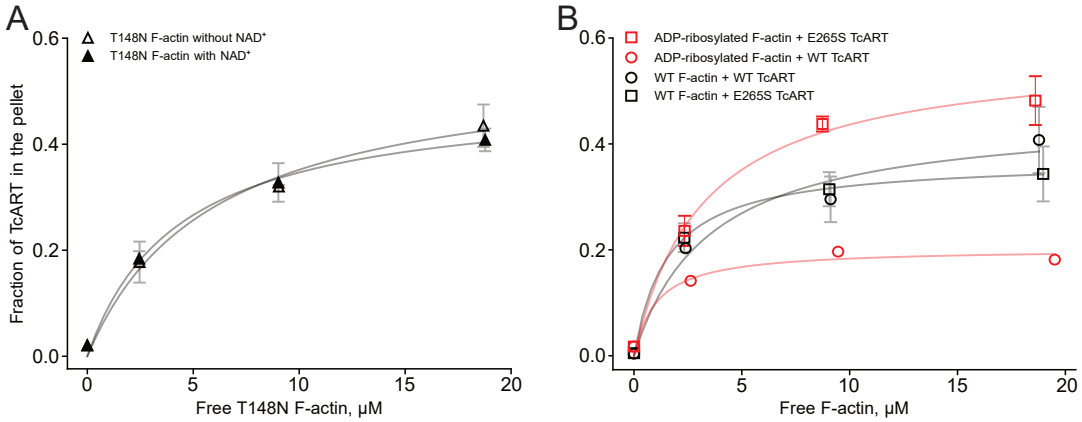

Fig. A.16.: Co-sedimentation of TcART with F-action. Phalloidin-stabilized F-actin (either T148N (non-modifiable): human cytosolic β-actin, or wt: rabbit muscle α-actin) and 3 µM TcART were co-sedimented and analyzed by SDS-PAGE. (**A**) Fractions of TcART that co-sediment with T148N F-actin in the presence or absence of NAD^+ plotted against F-actin concentrations. (**B**) Fractions of TcART (wt or E265S) that co-sediment with F-actin (wt or pre-ADP-ribosylated) plotted against F-actin concentrations. The data were quantified by densitometry and are presented as mean values, the error bars correspond to standard deviations of 3 independent experiments. Error bars smaller than the marker diameter are omitted. Experiments were conducted by Alexander Belyy at the Max Planck Institute of Molecular Physiology in Dortmund, data are published under [2].

Tab. A.1.: Further H^N-H^N contacts of through-space spectra recorded on TasA. List of additional H^N-H^N contacts derived from the hNHH and hNhhNH spectra which are not included in Fig. 4.8 on page 71. Symmetric cross-peak patterns arising from β-sheet contacts are shown on the left, those with sequential proximity in the right columns.

β-sheet		**sequential**							
K49	Q71	F29	D30	F105	L106	G166	K167	G235	L236
N51	D69	D31	I32	S107	Q108	V170	A171	T237	I238
G63	F200	I32	K33	S107	F109	T172	I173	K239	K240
K61	D64	K33	S34	Q108	F109	E180	Y181	K240	D241
L66	I198	S34	K35	T112	L113	Y181	D182	K240	H242
K68	M196	K35	D36	I129	L130	K186	T187	K240	T243
F70	V194	A37	T38	D132	A133	V194	Q195	D241	H242
L84	E229	F39	A40	N134	L135	Q195	M196	D241	T243
A86	S227	A40	S41	L135	K136	F200	K201	H242	T243
N88	Q225	S41	G42	K136	D137	K201	D202	D244	K245
G90	K223	A48	K49	D137	L138	K204	T205	D244	G246
K93	S221	N51	S52	D137	Y139	T205	K206	K245	G247
V111	A133	A53	S54	L138	Y139	K206	D207	G247	Y248
V168	M85*	S54	V55	L138	L140	D207	E208	V249	K250
V170	V83*	V55	N56	Y139	L140	D207	K209	V249	E151
		N56	L57	Y139	M141	D207	G210	K250	E251
		L57	S58	L140	M141	D207	L211	E251	N252
		S58	N59	M141	S142	E208	K209	H256	S257
		N59	L60	M141	A143	E208	G210	S257	E258
		G63	D64	S142	A143	E208	L211	E258	D259
		K65	L66	A143	K144	K209	G210	D259	K260
		F70	Q71	K144	N145	K209	L211		
		N74	N75	K144	D146	G210	L211		
		N75	G76	D146	A147	N215	K216		
		S77	L78	A147	A148	K216	Y217		
		K81	E82	A147	A149	K216	Q218		
		N88	Y89	A148	A149	Y217	Q218		
		Y89	G90	A148	A150	Q218	N220		
		G90	D91	A149	A150	G219	N220		
		G90	F92	A150	E151	S221	I222		
		D91	F92	E151	K152	I222	K223		
		F92	K93	E151	I153	K223	L224		
		K93	A94	K152	I153	L224	Q225		
		G96	G97	I153	K154	Q225	F226		
		S98	N99	I153	K155	F226	S227		
		T100	S101	K154	K155	S227	F228		
		E103	D104	K155	Q156	A230	T231		
		D104	F105	Q156	I157	T231	Q232		
		D104	L106	A164	S165	Q232	W233		
		F105	E103	S165	G166	N234	G235		

*M85 / V83 show peak doubling and only one peak has a symmetric cross-peak pattern

Tab. A.2.: Summary of structural statistics for the NMR ensemble.

Completeness of resonance assignment	
Backbone[a]	91 %
Sidechain non-H	74 %
Sidechain H	89 %
Aromatic	89 %
Conformationally restricting restraints (consensus ensembles)	
Distance restraints	
Total	3315
Intraresidue (i=j)	0
Sequential (|i-1|=1)	834
Medium range (1<|i-j|<5)	460
Long range (|i-j|>=5)	1297
Ambiguous	724
Dihedral angle restr.	306
Hydrogen-bond restr.	37
Residual restraint violations (average of 2 consensus runs)	
Average no. of distance viol. per structure	
0.1 - 0.3 Å	209
0.3 - 0.5 Å	165
> 0.5 Å	270
Average no. of dihedral viol. per structure	
> 5°	3.6
Model quality[b]	
RMSD backbone atoms (Å)[c]	0.5
RMSD heavy atoms (Å)[c]	0.8
RMSD bond lenghts (Å)	0.005
RMSD bond angles (°)	0.7
Molprobity Ramachandran Statistics[b,c]	
Most favored regions (%)	97.2
Allowed regions (%)	2.7
Disallowed regions (%)	0.1
Global quality scores (raw/Z scores)[b]	
Verify3D	(0.25/-3.37)
Prosall	(0.60/-0.21)
Procheck (Φ-Ψ)[c]	(-0.42/-1.34)
Procheck (all)[c]	(-0.35/-2.07)
MolProbity clash score	(24.08/-2.61)
Model contents	
Ordered residue ranges[b]	103-200, 207-231, 237-258, 260-280
Total no. of residues	180
BMRB accession numbers	34717, 51438
PDB ID Code	7ZBQ

[a]CO resonances were only assigned for the deuterated protein
[b]calculated using the PSVS webserver (https://montelionelab.chem.rpi.edu/PSVS/PSVS/)
[c]for all ordered residues from PSVS based on dihedral order parameter S(Φ)+S(Ψ)>=1.8

Tab. A.3.: Complete peak estimates of arginine ^{1}H-$^{15}N\varepsilon$ integral regions for TcB-TcC.

ref. area	peak count	area	abs. integral	est. peaks	total
1	**30**	**1**	**$5.50 \cdot 10^{12}$**	**30.00**	
		2	$2.16 \cdot 10^{13}$	117.55	
		-	exterior Peaks	2	Σ 149.55
		3	$1.62 \cdot 10^{12}$	8.84	
		4	$2.87 \cdot 10^{12}$	15.64	
		5	$1.80 \cdot 10^{13}$	98.13	
		6	$1.19 \cdot 10^{12}$	6.49	
		7	$6.46 \cdot 10^{11}$	3.52	
		-	exterior Peaks	12	Σ 144.62
3	**9**	1	$5.50 \cdot 10^{12}$	30.55	
		2	$2.16 \cdot 10^{13}$	119.69	
		-	exterior Peaks	2	Σ 152.24
		3	**$1.62 \cdot 10^{12}$**	**9.00**	
		4	$2.87 \cdot 10^{12}$	15.92	
		5	$1.80 \cdot 10^{13}$	99.92	
		6	$1.19 \cdot 10^{12}$	6.61	
		7	$6.46 \cdot 10^{11}$	3.59	
		-	exterior Peaks	12	Σ 147.04
4	**16**	1	$5.50 \cdot 10^{12}$	30.70	
		2	$2.16 \cdot 10^{13}$	120.27	
		-	exterior Peaks	2	Σ 152.97
		3	$1.62 \cdot 10^{12}$	9.04	
		4	**$2.87 \cdot 10^{12}$**	**16.00**	
		5	$1.80 \cdot 10^{13}$	100.41	
		6	$1.19 \cdot 10^{12}$	6.65	
		7	$6.46 \cdot 10^{11}$	3.60	
		-	exterior Peaks	12	Σ 147.70

List of Figures

List of Tables

References

[1] Yvette Roske et al. “TapA acts as specific chaperone in TasA filament formation by strand complementation”. In: *Proceedings of the National Academy of Sciences* 120.17 (2023). DOI: 10.1073/pnas.2217070120.

[2] Alexander Belyy et al. “Mechanism of threonine ADP-ribosylation of F-actin by a Tc toxin”. In: *Nature Communications* 13.1 (2022). DOI: 10.1038/s41467-022-31836-w.

[3] Alexander Fleming. “On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*.” eng. In: *British Journal of Experimental Pathology* 10 (8 1929), pp. 226–236. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566493/.

[4] Johan P. Mackenbach and Caspar W. N. Looman. “Secular Trends of Infectious Disease Mortality in the Netherlands, 1911-1978: Quantitative Estimates of Changes Coinciding with the Introduction of Antibiotics”. In: *International Journal of Epidemiology* 17.3 (1988), pp. 618–624. DOI: 10.1093/ije/17.3.618.

[5] Julian Davies and Dorothy Davies. “Origins and Evolution of Antibiotic Resistance”. In: *Microbiology and Molecular Biology Reviews* 74.3 (2010), pp. 417–433. DOI: 10.1128/mmbr.00016-10.

[6] Alessandro Cassini et al. “Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis”. In: *The Lancet Infectious Diseases* 19.1 (2019), pp. 56–66. DOI: 10.1016/s1473-3099(18)30605-4.

[7] *WHO publishes list of bacteria for which new antibiotics are urgently needed*. Feb. 2017. URL: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed (visited on 06/10/2023).

[8] Eric D. Brown and Gerard D. Wright. “Antibacterial drug discovery in the resistance era”. In: *Nature* 529.7586 (2016), pp. 336–343. DOI: 10.1038/nature17042.

[9] Rhythm Shukla et al. “Teixobactin kills bacteria by a two-pronged attack on the cell envelope”. In: *Nature* 608.7922 (2022), pp. 390–396. DOI: 10.1038/s41586-022-05019-y.

[10] Nicola Principi, Ettore Silvestri, and Susanna Esposito. “Advantages and Limitations of Bacteriophages for the Treatment of Bacterial Infections”. In: *Frontiers in Pharmacology* 10 (2019). DOI: 10.3389/fphar.2019.00513.

[11] Sofia Arnaouteli et al. “Bifunctionality of a biofilm matrix protein controlled by redox state”. In: *Proceedings of the National Academy of Sciences* 114.30 (July 2017), E6184–E6191. DOI: 10.1073/pnas.1707687114.

[12] Daniel Roderer et al. “Towards the application of Tc toxins as a universal protein translocation system”. In: *Nature Communications* 10.1 (2019). DOI: 10.1038/s41467-019-13253-8.

[13] Daniel López, Hera Vlamakis, and Roberto Kolter. “Biofilms”. In: *Cold Spring Harbor perspectives in biology* 2.7 (2010), a000398. DOI: 10.1101/cshperspect.a000398.

[14] Adina Schulze et al. “Biofilms by bacterial human pathogens: Clinical relevance - development, composition and regulation - therapeutical strategies”. In: *Microbial Cell* 8.2 (2021), pp. 28–56. DOI: 10.15698/mic2021.02.741.

[15] Sapna Chitlapilly Dass et al. “Impact of mixed biofilm formation with environmental microorganisms on *E. coli* O157:H7 survival against sanitization”. In: *npj Science of Food* 4.1 (2020). DOI: 10.1038/s41538-020-00076-x.

[16] Rodney M Donlan. “Biofilms: microbial life on surfaces”. In: *Emerging infectious diseases* 8.9 (2002), p. 881. DOI: 10.3201/eid0809.020063.

[17] Evelina N. Lagamayo. “Antimicrobial resistance in major pathogens of hospital-acquired pneumonia in Asian countries”. In: *American journal of infection control* 36.4 (2008), S101–S108. DOI: 10.1016/j.ajic.2007.10.020.

[18] Hera Vlamakis et al. “Sticking together: building a biofilm the *Bacillus subtilis* way.” In: *Nature reviews. Microbiology* 11 (3 Mar. 2013), pp. 157–168. ISSN: 1740-1534. DOI: 10.1038/nrmicro2960.

[19] Diego Romero et al. “An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms.” In: *Molecular microbiology* 80 (5 June 2011), pp. 1155–1168. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2011.07653.x.

[20] Kazuo Kobayashi and Megumi Iwano. “BslA (YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms”. In: *Molecular Microbiology* 85.1 (2012), pp. 51–66. DOI: 10.1111/j.1365-2958.2012.08094.x.

[21] Laura Hobley et al. “BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm.” In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (33 Aug. 2013), pp. 13600–13605. ISSN: 1091-6490. DOI: 10.1073/pnas.1306390110.

[22] Diego Romero et al. “Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms.” In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (5 Feb. 2010), pp. 2230–2234. ISSN: 1091-6490. DOI: 10.1073/pnas.0910560107.

[23] Matthew R. Amos et al. “Influence of the *Photorhabdus luminescens* Phosphomannose Isomerase Gene, *manA*, on Mannose Utilization, Exopolysaccharide Structure, and Biofilm Formation”. In: *Applied and Environmental Microbiology* 77.3 (2011), pp. 776–785. DOI: 10.1128/aem.02326-10.

[24] Nick R. Waterfield, Todd Ciche, and David Clarke. “*Photorhabdus* and a Host of Hosts”. In: *Annual Review of Microbiology* 63.1 (2009), pp. 557–574. DOI: 10.1146/annurev.micro.091208.073507.

[25] Mahfouz M. M. Abd-Elgawad. “*Photorhabdus* spp.: An Overview of the Beneficial Aspects of Mutualistic Bacteria of Insecticidal Nematodes”. In: *Plants* 10.8 (2021), p. 1660. DOI: 10.3390/plants10081660.

[26] Paramaporn Muangpat et al. “Antibacterial activity of *Xenorhabdus* and *Photorhabdus* isolated from entomopathogenic nematodes against antibiotic-resistant bacteria”. In: *PLOS ONE* 15.6 (2020). Ed. by Filippo Giarratana, e0234129. DOI: 10.1371/journal.pone.0234129.

[27] Ioannis Eleftherianos et al. “An antibiotic produced by an insect-pathogenic bacterium suppresses host defenses through phenoloxidase inhibition”. In: *Proceedings of the National Academy of Sciences* 104.7 (2007), pp. 2419–2424. DOI: 10.1073/pnas.0610525104.

[28] P. Garbeva, J.A. van Veen, and J.D. van Elsas. “Predominant *Bacillus* spp. in Agricultural Soil under Different Management Regimes Detected via PCR-DGGE”. In: *Microbial Ecology* 45.3 (2003), pp. 302–316. DOI: 10.1007/s00248-002-2034-8.

[29] M. A. Andersson et al. “A Novel Sensitive Bioassay for Detection of *Bacillus cereus* Emetic Toxin and Related Depsipeptide Ionophores”. In: *Applied and Environmental Microbiology* 64.4 (1998), pp. 1338–1343. DOI: 10.1128/aem.64.4.1338-1343.1998.

[30] Steve M. Blevins and Michael S. Bronze. “Robert Koch and the ‘golden age’ of bacteriology”. In: *International Journal of Infectious Diseases* 14.9 (2010), e744–e751. DOI: 10.1016/j.ijid.2009.12.003.

[31] Mohamed A. Ibrahim et al. “*Bacillus thuringiensis*”. In: *Bioengineered Bugs* 1.1 (2010), pp. 31–50. DOI: 10.4161/bbug.1.1.10519.

[32] Matin Qaim and David Zilberman. “Yield Effects of Genetically Modified Crops in Developing Countries”. In: *Science* 299.5608 (2003), pp. 900–902. DOI: 10.1126/science.1080609.

[33] Peter T. McKenney, Adam Driks, and Patrick Eichenberger. “The Bacillus subtilis endospore: assembly and functions of the multilayered coat”. In: *Nature Reviews Microbiology* 11.1 (2012), pp. 33–44. DOI: 10.1038/nrmicro2921.

[34] Yuan Su et al. “*Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine”. In: *Microbial Cell Factories* 19.1 (2020). DOI: 10.1186/s12934-020-01436-8.

[35] Nur Aliyah Mohd Azrin et al. “Versatility of subtilisin: A review on structure, characteristics, and applications”. In: *Biotechnology and Applied Biochemistry* 69.6 (2022), pp. 2599–2616. DOI: 10.1002/bab.2309.

[36] Mónica Serrano et al. “A *Bacillus subtilis* secreted protein with a role in endospore coat assembly and function”. In: *Journal of bacteriology* 181.12 (1999), pp. 3632–3643. DOI: https://doi.org/10.1128/jb.181.12.3632-3643.1999.

[37] Axel G. Stöver and Adam Driks. “Secretion, localization, and antibacterial activity of TasA, a *Bacillus subtilis* spore-associated protein”. In: *Journal of bacteriology* 181.5 (1999), pp. 1664–1672. DOI: https://doi.org/10.1128/jb.181.5.1664-1672.1999.

[38] Steven S. Branda et al. “A major protein component of the *Bacillus subtilis* biofilm matrix.” In: *Molecular microbiology* 59 (4 Feb. 2006), pp. 1229–1238. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.2005.05020.x.

[39] Harold Tjalsma et al. “Functional analysis of the secretory precursor processing machinery of *Bacillus subtilis*: identification of a eubacterial homolog of archaeal and eukaryotic signal peptidases”. In: *Genes & Development* 12.15 (1998), pp. 2318–2331. DOI: 10.1101/gad.12.15.2318.

[40] Axel G. Stöver and Adam Driks. “Regulation of Synthesis of the *Bacillus subtilis* Transition-Phase, Spore-Associated Antibacterial Protein TasA”. In: *Journal of Bacteriology* 181.17 (1999), pp. 5476–5481. DOI: 10.1128/jb.181.17.5476-5481.1999.

[41] Ines Mandic-Mulec, Laurance Doukhan, and Issar Smith. “The Bacillus subtilis SinR protein is a repressor of the key sporulation gene spo0A”. In: *Journal of Bacteriology* 177.16 (1995), pp. 4619–4627. DOI: 10.1128/jb.177.16.4619-4627.1995.

[42] Daniel B. Kearns et al. “A master regulator for biofilm formation by *Bacillus subtilis*”. In: *Molecular Microbiology* 55.3 (2004), pp. 739–749. DOI: 10.1111/j.1365-2958.2004.04440.x.

[43] Tantan Gao et al. “Alternative modes of biofilm formation by plant-associated *Bacillus cereus*”. In: *MicrobiologyOpen* 4.3 (2015), pp. 452–464. DOI: 10.1002/mbo3.251.

[44] Joaquín Caro-Astorga et al. “A genomic region involved in the formation of adhesin fibers in *Bacillus cereus* biofilms”. In: *Frontiers in Microbiology* 5 (2015). DOI: 10.3389/fmicb.2014.00745.

[45] Daniel López and Roberto Kolter. “Extracellular signals that define distinct and co-existing cell fates in *Bacillus subtilis*”. In: *FEMS microbiology reviews* 34.2 (2010), pp. 134–149. DOI: 10.1111/j.1574-6976.2009.00199.x.

[46] Benjamin Mielich-Süss and Daniel Lopez. “Molecular mechanisms involved in *Bacillus subtilis* biofilm formation”. In: *Environmental microbiology* 17.3 (2015), pp. 555–565. DOI: 10.1111/1462-2920.12527.

[47] Rachel Bleich et al. “Thiopeptide antibiotics stimulate biofilm formation in *Bacillus subtilis*”. In: *Proceedings of the National Academy of Sciences* 112.10 (2015), pp. 3086–3091. DOI: 10.1073/pnas.1414272112.

[48] Maryam Khezri, Gholamreza Salehi Jouzani, and Masoud Ahmadzadeh. “*Fusarium culmorum* affects expression of biofilm formation key genes in *Bacillus subtilis*”. In: *Brazilian Journal of Microbiology* 47.1 (2016), pp. 47–54. DOI: 10.1016/j.bjm.2015.11.019.

[49] Laura Hobley et al. “Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes”. In: *FEMS Microbiology Reviews* 39.5 (2015). Ed. by Sonja-Verena Albers, pp. 649–669. DOI: 10.1093/femsre/fuv015.

[50] Diego Romero et al. “Functional analysis of the accessory protein TapA in *Bacillus subtilis* amyloid fiber assembly.” In: *Journal of bacteriology* 196 (8 Apr. 2014), pp. 1505–1513. ISSN: 1098-5530. DOI: 10.1128/JB.01363-13.

[51] Anna Dragoš, Ákos T Kovács, and Dennis Claessen. “The Role of Functional Amyloids in Multicellular Growth and Development of Gram-Positive Bacteria”. In: *Biomolecules* 7.3 (2017), p. 60. DOI: 10.3390/biom7030060.

[52] Nadia El Mammeri et al. “Molecular architecture of bacterial amyloids in *Bacillus* biofilms”. In: *The FASEB Journal* 33.11 (2019), pp. 12146–12163. DOI: 10.1096/fj.201900831r.

[53] Andrea Hawe, Marc Sutter, and Wim Jiskoot. “Extrinsic Fluorescent Dyes as Tools for Protein Characterization”. In: *Pharmaceutical Research* 25.7 (2008), pp. 1487–1499. DOI: 10.1007/s11095-007-9516-9.

[54] Liraz Chai et al. “Isolation, Characterization, and Aggregation of a Structured Bacterial Matrix Precursor”. In: *Journal of Biological Chemistry* 288.24 (2013), pp. 17559–17568. DOI: 10.1074/jbc.m113.453605.

[55] Elliot Erskine et al. “Formation of functional, non-amyloidogenic fibres by recombinant *Bacillus subtilis* TasA”. In: *Molecular Microbiology* 110.6 (2018), pp. 897–913. DOI: 10.1111/mmi.13985.

[56] Jan Böhning et al. “Donor-strand exchange drives assembly of the TasA scaffold in *Bacillus subtilis* biofilms”. In: *Nature Communications* 13.1 (2022). DOI: 10.1038/s41467-022-34700-z.

[57] David Fernández-Chapa, Jesica Ramírez-Villalobos, and Luis Galán-Wong. "Toxic Potential of *Bacillus thuringiensis*: An Overview". In: *Protecting Rice Grains in the Post-Genomic Era* (2019). DOI: 10.5772/intechopen.85756.

[58] Richard H. Ffrench-Constant, Andrea Dowling, and Nicholas R. Waterfield. "Insecticidal toxins from *Photorhabdus* bacteria and their potential use in agriculture". In: *Toxicon* 49.4 (2007), pp. 436–451. DOI: 10.1016/j.toxicon.2006.11.019.

[59] Wellington Junior da Silva et al. "The great potential of entomopathogenic bacteria *Xenorhabdus* and *Photorhabdus* for mosquito control: a review". In: *Parasites & Vectors* 13.1 (2020). DOI: 10.1186/s13071-020-04236-6.

[60] David Bowen et al. "Insecticidal Toxins from the Bacterium *Photorhabdus luminescens*". In: *Science* 280.5372 (1998), pp. 2129–2132. DOI: 10.1126/science.280.5372.2129.

[61] Michael B. Blackburn et al. "The Occurrence of *Photorhabdus*-Like Toxin Complexes in *Bacillus thuringiensis*". In: *PLOS ONE* 6.3 (2011). Ed. by Herman Tse, e18122. DOI: 10.1371/journal.pone.0018122.

[62] Timothy D. Hey et al. "Mixing and matching TC proteins for pest control". In: *US Patent* US 8,084,418 B2 (2009). URL: https://patents.google.com/patent/US8084418B2/en.

[63] Michelle C. Hares et al. "The *Yersinia pseudotuberculosis* and *Yersinia pestis* toxin complex is active against cultured mammalian cells". In: *Microbiology* 154.11 (2008), pp. 3503–3517. DOI: 10.1099/mic.0.2008/018440-0.

[64] Daniel Roderer and Stefan Raunser. "Tc Toxin Complexes: Assembly, Membrane Permeation, and Protein Translocation". In: *Annual Review of Microbiology* 73.1 (2019), pp. 247–265. DOI: 10.1146/annurev-micro-102215-095531.

[65] Jingxian Li et al. "Structural Basis for Teneurin Function in Circuit-Wiring: A Toxin Motif at the Synapse". In: *Cell* 173.3 (2018), 735–748.e15. DOI: 10.1016/j.cell.2018.03.036.

[66] Zubing Zhan, Xuehong Qiu, and Richou Han. "Horizontal transfer of the C-termini of *tccC* genes in *Photorhabdus* and *Xenorhabdus*". In: *Genes & Genomics* 38.8 (2016), pp. 685–692. DOI: 10.1007/s13258-016-0410-x.

[67] Richard Ffrench-Constant and Nicholas Waterfield. "An ABC Guide to the Bacterial Toxin Complexes". In: *Advances in Applied Microbiology* (2005), pp. 169–183. DOI: 10.1016/s0065-2164(05)58005-5.

[68] Eric Duchaud et al. "The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*". In: *Nature Biotechnology* 21.11 (2003), pp. 1307–1313. DOI: 10.1038/nbt886.

[69] N. Waterfield et al. "Potentiation and cellular phenotypes of the insecticidal Toxin complexes of *Photorhabdus* bacteria". In: *Cellular Microbiology* 7.3 (2005), pp. 373–382. DOI: 10.1111/j.1462-5822.2004.00467.x.

[70] Dominic Meusch et al. “Mechanism of Tc toxin action revealed in molecular detail”. In: *Nature* 508.7494 (2014), pp. 61–65. DOI: 10.1038/nature13015.

[71] Alexander E. Lang et al. “*Photorhabdus luminescens* Toxins ADP-Ribosylate Actin and RhoA to Force Actin Clustering”. In: *Science* 327.5969 (2010), pp. 1139–1142. DOI: 10.1126/science.1184557.

[72] Jason N. Busby et al. “The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device”. In: *Nature* 501.7468 (2013), pp. 547–550. DOI: 10.1038/nature12465.

[73] Nan Song et al. “Genome-wide dissection reveals diverse pathogenic roles of bacterial Tc toxins”. In: *PLOS Pathogens* 17.2 (2021). Ed. by William Navarre, e1009102. DOI: 10.1371/journal.ppat.1009102.

[74] Daniel Roderer et al. “Glycan-dependent cell adhesion mechanism of Tc toxins”. In: *Nature Communications* 11.1 (2020). DOI: 10.1038/s41467-020-16536-7.

[75] Daniel Roderer et al. “Structure of a Tc holotoxin pore provides insights into the translocation mechanism”. In: *Proceedings of the National Academy of Sciences* 116.46 (2019), pp. 23083–23090. DOI: 10.1073/pnas.1909821116.

[76] Nathan C. Simon, Klaus Aktories, and Joseph T. Barbieri. “Novel bacterial ADP-ribosylating toxins: structure and function”. In: *Nature Reviews Microbiology* 12.9 (2014), pp. 599–611. DOI: 10.1038/nrmicro3310.

[77] Qing Deng and Joseph T. Barbieri. “Molecular Mechanisms of the Cytotoxicity of ADP-Ribosylating Toxins”. In: *Annual Review of Microbiology* 62.1 (2008), pp. 271–288. DOI: 10.1146/annurev.micro.62.081307.162848.

[78] L. Aravind et al. “The Natural History of ADP-Ribosyltransferases and the ADP-Ribosylation System”. In: *Endogenous ADP-Ribosylation*. Springer International Publishing, 2014, pp. 3–32. DOI: 10.1007/82_2014_414.

[79] Toru Yoshida and Hideaki Tsuge. “Substrate N^2 atom recognition mechanism in pierisin family DNA-targeting, guanine-specific ADP-ribosyltransferase ScARP”. In: *Journal of Biological Chemistry* 293.36 (2018), pp. 13768–13774. DOI: 10.1074/jbc.ac118.004412.

[80] Penelope E. Stein et al. “Structure of a pertussis toxin-sugar complex as a model for receptor binding”. In: *Nature Structural & Molecular Biology* 1.9 (1994), pp. 591–596. DOI: 10.1038/nsb0994-591.

[81] Argentina Becker et al. “Structure of CARDS toxin, a unique ADP-ribosylating and vacuolating cytotoxin from *Mycoplasma pneumoniae*”. In: *Proceedings of the National Academy of Sciences* 112.16 (2015), pp. 5165–5170. DOI: 10.1073/pnas.1420308112.

[82] Songyu Dong et al. “*Photorhabdus luminescens* TccC3 Toxin Targets the Dynamic Population of F-Actin and Impairs Cell Cortex Integrity”. In: *International Journal of Molecular Sciences* 23.13 (2022), p. 7026. DOI: 10.3390/ijms23137026.

[83] Alexander E. Lang et al. “Actin ADP-ribosylation at Threonine148 by *Photorhabdus luminescens* toxin TccC3 induces aggregation of intracellular F-actin”. In: *Cellular Microbiology* 19.1 (2016), e12636. DOI: 10.1111/cmi.12636.

[84] Christos Gatsogiannis et al. “Tc toxin activation requires unfolding and refolding of a β-propeller”. In: *Nature* 563.7730 (2018), pp. 209–213. DOI: 10.1038/s41586-018-0556-6.

[85] P. J. Hore. *Nuclear Magnetic Resonance*. 2nd. Oxford Chemistry Primers, 2015. ISBN: 978-0-19-870341-9.

[86] Malcolm H. Levitt. *Spin dynamics : Basics of nuclear magnetic resonance*. 2nd ed. John Wiley & Sons Ltd, 2008. ISBN: 978-0-470-51118-3.

[87] David D. Laws, Hans-Marcus L. Bitter, and Alexej Jerschow. “Solid-State NMR Spectroscopic Methods in Chemistry”. In: *Angewandte Chemie International Edition* 41.17 (2002), pp. 3096–3129. DOI: 10.1002/1521-3773(20020902)41:17<3096::AID-ANIE3096>3.0.CO;2-X.

[88] Bernd Reif et al. “Solid–state NMR spectroscopy”. In: *Nature Reviews Methods Primers* 1.1 (2021). DOI: 10.1038/s43586-020-00002-1.

[89] David S. Wishart et al. “^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR”. In: *Journal of Biomolecular NMR* 6.2 (Sept. 1995), pp. 135–140. ISSN: 1573-5001. DOI: 10.1007/BF00211777.

[90] Peter W. Atkins and Julio de Paula. *Physikalische Chemie*. 5. Auflage. pp. 547-582. WILEY-VCH, 2013. ISBN: 978-3-527-33247-2.

[91] Andrew E. Bennett et al. “Heteronuclear decoupling in rotating solids”. In: *The Journal of Chemical Physics* 103.16 (1995), pp. 6951–6958. DOI: 10.1063/1.470372.

[92] Bernd Reif and Robert G. Griffin. “^{1}H detected ^{1}H,^{15}N correlation spectroscopy in rotating solids”. In: *Journal of Magnetic Resonance* 160.1 (2003), pp. 78 –83. ISSN: 1090-7807. DOI: https://doi.org/10.1016/S1090-7807(02)00035-6.

[93] Veniamin Chevelkov et al. “Ultrahigh Resolution in Proton Solid-State NMR Spectroscopy at High Levels of Deuteration”. In: *Angewandte Chemie International Edition* 45.23 (2006), pp. 3878–3881. DOI: 10.1002/anie.200600328.

[94] Ümit Akbey et al. “Optimum levels of exchangeable protons in perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy.” In: *Journal of biomolecular NMR* 46 (1 Jan. 2010), pp. 67–73. ISSN: 1573-5001. DOI: 10.1007/s10858-009-9369-0.

[95] Donghua H. Zhou et al. “Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning”. In: *Journal of the American Chemical Society* 129.38 (2007), pp. 11791–11801. DOI: 10.1021/ja073462m.

[96] Józef R. Lewandowski et al. “Enhanced Resolution and Coherence Lifetimes in the Solid-State NMR Spectroscopy of Perdeuterated Proteins under Ultrafast Magic-Angle Spinning”. In: *J. Phys. Chem. Lett.* 2 (2011), pp. 2205–2211. ISSN: 1948-7185. DOI: 10.1021/jz200844n.

[97] Andrew J. Nieuwkoop et al. “Sensitivity and resolution of proton detected spectra of a deuterated protein at 40 and 60 kHz magic-angle-spinning”. In: *Journal of Biomolecular NMR* 61.2 (2015), pp. 161–171. ISSN: 1573-5001. DOI: 10.1007/s10858-015-9904-0.

[98] Yusuke Nishiyama. “Fast magic-angle sample spinning solid-state NMR at 60-100 kHz for natural abundance samples.” In: *Solid state nuclear magnetic resonance* 78 (Sept. 2016), pp. 24–36. ISSN: 1527-3326. DOI: 10.1016/j.ssnmr.2016.06.002.

[99] Jochem Struppe et al. “Expanding the horizons for structural analysis of fully protonated protein assemblies by NMR spectroscopy at MAS frequencies above 100 kHz”. In: *Solid state nuclear magnetic resonance* 87 (2017), pp. 117–125. DOI: 10.1016/j.ssnmr.2017.07.001.

[100] Susanne Penzel et al. “Spinning faster: protein NMR at MAS frequencies up to 126 kHz”. In: *Journal of Biomolecular NMR* 73.1-2 (2019), pp. 19–29. DOI: 10.1007/s10858-018-0219-9.

[101] Eric Chung-Yueh Yuan et al. “Faster magic angle spinning reveals cellulose conformations in woods”. In: *Chemical Communications* 57.34 (2021), pp. 4110–4113. DOI: 10.1039/d1cc01149a.

[102] James Keeler. *Understanding NMR spectroscopy*. John Wiley & Sons, 2011.

[103] Petra Rovó et al. “Mechanistic Insights into Microsecond Time-Scale Motion of Solid Proteins Using Complementary ^{15}N and ^{1}H Relaxation Dispersion Techniques”. In: *Journal of the American Chemical Society* 141.2 (2019), pp. 858–869. DOI: 10.1021/jacs.8b09258.

[104] Petra Rovó. “Recent advances in solid-state relaxation dispersion techniques”. In: *Solid State Nuclear Magnetic Resonance* 108 (2020), p. 101665. DOI: 10.1016/j.ssnmr.2020.101665.

[105] Diego F. Gauto et al. “Functional control of a 0.5 MDa TET aminopeptidase by a flexible loop revealed by MAS NMR”. In: *Nature Communications* 13.1 (2022). DOI: 10.1038/s41467-022-29423-0.

[106] Gareth A. Morris and Ray Freeman. “Enhancement of nuclear magnetic resonance signals by polarization transfer”. In: *Journal of the American Chemical Society* 101.3 (1979), pp. 760–762. DOI: 10.1021/ja00497a058.

[107] A. Pines, M. G. Gibby, and J. S. Waugh. “Proton-Enhanced Nuclear Induction Spectroscopy. A Method for High Resolution NMR of Dilute Spins in Solids”. In: *The Journal of Chemical Physics* 56.4 (1972), pp. 1776–1777. DOI: 10.1063/1.1677439.

[108] K. Takegoshi, Shinji Nakamura, and Takehiko Terao. "^{13}C–^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR". In: *Chemical Physics Letters* 344.5 (2001), pp. 631 –637. ISSN: 0009-2614. DOI: https://doi.org/10.1016/S0009-2614(01)00791-6.

[109] Federica Castellani et al. "Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy". In: *Nature* 420.6911 (2002), p. 98. DOI: 10.1038/nature01070.

[110] Emeline Barbet-Massin et al. "Rapid proton-detected NMR assignment for proteins with fast magic angle spinning." In: *Journal of the American Chemical Society* 136 (35 Sept. 2014), pp. 12489–12497. ISSN: 1520-5126. DOI: 10.1021/ja507382j.

[111] Victoria A. Higman. "Solid-state MAS NMR resonance assignment methods for proteins". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 106-107 (2018), pp. 37–65. DOI: 10.1016/j.pnmrs.2018.04.002.

[112] Lauro Ribeiro de Souza Neto et al. "*In silico* Strategies to Support Fragment-to-Lead Optimization in Drug Discovery". In: *Frontiers in Chemistry* 8 (2020). DOI: 10.3389/fchem.2020.00093.

[113] Helen Berman, Kim Henrick, and Haruki Nakamura. "Announcing the worldwide Protein Data Bank". In: *Nature Structural & Molecular Biology* 10.12 (2003), pp. 980–980. DOI: 10.1038/nsb1203-980.

[114] Jonathan C. Brooks-Bartlett and Elspeth F. Garman. "The Nobel Science: One Hundred Years of Crystallography". In: *Interdisciplinary Science Reviews* 40.3 (2015), pp. 244–264. DOI: 10.1179/0308018815z.000000000116.

[115] *The Nobel Prize in Chemistry 1962*. 1962. URL: https://www.nobelprize.org/prizes/chemistry/1962/summary/ (visited on 06/10/2023).

[116] Alexander McPherson and Jose A. Gavira. "Introduction to protein crystallization". In: *Acta Crystallographica Section F Structural Biology Communications* 70.1 (2013), pp. 2–20. DOI: 10.1107/s2053230x13033141.

[117] Yifan Cheng. "Single-particle cryo-EM-How did it get here and where will it go". In: *Science* 361.6405 (2018), pp. 876–880. DOI: 10.1126/science.aat4346.

[118] D. J. de Rosier and A. Klug. "Reconstruction of Three Dimensional Structures from Electron Micrographs". In: *Nature* 217.5124 (1968), pp. 130–134. DOI: 10.1038/217130a0.

[119] G. McMullan, A.R. Faruqi, and R. Henderson. "Direct Electron Detectors". In: *Methods in Enzymology*. Elsevier, 2016, pp. 1–17. DOI: 10.1016/bs.mie.2016.05.056.

[120] Werner Kühlbrandt. "The Resolution Revolution". In: *Science* 343.6178 (2014), pp. 1443–1444. DOI: 10.1126/science.1251652.

[121] *The Nobel Prize in Chemistry 2017*. 2017. URL: https://www.nobelprize.org/prizes/chemistry/2017/summary/ (visited on 06/10/2023).

[122] *The Nobel Prize in Chemistry 2002*. 2002. URL: https://www.nobelprize.org/prizes/chemistry/2002/summary/ (visited on 06/13/2023).

[123] Elena Schmidt and Peter Güntert. "A New Algorithm for Reliable and General NMR Resonance Assignment". In: *Journal of the American Chemical Society* 134.30 (2012), pp. 12817–12829. DOI: 10.1021/ja305091n.

[124] Piotr Klukowski, Roland Riek, and Peter Güntert. "Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA". In: *Nature Communications* 13.1 (2022). DOI: 10.1038/s41467-022-33879-5.

[125] Rasmus Linser et al. "Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins." In: *Angewandte Chemie (International ed. in English)* 50 (19 May 2011), pp. 4508–4512. ISSN: 1521-3773. DOI: 10.1002/anie.201008244.

[126] Phineus R. L. Markwick, Thérèse Malliavin, and Michael Nilges. "Structural Biology by NMR: Structure, Dynamics, and Interactions". In: *PLOS Computational Biology* 4.9 (2008). Ed. by Johanna McEntyre, e1000168. DOI: 10.1371/journal.pcbi.1000168.

[127] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[128] Ken A. Dill et al. "The Protein Folding Problem". In: *Annual Review of Biophysics* 37.1 (2008), pp. 289–316. DOI: 10.1146/annurev.biophys.37.092707.153558.

[129] Andriy Kryshtafovych et al. "Critical assessment of methods of protein structure prediction (CASP)-Round XIV". In: *Proteins: Structure, Function, and Bioinformatics* 89.12 (2021), pp. 1607–1617. DOI: 10.1002/prot.26237.

[130] *AlphaFold*. 2021. URL: https://github.com/deepmind/alphafold (visited on 06/13/2023).

[131] Milot Mirdita et al. "ColabFold: making protein folding accessible to all". In: *Nature Methods* 19.6 (2022), pp. 679–682. DOI: 10.1038/s41592-022-01488-1.

[132] Richard Evans et al. "Protein complex prediction with AlphaFold-Multimer". In: (2021). DOI: 10.1101/2021.10.04.463034.

[133] Mihaly Varadi et al. "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models". In: *Nucleic Acids Research* 50.D1 (2021), pp. D439–D444. DOI: 10.1093/nar/gkab1061.

[134] Anastassis Perrakis and Titia K Sixma. "AI revolutions in biology". In: *EMBO reports* 22.11 (2021). DOI: 10.15252/embr.202154046.

[135] Andrew B. Ward, Andrej Sali, and Ian A. Wilson. "Integrative Structural Biology". In: *Science* 339.6122 (2013), pp. 913–915. DOI: 10.1126/science.1228565.

[136] Maximilian Zinke et al. "Architecture of the flexible tail tube of bacteriophage SPP1". In: *Nature Communications* 11.1 (2020). DOI: 10.1038/s41467-020-19611-1.

[137] Lolita Piersimoni et al. “Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions-A Method for All Seasons”. In: *Chemical Reviews* 122.8 (2021), pp. 7500–7531. DOI: 10.1021/acs.chemrev.1c00786.

[138] Indra D. Sahu and Gary A. Lorigan. “Electron Paramagnetic Resonance as a Tool for Studying Membrane Proteins”. In: *Biomolecules* 10.5 (2020), p. 763. DOI: 10.3390/biom10050763.

[139] Samuel Naudi-Fabra, Martin Blackledge, and Sigrid Milles. “Synergies of Single Molecule Fluorescence and NMR for the Study of Intrinsically Disordered Proteins”. In: *Biomolecules* 12.1 (2021), p. 27. DOI: 10.3390/biom12010027.

[140] Roberto Tejero et al. “AlphaFold Models of Small Proteins Rival the Accuracy of Solution NMR Structures”. In: *Frontiers in Molecular Biosciences* 9 (2022). DOI: 10.3389/fmolb.2022.877000.

[141] Nicholas J. Fowler and Mike P. Williamson. “The accuracy of protein structures in solution determined by AlphaFold and NMR”. In: *Structure* 30.7 (2022), 925–933.e2. DOI: 10.1016/j.str.2022.04.005.

[142] G. Bertani. “Studies on Lysogensis I”. In: *Journal of Bacteriology* 62.3 (1951), pp. 293–300. DOI: 10.1128/jb.62.3.293-300.1951.

[143] Donghua H. Zhou and Chad M. Rienstra. “High-performance solvent suppression for proton detected solid-state NMR”. In: *Journal of Magnetic Resonance* 192.1 (2008), pp. 167–172. DOI: 10.1016/j.jmr.2008.01.012.

[144] A. J. Shaka et al. “An improved sequence for broadband decoupling: WALTZ-16”. In: *Journal of Magnetic Resonance (1969)* 52.2 (1983), pp. 335–338. DOI: 10.1016/0022-2364(83)90207-X.

[145] Dominique Marion et al. “Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins”. In: *Journal of Magnetic Resonance* 85 (1989), pp. 393–399. DOI: 10.1016/0022-2364(89)90152-2.

[146] A. E. Bennett et al. “Chemical shift correlation spectroscopy in rotating solids: Radio frequency-driven dipolar recoupling and longitudinal exchange”. In: *The Journal of Chemical Physics* 96.11 (1992), pp. 8624–8627. DOI: 10.1063/1.462267.

[147] Bing-Man Fung, Aanatoly K Khitrin, and Konstantin Ermolaev. “An improved broadband decoupling sequence for liquid crystals and solids”. In: *Journal of Magnetic Resonance* 142.1 (2000), pp. 97–101. DOI: 10.1006/jmre.1999.1896.

[148] Józef R. Lewandowski et al. “Measurement of Site-Specific ^{13}C Spin-Lattice Relaxation in a Crystalline Protein”. In: *Journal of the American Chemical Society* 132.24 (2010), pp. 8252–8254. DOI: 10.1021/ja102744b.

[149] Günther Metz, X. L. Wu, and Steven O. Smith. “Ramped-amplitude cross polarization in magic-angle-spinning NMR”. In: *Journal of Magnetic Resonance, Series A* 110.2 (1994), pp. 219–227. DOI: 10.1006/jmra.1994.1208.

[150] Koichiro Tamura, Glen Stecher, and Sudhir Kumar. “MEGA11: Molecular Evolutionary Genetics Analysis Version 11”. In: *Molecular Biology and Evolution* 38.7 (2021). Ed. by Fabia Ursula Battistuzzi, pp. 3022–3027. DOI: 10.1093/molbev/msab120.

[151] M. A. Larkin et al. “Clustal W and Clustal X version 2.0”. In: *Bioinformatics* 23.21 (2007), pp. 2947–2948. DOI: 10.1093/bioinformatics/btm404.

[152] Christoph Gille et al. “Alignment-Annotator web server: rendering and annotating sequence alignments”. In: *Nucleic Acids Research* 42.W1 (2014), W3–W6. DOI: 10.1093/nar/gku400.

[153] Paul Schanda, Ēriks Kupče, and Bernhard Brutscher. “SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds”. In: *Journal of Biomolecular NMR* 33.4 (2005), pp. 199–211. DOI: 10.1007/s10858-005-4425-x.

[154] Michael Sattler, Jürgen Schleucher, and Christian Griesinger. “Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 34.2 (1999), pp. 93–158. DOI: 10.1016/s0079-6565(98)00025-9.

[155] Ad Bax, G. Marius Clore, and Angela M. Gronenborn. “^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins”. In: *Journal of Magnetic Resonance (1969)* 88.2 (1990), pp. 425–431. DOI: 10.1016/0022-2364(90)90202-k.

[156] M. Kadkhodaie et al. “Broadband homonuclear cross polarization using flip-flop spectroscopy”. In: *Journal of Magnetic Resonance* 91.2 (1991), pp. 437–443. DOI: 10.1016/0022-2364(91)90210-k.

[157] Lewis E. Kay et al. “A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain ^{1}H and ^{13}C Correlations in H_2O Samples of Proteins”. In: *Journal of Magnetic Resonance, Series B* 101.3 (1993), pp. 333–337. DOI: 10.1006/jmrb.1993.1053.

[158] Alexander S. Maltsev, Jinfa Ying, and Ad Bax. “Deuterium isotope shifts for backbone ^{1}H, ^{15}N and ^{13}C nuclei in intrinsically disordered protein α-synuclein”. In: *Journal of biomolecular NMR* 54.2 (2012), pp. 181–191. DOI: 10.1007/s10858-012-9666-x.

[159] Stephen W. Fesik and Erik R. P. Zuiderweg. “Heteronuclear three-dimensional nmr spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra”. In: *Journal of Magnetic Resonance (1969)* 78.3 (1988), pp. 588–593. DOI: 10.1016/0022-2364(88)90144-8.

[160] Simon P. Skinner et al. “CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis”. In: *Journal of Biomolecular NMR* 66.2 (2016), pp. 111–124. DOI: 10.1007/s10858-016-0060-y.

[161] Wim F. Vranken et al. “The CCPN data model for NMR spectroscopy: development of a software pipeline”. In: *Proteins: Structure, Function, and Bioinformatics* 59.4 (2005), pp. 687–696. DOI: 10.1002/prot.20449.

[162] Martin Ballaschk. “Kernmagnetresonanzspektroskopische Untersuchungen der Dynamik von HLA-B*27”. PhD thesis. Free University Berlin, 2016. URL: https://refubium.fu-berlin.de/handle/fub188/6488.

[163] VladimÍr Sklenář, Dennis Torchia, and Ad Bax. “Measurement of carbon-13 longitudinal relaxation using 1H detection”. In: *Journal of Magnetic Resonance (1969)* 73.2 (1987), pp. 375–379. DOI: 10.1016/0022-2364(87)90214-9.

[164] Lewis E. Kay, Dennis A. Torchia, and Ad Bax. “Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease”. In: *Biochemistry* 28.23 (1989), pp. 8972–8979. DOI: 10.1021/bi00449a003.

[165] Mike P. Williamson. “Using chemical shift perturbation to characterise ligand binding”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 73 (2013), pp. 1–16. DOI: 10.1016/j.pnmrs.2013.02.001.

[166] Wolfgang Rieping et al. “ARIA2: Automated NOE assignment and data integration in NMR structure calculation”. In: *Bioinformatics* 23.3 (2006), pp. 381–382. DOI: 10.1093/bioinformatics/btl589.

[167] Benjamin Bardiaux, Thérèse Malliavin, and Michael Nilges. “ARIA for Solution and Solid-State NMR”. In: *Methods in Molecular Biology*. Humana Press, 2011, pp. 453–483. DOI: 10.1007/978-1-61779-480-3_23.

[168] Axel T. Brunger. “Version 1.2 of the Crystallography and NMR system”. In: *Nature Protocols* 2.11 (2007), pp. 2728–2733. DOI: 10.1038/nprot.2007.406.

[169] Lena Buchner and Peter Güntert. “Increased Reliability of Nuclear Magnetic Resonance Protein Structures by Consensus Structure Bundles”. In: *Structure* 23.2 (2015), pp. 425–434. DOI: 10.1016/j.str.2014.11.014.

[170] Yang Shen et al. “TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts”. In: *Journal of Biomolecular NMR* 44.4 (2009), pp. 213–223. DOI: 10.1007/s10858-009-9333-z.

[171] Fabien Mareuil et al. “Improved reliability, accuracy and quality in automated NMR structure calculation with ARIA”. In: *Journal of Biomolecular NMR* 62.4 (2015), pp. 425–438. DOI: 10.1007/s10858-015-9928-5.

[172] Torsten Herrmann, Peter Güntert, and Kurt Wüthrich. “Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA”. In: *Journal of Molecular Biology* 319.1 (2002), pp. 209–227. DOI: 10.1016/s0022-2836(02)00241-3.

[173] Jens P. Linge et al. “Refinement of protein structures in explicit solvent”. In: *Proteins: Structure, Function, and Bioinformatics* 50.3 (2003), pp. 496–506. DOI: 10.1002/prot.10299.

[174] Michael Nilges et al. “Accurate NMR Structures Through Minimization of an Extended Hybrid Energy”. In: *Structure* 16.9 (2008), pp. 1305–1312. DOI: 10.1016/j.str.2008.07.008.

[175] Thomas A. Halgren et al. “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening”. In: *Journal of Medicinal Chemistry* 47.7 (2004), pp. 1750–1759. DOI: 10.1021/jm030644s.

[176] Daniel Stöppler et al. “Insight into small molecule binding to the neonatal Fc receptor by X-ray crystallography and 100 kHz magic-angle-spinning NMR”. In: *PLOS Biology* 16.5 (2018). Ed. by Ann Stock, e2006192. DOI: 10.1371/journal.pbio.2006192.

[177] Anne Diehl et al. “Structural changes of TasA in biofilm formation of *Bacillus subtilis*”. In: *Proceedings of the National Academy of Sciences* (2018). ISSN: 0027-8424. DOI: 10.1073/pnas.1718102115.

[178] Michael Levitt and Max F. Perutz. “Aromatic rings act as hydrogen bond acceptors”. In: *Journal of Molecular Biology* 201.4 (1988), pp. 751–754. DOI: 10.1016/0022-2836(88)90471-8.

[179] Kumaran Baskaran et al. “Anomalous amide proton chemical shifts as signatures of hydrogen bonding to aromatic sidechains”. In: *Magnetic Resonance* 2.2 (2021), pp. 765–775. DOI: 10.5194/mr-2-765-2021.

[180] Kurt Wüthrich, Martin Billeter, and Werner Braun. “Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances”. In: *Journal of Molecular Biology* 180.3 (1984), pp. 715–740. DOI: 10.1016/0022-2836(84)90034-2.

[181] Hao-Bo Guo et al. “AlphaFold2 models indicate that protein sequence determines both structure and dynamics”. In: *Scientific Reports* 12.1 (2022). DOI: 10.1038/s41598-022-14382-9.

[182] Manuela K. Hospenthal et al. “The Cryoelectron Microscopy Structure of the Type 1 Chaperone-Usher Pilus Rod”. In: *Structure* 25.12 (2017), 1829–1838.e4. DOI: 10.1016/j.str.2017.10.004.

[183] Natalia Pakharukova et al. “Archaic chaperone–usher pili self-secrete into superelastic zigzag springs”. In: *Nature* 609.7926 (2022), pp. 335–340. DOI: 10.1038/s41586-022-05095-0.

[184] Victoria A. Higman et al. “Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively ^{13}C-labelled proteins”. In: *Journal of biomolecular NMR* 44.4 (2009), pp. 245–260. DOI: 10.1007/s10858-009-9338-7.

[185] Linda J. Ball et al. “Recognition of Proline-Rich Motifs by Protein-Protein-Interaction Domains”. In: *Angewandte Chemie International Edition* 44.19 (2005), pp. 2852–2869. DOI: 10.1002/anie.200400618.

[186] Franziska Leidreiter et al. “Common architecture of Tc toxins from human and insect pathogenic bacteria”. In: *Science Advances* 5.10 (2019). DOI: 10.1126/sciadv.aax6497.

[187] Alexey Drozdetskiy et al. “JPred4: a protein secondary structure prediction server”. In: *Nucleic Acids Research* 43.W1 (2015), W389–W394. DOI: 10.1093/nar/gkv332.

[188] Dene R. Littler et al. “Structure-function analyses of a pertussis-like toxin from pathogenic *Escherichia coli* reveal a distinct mechanism of inhibition of trimeric G-proteins”. In: *Journal of Biological Chemistry* 292.36 (2017), pp. 15143–15158. DOI: 10.1074/jbc.m117.796094.

[189] Diego Romero and Roberto Kolter. “Functional amyloids in bacteria.” In: *International microbiology : the official journal of the Spanish Society for Microbiology* 17 (2 June 2014), pp. 65–73. ISSN: 1139-6709. DOI: 10.2436/20.1501.01.208.

[190] Anna Dragoš et al. “Division of Labor during Biofilm Matrix Production”. In: *Current Biology* 28.12 (2018), 1903–1913.e5. DOI: 10.1016/j.cub.2018.04.046.

[191] Anna Dragoš et al. “Collapse of genetic division of labour and evolution of autonomy in pellicle biofilms”. In: *Nature Microbiology* 3.12 (2018), pp. 1451–1460. DOI: 10.1038/s41564-018-0263-y.

[192] Karel H. M. van Wely et al. “Translocation of proteins across the cell envelope of Gram-positive bacteria”. In: *FEMS Microbiology Reviews* 25.4 (2001), pp. 437–454. DOI: 10.1016/s0168-6445(01)00062-6.

[193] Sean-Paul Nuccio and Andreas J. Bäumler. “Evolution of the Chaperone/Usher Assembly Pathway: Fimbrial Classification Goes Greek”. In: *Microbiology and Molecular Biology Reviews* 71.4 (2007), pp. 551–575. DOI: 10.1128/mmbr.00014-07.

[194] Satoshi Shibata et al. “Structure of polymerized type V pilin reveals assembly mechanism involving protease-mediated strand exchange”. In: *Nature Microbiology* 5.6 (2020), pp. 830–837. DOI: 10.1038/s41564-020-0705-1.

[195] Jan Böhning et al. “Architecture of the biofilm-associated archaic Chaperone-Usher pilus CupE from *Pseudomonas aeruginosa*”. In: *PLOS Pathogens* 19.4 (2023). Ed. by Matthew A. Mulvey, e1011177. DOI: 10.1371/journal.ppat.1011177.

[196] Sivan Elias and Ehud Banin. “Multi-species biofilms: living with friendly neighbors”. In: *FEMS Microbiology Reviews* 36.5 (2012), pp. 990–1004. DOI: 10.1111/j.1574-6976.2012.00325.x.

[197] Thomas D. Pollard. “Actin and Actin-Binding Proteins”. In: *Cold Spring Harbor Perspectives in Biology* 8.8 (2016), a018226. DOI: 10.1101/cshperspect.a018226.

[198] Klaus Aktories et al. “Actin as target for modification by bacterial protein toxins”. In: *FEBS Journal* 278.23 (2011), pp. 4526–4543. DOI: 10.1111/j.1742-4658.2011.08113.x.

[199] Sabrina Pospich et al. “High-resolution structures of the actomyosin-V complex in three nucleotide states provide insights into the force generation mechanism”. In: *eLife* 10 (2021). DOI: 10.7554/elife.73724.

[200] Alexander Belyy et al. “Mechanism of actin-dependent activation of nucleotidyl cyclase toxins from bacterial human pathogens”. In: *Nature Communications* 12.1 (2021). DOI: 10.1038/s41467-021-26889-2.

[201] Alexander Belyy et al. “Structure of the Lifeact-F-actin complex”. In: *PLOS Biology* 18.11 (2020). Ed. by Carole A. Parent, e3000925. DOI: 10.1371/journal.pbio.3000925.

[202] Kurt Peter C. Vollhardt and Neil E. Schore. *Organische Chemie*. ger. Ed. by Holger Butenschön. 6th ed. pp. 266-314. Weinheim: Wiley-VCH, 2020. ISBN: 978-3-527-34582-3.

[203] Michael S. Cohen and Paul Chang. “Insights into the biogenesis, function, and regulation of ADP-ribosylation”. In: *Nature Chemical Biology* 14.3 (2018), pp. 236–243. DOI: 10.1038/nchembio.2568.

[204] Toshiharu Tsurumura et al. “Arginine ADP-ribosylation mechanism based on structural snapshots of iota-toxin and actin complex”. In: *Proceedings of the National Academy of Sciences* 110.11 (2012), pp. 4267–4272. DOI: 10.1073/pnas.1217227110.

[205] Leif Schröder, Christian Schmitz, and Peter Bachert. “Cumulative “roof effect” in high-resolution *in vivo* ^{31}P NMR spectra of human calf muscle and the Clebsch-Gordan coefficients of ATP at 1.5 T”. In: *Journal of Magnetic Resonance* 174.1 (2005), pp. 68–77. DOI: 10.1016/j.jmr.2005.01.011.

[206] Katie L. Thoren et al. “Lethal factor unfolding is the most force-dependent step of anthrax toxin translocation”. In: *Proceedings of the National Academy of Sciences* 106.51 (2009), pp. 21555–21560. DOI: 10.1073/pnas.0905880106.

[207] Bryan A. Krantz et al. “Acid-induced Unfolding of the Amino-terminal Domains of the Lethal and Edema Factors of Anthrax Toxin”. In: *Journal of Molecular Biology* 344.3 (2004), pp. 739–756. DOI: 10.1016/j.jmb.2004.09.067.

[208] Patrick Günther et al. “Structure of a bacterial Rhs effector exported by the type VI secretion system”. In: *PLOS Pathogens* 18.1 (2022). Ed. by Karla J.F. Satchell, e1010182. DOI: 10.1371/journal.ppat.1010182.

[209] Dukas Jurėnas et al. “Mounting, structure and autocleavage of a type VI secretion-associated Rhs polymorphic toxin”. In: *Nature Communications* 12.1 (2021). DOI: 10.1038/s41467-021-27388-0.

[210] Maarten Schledorn et al. “Protein NMR Spectroscopy at 150 kHz Magic-Angle Spinning Continues To Improve Resolution and Mass Sensitivity”. In: *ChemBioChem* 21.17 (2020), pp. 2540–2548. DOI: 10.1002/cbic.202000341.

[211] Zdeněk Tošner et al. “Maximizing efficiency of dipolar recoupling in solid-state NMR using optimal control sequences”. In: *Science Advances* 7.42 (2021). DOI: 10.1126/sciadv.abj5913.

[212] Morgane Callon et al. “Biomolecular solid-state NMR spectroscopy at 1200 MHz: the gain in resolution”. In: *Journal of Biomolecular NMR* 75.6-7 (2021), pp. 255–272. DOI: 10.1007/s10858-021-00373-x.

[213] Lisa Maria Gerland. “Investigation of Protonation-Related Processes in Light-Absorbing and Converting Proteins by NMR”. PhD thesis. Free University Berlin, 2022. URL: https://refubium.fu-berlin.de/handle/fub188/38813.

[214] Joren Sebastian Retel. “Structure Determination of Outer Membrane Protein G in Native Lipids by Solid-State NMR Spectroscopy”. PhD thesis. Free University Berlin, 2016. URL: https://refubium.fu-berlin.de/handle/fub188/9261.

Acknowledgments

As much as science is a team effort, so is life. There probably is an impossibly large number of people that I would like to thank for guiding, helping, and supporting me on my PhD journey. The biggest thanks go to my girlfriend Naphat, my family, and my friends that have accompanied me over these more than 10 years of study.

I would like to thank Hartmut Oschkinat for supervising me through these exciting years of doctoral study and being the first reviewer of this thesis. Furthermore, I would like to thank Christian Freund for agreeing to be the second reviewer of my work.

Again as in science, so in life, there is often not a clear cutoff where one characteristic ends and another one starts. Many of my scientific co-workers I do also consider valuable companions or friends. I would like to thank Hartmut Oschkinat for personal as well as scientific support and for giving me the freedom to pursue scientific approaches that I deemed relevant. Another big chunk of thanks goes to my long-term office-mate Lisa Gerland. Thank you very much for continued 'Kicktipp' assistance, a great office atmosphere, and for accompanying me for during my PhD. On this note, I want to thank my former colleagues that worked in the same office as myself: Daniel Friedrich (né Stöppler), Michel Geiger, and Amira Gutmann-Trieb. For invaluable scientific advice, I want to thank Wing Ying Chow, Miguel Arbesú, and Barth van Rossum. A very big THANK YOU goes to Matthias Herrera Glomm (the scientist formerly known as Hiller) for scientific help, personal counsel, and maintaining the solid–state NMR facility in a breathtaking manner. You have given me the most heartfelt counsel during these years. I also would like to thank Peter Schmieder and everyone maintaining the solution NMR facility.

Although the scientific contribution was minor, I want to thank Andrea Steuer for her constant, reliable duty as secretary and diligent maintenance of Hartmut's calendar. You definitely had a significant non-scientific contribution! I also would like to thank everyone working in the wet lab. Of course, Anne Diehl for her supervising role there and Martina Leidert (who spent years of her time preparing TasA samples - Thank You!), Natalja Erdmann (for sample assistance), and Nils Cremer (for sample assistance as well but also for small talk).

Furthermore, I also would like to thank our collaborators on both projects. For the TasA project, thank you to Yvette Roske and Kürşad Turgay. For the TccC3 project, big thanks to Alexander Belyy, Tillmann Utesch, Benjamin Bardiaux, Jonas Protze, Daniel Friedrich, Peter Schmieder, Daniel Roderer and Stefan Raunser. With you as a co-author Alex, science didn't seem so daunting. Thank you for all your postdoctoral advice.

Lastly, I want to thank everyone who helped me in the creation of the written thesis. Thank you to Barth van Rossum and Lisa Gerland for diligent error-checking and advice on figure coloring. Thank you Barth especially for providing me with Fig. 5.2 and a big chunk of aesthetic advice, which really helped making the figures more intuitive. In addition, thanks to Alexander Belyy and Daniel Roderer for supplying me with figure templates. I would also thank everyone that contributed further to error-checking: Samuel Naudi-Fabra (I really asked you a lot), Matthias Herrera Glomm, Naphat, Duc Khoi Do, Dallas Bell, and Leon Holly.

Curriculum Vitae

For reasons of data protection, the curriculum vitae is not published in the online version.