Supramolecular biological protein assemblies study using solid-state NMR

Nadia El Mammeri

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

SPÉCIALITÉ CHIMIE PHYSIQUE

Par Nadia EL MAMMERI

Étude d'assemblages protéiques biologiques supramoléculaires par RMN du solide

Sous la direction de : Dr. Antoine LOQUET

Soutenue le 04/12/2020

Membres du jury :

M. DUFOURC, Erick J.	Directeur de Recherche, CNRS, Bordeaux	Président
Mme BÖCKMANN, Anja	Directrice de Recherche, CNRS, Lyon	Rapporteur
Mme FRANCETIC, Olivera	Directrice de Recherche, Institut Pasteur, Paris	Rapporteur
M. SAUPE, Sven	Directeur de Recherche, CNRS, Bordeaux	Examinateur
M. MACKERETH, Cameron	Directeur de Recherche, INSERM, Bordeaux	Examinateur
M. KOVACS, Ákos T.	Professeur, DTU, Danemark	Examinateur
M. GALAN, Jorge E.	Professeur, Université Yale, USA	Examinateur
M. LOQUET, Antoine	Directeur de Recherche, CNRS, Bordeaux	Directeur de thèse

To my parents

And Reda, Sophia, Amine,
Anissa, Salma, Amel, Lina, Adam

Titre : Étude d'assemblages protéiques biologiques supramoléculaires par RMN du solide

Résumé : Ma thèse a pour buts l'élucidation fonctionnelle et l'étude structurale de nano-machines biologiques en utilisant principalement la RMN du solide : (i) La protéine SesB trouvée chez *Nectria haematococca*, dont l'assemblage en fibrilles amyloïdes, serait impliquée dans les mécanismes de signalisation de mort cellulaire programmée. En utilisant la RMN du solide avec rotation à l'angle magique, un nouveau modèle structural de cet amyloïde a été établi pour les fibrilles amyloïdes SesB de *Nectria haematococca*. (ii) Des protéines appelées TasA, TapA et CalY dont l'assemblage en fibrilles amyloïdes est impliqué dans la formation et l'intégrité des biofilms bactériens chez *Bacillus subtilis* et *Bacillus cereus*. Des différences structurales ont été observées entre ces deux espèces. *Bacillus subtilis* et quelques mutants (mutations de gènes impliqués dans la formation de biofilms) ont été étudiés en utilisant la RMN du solide sur cellules entières pour comprendre l'impact de la délétion de ces composants sur la paroi cellulaire globale et la composition de la matrice. (iii) HET-s, l'amyloïde fonctionnelle fongique aujourd'hui très étudié, est utilisé ici comme système modèle dans une stratégie visant à utiliser les avantages combinés de la DNP, de la rotation très rapide à l'angle magique et de la dilution de spin dans le contexte d'études de biologie structurale. Il a été prouvé qu'une utilisation pertinente de ces techniques constitue une stratégie potentiellement robuste pour la caractérisation structurale des assemblages supramoléculaires pour augmenter la sensibilité des analyses par RMN. (iv) Des polymères protéiques inhabituels, en forme de ruban, appelés R-bodies (« refractile bodies » de type 51), que l'on trouve chez de nombreuses espèces bactériennes, comme *Caedibacter* et *Pseudomonas*. Une attribution complète des signaux RMN des R-bodies a été réalisée en utilisant la RMN du solide avec rotation à l'angle magique à très grande vitesse. Un modèle structurel des monomères est établi et tient compte de leur changement conformationnel en fonction du pH, ainsi que de leur capacité à relarguer des biomolécules.

Mots clés : assemblages supramoléculaires, biologie structurale, RMN du solide, machine bactérienne

Title: Supramolecular biological protein assemblies study using solid-state NMR

Abstract: I have worked on elucidating and studying biological nanomachines using mainly solid-state NMR (SSNMR): (i) The protein SesB found in *Nectria haematococca* whose assembly into amyloid fibrils is thought to be involved in programmed-cell death signalling mechanisms. Using magic-angle spinning SSNMR, a novel structural amyloid fold model for *Nectria haematococca* amyloid fibrils was established. (ii) Proteins called TasA, TapA, and CalY found in *Bacillus subtilis* and *Bacillus cereus* whose assembly into amyloid fibrils is involved in biofilm formation and integrity. *Bacillus subtilis* and its mutants have been studied using whole-cell SSNMR to understand the impact of biofilm matrix components deletion in the overall cell wall, and matrix composition. The protein TasA found in *Bacillus subtilis* has been observed to perturb liposomes as membrane models. (iii) HET-s, the well documented fungal functional amyloid, used here as a model system in a scheme to use the combined advantages of DNP, fast MAS, and spin dilution in the context of structural biology studies. A relevant use of DNP in combination with fast MAS and specific labelling has been proven to be a potential robust strategy for structural characterization of supramolecular assemblies. (iv) Unusual, ribbon-like protein polymers called R-bodies (Type 51 refractile bodies) found in many bacterial species, such as *Caedibacter* and *Pseudomonas*. A complete resonance assignment of R- bodies was achieved using very fast MAS SSNMR. A structural model of the monomers was established and accounts for their interesting pH-dependent switch, as well as the ability to deliver biomolecules.

Keywords: Supramolecular assemblies, structural biology, solid-state NMR, bacterial machine

---

UNITE DE RECHERCHE

[CBMN, UMR 5248, 1 Allée Geoffroy Saint-Hilaire, 33600 Pessac]

ACKNOWLEDGMENTS / REMERCIEMENTS

First, I would like to express my most sincere gratitude to Antoine Loquet for allowing me to conduct this research under his auspices. I am especially grateful for his trust, advice and the independence he allowed me throughout the years. I take the time to also gratefully thank Erick Dufourc, who supported me in all stages of my PhD research and after. He always gave me constant encouragement, and support despite his busy schedule.

I am deeply grateful to all the members of the jury. I thank Anja Böckmann, and Olivera Francetic for agreeing to read the manuscript and evaluate my thesis, and Erick J. Dufourc, Sven Saupe, Cameron Mackereth, Ákos Kovacs and Jorge E. Galán for accepting to participate in the defense of this thesis.

I would like to acknowledge Professor Jorge E. Galán who welcomed me in his team in Yale University for five months for a fruitful and enriching collaboration. This experience has been a major part of my personal and professional development. I wish to thank the Fulbright Commission Franco-américaine organization for sponsoring my stay with Professor Jorge E. Galán at Yale University, New Haven, USA.

My thesis work was made possible through numerous collaborations with brilliant researchers. I would like to thank all participants of the following work with whom we had fascinating scientific discussions. I am honoured to have worked with Prof. Jorge E. Galán, Dr. Maria Lara-Tejero, Dr. Guido Pintacuda, Dr. Anne Lesage, Dr. Diego Romero, Professor Justin Kollman, Dr. Sven Saupe, Dr. Brice Kauffmann, Dr. Sophie Lecomte, and Prof. Erick J. Dufourc.

I express my most sincere appreciation to Axelle Grélard, Estelle Morvan, Mélanie Berbon, Antoine Dutour, and Mathilde Bertoni who helped me to learn biochemical and biophysical techniques in my laboratory. They have been a constant support in the past three years. I warmly thank Birgit Habenstein for her support and encouragements. It was a real pleasure to work with Axelle Raymond, and I thank her for her help in the biological lab. I am glad to have met so many interesting and pleasant people in the three years that I have spent in the Loquet team. I have learned from all of you. Thank you!

Gaëlle, Arpita, Mathilde, we'll always have teatime. Gaëlle, Jaya, we'll always have Paris.

Pony, thanks for your time, support, and conversation. After five years, we've finally made it.

Finally, to my family, my parents, Sophia, Reda, Amine, Anissa, Salma, Amel, Lina, Adam and Robert Porifera, it goes without saying, thank you for everything. No need to read any further. I'm even surprised you got this far!

Table of contents

List of Publications

(1) Jan Stanek, Loren B Andreas, Kristaps Jaudzems, Diane Cala, Daniela Lalli, Andrea Bertarello, Tobias Schubeis, Inara Akopjana, Svetlana Kotelovica, Kaspars Tars, Andrea Pica, Serena Leone, Delia Picone, Zhi- Qiang Xu, Nicholas E Dixon, Denis Martinez, Melanie Berbon, **Nadia El Mammeri***, Abdelmajid Noubhani, Sven Saupe, Birgit Habenstein, Antoine Loquet, Guido Pintacuda. (2016) NMR spectroscopic assignment of backbone and side-chain protons in fully protonated proteins: microcrystals, sedimented assemblies, and amyloid fibrils. Angewandte Chemie International Edition, 55: 15504- 15509.

(2) Antoine Loquet, **Nadia El Mammeri***, Jan Stanek, Mélanie Berbon, Benjamin Bardiaux, Guido Pintacuda, Birgit Habenstein. (2018) 3D structure determination of amyloid fibrils using solid-state NMR spectroscopy. Methods, 138: 26-38.

(3) Birgit Habenstein, **Nadia El Mammeri***, James Tolchard, Gaëlle Lamon, Arpita Tawani, Mélanie Berbon, Antoine Loquet. (2019) Structures of type III secretion system needle filaments. Springer, Berlin, Heidelberg book chapter, 1- 23.

(4) **El Mammeri N.***, Hierrezuelo J., Tolchard J., Cámara-Almirón J., Caro-Astorga J., Álvarez-Mena A., Dutour A., Berbon M., Shenoy J., Morvan E., Grélard A., Kauffmann B., Lecomte S., de Vicente A., Habenstein B., Romero D. and Loquet A. (2019), Molecular architecture of bacterial amyloids in Bacillus biofilms. The FASEB Journal, 33: 12146- 12163.

(5) J. Shenoy, **N. El Mammeri***, A. Dutour, M. Berbon, A. Saad, A. Lends, E. Morvan, A. Grélard, S. Lecomte, B. Kauffmann, F.-X. Theillet, B. Habenstein, A. Loquet. (2019) Structural dissection of amyloid aggregates of TDP-43 and its C-terminal fragments TDP-35 and TDP-16. FEBS J, 287: 2449-2467.

(6) A. Daskalov, D. Martinez, V. Coustou, **N. El Mammeri***, M. Berbon, L.B. Andreas, B. Bardiaux, J. Stanek, A. Noubhani, B. Kauffmann, J.S. Wall, G. Pintacuda, S.J. Saupe, B. Habenstein, A. Loquet. (2020) Structural and molecular basis of cross- seeding barriers in amyloids. BioRxiv

Disclaimer

The text and figures of the Introduction section and Chapter 3 have been taken or adapted from published articles. (2) Loquet *et al.*, 3D structure determination of amyloid fibrils using solid-state NMR spectroscopy. Methods, 2018; (4) El Mammeri *et al.*, Molecular architecture of bacterial amyloids in *Bacillus* biofilms. The FASEB Journal, 2019.
For clarity purposes, figures are numbered independently in each chapter and comprise the chapter section number.

Introduction

1.1. SSNMR as a tool for structural biology

1.1.1. Biological assemblies

Supramolecular assemblies of proteins, nucleic acids, lipids, sugars, etc, are very common in most biological cell functions. Solid-state NMR spectroscopy (SSNMR) has been playing a key role in understanding and studying biological samples that are insoluble, non-crystalline, or even intrinsically disordered (1-5). In the past decades a wide range of biological systems has been put through SSNMR scrutiny: viral capsids (6), filaments (7), amyloid fibrils (8), membrane proteins in native environments (9-15), large supramolecular assemblies (16), etc. With crucial milestones including the molecular conformation of Aβ fibrils from Robert Tycko's work (17-20), transthyretin fibril structure from Bob Griffin's laboratory (21, 22), or the structure of the functional amyloid HET-s by Beat Meier (23), SSNMR has now become a major actor in structural biology. SSNMR offers a unique awareness of heterogeneous and "imperfect" biological samples (2) thus, in combination with cryo-electron microscopy (cryo-EM), giving access to a wide scope of biological events ranging from highly ordered assemblies (SSNMR: up to about ~25 kDa; cryo-EM: even higher) to highly polymorphic entities. In addition to the structural information, SSNMR also allows the study of interactions of molecules of different sizes and natures, dynamics, local and overall order, local motions, etc (24-26). SSNMR is often used in synergy with X-ray crystallography, solution NMR, and cryo- or conventional EM in biological studies (27). To date, 135 protein/peptide 3D structures

determined by SSNMR have been deposited at the PDB (Protein Data Bank), including different types of protein complexes such as membrane-associated systems, macromolecular assemblies (filaments, fibrils, capsids), or microcrystalline proteins.

SSNMR methodologies require the use of isotopic $^{13}C/^{15}N$ enrichment of proteins through labelled precursors introduced in the cell cultures prior to protein expression (28-35). Indeed, at natural abundance, cells contain native matter made of ^{1}H at 99.98%, ^{13}C at 1.11%, and ^{15}N at 0.36%, these isotopes being active in NMR. In such conditions, NMR experiments which depend on specific nuclear properties of carbon and nitrogen atoms might be difficult to implement. Various labelling schemes have been developed in the past decades ranging from uniformly labelled samples to highly site-specific $^{13}C/^{15}N$ enrichment (see section 1.2).

Another type of SSNMR application is the study of entire cells, bacteria, fungi, etc, called whole-cell NMR (36). In such studies, isotopic enrichment is applied on the entire cells. Many studies report the detailed analysis of bacterial or plant cell wall components, inclusion bodies, live cells, etc (37-43). With the use of isotopic labelling, multidimensional spectroscopy, and ^{13}C, ^{1}H, ^{15}N, ^{31}P detection, SSNMR groups have reported detailed accounts of the carbohydrate, protein, lipid, and other molecules' composition within complex organisms.

1.1.2. The interesting case of amyloid fibrils

Amyloid-forming proteins have long been a target of choice in the field of SSNMR as their intrinsic polymorphism had long hindered their study at atomic resolution by other biophysical methods (44-51). The term amyloid is by definition suitable for any protein-based supramolecular assembly with birefringence features and the propensity to bind the Congo red dye, displaying non-branched fibrils of approximately 10 nm in diameter (52). The characterization of such unique entities has then been extended to the presence of the so-called cross-β arrangement consisting in stacked β-strands perpendicular to the fibril axis that arrange as intermolecular β-sheets along the fibril axis (**Figure 0-1A**). This structural hallmark, presently considered to be a much more robust portrayal of the amyloid family, is adequately pictured through typical X-ray diffraction patterns (53, 54), in which are inferred some structural characteristics, as typical inter-strand and inter-sheet distances of 4.7 Å and 10 Å, respectively (**Figure 0-1B-C**).

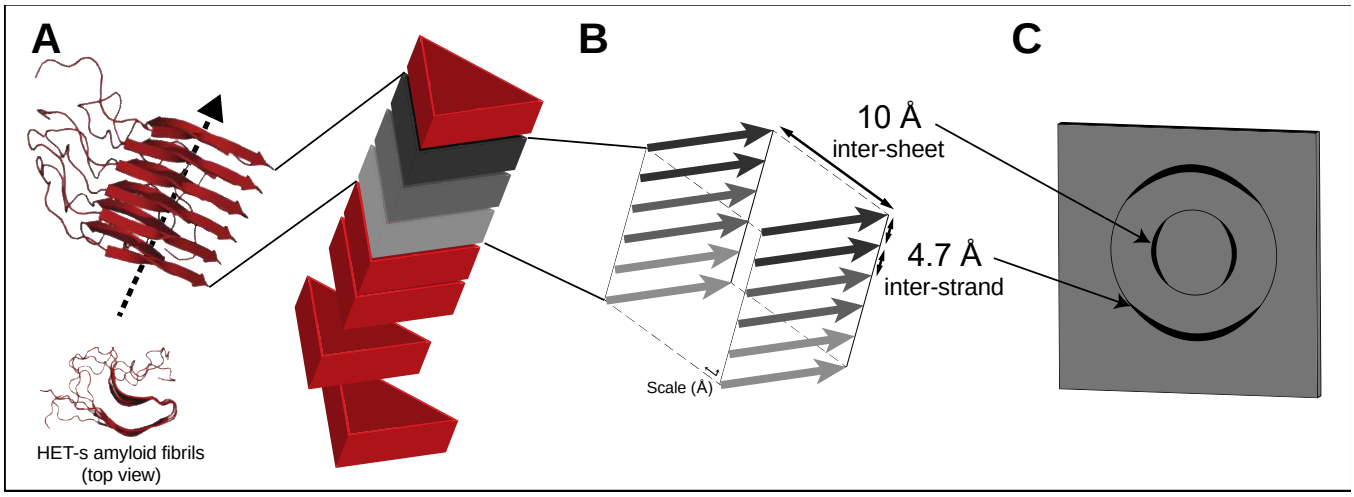


Figure 0-1: Structural features of amyloid fibrils. (A) General morphology and monomer stacking of an amyloid fibril with HET-s fibrils (55) here used as an illustration with a singular two-layer β-solenoid amyloid fold; (B) representation of the so-called cross-β quaternary structure in which β-sheets are stacked along the fibril axis with a 4.7 Å spacing distance as inter-strand interactions, and 10 Å as inter-sheet contacts ; (C) schematic representation of a typical X-ray diffraction pattern, highly representative of the mentioned structural aspects associated with amyloid proteins.

Interestingly, the physician Rudolph Virchow first introduced the word amyloid in 1854, while describing brain-related cellulose-like materials, called *corpora amylacea* (56). It was later demonstrated that amyloid entities were in fact a novel class of proteins, which displays the ability to undergo conformational change and adopt the cross-β architecture, thus forming amyloid fibrils. Such observations paved the way for ground-breaking discoveries in a multitude of research fields. This considerable attention fundamentally arises from their distinct association with a series of deadly (neuro-)degenerative disorders, called amyloidosis, such as Alzheimer's, Parkinson's diseases and type 2 diabetes (49, 57-62).

However, more recently it has become evident that the amyloid family can actually be split into two sub-classes, regarding their pathological or beneficial effects. These amyloids, executing native cellular functions, have been identified in a range of organisms from bacteria to mammals and are termed "functional amyloids" (63-69). Biofilm formation, hormone storage, or cellular signal transduction are some of many examples illustrating the presence of amyloid-like systems involved in biological processes (70-77).

Such duality of effects, non-native aggregation and functional assembly, reveals the complexity of the amyloid fold. Several reports suggest that virtually any protein, when set in the proper conditions/environment, could adopt an amyloid-like β-strand-rich conformation, thus advocating the amyloid fold as a universal global free-energy minimum for polypeptide chains (78-80). Nonetheless, the implication of functional amyloids in selective biological

processes such as cellular signal transduction (71, 74-76) indicate that the recruitment strategies of amyloid monomers can be very specific and not necessarily related to misfolding events.

The non-crystalline and insoluble character of amyloid fibrils poses many obstacles for the structural investigations of these systems at the atomic level, based on well-established techniques such as X-ray crystallography or solution NMR spectroscopy. Additionally, a certain level of structural heterogeneity on the mesoscopic level, often encountered in amyloid fibrils, has hampered the use of cryo-EM until recently (81, 82). SSNMR is a method of choice to achieve keen characterization of supramolecular assemblies in general (4, 5, 25, 83-86), and more specifically of amyloid fibrillar assemblies at high resolution (87-90). This is illustrated by the first structure determination of prion amyloid fibrils that Meier and co-workers achieved in 2008 (55), or by the discovery of the 3D fold of amyloid fibrils of α-synuclein (91) and Aβ (92-96), solved by SSNMR techniques. SSNMR capacity to probe the local order of "imperfect" inhomogeneous biological assemblies stands behind its efficiency over amyloid fibrils (97). Most high-resolution structures of amyloid fibrils have been elucidated using SSNMR as the main method: to date, 35 amyloid protein/peptide structures or structural models have been deposited in the PDB. **Figure 0-2** displays some of the most relevant structures of amyloid fibrils elucidated using SSNMR.

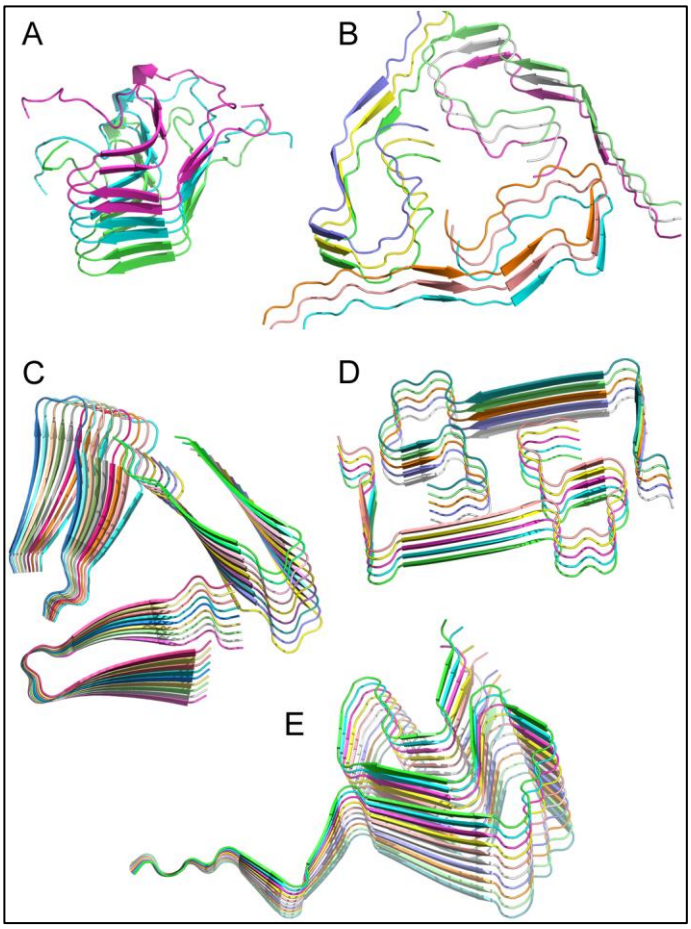


Figure 0-2: Selection of amyloid fibrils atomic structures determined by solid-state NMR. (A) HET-s(218-289) prion fibrils, automated calculation with CYANA (PDB 2KJ3) (23); (B) $A\beta_{1-40}$ fibrils in Alzheimer's disease brain tissue, manual calculation with XPLOR-NIH (PDB 2M4J) (94); (C) $A\beta_{1-40}$ D23N "Iowa" mutant fibrils, manual calculation with Rosetta (PDB 2MPZ) (98); (D) $A\beta_{1-40}$ E22Δ "Osaka" mutant fibrils, automated calculation with CYANA (PDB 2MVX) (99); (E) Human α-synuclein fibrils, manual calculation with XPLOR-NIH (PDB 2N0A) (91). Figure was generated with the PyMOL Molecular Graphics System (Schrödinger, LLC).

1.2. Biochemical methodology

1.2.1. Sample preparation

Both chemical synthesis and recombinant protein production in *Escherichia coli* are used for structure determination of supramolecular assemblies. For whole-cell NMR, the introduction of labelled isotopes can be achieved by providing $^{13}C/^{2}H/^{15}N$ sources in the culture medium. Most small protein/peptide samples (up to about 30-50 residues) are prepared based on well-known Fmoc-protection solid-phase chemical synthesis methods (100, 101), and HPLC-based purification. Higher molecular-weight proteins, however, are usually produced by *E. coli*-based recombinant expression followed by purification steps, although expression in the native organism has also been used. The heterologous expression has the advantage that isotope labeling (^{2}H, ^{13}C, ^{15}N) can be easily implemented by introducing the labeled precursors in the bacterial culture medium. Furthermore, an arsenal of different selective labeling strategies

based on heterologous expression has been developed and successfully applied to determine structures. In the case of amyloid proteins, during bacterial protein production, monomers generally accumulate in inclusion bodies and need to be extracted (*e.g.*, using 8 M urea, or 6 M guanidinium chloride), followed by an enrichment and purification process, which could employ one or several affinity- and/or size-exclusion- and/or hydrophobic-chromatography procedures. Note that the optimal protein production conditions may vary between the production in rich, unlabeled and in poor bacterial expression medium that can be supplemented with labeled or unlabeled ^{2}H, ^{13}C, ^{15}N sources. It is therefore beneficial in terms of cost and time effort to optimize the production in unlabeled poor medium before using isotope-labeled sources.

1.2.2. Assembly and polymorphism

For biological systems that require an *in vitro* post-purification assembly, prior to actual NMR data recording and analysis, the polymerization process represents a significant step in the sample preparation because the conditions during the assembly play a major role in driving the protein subunits to adopt one specific homogeneous molecular structure, a heterogeneous variant or several structural polymorphs in the fibrillar sample. Incubation temperature and duration, pH, protein concentration, nature of the buffer, presence/quantity of additional compounds (*e.g.*, metal ions, chaotropic agents, chelating agents, redox reagents, antibacterial molecules, etc.), protein state upon rotor-packing (*i.e.* hydrated or lyophilized sample), are some of the numerous conditions that need to be thoroughly optimized with both SSNMR and transmission electron microscopy as control tools. The optimization steps are crucial since conformational and structural differences in the fibril arrangement can be observed following only slight changes in sample preparation. Indeed, obtaining highly homogeneous samples is a compelling task in the field of SSNMR and the polymorph selection is often guided by the assembly conditions. The sample heterogeneity and polymorphism observed for amyloid fibrils, for example, is under constant discussion (47, 87, 102), and polymorphism as well as fold instability might carry biological significance. Different structural polymorphs in disease-related amyloids lead to strains that might cause different clinical and pathological phenotypes (103-106) and that can vary in cellular toxicity (107-109). SSNMR spectroscopy serves as a probe of the molecular order at the atomic level, as it can detect slight differences in local conformational states in the assembly. Indeed, the level of sample homogeneity is directly

reflected in the SSNMR line width and has thus a significant impact on the structural analysis (88-90, 97).

1.2.3. Isotope labeling

SSNMR approaches require ^{15}N and ^{13}C isotopic enrichment of biological samples to enhance the signal abundance and allow site-specific studies. Protocols for supramolecular assemblies preparation for SSNMR resemble those employed in solution NMR studies, but differ in that the final sample needs to be assembled from the monomeric subunits or kept assembled (4). The development of isotopic labelling strategies has greatly enhanced the power of SSNMR in structural studies as the labelling schemes allow to selectively retrieve the desired information (*i.e.* a specific position of labelled ^{13}C and ^{15}N nuclei within the amyloid proteins), to reduce dipolar truncation by spin dilution (distribution of labelled ^{13}C sites based on metabolic pathways of the used precursor), and to alleviate spectral congestion (110-113).

^{13}C-^{15}N enrichment can either derive from chemically labelled precursors during solid-phase chemical synthesis, for which the limitations in labelling combination are of chemical nature, or from the carbon sources that are used during recombinant expression. During recombinant protein production, metabolic pathways of the host introduce the ^{13}C-labeling to specific positions in the amino acids (111). The same considerations apply when isotopically enriching whole cells. Numerous selective ^{13}C-labeling strategies have been exploited for SSNMR, such as using [U-^{13}C]-, [1-^{13}C]-, [2-^{13}C]-, [1,3-$^{13}C_2$]- glucose or glycerol derivatives. **Figure 3** displays some of these labelling systems that can be used either for intra-molecular or inter-molecular interaction determination (**Figure 3 -** illustrated with the use of a fibril for clarity purposes). Another production and labelling strategy is the use of cell-free techniques, which is an *in vitro* tool that mimics a bacterial system and is used to produce proteins while keeping a relevant control over $^{13}C/^{15}N$ enrichment (114, 115).

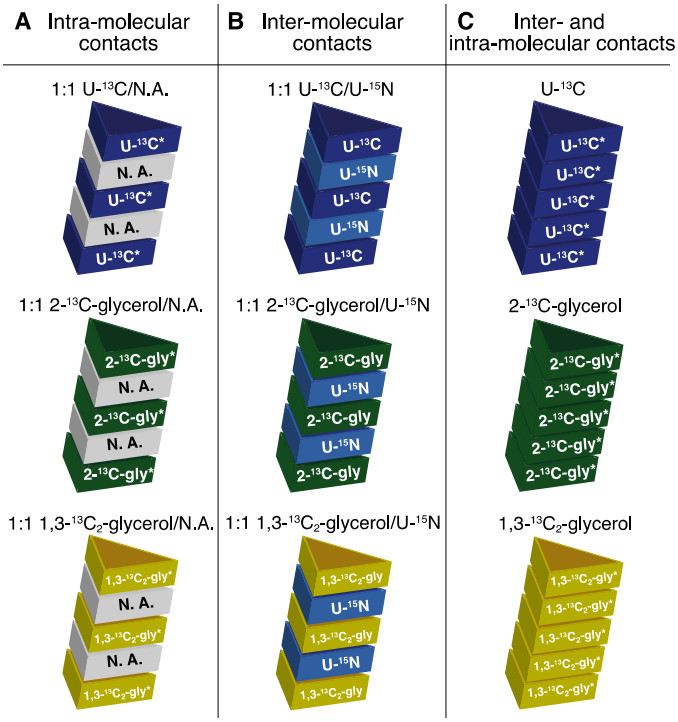


Figure 0-3: Labelling schemes utilized for amyloid structure determination using SSNMR (NB: * is added when ^{15}N-labelling can be added if ^{15}N-detected SSNMR experiments are required). (A) any mix between ^{13}C-labelled and unlabelled (natural abundance, N.A.) proteins will distinctly prevent inter-monomer ^{13}C-^{13}C correlations, thus allowing the unambiguous discrimination between inter- and intra-molecular interactions; (B) any mix between ^{13}C-labelled and ^{15}N-labelled proteins is extremely helpful in determining the fibril stacking registry (*i.e.* in or out); (C) both inter- and intra-monomer interactions can be detected, while specific ^{13}C-glycerol derivatives allow spin dilution.

^{1}H-detected SSNMR is an attractive alternative to ^{13}C-detected methods, which could in perspective alleviate the need for isotopic enrichment. However, ^{1}H-detected SSNMR methods at moderate to fast MAS (Magic Angle Spinning; approx. 20-60 kHz; see section 1.3) need to be coupled not only to ^{13}C- and ^{15}N-labelling to increase the separation of resonances in multidimensional spectra, but also to different levels of deuteration to ensure narrow linewidths in the ^{1}H dimension and sensitivity of detection (*i.e.*, ^{1}H dipolar coupling broadens NMR signals at low MAS). For example, extreme levels of proton dilution in side-chains are obtained by the expression in D_2O/H_2O mixtures (*e.g.*, the ratio of 9:1) and the use of [^{2}H/^{13}C-]glucose as the sole carbon source, followed by a purification in D_2O (116, 117). Alternatively, an excess of D_2O can be used for the expression, and a mixture of D_2O/H_2O can be used during the purification/assembly step to partially reintroduce only exchangeable amide protons in contact with the aqueous buffer (118, 119). However, a high dilution of protons severely limits the probability of observing ^{1}H-^{1}H contacts (120), therefore either a full back-exchange (in 100% H_2O) or a selective protonation in side-chains with a 100% occupancy (*e.g.*, at methyl

sites) has been proposed to record amide-amide (121-123), methyl-amide and methyl-methyl (124-128) interproton distance restraints. However, the amount of potentially available structural information is also significantly reduced, which can impair the high-resolution structure determination. This approach can be used with automation for small model proteins (126), but not on a larger molecular scale if not supplemented by other long-range restraints (129, 130). Also, incomplete exchange of amide ^{1}H can severely hamper the spectral analysis by ^{1}H-detected SSNMR methods, particularly if a refolding protocol for the protein is lacking (119, 131-133). With the advent of fast spinning MAS probes (> 60 kHz), intermediate approaches were proposed, yielding a significantly lower deuteration level in side chains (and in particular, a full occupancy at Hα sites), such as "inverse fractional deuteration" (132) where ^{2}H,^{13}C-glucose and 100% H_2O is used for expression, and either D_2O or H_2O at the purification step. Recently, Meier and co-workers employed partial (about 75%) deuteration for ^{1}H-detection of amyloid fibrils of HET-s(218-289) at 60 kHz MAS, where a H_2O:D_2O ratio of 1:3 was used for expression, and either pure D_2O or H_2O was used for washing the resulting fibrils (134). While this approach favours resolution of $^{1}H^{N}$ and $^{1}H\alpha$ resonances over occupancy of $^{1}H\alpha$ sites, the non-uniform ^{2}H-induced shifts reduce the resolution in the ^{13}C dimension or, for a given sample, cause systematic inconsistencies of $^{13}C\alpha$ and $^{13}C\beta$ shifts in different spectra (or in different regions of the same spectrum) (134). A number of studies have recently demonstrated that in the ultrafast MAS regime (100-111 kHz) a large set of ^{1}H-^{1}H restraints can be obtained for fully-protonated microcrystalline proteins, viral assemblies or membrane-embedded proteins, thus alleviating all issues related to deuteration (133, 135-138). A detailed investigation on ^{1}H linewidths showed that at 111 kHz MAS a minor resolution gain is expected even with a high level of deuteration for microcrystalline proteins (139), while partial deuteration remains beneficial for more dynamic systems such as membrane proteins (140) or fibrils.

The distinction of inter- and intramolecular contacts in ^{1}H-detected SSNMR remains a challenge for sensitivity reasons, and until now is possible with a (uniform) mixed labelling with ^{15}N or ^{13}C, or by deuterating one part of the assembly (141).

1.3. Biophysical methodology

1.3.1. SSNMR experiments for biomolecular assemblies

SSNMR has emerged as the method of choice for investigation of insoluble non-crystalline biopolymers at atomic resolution. The method is mainly based on two tools to

circumvent issues such as line broadening and low signal to noise ratios associated with ^{13}C and ^{15}N. (i) Proteins in solid state move very slowly, so that the spectral lines are broadened considerably compared to solution NMR spectra on soluble proteins, for which the isotropic motion averages out several anisotropic NMR interactions. The use of MAS allows for the recovery of spectral resolution. Indeed, the sample is spun at the magic angle (*i.e.*, 54.7°; this angle averages out several anisotropic interactions that are orientation-dependent) relative to the magnetic field. Therefore, many line-broadening interactions, such as CSA, quadrupolar and dipolar interactions, are partially averaged out by the MAS, leading to significant line-narrowing and an increase in resolution and signal to noise ratio. (ii) Cross-polarization (CP) is a well-established SSNMR experiment used to increase sensitivity of ^{13}C and ^{15}N detection. Indeed, SSNMR usually uses ^{13}C and ^{15}N nuclei, with low gyromagnetic ratios, which require long experimental times to provide suitable signal. CP allows magnetization to be transferred from ^{1}H (high gyromagnetic ratio) to heteroatoms (*e.g.* ^{13}C and 15N) that are dipolar coupled to enhance the signal and decrease the experimental time. To date, mostly ^{13}C-based structure elucidation strategies have been employed on supramolecular assemblies, but ^{1}H-detected fast MAS are increasingly reported for this kind of systems.

1.3.2. Data analysis

1D, 2D, and 3D SSNMR spectra of supramolecular assemblies are usually firstly recorded based on ^{13}C-^{13}C, ^{13}C-^{15}N, ^{13}C-^{15}N-^{13}C, ^{15}N-^{13}C-^{13}C or (^{1}H-)^{15}N-^{1}H correlations in the aim of assigning different resonances to the protein atoms, as chemical shift assignments are a prerequisite for extracting structural information using NMR spectroscopy. The level of complexity during the resonance assignment process typically depends on the diversity of the amino acid composition, the presence or absence of repetitive sequence motifs, the extent of undetectable residues (due to static or dynamic disorder), structural homogeneity in the sample, available SSNMR technology and methodology and also on the availability of other biophysical data. Assigned SSNMR resonances (mainly $^{13}C\alpha$, $^{13}C\beta$) allow the secondary structure determination of all detected residues, based on the use of secondary chemical shifts (142).

When aiming at robust and high-resolution analysis, spectral crowdedness can hamper peak observation, and thus the 3D structural information collection, which is achieved by measuring inter- and intra-molecular distance restraints, obtained using polarization transfers between two atoms (usually ^{1}H, ^{13}C, or ^{15}N), located below ~2-10 Å in space. Spin dilution

achieved through the above-mentioned labelling strategies (*i.e.* ^{13}C-glucose and ^{13}C-glycerol-derived precursors) offers higher spectral resolution by reducing scalar one-bond ^{13}C-^{13}C couplings. In addition, the ^{13}C spin dilution improves the efficiency of polarization transfer between through-space coupled ^{13}C-^{13}C pairs, by decreasing the phenomenon of dipolar truncation (143-145). [1-^{13}C]-, [2-^{13}C]-, [1,3-$^{13}C_2$]- glucose (or glycerol) labelling strategies allow the almost complete removal of one-bond dipolar, and J-couplings, thus considerably improving the spectral resolution, by a factor of ~2 or larger (85, 112, 146, 147). Indeed, the reduced number of ^{13}C-labeled sites lessens the spectral overlap and permits the sequential assignment of short- and medium-range correlations. Many labelling combinations have been developed over the years, in order to improve the data quality, and to target specific structural details. Both intra- and inter-molecular restraints are gathered prior to model calculations as described below.

1.3.3. Restraint detection and identification

For ^{13}C/^{15}N based restraint identification, several labelling and monomer-interaction combinations can be useful (**Figure 3**): (i) mix between ^{13}C-labeled (or ^{13}C-^{15}N) and natural-abundance proteins for intra-molecular contacts (146) only (from 1:1 to 1:4 molar ratios) (**Figure 3A**); (ii) mix between two different labelling schemes (usually contains 50% of U-^{15}N with either 50% of [U-^{13}C]-glucose (148), [2-^{13}C]-, or [1,3-$^{13}C_2$]-glucose (147) or glycerol (149)) for inter-molecular contacts only (usually 1:1 molar ratio) (**Figure 3B**); (iii) [U-^{13}C]-glucose, [2-^{13}C]-, or [1,3-$^{13}C_2$]-glycerol residue labelling for all types of through-space contact detection (**Figure 3C**), based on 2D and 3D spectroscopy (150, 151).

The specific 1:1 [1-^{13}C]-glucose / [2-^{13}C]-glucose mix has been employed, for the measurement of definite intermolecular long-range distances (150, 152). For example, using a 1:1 U-^{15}N/U-^{13}C mixed labelling scheme can reveal the amyloid fibril register (*i.e.*, β-strand packing), by comparing an NCA experiment (usually displaying the backbone correlation ^{15}N/^{13}Cα of a residue) of a U-^{13}C-^{15}N sample and the one of the 1:1 U-^{15}N/U-^{13}C mix. Indeed, an in-register stacking will show perfect overlap of the two experiments, and the ^{13}Cα of a residue *i* will be in direct contact with the backbone ^{15}N labelling of the same residue on the adjacent monomer $i \pm 1$, thus showing the same ^{15}N/^{13}Cα correlations as the one observed in a uniformly ^{13}C-^{15}N labelled sample. Generally, most long-range through-space correlations will appear on the [2-^{13}C]-glycerol and [1,3-$^{13}C_2$]-glycerol samples, and the same labelling scheme diluted in natural-abundance monomers (*i.e.* mixed prior to assembly) should suffice to

discriminate inter- from intra-molecular residue contacts (**Figure 3C**); the applied molar ratio between labelled and unlabelled monomers should reflect a compromise, to prevent both ^{13}C-^{13}C inter-molecular correlations and data with too low signal (*e.g*. having only ~50% of the sites ^{13}C-labeled in the supramolecular assemblies will reduce the signal to noise ratio by ~4).

The ^{13}C-^{13}C correlations, displayed as peaks on the spectra, are usually considered as four different types of contacts: (i) intra-residual ^{13}C-^{13}C correlations, which account for ^{13}C-labeled sites of the same residue; (ii) short-range correlations, which comprise inter-residual interactions (between a residue i and the residues i+1) (also referred to as "sequential correlations"); (iii) medium-range correlations include the correlations from i to i+2 to i+4; (iv) long-range ^{13}C-^{13}C correlations contain i to >i+4 contacts, crucial for determining long-range through-space restraints further used during structure calculations. Indeed, in optimal conditions (*e.g.,* high spectral resolution, high signal to noise ratios, no polymorphism, etc.) a complete and thorough analysis of the SSNMR data should provide a list of ^{13}C-^{13}C or ^{1}H-^{1}H long-range correlations subdivided into four groups: unambiguous inter-molecular, ambiguous inter-molecular, unambiguous intra-molecular, and ambiguous intra-molecular distances; both the quantity and quality of restraints are to be considered. **Figure 4** illustrates such data set with the example of HET-s fibrils (55), in which reported distances involving T233 and I231 are displayed.

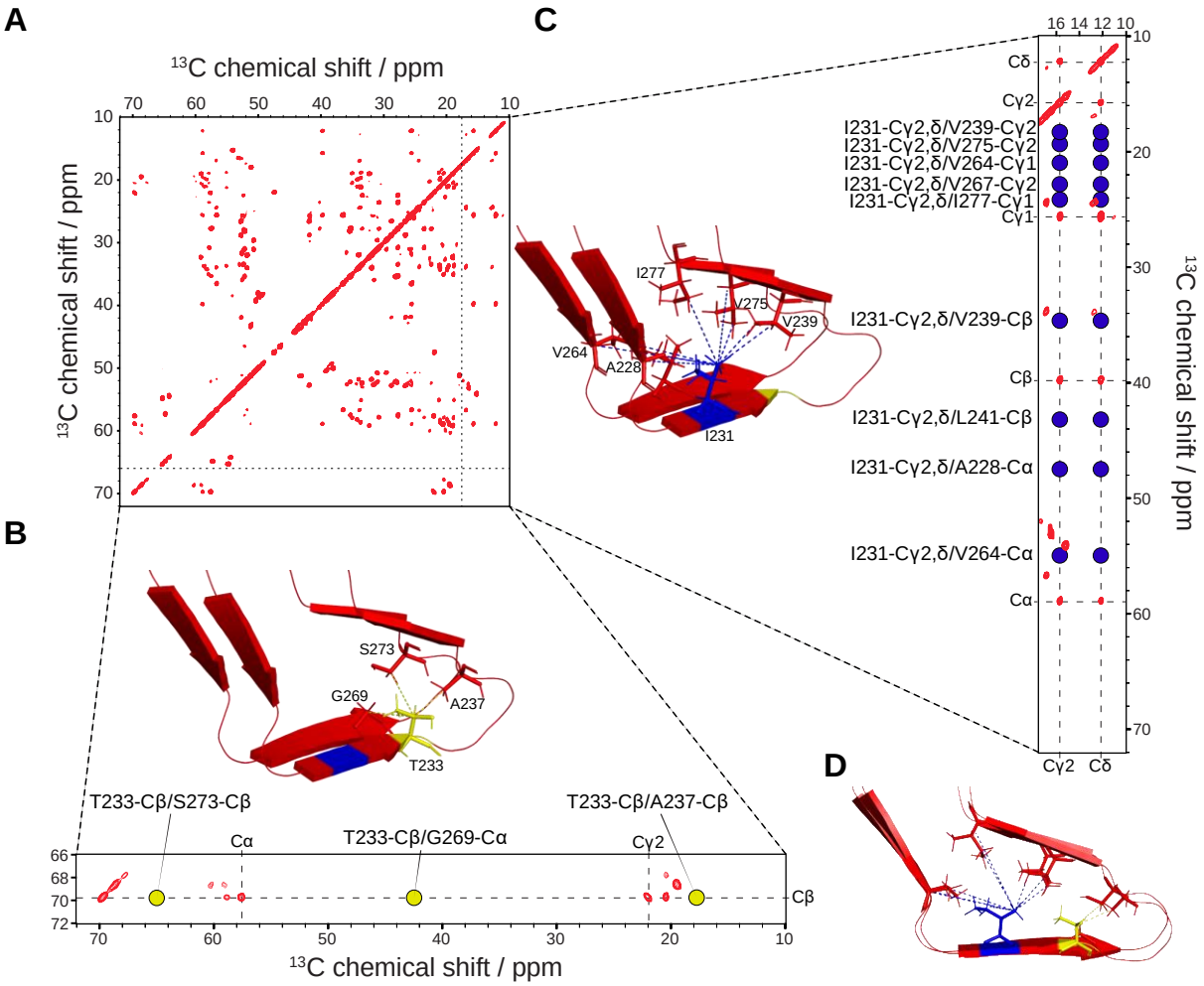


Figure 0-4: Illustration of long range through-space distance determination, based on HET-s(218-289) amyloid core (55). (A) 2D-^{13}C-^{13}C PDSD spectrum (50ms mixing time) of HET-s(218-289) uniformly labelled (^{13}C/^{15}N) recorded on a 800MHz spectrometer (proton frequency). (B) Excerpts of the PDSD spectrum shown in A, displaying only threonine correlations. T233 resonances are highlighted with dashes (Cα, Cβ, Cγ2). Reported distances regarding T233 are displayed with yellow dots and illustrated on the HET-s(218-289) structure (PDB ID: 2RNM). (C) Excerpts of the PDSD spectrum shown in A, (ppm window: 10-17.5). I231 resonances are highlighted with dashes (Cα, Cβ, Cγ1, Cγ2, Cδ). Reported distances regarding I231 are displayed with blue dots and illustrated in HET-s(218-389) structure (PDB ID: 2RNM). For clarity purposes, overlapping peaks were disregarded. (D) HET-s structure (55), as shown in B-C, from a lateral point of view. All mentioned distances were determined based on either PDSD long-mixing time experiments (with [2-^{13}C]- or [1,3-$^{13}C_2$]-glycerol labelling schemes), or CHHC spectra (with a fully labelled HET-s amyloid core sample).

Interestingly, the 3D structure model construction of supramolecular assemblies often requires hybrid approaches, which employ data acquired using other techniques, such as X-ray diffraction, or scanning transmission electron microscopy (STEM), which give information about the mass-per-unit-lengths parameters of filamentous objects (5, 92, 153, 154). From such bio-objects, a data set, comprising knowledge about the molecular atomic structure from SSNMR experiment and fibril morphology with STEM analysis, will be the basis for the complex model calculation and energy minimization processes.

1.3.4. ^{1}H-detection for structural investigations

^{1}H detection boosts sensitivity and allows the direct observation of ^{1}H-^{1}H proximities. ^{1}H-^{1}H proximities represent natural restraints for structure calculation, easy to interpret, and in contrast to ^{13}C-^{13}C interactions, are a direct indication of physical close contacts, especially relevant across β-strands or within hydrophobic patches.

The capabilities of ^{1}H-detected techniques with respect to a full structure determination in the solid state have been demonstrated in a number of microcrystalline systems (121, 122, 124-126, 155). MAS rates in the 40-60 kHz range were employed, aided by extensive sample deuteration and selective reintroduction of protons (see section 2.1), allowing a significant reduction in the sample amount required, as well as faster spectral acquisition and unbiased automated data analysis. The approach has recently been extended toward the determination of a membrane-embedded target, in combination with ^{13}C-^{13}C constraints (123, 128). Notably, this approach considerably accelerates resonance assignment, which represent the first but crucial phase of an NMR structure determination and provides a quick evidence of the secondary structure adopted by the protein. In the field of amyloid proteins and in a pioneering study, Reif and coworkers demonstrated high-resolution ^{1}H-detected spectra for Alzheimer disease β-amyloid peptide Aβ (1-40), using 20 kHz MAS and a deuterated sample back-reprotonated at the amide sites at 25-50% (119). There, 2D-^{1}H,^{15}N fingerprint correlation spectra were used as a clear and rapid readout of sample homogeneity, enabling the optimization of fiber preparation through multiple seeding cycles. Recently, Meier and coworkers showed that ^{1}Hα,^{15}N,^{13}Cα inter-residue correlations can be used to transfer available ^{13}C and ^{15}N assignments to α-protons in a partially (above 75%) deuterated sample of HET-s(218-289) at 60 kHz MAS (134).

The use of MAS rates beyond 100 kHz allows to extend the approach and removes the requirement for proton dilution by deuteration, opening the way to the sensitive detection of amide and aliphatic protons in fully protonated proteins. In particular, a set of experiment was proposed for sequence-specific backbone and aliphatic side-chain resonance assignments. Only 500 μg of sample and a few days of data acquisition were used in this study (135). Overall, this opens the perspective of determining tertiary and quaternary structures of all supramolecular assemblies by directly probing side-chain-to-side-chain and/or backbone-to-side-chain ^{1}H-^{1}H proximities (133).

1.4. Structure calculation

Despite their topological peculiarities, high-resolution 3D structures can be determined from SSNMR data using the same general procedure as for solution NMR, *i.e.*: (i) conversion of assigned NMR observables into structurally meaningful restraints, (ii) calculation of hundreds of conformers by restrained molecular-dynamics simulation and (iii) selection of best fitting conformers as final structure bundle. Detected cross-peak signals in relevant hetero- or homo-nuclear spectra are first converted into restraints restricting the distance between the different assigned nuclei. In SSNMR studies, distance restraints often take the form of an upper-bound distance that can be either estimated from peak intensity and known reference distances or from L-shape curves obtained from multiple structure calculation trials with increasing upper-limits (23). It is worth noting that the success of structure calculation strongly relies on a critical number of long-range (between residues separated by more than 5 residues in sequence) and inter-molecular restraints (between different subunits).

The restraint set for structure calculation can be supplemented by dihedral angle restraints on φ/Ψ backbone angles, predicted from backbone secondary chemical shifts using tools like TALOS+ (156). On the basis on the collected restraints, several conformers (usually hundreds) are then calculated using dedicated software, such as CYANA (157), UNIO (158), ARIA (159), XPLOR-NIH (160) or Rosetta (161) that perform restrained molecular modeling.

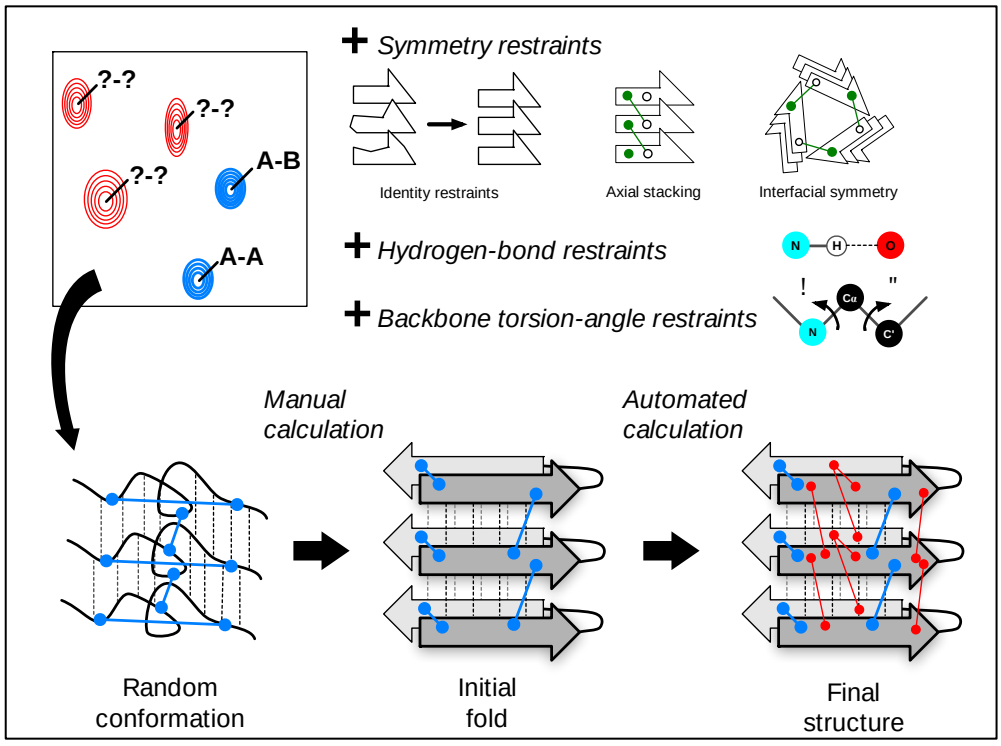


Figure 0-5: General procedure for 3D structure calculation of amyloid fibrils structures from SSNMR. First, distance restraints derived from unambiguously assigned cross-peaks or peaks with low spectral ambiguity are used for structural calculation to establish the initial fold of the fibrils (*Manual calculation*). Next, the restraint set is supplemented with all identified cross-peaks (generally with high spectral ambiguity) that will be assigned automatically and yielding a high precision bundle of conformers as the final structure (*Automated calculation).* Owing to the highly ordered nature of amyloid fibrils, symmetry restraints are employed to ensure a symmetric arrangement of subunits during calculation (identity restraints, axial stacking and interfacial symmetry, if applicable). Additionally, backbone dihedral angle restraints, derived for secondary chemical shifts, and distance restraints enforcing hydrogen-bonds are implemented in order to impose a regular conformation and proper stacking of stands.

1.4.1. Manual structure calculation

Direct application of automated NMR cross-peak assignment and structure calculation tools used routinely for solution NMR data is pragmatically limited for amyloid fibrils due to not only the lower spectral resolution of SSNMR but also the difficulty to disentangle intra- and inter-molecular restraints, the latter potentially originating from axial (along the fibril axis) or lateral (between protofibrils) contacts. It is thus common practice to first perform structure calculation using solely SSNMR restraints that could be assigned unambiguously or with very limited spectra ambiguity that can be identified manually by a trained operator (ca. a few dozens of long-range and a handful of inter-molecular restraints at least). Such *manual structure calculation* (**Figure 5**) generally determines an initial fold with reasonable precision that will serve to identify new compatible restraints for additional rounds of manual calculation. At this stage, it is important to carefully replicate restraints for all equivalent subunits to respect the symmetric nature of the fibrils.

1.4.2. Automated structure calculation

Alternatively, one can proceed with *automated structure calculation* (**Figure 5**) using all remaining cross-peaks for which unambiguous assignment cannot be safely obtained manually. For instance, the CYANA software can assist in automatically collecting assignment possibilities by matching cross-peaks positions with the list of assigned chemical shifts according to user defined tolerance windows. These tolerances for assignment of cross-signals should be reflective of the linewidth in the different dimensions of the originating spectrum. CYANA then carries out an iterative protocol by alternating calculation of conformer bundles and selection of the most probable assignments compatible with the structures, hence reducing human bias in assigning cross-peaks.

1.5. Thesis objectives

The following thesis manuscript is divided into six independent chapters based on the different projects that my PhD work encompasses.

Chapter 1 relays a methodological approach to use DNP (Dynamic Nuclear Polarization) in the context of structural biology with the combined advantages of fast MAS (up to 40 kHz MAS) and selective labelling. We performed such methodological development in collaboration with Dr Guido Pintacuda and Dr Anne Lesage from Lyon ENS in an attempt to set up a proof-of-concept of ^{1}H detection of proteins under DNP conditions and further the knowledge of DNP applications to biological samples.

Chapter 2 encompasses the study of a novel functional amyloid family involved in programmed cell-death of *Nectria haematococca*, highly reminiscent of the HET-s paradigm in *Podospora anserina*. HET-s is a model protein for functional amyloids and prions as it stands as a fascinating amyloid protein whose assembly is thought to be part of a signalling cascade. Dr. Sven Saupe, from the University of Bordeaux, identified a three-gene system encoding for three proteins called SesA, SesB, and het-eN, which share a short segment in their N- or C-terminal ends with high sequence homology. The purpose of the study is to characterize the activity *in vivo* and the structural features *in vitro*, in close collaboration with Dr. Sven Saupe. Identifying a new functional amyloid is of great interest in the mystery of functional *vs* pathological amyloids. An acute understanding of how functional amyloid differ from pathological misfolded proteins is still lacking. We used SSNMR to investigate the structural fold of SesB amyloid motif and found that SesB fold seems to constitute a new functional amyloid family.

Chapter 3 mentions our study of another functional amyloid entity called TasA. TasA is involved in *Bacillus subtilis* and *Bacillus cereus* biofilm matrix composition. Biofilms are surface-attached bacterial communities in which bacteria an extracellular matrix (ECM) is secreted. The ECM is composed of exopolysaccharides (EPS), proteins, DNA, etc. This work has been done in collaboration with Dr. Diego Romero, from the University of Malaga. The goal of this study is to analyse yet another functional amyloid. Several reports have already proposed TasA to form amyloid fibrils within the extracellular matrix of the biofilm thus being responsible for the biofilm integrity and protect bacteria from external and mechanical stresses. In addition to TasA, the proteins TapA and CalY, the partner proteins in *Bacillus subtilis*, and *Bacillus cereus*, respectively. This broad functional versatility and their remarkable ability to form robust filamentous self-assemblies contribute to general interesting and wonder surrounding functional amyloids in the bacterial kingdom. This chapter combines a

biochemical/biophysical approach to a microbiological one as we report the differences displayed by the species *Bacillus subtilis* and the pathological *Bacillus cereus*.

Chapter 4 resumes such focus on biofilm formation as we investigate the impact of biofilm-related mutations, in *eps*, *tasA*, and *tapA*, on the overall biofilm matrix composition and on the general dynamical features of the biofilm. This chapter introduces the use of SSNMR multidimensional techniques as a powerful tool to obtain the composition of cell wall molecules in intact *Bacillus* biofilms. Changes in molecular composition of the cell wall has been studied for three different mutants and compared to wild type biofilms to decipher the impact of mutations on the cell wall composition.

Chapter 5 focuses on the inner rod proteins of the type three secretion system of Gram-negative bacteria. Type 3 secretion systems (T3SSs) are in the centre of Gram-negative bacterial ability to directly deliver effector proteins from their cytoplasm to that of eukaryotic cells via a protein-based needle complex, through a most conserved battery of homologous genes required to assemble a nanomachine, also called the "injectisome". This appendage passes through the bacterial barrier composed of plasma membranes, peptidoglycan layer and the extracellular space. Such work has been done in collaboration with Prof. Jorge Galán from Yale University. Following the elucidation of the needle complex, we set to characterize the inner rod, which constitutes the basis of the needle by connecting it to the basal body. We wish to study two proteins, PscI in *Pseudomonas*, and PrgJ in *Salmonella,* which form the inner rod. In the context of this study, I was able to work with Prof. Jorge Galán for five months. Chapter 5 thus displayed the work we have done which focuses on the *in vitro* characterization of the inner rod proteins, as we discuss the structural and sequence conservation of the inner rod proteins.

Chapter 6 relates the work we have done on a complex and mysterious bacterial supramolecular assembly called refractile bodies (a.k.a. R bodies). R bodies (in this study, type 51 refractile bodies) are unusual, ribbon-like protein polymers found in many bacterial species, such as *Caedibacter* and *Pseudomonas*. In 1938, Tracy M. Sonneborn observed that some *Paracemium aurelia* strains were able to kill other sensitive strains of the same species. This ability was initially linked to an extranuclear factor, called kappa, which was associated with the production and release of a toxin in the medium. R bodies are large protein biopolymers that look like ribbons and have the ability to roll and unroll in a pH-dependent fashion. Such work is done in collaboration with Dr. Justin Kollman, from Seattle University, USA. No structural has been uncovered so far, and we set to characterize such unusual structural arrangement.

Chapter 1: High-resolution ^{1}H-detected study of biomolecules by Dynamic Nuclear Polarization Enhanced very fast MAS NMR

1.1. Context

As SSNMR methods are increasingly employed to characterize non-crystalline and insoluble biomaterials (84, 162-170), recent and on-going advances regarding both sensitivity improvement (171-176) and proton-detected approaches (136, 177-186) are expected to boost MAS SSNMR tremendous role in understanding supramolecular assemblies (1, 2, 187). For more than half a century, SSNMR has been favorably used in the study of molecular structure, organization and dynamics of non-crystalline and insoluble samples, from chemistry to biology. Consequently, improvements in NMR sensitivity are continually sought to aid in the observation of dilute moieties and to accelerate data acquisition.

An interesting challenge of NMR-based methods in general is the signal sensitivity (171, 173, 188, 189). For decades, the fields of non-biological materials have used Dynamic Nuclear Polarization (DNP) (190-193) for improving such issue and reduce the samples concentrations as well as experimental times. DNP is based on the hyperpolarization of nuclei by transferring large electron polarization to the surrounding nuclei (172-174, 188, 194-198). Such transfer is achieved through continuous microwave (MW) radiations that saturate the electron transition. This phenomenon requires the use of cryogenic temperature (below 200 K) and organic radicals as electron sources. High power and frequency gyrotrons are used and so far, NMR

signals have been enhanced by a factor of ~10-200 for biological samples in numerous studies. The use of such remarkable technology still requires the optimisation of several parameters: MW strength, temperature, the type of polarization agent used, overall radical concentration and insertion into the sample, MAS frequency, magnetic field strength, etc (172).

Dynamic nuclear polarization (DNP) has revolutionized SSNMR for material sciences (174, 196, 199-206), wherein microwave (μw) irradiation of the sample drives electron-nuclear spin polarization to enhance the observable NMR signal which, for solid samples, with a theoretical enhancement limit of ~660 for protons (192). Here, we report an approach combining fast MAS, ^{1}H-detected SSNMR, and DNP coupled with site-specific labelling in the analysis of a biological system. Altogether with severe spin dilution, three-dimensional proton detected SSNMR spectroscopy becomes possible for macromolecular protein assemblies while allowing improvements in sensitivity. The detection of SSNMR signals encoding for proton-proton proximities is demonstrated and paves the way for future structure elucidation of protein assemblies otherwise intractable by conventional biophysical techniques.

The use of DNP on proteins and biological assemblies (172, 197, 207) has been in constant increase through the past two decades as new SSNMR methodologies and instrumentations have been developed (208-212), and the chemical design of more efficient polarizing agents (213-218). Many biological objects have been studied with the help of DNP-enhanced SSNMR, including large complexes (219-226), fibrils (227-232), microcrystals (176), viral particles (233-235), membrane proteins (16, 236-241), entire organisms (194, 242-246), cellular cell walls (38, 247, 248), or lipidic systems (198, 249-253).

DNP-based SSNMR is now at a critical stage where proton-detection has yet to be implemented in such experiments, as signal enhancement combined with high sensitivity nuclei such as protons, is expected to be extremely useful. Similar to the race for proton detection “traditional” SSNMR has witnessed in the past two decades, DNP-enhanced SSNMR is now experiencing the same experimental trends. Although the current state of SSNMR advances gives access to more than 100 kHz magic-angle spinning, no DNP-equipped SSNMR setup is commercially available with the ultra-fast MAS technology. The current theoretical understanding of cross-effect (CE) DNP and empirical DNP observations of increasing MAS speed also raise questions regarding the potential efficiency of hyperpolarization with ultra-fast MAS (197, 254, 255). The very-fast MAS SSNMR probes (~60 kHz) with cryogenic cooling which have been constructed and paired with the obligatory DNP apparatus (a gyrotron microwave source and waveguide) therefore remain pertinent to SSNMR proton detection (256-258). Inherently lower

peak resolution at these MAS frequencies, compared to ultra-fast MAS SSNMR, typically lead to detrimental peak overlap for the dense aliphatic ^{1}H region and commonly require partially or specifically protonated samples. In the context of DNP-enhanced SSNMR, direct proton-detection is also practically more favorable as it exploits the greater NMR sensitivity of ^{1}H nuclei, allowing for shorter experimental acquisition times and reduced sample amounts, in addition to signal enhancement.

We present an approach that combines dynamic nuclear polarization (DNP) and proton-detection with methyl isotopic labelling and fast magic-angle spinning in order to enable an unprecedented sensitivity and resolution improvement of SSNMR experiments of biomolecules. 2D ^{13}C-^{1}H and 3D ^{13}C-^{1}H-^{1}H experiments are presented on the prion-forming domain of HET-s in its amyloid state, allowing the extraction of ^{1}H-^{1}H methyl restraints. Our approach opens a new avenue for breaking the sensitivity boundaries in SSNMR-based structural biology.

1.2. Methodology

1.2.1. Isotopically enriched protein expression

BL21 (DE3) *E. coli* cells were transformed via electroporation with a 2-pET21 vector containing a C-terminally hexa-His tagged construct of HET-s 218-289. Cells were plated and grown overnight on LB-ampicillin agar media before a single transformed colony was picked and used to inoculate a 1 mL LB-ampicillin (100 µg/mL) pre-culture. Following 4 hours of growth at 37 °C, cells were harvested by centrifugation (15 min, 1000g), resuspended in 10 mL of standard ^{15}N ammonium chloride M9 media (100% H_2O, 100 µg/mL ampicillin) and incubated at 37 °C. When cell density had reached an OD_{600nm} value of 0.6 - 0.8, cells were harvested by centrifugation as before and resuspended in 25 mL deuterated ^{15}N ammonium chloride M9 media (100% D_2O, 100 µg/mL ampicillin). For D_2O conditioning, cells were grown at 30 °C overnight to an OD_{600nm} of ~2. The pre-culture was then diluted to 250 mL with D7-^{13}C-glucose, ^{15}N M9 media (100% D_2O, 100 µg/mL ampicillin) and grown at 37 °C. The rate of bacterial growth was estimated according to initial OD_{600nm} measurements and 1 hour prior to induction the culture was supplemented by 240 µg/mL 2-(^{13}C,D_2) methyl-4-(D_3) acetolactate (NMRBio, France) followed by 60 µg/mL 2-oxo-3.3.4.4-(D_4)-(^{13}C) butyrate (NMRBio, France) 40 minutes later. Overexpression was induced by the addition of 1 mM IPTG when cell density reached an OD_{600nm} ~0.7. Protein production occurred for a further 4 hours at 37 °C before harvesting by centrifugation (30 mins, 6000g).

1.2.2. Protein purification and assembly

All cell lysis and purification steps were carried out in 100% H_2O buffers to allow back-exchange of the labile protons. Initial cell lysis was carried out after thoroughly resuspending the bacterial cell pellet in 25 mL buffer A (50 mM Tris and 150 mM NaCl) by three rounds of sonication (30 sec, 12W), cooling on ice between each successive step. Inclusion bodies and cell debris were collected by centrifugation (1hr, 15000g), resuspended (20 ml Buffer A + 2% Triton X-100) and incubated at 37 °C for 20 mins. Insoluble components were then again isolated by centrifugation (10 min, 50000g), washed in 25 mL Buffer A and incubated for 1.5 hr at 37 °C. After a second identical centrifugation step, the resulting pellet was resuspended in <10 mL of 50 mM Tris and 0.5 M NaCl and supplemented with dry Gu HCl until saturated (final volume ~25 mL). The sample was then left at 60 °C overnight and sonicated the following morning three times (as before) to fully solubilise the protein. A final ultracentrifugation step (1hr, 250000g) was then used to clear the supernatant of remaining contaminants. The final HET-s containing supernatant was collected and purified on a 1 mL His trap nickel-affinity column (pre-equilibrated in 8 M urea) using a standard imidazole linear gradient elution. Lastly, a Hi-prep 26/10 SEC column (GE Healthcare) was used to remove any remaining imidazole and urea and the final protein yield was estimated to be 3.5 mg. Protein purity was assessed by SDS-PAGE. The sample was concentrated to ~1mM, adjusted to pH 7 and left to fibrillise at room temp over one week. Fibrillar HET-s was then resuspended in 200 µl of 'DNP juice' (deuterated D^8-glycerol, D_2O, H_2O; 60/30/10%, respectively) containing 10 mM AMUPol (CORTECNET, France) and incubated overnight prior to experimentation.

1.2.3. NMR spectroscopy and DNP

The MAS DNP NMR experiments were performed on Bruker Avance III 800 MHz wide bore spectrometer, equipped with triple resonance 1.3 mm low-temperature MAS probes. DNP was achieved by irradiating the sample with high-power microwaves at a frequency of 527 GHz (18.8 T / 800 MHz), generated by gyrotrons that were operating continuously during the DNP experiments (stability of better than ±1%). Acquired spectra used in-house variants of Bruker pulse-programs. Multidimensional spectra were processed using NMRPipe and spectral analysis was carried out with CCPNMR analysis.

1.3. Results and Discussion

1.3.1. Model system

In addition to very fast MAS SSNMR, we used a $^{13}CHD_2$-ILV methyl labelling approach, allowing a severe level of spin dilution. The deuteration of proteins improves the signal enhancement as obtained by dynamic nuclear polarization (259). In this study we scrutinized the 8.6 kDa prion-forming domain (residues 218-289) of the HET-s functional amyloid from the filamentous fungus *Podospora anserina*. The native HET-s functions in the heterokaryon incompatibility – a none-self cell recognition event, which facilitates programmed cell death upon fusion with genetically dissimilar fungi (260-263). The HET-s fibrillar state has been studied in detail, in particular by SSNMR which has revealed its two-fold β-solenoid structural arrangement (55) and explored its dynamic behavior (264-268). HET-s SSNMR spectra are quite remarkably resolved, and its singular molecular amyloid structure has made it an exemplary biological non-crystalline SSNMR system in many studies. The most studied HET-s construct is the truncated HET-$s_{218-289}$, comprising 79 amino acids (with 6 Ile, 3 Leu, 8 Val in the entire HET-$s_{218-289}$ construct – 2 Ile, 2 Leu, 5 Val within the known rigid core) and the His_6 tag, theoretically counting for 595 hydrogen positions in total. Using perdeuterated media during production, and 100% ^1H back-exchange after purification, one can expect a single side-chain methyl proton per I/L/V residue in addition to the 154 exchangeable hydrogen positions (backbone and sidechain), thus resulting in a ~29% ^1H-protonation level (with 424 unexchangeable ^2H positions). We aimed at exploiting such spin dilution in order to increase the spectral resolution and the signal-to-noise ratio through the use of DNP very-fast MAS (168, 269-271).

1.3.2. Polarizing source concentration optimization

First, the choice of water-soluble polarizing source was easily made as AMUPoL (213) was developed to be rigid and to display long electron relaxation times and efficient electron-electron couplings (257). We set up a series of experiments, with and without microwaves irradiation, assessing several parameters, such as radical concentration and MAS rates, and their influence over DNP enhancement, and data quality as recorded on a DNP-equipped 800 MHz spectrometer, using 1.3 mm diameter rotors (**Figure 1**). 1D (^1H)-^{13}C-^1H CP-MAS spectra of fully back-exchanged $^{13}CHD_2$-ILV-labelled HET-s fibrils, in the presence of 5, 10, 20, or 40 mM AMUPoL, at 40 kHz MAS, were first acquired with and without microwave irradiation using T1-optimised recycle delays (**Figure 1A**) in an attempt to assess and determine the

optimal concentration for such a sample. Approximatively, we observed signal enhancements reaching ~20 for all four samples. However, based on saturation recovery experiments, spin-lattice relaxation rates for all samples suggest that the presence of higher concentrations of radicals led to drastically decreased T1 time, while sensitivity enhancement (*i.e.,* T1-corrected) factors increase (**Figure 1B**). Indeed, the radical concentration has been reported to alter both longitudinal and transverse relaxation, thus having an effect on overall NMR signal-to-noise ratio, and signal linewidths (5-10 mM: ~0.55 ppm; 20-40 mM: ~0.75 ppm) (272).

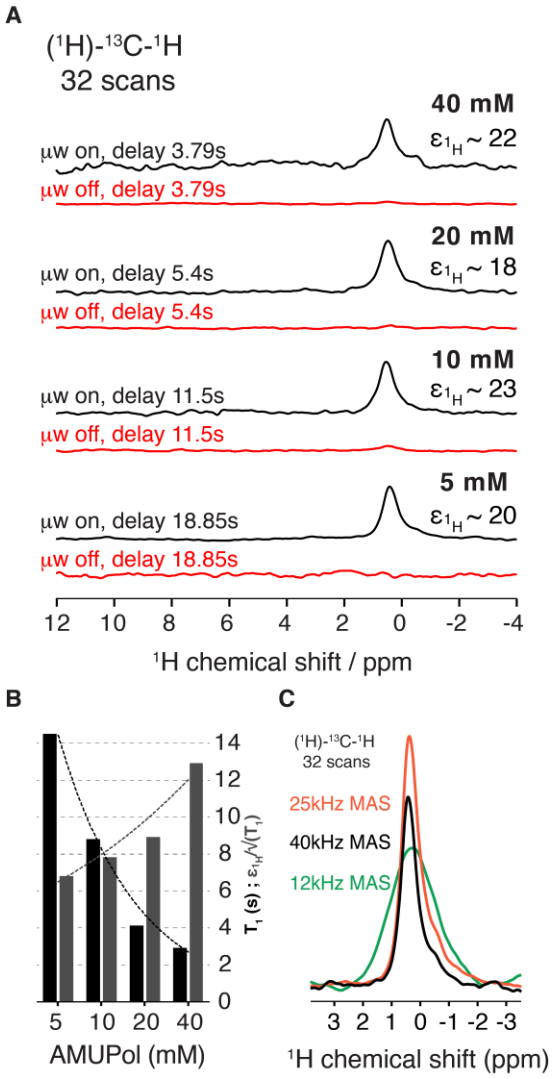


Figure 1-1: (A) Comparisons of 1D (^{1}H)-^{13}C-^{1}H CP-MAS spectra of $^{13}CHD_2$-ILV labelled and back-exchange protonated HET-s fibrils, in the presence of 5, 10, 20, or 40 mM AMUPoL, with and without microwave irradiation (17.8 kV, 170 mA) using T1-optimised recycle delays. (B) Dependence of the enhancement εDNP (=$\varepsilon^{1}H/\sqrt{T1}$) (black) and T1 relaxation time (grey) on the overall AMUPoL concentration, recorded on $^{13}CHD_2$-ILV labelled and back-exchange protonated HET-s fibrils. (C) Comparisons of 1D (^{1}H)-^{13}C-^{1}H CP-MAS spectra of $^{13}CHD_2$-ILV labelled and back-exchange protonated HET-s fibrils, in the presence of 5 mM AMUPoL, with microwave irradiation (17.8 kV, 170 mA), at different MAS rates: 12, 25, and 40 kHz.

1.3.3. Effect of MAS rate on DNP enhancement

Additionally, the same 1D experiment was performed on the 5 mM AMUPoL sample, at different MAS rates: 12 and 25 kHz (**Figure 1C**). As expected, a linewidth broadening effect (40 kHz: 0.59 ppm; 25 kHz: 0.73 ppm; 12 kHz: 1.82 ppm) is observed as the MAS rate decreases, while MAS-dependent signal enhancements εDNP and T1 times are obtained at different spinning frequencies. Both T1 times and εDNP decrease with higher MAS velocity: 14.5 s / εDNP ~20 at 40 kHz, 12.5 s / εDNP ~18 at 25 kHz MAS, and 9.9 s / εDNP ~15 at 12 kHz.

1.3.4. Two-dimensional SSNMR

Further 2D proton-detected (^{1}H)-^{13}C-^{1}H CP spectrum acquisition of HET-s selectively-labelled fibrils under DNP conditions resulted in a well-resolved spectrum comprising all 9 expected ^{1}H-^{13}C correlations from the I/L/V residues within HET-s$_{(218-289)}$ primary sequence (**Figure 2A**). The spectrum was recorded with a 10-second recycle delay, over 52 minutes, while obtained resolution, spectral dispersion and signal-to-noise levels allowed residual assignment based on reported chemical shifts. The 2D (^{1}H)-^{15}N-^{1}H CP spectrum (**Figure 2B**) also highlights the impressive DNP signal enhancement witnessed in the 1D data, at 10 mM AMUPoL, while informing on the overall amyloid fold as the spectral fingerprint displays a rather dispersed and structurally organized pattern. The resulting amide chemical shift fingerprint is indeed in good agreement with that previously published (273) and therefore confirms the expected molecular arrangement of AMUPoL-containing sample, as a prerequisite for further structural investigations.

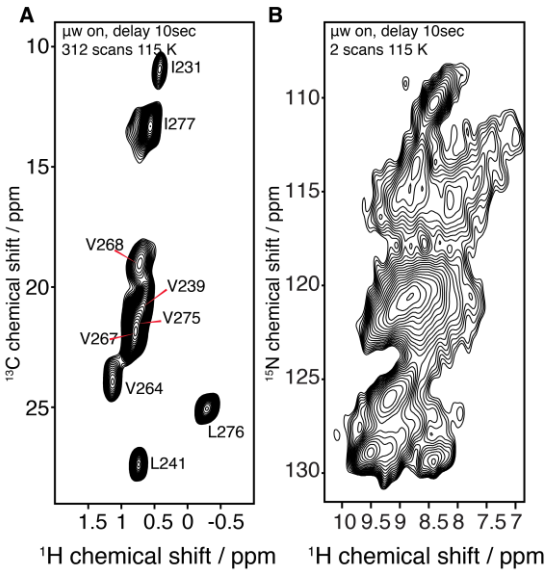


Figure 1-2: (A) 2D (^{1}H)-^{13}C-^{1}H proton-detected CP-MAS spectrum of HET-s fibrils under DNP conditions, with 10 mM AMUPoL, using a recycle delay of 10 seconds (experiment time: 52 minutes). Spectral resolution, dispersion and signal to noise were sufficient to assign the nine expected ^{1}H-^{13}C methyl correlations of the HET-s hydrophobic core. (B) 2D (^{1}H)-^{15}N-^{1}H proton-detected CP-MAS spectra of the HET-s$_{(218-289)}$ fibrils under DNP conditions, in the presence of 10 mM AMUPoL (experiment time: 18 minutes). All experiments are to scale and were acquired at 800 MHz with 40 kHz MAS.

1.3.5. Proof of concept

As proof of concept, both the site-specific labelling scheme and the method relevance on other types of biomaterials were validated through a series of experiments on ^{13}C-^{15}N uniformly labelled HET-s fibrils and on the MU phage tail (274-276). $^{13}CHD_2$-ILV labelling has indeed been developed in the context of structural inquiries with the goal of determining long-range through-space distance restraints in hydrophobic-residue-rich biological systems, as only side-chain atoms are ^{13}C- and ^{1}H-labelled (277). The benefit of such site-specific labelling scheme within a DNP-based proton-detection approach is easily perceived when comparing the spectral resolution and dispersion to that of a uniformly ^{13}C-^{15}N labelled sample, under DNP condition at 10 mM AMUPoL (**Figure 3A**). Additionally, the perdeuteration level itself has been shown to be correlated to longitudinal relaxation times (259, 272). Although the DNP signal enhancement is still remarkably pertinent (εDNP ~15), with a recycle delay of 11.3 seconds (~14.6-hour experiment), the 2D (^{1}H)-^{13}C-^{1}H CP spectrum displays broadened (linewidth: >2 ppm) and crowded ^{1}H-^{13}C correlations. Additionally, HET-s being one of the most studied amyloid-based supramolecular assemblies using SSNMR, the question of sample relevance comes to mind, as such unprecedented approach might not satisfy larger assemblies

with higher I/L/V contents. Assessing such limitations, a rather different biological system that is the phage tail MU (118 amino-acids: 9 isoleucines, 5 leucines, and 7 valines), analyzed with $^{13}CHD_2$-ILV-labelling under DNP conditions at 10 mM AMUPoL, endorses the use of DNP in the context of proton detection for larger supramolecular assemblies through two observations: (i) a 13-fold DNP signal enhancement with a 13.8 second recycle delay as observed on the 1D (^{1}H)-^{13}C-^{1}H CP-MAS spectrum of fully back-exchanged $^{13}CHD_2$-ILV-labelled MU (linewidth: ~0.75 ppm), (ii) a well-dispersed and quite well-resolved 2D (^{1}H)-^{13}C-^{1}H CP spectrum recorded for ~13 hours (**Figure 3B**).

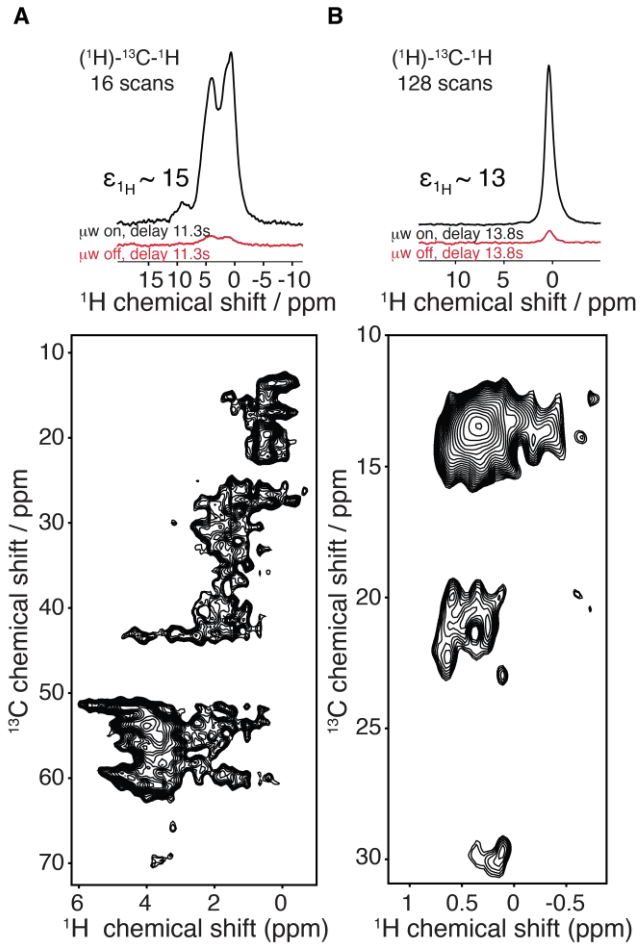


Figure 1-3: 1D (^{1}H)-^{13}C-^{1}H proton-detected CP-MAS spectrum, with 10 mM AMUPoL, with and without microwave irradiation (17.8 kV, 170 mA) and corresponding 2D (^{1}H)-^{13}C-^{1}H proton-detected CP-MAS spectrum, under DNP conditions, with 10 mM AMUPoL, as recorded on (A) HET-s fibrils (^{13}C-^{15}N uniformly labelled) and (B) MU assemblies ($^{13}CHD_2$-ILV labelled).

1.3.6. Structural analysis

Lastly, we were able to acquire a 3D hCCH proton detected DNP experiment for the HET-s construct, site-specific labelled. Using only 16 scans, with 1184 x 46 x 24 complex points, and

a 2.5 s recycle delay (experiment time 12 hours 8 minutes) we were able to acquire sufficient signal and resolution to observe all of our labelled methyl groups and a range of through space correlations to both amide and aliphatic protons. From comparisons to the known HET-s peak assignments we were also able to assign specific restraints corresponding to intra and inter-nuclear correlations up to 6 Å in distance (**Figure 4A-B**). A complete assignment of the correlated proton peaks was not possible during this work; yet comprehensive future studies can easily be envisaged by employing classical backbone and side-chain assignment strategies under DNP conditions.

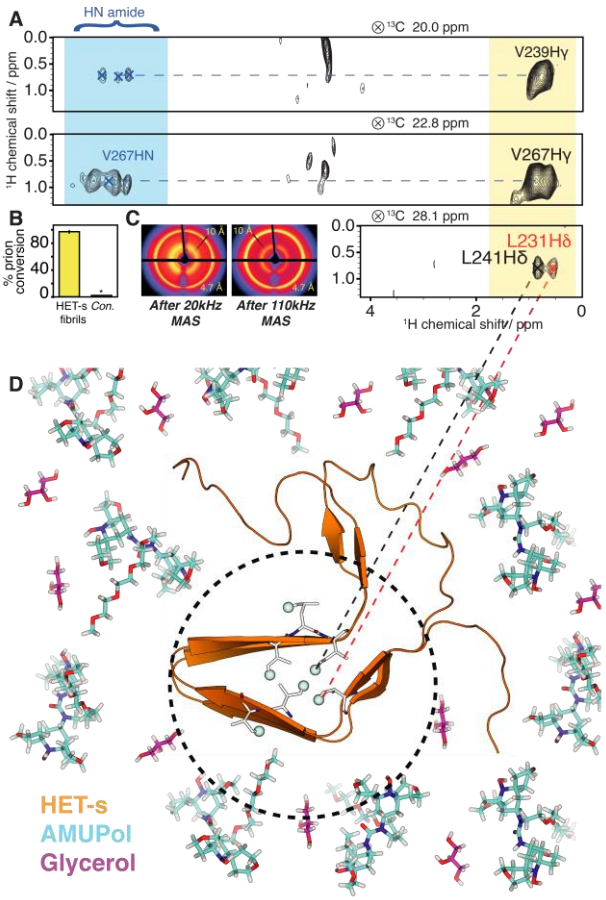


Figure 1-4: (A) Excerpts of 2D ^{1}H-^{1}H planes extracted from the 3D (^{1}H)-^{13}C-^{13}C-^{1}H proton-detected CP-MAS spectrum, under DNP conditions (experiment time: ~12 hours), showing the through-space ^{1}H-^{1}H correlations for the $^{13}CHD_2$-ILV methyl groups of V239 (top), V267 (mid) and L241 (bottom). The methyl-methyl L241H$_\delta$-L231H$_\delta$ contact is highlighted on the HET-s structure. (B) HET-s fibrils prion infectivity assessment, as measured following 40 kHz MAS, for HET-s fibrils and control sample (without protein). (C) X-ray diffraction patterns of HET-s fibrils following 20 kHz MAS (left) and 110 kHz MAS (right). (D) Schematic view of the spin diluted sample of HET-s amyloid fibrils, consisting of the aggregated HET-s$_{(218-289)}$ prion domain mixed with AMUPoL in DNP juice. For the sake of clarity, only the methyl proton labelled on the R2 domain (residues 262-278) are

shown as spheres. AMUPoL (cyan) and glycerol (magenta) molecules are shown as sticks. A sphere of ~15 Å centered on the HET-s hydrophobic core is shown.

In contrast to the aforementioned 2D spectra, the hCCH experiment highlights the fact that >2D multidimensional experiments are perhaps best suited to the advantages gained from DNP enhanced proton detected SSNMR. Although the impressive gains in signal intensity from DNP can be readily appreciated in a simple sense, when observing crowded spectral regions, 2D data suggests that the incorporation of dilute labelling schemes is practically obligatory in order to extract meaningful information. However, in the case of the 3D hCCH experiment, the addition of an extra experimental dimension diminishes this complication and demonstrates that the achievable resolution is largely sufficient in identifying inter-nuclear correlations.

1.3.7. Biological relevance

Such tremendous outcome might be cross-evaluated in order to determine whether or not the DNP-based very fast MAS could be considered as bio-friendly. In addition to rotor spinning, the DNP temperature and microwave irradiation might cause damage to the assemblies and/or surrender the sample as biologically irrelevant. We tested both the prion activity and the cross-β arrangement resistance to such treatment. Following the 40 kHz magic-angle spinning under DNP conditions used within this study, the ability of HET-s$_{(218-289)}$ to convey its amyloid fold *in vivo* (*i.e.,* prion conversion) was assessed, positively so (**Figure 4C**). Furthermore, the typical cross-β arrangement as detected using X-ray diffraction has been verified after 20 kHz, and 110 kHz. The distinction of the 4.7 Å and the 10 Å signals confirms both DNP conditions and ultra-fast spinning do not impact HET-s$_{(218-289)}$ biological relevance (**Figure 4D**).

1.4. Conclusion

Throughout the study, we tested and confirmed the relevance and advantages brought by MAS-DNP in the context of proton-detection SSNMR for biomaterials. Outstanding signal and sensitivity enhancements were obtained using AMUPoL as the electron-providing radical. Both the MAS and radical concentration effects were thoroughly studied. Remarkable 2D spectral resolutions were achieved with the highly specific $^{13}CHD_2$-ILV labelling, as well as through-space distance-determining 3D correlations. Our approach was reinforced by submitting it to uniformly labelled HET-s fibrils, as, although rather good signal enhancements were obtained, only the $^{13}CHD_2$-ILV-labelled sample introduces tremendous 2D and 3D spectral resolution.

Additionally, $^{13}CHD_2$-ILV-labelled MU phage tail comprising higher I/L/V contents and different structural arrangements, validates the method on other bio-solids. Lastly, HET-s fibrils biological relevance upon ultra-fast spinning and DNP conditions was approved as both the ability to convey prion behavior and the cross-β arrangement, typical of amyloid assemblies, are conserved after fast MAS-DNP SSNMR experiments.

Combined with a highly relevant site-specific labelling strategy, our approach uses very-fast MAS-DNP SSNMR in the context of biological structural investigation of non-crystalline samples. Such method fuses the advantages of state-of-the-art (i) SSNMR (small amount of required sample for 1.3 mm rotors, 40-60 kHz MAS, proton detection), (ii) DNP-MAS (AMUPoL as an efficient radical source, tremendous signal enhancement), and (iii) site-specific labelling strategies (ILV labelling proved to be highly efficient for restraint determination processes within supramolecular assemblies).

Chapter 2: A novel fungal functional amyloid family involved in programmed cell-death in *Nectria haematococca*

2.1. Context

In the past decades, higher-order complexes (260, 262, 268, 278-282) have emerged as a subclass of signal transduction events, and more specifically in immunity-associated signaling cascades. Higher-order assembly has thus been linked to a nucleated polymerization process which is responsible for signal transduction through amyloid formation. An illustrating system of such paradigm is the one of the prion proteins, HET-s, found in the fungus *Podospora anserina* (260, 262, 282). The [Het-s] prion is involved in a programmed cell death (PCD) reaction termed heterokaryon incompatibility (HI). HI is observed when cells between genetically distinct fungal individuals fuse in order to prevent the formation of a common syncytial network, thus avoiding the spread of mycoviruses in an amyloid-based specific defense mechanism. In that regard, the [Het-s] prion is considered as a functional amyloid (283). The system found in *Podospora anserina* also involves an NLR (Nod-like receptor; NWD2) component as HET-s is considered to be the cell death execution protein through its globular HeLo domain. Both proteins are encoded by adjacent genes and contain a short

amyloid-forming motif. The assembly of the prion-forming domain (PFD, residue 218 to 289) is believed to induce the refolding of the HeLo domain which in turn triggers its pore-forming activity (268). The [Het-s] prion is a 289 amino acid protein, termed HET-s (small "s") [92]. The HET-s protein can exist in a monomeric state termed [Het-s*] (small "s star") or in the prion aggregate state [Het-s]. Strains of the het-s genotype are either prion-free or prion-infected. Only prion-infected [Het-s] strains produce HI cell death with strains from the het-S (large "S") genotype which constitute another allelic variant. Thus, the HI phenomenon is described by an allelic incompatibility between the het-s and het-S alleles, which encode for HET-s and HET-S. These proteins differ at only 13 aa positions and share the same domain architecture. Using SSNMR, the structure of PFD amyloid fibrils has been elucidated in 2008. The PFD assembly is necessary both for amyloid formation and for prion propagation (284-288). PFD amyloid fibrils are formed by two amyloid repeats per molecule, stacked along the fibril axis. The two pseudo repeats are termed R1 and R2, and a 15-amino acid flexible loop connects the β-sheets. Each repeat comprises four β-strands, three of which form a triangular hydrophobic core. This atypical building, known as a β-solenoid core, displays a number of crucial structural elements linked to its function. More recently, a second fungal amyloid signaling motif family termed PP was described in *Chaetomium globosum*. This prion-motif is also found at the N-terminus of an NLR homolog termed PNT1 and allows activation of two distinct proteins, HELLP and SBP, encoded by adjacent genes. HELLP contains a cell death-inducing HeLo-like domain and SBP is a putative lipase. Remarkably, HELLP shows homology to MLKL, the terminal cell death execution protein in mammalian necroptosis indicating a trans-kingdom evolutionary conservation of programmed cell death mechanisms (262).

The present project was designed to characterize the structure of a third amyloid motif (called Sigma) involved in fungal programmed cell death reaction. Another inheritable phenotypic switch is observed in the species *Nectria haematococca.* A spontaneous and transmissible growth alteration occurs: the so-called Secteur (289). The Secteur phenotype is described at the growing rim of the fungal colony and is characterized by an increase in pigment production and reduced hyphal growth rate (**Figure 1B**). Such phenotype is transmissible to normal growth cultures through hyphal fusions and is due to a cytoplasmic infectious element termed σ. Propagation of σ depends on nuclear genes and mutations that specifically alter σ formation

have been isolated in a single locus termed *ses*. Two genes were identified in the *ses* locus: *sesA* and *sesB* (289).

sesA encodes a 208 amino acid long protein with an N-terminal domain showing a remote homology to the HeLo domain. *sesB* encodes a 386 amino acid long putative lipase (**Figure 1A-C**). Interestingly, reminiscent of the HET-s system in *Podospora anserina*, SesA and SesB share a C-terminal 40-50 amino acid long region of homology corresponding to the sigma motif (**Figure 1A**). Adjacent to the *sesA* and *sesB* is a gene termed *het-eN* encoding a Nod-like receptor protein showing homology to the *P. anserina Nwd*-gene family. In the *het-eN* product the same conserved motif common to SesA and SesB is also found but located in an N-terminal position. These resemblances in domain and genome-architecture between the *Ses* locus and, the *het-S* locus and the PP amyloid-motif gene cluster suggest that the *het-eN/sesA/sesB* gene complex could function analogously to the *Nwd2/het-S* and PP-motif clusters (282). Specifically, we have proposed that the sigma motif region might serve as an amyloid-forming domain that controls activation of the SesA and SesB N-terminal domains in response to activation of HET-eN and that the σ element might correspond to the prion form of SesA and/or SesB.

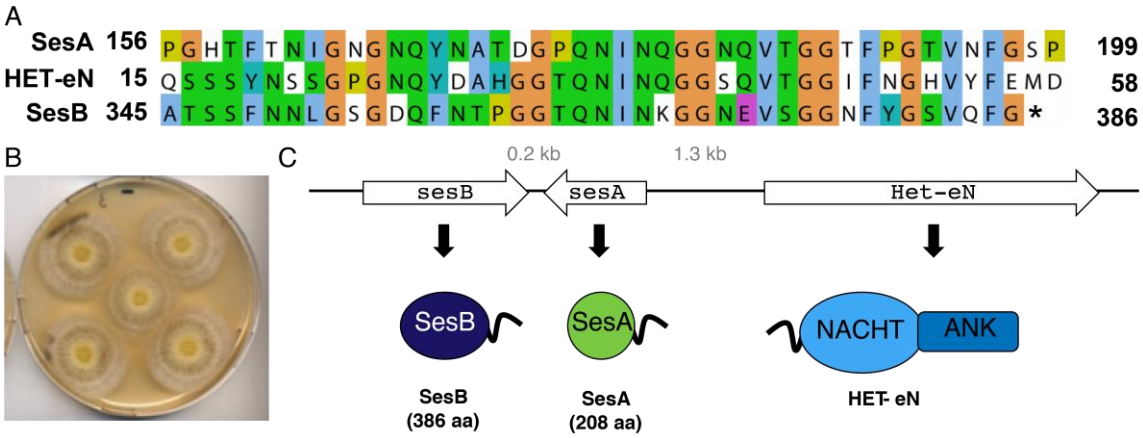


Figure 2-1: The sigma motif is conserved in the proteins SesA, SesB, and het-eN in *Nectria haematococca* encoded in a single locus. (A) Sequence alignment of the C-terminal moieties of SesA and SesB and the N-terminal segment of HET-eN, as called sigma. (B) The 'Secteur' phenotype as observed in some *Nectria haematococca* subcultures, characterized as an increase of pigment production and a reduced hyphal growth rate. (C) Representation of the *ses* locus comprising the three genes *sesB*, *sesA*, *Het-eN* encoding the three proteins of the same name.

Herein, we describe the biochemical and structural characterization of the sigma motif. We show that the sigma motif regions of SesA, SesB and HET-eN form amyloids *in vitro*. In parallel, experimentation performed in the group of Sven Saupe (Bordeaux) showed SesA and SesB to propagate as prions *in vivo* when expressed in *Podospora anserina*. They find that

interaction of the SesB lipase with the prion form of the sigma motif leads to cell death reaction reminiscent of [Het-s]/HET-S incompatibility (manuscript in preparation). We studied the structural arrangements of SesB sigma motif: $SesB_{327-386}$. Using SSNMR, we analysed isotopically labelled $SesB_{327-386}$ fibrils spontaneously formed *in vitro*. Very-fast MAS SSNMR, and the use of specific labelling schemes such as [^{13}C]-1,3-glycerol has allowed us to determine $SesB_{327-386}$ secondary structure in its amyloid fibrillar form. Interestingly $SesB_{327-386}$ seems to be constituted of 5 β strands, which support the idea that the σ motif is an entirely new fungal functional amyloid family. We report that a shorter version of SesB (*i.e.*, $SesB_{340-386}$) shows the same structural arrangements as assessed by SSNMR. SesA full-length also could share such ordered amyloid fold as was observed in our study. This work established the sigma motif as a novel example of fungal prion-like signaling motif.

2.2. Methodology

2.2.1. Bacterial strains and culture conditions

P. anserina strains used in this study were wild-type het-s, het-S, Δhet-s strains. The experiments involving *P. anserina* were performed by Dr Sven Saupe and all related figures were kindly provided by Dr Sven Saupe. For cloning and plasmid replication, *E. coli* DH5α was used. *Escherichia coli* BL21(DE3) pLysS (Merck) was used for protein production and purification. All strains were grown in Luria-Bertani (LB) medium: 1% tryptone (Oxoid), 0.5% yeast extract (Oxoid) and 0.5% NaCl. M9 minimal medium (composition for 1 liter was: 48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.01 mM $ZnCl_2$, 0.001 mM $FeCl_3$, 0.1 mM $CaCl_2$, 18.7 mM NH_4Cl, 0.2 % glucose and 10 mL of 100 x MEM vitamins solution (Merck)) was used for protein expression for NMR studies. M9 was enriched in$^{15}NH_4Cl$ and ^{13}C labeled carbon sources ([U-^{13}C] D-glucose) in order to produce labeled recombinant proteins.

The antibiotics used and their final concentrations were: 100 µg/ml ampicillin, 50 µg/ml kanamycin, 10 µg/ml kanamycin and tetracycline (10 µg/ml).

2.2.2. Cell growth and protein expression

A single freshly transformed BL21(DE3) pLysS *E. coli* was picked and cultured in 10mL of Luria-Bertani (LB) media with 50 µg/mL kanamycin at 37°C under shaking condition for 5 h to reach exponential growth phase. Cells were harvested (1000 g, 10 min) and resuspended in 100 mL of labeled minimal medium (1 g/L $^{15}NH_4Cl$, 2 g/L [U-^{13}C] D-glucose) supplemented with antibiotics and grown overnight at 30°C, 200 rpm until stationary growth phase. 10% (v/v) of the culture was used to inoculate 1 L of labeled minimal medium supplemented with

antibiotics, cultured at 37°C, 220 rpm until OD_{600nm} reach 0.8 and induced with 0.75 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The proteins were expressed at 30°C for 20 h Cells were harvested (6000 g, 30 min, 4°C, Beckman Coulter JLA 8.1 rotor) and frozen at -80°C until purification.

2.2.3. Cell growth and protein expression in D_2O

BL21 (DE3) *E. coli* cells were transformed via electroporation with a 2-pET24 vector containing a C-terminally hexa-His tagged construct of SesB. Cells were plated and grown overnight on LB-ampicillin agar media before a single transformed colony was picked and used to inoculate a 1 mL LB-ampicillin (100 µg/mL) pre-culture. Following 4 hours growth at 37 °C, cells were harvested by centrifugation (15 min, 1000 g), resuspended in 10 mL of standard 15N ammonium chloride M9 media (100% H_2O, 100 µg/mL ampicillin) and incubated at 37 °C. When cell density had reached an OD_{600nm} value of 0.6-0.8 cells were harvested by centrifugation as before and resuspended in 25 mL deuterated ^{15}N ammonium chloride M9 media (100% D_2O, 100 µg/mL ampicillin). For D_2O conditioning, cells were grown at 30 °C overnight to an OD_{600nm} of ~2. The pre-culture was then diluted to 250 mL with D_7-^{13}C-glucose, ^{15}N M9 media (100% D_2O, 100 µg/mL ampicillin) and grown at 37 °C. The rate of bacterial growth was estimated according to initial OD_{600nm} measurements and 1 hour prior to induction the culture was supplemented by 240 µg/mL 2-(13C,D_2) methyl-4-(D_3) acetolactate (NMRBio, France) followed by 60 µg/mL 2-oxo-3.3.4.4-($_{D4}$)-(^{13}C) butyrate (NMRBio, France) 40 minutes later. Overexpression was induced by the addition of 1 mM IPTG when cell density reached an OD_{600nm} ~0.7. Protein expression occurred for a further 20 hours at 30 °C before harvesting by centrifugation (30 minutes, 6000 g).

2.2.4. Protein purification and protein assembly

Cells were resuspended in buffer A (50 mM Tris, 150 mM NaCl, pH 8) and sonicated on ice (3x45 sec, 40%, Bandelin sonopuls VS 70T probe) and centrifuged (15 000 g, 60 min, 4°C, Eppendorf F34-6-38 rotor). The supernatant was discarded as proteins were mainly expressed in inclusion bodies. The pellet was resuspended in buffer A supplemented with 2% Triton X-100, incubated at 37°C under shaking condition for 20 min and centrifuged (50 000 g, 10 min, 4°C, Beckman Coulter JA 25.50 rotor). The pellet was extensively washed with buffer A, centrifuged (50 000 g, 10 min, 4°C), resuspended in denaturing condition (Tris 50 mM NaCl 500 mM, 6 M GuHCl) and incubated at 60 °C overnight until complete solubilization. Lysate was clarified by centrifugation (200 000 g, 1h, 16°C, Beckman Coulter 50.2 Ti rotor). Solubilized inclusion bodies were loaded on a HisTrap HP 5 mL column (GE Healthcare)

previously equilibrated in the binding buffer (50 mM Tris, 0.5 M NaCl, 20 mM imidazole, 8 M urea, pH 8). Proteins were eluted from the column with a linear gradient of elution buffer (50 mM Tris, 0.5 M NaCl, 500 mM imidazole, 8 M urea, pH 8). After the affinity chromatography step, proteins were loaded on a HiPrep 26/10 desalting column (GE Healthcare) to exchange buffer for 1 % acetic acid pH 3. Purification fractions were loaded on a 13% Tris-Tricine SDS-PAGE to check purity.

2.2.5. X-ray diffraction

Fiber diffraction patterns were measured at 4°C on a Rigaku FRX rotating anode X-Ray generator equipped with a Pilatus 200 K hybrid pixel detector at the copper wavelength. The concentrated hydrated samples were mounted in a MicroLoops™ from Mitegen on a goniometer head under the cold nitrogen flow. Each diffraction pattern corresponds to a 360° rotation along the phi axis with an exposure time of 720 s after subtraction of a "blank" image of the same exposure time with only the loop on the goniometer head.

2.2.6. Solid-state NMR

Experiments were performed at a MAS frequency of 11 or 60 kHz on triple resonance 4 mm, 3.2 mm or 1.3 mm MAS probes using 600 MHz and 900 MHz ^{1}H Larmor frequency spectrometers (Bruker Biospin). The sample temperature was set to ~276 K and chemical shifts were calibrated using DSS as an internal reference.

Intra-residue correlations were probed using PDSD experiments with a mixing time of 50 ms for acquisition times of 20 ms (direct) and ~6-10 ms (indirect), leading to a total experimental time of ~4 days, per sample. A ^{1}H-^{13}C cross-polarization contact time of 500-1000 μs. Two-dimensional ^{1}H-^{13}C INEPT were recorded using acquisition times of 20 ms and ~6-8 ms in direct and indirect dimensions respectively for a total experimental time of ~24h, per sample.

For proton detection experiments, SSNMR spectra were recorded on a 900 MHz (^{1}H frequency) spectrometer (Bruker Biospin, Germany) equipped with a 1.3 mm triple resonance (^{1}H, ^{13}C, ^{15}N) MAS probe in collaboration with Dr Paul Schanda in CEA Grenoble, France. Sample spinning frequency was 60 kHz. For resonance assignment, we used a set of four 3D ^{1}H detected experiments: (H)CANH [8 scans, 6 ms (t3) x 8 ms (t2) x 20 ms (t1)], (H)CONH [8 scans, 12 ms (t3) x 8 ms (t2) x 20 ms (t1)], (H)CO(CA)NH [24 scans, 12 ms (t3) x 8 ms (t2) x 20 ms (t1)], (HCO)CA(CO)NH [24 scans, 5 ms (t3) x 8 ms (t2) x 20 ms (t1)]. The combination of these experiments allowed the connectivities between intra-residual or sequential CA, CB, CO, N, and HN resonances, necessary to perform the entire backbone assignment. Experimental data were processed using TopSpin and analyzed using CCPNMR.

2.2.7. Protein preparation for electron microscopy

SesA, SesB, SesA(148-210), SesB(327-386) and HET-eN(1-74) proteins were expressed in *E. coli* BL21-DE3 plysS cells and purified using an ÄKTA purifier core system (GE Healthcare). Proteins had a C-terminal 6-his tag and all protein except SesB were purified from inclusion bodies under denaturing conditions (6 M GuHCl, 50mM Tris–HCl pH 8) using Qiagen columns. Proteins were eluted in 8 M urea, 50mM Tris–HCl pH 8, 150mM NaCl, 200 mM imidazole. Elution buffer was replaced by overnight dialyse at 4°C in 50mM Tris–HCl pH 8, 150mM NaCl. SesB was purified under native conditions (50mM Tris–HCl pH 8, 150mM NaCl). For fibril formation, proteins were incubated overnight at a concentration of 1 mg/mL at room temperature with no agitation in 50mM Tris–HCl pH 8, 150mM NaCl.

2.2.8. Negative-staining electron microscopy

Electron micrographs shown in the following chapter were kindly provided by Dr Sven Saupe. *P. anserina* hyphae were inoculated on solid medium and cultivated for 24 to 72 h at 26°C. The medium was then cut out, placed on a glass slide and examined with a Leica DMRXA microscope equipped with a Micromax CCD (Princeton Instruments) controlled by the Metamorph 5.06 software (Roper Scientific). The microscope was fitted with a Leica PL APO 100X immersion lens. For electron microscopy, 400-mesh copper electron microscopy grids coated with a plastic film (Formvar) were used. A fraction of the protein suspension (at 1 mg/mL) was put onto the grid and sedimented during 10 to 30 min in a moist Petri dish to avoid rapid desiccation. Grids were then rinsed with 15–20 drops of freshly prepared 2% uranyl acetate in water and filtered with 0.22 μm Millipore, dried with filter paper, and observed with a Phillips TECNAI 12 Biowin electron microscope at 80 kV.

2.3. Results

2.3.1. SesA, SesB and the sigma motif region of HET-eN form fibrils in vitro

With regard to the known signalling pathways involving the functional amyloids HET-s and HELLP, the *ses* locus seems to embody a novel family of amyloid-forming region (**Figure 1**). The three proteins were recombinantly overexpressed in *E. coli* with the purpose of characterizing such system. SesA full-length (FL) formed inclusion bodies in *E. coli* while SesB FL was found in the soluble fraction. Following purification and removal of any denaturing agent, the proteins spontaneously formed fibrils at room temperature in a neutral buffer. Interestingly, SesA FL fibrils were longer and less prone to form bundles, when compared to SesB. SesA FL and SesB FL filaments displayed diameters of ~8-12 nm and ~14-18 nm, respectively (**Figure 2A**). Knowing that the two proteins share a high sequence

homology on their C-terminal moieties (*i.e.*, the so-called sigma motif), we set to analyse their potential to form fibrils *in vitro*. Thus, two constructs were produced, $SesA_{148-210}$ and $SesB_{327-386}$, and were similarly found in *E. coli* inclusion bodies. In the same assembly conditions as FL constructs (*i.e.*, buffer, pH and temperature), the shorter version of SesA and SesB also formed fibrillary aggregates (**Figure 2A**). As one would expect, $SesA_{148-210}$ and $SesB_{327-386}$ filamentous assemblies show significantly lower widths of ~4-6 nm (**Figure 2A**). Interestingly, $SesA_{148-210}$ forms larger bundles of fibrils. The third partner protein HET-eN_{1-74} also showed the ability to spontaneously self-assemble into filaments of ~4-6 nm in width, which strongly interacted with each other forming large bundles. We therefore conclude that the three *Nectria haematococca* proteins, namely SesA, SesB and HET-eN encompass the sigma motif which seem to be responsible for their self-assembly into fibrils *in vitro*.

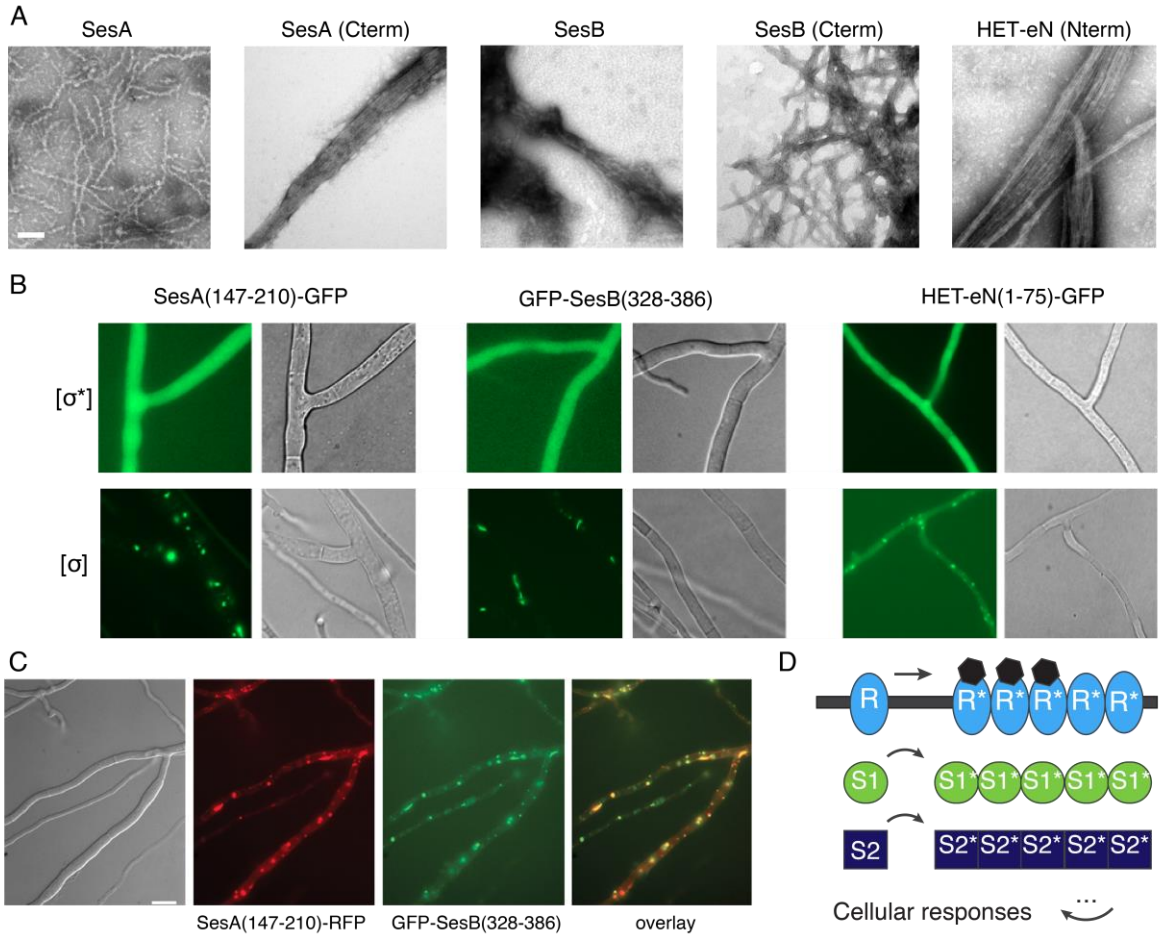


Figure 2-2: *In vitro* sigma segments spontaneously form fibrils while they can aggregate when expressed in *Podospora anserina*. (A) Negative-staining electron micrographs of fibrils of recombinant SesA, SesA-Cterm, SesB, SesB-Cterm, and HET-eN-Cterm. (B) Sigma-motif regions of SesA, SesB, and HET-eN behave as a prion forming domain *in vivo* as shown on micrographs *of P. anserina* strains expressing different molecular fusions as indicated above each micrograph, (scale bar is 5 µm). (C) Micrographs of fusion cells co-expressing SesA(147-210)-RFP GFP-SesB(328-386) in the [σ] aggregated state. (D) Schematic representation of the hypothesized cell-death signalling cascade though complex protein assembly.

2.3.2. The sigma motif region of SesA, SesB and HET-eN form prions

Further analysis was required to test whether such behaviour can be observed *in vivo*. We therefore expressed the genes encoding sigma motifs of the three partner proteins *in vivo*. We decided to express them in *Podospora anserina*, a well-known fungal system used to analyse prion propagation. $SesA_{147-210}$-GFP, GFP-$SesB_{328-386}$, HET-eN_{1-75}-GFP fusion proteins were introduced into *P. anserina* and transformants were monitored for the formation of fluorescent foci. We observed the formation of foci in a fraction of transformants containing the mentioned constructs, while fluorescence was diffuse in other cases (**Figure 2B**). We set out to determine whether such foci phenotype was transmissible between a donor strain showing the foci and a recipient strain displaying diffuse fluorescence. We found that donors efficiently converted the recipient strains to the aggregated state. Spontaneous conversion was also observed albeit with much lower frequency than when confronted to donor strains. We conclude that the sigma motifs exist in two states described as diffused fluorescence ([σ*]) and foci formation ([σ]), following the nomenclature adopted for [Het-s] and [π] prions (262). Based on the fact that phenotype transmission between the two states is possible, the sigma motif alone seems to behave as a prion element when expressed in *Podospora anserina*. Additionally, using different fluorescence labelling of SesB and SesA sigma motifs, co-localization seems to be observed in the [σ] states (**Figure 2C**, thus suggesting a high-order-based signalling mechanism (**Figure 2D**).

2.3.3. Recombinant SesB sigma motif fibrils display amyloid features

The study focuses on the expected lipase SesB, and more specifically its prion-forming segment *sigma*. $SesB_{316-386}$ and $SesA_{133-197}$ moieties share a 39% sequence identity (**Figure 1A**). The formed fibrils were then analysed using X-ray diffraction (**Figure 3B**). The so-called cross-β arrangement was detected with the typical X-ray diffraction pattern as the ultimate feature of amyloid fibrils (**Figure 3A**). Such pattern displays the signature repetitive distances of amyloid arrangements (*i.e.*, 4.7 Å and 10 Å for intra-strand and intra-sheet spaces, respectively). Clear diffraction signals at 4.7 and 10 Å demonstrate the presence of a cross-β structure within $SesB_{327-386}$ fibrils and thus seem to be yet another example of functional amyloid (**Figure 3B**).

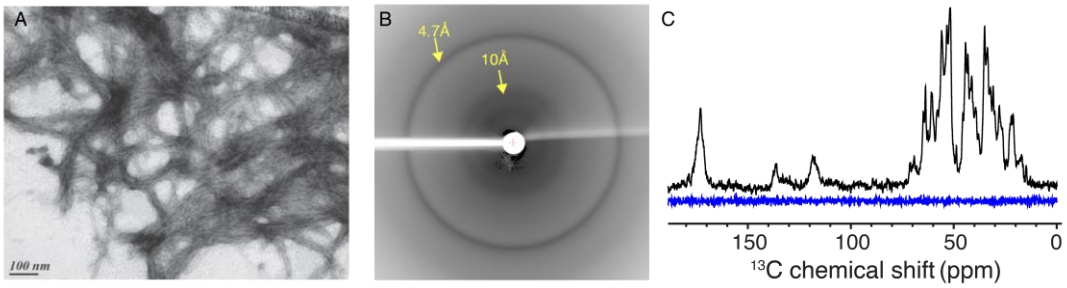


Figure 2-3: $SesB_{328-386}$ displays typical amyloid features in its fibrillary state. (A) Electron micrograph of recombinantly produced $SesB_{328-386}$ spontaneously self-assembled *in vitro* at room temperature. (B) X-ray diffraction patterns of $SesB_{328-386}$ filaments (reflections at ~4.7 Å and ~10 Å are highlighted). (C) 1D ^{13}C CP (black) and 1D ^{13}C INEPT spectra of $SesB_{328-386}$ filaments recorded on an 800 MHz spectrometer (proton frequency) at 20 kHz MAS at 276 K.

By combining high-field NMR, MAS and proton decoupling, high sensitivity and high-resolution NMR spectra of low gyromagnetic ratio nuclei like ^{13}C or ^{15}N can be recorded routinely for solid biosamples. We carried out experiments on U-[^{13}C/^{15}N]-labelled $SesB_{327-386}$ using both through space (*i.e.*, cross polarization) and through-bond (*i.e.*, INEPT) polarization transfers to detect and assess rigid and mobile segments of the protein amyloid fibrils (**Figure 3C – blue spectrum**). The one-dimensional (1D) ^{13}C-detected CP-based SSNMR experiment revealed a rather well-ordered arrangements of the $SesB_{327-386}$ monomers when stacked into fibrils as displayed by sharp NMR peaks (**Figure 3C- black spectrum**). However, the 1D ^{13}C-detected INEPT-based SSNMR experiment seem to suggest $SesB_{327-386}$ to lack mobile and flexible residues when engaged in filamentous objects (**Figure 3A**). Such observation would suggest the designed construct to completely omit any residue from the expected globular moiety of the protein.

2.3.4. The sigma motif is a ~50 amino acid-long ordered segment

To further investigate the structure of the rigid core of $SesB_{327-386}$, we employed multidimensional ^{13}C-^{13}C SSNMR with through-space mixing on a uniformly [^{13}C]/^{15}N-labelled sample. The initial 25 ms-long mixing time 2D ^{13}C-^{13}C PDSD experiment presents intra-residual correlations (**Figure 4A**). A 2D ^{13}C-^{13}C DREAM experiment recorded at 18 kHz MAS rate is displayed in **Figure 4B** and shows the amino acids with the highest local order. All recorded spectra display high signal and resolution features (^{13}C line width: 80-100 Hz) which underlines a rather well-organized assembly. The assignment process consists in identifying an NMR signal in relation to its related spins (*i.e.*, its atoms). We used both conventional and very-fast MAS to perform such task (*i.e.*, 11 and 60 kHz MAS).

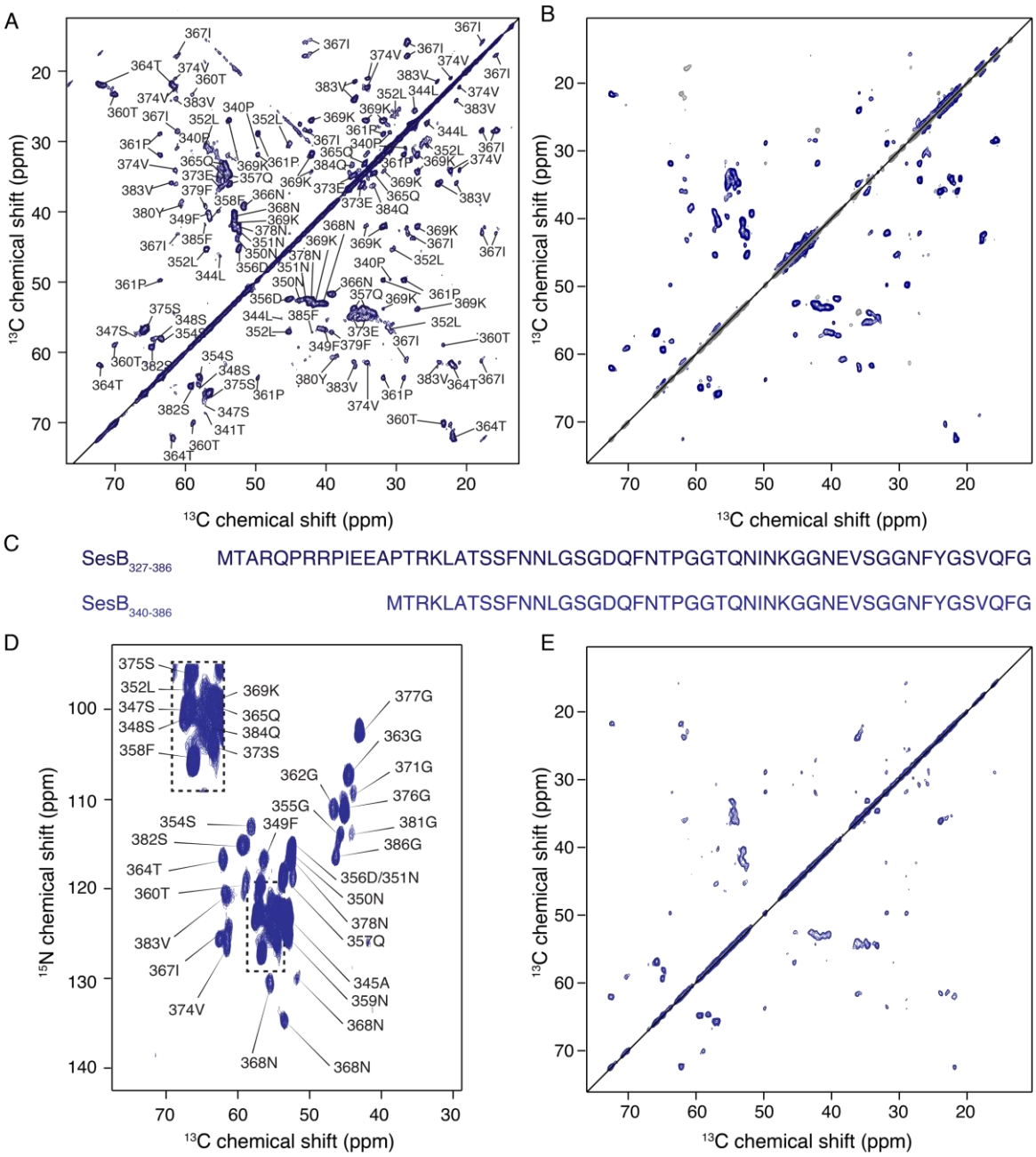


Figure 2-4: SesB$_{328-386}$ and SesB$_{340-386}$ share a similar rigid amyloid fold comprising ~50 residues. (A) 2D ^{13}C-^{13}C PDSD with a 50 ms-mixing time displaying intra-residual correlations of SesB$_{328-386}$ (peak assignments are shown). (B) 2D ^{13}C-^{13}C DREAM spectrum recorded on SesB$_{328-386}$ fibrils and exhibiting the most rigid amino acids. (C) Primary sequences of SesB$_{328-386}$ and SesB$_{340-386}$. (D) 2D ^{13}C-^{15}N projection of a 3D ^{1}H-^{13}C-^{15}N hCANH of perdeuterated and ^{1}H-back-exchanged SesB$_{340-386}$ fibrils (peak assignments are shown). (E) 2D ^{13}C-^{13}C PDSD with a 50 ms-mixing time displaying intra-residual correlations of SesB$_{340-386}$ which shows a spectral fingerprint very similar to that of SesB$_{328-386}$. The 2D spectra were recorded on an 800 MHz (proton frequency) using a 11 (*i.e.*, PDSD) or 18 (*i.e.*, DREAM) kHz MAS and a 3.2 mm probe (3-day experiments), and the 3D spectrum was recorded on a 900 MHz spectrometer (proton frequency) using a 60 kHz MAS and a 1.3 mm probe (1-day experiment)

Using a 2D ^{13}C-^{13}C PDSD experiment, the spectral analysis of the sigma motif (*i.e.*, $SesB_{327-386}$) led to the detection of about 95% of all residues (57 out of 60) and quick sequential assignments performed using a 150 ms mixing time on the same experiment led to the observation that the first ~15 amino acids were not or poorly detected within the 2D ^{13}C-^{13}C PDSD spectrum (**Figure 4A**). Those residues could otherwise be expected to be either engaging in a static disorder state (also called polymorphism) or within an intermediate dynamic regime (neither rigid nor flexible), which would explain their absence in both INEPT (used to detect flexible residues) and PDSD experiments (**Figure 3C**). Therefore, a 13-residue shorter version of the sigma motif, called $SesB_{340-386}$, was examined (**Figure 4C**). Against probable discrepancy, $SesB_{340-386}$ shared the same fold as $SesB_{327-386}$, as observed on a short mixing time 2D ^{13}C-^{13}C PDSD (**Figure 4E**). Such absolute superposition suggests that all spin systems are within the exact same environment, highlighting that assembly is not prevented nor altered following truncation. However, such observation does not necessarily rule out the possible implication of the 13 truncated residues in the amyloid core of SesB *in vivo*.

We used very-fast MAS on $SesB_{340-386}$ in parallel with conventional NMR recorded on $SesB_{327-386}$. A perdeuterated sample of $SesB_{340-386}$ was prepared and 3D ^{1}H-^{13}C-^{15}N hCANH, hCONH, hCOcaNH spectra were recorded on a 900 MHz spectrometer (proton frequency) using a 1.3 mm SSNMR probe with 60 kHz MAS (**Figure 4D**) allowing us to sequentially assign different spin systems. Sample preparation for the latter included washing steps using water, thus allowing $^{2}H/^{1}H$ exchanges for highly labile protons (e.g., backbone amide $^{15}N_H$). All side-chain chemical shifts were identified using the 2D ^{13}C-^{13}C PDSD recorded on $^{13}C/^{15}N$-labelled $SesB_{327-386}$. With 57 spin systems detected in total, all recorded spectra considered, (10 unassigned, 2 only assigned in $SesB_{327-386}$), 47 are now sequentially assigned, it can be inferred that the remaining unassigned detected residues are part of a solid and compact amyloid core. Those spin systems have yet to be identified and are listed as ambiguities: 2 Arg (R), 3 Glu/Gln (Q/E), 1 Pro (P) and 1 Gly (G), based on their chemical shifts. Interestingly, these spin systems seem to match the residue primary sequence found within the N-terminal moiety of SesB sigma motif. Some peaks remain unassigned on the 3D ^{1}H-^{13}C-^{15}N hCANH which can be explained by the fact that most residue backbone ^{15}N atoms are within a quite narrow resonance range from 115 to 130 ppm (with the notable exception of glycine backbone ^{15}N which are remarkably and easily identifiable around 110 ppm).

The sigma motif seems to respond favourably to SSNMR conditions thus suggestive of a highly ordered assembled system, as confirmed by the apparent lack of general polymorphism, expect for peak doubling of some residues (*i.e.*, 340Pro, 344Lys, 361Pro, 380Tyr, 383V). Peak

doubling could be explained by a possible plurality of the residues' conformation into two different but similar (due to the low differences between chemical shift values) states from an atomic-level point of view. We emit the hypothesis that the N-terminal part of $SesB_{327-386}$ is not entirely involved in the rigid core of the fibril, since all 15 missing residues turned out to be within that N-terminal moiety. Interestingly, other spin systems presented different signal levels upon truncation (*e.g.*, 354S, 360T, 367I, 366N, 374V, 380Y). The missing C-terminal region could either affect the conformation of those residues or even be in close proximity, thus explaining those changes. Such observations should be considered when examining a structural model of the amyloid core of SesB sigma motif.

2.3.5. SesB fibrils constitute a novel amyloid core, different from other functional amyloids

Sequential backbone and side-chain resonance assignment allows for an assessment of the secondary structure of the biomolecule. It is indeed achieved using an empirical calculation based on the assigned chemical shifts (290) (**Figure 5**). The difference between measured $^{13}C\alpha$ (and $^{13}C\beta$) chemical shifts and corresponding random coil $^{13}C\alpha$ (and $^{13}C\beta$) chemical shifts gives leave to the analysis of the so-called chemical shifts, as denoted $\Delta\delta C\alpha$-$\Delta\delta C\beta$ (291).

As a result, 5 β-sheet structural elements have been identified in the C-terminal amyloid moiety of both $SesB_{340-386}$ and $SesB_{327-386}$ (**Figure 5**). Interestingly, using the TALOS+ databank approach, very similar predictions were calculated as for the nature and location of the secondary structure (data not shown). The protein ψ and ϕ backbone angles of $SesB_{340-386}$ assigned residues were extracted from this reliable TALOS+ approach and are pending structural calculations. From such analysis, a comparison with all known amyloid folds (both pathogenic and functional) was attempted. Remarkably no match was found, as SesB amyloid core seems to display a novel previously unobserved amyloid 3D-structure. Comparing SesB secondary structure propensity to other functional or pathological amyloid structures suggests that the sigma motif belongs to a new family of functional amyloids involved in programmed cell death, and more generally, in biological signalling phenomena.

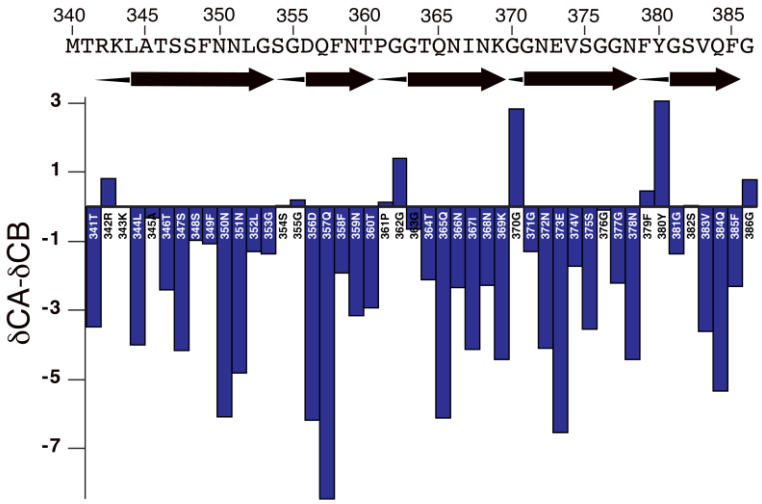


Figure 2-5: The amyloid sigma motif is composed of 5 β-strands. SSNMR-based secondary structure assignment of $SesB_{340-386}$ as a function of the amino acid sequence. Negative and positive values indicate β-strand or α-helix conformation, respectively.

An accurate model should be obtained based on *ab initio* information such as the determination of atomic-level distances between residue-based ^{13}C. Distance measuring experiments require high mixing time during the 2D ^{13}C-^{13}C PDSD pulse sequence, in order to achieve and observe long-range through space ^{13}C-^{13}C interactions (*i.e.*, PDSD 600 ms). A detailed structural investigation is performed using different labelling strategies for distance accuracy and peak dilution purpose. Two labelling strategies have been employed so far in this investigation: ^{13}C-2-glycerol- and ^{13}C-1,3-glycerol-based samples (**Figure 6A-B**). The ^{13}C-2-glycerol-based labelling scheme inserts mainly ^{13}C atoms at the α-position of the residue backbone; therefore, a long mixing time (*i.e.*, 600 ms) will allow the detection of long-range through-space interactions between the different labelled residues (292). Oppositely to ^{13}C-2-glycerol, ^{13}C-1,3-glycerol mostly permits ^{13}C-labelling of aliphatic carbons in different residues (*e.g.*, $^{13}C\beta$, $^{13}C\gamma$, $^{13}C\delta$, etc). Significant is the fact that the labelling patterns are not that straightforward, and are actually residue-dependent (293). For example, leucine is $^{13}C\alpha$-labelled in ^{13}C-1,3-glycerol-containing medium, and $^{13}C\beta$-labelled in ^{13}C-2-glycerol-containing medium. Peaks on those two spectra (*i.e.*, 2D ^{13}C-^{13}C PDSD with a 600ms mixing time) were identified as either unambiguous or ambiguous, depending on the level of certainty arising from their recognition (**Figure 6C**).

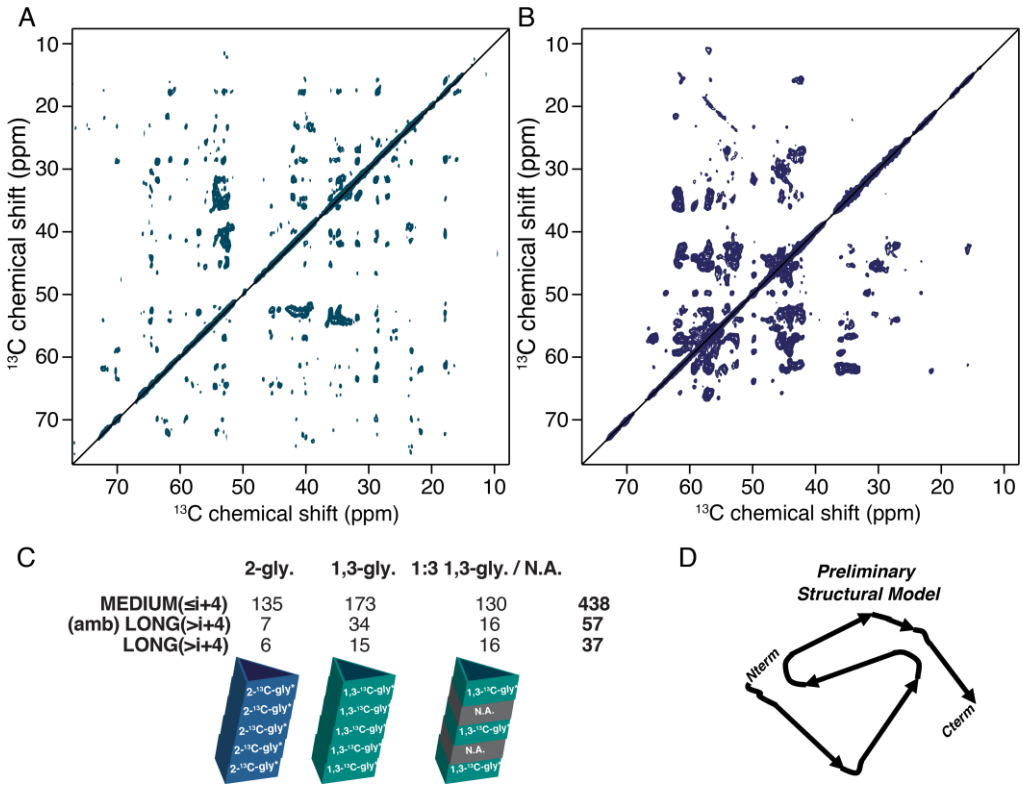


Figure 2-6: The sigma motif seems to constitute a novel functional amyloid family based on its structural singularity. (A) 2D ^{13}C-^{13}C PDSD with a 600 ms-mixing time displaying short- to long-range ^{13}C-^{13}C correlations of [$^{13}C_2$]-1,3-glycerol specifically labelled SesB$_{340-386}$ fibrils. (B) 2D ^{13}C-^{13}C PDSD with a 600 ms-mixing time displaying short- to long-range ^{13}C-^{13}C correlations of [^{13}C]-2-glycerol specifically labelled SesB$_{340-386}$ fibrils. (C) Overview of collected medium- and long-range distances based on three labelled samples: those mentioned in panel A and B, in addition to a [$^{13}C_2$]-1,3-glycerol/N.A. (natural abundance) mix with a 1:2 ratio with the purpose of removing the ambiguities of inter-molecular contacts. The 2D spectra were recorded on an 800 MHz (proton frequency) using a 11 kHz MAS and a 3.2 mm probe, for 7 to 14 days for each sample.

So far, 37 unambiguous long-range (|i - j| > 4) ^{13}C-^{13}C correlations have been successfully identified (**Figure 6C**). We report 438 medium-range correlations ((|i - j| < 4). Based on the unambiguous distances only, a preliminary model has been developed and drawn in **Figure 6D**. It now serves as a basis for current investigations. The 37 unambiguous distances have accordingly been identified as "intra-monomer" and "inter-monomer" correlations comparing the two samples (i) ^{13}C-1,3-glycerol and (ii) ^{13}C-1,3-glycerol mixed with natural abundance SesB$_{340-386}$ at a 1:2 molar ratio (**Figure 6C**). Inter-monomer interactions are indeed removed in the diluted sample. A flexible loop (unclearly determined at this point of the analysis) could join the amyloid rigid core to the globular domain. However, it is still unclear if the 13 truncated residues are somehow involved in the amyloid fold. At this point, residues 327 to 340 are thought to be involved in the amyloid core, as a result of their observation in the data (prior to truncation into SesB$_{340-386}$, some of the 13 residues were indeed detected). However, based on

the analysed data, it could be conjectured that those residues are part of a slightly more flexible region, nonetheless involved in the amyloid fold of SesB.

2.3.6. The sigma motif fold is conserved in the partner protein SesA

In the same perspective, we analysed SesA FL in its fibrillary state. We overexpressed $SesA_{1-208}$ in *E. coli* and prepared a uniformly [^{13}C]/^{15}N labelled sample to perform MAS SSNMR using a 4 mm SSNMR probe at 11 kHz magic angle spinning: a 2D ^{13}C-^{13}C PDSD was recorded for about 6 days and the displayed spectrum shows remarkable resemblance to the $SesB_{327-386}$ spectral fingerprint (**Figure 7A**). Although the presence of the globular part hinders the spectral resolution (*i.e.*, strong polymorphism observed), as observed in full-length HET-s (294), we noticed a rather well-resolved spectrum, highly indicative of an ordered arrangement. Such observation strongly suggests that both SesA and SesB share the same amyloid core through their very similar sigma motifs, from a structural point of view. Moreover, a 2D ^{1}H-^{13}C INEPT recorded on $SesA_{1-208}$ allows us to probe the globular part of SesA (*i.e.*, not engaged in the amyloid core) (**Figure 7B**). In this experiment, all flexible ^{1}H-^{13}C bonds are detected and thus the detection of all residue types suggests the globular part to be in a random coil conformation.

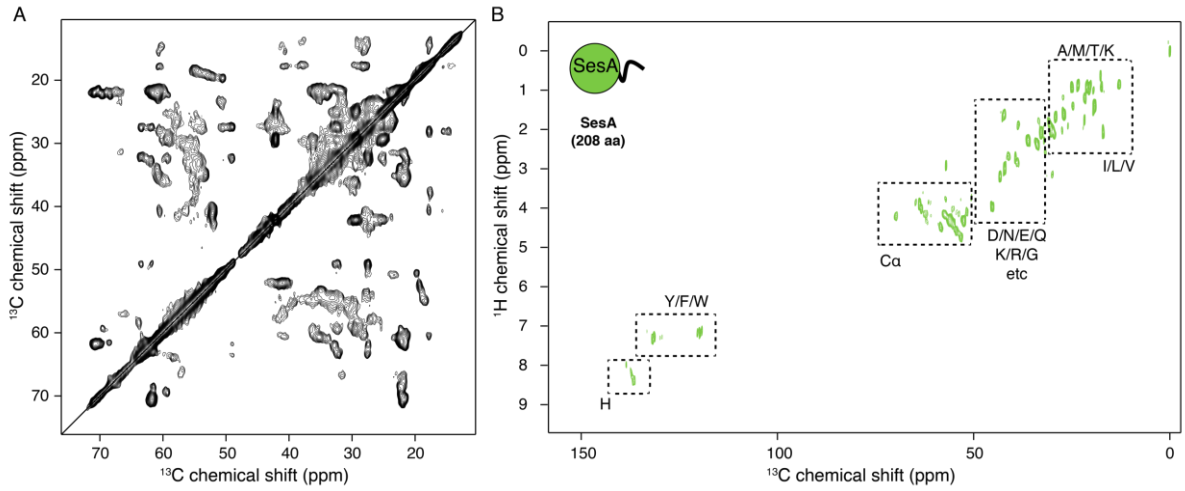


Figure 2-7: The partner protein SesA displays a similar ^{13}C-^{13}C spectral fingerprint in its assemble form suggesting a common structural arrangement. (A) 2D ^{13}C-^{13}C PDSD with a 50 ms-mixing time displaying short-range ^{13}C-^{13}C correlations of $SesA_{1-208}$ fibrils' rigid core. (B) 2D ^{1}H-^{13}C J-based INEPT displaying flexible segments of $SesA_{1-208}$ fibrils, probably accounting for the globular unstructured part.

2.4. Discussion

We studied the fungal prion SesB in its involvement in the "secteur" phenotype in *Nectria haematococca*. SesB in encoded by the *ses* locus alongside with SesA and het-eN. The three proteins share a region of ~50 amino acids in their N-terminal (SesA and SesB) or C-terminal

(het-eN) parts. This common region has been called "sigma motif" (**Figure 1A**). We report that transformants of *Podospora anserina* expressing these conserved regions fused with GFP or RFP can show distinct fluorescent foci as opposed to others displaying diffused fluorescence (**Figure 2B**), and that such phenotype can be transmitted to recipient strains by confronting them to donor strains. The sigma motif has thus been linked to prion activity and infectivity features. We furthermore prove that the three proteins, in their full-length or sigma moieties, can spontaneously self-assemble *in vitro* at room temperature in a neutral buffer, as observed using negative-staining electron microscopy (**Figure 2A**). Following such observations, we set to study SesB, the protein hypothesized to exhibit lipase activity, from a structural point of view. Like many functional biological systems with similar features, we prove that SesB is a functional amyloid as observed via X-ray diffraction and by its ability to form fibrils (**Figure 3**). We used SSNMR in an attempt to determine such amyloid fold. Using conventional and very-fast MAS, we showed that SesB sigma motif is highly rigid and ordered as indicated by narrow NMR linewidths (*i.e.*, 80-100 Hz). We were able to assign NMR resonances and identify a ~46 residue-long region which displays β-strand propensity. The secondary structure of $SesB_{340-386}$ was assessed using an empirical calculation based on Cα and Cβ chemical shifts (**Figure 4**). We report that SesB seems to fold into a 5-β-strand long amyloid core and, to the best of our knowledge, belongs to a novel and unobserved functional amyloid family involved in fungal signaling (**Figure 5**). Considering the other highly similar systems such as HET-s, HELLF, HELLP, RIP, etc., which were rather well documented in the past, the discovery of yet another system highlights the importance and abundance of such amyloid-based systems in nature. Historically, the ever-increasing interest invested in such supramolecular filaments stems from their resemblance to the pathological amyloids found in Parkinson's and Alzheimer's patients' brains, for example. Fascinating is the fact that amyloids which were long considered to be only linked to misfolding events, are now uncovered in signaling cascades or other biological functions.

2.5. Conclusion and perspectives

We report a novel functional fungal amyloid family which will join the already known HET-s, HELLF and RIP paradigms. A preliminary structural investigation is ongoing and several goals have yet to achieved. A structural calculation based on these data is being performed in collaboration with Dr. Benjamin Bardiaux and his group from Institut Pasteur in Paris, France, and will soon give rise to a more complete structural characterization of the sigma motif. We aim to determine the mass-per-length of such objects using scanning transmission electron

microscopy. Furthermore, using SSNMR, the register of the β-strand stacking could be inferred. Using a 1:1 $^{13}C/^{15}N$ mix sample of $SesB_{340-386}$, we will test the stacking properties of the cross-β arrangement of sigma amyloid fibrils, in other words, assess whether each amino acid interacts with itself in the adjacent monomer, unlike the well-known two-fold solenoid observed in HET-s (55). SesA and het-eN will also be part of this large and complex study. An interesting point would be studying the molecular conformation of the SesB amyloid motif in the context of a truncated amyloid motif in isolation and in the context of the full-length protein aggregated into fibrils. Past observations on other amyloids have reported two major behaviours: in the case of HET-s, the globular domain remains folded as a molten globular structure in the context of full-length HET-s amyloid fibrils (266), while in the case of the prion Ure2, the globular domain has been observed to be part of the rigid core of the fibrils (295). We will investigate whether the non-amyloid domain of SesB (and later for SesA) is involved in the structural scaffolding of full-length protein into amyloid. It will help us to understand if the non-amyloid domain of SesB, in analogy to the HeLo domain in HET-s, can undergo a trans-conformational change during aggregation to help relocating SesB at the membrane. The hypothesis that SesB exhibits lipase activity might thus be confirmed and might pinpoint mutations that could help in understanding this amyloid fold and its differences with HET-s.

Chapter 3: Molecular architecture of bacterial amyloids in *Bacillus* biofilms

3.1. Nonstandard abbreviations

ATR-FTIR: Attenuated total reflection Fourier-transform Infra-red

DLS: Dynamic light scattering

INEPT: Insensitive nuclei enhanced polarization transfer

MAS: Magic angle spinning

NMR: Nuclear magnetic resonance

TEM: Transmission electron microscopy

ThT: Thioflavin T

3.2. Abstract

The formation of biofilms provides structural and adaptive bacterial response to the environment. In *Bacillus* species, the biofilm extracellular matrix is composed of exopolysaccharides, hydrophobins and several functional amyloid proteins. We report, using multi-scale approaches such as solid-state NMR, electron microscopy, X-ray diffraction, DLS, ATR-FTIR and immune-gold labeling, the molecular architecture of *Bacillus subtilis* and pathogenic *Bacillus cereus* functional amyloids. SSNMR data reveal that the major amyloid component TasA in its fibrillar amyloid form contain β-sheet and α-helical secondary structure, suggesting a non-typical amyloid architecture in *Bacillus subtilis*. Proteinase K digestion experiments indicate the amyloid moiety is approximately ~100 amino acids long, and subsequent SSNMR and FTIR signatures for *Bacillus subtilis* and *Bacillus cereus* TasA filaments highlight a conserved amyloid fold albeit with substantial differences in structural polymorphism and secondary structure composition. Structural analysis and co-assembly data on the accessory protein TapA in *Bacillus subtilis* and its counterpart CalY in *Bacillus cereus* reveal a catalyzing effect between the functional amyloid proteins and a common structural architecture, suggesting a co-assembly in the context of biofilm formation. Our findings highlight non-typical amyloid behavior of these bacterial functional amyloids, underlining structural variations between biofilms even in closely related bacterial species.

3.3. Keywords

Bacillus - Biofilms - Functional amyloids - Extracellular matrix - Solid-state NMR

3.4. Introduction

Functional amyloids (296) are protein aggregates that have emerged as one of the most outstanding contributors to bacterial physiology and behavior (297-299). Amyloid functions

(300) can range from purely structural roles (69), to supporting multicellular bacterial communities (301), mediating interaction with surfaces (302), driving assembly of the extracellular matrix (65, 303), and more recently in neurodegeneration (304). This broad functional versatility and their remarkable ability to form filamentous self-assemblies, robust and resistant to most denaturing or environmental conditions (305-307), contribute to functional amyloids being considered as a molecular swiss army knife in the bacterial kingdom.

A fascinating step in the bacterial life cycle is the assembly of bacterial communities called biofilms. Biofilms are embedded in an insoluble, non-crystalline supramolecular matrix mostly composed of water, proteins, lipids and polysaccharides. The coordinated expression of a pool of structural elements permits the building of the extracellular matrix, a sort of bacterial tissue that provides physical support to the biofilm, protection to the cells, mediates interactions with the environment, and regulates the flow of nutrients and other signaling molecules (308). Functional amyloid proteins are one of these structural components, and successive studies are highlighting their role in different steps of the biofilm life cycle: adhesion, extracellular matrix construction, and motility (309-311).

Bacillus subtilis, a soil-dwelling bacterium, is one the most studied microbes that has vastly contributed to our understanding of the genetic regulatory pathways and the structural elements of biofilms (312). A previous study of the TasA protein, a major protein component of the extracellular matrix of B. *subtilis* biofilms (313), provided biochemical evidence for its amyloid-like nature: TasA forms filaments with dye-binding activity of Thioflavin T or Congo Red, and can repolymerize after disaggregation with acid treatments (314). A double mutant of *B. subtilis, eps sinR,* served as a natural overproducer of TasA that was purified mostly as aggregates and co-purified with a small amount of TapA, an accessory protein which also contributes to the biofilm assembly (315). Proteins orthologous to TasA are widespread in the *Bacillus* genus, and we recently showed the enrolment of the orthologues TasA and CalY in *B. cereus* biofilm formation and cell-surface fiber formation (316).

Bacterial amyloids offer a paradigm to understand how the formation of these well-ordered objects at the mesoscopic scale can counterbalance potential amyloid-associated deleterious effects. Moreover, while most of the disease-related amyloid systems are found in the intracellular space, the vast majority of bacterial functional amyloids evolved at the bacterial surface and need to sustain variable and heterogeneous external conditions to fulfill their functional roles, as is the case for *Bacillus* TasA filaments found in biofilms. A detailed structural description of these functional amyloids, in their relevant fibrillar state, is required

to understand how their conformation responds to environmental stimuli in the context of biofilm formation and evolution.

Two recent structural studies (317, 318) have focused on *B. subtilis* TasA in its monomeric and fibrillar state. Monomeric $TasA_{261}$ (amino acids 28-261), crystallized at pH 4.6, has revealed a jellyroll fold (317) composed of α-helical and β-sheet secondary structure, and fibrils with a predominant β-sheet conformation from SSNMR analysis that exhibit a typical cross-β signature by X-ray diffraction. Conversely, another study (318) demonstrated an alternative conformation of TasA fibrils lacking a canonical cross-β signature by X-ray fiber diffraction. Although we have substantiated the ability of TasA to form β-sheet-rich fibrils through the observation of straight unbranched fibrils and its binding to Thioflavin T and Congo red dyes (315), so far, only scarce information has been reported on the structural organization of TasA within its fibrillar form.

We hereafter aimed to investigate inter-species structural conservation of these biofilm forming TasA fibrils, comparing *B. subtilis* and *B. cereus*, in addition to analyzing the singular accessory protein TapA, or CalY respectively. Using a combination of recombinant expression, *in vitro* polymerization, X-ray diffraction, electron microscopy, SSNMR and Fourier-transformed Infrared spectroscopy (FTIR), we here studied the molecular conformation of *B. subtilis* and *B. cereus* TasA and the accessory proteins TapA (for *B. subtilis*) or CalY (for *B. cereus*) in their filamentous amyloid state. We found that TasA proteins from both species, as well as TapA and CalY, form amyloid-like fibers exhibiting canonical cross-β organization, while a part of the sequence maintains a well-ordered α-helical fold within the filaments in a bacterial-dependent manner. Substantial differences in aggregation kinetics, FTIR, SSNMR line broadening and chemical shift-based secondary structure suggest that *B. subtilis* and *B. cereus* TasA filaments adopt virtually the same three-dimensional protein fold architecture in the mature filaments, while having significant structural differences based on their secondary structure composition. Furthermore, we used co-assembly experiments to compare the molecular architecture of *B. subtilis* and *B. cereus* TasA amyloids in the presence of their accessory proteins. The data suggest that a putative co-assembly is structurally possible in the context of the biofilm formation. Our results uncover complex interdependent assembly mechanisms between TasA and its accessory protein that differ between the two *Bacillus* species.

3.5. Materials and methods

3.5.1. Bacterial strains and culture conditions

The strains used in this study were *Bacillus subtilis* NCIB3610, an undomesticated prototrophic strain from the laboratory collection of Dr. Diego Romero, *Bacillus cereus* ATCC14579, and derivative mutant strains in the *tasA* gene CA017 (*Bacillus subtilis* NCIB3610 *tasA*::Km) (314) and JCA111 (*Bacillus cereus* CECT148 *tasA*::Km) (316). For TapA immunolocalization assays, a *B. subtilis tasA eps* double mutant was used (*tasA*::*km*, *epsA-O*::*tet*). For cloning and plasmid replication, *E. coli* DH5α was used. *E. coli* BL21(DE3) pLysS (Merck) was used for protein expression and purification. All strains were grown in Luria-Bertani (LB) medium: 1% tryptone (Oxoid), 0.5% yeast extract (Oxoid) and 0.5% NaCl. MOLP (Medium Optimized for Lipopeptide Production) medium was used for extracellular complementation assays in experiments performed with *B. subtilis* and was prepared as described (319). TY medium: 1% tryptone, 0.5% yeast extract, 0.5% NaCl, 10 mM $MgSO_4$, and 1 mM $MnSO_4$ was used for extracellular complementation assays in experiments performed with *B. cereus*. Media were solidified by addition of bacteriological agar (Oxoid) to a final concentration of 1.5%. M9 minimal medium (composition for 1 liter was: 48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.01 mM $ZnCl_2$, 0.001 mM $FeCl_3$, 0.1 mM $CaCl_2$, 18.7 mM NH_4Cl, 0.20% glucose and 10 mL of 100x MEM vitamins solution (Merck)) was used for protein expression for NMR studies. M9 was enriched in $^{15}NH_4Cl$ and ^{13}C labeled carbon sources ([U-^{13}C] D-glucose) in order to produce labeled recombinant proteins.

The antibiotics used and their final concentrations were: 100 μg/ml ampicillin, 50 μg/ml kanamycin, 10 μg/ml kanamycin and tetracycline (10 μg/ml).

3.5.2. Plasmid construction

Plasmid pDFR6 (pET22b-*tasA*), containing the open reading frame of the *tasA* gene from *B. subtilis* NCIB3610 without the signal peptide or the stop codon, was constructed as previously described (315) and used for TasA (*B. subtilis*) expression and purification. Plasmid pJC1 (pDEST17-*tapA*) was constructed using the Gateway Cloning System (ThermoFisher). Primers TapA–attb1–F (5'–ggggacaagtttgtacaaaaaagcaggcttgATATGCTTACAATTTTCCGATGAT-3') and TapA–attb2–R (5'–ggggaccactttgtacaagaaagctgggtcTTACTGATCAGCTTCATTGCT-3') contain the recognition sites for BP clonase enzyme and were used to amplify the open reading frame of the *tapA* gene. The insert was then cloned into pDONR207 by a BP recombination reaction. The plasmid pDONR207-tapA was purified from the entry clone and then, the insert was cloned into the

pDEST17 vector by a LR recombination reaction, creating the final expression vector. The 6his-tag is located at the N-terminal region of the *tapA* sequence. Plasmid pJQN1 (pET30a-*tasA*) and pJQN2 (pET30a-*calY*) were constructed by amplifying the open reading frame of the *tasA* or *calY* genes respectively from *B. cereus* CECT148, without the signal peptide or the stop codon, using the primers NdeI.TasA.expr.Fw (5'-tttttcatatgTTTTTCAGTGATAAAGAAGTGTCAAAC-3') and XhoI.TasA.expr.Rv (5'-tttttctcgagAGAACCGCGTGGCACCAGTTTTTCTTCACCTGCTGTTTG-3') for *tasA* or NdeI.CalY.expr.Fw (5'-tttttcatatgGCATTGGGGTTAGCTTTAATTGG-3') and XhoI.CalY.exp.Rv (5'-tttttctcgagAGAACCGCGTGGCACCAGTTTTTCTTCCCCAGCTTCTTG-3') for *calY*. The fragments were digested with NdeI and XhoI and cloned into pET30a vector (Merck) generating an in-frame fusion with the 6his-tag. The reverse primer includes the sequence corresponding to the thrombin recognition site.

3.5.3. Cell growth and protein expression

A single freshly transformed BL21(DE3) pLysS *E. coli* was picked and cultured in 10mL of LB media with 100 ug/mL ampicillin for TasA-bs and TapA, and 30 ug/mL kanamycin for TasA-bc and CalY at 37°C under shaking condition for 5 h to reach exponential growth phase. Cells were harvested (1000 g, 10 min) and resuspended in 100 mL of labeled minimal medium (1 g/L $^{15}NH_4Cl$, 2 g/L [U-^{13}C] D-glucose) supplemented with antibiotics and grown overnight at 30°C, 200 rpm until stationary growth phase. 10% (v/v) of the culture was used to inoculate 1 L of labeled minimal medium supplemented with antibiotics, cultured at 37°C, 220 rpm until OD_{600nm} reached 0.8 and induced with 0.75 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). TapA was expressed at 18°C for 20 h and TasA-bs, TasA-bc and CalY at 30°C for 20 h. Cells were harvested (6000 g, 30 min, 4°C, Beckman Coulter JLA 8.1 rotor) and frozen at -80°C until purification.

3.5.4. Protein purification

Cells were resuspended in buffer A (50 mM Tris, 150 mM NaCl, pH 8), sonicated on ice (3x45 sec, 40%, Bandelin sonopuls VS 70T probe) and centrifuged (15 000 g, 60 min, 4°C, Eppendorf F34-6-38 rotor). The supernatant was discarded as proteins were mainly found in inclusion bodies. The pellet was resuspended in buffer A supplemented with 2% Triton X-100, incubated at 37°C under shaking condition for 20 min and centrifuged (50 000 g, 10 min, 4°C, Beckman Coulter JA 25.50 rotor). The pellet was extensively washed with buffer A, centrifuged (50 000 g, 10 min, 4°C), resuspended in denaturing buffer (Tris 50 mM NaCl 500 mM, 6 M GuHCl)

and incubated at 60 °C overnight until complete solubilization. Lysate was clarified by centrifugation (200 000 g, 1h, 16°C, Beckman Coulter 50.2 Ti rotor). Solubilized inclusion bodies were loaded on a HisTrap HP 5 mL column (GE Healthcare) previously equilibrated in the binding buffer (50 mM Tris, 0.5 M NaCl, 20 mM imidazole, 8 M urea, pH 8). Proteins were eluted from the column with a linear gradient of elution buffer (50 mM Tris, 0.5 M NaCl, 500 mM imidazole, 8 M urea, pH 8). After the affinity chromatography step, proteins were loaded on a HiPrep 26/10 desalting column (GE Healthcare) to exchange buffer for 1 % acetic acid pH 3. Purification fractions were loaded on a 13% Tris-Tricine SDS-PAGE to check purity.

3.5.5. Biofilm formation and extracellular complementation assays

These assays were performed in the laboratory of Dr. Diego Romero. Biofilm experiments of *B. subtilis* NCIB3610 were done by adding 10 µL of a starting culture to 1 mL of MOLP medium in 24-wells plates (314). The plates were incubated at 28°C without agitation and the formation of wrinkly pellicles at the air-liquid interphase were recorded after 48 h of incubation. For *B. cereus* ATCC14579, 10 µL of starting culture were added to 1 mL of TY medium and the plates were incubated under the same conditions for 48 h. The formation of biofilm was monitored by observing the formation of bacterial biomass adhered to the wall of the wells, further revealed by staining with 1% solution of crystal violet. Extracellular complementation experiments were performed by inoculating the *tasA* or *tapA* single mutants of *B. subtilis* NCIB3610 or *tasA* single mutant of *B. cereus* ATCC14579 in 1 mL of the corresponding media as described above. A total amount of 80 µg (3 µM) or 160 µg (6 µM) of freshly purified *B. subtilis* TasA or *B. cereus* TasA were added to the corresponding culture. Photographs were taken at different time points over the course of 72 h. The same volume of protein buffer (10 mM Tris, 100 mM NaCl), or inclusion bodies, were additionally included as controls.

3.5.6. Assembly of filaments

The concentration of the protein recovered after HPLC purification was measured using the absorbance of aromatic amino acids at 280 nm. The sample was then concentrated in Amicon Ultra-15 (cut-off 3 kDa) centrifugal filter units (Merck Millipore) to 0.25 mM and dialysed against a 300 mM Tris, 1% acetic acid, 0.02% azide, pH 7.4 solution. Self-assembly was then promoted under agitation for 1 week. Fibrils were then extensively washed in water, ultracentrifuged (300 000 g, 1 h, 4°C, Beckman Coulter TLA 120.1 rotor) and transferred to the SSNMR rotor.

3.5.7. ThT measurements

The experiment was initiated by placing the 96-well plate at 37 °C and shaking at 100 rpm in a plate reader (TECAN Infinite® M1000 PRO). The ThT fluorescence was measured through the top of the plate every 1 min (with excitation filter 440 nm and emission filter 480 nm) with continuous shaking at 100 rpm between reads. Every measurement was recorded in triplicate (15 μM ThT and 10, 20, 25, 40 or 50 μM protein homogenate). For co-assembly experiments, a protein concentration of 50 μM was used.

3.5.8. Dynamic light scattering experiments

Performed in the laboratory of Dr. Diego Romero. A 50 μM solution of TasA in 1% acetic acid was treated with NaOH 10% to adjust the pH to 7.0 and filtered through a 0.46 μm syringe filter. The size measurements were carried out in a Malvern Zetasizer Nano ZS with a 632.8 nm laser wavelength as the excitation source and using a 1 cm pathway polystyrene cuvette as the sample holder. One milliliter of the protein solution was transferred to the cuvette and kept in the system the whole experiment without stirring. The DLS signal was monitored at 25ºC for each protein at 60 min intervals and a count rate of (2-4) x 103 counts/s for TasA-bs and 30 min intervals and (1-17) x 103 counts/s for TasA-bc. Zetasizer software version 6.34 (Malvern Instruments Ltd) was used in data collection and processing.

3.5.9. Co-assembly SSNMR experiments

Soluble $^{13}C/^{15}N$-labeled proteins were mixed with natural abundance proteins in a 4:1 ratio (^{13}C-^{15}N TasA-bs:TapA / ^{13}C-^{15}N TapA:TasA-bs / ^{13}C-^{15}N TasA-bc:CalY / ^{13}C-^{15}N CalY:TasA-bc), and then concentrated, dialyzed and assembled as described above.

3.5.10. Proteinase K digestion assay

Performed in the laboratory of Dr. Diego Romero. A 50-μg solution of aggregated TasA-bs protein was treated for 1 h at 37°C with 0.02 mg/mL of proteinase K in 50 mM Tris pH 8, 150 mM NaCl for various times (1 min, 5 min, 15 min, 20 min, 30 min, 45 min or 1 h). Reactions were stopped by addition of 1 vol. of Laemmli buffer and heated at 100°C for 5 min. The resulting samples were separated by SDS-PAGE in 12% acrylamide gels and subsequently stained with Coomassie brilliant blue.

The sequence corresponding to the amyloid core of TasA-bs was identified by tandem mass spectrometry using a “nano” ion trap system (HPLC-ESI-MS/MS). Briefly, the bands obtained after electrophoresis were cut, washed and destained. Subsequently, disulfide bridges were reduced with DTT, cysteines were alkylated by the use of iodoacetamide and trypsin digestion was performed in-gel to extract the peptides corresponding to the protein samples. All this

process was carried out automatically using an automatic digester (DigestPro, Intavis Bioanalytical Instruments). The peptides were then concentrated and desalted using a capture column C18 ZORBAX 300SB-C18 (Agilent Technologies, Germany), 5x0.3 mm, with 5 µm particle diameter and 300 Å pore size, using a gradient of 98% H_2O:2% acetonitrile (ACN)/0.1% formic acid (FA) with a flow rate of 20 µL/min for 6 min. The capture column was connected in line with a ZORBAX 300SB-C18 analytical column (Agilent Technologies), 150x0.075 mm, with 3.5 µm particle diameter and 300 Å pore size, through a 6-port valve. Elution of the samples from the capture column was performed in gradient, using FA 0.1% in water as mobile phase A, and FA 0.1% in ACN 80%/water 20% as mobile phase B. The LC system was coupled through a nanospray source (CaptiveSpray, Bruker Daltonics) to a 3D ion trap mass spectrometer (amaZon speed ETD, Bruker Daltonics) operating in positive mode with a capillary voltage set to 1500 V and a sweep range: m/z 300-1,500. A "data-dependent" acquisition was carried out in automatic mode that allowed obtaining sequentially a MS spectrum in "full scan" (m / z 300_1400) followed by a MS spectrum in tandem by CID of the eight most abundant ions. For identification, the software ProteinScape 3 (Bruker Daltonics) coupled to the search engine Mascot 3.1 (Matrix Science) was used, matching MS/MS data against the databases Swiss-Prot and NCBInr.

3.5.11. X-ray diffraction

Fiber diffraction patterns were measured at 4°C on a Rigaku FRX rotating anode X-Ray generator equipped with a Pilatus 200 K hybrid pixel detector at the copper wavelength. The concentrated hydrated samples were mounted in a MicroLoops™ from Mitegen on a goniometer head under the cold nitrogen flow. Each diffraction pattern corresponds to a 360° rotation along the phi axis with an exposure time of 720 s after subtraction of a “blank” image of the same exposure time with only the loop on the goniometer head.

3.5.12. Solution NMR

Samples were concentrated to ~1 mM and supplemented with 10% D_2O and ~50 µM DSS as an internal ^{1}H chemical shift standard. Indirect dimensions were referenced to DSS according to their gyromagnetic ratios and specific nuclear observation frequencies. ^{1}H-^{15}N SOFAST-HMQC spectra were acquired at 700 and 800 MHz ^{1}H Larmor frequency spectrometers (Bruker Biospin), 310 K, using a triple resonance room temperature probe and a triple resonance cryo-probe. Backbone assignment was performed using HNCACB, HNCO, HNCA, HNcaCO, HNcoCA and CBCAcoNH experiments with 20 or 25% non-uniform sampling (320) at 700 MHz at a temperature of 310 K.

3.5.13. Solid-state NMR

Experiments were performed at a MAS frequency of 11 kHz on triple resonance 4 mm and 3.2 mm MAS probes using 600 MHz and 800 MHz ^{1}H Larmor frequency spectrometers (Bruker Biospin). The sample temperature was set to ~278 K and chemical shifts were calibrated using DSS as an internal reference. A cross-polarization contact time of 1 ms was used for a decoupling strength of 90 kHz using SPINAL-64. The rigid part of the fibrils was probed using two-dimensional ^{13}C-^{13}C correlation spectra using a proton-driven spin-diffusion (PDSD) mixing scheme. Intra-residue correlations were probed using a PDSD mixing time of 50 ms for a total experimental time of ~3 days, per sample. Two-dimensional ^{1}H-^{13}C INEPT were recorded for a total experimental time of ~18 h, per sample.

3.5.14. Secondary structure analysis of fibrils by solid-state NMR

Conformation propensity assessment was determined using secondary structure assignment calculations (*i.e.* differences of secondary chemical shifts $\Delta\delta Csec = \delta C\alpha - \delta C\beta$, with δCx = Cx(experimental) - Cx(random coil - BMRB)). Negative values therefore are considered to be highly indicative of a β-strand-like conformation. For each residue type (*i.e.* alanine, serine, and threonine), the total peak volume was calculated on Sparky using integration tools and used for percentage assessment of α/β conformation. Error bars were determined using measurements on both sides the 2D ^{13}C-^{13}C PDSD diagonal. The theoretical α/β peak distribution on the spectrum is based on cross-calculations between reported secondary structure-based chemical shifts (321) and secondary structure assignment.

3.5.15. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR-FTIR)

ATR-FTIR spectra were recorded on a Nicolet 6700 FT-IR spectrometer (Nicolet Instrument, Madison,WI) equipped with a liquid nitrogen cooled mercury–cadmium–telluride detector (ThermoFisher Scientific, San Jose, CA, USA), with a spectral resolution of 4 cm^{-1} and a one-level zero filling. Concentrated and hydrated samples were deposited on a germanium ATR crystal in a 10 µL droplet. The droplet was dried and the spectrum (200 scans) of the proteins was recorded. To determine the secondary structure elements of each protein, spectra were analyzed with an algorithm based on a second-derivative function and a self-deconvolution procedure (GRAMS and OMNIC softwares, ThermoFisher Scientific) to determine the number and wavenumber of individual bands within the spectral range of the amide I band (1700-1600 cm^{-1}).

3.5.16. Electron microscopy and immunoelectron microscopy

Negative staining micrographs were recorded using one droplet of 2-fold serial dilutions of TasA-bs, TasA-bc, TapA, and CalY assemblies stained with a 2% uranyl acetate (w/v) solution for 1 min and dried in the dark. Samples were observed using a FEI CM120 transmission electron microscope at an accelerating voltage of 120 kV under TEM low-dose mode at a magnification of 40.000X for TasA-bs/TasA-bc and 25.000X for TapA/CalY. TEM images were recorded using a Gatan USC1000 2k x 2k camera.

Performed in the laboratory of Dr. Diego Romero. For immunolabeling assays, samples from cells growing under biofilm inducing conditions were taken and incubated over copper grids for 2 h. Grids were then floated on blocking buffer (1% non-fat dry milk in PBS with 0.1% Tween20) for 30 min and on anti-TapA of *B. subtilis* at a 1:150 dilution for 2 h and rinsed in PBST. Then, the grids were floated in goat-antirabbit 20 nm gold-conjugated secondary antibody (TedPella) at a 1:50 dilution for 1 h and rinsed in PBST and, finally, in ultrapure water four times for 5 min each. Samples were dried at RT and negatively stained with 1% uranyl acetate for 30 s followed by two washes in ultrapure water. For *B. cereus* TasA and CalY, the monoclonal anti-TasA or anti-CalY primary antibodies were raised against the peptide KGISAGKSDKFK from the TasA sequence, or the peptide QSEPVYETTLAD from the CalY sequence (NovoPro) and used at a 1:50 dilution. Goat-antirabbit 20 or 30 nm gold-conjugated secondary antibodies were used to visualize the signal associated to each protein. Samples were visualized under a FEI Tecnai G2 20 TWIN Transmission Electron Microscope at an accelerating voltage of 80 kV. The images were taken using a side-mounted CCD Olympus Veleta with 2k x 2k Mp.

3.6. Results

3.6.1. *E. coli* expression of TasA and functionality

The TasA protein is well conserved across the *Bacillus* genus but can be phylogenetically separated into two distinct groups: firstly in *B. subtilis* (hereafter mentioned as TasA-bs) and other non-pathogenic species such as *B. amyloliquefaciens*, the protein is approximately 237 amino acids long (261 amino acids for TasA-bs). Secondly, in pathogenic species such as *B. cereus* (hereafter mentioned as TasA-bc), the protein is ~170 amino acids long (197 amino acids for TasA-bc). TasA-bs and TasA-bc share ~35 % sequence homology, mostly located in the N-terminal half (**Figure 1**), and this seems to be sufficient for function in bacterial multicellularity. In *B. subtilis* (NCBI 3610), TasA-bs is known to participate in the formation of wrinkly pellicles at the air-liquid interface (313), and in *B. cereus* (ATCC 14567) TasA-bc

stabilizes biofilm surface-adherence (316). *B. cereus* TasA and CalY share a significant sequence identity (~62%), while interestingly, only ~33% is observed between TasA-bs and TapA (**Supplemental Figure 1**). Unlike the TasA proteins, CalY and TapA do not share high sequence identity (~21%).

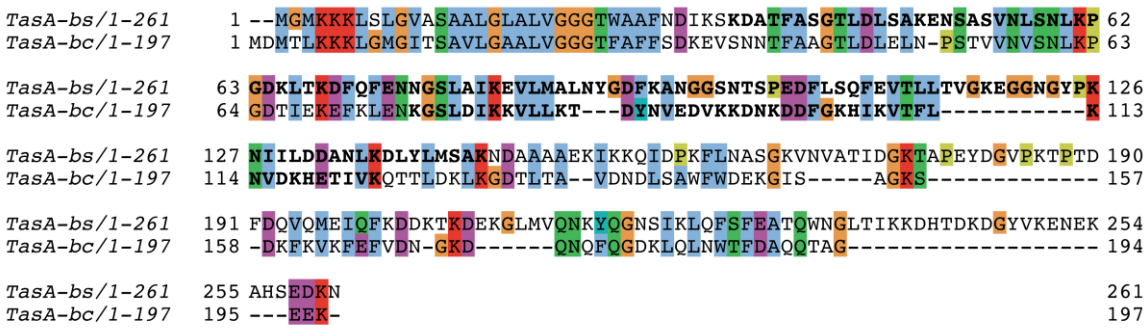


Figure 3-1: Amino acid sequence alignments of the two TasA homologs (*B. subtilis* $TasA_{1-261}$ and *B. cereus* $TasA_{1-197}$).

In our effort to engage in structural characterization of these filaments, we developed and optimized protocols for over-expression and purification of both TasA-bs and TasA-bc. Heterologous TasA expression allowed us to investigate the aggregation and conformation of filaments formed solely by TasA, in the absence of the TapA accessory protein involved in TasA-bs polymerization (315). Following the overexpression and purification of TasA-bs and TasA-bc, both samples were kept soluble in acidic buffer at 4ºC. In order to demonstrate the functionality of the recombinant TasA-bc and TasA-bs proteins, complementation assays were performed, as previously described (314). TasA-bc protomers rescued the wrinkly pellicle of the Δ*tasA* mutant in a concentration dependent manner (**Figure 2, top row**). In a similar way, the addition of TasA-bc to a culture of *B. cereus* Δ*tasA* mutant led to the formation of a sort of pellicle in the air-liquid interphase (**Figure 2, bottom row**). Given that inclusion bodies (IBs) are insoluble and composed of aggregated and non-functional form of the purified protein, 20 µl of the isolated IBs suspension (OD_{600} ~ 0.5) were added to 1 ml of the medium in a 24-well microplate. Interestingly, inclusion bodies made of each protein failed to restore the wildtype phenotype in both mutant strains. Such results acquired on *B. cereus* are interestingly quite similar to those reported on *B. subtilis*, regarding external complementation experiments (314).

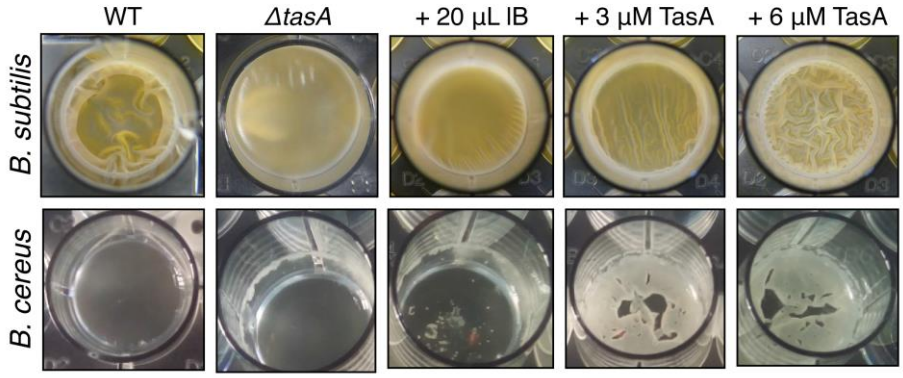


Figure 3-2: TasA expressed in *E. coli* retains functionality. TasA purified from *E. coli* rescues the phenotype of Δ*tasA* mutant of *B. subtilis* (top row) or *B. cereus* (bottom row) in a concentration dependent manner. Inclusion bodies (IB) made of TasA are not functional.

3.6.2. *B. cereus* TasA and CalY maintain an unfolded conformation in solution

We first investigated the behavior of TasA-bc and CalY in their monomeric forms. The 2D ^{1}H-^{15}N solution NMR spectrum of TasA-bc (**Figure 3A and Supplemental Figure 2A-B**) shows markedly narrow ^{1}H dispersion (~1 ppm) across the backbone amide HN spectral region – a spectral feature typically associated with unfolded proteins (322). Spectra showed the approximate number of expected H_N peaks and no signs of degradation were observed for TasA-bc during the course of additional experimentation (>4 weeks). Using inter- and intra-residue 3D triple-resonance solution NMR experiments, chemical shifts for 175 of the TasA-bc residues were sequentially assigned. Backbone and ^{13}Cβ chemical shifts are strong reporters on protein secondary structure and typical NMR analyses report these values as referenced to random-coil chemical shifts in order to assess the α-helical or β-strand populations (323). However, transient or low-propensity secondary structure can be difficult to assess by secondary chemical shift-based methods due to variations in the secondary structure sensitivity of different nuclei and residue types, as well as the magnitude of the structure-associated chemical shift differences. To this end, TasA-bc chemical shifts were evaluated using secondary structure propensity scores (SSP) (324), using only ^{13}Cα and ^{13}Cβ chemical shifts (**Figure 3B**). In good agreement with the limited H_N dispersion, the SSP analysis for TasA-bc primarily indicates that the entire protein is significantly unstructured, with SSP scores within the range of ± 0.3 for the majority of residues.

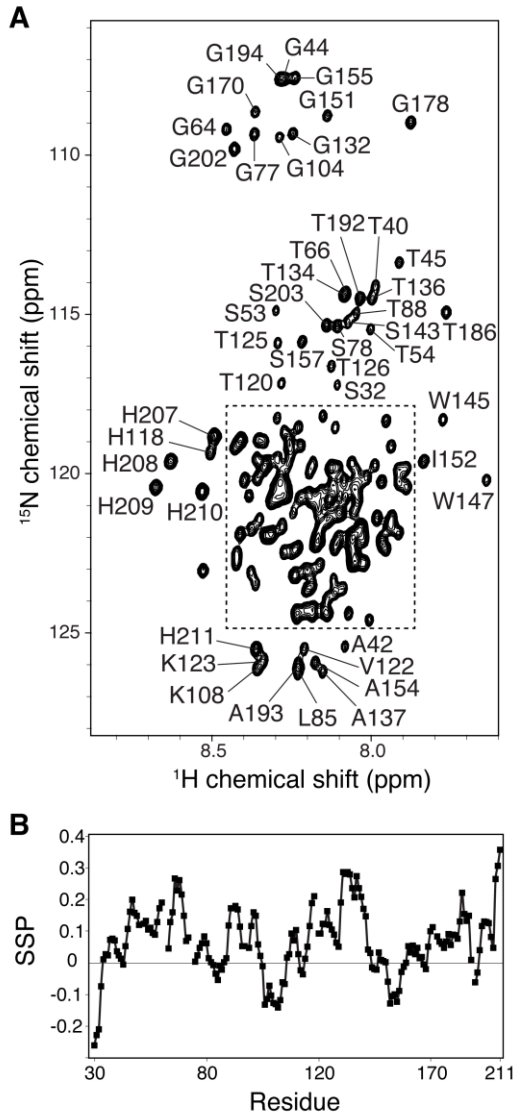


Figure 3-3: *B. cereus* TasA is monomeric and unfolded in solution. (A) 2D ^{1}H–^{15}N solution NMR spectrum of TasA-bc, (B) The secondary structure propensity (SSP) score, as calculated from $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts, using SSP with default parameters. SSP scores of +1 and -1 represent fully formed α-helix and β-strand, respectively. For clarity, sequential assignments within the dashed line area are reported separately in **Supplemental Figure 2A-B**.

In comparison to TasA-bc, the 2D ^{1}H-^{15}N solution NMR fingerprint spectrum of CalY (**Supplemental Figure 2C**) displays equally limited ^{1}H dispersion (~1 ppm), also indicating a predominantly unstructured conformation. Such observations are quite interesting as no data has yet been reported for the *B. cereus* biofilm-forming proteins TasA-bc and CalY (in these conditions) in their monomeric form and therefore the above characterization suggests drastic differences between *B. cereus*, and *B. subtilis*, whereby both TasA-bs and TapA show a structured fold and a only partially disordered one, respectively (317, 325).

3.6.3. *B. subtilis* and *B. cereus* TasA functional amyloids show various macroscopic morphologies

TasA filaments, produced by heterologous expression in *E. coli* and assembled *in vitro* (at 0.25 mM protein concentration, pH 7.4), were analyzed using negative-stain transmission electron microscopy (TEM) (**Figure 4A-B**). Both TasA-bs and TasA-bc assemblies reveal macroscopic filamentous morphologies following polymerization over one week at room temperature. TasA-bs filaments show a straight, uniform fibrillar morphology (**Figure 4A**). The protofilaments (*i.e.* the smallest fibrillar entity of the quaternary architecture) arrange in a supramolecular bundle-like organization, made of laterally associated stacked proto-filaments. The diameter of the protofilaments was ~5 nm, consistent with previous observations of TasA filaments in biofilm culture (314), and their length was variable, up to several microns. Interestingly, a symmetrical arrangement of uniform repeats, perpendicular to the primary axis, was observed along the filament. The axial translation per repeat was ~5 nm. Conversely, electron micrographs of TasA-bc filaments revealed filamentous nano-objects, but with less structural organization compared to *B. subtilis* filaments (**Figure 4B**).

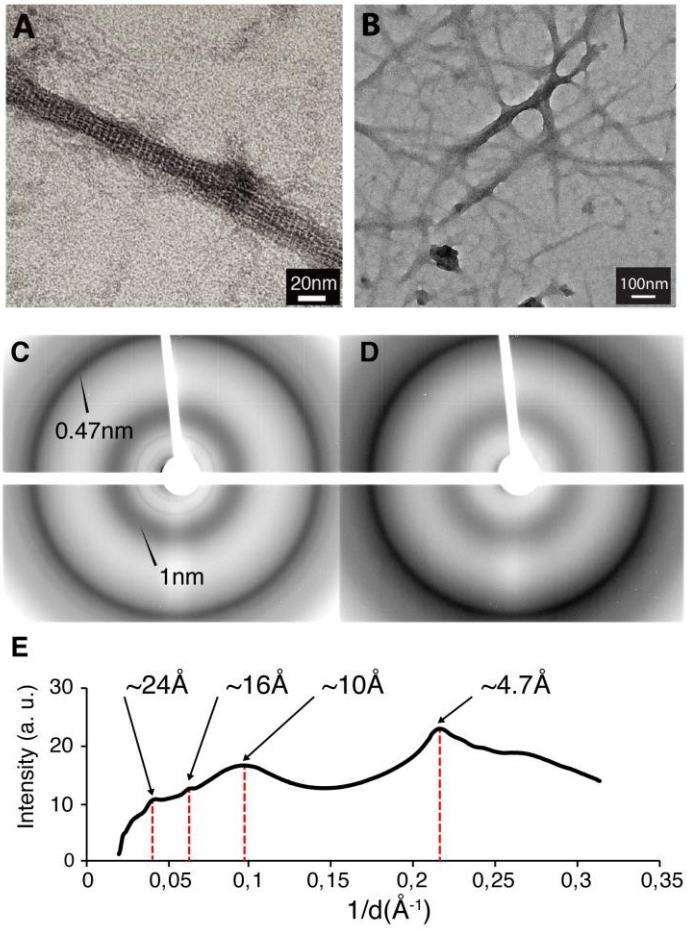


Figure 3-4: Transmission electron micrographs of negatively stained *B. subtilis* TasA (A) and *B. cereus* TasA amyloids (B). The average TasA-bs filament width was ~5 nm. X-ray diffraction patterns of TasA-bc functional amyloids (C) (reflections at ~4.7 Å and ~10 Å are highlighted) and TasA-bs functional amyloids (D). (E) Diffraction intensity plot of TasA-bc filaments, diffraction signals at 24, 16, 10 and 4.7 Å are highlighted.

3.6.4. *B. subtilis* and *B. cereus* TasA functional amyloids display a cross-β architecture

One structural hallmark of amyloid protein assemblies is their cross-β arrangement in filaments, a feature reflected in the observation of typical X-ray diffraction patterns, with the signature of inter-strand and inter-sheet repetitive distances at 4.7 Å and 10 Å, respectively (326). We have previously demonstrated that filaments of TasA purified from *B. subtilis* can bind thioflavin T and Congo Red dyes (314), suggesting TasA assembles into functional amyloid filaments, however detailed structural evidence of the amyloid architecture was missing. X-ray diffraction patterns were recently reported for recombinant TasA-bs fibrils by two independent laboratories (317, 318), controversially showing different diffraction rings, encoding (317) or not (318) for a canonical cross-β arrangement. Here, we performed X-ray diffraction experiments on TasA-bs (**Figure 4C**) and TasA-bc **(Figure 4D**) fibrils to

characterize the amyloid nature of our sample preparations. Clear diffraction signals at 4.7 and 10 Å demonstrate the presence of a cross-β structure within TasA filaments from both bacterial species. These data demonstrate that both bacterial filaments exhibit the structural hallmark of amyloid filaments, possessing the canonical cross-β structure, and support the amyloid-like nature of the TasA proteins in their fibrillar forms. Interestingly, additional X-ray diffraction reflections at ~24 and ~16 Å are observed in the *B. cereus* sample (**Figure 4E**) (not visible for TasA-bs) (background diffraction signal shown in **Supplemental Figure 2D-E**). Diffractions at similar values have been reported by Erskine *et al.* (318) (at 21 and 14.8 Å), however the canonical cross-β reflections at 4.7 and 10 Å were not observed in their diffraction data.

3.6.5. *B. subtilis* and *B. cereus* TasA functional amyloid filaments reveal different assembly kinetics

An intrinsic feature of amyloid proteins is their transition from the monomeric to the fibrillar state, concomitant with enrichment in β-sheets which confer to amyloid proteins the ability to bind and react, in a very well-defined way to specific dyes such as Congo red or thioflavin T. We previously demonstrated that acid-disaggregated TasA solutions purified from *B. subtilis* could re-polymerize into fibers following a typical amyloid polymerization curve using the ThT fluorescence assay (314). To reproduce this finding, we took advantage of the fact that freshly prepared TasA homogenates from *E. coli* could be maintained, after the purification process, mostly as monomers in acidic conditions. After switching to physiological pH, the ThT fluorescence curves for TasA-bs and TasA-bc at different concentrations (20, 40 and 50 μM) demonstrate that both proteins specifically bind to ThT. The fluorescent signal at the amyloid-specific 480-484 nm wavelength delineated polymerization curves similar to those of other amyloids proteins (**Figure 5**) (314). Several pieces of evidence, however, suggested that the polymerization of TasA-bc was more efficient than that of TasA-bs (**Figure 5A**): i) the emission of ThT fluorescence during TasA-bc polymerization was rapid and without a lag-phase, regardless of the concentration used in our studies, ii) the slope of the exponential phase of polymerization was more pronounced in TasA-bc and iii) the intensity of the fluorescent signal saturated earlier in the presence of TasA-bc (after 5 hours of incubation) compared to the 15 hours required for TasA-bs. Additionally, both TasA-bs and TasA-bc exhibited a linear dependence between the monomer concentration and the half-time of completion values (**Figure 5B**).

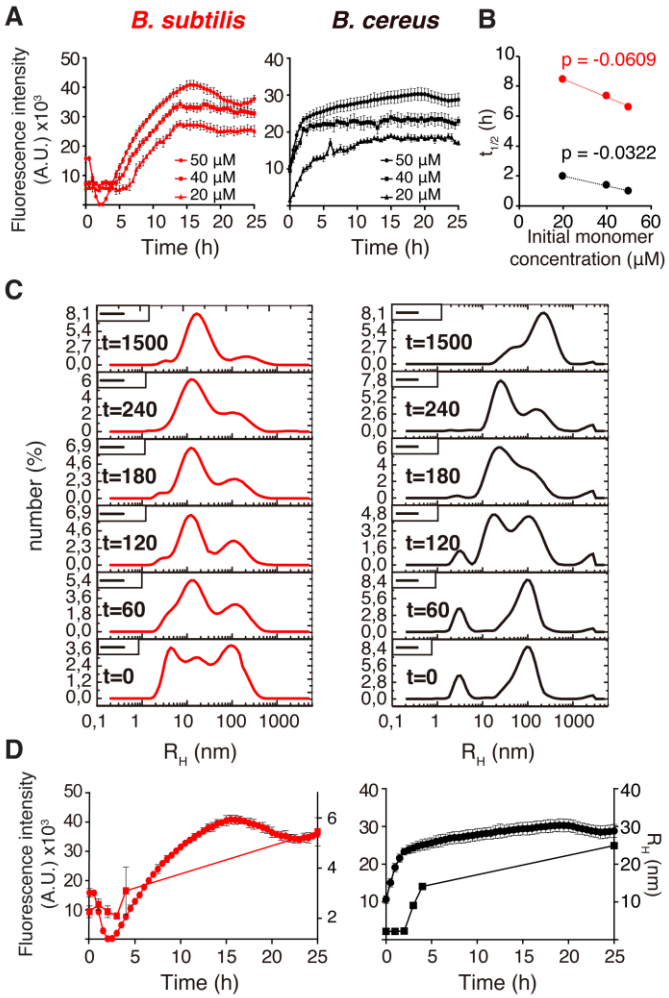


Figure 3-5: Aggregation kinetics of *B. subtilis* and *B. cereus* TasA proteins as followed using ThT fluorescence measurements (A), a plot of half-time values of ThT fluorescence against the initial monomer concentration of TasA-bs (red) or TasA-bc (black) (slope SD: ± 0.003) (B), DLS experiments (time values are minutes) (C), and (D) ThT fluorescence (circles) and DLS (squares) dynamics of 50µM solution of TasA incubated at RT. Similar kinetics profiles can be observed by both techniques for each protein: Left, *B. subtilis*, and right *B. cereus.*

The evolution of monomers to insoluble fibers implies the formation of intermediate aggregates which may function as nucleators that enhance polymerization. The simplest interpretation of the immediate initiation of polymerization for TasA-bc, as compared to TasA-bs, was a faster and more stable transition of monomers to aggregates or stable nuclei prone to polymerization. To test this hypothesis, we characterized the progression and composition of both protein solutions, monomer and aggregates and their sizes, over the first 24 hours by dynamic light scattering (DLS) assays. We started with monomeric protein in solution (50 µM, 1% acetic acid, pH 3.0) and after adjusting the solution pH to 7.0 with 10% NaOH. To achieve a better view of the oligomers size distribution, we looked at the signal intensity for each hydrodynamic

radius over time (**Figure 5C and Supplemental Figure 2F**). The most representative population of particles of TasA-bs reached 10 nm of mean hydrodynamic radius after 24 hours of incubation, while the highest scattering intensity of TasA-bc was associated to a particle population one order of magnitude larger (**Figure 5C**). Interestingly, the evolution of the average hydrodynamic radius R_H (**Figure 5D**, *squares*) over time exhibits a comparable profile when compared to the fluorescence with ThT measurements (**Figure D**, *circles*). Previous work done with TasA purified from *B. subtilis*, established the size of the purified protein in monomeric form at R_H of approximated 2.3 nm (327). In harmony with these findings, the data collected in initial stages of TasA homogenates, expressed in *E. coli*, rendered a mean hydrodynamic radius around 2 nm. Remarkable differences were however observable in further evolution of TasA-bs or TasA-bc. Focusing on the earlier time points, TasA-bc displayed a higher increase after 2 hours of incubation while hydrodynamic radius of TasA-bs remained almost invariable along the whole experiment in a similar way that can be observed in the corresponding ThT fluorescence measurement. Even that slight increase might be noted after 24 hours of incubation, the final average size of TasA-bs particles was 5 nm, notably smaller than the 25 nm particles characterizing TasA-bc.

3.6.6. *B. subtilis* and *B. cereus* TasA functional amyloids share the same structural fold but different local structural polymorphism

SSNMR is an exquisite tool to monitor the local structural order and observe the presence of dynamic or static disorder and polymorphism in fibrillar biological samples (1, 3, 48, 328, 329). We overexpressed TasA-bs and TasA-bc proteins in *E. coli* using isotopically ^{13}C-^{15}N-labeled sources to obtain uniformly $^{13}C/^{15}N$-labeled TasA filaments. We carried out SSNMR experiments using through-space (*i.e.* cross-polarization) and through-bond polarization transfers (*i.e.* INEPT) to respectively detect rigid and mobile protein moieties in the amyloid filamentous state. 1D ^{13}C-detected SSNMR experiment performed at 278 K (**Supplemental Figure 3A-B**) on *B. subtilis* TasA filaments reveal well-resolved NMR resonances, indicating a high local structural order of TasA protein subunit within the filaments. Temperature-dependent experiments performed at room temperature (298 K) and in the frozen state (263 K) show a similar fingerprint (**Supplemental Figure 3A-B**), indicating that TasA subunits maintain the same conformation in the filaments at various temperatures, although a significant broadening is observed at -10°C (263 K) reflecting the presence of multiple local conformations frozen out at low temperatures, as already observed by SSNMR for solid

proteins (330), including amyloid fibrils (331). In addition to the previous through-space experiments, through-bond mixing was used to detect mobile protein moieties at different temperatures (**Supplemental Figure 3A-B**). We observed a gradual increase of the signal with the temperature, indicating increasing mobility of TasA with increasing temperature.

We employed multidimensional ^{1}H-^{13}C experiments for TasA-bs and TasA-bc using through-bond mixing (**Supplemental Figure 3C-D**) to assign SSNMR correlations for mobile residues. We observed correlations for all amino-acid types in a random-coil conformation, clearly indicating the presence of non-assembled TasA molecules within the ultra-centrifugation pellet sample. The observation of monomeric TasA in the filament sample is consistent with previous observations on other amyloid fibrils such as Tau pair-helical filaments (332) or α-synuclein fibrils (333). Together with our observation that TasA-bs and TasA-bc form fibrillar objects as seen by EM analysis, these SSNMR data indicate that TasA-bs and TasA-bc filaments are formed by well-ordered TasA subunits, while a part of the protein subunits stays in a non-assembled (mobile, unfolded) conformation. Such observations could suggest the assembly process attains an equilibrium state between the monomeric and the fibrillar (*i.e.* assembled) TasA state.

To further investigate the structure of the rigid amyloid core in TasA filaments of both *B. subtilis* and *B. cereus*, we employ multidimensional ^{13}C-^{13}C SSNMR with through-space mixing on uniformly ^{13}C-labeled TasA-bs and TasA-bc samples (**Figure 6A**). A preliminary analysis of such data already allows for a global comparison of the structures adopted by TasA-bs and TasA-bc filaments, concerning two different aspects. (i) We observe that the spectral resolution (*i.e.* ^{13}C line width) displayed by the two TasA filaments is strikingly different under the same experimental conditions. TasA-bc SSNMR data indeed show strong peak broadening (full-width at half-height FWHH ~100-200 Hz), implying a large conformational variability at the atomic level (*i.e.* structural polymorphism). In contrast, TasA-bs filaments exhibit narrower ^{13}C line widths (FWHH ~ 30-100 Hz), thus accounting for highly ordered local arrangements of TasA subunits within the filaments. Interestingly although TasA-bs in its mature form comprises 233 amino acids, only ~70 NMR spin systems (*i.e.* residues) were detected in the rigid core using through-space mixing, suggesting the amyloid rigid core contains only part of the TasA sequence (**Supplemental Table S1**). However, the spectrum of TasA-bc amyloid fibrils only displays ~40 residues. More amino acids may nonetheless belong to the rigid core, as the spectral resolution does not permit a more accurate inventory. (ii) SSNMR chemical shift values are a valuable tool to probe the local molecular conformation (secondary structure

and structural environment). When superimposed, the two TasA filaments feature similar ^{13}C-^{13}C spectral fingerprints despite the difference in spectral resolution. Such remarkable spectral superimposition strongly indicates a high structural similarity of TasA in the rigid core of TasA-bs and TasA-bc subunits, even though the different spectral resolution points to noticeable differences in their structural homogeneity. Nevertheless, such data imply a very similar 3D structural fold, whereas the two TasA filaments seem to rely on different physical properties during assembly reflected by their various extent of polymorphism.

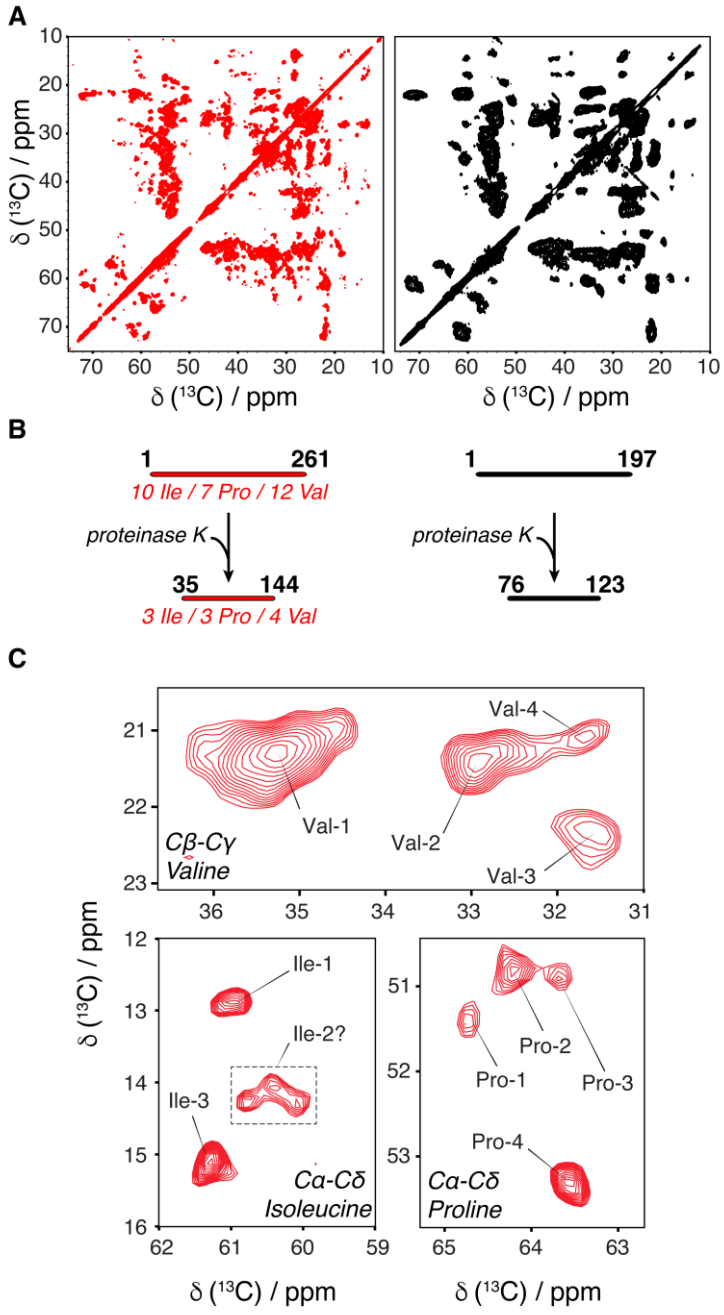


Figure 3-6: SSNMR analysis of *B. subtilis* and *B. cereus* TasA functional amyloids. (A) 2D ^{13}C-^{13}C SSNMR experiments of TasA-bs (red), and TasA-bc (black) filaments. (B) Proteinase K digestion resistant segments (TasA-bs and TasA-bc) displaying Ile, Pro, and Val contents of full-length TasA-bs (TasA-bs$_{1\text{-}261}$) and proteinase-digested TasA-bs filaments (TasA-bs$_{35\text{-}144}$), enabling the identification of resistant fragments K35-K144 for TasA-bs and K76-K123 for TasA-bc. (C) Excerpts of the 2D ^{13}C-^{13}C experiment of TasA-bs fibrils showing chemical shift ranges covering the different Ile Cδ_1-Cα, Pro Cδ-Cα, and Val C$\gamma_{1\text{-}2}$-Cβ cross-peaks.

In line with SSNMR results, the use of proteinase K digestion led to the identification of a 110-amino-acid long resistant segment (K35 to K144) for TasA-bs and 47-amino-acid long for TasA-bc (K76 to K123). Upon digestion steps with varying incubation duration, the remaining proteinase-resistant TasA-bs and TasA-bc segments were observed and identified using SDS-PAGE gel analysis (data not shown). The 2D ^{13}C-^{13}C SSNMR intra-residual correlation experiments as the number of identified spin systems per residue type (*i.e.* isoleucine, proline, valine) match the putative amyloid core amino-acid content rather well (**Figure 6B-C**). The TasA moiety identified by proteinase K digestion indeed encompasses 3 isoleucines, 3 prolines, and 4 valines (with 10, 7, and 12, respectively, in the full-length TasA-bs) (**Figure 6B – Supplemental Table S1**), and interestingly 2D ^{13}C-^{13}C PDSD of TasA-bs reveals 4 cross-peaks in the Cβ-Cγ spectral region of valines, 3-4 (depending on potential polymorphism) cross-peaks in the Cα-Cδ region of isoleucines, and 4 cross-peaks in the Cα-Cδ region of prolines (**Figure 6C**). TasA-bc however displays a rather different behavior when put through proteinase K digestion. Indeed, the resistant segment is a ~50-residue long moiety which correlates quite well with SSNMR observations. The greater polymorphism shown for TasA-bc prevents us from accurately and relevantly comparing the amino-acid contents from SSNMR and proteinase K digestion experiments. Interestingly the corresponding digestion resistant segments are found within the same primary sequence region in a sequence alignment display (shown in bold characters in **Figure 1**).

3.6.7. Comparison of secondary structure propensity of TasA functional amyloids between *B. subtilis* and *B. cereus*

Cα and Cβ SSNMR chemical shift values provide reliable data for establishing the secondary structure of TasA in the amyloid state. We engaged in a conformational analysis of the secondary structure content in TasA-bs and TasA-bc filaments (**Figure 7A**). For purposes of accuracy, we mainly focused on the amino-acids alanine (A), serine (S), and threonine (T), as their $^{13}C\alpha$ -$^{13}C\beta$ correlation signals appear in isolated regions and could be identified

unambiguously with respect to other ^{13}C signal contributions. The mature *B. subtilis* TasA construct (233 residues in total) comprises 19 alanine (Ala), 15 serine (Ser), and 14 threonine (Thr) residues, while mature *B. cereus* TasA (177 residues in total) encompasses 8 Ala, 8 Ser, and 13 Thr. For clarity purposes, **Figure 7A** only shows the theoretical spectral regions encoding for alanine residues, based on the probability analysis of NMR chemical shifts reported by Jardetzky (321) (i.e. ~48−56 ppm for $^{13}C\alpha$ and ~16-25 ppm for $^{13}C\beta$). Through the analysis of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts, NMR spectroscopy indeed allows one to determine with high accuracy the secondary structural arrangement displayed by any given residue. In other words, depending on the peak position within the spectrum, the secondary structure can be easily inferred. The theoretical alanine-α/β secondary structure distributions are highlighted in light red and light blue, respectively. We observed that *B. subtilis* TasA amyloids display a pronounced α-helical propensity while TasA-bc amyloids show very few peaks in the spectral regions reflecting α-helical conformation (**Figure 7A**). This indicates that the extent of α-helical conformation is substantially higher in TasA-bs filaments. Evaluating peak volume ratios for each residue type, we pursued a more rigorous assessment of secondary structure content. Based on secondary chemical shift analysis (291), probability-based secondary structure identification has been performed to compare TasA-bs and TasA-bc filaments. Conformation propensity was assessed using secondary chemical shift ($\Delta\delta C_{sec} = \delta C\alpha - \delta C\beta$, with $\delta Cx = Cx_{(experimental)} - Cx_{(random\ coil)}$). Negative values are considered to be highly indicative of a β-strand conformation. $\Delta\delta C_{sec}$ values ranging from -1 to 1 were hereafter regarded as random coil-like arrangements, as using a less strict segregation would only result in an under-estimation of residues arranged in a random coil manner. For each studied residue type (*i.e.* Ala, Ser and Thr), the SSNMR peak volumes were used for estimating the percentage of the secondary structure content. Both TasA-bs and TasA-bc show a considerable tendency to fold in a β-strand-like manner, as expected considering the typical amyloid X-ray diffraction patterns observed in our study. On average, ~60% of examined Ala, Thr and Ser residues in TasA-bs and TasA-bc are involved in a β-sheet conformation (**Figure 7B**). Both bacterial strains however drastically diverge regarding α-helix propensity. It is clear that few to no residues in TasA-bc display clear α-helix conformation within the theoretical α/β peak distribution, in comparison to TasA-bs (**Figure 7B**). On average, ~20% of examined residues in *B. subtilis* TasA are involved in a α-helical conformation. Interestingly, amyloid proteins in their fibrillar state are known to contain very high β-sheet content (334), and are very often uniquely composed of this type of secondary structure. However, several examples of amyloid

fibrils containing a marked amount of α-helical secondary structure have been reported, *e.g.* the prion HET-s in its full-length form (55) or full-length Ure2p (335), both comprising a β-sheet-rich amyloid core and containing folded or molten globular α-helical moieties. The remarkable presence of α-helical conformation within *B. subtilis* TasA encoded in SSNMR chemical shifts is a proof of its structural singularity. Interestingly, we compared our observations with the recently reported crystal structure of TasA-bs (317) (**Figure 7C**), which shows a jellyroll fold comprising two antiparallel β-sheets (~30% of secondary structure propensity considering Thr, Ser and Ala residues) surrounded with six short α-helices (~25%) and several loops (~45%) (**Figure 7C**). Using SPARTA (336), Cα and Cβ chemical shifts were predicted from the TasA-bs crystal structure and the 2D ^{13}C-^{13}C patterns correlate relatively well (**Supplemental Figure 4**). Acknowledging the limitations of such prediction methods and the very different nature of samples (crystalline monomeric protein vs. assembled insoluble proteins), the experimental data and the predicted chemical shifts can hardly be expected to be completely alike; however, the general peak fingerprint seems to be conserved, and all observations considered, one might speculate on a partial conservation of part of the globular structure within the amyloid fold.

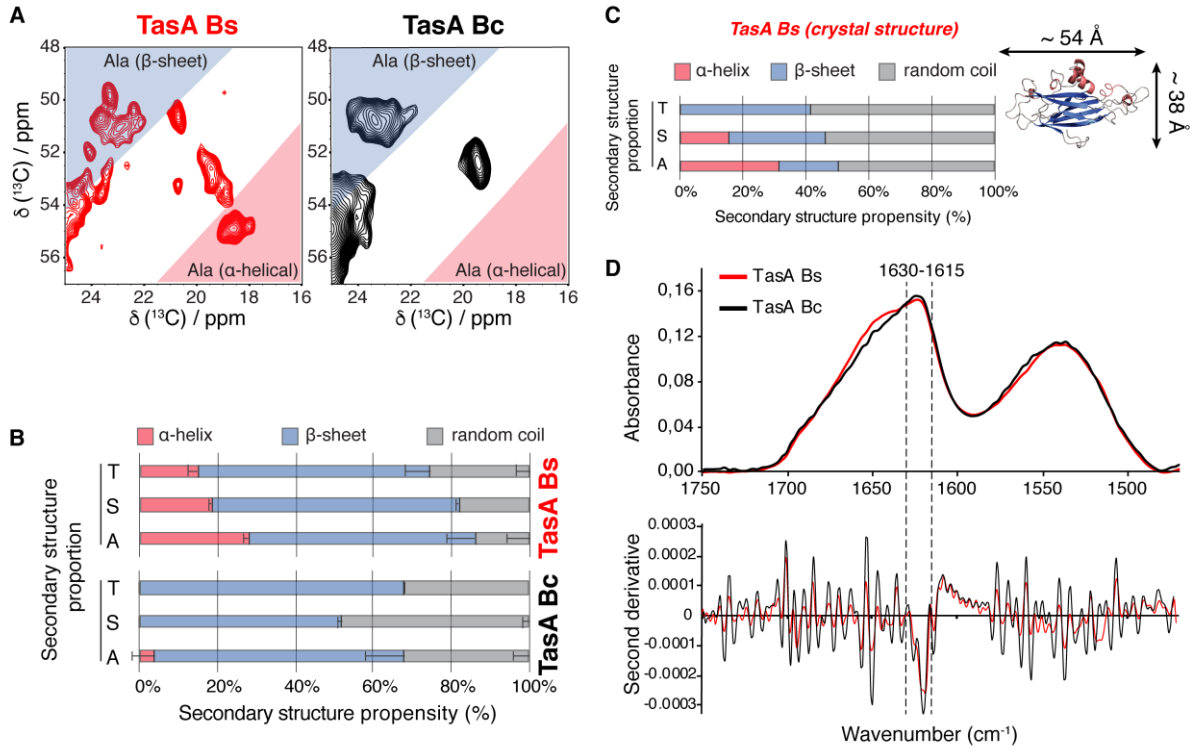


Figure 3-7: TasA functional amyloid filaments from *B. subtilis* and *B. cereus* display different secondary structure composition. (A) Excerpts of the 2D ^{13}C-^{13}C experiments are shown for TasA-bs in red (left panel) and for TasA-bc in black (right panel); for a chemical shift range covering the different alanine (Ala) conformations. Typical

expected areas for α-helix and β-sheet conformations are highlighted in light red and light blue respectively. (B) Secondary structure propensity of Ala, Ser, and Thr as assessed by analyzing secondary chemical shift for TasA-bs (upper panel) and TasA-bc (lower panel) ^{13}C-^{13}C correlations and using peak volume ratios for quantitative assessment. α-helix and β-strand conformations are highlighted in light red and light blue respectively. (C) Secondary structure propensity of Ala, Ser, and Thr as assessed by analyzing the monomeric crystal structure of TasA-bs (317). α-helix and β-strand conformations are highlighted in light red and light blue respectively. (D, Top) ATR-FTIR spectra of TasA-bs (red) and TasA-bc (black) filaments (natural abundance) in the amide I and II range and (D, bottom) second derivative of the amide I band to better resolve the absorbance components in the whole spectra. The *dashed lines* delineate the position of typical β-sheet contribution.

Then, we conducted ATR-FTIR analysis to corroborate the secondary structure elements of TasA-bs and TasA-bc amyloids. Both TasA filaments have an absorption spectrum with strong amide I and amide II bands at ~1620 cm^{-1} and 1535 cm^{-1}, respectively (**Figure 7D top**). Although the two FTIR spectra display a similar absorption pattern centered around amide I band that is better resolved on the second derivative spectra (Fig. 7D bottom) encoding for β-sheet conformation in amyloid fibrils (1611-1630 cm^{-1}) (337), substantial differences are observed at intensity ~1640-1660 cm^{-1} suggesting that the two filaments do not have the same secondary structure content. We used a deconvolution approach to extract the percentage of secondary structure elements in TasA-bs and TasA-bc, namely parallel β-sheet (1634 cm^{-1}), anti-parallel β-sheet (1620 and 1690 cm^{-1}), α-helix (1658 cm^{-1}), turns (1675 cm^{-1}) and random coil (1646 cm^{-1}), according to well-known assignments (338) (**Table 1**). We observed that the total β-sheet content (including parallel and anti-parallel β-sheets) is conserved between the two TasA filaments, with 56% and 60% for TasA-bs and TasA-bc amyloids respectively. In line with the SSNMR secondary chemical shift analysis, the α-helical contribution is significantly higher for TasA-bs compared to TasA-bc, with 21% and 13%, respectively. The SSNMR and the ATR-FTIR analysis show comparable results for relative α-helical content in the two filaments, with TasA-bs exhibiting ~ 1.5 to 2 times more α-helical content compared to TasA-bc; however, we note that the absolute contributions differ. This effect can be attributed to the fact that dipolar-based polarization transfer in SSNMR (*i.e.* in the PDSD experiment) selectively probes the rigid protein part of the filaments, while ATR-FTIR analysis probes the overall protein contribution. Therefore ATR-FTIR reveals the secondary structure composition of both rigid and mobile moieties, as well as of monomeric TasA proteins in the sample. Both mobile moieties and non-assembled monomeric TasA would only be observable in the J-based experiment (*i.e.* the INEPT experiment) by SSNMR and were therefore not included in the secondary chemical shift analysis.

Secondary structure element	Wavenumbers (cm^{-1})	Percentage of structural elements			
		TasA-bs	TasA-bc	TapA	CalY
β-sheet anti //	1620 1690	37 %	41 %	20 %	23 %
β-sheet //	1634	19 %	19 %	28 %	33 %
Random coil	1646	14 %	18 %	19%	10%
α-Helix	1658	21 %	13 %	24 %	24 %
Turn	1675	9 %	9 %	10 %	10 %

Table 3-1: Secondary structure propensity of TasA-bs, TasA-bc, TapA and CalY amyloids probed by ATR-FTIR analysis.

3.6.8. Accessory proteins TapA and CalY self-assemble and adopt the same structural fold as observed for TasA filaments

B. subtilis TapA is an accessory protein present at a molar ratio of ~1/100 in extracellular biofilms and is considered to be a minor biofilm protein. In *B. cereus* its homologue is called CalY and is present at a molar ratio of ~1/1 (Romero et al., unpublished). Functional aggregation of TasA in *B. subtilis* requires the *tapA-sipW-tasA* operon (313), and a putative role in cell surface anchoring during TasA aggregation has been proposed for TapA (314, 315). TasA-bs and TapA share 41% of sequence identity (**Figure 1**), and *tapA* null mutant shows unstable TasA secretion and polymerization (315), this effect being linked to the N-terminal half of TapA (315). The precise role of TapA and its molecular impact on TasA assembly is still unclear and several functions of TapA have been proposed (315, 316), namely a TapA-dependent TasA assembly during biofilm formation and the formation of TapA multimers involved in TasA assembly. However, it remains unclear if TasA and TapA co-assemble in the context of *in vivo* biofilm assembly (316). To gain insight into the molecular structure of TapA (and its counterpart in *B. cereus*, CalY), we employed SSNMR and ATR-FTIR to investigate *in vitro* TapA and CalY assemblies. We observed that recombinant TapA and CalY proteins self-assemble at pH 7.4 under the same conditions used for TasA-bs and TasA-bc aggregation. Upon assembly, TapA and CalY exhibit a previously unknown ability to form aggregates in the absence of TasA, as demonstrated by the observation of (i) small supramolecular entities using transmission electron microscopy (**Figure 8A-B**), and (ii) the typical cross-β diffraction pattern using X-ray diffraction (**Figure 8C-D**). The resulting TapA and CalY macroscopic assemblies do not exhibit an unbranched fibrillar shape as observed for TasA-bs and TasA-bc, but rather an amorphous aggregated morphology, with a diameter varying between ~ 5 to 7

nm, which is comparable to the diameter measured for TasA-bs and TasA-bc filaments. The morphology of TapA and CalY assemblies is comparable to small amorphous amyloid aggregates that have been observed for several amyloid proteins depending the experimental conditions during the *in vitro* aggregation process (46, 339). The X-ray diffraction analysis (**Figure 8C-D**) reveals reflections that are characteristic of cross-β arrangement with strong signals at 4.7 and 10 Å, implying the amyloid nature of TapA and CalY assemblies.

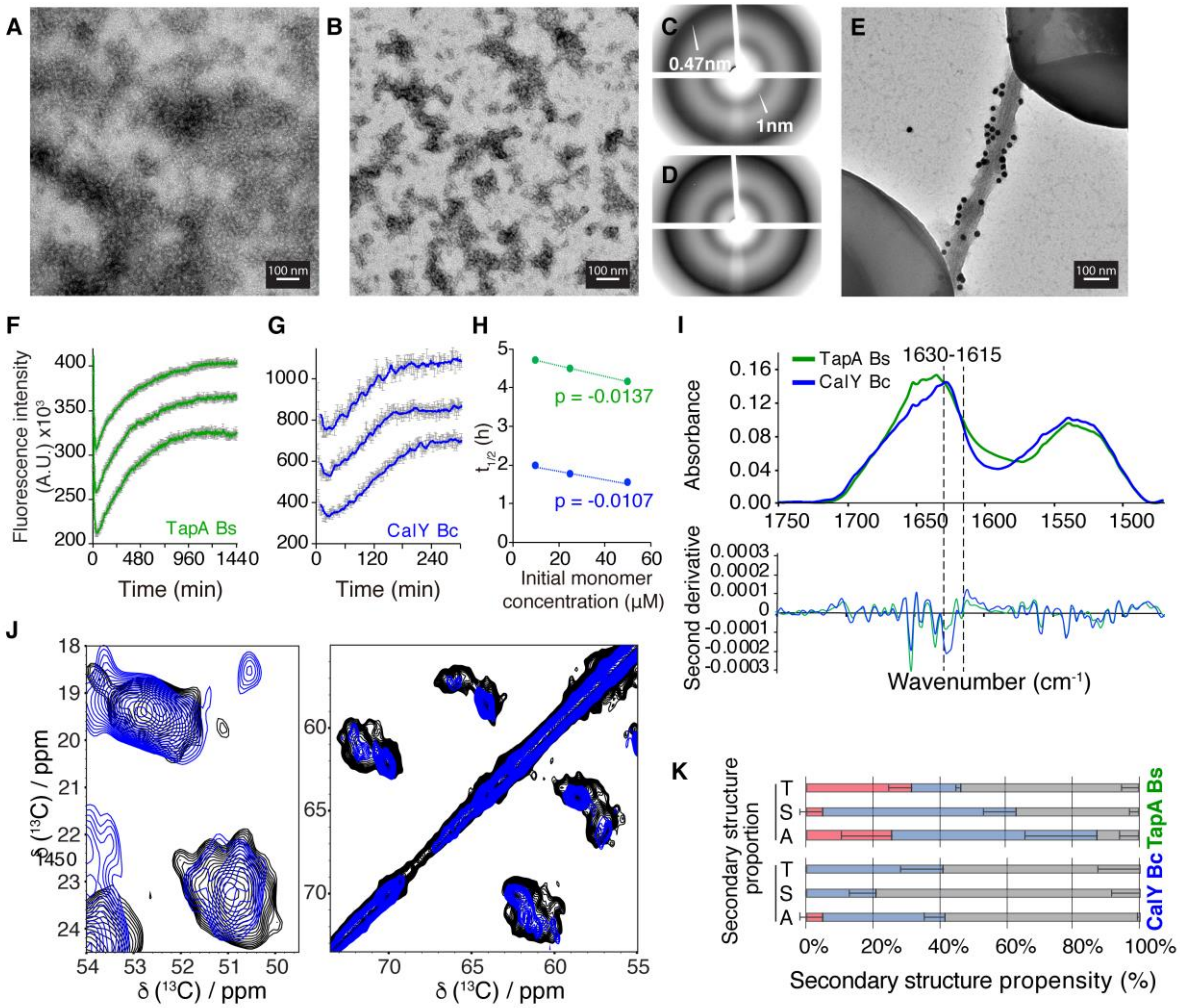


Figure 3-8: *B. subtilis* TapA and *B. cereus* CalY self-assemble to into amyloids. (A-B) Negative-staining micrographs of TapA (A) and CalY (B) assemblies. (C-D) X-ray diffraction patterns of TapA (C) and CalY (D) assemblies. (E) Transmission electron micrograph of *eps tasA* double mutants immunolabelled with anti-TapA antibodies. (F-G) Aggregation kinetics of *B. subtilis* TapA (green) and *B. cereus* CalY (blue) as followed using ThT fluorescence. (H) Plot of half-time values of ThT fluorescence against the initial monomer concentration of TapA (green) or CalY (blue) (slope SD: ± 7.10^{-5} for TapA and ± 2.10^{-4} for CalY). (I) ATR-FTIR spectra of TapA and CalY assemblies (top) and second derivative (bottom); the *dashed lines* delineate the position of typical β-sheet contribution. (J) Overlay of 2D SSNMR ^{13}C-^{13}C spectra of CalY (blue) and TasA-bc fibrils, the alanine Cβ-

Cα (left) and threonine and serine Cα-Cβ (right) regions are showed. (K) Secondary structure conformational distribution in TapA and CalY, as analyzed using SSNMR.

We investigated the putative formation of TapA fibrils *in vivo* in *B. subtilis* in a Δ*tasA* Δ*eps* strain. This strain shows induced expression of extracellular matrix genes while lacking both the EPS and TasA, which are the main components of the extracellular matrix. These chemical features make this strain a suitable candidate to study TapA in biofilm-inducing conditions in MSgg medium. We studied TapA localization in 72-hour mature biofilms by TEM and immunogold-conjugated secondary antibodies and found large concentrations of TapA along the edge of fibrillar assemblies that are not composed of TasA (given the Δ*tasA* background) (**Figure 8E**) suggesting that indeed, in agreement to our results obtained *in vitro*, TapA is able to polymerize into fibril-shaped aggregates in the absence of TasA. Likewise, we performed double labelling experiments with secondary antibodies conjugated to gold particles of different sizes using TasA and CalY primary antibodies in mature biofilms from *B. cereus* and found signal colocalization corresponding to both proteins in fibrillar structures (**Supplemental Figure 5**).

The aggregation kinetics of TapA and CalY were studied by means of ThT fluorescence. Interestingly, TapA and CalY show rather similar assembly kinetics behavior, as they both display clear concentration dependence regarding the half-time of completion and similar slopes (slope: ~-0.01 min^{-1}) (**Figure 8H**). TapA show $t_{1/2}$ values around ~4-5 hours (**Figure 8F**) while CalY shows the half-completion after ~1-2 hours (**Figure 8G**). We then conducted ATR-FTIR to investigate the secondary structure of mature TapA and CalY assemblies (**Figure 8I**). Both protein assemblies have an absorption spectrum with a broad amide I band with a maximum at 1630 cm^{-1}, strongly indicating that TapA and CalY are, at the molecular level, organized into β-sheet-rich structure. Deconvolution analysis (**Table 1**) suggests that the total amount of β-sheet secondary structure in TapA and CalY is comparable to TasA-bs/bc (with ~50%), but the proportion of anti-parallel and parallel β-sheets is different. Interestingly, the α-helical content is comparable between TapA and CalY (with ~25%), suggesting a less pronounced secondary structure variability for the accessory proteins compared to their TasA counterparts.

In order to obtain more site-specific insights into the TapA and CalY assemblies, we recorded SSNMR ^{13}C-^{13}C spectra and observed a noticeable level of structural polymorphism reflected in broad lines. When superimposed, the overall 2D ^{13}C-^{13}C patterns observed for TapA and

CalY are comparable, suggesting a common architecture for the accessory proteins in their assembled form (**Supplemental Figure 6A**).

By examining the alanine and serine ^{13}C-^{13}C regions encoding Cβ-Cα correlations, we observed a rather striking resemblance between CalY and TasA-bc. **Figure 8J** illustrates this apparently similar structural fold of the two amyloid-like assemblies, as very similar Cα-Cβ chemical shift patterns for alanine, serine and threonine are observed. We used the same approach based on chemical shifts to derive the secondary structure propensity of TapA and CalY assemblies (**Figure 8K**). Overall, we observed a higher content in random coil conformation for CalY compared to TapA, in line with the ATR-FTIR analysis.

3.6.9. *In vitro* co-assembly between TasA and its accessory protein does not perturb the global filament architecture but catalyse its fibrillization

The similarity of the SSNMR fingerprints of TasA-bs, TasA-bc, TapA and CalY proteins in their assembled state indicates a similar 3D structure of their subunits. Therefore, the question arises if (i) the accessory proteins might influence the kinetics of TasA polymerization and (ii) if TasA and its accessory proteins could co-assemble during the biofilm formation. The first issue was successfully addressed in previous studies of the aggregation kinetics of a sample containing TasA-bs and TapA in different molar ratios. It was then demonstrated that the 3:1 ratio is the most effective to enhance TasA polymerization in *B. subtilis* (340). Here, we used the same procedure to investigate the effect of CalY on TasA-bc, by studying the aggregation kinetics of TasA-bc at different CalY concentrations. Interestingly, the same procedure using a mix of TasA-bc and CalY led to similar results in our experiments. A decrease in half-time of completion values was observed when adding CalY at different protein ratios (ratio CalY:TasA-bc) as compared to the half-time of completion displayed by TasA-only (~14 hours): 30:70 ($t_{1/2}$ ~10 hours), 50:50 ($t_{1/2}$ ~7.1 hours) and 70:30 ratio ($t_{1/2}$ ~4.3 hours) (ratio CalY:TasA-bc) as compared to the half-time of completion displayed by TasA-only (~14 hours) (**Figure 9A**). Such observations might suggest that CalY increases the assembly kinetics of TasA-bc, and thus act as a catalyzer.

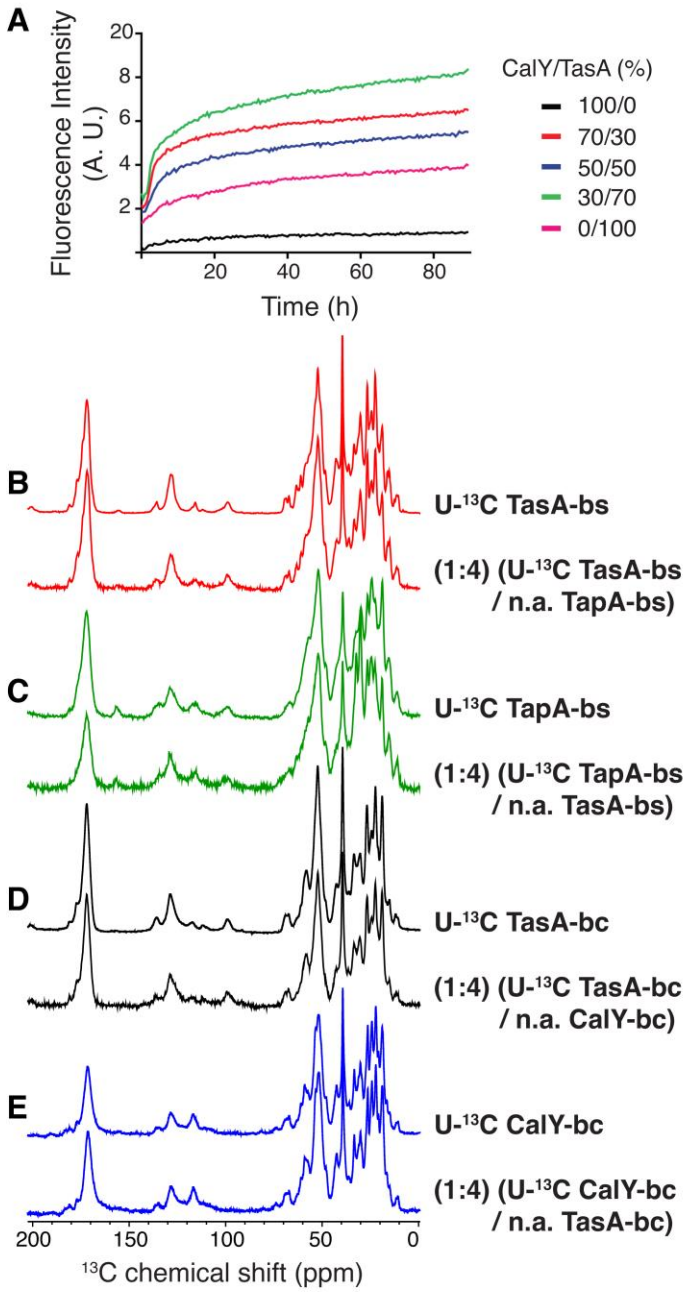


Figure 3-9: The effect of accessory proteins TapA and CalY during the assembly process of TasA. (A) Polymerization kinetics of different CalY:TasA-bc molar ratios assessed using Tht fluorescence emission over time. (B-E) 1D ^{13}C-detected experiment of co-assembled (B) (1:4) (U-^{13}C TasA-bs / n.a TapA), (C) (1:4) (U-^{13}C TapA-bs / n.a TasA-bs), (1:4) (D) (U-^{13}C TasA-bc / n.a CalY), (1:4) (E) (U-^{13}C CalY / n.a TasA-bc).

Next, we tested whether the TasA fold was altered in presence of the accessory protein during the assembly process. To address this question, we designed an SSNMR-based experiment that consists in the co-assembly of soluble TasA-bs and TapA (and TasA-bc with CalY) proteins prior to the assembly process. Only one protein was chosen to be isotopically labeled with ^{13}C

isotopes, while the other protein was at natural abundance. The co-assembled sample was then analyzed by ^{13}C-detected SSNMR to monitor chemical shift perturbation on the labeled proteins. We prepared four samples in total, consisting of ^{13}C-labelled TasA-bs, TasA-bc, TapA and CalY, co-assembled with their unlabeled partner (molar ratio of 1:4), ensuring the spectral observation of a single protein, and of possible structural differences using SSNMR (n.a.: natural abundance): (1:4) (U-^{13}C TasA-bs / n.a. TapA), (1:4) (U-^{13}C TapA-bs / n.a. TasA-bs) (**Figure 9B-C**), (1:4) (U-^{13}C TasA-bc / n.a. CalY), (1:4) (U-^{13}C CalY / n.a. TasA-bc) (**Figure 9D-E**). No drastic changes were observed in the co-assembled samples, when compared to a sample composed of a unique protein during assembly. This suggests that the structural arrangement of mature TasA-bs and TasA-bc self-assemblies are not altered by the presence of their corresponding accessory proteins (**Figure 9B and D**) during the polymerization process. A possible explanation could have been that the accessory proteins TapA and CalY might perfectly accommodate the TasA fold *in vitro* by intercalating between TasA subunits along the amyloid fibril axis, not modifying its structural arrangement. We however ruled out this hypothesis based on the observation of similar 1D ^{13}C signals between U-^{13}C TapA and (1:4) (U-^{13}C TapA-bs / n.a TasA-bs) (**Figure 9C**) (similar observation was made on CalY, **Figure 9E**). It indicates that the well-defined (*i.e.* higher ^{13}C-spectral resolution) TasA-bs fold could not be propagated to TapA subunits during polymerization (*i.e.* TapA when intercalated with TasA-bs would display the same ^{13}C-spectral resolution as TasA-bs alone).

3.7. Discussion

In this work, we describe the structural comparison of TasA functional amyloid fibrils from *B. subtilis* and *B. cereus*, alongside their accessory proteins, TapA and CalY, respectively. These results, together with previous studies (313-318) allow us to establish a mechanistic picture of TasA assembly in both *Bacillus* species (**Figure 10**).

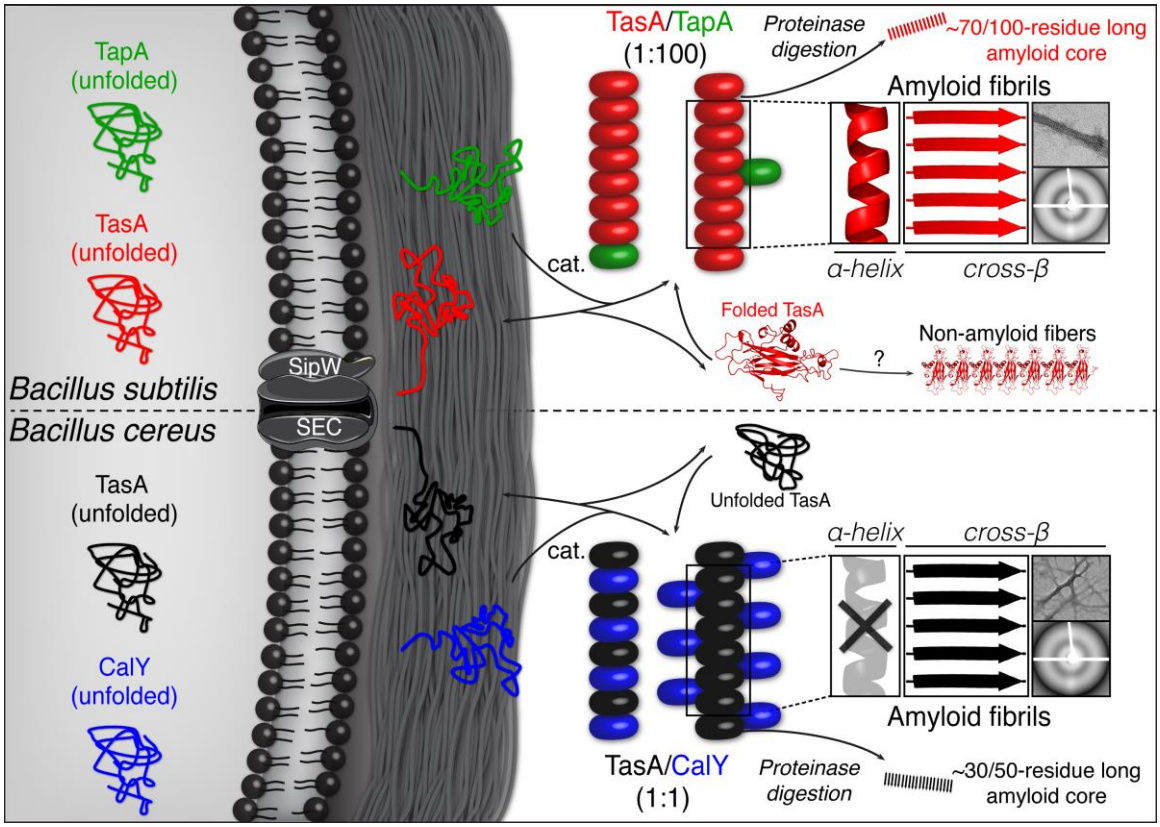


Figure 3-10: The current understanding of the mechanisms involved in TasA/CalY/TapA secretion and *Bacillus* biofilm formation in both *B. subtilis* and *B. cereus*.

Recent studies carried out by two independent groups revealed substantial differences regarding the assembly and the structural arrangement of *B. subtilis* TasA fibrils. Oschkinat and coworkers (317) reported the high-resolution crystal structure of *B. subtilis* TasA in its monomeric form, and demonstrated by solution and solid-state NMR a structural transition between a soluble, globular state to filaments with an unambiguous β-sheet SSNMR signature. Controversially, Erskine *et al.* reported that *B. subtilis* TasA may also assemble into a non-amyloid form (as seen by the lack of cross-β reflections in X-ray fiber diffraction experiments) and proposed a supramolecular arrangement made of globular TasA subunits (318). Our results incompatible with the findings of Erskine et al., support the observations made by Oschkinat and coworkers, and the model suggested by Kolter and coworkers (314, 315, 317). We demonstrate here that recombinant TasA-bs and TasA-bc form amyloid filaments upon assembly at pH 7. Both TasA fibrils, when analyzed with X-ray diffraction, display typical reflections at 4.7 Å and 10 Å highly indicative of a canonical cross-β amyloid fold. This suggests that TasA filaments from the two species indeed adopt a cross-β arrangement with the fibrillar structure, in agreement with previous observations of Oschkinat and coworkers (317).

Taken together, these data strongly suggest that both bacterial strains use the highly stable amyloid arrangement of TasA to ensure its biofilm integrity. Interestingly, both TasA fibrils display dissimilar structural polymorphism levels. Indeed, solid-state NMR linewidths indicate that TasA-bs fibrils exhibit a high structural order at the molecular level, while TasA-bc fibrils show a higher polymorphism.

Further SSNMR chemical shift analyses revealed other differences between the two biofilm-forming bacterial strains. TasA-bs in its fibrillar form displays an α-helical propensity alongside the β-sheet conformation within its rigid core. Such unique features are however not found in TasA-bc fibrils, suggesting the presence of additional structural elements in *B. subtilis* biofilm molecular architectures. Interestingly, 2D ^{13}C-^{13}C SSNMR fingerprints for TasA-bs and TasA-bc superimpose quite well, indicating their similar overall structural fold. The crystal structure (317) uncovered TasA globular monomeric form, which consists of a jellyroll fold comprising two antiparallel β-sheets surrounded with six short α-helices and several loops. Fascinating is the idea that TasA-bs fibrils might, upon assembly, conserve a part of TasA globular form while adjusting β-strands in a cross-β stacking manner within its amyloid core. Chemical shift values predicted from the published globular structure were compared to our experimental results reported here. We superposed predicted peaks upon SSNMR ^{13}C-^{13}C fingerprint spectra of TasA-bs filaments (**Supplemental Figure 4**). Such comparison shows a partial overlap, suggesting that a part of the TasA-bs globular fold may be conserved upon fibril formation, while the majority of the secondary structure content would shift to a β-sheet conformation.

Work from Chai *et al.* reported a partially disordered conformation for TapA in solution (325). In this work, we have characterized TasA-bc and CalY as disordered and unfolded in solution, presenting an additional variability between *B. cereus* and *B. subtilis* for which both TasA-bs and TapA show folded or partially folded monomers (317, 325). We investigated the accessory proteins associated with TasA biofilm-forming activities in *B. subtilis* and *B. cereus*, known as TapA and CalY, respectively. Both TapA and CalY have the ability to form supramolecular assemblies *in vitro* observed with negative-staining EM. The key observation that both accessory proteins can feature a typical cross-β X-ray diffraction pattern, puts into perspective the actual involvement of those auxiliary proteins within the biofilm (303, 314, 315). As the two mentioned TasA filaments, TapA and CalY display a comparable 2D ^{13}C-^{13}C fingerprint in solid-state NMR, and FTIR analysis shows comparable behavior with intense signal at 1630 cm^{-1}, as would be expected for β-rich contents. In addition to striking similarities between the

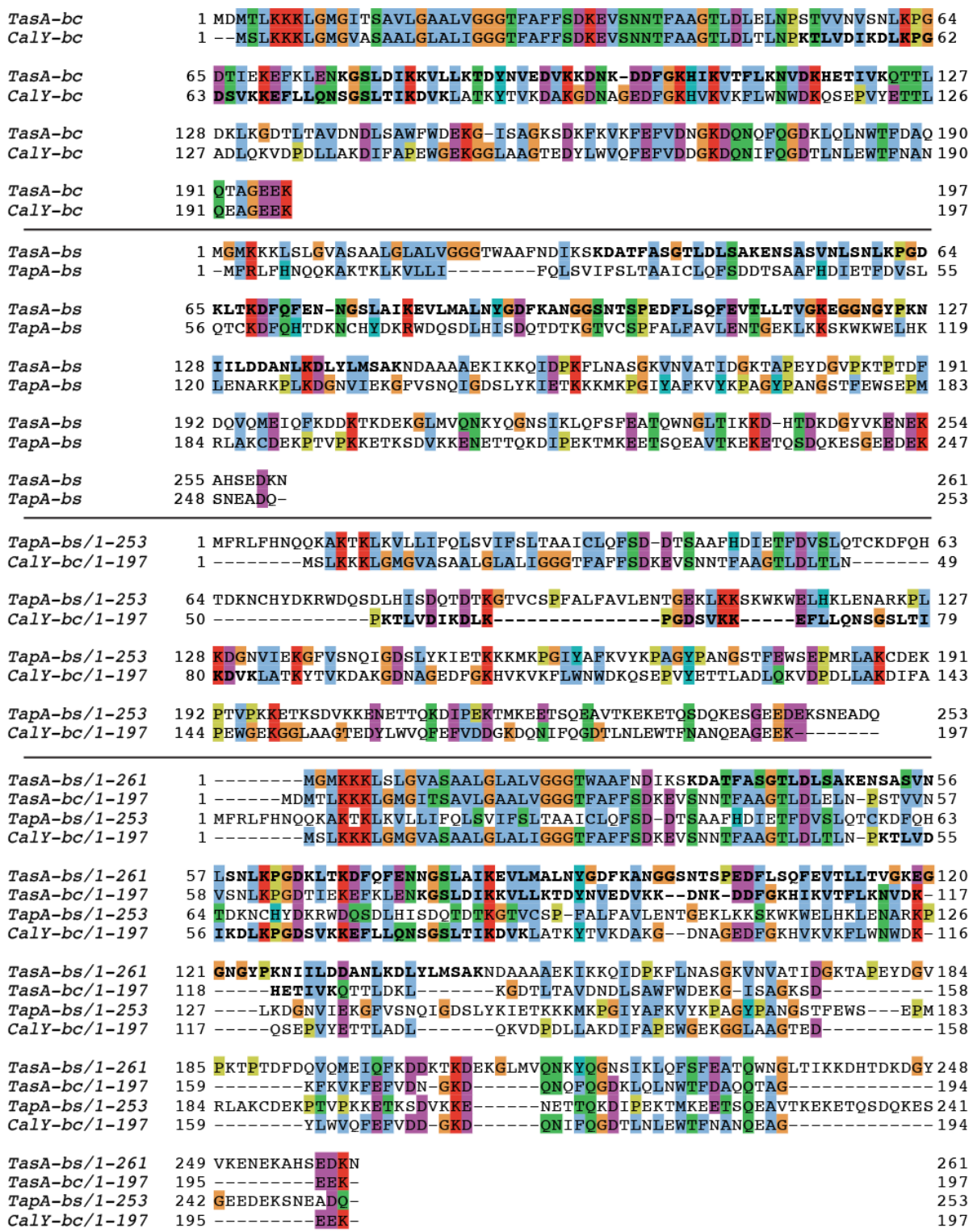
two accessory proteins, both also show comparable SSNMR 2D ^{13}C-^{13}C fingerprints to their respective TasA spectra (**Supplemental Figure 6B-C**), thus suggesting a shared structural fold.

Seeding experiments on all mentioned *Bacillus* proteins indicate that the presence of the accessory protein does not structurally affect TasA fibril formation *in vitro*, at least from a structural point of view. However, kinetics experiments strongly suggest a catalyzing effect of TapA on TasA fibril formation. Additionally, the observation of *in vivo* TapA filaments in Δ*tasA* mutants might be indicative that TapA seeds TasA filaments within *B. subtilis* biofilms and is actually present within the matrix (ratio TapA:TasA = 1:100 (315)), inducing the TasA amyloid fold. The observation of heterogeneous aggregates by electron microscopy and the presence of broad SSNMR line widths indicate that the accessory proteins TapA and CalY lack the ability to self-assemble into homogeneous amyloid fibrils on their own and therefore suggests that the accessory proteins are not involved in biofilm integrity but rather assist the TasA proteins in the process. Indeed, the biological role of TapA in the *B. subtilis* system might be functionally comparable to the CsgB protein during Curli amyloid fiber biogenesis in *E. coli*. In this system, polymerization of the main protein component of the fiber, CsgA, is aided by a minor curli subunit protein CsgB by acting as a nucleator that, additionally, is able to polymerize by itself exhibiting amyloid properties (341). However, contrary to what occurs for TapA and TasA-bs, CsgB and CsgA do not need to be synthesized in the same cell to assemble a curli fiber and a functioning biofilm (341). On the other hand, the fact that TasA-bc exhibits a higher degree of structural polymorphism might influence the biological function of this protein, as it is known for other amyloid proteins that different structural assemblies can modulate the interaction with biological entities (342) (50). The role of CalY in this process is determinant, as it appears to be catalyzing the polymerization of TasA-bc, giving rise to different populations of aggregates and fibrils. Interestingly, a role in virulence and adhesion of bacterial cells to the host has been reported for CalY, in which this amyloid protein is rendered monomeric in early stationary-phase and located on the bacterial surface mediating the initial contact with the host cells. During biofilm formation, however, CalY is secreted and integrated in the matrix amyloid fibers, presumably, favoring TasA polymerization and conferring *B. cereus* biofilms with the versatility needed to colonize the host or interact with abiotic surfaces, and maintain the structural stability of the entire bacterial population (343).

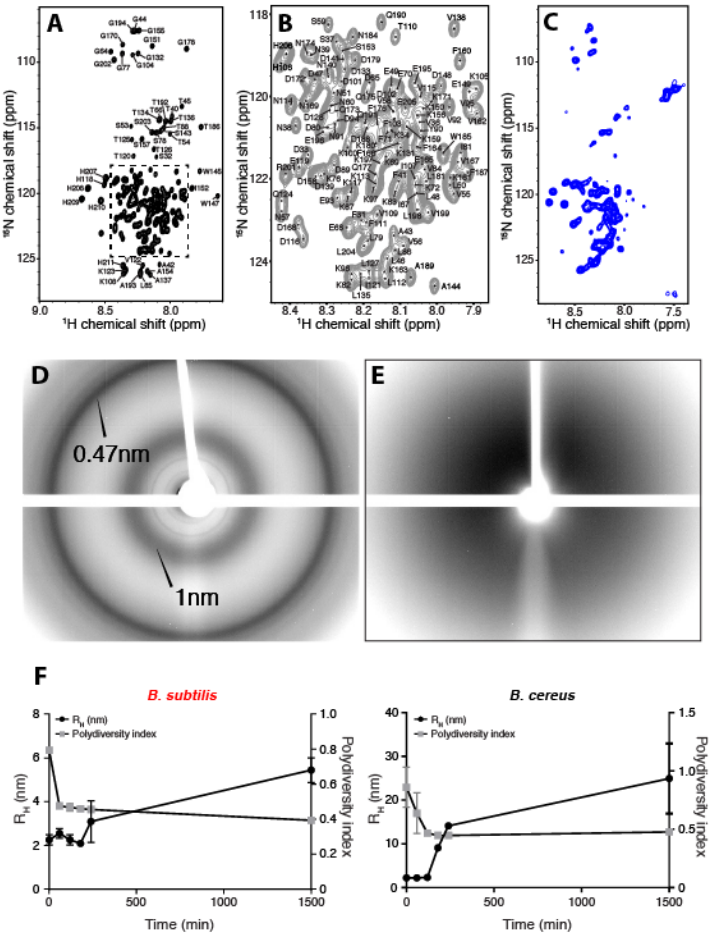
Our results contribute to the ongoing investigations of *Bacillus* strain biofilm properties and offer a unique comparative analysis of *B. subtilis* and *B. cereus* in their fibrillar disclosed amyloid form, providing new insights regarding the potential involvement of the accessory

proteins TapA and CalY in biofilm design. We corroborate the main understanding of *Bacillus* biofilm as amyloid-based entities and add a species comparison to the current knowledge, as TasA-bc functional amyloids are now proven to be similar in fold and shape to TasA-bs, while the two biofilms display different local structural properties and assembly kinetics. More studies of both systems will tell us how these local differences have been shaped by the evolution of both bacterial species to permit: i) efficient interactions with complementary species-specific ECM components, and thus define the overall chemical features of the biofilm and ii) effective response to the variety of niches that these bacteria species encounter during their life cycle.

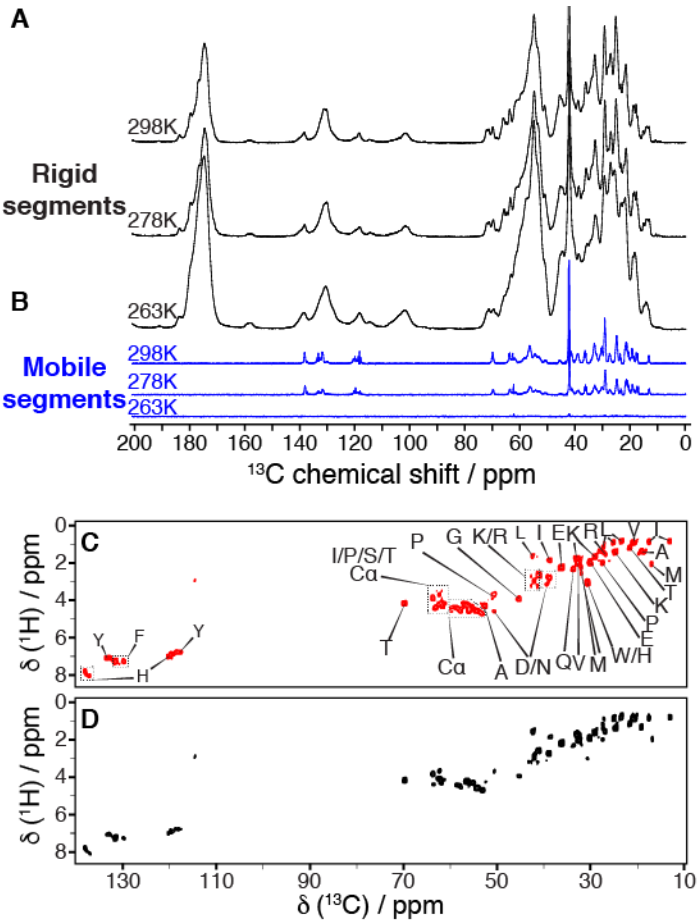
3.8. Supplementary data



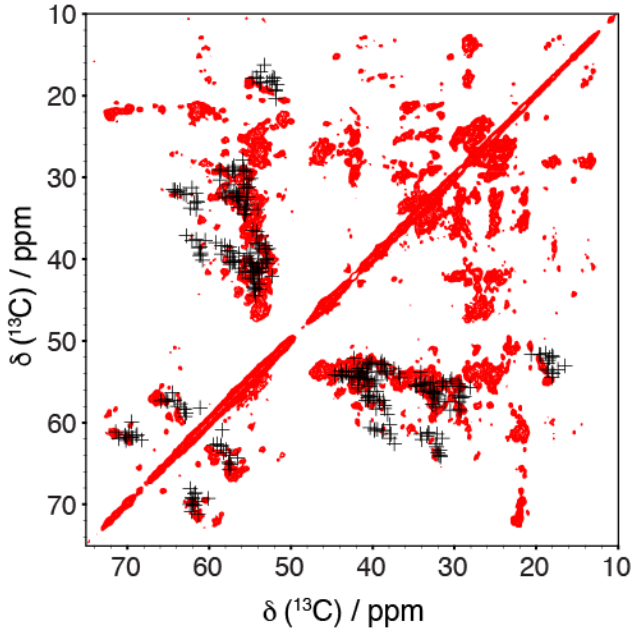
Supplemental Figure 3-1: Amino acid sequence alignments of (from top to bottom): (i) *B. cereus* TasA and *B. cereus* CalY, (ii) *B. subtilis* TasA and *B. subtilis* TapA, (iii) *B. subtilis* TapA and *B. cereus* CalY and (iv) all four sequences. The sequence alignments were generated by Jalview with Clustal. Proteinase K-resistant segments are highlighted in bold characters (see Section "*B. subtilis and B. cereus TasA functional amyloids share the same structural fold but different local structural polymorphism"*).



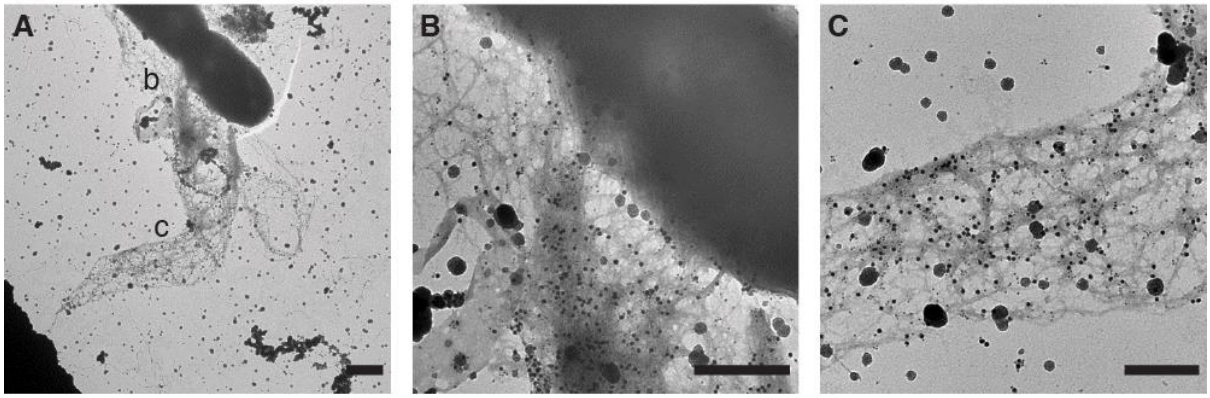
Supplemental Figure 3-2: (A) 2D ^{1}H–^{15}N SOFAST–HMQC spectrum of *B. cereus* TasA. (B) Excerpt of the HMQC spectrum showing the NMR assignment. (C) 2D ^{1}H–^{15}N SOFAST–HMQC spectrum of *B. cereus* CalY. (D) X-ray diffraction patterns of *B. cereus* TasA filaments showing reflections at ~4.7 Å and ~10 Å, (E) background diffraction. (F) Evolution of the mean hydrodynamic radius and polydispersity index of a solution of TasA (50µM) in acid solution after treatment with 10% NaOH to a final pH of 7.0 measured over the course of 24 h at room temperature by DLS.



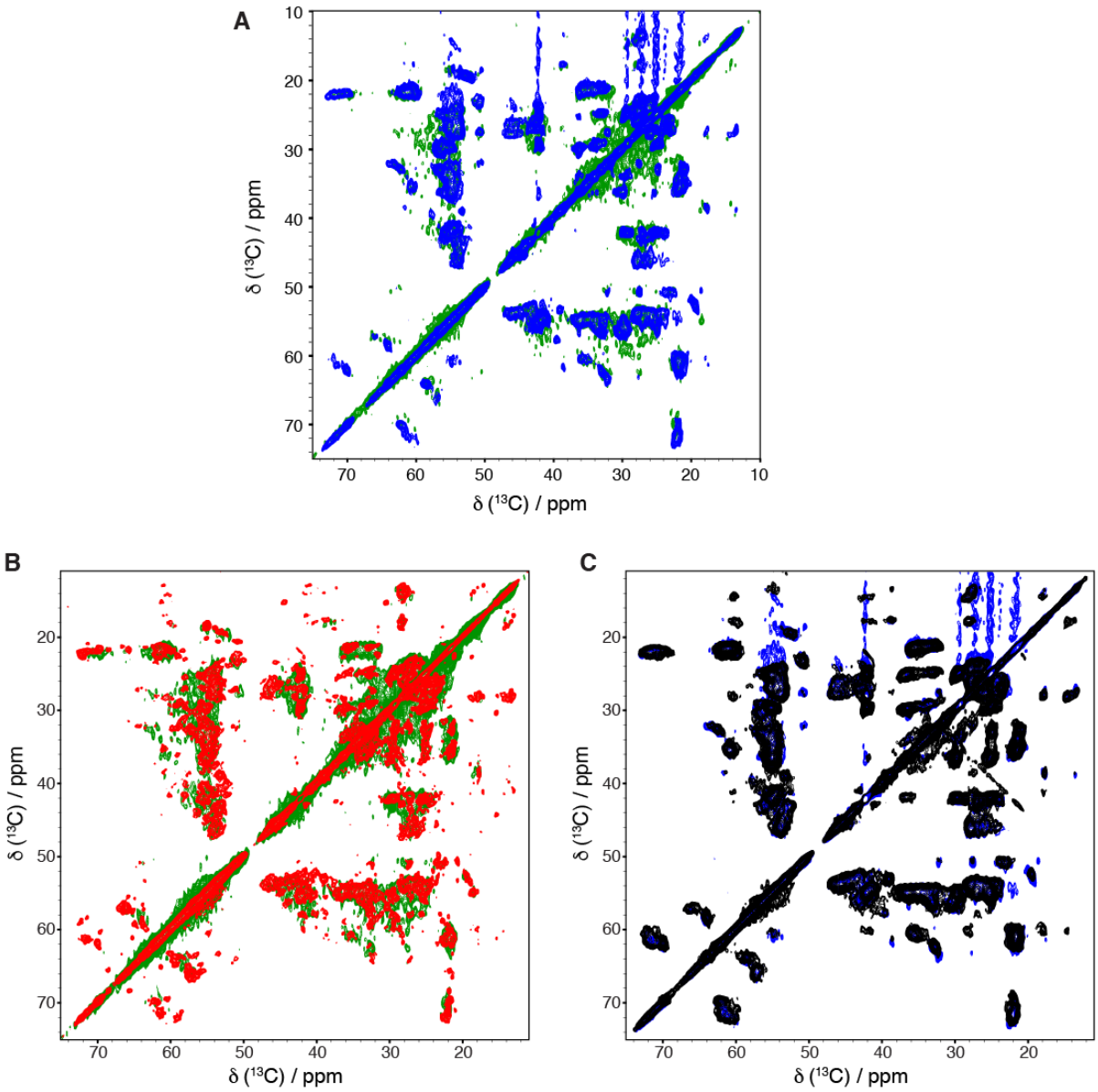
Supplemental Figure 3-3: (A) Through-space cross-polarization experiments (in black) reveal rigid protein segments in the fibrillar assembly of *B. subtilis* TasA. (B) Through-bond INEPT experiments (in blue) reflect mobile segments. Two-dimensional SSNMR ^{1}H-^{13}C correlation experiment using through-bond mixing (INEPT) to detect the mobile segments of (C) *B. subtilis* TasA filaments (red) and (D) *B. cereus* TasA filaments (black). Both correlation spectra show the same amino acid composition.



Supplemental Figure 3-4: Comparison of SSNMR 2D ^{13}C-^{13}C correlation experiment (PDSD, 50 ms mixing time) of TasA-bs to chemical shift predictions.



Supplemental Figure 3-5: (A) Transmission electron micrographs of negatively stained *B. cereus* cells decorated with fibers reactive to anti-TasA antibodies (10 nm particles) and anti-TasA antibodies (20 nm particles). (B) and (C) are closer views of "c" or "b" areas in panel (A). Scale bars are 500 nm.



Supplemental Figure 3-6: (A) Superposition of 2D ^{13}C-^{13}C SSNMR experiments *B. subtilis* TapA (green), and *B. cereus* CalY (blue) amyloid-like aggregates. Overlay of 2D SSNMR ^{13}C-^{13}C spectra of (B) TapA (green) and TasA-bc fibrils (red), and of (C) CalY (blue) and TasA-bc fibrils (black).

RES. TYPE	TasA-bs			TasA-bc		
	SSNMR	Proteinase	Sequence	SSNMR	Proteinase	Sequence
Ala	~13	9	24	~4	0	11
Arg	0	0	0	1	0	0
Asn	4	11	20	1	3	13
Asp	4	9	23	4	7	21
Cys	0	0	0	0	0	0
Gln	1	2	10	1	0	7
Glu	2	6	14	2	2	11
Gly	/	10	23	/	2	16
His	0	0	2	0	2	2
Ile	3 or 4	3	10	2	3	6
Leu	2	15	24	2	4	19
Lys	4	11	34	2	11	29
Met	0	2	6	0	0	3
Phe	1	6	12	1	2	13
Pro	3 or 4	3	7	0	0	2
Ser	~15	10	17	~4	1	9
Thr	~10	6	15	~6	3	16
Trp	0	0	2	0	0	3
Tyr	0	3	6	0	1	1
Val	4	4	12	2	6	15
TOT	~68	110	261	~32	47	197

Supplemental Table 3-1: Amino-acid content of the amyloid core as detected in the 2D ^{13}C-^{13}C SSNMR experiments.

Chapter 4: *Bacillus subtilis* biofilm molecular organization is modulated by the expression of *tasA* and extracellular polymeric substances

4.1. Nonstandard abbreviations

INEPT: Insensitive nuclei enhanced polarization transfer

MAS: Magic angle spinning

SSNMR: Solid-state nuclear magnetic resonance

TEM: Transmission electron microscopy

EPS: Exopolysaccharides

DP: Direct polarization

ECM: Extracellular matrix

GlcNAc/MurNAc: N-acetylglucosamine/N-acetylmuramic acid

A2pm: Diaminopimelic acid

Glc/Gal/Rib/Fur/Pyr: Glucose/Galactose/Ribose/Furanose/Pyranose

UDP: Uridine diphosphate

4.2. Context

Bacteria have a wide range of adaptive behaviours to overcome extreme environmental conditions. Amongst such responses, biofilm formation remains today one of the most fascinating as it implies the development of complex homo- or hetero-bacterial communities which are surface-associated (344-347). A common feature of bacterial biofilms is the production of an extracellular matrix (348) composed of exopolysaccharides (EPS), adhesins,

proteins, extracellular DNA, etc, which facilitates nutrients distribution (349) within the community and also mechanically protects from external invasion. In *Bacillus subtilis*, a sporulating gram-positive bacterial species (350, 351), such gigantic organization also involves the production of two proteins, TasA and TapA (315), which have been proven crucial for intact biofilm development. The two proteins are encoded by the *tapA-sipW-tasA* operon (313, 352, 353). TasA is now considered to be a functional amyloid (301, 303, 314, 354, 355), whose folding into cross-β fibres participates in protecting the biofilm integrity. TapA is known as TasA partner protein and is believed to help in TasA self-assembly into a functional amyloid. The protein BslA forms a hydrophobic overcoating at the surface of the biofilm (72, 356-358). In laboratory conditions, *B. subtilis* biofilms are observed as pellicles at the air-liquid surface when grown in static conditions in liquid cultures and as wrinkled bacterial lawns when incubated on agar media (359).

Following the activation of specific biofilm formation signalling pathways (360), *B. subtilis* starts to produce and secrete a variety of protein, carbohydrate, or nucleic acid biomolecules. The EPS is produced by enzymes encoded in a large operon of 15 genes. *eps*-deleted mutants show drastically different morphologies when grown in biofilm-forming conditions. Both TasA and EPS contribute to biofilm hydrophobicity and repellence (361) and are thus considered as the biofilm skeleton. EPS produced in *B. subtilis* have a wide range of sizes, from 0.57 to 128 kDa and are of different chemical nature. There are different classes of EPS (360): structural, sorptive, surface-active, active, informative, redox-active and nutritive. Structural EPS are levan type I and II. Both consists of β-2,6- linked D-fructose units with a glucose subunit α-glycosylated on the terminal fructose in levan II. Their biosynthesis has been linked to the genes *yhxB*, *yveK*, *yvfF*, *yveQ* (also called *epsG*), and *yveR* (also called *epsR*). This type of EPS is suggested to protect the cells from external stresses and handles solutes distribution within the biofilm. Sorptive EPS include the non-toxic, biodegradable and viscous polymer of D- and L-glutamate called poly-γ-glutamate (γ-PGA). γ-PGA mass is over 10000 kDa. Their biosynthesis is ATP-dependent and linked to the genes *pgsBCA*. Surface-active EPS comprise all lipopeptides such as surfactin. The active EPS are all enzymes found in the extracellular matrix such as kinases, proteases, peptidases, etc. Another category contains saccharides-related enzymes such as transferases, racemases, levanases, lyases, etc. It should be noted that many studies report the presence of DNA in the biofilm matrix, however very little is known about the role of such entities.

We aimed to focus on the composition alteration and the structural change upon *eps*, *tasA*, and *tapA*, mutations. Very little is known about how the EPS, TasA, and TapA influence the extracellular matrix and the cell wall composition of *Bacillus subtilis* in the context of biofilm formation. SSNMR is used as a non-invasive method to assess and probe whole bacterial cells in near native conditions (37). Recent applications of the technique have been reported to study the cell wall architecture of *Aspergillus fumigatus* and plant cells (38, 43, 362). Here our main goal was to explore how SSNMR performed on intact *Bacillus* cells could help to understand the effect of several mutations on the cell wall structure. We analysed and compared by SSNMR *B. subtilis* WT and *eps*, *tasA*, and *tapA* mutants involved in biofilm formation. In whole-cell samples and in the context of fully $^{13}C/^{15}N$-labelled samples, all biomolecules were enriched in $^{13}C/^{15}N$ isotopes. A detailed assignment of NMR signals could be tricky to obtain and to disseminate the structural composition of *B. subtilis* biofilms, considering the complexity of such native systems, containing whole bacteria and all extracellular matrix components. We favoured the use of spin dilution in an attempt to reduce the number of NMR correlations and make the analysis as accurate and exhaustive as possible. We thus used $^{13}C_2$-1,3-glycerol-containing media to cultivate the different bacterial biofilms used in this study. Such labelling schemes has been documented in the past mostly in *E. coli (363)* or *Legionella pneumophila* (364).

SSNMR offers a large variety of pulse sequences probing a large range of dynamical states (*i.e.*, hereafter, mainly ^{13}C atoms in flexible and rigid states) as the choice of polarization transfer scheme will purposely filter out molecular components with specific dynamical properties. CP-based experiments are designed to observed ordered and rigid ^{13}C atoms (motion in the ms range), while INEPT- and DP-based experiments will provide insights into more flexible molecules (motion the μs-ns range). All spectra were recorded on a 600 MHz spectrometer (proton frequency) at 276 K using a 4 mm probe, except for 2D ^{13}C-^{13}C INADEQUATE spectra which were recorded on 500 MHz (proton frequency) at 284 K using a 4 mm probe. Considering the slight heating rendered by magic angle spinning at 11 kHz, the final sample temperature borders 4-5°C.

We analysed four *B. subtilis* strains: WT NCIB 3610, Δ*eps*, Δ*eps* Δ*tasA*, Δ*eps* Δ*tapA*. Our results encompass the analysis of the entire bacterial biofilm including the extracellular matrix and the bacteria, and the effects of biofilm-related gene deletions on such compositions. Although the mutants show very different bacterial behaviour after static cultures in MSgg, we refer to all samples as biofilms. We observed NMR signals for lipids, amino acids, nucleic

acids, mono- and polysaccharides. Interestingly the cell wall peptidoglycan was clearly identified. We note a general flexible and mobile feature associated with *B. subtilis* WT biofilms, and mutations altered the rigid/flexible ratios and contributions.

4.3. Methodology

4.3.1. Bacterial strains and culture conditions

The strains used in this study were *B. subtilis* NCIB3610, an undomesticated prototrophic strain from our laboratory collection. *B. subtilis tasA/eps* and *tapA/eps* double mutant were used (*B. subtilis* NCIB3610, *tasA*::spc, *eps*::tet; *Bacillus subtilis* NCIB3610 *tapA*::cat, *eps*::tet). Biofilms were grown in MSgg medium: 100 mM morpholinepropane sulfonic acid (MOPS) (pH 7), 0.5% [1,3-$^{13}C_2$]-glycerol, 0.5% glutamate, 5 mM potassium phosphate (pH 7), 50 µg/mL tryptophan, 50 µg/mL phenylalanine, 50 µg/ml threonine, 2 mM $MgCl_2$, 700 µM $CaCl_2$, 50 µM FeCl3, 50 µM MnCl2, 2 µM thiamine, 1 µM $ZnCl_2$. MSgg was enriched in ^{15}N labelled L-glutamic acid and ^{13}C labelled carbon sources ([1,3-$^{13}C_2$]-glycerol) for SSNMR purposes. The antibiotics used and their final concentrations were: 100 µg/mL spectinomycin, 5 µg/mL chloramphenicol and 10 µg/ml tetracycline. Biofilms were grown in liquid static culture at 30°C for 48 hours to obtain *B. subtilis* pellicle biofilms at the air-media interface. Samples were kindly provided by Dr Diego Romero's laboratory from Málaga University (Spain).

4.3.2. Solid-state NMR

B. subtilis native colonies were filled in 4 mm SSNMR rotors. Experiments were performed on a 600 and 500 MHz Bruker Biospin spectrometers (proton frequency) equipped with 4 mm Efree and HFX (respectively) probes at a magic-angle spinning frequency of 11 kHz. 1-dimensional spectra were recorded using 4k scans at 298 K. 1D ^{13}C CP experiments were recorded with a contact time of 1 ms. The 2D ^{13}C-^{13}C J-INADEQUATE experiments were recorded with a 3-second recycle delay using 256 scans and acquisition times of 4 ms (indirect) and 20 ms (direct) at 287 K. The 2D ^{1}H-^{13}C INEPT experiments were recorded using 80 scans and acquisition times of 30 ms (indirect-^{1}H) and 15 ms (direct-^{13}C).

4.4. Results

4.4.1. Polysaccharides in the extracellular matrix and the cell wall are flexible

A WT *B. subtills* strain was grown in the group of D. Romero (University of Malaga, Spain), using $^{13}C/^{15}N$ liquid MSgg medium in static conditions for several days. For SSNMR analysis, the obtained biofilms were simply retrieved without any treatment, and packed into SSNMR

rotor in native conditions. SSNMR experiments were recorded in the following days. Using a battery of ^{13}C- and ^{15}N-recorded 1D experiments detecting flexible and rigid molecules within the sample, we first assessed the signal and dynamics levels of the $^{13}C/^{15}N$-labelled WT *B. subtilis* biofilms. The 1D ^{13}C DP experiment (*i.e.*, single pulse excitation) is a rather accurate way of quantitatively assessing the samples content and accounts for both rigid and flexible regions. It should be noted that such experiment, although in theory insensitive to dynamics, can only be used for quantitative analyses when sufficiently long relaxation periods are provided between scans (365) (**Figure 1A**, middle panel). As a first rough analysis and assessment of the sample, a basic 2 s recycle delay was used for all 1D experiment regardless of the probed ^{13}C dynamical states. Interestingly, the spectra display peaks ranging from ~10 to ~190 ppm thus accounting for a variety of biomolecules. In all spectra shown in this chapter, biomolecule types are exhibited using a coloured scheme: residue chemical shifts in black, lipids' in orange, saccharides' in blue, and other aromatics peaks in dark red. ^{13}C resonances range include CH_2-CH_3 chemical groups (~10-40 ppm), C-N bonds (45-65 ppm), C-O bonds (~60-110 ppm), C=C/C=N double bonds (110-160 ppm), C=O double bonds (170-200 ppm). The 1D ^{13}C-INEPT filters out ordered ^{13}C atoms and interestingly seems to display a similar spectral fingerprint as the direct polarization (with the exception of ^{13}C atoms unbonded to ^{1}H nuclei), suggesting the majority of biomolecules are in a flexible state (**Figure 1A**, upper panel). Interestingly, the 1D ^{13}C-CP clearly shows lower signal levels, specifically for non-protein biomolecules (**Figure 1A**, lower panel). Most rigid ^{13}C atoms seem to correspond to side-chains or lipidic CH_2-CH_3 chemical groups.

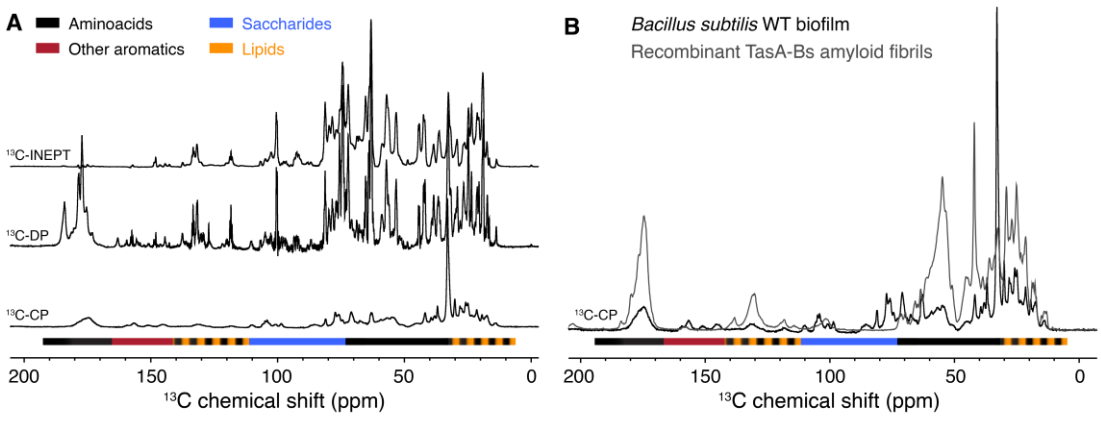


Figure 4-1: Mobility-based analysis of *B. subtilis* WT biofilms. (A) 1D ^{13}C CP (lower spectrum), DP (middle spectrum), INEPT (upper spectrum) of *B. subtilis* WT biofilms as observed using a 600 MHz NMR spectrometer (proton frequency) at 284 K. (B) Superimposition of the 1D ^{13}C CP spectra of the *B. subtilis* WT biofilms (black), and TasA $^{13}C/^{15}N$ recombinant amyloid fibrils (light grey) recorded on a 600 MHz NMR spectrometer (proton

frequency) at 276 K. In both panels, biomolecules are distributed within the spectra depending on their chemical nature: lipids in dark yellow, amino acids in black, saccharides in blue, and aromatic molecules in dark red.

Given the prior knowledge that TasA assembles (314) into tight amyloid fibrils within the extracellular matrix of *B. subtilis* biofilms, one might think that TasA amyloid fibrils are thus detected in the 1D ^{13}C-CP experiment (354, 366). To test such hypothesis, we compared the spectrum to that of ^{13}C/^{15}N-labelled recombinant TasA amyloid fibrils (354). **Figure 1B** shows the superimposition of spectra obtained from the two samples: WT *B. subtilis* biofilms in black, recombinant TasA-bs fibrils in light grey (**Figure 1B**). Interestingly, the residue side-chain spectral area (*i.e.*, 0-50 ppm) seems quite similar in terms of general chemical shift fingerprint, with one exception for a peak at ~33 ppm which seems to display a higher signal level (**Figure 1B**). Nevertheless, only side chain ^{13}C atoms seem to be rigidified within the biofilm, which is explained by the fact that the $^{13}C_2$-1,3-glycerol scheme mostly enriches for side chain carbons, as opposed to $^{13}C\alpha$ peaks (*i.e.*, 45-65 ppm). Surprisingly, few polysaccharide signals are detected in the CP-based experiment thus pointing towards the fact that both the cell wall and EPS in the biofilm matrix are not completely ordered. A further 2D CP-based ^{13}C-^{13}C experiment, showed the absence of a diagonal signal, suggesting an overall lack of "higher-order" rigidity and the presence of intermediate dynamic regimes, difficult to detect using CP polarization transfer.

4.4.2. WT *B. subtilis* biofilm content analysed by SSNMR

The main ECM components (367) are synthesized by the products of the *epsABCDEFGHIJKLMNO* operon. Judging from the previous results, we turned to INEPT- and J-coupling-based experiments to gain access to polysaccharides. The 2D ^{1}H-^{13}C INEPT shows the multiple spectral areas as displayed in **Figure 2A**. As the spectra are rather overcrowded, it is quite tricky to unambiguously assign these peaks. However, they give a good evaluation of the kind of molecules the most present in WT *B. subtilis* biofilms in the flexible state. WT biofilms seem to contain unordered proteins in the extracellular biofilm matrix (**Figure 2A**). In the same aromatic spectral region, most signal can be assigned to residues such as phenylalanine, tyrosine, histidine, and tryptophan (below ~140 ppm ^{13}C chemical shifts). Interestingly, the peaks presenting ^{1}H chemical shift around ~8 ppm non-protein molecules which could be nucleotides or even DNA and RNA (360) (**Figure 2A**). The presence of non-protein acyl ^{13}C can be mentioned, however no unsaturated ^{13}C can be detected, thus suggesting the sole presence of saturated lipids, if any. Many lipid contributions are found in the 2D ^{1}H-

^{13}C INEPT spectrum from ~20 to ~40 ppm in the ^{13}C dimensions, and ~2 ppm in the ^{1}H dimensions (**Figure 2A**). Furthermore, several carbohydrates thus seem to be detected in our sample judging from the number of peaks from 85-105 ppm (^{13}C dimension) which usually corresponds to C1 atoms (368-371). Looking closer to the carbohydrate region, several types of poly- and monosaccharides are detected (**Figure 2B**), both α- and β-anomers.

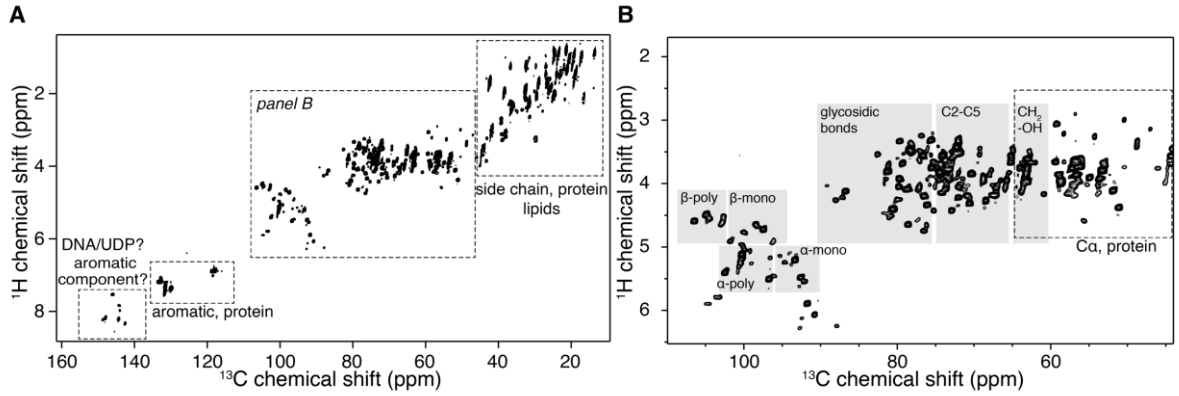


Figure 4-2: *B. subtilis* WT biofilms are made of highly flexible biomolecule types. (A) 2D ^{1}H-^{13}C through-bond INEPT experiment recorded on *B. subtilis* WT biofilms using a 600 MHz SSNMR spectrometer (proton frequency) at 276 K. Biomolecule areas are delimited in the figure. (B) Excerpt of the carbohydrate region in the 2D ^{1}H-^{13}C through-bond INEPT experiment as shown in panel A. For clarity purposes, the expected areas for poly- and monosaccharides are highlighted in light grey.

The 2D ^{13}C-^{13}C J-INADEQUATE experiment (**Figure 3A**) is very useful for full assignment of complex samples as it connects directly bonded ^{13}C atoms and allow for the rapid identification of the molecules. Double-quantum (DQ) ^{13}C coherences were excited through ^{13}C-^{13}C J-coupling and correlated to single-quantum (SQ) signals. The cell wall components, which are peptidoglycan (372, 373), of *B. subtilis* was detected quite unambiguously. Indeed, peptidoglycan molecules are polymers of glycans cross-linked by peptides. The glycan found in most Gram-positive bacteria is made of GlcNAc and MurNAc (1:1 ratio) (**Figure 3B, upper left panel**). Several GlcNAc and MurNAc signals are unequivocally assigned within the spectrum with quite high signal to noise ratios: C1 ~92.6/93.6 ppm, C2 ~56.9 ppm, C3 ~74.6/72.9 ppm, C7 ~177.1 ppm, C8 ~24.9, C9 ~80.8 ppm, C10 ~185.5 ppm, C11 ~21.5 ppm. Interestingly, additional N-acetylated sugars are observed in the figure and could be associated with diamino sugars that are found in *B. subtilis* (374). Signals related to the peptidyl chains are found within the amino acid region of the 2D ^{13}C-^{13}C J-INADEQUATE spectrum, as assigned to L-alanine, D-glutamic acid, diaminopimelic acid (A2pm), D-alanine (**Figure 3B,**

upper right panel). Interestingly, other residues are detected with a quite high intensity and could be assigned to any protein or peptide present in the biofilm matrix and displaying a random coil-like conformation. The case of A2pm is however quite unique as such amino acid is not naturally found in protein primary sequences. Its spin system clearly displays two Cα-like (*i.e.*, 56 ppm and 57.6 ppm) and Cβ-like (*i.e.*, 33.1 ppm and 32.5 ppm) chemical shifts connected to single Cγ- (*i.e.*, 23.5 ppm) and CO-like (*i.e.*, 177.4 ppm) ^{13}C atoms. All chemical shifts are shown in **Table 1**. **Figure 3C** displays the chemical structures of unusual saccharides and the peptidyl chain of peptidoglycans.

	Sites	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**
Cell Wall												
	GlcNAc	93.6	56.9	74.6/73.4	65.2			177.1	24.9			
	MurNAc	92.6	56.9	72.9	65.2			177.1	24.9	80.8	185.5	21.5
	L-Ala		53/53.6-CA	18.8/19.3-CB								
	D-Glu	178.1-CO	57.4-CA	29.5-CB	36-CG	184-CG						
	A2pm	177.4-CO	56/57.6-CA	33.1/32.5-CB	23.5-CG							
	D-Ala		53/53.6-CA	18.8/19.3-CB								
Structural EPS												
	Levan	106.6/105.1	75.7	78.3								
	Pyr	100.4	74.3	75.8		74.6	63.2					
	GluA					76.8/78.8	179.8/183.7					
	O-acetyl							184	26			
Sorptive EPS												
	PGA	178.1-CO	57.4-CA	29.5-CB	36-CG	184-CD						
Miscellaneous												
	Fur	91.8	76.3	72.1								
	Uridine				168.6	104.9/103.8	144.5/146.3					

Amino Acids												
	Val		63.6-CA	31.7-CB	20.6-CG							
	Ile		62.3-CA	38.5-CB	27.1-CG1	17.3-CG2	13.9-CD					
	Leu		56.1-CA	42.5-CB	26.7-CG	24.8-CD						
	Arg				29.1-CG	41.9-CD						
	Thr	177-CO	63.2-CA	68.7-CB	22.1-CG2							
	Trw			26.1-CB	110.3 -CG							
	Phe				137.8 -CG	130.6 -CD1	131.8 -CD2		138.5 -CE2			

Table 4-1: Assigned ^{13}C chemicals shifts as identified in all samples.

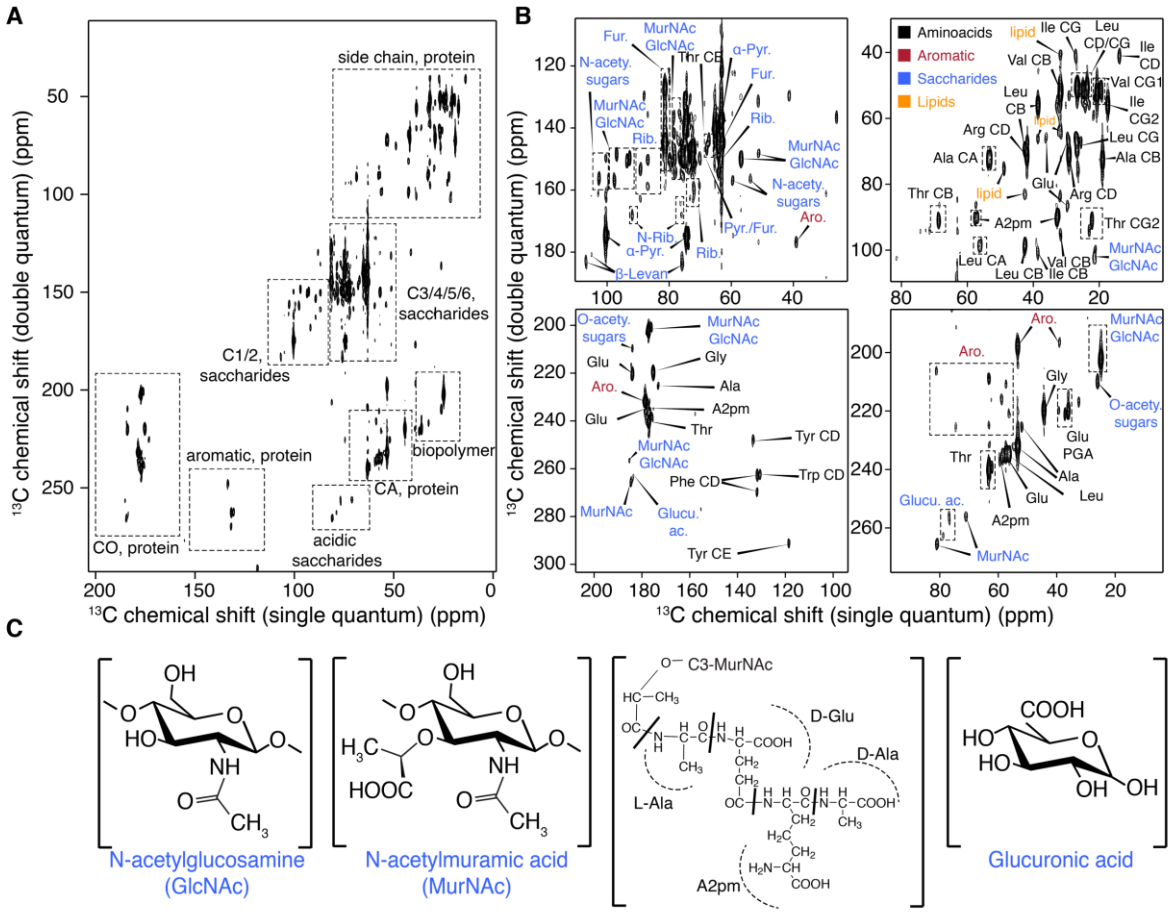


Figure 4-3: Both the cell wall and the biofilm extracellular matrix are detected as flexible moieties. (A) 2D ^{13}C-^{13}C J-INADEQUATE recorded on a 500 MHz SSNMR spectrometer (proton frequency) at 284 K, which resolves the ^{13}C through-bond connectivity for each biomolecule. (B) Excerpts of the ^{13}C J-INADEQUATE. The upper

left area usually accounts for the carbohydrates, the amino acids are found mostly in the upper right regions. (C) Peptidoglycan and glucuronic acid elements are illustrated.

The vast majority of carbohydrates detected in our experiments are related to the biofilm matrix itself and must therefore be part of the extracellular matrix. **Figure 3B** (upper left panel) shows the carbohydrate region of the 2D ^{13}C-^{13}C J-INADEQUATE experiment and we based our analysis on the literature chemical shifts and matrix content to assign most signals. Amongst structural EPS in *B. subtilis*, levans, which are polymers of β-D-(2→6)-Fru*f*, are here detected within the biofilm. Levans are referred to as β-Levan in **Figure 3B** (upper left panel). When compared to other carbohydrate spin systems, levan-related peaks seem to display less intensity in J-coupling NMR schemes, perhaps accounting for a slightly more rigid state. Several studies indeed report levan biopolymers to form a capsule to protect bacteria and prevent cell desiccation by attracting solutes to the cells through an osmotic effect (375, 376). Another category of structural EPS counts several exopolysaccharides composed of glucose, galactose, fucose, glucuronic acid, and other O-acetylated sugars (approximate molar ratio 2:2:1:1:1.5) (377). With the exception of fucose, all mentioned saccharides were probed by SSNMR with the expected ratios. Glucose and galactose showed however higher intensity than any other spin system and seem to display a D-α-pyranose confirmation (hereafter mentioned as α-Glc*p* and α-Gal*p*). In the figure, peaks assigned as α-Pyr refer to those molecules (**Figure 3B,** upper right and lower left panels). In addition to structural EPS, sorptive EPS have been documented in the literature. Poly-γ-glutamate has been the main focus of study and has been reported to be non-toxic, anionic, and biodegradable (378-380). Its mass can reach up to 10 000 kDa with both L- and D-monomers. Its synthesis in *B. subtilis* is ATP-dependant and therefore raises the question of its role in biofilm formation. In *B. anthracis* it has been linked to virulence (381). In our samples, such polymers might be difficult to unambiguously assign as peptidoglycan of most Gram-positive bacteria contains D-glutamic acid as well. The complexity of the system overcrowds such spectrum and limits the identification process. However, in addition to all mentioned carbohydrates, several N-acetylated and furanose signals were assigned. As levan is made of β-D-(2→6)-Fru*f* the additional peaks should be considered as the remaining ^{13}C atoms. As for additional N-acetylated carbohydrate-like signals, the potential polymorphism of peptidoglycan, and more specifically GlcNAc and MurNAc, can be contemplated.

An interesting spin system is the one assigned to ribose, which is quite isolated and displays a C1-like atom corresponding to nitrogen-linked chemical shifts **(Figure 3B**, referred to as N-

Rib). Such set of peaks is highly reminiscent of ribose signals in nucleotides. An additional spin system comprising chemical shifts is observed with very low signal intensity: 144.5-146.3 ppm, 104.9-103.8 ppm (382). Both aromatic ^{13}C are identified in the aromatic region as well in the 2D ^{1}H-^{13}C INEPT with ^{1}H chemical around 8 ppm. Additional peaks in the same region might correspond to DNA-related spin systems as one might expect in extracellular biofilm matrix (383, 384). Several amino acids are also detected with a high intensity, such as Val, Arg, Leu, Ile, Ala, and Thr (Figure 3B, upper right panel).

4.4.3. Biofilms of *eps*-, *tasA*-, *tapA*-deleted mutants are dynamically heterogeneous

We tested three *B. subtilis* mutants to see the effect of deleting the *epsABCDEFGHIJKLMNO* operon, and the *tasA* or *tapA* genes on the biofilm composition and properties. **Figure 4A** displays the 1D ^{13}C-DP spectra recorded on the four *Bacillus* strains (including the already displayed WT spectrum): WT in black, Δ*eps* in blue, Δ*eps* Δ*tasA* in red, Δ*eps* Δ*tapA* in green. **Figure 4C** displays the biofilms morphology as obtained from the WT and mutant strains. WT biofilms show the typical wrinkled pellicle as obtained after 48 hours of incubation in MSgg medium, while the Δ*eps* strain form fragile and flat pellicles. Both double mutants show drastically different bacterial communities which seem to lack integrity (**Figure 4C**). The samples show quite similar ^{13}C NMR fingerprints with the noted exception of a peak around 113 ppm found in the two double mutants. Considering the signal level compared to other ^{13}C contributions and the large linewidths, those peaks are most likely to be a contamination of the sample, or a component of the buffer. Fortunately, this peak does not hinder the spectra analysis as it is found isolated from the biological signals. The 1D ^{13}C DP recorded on the WT *B. subtilis* biofilm was utilized to attempt to correct the differences in signal levels (*i.e.*, not to consider the sample quantity in each rotor) in all 1D spectra. We indeed calculated simple signal ratios based on absolute integration between the DP spectra of the WT sample compared to mutants, or using peaks assigned to a cell wall component. In the same conditions, such approximation should give a relevant comparison of sample quantities within the SSNMR rotors. Comparing to the WT 1D ^{13}C DP experiment, Δ*eps* displays ~85% of the signal in WT, Δ*eps* Δ*tasA* ~13%, and Δ*eps* Δ*tapA* ~11% (**Table 2**). Calculating the signal ratios based on the cell wall peaks, we obtained very similar values.

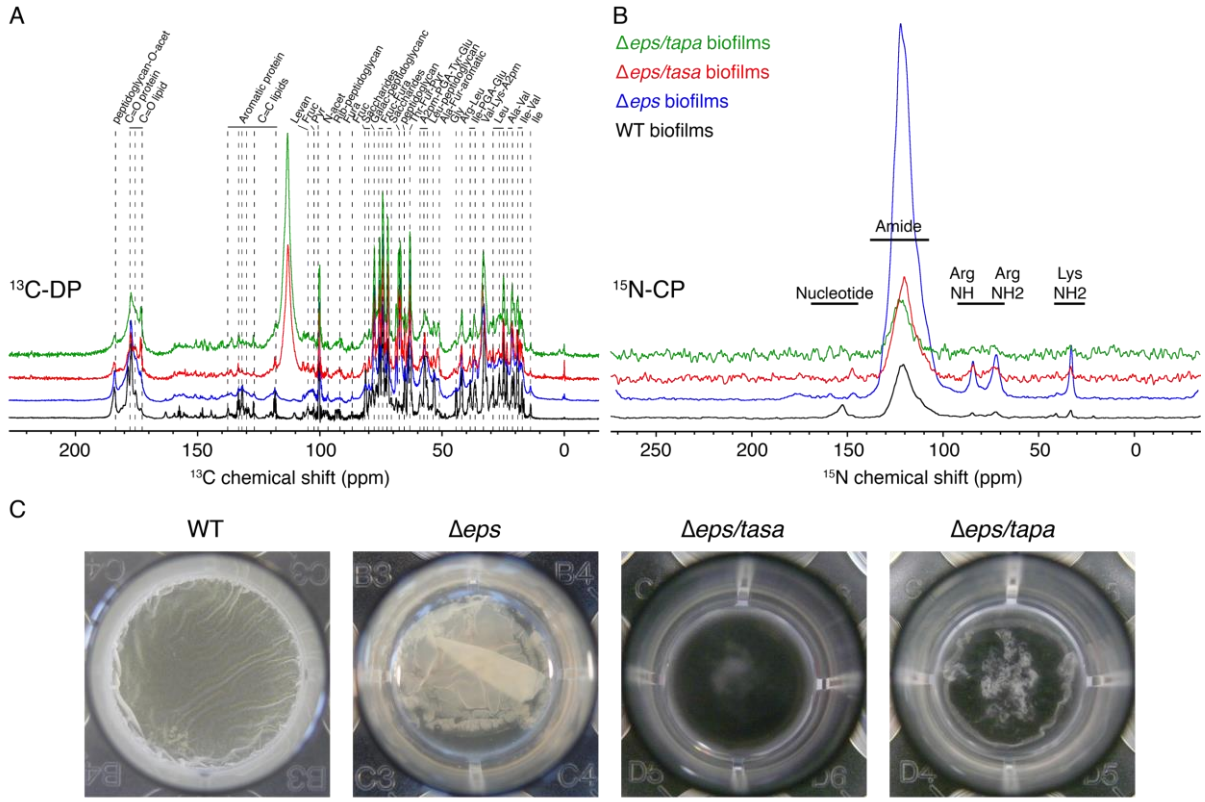


Figure 4-4: The *eps*-, *eps/tasA*-, and *eps/tapA* deletion impact the overall dynamic of the biofilms (I). (A) 1D ^{13}C DP of the four native biofilms, with previous assignments. (B) 1D ^{15}N CP spectra recorded on a 600 MHz spectrometer (proton frequency) and showing the rigid ^{15}N atoms. From bottom to top, WT in black, Δ*eps in blue*, Δ*eps* Δ*tasA* in red, and Δ*eps* Δ*tapA* in green.

	DP_{mut} /DP_{WT}	**CP^*_{mut}/ CP_{WT}***	**CP/DP * ratio**	**J_{mut}*/J_{WT} (0-60 ppm)**	**J_{mut}*/J_{WT} (60-150 ppm)**	**J/DP ratio overall**	**CP/J ratio overall**	**CP/J ratio 0-60 ppm**	**CP/J ratio 60-150 ppm**
WT	1	1	0.3	1	1	0.7	0.4	0.6	0.2
Δ*eps*	0.85	2.5	0.6	0.7	1.5	0.7	0.97	3.4	0.2
Δ*eps/tasA*	0.13	1.5	0.4	0.7	1.2	0.6	0.7	1.9	0.05
Δ*eps/tapA*	0.11	1.1	0.3	0.5	0.5	0.3	0.9	1.6	0.3

Table 4-2: Absolute signal integration ratios calculated using DMFIT. Mutant integration values marked with * were multiplied by DP_{WT}/DP_{mut} to correct the sample quantity differences and thus accurately assess the dynamical differences.

In an attempt to assess the nitrogen content of the biofilms, we recorded 1D ^{15}N CP spectra on the four samples where Δ*eps* biofilms show drastically more rigid content. Interestingly, lysine and arginine signals were stronger in Δ*eps* biofilms (**Figure 4B**). These two amino acids have

been reported to play a key role in biofilm formation of different bacteria (385-390). **Figure 5A-B** displays the 1D ^{13}C CP and 1D ^{13}C INEPT spectra of the four biofilm samples. Signal ratios were calculated and multiplied by a correction coefficient of DP_{WT}/DP_{mutant}. **Table 2** comprises all signal ratios calculated on all samples. Interestingly, all mutants displayed rather different behaviour in dynamic-based experiments: in the same conditions, the Δ*eps* and Δ*eps*/Δ*tasA* mutants seemed to have more rigid content than the WT stain (*i.e.*, 2.5 and 1.5 times more, respectively). As for the flexible biomolecules, lipids and proteins signals (0-60 ppm) seem to be ~60% that of WT levels. However, for carbohydrate and aromatic molecules, the tendency to be flexible seems higher in the Δ*eps* and Δ*eps*/Δ*tasA* mutants (*i.e.*, 1.5 and 1.2 times more, respectively). The Δ*eps*/Δ*tapA* mutant seems to present similar signal levels in CP-based experiments, while showing drastic decrease in overall flexibility. The observed differences in signal intensity thus seem to account for an alteration of the general dynamics within *B. subtilis* biofilms upon mutations in the same experimental conditions (**Table 2**).

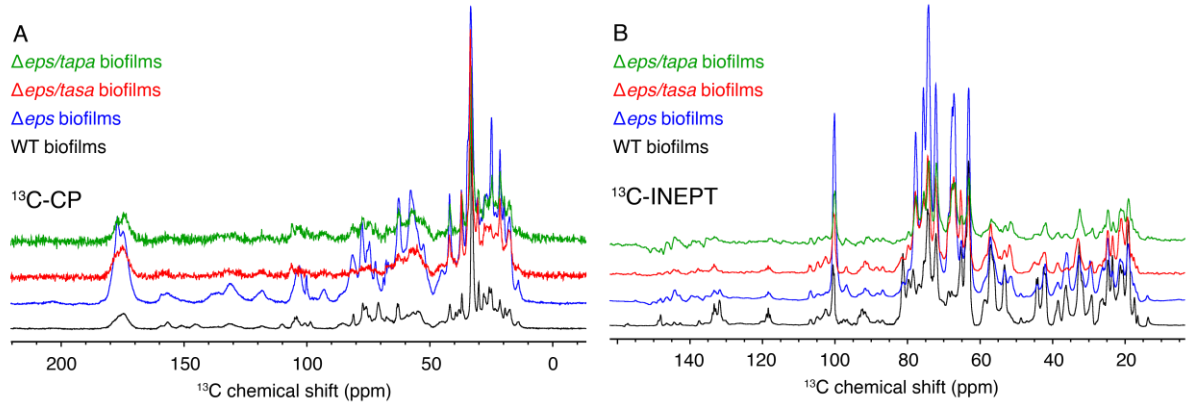


Figure 4-5: The *eps*-, *eps/tasA*-, and *eps/tapA* deletion impact the overall dynamic of the biofilms (II). (A) 1D ^{13}C CP of the four biofilms probing ordered molecules. (B) 1D ^{13}C INEPT of the four biofilms probing flexible molecules. All spectra were recorded on a 600 MHz spectrometer (proton frequency). From bottom to top, WT in black, Δ*eps in blue*, Δ*eps* Δ*tasA* in red, and Δ*eps* Δ*tapA* in green.

4.4.4. *eps*-, and *tasA*-, *tapA*-deleted double mutant biofilms display drastic changes in biomolecule content

The deletion of the complete *eps* operon led to drastic changes in the spectral fingerprints. **Figure 6A** shows the superposition of the 2D ^{13}C-^{13}C double quantum J-INADEQUATE experiments recorded on WT (as previously displayed in **Figure 3A-B**) and the Δ*eps*. Peaks found in the Δ*eps* sample spectrum are assigned in blue, regardless of the biomolecule type. Several sugars are found with lower signal intensity upon deletion of the *eps* operon: β-levan,

several N-acetylated carbohydrates, GlcNAc/MurNAc, for example. One might have expected such occurrence based on the previous observation that the Δ*eps* mutant contains more rigid components. Some peaks were only found in the Δ*eps* spectrum in the carbohydrate area as was observed in the 1D ^{13}C DP spectrum (**Figure 4A**). Noteworthy is the fact that all new signals are below ~75 ppm and thus seem to correspond to monosaccharides (*i.e.*, no glycosidic bonds). Most peaks in the amino acid region were conserved, with the same signal intensity. However, lipids and some hydrophobic amino acids were absent in the J-based 2D spectrum. As the 1D ^{13}C DP of WT and Δ*eps* biofilms are rather similar (both in signal intensity, and peak content) in the 0-50 ppm region, those molecules are present but more rigid in the *eps*-deleted biofilms (**Figure 4A**).

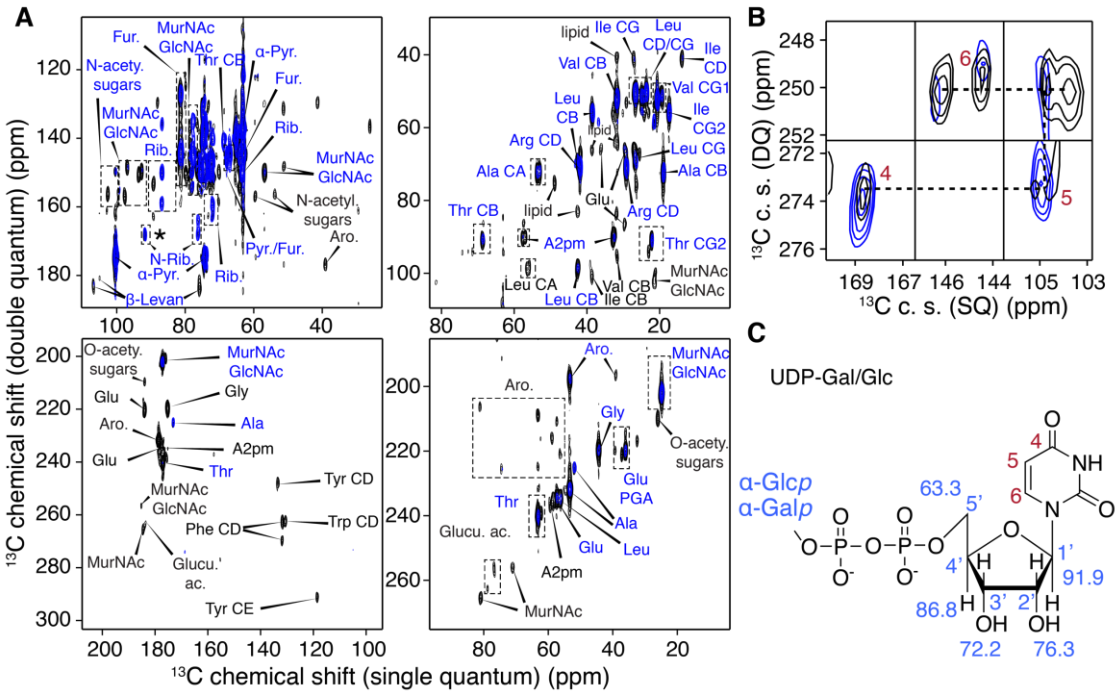


Figure 4-6: The *eps* deletion leads to an accumulation of UDP-Gal/UDP-Glc in the biofilm extracellular matrix. (A) Excerpts of the 2D ^{13}C-^{13}C J-INADEQUATE (284 K) with the WT (black) and *eps*-deleted strain (blue) as shown previously. Biomolecules assigned in black are only found in the WT data. (B) Excerpts a single aromatic spin system recorded on the same experiments assigned to UDP. (C) Chemical structure of UDP-Gal and UDP-Glc.

In the aromatic region of the 2D ^{13}C-^{13}C double quantum J-INADEQUATE experiment, the spin system mentioned above appears with more intensity and an additional chemical shift is identified at ~168.6 ppm (**Figure 6B**). Moreover, α-pyranose and amino-ribose molecules seem more abundant. We hypothesize that the chemical shifts identified as α-pyranose, amino-

ribose and aromatic molecules are part of UDP-Gal or UDP-Glc (**Figure 6A-C**). UDP-Gal was indeed proposed to be the basic toxic sugar metabolite used for EPS biosynthesis (345, 391-393). We thus suggest that in absence of *eps*-associated proteins and enzymes, UDP-Gal accumulates within the biofilms, and might be converted to UDP-Glc by the enzyme UDP-glucose 4-epimerase GalE which is crucial for *B. subtilis* to grow on galactose as it counters UDP-Gal toxic accumulation (391). In complex samples such as complete and native biofilms, it is quite difficult to unambiguously segregate and identify galactose from glucose.

We also studied the double mutants to assess the effect of the lack of both *eps* and *tasA* or *tapA* genes in biofilm formation. Again, using the J-based 2D INADEQUATE double quantum experiment, as very little signal is obtained in CP-based polarization schemes, the Δ*eps*/Δ*tasA* deletion mutant showed less NMR signals than the Δ*eps* biofilms with most peaks found in the carbohydrate and hydrophobic and charged (Leu, Val, Arg, Val, Ile) amino acid regions (**Figure 7A**). Similar to the Δ*eps* mutant, the Δ*eps*/Δ*tasA* biofilms display strong correlations for D-α-pyranose and N-ribose sugars. We recorded a 2D ^{1}H-^{13}C INEPT on both the double mutants to assess more flexible molecules and compare all four samples. Traces of β-levan were found in the double mutant lacking the *tasA* gene, whereas none was detected in the Δ*eps*/Δ*tapA* mutant (**Figure 7B**). Interestingly, high levels of α-pyranose were assessed in the four strains, and even regarding the Δ*eps*/Δ*tapA* deleted strain which also shows traces of N-ribose (**Figure 7B**).

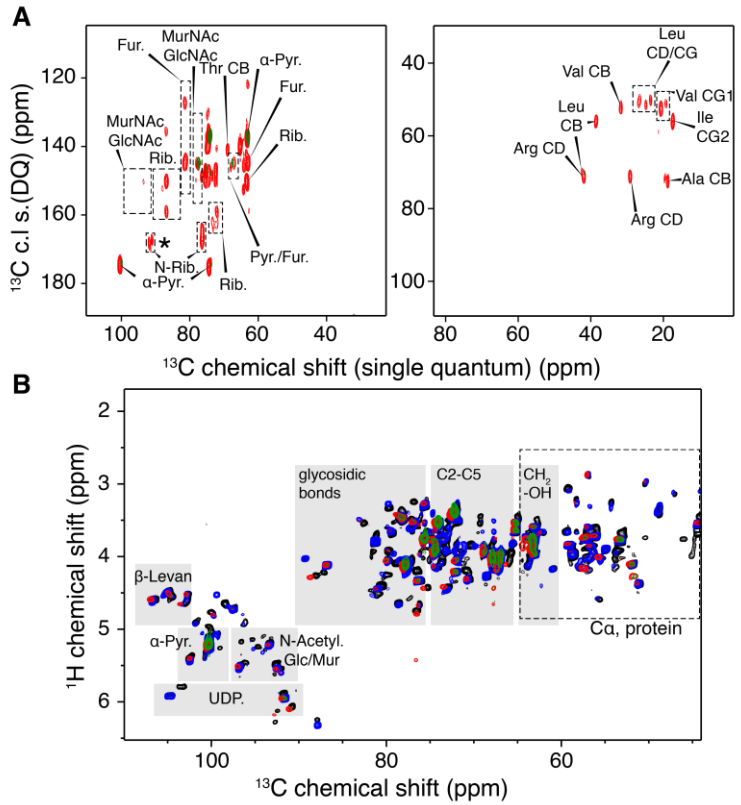


Figure 4-7: Double mutants Δ*eps/tasA* and Δ*eps/tapA* have additive effects on the carbohydrate composition compared to the *eps*-deletion only. (A) Focus on the carbohydrate and amino acids regions of the 2D ^{13}C-^{13}C double quantum J-INADEQUATE recorded on the two *B. subtilis* double mutants. Δ*eps/tasA* biofilms displayed in red and Δ*eps/tapA* displayed in green. (B) Focus on the carbohydrate region of the 2D ^{1}H-^{13}C INEPT recorded on the four *B. subtilis* strains. WT in black, Δ*eps in blue,* Δ*eps* Δ*tasA* in red, and Δ*eps* Δ*tapA* in green.

4.5. Discussion

In this study, we analysed native *B. subtilis* biofilms of WT and mutant strains using SSNMR. The cells were grown in ^{13}C/^{15}N-labelled minimal media under static conditions. Performing J-coupled pulse schemes, we identified several biomolecules within the biofilms, including the peptidoglycan elements such as GlcNAc and MurNAc and the L-Ala-G-Glu-A2pm-D-Ala peptidyl chain of the cell wall. The extracellular matrix is known to contain many biomolecules, including polysaccharides, monosaccharides, DNA, enzymes, lipids, the proteins TasA and TapA, etc (360). We were able to mostly detect (i) saccharides (*e.g.*, D-α-pyranose, ribose, β-Levan, glucose, galactose, glucuronic acid, O-acetylated saccharides), (ii), amino acids (*e.g.*, mostly, phenylalanine, tyrosine, histidine, and tryptophan, alanine, valine, arginine, leucine, isoleucine, and threonine) (**Figure 2-3**).

We analysed the impact of *eps*, *eps/tasA* and *eps/tapA* deletions on biofilm formation. A first observation was the drastic change of dynamics in the mutants when compared to WT *B. subtilis* biofilms. All mutants indeed seemed stiffer (**Table 2, Figure 4-5**). Both the *Δeps* and the Δ*eps*/Δ*tasA* double mutant gave more flexible signals in the saccharides spectral regions, and thus their analysis in multidimensional NMR allowed several observations: (i) in the *Δeps* mutant, levels of β-Levan seemed lower while both D-α-pyranose, furanose and ribose signals increased, (ii) in the same mutant, the amino acids content seemed to be the same as in WT biofilms while showing higher intensity, (iii) the double mutant Δ*eps*/Δ*tasA* did not seem to produce β-Levan altogether however D-α-pyranose, furanose and ribose signals were among the highest, strongly suggesting the accumulation of monosaccharide building blocks in absence of required *eps*-related enzymes. In both mutants, in addition to ribose, we observed an aromatic three-^{13}C spin system strongly resembling the pyrimidine-derived nucleic acid uracil. Such observation seems to support the previously reported hypothesis that UDP-galactose and UDP-glucose are involved in the formation of EPS in the extracellular matrix in biofilms (391). Thus, in the absence of the *eps* genes, the substrates accumulate (**Figure 6-7**). While the double mutant Δ*eps*/Δ*tapA* does not show strong correlations in the 2D ^{13}C-^{13}C double quantum J-INADEQUATE, peaks with the highest intensity are assigned to D-α-pyranose saccharides. We report a strong accumulation of ribose, pyranose saccharides, pyrimidine bases, thus strongly suggesting UDP-galactose and UDP-glucose to be active monosaccharides used during EPS production. Since UDP-galactose is quite lethal to *B. subtilis*, it is more likely that UDP-glucose is accumulated in the absence of *eps*-related enzymes.

We also note the abundance of random-coiled or monomeric amino acids within the matrix such as valine (394), arginine (385-389), leucine, and isoleucine. The roles of such specific residues have yet to be determined. In 2019, Pisithkul *et al.* reported a full analysis of the metabolism remodelling during the first two days of biofilm development (395). They observed a strong accumulation of tryptophan and histidine over time, correlating with the strong NMR signal we observe. Valine, arginine, leucine, isoleucine and lysine display an interesting abundance variation over time. As our samples assess the final state of biofilm formation, the high molar concentration of theses amino acids as opposed to any other might be important in the general understanding of biofilm development and more specifically its sustainability. In the Pisithkul *et al.* study, UDP-Glc was more abundant after 16 hours and later dropped, probably owing to the incorporation of such activated saccharides into exopolysaccharides, and

explaining their accumulation in Δ*eps*, Δ*eps/tasA* and Δ*eps/tapA* mutants. In some bacteria and in mixed species contexts, valine is secreted to prevent the development of other valine-sensitive bacterial species, such as certain *E. coli* (396). Furthermore, racemase activity has been linked to a natural biofilm disassembly process leading to the accumulation of D-leucine, D-methionine, D-tyrosine, and D-tryptophan (397).

Interestingly, using conventional SSNMR at 11 kHz MAS, WT biofilms were found to display very little signal in CP-based experiments, thus highlighting the overall lack of rigidity and order in the biofilms. More intense signals seemed to be observed in the protein area of the 1D ^{13}C CP spectrum, reminiscent of those of a recombinant TasA amyloid fibrils sample (**Figure 1**). Such low rigidity hindered the possibility of applying multidimensional CP-based NMR. The differences in overall dynamics observed in the four *B. subtilis* strains (**Table 2**) highlight the role of EPS and protein networks in water accessibility within the biofilm communities. Many reports have focused on the characterization of bacterial biofilms from a soft matter perspective. *B. subtilis* WT biofilms are regarded as colloidal hydrogel in which the bacteria are composites of colloids embedded within a polymeric cross-linked network of exopolysaccharides and protein fibrils (398-400). Moreover, the biomass only accounts for 20-30% in dry weight. Chai and co-workers reported the water content in WT, Δ*eps*, Δ*tasA* biofilms. ~80% of water in found in WT and Δ*tasA* biofilms, compared to ~75% in Δ*eps* biofilms (399). The authors surprisingly state that all water is bound (*i.e.*, to molecules, solutes, etc) as no free water was found. In absence of EPS and TasA, very few differences are observed in water states and content thus suggesting that saccharides and proteins have little to do with water distribution. Further analyses of salts, water, and culture conditions imply that biofilms accumulate salts and molecules from the medium and that such solutes are solely responsible for water binding in laboratory conditions. However, swelling experiments in which water uptake is analysed show that EPS and TasA are crucial for maintaining water inside the biofilms (399).

4.6. Conclusion

Our analyses of the differences between *B. subtilis* biofilms show the importance of EPS and the proteins TasA and TapA in the overall structural dynamic of the bacterial community. We noticed an accumulation of monosaccharides such as pyranose and ribose, which strongly point to UDP-Glc as the building block for EPS biosynthesis. The Δ*eps*, Δ*eps-tasA*, and Δ*eps-tapA*

mutants show different dynamical properties with higher order and rigidity. Using SSNMR, we demonstrate high abundance of specific amino acids such as valine, arginine, and isoleucine within the biofilms in WT and mutant strains. Our results, combined with the general focus on *B. subtilis* as a model for understanding biofilm formation, highlight the many roles of EPS and functional amyloids in water accessibility, biofilm resistance to mechanical stress, overall dynamics, saccharide content, etc. Interestingly, the double mutants Δ*eps-tasA* and Δ*eps-tapA* indicate a synergic collaboration of EPS and TasA/TapA, as changes in Δ*eps* mutant biofilms are exacerbated in the Δ*eps-tasA* and Δ*eps-tapA* double mutants.

Chapter 5: The inner rod proteins of the type III secretion system in *Salmonella* and *Pseudomonas* strains

5.1. Context

Type 3 secretion systems (T3SSs) are in the centre of Gram-negative bacterial ability to directly deliver effector proteins from their cytoplasm to that of eukaryotic cells, through a most conserved battery of homologous proteins required to assemble a nanomachine, also called the "injectisome" (401). This appendage passes through the bacterial barrier composed of plasma and outer membranes, peptidoglycan layer and the extracellular space. T3SSs were first observed in 1998 by Galán, Kubori and co-workers (402), who described a supramolecular assembly spanning the inner and outer membranes of flagellated and non-flagellated Gram-negative strains of *Salmonella typhimurium*. The first observation of the T3SS needle filament by electron microscopy opened the way for structure elucidation of this and several other T3SS subcomplexes (7, 403, 404).

T3SSs are composed of several subcomplexes: a cytosolic ATPase complex, a cytoplasmic ring (C-ring), an inner membrane export apparatus, a basal body (both in inner and outer membranes) encircling an inner rod from which arises the needle and a translocation pore targeting the host cell membrane (**Figure 1C**). The T3SS needle is an extracellular protein filament, tens of nanometres in length and with an outer diameter of ~8 nm. Several structural models described an inner diameter of ~2.0-2.5 nm (7, 405-407), however a recent cryo-EM

model by Strynadka and co-workers showed a smaller inner lumen with a 1.5 nm-diameter (403).

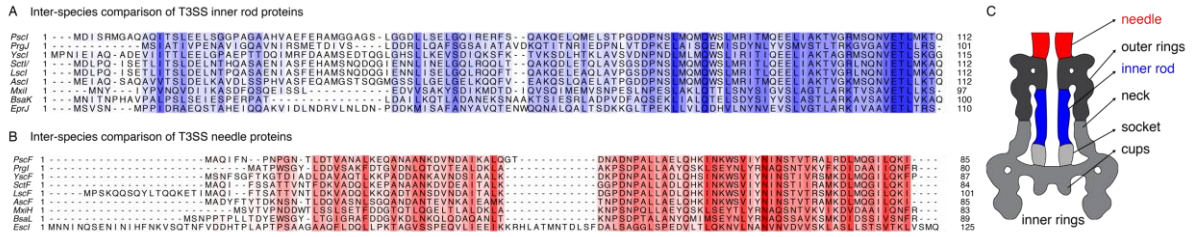


Figure 5-1: Both the inner rod and needle sequences are highly conserved. Sequence alignment of inner rod (A) and needle (B) proteins of the type three secretion system from Gram-negative bacteria. Schematic representation of the *Salmonella* T3SS displaying the general arrangements of all T3SS components (C).

Understanding the T3SSs structure could lead to better understating of their cellular activities and delivery pathways in Gram-negative bacteria. Although most structural components of the injectisomes are known, the lack of intel about their interactions within the nanomachine prevent the complete appreciation of how proteins enter the host cytosol, how the sorting platform functions, or even how the overall secretory hierarchy is tightly maintained. The detailed molecular description of these injectisomes, from the basal body to the inner rod and the tip of the needle in the host cell cytoplasm, might help elucidate the effector delivery mechanisms, and more specifically the role of the needle and the inner rod in this process. Both SSNMR (7, 168, 408-410) and, more recently, cryo-EM (403, 411, 412) have proven extremely helpful in the elucidation of the mysteries behind this fascinating machinery.

The inner rod is found within the basal body which is a conduit for protein through periplasm (**Figure 1C**). It is formed by a unique protein: PrgJ in *Salmonella*, MxiI in *Shigella*, YscI in *Yersinia* and PscI in *Pseudomonas* stains. All inner rod proteins present high sequence homology, more specifically their C-terminal ends are conserved in several bacterial strains (**Figure 1A**), which is also the case for needle proteins (**Figure 1B**). Mutants lacking PrgJ are non-invasive (413) underlining its role in the virulence process. It was recently shown that the C-terminal domain of PrgJ interacts with the receptor NLRC4 in order to manipulate the host inflammatory response (414). Cryo-EM studies of the isolated T3SS base have long only been able to provide the shape of the inner rod (404, 415, 416), however in 2019 the first cryo-EM structures of the inner rod complex were reported (411, 417, 418). Before that, circular

dichroism and solution NMR have shown that monomeric PrgJ in solution lacks tertiary structure (419) and only a short segment of the C-terminal domain is folded in an α-helical conformation. As already observed for the T3SS needle (7), we suppose that drastic conformational changes in the subunit protein PrgJ/PscI accompany its assembly into the inner rod substructure. So far, very little information connects structural properties to functional abilities, and it is precisely one of my thesis goals. We hereafter focus on PrgJ from *Salmonella* and PscI from *Pseudomonas* to characterize the inner rod structural order and *in vitro* behaviours. PrgJ structure within the needle complex being already known, we use SSNMR to characterize and probe PrgJ and PscI *in vitro* assemblies in an attempt to compare the two proteins by assessing their local order, general polymorphism levels, sensitivity to self-assembly conditions, mass-per-length value (PscI), flexible contributions. Both proteins self-assemble *in vitro* and form filamentous objects. SSNMR experiments demonstrate a high level of structural polymorphism, as opposed to all needle proteins studied using SSNMR, and both show a rigid C-terminus as confirmed by the assembly of truncated versions of the inner rod proteins, PrgJ(51-101) and PscI(50-112). We also designed chimera combining N-terminal segments of PrgI, the needle protein in *Salmonella*, and C-terminal moieties of PrgJ. We show such constructs to have lost the ability to form homogeneous filamentous tubes thus indicating a strong interaction between PrgI N-terminal and PrgJ C-terminal moieties.

5.2. Methodology

5.2.1. Bacterial material

Bacterial strains are derivatives of either SJW2941 (420) or SL1344 (421) strains. Allelic exchange was employed to generate Δ*prgI*, Δ*prgJ*, Δ*invG*, *malE-prgH*, as previously described (422). *Salmonella* strains were part of Professor Galán's collection. For recombinant protein expression, the strain used in this study is BL21 (DE3)-Gold pLysS, derivative of *E. coli*. All strains were grown in LB medium for expression and purification optimization: 1% tryptone (Oxoid), 0.5% yeast extract (Oxoid) and 0.5% NaCl. Media were solidified by addition of bacteriological agar (Oxoid) to a final concentration of 1.5%. M9 minimal medium (composition for 1 liter was: 48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.01 mM $ZnCl_2$, 0.001 mM $FeCl_3$, 0.1 mM $CaCl_2$, 18.7 mM NH_4Cl, 0.20% ^{13}C labelled carbon source and 10 mL of 100x MEM vitamins solution (Merck)) was used for protein expression for NMR studies. M9 was enriched in $^{15}NH_4Cl$ and ^{13}C labeled carbon sources ([U-^{13}C] D-glucose, [2-^{13}C] glycerol, [1,3-^{13}C] glycerol) in order to produce labeled recombinant proteins.

The antibiotic used was kanamycin and its final concentration was 50 μg/mL. Genes were inserted in pET24 plasmids with kanamycin resistance. 6-His tags were introduced at the N-terminal ends of PscI and PrgJ, and at position 251 in SpaR for the *spaPQR* plasmid.

5.2.2. Cell growth and protein expression

A single freshly transformed BL21(DE3) pLysS *E. coli* was picked and cultured in 10 mL of LB media with 50 μg/mL kanamycin at 37°C under shaking condition for 5 h to reach exponential growth phase. Cells were harvested (1000 g, 10 min) and resuspended in 100 mL of labeled minimal medium (1 g/L $^{15}NH_4Cl$, 2 g/L [U-^{13}C] D-glucose or other carbon source) supplemented with antibiotics and grown overnight at 30°C, 200 rpm until stationary phase. 10% (v/v) of this culture was used to inoculate 1 L of labeled minimal medium supplemented with antibiotics, cultured at 37°C, 220 rpm until OD_{600nm} of 0.8 and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Expression of genes encoding PscI and PrgJ was induced at 30°C for 20 h. Cells were harvested (6000 g, 30 min, 4°C, Beckman Coulter JLA 8.1 rotor) and frozen at -80°C until purification.

5.2.3. PrgJ/PscI purification

Cells were resuspended in 50 mL of lysis buffer (100 mM $NaPO_4^{2-}$, 10 mM Tris, 8 M urea, pH 8.0) and stirred for 48 hours at room temperature. Cells were lysed by sonication on ice and lysate was clarified at 33 000 g during 30 min at 16°C. The supernatant was stored for further affinity chromatography purification. Solubilized inclusion bodies were loaded on a HisTrap HP 5 mL column (GE Healthcare) previously equilibrated in the binding buffer (50 mM Tris, 0.5 M NaCl, 20 mM imidazole, 8 M urea, pH 8). Proteins were eluted from the column with a linear gradient of elution buffer (50 mM Tris, 0.5 M NaCl, 500 mM imidazole, 8 M urea, pH 8). After the affinity chromatography step, proteins were loaded on a HiPrep 26/10 desalting column (GE Healthcare) to exchange buffer for 20 mM pH 8.5. The 6-His tag was cleaved using TEV protease at 3/100 (w/w) ratio in cleavage buffer (20 mM MES, 1 mM DTT, 0.5 mM EDTA, 200 mM PMSF, pH 8.5) and agitated 48h at 4°C. 4 M of urea was added to solubilize proteins and increase separation of the different components. The TEV mix was loaded into a HisTrapTM 5 mL column (GE Healthcare Life Science©). The cleaved protein was recovered in the flow-through fraction and concentrated at 2 mg/mL using centrifugation filter units (Amicon Ultra, Merck Millipore) of 2 kDa cut-off (after removing urea to avoid protein loss). The concentrated sample was loaded on a HiLoadTM 26/600 SuperdexTM 75 prep grade size exclusion column (GE Healthcare Life Science©) in a buffer containing (20 mM MES, pH8.5).

The monomer-containing fractions were dialyzed against 20 mM MES, pH5.5 and left to self-assemble for 73 hours.

5.2.4. The SpaPQR complex purification

The protocol was adapted from a report of Kuhlen *et al.* in 2018 (412). Briefly, the cells were resuspended in buffer A (100 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 8) and lysed using a pressure-based system. Cell debris were removed by centrifuging at 21,000g for 25 min. Cell membranes were isolated from the cleared lysate by ultracentrifugation at 45,000g for 1h. The obtained membranes pellet was put in buffer A with detergent DDM 1% (w/v) and gently stirred for 1 hour at 4°C. The remaining insolubilized material was handled using manual homogenization with a Dounce Grinder system. Ni-NTA beads were added and the mix was incubated overnight at 4°C. The beads are washed with 50 mL of buffer A with 0.1% (w/v) detergent and 20 mM imidazole and the complex of interest was eluted in 200-300 µL of buffer A with 0.02% DDM and 300 mM imidazole. After the size-exclusion chromatography (S200 10/300 increase, GE Healthcare) step, two peaks were retrieved (elution volumes 8 mL and 10-11 mL). The one at 11 mL was the expected peak corresponding to ~191 kDa expected to contain 6 SpaP, 1 SpaR, and 4 SpaQ molecules. For homogeneity purposes, only the leading edge go the "11 mL" peak was kept. The sample was concentrated and kept at 4°C for two weeks without losing the complex stability.

5.2.5. Isolation of the needle complex substructures

The needle complex purification from the different *S. Typhimurium* strains was developed in Professor Galán's laboratory and the following text has been reported in 2019 (423). Briefly, an MBP-tagged PrgH and a Δ*invG* allele (to remove the outer rings and neck of the needle complex) were introduced into the *S. Typhimurium* strain. Two litres of LB containing 0.3 M NaCl, 100 µg/mL of ampicillin, and 0.1% arabinose were inoculated with the different strains and grown for ~10 hours under gentle (100 rpm) shaking. Cells were recovered by centrifugation at 6,000 rpm for 20 minutes, resuspended in 10 mL of lysis buffer [200 mM Tris pH 7.5, 20% sucrose, 1 mM EDTA, 0.25mg/mL of lysozyme and cOmplete™ EDTA-free protease inhibitor cocktail (Sigma 4693159001)] and incubated on ice for 1 hour. Cells were subsequently incubated for 5 min at 37ºC and lysed by the addition of 0.5% N-Dodecyl ß-D-maltoside (DDM) (Anatrace D310S). Cells were incubated at 37ºC for additional 5 to 10 min while monitoring lysis, transferred to ice, and further incubated for 1 hour. Debris were removed by centrifugation at 14,000 rpm for 1 hour and the clarified lysate was transferred to a fresh tube. Two hundred microliters of amylose resin (NEB E8021) were added and the

suspension was incubated O/N at 4ºC under rocking conditions. Beads were then washed 4x with 10 mL of washing buffer (20 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA) and finally resuspended in 50 µL of washing buffer containing 20 mM maltose. After 1-hour incubation on ice with occasional tapping, beads were removed by centrifugation (3,000 rpm for 5 min) and the needle complex-containing supernatant was transferred to a fresh tube for further analysis.

5.2.6. Solid-state NMR

Experiments were performed at a MAS frequency of 11 or 60 kHz on triple resonance 4 mm, 3.2 mm or 1.3 mm MAS probes using 600 and 800 MHz ^{1}H Larmor frequency spectrometers (Bruker Biospin). The sample temperature was set to ~276 K and chemical shifts were calibrated using DSS as an internal reference.

The conformational arrangements were probed using two-dimensional ^{13}C-^{13}C correlation spectra using a proton-driven spin-diffusion (PDSD) mixing scheme at 11 kHz MAS. Intra-residue correlations were probed using a PDSD mixing time of 50 ms for acquisition times of 20 ms (direct) and ~6-12 ms (indirect), leading to a total experimental time of ~4 days, per sample. A ^{1}H-^{13}C cross-polarization contact time of 500-1000 µs. Two-dimensional ^{1}H-^{13}C INEPT were recorded using acquisition times of 20 ms in direct and ~6-8 ms in indirect dimensions respectively for a total experimental time of ~24h, per sample.

For proton detection experiments on PrgJ, SSNMR spectra were recorded on a 600 MHz (^{1}H frequency) spectrometer (Bruker Biospin, Germany) equipped with a 1.3 mm triple resonance (^{1}H, ^{13}C, ^{15}N) MAS probe. Sample spinning frequency was set to 60 kHz. For resonance assignment, we used a set of 4 3D ^{1}H detected experiments: (H)CANH [40 scans, 6 ms (t3) x 6.5 ms (t2) x 50 ms (t1)], (HCA)CB(CA)NH [64 scans, 5 ms (t3) x 5 ms (t2) x 50 ms (t1)], (HCO)CA(CO)NH [72 scans, 6 ms (t3) x 4 ms (t2) x 50 ms (t1)], (HCA)CB(CACO)NH [28 scans, 7 ms (t3) x 4 ms (t2) x 25 ms (t1)]. The combination of these experiments allowed the connectivities between intra-residual or sequential CA, CB, CO, N, HN and HA resonances, necessary to perform the entire backbone assignment. Experimental data were processed using TopSpin and analyzed using CCPNMR.

5.2.7. Secondary structure analysis by solid-state NMR

Conformation propensity assessment was determined using secondary chemical shift analysis calculations (*i.e.* differences of secondary chemical shifts $\Delta\delta Csec = \delta C\alpha - \delta C\beta$, with δCx = Cx(experimental) - Cx(random coil - BMRB) or δHx). Positive values for $\Delta\delta Csec$ and negative for δHx therefore are considered to be highly indicative of a α-helical conformation.

5.2.8. Negative-staining electron microscopy

Negative staining micrographs were recorded using one droplet of fibrils at 0.5 mg/mL stained with a 2% uranyl formate (for high pH) (w/v) solution for 1 min and dried in the dark. Samples were observed using a FEI CM120 transmission electron microscope at an accelerating voltage of 120 kV under TEM low-dose mode at a magnification of ~15-40kX. TEM images were recorded using a Gatan USC1000 2k x 2k camera and processed using the ImageJ software.

5.3. Results and Discussion

5.3.1. PscI from *Pseudomonas*

With the purpose of characterizing PscI, we used recombinant overexpression and production in *E. coli*. Our main goal was to obtain a structural model for PscI in its assembled form and characterize its role within the injectisome machinery. First, we characterized the inner rod constituent of *Pseudomonas* in its assembled form using negative-staining EM in order to assess the sample quality and homogeneity prior to SSNMR analysis (**Figure 2A**). The assembly buffer (20 mM MES) used for all native purification steps was thus optimized for SSNMR analysis (see **Figure 2**).

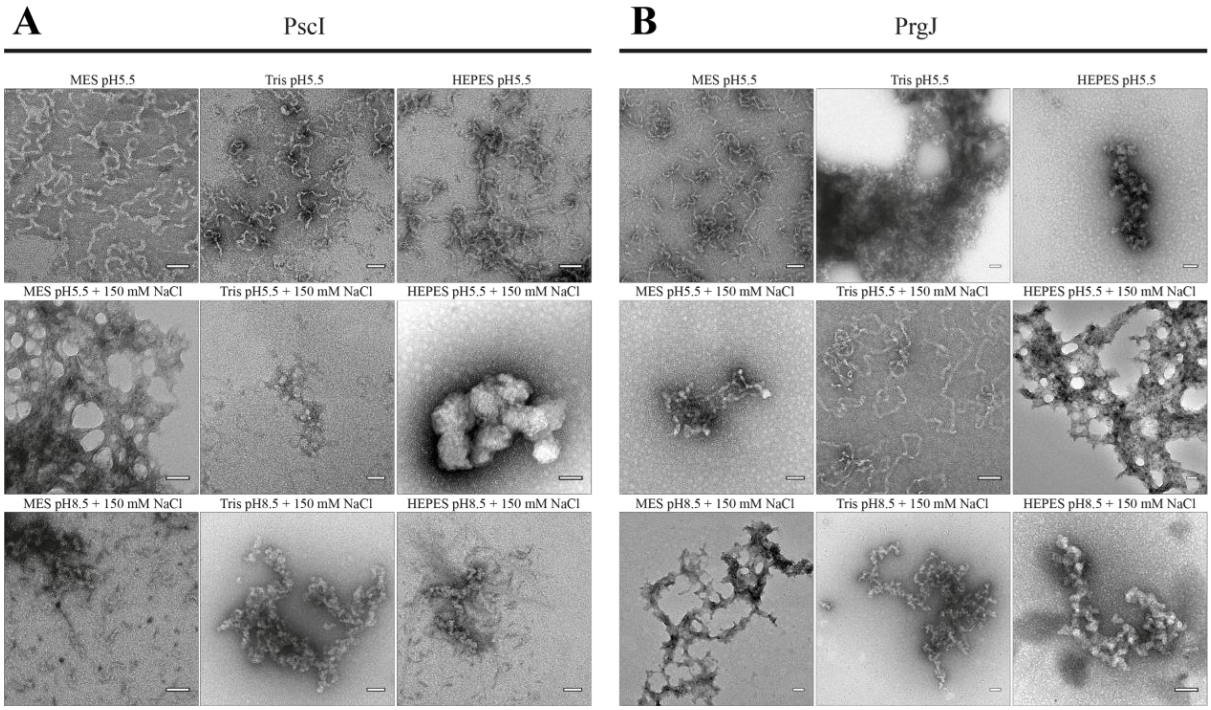


Figure 5-2: Inner rod protein assembly depends on the buffer used to promote polymerization. Electron micrographs of negatively stained samples of PscI (panel A) and PrgJ (panel B), showing assembled or aggregated fibres. [Scale bars, 50 nm.]

As previously observed (424), PscI assembled *in* vitro and formed small tube-like objects displaying a diameter of ~5-10 nm (**Figure 3A**). Such samples met the macromolecular homogeneity standards for SSNMR analyses, and we used an 800 MHz spectrometer (proton frequency) to obtain a 2D ^{13}C-^{13}C spectral fingerprint (**Figure 3C**). This allowed us to assess several features of the nano-objects: rigidity, order at the atomic level, polymorphism, amino-acid composition of the rigid moiety, and secondary structure of the ordered parts. The filaments showed a rather good level of order, while displaying an obvious polymorphism at the atomic level. Further analysis of the full-length construct of PscI using different specific ^{13}C-isotopic labelling strategies were useful in the assignment of a 16-amino-acid-long section of the C-terminal domain (**Figure 3C-E-G**). Such segment is found at the C-terminal region of PscI and thus suggests the C-terminal moiety to be involved in inter-monomer interactions in the rigid core of the type III secretion system inner rod. With this level of polymorphism as observed through the broadening of the peaks, the finite number of rigid residues could not be determined unambiguously. As such, the 16 residues might not be the only ordered segment in PscI filamentous objects.

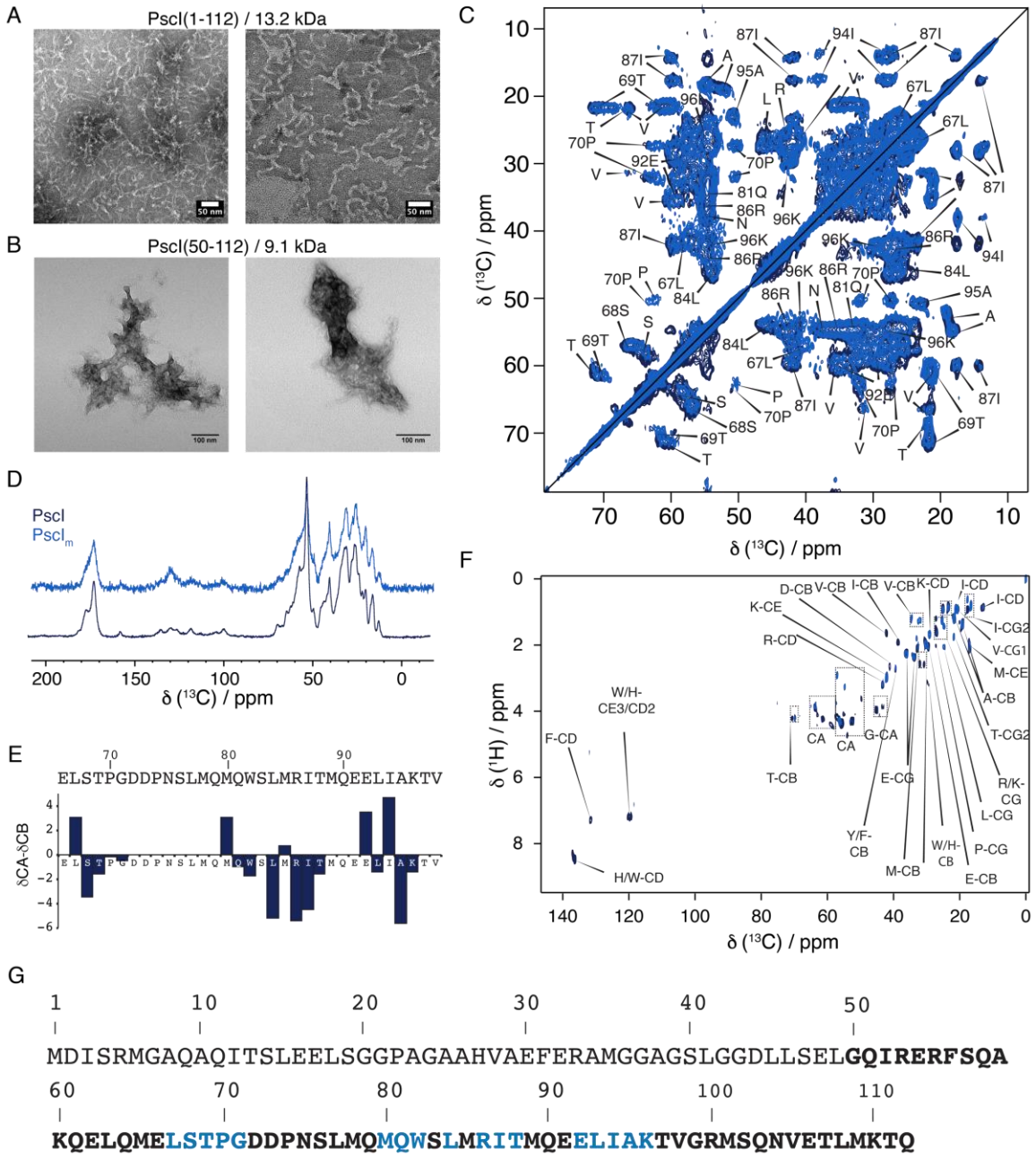


Figure 5-3: PscI display polymorphism in SSNMR spectral fingerprint and its C-terminal moiety is ordered. (A) Negative-staining micrographs of assembled PscI(1-112) in 20 mM MES pH 5.5. (B) Negative-staining micrographs of assembled PscI(50-112). (C) Superimposition of 2D 13C-13C PDSD spectra with a 50 ms mixing time using a 800 MHz spectrometer (proton frequency) of PscI(1-112) in dark blue and PscI(50-112) in light blue; sequential assignment are displayed. (D) 1D ^{13}C CP of PscI(1-112) in dark blue and PscI(50-112) in light blue showing a highly similar fold. (E) SSNMR-based secondary structure assignment of PscI(1-112) as a function of the amino acid sequence. Negative and positive values indicate β-strand or α-helix conformation, respectively. (F) 2D ^{1}H-^{13}C INEPT spectra of of PscI(1-112) in dark blue and PscI(50-112) in light blue. (G) Sequence of PscI(1-112) displaying the assigned amino acids in blue, and the PscI(50-112) moiety in bold characters.

We then engineered and produced a smaller construct of PscI, called PscI(50-112), with the goal of determining the shortest segment displaying the ability to self-assemble while keeping the same structural conformation (**Figure 3B-G**, highlighted in bold characters). This ~60-

amino-acid-long polypeptide comprises the conserved C-terminal moiety. Interestingly, such construct exhibits an almost identical ^{13}C-^{13}C spectral fingerprint compared to the full-length protein (**Figure 3C**) as observed in both the 2D ^{13}C-^{13}C PDSD and the 1D ^{13}C CP spectra (**Figure 3D**) suggesting that the N-terminal part of PscI does not have an important role in the overall structural arrangement. In collaboration with Joe Wall in Brookhaven National Laboratory in Upton, New York, we used scanning transmission electron microscopy to measure the mass-per-length of PscI tubes. Based on 64 recorded micrographs, one of which is shown in **Figure 4A**, an average of 1.36 kDa/Å has been calculated (**Figure 4B**). The detailed statistical analysis shows the number of measurements taken in each of the 64 micrographs (**Figure 4C**), and the average mass-per-length calculated (**Figure 4D**). PscI is a 112 residue-long protein of about ~12.17 kDa. This would suggest that the axial rise per subunit is ~8.9 Å. Interestingly, the axial rise per subunit of needle proteins is on average ~4.2 Å.

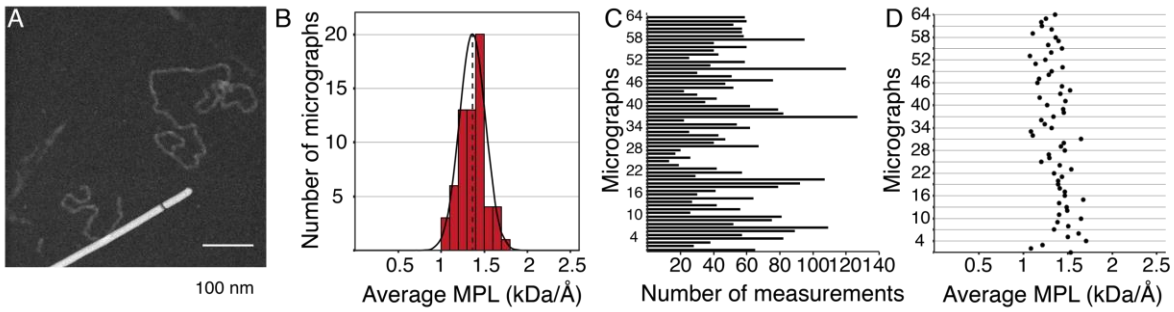


Figure 5-4: PscI mass-per-length analysis based on STEM 64 electron micrographs (one is shown in panel A). An overview is given corresponding to the number of measurements per image (C), the obtained average M/L value per image (D), and the distribution of averages (B) fitted into a gaussian distribution. The average MPL value is 1.36 kDa/Å. (Scale bar: 100 nm)

5.3.2. PrgJ from *Salmonella*

Very similarly to PscI, PrgJ was assessed in its assembled form using negative-staining electron microscopy (**Figure 5A**) and SSNMR. As mentioned above, about ~15 mg of $^{13}C/^{15}N$ labelled PrgJ were used for solid-state analysis, in a 4 mm rotor (**Figure 5C-D**). The 50-ms mixing time PDSD spectrum show a rather similar spectral resolution as demonstrated by a FWHM (*i.e.*, full width half maximum) of about 80-130 Hz for PrgJ. It should be noted that the remaining polymorphism level observed in some areas of the spectra **(Figure 5C-D**) might underestimate the number of rigid residues within the sequences.

Experiments performed with a 1.3 mm probe on a perdeuterated sample of PrgJ were recorded. The 2D ^{13}C-^{15}N projection of a 3D ^{1}H-^{13}-^{15}N hCANH spectrum is shown in **Figure 5E**. In general, broad peaks were observed indicating a lack of high-level order. Using several 3D

sequential spectra, we were able to identify and assign ~10 residues in the N-terminal part of PrgJ (highlighted in **Figure 5F**), as the rest of the NMR signals showed too much ambiguity to be assigned unequivocally. With the same purpose as before, we designed a smaller construct of PrgJ, called PrgJ(51-101). Produced and purified according to the protocol used for the full-length protein, the truncated PrgJ self-assembled into small and straight filaments 20 to 50 nm long and ~6-8 nm in diameter (**Figure 5B**). Such drastic change upon truncation was more precisely analysed using SSNMR. Surprisingly, the 2D ^{13}C-^{13}C spectrum shows the same polymorphism and spectral heterogeneity as that of full-length PrgJ (**Figure 5C**). Moreover, both seem to share the same structural arrangement.

We recorded 2D ^{1}H-^{13}C J-INEPT experiments on both samples and to probe residues that are within a dynamic and flexible regime. In our study, both proteins showed mobile residues and not all residues were present, suggesting the absence of free monomers in the SSNMR rotor. Indeed, the presence of free monomers within the sample would have resulted in the observation of all residue types (*i.e.*, ~20 amino acids) in a flexible state. Our data do not support such hypothesis, as no aromatic residues were found within the INEPT spectra (for both constructs). Such observations also imply that all aromatic residues are in a rigid and ordered state (*i.e.*, two tyrosine residues in the C-terminal end) (**Figure 5G**). The dynamic domain is dominated by amino acid residues that are principally leucine, isoleucine and lysine (**Figure 5G**).

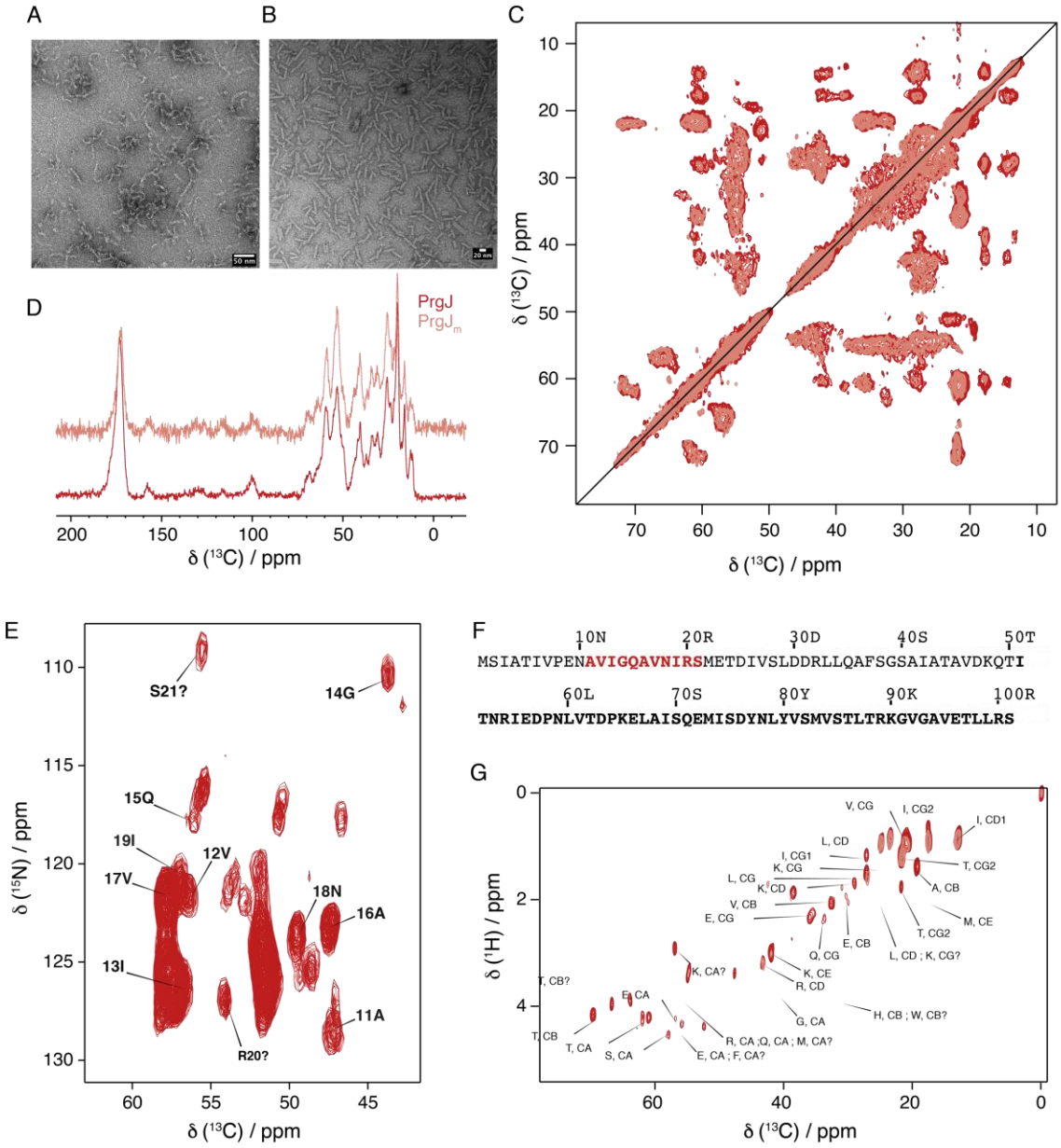


Figure 5-5: PrgJ displays polymorphism in SSNMR spectral fingerprint and its C-terminal moiety is ordered. (A) Negative-staining micrographs of assembled PrgJ(1-101) in 20 mM MES pH 5.5. (B) Negative-staining micrographs of assembled PrgJ(51-101). (C) Superposition of 2D ^{13}C-^{13}C PDSD spectra with a 50 ms mixing time using a 600 MHz spectrometer (proton frequency) of PrgJ(1-101) in dark red and PrgJ(51-101) in light coral. (D) 1D ^{13}C CP of PrgJ(1-101) in dark red and PrgJ(51-101) in light coral showing a highly similar fold. (E) ^{13}C-^{15}N 2D projection of a 3D hCANH spectrum recorded at 60 kHz MAS with a 600 kHZ spectrometer (proton frequency) recorded on of ^{2}H-^{13}C-^{15}N 100% back-exchanged in H_2O PrgJ (1-101). Assigned residues are displayed. (F) Sequence of PrgJ(1-101) showing the assigned amino acids in red, and the PrgJ(51-101) moiety in bold characters. (G) 2D ^{1}H-^{13}C INEPT spectra of PrgJ(1-101) in red blue and PrgJ(51-101) in light coral.

Following the observation that PrgJ(51-101) forms such homogeneous and straight segmented filaments *in vitro* in negative-staining electron micrographs (**Figure 5B**), we designed chimeras of PrgJ which encompass moieties of PrgI primary sequence in the N-terminus. Based on the

sequence alignment of the partner proteins in *Salmonella* (**Figure 1B**), we constructed three chimeras called PrgJ(I-13), PrgJ(I-25), and PrgJ(I-35). PrgJ(I-13) combines the (1-13) residue section of PrgI and residues (18-101) from PrgJ primary sequence. PrgJ(I-25) combines the (1-25) residue section of PrgI and residues (32-101) from PrgJ primary sequence. PrgJ(I-35) combines the (1-35) residue section of PrgI and residues (51-101) from PrgJ primary sequence. The three chimeras interestingly show an inability to form proper filaments under the same conditions as PrgJ. The **Figure 6A-B-C** shows the aggregates formed by the three chimeras self-assembly. Additionally, the figure displays the region of PrgI used to design PrgJ(I-13), PrgJ(I-25), and PrgJ(I-35) in a structural model of PrgI (PDB ID: 2LPZ).

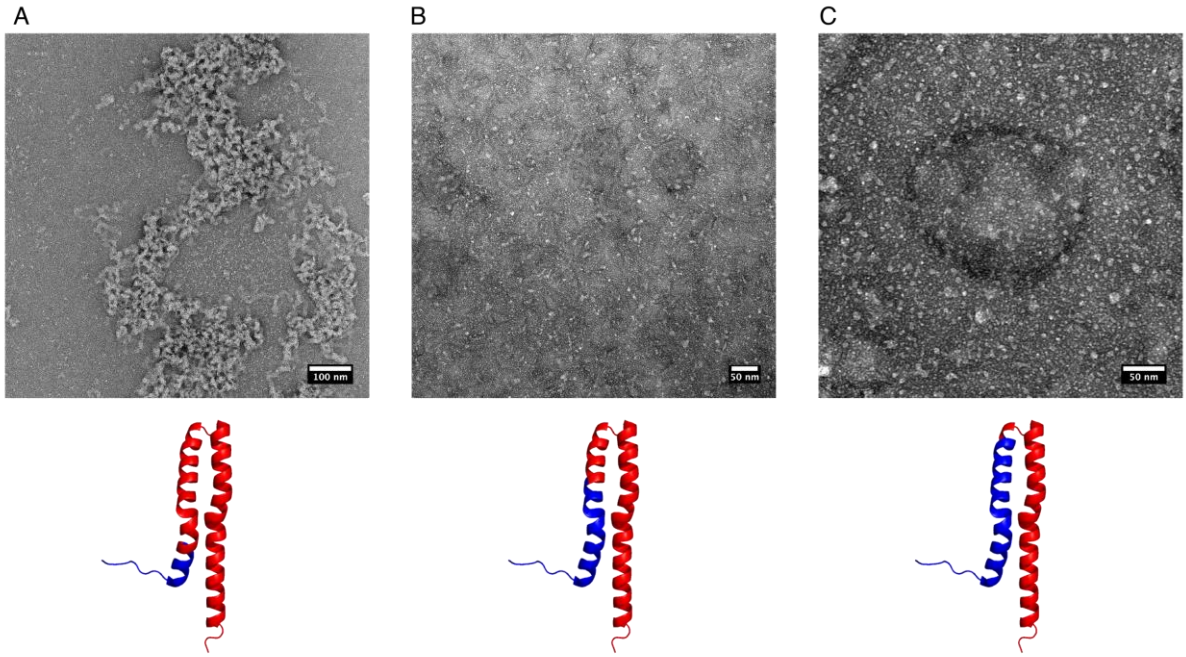


Figure 5-6: Negative staining electron micrographs of the chimeras (A) PrgJ(I-13), (B) PrgJ(I-25), and (C) PrgJ(I-35) with region of PrgI conserved a shown in a structural model of PrgI (the regions coloured in blue were used in PrgJ-I the chimeras based on PDB entry 2LPZ).

We set out to test whether PrgJ would interact with the purified needle complex and export apparatus *in vitro*. Indeed, the needle complexes of a Δ*invG*, Δ*prgI/J Salmonella* strain with a MBP label on PrgH were purified, and separately purified PrgJ and PrgI were added in a 1:3 molar ratio. The resulting electron micrographs are shown in **Figure 7A-B**. Our particles show similar measurements to those reported in the 2019 cryo-EM structure of the type III secretion system (417). Addition of purified PrgJ and PrgI monomers did not seem to alter the general dimensions and morphology of the needle complex. The structure reports a ~24.4 nm width and a ~12.0 nm in height. Based on the measurements of 30 particles, our images seem to correspond to such a description and PrgJ and PrgI do not perturb such standards. The apparent

hydrophobicity of the needle complex, coupled with a difficult control of the orientation of the nano-machines on the EM grids, led us to turn to the SpaPQR export apparatus to study the *in vitro* assembly. The same experiment was conducted, and the recorded images are shown in **Figure7C-D**. 80 export apparatus objects were measured before and after adding PrgJ only. The height of SpaPQR seems to be significantly increased after adding PrgJ. These data seem to suggest that PrgJ can bind to SpaPQR without the rest of the machinery. PrgJ seems to form less fibrillary tubes in the presence of SpaPQR (418, 425) (**Figure 7C-D**). In the absence of additional factors such as chaperones, PrgJ seems to bind spontaneously to the SpaPQR complex, however this need to be confirmed using another approach.

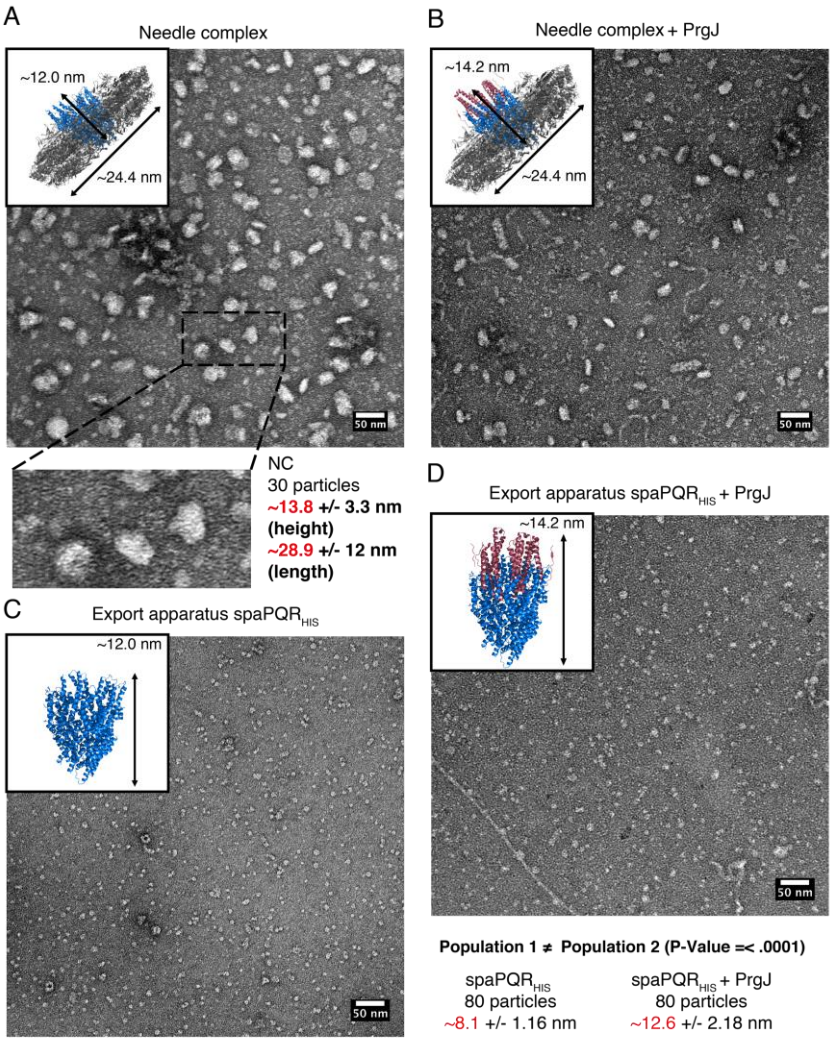


Figure 5-7: PrgJ in vitro interaction assays. (A-B) Negative staining of extracted needle complexes of a Δ*invG*, Δ*prgI/J Salmonella* strain with an MBP label on PrgH without (A) and with (B) purified PrgJ added in a 1:3 molar ratio. Experimental size statistics of needle complex are displayed in panel A. (C-D) Negative staining of purified recombinant SpaPQR export apparatus without (C) and with (D) purified PrgJ. Experimental size statistics of the export apparatus populations are displayed in panel D. In all panels, the structures of the expected sample

composition (*i.e.*, NC with and without PrgJ, and SpaPQR with and without PrgJ) are displayed as reported by Strynadka and co-workers. The structures were generated by PyMOL.

5.4. Discussion

In this study, we report a biophysical analysis of the two inner rods formed by PrgJ and PscI. We compare the two proteins in their assembled form. PscI from *Pseudomonas* and PrgJ from *Salmonella* show similar behaviours: (i) as shown in **Figure 1**, inner rod proteins of several Gram-negative bacteria share a high sequence homology in their C-terminal end; (ii) both PscI and PrgJ have the ability to form fibrils of ~5-8 nm diameter *in vitro* and they form much shorter objects *in vivo* at the base of the T3SS needle; (iii) the structural arrangement of the *in vitro* filamentous tubes of both proteins as probed by SSNMR show a remarkable resemblance (**Figure 8**), and both display a certain level of polymorphism and local disorder.

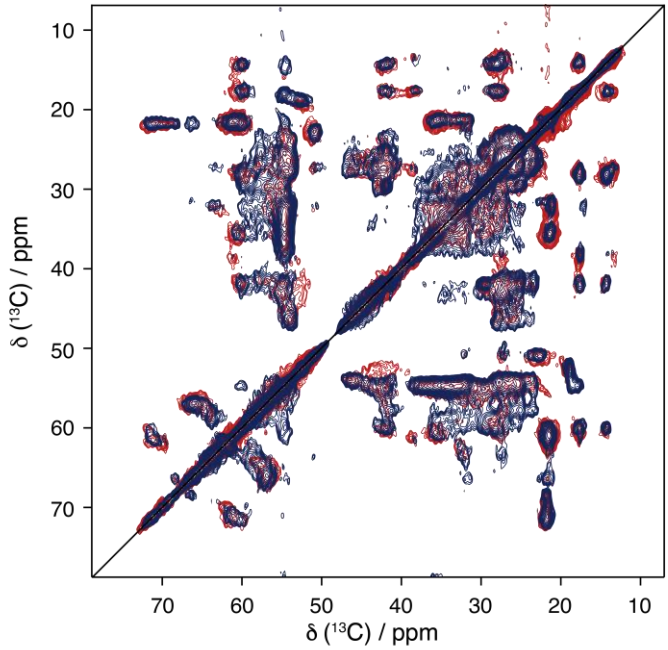


Figure 5-8: Superposition of 2D ^{13}C-^{13}C PDSD spectra with a 50 ms mixing time using a 600 MHz spectrometer (proton frequency) of PrgJ(1-101) in dark red and PscI(1-112) in dark blue.

Another interesting feature is that truncated versions of both PscI and PrgJ, called PscI(50-112) and PrgJ(51-101), show identical ^{13}C-^{13}C spectral fingerprints (**Figure 3, Figure 5**), strongly suggesting that these C-terminal moieties are mainly responsible for the conformation of the T3SS inner rod. In the case of PrgJ, experiments performed using a 1.3 mm probe at 60 kHz MAS and proton detection, showed that the 11 residue-long segment of PrgJ(11-21) indeed encompassed an ordered conformation. Interestingly, studies have suggested the importance of the N-terminus in structural and functional integrity.

Taken together, our results are in line with several reported inner rod structures and models of the T3SS. Inner rod proteins are now known to form the first helical turn that anchors the needle to SpaPQR (417, 426). As reported by Strynadka and co-workers in 2019 (417), five PrgJ subunits are observed for the first time in the injectisome needle complex (structure displayed in **Figure 7A-B-C-D**). PrgJ is seen interacting with a couple monomers of PrgI and the SpaPQR export apparatus. Our results show a stronger order on the C-terminal segment of PrgJ, in agreement with their results. Interestingly, *in vitro* PrgJ assemblies are quite long filamentous objects when compared to their *in vivo* states, suggesting the need of an assembly control. Similar results have been reported for the *Shigella* needle complex (411). The three chimeras designed in this work, PrgJ(I-13), PrgJ(I-25), and PrgJ(I-35), interestingly cannot form proper filaments under the same conditions as PrgJ, which indicates that the strong PrgJ/PrgI interaction hinders or alters PrgJ assembly (**Figure 6**).

Based on previous reports and our limited structural characterization of inner rod proteins, PscI would behave in a very similar fashion in *Pseudomonas,* suggesting that the inner rod structural arrangement is conserved in other bacterial species (427, 428). This would confirm the strong conservation not only of primary sequences, but also of the structural arrangement of needle complexes in T3S injectisomes, and the role of inner rod as a bridge between the export apparatus and the needle (416, 419, 429).

5.5. Conclusion and Perspectives

The inner rod proteins are important in host-pathogen interaction and infection process (Monlezun et al. 2015, Dietsche et al. 2016). Our results confirm and complement the current knowledge of the inner rod arrangement. We show a striking resemblance of PscI and PrgJ *in vitro* assemblies by SSNMR confirming a structural conservation between the *Salmonella* and *Pseudomonas* species. Contrary to the needle filament, assembly of purified inner rod monomers does not lead to homogeneous objects, implying a less robust self-assembly process for the inner rod compared to the needle filament. We speculate that a more stringent assembly process, driven by non-covalent interactions between the inner rod proteins and other proteins of the T3SS machinery, is controlling the early step of the inner rod assembly.

Chapter 6: Structural insights into the 'killer trait' conveyed by supramolecular refractile bodies with solid-state NMR

6.1. Nonstandard abbreviations

INEPT: Insensitive nuclei enhanced polarization transfer

MAS: Magic angle spinning

SSNMR: Solid-state nuclear magnetic resonance

TEM: Transmission electron microscopy

UDP: Uridine diphosphate

R bodies: Refractile bodies

SDS: Sodium dodecylsulfate

CP: Cross polarization

PDSD: Proton-driven spin diffusion

6.2. Context

R bodies (in this study, type 51 refractile bodies) are unusual, ribbon-like protein polymers found in many bacterial species, such as *Caedibacter* and *Pseudomonas*. In 1938, Tracy M. Sonneborn observed that some *Paracemium aurelia* strains were able to kill other sensitive strains of the same species (430-433). This ability was initially linked to an extranuclear factor, called kappa, which was associated with the production and release of a toxin in the medium. Several subsequent studies (434) reported the possibility that the kappa factor had a bacterial behaviour and was thus identified as bacterial endosymbiont, after many years of controversial research (435). In the first 30 years after Sonneborn's discovery, scientific efforts focusing on R bodies provided historically important contributions to the fields of cytoplasmic inheritance, endosymbiosis, and the hereditary/infection dilemma, as documented in *Paramecium* genetic studies between 1945 and 1965 (436).

Several types of symbionts were found in *Paracemium aurelia* and most were observed to release toxic biomolecules in the medium (437). Upon exposing the sensitive strains, several pre-lethal symptoms were identified depending on the bacterial strain: spinning, hump formation, paralysis, vacuolization and rapid lysis (431). Closer examinations of the kappa particles led to the discovery of refractile bodies (aka R bodies) which were found in some bacterial symbionts within the unicellular ciliates mostly found in freshwater environments. These objects were first described as a roll of tape ~0.5 μm in diameter, which could in certain strains fully extend into a ribbon (438). The R bodies were soon linked to many toxicity pathways, as the entry of R bodies-containing bacteria into target *Paramecia* strains somehow led to the killing of "infected" stains. Five classes of R bodies (*i.e.*, type 51, 7, Cc, Pt, and Pa) were described in the literature, based on dimensions, behaviour, and morphology. For instance, the type 51 studied in the following report was originally observed in the strain 51 of cytoplasmic kappa *Paracemium* symbionts. Today it is identified as *Caedibacter taeniospiralis* 51 and the related R bodies are generally 20 μm in length maximum, and 0.4 μm in width. Another main description point is the morphologyof the ribbon termini (*i.e.*, blunt or tapered). In R bodies 51, ribbon termini are tapered and the ribbons seem to unroll from the inside in a telescopic manner. For R bodies 51, the related killing function was described as 'humping' (439). The unrolling of type 51 R bodies is pH dependent and reversible, with an elongated morphology in environments below pH 6.5. In addition to the *Caedibacter taeniospiralis* species, several other kappa strains were reported (*e.g.*, *C. tauniospiralis*, *C. varicaedens*, *C. pseudmutans*, *C. paraconjugatus*, or *C. caryophila*) with numerous free-living species such as

Pseudomonas avenae and *Pseudomonas taeniospiralis*. It is believed that killer *Paramecia* release kappa particles into their environment and sensitive strains are killed after ingestion and toxin release, most likely of a protein nature (439).

Clear evidence unambiguously proved that R body assemblies are non-toxic, and it is now generally considered that R bodies act as pH-dependent toxin delivery devices. R bodies are nowadays among the largest supramolecular biological assemblies ever observed, as these complexes unroll upon pH switch from a 500 nm –diameter coil (*i.e.*, contracted form), into a 165 nm –diameter tube (up to 2 um long) (*i.e.*, extended form). At least four genes are required for the production of R bodies: *rebA, rebB, rebC, rebD*. Little information has been gathered on their biophysical and biochemical features, and no high-resolution structures or conformational models have yet been proposed. Interestingly, R bodies show high resistance to generally denaturing conditions such as 8M urea, 1 to 10% SDS, and up to 5% 2-mercaptoethanol or dithiothreitol, 1-hour treatments at 100°C, 6M guanidine thiocyanate. Boiling in hydrochloric acid was also tested in an unsuccessful attempt to dissociate such unusual assemblies (439). It is thus suggested that covalent bonds other than disulphide bridges between protein monomers are involved in R bodies supramolecular assemblies (440). In 2016, Polka *et al.* recombinantly produced R bodies in *E. coli* and used the machinery as a 'pH tunable' synthetic device for toxin delivery and acidification-dependent release (441). Several point mutations in RebA in the residue area 80 to 100 were designed to modify the pH at which 50% of bodies are extended. A preliminary secondary structure analysis assessed ~55% of helices and ~20% of stands within R bodies using circular dichroism.

We used state-of-the-art SSNMR with cutting-edge proton detection at 100 kHz MAS to study the general morphology and characterize the local order of functional R bodies (178-180, 442). We have succeeded in identifying the R bodies structural core and concluded on the sole presence of RebA and RebB? within the final supramolecular architecture. The secondary structure of the assembled moieties was determined, and we used the server software iTASSER based on NMR data to develop a structural model for both monomeric RebA and monomeric RebB which seem to have helical structures with a helix-loop-helix conformation. We continued to investigate the mechanistic aspects of the pH switch and have pinpointed key residues within the ordered core involved in the massive structural remodelling that R bodies undergo at molecular level. Additional solid-state and solution NMR studies have allowed better understanding on how R bodies can be used for the transport of biomolecules *in vivo*, and how the whole machinery acts as an ATP-independent delivery nanomachine.

6.3. Methodology

6.3.1. Bacterial material

The strain used in this study is BL21 (DE3)-Gold pLysS, derivative of *E. coli*. All strains were grown in LB medium for expression and purification optimization: 1% tryptone (Oxoid), 0.5% yeast extract (Oxoid) and 0.5% NaCl. Media were solidified by addition of bacteriological agar (Oxoid) to a final concentration of 1.5%. M9 minimal medium (composition for 1 liter was: 48 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 1 mM $MgSO_4$, 0.01 mM $ZnCl_2$, 0.001 mM $FeCl_3$, 0.1 mM $CaCl_2$, 18.7 mM NH_4Cl, 0.20% glucose and 10 mL of 100x MEM vitamins solution (Merck)) was used for protein production for NMR studies. M9 was enriched in$^{15}NH_4Cl$ and ^{13}C labeled carbon sources ([U-^{13}C] D-glucose, [2-^{13}C] glycerol, [1,3-^{13}C] glycerol) in order to produce labeled recombinant proteins. The antibiotic used was kanamycin and its final concentration was 50 μg/mL. The plasmid containing *rebA-rebB-rebC-rebD* in a single operon was kindly provided by Dr Justin Kollman (University of Washington, Seattle, WA, USA), and the protocol described below was established in his laboratory.

6.3.2. Cell growth and protein expression

A single freshly transformed BL21(DE3) pLysS *E. coli* was picked and cultured in 10 mL of LB media with 50 μg/mL kanamycin at 37°C with shaking for 5 h to reach exponential growth phase. Cells were harvested (1000 g, 10 min) and resuspended in 100 mL of labeled minimal medium (1 g/L $^{15}NH_4Cl$, 2 g/L [U-^{13}C] D-glucose or other carbon source) supplemented with antibiotics and grown overnight at 30°C, 200 rpm until stationary growth phase. 10% (v/v) of the culture was used to inoculate 1 L of labeled minimal medium supplemented with antibiotics, cultured at 37°C, 220 rpm until OD_{600nm} of 0.8 and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Genes encoding R bodies were expressed at 20°C for 20 h. Cells were harvested (6000 g, 30 min, 4°C, Beckman Coulter JLA 8.1 rotor) and frozen at -80°C until purification.

6.3.3. R bodies extraction

Cells were resuspended in 50 mL of lysis buffer (25 mM Tris, 100 mM KCl, 2 mM EDTA, pH 7.0). 50 μL of freshly solubilized lysozyme solution were added from solution stock at 15 mg/mL, and the whole purification solution was incubated at 37°C for 2 hours. 500 μL of 1 M $MgCl_2$ and 1 M $CaCl_2$ solutions were added, as well as 50 μL of DNase at 15 mg/mL and the lysis bulk was further incubated for 30 minutes at 37°C. 6 mL of 10% SDS were mixed into the solution and span at 4000 g for 20 minutes. The supernatant was removed, and the pellet was resuspended in 50 mL of water. Such washing step was repeated twice, and the final pellet

was resuspended in the storage buffer (25 mM Tris, 100 mM KCl, 15% glycerol, pH 7.5). Samples were either stored at 4°C or flash frozen. Purification fractions were analyzed on 15% Tris-Tricine SDS-PAGE to check purity.

6.3.4. ^{31}P NMR

Liquid-state ^{31}P NMR experiments were carried out on a Bruker Avance III-HD 400 MHz (proton frequency) SB spectrometer (Wissembourg, France) equipped with a 5mm broadband SmartProbe at 25°C. ^{31}P-NMR spectra were acquired at 161.98 MHz by using a one-pulse sequence with proton decoupling during the acquisition ($\pi/2$ pulse width of 8 µs, recycling delay of 10 s, acquisition time of 0.7 s, spectral window of 300 ppm and ~4k scans). Solid-state ^{31}P-NMR spectra were acquired at 161.98 MHz by using a proton decoupled Hahn-echo pulse sequence. Typical acquisition parameters were as follows: spectral window of 400 ppm, $\pi/2$ pulse width of 8 µs, interpulse delay of 40 µs and recycle delay of 5 s and typically 9k scans were accumulated. Phosphorus chemicals shifts were calibrated relative to TPP (triphenylphosphate) in acetone (-17.8 ppm).

6.3.5. Solid-state NMR

Prior to SSNMR experiments, R bodies were washed extensively to remove glycerol from the storage buffer, and packed into NMR rotors after ultracentrifugation at ~50 000 g. Experiments were performed at a MAS frequency of 11 or 100 kHz on triple resonance 4 mm, 3.2 mm or 0.7 mm MAS probes using 600 MHz and 1000 MHz ^{1}H Larmor frequency spectrometers (Bruker Biospin). The sample temperature was set to ~276 K and chemical shifts were calibrated using DSS as an internal reference.

The conformational changes upon extension and re-contraction were probed using two-dimensional ^{13}C-^{13}C correlation spectra using a proton-driven spin-diffusion (PDSD) mixing scheme at 11 kHz MAS. Intra-residue correlations were probed using a PDSD mixing time of 50 ms for acquisition times of 20 ms (direct) and ~10-14 ms (indirect), leading to a total experimental time of ~4 days, per sample. A ^{1}H-^{13}C cross-polarization contact time of 500-1000 µs. Two-dimensional ^{1}H-^{13}C INEPT were recorded using acquisition times of 20 ms in direct and ~6-8 ms in indirect dimensions respectively for a total experimental time of ~24h, per sample. Two-dimensional ^{13}C-^{15}N CP-NCA were recorded using acquisition times of 20 ms and ~15 ms in direct and indirect dimensions respectively for a total experimental time of ~24h, per sample. A ^{1}H-^{15}N cross-polarization contact time of 500-1000 µs was used, while the ^{13}C-^{15}N double cross-polarization contact time was 3 ms. The 2D ^{13}C-^{13}C J-INADEQUATE

experiments were recorded with a 2-second recycle delay using 96 scans and acquisition times of 5 ms (indirect) and 18 ms (direct) at 276 K.

For proton detection experiments, SSNMR spectra were recorded on a 23.5 T (1 GHz ^{1}H frequency) spectrometer (Bruker Biospin, Germany) equipped with a 0.7 mm triple resonance (^{1}H, ^{13}C, ^{15}N) MAS probe. Sample spinning frequency was 100 kHz. Spectra were referenced according to 4,4-dimethyl-4-silapentane-1-sulphonic acid (DSS) signals. For resonance assignment, we used a set of ten 3D 1H detected experiments: (HCA)CB(CA)NH [8 scans, 9 ms (t3) x 4 ms (t2) x 20 ms (t1)], (HCA)CB(CACO)NH [8 scans, 9 ms (t3) x 4 ms (t2) x 20 ms (t1)], (H)CANH [8 scans, 10 ms (t3) x 6 ms (t2) x 20 ms (t1)], (H)CONH [20 scans, 10 ms (t3) x 6 ms (t2) x 20 ms (t1)], (H)CO(CA)NH [16 scans, 10 ms (t3) x 6 ms (t2) x 20 ms (t1)], (HCO)CA(CO)NH [8 scans, 10 ms (t3) x 6 ms (t2) x 20 ms (t1)], (HCA)CBCAHA [4 scans, 3.6 ms (t3) x 3.5 ms (t2) x 20 ms (t1)], (H)COCAHA [8 scans, 11 ms (t3) x 11 ms (t2) x 20 ms (t1)], (H)N(CO)CAHA [8 scans, 10 ms (t3) x 5 ms (t2) x 20 ms (t1)] and (H)NCAHA [8 scans, 10 ms (t3) x 5 ms (t2) x 20 ms (t1)]. The combination of these experiments allowed the connectivities between intra-residual or sequential CA, CB, CO, N, HN and HA resonances, necessary to perform the entire backbone assignment. Side-chain proton assignment was performed using a ^{1}H-^{13}C CP-based sequence followed by a WALTZ mixing step (H)CwaltzCH [4 scans, 5 ms (t3) x 5 ms (t2) x 20 ms (t1)]. We used as a starting point the previously assigned CA and HA chemical shifts to correlate unambiguously all the carbons and protons of the side chains step by step. Experimental data were processed using TopSpin and analyzed using CCPNMR.

6.3.6. Secondary structure analysis by solid-state NMR

Conformation propensity assessment was determined using secondary structure assignment calculations (*i.e.* differences of secondary chemical shifts $\Delta\delta$Csec = δCα- δCβ, with δCx = Cx(experimental) - Cx(random coil - BMRB) or δHx). Positive values for $\Delta\delta$Csec and negative for δHx therefore are considered to be highly indicative of a α-helical conformation.

6.3.7. Negative-staining electron microscopy

Negative staining micrographs were recorded using one droplet of 5-fold dilutions of R bodies in the storage buffer in a buffer at pH 7 (10 mM HEPES, 100 mM NaCl, pH 7) or at pH 5 (10 mM MES, 100 mM NaCl, pH 5) stained with a 2% uranyl acetate (for low pH) or formate (for high pH) (w/v) solution for 1 min and dried in the dark. Samples were observed using a FEI CM120 transmission electron microscope at an accelerating voltage of 120 kV under TEM

low-dose mode at a magnification of ~2-5kX. TEM images were recorded using a Gatan USC1000 2k x 2k camera and processed using the software ImageJ.

6.4. Results

6.4.1. *E. coli* expression of R bodies and pH-dependent macroscopic morphologies

R bodies are natural bio-objects that require four proteins to be formed: RebA, RebB, RebC, and RebD (**Figure 1A**). RebA and RebB are the major structural components (439). These four proteins were co-expressed in *E. coli* BL21 (DE3) pLysS using a pETM11 plasmid containing a single operon of four open reading frames (*i.e.*, *rebA-D*) and the whole machinery was purified using lysozyme digestion, DNase treatment, SDS solubilization of membranes, and several low-g centrifugation steps. The refractile bodies display unique sedimentation characteristics and are thus rapidly separated from the rest of the lysate for purification purposes. Interestingly, as reported by Polka *et al.* in 2016, the R bodies solubility depends on their extension state. A neutral pH will allow the structures to sediment after several hours at room temperature, whereas R bodies will remain in solution in a more acidic pH (*i.e.*, pH ~5). Recombinant R bodies seem to be functional and display the expected contraction/extension behaviour as observed by negative-staining electron microscopy: at neutral pH they are coiled ribbon-like objects (**Figure 1B**) that uncoil into hollow tubes with sharp ends when put in low-pH buffers (**Figure 1C**). The average width of the ribbon is ~217 nm (± 7.2 (3.3%) nm; n=21 measurements, 95% confidence level). The assembly seems to display an average thickness of ~93 nm (± 7.9 (8.5%) nm; n=21 measurements, 95% confidence level) when coiled, with a total diameter of ~446 nm (± 17.5 (3.9%) nm; n=10 measurements, 95% confidence level). **Figure 1B** (right panel, top view of a single R body) exhibits the clear layers that form the coiled R bodies, and the micrograph image resolution allows us to assess the layers thickness to ~14 nm (±0.708 (5.11%) (n=30 measurements, 95% confidence level), which would indicate an average of ~6-7 turns per coil. Once extended, R bodies can be up to 4-6 μm long, but not systematically. Considering the respective standard deviations of less than 10% of the mean values, recombinant R bodies seem to be quite homogenous in size and shape.

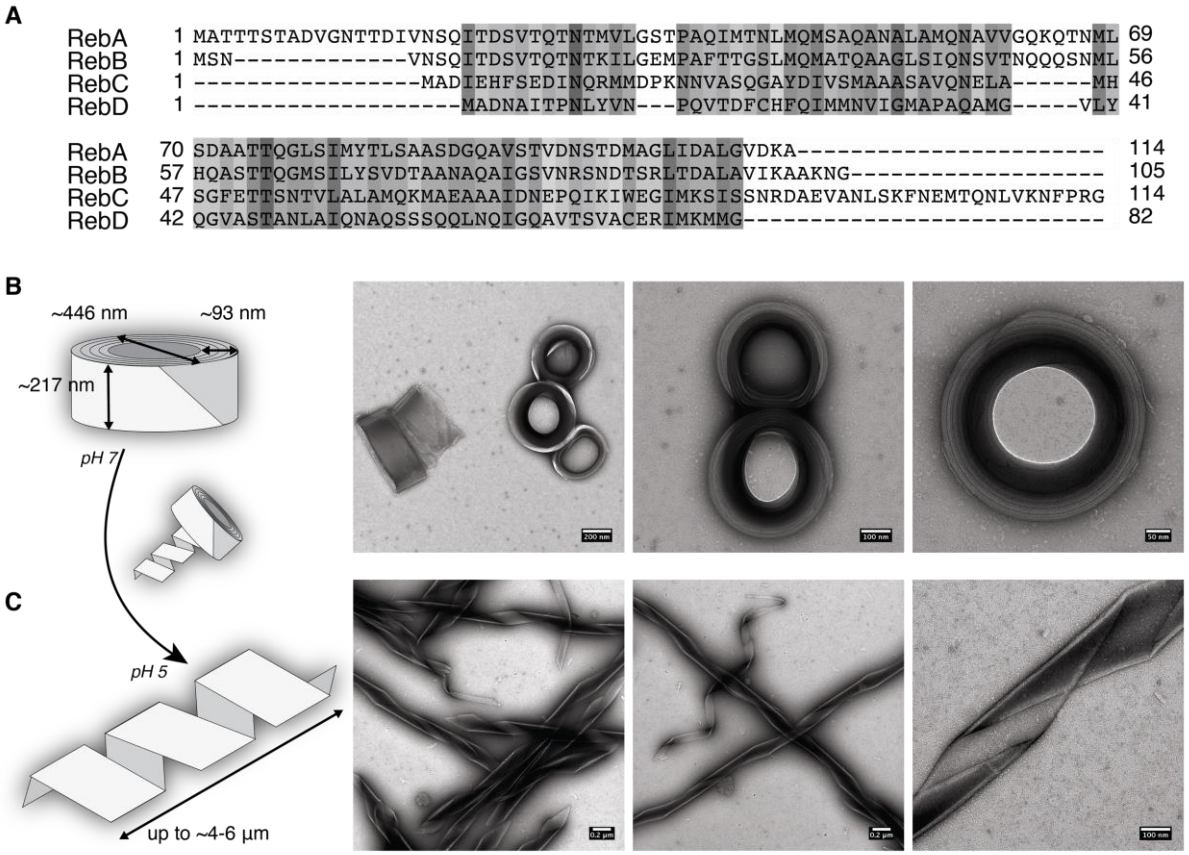


Figure 6-1: (A) Sequence alignment of RebA, RebB, RebC, RebD (generated with Clustal, and represented with Jalview). (B) Morphology schematic and negative-staining electron micrographs of R bodies in HEPES 10 mM, NaCl 100 mM, pH 7. (C) Morphology schematic and negative-staining electron micrographs of R bodies in MES 10 mM, NaCl 100 mM, pH 5.

6.4.2. Very-fast MAS SSNMR characterization of R bodies at pH 7

The sequence homology between RebA, RebB, RebC, and RebD (**Table 1**, **Figure 1A**) can be regarded as a hindrance for chemical shift assignment, as it is at this point unclear which proteins are part of R bodies ordered structure and will thus be detected in SSNMR experiments. Specifically, RebA and RebB share a 53.5% sequence identity. Furthermore, all proteins are isotopically labelled at the same time and thus any acquired spectrum might display a certain level of peak overcrowding. The level of complexity rendered by the potential presence of several proteins strongly favoured the use of 3D SSNMR and more specifically proton-detected very-fast MAS SSNMR (*i.e.*, up to 111 kHz MAS with a 0.7 mm rotor) for such challenging system. In all cases, introduction of an additional dimension probing a different nucleus allows better segregation of signals.

	RebA	*RebB*	*RebC*	*RebD*
RebA	100	53.5	21	32 .9
RebB	X	100	17.6	31.1
RebC	X	X	100	21.7
RebD	X	X	X	100

Table 6-1: Sequence identity percentages.

We recorded a CP-based 3D ^{1}H-^{13}C-^{15}N hCANH and obtained a very well-resolved spectrum, suggesting a very homogeneous structure within the rigid assembly architecture (*i.e.*, the protein structure is well-defined). Such high-quality data usually indicate the occurrence of a single molecular conformation (*i.e.*, no polymorphism at the local and monomeric level) of all components of the rigid core (**Figure 2A**).

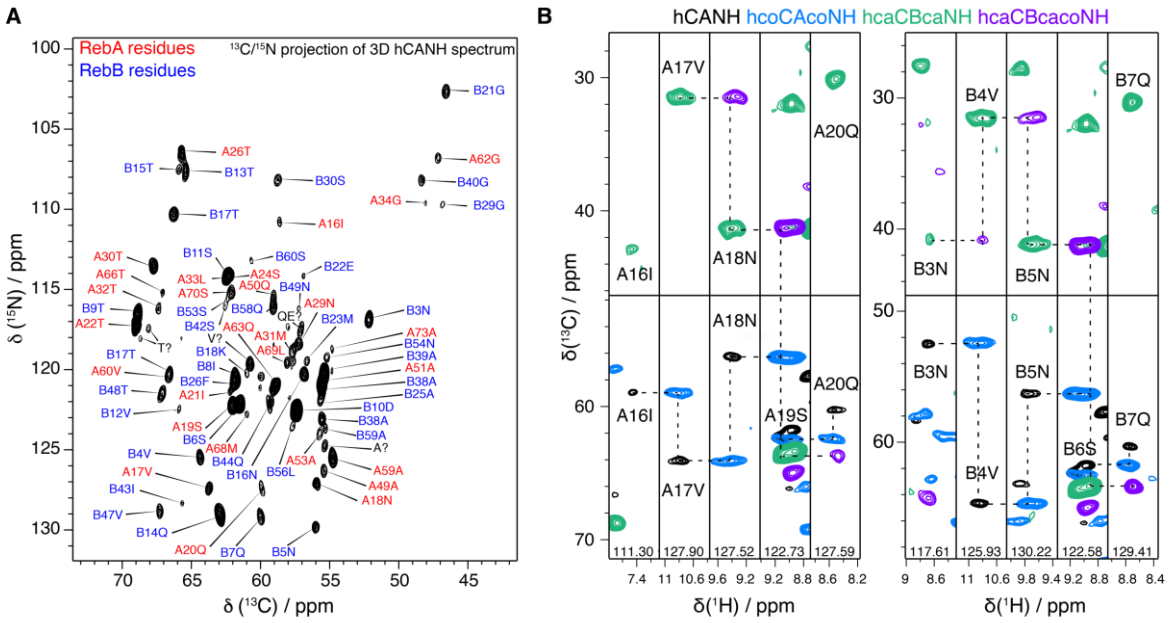


Figure 6-2: (A) ^{13}C-^{15}N 2D projection of a 3D ^{1}H-^{13}C-^{15}N hCANH experiment connecting intra-residual Cα/N/H_N connectivities, recorded on a 1 GHz NMR spectrometer (proton frequency), with a 0.7 mm probe, at 100 kHz MAS. Each peak is the correlation of the backbone ^{15}N with its ^{13}Cα. Assigned residues are shown in red for RebA and in blue for RebB. (B) Extracts of SSNMR spectra for ^{1}H, ^{13}C and ^{15}N sequential assignments (^{15}N chemical shifts are displayed at the bottom of each extract). A combination of hcaCBcaNH (green intra-residual), hcoCAcoNH (light blue, inter-residual), hCANH (black, intra-residual), and hcaCBcacoNH (purple, inter-residual) was used to assign R bodies assemblies at pH 7.

We were able to determine the number of detected spin systems (*i.e.*, residues), identifying their backbone/side-chain carbon chemical shifts. Less than 100 residues seem to be involved in the rigid/well-ordered R-bodies architecture as observed in the 3D ^{1}H-^{13}C-^{15}N hCANH spectrum. Considering that the four proteins sum up to 415 residues (RebA, 114 residues; RebB, 105 residues; RebC, 114 residues; RebD, 82 residues), the entire sequences of all four proteins cannot contribute to the R bodies rigid protein core. However, based on recorded SSNMR data, most NMR peaks show equivalent intensity, which suggest that if more than one protein is indeed detected, it would seem to be equimolar.

All proton-detected spectra were acquired on a 1 GHz spectrometer (proton frequency), using a 0.7 mm SSNMR probe, in collaboration with the group of Dr. Guido Pintacuda (180). The assignment process required the use of intra- and inter-residual experiments: for example, the hcaCBcaNH (green) and hCANH (black) experiments detect intra-residual $^{13}C\beta_i/^{13}C\alpha_i/^{15}N_i/^{1}H_i$ atoms while the hcaCBcacoNH (purple) and hcoCAcoNH (light blue) spectra connect $^{13}C\beta_i/^{13}C\alpha_I$ to $^{15}N_{i+1}/^{1}H_{i+1}$. The assignment of ~86% of the peaks was achieved through the analysis of eight 3D spectra, of which four are hereafter shown (**Figure 2B**). The sequential assignment process is illustrated in **Figure 2B** for the regions $RebA_{16\text{-}20}$ and $RebB_{3\text{-}7}$, which are part of the 14-amino acid long sequence identical in both proteins (*i.e.*, $RebA_{17\text{-}30}$ and $RebB_{4\text{-}17}$). In total, 39 out of 114 amino acids were detected and sequentially assigned in RebA, and 49 out of 105 for RebB. Rigid moieties begin with the regions showing high sequence homology in both proteins and are about 60-residue long (*i.e.*, $RebA_{15\text{-}75}$ and $RebB_{1\text{-}60}$). These assigned regions in RebA and RebB are shown in bold characters in **Figure 3A**. Interestingly, the N-terminal 15-residues of RebA remained undetected under CP conditions. The same was observed for the 39- and 45-residue long C-terminal moieties of RebA and RebB, respectively, for which no NMR signal has been assigned. At this stage, no NMR signal is identified as RebC or RebD amino acids.

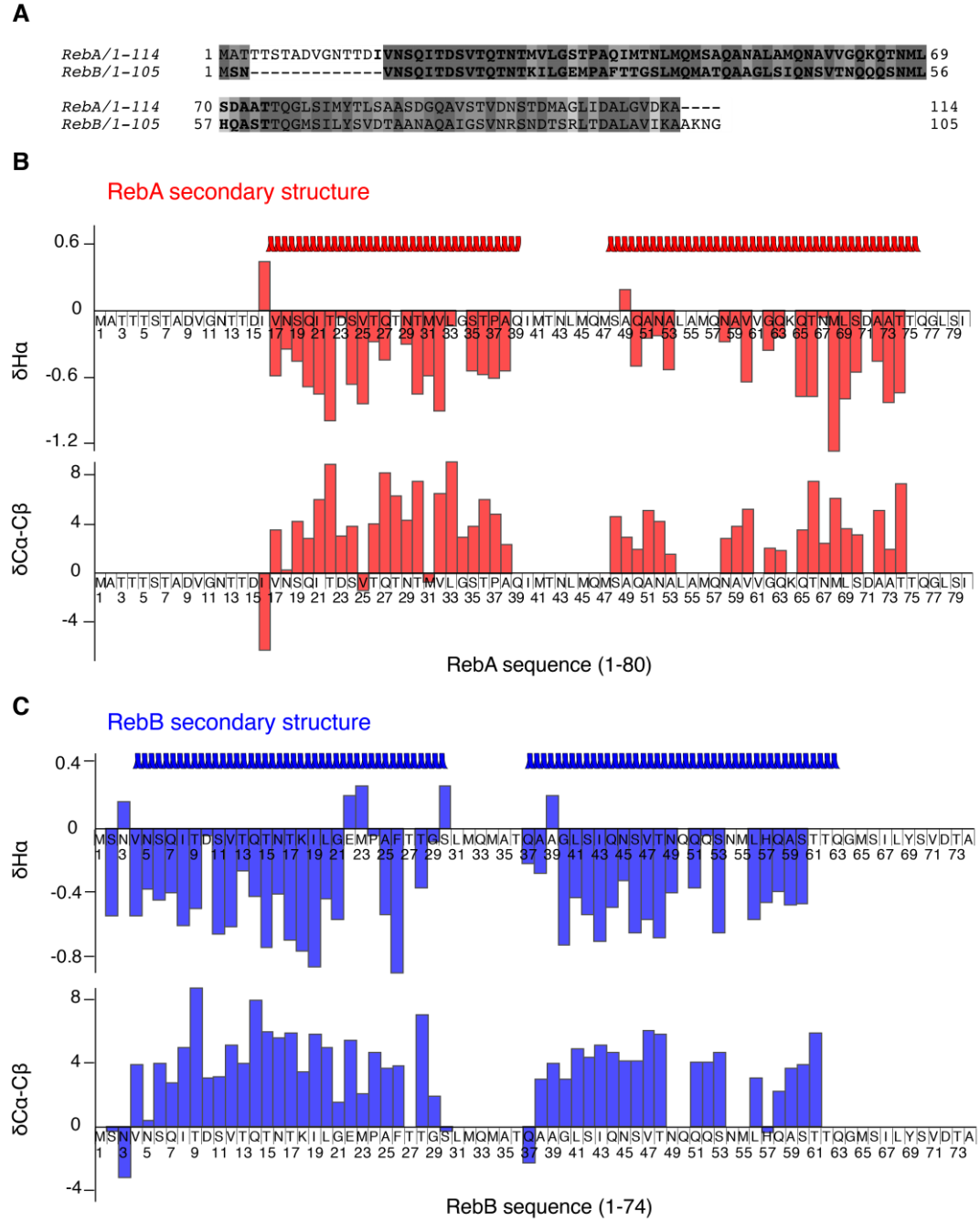


Figure 6-3: (A) Sequence alignment of RebA, RebB (generated with Clustal, and represented with Jalview). For clarity purposes, the assigned residues shown in the Figure 2A are highlighted in bold characters. (B) δ(Cα-Cβ), δHα and secondary structure of assigned residue in RebA (142). (C) δ(Cα-Cβ), δHα and secondary structure of assigned residue in RebB (142).

Such motion-induced invisibility clearly shows that these proteins consist of different regions which are in distinctive order states. Parts of RebA and RebB are clearly more flexible and probably not part of the rigid and ordered core. The remaining ~14% are residues that are

detected in the shown 2D projection of hCANH fingerprint spectrum but are not observed in other sequential spectra, probably due to polarization transfer loss (*i.e.*, local disorder of the amino acids leading to fast relaxation). There are 17 detected and unassigned residues within the spectrum: 1 Ala, 4 Val, 1 Thr, 3 Ile, 1 Leu, 7 Met/Gln/Glu. These residues could belong to the flexible C-terminal regions of RebA and RebB, in which no residue has been sequentially assigned so far, or even RebC or RebD. It should be noted that their initial detection in CP-based experiments unequivocally places them in a somewhat rigid state, as a fully flexible protein residue (dynamic processes in the ns to μs timescale) would only be probed in an INEPT-based pulse sequence.

6.4.3. Monomeric RebA and RebB are composed of helical structural elements

Based on the Hα, Cα and Cβ chemical shift values, the secondary structure of each assigned residue of R bodies at pH 7 was determined. Such empirical calculations are made possible through experimental chemical shift values and the theoretical random coil ones. For this purpose, the δHα values are negative when representing α-helical conformation while the δ(Cα-Cβ) values are positive for the same structural conformation. Most amino acids display an α-helical conformation as observed for both δ(Cα-Cβ) and δHα calculations. RebA and RebB appear constituted of two to three helical structural elements, clearly intercalated with breakers (*i.e.*, δHα values positive or around 0) (**Figure 3B, 3C**). The online server I-TASSER has been used to predict a structural model of RebA and RebB considering the secondary structure propensity as determined using SSNMR, and looking for structure templates within the PDB library and comparing with those found. The resulting models are quite similar, preducting a helix-loop-helix arrangement for the two proteins (**Figure 4A, 4B**). The presence of a single proline, which could play the role of a helix breaker in putative loop regions, corroborate such predictions.

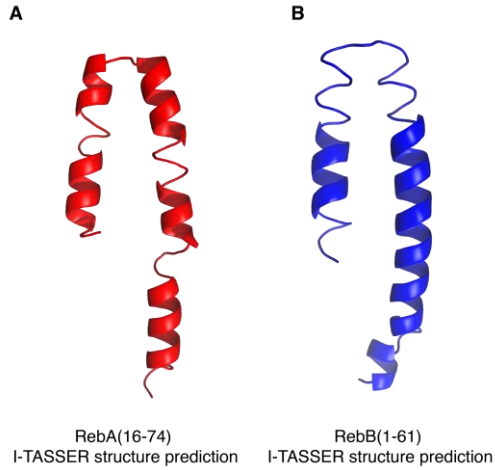


Figure 6-4: (A) Structural model of RebA(16-74) predicted using the server I-TASSER and experimentally determined secondary structures. (B) Structural model of RebB(1-61) predicted using the server I-TASSER and experimentally determined secondary structures.

Detailed structural investigation was furthered using different labelling strategies. Sparsely ^{13}C-labelled proteins display thinner linewidths as a result of the spin dilution phenomenon. Additionally, the crowdedness of spectra of highly complex system, as is the case for our study, can be significantly reduced in an effort to alleviate some spectral ambiguities. Two labelling strategies have thus been employed: ^{13}C-2-glycerol- and ^{13}C-1,3-glycerol-based samples (data not shown). 2D ^{13}C-^{13}C PDSD experiments with a 600 ms mixing time were recorded on an 800 MHz spectrometer (proton frequency) on specifically labelled samples. The longer mixing time allows for the magnetization to be transferred between ^{13}C atoms that are further away in space (*i.e.*, through-space dipolar interactions), as compared to a 25-50 ms mixing time which would only show intra-residual correlations Unambiguous long-range ^{13}C-^{13}C restraints were unfortunately scarce at this time. The spectral ambiguity resulting from the peaks over-crowdedness and the complexity of the R bodies system proved to be quite challenging to overcome. The already obtained distance restraints, from the [2-^{13}C]-glycerol and [1,3-^{13}C]-glycerol labelled samples, seem to agree with those models, however the number of detected residues maintained a certain level of ambiguities with regards to the structural data. Additionally, the aforementioned detection of unassigned residues adds a level of complexity and uncertainty to the analysis of distance-probing experiments. At this stage of the study, the complete structural elucidation of R bodies in their quaternary structures (*i.e.*, the supramolecular arrangements of RebA and RebB monomers and their spatial interaction, if any) remains a mystery. SSNMR which can probe the local order of protein supramolecular

assemblies cannot alone reveal such multipart systems. Additional data using X-ray or cryo-electron tomography could be useful in gathering more insights on these fascinating elements. As a way to reduce the ambiguity of our structural data, we thought of using selective labelling. We have indeed prepared two different R body samples: CH_3-ILV, CD_2H-AI. For the ILV and the AI labelling schemes a single carbon position per Isoleucine (Cδ), Leucine (CδproS), Valine (CγproS), and Alanine (Cβ) is labelled as ^{13}C. Such samples have been used in the purpose of determining more unambiguous distance restraints. A PDSD with longer mixing time (700 ms) is generally used. However, no cross-peaks were observed, probably due to a peculiar dynamic state of RebA and RebB subunits, which renders any long range through-space contacts tricky to probe using SSNMR (data not shown).

6.4.4. R bodies contain non-protein biomolecules

In an attempt to detect and get more insights on the residues lacking in the CP-based experiments, we recorded 2D ^{13}C-^{13}C DP-INADEQUATE and 2D ^{1}H-^{13}C INEPT spectra (**Figure 5A, 5B, respectively**). The 2D ^{13}C-^{13}C DP-INADEQUATE is designed to detect mobile ^{13}C as it uses a combination of ^{13}C DP excitation, ^{13}C-^{13}C J-coupling for polarization transfer, and a short recycle delay of 2 s (*i.e.*, filtering out slow ^{13}C T1 relaxation, typical of ^{13}C atoms lacking motion) to enhance the signals of mobile biomolecules. We observed several spin systems within both spectra, mainly indicating the presence of amino acids, acyl chains, and saccharides in a flexible state:

- In such experiments, the amino acid signals are representative of random coiled protein moieties and are therefore difficult to assign (**Figure 5A, 5B, as shown in purple**). Interestingly, only Ala, Val, Thr, Lys, Leu ^{13}C signals were detected in this 2D ^{13}C-^{13}C DP-INADEQUATE when compared to the 2D ^{1}H-^{13}C INEPT which detects flexible ^{1}H-^{13}C correlations for at least 14 residue types (*i.e.*, which might account for flexible loops, unassigned missing residues in CP-based experiments, and potential unstructured monomers within the SSNMR sample).
- Within the ^{13}C acyl spin systems (*i.e.*, lipid-like molecules), a single unsaturation is detected around 130 ppm (^{13}C chemical shift) and seems to be separated from a $^{13}CH_3$ group by five ^{13}C-^{13}C correlations within the 2D ^{13}C-^{13}C DP-INADEQUATE, thus suggesting a ω-6 acyl chain. On the other hand, the C1-C2 correlation peaks regions (*i.e.*, C1/CO ~177 ppm and C2/CH_2 ~35 ppm) seem to contain more than one conformation, or molecule (*i.e.*, 3-4 peaks). At this stage of the analysis, it is unclear

whether these acyl chains are esterified onto glycerol-based groups (*e.g.*, phospholipids, TAG, DAG, etc).

- Noteworthy is the presence of saccharide signals in both SSNMR experiments shown in **Figure 5A and 5B**. Several carbohydrates (referred hereafter as *CH-Xn*) were detected as inferred by the number of C1-C2 correlations observed. The few carbohydrates had the following chemical shifts (unexhaustive list): *CH-X1* (C1 93.2 ppm, C2 75.5 ppm, C3 72.1 ppm, C4 81.8 ppm, C5 64.4 ppm), *CH-X2* (C1 102.3 ppm, C2 74.3, C3 75.9 ppm, C4 79 ppm, C5 77.2 ppm, C6 63.2 ppm), *CH-X3* (C1 98.4 ppm, C2 76.7 ppm), *CH-X4* (C3/4 79.2 ppm, C4/5 73.7 ppm, C5/6 63.1 ppm). Additional ^{13}C-^{13}C correlations were observed around 73-80 ppm, but these were uncorrelated to C1-like peaks, and were thus ambiguous.

  It should be noted that the sample saccharide content (*i.e.*, number and nature of observed molecules in each sample), as assessed using J-coupling-based experiments, seemed to be the same in distinct samples at pH 7. Indeed, in total three labelled R bodies pH 7 samples were distinctly prepared and compared to assess differences in content (data not shown). *CH-X1* and *CH-X2* are usually partially detected in CP-based experiments (*i.e.*, not all ^{13}C-^{13}C correlations are seen in three different samples) (**Figure 5C**). In a 2D CP-based experiment (**Figure 6** upper panel - experiment introduced in the next *section Structural basis of the pH-dependent macro-morphological switch*, data not shown for saccharide spectral region), the signal intensities of *CH-X1* and *CH-X2* peaks compared with protein signals indicated a proportion of approximately 1:90 – 3:90 (*i.e.*, signals appear with approximately the same intensity for residue types and sugar moieties in the rigid state). Interestingly, *CH-X1* was *once fully* (when only partially detected in the two others) detected in a 2D ^{13}C-^{13}C PDSD, and in such spectrum, additional ^{13}C correlations were observed in the aromatic region thus completing this spin system: C'X 158.8, C'X 152.9, C1 93.2 ppm, C2 75.5 ppm, C3 72.1 ppm, C4 81.8 ppm, C5 64.4 ppm. One might notice that CH-X1 ^{13}C chemical shifts are near expected values of nucleotide carbons in uridine (443): C'1~87-94 ppm, C'2/C'3~70-78 ppm, C'4~82-86 ppm, C'5~63-68 ppm, C2~145-150 ppm, C4~155-165). Interestingly those chemical shifts might be reminiscent of that of ribose and more specifically RNA.

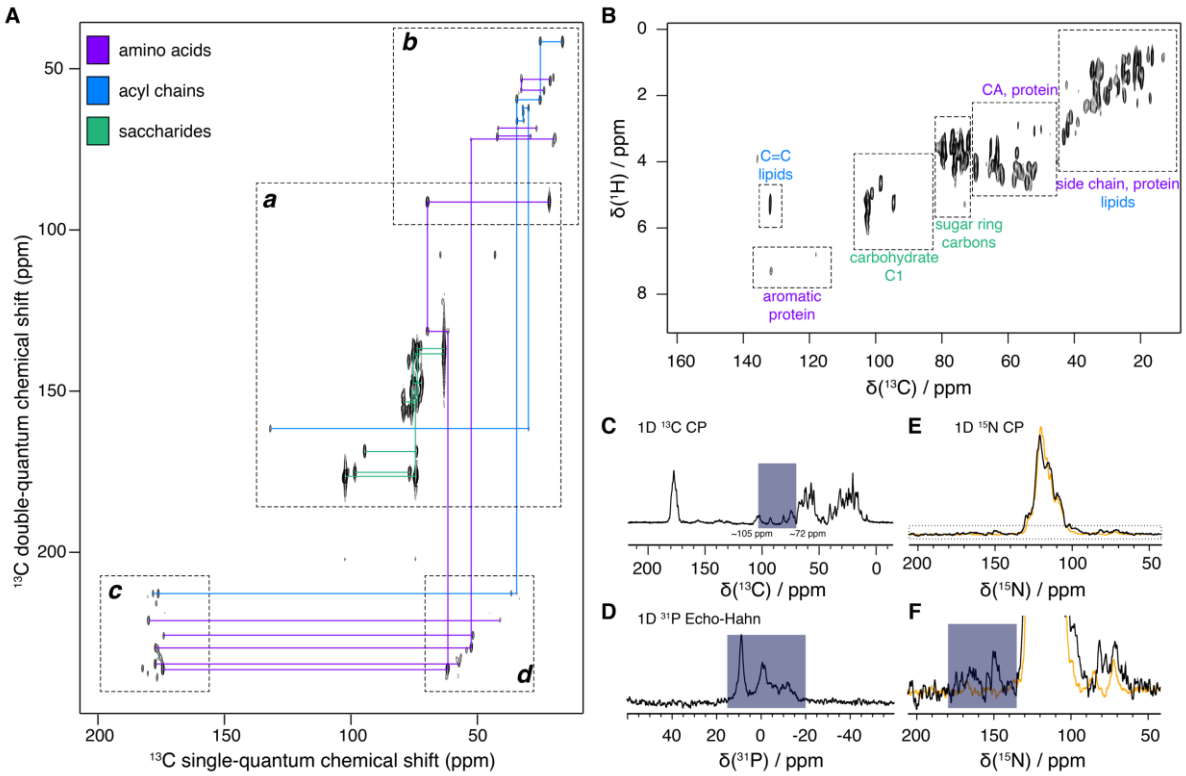


Figure 6-5: R bodies contains different biomolecule types. (A) 2D ^{13}C-^{13}C DP-INADEQUATE recorded on R bodies at pH 7 with a recycle time of 2 s. (B) 2D ^{1}H-^{13}C J-INEPT. For clarity purposes, amino acids, acyls chains, and saccharide peaks and regions are highlighted in purple, blue, and green, respectively. (C) 1D ^{13}C CP experiment of R bodies at pH 7. (D) 1D ^{31}P Echo-Hahn experiment of R bodies at pH 7. (E) 1D ^{15}N CP experiment of R bodies at pH 7 (zoom of area highlighted in dashed square in panel F). For clarity purposes, the spectral areas of interest are highlighted in blue when necessary.

Static ^{31}P NMR was used to assess the potential presence of phosphate groups within the solid sample (**Figure 5D**). Indeed, using a 400 MHz spectrometer (proton frequency), ^{31}P signals were observed using a ^{31}P Echo-Hahn pulse sequence. Interestingly, the detected signals were broad and ranging from ~-20 to ~20 ppm. The selected pulse sequence allowed us to probe the dynamic state of the nuclei, and such large signals would suggest the phosphorus atoms to be in a rather rigid state. The present resolution hinders any detailed assignment of the signals, however sharp peaks at ~-11 ppm, ~-0.4 ppm and ~9 ppm could demonstrate the presence of both phosphate and phosphonate groups within the solid sample. Interestingly, published work about ^{31}P NMR of RNA molecules have reported ^{31}P chemical shifts ranging from ~-2 to ~30 ppm (444). A 1D ^{15}N CP (**Figure 5E**) showed NMR signals ranging from ~110 to ~130 ppm as would be expected for protein amide backbone interactions. Noteworthy is the presence of

^{15}N NMR peaks near 150 ppm (**Figure 5F**), which interestingly could correspond to ^{15}N nucleotide signals (445).

6.4.5. Structural basis of the pH-dependent macro-morphological switch

The differences, from a molecular point of view, between the two forms of R bodies have been studied as well. We have used negative-staining electron microscopy to check the sample quality and homogeneity for isotopically labelled R bodies (**Figure 1B, 1C**). Theoretically, SSNMR allows us to identify any structural change as it would result in a chemical shift perturbation from an atomic point of view (*i.e.*, chemical shift perturbation) or a difference of polymorphism levels from a molecular point of view (*i.e.*, different peak width of peak intensity). Three samples were prepared from a common fully ^{13}C/^{15}N-labelled refractile bodies extracted form *E. coli* as previously described and using MES pH 5 (10 mM, NaCl 100 mM), and HEPES pH 7 (10 mM, NaCl 100 mM) as final buffers: R bodies pH 7, pH 5, pH 7$_{back}$. At this stage of the analysis, it should be noted that each sample was extensively washed with the proper buffer and pelleted several times prior to packing in a rotor.

We first recorded 2D ^{13}C-^{13}C PDSD spectra with a 50 ms mixing time allowing us to probe all intra-residual ^{13}C-based interactions (**Figure 6A-B-C**). The *general* structural arrangement seems to be kept in the three stages of the unfolding/folding process as very similar ^{13}C-^{13}C spectral fingerprint are shown. It would seem that R bodies retrieve the same structural arrangement when folding back, as both pH 7 and pH 7$_{back}$ samples shown highly similar spectral fingerprint. Most differences are indeed observed at pH 5. The first obvious difference is the loss of all carbohydrate spin systems at such pH (saccharides spectral region is highlighted in blue in **Figure 6A-B-C**). The second observation is of structural nature.

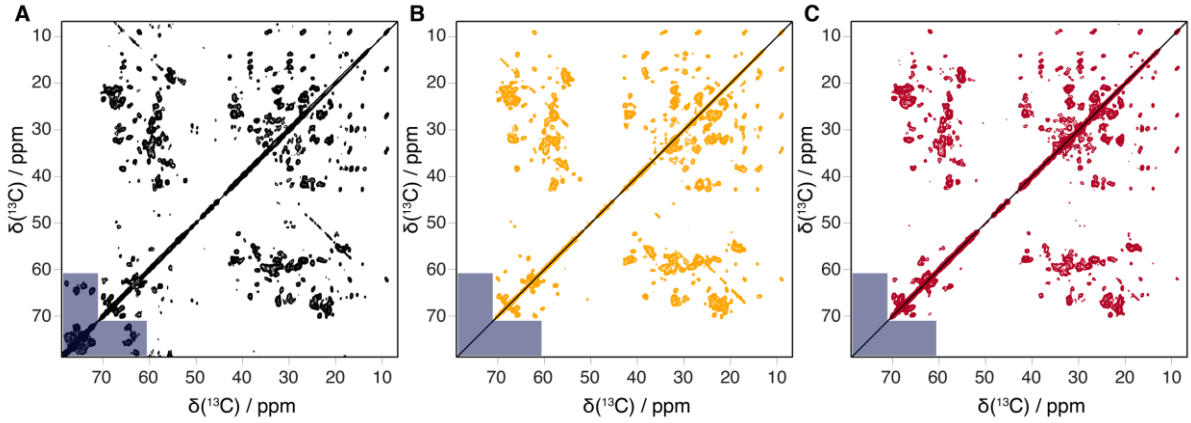


Figure 6-6: R bodies keep the same general structural architecture as they extend at pH 5, and contract again at pH 7. 2D ^{13}C-^{13}C PDSD recorded on R bodies at pH 7 (A, black), pH 5 (B, dark yellow) and pH 7$_{back}$ (C, dark

red) with a 50 ms mixing time. For clarity purposes, the spectral areas of interest are highlighted in blue when necessary.

Not considering chemical shift perturbation and peak doubling, we note the detection of previously unobserved ordered amino-acids at pH 5 (additional protein peaks are circled in **Figure 7A**) when comparing in greater details R bodies 2D ^{13}C-^{13}C spectral fingerprints at pH 7 and 5. A further analysis of the spectrum acquired at pH 5, and a preliminary identification of the amino acids only detected in this sample led to the hypothesis that the N-terminal moiety of RebA is involved in the structural switch between the two pH values. Indeed, the new residues were 1- 2 Ile, 1-2 Ala, 1-2 Thr, 2 Ser, 1-2 Asp or Asn, and 1-2 Val. The 15-aminoacid long region of RebA N-terminal could correspond to such a description (**Figure 3A**). In their report od 2016; Polka *et al.* discuss the possibility that the structural change is based on the motion of a α-helix in RebA as our data seem to suggest (441).

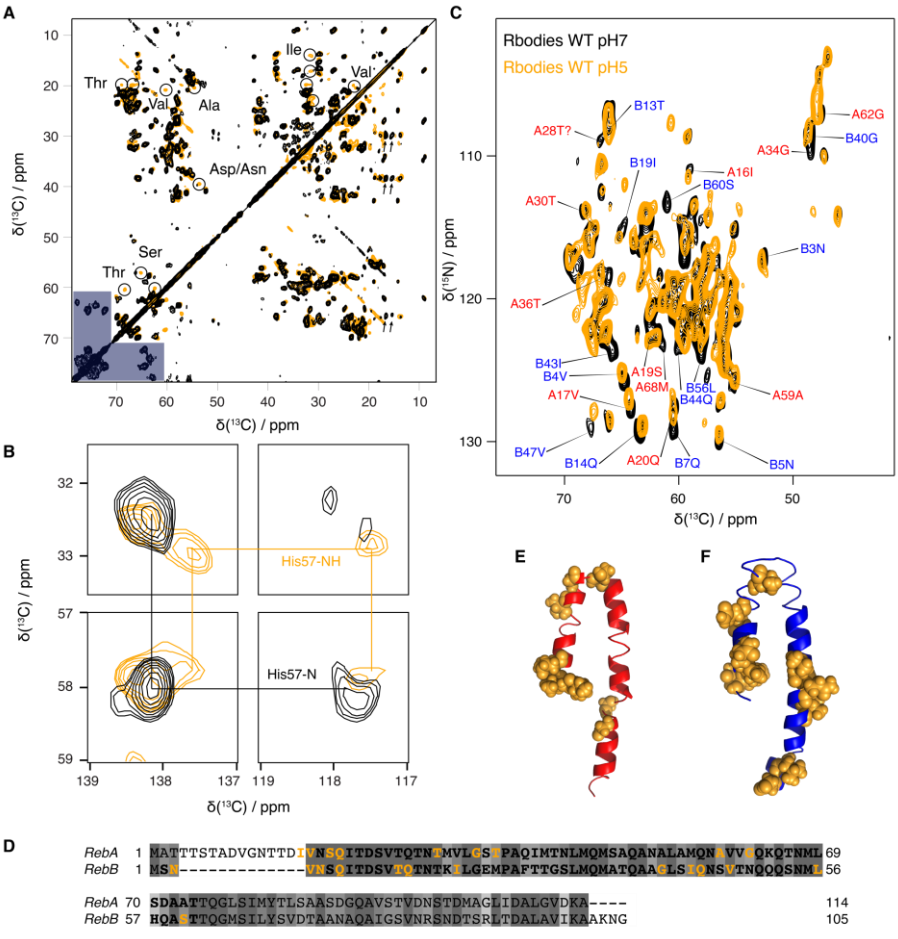


Figure 6-7: (A) Superimposed 2D ^{13}C-^{13}C PDSD recorded on R bodies at pH 7 (black), pH 5 (dark yellow) with a 50 ms mixing time. For clarity purposes, the spectral areas of interest are highlighted in blue when necessary. Peaks of interest are circled in black, are highlighted with black arrows. (B) Excerpts from the 2D ^{13}C-^{13}C PDSD SSNMR experiments of R bodies at pH 7 (black) and pH 5 (dark yellow) displaying the spectral areas for the residue RebB-H_{57}. (C) Superimposed 2D ^{13}C-^{15}N NCA SSNMR experiment recorded on R bodies at pH 7 (black), pH 5 (dark yellow). For clarity purposes, assignments (RebA in red, RebB in blue) are only shown for peaks of interest. (D) Sequence alignment of RebA, RebB (generated with Clustal, and designed with Jalview). For clarity purposes, the assigned residues shown in the Fig. 2A are highlighted in bold and assigned peaks in panel B are coloured in yellow. (E) Structural model of RebA(16-74) predicted using the server I-TASSER and experimentally determined secondary structures with residues highlighted in panel C coloured on yellow. (F) Structural model of RebB(1-61) predicted using the server I-TASSER and experimentally determined secondary structures with residues highlighted in panel C coloured in yellow.

Additionally, RebB-H_{57} seems to now have different conformations (peak doubling), or to be of different chemical nature (*e.g.*, protonation) (**Figure 7B**). RebB-H_{57} aromatic carbons seem to display different chemical shifts depending on the pH as one would expect since histidine pKa is ~6.45 and both nitrogens are likely to be protonated below such pH value (446). RebB-H_{57} is the only histidine in the primary sequences of both RebA and RebB, and thus seems to be highly important in the studied phenomenon. The differences in chemical shifts are quite similar to those previously documented in the literature: $\Delta\delta$ = -0.4 ppm (upfield) for Cα (**Figure 7B** lower panels), $\Delta\delta$ = 0.7 ppm (downfield) for Cβ (**Figure 7B** upper panels**)**, $\Delta\delta$ = -0.3 (upfield) for Cδ2 (**Figure 7B** right panels**)**, $\Delta\delta$ = -0.6 ppm (upfield) for Cε1(**Figure 7B** left panels). RebB-H_{57}, positively charged at pH 5, might interact with the two negative charges from RebA N-terminal domain (*i.e.*, RebA-D_9 and RebA-D_{15}), thus (i) keeping RebA N-terminal moiety in an orderly fashion and (ii) rearranging the overall supramolecular architecture of the nanomachine.

We also performed 2D ^{13}C-^{15}N NCA SSNMR experiments probing backbone atoms (**Figure 7C**). In such spectra, each peak connects intra-residual $^{13}C\alpha$ and ^{15}N. Comparing the two fingerprints, several peaks had shifted upon pH change. Such residues are specifically shown in the spectra: in RebB, B3N, B4V, B5N, B7Q, B13T, B14Q, B19I, B40G, B43I, B44Q, B56L, B60S; in RebA, A17V, A16I, A19S, A20Q, A28T, A30T, A34G, A36T, A59A, A62G (**Figure 7C**), the sequence alignment (**Figure 7D**), and the structural models of RebA and RebB (**Figure 7E-F**). Amongst all impacted residues, RebB-I_{43} also showed some level of polymorphism at pH 7 in 2D ^{13}C-^{13}C PDSD experiments (*i.e.*, two sets of peaks were observed for this residue), and interestingly only one set of peaks was detected in experiments performed

at pH 5 (*i.e.*, some peaks disappear). Such alterations are highlighted with black arrows in **Figure 7A**. Similar to RebB-H_{57}, RebB-I_{43} is one of the few amino acids in R bodies displaying a drastic change of structural arrangement upon pH change. Both are highlighted in yellow on the RebB structural model (**Figure 7F**).

6.4.6. R bodies can store and release biomolecules and thus act as a delivery nano-machine

The loss of ordered carbohydrate ^{13}C signals raises an interesting interrogation about of its presence and its potential role. A covalent link between an amino acid and a carbohydrate is unlikely as any post-translational modification is rare and tricky to obtain when using a bacterial recombinant overexpression system. Therefore, the presence of carbohydrates in our samples would be better explained by a potential strong interaction with the proteins in R bodies, or by random export of *E. coli* content by these massive nano-machines. At this point, many hypotheses could be challenged: (i) the functional involvement of the carbohydrate in the structural integrity of the R bodies assemblies, (ii) the potential presence of carbohydrate-containing molecules as opposed to a monomeric sugar (*e.g.*, RNA, DNA, NAD, *etc.*), (iii) the possibility that the carbohydrate, although found in a rather high concentration, is an artefact of the purification protocol.

In order to gain more insights on the different non-protein biomolecules found in recombinant R bodies we set up a battery of experiments using both solution and SSNMR with different nuclei (^{31}P, ^{1}H, ^{13}C, ^{15}N). To this end, we prepared different samples in order to examine the compositions and dynamic states of R bodies during one extension/contraction cycle. From a single initial production of $^{13}C/^{15}N$ doubly labelled R bodies, one extension/contraction cycle was performed and a total of five samples were thus extracted at each step (chronologically): (i) R bodies at pH 7 (solid sample), (ii) R bodies at pH 5 (solid state), (iii) supernatant retrieved after pelleting the R bodies pH 5 sample (liquid sample), (iv) R bodies at pH 7 after pH 5 (referred to as pH $7_{back)}$, (v) supernatant retrieved after pelleting the R bodies pH 7_{back} (liquid sample).

As mentioned previously, no trace of any carbohydrate/nucleotide was found in both 1D (**Figure 8A**) and 2D CP-based (**Figure 6**) experiments of R bodies at pH 5. The same has been observed for R bodies at pH 7_{back}. We performed 1D (**Figure 8B**) and 2D INEPT experiments (**Figure 8C**), designed to detect any ^{1}H-^{13}C flexible bond, to try and detect the missing signal as part of a more flexible region. Such data clearly shows carbohydrate signals at pH 7 and pH

5, however, almost no trace of saccharides is found in R bodies when returning to a coiled arrangement in rigid nor flexile states. We might infer that some carbohydrates remain in the sample in a flexible state at pH 5, within bulk water as the sample is still hydrated. Interestingly, both rigidity- and flexibility-detecting experiments do not display carbohydrate-like NMR signal after the first pH 7/pH 5 round (**Figure 8A-B-C**). Such results would suggest that the sugars have been completely discarded following the second switch to pH 7_{back} (**Figure 8D**). Additionally, all non-protein (*i.e.*, acyl chains, saccharides) ^{13}C-^{13}C signals observed in the 2D DP-INADEQUATE experiments are missing for the sample at pH 5 (**Figure 8E**). Supernatants as retrieved from pelleting the solid samples in between buffer exchanges were analysed using ^{31}P NMR. It should be noted that liquid samples were probed using ^{31}P direct polarization experiment, while ordered ^{31}P were observed with an Echo-Hahn pulse sequence. (**Figure 8F**). Most signals disappear in the ^{31}P Echo-Hahn at pH 5 when compared to the ^{31}P Echo-Hahn recorded for R bodies at pH 7 (**Figure 8F**, black spectrum), leaving a single signal around -0.5-0.7 ppm (**Figure 8F,** yellow spectra), which is missing when switching back to pH 7_{back} (**Figure 8F,** red spectra). For ^{31}P DP experiments of supernatants, the same was observed as the peak shown at pH 5 completely disappeared when switching back to pH 7_{back} (**Figure 8F,** red spectra). Additionally, the mentioned ^{15}N peaks detected in a 1D CP around 150 ppm seemed to disappear at pH 5 (**Figure 5E-F**).

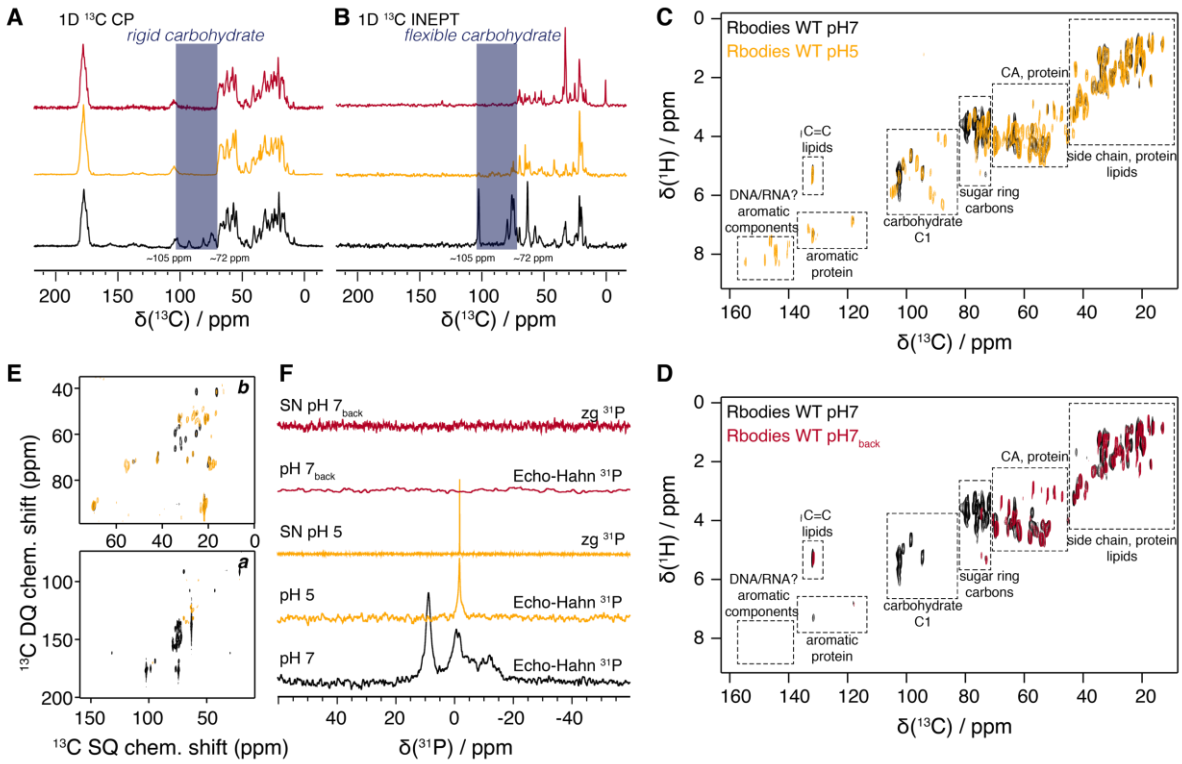


Figure 6-8: (A) 1D ^{13}C CP experiment of R bodies at pH 7 (black), pH 5 (dark yellow), pH 7_{back} (dark red). (B) 1D ^{13}C INEPT experiment of R bodies at pH 7 (black), pH 5 (dark yellow), pH 7_{back} (dark red). (C) Superimposed 2D ^{1}H-^{13}C J-INEPT experiments of R bodies at pH 7 (black), pH 5 (dark yellow). (D) Superimposed 2D ^{1}H-^{13}C J-INEPT experiments of R bodies at pH 7 (black), pH 7_{back} (dark red). (E) Excerpts of 2D ^{13}C-^{13}C DP-INADEQUATE recorded on R bodies at pH 7 (black) and pH 5 (dark yellow) showing the sidechain / acyl chains (upper) and the carbohydrate area (lower). (F) 1D ^{31}P Echo-Hahn experiments of or 1D ^{31}P DP experiments of R bodies at pH 7/5/7_{back} and retrieved supernatants. The spectral areas of interest are highlighted in blue when necessary. For all panels, samples at pH 7, pH 5 and pH 7_{back} are displayed in black, dark yellow and dark red, respectively.

At this point, all data (*i.e.*, the ^{13}C and ^{31}P chemical shifts) support the current hypothesis that uridine/uracil molecules are entrapped in closed enrolled R bodies (444). These observations strongly suggest that a nucleotide-like molecule (*i.e.*, presence of ribose, aromatic/basic ^{13}C, and phosphate groups, purine-like ^{15}N) is first entrapped in the R bodies peculiar shape and released upon unwinding of the nanomachine. At this stage of the analysis, the presence of these molecules could be an artefact of the production and extraction protocol, regardless of the numerous washing steps prior to any SSNMR experiments. After several pH 7/pH 5 rounds, the sugar/phosphate-containing molecules could be discharged from the assemblies. Additionally, the specific uptake and delivery of RNA or nucleotides as part the R body machinery could also be considered as a functional observation, as they seem to be entrapped within the R bodies and can only be removed by mechanical changes within the assemblies. Such hypothesis can be substantiated by the extensive analysis of RNA and DNA within R bodies (430, 439, 447). However, regardless of the bio-relevance of molecules observed in our study, the mere fact that recombinant R bodies were able to store different biomolecules from *E. coli* cytoplasm and keep them throughout the entire purification and wash processes shows an interesting feature of R bodies in line with the current hypothesis.

6.5. Conclusion

We studied R bodies type 51 supramolecular assemblies using ultra-fast and regular MAS SSNMR experiments. We assigned the rigid core of the assemblies and identified RebA and RebB to be solely involved in the structural arrangement of R bodies. Both proteins engage their N terminal ends, $RebA_{15\text{-}75}$ and $RebB_{1\text{-}60}$. Preliminary structural models suggest a helix-loop-helix arrangement. The C terminal moieties have not been unambiguously assigned, however, NMR ^{13}C signals detected in J-based (*i.e.*, flexible) could account for the missing

residues. Interestingly no NMR resonances have been assigned to RebC and RebD even if their co-expression is indispensable for the production and assembly of fully functional R bodies. A complete 3D supramolecular structure is still to be determined as there is no information at the moment regarding the quaternary arrangement of RebA and RebB. The precise role and actions of RebC and RebD remains to be documented, however, we suspect them to participate in the complex assembly of these nanomachines as chaperones.

We identified residues within RebA and RebB primary sequences which seemed to be impacted by the extension around pH 5. Interestingly, several residues show chemical shift perturbation upon pH switch within the start/end of the rigid sequence and close to RebB-H_{57} and RebB-I_{43}. We emit the hypothesis that the 15-residue long N terminal moiety of RebA is fully engaged in the extension, and that this phenomenon is triggered by RebB-H_{57} protonation at lower pH. Another interesting feature is the ability of R bodies to store a variety of biomolecules, such as lipids, saccharides, and phosphate groups suggesting the potential presence of RNA or DNA, and to deliver such cargo to the medium upon extension. Given that R bodies are not toxic themselves, it has long been postulated that they are involved in carrying a toxin to the sensitive strain. Such toxin has yet to be discovered. One might wonder how the toxin is specifically attached to R bodies, as many biomolecule types were carried out of the *E. coli* cytoplasm in our experiment with quite high abundance.

After several decades of research of theses fascinating objects, the remarkable performance of R bodies is now being considered in various potential engineering applications as tunable research tools, or even in clinical applications as programmable release/penetration devices targeting membrane compartments (441). A large-scale and accurate structural characterization of R bodies should tremendously impact ongoing research and lead to a better understanding of the so-called “killer” trait conferred by R bodies.

Table of figures

Conclusions and Perspectives

The present manuscript consists of six independent chapters designed as research articles, summarizing my research work between 2017 and 2020.

Chapter 1 presents the development of an SSNMR methodological approach to use DNP in the context of structural investigation of protein assemblies, with the combined advantages of fast MAS and selective isotope labelling to enable ^{1}H-detected experiments. The approach relies on the use of a protein sample with a minimal ^{13}C and ^{1}H isotope labelling, namely a single ^{13}C-^{1}H couple solely in isoleucine, leucine, and valine side chains. Such scheme combines the effects of spin dilution (*i.e.*, selective labelling) and DNP (*i.e.*, signal enhancement) on the overall signal sensitivity spectral resolution. We demonstrated the recording of ^{1}H-detected SSNMR multidimensional experiments on protein samples, with impressive sensitivity provided by the DNP enhancement. The proof-of-concept is demonstrated on HET-s amyloid fibrils and will be ready to apply to various biological systems of interest.

Chapter 2 encompasses the study of a novel functional amyloid family involved in programmed cell death of *Nectria haematococca*, highly reminiscent of the HET-s paradigm in *Podospora anserina*. We focused on a single protein called SesB, part of a three-protein system, discovered by Dr Sven Saupe with whom this study has been conducted. The three proteins possess a conserved amyloid motif, which is called the sigma motif. We used SSNMR to investigate the structural fold of SesB amyloid motif and found that SesB fold seems to constitute a new functional amyloid family. We determined the SesB secondary structure, and a three-dimensional structural calculation is in progress. Contrary to the initial hypothesis that *Nectria haematococca* shares the same molecular signalling system as *Podospora anserina*, SesB amyloid core seems to be formed of five β-strands, and seems to represent a new structural family compared to known functional and pathological amyloid folds. Preliminary data indicate that the same structural fold in conserved in SesA, the other partner protein in the same system. We predict the three proteins SesA, SesB, het-eN, share the same structural amyloid fold. Another perspective would be to study the co-assembly of these proteins and assess their potential interactions. SesB lipase activity remains to be characterized. Elucidating the cell death mechanism is important to understand the function of this system compared to other known high-order signalling complexes.

Chapter 3 is based on the published report (354) of our collaboration with Dr Diego Romero at the University of Malaga in Spain. Combining a large set of biophysical and biological methods, we analysed TasA proteins, which are found in large abundance in the biofilm matrix of *Bacillus subtilis* and the pathogenic strain *Bacillus cereus*. TasA had indeed been characterized as a functional amyloid responsible for the biofilm matrix integrity. We compared the structural amyloid arrangements of the two TasA proteins in their assembled and amyloid forms as observed using X-ray diffraction. We show using SSNMR that the two species share the same structural arrangement, although different local orders are observed. We continued by assessing their binding rates to ThT and concluded that the difference in polymorphism levels might be explained by their very different self-assembly kinetics. In a co-assembly assay, we tried to assess the roles of TapA and CalY, two partner proteins found in *B. subtilis* and *B. cereus*, respectively, and found that although they do not alter TasA fold in vitro, the partner proteins might have a catalytic function during the biofilm matrix formation. We are continuing this study and focusing on the potential interaction of TasA with membranes. Preliminary data have been obtained and highlight the membrane perturbation effect of TasA in *B. subtilis*.

Chapter 4 is linked to Chapter 3 as it describes a whole-cell analysis of native *B. subtilis* biofilms. Combining spin dilution and selective labelling, we conducted SSNMR analyses on rather complex native mixtures, that are biofilms. We assessed and probed the matrix compositions and found most EPS-related biomolecules. The next step was to achieve a similar biophysical assessment on different *B. subtilis* biofilm-related mutants: Δ*eps*, Δ*eps/tasA*, Δ*eps/tapA*. The mutants presented drastically different dynamical properties, thus changing the biophysical features of this natural gel. Δ*eps*, Δ*eps-tasA*, and Δ*eps-tapA* mutants show different dynamic properties with higher order and rigidity. We report an accumulation of monosaccharides such as pyranose and ribose, which strongly suggests the presence of UDP-Glc as the building block for EPS biosynthesis. Our results, combined with the general focus on *B. subtilis* as a model for understanding biofilm formation, highlights the many roles of EPS and functional amyloids in water accessibility, biofilm resistance to mechanical stress, overall dynamics, saccharide content, etc. Such studies can be continued with other SSNMR experiments such as water-edited analysis, which assesses the water accessibility of the different biomolecule types to have a better understanding of *B. subtilis* biofilm architecture.

The ultra-fast MAS SSNMR technology might also be used for a more exhaustive analysis of the biological system.

Chapter 5 focuses on the study of the type III secretion system in Gram-negative bacteria and more specifically, on assembly of the inner rod, which connects the needle to the basal body of the nano-machine. We characterized the *in vitro* self-assembly features of inner rod subunits PscI from *Pseudomonas* species and PrgJ from *Salmonella* that form filamentous tubes *in vitro*. Our results provide a preliminary structural study of such polymorphic objects and show a difference between the local disorder displayed by the N-terminal moiety of the proteins and the highly ordered C-terminus which shows a distinct level of 'imperfect' homogeneity. Both proteins keep the ability to form tubes *in vitro* after the truncation of their N-terminus, and such truncated constructs show almost identical structural arrangements when probed with SSNMR. We analysed the mass-per-length of PscI fibrils, in collaboration with Joe Wall from the Brookhaven National Laboratory and showed that PscI has a ~9 Å axial rise per subunit. Our results indicate that, contrary to the needle filament, the inner rod lacks the ability to assemble into homogeneous objects in absence of the T3SS base, highlighting that the T3SS machinery assembly proceeds in regulated sequential steps.

Chapter 6 encloses the fascinating study of R bodies, the polymeric protein inclusions found in the cytoplasm of bacteria that have the ability to contract and extend into tape-like ribbons and hollow tubes in an ATP-independent and pH-dependent fashion. We used ultra-fast MAS to study R bodies extracted from *E. coli* cells. Although four proteins are required for the in vivo assembly of R bodies (*i.e.*, RebA, RebB, RebC, RebD), only RebA and RebB were found in the highly ordered and rigid core of the R bodies. The two proteins show an α-helical propensity and a preliminary model of the two monomers supports a helix-turn-helix conformation. The inter-molecular interactions and arrangements between RebA and RebB have yet to be determined. We also assessed the R bodies ability to store a range of biomolecules, such as DNA/RNA, lipids, and saccharides, as such cargo is released upon extension of R bodies at a pH around 5. We investigated the amino acids impacted by such a drastic structural rearrangement and have pinpointed numerous regions in RebA and RebB which are altered upon pH switch. Last, we emit the hypothesis that a histidine residue in RebB is involved in the unrolling of the ribbon by attracting a ~15 residue-long segment at the N-terminus of RebA which is negatively charged, as the histidine gets protonated at lower pH

(histidine pKa = ~6). A more detailed structural description has yet to be done as our study only encompasses a preliminary exploration. Additional inter-molecular interactions and arrangements have to be determined as, so far, there is no accurate data on the general symmetry of the R bodies ribbon.

Résumé de la thèse (en français)

Le chapitre 1 discute une approche méthodologique pour utiliser la DNP (*Dynamic Nuclear Polarization* ou Polarisation Nucléaire Dynamique) dans le contexte de la biologie structurale en combinant les avantages combinés de la rotation à l'angle magique rapide (jusqu'à 40 kHz) et du marquage isotopique sélectif. Nous avons réalisé ce développement méthodologique en collaboration avec Dr Guido Pintacuda et Dr Anne Lesage de l'ENS de Lyon pour tenter de mettre en place une méthode de détection ^{1}H des protéines dans des conditions de Polarisation Nucléaire Dynamique et d'approfondir les connaissances sur les applications de cette méthode aux échantillons biologiques. L'approche repose sur l'utilisation d'un échantillon de protéine avec un marquage minimal des isotopes ^{13}C et ^{1}H uniquement dans les chaînes latérales des isoleucines, leucines et valines. Ce schéma combine les effets du phénomène appelé *spin dilution* (*i.e.*, marquage isotopique sélectif) et de la Polarisation Nucléaire Dynamique (*i.e.*, augmentation du signal) sur la résolution spectrale et la sensibilité globale du signal. Nous avons démontré la possibilité d'obtenir des expériences multidimensionnelles de RMN du solide détectées en ^{1}H sur des échantillons de protéines. Nous observons une sensibilité impressionnante et une résolution suffisante même à de très basses températures. Le concept est démontré sur des fibres amyloïdes de la protéine HET-s et peut être appliqué à divers systèmes biologiques d'intérêt.

Le chapitre 2 porte sur l'étude d'une nouvelle famille d'amyloïdes fonctionnelles impliquées dans la mort cellulaire programmée de *Nectria haematococca*, qui rappelle le paradigme HET-s de *Podospora anserina*. HET-s est une protéine modèle pour les amyloïdes fonctionnelles et les prions, car elle se présente comme une protéine amyloïde fascinante dont l'assemblage fait partie d'une cascade de signalisation. Dr Sven Saupe, de l'Université de Bordeaux, a identifié un système de trois gènes codant pour trois protéines appelées SesA, SesB et HET-eN, qui partagent un court segment à leur extrémité N- ou C-terminale. Ces segments présentent une homologie de séquence élevée. Le but de l'étude est de caractériser l'activité *in vivo* et les éléments structurels *in vitro*, en étroite collaboration avec Dr Sven Saupe. L'identification d'une nouvelle fibre amyloïde fonctionnelle est d'un grand intérêt dans le domaine des amyloïdes pathologiques et leur implication dans de nombreuses maladies. Aujourd'hui une compréhension détaillée des différences entre les protéines amyloïdes fonctionnelles et pathologiques reste à construire. Nous avons utilisé la RMN des solides pour étudier la

structure et le repliement du motif amyloïde SesB. SesB semble constituer une nouvelle famille d'amyloïdes fonctionnelles. Nous avons déterminé la structure secondaire de SesB formé de cinq brins béta, et un calcul structurel tridimensionnel est en cours. Contrairement à l'hypothèse initiale selon laquelle *Nectria haematococca* partage le même système de signalisation moléculaire que *Podospora anserina*, notre étude semble au contraire indiquer une nouvelle famille structurelle qui comprendrait SesA, SesB, et HET-eN. Des données préliminaires indiquent que la structure est conservée pour SesA, l'autre protéine partenaire. Nous prévoyons que les trois protéines SesA, SesB, HET-eN, partagent la même structure amyloïde. Une autre perspective consisterait à étudier le co-assemblage de ces protéines et à évaluer leurs interactions potentielles. De plus, l'activité de la lipase SesB reste à caractériser. L'élucidation du mécanisme de mort cellulaire est importante pour comprendre la fonction de ce système par rapport à d'autres complexes de signalisation d'ordre supérieur connus.

Le chapitre 3 est une étude d'une autre protéine amyloïde fonctionnelle appelée TasA. TasA est impliquée dans la composition de la matrice du biofilm de *Bacillus subtilis* et de *Bacillus cereus*. Les biofilms sont des communautés bactériennes fixées en surface dans lesquelles les bactéries sécrètent une matrice extracellulaire (MEC). La MEC est composée principalement d'exopolysaccharides (EPS), de protéines, d'ADN, etc. Ce travail a été réalisé en collaboration avec Dr Diego Romero, de l'Université de Malaga. L'objectif de cette étude est d'analyser une nouvelle protéine amyloïde fonctionnelle. Plusieurs rapports ont déjà proposé que TasA forme des fibres amyloïdes au sein de la matrice extracellulaire du biofilm, étant ainsi responsable de l'intégrité structurelle du biofilm et protégeant les bactéries des stresses externes et mécaniques. En plus de TasA, on retrouve les protéines TapA et CalY qui sont des protéines partenaires chez *Bacillus subtilis* et *Bacillus cereus*, respectivement. Leur grande polyvalence fonctionnelle et leur remarquable capacité à former des assemblages filamenteux robustes contribuent à l'intérêt général et à l'émerveillement entourant les amyloïdes fonctionnelles dans le règne bactérien. Ce chapitre combine une approche biochimique/biophysique à une approche microbiologique, car nous présentons les différences entre les espèces *Bacillus subtilis* et *Bacillus cereus*. Ce chapitre 3 est basé sur l'article publié (354) de notre collaboration avec Dr Diego Romer. En combinant un large ensemble de méthodes biophysiques et biologiques, nous avons analysé les protéines TasA, qui se trouvent en grande abondance dans la matrice du biofilm de *Bacillus subtilis* et de la souche pathogène *Bacillus cereus*. Nous avons comparé les arrangements structurels amyloïdes des deux protéines TasA dans leurs formes assemblées et amyloïdes, tels qu'observés par diffraction des rayons X. Nous avons montré par RMN des

solides que les deux espèces partagent le même arrangement structurel, bien que des ordres locaux différents soient observés. Nous avons poursuivi en évaluant leur taux de liaison au ThT et avons conclu que la différence des niveaux de polymorphisme pourrait s'expliquer par leur cinétique d'assemblage très différente. Dans une expérience de co-assemblage, nous avons essayé d'évaluer les rôles de TapA et de CalY et avons trouvé que bien qu'elles n'altèrent pas la structure de TasA *in vitro*, les protéines partenaires pourraient avoir une fonction catalytique pendant la formation de la matrice du biofilm. Nous poursuivons actuellement cette étude et nous nous concentrons sur l'interaction potentielle de TasA avec les membranes. Des données préliminaires ont été obtenues et mettent en évidence l'effet de perturbation membranaire de TasA chez *B. subtilis*.

Le chapitre 4 se concentre sur la formation du biofilm et étudie l'impact des mutations des gènes *eps*, *tasA* et *tapA*, sur la composition globale de la matrice du biofilm et sur ses caractéristiques dynamiques générales. Ce chapitre présente l'utilisation des techniques multidimensionnelles de la RMN des solides comme un outil puissant pour obtenir la composition de la MEC dans les biofilms intacts de *Bacillus subtilis*. Les modifications de la composition moléculaire du biofilm ont été étudiées pour trois mutants différents et comparées à des biofilms de type sauvage afin d'entrevoir l'impact des mutations sur la composition de la matrice. En combinant le phénomène de *spin dilution* et le marquage isotopique sélectif, nous avons sondé la matrice extracellulaire et avons détecté la plupart des biomolécules attendues dans le biofilm natif. L'étape suivante consistait à réaliser une évaluation biophysique similaire sur différents mutants liés au biofilm de *B. subtilis* : Δ*eps*, Δ*eps/tasA*, Δ*eps/tapA*. Les mutants présentaient des propriétés dynamiques radicalement différentes, modifiant ainsi les caractéristiques biophysiques de ce gel naturel : propriétés dynamiques différentes avec un ordre et une rigidité plus élevés. Nous observons une accumulation de monosaccharides tels que le pyranose et le ribose, ce qui suggère fortement la présence de l'UDP-Glc comme élément de base de la biosynthèse des EPS. Nos résultats mettent en évidence les nombreux rôles de la matrice extracellulaire et des protéines amyloïdes fonctionnelles dans l'accessibilité à l'eau, la résistance du biofilm aux contraintes mécaniques, la dynamique globale, la teneur en saccharides, etc. Ces études peuvent être poursuivies avec d'autres expériences telles que l'analyse de l'accessibilité à l'eau des différents types de biomolécules afin de mieux comprendre l'architecture du biofilm de *B. subtilis*. La technologie de la RMN des solides

combinée à la rotation à l'angle magique ultra-rapide pourrait également être utilisée pour une analyse plus exhaustive du système biologique.

Le chapitre 5 se concentre sur les protéines de la tige interne (*inner rod*) du système de sécrétion de type 3 des bactéries Gram-négatives. Les systèmes de sécrétion de type 3 (T3SS) sont responsables de la capacité de certaines bactéries Gram-négatives à délivrer directement des protéines effectrices de leur cytoplasme à celui des cellules eucaryotes via le complexe appélé *needle* (aiguille), par le biais d'une batterie de gènes nécessaires à l'assemblage d'une nanomachine, également appelée "injectisome". Cet appendice passe à travers la barrière bactérienne composée de membranes plasmiques, de la couche de peptidoglycanes et de l'espace extracellulaire. Ce travail a été réalisé en collaboration avec le professeur Jorge E. Galán de l'université de Yale (New Haven, Connecticut, USA). Après avoir élucidé le complexe de la *needle*, nous avons entrepris de caractériser la tige interne, qui constitue la base de l'aiguille en la reliant au corps basal. Nous souhaitons étudier deux protéines, PscI chez *Pseudomonas* spp., et PrgJ chez *Salmonella* spp., qui forment la tige interne. Dans le cadre de cette étude, j'ai pu travailler avec le Professeur Jorge Galán pendant cinq mois. Nous avons caractérisé *in vitro* les caractéristiques d'auto-assemblage des sous-unités de la tige interne PscI et PrgJ qui forment des tubes filamenteux *in vitro*. Nos résultats fournissent une étude structurelle préliminaire de ces objets polymorphes et montrent une différence entre le désordre local de la partie N-terminale des protéines alors que la partie C-terminale hautement ordonnée présente un niveau distinct d'homogénéité "imparfaite". Les deux protéines conservent la capacité de former des tubes *in vitro* après la troncation de leur extrémité N-terminale, et ces constructions tronquées présentent des arrangements structurels presque identiques lorsqu'elles sont étudiées par SSNMR. Nous avons analysé la *mass-per-length* des fibres de PscI, en collaboration avec Joe Wall du Brookhaven National Laboratory, et avons montré que la protéine PscI a une élévation axiale (*axial rise*) de ~9 Å par sous-unité. Nos résultats indiquent que, contrairement au filament de l'aiguille, la tige interne n'a pas la capacité de s'assembler en objets homogènes en l'absence de la base du T3SS, ce qui suggère que l'assemblage de la machine du T3SS se fait par étapes séquentielles régulées.

Le chapitre 6 relate les travaux que nous avons réalisés sur un assemblage bactérien supramoléculaire complexe et mystérieux appelé corps réfractaires (*refractile bodies*, aussi appelés corps R). Les corps R (dans cette étude, type 51) sont des polymères protéiques inhabituels, en forme de ruban, que l'on trouve dans de nombreuses espèces bactériennes, telles

que *Caedibacter* spp. et *Pseudomonas* spp. En 1938, Tracy M. Sonneborn a observé que certaines souches de *Paracemium aurelia* étaient capables de tuer d'autres souches sensibles de la même espèce. Cette capacité était initialement liée à un facteur extranucléaire, appelé kappa, qui était associé à la production et à la libération d'une toxine dans le milieu. Les corps R sont de grands biopolymères protéiques qui ressemblent à des rubans et qui ont la capacité de s'enrouler et de se dérouler en fonction du pH. Ce travail est réalisé en collaboration avec Dr Justin Kollman, de l'université de Seattle, aux États-Unis. Aucune structure n'a été découverte jusqu'à présent, et nous avons entrepris de caractériser cet arrangement structurel inhabituel. Le chapitre 6 présente l'étude fascinante des corps R, ces inclusions de protéines polymères que l'on trouve dans le cytoplasme des bactéries et qui ont la capacité de se contracter et de s'étendre en rubans et en tubes creux indépendemmant de l'ATP et uniquement en fonction du pH. Nous avons utilisé la rotation à l'angle magique ultra-rapide pour étudier les corps R extraits des cellules d'*E. coli*. Bien que quatre protéines soient nécessaires pour l'assemblage *in vivo* des corps R (c'est-à-dire RebA, RebB, RebC, RebD), seules RebA et RebB ont été trouvées dans le noyau hautement ordonné et rigide des corps R. Les deux protéines présentent un arrangement en hélices alpha et un modèle préliminaire des deux monomères permet d'établir une conformation en hélice-boucle-hélice. Les interactions et les arrangements intermoléculaires entre RebA et RebB restent à déterminer. Nous avons également évalué la capacité des corps R à stocker une série de biomolécules, telles que l'ADN/ARN, les lipides et les saccharides, et ce chargement est libéré lors de l'extension des corps R à un pH d'environ 5. Nous avons étudié les acides aminés affectés par le réarrangement structurel et avons identifié de nombreuses régions dans les RebA et RebB sont altérées lors du changement de pH. D'autres interactions et arrangements intermoléculaires doivent être déterminés car, jusqu'à présent, il n'existe pas de données précises sur la symétrie générale du ruban des corps R.

References

Uncategorized References

1. Linser, R. (2017) Solid-state NMR spectroscopic trends for supramolecular assemblies and protein aggregates. *Solid State Nucl Magn Reson* **87**, 45-53
2. Demers, J.-P., Fricke, P., Shi, C., Chevelkov, V., and Lange, A. (2018) Structure determination of supra-molecular assemblies by solid-state NMR: Practical considerations. *Progress in Nuclear Magnetic Resonance Spectroscopy* **109**, 51-78
3. van der Wel, P. C. A. (2017) Insights into protein misfolding and aggregation enabled by solid-state NMR spectroscopy. *Solid State Nucl Magn Reson* **88**, 1-14
4. Loquet, A., Tolchard, J., Berbon, M., Martinez, D., and Habenstein, B. (2017) Atomic Scale Structural Studies of Macromolecular Assemblies by Solid-state Nuclear Magnetic Resonance Spectroscopy. *Journal of Visualized Experiments*, e55779
5. Habenstein, B., and Loquet, A. (2016) Solid-state NMR: An emerging technique in structural biology of self-assemblies. *Biophysical Chemistry* **210**, 14-26
6. Morag, O., Sgourakis, N. G., Baker, D., and Goldbourt, A. (2015) The NMR-Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proc Natl Acad Sci U S A* **112**, 971-976
7. Loquet, A., Sgourakis, N. G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., Griesinger, C., Kolbe, M., Baker, D., Becker, S., and Lange, A. (2012) Atomic model of the type III secretion system needle. *Nature* **486**, 276-279
8. Meier, B. H., and Bockmann, A. (2015) The structure of fibrils from 'misfolded' proteins. *Curr Opin Struct Biol* **30**, 43-49
9. Sperling, L. J., Tang, M., Berthold, D. A., Nesbitt, A. E., Gennis, R. B., and Rienstra, C. M. (2013) Solid-state NMR study of a 41 kDa membrane protein complex DsbA/DsbB. *J Phys Chem B* **117**, 6052-6060
10. Murray, D. T., Das, N., and Cross, T. A. (2013) Solid State NMR Strategy for Characterizing Native Membrane Protein Structures. *Acc Chem Res*
11. Das, N., Murray, D. T., and Cross, T. A. (2013) Lipid bilayer preparations of membrane proteins for oriented and magic-angle spinning solid-state NMR samples. *Nat Protoc* **8**, 2256-2270
12. Cady, S. D., Wang, J., Wu, Y., DeGrado, W. F., and Hong, M. (2011) Specific binding of adamantane drugs and direction of their polar amines in the pore of the influenza M2 transmembrane domain in lipid bilayers and dodecylphosphocholine micelles determined by NMR spectroscopy. *J. Am. Chem. Soc.* **133**, 4274-4284
13. Hong, M., Hu, F. H., Luo, W. B., and Cady, S. D. (2011) Conformational plasticity of the influenza A M2 transmembrane helix in lipid bilayers under varying pH, drug binding, and membrane thickness. *Biochimica Et Biophysica Acta-Biomembranes* **1808**, 415-423
14. Abdine, A., Verhoeven, M. A., Park, K. H., Ghazi, A., Guittet, E., Berrier, C., Van Heijenoort, C., and Warschawski, D. E. (2010) Structural study of the membrane protein MscL using cell-free expression and solid-state NMR. *J Magn Reson* **204**, 155-159
15. Ahmed, M. A., Bamm, V. V., Harauz, G., and Ladizhansky, V. (2010) Solid-state NMR spectroscopy of membrane-associated myelin basic protein--conformation and dynamics of an immunodominant epitope. *Biophys J* **99**, 1247-1255
16. Kaplan, M., Cukkemane, A., van Zundert, G. C. P., Narasimhan, S., Daniëls, M., Mance, D., Waksman, G., Bonvin, A. M. J. J., Fronzes, R., Folkers, G. E., and Baldus, M. (2015) Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nature Methods* **12**, 649-652

17. Antzutkin, O. N., Leapman, R. D., Balbach, J. J., and Tycko, R. (2002) Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* **41**, 15436-15450.
18. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16742-16747
19. Balbach, J. J., Petkova, A. T., Oyler, N. A., Antzutkin, O. N., Gordon, D. J., Meredith, S. C., and Tycko, R. (2002) Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance. *Biophys J* **83**, 1205-1216.
20. Balbach, J. J., Ishii, Y., Antzutkin, O. N., Leapman, R. D., Rizzo, N. W., Dyda, F., Reed, J., and Tycko, R. (2000) Amyloid Fibril Formation by Abeta16-22, a Seven-Residue Fragment of the Alzheimer's beta-Amyloid Peptide, and Structural Characterization by Solid State NMR. *Biochemistry* **39**, 13748-13759
21. Jaroniec, C. P., MacPhee, C. E., Bajaj, V. S., McMahon, M. T., Dobson, C. M., and Griffin, R. G. (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 711-716
22. Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M., and Griffin, R. G. (2002) Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16748-16753
23. Van Melckebeke, H., Wasmer, C., Lange, A., Ab, E., Loquet, A., Böckmann, A., and Meier, B. H. (2010) Atomic-Resolution Three-Dimensional Structure of HET-s(218-289) Amyloid Fibrils by Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **132**, 13765-13775
24. Yan, S., Suiter, C. L., Hou, G., Zhang, H., and Polenova, T. (2013) Probing Structure and Dynamics of Protein Assemblies by Magic Angle Spinning NMR Spectroscopy. *Acc Chem Res*
25. Sun, S., Han, Y., Paramasivam, S., Yan, S., Siglin, A. E., Williams, J. C., Byeon, I. J., Ahn, J., Gronenborn, A. M., and Polenova, T. (2012) Solid-state NMR spectroscopy of protein complexes. *Methods in Molecular Biology* **831**, 303-331
26. Yang, J., Tasayco, M. L., and Polenova, T. (2008) Magic angle spinning NMR experiments for structural studies of differentially enriched protein interfaces and protein assemblies. *J. Am. Chem. Soc.* **130**, 5798-5807
27. Cuniasse, P., Tavares, P., Orlova, E. V., and Zinn-Justin, S. (2017) Structures of biomolecular complexes by combination of NMR and cryoEM methods. *Curr Opin Struct Biol* **43**, 104-113
28. Sperling, L. J., Berthold, D. A., Sasser, T. L., Jeisy-Scott, V., and Rienstra, C. M. (2010) Assignment strategies for large proteins by magic-angle spinning NMR: the 21-kDa disulfide-bond-forming enzyme DsbA. *J Mol Biol* **399**, 268-282
29. Curtis-Fisk, J., Spencer, R. M., and Weliky, D. P. (2008) Isotopically labeled expression in E. coli, purification, and refolding of the full ectodomain of the influenza virus membrane fusion protein. *Protein Expr Purif* **61**, 212-219
30. Tycko, R., and Ishii, Y. (2003) Constraints on supramolecular structure in amyloid fibrils from two-dimensional solid-state NMR spectroscopy with uniform isotopic labeling. *J. Am. Chem. Soc.* **125**, 6606-6607
31. Egorova-Zachernyuk, T. A., Shvets, V. I., Versluis, K., Heerma, W., Creemers, A. F., Nieuwenhuis, S. A., Lugtenburg, J., and Raap, J. (1996) Preparation of site-specific

isotopically labelled zervamicins, the antibiotic peptaibols produced by Emericellopsis salmosynnemata. *J Pept Sci* **2**, 341-350
32. Ikura, M., Marion, D., Kay, L. E., Shih, H., Krinks, M., Klee, C. B., and Bax, A. (1990) Heteronuclear 3D NMR and isotopic labeling of calmodulin. Towards the complete assignment of the 1H NMR spectrum. *Biochem Pharmacol* **40**, 153-160
33. Ikura, M., Krinks, M., Torchia, D. A., and Bax, A. (1990) An efficient NMR approach for obtaining sequence-specific resonance assignments of larger proteins based on multiple isotopic labeling. *FEBS Lett* **266**, 155-158
34. Mack, J. W., Torchia, D. A., and Steinert, P. M. (1988) Solid-state NMR studies of the dynamics and structure of mouse keratin intermediate filaments. *Biochemistry* **27**, 5418-5426
35. Kinsey, R. A., Kintanar, A., and Oldfield, E. (1981) Dynamics of amino acid side chains in membrane proteins by high field solid state deuterium nuclear magnetic resonance spectroscopy. Phenylalanine, tyrosine, and tryptophan. *J Biol Chem* **256**, 9028-9036
36. Siegal, G., and Selenko, P. (2019) Cells, drugs and NMR. *Journal of Magnetic Resonance* **306**, 202-212
37. Zhao, W., Fernando, L. D., Kirui, A., Deligey, F., and Wang, T. (2020) Solid-state NMR of plant and fungal cell walls: A critical review. *Solid State Nuclear Magnetic Resonance* **107**, 101660
38. Kang, X., Kirui, A., Dickwella Widanage, M. C., Mentink-Vigier, F., Cosgrove, D. J., and Wang, T. (2019) Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. *Nature Communications* **10**, 347
39. Romaniuk, J. A. H., and Cegelski, L. (2015) Bacterial cell wall composition and the influence of antibiotics by cell-wall and whole-cell NMR. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20150024
40. Kim, S. J., Cegelski, L., Preobrazhenskaya, M., and Schaefer, J. (2006) Structures of Staphylococcus aureus cell-wall complexes with vancomycin, eremomycin, and chloroeremomycin derivatives by 13C{19F} and 15N{19F} rotational-echo double resonance. *Biochemistry* **45**, 5235-5250
41. Bouvier, G., Simenel, C., Jang, J., Kalia, N. P., Choi, I., Nilges, M., Pethe, K., and Izadi-Pruneyre, N. (2019) Target Engagement and Binding Mode of an Antituberculosis Drug to Its Bacterial Target Deciphered in Whole Living Cells by NMR. *Biochemistry* **58**, 526-533
42. Thamarath, S. S., Alia, A., Daviso, E., Mance, D., Golbeck, J. H., and Matysik, J. (2012) Whole Cell Nuclear Magnetic Resonance Characterization of Two Photochemically Active States of the Photosynthetic Reaction Center in Heliobacteria. *Biochemistry* **51**, 5763-5773
43. Arnold, A. A., Bourgouin, J.-P., Genard, B., Warschawski, D. E., Tremblay, R., and Marcotte, I. (2018) Whole cell solid-state NMR study of Chlamydomonas reinhardtii microalgae. *Journal of Biomolecular NMR* **70**, 123-131
44. Seuring, C., Verasdonck, J., Ringler, P., Cadalbert, R., Stahlberg, H., Bockmann, A., Meier, B. H., and Riek, R. (2017) Amyloid Fibril Polymorphism: Almost Identical on the Atomic Level, Mesoscopically Very Different. *J Phys Chem B* **121**, 1783-1792
45. Landreh, M., Sawaya, M. R., Hipp, M. S., Eisenberg, D. S., Wuthrich, K., and Hartl, F. U. (2016) The formation, function and regulation of amyloids: insights from structural biology. *J Intern Med* **280**, 164-176
46. Vetri, V., and Fodera, V. (2015) The route to protein aggregate superstructures: Particulates and amyloid-like spherulites. *FEBS Lett* **589**, 2448-2463

47. Tycko, R. (2015) Amyloid polymorphism: structural basis and neurobiological relevance. *Neuron* **86**, 632-645
48. Tycko, R., and Wickner, R. B. (2013) Molecular Structures of Amyloid and Prion Fibrils: Consensus versus Controversy. *Acc Chem Res* **46**, 1487-1496
49. Eisenberg, D., and Jucker, M. (2012) The amyloid state of proteins in human diseases. *Cell* **148**, 1188-1203
50. Stefani, M. (2010) Structural Polymorphism of Amyloid Oligomers and Fibrils Underlies Different Fibrillization Pathways: Immunogenicity and Cytotoxicity. *Current Protein & Peptide Science* **11**, 343-354
51. Fandrich, M., Meinhardt, J., and Grigorieff, N. (2009) Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion* **3**, 89-93
52. Sipe, J. D., Benson, M. D., Buxbaum, J. N., Ikeda, S. I., Merlini, G., Saraiva, M. J., and Westermark, P. (2016) Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid* **23**, 209-213
53. Eichner, T., and Radford, S. E. (2011) A diversity of assembly mechanisms of a generic amyloid fold. *Molecular Cell* **43**, 8-18
54. Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of Molecular Biology* **273**, 729-739
55. Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R., and Meier, B. H. (2008) Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319**, 1523-1526
56. Sipe, J. D., and Cohen, A. S. (2000) Review: History of the amyloid fibril. *Journal of Structural Biology* **130**, 88-98
57. Knowles, T. P., Vendruscolo, M., and Dobson, C. M. (2014) The amyloid state and its association with protein misfolding diseases. *Nature Reviews Molecualr Cell Biology* **15**, 384-396
58. Eisele, Y. S., Monteiro, C., Fearns, C., Encalada, S. E., Wiseman, R. L., Powers, E. T., and Kelly, J. W. (2015) Targeting protein aggregation for the treatment of degenerative diseases. *Nature Reviews Drug Discovery* **14**, 759-780
59. Selkoe, D. J., and Hardy, J. (2016) The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine* **8**, 595-608
60. Kotler, S. A., Walsh, P., Brender, J. R., and Ramamoorthy, A. (2014) Differences between amyloid-[small beta] aggregation in solution and on the membrane: insights into elucidation of the mechanistic details of Alzheimer's disease. *Chemical Society Reviews* **43**, 6692-6700
61. Patel, H. R., Pithadia, A. S., Brender, J. R., Fierke, C. A., and Ramamoorthy, A. (2014) In Search of Aggregation Pathways of IAPP and Other Amyloidogenic Proteins: Finding Answers through NMR Spectroscopy. *The Journal of Physical Chemistry Letters* **5**, 1864-1870
62. Brender, J. R., Salamekh, S., and Ramamoorthy, A. (2012) Membrane Disruption and Early Events in the Aggregation of the Diabetes Related Peptide IAPP from a Molecular Perspective. *Accounts of Chemical Research* **45**, 454-462
63. Romero, D., and Kolter, R. (2014) Functional amyloids in bacteria. *International Microbiology* **17**, 65-73
64. Pham, C. L., Kwan, A. H., and Sunde, M. (2014) Functional amyloid: widespread in Nature, diverse in purpose. *Essays in Biochemistry* **56**, 207-219
65. Otzen, D. (2010) Functional amyloid: turning swords into plowshares. *Prion* **4**, 256-264

66. Maury, C. P. (2009) The emerging concept of functional amyloid. *Journal of Internal Medicine* **265**, 329-334
67. Fowler, D. M., Koulov, A. V., Balch, W. E., and Kelly, J. W. (2007) Functional amyloid--from bacteria to humans. *Trends in Biochemical Sciences* **32**, 217-224
68. Fowler, D. M., Koulov, A. V., Alory-Jost, C., Marks, M. S., Balch, W. E., and Kelly, J. W. (2006) Functional amyloid formation within mammalian tissue. *PLoS Biology* **4**, e6
69. Riek, R., and Eisenberg, D. S. (2016) The activities of amyloids from a structural perspective. *Nature* **539**, 227-235
70. Macedo, B., Sant'Anna, R., Navarro, S., Cordeiro, Y., and Ventura, S. (2015) Mammalian prion protein (PrP) forms conformationally different amyloid intracellular aggregates in bacteria. *Microbial Cell Factories* **14**, 174
71. Li, J., McQuade, T., Siemer, A. B., Napetschnig, J., Moriwaki, K., Hsiao, Y. S., Damko, E., Moquin, D., Walz, T., McDermott, A., Chan, F. K., and Wu, H. (2012) The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell* **150**, 339-350
72. Di Martino, P. (2016) Bap: A New Type of Functional Amyloid. *Trends in Microbiology* **24**, 682-684
73. Maji, S. K., Perrin, M. H., Sawaya, M. R., Jessberger, S., Vadodaria, K., Rissman, R. A., Singru, P. S., Nilsson, K. P. R., Simon, R., Schubert, D., Eisenberg, D., Rivier, J., Sawchenko, P., Vale, W., and Riek, R. (2009) Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science* **325**, 328-332
74. Daskalov, A., Habenstein, B., Sabate, R., Berbon, M., Martinez, D., Chaignepain, S., Coulary-Salin, B., Hofmann, K., Loquet, A., and Saupe, S. J. (2016) Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. *Proceedings of the National Academy of Sciences* **113**, 2720-2725
75. Daskalov, A., Habenstein, B., Martinez, D., Debets, A. J., Sabate, R., Loquet, A., and Saupe, S. J. (2015) Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. *PLoS Biology* **13**, e1002059
76. Saupe, S., and Daskalov, A. (2012) The [Het-s] Prion, an Amyloid Fold as a Cell Death Activation Trigger. *PloS Pathogens* **8**
77. Hartman, K., Brender, J. R., Monde, K., Ono, A., Evans, Margery L., Popovych, N., Chapman, M. R., and Ramamoorthy, A. (2013) Bacterial curli protein promotes the conversion of PAP(248-286) into the amyloid SEVI: cross-seeding of dissimilar amyloid sequences. *PeerJ* **1**, e5
78. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C. M. (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proceedings of the National Academy of Sciences* **96**, 3590-3594
79. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426**, 884
80. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D., and Dobson, C. M. (1998) Amyloid fibril formation by an SH3 domain. *Proceedings of the National Academy of Sciences* **95**, 4224-4228
81. Fitzpatrick, A. W. P., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., Crowther, R. A., Ghetti, B., Goedert, M., and Scheres, S. H. W. (2017) Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185-190
82. Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R. B. G., Tusche, M., Lopez-Iglesias, C., Hoyer, W., Heise, H., Willbold, D., and Schröder, G. F. (2017) Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. *Science* **358**, 116-119

83. Linser, R. (2017) Solid-state NMR spectroscopic trends for supramolecular assemblies and protein aggregates. *Solid State Nuclear Magnetic Resonance* **87**, 45-53
84. Weingarth, M., and Baldus, M. (2013) Solid-state NMR-based approaches for supramolecular structure elucidation. *Accounts Of Chemical Research* **46**, 2037-2046
85. Loquet, A., Habenstein, B., and Lange, A. (2013) Structural Investigations of Molecular Machines by Solid-State NMR. *Accounts Of Chemical Research* **46**, 2070-2079
86. Andreas, L. B., Le Marchand, T., Jaudzems, K., and Pintacuda, G. (2015) High-resolution proton-detected NMR of proteins at very fast MAS. *Journal of Magnetic Resonance* **253**, 36-49
87. Tycko, R., and Wickner, R. B. (2013) Molecular structures of amyloid and prion fibrils: consensus versus controversy. *Accounts Of Chemical Research* **46**, 1487-1496
88. Tycko, R. (2011) Solid-state NMR studies of amyloid fibril structure. *Annual Review of Physical Chemistry* **62**, 279-299
89. Meier, B. H., and Böckmann, A. (2015) The structure of fibrils from 'misfolded' proteins. *Current Opinion in Structural Biology* **30**, 43-49
90. Comellas, G., and Rienstra, C. M. (2013) Protein structure determination by magic-angle spinning solid-state NMR, and insights into the formation, structure, and stability of amyloid fibrils. *Annual Review of Biophysics* **42**, 515-536
91. Tuttle, M. D., Comellas, G., Nieuwkoop, A. J., Covell, D. J., Berthold, D. A., Kloepper, K. D., Courtney, J. M., Kim, J. K., Barclay, A. M., Kendall, A., Wan, W., Stubbs, G., Schwieters, C. D., Lee, V. M., George, J. M., and Rienstra, C. M. (2016) Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nature Structural & Molecular Biology* **23**, 409-415
92. Walti, M. A., Ravotti, F., Arai, H., Glabe, C. G., Wall, J. S., Bockmann, A., Guntert, P., Meier, B. H., and Riek, R. (2016) Atomic-resolution structure of a disease-relevant Abeta(1-42) amyloid fibril. *Proceedings of the National Academy of Sciences* **113**, E4976-4984
93. Colvin, M. T., Silvers, R., Ni, Q. Z., Can, T. V., Sergeyev, I., Rosay, M., Donovan, K. J., Michael, B., Wall, J., Linse, S., and Griffin, R. G. (2016) Atomic Resolution Structure of Monomorphic Abeta42 Amyloid Fibrils. *Journal of the American Chemical Society* **138**, 9663-9674
94. Lu, J. X., Qiang, W., Yau, W. M., Schwieters, C. D., Meredith, S. C., and Tycko, R. (2013) Molecular structure of beta-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **154**, 1257-1268
95. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. (2002) A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences* **99**, 16742-16747
96. Paravastu, A. K., Leapman, R. D., Yau, W., and Tycko, R. (2008) Molecular structural basis for polymorphism in Alzheimer's beta -amyloid fibrils. *Proceedings of the National Academy of Sciences* **105**, 18349-18354
97. Meier, B. H., Riek, R., and Bockmann, A. (2017) Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR. *Trends in Biochemical Sciences*
98. Sgourakis, N. G., Yau, W. M., and Qiang, W. (2015) Modeling an in-register, parallel "iowa" abeta fibril structure using solid-state NMR data from labeled samples with rosetta. *Structure* **23**, 216-227
99. Schutz, A. K., Vagt, T., Huber, M., Ovchinnikova, O. Y., Cadalbert, R., Wall, J., Guntert, P., Bockmann, A., Glockshuber, R., and Meier, B. H. (2015) Atomic-

resolution three-dimensional structure of amyloid beta fibrils bearing the Osaka mutation. *Angewandte Chemie International Edition* **54**, 331-335
100. Raibaut, L., El Mahdi, O., and Melnyk, O. (2015) Solid Phase proteins Chemical Synthesis. *Topics in Current Chemistry* **363**, 103-154
101. Guan, X., Chaffey, P. K., Zeng, C., and Tan, Z. (2015) New Methods for Chemical Protein Synthesis. *Topics in Current Chemistry* **363**, 155-192
102. Eisenberg, D. S., and Sawaya, M. R. (2017) Structural Studies of Amyloid Proteins at the Molecular Level. *Annual Review of Biochemistry* **86**, 69-95
103. Qiang, W., Yau, W.-M., Lu, J.-X., Collinge, J., and Tycko, R. (2017) Structural variation in amyloid-β fibrils from Alzheimer's disease clinical subtypes. *Nature* **541**, 217
104. Cohen, M. L., Kim, C., Haldiman, T., ElHag, M., Mehndiratta, P., Pichet, T., Lissemore, F., Shea, M., Cohen, Y., Chen, W., Blevins, J., Appleby, B. S., Surewicz, K., Surewicz, W. K., Sajatovic, M., Tatsuoka, C., Zhang, S., Mayo, P., Butkiewicz, M., Haines, J. L., Lerner, A. J., and Safar, J. G. (2015) Rapidly progressive Alzheimer's disease features distinct structures of amyloid-β. *Brain* **138**, 1009-1022
105. Collinge, J., and Clarke, A. R. (2007) A General Model of Prion Strains and Their Pathogenicity. *Science* **318**, 930-936
106. Safar, J., Wille, H., Itri, V., Groth, D., Serban, H., Torchia, M., Cohen, F. E., and Prusiner, S. B. (1998) Eight prion strains have PrPSc molecules with different conformations. *Nature Medicine* **4**, 1157
107. Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W.-M., Mattson, M. P., and Tycko, R. (2005) Self-Propagating, Molecular-Level Polymorphism in Alzheimer's ß-Amyloid Fibrils. *Science* **307**, 262-265
108. Gath, J., Bousset, L., Habenstein, B., Melki, R., Böckmann, A., and Meier, B. H. (2014) Unlike Twins: An NMR Comparison of Two α-Synuclein Polymorphs Featuring Different Toxicity. *PLoS One* **9**, e90659
109. Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., Madiona, K., Olieric, V., Böckmann, A., Meier, B. H., and Melki, R. (2013) Structural and functional characterization of two alpha-synuclein strains. *Nature Communications* **4**, 2575
110. Lacabanne, D., Meier, B. H., and Böckmann, A. (2017) Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. *Journal of Biomolecular NMR*
111. Higman, V. A., Flinders, J., Hiller, M., Jehle, S., Markovic, S., Fiedler, S., van Rossum, B. J., and Oschkinat, H. (2009) Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins. *Journal of Biomolecular NMR* **44**, 245-260
112. Hong, M. (1999) Determination of Multiple φ-Torsion Angles in Proteins by Selective and Extensive 13C Labeling and Two-Dimensional Solid-State NMR. *Journal of Magnetic Resonance* **139**, 389-401
113. Lundstrom, P., Teilum, K., Carstensen, T., Bezsonova, I., Wiesner, S., Hansen, D. F., Religa, T. L., Akke, M., and Kay, L. E. (2007) Fractional 13C enrichment of isolated carbons using [1-13C]- or [2- 13C]-glucose facilitates the accurate measurement of dynamics at backbone Calpha and side-chain methyl positions in proteins. *Journal of Biomolecular NMR* **38**, 199-212
114. Agarwal, V., Penzel, S., Szekely, K., Cadalbert, R., Testori, E., Oss, A., Past, J., Samoson, A., Ernst, M., Bockmann, A., and Meier, B. H. (2014) De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy. *Angew Chem Int Ed Engl* **53**, 12253-12256

115. Noirot, C., Habenstein, B., Bousset, L., Melki, R., Meier, B. H., Endo, Y., Penin, F., and Bockmann, A. (2010) Wheat-germ cell-free production of prion proteins for solid-state NMR structural studies. *N Biotechnol*
116. Paulson, E. K., Morcombe, C. R., Gaponenko, V., Dancheck, B., Byrd, R. A., and Zilm, K. W. (2003) Sensitive High Resolution Inverse Detection NMR Spectroscopy of Proteins in the Solid State. *Journal of the American Chemical Society* **125**, 15831-15836
117. Asami, S., Schmieder, P., and Reif, B. (2010) High Resolution 1H-Detected Solid-State NMR Spectroscopy of Protein Aliphatic Resonances: Access to Tertiary Structure Information. *Journal of the American Chemical Society* **132**, 15133-15135
118. Chevelkov, V., Rehbein, K., Diehl, A., and Reif, B. (2006) Ultrahigh Resolution in Proton Solid-State NMR Spectroscopy at High Levels of Deuteration. *Angewandte Chemie International Edition* **45**, 3878-3881
119. Linser, R., Dasari, M., Hiller, M., Higman, V., Fink, U., Lopez del Amo, J.-M., Markovic, S., Handel, L., Kessler, B., Schmieder, P., Oesterhelt, D., Oschkinat, H., and Reif, B. (2011) Proton-Detected Solid-State NMR Spectroscopy of Fibrillar and Membrane Proteins. *Angewandte Chemie International Edition* **50**, 4508-4512
120. Paulson, E. K., Morcombe, C. R., Gaponenko, V., Dancheck, B., Byrd, R. A., and Zilm, K. W. (2003) High-Sensitivity Observation of Dipolar Exchange and NOEs between Exchangeable Protons in Proteins by 3D Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **125**, 14222-14223
121. Zhou, D. H., Shea, J. J., Nieuwkoop, A. J., Franks, W. T., Wylie, B. J., Mullen, C., Sandoz, D., and Rienstra, C. M. (2007) Solid-State Protein Structure Determination with Proton-Detected Triple Resonance 3D Magic-Angle Spinning NMR Spectroscopy. *Angewandte Chemie International Edition* **46**, 8380-8383
122. Knight, M. J., Webber, A. L., Pell, A. J., Guerry, P., Barbet-Massin, E., Bertini, I., Felli, I. C., Gonnelli, L., Pierattelli, R., Emsley, L., Lesage, A., Herrmann, T., and Pintacuda, G. (2011) Fast Resonance Assignment and Fold Determination of Human Superoxide Dismutase by High-Resolution Proton-Detected Solid-State MAS NMR Spectroscopy. *Angewandte Chemie International Edition* **50**, 11697-11701
123. Retel, J. S., Nieuwkoop, A. J., Hiller, M., Higman, V. A., Barbet-Massin, E., Stanek, J., Andreas, L. B., Franks, W. T., van Rossum, B. J., Vinothkumar, K. R., Handel, L., de Palma, G. G., Bardiaux, B., Pintacuda, G., Emsley, L., Kuhlbrandt, W., and Oschkinat, H. (2017) Structure of outer membrane protein G in lipid bilayers. *Nature Communications* **8**, 2073
124. Linser, R., Bardiaux, B., Higman, V., Fink, U., and Reif, B. (2011) Structure Calculation from Unambiguous Long-Range Amide and Methyl 1H−1H Distance Restraints for a Microcrystalline Protein with MAS Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **133**, 5905-5912
125. Huber, M., Hiller, S., Schanda, P., Ernst, M., Böckmann, A., Verel, R., and Meier, B. H. (2011) A Proton-Detected 4D Solid-State NMR Experiment for Protein Structure Determination. *ChemPhysChem* **12**, 915-918
126. Agarwal, V., Penzel, S., Szekely, K., Cadalbert, R., Testori, E., Oss, A., Past, J., Samoson, A., Ernst, M., Böckmann, A., and Meier, B. H. (2014) De Novo 3D Structure Determination from Sub-milligram Protein Samples by Solid-State 100 kHz MAS NMR Spectroscopy. *Angewandte Chemie International Edition* **53**, 12253-12256
127. Sinnige, T., Daniëls, M., Baldus, M., and Weingarth, M. (2014) Proton Clouds to Measure Long-Range Contacts between Nonexchangeable Side Chain Protons in Solid-State NMR. *Journal of the American Chemical Society* **136**, 4452-4455

128. Andreas, L. B., Reese, M., Eddy, M. T., Gelev, V., Ni, Q. Z., Miller, E. A., Emsley, L., Pintacuda, G., Chou, J. J., and Griffin, R. G. (2015) Structure and Mechanism of the Influenza A M218–60 Dimer of Dimers. *Journal of the American Chemical Society* **137**, 14877-14886
129. Knight, M. J., Pell, A. J., Bertini, I., Felli, I. C., Gonnelli, L., Pierattelli, R., Herrmann, T., Emsley, L., and Pintacuda, G. (2012) Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. *Proceedings of the National Academy of Sciences* **109**, 11095-11100
130. Knight, M. J., Felli, I. C., Pierattelli, R., Bertini, I., Emsley, L., Herrmann, T., and Pintacuda, G. (2012) Rapid Measurement of Pseudocontact Shifts in Metalloproteins by Proton-Detected Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **134**, 14730-14733
131. Shi, L., Kawamura, I., Jung, K.-H., Brown, L. S., and Ladizhansky, V. (2011) Conformation of a Seven-Helical Transmembrane Photosensor in the Lipid Environment. *Angewandte Chemie International Edition* **50**, 1302-1305
132. Medeiros-Silva, J., Mance, D., Daniëls, M., Jekhmane, S., Houben, K., Baldus, M., and Weingarth, M. (2016) 1H-Detected Solid-State NMR Studies of Water-Inaccessible Proteins In Vitro and In Situ. *Angewandte Chemie International Edition* **55**, 13606-13610
133. Andreas, L. B., Jaudzems, K., Stanek, J., Lalli, D., Bertarello, A., Le Marchand, T., Cala-De Paepe, D., Kotelovica, S., Akopjana, I., Knott, B., Wegner, S., Engelke, F., Lesage, A., Emsley, L., Tars, K., Herrmann, T., and Pintacuda, G. (2016) Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proceedings of the National Academy of Sciences* **113**, 9187-9192
134. Smith, A. A., Ravotti, F., Testori, E., Cadalbert, R., Ernst, M., Bockmann, A., and Meier, B. H. (2017) Partially-deuterated samples of HET-s(218-289) fibrils: assignment and deuterium isotope effect. *Journal of Biomolecular NMR* **67**, 109-119
135. Stanek, J., Andreas, L. B., Jaudzems, K., Cala, D., Lalli, D., Bertarello, A., Schubeis, T., Akopjana, I., Kotelovica, S., Tars, K., Pica, A., Leone, S., Picone, D., Xu, Z. Q., Dixon, N. E., Martinez, D., Berbon, M., El Mammeri, N., Noubhani, A., Saupe, S., Habenstein, B., Loquet, A., and Pintacuda, G. (2016) NMR Spectroscopic Assignment of Backbone and Side-Chain Protons in Fully Protonated Proteins: Microcrystals, Sedimented Assemblies, and Amyloid Fibrils. *Angewandte Chemie International Edition* **55**, 15504-15509
136. Jain, M. G., Lalli, D., Stanek, J., Gowda, C., Prakash, S., Schwarzer, T. S., Schubeis, T., Castiglione, K., Andreas, L. B., Madhu, P. K., Pintacuda, G., and Agarwal, V. (2017) Selective 1H–1H Distance Restraints in Fully Protonated Proteins by Very Fast Magic-Angle Spinning Solid-State NMR. *The Journal of Physical Chemistry Letters* **8**, 2399-2405
137. Lalli, D., Idso, M. N., Andreas, L. B., Hussain, S., Baxter, N., Han, S., Chmelka, B. F., and Pintacuda, G. (2017) Proton-Based Structural Analysis of a Heptahelical Transmembrane Protein in Lipid Bilayers. *Journal of the American Chemical Society* **139**, 13006-13012
138. Struppe, J., Quinn, C. M., Lu, M., Wang, M., Hou, G., Lu, X., Kraus, J., Andreas, L. B., Stanek, J., Lalli, D., Lesage, A., Pintacuda, G., Maas, W., Gronenborn, A. M., and Polenova, T. (2017) Expanding the horizons for structural analysis of fully protonated protein assemblies by NMR spectroscopy at MAS frequencies above 100 kHz. *Solid State Nuclear Magnetic Resonance* **87**, 117-125

139. Cala-De Paepe, D., Stanek, J., Jaudzems, K., Tars, K., Andreas, L. B., and Pintacuda, G. (2017) Is protein deuteration beneficial for proton detected solid-state NMR at and above 100 kHz magic-angle spinning? *Solid State Nuclear Magnetic Resonance*
140. Schubeis, T., Le Marchand, T., Andreas, L. B., and Pintacuda, G. (2018) 1H NMR in solids evolves as a powerful new tool for membrane proteins. *Journal of Magnetic Resonance*
141. Asami, S., Rakwalska-Bange, M., Carlomagno, T., and Reif, B. (2013) Protein–RNA Interfaces Probed by 1H-Detected MAS Solid-State NMR Spectroscopy. *Angewandte Chemie International Edition* **52**, 2345-2349
142. Luca, S., Filippov, D. V., van Boom, J. H., Oschkinat, H., de Groot, H. J. M., and Baldus, M. (2001) Secondary chemical shifts in immobilized peptides and proteins: A qualitative basis for structure refinement under Magic Angle Spinning. *Journal of Biomolecular NMR* **20**, 325-331
143. Bayro, M. J., Huber, M., Ramachandran, R., Davenport, T. C., Meier, B. H., Ernst, M., and Griffin, R. G. (2009) Dipolar truncation in magic-angle spinning NMR recoupling experiments. *The Journal of Chemical Physics* **130**, 114506
144. Grommek, A., Meier, B. H., and Ernst, M. (2006) Distance information from proton-driven spin diffusion under MAS. *Chemical Physics Letters* **427**, 404-409
145. Hodgkinson, P., and Emsley, L. (1999) The accuracy of distance measurements in solid-state NMR. *Journal of Magnetic Resonance* **139**, 46-59
146. Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., and Oschkinat, H. (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98
147. Loquet, A., Giller, K., Becker, S., and Lange, A. (2010) Supramolecular Interactions Probed by 13C−13C Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **132**, 15164-15166
148. Etzkorn, M., Bockmann, A., Lange, A., and Baldus, M. (2004) Probing molecular interfaces using 2D magic-angle-spinning NMR on protein mixtures with different uniform labeling. *Journal of the American Chemical Society* **126**, 14746-14751
149. Debelouchina, G. T., Platt, G. W., Bayro, M. J., Radford, S. E., and Griffin, R. G. (2010) Intermolecular Alignment in β(2)-Microglobulin Amyloid Fibrils. *Journal of the American Chemical Society* **132**, 17077-17079
150. Castellani, F., van Rossum, B.-J., Diehl, A., Rehbein, K., and Oschkinat, H. (2003) Determination of Solid-State NMR Structures of Proteins by Means of Three-Dimensional 15N−13C−13C Dipolar Correlation Spectroscopy and Chemical Shift Analysis. *Biochemistry* **42**, 11476-11483
151. Habenstein, B., Loquet, A., Giller, K., Becker, S., and Lange, A. (2013) Structural characterization of supramolecular assemblies by 13C spin dilution and 3D solid-state NMR. *Journal of Biomolecular NMR* **55**, 1-9
152. Loquet, A., Lv, G., Giller, K., Becker, S., and Lange, A. (2011) 13C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies. *Journal of the American Chemical Society* **133**, 4722-4725
153. Sen, A., Baxa, U., Simon, M. N., Wall, J. S., Sabate, R., Saupe, S. J., and Steven, A. C. (2007) Mass analysis by scanning transmission electron microscopy and electron diffraction validate predictions of stacked beta-solenoid model of HET-s prion fibrils. *Journal of Biological Chemistry* **282**, 5545-5550
154. Goldsbury, C., Baxa, U., Simon, M. N., Steven, A. C., Engel, A., Wall, J. S., Aebi, U., and Muller, S. A. (2011) Amyloid structure and assembly: insights from scanning transmission electron microscopy. *Journal of Structural Biology* **173**, 1-13

155. Linser, R., Bardiaux, B., Andreas, L. B., Hyberts, S. G., Morris, V. K., Pintacuda, G., Sunde, M., Kwan, A. H., and Wagner, G. (2014) Solid-State NMR Structure Determination from Diagonal-Compensated, Sparsely Nonuniform-Sampled 4D Proton–Proton Restraints. *Journal of the American Chemical Society* **136**, 11002-11010
156. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular Nmr* **44**, 213-223
157. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program Dyana11Edited by P. E. Wright. *Journal of Molecular Biology* **273**, 283-298
158. Guerry, P., and Herrmann, T. (2012) Comprehensive Automation for NMR Structure Determination of Proteins. In *Protein NMR Techniques* (Shekhtman, A., and Burz, D. S., eds) pp. 429-451, Humana Press, Totowa, NJ
159. Bardiaux, B., Malliavin, T., and Nilges, M. (2012) ARIA for Solution and Solid-State NMR. In *Protein NMR Techniques* (Shekhtman, A., and Burz, D. S., eds) pp. 453-483, Humana Press, Totowa, NJ
160. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Marius Clore, G. (2003) The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **160**, 65-73
161. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and André, I. (2011) Modeling Symmetric Macromolecular Structures in Rosetta3. *PLoS One* **6**, e20450
162. Ashbrook, S. E., Griffin, J. M., and Johnston, K. E. (2018) Recent Advances in Solid-State Nuclear Magnetic Resonance Spectroscopy. *Annual Review of Analytical Chemistry* **11**, 485-508
163. Zech, S. G., Wand, A. J., and McDermott, A. E. (2005) Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *J. Am. Chem. Soc.* **127**, 8618-8626
164. Luca, S., Heise, H., and Baldus, M. (2003) High-resolution solid-state NMR applied to polypeptides and membrane proteins. *Acc. Chem. Res.* **36**, 858-865
165. Heise, H., Seidel, K., Etzkorn, M., Becker, S., and Baldus, M. (2005) 3D NMR spectroscopy for resonance assignment and structure elucidation of proteins under MAS: novel pulse schemes and sensitivity considerations. *Journal of Magnetic Resonance* **173**, 64-74
166. Renault, M., Cukkemane, A., and Baldus, M. (2010) Solid-state NMR spectroscopy on complex biomolecules. *Angew Chem Int Ed Engl* **49**, 8346-8357
167. Baldus, M. (2006) Solid-state NMR spectroscopy: molecular structure and organization at the atomic level. *Angew. Chem. Int. Ed.* **45**, 1186-1188
168. Loquet, A., Habenstein, B., and Lange, A. (2013) Structural Investigations of Molecular Machines by Solid-State NMR. *Acc Chem Res*
169. Heise, H., Seidel, K., Etzkorn, M., Becker, S., and Baldus, M. (2005) 3D NMR spectroscopy for resonance assignment and structure elucidation of proteins under MAS: novel pulse schemes and sensitivity consideration. *J. Magn. Reson.* **173**, 64-74
170. Tang, M., Comellas, G., and Rienstra, C. M. (2013) Advanced Solid-State NMR Approaches for Structure Determination of Membrane Proteins and Amyloid Fibrils. *Acc Chem Res*
171. Venkatesh, A., Ryan, M. J., Biswas, A., Boteju, K. C., Sadow, A. D., and Rossini, A. J. (2018) Enhancing the Sensitivity of Solid-State NMR Experiments with Very Low Gyromagnetic Ratio Nuclei with Fast Magic Angle Spinning and Proton Detection. *The Journal of Physical Chemistry A* **122**, 5635-5643

172. Jaudzems, K., Polenova, T., Pintacuda, G., Oschkinat, H., and Lesage, A. (2019) DNP NMR of biomolecular assemblies. *Journal of Structural Biology* **206**, 90-98
173. Mandala, V. S., and Hong, M. (2019) High-sensitivity protein solid-state NMR spectroscopy. *Current Opinion in Structural Biology* **58**, 183-190
174. Rankin, A. G. M., Trébosc, J., Pourpoint, F., Amoureux, J.-P., and Lafon, O. (2019) Recent developments in MAS DNP-NMR of materials. *Solid State Nuclear Magnetic Resonance* **101**, 116-143
175. McDermott, A. E. (2009) Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR. *Annu. Rev. Biophys.* **38**, 385
176. Rossini, A. J., Zagdoun, A., Hegner, F., Schwarzwalder, M., Gajan, D., Coperet, C., Lesage, A., and Emsley, L. (2012) Dynamic nuclear polarization NMR spectroscopy of microcrystalline solids. *J Am Chem Soc* **134**, 16899-16908
177. Tolchard, J., Pandey, M. K., Berbon, M., Noubhani, A., Saupe, S. J., Nishiyama, Y., Habenstein, B., and Loquet, A. (2018) Detection of side-chain proton resonances of fully protonated biosolids in nano-litre volumes by magic angle spinning solid-state NMR. *J Biomol NMR* **70**, 177-185
178. Schubeis, T., Le Marchand, T., Andreas, L. B., and Pintacuda, G. (2018) 1H magic-angle spinning NMR evolves as a powerful new tool for membrane proteins. *Journal of Magnetic Resonance* **287**, 140-152
179. Zhang, R., Mroue, K. H., and Ramamoorthy, A. (2017) Proton-Based Ultrafast Magic Angle Spinning Solid-State NMR Spectroscopy. *Accounts of Chemical Research* **50**, 1105-1113
180. Stanek, J., Andreas, L. B., Jaudzems, K., Cala, D., Lalli, D., Bertarello, A., Schubeis, T., Akopjana, I., Kotelovica, S., Tars, K., Pica, A., Leone, S., Picone, D., Xu, Z. Q., Dixon, N. E., Martinez, D., Berbon, M., El Mammeri, N., Noubhani, A., Saupe, S., Habenstein, B., Loquet, A., and Pintacuda, G. (2016) NMR Spectroscopic Assignment of Backbone and Side-Chain Protons in Fully Protonated Proteins: Microcrystals, Sedimented Assemblies, and Amyloid Fibrils. *Angew Chem Int Ed Engl* **55**, 15504-15509
181. Sinnige, T., Daniels, M., Baldus, M., and Weingarth, M. (2014) Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J Am Chem Soc* **136**, 4452-4455
182. Barbet-Massin, E., Pell, A. J., Jaudzems, K., Franks, W. T., Retel, J. S., Kotelovica, S., Akopjana, I., Tars, K., Emsley, L., Oschkinat, H., Lesage, A., and Pintacuda, G. (2013) Out-and-back 13C-13C scalar transfers in protein resonance assignment by proton-detected solid-state NMR under ultra-fast MAS. *J Biomol NMR* **56**, 379-386
183. Zhou, D. H., Nieuwkoop, A. J., Berthold, D. A., Comellas, G., Sperling, L. J., Tang, M., Shah, G. J., Brea, E. J., Lemkau, L. R., and Rienstra, C. M. (2012) Solid-state NMR analysis of membrane proteins and protein aggregates by proton detected spectroscopy. *J Biomol NMR*
184. Knight, M. J., Felli, I. C., Pierattelli, R., Bertini, I., Emsley, L., Herrmann, T., and Pintacuda, G. (2012) Rapid measurement of pseudocontact shifts in metalloproteins by proton-detected solid-state NMR spectroscopy. *J Am Chem Soc* **134**, 14730-14733
185. Ward, M. E., Shi, L., Lake, E., Krishnamurthy, S., Hutchins, H., Brown, L. S., and Ladizhansky, V. (2011) Proton-detected solid-state NMR reveals intramembrane polar networks in a seven-helical transmembrane protein proteorhodopsin. *J Am Chem Soc* **133**, 17434-17443
186. Linser, R., Dasari, M., Hiller, M., Higman, V., Fink, U., Lopez del Amo, J. M., Markovic, S., Handel, L., Kessler, B., Schmieder, P., Oesterhelt, D., Oschkinat, H.,

and Reif, B. (2011) Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins. *Angew Chem Int Ed Engl* **50**, 4508-4512
187. Hansen, M. R., Graf, R., and Spiess, H. W. (2013) Solid-state NMR in macromolecular systems: insights on how molecular entities move. *Acc Chem Res* **46**, 1996-2007
188. Lee, D., Hediger, S., and De Paëpe, G. (2015) Is solid-state NMR enhanced by dynamic nuclear polarization? *Solid State Nuclear Magnetic Resonance* **66-67**, 6-20
189. Griffin, R. G., and Prisner, T. F. (2010) High field dynamic nuclear polarization—the renaissance. *Physical Chemistry Chemical Physics* **12**, 5737-5740
190. Can, T. V., Ni, Q. Z., and Griffin, R. G. (2015) Mechanisms of dynamic nuclear polarization in insulating solids. *Journal of Magnetic Resonance* **253**, 23-35
191. Ni, Q. Z., Daviso, E., Can, T. V., Markhasin, E., Jawla, S. K., Swager, T. M., Temkin, R. J., Herzfeld, J., and Griffin, R. G. (2013) High Frequency Dynamic Nuclear Polarization. *Accounts of Chemical Research* **46**, 1933-1941
192. Maly, T., Debelouchina, G. T., Bajaj, V. S., Hu, K. N., Joo, C. G., Mak-Jurkauskas, M. L., Sirigiri, J. R., van der Wel, P. C., Herzfeld, J., Temkin, R. J., and Griffin, R. G. (2008) Dynamic nuclear polarization at high magnetic fields. *J Chem Phys* **128**, 052211
193. Mentink-Vigier, F., Akbey, Ü., Oschkinat, H., Vega, S., and Feintuch, A. (2015) Theoretical aspects of Magic Angle Spinning - Dynamic Nuclear Polarization. *Journal of Magnetic Resonance* **258**, 102-120
194. Narasimhan, S., Scherpe, S., Lucini Paioni, A., van der Zwan, J., Folkers, G. E., Ovaa, H., and Baldus, M. (2019) DNP-Supported Solid-State NMR Spectroscopy of Proteins Inside Mammalian Cells. *Angewandte Chemie International Edition* **58**, 12969-12973
195. Mentink-Vigier, F., Mathies, G., Liu, Y., Barra, A.-L., Caporini, M. A., Lee, D., Hediger, S., G. Griffin, R., and De Paëpe, G. (2017) Efficient cross-effect dynamic nuclear polarization without depolarization in high-resolution MAS NMR. *Chemical Science* **8**, 8150-8163
196. Lilly Thankamony, A. S., Wittmann, J. J., Kaushik, M., and Corzilius, B. (2017) Dynamic nuclear polarization for sensitivity enhancement in modern solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **102-103**, 120-195
197. Akbey, Ü., and Oschkinat, H. (2016) Structural biology applications of solid state MAS DNP NMR. *Journal of Magnetic Resonance* **269**, 213-224
198. Smith, A. N., Twahir, U. T., Dubroca, T., Fanucci, G. E., and Long, J. R. (2016) Molecular Rationale for Improved Dynamic Nuclear Polarization of Biomembranes. *The Journal of Physical Chemistry B* **120**, 7880-7888
199. Manjunatha Reddy, G. N., Gerbec, J. A., Shimizu, F., and Chmelka, B. F. (2019) Nanoscale Surface Compositions and Structures Influence Boron Adsorption Properties of Anion Exchange Resins. *Langmuir*
200. Sangodkar, R. P., Smith, B. J., Gajan, D., Rossini, A. J., Roberts, L. R., Funkhouser, G. P., Lesage, A., Emsley, L., and Chmelka, B. F. (2015) Influences of Dilute Organic Adsorbates on the Hydration of Low-Surface-Area Silicates. *Journal of the American Chemical Society* **137**, 8096-8112
201. Lelli, M., Gajan, D., Lesage, A., Caporini, M. A., Vitzthum, V., Miéville, P., Héroguel, F., Rascón, F., Roussey, A., Thieuleux, C., Boualleg, M., Veyre, L., Bodenhausen, G., Coperet, C., and Emsley, L. (2011) Fast Characterization of Functionalized Silica Materials by Silicon-29 Surface-Enhanced NMR Spectroscopy Using Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **133**, 2104-2107

202. Ong, T.-C., Mak-Jurkauskas, M. L., Walish, J. J., Michaelis, V. K., Corzilius, B., Smith, A. A., Clausen, A. M., Cheetham, J. C., Swager, T. M., and Griffin, R. G. (2013) Solvent-Free Dynamic Nuclear Polarization of Amorphous and Crystalline ortho-Terphenyl. *The Journal of Physical Chemistry B* **117**, 3040-3046
203. Lesage, A., Lelli, M., Gajan, D., Caporini, M. A., Vitzthum, V., Mieville, P., Alauzun, J., Roussey, A., Thieuleux, C., Mehdi, A., Bodenhausen, G., Coperet, C., and Emsley, L. (2010) Surface enhanced NMR spectroscopy by dynamic nuclear polarization. *J Am Chem Soc* **132**, 15459-15461
204. Rossini, A. J., Zagdoun, A., Lelli, M., Canivet, J., Aguado, S., Ouari, O., Tordo, P., Rosay, M., Maas, W. E., Coperet, C., Farrusseng, D., Emsley, L., and Lesage, A. (2012) Dynamic nuclear polarization enhanced solid-state NMR spectroscopy of functionalized metal-organic frameworks. *Angew Chem Int Ed Engl* **51**, 123-127
205. Vitzthum, V., Mieville, P., Carnevale, D., Caporini, M. A., Gajan, D., Coperet, C., Lelli, M., Zagdoun, A., Rossini, A. J., Lesage, A., Emsley, L., and Bodenhausen, G. (2012) Dynamic nuclear polarization of quadrupolar nuclei using cross polarization from protons: surface-enhanced aluminium-27 NMR. *Chem Commun (Camb)* **48**, 1988-1990
206. Rossini, A. J., Zagdoun, A., Hegner, F., Schwarzwälder, M., Gajan, D., Copéret, C., Lesage, A., and Emsley, L. (2012) Dynamic Nuclear Polarization NMR Spectroscopy of Microcrystalline Solids. *Journal of the American Chemical Society* **134**, 16899-16908
207. Barnes, A. B., Paepe, G. D., van der Wel, P. C., Hu, K. N., Joo, C. G., Bajaj, V. S., Mak-Jurkauskas, M. L., Sirigiri, J. R., Herzfeld, J., Temkin, R. J., and Griffin, R. G. (2008) High-Field Dynamic Nuclear Polarization for Solid and Solution Biological NMR. *Appl Magn Reson* **34**, 237-263
208. Rosay, M., Weis, V., Kreischer, K. E., Temkin, R. J., and Griffin, R. G. (2002) Two-Dimensional 13C−13C Correlation Spectroscopy with Magic Angle Spinning and Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **124**, 3214-3215
209. Rosay, M., Tometich, L., Pawsey, S., Bader, R., Schauwecker, R., Blank, M., Borchard, P. M., Cauffman, S. R., Felch, K. L., Weber, R. T., Temkin, R. J., Griffin, R. G., and Maas, W. E. (2010) Solid-state dynamic nuclear polarization at 263 GHz: spectrometer design and experimental results. *Physical Chemistry Chemical Physics* **12**, 5850-5860
210. van der Wel, P. C., Hu, K. N., Lewandowski, J., and Griffin, R. G. (2006) Dynamic nuclear polarization of amyloidogenic peptide nanocrystals: GNNQQNY, a core segment of the yeast prion protein Sup35p. *J Am Chem Soc* **128**, 10840-10846
211. Rosay, M., Lansing, J. C., Haddad, K. C., Bachovchin, W. W., Herzfeld, J., Temkin, R. J., and Griffin, R. G. (2003) High-frequency dynamic nuclear polarization in MAS spectra of membrane and soluble proteins. *J Am Chem Soc* **125**, 13626-13627
212. Bajaj, V. S., Farrar, C. T., Hornstein, M. K., Mastovsky, I., Vieregg, J., Bryant, J., Eléna, B., Kreischer, K. E., Temkin, R. J., and Griffin, R. G. (2003) Dynamic nuclear polarization at 9T using a novel 250GHz gyrotron microwave source. *Journal of Magnetic Resonance* **160**, 85-90
213. Sauvée, C., Rosay, M., Casano, G., Aussenac, F., Weber, R. T., Ouari, O., and Tordo, P. (2013) Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angewandte Chemie International Edition* **52**, 10858-10861
214. Zagdoun, A., Casano, G., Ouari, O., Schwarzwälder, M., Rossini, A. J., Aussenac, F., Yulikov, M., Jeschke, G., Copéret, C., Lesage, A., Tordo, P., and Emsley, L. (2013)

Large Molecular Weight Nitroxide Biradicals Providing Efficient Dynamic Nuclear Polarization at Temperatures up to 200 K. *Journal of the American Chemical Society* **135**, 12790-12797
215. Haze, O., Corzilius, B., Smith, A. A., Griffin, R. G., and Swager, T. M. (2012) Water-Soluble Narrow-Line Radicals for Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **134**, 14287-14290
216. Kiesewetter, M. K., Corzilius, B., Smith, A. A., Griffin, R. G., and Swager, T. M. (2012) Dynamic Nuclear Polarization with a Water-Soluble Rigid Biradical. *Journal of the American Chemical Society* **134**, 4537-4540
217. Dane, E. L., Corzilius, B., Rizzato, E., Stocker, P., Maly, T., Smith, A. A., Griffin, R. G., Ouari, O., Tordo, P., and Swager, T. M. (2012) Rigid Orthogonal Bis-TEMPO Biradicals with Improved Solubility for Dynamic Nuclear Polarization. *The Journal of Organic Chemistry* **77**, 1789-1797
218. Song, C., Hu, K. N., Joo, C. G., Swager, T. M., and Griffin, R. G. (2006) TOTAPOL: a biradical polarizing agent for dynamic nuclear polarization experiments in aqueous media. *J Am Chem Soc* **128**, 11385-11390
219. Gelis, I., Vitzthum, V., Dhimole, N., Caporini, M. A., Schedlbauer, A., Carnevale, D., Connell, S. R., Fucini, P., and Bodenhausen, G. (2013) Solid-state NMR enhanced by dynamic nuclear polarization as a novel tool for ribosome structural biology. *Journal of Biomolecular NMR* **56**, 85-93
220. Smith, A. N., Märker, K., Piretra, T., Boatz, J. C., Matlahov, I., Kodali, R., Hediger, S., van der Wel, P. C. A., and De Paëpe, G. (2018) Structural Fingerprinting of Protein Aggregates by Dynamic Nuclear Polarization-Enhanced Solid-State NMR at Natural Isotopic Abundance. *Journal of the American Chemical Society* **140**, 14576-14580
221. Wiegand, T., Liao, W.-C., Ong, T. C., Däpp, A., Cadalbert, R., Copéret, C., Böckmann, A., and Meier, B. H. (2017) Protein–nucleotide contacts in motor proteins detected by DNP-enhanced solid-state NMR. *Journal of Biomolecular NMR* **69**, 157-164
222. Lehnert, E., Mao, J., Mehdipour, A. R., Hummer, G., Abele, R., Glaubitz, C., and Tampé, R. (2016) Antigenic Peptide Recognition on the Human ABC Transporter TAP Resolved by DNP-Enhanced Solid-State NMR Spectroscopy. *Journal of the American Chemical Society* **138**, 13967-13974
223. Fricke, P., Mance, D., Chevelkov, V., Giller, K., Becker, S., Baldus, M., and Lange, A. (2016) High resolution observed in 800 MHz DNP spectra of extremely rigid type III secretion needles. *Journal of Biomolecular NMR* **65**, 121-126
224. Singh, C., Rai, R. K., Aussenac, F., and Sinha, N. (2014) Direct Evidence of Imino Acid–Aromatic Interactions in Native Collagen Protein by DNP-Enhanced Solid-State NMR Spectroscopy. *The Journal of Physical Chemistry Letters* **5**, 4044-4048
225. Fricke, P., Demers, J.-P., Becker, S., and Lange, A. (2014) Studies on the MxiH Protein in T3SS Needles Using DNP-Enhanced ssNMR Spectroscopy. *ChemPhysChem* **15**, 57-60
226. Ong, Y. S., Lakatos, A., Becker-Baldus, J., Pos, K. M., and Glaubitz, C. (2013) Detecting Substrates Bound to the Secondary Multidrug Efflux Pump EmrE by DNP-Enhanced Solid-State NMR. *Journal of the American Chemical Society* **135**, 15754-15762
227. Bayro, M. J., Debelouchina, G. T., Eddy, M. T., Birkett, N. R., MacPhee, C. E., Rosay, M., Maas, W. E., Dobson, C. M., and Griffin, R. G. (2011) Intermolecular structure determination of amyloid fibrils with magic-angle spinning and dynamic nuclear polarization NMR. *J Am Chem Soc* **133**, 13967-13974

228. Uluca, B., Viennet, T., Petrović, D., Shaykhalishahi, H., Weirich, F., Gönülalan, A., Strodel, B., Etzkorn, M., Hoyer, W., and Heise, H. (2018) DNP-Enhanced MAS NMR: A Tool to Snapshot Conformational Ensembles of α-Synuclein in Different States. *Biophysical Journal* **114**, 1614-1623
229. Frederick, K. K., Michaelis, V. K., Caporini, M. A., Andreas, L. B., Debelouchina, G. T., Griffin, R. G., and Lindquist, S. (2017) Combining DNP NMR with segmental and specific labeling to study a yeast prion protein strain that is not parallel in-register. *Proc Natl Acad Sci U S A* **114**, 3642-3647
230. Lopez Del Amo, J. M., Schneider, D., Loquet, A., Lange, A., and Reif, B. (2013) Cryogenic solid state NMR studies of fibrils of the Alzheimer's disease amyloid-beta peptide: perspectives for DNP. *J Biomol NMR*
231. Debelouchina, G. T., Bayro, M. J., Fitzpatrick, A. W., Ladizhansky, V., Colvin, M. T., Caporini, M. A., Jaroniec, C. P., Bajaj, V. S., Rosay, M., Macphee, C. E., Vendruscolo, M., Maas, W. E., Dobson, C. M., and Griffin, R. G. (2013) Higher order amyloid fibril structure by MAS NMR and DNP spectroscopy. *J Am Chem Soc* **135**, 19237-19247
232. Potapov, A., Yau, W.-M., Ghirlando, R., Thurber, K. R., and Tycko, R. (2015) Successive Stages of Amyloid-β Self-Assembly Characterized by Solid-State Nuclear Magnetic Resonance with Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **137**, 8294-8307
233. Gupta, R., Zhang, H., Lu, M., Hou, G., Caporini, M., Rosay, M., Maas, W., Struppe, J., Ahn, J., Byeon, I.-J. L., Oschkinat, H., Jaudzems, K., Barbet-Massin, E., Emsley, L., Pintacuda, G., Lesage, A., Gronenborn, A. M., and Polenova, T. (2019) Dynamic Nuclear Polarization Magic-Angle Spinning Nuclear Magnetic Resonance Combined with Molecular Dynamics Simulations Permits Detection of Order and Disorder in Viral Assemblies. *The Journal of Physical Chemistry B* **123**, 5048-5058
234. Sergeyev, I. V., Itin, B., Rogawski, R., Day, L. A., and McDermott, A. E. (2017) Efficient assignment and NMR analysis of an intact virus using sequential side-chain correlations and DNP sensitization. *Proceedings of the National Academy of Sciences* **114**, 5171-5176
235. Gupta, R., Lu, M., Hou, G., Caporini, M. A., Rosay, M., Maas, W., Struppe, J., Suiter, C., Ahn, J., Byeon, I.-J. L., Franks, W. T., Orwick-Rydmark, M., Bertarello, A., Oschkinat, H., Lesage, A., Pintacuda, G., Gronenborn, A. M., and Polenova, T. (2016) Dynamic Nuclear Polarization Enhanced MAS NMR Spectroscopy for Structural Analysis of HIV-1 Protein Assemblies. *The Journal of Physical Chemistry B* **120**, 329-339
236. Petkova, A. T., Baldus, M., Belenky, M., Hong, M., Griffin, R. G., and Herzfeld, J. (2003) Backbone and side chain assignment strategies for multiply labeled membrane peptides and proteins in the solid state. *J. Magn. Reson.* **160**, 1-12
237. Reggie, L., Lopez, J. J., Collinson, I., Glaubitz, C., and Lorch, M. (2011) Dynamic Nuclear Polarization-Enhanced Solid-State NMR of a 13C-Labeled Signal Peptide Bound to Lipid-Reconstituted Sec Translocon. *Journal of the American Chemical Society* **133**, 19084-19086
238. Maciejko, J., Mehler, M., Kaur, J., Lieblein, T., Morgner, N., Ouari, O., Tordo, P., Becker-Baldus, J., and Glaubitz, C. (2015) Visualizing Specific Cross-Protomer Interactions in the Homo-Oligomeric Membrane Protein Proteorhodopsin by Dynamic-Nuclear-Polarization-Enhanced Solid-State NMR. *Journal of the American Chemical Society* **137**, 9032-9043
239. Salnikov, E., Rosay, M., Pawsey, S., Ouari, O., Tordo, P., and Bechinger, B. (2010) Solid-State NMR Spectroscopy of Oriented Membrane Polypeptides at 100 K with

Signal Enhancement by Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **132**, 5940-5941
240. Smith, A. N., Caporini, M. A., Fanucci, G. E., and Long, J. R. (2015) A Method for Dynamic Nuclear Polarization Enhancement of Membrane Proteins. *Angewandte Chemie International Edition* **54**, 1542-1546
241. Mao, J., Aladin, V., Jin, X., Leeder, A. J., Brown, L. J., Brown, R. C. D., He, X., Corzilius, B., and Glaubitz, C. (2019) Exploring Protein Structures by DNP-Enhanced Methyl Solid-State NMR Spectroscopy. *Journal of the American Chemical Society*
242. Renault, M., Pawsey, S., Bos, M. P., Koers, E. J., Nand, D., Tommassen-van Boxtel, R., Rosay, M., Tommassen, J., Maas, W. E., and Baldus, M. (2012) Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew Chem Int Ed Engl* **51**, 2998-3001
243. Costello, W. N., Xiao, Y., and Frederick, K. K. (2019) Chapter Ten - DNP-Assisted NMR Investigation of Proteins at Endogenous Levels in Cellular Milieu. In *Methods in Enzymology* (Wand, A. J., ed) Vol. 615 pp. 373-406, Academic Press
244. Yamamoto, K., Caporini, M. A., Im, S.-C., Waskell, L., and Ramamoorthy, A. (2017) Transmembrane Interactions of Full-length Mammalian Bitopic Cytochrome-P450-Cytochrome-b5 Complex in Lipid Bilayers Revealed by Sensitivity-Enhanced Dynamic Nuclear Polarization Solid-state NMR Spectroscopy. *Scientific Reports* **7**, 4116
245. Yamamoto, K., Caporini, M. A., Im, S.-C., Waskell, L., and Ramamoorthy, A. (2015) Cellular solid-state NMR investigation of a membrane protein using dynamic nuclear polarization. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1848**, 342-349
246. Takahashi, H., Ayala, I., Bardet, M., De Paëpe, G., Simorre, J.-P., and Hediger, S. (2013) Solid-State NMR on Bacterial Cells: Selective Cell Wall Signal Enhancement and Resolution Improvement using Dynamic Nuclear Polarization. *Journal of the American Chemical Society* **135**, 5105-5110
247. Wang, T., Park, Y. B., Caporini, M. A., Rosay, M., Zhong, L., Cosgrove, D. J., and Hong, M. (2013) Sensitivity-enhanced solid-state NMR detection of expansin's target in plant cell walls. *Proceedings of the National Academy of Sciences* **110**, 16444-16449
248. Wylie, B. J., Dzikovski, B. G., Pawsey, S., Caporini, M., Rosay, M., Freed, J. H., and McDermott, A. E. (2015) Dynamic nuclear polarization of membrane proteins: covalently bound spin-labels at protein–protein interfaces. *Journal of Biomolecular NMR* **61**, 361-367
249. McCarney, E. R., Armstrong, B. D., Kausik, R., and Han, S. (2008) Dynamic Nuclear Polarization Enhanced Nuclear Magnetic Resonance and Electron Spin Resonance Studies of Hydration and Local Water Dynamics in Micelle and Vesicle Assemblies. *Langmuir* **24**, 10062-10072
250. Salnikov, E. S., Sarrouj, H., Reiter, C., Aisenbrey, C., Purea, A., Aussenac, F., Ouari, O., Tordo, P., Fedotenko, I., Engelke, F., and Bechinger, B. (2015) Solid-State NMR/Dynamic Nuclear Polarization of Polypeptides in Planar Supported Lipid Bilayers. *The Journal of Physical Chemistry B* **119**, 14574-14583
251. Salnikov, E. S., Aisenbrey, C., Aussenac, F., Ouari, O., Sarrouj, H., Reiter, C., Tordo, P., Engelke, F., and Bechinger, B. (2016) Membrane topologies of the PGLa antimicrobial peptide and a transmembrane anchor sequence by Dynamic Nuclear Polarization/solid-state NMR spectroscopy. *Scientific Reports* **6**, 20895
252. Salnikov, E. S., Abel, S., Karthikeyan, G., Karoui, H., Aussenac, F., Tordo, P., Bechinger, B., and Ouari, O. (2017) Dynamic Nuclear Polarization/Solid-State NMR

Spectroscopy of Membrane Polypeptides: Free-Radical Optimization for Matrix-Free Lipid Bilayer Samples. *ChemPhysChem* **18**, 2103-2113
253. Salnikov, E. S., Aussenac, F., Abel, S., Purea, A., Tordo, P., Ouari, O., and Bechinger, B. (2019) Dynamic Nuclear Polarization / solid-state NMR of membranes. Thermal effects and sample geometry. *Solid State Nuclear Magnetic Resonance* **100**, 70-76
254. Rosay, M., Weis, V., Kreischer, K. E., Temkin, R. J., and Griffin, R. G. (2002) Two-dimensional (13)C-(13)C correlation spectroscopy with magic angle spinning and dynamic nuclear polarization. *J Am Chem Soc* **124**, 3214-3215
255. Hu, K. N., Yu, H. H., Swager, T. M., and Griffin, R. G. (2004) Dynamic nuclear polarization with biradicals. *J Am Chem Soc* **126**, 10844-10845
256. Chaudhari, S. R., Berruyer, P., Gajan, D., Reiter, C., Engelke, F., Silverio, D. L., Copéret, C., Lelli, M., Lesage, A., and Emsley, L. (2016) Dynamic nuclear polarization at 40 kHz magic angle spinning. *Physical Chemistry Chemical Physics* **18**, 10616-10622
257. Jaudzems, K., Bertarello, A., Chaudhari, S. R., Pica, A., Cala-De Paepe, D., Barbet-Massin, E., Pell, A. J., Akopjana, I., Kotelovica, S., Gajan, D., Ouari, O., Tars, K., Pintacuda, G., and Lesage, A. (2018) Dynamic Nuclear Polarization-Enhanced Biomolecular NMR Spectroscopy at High Magnetic Field with Fast Magic-Angle Spinning. *Angewandte Chemie International Edition* **57**, 7458-7462
258. Chaudhari, S. R., Wisser, D., Pinon, A. C., Berruyer, P., Gajan, D., Tordo, P., Ouari, O., Reiter, C., Engelke, F., Copéret, C., Lelli, M., Lesage, A., and Emsley, L. (2017) Dynamic Nuclear Polarization Efficiency Increased by Very Fast Magic Angle Spinning. *Journal of the American Chemical Society* **139**, 10609-10612
259. Akbey, m., Franks, W. T., Linden, A., Lange, S., Griffin, R. G., van Rossum, B.-J., and Oschkinat, H. (2010) Dynamic nuclear polarization of deuterated proteins. *Angewandte Chemie (International ed in English)* **49**, 7803-7806
260. Daskalov, A., Habenstein, B., Martinez, D., Debets, A. J., Sabate, R., Loquet, A., and Saupe, S. J. (2015) Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. *PLoS Biol* **13**, e1002059
261. Daskalov, A. (2016) On the evolutionary trajectories of signal-transducing amyloids in fungi and beyond. *Prion* **10**, 362-368
262. Daskalov, A., Habenstein, B., Sabate, R., Berbon, M., Martinez, D., Chaignepain, S., Coulary-Salin, B., Hofmann, K., Loquet, A., and Saupe, S. J. (2016) Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. *Proc Natl Acad Sci U S A* **113**, 2720-2725
263. Loquet, A., and Saupe, S. J. (2017) Diversity of Amyloid Motifs in NLR Signaling in Fungi. *Biomolecules* **7**
264. Siemer, A. B., Arnold, A. A., Ritter, C., Westfeld, T., Ernst, M., Riek, R., and Meier, B. H. (2006) Observation of highly flexible residues in amyloid fibrils of the HET-s prion. *J. Am. Chem. Soc.* **128**, 13224-13228
265. Wasmer, C., Soragni, A., Sabate, R., Lange, A., Riek, R., and Meier, B. H. (2008) Infectious and noninfectious amyloids of the HET-s(218-289) prion have different NMR spectra. *Angew. Chem. Int. Ed. Engl.* **47**, 5839-5841
266. Wasmer, C., Schutz, A., Loquet, A., Buhtz, C., Greenwald, J., Riek, R., Bockmann, A., and Meier, B. H. (2009) The molecular organization of the fungal prion HET-s in its amyloid form. *Journal of Molecular Biology* **394**, 119-127

267. Van Melckebeke, H., Schanda, P., Gath, J., Wasmer, C., Verel, R., Lange, A., Meier, B. H., and Bockmann, A. (2011) Probing water accessibility in HET-s(218-289) amyloid fibrils by solid-state NMR. *J Mol Biol* **405**, 765-772
268. Seuring, C., Greenwald, J., Wasmer, C., Wepf, R., Saupe, S. J., Meier, B. H., and Riek, R. (2012) The mechanism of toxicity in HET-S/HET-s prion incompatibility. *PLoS Biol* **10**, e1001451
269. Habenstein, B., Loquet, A., Giller, K., Becker, S., and Lange, A. (2013) Structural characterization of supramolecular assemblies by (13)C spin dilution and 3D solid-state NMR. *J Biomol NMR* **55**, 1-9
270. Loquet, A., Lv, G., Giller, K., Becker, S., and Lange, A. (2011) (13)C Spin Dilution for Simplified and Complete Solid-State NMR Resonance Assignment of Insoluble Biological Assemblies. *J. Am. Chem. Soc.* **133**, 4727-4725
271. Lange, A., Loquet, A. L., A., Lv, G., Giller, K., and Becker, S. (2011) (13)C Spin Dilution for Simplified and Complete Solid-State NMR Resonance Assignment of Insoluble Biological Assemblies. *Journal of the American Chemical Society* **133**, 4722-4725
272. Lange, S., Linden, A. H., Akbey, Ü., Trent Franks, W., Loening, N. M., Rossum, B.-J. v., and Oschkinat, H. (2012) The effect of biradical concentration on the performance of DNP-MAS-NMR. *Journal of Magnetic Resonance* **216**, 209-212
273. Smith, A. A., Ravotti, F., Testori, E., Cadalbert, R., Ernst, M., Böckmann, A., and Meier, B. H. (2017) Partially-deuterated samples of HET-s(218–289) fibrils: assignment and deuterium isotope effect. *Journal of Biomolecular NMR* **67**, 109-119
274. Admiraal, G., and Mellema, J. E. (1976) The structure of the contractile sheath of bacteriophage Mu. *J Ultrastruct Res* **56**, 48-64
275. Morgan, G. J., Hatfull, G. F., Casjens, S., and Hendrix, R. W. (2002) Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus11Edited by M. Gottesman. *Journal of Molecular Biology* **317**, 337-359
276. Büttner, C. R., Wu, Y., Maxwell, K. L., and Davidson, A. R. (2016) Baseplate assembly of phage Mu: Defining the conserved core components of contractile-tailed phages and related bacterial systems. *Proceedings of the National Academy of Sciences* **113**, 10174-10179
277. Sibille, N., Hanoulle, X., Bonachera, F., Verdegem, D., Landrieu, I., Wieruszeski, J.-M., and Lippens, G. (2009) Selective backbone labelling of ILV methyl labelled proteins. *Journal of Biomolecular NMR* **43**, 219-227
278. Kagan, J. C., Magupalli, V. G., and Wu, H. (2014) SMOCs: supramolecular organizing centres that control innate immunity. *Nature Reviews Immunology* **14**, 821-826
279. Wu, H. (2013) Higher-Order Assemblies in a New Paradigm of Signal Transduction. *Cell* **153**, 287-292
280. Cai, X., and Chen, Z. J. (2014) Prion-like polymerization as a signaling mechanism. *Trends in Immunology* **35**, 622-630
281. Cai, X., Chen, J., Xu, H., Liu, S., Jiang, Q.-X., Halfmann, R., and Chen, Zhijian J. (2014) Prion-like Polymerization Underlies Signal Transduction in Antiviral Immune Defense and Inflammasome Activation. *Cell* **156**, 1207-1222
282. Daskalov, A., Paoletti, M., Ness, F., and Saupe, S. J. (2012) Genomic clustering and homology between HET-S and the NWD2 STAND protein in various fungal genomes. *PLoS ONE* **7**, e34854
283. Riek, R., and Saupe, S. J. (2016) The HET-S/s Prion Motif in the Control of Programmed Cell Death. *Cold Spring Harbor Perspectives in Biology* **8**

284. Wickner, R. B. (2016) Yeast and Fungal Prions. *Cold Spring Harb Perspect Biol* **8**
285. Aguzzi, A., and Falsig, J. (2012) Prion propagation, toxicity and degradation. *Nat Neurosci* **15**, 936-939
286. Greenwald, J., Buhtz, C., Ritter, C., Kwiatkowski, W., Choe, S., Maddelein, M. L., Ness, F., Cescau, S., Soragni, A., Leitz, D., Saupe, S. J., and Riek, R. (2010) The mechanism of prion inhibition by HET-S. *Mol Cell* **38**, 889-899
287. Wickner, R. B., Edskes, H. K., Shewmaker, F., and Nakayashiki, T. (2007) Prions of fungi: inherited structures and biological roles. *Nat Rev Microbiol* **5**, 611-618
288. Wickner, R. B., Taylor, K. L., Edskes, H. K., Maddelein, M. L., Moriyama, H., and Roberts, B. T. (2000) Prions of yeast as heritable amyloidoses. *J Struct Biol* **130**, 310-322
289. Graziani, S., Silar, P., and Daboussi, M. J. (2004) Bistability and hysteresis of the 'Secteur' differentiation are controlled by a two-gene locus in Nectria haematococca. *BMC Biol* **2**, 18
290. Wang, Y., and Jardetzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **11**, 852-861
291. Luca, S., Filippov, D. V., van Boom, J. H., Oschkinat, H., de Groot, H. J., and Baldus, M. (2001) Secondary chemical shifts in immobilized peptides and proteins: a qualitative basis for structure refinement under magic angle spinning. *J. Biomol. NMR* **20**, 325-331
292. Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., and Oschkinat, H. (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98-102
293. Higman, V. A., Flinders, J., Hiller, M., Jehle, S., Markovic, S., Fiedler, S., van Rossum, B. J., and Oschkinat, H. (2009) Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins. *J Biomol NMR* **44**, 245-260
294. Wasmer, C., Schütz, A., Loquet, A., Buhtz, C., Greenwald, J., Riek, R., Böckmann, A., and Meier, B. H. (2009) The Molecular Organization of the Fungal Prion HET-s in Its Amyloid Form. *Journal of Molecular Biology* **394**, 119-127
295. Meier, B. H., Loquet, A., Bousset, L., Gardiennet, C., Sourigues, Y., Wasmer, C., Habenstein, B., Schutz, A., Melki, R., and Bockmann, A. (2009) Prion Fibrils of Ure2p Assembled under Physiological Conditions Contain Highly Ordered, Natively Folded Modules. *Journal of Molecular Biology* **394**, 108-118
296. Loquet, A., Saupe, S. J., and Romero, D. (2018) Functional Amyloids in Health and Disease. *J Mol Biol* **430**, 3629-3630
297. Van Gerven, N., Van der Verren, S. E., Reiter, D. M., and Remaut, H. (2018) The Role of Functional Amyloids in Bacterial Virulence. *J Mol Biol* **430**, 3657-3684
298. Deshmukh, M., Evans, M. L., and Chapman, M. R. (2018) Amyloid by Design: Intrinsic Regulation of Microbial Amyloid Assembly. *J Mol Biol* **430**, 3631-3641
299. Rouse, S. L., Matthews, S. J., and Dueholm, M. S. (2018) Ecology and Biogenesis of Functional Amyloids in Pseudomonas. *J Mol Biol* **430**, 3685-3695
300. Pham, C. L., Kwan, A. H., and Sunde, M. (2014) Functional amyloid: widespread in Nature, diverse in purpose. *Essays Biochem* **56**, 207-219
301. Dragos, A., Kovacs, A. T., and Claessen, D. (2017) The Role of Functional Amyloids in Multicellular Growth and Development of Gram-Positive Bacteria. *Biomolecules* **7**, 60
302. Zhou, Y., Blanco, L. P., Smith, D. R., and Chapman, M. R. (2012) Bacterial amyloids. *Methods Mol Biol* **849**, 303-320

303. Romero, D., and Kolter, R. (2014) Functional amyloids in bacteria. *Int Microbiol* **17**, 65-73
304. Friedland, R. P., and Chapman, M. R. (2017) The role of microbial amyloid in neurodegeneration. *PLoS Pathog* **13**, e1006654
305. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D., and Dobson, C. M. (1998) Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* **95**, 4224-4228
306. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C. M. (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci U S A* **96**, 3590-3594
307. Ramirez-Alvarado, M., Merkel, J. S., and Regan, L. (2000) A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc Natl Acad Sci U S A* **97**, 8979-8984
308. Dragos, A., and Kovacs, A. T. (2017) The Peculiar Functions of the Bacterial Extracellular Matrix. *Trends Microbiol* **25**, 257-266
309. Blanco, L. P., Evans, M. L., Smith, D. R., Badtke, M. P., and Chapman, M. R. (2012) Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol* **20**, 66-73
310. Taylor, J. D., and Matthews, S. J. (2015) New insight into the molecular control of bacterial functional amyloids. *Front Cell Infect Microbiol* **5**, 33
311. Van Gerven, N., Klein, R. D., Hultgren, S. J., and Remaut, H. (2015) Bacterial amyloid formation: structural insights into curli biogensis. *Trends Microbiol* **23**, 693-706
312. Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001) Fruiting body formation by Bacillus subtilis. *Proc Natl Acad Sci U S A* **98**, 11621-11626
313. Branda, S. S., Chu, F., Kearns, D. B., Losick, R., and Kolter, R. (2006) A major protein component of the Bacillus subtilis biofilm matrix. *Mol Microbiol* **59**, 1229-1238
314. Romero, D., Aguilar, C., Losick, R., and Kolter, R. (2010) Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. *Proc Natl Acad Sci U S A* **107**, 2230-2234
315. Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2011) An accessory protein required for anchoring and assembly of amyloid fibres in B. subtilis biofilms. *Mol Microbiol* **80**, 1155-1168
316. Caro-Astorga, J., Perez-Garcia, A., de Vicente, A., and Romero, D. (2014) A genomic region involved in the formation of adhesin fibers in Bacillus cereus biofilms. *Front Microbiol* **5**, 745
317. Diehl, A., Roske, Y., Ball, L., Chowdhury, A., Hiller, M., Moliere, N., Kramer, R., Stoppler, D., Worth, C. L., Schlegel, B., Leidert, M., Cremer, N., Erdmann, N., Lopez, D., Stephanowitz, H., Krause, E., van Rossum, B. J., Schmieder, P., Heinemann, U., Turgay, K., Akbey, U., and Oschkinat, H. (2018) Structural changes of TasA in biofilm formation of Bacillus subtilis. *Proc Natl Acad Sci U S A* **115**, 3237-3242
318. Erskine, E., Morris, R. J., Schor, M., Earl, C., Gillespie, R. M. C., Bromley, K. M., Sukhodub, T., Clark, L., Fyfe, P. K., Serpell, L. C., Stanley-Wall, N. R., and MacPhee, C. E. (2018) Formation of functional, non-amyloidogenic fibres by recombinant Bacillus subtilis TasA. *Molecular Microbiology* **110**, 897-913
319. Ahimou, F., Jacques, P., and Deleu, M. (2000) Surfactin and iturin A effects on Bacillus subtilis surface hydrophobicity. *Enzyme and Microbial Technology* **27**, 749-754

320. Hyberts, S. G., Arthanari, H., and Wagner, G. (2012) Applications of Non-Uniform Sampling and Processing. In *Novel Sampling Approaches in Higher Dimensional NMR* (Billeter, M., and Orekhov, V., eds) pp. 125-148, Springer Berlin Heidelberg, Berlin, Heidelberg
321. Wang, Y., and Jardetzky, O. (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11**, 852–861
322. Yao, J., Dyson, H. J., and Wright, P. E. (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Letters* **419**, 285-289
323. Spera, S., and Bax, A. (1991) Empirical Correlation between Protein Backbone Conformation and Ca and Cb ^{13}C Nuclaer Magnetic Resonance Chemical Shifts. *J. Am. Chem. Soc.* **113**, 5490-5492
324. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science* **15**, 2795-2804
325. Abbasi, R., Mousa, R., Dekel, N., Amartely, H., Danieli, T., Lebendiker, M., Levi-Kalisman, Y., Shalev, D. E., Metanis, N., and Chai, L. (2019) The Bacterial Extracellular Matrix Protein TapA Is a Two-Domain Partially Disordered Protein. *ChemBioChem* **20**, 355-359
326. Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273**, 729-739
327. Chai, L., Romero, D., Kayatekin, C., Akabayov, B., Vlamakis, H., Losick, R., and Kolter, R. (2013) Isolation, Characterization, and Aggregation of a Structured Bacterial Matrix Precursor. *Journal of Biological Chemistry* **288**, 17559-17568
328. Meier, B. H., Riek, R., and Bockmann, A. (2017) Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR. *Trends Biochem Sci* **42**, 777-787
329. Loquet, A., El Mammeri, N., Stanek, J., Berbon, M., Bardiaux, B., Pintacuda, G., and Habenstein, B. (2018) 3D Structure Determination of Amyloid Fibrils using Solid-State NMR Spectroscopy. *Methods* **138-139**, 26-38
330. Linden, A. H., Franks, W. T., Akbey, U., Lange, S., van Rossum, B. J., and Oschkinat, H. (2011) Cryogenic temperature effects and resolution upon slow cooling of protein preparations in solid state NMR. *J Biomol NMR* **51**, 283-292
331. Bauer, T., Dotta, C., Balacescu, L., Gath, J., Hunkeler, A., Bockmann, A., and Meier, B. H. (2017) Line-Broadening in Low-Temperature Solid-State NMR Spectra of Fibrils. *J Biomol NMR* **67**, 51-61
332. Daebel, V., Chinnathambi, S., Biernat, J., Schwalbe, M., Habenstein, B., Loquet, A., Akoury, E., Tepper, K., Muller, H., Baldus, M., Griesinger, C., Zweckstetter, M., Mandelkow, E., Vijayan, V., and Lange, A. (2012) beta-Sheet Core of Tau Paired Helical Filaments Revealed by Solid-State NMR. *J. Am. Chem. Soc.* **134**, 13982-13989
333. Heise, H., Hoyer, W., Becker, S., Andronesi, O. C., Riedel, D., and Baldus, M. (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length *Proc. Natl. Acad. Sci. USA* **102**, 15871–15876
334. Eisenberg, D. S., and Sawaya, M. R. (2017) Structural Studies of Amyloid Proteins at the Molecular Level. *Annu Rev Biochem* **86**, 69-95
335. Loquet, A., Bousset, L., Gardiennet, C., Sourigues, Y., Wasmer, C., Habenstein, B., Schutz, A., Meier, B. H., Melki, R., and Bockmann, A. (2009) Prion fibrils of Ure2p assembled under physiological conditions contain highly ordered, natively folded modules. *J. Mol. Biol.* **394**, 108-118

336. Shen, Y., and Bax, A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR* **38**, 289-302
337. Zandomeneghi, G., Krebs, M. R., McCammon, M. G., and Fandrich, M. (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Sci* **13**, 3314-3321
338. Goormaghtigh, E., Cabiaux, V., and Ruysschaert, J. M. (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. III. Secondary structures. *Subcell Biochem* **23**, 405-450
339. Yoshimura, Y., Lin, Y., Yagi, H., Lee, Y. H., Kitayama, H., Sakurai, K., So, M., Ogi, H., Naiki, H., and Goto, Y. (2012) Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proc Natl Acad Sci U S A* **109**, 14446-14451
340. Romero, D., Vlamakis, H., Losick, R., and Kolter, R. (2014) Functional Analysis of the Accessory Protein TapA in Bacillus subtilis Amyloid Fiber Assembly. *Journal of Bacteriology* **196**, 1505-1513
341. Hammer, N. D., Schmidt, J. C., and Chapman, M. R. (2007) The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proceedings of the National Academy of Sciences* **104**, 12494-12499
342. Fandrich, M. (2007) On the structural definition of amyloid fibrils and other polypeptide aggregates. *Cell Mol Life Sci* **64**, 2066-2078
343. Candela, T., Fagerlund, A., Buisson, C., Gilois, N., Kolstø, A.-B., Økstad, O. A., Aymerich, S., Nielsen-Leroux, C., Lereclus, D., and Gohar, M. (2018) CalY is a major virulence factor and a biofilm matrix protein. *Molecular Microbiology*
344. van Gestel, J., Vlamakis, H., and Kolter, R. (2015) From Cell Differentiation to Cell Collectives: Bacillus subtilis Uses Division of Labor to Migrate. *PLOS Biology* **13**, e1002141
345. Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. (2013) Sticking together: building a biofilm the Bacillus subtilis way. *Nature Reviews Microbiology* **11**, 157-168
346. Shank, E. A., Klepac-Ceraj, V., Collado-Torres, L., Powers, G. E., Losick, R., and Kolter, R. (2011) Interspecies interactions that result in Bacillus subtilis forming biofilms are mediated mainly by members of its own genus. *Proceedings of the National Academy of Sciences* **108**, E1236-E1243
347. O'Toole, G., Kaplan, H. B., and Kolter, R. (2000) Biofilm formation as microbial development. *Annu Rev Microbiol* **54**, 49-79
348. Flemming, H.-C., and Wingender, J. (2010) The biofilm matrix. *Nature Reviews Microbiology* **8**, 623-633
349. Sutherland, I. W. (2001) The biofilm matrix – an immobilized but dynamic microbial environment. *Trends in Microbiology* **9**, 222-227
350. Lopez, D., Vlamakis, H., and Kolter, R. (2008) Generation of multiple cell types in Bacillus subtilis. *FEMS Microbiology Reviews* **33**, 152-163
351. Stöver, A. G., and Driks, A. (1999) Secretion, Localization, and Antibacterial Activity of TasA, a Bacillus subtilis Spore-Associated Protein. *Journal of Bacteriology* **181**, 1664-1672
352. Chu, F., Kearns, D. B., McLoon, A., Chai, Y., Kolter, R., and Losick, R. (2008) A novel regulatory protein governing biofilm formation in Bacillus subtilis. *Molecular Microbiology* **68**, 1117-1127

353. Chu, F., Kearns, D. B., Branda, S. S., Kolter, R., and Losick, R. (2006) Targets of the master regulator of biofilm formation in Bacillus subtilis. *Mol Microbiol* **59**, 1216-1228
354. El Mammeri, N., Hierrezuelo, J., Tolchard, J., Cámara-Almirón, J., Caro-Astorga, J., Álvarez-Mena, A., Dutour, A., Berbon, M., Shenoy, J., Morvan, E., Grélard, A., Kauffmann, B., Lecomte, S., Vicente, A. d., Habenstein, B., Romero, D., and Loquet, A. (2019) Molecular architecture of bacterial amyloids in Bacillus biofilms. *The FASEB Journal* **0**, fj.201900831R
355. Erskine, E., Morris, R. J., Schor, M., Earl, C., Gillespie, R. M. C., Bromley, K., Sukhodub, T., Clark, L., Fyfe, P. K., Serpell, L. C., Stanley-Wall, N. R., and MacPhee, C. E. (2018) Formation of functional, non-amyloidogenic fibres by recombinant Bacillus subtilis TasA. *Mol Microbiol*
356. Kovács, Á. T., van Gestel, J., and Kuipers, O. P. (2012) The protective layer of biofilm: a repellent function for a new class of amphiphilic proteins. *Molecular Microbiology* **85**, 8-11
357. Lasa, I., and Penadés, J. R. (2006) Bap: A family of surface proteins involved in biofilm formation. *Research in Microbiology* **157**, 99-107
358. Kobayashi, K., and Iwano, M. (2012) BslA(YuaB) forms a hydrophobic layer on the surface of Bacillus subtilis biofilms. *Mol Microbiol* **85**, 51-66
359. Romero, D. (2013) Bacterial determinants of the social behavior of Bacillus subtilis. *Research in Microbiology* **164**, 788-798
360. Marvasi, M., Visscher, P. T., and Casillas Martinez, L. (2010) Exopolymeric substances (EPS) from Bacillus subtilis : polymers and genes encoding their synthesis. *FEMS Microbiology Letters* **313**, 1-9
361. Epstein, A. K., Pokroy, B., Seminara, A., and Aizenberg, J. (2011) Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration. *Proceedings of the National Academy of Sciences* **108**, 995-1000
362. Kang, X., Kirui, A., Muszyński, A., Widanage, M. C. D., Chen, A., Azadi, P., Wang, P., Mentink-Vigier, F., and Wang, T. (2018) Molecular architecture of fungal cell walls revealed by solid-state NMR. *Nature Communications* **9**, 2747
363. Thakur, C. S., Sama, J. N., Jackson, M. E., Chen, B., and Dayie, T. K. (2010) Selective 13C labeling of nucleotides for large RNA NMR spectroscopy using an E. coli strain disabled in the TCA cycle. *Journal of Biomolecular NMR* **48**, 179-192
364. Häuslein, I., Manske, C., Goebel, W., Eisenreich, W., and Hilbi, H. (2016) Pathway analysis using 13C-glycerol and other carbon tracers reveals a bipartite metabolism of Legionella pneumophila. *Molecular Microbiology* **100**, 229-246
365. Matlahov, I., and van der Wel, P. C. A. (2018) Hidden motions and motion-induced invisibility: Dynamics-based spectral editing in solid-state NMR. *Methods*
366. Cámara-Almirón, J., Navarro, Y., Díaz-Martínez, L., Magno-Pérez-Bryan, M. C., Molina-Santiago, C., Pearson, J. R., de Vicente, A., Pérez-García, A., and Romero, D. (2020) Dual functionality of the amyloid protein TasA in Bacillus physiology and fitness on the phylloplane. *Nature Communications* **11**, 1859
367. Kesel, S., von Bronk, B., Falcón García, C., Götz, A., Lieleg, O., and Opitz, M. (2017) Matrix composition determines the dimensions of Bacillus subtilis NCIB 3610 biofilm colonies grown on LB agar. *RSC Advances* **7**, 31886-31898
368. Buda, S., Nawój, M., and Mlynarski, J. (2016) Chapter Four - Recent Advances in NMR Studies of Carbohydrates. In *Annual Reports on NMR Spectroscopy* (Webb, G. A., ed) Vol. 89 pp. 185-223, Academic Press
369. Espinosa, J. F., Asensio, J. L., Garcia, J. L., Laynez, J., Bruix, M., Wright, C., Siebert, H. C., Gabius, H. J., Canada, F. J., and Jimenez-Barbero, J. (2000) NMR

investigations of protein-carbohydrate interactions binding studies and refined three-dimensional solution structure of the complex between the B domain of wheat germ agglutinin and N,N', N"-triacetylchitotriose. *Eur J Biochem* **267**, 3965-3978
370. Duus, J. Ø., Gotfredsen, C. H., and Bock, K. (2000) Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations. *Chemical Reviews* **100**, 4589-4614
371. Bubb, W. A. (2003) NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. *Concepts in Magnetic Resonance Part A* **19A**, 1-19
372. Kern, T., Hediger, S., Muller, P., Giustini, C., Joris, B., Bougault, C., Vollmer, W., and Simorre, J. P. (2008) Toward the characterization of peptidoglycan structure and protein-peptidoglycan interactions by solid-state NMR spectroscopy. *J Am Chem Soc* **130**, 5618-5619
373. Hayhurst, E. J., Kailas, L., Hobbs, J. K., and Foster, S. J. (2008) Cell wall peptidoglycan architecture in Bacillus subtilis. *Proceedings of the National Academy of Sciences* **105**, 14603-14608
374. Sharon, N. (2007) Celebrating the golden anniversary of the discovery of bacillosamine, the diamino sugar of a Bacillus*,†. *Glycobiology* **17**, 1150-1155
375. Molina-Santiago, C., Pearson, J. R., Navarro, Y., Berlanga-Clavero, M. V., Caraballo-Rodriguez, A. M., Petras, D., García-Martín, M. L., Lamon, G., Haberstein, B., Cazorla, F. M., de Vicente, A., Loquet, A., Dorrestein, P. C., and Romero, D. (2019) The extracellular matrix protects Bacillus subtilis colonies from Pseudomonas invasion and modulates plant co-colonization. *Nature Communications* **10**, 1919
376. Paton, A. M. (1960) THE ROLE OF PSEUDOMONAS IN PLANT DISEASE. *Journal of Applied Bacteriology* **23**, 526-532
377. Morita, N., Takagi, M., and Murao, S. (1967) A New Gel-Forming Polysaccharide Produced by Bacillus subtilis FT-3 Its Structure and its Physical and Chemical Characteristics. *Agricultural and biological chemistry* **31**, 27-32
378. Ashiuchi, M., Soda, K., and Misono, H. (1999) A Poly-γ-glutamate Synthetic System of Bacillus subtilis IFO 3336: Gene Cloning and Biochemical Analysis of Poly-γ-glutamate Produced by Escherichia coli Clone Cells. *Biochemical and Biophysical Research Communications* **263**, 6-12
379. Ashiuchi, M., Nawa, C., Kamei, T., Song, J.-J., Hong, S.-P., Sung, M.-H., Soda, K., Yagi, T., and Misono, H. (2001) Physiological and biochemical characteristics of poly γ-glutamate synthetase complex of Bacillus subtilis. *European Journal of Biochemistry* **268**, 5321-5328
380. Morikawa, M., Kagihiro, S., Haruki, M., Takano, K., Branda, S., Kolter, R., and Kanaya, S. (2006) Biofilm formation by a Bacillus subtilis strain that produces γ-polyglutamate. *Microbiology* **152**, 2801-2807
381. Candela, T., and Fouet, A. (2006) Poly-gamma-glutamate in bacteria. *Molecular Microbiology* **60**, 1091-1098
382. Breazeale, S. D., Ribeiro, A. A., and Raetz, C. R. H. (2002) Oxidative decarboxylation of UDP-glucuronic acid in extracts of polymyxin-resistant Escherichia coli. Origin of lipid a species modified with 4-amino-4-deoxy-L-arabinose. *J Biol Chem* **277**, 2886-2896
383. Vilain, S., Pretorius, J. M., Theron, J., and Brözel, V. S. (2009) DNA as an Adhesin: Bacillus cereus Requires Extracellular DNA To Form Biofilms. *Applied and Environmental Microbiology* **75**, 2861-2868
384. Spoering, A. L., and Gilmore, M. S. (2006) Quorum sensing and DNA release in bacterial biofilms. *Current Opinion in Microbiology* **9**, 133-137

385. Huang, X., Zhang, K., Deng, M., Exterkate, R. A. M., Liu, C., Zhou, X., Cheng, L., and ten Cate, J. M. (2017) Effect of arginine on the growth and biofilm formation of oral bacteria. *Archives of Oral Biology* **82**, 256-262
386. He, J., Hwang, G., Liu, Y., Gao, L., Kilpatrick-Liverman, L., Santarpia, P., Zhou, X., and Koo, H. (2016) L-Arginine Modifies the Exopolysaccharide Matrix and Thwarts Streptococcus mutans Outgrowth within Mixed-Species Oral Biofilms. *Journal of Bacteriology* **198**, 2651-2661
387. Jakubovics, N. S., Robinson, J. C., Samarian, D. S., Kolderman, E., Yassin, S. A., Bettampadi, D., Bashton, M., and Rickard, A. H. (2015) Critical roles of arginine in growth and biofilm development by Streptococcus gordonii. *Molecular Microbiology* **97**, 281-300
388. Kolderman, E., Bettampadi, D., Samarian, D., Dowd, S. E., Foxman, B., Jakubovics, N. S., and Rickard, A. H. (2015) L-Arginine Destabilizes Oral Multi-Species Biofilm Communities Developed in Human Saliva. *PLOS ONE* **10**, e0121835
389. Burrell, M., Hanfrey, C. C., Murray, E. J., Stanley-Wall, N. R., and Michael, A. J. (2010) Evolution and Multiplicity of Arginine Decarboxylases in Polyamine Biosynthesis and Essential Role in Bacillus subtilis Biofilm Formation. *Journal of Biological Chemistry* **285**, 39224-39238
390. Reverdy, A., Chen, Y., Hunter, E., Gozzi, K., and Chai, Y. (2018) Protein lysine acetylation plays a regulatory role in Bacillus subtilis multicellularity. *PLOS ONE* **13**, e0204687
391. Chai, Y., Beauregard, P. B., Vlamakis, H., Losick, R., and Kolter, R. (2012) Galactose Metabolism Plays a Crucial Role in Biofilm Formation by Bacillus subtilis. *mBio* **3**, e00184-00112
392. Branda, S. S., González-Pastor, J. E., Dervyn, E., Ehrlich, S. D., Losick, R., and Kolter, R. (2004) Genes Involved in Formation of Structured Multicellular Communities by Bacillus subtilis. *Journal of Bacteriology* **186**, 3970-3979
393. Lazarevic, V., Soldo, B., Médico, N., Pooley, H., Bron, S., and Karamata, D. (2005) Bacillus subtilis α-Phosphoglucomutase Is Required for Normal Cell Morphology and Biofilm Formation. *Applied and Environmental Microbiology* **71**, 39-45
394. Valle, J., Da Re, S., Schmid, S., Skurnik, D., D'Ari, R., and Ghigo, J.-M. (2008) The Amino Acid Valine Is Secreted in Continuous-Flow Bacterial Biofilms. *Journal of Bacteriology* **190**, 264-274
395. Pisithkul, T., Schroeder, J. W., Trujillo, E. A., Yeesin, P., Stevenson, D. M., Chaiamarit, T., Coon, J. J., Wang, J. D., and Amador-Noguez, D. (2019) Metabolic Remodeling during Biofilm Development of Bacillus subtilis. *mBio* **10**, e00623-00619
396. Rendueles, O., and Ghigo, J.-M. (2012) Multi-species biofilms: how to avoid unfriendly neighbors. *FEMS Microbiology Reviews* **36**, 972-989
397. Kolodkin-Gal, I., Romero, D., Cao, S., Clardy, J., Kolter, R., and Losick, R. (2010) D-Amino Acids Trigger Biofilm Disassembly. *Science* **328**, 627-629
398. Wilking, J. N., Angelini, T. E., Seminara, A., Brenner, M. P., and Weitz, D. A. (2011) Biofilms as complex fluids. *MRS Bulletin* **36**, 385-391
399. Ido, N., Lybman, A., Hayet, S., Azulay, D. N., Ghrayeb, M., Liddawieh, S., and Chai, L. (2020) Bacillus subtilis biofilms characterized as hydrogels. Insights on water uptake and water binding in biofilms. *Soft Matter* **16**, 6180-6190

400. Huang, J., Liu, S., Zhang, C., Wang, X., Pu, J., Ba, F., Xue, S., Ye, H., Zhao, T., Li, K., Wang, Y., Zhang, J., Wang, L., Fan, C., Lu, T. K., and Zhong, C. (2019) Programmable and printable Bacillus subtilis biofilms as engineered living materials. *Nature Chemical Biology* **15**, 34-41
401. Notti, R. Q., and Stebbins, C. E. (2016) The Structure and Function of Type III Secretion Systems. *Microbiology Spectrum* **4**
402. Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J. E., and Aizawa, S. I. (1998) Supramolecular structure of the Salmonella typhimurium type III protein secretion system. *Science* **280**, 602-605
403. Hu, J., Worrall, L. J., Hong, C., Vuckovic, M., Atkinson, C. E., Caveney, N., Yu, Z., and Strynadka, N. C. J. (2018) Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nat Commun* **9**, 3840
404. Schraidt, O., and Marlovits, T. C. (2011) Three-dimensional model of Salmonella's needle complex at subnanometer resolution. *Science* **331**, 1192-1195
405. Galkin, V. E., Schmied, W. H., Schraidt, O., Marlovits, T. C., and Egelman, E. H. (2010) The structure of the Salmonella typhimurium type III secretion system needle shows divergence from the flagellar system. *J. Mol. Biol.* **396**, 1392-1397
406. Fujii, T., Cheung, M., Blanco, A., Kato, T., Blocker, A. J., and Namba, K. (2012) Structure of a type III secretion needle at 7-A resolution provides insights into its assembly and signaling mechanisms. *Proc Natl Acad Sci U S A* **109**, 4461-4466
407. Demers, J. P., Sgourakis, N. G., Gupta, R., Loquet, A., Giller, K., Riedel, D., Laube, B., Kolbe, M., Baker, D., Becker, S., and Lange, A. (2013) The common structural architecture of shigella flexneri and salmonella typhimurium type three secretion system needles. *PLoS Pathog* **accepted**
408. Verasdonck, J., Shen, D. K., Treadgold, A., Arthur, C., Bockmann, A., Meier, B. H., and Blocker, A. J. (2015) Reassessment of MxiH subunit orientation and fold within native Shigella T3SS needles using surface labelling and solid-state NMR. *J Struct Biol* **192**, 441-448
409. Demers, J. P., Habenstein, B., Loquet, A., Kumar Vasa, S., Giller, K., Becker, S., Baker, D., Lange, A., and Sgourakis, N. G. (2014) High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat Commun* **5**, 4976
410. Demers, J. P., Sgourakis, N. G., Gupta, R., Loquet, A., Giller, K., Riedel, D., Laube, B., Kolbe, M., Baker, D., Becker, S., and Lange, A. (2013) The common structural architecture of Shigella flexneri and Salmonella typhimurium type three secretion needles. *PLoS Pathog* **9**, e1003245
411. Lunelli, M., Kamprad, A., Bürger, J., Mielke, T., Spahn, C. M. T., and Kolbe, M. (2020) Cryo-EM structure of the Shigella type III needle complex. *PLOS Pathogens* **16**, e1008263
412. Kuhlen, L., Abrusci, P., Johnson, S., Gault, J., Deme, J., Caesar, J., Dietsche, T., Mebrhatu, M. T., Ganief, T., Macek, B., Wagner, S., Robinson, C. V., and Lea, S. M. (2018) Structure of the core of the type III secretion system export apparatus. *Nature Structural & Molecular Biology* **25**, 583-590
413. Klein, J. R., Fahlen, T. F., and Jones, B. D. (2000) Transcriptional organization and function of invasion genes within Salmonella enterica serovar Typhimurium pathogenicity island 1, including the prgH, prgI, prgJ, prgK, orgA, orgB, and orgC genes. *Infect Immun* **68**, 3368-3376
414. Miao, E. A., Mao, D. P., Yudkovsky, N., Bonneau, R., Lorang, C. G., Warren, S. E., Leaf, I. A., and Aderem, A. (2010) Innate immune detection of the type III secretion

apparatus through the NLRC4 inflammasome. *Proc Natl Acad Sci U S A* **107**, 3076-3080
415. Marlovits, T. C., Kubori, T., Lara-Tejero, M., Thomas, D., Unger, V. M., and Galan, J. E. (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441**, 637-640
416. Marlovits, T. C., Kubori, T., Sukhan, A., Thomas, D. R., Galan, J. E., and Unger, V. M. (2004) Structural insights into the assembly of the type III secretion needle complex. *Science (New York, N Y)* **306**, 1040-1042
417. Hu, J., Worrall, L. J., Vuckovic, M., Hong, C., Deng, W., Atkinson, C. E., Brett Finlay, B., Yu, Z., and Strynadka, N. C. J. (2019) T3S injectisome needle complex structures in four distinct states reveal the basis of membrane coupling and assembly. *Nature Microbiology* **4**, 2010-2019
418. Torres-Vargas, C. E., Kronenberger, T., Roos, N., Dietsche, T., Poso, A., and Wagner, S. (2019) The inner rod of virulence-associated type III secretion systems constitutes a needle adapter of one helical turn that is deeply integrated into the system's export apparatus. *Molecular Microbiology* **112**, 918-931
419. Zhong, D., Lefebre, M., Kaur, K., McDowell, M. A., Gdowski, C., Jo, S., Wang, Y., Benedict, S. H., Lea, S. M., Galan, J. E., and De Guzman, R. N. (2012) The Salmonella Type III Secretion System Inner Rod Protein PrgJ Is Partially Folded. *J Biol Chem* **287**, 25303-25311
420. Yamaguchi, S., Fujita, H., Ishihara, A., Aizawa, S., and Macnab, R. M. (1986) Subdivision of flagellar genes of Salmonella typhimurium into regions responsible for assembly, rotation, and switching. *Journal of Bacteriology* **166**, 187-193
421. Hoiseth, S. K., and Stocker, B. A. D. (1981) Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature* **291**, 238-239
422. Kaniga, K., Bossio, J. C., and Galán, J. E. (1994) The Salmonella typhimurium invasion genes invF and invG encode homologues of the AraC and PulD family of proteins. *Molecular Microbiology* **13**, 555-568
423. Butan, C., Lara-Tejero, M., Li, W., Liu, J., and Galán, J. E. (2019) High-resolution view of the type III secretion export apparatus in situ reveals membrane remodeling and a secretion pathway. *Proceedings of the National Academy of Sciences* **116**, 24786-24795
424. Monlezun, L., Liebl, D., Fenel, D., Grandjean, T., Berry, A., Schoehn, G., Dessein, R., Faudry, E., and Attree, I. (2015) PscI is a type III secretion needle anchoring protein with in vitro polymerization capacities. *Molecular Microbiology* **96**, 419-436
425. Dietsche, T., Tesfazgi Mebrhatu, M., Brunner, M. J., Abrusci, P., Yan, J., Franz-Wachtel, M., Schärfe, C., Zilkenat, S., Grin, I., Galán, J. E., Kohlbacher, O., Lea, S., Macek, B., Marlovits, T. C., Robinson, C. V., and Wagner, S. (2016) Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus. *PLOS Pathogens* **12**, e1006071
426. Zilkenat, S., Franz-Wachtel, M., Stierhof, Y.-D., Galán, J. E., Macek, B., and Wagner, S. (2016) Determination of the Stoichiometry of the Complete Bacterial Type III Secretion Needle Complex Using a Combined Quantitative Proteomic Approach. *Molecular & Cellular Proteomics* **15**, 1598-1609
427. Sal-Man, N., Deng, W., and Finlay, B. B. (2012) EscI: a crucial component of the type III secretion system forms the inner rod structure in enteropathogenic Escherichia coli. *The Biochemical journal* **442**, 119-125
428. Wood, S. E., Jin, J., and Lloyd, S. A. (2008) YscP and YscU switch the substrate specificity of the Yersinia type III secretion system by regulating export of the inner rod protein YscI. *J Bacteriol* **190**, 4252-4262

429. Lefebre, M. D., and Galan, J. E. (2014) The inner rod protein controls substrate switching and needle length in a Salmonella type III secretion system. *Proc Natl Acad Sci U S A* **111**, 817-822
430. Preer, J. R., Jr., Preer, L. B., and Jurand, A. (1974) Kappa and other endosymbionts in Paramecium aurelia. *Bacteriol Rev* **38**, 113-163
431. Sonneborn, T. M. (1959) Kappa and Related Particles in Paramecium. In *Advances in Virus Research* (Smith, K. M., and Lauffer, M. A., eds) Vol. 6 pp. 229-356, Academic Press
432. Sonneborn, T. M. (1959) Kappa and Related Particles in Paramecium11Contribution 657 from the Zoological Laboratory, Indiana University. Written with the aid of support from the National Science Foundation during sabbatical leave at the Institute for Advanced Study, Princeton. Much of the work of the author and his co-workers which is included in this review was aided by grants from the Rockefeller Foundation, the American Cancer Society, and the Atomic Energy Commission (Contract No. AT(11-1)235). In *Advances in Virus Research* (Smith, K. M., and Lauffer, M. A., eds) Vol. 6 pp. 229-356, Academic Press
433. Sonneborn, T. M. (1939) Paramecium aurelia: Mating Types and Groups; Lethal Interactions; Determination and Inheritance. *The American Naturalist* **73**, 390-413
434. Dippell , R. V. (1958) The Fine Structure of Kappa in Killer Stock 51 of Paramecium aurelia. Preliminary Observations. *The Journal of Biophysical and Biochemical Cytology* **4**, 125-128
435. Preer, J. R. (1975) The hereditary symbionts of Paramecium aurelia. *Symp Soc Exp Biol*, 125-144
436. Preer, J. R. (1997) Whatever Happened To Paramecium Genetics? *Genetics* **145**, 217-225
437. BEALE, G. H., JURAND, A., and PREER, J. R. (1969) THE CLASSES OF ENDOSYMBIONT OF PARAMECIUM AURELIA. *Journal of Cell Science* **5**, 65-91
438. Preer, J. R., and Jurand, A. (1968) The relation between virus-like particles and R bodies of Paramecium aurelia. *Genetical Research* **12**, 331-340
439. Pond, F. R., Gibson, I., Lalucat, J., and Quackenbush, R. L. (1989) R-body-producing bacteria. *Microbiological Reviews* **53**, 25-67
440. Harding, J. J. (1985) Nonenzymatic covalent posttranslational modification of proteins in vivo. *Advances in protein chemistry* **37**, 247-334
441. Polka, J. K., and Silver, P. A. (2016) A Tunable Protein Piston That Breaks Membranes to Release Encapsulated Cargo. *ACS Synthetic Biology* **5**, 303-311
442. Loquet, A., El Mammeri, N., Stanek, J., Berbon, M., Bardiaux, B., Pintacuda, G., and Habenstein, B. (2018) 3D Structure Determination of Amyloid Fibrils using Solid-State NMR Spectroscopy. *Methods*
443. Huang, W., Bardaro, M. F., Jr., Varani, G., and Drobny, G. P. (2012) Preparation of RNA samples with narrow line widths for solid state NMR investigations. *J Magn Reson* **223**, 51-54
444. Cade-Menun, B. J. (2015) Improved peak identification in 31P-NMR spectra of environmental samples with a standardized method and peak library. *Geoderma* **257-258**, 102-114
445. Kupferschmitt, G., Schmidt, J., Schmidt, T., Fera, B., Buck, F., and Riiterjans, H. (1987) 15 N labeling of oligodeoxynucleotides for NMR studies of DNA-ligand interactions. *Nucleic Acids Research* **15**, 6225-6241

446. Platzer, G., Okon, M., and McIntosh, L. P. (2014) pH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *Journal of Biomolecular NMR* **60**, 109-129
447. Gibson, I. A. N., Chance, M., and Williams, J. (1971) Extranuclear DNA and the Endosymbionts of Paramecium aurelia. *Nature New Biology* **234**, 75-77