THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : PHYS - Physique
Spécialité : Physique pour les Sciences du Vivant
Unité de recherche : Institut de Biologie Structurale

Caractérisation structurale et dynamique du désordre intrinsèque, des assemblages fibrillaires et des propriétés de gélification d'une protéine de tardigrade les protégeant de la déshydratation

Structural and dynamic characterisation of the intrinsic disorder, fibrillar assemblies and gelation properties of a dehydration-protective tardigrade protein

Présentée par :

Anas MALKI

Direction de thèse :

Martin BLACKLEDGE — Directeur de thèse
DIRECTEUR DE RECHERCHE, CEA - Centre de Grenoble

Rapporteurs :

Christian GRIESINGER
PROFESSEUR, University of Göttingen
Roland RIEK
FULL PROFESSOR, ETH Zürich

Thèse soutenue publiquement le **18 décembre 2024**, devant le jury composé de :

Christian GRIESINGER, — Rapporteur
PROFESSEUR, University of Göttingen
Roland RIEK, — Rapporteur
FULL PROFESSOR, ETH Zürich
Montserrat SOLER LOPEZ, — Examinatrice
SCIENTIST, European Synchrotron Radiation Facility
Franz BRUCKERT, — Président du jury
PROFESSEUR DES UNIVERSITES, Grenoble INP - UGA

Invités :

Laurent BLANCHOIN
Directeur de Recherche, CNRS Délégation Alpes

Acknowledgments

I am deeply grateful to Martin for giving me the opportunity to do my PhD in his team. Thank you for trusting me and giving me the immense freedom to explore, to learn and to grow, for your many corrections, advices, teachings and scientific rigour. I will benefit from them for years to come.

I also thank Nicola for all your teaching, for having always being willing to help, and for your sense of humour. You were a model and I won't forget the IBS day 2022 anytime soon.

I am grateful towards the jury members Christian Griesinger, Roland Riek, Montserrat Soler Lopez, Laurent Blanchoin and Franz Brukert and for accepting to review this thesis and in particular, for their time and for their expertise.

I also thank Robert for your corrections and insightful comments. You may have (re-) joined the group at the end of my journey, but you certainly contributed to it.

Thank you to all the FDP group members, in particular for the great and particular lab culture we developed together. "Brain covid" is certainly an unforgettable expression. Aux collègues Vincent, Damien et Thibault, nos échanges, tant scientifiques que culturels ou sportifs, étaient précieux et stimulants. On a beaucoup appris et rigolé aux fils des années.

A mention first to the team member that were already there before me and helped me getting started. Malene, Sigrid and Torsten whose insights were always precious, Aldo and Luiza, the postdocs at that time and of course my fellow PhD students Wiktor and Sima. Also the ones that started more or less at the same time as me and were there daily, Laura, Maud and Emmi. This pre-covid team always felt special and so were the team members that started after, Lenette, Maiia, Selin and then Lefan. We sure had some great times in the lab. And of course, I do not forget the ones that came at the end, Elise, Jacqueline and Shivangi.

Et bien sur tous les collègues de l'IBS, JB, Laura, Sami, Guillaume, Sylvain et tous les autres sans qui les soirées aux labo auraient été bien vides.

I am deeply grateful to the precious collaborators that helped and taught me a lot during these years. Without you, this thesis would have been much smaller and limited.

Merci à vous Jean-Luc et Jean-Marie pour avoir été là, pour votre disponibilité et votre amitié depuis la toute première année jusqu'à la toute fin. On en aura passé du temps devant l'AFM à regarder des fibres se former !

Thank you Max for always have been so willing to shoot crystals and to teach. I don't know how many hundreds of terrible crystals I brought you before we finally got something!

Merci à toi Guy aussi pour avoir suivi ma thèse dans le TAC, pour tes conseils et ta disponibilité. J'aurais aimé t'apporter de meilleurs échantillons !

Thank you also to Guido for following my thesis in the TAC, for your precious advices and willingness to continue trying. Et merci à toi aussi Tanguy, je ne compte plus combien d'heures on a pu passer à se battre avec le gel et les fibres pour packer ces rotors. Une mention également à Loïc et son équipe si sympathique. J'ai beaucoup apprécié mes passages semi-réguliers à Lyon au CRMN.

Merci également à Ghina, à Laurent, Jérémie et Christophe pour ces 2 ans à injecter des cellules. J'ai beaucoup appris grâce à vous et ai apprécié mes passages réguliers dans votre équipe. Je regrette de ne pas avoir pu inclure ces résultats dans ce manuscrit.

I'm also thankful to all the people maintaining and helping with the plateformes and beamlines. En particulier Caroline, Alicia et Adrien que j'ai souvent embêté et qui ont toujours été disponible. Merci aussi à Philippe pour ton aide au CIBB and Mark and Petra at the beamline.

Looking back in time, I probably would not have done this thesis without Alain Milon and Evert Haanappel in Toulouse who transmitted me their passion respectively for NMR and for biophysics. I learnt a lot in Göttingen, and I am particularly grateful to Kumar who taught me all I needed to succeed in a thesis and whose example I still remember years later, to Loren for his teachings and for having me in his lab as well as to Marcel, Stefan, Kai, Riza, Hessam and all the all the other friends from the MPI with whom I discussed a lot.

Je pense aussi à Wafa, Nikola, Sam, Tim, Thomas, Ray, Fanjo et tous les autres collègues de la fac ainsi qu'aux ô combien précieux P., P. et L.

Enfin, cette thèse et les études qui l'ont précédées n'ont été possible que grâce au soutien et à l'aide de mes parents et de ma famille.

Résumé

Les tardigrades sont des animaux microscopiques capables de survivre à des conditions de stress extrêmes et tout particulièrement à une déshydratation sévère et prolongée. Récemment, une famille de protéines intrinsèquement désordonnées, les protéines CAHS (Cytosolic Abundant Heat-Soluble), a été démontrée avoir un rôle critique dans la survie de tardigrades de l'espèce *Hybsibius exemplaris*. De plus, l'expression de ces protéines par des cellules autres que celles de tardigrades, par exemple des levures ou des cellules humaines, leur confère à leur tour une résistance accrue à la déshydratation. La compréhension des mécanismes biophysiques permettant cette survie à la déshydratation présente donc des intérêts majeurs pour la survie de cultures agricoles pendant des sécheresse, pour la préservation d'aliments, de médicaments ou même de tissus biologiques. Cette thèse porte donc sur l'élucidation des mécanismes biophysiques permettant à l'une des protéines ayant montré le plus d'effet protecteur, la protéine CAHS-8 de *Hybsibius exemplaris*, de s'adapter à la déshydratation et d'en protéger son organisme hôte.

Cette thèse débute par une présentation des différentes notions nécessaire et de l'état des connaissances bibliographiques sur ces sujets. Sont traitées tout d'abord les protéines intrinsèquement désordonnées et leur spécificité, puis l'eau, sa structure et sa dynamique ainsi que les effets de la déshydratation et les dégâts qu'elle engendre. Les stratégies de défense cellulaire contre la déshydratation sont ensuite étudiées et finalement, les tardigrades, leur incroyable résistance aux stress et leur protéines désordonnées uniques sont présentés. Ensuite une introduction théorique des techniques les plus utilisées dans cette thèse où l'accent est mis sur les utilisations et cas, parfois rares et complexes, rencontrés dans les résultats présentés ci-après.

Dans la première partie des résultats, la structure secondaire et tertiaire de CAHS-8 est caractérisée par RMN et son ensemble conformationnel est obtenu, démontrant sa nature de protéiné intrinsèquement désordonné et la présence d'un domaine hélicoïdal partiel.

L'étude de la relaxation des spins y indique un mécanisme complexe d'échange chimique avec ce qui paraît être une forme oligomérique transitoire de type superhélice (coiled-coil) intermédiaire entre la forme monomérique de CAHS-8 et des assemblages fibreux qui ont pu être observé par plusieurs techniques et dont l'enchevêtrement forme un hydrogel. Le diagramme de phase ainsi dressé montre une fibrillation et gélation dépendante de la concentration et de la température. Dans un second temps, ces fibres sont caractérisées et l'études de formes tronquées de CAHS-8 permet l'identification des rôles des deux domaines flexibles de la protéine. Enfin, la manipulation de la fibrillation à l'aide de plusieurs composés est explorée.

Dans la seconde partie des résultats, les gels formés par CAHS-8 sont étudiés et caractérisés à l'aide de plusieurs techniques, en particulier l'hydrogel qui se révèle être un mélange non seulement visqueux mais également complexe. Ainsi, la dynamique de différente protéines exogènes incorporées dans le gel, tant des protéines structurées que désordonnées, est étudiée et montre un clair ralentissement de leur diffusion translationnelle et une perturbation de leur diffusion rotationnelle ainsi qu'un apparent effet de confinement spatial. Ces expériences suggèrent que CAHS-8 pourrait former des gels in vivo, dans la cellule des tardigrades pendant la déshydrations et ces expériences renseignent donc sur les effets sur les constituants cellulaires qui pourrait expliquer la remarquable résistance des tardigrades.

Summary

Tardigrades are microscopic animals able to survive extreme stress conditions, and in particular severe and prolonged dehydration. Recently, a family of intrinsically disordered proteins, the Cytosolic Abundant Heat-Soluble (CAHS) proteins, were shown to have a critical role in the survival of tardigrades of the species *Hybsibius exemplaris*. Indeed, the expression of these proteins by cells other than tardigrades, such as yeast or human cells, confers enhanced resistance to dehydration. Understanding the biophysical mechanisms that enable this survival to dehydration is therefore of major interest for the survival of agricultural crops during drought, for the preservation of food, drugs or even biological tissues. Thus, this thesis aims at studying the biophysical mechanisms that enable one of the most protective proteins, CAHS-8 from Hybsibius exemplaris, to adapt to dehydration and protect its host organism.

This thesis begins with a presentation of the various concepts used and the state of bibliographical knowledge on these subjects. First, intrinsically disordered proteins and their specificity are discussed, followed by water, its structure and dynamics, and the effects of dehydration and the damage it causes. Cellular defense strategies against dehydration are then studied and finally, tardigrades, their incredible resistance to stress and their unique disordered proteins are presented. This is followed by a theoretical introduction to the techniques most widely used in this thesis, focusing on the sometimes rare and complex uses and cases encountered in the results presented below.

In the first part of the results, the secondary and tertiary structure of CAHS-8 is characterised by NMR and its conformational ensemble is obtained, demonstrating its nature as an intrinsically disordered protein and the presence of a partial helical domain. Spin relaxation studies indicate a complex mechanism of chemical exchange with what appears to be a transient oligomeric coiled-coil form intermediate between the monomeric form of CAHS-8 and the fibrous assemblies observed by several techniques and whose entanglement forms a hydrogel. The resulting phase diagram shows

concentration- and temperature-dependent fibrillation and gelation. These fibers were then characterised, and studies of truncated forms of CAHS-8 identified the roles of the protein's two flexible domains. Finally, the manipulation of fibrillation using several compounds is explored.

In the second part of the results, the gels formed by CAHS-8 are studied and characterized using several techniques, in particular the hydrogel, which is not only viscous but also a complex mixture. The dynamics of various exogenous proteins incorporated in the gel, both folded and disordered, are studied and show a clear slowing of their translational diffusion and a disruption of their rotational diffusion, as well as an apparent spatial confinement effect. These experiments suggest that CAHS-8 may form gels in vivo, in the cell of tardigrades during dehydration, and thus provide insights into the effects on cellular constituents that could explain the remarkable resistance of tardigrades.

Foreword

When this thesis started, in October 2019, not much was known about the tardigrade CAHS proteins. The seminal paper announcing the discovery of the CAHS family of protein was published 2012 by Yamaguchi and co-workers from the Kunieda team, and in 2017, Boothby and co-workers demonstrated their role in desiccation protection in tardigrades, with these two papers essentially constituting the then entire bibliography on these proteins.

Thus, the initial objective of this thesis was to uncover the molecular mechanisms of tardigrade stress resistance, with everything having to be established and discovered. Four years later, there is a clearer idea of the properties of CAHS proteins thanks to several articles published by these two teams, ours, as well as new teams having started to work on the subject. As such, concepts that are established and central now in the CAHS field, such as the fibrillation of these proteins, was not even a hypothesis four years ago. This also explains why some of the work presented in the next chapters has since then been published elsewhere by other teams.

The objectives of this thesis manuscript are therefore: to give a clear picture of the different state of CAHS-8, from a monomeric intrinsically disordered protein to gels, via dimers, fibrils and fibres, and how this may be related to the desiccation-protective abilities of this protein.

The structure of the thesis is the following:

The chapter 1 constitute the general introduction and starts by presenting intrinsically disordered proteins. Then, water, its importance for life and proteins, as well as dehydration, its nefarious effects and strategies to avoid it are discussed. Finally, tardigrades and their remarkable abilities are introduced.

The second chapter is the theoretical introduction and presents the main techniques used.

The chapter 3 gives a general overview of the protein and its phase diagram, which eases the comprehension of the results as they are presented in a hierarchical way.

The chapter 4 is the reproduction of the published article.

The results are then separated in two parts:

> The part 1 is constituted of the chapter 5 and 6, which cover respectively the helical domain of CAHS-8 and its dimerisation for the former, and the disordered chains of CAHS-8 and its fibrillation for the later.
>
> The part 2 is constituted of the chapter 7 and 8 which cover respectively the gels of CAHS-8 and its effects on other proteins entrapped in it.

Each of these two parts begins with an introduction of the specific concepts used in their chapters and concludes with a discussion. As such, the introduction of the first part presents α-helices, coiled-coils and fibres, while the introduction of the first part discusses gels, gelation and dynamics in a complex environment.

Finally, the chapter 9 is the conclusion of the thesis and discusses its perspectives.

Glossary

Abbreviation	Meaning
AFM	Atomic Force Microscopy
CAHS	Cytosolic Abundant Heat Soluble
CPMG	Carr-Purcell-Meiboom-Gill
cryoEM	Cryo Electron Microscopy
CSA	Chemical Shift Anisotropy
DCC	Dimerisation-driving Coiled-Coil
DD	Dipole-Dipole
DLS	Dynamic Light Scattering
η_{xy}	Transversal cross-correlated relaxation rates
η_z	Longitudinal cross-correlated relaxation rates
FCC	Fibrillation-driving Coiled-Coil
FT-IR	Fourier-Transform Infrared
H.ex	*Hypsibius exemplaris*
HD	Helical Domain construct
HDCTD	Helical Domain and C-Terminal disordered chain construct
IF	Intermediary Filament
LEA	Late Embryogenesis Abundant
LEAM	Late Embryogenesis Abundant Mitochondrial
LLPS	Liquid-Liquid Phase Separation
MAHS	Mitochondrial Abundant Heat Soluble
MD	Molecular Dynamics
NMR	Nuclear Magnetic Resonance
nOe	Nuclear Overhauser Effect
NS-TEM	Negative-Stain Transmission Electron Microscopy
NTDHD	N-Terminal disordered chain and Helical Domain + construct
PRE	Paramagnetic Relaxation Enhancement
R.va	*Ramazzottius varieornatus*
R_1	Longitudinal autorelaxation rate
$R_{1\rho}$	Longitudinal relaxation rate in the rotating frame
R_2	Transverse autorelaxation rate
RDC	Residual Dipolar Coupling
R_{ex}	Exchange contribution to the transverse autorelaxation rate (R_2)
RH	Relative Humidity
SAHS	Secreted Abundant Heat Soluble
SAXS	Small-Angles X-ray Scattering
ssNMR	Solid-State NMR
TDP	Tardigrade-specific Disordered Proteins
TFE	2,2,2-Trifluoroethanol
TMAO	Trimethylamine N-oxide

Table of contents

Résumé 3

Summary 5

Foreword 7

Glossary 9

Table of contents 10

Table of figures 16

Chapter 1: Disordered proteins, water and tardigrades 19

1.1 Intrinsically Disordered proteins 20

1.1.1 Of structure and disorder 20

1.1.1.1 Folded proteins and intrinsically disordered proteins 20

1.1.1.2 Energy landscape of proteins 21

1.1.2 Interaction of IDPs 22

1.1.3 Advantages of disorder 23

1.2 Water 25

1.2.1 Description of water 26

1.2.1.1 Bulk water 26

1.2.1.2 Hydration layer 27

1.2.1.3 Solvent dependent dynamics 28

1.2.2 Dehydration 30

1.2.2.1 Cellular effects 30

1.2.2.2 Practical aspects 33

1.2.3 Anhydrobiosis 34

1.3 Tardigrades 39

1.3.1 Biology and ecology 39

1.3.2 Extremotolerance 41

1.3.2.1 The tun state 41

1.3.2.2 Anhydrobiotic abilities and stress resistance 42

1.3.2.3 Earlier hypotheses: Trehalose vitrification and LEA proteins, again? 44

1.3.3 Tardigrade specific IDPs 45

1.3.3.1 TDPs 45

1.3.3.2 DSup 48

1.3.3.3 LEA proteins 49

1.3.3.4 Beyond the current nomenclature 50

Chapter 2: Theoretical background *51*

2.1 Nuclear Magnetic Resonance 52

2.1.1 Of Spin 52

2.1.2 Classical view of NMR 53

2.1.2.1 Zeeman effect 53

2.1.2.2 Larmor precession 55

2.1.2.3 Bloch equations 56

2.1.3 Quantum view of NMR 59

2.1.3.1 Mixed states and density operator 59

2.1.3.2 Bloch-Wangsness-Redfield theory 62

2.1.3.3 Spectral density function and auto-correlation function 67

2.1.3.4 The relaxation superoperator 69

2.1.4 Relaxation mechanisms 72

2.1.4.1 Dipole-dipole induced relaxation 72

2.1.4.2 Chemical-shift anisotropy induced relaxation 74

2.1.5 Relaxation and dynamics of the protein backbone amide ^{15}N 75

2.1.5.1 Relaxation rates 75

2.1.5.2 Residual dipolar couplings 79

2.1.5.3 Rotational diffusion 80

2.1.5.4 Model-free analysis of folded proteins 84

2.1.5.5 Model-free analysis of intrinsically disordered proteins 86

2.1.6 Chemical exchange 87

2.2 Scattering 93

2.2.1 The wave scattering phenomenon 93

2.2.2 Small-Angle X-Ray Scattering 94

2.2.2.1 Why small angles? 94

2.2.2.2 Pair distance distribution electron 96

2.2.2.3 Radius of gyration 97

2.2.2.4 Distinguishing compact and flexible objects 98

2.2.3 X-Ray Diffraction crystallography 100

2.2.3.1 Crystals 100

2.2.3.2 Diffraction by crystals 100

2.2.3.3 The phase problem 103

2.2.3.4 Non-crystallographic symmetries and crystal defects 107

2.2.3.5 The B-factor 108

2.2.4 Dynamic Light Scattering 109

2.2.4.1 Scattering considerations 109

2.2.4.2 Fluctuations and autocorrelation function 109
2.2.4.3 DLS in non-ergodic media 112
2.2.4.4 Hydrodynamic radius 114

2.3 Atomic Force Microscopy 115
2.3.1 Tip-sample interaction 115
2.3.2 Different modes, different experiments 116
2.3.3 Tip-sample convolution 117

Chapter 3: Overview of CAHS-8 and its different forms 119

Chapter 4: Published article 120

Part 1: CAHS-8 – Monomer, coiled-coils and fibres 140

Chapter 5: CAHS-8 and its helices 141

Introduction 142
α-helices 142
Coiled-coils 143
Canonical double stranded coiled-coils 143
Discontinuities and non-canonical repeats 145
Geometrical analysis of coiled-coils 147
Coiled-coils based fibrils 148

5.1 Results 151
5.1.1 CAHS-8, an IDP with a long helical domain 151
5.1.1.1 Oscillating patterns 154
5.1.1.2 Sequence analysis 158
5.1.1.3 Evidence of oligomerisation 161
5.1.1.4 A high temperature dependence 163
5.1.2 Study of the helical domain 168
5.1.2.1 Similarities and differences in spin relaxation 168
5.1.2.2 Crystals of the coiled-coil dimer 171
5.1.2.3 HD has a reduced ability to fibrillate 185
5.1.2.4 Model-free analysis of the helices 189

Chapter 6: The fibres and the disordered chains 194

6.1 Results 195
6.1.1 Structure of the fibrillar assemblies of CAHS-8 195
6.1.1.1 Fibrils in the lattice 195
6.1.1.2 A second coiled-coil drives the fibrillation 201
6.1.1.3 The fibrillation can be repressed by perturbating the FCC 206

6.1.2 Role of the disordered termini in the fibre assembly 209
6.1.2.1 Model-free analysis of the disordered termini 209
6.1.2.2 One-armed constructs 216
6.1.3 CAHS-8 can form fibres-like crystals of the fibrils 223
6.1.3.1 The dehydration rate controls the assembly of CAHS-8 223
6.1.3.2 An assembly of crystalline fibrils 227
6.1.4 Manipulation of the fibrillation 234
6.1.4.1 TFE 234
6.1.4.2 TMAO 236
6.1.4.3 Gadolinium 238

6.2 Discussion 239

6.2.1 A first characterisation of CAHS proteins 240
6.2.2 Stutters, hendecads and an antiparallel orientation 241
6.2.3 A transient coiled-coil and possible sequence of events 242
6.2.4 Relevance for CAHS and LEA proteins 246
6.2.5 Crystallisation of CAHS-8 248
6.2.6 The fibril structure and assembly mode 250
6.2.7 Assembly of the fibres 253
6.2.8 The crystalline fibres: a saturation-induced physiological state? 255

6.3 Material and methods 260

6.3.1 Sample expression and purification 260
6.3.2 Sequence alignment and coiled-coil prediction 261
6.3.3 Solution-state NMR 262
6.3.3.1 Assignment 262
6.3.3.2 Secondary chemical shifts 262
6.3.3.3 Relaxation measurement 263
6.3.3.4 Chemical Shifts Perturbation 264
6.3.3.5 Paramagnetic Relaxation Enhancement 264
6.3.3.6 Residual Dipolar Couplings 264
6.3.3.7 Analysis of the relaxation rates 265
6.3.3.8 CPMG relaxation dispersion 266
6.3.4 Solid-state NMR 266
6.3.4.1 Microcrystals 266
6.3.4.2 Crystallised fibres 267
6.3.5 Dynamic Light Scattering 267
6.3.6 Small-angle X-Ray Scattering 267
6.3.7 X-Ray diffraction crystallography 268
6.3.8 Fibre diffraction 269
6.3.9 Atomic Force Microscopy 270

6.3.10 Electron Microscopy 270
6.3.10.1 Negative-stain Transmission Electron Microscopy 270
6.3.10.2 Cryo Transmission Electron Microscopy 271
6.3.11 Manipulation of the fibrillation 271
6.3.11.1 TFE 271
6.3.11.2 TMAO 271
6.3.11.3 Solvent PREs 272

Part 2: CAHS-8 – Its gels and their properties 273

Chapter 7: The gels of CAHS-8 274

7.1 Introduction 275

7.1.1 Gels 275
7.1.1.1 Gels and gelation 275
7.1.1.2 Hydrogels, water and dry gels 278
7.1.1.3 Gels, fibres, crystals and glasses 279
7.1.2 Dynamics in complex environments 280
7.1.2.1 Shear and bulk viscosities 280
7.1.2.2 Crowding and confinement 282

7.2 Results 291

7.2.1 The fibrous hydrogel and xerogel 291
7.2.1.1 The entangled fibrils form a gel 291
7.2.1.2 The disordered terminal chains of the fibres are free in the gel state 296
7.2.1.3 No exchange of the free CAHS-8 with the fibres could be characterised 301
7.2.2 The fibrous aerogel 302

Chapter 8: Effects of the CAHS-8 hydrogel on entrapped proteins 306

8.1 Results 307

8.1.1 Proteins and macromolecular assemblies can be encapsulated in the gels 307
8.1.2 The gel is a viscous environment in which diffusion is slowed and restricted 309
8.1.3 The disordered termini of CAHS-8 are in contact with entrapped proteins 315
8.1.4 Interactions are repressed in the hydrogel 320

8.2 Discussion 325

8.2.1 The gelation of CAHS-8 327
8.2.2 Formation and crystallinity of the aerogel 328
8.2.3 Organisation, viscosity, crowding and confinement in the hydrogel pores 329
8.2.4 Effects of the pore environment on encapsulated proteins 331
8.2.5 Slow-down of the interactions in the hydrogel 332
8.2.6 Model of the molecular basis of the desiccation resistance of tardigrades 334

8.3 Material and methods 337

8.3.1 Sample expression and purification 337

8.3.2 Negative-stain Transmission Electron Microscopy 337

8.3.3 Atomic Force Microscopy 338

8.3.4 Dynamic Light Scattering 339

8.3.5 Differential Scanning Calorimetry 340

8.3.6 Nuclear Magnetic Resonance 340

8.3.6.1 Solid-state NMR 340

8.3.6.2 Solution-state NMR 341

8.3.7 X-Ray Diffraction 343

8.3.8 Small-Angle X-Ray Scattering 344

Chapter 9: Conclusion and perspectives 345

References 348

Annex 375

Model free analysis reproduction of the experimental rates 375

Table of figures

Figure 2.2.1: Examples of spectral density function and associated normalised autocorrelation function. 70
Figure 0.2: Dependence of the relaxation rates on the correlation time. 77
Figure 2.3: Effects of the exchange rates, populations and $R2B$ on the linewidths. 91
Figure 2.4: DLS autocorrelation functions for ergodic and non-ergodic media. 112
Figure 5.1: CAHS-8 is an intrinsically disordered protein with a central helical domain. 152
Figure 5.2: Absence of tertiary structure elements in CAHS-8. 153
Figure 5.3: Relaxation of CAHS-8 at 293K 155
Figure 5.4: The helices of CAHS-8 present local variations. 157
Figure 5.5: The helical domain sequence is highly conserved amongst CAHS proteins. 159
Figure 5.6: Prediction and evidence of concentration-dependent coiled-coil oligomerisation of CAHS-8. 162
Figure 5.7: Temperature-dependent behaviour of CAHS-8 164
Figure 5.8: Relaxation of CAHS-8 at 303K. 165
Figure 5.9: Relaxation of CAHS-8 at 278 K. 166
Figure 5.10: Temperature variation of the relaxation of the helical domain of CAHS-8. 167
Figure 5.11: Comparison of the HD construct with the helical domain of the full length CAHS-8. 169
Figure 5.12: Comparison of the relaxation of HD with the corresponding helical domain of CAHS-8 at 293 K and 600 MHz. 170
Figure 5.13: Comparison of the relaxation of HD with the corresponding helical domain of CAHS-8 at 293 K and 600 MHz. 171
Figure 5.14: Tentative ssNMR study of HD microcrystals. 172
Figure 5.15: Crystallographic structure of the antiparallel coiled-coil dimer of the HD construct of CAHS-8. 175
Figure 5.16: Coiled-coil metrics of the structure. 178
Figure 5.17: The coiled-coil crystal structure presents particular deformations and is present in solution. 182
Figure 5.18: HD has a reduced ability to form fibrillar assemblies compared to CAHS-8. 186
Figure 5.19: The deuterated HD at 278 K shows signs of the coiled-coil formation. 188
Figure 5.20: The model-free analysis treatment of the spin relaxation rates of CAHS-8 and HD shows different exchange profile. 190
Figure 6.1: The outer face of the C-terminal half of the helical domain of CAHS-8 is apolar. 196

Figure 6.2: Structure of the fibrils and fibres of CAHS-8. 199
Figure 6.3: Cryo Electron Microscopy study of the fibrillar assemblies of CAHS-8..... 200
Figure 6.4: The coiled-coil dimers assemble with each other via a second antiparallel coiled-coil to form fibrils. 202
Figure 6.5: The fibrillation-driving coiled-coil is stabilised by salt-bridges mediated by four lysines in core positions. 205
Figure 6.6: A single mutation in the fibrillation-driving coiled-coil is enough to repress the fibrillation. 207
Figure 6.7: Relaxation of the N-terminal disordered chain of CAHS-8 and model-free analysis. 211
Figure 6.8: Relaxation of the C-terminal disordered chain of CAHS-8 and model-free analysis. 211
Figure 6.9: The different assemblies of CAHS-8 result in physical restrictions of the disordered termini which show various level of conservation. 214
Figure 6.10: Comparison of the relaxation of the different constructs of CAHS-8 at 600 MHz, 350 µM and 303 K. 217
Figure 6.11: Comparison of the relaxation of the different constructs of CAHS-8 at 600 MHz, 350 µM and 293 K. 219
Figure 6.12: Comparison of the relaxation of the different constructs of CAHS-8 at 600 MHz, 350 µM and 278 K. 220
Figure 6.13: The disordered termini are important for the gel formation and the morphology of the fibrillar assemblies of CAHS-8..... 222
Figure 6.14: CAHS-8 assembles into crystalline fibrillar morphologies. 225
Figure 6.15: The assembly of HD into crystalline fibres depends on the RH and concentration. 226
Figure 6.16: The full-length CAHS-8 can assemble into crystalline fibres and dendritic crystals. 228
Figure 6.17: The different constructs of CAHS-8 form crystalline fibres that orient under magnetic field and are composed of c-axis grown crystals. 230
Figure 6.18: Fibre diffraction of the crystalline fibres grown under the magnet of CAHS-8 and HD. 232
Figure 6.19: The magnetically oriented crystalline fibres of CAHS-8 are composed of a lattice of crystallised twisting fibrils..... 233
Figure 6.20: TFE induce a gelation and a particular fibrillation of CAHS-8 and HD. 235
Figure 6.21: TFE induces a gelation and a particular fibrillation of CAHS-8 and HD. 237
Figure 7.1: The fibrous gels of CAHS-8 observed by EM and AFM. 292
Figure 7.2: Phase diagram of CAHS-8 and characterisation of its gelation and melting. 293
Figure 7.3: Solid-state NMR spectra of the CAHS-8 hydrogel. 296

Figure 7.4: The gel-forming fibres of CAHS-8 have free disordered termini and coexist with free CAHS-8 species. 298
Figure 7.5: Comparison of the relaxation of CAHS-8 in its soluble and hydrogel forms at 293 K. 299
Figure 7.6: The aerogel of CAHS-8 is highly porous and constituted of fibres. 303
Figure 7.7: The aerogel of CAHS-8 displays crystallinity but dubious glass properties. 305
Figure 8.1: AFM imaging of proteins and macromolecular assemblies entrapped in the CAHS-8 hydrogel. 308
Figure 8.2: Proteins can tumble in the CAHS-8 hydrogel without significant change of their fold or conformational sampling but their translational diffusion is slowed down. 310
Figure 8.3: The relaxation of Sendai NTail is modified in the CAHS-8 hydrogel. 312
Figure 8.4: The cavities of the CASH-8 hydrogel are viscous environments in which the diffusion is confined. 314
Figure 8.5 : The relaxation of Sendai NTail in the CAHS-8 is not only an effect of viscosity. 317
Figure 8.6: The hysteresis of CAHS-8 shows effects on Sendai NTail specific to the gel state. 318
Figure 8.7: Interactions are perturbed in the CAHS-8 hydrogel. 322
Figure 8.8: The hysteresis of CAHS-8 demonstrates that the hydrogel phase has a specific effect on the interactions. 324

Chapter 1: Disordered proteins, water and tardigrades

"Water is a very good servant, but it is a cruell maister."
William Bullein,
Bulwarke of Defence against All Sicknesses, 1562

1.1 Intrinsically Disordered proteins

1.1.1 Of structure and disorder

1.1.1.1 Folded proteins and intrinsically disordered proteins

Proteins are biomolecules of prime importance for living organisms. They are involved in a broad range of biological processes, including metabolic biochemical reactions, passing signalling information, shaping cells or protecting organisms from harmful infections. As such, proteins are essential for life and they must be diverse enough to successfully carry out those different functions. Consequently, disruption of their properties is associated with a vast number of pathologies.

Their exact composition is encoded in the DNA, in the corresponding gene. This gene's codons are translated into amino acids, which form the base unit of proteins and are assembled by peptide bonds into the unbranched polypeptide chain that forms the biological heteropolymer we define as protein. The 22 types of amino acids, or residues, found in proteins vary in charge, hydrophobicity, size or bulkiness and their position and abundance in the chain give their properties to proteins[1].

For decades, the paradigm in the study of proteins was that each protein's amino-acid sequence corresponds to a well-defined three-dimensional structure which in turn corresponds to a particular biological function[2]. However, proteins able to carry a biological function while being in an unstructured, disordered state were already observed in the 1950s, and it became apparent in the late 1990s and 2000s that a large part of the proteome is actually made of Intrinsically Disordered Proteins (IDPs), i.e. proteins lacking a stable three-dimensional fold, but which, nevertheless, do have biological functions, breaking the hitherto held structure-function duality[3–9].

Of course, this disorder can vary from a simple loop between folded domains to entire chains being unfolded. Therefore, proteins that are mostly disordered are called IDPs while specific disordered regions within IDPs or folded proteins are commonly termed Intrinsically Disordered Regions (IDRs)[8,10].

1.1.1.2 Energy landscape of proteins

A protein can theoretically exist in a large number of possible conformations defined by its sequence, i.e. the order and properties of its component amino acids. As a consequence, the structure of a folded proteins corresponds to a thermodynamically stable conformation amongst all those possible conformers, that populates a deep minimum of the free-energy landscape of the protein fold in its physiological conditions as the enthalpy gained by reaching this structure must be high enough to compensate the loss of conformational entropy. This landscape, and the corresponding structure, is therefore dependent on the properties of the amino-acid sequence of the proteins as several forces lead to non-covalent interactions or conversely to repulsions between atoms and are ultimately responsible for the protein fold[11,12].

Indeed, as several amino acids have a dipole moment, whether induced or permanent, the collectively termed short-range van der Waals interaction can lead to attraction or dispersion between those dipoles, and as several sidechains have a net charge, the electrostatic ionic bonding or repulsion can play a significant role in deciding the ultimate structure. Moreover, hydrogen bonds, i.e. the non-covalent attraction between a hydrogen atom covalently bound to an electronegative donor atom and the lone electron pair of an electronegative acceptor one, is of particular importance for protein structure as a stabilizing force of secondary structures while the combination of a hydrogen and a ionic bond, that is a salt bridge, is also a strong stabilizing attractive interaction. Last but not least, the interaction between solvent water and proteins is crucial to the structure of the protein, as the hydrophobic effect is the main driver of protein folding. Indeed, a number of amino-acid sidechains are hydrophobic and thus it is highly energetically favourable to bury them in a core shielded from water which in turn requires the cooperative folding of more hydrophilic residues on the solvent exposed surface[11–15].

Intrinsically Disordered Proteins, in turn, lack a well-defined minimum in their free-energy landscape and have instead multiple local minima corresponding to an ensemble of conformations which are continuously thermally sampled by the protein and whose populations and structures vary over time. Therefore, instead of describing a single structure, the notion of an ensemble of conformers is more adapted to describe the

conformational sampling of IDPs. This flatter free-energy landscape leading to disordered conformations is therefore a direct consequence of their amino-acid composition, which is characterised by an enrichment in polar residues and in prolines as well as a depletion in aromatic and hydrophobic ones, while often presenting a high net charge. This compositional bias is therefore used to predict the disorder level of a sequence which is therefore encoded in the protein sequence itself, hence the adjective "intrinsic"[4,8,16–18].

1.1.2 Interaction of IDPs

When their function is not disorder itself, as is the case for entropic chains, IDPs are often involved in interactions and typically bind to their partners through short recognition motifs[19–22], with their binding mechanisms being commonly separated into the two conceptual mechanisms of *conformational selection* and *induced fit* mechanism, although the binding of IDP is typically a combination of these two mechanisms, and with multi-step binding processes being difficult to categorise by such simplistic models[22–26].

IDPs sample an extended range of conformations which may include the bound-state one, which would therefore exist before binding. Conformational selection means that amongst the many thermally accessible conformers of the IDP landscape, only one or several particular *conformations* can bind and are therefore *selected*, even if they are weakly populated. As such, the classical "lock and key" model of protein binding amounts to the particular subcase of conformational selection when the selected conformer is the one with the lowest conformational energy[25–29].

In IDPs, the induced fit mechanism, refers to a situation in which the protein binds to its partner in a mostly disordered state and only then adopts a more structured and final conformation within the complex. This is typically observed when the bound conformation of the IDP involves a tertiary structure unlikely to be spontaneously populated in solution given the absence of any hydrophobic core. As such, the conformational selection picture would therefore not be able to explain the observed kinetics as the bound conformation does not exist before binding[23,25,26,30]. Moreover, IDPs can fold upon binding differently depending on the binding partner. For example, the

C-terminal IDR of the tumour suppressor p53 can fold into an α-helix, a β-strand or two different irregular structures depending on its partner[31].

These two mechanisms can also coexist. Indeed, as IDP binding often occurs stepwise, an initial step of conformational selection for the initial stabilisation of the complex and then its finalisation through the induced fit mechanism may well be thermodynamically favourable. As an example, the binding site of the free disordered C-terminus tail of the Sendai virus nucleoprotein samples different conformations ranging from totally disordered to highly helical[32]. Upon interaction with its phosphoprotein partner, one of these helical substates is first selected in a nonspecific encounter complex before locating into its binding site, an inter-helical groove on the surface of its partner, at a rate coincident with the groove's breathing motions, while it undergoes a modest conformational change[33].

On the other hand, bound IDPs can also continue to sample different conformations and form a dynamic or fuzzy complex[34]. Examples include the interaction of the IDR regulatory domain of the mitogen-activated protein kinase (MAPK) MKK4 with its partner, the MAPK p38α, through a docking motif that remains rigid in the complex, while the interaction motif remains highly dynamic and samples several conformations on the surface of its partner[35]. Another example is the picomolar affinity interaction between the partly disordered H1 histone and the IDP prothymosin-α chaperone, with both partners remaining disordered in an electrostatically-driven and highly dynamic complex[36].

Finally, multivalent binding is not uncommon in IDPs thanks to their flexibility and short linear motifs. For example, the phenylalanine and glycine rich (FG) nucleoporin Nup153 shows highly plastic and fast low-affinity multivalent binding to importin-β through its multiple FG repeats which remain disordered[37].

1.1.3 Advantages of disorder

Disorder is a great advantage for interactions. Indeed, the Einstein-Smoluchowski diffusion limit of the association rate of the interaction of globular proteins in the absence

of external forces is estimated to 10^9 M^{-1} s^{-1}, but this limit is hardly ever reached in the absence of an electrostatic attraction component, in which case the limit is estimated to be approximately 10^6 M^{-1} s^{-1}. However, IDPs have been shown to display higher than predicted on-rates and to approach this limit, which may be due to their intrinsic disorder [33,37–40]. Moreover, this high flexibility also permits them to avoid steric hindrance with their partner as seen in the induced fit mechanism, permitting them to physically reach interaction cavities at the right conformation more efficiently than folded proteins[20,41,42], and in general, allows them the diversity of binding modes previously mentioned.

Moreover, IDPs are also well suited for stress survival and stress response. Indeed, by definition, a disordered chain cannot be denatured even in harsh conditions of temperature, salt concentration or pH and therefore may remain functional provided it does not aggregate, all while secondary structure elements in IDPs can undergo reversible thermally-induced folding or unfolding. In turn, even if the IDP is not functional in these stress conditions, if, for instance, a high salt concentration prevents the IDP from binding to a partner through electrostatic interactions , the return to physiological conditions will reactivate this interaction while a folded protein may have been denatured or have aggregated and therefore may not be functional anymore[43,44].
Conversely, IDPs can be non-functional in physiological conditions but may undergo a structural change, whether folding or unfolding, in response to stress conditions or to stress-induced post-translational modifications that will make them active[44–46].

Responses to changes in conditions by IDPs are well illustrated by liquid-liquid phase separation (LLPS) that can be triggered by changes in local protein or salt concentration, pH, temperature, post-translational modification or interactions[47]. For example, the IDRs comprising plant proteins ELF3 and FLOE1 undergo LLPS in response to respectively changes of temperature[48,49] and to dehydration stress[50], while the IDR of the yeast proteins Sup35 and Pab1 respectively drive and regulate their phase separation into gel-like droplets in response to changes of pH[51,52].

1.2 Water

"Grain was scant,
Kernels were dried up,
Scarce was every kind of food.
Every man robbed his twin,
Those who entered did not go.
Children cried,
Youngsters fell,
The hearts of the old were grieving;
Legs drawn up, they hugged the ground,
Their arms clasped about them.
Courtiers were needy,
Temples were shut,
Shrines covered with dust,
Everyone was in distress."[53]

It is by these words that the so-called Famine Stela, erected around 200 BC in ancient Egypt, describes an unforgettable seven-year drought that happened approximately 2400 years before it was carved[54]. Indeed, droughts have been feared by humanity since the dawn of time, as without water, there can be no life, crops wither, animals starve and, ultimately, humans die.

The lack of water can therefore have catastrophic consequences, as exemplified by the so-called Late Bronze Age collapse that set back civilisation by hundreds of years and is currently thought to have been caused, at least partly, by persistent droughts[55–58]. It is therefore not surprising that humanity's oldest known story, the Sumerian epic of Gilgamesh dated from the 2nd millennium BC, has the eponymous hero defeating drought materialised as a bull whose snorts dries and cracks Earth, in which people are swallowed[57,59,60].

Unfortunately, as climate changes and the frequency of droughts increases if the current trend continues, water will become an increasingly scarce and important resource[61].

1.2.1 Description of water

Without water, there is no life. Second only to oxygen in importance, water is needed for many essential biochemical reactions, amongst them cell respiration and photosynthesis which provide cells with energy[62]. Indeed, most constituents of living organisms evolve in an aqueous medium and as such, water make up for more than 70% of the mass of cells[62], constituting overall more than 80% of the mass of plants[63] and more than 65% of the mass of an average human adult[64].

1.2.1.1 Bulk water

The water molecules that are not influenced by hydrated molecules are referred to collectively as the bulk water. The most energetically favourable configuration for each of these water molecules is to form a tetrahedral geometry, with each central H_2O bonding with four others located at the corner of the tetrahedron with two giving and two receiving hydrogen bonds, which is the structure water adopts in its ice state[65–67]. However, in practice, this bonding is dynamic as it is constantly reshaped due to the constant breaking and reformation of hydrogen-bonds with different molecules on a few picoseconds timescale resulting from the so-called angular jump of water molecules[68–74]. This mechanism, first described in Molecular Dynamics (MD) simulations and later supported by 2D Fourier-Transform Infrared (FT-IR) spectroscopy consists in a large-amplitude reorientation of the OH groups as the molecule changes its hydrogen-bonding partners[68–75], in an evolution of the historical Debye picture of small-amplitude reorientations driven by rotational diffusion of the water molecules[68,73,76]. This “jump” mechanism is cooperative as it involves locally at least three water molecules, the departure of at least partner and the approach of at least one previously indirectly connected molecule[68–70,73]. As such, bulk water can be described on the femtosecond timescale as a constantly reorganising three dimensional network of hydrogen-bond clusters spanning the entire bulk[67,77–85]. Moreover, it gives rise to short-lived transient "dangling" non-bonded states[73,86] and thus facilitates the diffusion of the water

molecules, as this constant cooperative rearrangement allows them to rotate and translate without ever experiencing a totally unbonded state[67,87,88].

These angular jump motions of the bulk water molecules and the reshaping of their connectivity coexists with other different dynamic motions occurring over hundreds of femtoseconds that include librational motions of one, a few or an entire cluster of hydrogen-bonded H_2O molecules[68,89–91] as well as stretching and bending vibrations of the intermolecular hydrogen-bond network[89,90,92].

1.2.1.2 Hydration layer

Proteins are hydrated in water, which means that they have a hydration layer defined as a sheath of water molecules covering them. These H_2O molecules interact with the protein and their local structure and dynamics differ from those of the bulk water to the point that they have been referred to as "bound" water[93,94]. This term can be deceptive, as the hydration layer is not static and individual hydration shell molecules are constantly experiencing an exchange process from a "bound" to a "free" state on the picosecond timescale, while others are experiencing the reverse process due to the constant translational diffusion of water[88,90,95].

As these interfacial water molecules form hydrogen bonds with the proteins, they depart from the previously described coordinated tetrahedral hydrogen-bond network of bulk water that facilitates their rotational diffusion, which is therefore slowed down due to the smaller number of hydrogen-bonding opportunities and the local confinement imposed by the protein structure[69,96,97]. Indeed, the rotational diffusion of the hydration layer has been found to be, on average, approximately two times slower than that of bulk water[97,98], while its translational diffusion is also perturbed, being about two to three times slower than in the bulk[99,100]. However, the exact range of this layer and therefore of this perturbation has been subject to discussion and has been recently estimated to amount to one molecule of water[101].

Of course, water is critical to the structure of proteins due to its role in protein folding[85,102,103], but some water molecules can even be considered part of the protein

structure. Technically part of the hydration layer but considered differently, internal waters are the water molecules that localize in a consistent way in cavities of globular proteins, cannot be connected to the external water by a continuous chain and are necessary for the functions of these proteins[104,105]. Indeed, this internal water appears in X-Ray crystallographic structures and can be probed by NMR, to the point that it can be seen as a part of a protein's structure and is therefore important for its function[90,104,105]. Like molecules in the hydration shell, internal water molecules are not static and do exchange with other water molecules, although on slower, nanosecond and longer timescales[106–108].

Because of their intrinsic disorder and high hydrophilicity, IDPs and IDRs tend to have a bigger hydration shell relative to their molecular weight than folded proteins, as their solvent-exposed surface is larger. However, due to the relative absence of particular topography, the more unfolded a protein is, the less the local dynamics of its hydration shell is perturbed in comparison that of the bulk water[109–113].

1.2.1.3 Solvent dependent dynamics

Hydrated proteins at low temperature exist in a glassy state devoid of conformational flexibility in which they are frozen in one of their conformational substates, and therefore, their biological function is severely impaired, if not deactivated[13,114,115]. Indeed, from approximately 180 K to 220 K, proteins are biologically inactive and exist in a harmonic motional regime dominated by vibrations, while 220 K is the approximate onset of an apparent “dynamic transition” at which a gradually increasing renewed anharmonic motional flexibility is observed by Mössbauer spectroscopy[116], inelastic[117,118] and elastic incoherent[119,120] neutron scattering, Terahertz spectroscopy[121], Differential Scanning Calorimetry (DSC)[122], solid-state NMR (ssNMR)[123], X-ray photon correlation spectroscopy (XPCS)[124] and molecular dynamics (MD) simulations[95,125,126].

However, this recovery of motions and apparent dynamic transition is not observed in dry proteins even at room temperature[118,126], highlighting the role of the water in the protein hydration shell which also exhibits an apparent dynamic transition at 220 K[117,120,127,128]. As such, the onset of protein dynamics at this temperature appears to be

triggered by the concomitant restoration of translational diffusion in the hydration shell[95,100,116,120,129,130]. This is also supported by the fact that DNA[131], tRNA[132] and polymers[133] also exhibit a similar transition at about the same temperature. In turn, this gradual activation of protein motions is shifted to higher temperature if the solvent viscosity is increased[130,134,135]. On the other hand, inhibiting solvent translational diffusion, but not its rotational and librational mobility, has a similar effect on protein dynamics as dehydration[95].

Water and protein dynamics therefore appear to be linked and as such, the motions of globular proteins were explained through fluctuations of the hydration shell or cooperative motions between protein segments and their bound water[124,130,136–139].
Direct observation of this protein-solvent coupling by ssNMR revealed its onset at 160 K, the temperature above which rotational diffusion of water resumes[122,123]. As the temperature increases, different transitions could be observed which reported on the hierarchical thermal activation of motional processes[123]. Indeed, a first transition appears around 195 K where protein sidechains start to experience exchange between rotameric states as hydrogen bonds with the hydration shell are formed and broken [123,140]. A second transition occurs at about 220 K, when translational diffusion resumes in the hydration shell, which permits the activation of large-amplitude anisotropic motions occurring on the nanosecond timescale in the protein backbone which have been observed by other techniques (vide supra). Finally, a third transition is observed around 250 K, corresponding to large amplitude motions of the sidechains occurring in the pico- to nanosecond timescale[123]. These conclusions were substantiated by subsequent field dependent studies that exploited the different dependence of longitudinal relaxation on motional timescales as a function of static magnetic field strength[141].

As previously discussed, intrinsically disordered proteins have a larger solvent-exposed surface than folded proteins and perturb the dynamics of their hydration shell less, and thus, their dynamics is expected to be highly correlated with that of the solvent[113]. Indeed, the observed activation of dynamic processes also happens in unfolded proteins, although at a slightly higher temperature than in folded ones[142], and a strong coupling between water and IDP dynamics is also observed by elastic incoherent neutron scattering[111].

This strong effect of water on IDP dynamics was clearly demonstrated by the correlation of the dynamics of IDP segmental modes with solvent friction in viscous conditions[143] or by the increasing tendency of α-synuclein to aggregate if the solvent mobility is restrained[144]. As such, the accuracy of the water model in ensemble-based molecular dynamics (MD) simulation is crucial to reproduce dimensions of IDPs[145,146] as well as the segmental dynamics of IDP motional modes occurring from tens of picoseconds to tens of nanoseconds observed by NMR[147,148].

1.2.2 Dehydration

Throughout this thesis, the term dehydration is used to mean a critical loss of water content, as opposed to being thirsty, and desiccation signifies the highest level of dehydration.

Desiccation can result from different abiotic conditions, the most evident being atmospheric dehydration, which results from dry air or high temperatures that make moisture evaporate. Although cells are able to maintain homeostasis within a certain range of osmotic conditions using osmolytes, a highly hyperosmotic environment leads to the cellular water being drawn out of the cell and therefore dehydration, although there are some differences with atmospheric dehydration[149–153]. Another maybe counter-intuitive source of dehydration but common in nature is freezing. Indeed, freezing does not occur instantly and ubiquitously but instead, ice crystals have to form first and then grow by drawing liquid water molecules from their surroundings. Therefore, even if cells themselves do not freeze, the presence of extracellular ice crystals is enough to dehydrate them[154–156]. However, freezing-induced dehydration is not as extensive and therefore not as damaging as atmospheric dehydration because the non-freezable water remains liquid and available, a water which is lost during desiccation[154,155].

1.2.2.1 Cellular effects

Desiccation has many nefarious cellular effects which usually result in death. On the cellular scale, the loss of cellular water results in a loss of volume which creates a mechanical constraint upon the cell structure, in particular the plasma membrane as well

as the cell wall if present[153,157–160]. Additionally, this cell shrinkage also causes organelles, amongst them the mitochondria, to lose their shape and organisation[153,161,162]. Within the cell, the damages are even more numerous.

1.2.2.1.1 Damages to proteins

As previously emphasized, water plays a fundamental role in maintaining the structure, dynamics and functionality of proteins[85,138,148,163,164].
When dehydration occurs, the bulk cytoplasmic water is gradually removed until the hydration shell surrounding proteins starts to get severely altered[93,138,165,166], thereby disturbing the hydrogen-bond network with the solvent, thereby prompting changes of protein dynamics which can have perturbating effects on their functions[103,119,136] but can also trigger denaturation[160,165,167,168] and aggregation[103,160,165,168,169]. Additionally, the decrease in water content causes an increase of the crowding and of salt concentration, which can both result in additional denaturation and aggregation, through respectively perturbations induced by non-specific interactions[170–172] and through the salting-out effect[173,174].

These aberrant conformations and aggregates do not merely lead to mostly irreversible loss of protein functions, thus preventing them from accomplishing their role in cellular metabolism, but they may also be toxic themselves[175–177] and contribute to increasing cell crowding and viscosity even further. A chaotic vicious circle ensues, which further negatively impacts cell metabolism[178] by increasing the excluded volume inaccessible to metabolites[179,180], limiting their already hindered diffusion[180–182] and spatially segregating them[180,183–185].

Besides, as many proteins serve essential roles in shaping organelles and stabilising their structures[186], their loss of function further exacerbates the damages already inflicted by the mechanical strain upon these macromolecular structures[153,161,162].

1.2.2.1.2 Membrane damages

Membranes are also severely damaged by dehydration. When hydrated, phospholipid membranes form a lamellar bilayer in a liquid-crystalline phase characterized by high fluidity due to the spacing between the polar lipid headgroups[187,188]. The dehydration-induced loss of cell volume causes a lateral compression of the membrane, which results in demixing of membrane components due to different susceptibilities of different phospholipids, which in turn leads to local exclusions of membrane proteins and local enrichment in some lipids, resulting in irreversible damage[166,189,190]. The increased packing of lipids, in particular phosphatidylcholine, also causes an increase of van der Waals interactions between the acyl chains, leading to the transition from the lamellar liquid-crystalline phase to a gel phase[166,190–193] and therefore to an overall decrease of membrane fluidity due to the effective immobilisation of lipid chains, which translates to an increase of the permeability of the membrane[166,189,192–194].
The dehydration-induced lateral pressure may also trigger a local transition to the inverse hexagonal (HII) phase, in particular in areas enriched in phosphatidylethanolamine, where the bending stress on the membrane results in breaking of the bilayer into several cylinders, with the polar heads of the phospholipids pointing inside while their acyl chains get exposed to the outside[195–198].

These changes create irreversible damage upon rehydration, in particular if it is too fast for the cell to compensate. Since the loss of integrity results in the leakage of intracellular compounds, these damages culminate, unless the leakage is stopped, in cell death[158,190,199–204].

1.2.2.1.3 Oxidative damage

Desiccation-induced perturbation of the different metabolic pathways, notably within the intricate machinery of mitochondrial electron transport chains, can lead to increased production of reactive oxygen species (ROS). In turn, these ROS may yield significant damages to the cell by catalysing processes such as peroxidation[204–207] or de-esterification of membrane lipids[206–209], thereby further destabilising membranes, including mitochondrial membranes, which causes cell death[166,206,208,209]. Furthermore,

the deleterious impact of ROS also includes oxidative damage of proteins which can continuously occur even during the desiccated still-state[210–212], thus exacerbating cellular dysfunction even more and leading to death upon rehydration.

1.2.2.2 Practical aspects

1.2.2.2.1 Food preservation and storage

Desiccation has been used to preserve and store food as early as 12 000 BC as it effectively greatly reduces the speed of food spoilage due to the dehydration-induced suspension of biochemical reactions preventing bacteria and fungi infestation as well as processes for breakdown of food[213]. This dehydration was typically achieved by curing the food with salt and letting it dry in the open[213].

Due to these long-term storage properties as well as reduced weight and volume due to the reduced water content, desiccated food was also commonly used in manned space travel, with the food being dehydrated by freeze-drying and rehydrated upon consumption[214]. These characteristics also make dried food popular for hiking and trails.

1.2.2.2.2 Tissue preservation

Due to the suspension of biochemical reactions, desiccation of dead tissues effectively halts putrefaction and therefore preserves the aspect of dead tissues. This is one of the mechanisms that lead to mummification, along with freezing, anaerobic or acidic conditions[215].

This desiccation can spontaneously occur due to a naturally dry environment, or as the result of some voluntary water-removal processes. Examples of spontaneous human mummies include the Beauty of Loulan, one of the Tarim Basin mummies, or the Incans Children of Llullaillaco, which were preserved thanks to their exceedingly dry resting environments, respectively near the Lop Nur dried salt lake in the Lop Desert[216] and near the summit of the Llullaillaco volcano in the Atacama desert[215]. In turn, the ancient Egyptian mummies are the most famous example of artificial mummification, undergoing

a complex embalming process that includes desiccation by placing them in natron salt for 70 days[217–219].

Self-mummification of living persons is also possible and is a well-documented, albeit rare, practice in Asia amongst Buddhists and Shugendō monks. Most cases have been recorded in Japan where it is called Sokushinbutsu and is performed by merely eating pine needles, resin and bark, salts, roots and seeds for years while continuously decreasing the water intake to the minimum, effectively slowly dehydrating the body until death, after which the body does not decay due to its desiccation[220].

Beyond their values as anecdotes, those cases show that desiccation, by suspending biochemical reactions, has extended effects that include the preservation of biological tissues for centuries. However, in these cases, death of the organism precedes or is concomitant with desiccation and as such, rehydrating these biological tissues cannot revive the organism. Would that be possible if the desiccation occurred while the organism remains alive?

1.2.3 Anhydrobiosis

Some animals and plants, in particular those living in deserts, are adapted to live under constant water stress[221]. However, most of these survival strategies consist in limiting the loss of water for both animals and plants, while the latter may also be able to constantly harvest ambient water and store it[222]. It effectively *avoids* reaching a state of critical dehydration. Such organisms are commonly referred to as drought tolerant. However, another adaptation is also possible: *surviving* desiccation. This is called *anhydrobiosis*, i.e. life without water, and organisms able to achieve it are called desiccation tolerant[162,221].

First described in 1702 by the Dutch microbiologist van Leeuwenhoek who confessed “I never thought that there could be any living creature in a substance so dried as this was”[223] after observing what were most likely bdelloid rotifers waking up after having been desiccated[224], anhydrobiosis is the desiccation-induced category of *cryptobiosis*, i.e. hidden life, defined as “the state of an organism when it shows no visible signs of life and

when its metabolic activity becomes hardly measurable, or comes reversibly to a standstill"[225,226].
Indeed, in the anhydrobiotic state, these organisms lose almost all their water content, have their metabolism suspended and as such, there is barely any sign of them being alive[226,227]. However, unlike the desiccated yet dead state as in the above-mentioned mummies, anhydrobiosis is reversible by rehydration. For instance, bdelloid rotifers have been reanimated after 9 years of dormancy[228], while ancient lotus or date seeds have germinated after respectively 1300[229] and 2000 years of dormancy[230].

Remarkably, organisms in such a desiccation-induced dry state also commonly exhibit a significant cross-resistance to other environmental stresses, such as extreme pressure, temperature or radiation, indicating that the protective mechanisms necessary to attain an anhydrobiotic state also protect against damages induced by other type of stress[231–235].

1.2.3.1.1 Proposed mechanisms

1.2.3.1.1.1 Disaccharides

Several desiccation-resistant bacteria, yeast, fungi, plants and small animals have been shown to accumulate relatively high quantities of low-molecular weight soluble disaccharide sugars upon dehydration stress, in particular sucrose and trehalose[236–240]. For instance, the yeast *Saccharomyces cerevisiae* accumulates trehalose to up to 25% of the mass of the cell in response to stress[241,242], trehalose represents up to 7% of the dry mass of the spores of the slime mould *Dictyostelium mucoroides*[243], and the anhydrobiotic nematode *Aphelenchus avenae* synthesizes trehalose to up to 10% of its dry mass when slowly dehydrated[244].

Because of the clear correlation between desiccation stress and its accumulation, trehalose and other disaccharides were inferred to be cell protectants. This was supported by many in vitro studies about these protective effects (vide infra) and was for instance demonstrated in the case of larvae of the *Caenorhabditis elegans* nematode. These are strong anhydrobiotes that accumulate trehalose in response to drought, but

their mutants that are unable to synthesise trehalose do not survive even mild dehydration[245].

Mechanisms have been proposed to explain this protective effect. The first one is the vitrification hypothesis which states that these sugars, chiefly trehalose, vitrify upon desiccation, thereby forming a glassy matrix that fills the cell. In turn, this amorphous formation would be able to entrap all the contents of the cell as well as physically separating them and restricting their motions, resulting in suspension of biochemical reactions and mechanical prevention of both aggregation and denaturation, effectively amounting to suspension of cell metabolism in a still-state[239,246–248]. Indeed, trehalose is known to vitrify upon dehydration, and several studies have linked anhydrobiosis survival to the observation of a glass transition[249–251].
This model may appear paradoxical, as dehydrated proteins typically form an amorphous solid themselves upon dehydration[114]. However, the presence of a glassy matrix is not the sole requirement because, as discussed earlier, dehydration causes the denaturation and aggregation of proteins. Therefore, the presence of additional molecules is necessary to provide the proteins with spatial segregation and immobilisation[246].

The other hypothesis is the water replacement theory which states that trehalose, or any other protectant, would form hydrogen bonds with proteins when water is removed This would allow them to keep their conformation and prevent their aggregation by replacing the lost bonds with water[252,253]. This attractive model does not exclude the vitrification theory, as this matrix-protein hydrogen-bonding can happen in the glass state and has been initially supported by experimental Fourier-Transform Infrared spectroscopy (FT-IR) evidence[252,254]. However, recently, evidence has accumulated against this theory, favouring the alternative preferential hydration model[255,256].

This increasingly accepted preferential hydration theory proposes that, instead of forming hydrogen bonds with the protein directly, the protectant preserves or favours the water hydration layer needed by the proteins, which amounts to indirectly protecting them[257–259]. Indeed, water was observed to bond preferentially with protein rather than trehalose first using FT-IR[260]. This preference was later supported by neutron scattering[256] and MD simulation studies of multiple sugars that showed stabilising

effects of neighbouring proteins and highlighted the indirect nature of this mechanism[261–263].

This protective effect is stronger with trehalose than in other disaccharides and has long been attributed to an increased perturbation of the solvent[264–267]. However, NMR spin relaxation studies by the Halle group did not find evidence that trehalose perturbs water properties significantly more than other disaccharides[268] and therefore, the exact reasons for this increased desiccation protection conferred by trehalose remain disputed[269–272].

Moreover, trehalose and other disaccharides have also been demonstrated to protect membranes by reducing lateral compression in the membrane plane, which in turn prevents not only lipid demixing by reducing repulsion between membranes, but also avoids phase transitions by lowering the gel transition temperature to maintain the membrane in its liquid crystalline phase even at very low water content[191,273–277].

However, trehalose does not typically accumulate in high enough quantities in plants to protect them against dehydration stress on its own, highlighting the presence of another mechanism or protectant[240].

1.2.3.1.1.2 Late Embryogenesis Abundant proteins

First identified in 1981 by the work of Dure[278,279] as accumulating in cotton seeds facing dehydration stress, proteins belonging to the Late Embryogenesis Abundant (LEA) family are cytosolic and mitochondrial proteins found in high quantities in the last stages of plant seed development and which have been shown to protect them from the nefarious effects of dehydration[166,280]. Indeed, seeds cannot move in search of water and may have to endure months if not years of dehydration induced by drought, high salinity or freezing temperatures in the soil. It is therefore critical for them to be able to withstand desiccation.

Once thought to be exclusive to plants, and in particular to seeds, these LEA proteins have been identified over the years in bacteria[281,282] and fungi[283] but also in small animals like bdelloid rotifers[284,285], nematodes[286,287], brine shrimp cysts[288,289] or larvae of the

only known anhydrobiotic insect, the sleeping chironomid midge[290]. It is likely that over the next years, LEA proteins will be identified in even more organisms.

LEA proteins were shown to be upregulated in response to dehydration[280,287,291,292] but also to low temperatures[292,293] or oxidative stress[280,294]. In turn, they were demonstrated to have a protective effect on the activities of dehydration-sensitive enzymes after desiccation and rehydration[295–300] as well as generally preventing proteins from aggregating during dehydration[295,297,301–304]. They were proposed to act either as chaperones or as *molecular shields*, a mechanism in which they would crowd proteins and physically prevent them from aggregating [295,297,305,306], in an analogous way to the depletion stabilisation effect of polymers on colloids, where the presence of a low concentrated polymer results in local concentration gradient resulting from the depletion of the polymer between the colloids, thereby leading to a differential osmotic pressure that effectively repels the particles and therefore, inhibits their aggregation[305,307,308]. Moreover, LEA proteins also exhibit a stabilising effect on lipid membranes[300,309–311], a feature particularly observed for the many LEA proteins that localise around mitochondria and that are therefore commonly referred to as LEAM proteins[312–315].

Synergistic protective effects of LEA proteins have been observed with sugars, in particular trehalose and sucrose, and are thought to arise from the complementarity of the LEA proteins' own molecular shielding effects with the apparent mechanisms of trehalose of preferential hydration and glassy matrix formation. Trehalose appears to also interact with and be stabilised by LEA proteins [300,316], resulting in an increase of its glass transition temperature[316].

Depending on the classification system used, there are between 5 to 8 sub-groups of LEA proteins identified by different motifs[306,317]. Most of them are highly hydrophilic[318–320] and predicted to be disordered due to sequences rich in glycines, serines and alanines as well as depletion in hydrophobic residues, whose particular repetitive distribution led to predictions of formation of amphiphilic α helices and coiled-coils[318,321].

Experimental studies of LEA proteins using circular dichroism (CD), Fourier-Transform Infrared (FT-IR) or NMR spectroscopies confirmed their intrinsically disordered nature

and often showed a slight tendency to form α-helices in the hydrated state but also, as per CD and FT-IR, an increasing amount of secondary structure upon drying[293,295,309,312,316,322–329]. Specific examples include the formation of α-helices in a bulrush *Typha latifolia* pollen[316] and for the soybean *Glycine max* GmPM16[330] LEA proteins upon dehydration with sugars α-helices for the brine shrimp *Artemia franciscana* embryo AfrLEA6 protein[296] and the PvLEA4 protein from the sleeping midge *Polypedilum vanderplanki*[295], or α-helices and coiled-coils for the nematode *Aphelenchus avenae* AavLEA1 protein[324]. This latter protein was also investigated using MD simulations at different water content, which resulted in the protein progressively forming α-helices and β-sheets as the water content decreased, and eventually forming an intra-molecular coiled-coil at the lowest hydration levels[325].
It is however not clear to which extent these secondary structures formed upon dehydration are linked to their protective functions[305,306,331].

Finally, in 2005, the anhydrobiotic bdelloid rotifers were the first organisms reported to use LEA proteins for desiccation protection without simultaneously using disaccharide protectants[284,332], which would amount to a protection by intrinsically disordered proteins only. However, in 2015, putative trehalose synthase genes were identified and observed to be upregulated when rotifers enter desiccation, casting doubt on this claim[333]. Nevertheless, desiccation protection without disaccharides was reported soon afterwards in 2017 in a peculiar animal, the tardigrade.

1.3 Tardigrades

1.3.1 Biology and ecology

Tardigrades are microscopic metazoan animals with four pairs of legs, a variable number of claws and of size ranging from less than 0.5 mm to 1.2 mm[334,335]. They were first discovered in 1773 by the German zoologist Goeze who called them “small water bears”[336,337], due to their gait being somewhat similar to that of real bears’. They were later named tardigrades, i.e. slow walkers, by the Italian biologist Spallanzani in 1776 who noted their slow, tortoise-like walking motions[337,338].

Currently, more than 1300 known species are known in the Tardigrada phylum, which is separated in two main classes, *Eutardigrada* and *Heterotardigrada*[339,340]. Both have their body covered by a cuticle and have a stylet in the mouth that they use to pierce and feed on bacteria, algae and other small organisms[334,335,341]. There are, however, clear morphological differences between them with the most conspicuous being the often thicker and patterned cuticle with spikes of the highly diverse Heterotardigrades, which give them the appearance of having an armoured shell unlike the smoother Eutardigrades[335]. All tardigrade species mentioned from here on in this thesis belong to the Eutardigrades family which contains two orders: *Apochela* and *Parachela*[340].
A third class, *Mesotardigrada*, has also been described only once in 1937 in a hot spring[342] and is therefore nowadays considered to be unsubstantiated[343].

Tardigrades lack a proper respiratory and circulatory system[335,344]. Instead, their whole body cavity, which contains multiple free-floating storage cells, is filled with water which constantly diffuses in and out thanks to a highly permeable cuticle[345]. Tardigrades therefore rely on passive diffusion, facilitated by membrane-protein or by active transport of water, nutrients, electrolytes and gas exchange, specifically for the uptake of oxygen used for mitochondrial respiration[335].
As a consequence, tardigrades can only be active when their body is continuously surrounded by a film of water and therefore, either inhabit aquatic habitats or reside in close proximity to water sources to ensure an uninterrupted supply of water[345–347]. Due to the fluctuating availability of water, the latter tardigrade species are referred to as limnoterrestrial as they experience alternating periods of immersion and dehydration[345,348].

Aside from their distinctive appearance, *Heterotardigrada* and *Eutardigrada* species also often differ in terms of their ecological preferences. Heterotardigrades inhabit tidal, semi-terrestrial or marine habitats while Eutardigrades are predominantly semi- or limno-terrestrial. If most known tardigrade species inhabit limnoterrestrial freshwater habitats[347–349], many live in extreme conditions, with marine Heterotardigrades living in the deep-sea abysses[350,351] while Eutardigrades can be found in the deserts[352,353], Arctic[354] or Himalayan[355] cryoconite glacier holes or even Antarctica[356,357].

Tardigrades have historically been considered close to nematodes, which were also previously mentioned for their robust trehalose-mediated anhydrobiosis, on the basis for instance of whole-genome phylogeny[358], but this has been disputed as long branch attraction, which arises from distant lineages undergoing similar changes[345]. Indeed, the use of slow-evolving gene phylogeny, as well as the comparison of mitochondrial DNA or microRNA, indicated that the closest relatives of tardigrades are the velvet worms (*Onychophora*), which would make tardigrades panarthropods[340,359,360].
The oldest tardigrade fossil found is dated to 90 million years ago, and several were found dating from 60 to 80 million years ago[361]. However, estimation of a precise tardigrade apparition date is difficult as tardigrade fossils are relatively rare[334,362]. Nevertheless, the fact remains that tardigrades have existed for a long time and have survived numerous crisis and extinction events.

1.3.2 Extremotolerance

In 1776 already, Spallanzani noted the ability of tardigrade species found in his rooftop to survive desiccation and their apparent "revival after having died"[338]. Since then, limnoterrestrial tardigrades have been shown to alternate between active life and anhydrobiosis depending on water availability, whereas anhydrobiosis appears to be absent in purely marine tardigrades[349,363], apart from a relative resistance in some littoral species[349,364]. This supports the view of ancestral tardigrades as marine animals, with anhydrobiosis being necessary for their colonisation of terrestrial environments.

1.3.2.1 The tun state

When tardigrades face environmental stress, they enter a particular physiological state called the tun state in which their internal and external body structure is reorganised by muscular contraction, rearranging the internal organs and compacting their body by as much as 87% in some species, as the legs are withdrawn into the body[365]. During the formation of the tun, tardigrades rapidly lose most of their inner water over the course of a few minutes before displaying another phenomenon called the "permeability slump", which is the observed abrupt reduction of their transpiration, resulting in the end of this

fast water loss and therefore in the stabilisation of the remaining amount of internal water[366].

Overall, the water content of the tardigrade is significantly decreased, passing for instance from about 80% in the hydrated state to 1 to 3% in the dry state[367,368]. The reorganisation of the body into a tun as well as the permeability slump have been hypothesized to reduce the evaporation surface and rate of the tardigrade's cuticle in order to allow it to retain a necessary amount of water for survival[366].
Tun formation is found in both the *Eutardigrada* and *Heterotardigrada* classes and, contrary to earlier suggestions, tun formation is also found in marine species in response to changes in osmolarity[345,369,370], indicating that the tun state is a trait older than the division between these classes that precedes the apparition of anhydrobiosis. It is also noteworthy that forming a tun is a reaction initiated by the tardigrade in response to environmental stress and not an automatic reaction to it as anesthetised tardigrades do not form a tun but collapse into irregular and flattened forms they do not wake up from[367].

1.3.2.2 Anhydrobiotic abilities and stress resistance

Tardigrades in their dehydration-induced tun state show outstanding anhydrobiotic abilities. Indeed, they were shown to survive desiccation for as much as 20 years under controlled conditions[371].

Apart from this already incredible resistance to drought, tardigrades in the tun state also exhibit formidable cross-resistance to other kinds of stress. Indeed, while in anhydrobiosis, *Acutuncus antarcticus*, a tardigrade species closely related to *H.exemplaris* and *R.varieornatus*, has been shown to survive temperatures as cold as –30°C for 30 years[357], *H.exemplaris* and its close parent *Diphascon chilenense* survived –80°C for 6 years[372] and *Richtersius coronifer,* a more distant parent, even tolerated instant freezing in liquid nitrogen (–196°C) for up to 20 hours[373]. They are however more sensitive to high temperature[374] although many species are able to survive temperatures up to 80°C, with the very resistant *M.tardigradum* being barely affected by exposure up to 100°C[375].

This particular tardigrade species was also not only shown to survive immense hydrostatic pressure[376] of as much as 7.5 GPa[377], the equivalent of more than 74 000 times the standard atmospheric pressure, but also tolerated extremely low pressure in the form of space vacuum during a low Earth orbit flight[378]. In these conditions, *M.tardigradum* tuns were also simultaneously exposed to unfiltered cosmic and solar radiation and became the first reported animal species to survive the combined effect of these radiations and space vacuum, although the survival rate was low[378]. However, desiccated *M.tardigradum* had already been shown to survive gamma rays and heavy ion irradiation up to respectively 4400 and 5200 Gy[368], far more than the 8 Gy dose lethal for humans.
Unfortunately, it appears that tardigrades are not able to survive meteorite shocks, as desiccated *H.exemplaris* were only able to survive impact shocks up to 0.9 km s^{-1} or 3240 km h^{-1}, which amount to a shock pressure of 1.14 GPa[379].

However, not all anhydrobiotic tardigrades are equally resistant. While some species like *Milnesium tardigradum* or *Ramazzottius varieornatus* are known for their high tolerance to desiccation, other species like *Hypsibius exemplaris* are less resistant and can only survive slow desiccation. More than the tun state itself, the time necessary to display the permeability slump in the face of desiccation stress appears to be directly correlated with the survival ability of the different tardigrade species[366,380], although the exact link between these phenomena is currently not known.

As tardigrades are able to endure these stresses in a dormant state rather than their active state, and as this tolerance does not appear to be indefinitely sustainable, tardigrades are generally not considered to be true extremophiles but rather extremotolerants[381]. Nevertheless, some species also display stress resistance abilities already in their normal state, showing true extremophilic abilities. For instance, in the hydrated state, *Milnesium tardigradum* was shown to survive heavier doses of gamma rays and heavy ion irradiation than in the tun state[368], active *Macrobiotus richtersi* survived microgravity and cosmic rays in space[382], and some marine species show anoxybiotic abilities, i.e. survival in an oxygen-deprived environment, as tardigrades have been found in the known anoxic benthic zone of the Black Sea where the dissolved oxygen concentration is 10 times lower than in normal seawater[383].

This incredible resilience has made tardigrades both a subject of public curiosity and of research interest, since understanding it may allow to use this ability for instance to design crops more resistant to drought or to improve storage of food, medicines or even vaccines. Nevertheless, despite having been known for almost 250 years, the molecular mechanisms responsible for these abilities have hitherto remained elusive and have been the sources of many speculations.

1.3.2.3 Earlier hypotheses: Trehalose vitrification and LEA proteins, again?

A natural candidate due to its attested role in numerous plants and small organisms, the above-mentioned trehalose sugar was long sought in tardigrades, but despite years of research, no conclusive evidence was found for any involvement of trehalose in tardigrade anhydrobiosis. Indeed, although present in detectable quantity and at variable levels depending on the species – with a maximum in *Macrobiotidae* tardigrades where it can reach 3% of their dry weight, which is low in comparison with known trehalose-enriching anhydrobiotic species – , no enrichment of trehalose in the dehydrated state has been shown, nor has any evidence of direct involvement been found, even in species known to be particularly resistant to stress[384–387].

Some tardigrade species were reported to display an apparent glass transition as per DSC for *Macrobiotidae* tardigrades[250] and for *Hypsibius exemplaris*[388], which belongs to the *Hypsibiidae*. However, this claim is disputed as the observed transition may be due to residual water[389]. As species belonging to *Macrobiotidae* are the sole tardigrade species in which trehalose accumulation has been demonstrated so far[386,387], it is plausible that the apparent glass transition observed in these species may actually be that of trehalose, which would highlight the possibilities of different, and possibly complementary, survival strategies amongst the many species of tardigrades.

Classic LEA proteins, that is proteins presenting identified LEA motives, have also been found in tardigrades. For instance, *M.tardigradum* expresses at least one LEA protein in the hydrated state[390], and *R.varieornatus* embryos express a mitochondrial LEA protein[391], but evidence of involvement in anhydrobiosis remains scarce.

Tardigrades also possess Heat-Shock Proteins (Hsp). It was hypothesised that these chaperones would be involved in desiccation tolerance as a means to prevent protein aggregation or denaturation. However, overexpression of neither Hsp70 nor Hsp90 were shown to be triggered by desiccation in the tardigrade species *Paramacrobiotus richtersi*[392]. One isoform of Hsp70 appeared to be upregulated in dehydrating conditions in *Milnesium tardigradum*[393], while Hsp70 was observed to be expressed after stress, during the transition from the tun state to the active state in *Richtersius coronifer*[394]. These results seem to indicate a species-specific response and do not support a general role of Hsp proteins for tardigrade anhydrobiosis.

1.3.3 Tardigrade specific IDPs

1.3.3.1 TDPs

In 2012, while looking for tardigrade LEA proteins, researchers from the Kunieda group and collaborators identified two new families of heat soluble proteins in the resistant tardigrade species *Ramazzottius varieornatus*, the Cytoplasmic Abundant Heat Soluble (CAHS) and Secretory Abundant Heat Soluble (SAHS) families, which were shown to have similar properties to LEA proteins and also appeared to be intrinsically disordered[395]. They later identified a third family, the Mitochondrial Abundant Heat Soluble (MAHS) proteins as well as a mitochondrial tardigrade LEA protein, with both of them conferring a higher resistance to hyperosmotic shock, which mimics desiccation, upon expression in mammalian cells[391].

As previously discussed, *Hypsibius exemplaris* is a freshwater tardigrade and is one of the least resistant of the phylum, as it needs a slow drying to be able to survive desiccation, while for instance *Ramazzottius varieornatus* or *Milnesium tardigradum* can enter anhydrobiosis without any preconditioning. By comparing gene expression of *H.exemplaris* before and during slow drying, Boothby and collaborators were able to identify the proteins upregulated in response to the dehydration stress and therefore likely important to survive it. These proteins comprised the already discovered CAHS, SAHS and MAHS families and were collectively called the Tardigrade-specific Disordered Proteins or TDPs.

1.3.3.1.1 CAHS

The crucial importance of Cytoplasmic Abundant Heat Soluble proteins for tardigrade desiccation survival was demonstrated by Boothby and collaborators by knocking out individual expression using interfering RNA and observing significantly less survival[388]. In turn, the expression of individual CAHS proteins was shown to improve desiccation resistance in yeast cells[388], while mixing CAHS proteins with lactate dehydrogenase enzymes resulted in the protection of enzyme activity from irreversible inactivation due do dehydration, in an apparently more efficient way than by known protectors like trehalose[388].
In an essentially equivalent mechanism to trehalose, the proposed explanation is that CAHS protein vitrify upon dehydration, as an apparent glass transition was observed not only for the whole *H.exemplaris* tardigrade but also for dry CAHS protein alone[388]. However, no explanation has since been provided about the mechanisms by which vitrification would be achieved, and it is currently unclear if this transition is an artefact due to residual water[389]. Also, no other findings accumulated since then, in this thesis or elsewhere, support this hypothesis.

CAHS proteins appear to be neutral polyampholytes, with many charged residues but an overall neutral charge and without regions locally enriched in a particular residue type. They also have both hydrophilic and hydrophobic amino acids in non-negligible quantities, which has been inferred to allow CAHS proteins to adapt to all conditions. They were predicted from their discovery to be disordered with a helical propensity[395], which was supported by deuterium exchange experiments[388] and has been demonstrated experimentally for the first time in this thesis on a residue-by-residue basis[396].

The protein studied during this thesis, CAHS-8, belongs to *Hypsibius exemplaris*, a parachelan Eutardigrade previously known as *Hypsibius dujardini* but it recently become apparent that the modern *Hypsibius dujardini* was different from the historically described species and therefore had to be renamed . The protein was also called CAHS-94063 in its original discovery article[388] or CAHS D in later publications for unknown reasons.

1.3.3.1.2 MAHS

Mitochondria are critical for the survival of the cell, being involved in cellular energy production and elimination of ROS. Thus, protecting them is vital for an organism subjected to stress.
The Mitochondrial Abundant Heat Soluble (MAHS) protein was discovered by the Kunieda team and collaborators in *Ramazzottius varieornatus*[391]. The 25.4 kDa MAHS sequence comprises a mitochondrial signal motif. Mitochondrial localisation was demonstrated in human cells expressing fluorescently-labelled MAHS, which was also shown to improve the tolerance of these cells to desiccation-mimicking hyperosmotic shocks in comparison to native cells[391].

Tardigrade mitochondria have been shown to be smaller, with reduced cristae in the dehydrated state, and therefore MAHS have been inferred to be at least partly responsible for this reorganisation[398]. As previously discussed, mitochondrial membranes are significantly affected during desiccation stress, and therefore, the remodelling of the mitochondrial membrane may be responsible for or accompanied by a protection of said membrane.

Currently, not much is known about the MAHS protein, except that it has a homolog in *Hypsibius exemplaris*. Sequence prediction reports the presence of both IDRs and an amphipathic helix in the N-terminus, while the C-terminus may organise into a β-sheet with some IDRs.

1.3.3.1.3 SAHS

Secretory Abundant Heat Soluble proteins are typically found in the medium when expressed by cultivated cells. They appear to be mostly made of folded domains, may be more accurately described as folded proteins with some disordered segments. This is unlike what has been demonstrated in this thesis for a CAHS and what has been predicted for other CAHS and for the MAHS proteins.

This allowed for SAHS1 and SAHS4 of the tardigrade *Ramazzottius varieornatus* (Rv) to be crystallised by the Inoue team, revealing the canonical fatty acid binding protein (FABP) structure of an antiparallel β-barrel, with apparent binding domains located inside the barrel[399,400]. Using MD simulations, the Okumura team and collaborators predicted that the β-barrel of the RvSAHS1 protein is flexible, unlike human FABPs, and that the N-terminal region exchanges between a disordered state and an amphiphilic α-helix[401]. They later also proposed, on the basis of MD, that the β-barrel entrance properties are sensitive to the sequence of the substrate[402].

It is however currently unclear if SAHS have a protective role. However, they seem to be associated with the newly discovered special extracellular structure (SES) of tardigrades which appears to be exclusive to the tun state and is essentially akin to a plant cell wall[398]. They may however be implicated in establishment or stabilisation of the SES, which in turn may have a role in anhydrobiosis tolerance.

1.3.3.2 DSup

The damage suppressor protein (DSup) is a 42.8 kDa protein discovered in *Ramazzottius varieornatus* in 2016 by the Kunieda and Toyoda teams and collaborators[403], with a 33 kDa ortholog later identified in *Hypsibius exemplaris* as well[404,405]. Dsup is a highly positively charged protein, is predicted to be mainly disordered with a central helical domain, and comprises a nucleosome-binding site[403–405].

Ionizing radiation is harmful to life at high doses, since it breaks DNA and creates ROS, which in turn can create further damage[406]. DSup was demonstrated to protect human cells from these nefarious effects of irradiation by reducing damage to DNA by as much as 40%[403]. It was later demonstrated to also protect cells in which it was expressed from ROS-mediated damages[405,407,408].
It has therefore been proposed that DSup proteins bind to the nucleosome and physically shield it from damage, which is again similar to the proposed “molecular shield” mechanism of the LEA proteins as well as to the protection mechanism proposed in this thesis[405].

Although promising for DNA protection strategies, overexpression of DSup also shows some side effects. Indeed, most likely due to its positive charge, DSup was shown to bind non-specifically to DNA and RNA, which resulted in effects disturbing transcription by acting as a random transcriptional repressor[408].

1.3.3.3 LEA proteins

Discovered alongside the MAHS protein by the Kunieda team and collaborators in 2015 in the tardigrade species *Ramazzottius varieornatus* (Rv), the RvLEAM is a 30.5 kDa Late Embryogenesis Abundant Mitochondrial (LEAM) protein. It comprises multiple LEA-like motifs and shows sequence similarity with nematode LEA proteins[391]. It was shown to localise to mitochondria, is highly hydrophilic, and is predicted to be mostly disordered with a possible amphipathic helix in the regions containing the LEA motifs, indicating it most likely localises to the mitochondrial matrix rather than its membrane[391]. Like the MAHS protein, RvLEAM was shown to improve human cell tolerance to hyperosmotic shocks which mimic the effect of dehydration, although the mechanism by which this protection is achieved is currently not known[391].

The 25.5 kDa protein previously identified in 2017 as CAHS-7 of *Hypsibius exemplaris* (He) due to its cellular location[388] bears little sequence similarity to its other CAHS and has later been identified as homologous to RvLEAM and subsequently renamed HeLEA1 in 2024 by Li and collaborators[409]. They demonstrated that HeLEA1 is disordered under normal conditions, that it localises in the mitochondrial matrix and that it transitions to a helical conformation upon binding negatively charged membranes. In hyperosmotic stress conditions, the protein was shown to dissipate excess lateral tension in the membrane and to limit lipid packing defects, increasing the overall stability of the membrane in stress conditions, which translated to an enhanced tolerance to it in yeast expressing HeLEA1[409].

It is probable that RvLEAM also works in a similar way and that it has a synergistic effect with the MAHS protein.

1.3.3.4 Beyond the current nomenclature

The localisation-based nomenclature CAHS, MAHS and SAHS is probably not fully representative of the reality of the function of these proteins. Indeed, advances in prediction and a growing number of experimental evidence show that MAHS proteins are organised around a β-barrel structure[399,400] with some of them showing very little disorder apart from a few residues in the N-terminus and the residues between the β-strands. This contrasts with the clearly disordered nature of most other known TDPs and raises the question whether the SAHS family ought to remain included in the family of tardigrade-specific *disordered* proteins (TDP).

In the coming years, a better nomenclature will most probably emerge from a better understanding of the relationship between the CAHS, SAHS, MAHS, LEA and DSup families. It is noteworthy that multiple CAHS, SAHS, one putative MAHS and one putative LEAM genes were identified in multiple *Parachela* Eutardigrades, but only CAHS genes were identified in *Apochela* Eutardigrades and none of the CAHS, SAHS, MAHS and LEAM families were ever found in Heterotardigrades[410]. This supports the hypothesis of an apparition of CAHS proteins first after the split between Eu-and Heterotardigrades, while the SAHS, MAHS and LEAM families appear to be later parachelan adaptation .
It is also not improbable that more tardigrade-unique IDP families will be discovered, for instance in the newly discovered special extracellular structure or the understudied storage cells.

For the moment, for lack of a better alternative and for the purpose of clarity, the current established nomenclature will be used in this thesis.

Chapter 2: Theoretical background

"Theory attracts practice as the magnet attracts iron."
attributed to Carl Friedrich Gauss (1777-1855)[411].

1.4 Nuclear Magnetic Resonance

1.4.1 Of Spin

Atomic nuclei have intrinsic properties. One of them, the nuclear spin, can be approached as a quantum mechanical form of angular momentum and is of prime importance for NMR, as nuclei with non-zero spin also possess an intrinsic magnetic moment. This nuclear spin number I depends on the nucleon composition of the nucleus, i.e. its number of neutrons and protons, and therefore varies between isotopes of the same element. It is represented by a spin operator $\hat{\mathbf{I}}$ which is a vector with a quantized norm denoted $|\mathbf{I}|$. Using the Dirac notation, the squared norm of $\hat{\mathbf{I}}$ is given by[412–414]

$$\left|\hat{\mathbf{I}}^2\right| = \langle\psi|\hat{\mathbf{I}}^2|\psi\rangle = \hbar^2 \mathrm{I}(\mathrm{I}+1) \qquad 0.1$$

with ħ being the reduced Planck constant, $|\psi\rangle$ representing the state vector of the system, $\langle\psi|$ its complex conjugate and I the nuclear spin number. In turn, the magnitude of the angular momentum vector $\mathbf{I}$ is

$$|\mathbf{I}| = \hbar\sqrt{I(I+1)}\,. \qquad 0.2$$

The spatial components of $\hat{\mathbf{I}}$ can be arbitrarily chosen so its z axis is aligned with the z-axis in a fixed Cartesian frame commonly referred to as the *laboratory frame*. The value of angular momentum along the z-axis is

$$\mathrm{I}_z = m\hbar\,. \qquad 0.3$$

where m is the magnetic quantum number of spin and is defined as a discrete set of values varying between $-\mathrm{I}$ and $+\mathrm{I}$ in steps of 1 such as

$$m = \{-\mathrm{I}, -\mathrm{I}+1, \ldots, \mathrm{I}-1, \mathrm{I}\}\,.$$

This means that each of the $2\mathrm{I}+1$ possible values of m correspond to a different nuclear spin state whose associated spin wavefunctions are eigenfunctions of both $\hat{\mathbf{I}}^2$ and $\hat{\mathbf{I}}_z$ with eigenvalues of respectively ħI(I+1) and ħm that define the orthonormal basis of the associated Hilbert space. Finally, since the quarks making up the protons and neutrons composing the nuclei are electrically charged, the spin angular momentum generates a collinear magnetic dipole moment $\boldsymbol{\mu}$ proportional to the magnitude of the angular momentum $\mathbf{I}$ as

$$\boldsymbol{\mu} = \gamma\mathbf{I} \qquad 0.4$$

with its z-axis component being

$$\mu_z = \gamma\hbar m\,. \qquad 0.5$$

The gyromagnetic ratio γ depends on each nucleus and as indicated by its name, is a proportionality constant that relates the angular and magnetic momentum of the nucleus as the ratio of the two. It is therefore a magnetic moment per angular momentum, hence its expression in $rad \cdot s^{-1} \cdot T^{-1}$ in SI units.

Consider a nucleus with a spin I = $\frac{1}{2}$. These equations mean there is $2 \times \frac{1}{2} + 1 = 2$ possible eigenstates with possible values of the magnetic quantum number $m = +\frac{1}{2}$ and $m = -\frac{1}{2}$. which can be written in the Dirac notation as the ket-vector $\left|+\frac{1}{2}\right\rangle$ and $\left|-\frac{1}{2}\right\rangle$. For a spin with I = 1, there are 3 possible states of values m = +1, m = 0 and m = −1 and so on.
These states have different properties which gives rise to observable effects, notably the Zeeman effect.

1.4.2 Classical view of NMR

1.4.2.1 Zeeman effect

The most common case in biomolecular NMR is a large population of spins I = $\frac{1}{2}$ like ^{1}H, ^{13}C or ^{15}N. The bulk, or net, magnetisation vector **M** of this population of N spins i is simply the sum of all their individual dipole magnetic moments **μ**, that is[412,413,415]

$$\mathbf{M} = \sum_{i=1}^{N} \boldsymbol{\mu}_i = \sum_{i=1}^{N} \gamma \mathbf{I}_i \qquad 0.6$$

with $\mathbf{M} = (M_x, M_y, M_z)^T$. In the absence of any exterior magnetic field, the energy levels E associated with the magnetic quantum number m are degenerate, the spins are randomly oriented and thus $\mathbf{M} = 0$. Of course, this cannot be the case on Earth due to the effect Earth's own magnetic field of approximately 50 μT[416].

The state vector of the superposed and non-interacting spins is given by $|\Psi\rangle = a\left|-\frac{1}{2}\right\rangle + b\left|+\frac{1}{2}\right\rangle$ with a and b the complex coefficients representing the probability of finding these eigenstates, such that $|a|^2 + |b|^2 = 1$. Upon experiencing a static magnetic field $\mathbf{B}_0$, whose direction is arbitrarily chosen as the z-axis, the magnetic moments of the spins interact

with it and the degeneracy of their eigenstates is lifted, with their associated energies E given by

$$\mathrm{E} = -\mu_z \mathbf{B_0} = -\gamma \hbar m \mathbf{B_0} \,. \qquad 0.7$$

This is the nuclear *Zeeman effect*. As their energy level E depends on m, it is apparent that the two eigenstates do not have the same energy level. Indeed, an orientation of the nuclear magnetic moment in parallel to the magnetic field requires less energy than an antiparallel orientation (for $\gamma > 0$), hence the negative sign, and therefore the spins have a slight preference for the eigenstate with the lower energy level. The difference in the energy levels ΔE between those two states is given by

$$\Delta \mathrm{E} = +\frac{1}{2}\hbar\gamma \mathbf{B}_0 - \left(-\frac{1}{2}\hbar\gamma \mathbf{B_0}\right) = \hbar\gamma \mathbf{B}_0 \qquad 0.8$$

This means that the magnitude of the energy splitting depends directly on the strength of $\mathbf{B_0}$ and on the gyromagnetic ratio of the observed nuclei. The higher they are, the bigger the Zeeman splitting between the energy levels. As two spin states have different energy levels, they also have different relative populations and at thermal equilibrium, the relative ratio $\frac{\mathrm{N}_+}{\mathrm{N}_-}$ in their occupation follows a Boltzmann distribution as

$$\frac{\mathrm{N}_+}{\mathrm{N}_-} = \exp(-\Delta \mathrm{E}/k_B T) \qquad 0.9$$

with k_B the Boltzmann constant and T the temperature. Substituting the Zeeman equation 0.8 in this equation yields

$$\frac{\mathrm{N}_+}{\mathrm{N}_-} = \exp(\hbar\gamma \mathrm{B}_0/k_B T) \qquad 0.10$$

Consider a population of ^{1}H atoms experiencing a magnetic field of 14.07 T at 293 K. The ratio of the two populations can be calculated using the previous equation and is equal to 0.9999 while it is 0.9998 at 22.3 T due to its proportionality on $\mathbf{B}_0$. This small difference in the populations of the two spin states results in a *polarisation*, i.e. in the net magnetisation moment **M** detected by NMR. This is the reason why NMR spectroscopy is considered as not very sensitive, hence the need for higher magnetic fields.

It is important to note that at usual temperatures employed in NMR spectroscopy, the energies of the states defined in eq. 0.7 are much smaller than the surrounding thermal energy[412], i.e. $|\mathrm{E}| << k_B T$. This condition is known as the high-temperature limit and has important consequences that will be discussed later.

1.4.2.2 Larmor precession

The Planck equation relates the energy difference ΔE between quantum energy levels to the frequency υ of the corresponding electromagnetic radiation through the Planck constant h as[412]

$$\Delta E = h\upsilon \qquad 0.11$$

Substituting the Planck equation into the Zeeman equation 0.8 yields the Larmor equation, with the minus sign being a convention:

$$\upsilon = -\frac{\gamma \mathbf{B_0}}{2\pi} \qquad 0.12$$

This frequency υ is called the Larmor frequency, is expressed in Hz or s^{-1} and depends on the gyromagnetic ratio γ which is different for each nucleus.

Under a static magnetic field, the time derivative of the angular momentum **I** of one spin is the cross-product of its magnetic moment **μ** and $\mathbf{B}_0$[412,415,417], i.e.

$$\frac{d\mathbf{I}}{dt} = \boldsymbol{\mu} \times \mathbf{B}_0 = \boldsymbol{\tau} \qquad 0.13$$

which is equal to the torque **τ** experienced by the magnetic moment **μ** of the spin under the magnetic field $\mathbf{B_0}$. As implied by the cross-product, this torque **τ** can be seen as a perpendicular force seeking to align **μ** with $\mathbf{B_0}$. Substituting eq. 0.4 yields the first order differential equation

$$\frac{d\boldsymbol{\mu}}{dt} = \gamma \boldsymbol{\mu} \times \mathbf{B}_0 = \gamma \mu B_0 \sin(\theta) \qquad 0.14$$

with θ the angle between **μ** and $\mathbf{B_0}$ in the plane they define**.** As implied by this equation of motion, the magnitude μ of **μ** doesn't change over time, i.e. $d\mu/dt = 0$, therefore it is the sole direction of **μ** that is changing over time perpendicularly to the plane define by **μ** and $\mathbf{B_0}$. In other terms, as the magnetisation in the z-axis $\mathbf{I_z}$ is always smaller than the norm $|\hat{\mathbf{I}}|$ of the magnetisation vector, the torque causes the magnetic moment **μ** to rotate or *precess* around $\mathbf{B_0}$ at an angle θ, effectively drawing two cones around the z-axis and subtending an azimuthal angle ϕ in the x-y plane such as:

$$\frac{d\boldsymbol{\mu}}{d\phi} = \mu \sin(\theta)\ .$$

Putting those two equations together allows to express the subtended angle per time unit and thus, gives the angular velocity ω_0, that is the angular Larmor precession:

$$\frac{d\phi}{dt} = \gamma B_0 = -\omega_0$$

Bringing back the Larmor equation 0.12, the angular Larmor frequency ω_0 is therefore defined as

$$\omega_0 = -\gamma B_0 = 2\pi\upsilon \qquad 0.15$$

with ω in rad·s^{-1} and with the negative sign indicating a left-handed rotation around the z-axis due to ϕ being defined by convention in a right-handed fashion.

To return to our previous example, a population of ^{1}H atoms experiencing a static magnetic field of 14.09 T and with no interactions between the spins nor with their environment will theoretically all have the same angular Larmor frequency of 600 000 000 rotations per second or 600 MHz. Indeed, the commonly used radio-frequency (RF) to identify the NMR spectrometer instrument corresponds to the ^{1}H Larmor frequency instead of their magnetic field strength in Tesla. However, nuclear spins typically interact one with another or, due to Brownian motions, experience fluctuating smaller local magnetic fields from the electronic clouds of other atoms of the studied species, of water, of salts or of other molecules in their surroundings, which is historically called the *lattice*. This effectively slightly alters the Larmor precession frequency of each spin. Consequently, any inhomogeneities in $\mathbf{B}_0$ yield the same result.

1.4.2.3 Bloch equations

From the bulk magnetisation equation 0.6 and the torque equation 0.13, the time evolution of the net magnetisation is[415,417]

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma \mathbf{B}_0(t) \qquad 0.16$$

where $\mathbf{M}(t)$ is the bulk magnetization vector, γ is the gyromagnetic ratio of the observed nuclei and $\mathbf{B}_0(t)$ is the static magnetic field.

As previously described, the Zeeman effect gives rise to a net magnetisation moment along the z-axis defined by the direction of $\mathbf{B}_0$ while there is no magnetisation in the

transverse plane. As stated by the Planck equation 0.11, the Larmor frequency υ corresponds to the required frequency of an electromagnetic radiation to induce the transitions between the energy states i.e. to fulfil the *resonance* condition. This Larmor frequency typically lies in the radio-frequency (RF) range, and thus by applying a RF pulse at a frequency close to υ, it is possible to manipulate the net magnetisation around $\mathbf{B}_0$. Typically, an oscillating RF pulse at the Larmor frequency ω_0 is applied transversally for an arbitrary duration and at an arbitrary phase, resulting in a $\mathbf{B}_1$ magnetic field perpendicular to $\mathbf{B}_0$that can rotate the net magnetisation to any point of the Cartesian frame, i.e. at any angle – depending on duration and amplitude of the pulse – and on any axis – depending on the phase of the pulse. For instance, a particular, well-calibrated duration can trigger a 90° rotation of the net magnetisation onto the transverse plane. This effectively yields a *coherent*, in-phase precession of the spins around the z-axis, giving rise to a detectable signal in the x-y plane.

From Eq. 0.16, it would appear that the net magnetisation moment's magnitude and direction would remain the same forever. However, experiments show this is not the case and the thermal equilibrium is eventually restored after a finite period of time, i.e. spins *relax*.

1.4.2.3.1 Longitudinal and transverse relaxation

In order to phenomenologically explain the decay of the magnetisation, Bloch introduced in 1946 two relaxation mechanisms in his vector model[417]. The first mechanism describes the return of the z-component of the net magnetisation to equilibrium, i.e. the return of the distribution of the Zeeman effect-induced differences in the spin populations to the Boltzmann thermal equilibrium. This longitudinal relaxation can be described as[412,415,417]

$$\frac{d\mathrm{M}_z(t)}{dt} = \mathrm{R}_1[\mathrm{M}_0(t) - \mathrm{M}_z(t)] \qquad 0.17$$

with R_1 the longitudinal relaxation rate defined as the inverse of the longitudinal relaxation time T_1, and M_0 the magnitude of the z-component of the magnetisation at equilibrium. This equation describes a first-order process and therefore means that the magnetisation eventually returns to equilibrium in an exponential manner at a rate R_1. As

previously described, spins interact with their environment – the *lattice* or *heat bath*, which creates fluctuating local magnetic fields – through Brownian-motion induced collisions resulting in translation, rotations and direct interactions. This results in the return to the Boltzmann equilibrium due to the transfer of energy from the polarised spins to their surroundings in a lower energy state. Hence the historical, albeit somewhat misleading, name of *spin-lattice* relaxation for the longitudinal relaxation process.

A second mechanism is another first-order rate process called transverse relaxation which describes the exponential dephasing of the magnetisation in the x-y plane as

$$\begin{aligned} \frac{\mathrm{d}\mathrm{M}_x(t)}{\mathrm{d}t} &= -\mathrm{R}_2\mathrm{M}_x(t) \\ \frac{\mathrm{d}\mathrm{M}_y(t)}{\mathrm{d}t} &= -\mathrm{R}_2\mathrm{M}_y(t) \end{aligned} \qquad 0.18$$

with R_2 the transverse relaxation rate as $\mathrm{R}_2 = 1/\mathrm{T}_2$ with T_2 the transverse relaxation time. Due to the aforementioned local magnetic field perturbations or $\mathbf{B}_0$ inhomogeneities, the Larmor frequency of the spins slightly varies, which over time results in dephasing of the coherence of the spins in the x-y place and thus, in the exponential decay of the magnetisation in the transverse plane. Moreover, interactions between the spins themselves also result in a loss of coherence, hence the historical name of *spin-spin* relaxation for the transverse relaxation.

Combining eq. 0.16 with eq. 0.17 and 0.18 yields the phenomenological Bloch equations which describe the time-evolution of the three components of the net magnetisation in the absence of an RF pulse while taking into account the relaxation mechanisms:

$$\begin{aligned} \frac{\mathrm{d}\mathrm{M}_z(t)}{\mathrm{d}t} &= \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_z - \mathrm{R}_1[\mathrm{M}_0(t) - \mathrm{M}_z(t)\,] \\ \frac{\mathrm{d}\mathrm{M}_x(t)}{\mathrm{d}t} &= \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_x - \mathrm{R}_2\mathrm{M}_x(t) \\ \frac{\mathrm{d}\mathrm{M}_y(t)}{\mathrm{d}t} &= \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_y - \mathrm{R}_2\mathrm{M}_y(t)\,. \end{aligned} \qquad 0.19$$

These classical Bloch equations can describe only isolated, non-interacting spins. As such, when concerned only with the evolution of the bulk magnetisation, they provide an accurate view and were later expanded to account for chemical-exchange as the Bloch-McConnell equations[418] (section 1.4.6.1.1 page 88) or translational diffusion as the

Bloch-Torrey equations[419]. Although complemented notably by the Solomon equations to describe the dipole-dipole induced relaxation between two spins[420], it is more convenient and accurate to use a quantum mechanical approach to describe relaxation due to the quantum mechanical nature of spins and the different interactions they can experience.

1.4.3 Quantum view of NMR

1.4.3.1 Mixed states and density operator

The evolution of a quantum system is governed by the time-dependent Schrödinger equation[412,413,421,414]

$$\frac{\mathrm{d}}{\mathrm{d}t}|\Psi(t)\rangle = -\mathrm{i}\widehat{\mathcal{H}}(t)|\Psi(t)\rangle \qquad 0.20$$

With $\widehat{\mathcal{H}}$ the Hamilton operator describing the system in the natural units of angular frequency that will be used from now on – i.e. $\widehat{\mathcal{H}} = \hbar^{-1}\widehat{H}$ with $\widehat{H}$ the classical Hamiltonian in units of energy – and $|\Psi\rangle$ the state vector of the system in Dirac notation.

To continue with our example of N spins with I = $\frac{1}{2}$, the state vector $|\Psi\rangle$ of a defined 2^N dimensional Hilbert space can be decomposed as a linear combination of the 2^N eigenstates $|\psi_i\rangle$ that form its orthonormal basis such as

$$|\Psi(t)\rangle = \sum_{i=1}^{2^N} c_i(t)|\psi_i\rangle \qquad 0.21$$

with $c_i(t)$ the complex coefficients describing the time-dependent probability amplitude associated with the 2^N eigenstates of the system. This assumes that the Hilbert space is both finite and discrete.

However, as previously described, the different spin systems interact with each other as well as with their environment. In consequence, they are in a mixed state and therefore, it is more accurate and convenient to use a probability density P(Ψ) describing the statistical weight of each specific pure state $|\Psi\rangle$ to the ensemble. Due to the orthonormality of the basis eigenfunctions, the ensemble averaged expectation value of

any physical observable B represented by the Hermitian operator $\widehat{\mathrm{B}}$ within a spin ensemble with two eigenstates $|\psi_i\rangle$ and $|\psi_j\rangle$ would be

$$\begin{aligned}\langle B\rangle = \langle\Psi|\widehat{\mathrm{B}}|\Psi\rangle &= \int P(\Psi)\langle\Psi|\widehat{\mathrm{B}}|\Psi\rangle dt \\ &= \sum_{i,j}\int P(\Psi)c_i c_j^* dt(\Psi)\langle j|\widehat{\mathrm{B}}|\mathrm{i}\rangle = \sum_{i,j}\overline{c_i c_j^*}\langle \mathrm{j}\widehat{\mathrm{B}}|\mathrm{i}\rangle\end{aligned} \qquad 0.22$$

that is, a sum of eigenvalues of B weighted by the ensemble-average of the coefficients $c_i c_j^*$, hence the overbar. They define the density operator $\hat{\sigma}$ such as

$$\hat{\sigma}_{ij} = \overline{c_\imath c_j^*} = \langle\psi_i|\Psi\rangle\langle\Psi|\psi_j\rangle \qquad 0.23$$

which can be expressed in matrix form in the eigenbasis of $\mathcal{H}$, yielding

$$\hat{\sigma}_{ij} = \begin{pmatrix} c_i c_i^* & c_i c_j^* \\ c_i{}^* c_j & c_j c_j^* \end{pmatrix}. \qquad 0.24$$

Following the Born rule, the diagonal elements of this density matrix represent the population weighted probability of finding the system in each of the particular eigenstate within the spin system – i.e. their probability density – while its off-diagonal elements represent the coherences from the superposition of the two eigenstates which, if not null, implies a transverse magnetisation in the context of NMR.

As NMR processes are typically time-dependent, it is useful to have the density operator $\hat{\sigma}$ being time dependent. A general form of $\hat{\sigma}(t)$ is

$$\hat{\sigma}(t) = \int P(\Psi)|\Psi\rangle\langle\Psi|dt\,. \qquad 0.25$$

The density operator is a powerful tool. Indeed, the average expectation value of any arbitrary observable B represented by an operator $\widehat{\mathrm{B}}$ can be obtained from the trace of the product of B with the density operator $\hat{\sigma}$ as:

$$\langle\overline{B}\rangle = \mathrm{Tr}\big(\hat{\sigma}(t)\hat{B}\big)$$

Thus, it is therefore possible to describe the evolution of the whole spin system by describing the evolution of the density matrix. In particular, the time evolution of the density operator follows the Liouville-von-Neumann equation

$$\frac{d\sigma(t)}{dt} = -\mathrm{i}[\mathcal{H}, \sigma(t)] = \hat{L}\sigma(t) \qquad 0.26$$

with $\hat{L}$ the commutation super operator or Liouvillian $\hat{L}$ defined as $\hat{L}(t) = [\mathcal{H}(t),\cdot]$ which is convenient to use in order to describe the time evolution of the density operator.

This Liouville-von-Neumann equation is of fundamental importance in NMR. It means that, provided that the Hamiltonian $\mathcal{H}$ of the system is known, the time evolution of any observable element can be described.

The associated $(2^N)^2$ dimensional Liouville space basis is defined by the Cartesian Hermitian spin operators I_x, I_y and I_z representing x, y and z components of the spin angular momentum of each spin (the Pauli matrices), as well as the identity operator $\mathbb{I}$ as:

$$I_x = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix};\ I_y = \frac{1}{2}\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix};\ I_z = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix};\ \mathbb{I} = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \qquad 0.27$$

and with the extension to higher dimensions for more than one spin leads to product operators via the direct product of these individual spin operators. Another possible basis set are the shift operators defined as:

$$I^+ = I_x + iI_y = \frac{1}{2}(I_x + I_y);\ I^- = I_x - iI_y = \frac{1}{2i}(I_x - I_y)$$

together with I_z and $\mathbb{I}$ that are the same as in the Cartesian basis. This shift basis is particularly useful to study the changes in coherences during pulse sequences as the operators I^+ and I^- being referred to at the "raising" and "lowering" operators due to their effect on the order of the coherences.

As an example, for two spins-$\frac{1}{2}$ I and S, the basis of the Liouville space would be the product operators given in the Table 0.1 for the two sets of bases.

Table 0.1: Liouville space operator basis

Name	Cartesian operator basis	Shift operator basis
Longitudinal magnetisation	I_z, S_z	I_z, S_z
In-phase transverse magnetisation	I_x, I_y, S_x, S_y	I^+, I^-, S^+, S^-
Anti-phase Transverse magnetisation of spin I	$2I_xS_z, 2I_yS_z$	$2I^+S_z, 2I^-S_z$
Anti-phase Transverse magnetisation of spin S	$2I_zS_x, 2I_zS_y$	$2I_zS^+, 2I_zS^-$
Coherence of the two spins	$2I_xS_x, 2I_xS_y, 2I_yS_x, 2I_yS_y$	$2I^+S^+, 2I^-S^+, 2I^-S^+, 2I^-S^-$
Longitudinal two spin order	$2I_zS_z$	$2I_zS_z$
Unity operator	$\mathbb{I}$	$\mathbb{I}$

Unless stated otherwise, Liouville space will be used from now on.

1.4.3.2 Bloch-Wangsness-Redfield theory

The following section presents the NMR quantum theory developed from the work of Bloch[422], Wangsness[423] and Redfield[424,425] (BWR) that came to bear their names.

1.4.3.2.1 The Master equation

In the BWR theory of relaxation, the evolution of the density operator is estimated using perturbation theory[412,415,421,424–426]. For that, the Hamiltonian $\mathcal{H}$ of the system is decomposed in two term such as

$$\mathcal{H}(t) = \mathcal{H}_0 + \mathcal{H}_1(t) \qquad 0.28$$

with $\mathcal{H}_0$ the known time-independent Hamiltonian describing the spin system and $\mathcal{H}_1$ the time-dependent stochastic Hamiltonian describing the contributions of the random fluctuations of the system. It is to be stressed that despite the energy exchange between the spin systems and the lattice, the latter's temperature is assumed to not be altered by this exchange and to remain in statistical equilibrium.

The time independence of $\mathcal{H}_0$ implies the absence of any RF field and therefore arises mainly from the known Zeeman interaction with the $\mathbf{B}_0$ field while experiencing perturbations which are ensemble averaged to 0, i.e. $\overline{\mathcal{H}_1(t)} = 0$. Any time-dependent fluctuations that are not ensemble averaged to 0 are considered static and included in the static $\mathcal{H}_0$. Using the Liouville-von-Neumann equation, the corresponding time evolution of the density operator of this system becomes:

$$\frac{d\sigma(t)}{dt} = -i[\mathcal{H}_0 + \mathcal{H}_1(t), \sigma(t)] = -i\{\hat{L}_0 + \hat{L}_1(t)\}\sigma(t)\,. \qquad 0.29$$

As we are mainly interested in the stochastic part of the Hamiltonian $\mathcal{H}_1(t)$, it is useful to remove the dependence on the time-independent Zeeman interaction with the static magnetic field by transforming $\mathcal{H}_1(t)$ into another reference frame commonly called the *interaction frame* such as

$$\mathcal{H}_1{}^T(t) = e^{i\mathcal{H}_0 t}\,\mathcal{H}_1(t)\; e^{-i\mathcal{H}_0 t} = \hat{L}_0\mathcal{H}_1(t)$$

as well as the density operator $\sigma(t)$

$$\sigma^T(t) = e^{i\mathcal{H}_0 t}\,\sigma(t)\,e^{-i\mathcal{H}_0 t} = e^{+i\hat{L}_0 t}\,\sigma(t)$$

from which the Liouville-von-Neumann equation can be rewritten in this interaction frame as:

$$\frac{d\sigma^T(t)}{dt} = -i\hat{L}_1^T(t)\sigma^T(t) = -i[\mathcal{H}_1^T(t), \sigma^T(t)]\,. \qquad 0.30$$

This reflects the effects of the fluctuating perturbations in the interaction frame described by the stochastic Hamiltonian ${\mathcal{H}_1}^T$ on the system's evolution represented by $\sigma^T(t)$ from a known starting point $\sigma^T(t_0)$. As the local the local density matrices of remote part of the system can relax differently due to different random couplings with their surroundings, it is convenient to use an ensemble-averaged, denoted by the overbar, density matrix $\overline{\sigma^T(t)}$. Moreover, as these stochastic perturbations occur at much faster timescale than the evolution of the density matrix in the interaction frame, they can be considered to be statistically independent and therefore separated. Thus, approximate solution to eq. 0.30 can be found using perturbation theory where the contributions of the perturbations are approximated as the sum of increasingly small nth order corrections, usually up to the second, yielding the second order approximation

$$\begin{aligned}\overline{\sigma^T(t)} \approx\ & \overline{\sigma^T(t_0)} \\ & - i\int_{t_0}^{t}\left[\overline{{\mathcal{H}_1}^T(t_2), \sigma^T(t_0)}\right]dt_2 \\ & -\frac{1}{2}\int_{t_0}^{t}\int_{t_0}^{t}\left[\overline{{\mathcal{H}_1}^T(t_1), \left[{\mathcal{H}_1}^T(t_2), \sigma^T(t_0)\right]}\right]dt_1 dt_2\end{aligned} \qquad 0.31$$

whose time-derivation with respect to $\tau = t - t_1$ is

$$\frac{d\sigma^T(t)}{dt} = -i\left[{\mathcal{H}_1}^T(t_1), \sigma^T(0)\right] - \int_0^t \overline{\left[{\mathcal{H}_1}^T(t), \left[{\mathcal{H}_1}^T(t-\tau), \sigma^T(t) - \sigma_0\right]\right]}\,d\tau\,. \qquad 0.32$$

As stated earlier, we assume that $t \gg \tau_c$, i.e. the studied timescale t is significantly longer than the correlation time τ_c of the stochastic process described by the Hamiltonian ${\mathcal{H}_1}^T$, whose ensemble averaged value is null, i.e. $\overline{{\mathcal{H}_1}^T} = 0$ as any fluctuation that doesn't satisfy this is included in the static $\mathcal{H}_0$. This paves the way for the second assumption that ${\mathcal{H}_1}^T$ and $\sigma^T(t)$ are uncorrelated in this timescale and therefore can be treated separately, thus $\overline{\left[{\mathcal{H}_1}^T(t_1), \sigma^T(0)\right]} = \left[\overline{{\mathcal{H}_1}^T(t_1)}, \overline{\sigma^T(0)}\right] = 0$, which allows us to remove the first term of eq. 0.32. Moreover, this also implies that $\left[\overline{{\mathcal{H}_1}^T(t), \left[{\mathcal{H}_1}^T(t-\tau), \sigma^T(0)\right]}\right]$ and thus that the upper limit of the integrand can be extended to infinity.

Another significant assumption is that $\sigma^T(t) \approx \sigma^T(0)$ which is the zeroth order approximation to eq. 0.30 and means that the system must remain close to its starting

point. This can only be correct for a time t significantly smaller than the relaxation time of the system – but still longer than τ_c – i.e. $T_1, T_2 \gg t \gg \tau_c$ which effectively limits the range of possible t. The time t being significantly smaller than the relaxation times can be surprising as studying the relaxation of the spins is of prime importance in NMR. This apparent contradiction is solved by defining a time interval that is longer than T_1 and T_2 and made of multiple shorter timesteps that do satisfy the aforementioned conditions and which will be integrated stepwise in order to give the complete picture.

Finally, the last required assumption is that the system eventually reaches back the thermal equilibrium, i.e. $\sigma^T(t)$ has to be replaced ad hoc by $\sigma^T(t) - \sigma_0$ to ensure $\sigma^T(t)$ converges back to σ_0, which is a consequence of the lattice having a finite temperature. Altogether, this finally yields

$$\frac{d\sigma^T(t)}{dt} = -\int_0^\infty \left[\overline{\mathcal{H}_1{}^T(t), [\mathcal{H}_1{}^T(t-\tau), \sigma^T(t) - \sigma_0]}\right] d\tau \qquad 0.33$$

which is the final form of the master equation in the interaction frame.

It is however more convenient to express it in the laboratory frame and for this, the transformation properties of the stochastic Hamiltonian have to be identified. As much of spin relaxation results from random rotational modulations, $\mathcal{H}_1(t)$ can be expressed as the scalar contraction of a time-dependent spatial function $V(t)$ making use of spherical harmonics functions to represent the time-dependent reorientation of the interactions, and of an irreducible spherical tensor spin operator T of rank k and of coherence order q describing the nature of the interactions comprised in the Hamiltonian, as follows[412,427]

$$\mathcal{H}_1(t) = \sum_{q=-k}^{k} (-1)^q V_k^{-q}(t) T_k^q \qquad 0.34$$

Indeed, this scalar contraction of spherical tensor operators is convenient to describe the spin part of the Hamiltonian due to their well-defined transformation properties under rotations. Defining $V_k^{-q} \equiv (-1)^q V_k^{q*}$ and $T_k^{-q} \equiv (-1)^q T_k^{q\dagger}$ with * denoting the complex conjugate of the operator and † denoting the Hermitian conjugate. Each of these irreducible $\mathrm{T}_\mathrm{k}^\mathrm{q}$ can be further expanded as a linear combination of basis operators i.e.

$$\mathrm{T}_k^q = \sum_p T_{kp}^q$$

which in turn are eigenfunctions of $\hat{L}_0$ with eigenfrequencies $\omega_{kp}^q T_{kp}^q$ such as

$$\hat{L}_0\{T_{kp}^q\} = [\mathcal{H}_0, T_{kp}^q] = \omega_{kp}^q T_{kp}^q \qquad 0.35$$

with the subscript p introduced to distinguish operators of same order q but different eigenfrequencies. This implies that these basis operators are, in the interaction frame:

$$\begin{aligned} T_k^{qT} &= \exp(i\mathcal{H}_0 t) T_k^q \exp(-i\mathcal{H}_0 t) = \sum_p \exp(i\omega_p^q t) T_{kp}^q \\ T_k^{-qT} &= \sum_p \exp(-i\omega_p^q t) T_{kp}^{-q} \end{aligned} \qquad 0.36$$

with $\omega_p^{-q} = -\omega_p^q$. Assuming that the random spatial functions V_k^q and $V_k^{q'}$ are uncorrelated, any ensemble average of the two vanishes unless $q' = -q$. Consequently, putting together 0.36 and 0.34 in 0.33 yields

$$\begin{aligned} \frac{d\sigma^T(t)}{dt} = -\sum_{q=-k}^{k} \sum_{p,p'} \exp\{i(\omega_p^q - \omega_{p'}^q)t\} \left[T_{kp'}^{-q}, \left[T_{kp}^q, \sigma^T(t) - \sigma_0\right]\right] \\ \times \int_0^\infty \overline{V_k^q(t) V_k^{-q}(t-\tau) \exp(-i\omega_p^q \tau)}\, d\tau\,. \end{aligned} \qquad 0.37$$

If $|\omega_p^q - \omega_{p'}^q| >> 0$, the first term oscillates significantly faster than the common timescales of relaxation processes. Consequently, it will not affect the evolution of the density operator because it will have averaged out much earlier and can therefore be taken out of the equation. This is the *secular approximation* (with the neglected terms called *non-secular*), which is considered a fairly good approximation in the absence of quadrupolar couplings, whose consequences are important and will be discussed later. Moreover, in the absence of degeneracies in the eigenfrequencies, the remaining terms are non-secular and non-null only if $p' = p$. Altogether, eq. 0.37 becomes

$$\begin{aligned} \frac{d\sigma^T(t)}{dt} = -\sum_{q=-k}^{k} \sum_{p} \left[T_{kp}^{-q}, \left[T_{kp}^q, \sigma^T(t) - \sigma_0\right]\right] \\ \times \int_0^\infty \overline{V_k^q(t) V_k^{-q}(t-\tau)} \exp(-i\omega_p^q \tau)\, d\tau \end{aligned} \qquad 0.38$$

The right term is the product of the time-dependent spatial function $V_k^q(t)$ and its complex conjugate $V_k^{-q}(t)$, which can be recognised as a correlation function, while $\exp(-i\omega_p^q \tau)$ accounts for the evolution of the precession of the spins. In other words, the term $\overline{V_k^q(t) V_k^{-q}(t-\tau)}$ is the ensemble and time averaged correlation function $C(\tau)$ of the process described by the stochastic Hamiltonian which acts on the spin tensor operator to modulate its reorientation, ergo

$$C_q(\tau) = \langle \overline{V_k^q(t) V_k^{-q}(t-\tau)} \rangle \qquad 0.39$$

The real part of the integral in eq. 0.37 is therefore the *power spectral density function* $j^q(\omega_p^q)$:

$$j^q(\omega_p^q) = 2\,\mathrm{Re}\left\{ \int_0^\infty \overline{V_k^q(t) V_k^{-q}(t-\tau)} \exp(-i\omega_p^q \tau)\, d\tau \right\}$$

$$j^q(\omega_p^q) = \mathrm{Re}\left\{ \int_{-\infty}^\infty \overline{V_k^q(t) V_k^{-q}(t-\tau)} \exp(-i\omega_p^q \tau)\, d\tau \right\}$$

which is the Fourier transform of the autocorrelation functions, and is an even-valued function of ω while the imaginary part of the integral is $k^q(\omega_p^q)$

$$k^q(\omega_p^q) = \mathrm{Im}\left\{ \int_0^\infty \overline{V_k^q(t) V_k^{-q}(t-\tau)} \exp(-i\omega_p^q \tau)\, d\tau \right\}$$

which is an odd function of ω and thus, $k^{-q}(\omega_p^{-q}) = k^{-q}(-\omega_p^q) = -k^{-q}(\omega_p^q) =$ and $k^q(0) = 0$.

In the high temperature limit, the density matrix at equilibrium is proportional to the time independent static Hamiltonian, that is $\sigma_0 \propto \mathcal{H}_0$, and therefore, using eq. 0.35 allows to get rid of the σ_0 term. This allows to rewrite eq. 0.38 as

$$\frac{d\sigma^T(t)}{dt} = -\frac{1}{2} \sum_{q=-k}^{k} \sum_p \left[T_{kp}^{-q}, \left[T_{kp}^q, \sigma^T(t) - \sigma_0 \right] \right] j^q(\omega_p^q)$$

$$+ i \sum_{q=0}^{k} \sum_p \left[\left[T_{kp}^{-q}, T_{kp}^q \right], \sigma^T(t) \right] k^q(\omega_p^q)$$

which can be transformed back to the laboratory frame to obtain

$$\frac{d\sigma(t)}{dt} = -i[\mathcal{H}_0, \sigma(t)] - i[\Delta, \sigma(t)] - \hat{\Gamma}(\sigma(t) - \sigma_0)$$

With $\hat{\Gamma}$ the relaxation superoperator

$$\hat{\Gamma} = \frac{1}{2} \sum_{q=-k}^{k} \sum_p \left[T_{kp}^{-q}, \left[T_{kp}^q, \right] \right] j^q(\omega_p^q)$$

and Δ the dynamic frequency shift operator arising from the previously described imaginary part of the integral and responsible for second-order frequency shifts of the resonance lines

$$\Delta = -\sum_{q=0}^{k} \sum_p k^q(\omega_p^q) \left[T_{kp}^{-q}, T_{kp}^q \right]$$

which can be incorporated into $\mathcal{H}_0$ in order to obtain the final master equation.

$$\frac{d\sigma(t)}{dt} = -i[\mathcal{H}_0, \sigma(t)] - \hat{\Gamma}(\sigma(t) - \sigma_0) \qquad 0.40$$

The relaxation superoperator $\hat{\Gamma}$ is of fundamental importance for NMR spectroscopy. Indeed, one of its terms, the spectral density functions describe the probability density of observing relaxation at a given frequency ω of spins experiencing local stochastic magnetic field fluctuations. As these fluctuations arise from Brownian motion-induced perturbations, the power spectral density function is therefore a representation of the probability of rotational motions over a range of frequencies components (cf. section 1.4.5.3 page 80). As for the preceding term, the double commutation operator $\left[T_{kp}^{-q}, \left[T_{kp}^{q}, \right]\right]$, it effectively acts as a selection rule for the spectral density function and therefore allows relaxation only of the permitted transitions, i.e. in the case it doesn't vanish. In other terms, the relaxation superoperator describes how and when the spins relax.

1.4.3.3 Spectral density function and auto-correlation function

The previous equations included multiple spectral density functions $j^q(\omega)$ however, in the high temperature limit in isotropic fluids, the identity

$$j^q(\omega) = (-1)^q\, j^0(\omega) \equiv (-1)^q\, j(\omega)$$

is true and means that they can all be expressed as one sole spectral density function $(-1)^q\, j(\omega)$ instead of several[412,427].

As previously described, spin relaxation occurs due to stochastically modulated rotations. Therefore, the natural choice for the spatial functions in eq. 0.34 and beyond which describe the different relaxation mechanisms are traceless and symmetric rank 2 tensors V_2^q factored as a sum of normalized second-order spherical harmonics Y_2^q of the spherical coordinate. In turn, the time-dependence is expressed by the function $\Omega(t)$ describing the orientation of the principal direction vector **r** of the described interaction in the laboratory frame, with θ and ϕ respectively the polar and azimuthal angles. They are modulated by a spatial function coefficient c_0 unique to each relaxation interaction which arises from the physical constants and spatial variables at play in the interaction. Ergo

$$V_2^q(t) = c_0(t) Y_2^q[\Omega(t)]$$

with Y_2^q being the modified second order spherical harmonics functions obtained by multiplying the conventional spherical harmonics by a factor of $\sqrt{4\pi/2k+1}$ to get rid of the inverse term[428]. They are given in the following table:

Table 0.2 Modified spherical harmonics

q	$Y_2^q(\theta,\phi)$	$Y_2^{-q}(\theta,\phi) = (-1)^q Y_2^{q*}(\theta,\phi)$
0	$(3\cos^2\theta - 1)/2$	$(3cos^2\theta - 1)/2$
1	$-\sqrt{3/2}\sin\theta\cos\theta e^{i\phi}$	$\sqrt{3/2}\sin\theta\cos\theta e^{-i\phi}$
2	$\sqrt{3/8}\sin^2\theta e^{i2\phi}$	$\sqrt{3/8}\sin^2\theta e^{-i2\phi}$

In turn, the spectral density function $j(\omega)$ and the stochastic correlation function $C(\tau)$ are

$$j(\omega) = \mathrm{Re}\left\{\int_{-\infty}^{\infty} C(\tau)\exp\left(-i\omega_p^q\tau\right)d\tau\right\}$$
$$C(\tau) = \overline{c_0(t)c_0(t+\tau)Y_2^0[\Omega(t)]Y_2^0[\Omega(t+\tau)]}\,. \qquad 0.41$$

The terms $c(t)$ describe the nature of the relaxation-inducing mechanisms and are given in Table 0.3 for the three main mechanisms of relaxation, namely the dipole-dipole interaction, the chemical shift anisotropy (CSA) and the quadrupolar coupling. As the latter only concerns spins >1/2 which were not studied in the work presented in this thesis, we will only mention it briefly.

Mechanism of relaxation	c(t)
Dipole-dipole coupling	$\frac{-\sqrt{6}\mu_0\hbar\gamma_I\gamma_S}{4\pi} r_{IS}^{-3}(t)$
CSA	$\Delta\sigma\gamma_I B_0/\sqrt{3}$
Quadrupolar coupling	$e^2qQ/[4\hbar I(2I-1)]$

Table 0.3: Spatial coefficients
The spatial coefficients of the CSA and the quadrupolar coupling are given assuming respectively an axially symmetric chemical shift tensor, i.e. $\sigma_{zz} = \sigma_{\parallel}, \sigma_{xx} = \sigma_{yy} = \sigma_{\perp}$ and $\Delta\sigma = (\sigma_{\parallel} - \sigma_{\perp})$ for the former and an axially symmetric electric field gradient tensor for the latter i.e. with principal values $V_{zz} = eq$ and $V_{xx} = V_{yy}$ with e the electron charge and Q the nuclear quadrupole moment.

1.4.3.3.1 Rigid spherical molecule

The spatial function of a rigid spherical molecule undergoing Brownian motions is time-independent, i.e. $c(t) = c_0$. The autospectral density function is

$$j(\omega) = d_{00}J(\omega)$$

with $d_{00} = c_0^2$ and the orientation spectral density function $J(\omega)$ is

$$J(\omega) = \mathrm{Re}\left\{\int_{-\infty}^{\infty} C_{00}^2(\tau)\,\exp\left(-i\omega_p^q\tau\right)d\tau\right\}$$

with $C_{00}^2(\tau)$ the orientational correlation function:

$$C_{00}^2(\tau) = \overline{Y_2^0[\Omega(t)]Y_2^0[\Omega(t+\tau)]}\,. \qquad 0.42$$

If the spins are rigid rotors and the rotational diffusion is isotropic, $C_{00}^2(\tau)$ is[429]

$$C_{00}^2(\tau) = \frac{1}{5}e^{-\tau/\tau_c} \qquad 0.43$$

with τ_c the rotational *correlation time* defined as the average time for the molecule to rotate by 1 radian. The rotational diffusion phenomenon and this result will be treated in more detail in section 1.4.5.3 page 80. This correlation time τ_c is of prime importance for NMR spectroscopy and depends on a number of parameters including the molecular size of the rotating object, the viscosity of the solvent, and the temperature and is typically on the order of nanoseconds for biomolecules of the size of proteins. In turn, the associated spectral density function $J(\omega)$ is[415]

$$J(\omega) = \frac{2}{5}\frac{\tau_c}{(1+\omega^2\tau_c^2)} \qquad 0.44$$

which is a Lorentzian function. If the motion of the molecule is fast enough to satisfy $\omega^2\tau_c^2 \ll 1$, a condition known as the *extreme narrowing regime* in which case $J(\omega) \approx J(0)$; in the opposite case known as the *spin diffusion* or *slow tumbling regime* where $\omega^2\tau_c^2 \gg 1$ in which case $J(\omega) \propto \omega^{-2}$. Both the spectral density function and the orientational correlation function for different correlation times τ_c can be seen in Figure 2.2.1. In practice however, even a folded globular protein does not tumble as a perfectly rigid object, and both overall rotational motion of the protein and internal relative motions of the spins contribute to the overall stochastic modulation of the local magnetic fields. This point is treated in section 1.4.5.4 page 84.

1.4.3.4 The relaxation superoperator

In order to study further the relaxation superoperator, eq. 0.40 can be expanded in terms of the basis operators given in Table 0.1 to convert the master equation into its matrix form such as[412]

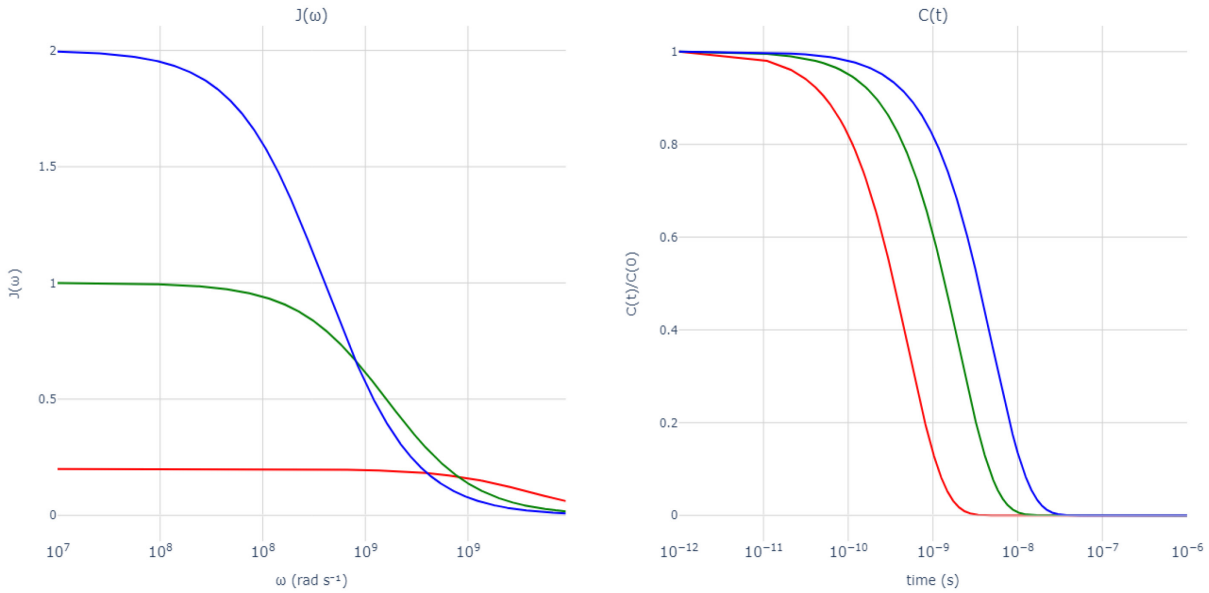


Figure 2.2.1: Examples of spectral density function and associated normalised autocorrelation function.

The correlation time used are $\tau_c = 0.5\,ns$ (red), $\tau_c = 2.5\,ns$ (green) or $\tau_c = 5\,ns$ (blue). The Lorentzian decay can clearly be seen in the spectral density function (left).

$$\frac{\mathrm{d}b_r(t)}{\mathrm{d}t} = \sum_s \{-i\Omega_{rs}b_s - \Gamma_{rs}[b_s(t) - b_{s0}]\} \qquad 0.45$$

With the characteristic frequencies Ω_{rs} defined as

$$\Omega_{rs} = \frac{\langle \mathbf{B}_r | [\mathcal{H}_0, \mathbf{B}_s] \rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle}$$

Indeed, the matrix form is more convenient to study the different relaxation terms and rates of the relaxation superoperator. Therefore, the relaxation between two operators $\mathbf{B}_r$ and $\mathbf{B}_s$ is defined by the rate constant $\Gamma_{rs} = \Gamma_{sr}$ in a normalized basis expressed as

$$\Gamma_{rs} = \frac{\langle \mathbf{B}_r | [\Gamma_{rs}\mathbf{B}_s] \rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} = \frac{1}{2}\sum_{q=-k}^{k}\sum_p \left[\frac{\left\langle \mathbf{B}_r \middle| \left[\mathbf{T}_{kp}^{-q}, \left[\mathbf{T}_{kp}^{q}, \mathbf{B}_s \right] \right] \right\rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} \right] j^q(\omega_p^q). \qquad 0.46$$

Finally, the application of the chosen projection superoperator onto the density matrix $\sigma(t)$ yields

$$b_r(t) = \frac{\langle \mathbf{B}_r | \sigma(t) \rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle}$$

In its matrix form, the diagonal elements Γ_{rr} of the relaxation matrix are the *auto-relaxation* rates of the operator $\mathbf{B}_r$ while the off-diagonal elements Γ_{rs} are the *cross-relaxation* rates between $\mathbf{B}_r$ and $\mathbf{B}_s$. As discussed in reaching eq. 0.37, only terms

satisfying $q' = -q$ are not null, i.e. cross-relaxation can only happen between operators of the same coherence order. For instance, cross-relaxation cannot happen between single and double quantum coherences. Moreover, cross-relaxation can happen only if the difference between the eigenfrequencies involved is small, i.e. as $|\omega_r - \omega_s| \to 0$ and therefore, under the secular approximation – which lead to the distinction between like and unlike spins – only degenerate transitions can give rise to cross-relaxation. This gives a distinctive kite-like shape to the matrix of the relaxation rates constants between the operators which is commonly known as the Redfield kite.

1.4.3.4.1 Cross-correlation between relaxation mechanisms

Typically, several mechanisms induce relaxation at the same time. Thus, their interference, that is their cross-correlation, is important to take into account to describe the system. In such cases, eq. 0.34 becomes

$$\mathcal{H}_1(t) = \sum_m \sum_{q=-k}^{k} (-1)^q V_{mk}^{-q}(t) T_{mk}^{q}$$

With m denoting the relaxation mechanism. We define the cross-correlated spectral density function j_{mn}^{q} *as*

$$j_{mn}^{q}(\omega) = \mathrm{Re}\left\{\int_{-\infty}^{\infty} \overline{V_{mk}^{q}(t) V_{nk}^{-q}(t-\tau)} \exp\left(-i\omega_p^q \tau\right) d\tau\right\} \qquad 0.47$$

and in turn, eq. 0.46 becomes

$$\Gamma_{rs} = \frac{1}{2}\sum_m \sum_{q=-k}^{k} \sum_p \left[\frac{\left\langle \mathbf{B}_r \middle| \left[\mathbf{T}_{mkp}^{-q}, \left[\mathbf{T}_{mkp}^{q}, \mathbf{B}_s\right]\right]\right\rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle}\right] j^q\left(\omega_p^q\right)$$

$$+ \frac{1}{2}\sum_{\substack{m,n \\ m\neq n}} \sum_{q=-k}^{k} \sum_p \left[\frac{\left\langle \mathbf{B}_r \middle| \left[\mathbf{T}_{mkp}^{-q}, \left[\mathbf{T}_{nkp}^{q}, \mathbf{B}_s\right]\right]\right\rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle}\right] j_{mn}^{q}\left(\omega_p^q\right) = \sum_m \Gamma_{rs}^{m} + \sum_{\substack{m,n \\ m\neq n}} \Gamma_{rs}^{mn}$$

with Γ_{rs}^{m} the auto-relaxation rate constant for the *m*th mechanism of relaxation while Γ_{rs}^{mn} is the cross-correlated rate constant between the *m*th and *n*th relaxation mechanism. Eq. 0.47 implies that the random processes $V_{mk}^{q}(t)$ and $V_{nk}^{q}(t)$ have to be correlated for interference to occur, otherwise $j_{mn}^{q}(\omega) = 0$ and thus $\Gamma_{rs}^{mn} = 0$, in which case the individual mechanisms simply contribute additively to the relaxation of the system. Interference effects typically result in "forbidden" cross-relaxation pathways, i.e. cross-

relaxation between operators which wouldn't be possible otherwise. This is cross-correlated cross-relaxation.

1.4.4 Relaxation mechanisms

Several physical mechanisms cause spin relaxation. The most frequently encountered are the dipolar, quadrupolar and scalar interactions as well the chemical shift anisotropy. In the presence of a paramagnetic species, paramagnetic relaxation is also an important contribution. This section introduces the relaxation mechanisms relevant for this thesis.

In the following, **I** and **S** refer to the spin operators of the spins I and S and γ_I and γ_S their respective gyromagnetic ratios.

1.4.4.1 Dipole-dipole induced relaxation

The local magnetic field of each nuclear spin arising from their magnetic moment affects the neighbouring spins. This results in a mutual through-space interaction called the magnetic dipole-dipole interaction. The Hamiltonian for an interaction between two spins I and S whose spin operators are **I** and **S** is given by[412,413]

$$\mathcal{H}^{dd} = d_{IS}(3(\mathbf{I}\cdot\mathbf{e}_{\mathrm{IS}})(\mathbf{S}\cdot\mathbf{e}_{\mathrm{IS}}) - \mathbf{I}\cdot\mathbf{S})$$

and with the dipolar coupling constant d already mentioned in Table 0.3 being

$$d_{IS} = -\frac{\mu_0\hbar\gamma_I\gamma_S}{4\pi}r_{IS}^{-3}$$

with **e** the unit vector parallel to the line joining the centres of the two nuclei I and S separated by an intermolecular distance r_{IS} and with μ_0 the permeability of free space constant. To continue with eq. 0.34, the dipole-dipole interaction Hamiltonian can further be decomposed as a product of the aforementioned rank-2 spherical harmonics described in the Table 0.2 and of rank-2 tensors operators such as

$$\mathcal{H}^{dd} = d_{IS}\sqrt{6}\sum_{q=-2}^{2}\sum_{p}(-1)^{q}\mathrm{Y}_2^{\mathrm{q}}(\Omega_{\mathrm{IS}})\mathrm{T}_{2p}^{q}$$

with the tensor operators T_{2p}^{q} and their characteristic frequencies $\omega_{2\mathrm{p}}^{\mathrm{q}}$ given in the following table.

q	p	T_{2p}^{q}		$T_{2p}^{-q} = (-1)^{q} T_{2p}^{q\dagger}$	ω_{2p}^{q}
0	0	$(2/\sqrt{6}) I_Z S_Z$		$(2/\sqrt{6}) I_Z S_Z$	0
0	-1	$-1/(2/\sqrt{6})\, I^- S^+$		$-1/(2/\sqrt{6})\, I^+ S^-$	$\omega_S - \omega_I$
0	1	$-1/(2/\sqrt{6})\, I^+ S^-$		$-1/(2/\sqrt{6})\, I^- S^+$	$\omega_I - \omega_S$
1	0	$-(1/2) I_Z S^+$		$(1/2) I_Z S^-$	ω_S
1	1	$-(1/2) I^+ S_Z$		$(1/2) I^- S_Z$	ω_I
2	0	$(1/2) I^+ S^+$		$(1/2) I^- S^-$	$\omega_I + \omega_S$

As the Zeeman interaction is the dominant interaction in NMR and its magnitude is proportional to the strength of the magnetic field, at high-field, the dipolar interaction Hamiltonian can be considered a perturbation of the Zeeman Hamiltonian. This is the high-field approximation. Therefore, only the parts of the dipolar coupling Hamiltonian that commute with the Zeeman Hamiltonian remain. Moreover, if the sum of the scalar and dipolar couplings between the two spins is much smaller than the magnitude of the Zeeman splitting ΔE (eq. 0.8), the secular approximation holds and only the secular parts of the Hamiltonian remain, which corresponds to the terms that satisfy $|\omega_I - \omega_S| >> d$. The remaining terms differ whether the interaction is homonuclear or heteronuclear. In the latter case, the remaining effective Hamiltonian is

$$\mathcal{H}_{secular}^{dd} = d_{IS} \frac{3\cos^2\theta - 1}{2} 2\mathbf{I_z S_z}$$

with $\mathbf{I_z}$ and $\mathbf{S_z}$ angular momentum of the spins I and S while θ is the angle of the internuclear vector between the spins I and S relative to the magnetic field. Therefore, the secular dipolar Hamiltonian directly depends on a second order Legendre polynomial of the orientation θ. Consequently, an angle θ between the internuclear vector and the z-axis of the magnetic field can be found at which $\mathcal{H}_{secular}^{dd} = 0$ which is therefore $\theta_0 = \arccos\left(\frac{1}{\sqrt{3}}\right) \cong 54.74°$. It is called the magic angle and is the basis for magic-angle spinning in solid-state NMR.

These equations make clear that the dipolar coupling constant d_{IS} depends on the gyromagnetic ratio of the nuclei, and on the distance r between them, evolving as function of r^{-3} and doesn't depend on the orientation. It is noteworthy that no term in this equation depends on the static magnetic field and thus, the dipole-dipole induced relaxation is field-independent.

Moreover, two dipolar-coupled spins can cross-relax, which effectively corresponds to the transfer of magnetisation from one spin to another through the dipole interaction. This is the nuclear Overhauser effect (nOe) discussed in section 1.4.5.1.2 page 78.

1.4.4.2 Chemical-shift anisotropy induced relaxation

Electrons have a spin of $\frac{1}{2}$ and therefore, a magnetic moment. Under a static magnetic field, their local magnetic fields are experienced by the nucleus they surround as well as the neighbouring nuclei and effectively modify the nuclear energy levels, thus shifting slightly their Larmor frequencies and allowing them to be spectroscopically distinguished. This is the *chemical shift*, whose Hamiltonian is[412]

$$\mathcal{H}^{CS} = \gamma B_0 \hbar \boldsymbol{I}\boldsymbol{\sigma} \qquad 0.48$$

with σ the shielding tensor of rank 2 describing the shielding effect of the electronic cloud on the nuclear spin. It is sometimes considered to be symmetric in its principal axis frame as described in Table 0.3[412].

Due to the geometry of the electronic orbitals, the chemical shift is generally not isotropic and consequently, its magnitude changes with the orientation of the spins with respect to the magnetic field. The chemical shift interaction is thus decomposed in the isotropic chemical shift contribution (CSI) and the traceless chemical shift anisotropy (CSA) contribution, so $\mathcal{H}^{CS} = \mathcal{H}^{CSI} + \mathcal{H}^{CSA}$. In isotropic solutions, the fast overall tumbling of the molecules effectively averages out the anisotropic chemical shift, i.e. $\langle \mathcal{H}^{CSA} \rangle = 0$, and therefore, only the isotropic component is detected which does not contribute to relaxation processes due to its invariance with respect to the aforementioned reorientational motions. It is thus included in the static Hamiltonian. This is not the case for the CSA which does contribute to relaxation[412].

Although usually assumed to be negligible for ^{1}H, the CSA has a quadratic dependence on the B_0 magnetic field strength and contributes significantly to the relaxation of ^{15}N or ^{13}C

amongst others. The T_{2p}^{q} and relaxation rate constants are given in the next table and assume, again, an axially symmetric CSA tensor[412].

q	p	T_{2p}^{q}	$T_{2p}^{-q} = (-1)^{q} T_{2p}^{q\dagger}$	ω_{p}^{q}
0	0	$(2/\sqrt{6})I_Z$	$(2/\sqrt{6})I_Z$	0
1	1	$-(1/2)I^{+}$	$(1/2)I^{-}$	ω_I

The CSA typically interferes with the dipolar coupling, resulting in the cross-correlated effects described later.

1.4.5 Relaxation and dynamics of the protein backbone amide ^{15}N

As relaxation depend on molecular motions, it is possible to interpret experimentally measured spin relaxation rates to characterise dynamic properties of the observed protein, either using a physical model describing the effect of motions on the relaxation rates or using a model free characterisation of the correlation function.

1.4.5.1 Relaxation rates

Throughout this section, we consider the relaxation rates arising from only the CSA, noted c, and dipole-dipole induced relaxation mechanisms, noted d, with the former on the amide ^{15}N atom and the latter between the ^{15}N and ^{1}H atoms. Their expressions vary in the literature and will be throughout this thesis[415]

$$c = (\sigma_{\parallel} - \sigma_{\perp})\omega_N \quad \text{and} \quad d = \frac{\mu_0 \hbar \gamma_N \gamma_H}{4\pi r_{NH}^3}$$

with $(\sigma_{\parallel} - \sigma_{\perp})$ the difference between the parallel and perpendicular components of the ^{15}N chemical shift tensor, ω_N the Larmor frequency of the ^{15}N nucleus, μ_0 the vacuum permeability, $\hbar$ the reduced Planck constant, γ_X the gyromagnetic ratio of the X nucleus and r_{NH} the interatomic distance between the ^{15}N and ^{1}H nuclei of the amide bond.

The $\frac{2}{5}$ factor in the spectral density function in eq. 0.44 is due to the normalisation of the modified spherical harmonics and is to be included or not depending on the exact

expression of the spin relaxation rates. It will be omitted from now on as the following expressions include it already,

1.4.5.1.1 Auto-relaxation rates

The equations derived from the Bloch-Wangness-Redfield theory for the longitudinal and transversal auto-relaxation rates R_1 and R_2 of the amide ^{15}N due to its CSA and dipolar coupling to the amide ^{1}H are[415]:

$$R_1 = \frac{d^2}{10}\left[3J(\omega_N) + 6J(\omega_H + \omega_N) + J(\omega_H - \omega_N)\right] + \frac{2c^2}{15}J(\omega_N) \qquad 0.49$$

$$R_2 = \frac{d^2}{20}[4J(0) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) + J(\omega_H - \omega_N)] + \frac{c^2}{45}[4J(0) + 3J(\omega_N)]$$

with R_1 describing the rate of the return to equilibrium of the bulk magnetisation along the B_0 field and R_2 the decay of the bulk magnetisation in the transverse, perpendicular plane to B_0. Both are the sum of the dipolar interaction contribution, noted d, and of that of the CSA noted c, modulated by the sensitivity of these different rates to motions occurring at different timescales; hence the different components of the spectral density function terms at 5 different eigenfrequencies, that is 0, ω_N, $\omega_H + \omega_N$, ω_H and $\omega_H - \omega_N$.

The examination of these equations shows that the biggest term in R_1 is its $6J(\omega_H + \omega_N)$ component which has the most important prefactor, indicating its increased sensitivity to motions typically occurring on the 100 ps to 1 ns timescale, while R_2 is typically dominated by its $J(0)$ component in proteins, indicating sensitivity to the total amount of dynamics. Therefore, R_2 is dominated by the motions occurring up to the slowest relaxation active timescale which is typically on the order of tens of nanoseconds.

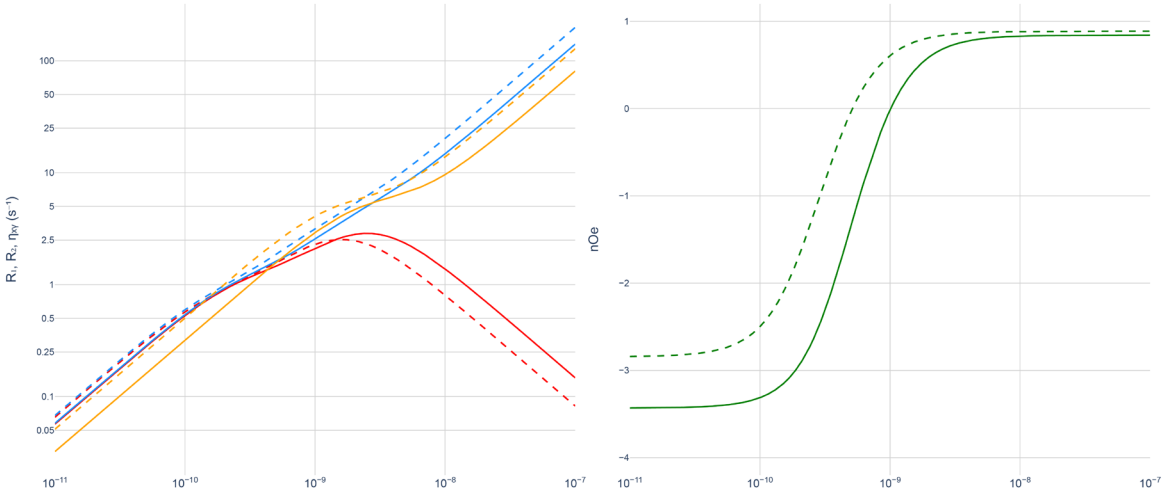


Figure 0.2: Dependence of the relaxation rates on the correlation time.
The rates are R_1 (red), R_2 (blue) and η_{xy} (yellow) on the left, and nOe (green) on the right. The solid lines are the calculated rates at 14.09 T (600 MHz) and the dashed lines at 22.3 T (950 MHz) for $(\sigma_{\parallel} - \sigma_{\perp}) = -172$ ppm, $r_{NH} = 1.023$ Å and $\theta = 22.5$ °.

1.4.5.1.1.1 $R_{1\rho}$

Due to the experimental imprecision and the exchange contribution R_2^{ex} to R_2 (cf. section 1.4.6 page 87), the ^{15}N longitudinal spin relaxation rate constant in the rotating frame $R_{1\rho}$ is often measured instead of R_2. It is expressed as[412]

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta \,.$$

To measure it, a continuous-wave RF field is applied in the transverse plane of the rotating frame which results in the spins experiencing an effective field arising from the vector sum of this spin-lock of strength ω_1 and the resonance offset $\Omega = \omega - \omega_0$ where ω is the carrier frequency and ω_0 the ^{15}N chemical shift of the observed peak. The tilt angle θ of the RF field experienced by the spins is therefore expressed as $\theta = \arctan(\omega_1/\Omega)$.

$R_{1\rho}$ is therefore the relaxation rate constant along the effective field, which can be used together with R_1 to back-calculate R_2. Because the value of $R_{1\rho}$ depends on the orientation and magnitude of this effective field, which in turn depends on the spin-lock strength ω_1 and the offset Ω, variations of those two parameters result in changes in the exact exchange contribution of R_2^{ex} to $R_{1\rho}$ and can therefore be used to characterise it[430].

1.4.5.1.2 Cross-relaxation rate and nOe

The cross-relaxation effect between the longitudinal coherences of the dipolar-coupled ^{15}N and ^{1}H spins gives rise to the cross relaxation rate σ_{NH} whose expression is[415]

$$\sigma_{NH} = \frac{d^2}{10}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] .$$

It is typically accessed experimentally by measuring R_1 and the steady-state nuclear Overhauser effect (nOe) enhancement obtained by saturating ^{1}H atoms using an RF field duration significantly longer than the T_1 of the spins to reach a steady state and comparing it to the magnetisation obtained without this saturation. The expression of the $^{15}N - \{^{1}H\}$ nOe relaxation is thus

$$\text{nOe} = 1 + \frac{\gamma_H}{R_1\gamma_N}\sigma_{NH} = 1 + \frac{\gamma_H}{R_1\gamma_N}\frac{d^2}{10}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

with R_1 being the longitudinal relaxation rate of the ^{15}N atom. As the Overhauser effect arises from the dipolar coupling, there is no contribution from the CSA except from this R_1 rate constant. As its sensitive frequencies are close to ω_H, nOe reports on fast motions occurring on the timescale of tens of picoseconds to hundreds of picoseconds.

1.4.5.1.3 CSA-DD cross-correlated relaxation

In a protein, cross-correlation occurs between the ^{1}H-^{15}N dipole-dipole induced relaxation of the backbone amide and the CSA of the ^{15}N, resulting in a longitudinal cross-correlated relaxation η_z and a transversal one η_{xy}. As such, their values depend on the angle θ between the principal axes of the CSA tensor and the dipolar interaction. Their expressions are

$$\begin{aligned} \eta_z &= \frac{1}{15}cd\ \mathrm{P}_2(\cos\theta)[6J(\omega_N)] \\ \eta_{xy} &= \frac{1}{15}cd\ \mathrm{P}_2(\cos\theta)[4J(0) + 3J(\omega_N)] \end{aligned} \qquad 0.50$$

with P_2 being the Legendre polynomial of second order. In a similar way to R_1, the longitudinal cross-correlated relaxation rate η_z is dominated by $J(\omega_N)$ while the transversal one η_{xy} is, like R_2, dominated by $J(0)$ and therefore reports on similar timescales, but differs by being free from exchange contributions.

1.4.5.2 Residual dipolar couplings

In isotropic liquids, dipolar couplings are normally averaged to zero due to the spins experiencing all possible orientations, while in the solid state, the entire dipolar interaction is present due to the absence of Brownian motion-induced orientational averaging. However, in weakly anisotropic liquids, molecules partially align relative to the static magnetic field, resulting in only a partial averaging of the dipolar interaction. The residual interaction leads to Residual Dipolar Couplings (RDCs). Although the full dipolar coupling is a strong interaction on the order of kHz, RDCs are typically on the order of tens of Hertz or less, which limits their broadening effects on the NMR spectra while rendering accessible their distance and angular information[431,432].

In folded proteins the dependence of RDCs on the different internuclear vector orientations with regards to a common reference frame is expressed in the alignment tensor which can be in turn used for structure determination. Moreover, the coupling strength of RDCs corresponds to a timescale of tens of milliseconds which allows them to be used as probes of motional averaging up to the chemical shift timescale limit of the order of the millisecond. As RDC values are uncorrelated to the rotational diffusion of the molecule, they report on motions that are both slower and faster than the correlation time τ_c of the molecule. Thus, for intrinsically disordered proteins, RDCs report on both the ensemble averaged distribution of dihedral angles between neighbouring amino-acids and long-range chain contacts which allows them to be used understand the ensemble of conformations sampled by the protein that interconvert on these timescales. As such, RDCs of protein bond vectors allow to determine the relative orientations of different structural elements in proteins, effectively probing their tertiary structure[16,432].

Under the high-field approximation (cf. section 1.4.4.1 page 72), the measured dipolar coupling between two spins I and S noted D_{IS} is[16,432]

$$D_{IS} = -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi^3} r_{IS}^{-3} \left\langle \frac{3\cos^2\Omega - 1}{2} \right\rangle$$

with Ω being the orientation of the internuclear vector between spins I and S relative to the static magnetic field and the angular bracket denoting the ensemble average of all possible orientations sampled by the spins at rates faster than the millisecond timescale.

Experimentally, partial alignment resulting from a weak anisotropy of the solution can be obtained by using filamentous phages with high magnetic susceptibility, liquid crystals based on alcohols or lipid bicelles, as well as mechanically stressed polyacrylamide gels, in each case weakly restricting the orientational sampling of the protein of interest so that it is no longer isotropic. However, as almost all molecules have an anisotropic magnetic susceptibility, natural alignment in respect to the static magnetic field is virtually always present, although it is negligible for diamagnetic species, but can become important in the presence of a repetitive order of structures with a substantial magnetic susceptibility like in the aforementioned phages[432]. In paramagnetic species, an anisotropic magnetic susceptibility induces pseudo-contact shifts[433].

1.4.5.3 Rotational diffusion

1.4.5.3.1.1 Physical mechanism

A protein in solution experiences constant collisions on the picosecond timescale with solvent molecules, salts and other proteins arising from the random thermal Brownian motions, which results in translational displacement and reorientational motions termed translational and rotational diffusion. Following the work of Stokes[434], Einstein[435], Smoluchowski[436] and Langevin[437], translational diffusion is directly linked to the mean-square displacement (MSD) of the particle such as

$$\langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \rangle = kD_T t \qquad 0.51$$

with $\boldsymbol{r}(t)$ being the time-dependent position of the protein such as $\boldsymbol{r}(t) = (x, y, z)^T$, k being the dimensional prefactor which is equal to 2, 4 or 6 for respectively one-, two- and three-dimensional diffusion and D_T the isotropic translational diffusion coefficient.

In turn, as per Stokes[434], Einstein[435] and Debye[438], rotational diffusion is linked to the mean-square angular displacement (MSAD) of the particle such as

$$\langle |\Theta(t) - \Theta(0)|^2 \rangle = 2D_R t \,,$$

with $\Theta(t)$ the single-axis time-dependent rotational angle, hence $k = 2$ and D_R the isotropic rotational diffusion coefficient.

However, the viscous drag resulting from the constant collisions also opposes these thermally driven transport and reorientation processes. This opposition can be seen as respectively a dissipative force and torque for translational and rotational diffusion. As such, they depend on the Stokes friction coefficient ξ with the solvent determined as

$$\xi_T = 6\pi\eta R_\mathrm{h} \qquad \text{and} \qquad \xi_R = 8\pi\eta {R_\mathrm{h}}^3$$

with η being the experienced viscosity and R_h the hydrodynamic radius of the protein. Thus, in the infinitely dilute limit, the diffusion translational and rotational diffusion coefficients amount to the ratio of the available thermal energy over the friction coefficients, i.e.

$$D_T = \frac{k_B T}{6\pi\eta R_\mathrm{h}} \qquad \text{and} \qquad D_\mathrm{R} = \frac{k_B T}{8\pi\eta {R_\mathrm{h}}^3} \qquad 0.52$$

which are respectively the Stokes-Einstein and Stokes-Einstein-Debye equations, with both of them assuming no-slip Navier's boundary conditions in an incompressible fluid and the particle to be a sphere. However, these equations are only valid as long as molecules do not perturb each other's hydrodynamic radii R_h, i.e. until the semidilute regime is reached[439–441].

However, many proteins are not globular and have shapes that cannot be approximated as a sphere. Moreover, intermolecular interactions can also result in an energetic barrier governing the geometric reorientation, such as a disk-shaped protein rotating more efficiently in the plane of the disk than in its transversal plane. These equations can then nevertheless be modified to quantitively take into account these geometric retardations for oblate or prolate spheroid shapes using the Perrin friction coefficients[442]. In such cases, the diffusion of the protein is not isotropic anymore but anisotropic, which has consequences on its autocorrelation and spectral density functions described below.

1.4.5.3.1.2 Study by NMR

As this section concerns rotational diffusion only, the rotational diffusion coefficient D_R is noted simply $\boldsymbol{D}$ for the sake of simplicity and consistency with the usual notations. It represents a rank-two cartesian tensor such as

$$\boldsymbol{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{My} & D_{zz} \end{pmatrix}$$

with the diagonal elements representing the eigenvalues of the diffusion tensor about the x, y and z principal axes frame and the six off-diagonal terms describing the correlations of the motions of each pair of principal directions. As per convention, $D_{xx} \leq D_{yy} \leq D_{zz}$ and for the case of an isotropic diffusion, $D_{xx} = D_{yy} = D_{zz} = D$ with the off-diagonal elements being null.

As discussed in the section 1.4.3.3.1 page 68, the spectral density functions of a rigid spherical top undergoing isotropic rotation is given by[443]

$$\mathrm{J}(\omega) = \frac{\tau_c}{(1 + \omega^2 \tau_c^2)}$$

with τ_c being, as per eq. 0.52,

$$\tau_c = \frac{1}{6D} = \frac{4\pi\eta {R_\mathrm{h}}^3}{3k_B T}. \qquad 0.53$$

However, in the case of a rigid top with fully anisotropic rotational diffusion, the spectral density function is the sum of 5 different components depending on the orientations of the considered N-H bond vector such as[443]

$$\mathrm{J}(\omega) = \sum_{i=1}^{5} A_i \frac{\tau_i}{(1 + \omega^2 \tau_i^2)}$$

with the coefficients A_i describing these different exponential processes being

$$A_1 = 3\mathbf{y}^2\mathbf{z}^2, \qquad A_2 = 3\mathbf{x}^2\mathbf{z}^2, \qquad A_3 = 3\mathbf{x}^2\mathbf{y}^2$$

$$\begin{aligned} A_4 = \frac{1}{4}[3(\mathbf{x}^4 + \mathbf{y}^4 + \mathbf{z}^4) - 1] \\ + \frac{1}{12}\left[\delta_x(3\mathbf{x}^4 + 6\mathbf{y}^2\mathbf{z}^2 - 1) + \delta_y(3\mathbf{y}^4 + 6\mathbf{x}^2\mathbf{z}^2 - 1)\right. \\ \left. + \delta_z(3\mathbf{z}^4 + 6\mathbf{y}^2\mathbf{x}^2 - 1)\right] \end{aligned}$$

$$\begin{aligned} A_5 = \frac{1}{4}[3(\mathbf{x}^4 + \mathbf{y}^4 + \mathbf{z}^4) - 1] \\ - \frac{1}{12}\left[\delta_x(3\mathbf{x}^4 + 6\mathbf{y}^2\mathbf{z}^2 - 1) + \delta_y(3\mathbf{y}^4 + 6\mathbf{x}^2\mathbf{z}^2 - 1)\right. \\ \left. + \delta_z(3\mathbf{z}^4 + 6\mathbf{y}^2\mathbf{x}^2 - 1)\right] \end{aligned}$$

and their associated time intervals τ_i being

$$\tau_1 = \left(4D_{xx} + D_{yy} + D_{zz}\right)^{-1}$$

$$\tau_2 = \left(D_{xx} + 4D_{yy} + D_{zz}\right)^{-1}$$

$$\tau_3 = \left(4D_{xx} + D_{yy} + 4D_{zz}\right)^{-1}$$

$$\tau_4 = \left(6D_{iso} + 6\left({D_{iso}}^2 - L^2\right)^{1/2}\right)^{-1}$$

$$\tau_5 = \left(6D_{iso} - 6\left({D_{iso}}^2 - L^2\right)^{1/2}\right)^{-1}$$

where

$$D_{iso} = \frac{1}{3}\left(D_{xx} + D_{yy} + D_{zz}\right)$$

$$L^2 = \frac{1}{3}\left(D_{xx}D_{yy} + D_{xx}D_{zz} + D_{yy}D_{zz}\right)$$

$$\delta_k = (D_{kk} - D_{iso})/\left({D_{iso}}^2 - L^2\right)^{1/2}$$

for $k = (x, y, z)$. In turn, $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are the direction cosines of the N-H bond vector in the principal axis frame of the diffusion tensor.

In the case of an axially symmetrical diffusion tensor, the spectral density function form is simplified and the tensor is described with $D_\perp$ and $D_\parallel$. If the tensor can be approximated as an oblate ellipsoid, then $D_\perp = D_{xx} = D_{yy}$ and $D_\parallel = D_{zz}$ whereas for a prolate ellipsoid approximation, $D_\perp = D_{zz} = D_{yy}$ and $D_\parallel = D_{xx}$[444]. As per the isotropic diffusion constant, it is $D_{iso} = (2D_\perp + D_\parallel)/3$ and the form of $\mathrm{J}(\omega)$ is now given by the sum of 3 components[443]:

$$\mathrm{J}(\omega) = \sum_{i=1}^{3} A_i \frac{\tau_i}{(1 + \omega^2\tau_i^2)}$$

where the coefficients A_i are

$$A_1 = \frac{1}{4}(3\cos^2\theta - 1)^2, \qquad A_2 = 3\sin^2\theta\,3\cos^2\theta\,, \qquad A_3 = \frac{3}{4}(\sin^4\theta)$$

and their associated time intervals τ_i being

$$\tau_1 = (6D_\perp)^{-1}, \qquad \tau_2 = (5D_\perp + D_\parallel)^{-1}, \qquad \tau_3 = (2D_\perp + 4D_\parallel)^{-1}$$

with θ is the angle between the N-H bond and the unique axis of the principal frame of the diffusion tensor.

Experimentally, the ratio R_2/R_1 is fitted against those equations to obtain the eigenvalues of the diffusion tensor and the Euler angles relating it to the molecular frame of the protein. Indeed, as aforementioned, R_2 and R_1 are respectively dominated by $\mathrm{J}(0)$ and $\mathrm{J}(\omega_N)$ and thus, are more sensitive to slow, reorientational motions such as rotational

diffusion assuming that internal motions are fast enough to scale the relaxation rates by an order parameter and are of very small amplitudes.

1.4.5.3.1.3 Diffusion of water

The rotational diffusion time of water is on the order of picoseconds, which falls in the extreme narrowing regime defined as $\omega^2\tau_c^2 \ll 1$ in section 1.4.3.3.1. Assuming the water molecules diffuse isotropically, their spectral density function is $J(\omega) = \frac{\tau_c}{(1+\omega^2\tau_c^2)}$ in which the denominator is ≈ 1. Thus, the spectral density function of the O-H bond vector of water and therefore its ^{1}H relaxation rates are directly proportional to its tumbling time τ_c such as $J(0) \approx J(\omega) \approx \tau_c$.
As per eq. 0.53, $\tau_c = 1/6D \propto \eta$, that is the correlation time τ_c and thus the water ^{1}H relaxation rates are inversely proportional to the experienced viscosity through the rotational diffusion coefficient D, making R_1 and R_2 of water protons good probes of the nano-viscosity as already noted by Abragam[415].

In this case, a solvent friction term $\rho(C)$ can be expressed as a function of the changes in nano-viscosity and thus of water R_1 regarding a reference solution such as

$$\rho(C) = \frac{(\eta_C - \eta_0)}{\eta_0} = \frac{\left(R_{1,C} - R_{1,0}\right)}{R_{1,0}}$$

with η_C and $R_{1,C}$ the viscosity and longitudinal water relaxation of the sample of interest and η_0 and $R_{1,0}$ the viscosity and longitudinal water relaxation of the sample of reference [439]. As such, longitudinal water ^{1}H relaxation provides a straightforward way of quantifying changes in nano-viscosity.

1.4.5.4 Model-free analysis of folded proteins

Different physical models have been developed to quantitively describe protein motions from relaxation measurements, but it is difficult to distinguish the best one statistically as they often fit the data equally well, but have to be simple enough due to the limited number of independent experimental measurement accessible.

In 1982, Lipari and Szabo[445–447] introduced which consists in not using any physical model of the motion, but instead describing the molecular motions responsible for relaxation in a *model-free* way. This is achieved by using a generalized order parameter S^2 to characterise the motions and an effective correlation time of the internal motions τ, as well as by assuming that the protein is undergoing two independent motions: a global motion, commonly identified as rotational diffusion, and internal motion. In turn, the statistical independence of these motional processes implies that they are decoupled[447]. The correlation function is thus the product of their correlation functions:

$$C(t) = C_{global}(t)\, C_{internal}(t)$$

As previously described in eq. 0.43, assuming that rotational diffusion is isotropic yields

$$C_{global}(t) = \mathrm{e}^{-t/\tau_c}$$

with τ_c being the overall rotational diffusion time. However, unlike $C_{global}(t)$, $C_{internal}(t)$ typically does not reach 0, but stabilises at a plateau which depends on how much the internal motions are restricted. Thence, the generalized order parameter S^2 is equal to the asymptote of this plateau, i.e. $S^2 = C_{internal}(\infty)$. Thus, if the internal motions are not restricted and can sample the whole spherical space, then $C_{internal}(t)$ fully decays and, as $0 \leq S^2 \leq 1$ because the correlation function is normalised, $S^2 = 0$. Conversely, if there are no internal motions, $C_{internal}(t)$ does not decay and $S^2 = 1$. This implies that $1 - S^2$ is amplitude of the contribution of the internal dynamics to the total relaxation. Altogether, this allows to express the protein's correlation function as

$$C(t) = e^{-t/\tau_c}[S^2 + (1 - S^2)\, C_{internal}(t)]\,.$$

Assuming that the internal correlation function is a single exponentially decaying function, the Fourier transform of $C(t)$ gives the corresponding spectral density function

$$\mathrm{J}(\omega) = S^2 \frac{\tau_c}{(1 + \omega^2\tau_c^2)} + (1 - S^2)\frac{\tau_i{}'}{(1 + \omega^2\tau_i'^2)} \qquad 0.54$$

Where the effective correlation time $\tau_i'^{-1} = \tau_c{}^{-1} + \tau_i{}^{-1}$ with τ_i being the internal correlation time and satisfying $\omega\tau_i \ll 1$. This equation is valid for an isotropic rotational diffusion, whereas in the presence of an anisotropic global motion, a related function can be usefully written (by analogy) [445,448]

$$\mathrm{J}(\omega) = \sum_{n=1}^{k} A_k \left[S^2 \frac{\tau_{c,k}}{\left(1 + \omega^2\tau_{c,k}^2\right)} + (1 - S^2)\frac{\tau'_{i,k}}{(1 + \omega^2\tau_k'^2)} \right] \qquad 0.55$$

where ${\tau'_{i,k}}^{-1} = {\tau_{c,k}}^{-1} + {\tau_i}^{-1}$ with $k = 5\ or\ 3$ in respectively the fully anisotropic rotational diffusion case and the axially symmetric case, with the coefficient A_k and the time intervals being given in the section 1.4.5.3.1.2 page 81.

In essence, the parameters in the model-free approach describe the relaxation inducing motions without any physical *a priori* as per their origin but can be later interpreted in a physical basis, for example by explaining the internal motions as resulting from the angular fluctuations around a cone of the N-H bond vector[449].

1.4.5.5 Model-free analysis of intrinsically disordered proteins

Unlike in folded proteins, the concept of an overall and common rotational motion is not relevant for IDPs due do the independence of their motions[450,451].
Assuming that k relaxation active motions are isotropic and statistically independent, the correlation function can be expressed as the sum of exponentially decaying terms representing these k different motions, i.e.

$$C(t) = \sum_k A_k \mathrm{e}^{-t/\tau_k}$$

with A_k representing the amplitudes associated with these motions. Therefore, the sum of their amplitude represents the total amplitude of relaxation inducing motion of the observed independent segment of the IDP chain, that is $\sum_k A_k = 1$. In turn, the spectral density function obtained by Fourier transformation of $C(t)$ is the sum of k Lorentzians:

$$\mathrm{J}(\omega) = \sum_k A_k \frac{\tau_k}{(1 + \omega^2 \tau_k^2)}.$$

1.4.5.5.1 Temperature and chain length dependence of IDP relaxation

In 2016, Abyzov, Salvi and co-workers from the Blackledge group[452] showed that using three components in the above equation accurately reproduced the backbone 15N relaxation of the intrinsically disordered C-terminal domain of the nucleoprotein of the Sendai virus, while only two modes were sufficient for the helix. They identified the physical origins of these three dynamic modes by measuring an extensive set of relaxation

rates over a broad temperature range (from 268 to 298 K) and by performing a model-free analysis of these relaxation rates together on a residue basis using an Arrhenius relationship for the correlation time τ_k as

$$\tau_k(T) = \tau_{k,\infty} \mathrm{e}^{-\frac{E_{ak}}{RT}}$$

for the number of dynamic modes k being comprised between 1 and 3, with $\tau_{k,\infty}$ the correlation time at infinite temperature, E_{ak} the energy of activation of each dynamic mode *k* for each residue, R the universal gas constant and T the temperature. The authors also performed and compared model-free analysis parameters for different chain lengths. They found that the faster-timescale contribution (~ 50 ps) exhibits almost no temperature dependence nor chain length dependence and thus was identified as originating from librational motions. The intermediate motions (~ 0.5 to 1.5 ns) display a flat amplitude and timescale in the disordered regions, deviating only in the helix and termini while showing very limited chain length dependence, suggesting that they report on local backbone sampling, an identification supported by the agreement between the derived energies of activation and the reported and predicted ones[452].

1.4.6 Chemical exchange

The *exchange* of nuclear spins between different environments can be monitored by NMR through its effect on relaxation, whether it is an intramolecular exchange associated with conformational changes or an intermolecular one associated with an interaction with other molecules.

Consider a (pseudo-) first-order chemical reaction or a two-site chemical exchange between two species noted A and B exchanging at a forward rate k_{AB} and a reverse rate k_{BA} such as[412]

$$\frac{d}{dt}\begin{bmatrix}[\mathrm{A}](t)\\ [\mathrm{B}](t)\end{bmatrix} = \begin{bmatrix}-k_{\mathrm{AB}} & k_{\mathrm{BA}}\\ k_{\mathrm{AB}} & -k_{\mathrm{BA}}\end{bmatrix}\begin{bmatrix}[\mathrm{A}](t)\\ [\mathrm{B}](t)\end{bmatrix}$$

with the concentration of the species being written [X] and their fractional population p_X. Their exchange rate is therefore $k_{\mathrm{ex}} = k_{\mathrm{AB}} + k_{\mathrm{BA}}$ with the exchange time constant $\tau_{\mathrm{ex}} = 1/k_{\mathrm{ex}}$. Consequently, the population of a state, e.g. B, is related to p_A by $p_B = 1 - p_A$ and to k_{ex} through $p_{\mathrm{B}} = k_{\mathrm{AB}}/k_{\mathrm{ex}}$.

Throughout this section, the system is considered to be reversible and at equilibrium, i.e. $\mathrm{A} \underset{k_{BA}}{\overset{k_{\mathrm{AB}}}{\rightleftarrows}} \mathrm{B}$ and $p_A k_{\mathrm{AB}} = p_{\mathrm{B}} k_{\mathrm{BA}}$, with the differences in chemical shift $\Delta\omega$ and in transverse relaxation ΔR_2 between the two states being $\Delta\omega = \omega_{\mathrm{B}} - \omega_{\mathrm{A}}$ and $\Delta R_2 = R_{2\mathrm{B}} - R_{2\mathrm{A}}$.

1.4.6.1.1 Bloch-McConnell Equations

Because of the stochastic transitions between the two states, the chemical shift Hamiltonian is time dependent and depends on $\Delta\omega$ such that $\omega(t) = \omega_{\mathrm{A}} + \Delta\omega$ if the observed species is in state B. In the case that $\Delta\omega$ and ΔR_2 are not null, a modified version of the classical Bloch equations discussed earlier called the Bloch-McConnell equations can be used to describe this chemical exchange process [412,418]. This requires the assumptions that the J-couplings of the observed spins are not perturbed by the exchange process, that the forward and reverse rates k_{AB} and k_{BA} do not depend on whether the spins are in state $+\frac{1}{2}$ or $-\frac{1}{2}$ and that the spin state does not change upon crossing the free-energy barrier of the reaction, which is justified as they typically occurs orders of magnitude faster than relaxation even in the overdamped diffusive regime[453–455].

Defining the exchange rate matrix $\widehat{\mathbf{K}}$ with its elements[412,448,456–458]

$$\widehat{\mathbf{K}}_{ij} = \mathrm{k}_{ji} \qquad (i \neq j)$$

$$\widehat{\mathbf{K}}_{ii} = -\sum_{\substack{j=1 \\ j\neq i}}^{N} \mathrm{k}_{ij} \,,$$

which, in the case of two states, yields

$$\widehat{\mathbf{K}} = \begin{bmatrix} -k_{\mathrm{AB}} & k_{\mathrm{BA}} \\ k_{\mathrm{AB}} & -k_{\mathrm{BA}} \end{bmatrix};$$

and the magnetisation vector $\mathbf{M}$ with M^{X} being the net magnetisation of states A and B as

$$\mathbf{M}(\boldsymbol{t}) = \begin{bmatrix} \mathbf{M}^{\mathrm{A}}(\boldsymbol{t}) \\ \mathbf{M}^{\mathrm{B}}(\boldsymbol{t}) \end{bmatrix},$$

the differential equations in the absence of relaxation can be written as

$$\frac{\mathrm{d}\mathbf{M}(\boldsymbol{t})}{\mathrm{dt}} = \widehat{\mathbf{K}}\mathbf{M}(\boldsymbol{t})$$

which can be seen as the exchange rate operator acting on the magnetisation and is equivalent, after separating the cartesian components of the magnetisation, to

$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix} \mathrm{E}/2 \\ M_{\mathrm{x}}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{\mathrm{y}}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{\mathrm{z}}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{\mathrm{x}}^{\mathrm{B}}(\boldsymbol{t}) \\ M_{\mathrm{y}}^{\mathrm{B}}(\boldsymbol{t}) \\ M_{\mathrm{z}}^{\mathrm{B}}(\boldsymbol{t}) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -k_{\mathrm{AB}} & 0 & 0 & k_{\mathrm{BA}} & 0 & 0 \\ 0 & 0 & -k_{\mathrm{AB}} & 0 & 0 & k_{\mathrm{BA}} & 0 \\ 0 & 0 & 0 & -k_{\mathrm{AB}} & 0 & 0 & k_{\mathrm{BA}} \\ 0 & k_{\mathrm{AB}} & 0 & 0 & -k_{\mathrm{BA}} & 0 & 0 \\ 0 & 0 & k_{\mathrm{AB}} & 0 & 0 & -k_{\mathrm{BA}} & 0 \\ 0 & 0 & 0 & k_{\mathrm{AB}} & 0 & 0 & -k_{\mathrm{BA}} \end{bmatrix} \begin{bmatrix} \mathrm{E}/2 \\ M_{\mathrm{x}}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{\mathrm{y}}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{\mathrm{z}}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{\mathrm{x}}^{\mathrm{B}}(\boldsymbol{t}) \\ M_{\mathrm{y}}^{\mathrm{B}}(\boldsymbol{t}) \\ M_{\mathrm{z}}^{\mathrm{B}}(\boldsymbol{t}) \end{bmatrix}$$

The exchange rate matrix $\widehat{\mathbf{K}}$ shows that there is no exchange interconversion between the different cartesian components x, y and z. Neglecting the longitudinal component and focusing only on the transverse magnetisation detected signal gives the complex transverse magnetisation as $\mathrm{M}_{+} = \mathrm{M}_{x} + \mathrm{iM}_{\mathrm{y}}$. Then, adding the matrix $\hat{\mathbf{L}}$ representing the chemical shifts offsets of the different states on the diagonal

$$\hat{\mathbf{L}} = \begin{bmatrix} \Omega_{\mathrm{A}} & \mathbf{0} \\ \mathbf{0} & \Omega_{\mathrm{B}} \end{bmatrix}$$

with Ω_x being the resonance frequencies in the rotating frame, that is the difference between the Larmor frequency of the state x and the frequency at which the RF field is applied, i.e. $\Omega_x = \omega_x - \omega_{RF}$. Finally, the relaxation matrix $\widehat{\mathbf{R}}$ is

$$\widehat{\mathbf{R}} = \begin{bmatrix} \mathrm{R}_{2\mathrm{A}} & \mathbf{0} \\ \mathbf{0} & \mathrm{R}_{2\mathrm{B}} \end{bmatrix}.$$

Altogether, this allows to write the Bloch-McConnell differential equation as[412,448,458]

$$\frac{\mathrm{d}\mathbf{M}(\boldsymbol{t})}{\mathrm{dt}} = \left[-\widehat{\mathbf{R}} + \mathrm{i}\hat{\mathbf{L}} + \widehat{\mathbf{K}}\right]\mathbf{M}(\boldsymbol{t}) \qquad 0.56$$

which is equivalent to

$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix} M_{+}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{+}^{\mathrm{B}}(\boldsymbol{t}) \end{bmatrix} = \begin{bmatrix} -\mathrm{R}_{2\mathrm{A}} + \mathrm{i}\Omega_{\mathrm{A}} - k_{\mathrm{AB}} & k_{\mathrm{BA}} \\ k_{\mathrm{AB}} & -\mathrm{R}_{2\mathrm{B}} + \mathrm{i}\Omega_{\mathrm{B}} - k_{\mathrm{BA}} \end{bmatrix}\begin{bmatrix} M_{+}^{\mathrm{A}}(\boldsymbol{t}) \\ M_{+}^{\mathrm{B}}(\boldsymbol{t}) \end{bmatrix}. \qquad 0.57$$

This Bloch-McConnell equation can be solved analytically for two states, allowing notably to obtain the exact eigenvalues of the matrix, or numerically for all cases, yielding the matrix exponential

$$\begin{bmatrix} M_{+}^{\mathrm{A}}(t) \\ M_{+}^{\mathrm{B}}(t) \end{bmatrix} = \exp\left\{\begin{bmatrix} -\mathrm{R}_{2\mathrm{A}} + \mathrm{i}\Omega_{\mathrm{A}} - k_{\mathrm{AB}} & k_{\mathrm{BA}} \\ k_{\mathrm{AB}} & -\mathrm{R}_{2\mathrm{B}} + \mathrm{i}\Omega_{\mathrm{B}} - k_{\mathrm{BA}} \end{bmatrix} t\right\}\begin{bmatrix} M_{+}^{\mathrm{A}}(0) \\ M_{+}^{\mathrm{B}}(0) \end{bmatrix} \qquad 0.58$$

with $M_{+}^{\mathrm{X}}(0)$ the initial transverse magnetisation of the state X at $t = 0$ which is the total initial magnetisation weighted by the population of the state, i.e. $M_{+}^{\mathrm{X}}(0) = p_{\mathrm{X}}M_{+}(0)$ and $M_{+}^{\mathrm{A}}(0) + M_{+}^{\mathrm{B}}(0) = M_{+}(0)$. Consequently, $M_{+}^{\mathrm{A}}(t) + M_{+}^{\mathrm{B}}(t) = M_{+}(t)$ which is detected in the FID of the system and whose Fourier transformation results in an NMR spectrum.

Fitting the line shapes of the peaks therefore allows to obtain information on the relaxation and exchange rates of the system. Indeed, eq. 0.57 shows the modulation of the transverse relaxation rates R_2 by the exchange process and therefore, the effective R_2^{eff} is the sum of the true R_2 and of the exchange contribution termed R_2^{ex}, i.e. $R_2^{eff} = R_2 + R_2^{ex}$. The exact contribution of R_2^{ex} depends on whether the exchange process is slow ($k_{ex}/\Delta\omega \ll 1$), intermediate ($k_{ex}/\Delta\omega \approx 1$) or fast ($k_{ex}/\Delta\omega \gg 1$) on the chemical shift time scale as well as on the relative distribution of the populations. Figure 2.3 shows simulated ^{1}H spectra obtained from numerical calculations of eq. 0.58 with a homemade Python script, allowing to observe the effects of the different parameters on the peak line shapes and on the chemical shift [412,448,457,458].

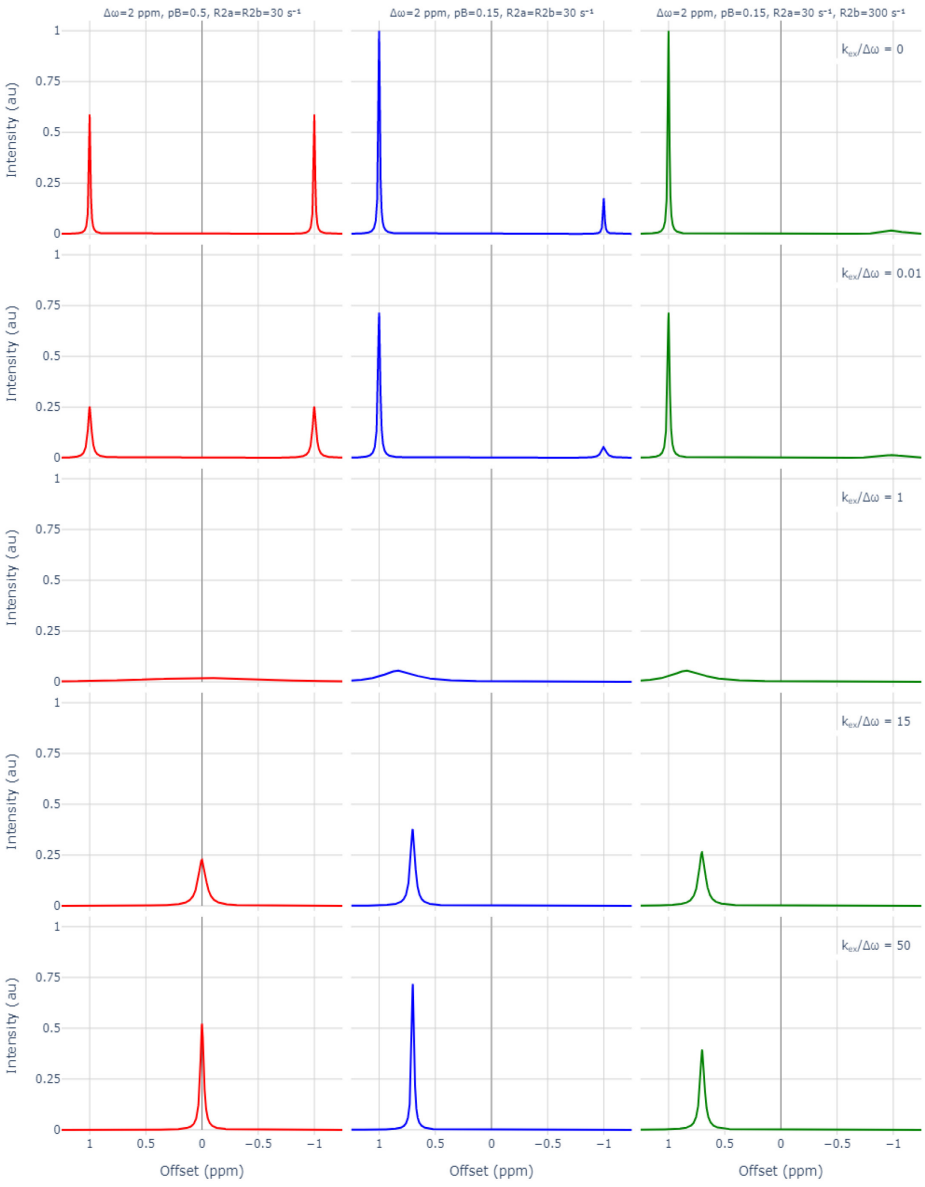


Figure 2.3: Effects of the exchange rates, populations and R_{2B} on the linewidths.
Simulated line shapes of a ^1H nucleus undergoing chemical exchange between two states A and B at 11.74 T with a chemical shift difference of 2 ppm ($\Delta\omega = 7540\ s^{-1}$) calculated using the Bloch-McConnell equations (eq. 2.58) with different $k_{ex}/\Delta\omega$ ratios indicated for each row and with $p_B = 0.5$, $R_{2A} = 30\ s^{-1}$ and $R_{2B} = 30\ s^{-1}$ on the left column (red); $p_B = 0.15$, $R_{2A} = 30\ s^{-1}$ and $R_{2B} = 30\ s^{-1}$ on the centre (blue); and $p_B = 0.15$, $R_{2A} = 30\ s^{-1}$ and $R_{2B} = 300\ s^{-1}$ on the right (green). Note the coalescence of the signal in the intermediate exchange regime (3rd row) and the difference in intensities between the middle and right columns in the fast exchange regime (5th row) due to the high R_{2B}.

1.4.6.1.2 Slow exchange

In the slow exchange regime, usually defined as $k_{\mathrm{ex}} \ll \Delta\omega$, the off-diagonal elements of the matrix of eq. 0.57, that is k_{BA} and k_{AB}, can be neglected because they are very small, as $k_{\mathrm{ex}} = k_{\mathrm{AB}} + k_{\mathrm{BA}}$. This yields the solution[412,458]

$$M_+(t) = M_+^{\mathrm{A}}(0)\mathrm{e}^{(\mathrm{i}\Omega_{\mathrm{A}} - \mathrm{R}_{2\mathrm{A}} - k_{\mathrm{AB}})t} + M_+^{\mathrm{B}}(0)e^{(\mathrm{i}\Omega_{\mathrm{B}} - \mathrm{R}_{2\mathrm{B}} - k_{\mathrm{BA}})t}.$$

This means that in the slow exchange regime, two peaks corresponding to A and B evolve independently at Ω_{A} and Ω_{B}, have an intensity proportional to their fractional populations p_{A} and p_{B} and have an $\mathrm{R}_2^{\mathrm{ex}}$ contribution to their linewidths being a broadening respectively equal to $\mathrm{R}_{2\mathrm{A}}^{\mathrm{ex,slow}} = k_{\mathrm{AB}}$ and $\mathrm{R}_{2\mathrm{B}}^{\mathrm{ex,slow}} = k_{\mathrm{BA}}$. This can be seen in Figure 2.3 on the second row from the top for equal populations (left) with $p_{\mathrm{A}} = p_{\mathrm{B}} = 0.5$ and skewed populations (centre) with $p_{\mathrm{B}} = 0.15$. Indeed, because the exchange is slow, τ_{ex} is long and therefore by the time a species originally in the A state jumps to the B state and comes back to the A state again, a significant time has passed and it has gone out of phase with the other species that stayed in state A.
Consequently, because $p_{\mathrm{B}} = k_{\mathrm{AB}}/k_{\mathrm{ex}}$, that $p_{\mathrm{A}} \gg p_{\mathrm{B}}$ implies $k_{\mathrm{BA}} \gg k_{\mathrm{AB}}$ and therefore, in the slow-to-intermediate exchange regime, the already smaller population weighted peak is impacted by even more severe broadening than the major peak, rendering it difficult to observe directly. This is seen by comparing the two top rows in the centre column of Figure 2.3.

1.4.6.1.3 Intermediate exchange

The above-described effect grows noticeably as k_{ex} approaches $\Delta\omega$, with correspondingly broader linewidths for smaller differences between them. In the conditions where $k_{\mathrm{ex}} < \Delta\omega$, which is sometimes referred to as the slow-intermediate regime, both peaks are significantly broadened until they ultimately coalesce and are essentially not observable

anymore when $k_{\mathrm{ex}} \approx \Delta\omega$, which is usually defined as the intermediate exchange regime[412,448,456,458]. This can be observed in the third row of Figure 2.3.

1.4.6.1.4 Fast exchange

In the fast exchange regime, defined as $k_{\mathrm{ex}} \gg \Delta\omega$, eq. 0.58 is simplified to[412,456–458]

$$M_+(t) \propto e^{\left(\mathrm{i}\bar{\Omega} - \bar{\mathrm{R}}_2 - \left(p_{\mathrm{A}} p_{\mathrm{B}} \Delta\omega^2 / k_{\mathrm{ex}}\right)\right)t}$$

with the population-weighted shift and transverse relaxation rate being respectively $\bar{\Omega} = p_{\mathrm{A}}\Omega_A + p_{\mathrm{B}}\Omega_B$ and $\bar{\mathrm{R}}_2 = p_{\mathrm{A}}\mathrm{R}_{2A} + p_{\mathrm{B}}\mathrm{R}_{2B}$. This means that a single peak is observed at a population-weighted average of the two original positions, i.e. $\omega_{obs} = p_{\mathrm{A}}\omega_A + p_{\mathrm{B}}\omega_B$ with a linewidth determined by $\bar{\mathrm{R}}_2$ by an additional exchange contribution $\mathrm{R}_2^{\mathrm{ex,fast}} = p_{\mathrm{A}} p_{\mathrm{B}} \Delta\omega^2 / k_{\mathrm{ex}}$ that becomes smaller, and thus the linewidth narrower, as k_{ex} increases, as seen in 4th and 5th row of Figure 2.3. Indeed, as opposed to the slow exchange case, because τ_{ex} gets increasingly smaller, the species interconvert so fast that there is no time for phase loss to build up as the chemical shift evolves, hence a single peak is observed[412,458]. However, because R_{2B} contributes to $\bar{\mathrm{R}}_2$, if R_{2B} is significantly high, even a small population would lead to significant loss of intensity of the single peak. This can be seen by comparing the bottom row of centre and right column of Figure 2.3. that have the same $p_{\mathrm{B}} = 0.15$ but respectively $\mathrm{R}_{2B} = 30\ \mathrm{s}^{-1}$ and $\mathrm{R}_{2B} = 300\ \mathrm{s}^{-1}$, which results in significantly different intensities and linewidths.

1.4.6.1.5 Exchange with no chemical shift difference

Although rare, it is possible that there is no chemical shift difference between the two states, i.e. $\Delta\omega = 0$, in which case the transverse relaxation exchange contribution is[458]

$$\mathrm{R}_2^{\mathrm{ex},\Delta\omega=0} \approx \frac{k_{\mathrm{AB}} \Delta\mathrm{R}_{2\mathrm{AB}}}{(k_{\mathrm{BA}} + \Delta\mathrm{R}_{2\mathrm{AB}})}$$

which shows that such a case causes line broadening. If $k_{\mathrm{ex}} \gg \Delta\mathrm{R}_{2\mathrm{AB}}$, the observed $\mathrm{R}_2^{\mathrm{eff}}$ decays in a similar way to the fast exchange case, that is $\mathrm{R}_2^{\mathrm{eff}} = p_{\mathrm{A}}\mathrm{R}_{2A} + p_{\mathrm{B}}$. However, if $k_{\mathrm{ex}} \ll \Delta\mathrm{R}_{2\mathrm{AB}}$, then in a conceptual equivalent to the slow exchange case, by the time the species originally in the A state returns to it after passing through the B state, its magnetisation will already have relaxed and will not contribute to the magnetisation of the A state. Therefore, the observed $\mathrm{R}_2^{\mathrm{eff}}$ is given by $\mathrm{R}_2^{\mathrm{eff}} = \mathrm{R}_{2\mathrm{A}} + k_{\mathrm{AB}}$.

1.5 Scattering

1.5.1 The wave scattering phenomenon

Maxwell's equations describe light as an electromagnetic wave oscillating in both space and time, with its oscillation direction being perpendicular to its propagation direction. As an incident electric field of frequency ν_i and wavelength λ, the incident light described by an angular wavevector $\mathbf{k}_i = 2\pi\nu_i = 2\pi/\lambda$ that interacts with the electric moments of particles it encounters, effectively distorting their charges distribution and causing the particles to emit a secondary, *scattered* wave $\mathbf{k}_\mathrm{s}$ to relax to equilibrium. If no exchange of energy between the photons and the particles occurs, the angular frequency and wavelength of both waves is the same and therefore so is their magnitude, i.e. $\mathrm{k_s} = \mathrm{k}_i$. This scattering is said to be *elastic*, while if energy is exchanged, i.e. $\nu_i \neq \nu_s$, the scattering is *inelastic* and therefore, $\mathrm{k_s} \neq \mathrm{k}_i$[459–462].

The scattering angle can be defined at 2θ while the scattering wavevector $\boldsymbol{q}$ is the difference between the scattered and the incident wave such as $\boldsymbol{q} = \mathbf{k}_\mathrm{s} - \mathbf{k}_\mathrm{i}$. In the case of an elastic scattering of a monochromatic incident beam, i.e. of a fixed wavelength λ, then $\mathbf{k}_\mathrm{i} = \mathbf{k}_\mathrm{s} = 2\pi/\lambda$ and the magnitude of the scattered angular wavevector q is given by[459–462]

$$q = 2\mathrm{k}_i \sin(\theta) = \frac{4\pi}{\lambda}\sin(\theta) \qquad 0.59$$

and its SI units are nm^{-1} although it is frequent to see q expressed in $\mathrm{Å}^{-1}$. Some books and articles divide θ by 2 but the form of the equation given above is the most widely used.

As apparent from eq. 0.59, the magnitude of the wavevector q depends on the inverse of λ, meaning that the longer the wavelengths, the smaller q and therefore the smaller the details can be resolved and the bigger the object can be probed. Thus, the scattering of X-Rays, which have a wavelength on the order of the a tenth of nm, will report on electron distributions, while visible light laser scattering, with a wavelength of hundreds of nm, will instead report on the distribution of individual macro-molecules. Together, they provide a complementary view for the study of proteins, polymers or colloids.

1.5.2 Small-Angle X-Ray Scattering

Small-Angle X-ray scattering (SAXS) is technique that consists in measuring the angle-dependent scattering intensity arising from the exposure of a sample in solution to an X-ray beam. Indeed, proteins, polymers or colloids have inhomogeneous electron density, and thus, those electrons resonate with the passing X-ray beam frequency to emit interfering secondary waves. The intensity of the scattered X-rays depends on the spatial distribution of the electron densities within the sample as well as the relative phases of the secondary waves due do their possible constructive or destructive interference[461,462].

1.5.2.1 Why small angles?

As scattering follows the reciprocity law described in eq. 0.59, there is an inverse relationship between the dimensions of the studied object and the scattering angle, hence the need to measure at small-angles due to the much bigger size of proteins with respect to the X-ray beam. In turn, this led to the development of ultra-small angle detection (USAXS) to study large assemblies of proteins or polymers.

Conveniently, the incoherent Compton scattering arising from the release of loosely bound electrons by the X-Ray beam is negligible at small angles, therefore we consider that the observed scattered waves are coherent, which implies that the relative position of the scattering electron can be derived from it. As the outer electrons are more likely to participate in Compton scattering due to their higher possibility of being released from the atom and to carry momentum by acquiring energy from the beam, their scattering probability increases with the scattering angle 2θ as well, and therefore, the observed scattering intensity is particularly noisy at higher q. Other sources of noise at high q include the contribution of both coherent and incoherent scattering from the buffer as well as inhomogeneities in the sample such as bubbles[459,461].

The following section describes the coherent X-ray scattering phenomenon known as Thompson scattering for diffusing particles in a solvent occurring in SAXS[459,461,463]. The related Thompson scattering case for a crystal is treated in the dedicated section 1.5.3.2 page on page 100.

As an X-Ray beam illuminates a sample made of N particles, they each scatter individually according their q-dependant atomic scattering factor $f_i(q)$. The q-dependent amplitude of the scattered wave in the reciprocal space $A(\boldsymbol{q})$ is therefore

$$A(\boldsymbol{q}) = \sum_{j=1}^{N} f_j(q) \exp\left(-i\boldsymbol{q} \cdot \boldsymbol{r_j}\right). \qquad 0.60$$

However, as the sample is made of a continuous distribution of these particles in a solvent, scattering occurs and can be quantified per unit of volume at a position **r** and noted as the electron density $\rho(\boldsymbol{r})$. The relative electron density $\Delta\rho(\boldsymbol{r})$ is therefore the difference scattering amplitude between the solvent and the particles, i.e. $\Delta\rho(\boldsymbol{r}) = \rho(\boldsymbol{r}) - \rho_{solvent}$. Thus, $A(\boldsymbol{q})$ is the Fourier transform of $\Delta\rho$, that is

$$A(\boldsymbol{q}) = \int \Delta\rho(\boldsymbol{r}) \exp(-i\boldsymbol{q} \cdot \boldsymbol{r})\, d\boldsymbol{r} \qquad 0.61$$

The amplitude cannot be directly measured but the intensity can and is determined by the spatial distribution of electron densities within the sample, as well as the relative amplitude and phase of each individual scattered wave. Indeed, constructive interference amplifies the scattering intensity when the scattered waves are in phase, while out-of-phase waves lead to destructive interference, resulting in reduced scattering intensity. Thus, the intensity is

$$I(\boldsymbol{q}) = A(\boldsymbol{q})A^*(\boldsymbol{q}) \qquad 0.62$$

The self-convolution of the relative electron density $\Delta\rho(\boldsymbol{r})$ defines the autocorrelation function $\gamma(\boldsymbol{r})$ which therefore describes the overlap or "shifted volume" between a particle and its "ghost", i.e. itself translated by a vector $\boldsymbol{u}$, that is

$$\gamma(\boldsymbol{r}) = \Delta\rho(\boldsymbol{r}) * \Delta\rho(-\boldsymbol{r}) = \int \Delta\rho(\boldsymbol{r}+\boldsymbol{u})\Delta\rho(\boldsymbol{r})\, d\boldsymbol{r}\,. \qquad 0.63$$

In turn, putting together the three previous equations allows to express the intensity results as a function of $\gamma(\boldsymbol{r})$ as

$$I(\boldsymbol{q}) = \int \gamma(\boldsymbol{r}) \exp(-i\boldsymbol{q} \cdot \boldsymbol{r})\, d\boldsymbol{r}\,.$$

It is worth noting that the autocorrelation function $\gamma(\boldsymbol{r})$ is effectively equivalent to a continuous version of the discrete Patterson function in X-Ray crystallography (cf. section 1.5.3.3.1 page 104), and its definition as the excess scattering over the background explains why, to access it, the solvent signal has to be subtracted from the measured data. If the scattering is isotropic, by noting that the ensemble-average of $\exp(-i\boldsymbol{q} \cdot \boldsymbol{r})$ over all

possible orientations is $\frac{\sin(qr)}{qr}$ and integrating in spherical coordinates, the ensemble-averaged isotropic intensity of the scattered wave can be written as

$$I(q) = 4\pi \int_0^{D_{max}} r^2 \gamma(r) \frac{\sin(qr)}{qr} dr \qquad 0.64$$

with D_{max} denoting the maximum distance r from the geometric centre of the scattering, which effectively corresponds to the maximum dimension of the scattering particle. It is used to limit the range of integration because, as the dimensions of a particle are physically limited.

1.5.2.2 Pair distance distribution electron

The term $r^2\gamma(r)$ in eq. 0.64 is the r^2 weighted autocorrelation function, which physically corresponds to the probability density of finding a pair of electrons at a given distance r from each other, with D_{max} being the maximum possible distance. Defining this the pair electron distance distribution function $p(r) = r^2\gamma(r)$ allows to rewrite eq. 0.64 as[459,461,463].

$$I(q) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin(qr)}{qr} dr\,. \qquad 0.65$$

When $q = 0$, the above equation reduces to

$$I(0) = 4\pi \int_0^{D_{\max}} p(r)\, dr \qquad 0.66$$

which can be interpreted as all the electrons scattering in coherence with their surroundings represented by the autocorrelation function $\gamma(r)$ in $p(r)$, i.e. all individual scattered waves being in phase[461].

It is apparent from eq. 0.65 that the measured signal is simply the Fourier transform of the $p(r)$. Indeed, the $p(r)$ function is fundamental to SAXS as it reports on the spatial correlation of the studied materials, that is its molecular arrangement and shape and thus governs the scattering. As such, the $p(r)$ function can be seen as a radially averaged equivalent of the Patterson function with peaks at the origins and at positions corresponding to pure crystallographic translations (cf. section 1.5.3.3.1 page 104).

However, because $I(q)$ is only known over the truncated range of q measured and is affected by experimental errors, it is an ill-conditioned problem, which is why the integration is limited by $D_{\max}$ and is solved by fitting the indirect Fourier transform of eq. 0.65[464,465].

As a distance probability distribution function, the aspect of the experimental $p(r)$ heavily depends on the geometric shape of the object and therefore allows straightforward estimation about its geometry while fitting procedures permit to gain precious insights about the dimensions of the object.

1.5.2.3 Radius of gyration

The radius of gyration R_g of an object is defined as the root-mean square distance between the parts of an object from its centre of gravity, and as such, corresponds to the root-mean square distance of all electron from its centre of gravity. The R_g therefore gives a geometrical measure of the size of the object. It can be obtained by different methods in SAXS[463,466].

A first method is the Guinier approximation[461,466,467]. Coming back to eq. 0.65 expressing the q-dependent intensity as a function of the pair distribution function $p(r)$ and plugging in the MacLaurin expansion $\frac{\sin(qr)}{qr} \approx 1 - \frac{(qr)^2}{3!} + \cdots$, if the values q are small enough to permit the truncation of the expansion to the second term, the scattered intensity can be expressed as

$$I(q) = 4\pi \int_0^{D_{max}} p(r)\left(1 - \frac{(qr)^2}{3!}\right) dr \,.$$

Using the definition of $I(0)$ in eq. 0.66 in the above equation yields

$$I(q) = I(0)\left(1 - \frac{(qr)^2}{3!}\right)$$

which allows, using again the MacLaurin expansion $e^x \approx 1 + x + \frac{x^2}{2!} + \cdots$ with x=$q^2R_g^2$, to obtain the Guinier equation

$$I(q) = I(0)\exp\left(-\frac{q^2R_g^2}{3}\right).$$

Therefore, the plot of ln $I(q)$ against q^2 is linear at small q where this approximation holds and has a slope of $R_g^2/3$ which is commonly used to obtain gyration radii for monodisperse samples, although other expressions can also be used depending on the shape of the particles if it is known.

However, the intensities at the smallest q are sensitive to the presence of even small oligomers or aggregates, as the reciprocal space relationship states that the smaller q is, the bigger the scattering objects that can be probed. Moreover, an attractive or repulsive regime also alters the shape of the curve at these small q values. The correct range of q values to use during a Guinier analysis has therefore to be carefully chosen to avoid over- or under-estimating the radius of gyration which should be compared to the one obtained from fitting the $p(r)$.

A second method for determining the radius of gyration of an object in SAXS is to use the pair distribution function. Indeed, R_g can be defined as the second moment of the pair density $p(r)$ around the centre of its scattering density distribution, yielding

$$R_g^2 = \frac{1}{2}\frac{\int_0^{D_{max}} p(r)r^2\,dr}{\int_0^{D_{max}} p(r)\,dr}. \qquad 0.67$$

Unlike the Guinier approximation, this equation makes use of the whole scattering curve instead of being limited to the small q regions where the curve is linear, and as such, is less sensitive to the presence of interactions or of a small fraction of oligomers. However, as it is accessed from $p(r)$ which is in turn accessed from a fitting procedure, the accuracy of this radius of gyration depends on the accuracy of the fitting. It is therefore common to use both methods and to compare the obtained gyration radii.

1.5.2.4 Distinguishing compact and flexible objects

1.5.2.4.1 Porod-Debye plot

The Porod[468] or Porod-Debye[469] regime is the scattering profile at high q which is dependent on the particular shape and surface-to-volume ratio of compact, geometrically

shaped objects with a sharp interface with the solvent. Indeed, the scattering of such objects at high q is given by[461,468,469]

$$\lim_{q \longrightarrow \infty} I(q) = \frac{2\pi(\Delta\rho)^2}{q^4} \frac{S}{V}.$$

With S the surface and V the volume of the object. As the scattering $I(q)$ scales with q^{-4}, which is known at Porod's law, plotting $I(q)q^4$ versus q^4 allows to observe a plateau at high q if the particles follow this law, which also allows to determine their surface-to-volume ratio. This is the Porod-Debye plot. In turn, the absence of such a plateau indicates the presence of flexibility or a particularly curved or fractal-like structure whose profile in such a plot at high q is approximately linear[459,461]. Thus, such representations allow to readily distinguish globular from intrinsically disordered proteins.

1.5.2.4.2 Kratky plots

The Kratky plot represents $q^2I(q)$ versus q and as such, highlights q-dependent deviations of the scattering intensity.

It allows to distinguish globular or proteins with a well-defined fold which show a bell-shaped profile, as the scattering intensity starts to decay at approximately $I(q) \approx q^{-4}$ due to their following of Porod's law. On the other hand, a fully disordered protein or ideal Gaussian chain polymer shows a plateau at high q due to the independence of their scattering from q values in the Porod regime which effectively translates to a plateau at $I(q) \approx q^{-2}$. Finally, a partially unfolded protein presents a profile in the Kratky-plot in-between these two cases, with the presence of a peak at small q followed by a plateau at higher q[459,470].

1.5.3 X-Ray Diffraction crystallography

In a crystal, the atoms are arranged in a periodic manner and therefore, upon exposition to an incident X-Ray beam, their scattered waves interfere with each other either constructively or destructively, which gives rise to a particular pattern that depends on the arrangement of these atoms. This is the *diffraction* phenomenon and therefore, its study allows to deduce the physical arrangement of these atoms, making X-Ray Diffraction (XRD) crystallography one of the most powerful and popular structural biology technique.

1.5.3.1 Crystals

Crystals are three-dimensional solids composed periodic repeats of an arrangement, which can consist of a single atom, molecules or several molecules forming its *lattice*, that is a three-dimensional array of points. The smallest identical repeating structure in this lattice is its *unit cell* and as such, the entire crystal can be described by a set of translation and rotation operations of the unit cell, repeating it infinitely in space. As a geometric unit, the unit cell is therefore defined by its length of its edges, noted a, b and c, as well as the angles between its axes, noted α, β and γ. However, the unit cell of a crystal typically contains several redundant structures linked by symmetry operations which are classified in different space groups. This smallest, non-redundant, pattern of a crystal is its *asymmetric unit* and applications of these symmetry operation generate the unit cell. Finally, in the context of protein crystals, the *biological assembly* refers to the functional unit of the protein and may be the same as the asymmetric unit, but may also be only a portion of it or, conversely, be composed of multiple asymmetric units[471–473].

In the following section, a unique crystal is considered infinite and without defects.

1.5.3.2 Diffraction by crystals

While SAXS or DLS notations use the angular scattering wavevector q defined in eq. 0.59, X-Ray diffraction crystallography notations typically use an ordinary wavevector **s** whose magnitude s is defined as

$$\mathrm{s} = \frac{2\sin(\theta)}{\lambda} = \frac{\mathrm{q}}{2\pi}.$$

Unlike the previously described SAXS experiment, X-Ray Diffraction (XRD) crystallography is not typically measured in solution and therefore, the position-dependent electron density $\rho(\boldsymbol{r})$ can be directly used to describe the scattering instead of the above-solvent relative electron density $\Delta\rho(\boldsymbol{r})$. Additionally, in the spherical polar coordinates centred around the atom, the electron cloud can be considered centrosymmetric around the atomic nucleus and therefore, the value of $\rho(\boldsymbol{r})$ is independent from the polar and azimuthal angles, i.e. $\rho(\boldsymbol{r}) = \rho(-\boldsymbol{r})$. As such, using s instead of q and $\rho(\boldsymbol{r})$ instead of $\Delta\rho(\boldsymbol{r})$ in eq. 0.61 yields the s-dependent amplitude of the scattered wave $A(\mathbf{s})$ for a single atom or electron scattering factor or form factor f_e is [471–473]

$$A(\mathbf{s}) = f_e = \int \rho(\boldsymbol{r}) \exp(2\pi i \mathbf{s}\cdot\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \qquad 0.68$$

and therefore, eq. 0.65 can be rewritten as a function of either s or q as

$$f_e = 4\pi \int_0^\infty r^2\rho(r)\frac{\sin(2\pi s r)}{2\pi s r}\mathrm{d}r = 4\pi \int_0^\infty r^2\rho(r)\frac{\sin(qr)}{qr}\mathrm{d}r\,. \qquad 0.69$$

Compared to eq. 0.65, $\rho(r)$ is directly used instead of the autocorrelation function $\gamma(r)$ because the particles of the crystal do not diffuse in a solvent and therefore, the scattering is function of their own electron density instead of their overlap, which is also why the integration does not have to be limited by the dimension of the particle.

In turn, the form factor f, also called the total X-ray scattering factor, of the crystal is the sum of the scattering f_e of all j atoms constituting it, that is[471–473]

$$f = \sum_j f_e\,. \qquad 0.70$$

As the crystals is made of regular repeats of a unit cell comprising j different atoms, their scattering can be described in the unit cell frame by a vector $\boldsymbol{f_j}$ such as

$$\boldsymbol{f_j} = f_e \exp(2\pi i \mathbf{s}\cdot\boldsymbol{r})$$

with the term $2\pi\mathbf{s}\cdot\boldsymbol{r}$ therefore representing the change in the phase angle of the scattered wave due to the frame origin being transferred from the atomic nucleus to the origin of

the unit cell[471–473]. In turn, the total scattering of a unit cell is the sum of the scattering of the different atom j being present in the cell and because it depends on the precise physical disposition of these atom, it is called the structure factor $\mathbf{F}(\mathbf{s})$, i.e.

$$\mathbf{F}(\mathbf{s}) = \sum_{j=1} f_e \exp\left(2\pi i \mathbf{s} \cdot \boldsymbol{r_j}\right).$$

It is noteworthy that as the electron density $\rho(r)$ is a real valued function and the scattering is normally symmetric due to the centrosymmetry of the electron-cloud, the structure factors at the wavevector $\mathbf{s}$ and $-\mathbf{s}$ are complex conjugate of each other's, i.e. $\mathbf{F}(\mathbf{s}) = \mathbf{F}(-\mathbf{s})^*$ and thus, that they have the same magnitude[471–473].

Finally, the crystal cell symmetry can be described in term of translation vector **a**, **b** and **c** necessary to pass from one unit cell to the other in the arbitrary directions a,b,c in which it contains respectively n_a, n_b and n_c unit cells whose positions can be described in term of coordinates t, u and v with regards to an origin arbitrarily chosen as the origin of any unit cell. The total wave scattered by the crystal $\mathbf{K}(\mathbf{s})$ is therefore the structure factor $\mathbf{F}(\mathbf{s})$ representing the scattering of each individual unit cell multiplied by the summation over every constituting unit cells $n_{a,b,c}$ in the a,b,c directions, ergo[471–473]

$$\mathbf{K}(\mathbf{s}) = \mathbf{F}(\mathbf{s}) \times \sum_{\boldsymbol{t=0}}^{n_a} \exp(2\pi i t \mathbf{s} \cdot \boldsymbol{a}) \times \sum_{\boldsymbol{u=0}}^{n_b} \exp(2\pi i u \mathbf{s} \cdot \boldsymbol{b}) \times \sum_{\boldsymbol{v=0}}^{n_c} \exp(2\pi i v \mathbf{s} \cdot \boldsymbol{c}).$$

In the case where the phase angle shifts are not multiple of 2π, i.e. if $\mathbf{s} \cdot \boldsymbol{a}$, $\mathbf{s} \cdot \boldsymbol{b}$ and $\mathbf{s} \cdot \boldsymbol{c}$ are not integers, the different scattered wavevectors point in different directions in the complex plane and consequently, will destructively interfere with each other, causing the overall sum, that is the diffraction at that **s** value, to tends towards zero. In other terms, the diffraction pattern will only show significant spots, also called reflections, when

$$\begin{cases} \mathbf{s} \cdot \boldsymbol{a} = h \\ \mathbf{s} \cdot \boldsymbol{b} = k \\ \mathbf{s} \cdot \boldsymbol{c} = l \end{cases} \qquad 0.71$$

with h, k and l being integers. These are known as the Laue conditions whose fulfilment result in the in-phase scattering of all the unit cells and therefore, yields the particular structure factor-dependent observed diffraction pattern. In turn, the h, k and l values are known as the Miller indices which identify the individual reflections and therefore allow their indexing[471–473].

1.5.3.3 The phase problem

The unit cell can be described by its volume V and by fractional coordinates x, y and z in the real space identifying the relative position within the cell, i.e. $0 \leq x, y, z < 1$. The structure factor $\mathbf{F}(\mathbf{s})$ can therefore be also expressed as an integration of the electron density over the unit cell in a similar way to the electron scattering factor in eq. 0.68 as

$$\mathbf{F}(\mathbf{s}) = \int_{\text{unit cell}} \rho(\boldsymbol{r}) \exp(2\pi i \boldsymbol{s} \cdot \boldsymbol{r})\, \mathrm{dv}$$

with $\mathrm{dv} = \mathrm{V} \cdot \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$. In turn, using again the translation vectors **a**, **b** and **c** instead of $\boldsymbol{r}$ to describe the coordinates, together with eq. 0.71, yields[471–473]

$$\begin{aligned} \mathbf{s} \cdot \boldsymbol{r} = \mathbf{s} \cdot (\boldsymbol{a} \cdot x + \boldsymbol{b} \cdot y + \boldsymbol{c} \cdot z) &= \mathbf{s} \cdot \boldsymbol{a} \cdot x + \mathbf{s} \cdot \boldsymbol{b} \cdot y + \mathbf{s} \cdot \boldsymbol{c} \cdot x \\ &= hx + ky + lz \end{aligned} \qquad 0.72$$

and thus, the **s**-dependent structure factor $\mathbf{F}(\mathbf{s})$ can be rewritten as a function of h,k,l as

$$\mathbf{F}(\mathrm{h\,k\,l}) = \mathrm{V} \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(\mathrm{x\,y\,z}) \exp[2\pi \mathrm{i}(hx + ky + lz)]\, dx\, dy\, dz\,. \qquad 0.73$$

This equation describes the structure factor $\mathbf{F}(\mathrm{h\,k\,l})$ as the Fourier transform of the electron density $\rho(\mathrm{x\,y\,z})$ and therefore, the inverse Fourier transform of eq. 0.73 gives

$$\rho(\mathrm{x\,y\,z}) = \frac{1}{\mathrm{V}} \sum_{h} \sum_{k} \sum_{l} \mathbf{F}(\mathrm{h\,k\,l}) \exp[-2\pi \mathrm{i}(hx + ky + lz)].$$

In turn, expressing the structure factor $\mathbf{F}(\mathrm{h\,k\,l})$ as dependent on its magnitude $F(\mathrm{h\,k\,l})$ and its phase angle α yields the final expression of the electron density of the unit cell as

$$\rho(\mathrm{x\,y\,z}) = \frac{1}{\mathrm{V}} \sum_{h} \sum_{k} \sum_{l} F(\mathrm{h\,k\,l}) \exp[-2\pi \mathrm{i}(hx + ky + lz) + \mathrm{i}\alpha(h\,k\,l)]. \qquad 0.74$$

This equation describes the electron density ρ for every coordinate x, y and z of the real space unit cell. This includes the electron density of the crystalised molecule and therefore allows the determination of its structure by modelling its obtained electron density[471–473].

Practically, the experimentally obtained h,k,l diffraction spots intensity $I(\mathrm{h\,k\,l})$ directly depends on the magnitude of the structure factor which, similarly to eq. 0.62, is equal to the product of the structure factor by its complex conjugate, i.e.

$$I(\mathrm{h\,k\,l}) \propto F(\mathrm{h\,k\,l})^2 = \mathbf{F}(\mathrm{h\,k\,l})\mathbf{F}^*(\mathrm{h\,k\,l}) \qquad 0.75$$

which thereby permits the estimation of $F(\mathrm{h\,k\,l})$**.** As previously discussed, the magnitude of structure factors of opposite wavevector **s**, and therefore opposite h,k,l indices, are equal which implies that their intensities is equal, i.e. $I(\mathrm{h\,k\,l}) = I(-\mathrm{h}\ -\mathrm{k}\ -\mathrm{l})$. This property is called Friedel's law[471–473].

In turn, the position of these h,k,l reflections is explained by the Laue conditions as function of the wavevector **s**, which in turn allows the estimation the volume of the unit cell V via the determination of its dimension. These correspond to the two first terms of this equation. However, the phase angle α cannot be measured by the detector and therefore this phase information is lost, which effectively prevents the calculation of the electron density $\rho(\boldsymbol{x\ y\ z})$ via the above equation. This is known as the phase "phase problem" of crystallography[471–473].

1.5.3.3.1 The Patterson function and method

The self-convolution of the electron density $\rho(\mathbf{u})$ yields an autocorrelation function called the Patterson function and noted $P(\mathbf{u})$, with $\boldsymbol{u}$ a vector representing any possible interatomic distance within the unit cell of volume V. This yields

$$P(\mathbf{u}) = \rho(\boldsymbol{r}) * \rho(-\boldsymbol{r}) = \int_V \rho(\boldsymbol{r}+\mathbf{u})\rho(\boldsymbol{r})\, d\boldsymbol{r}$$

which is the XRD crystallography equivalent of the SAXS autocorrelation function $\gamma(\boldsymbol{r})$ defined in eq. 0.63 in the dedicated section page 95. Moreover, as seen in eq. 0.75, the experimentally detected intensity of the h, k, l diffractions spots directly depend on the squared magnitude of the structure factor $F(\mathrm{h\,k\,l})^2$ whose inverse Fourier transform noted $\boldsymbol{FT^{-1}}$ is[471–473]

$$\boldsymbol{FT^{-1}}[F(\mathrm{h\,k\,l})^2] = \boldsymbol{FT^{-1}}[\mathbf{F}(\mathrm{h\,k\,l})\mathbf{F}^*(\mathrm{h\,k\,l})] = \rho(\boldsymbol{r}) * \rho(-\boldsymbol{r}) = P(\mathbf{u})$$

using eq. 0.63 which establishes the structure factor $\mathbf{F}(\mathrm{h\,k\,l})$ as the Fourier transform of the electron density $\rho(\mathrm{x\,y\,z}) = \rho(\boldsymbol{r})$**.** In other words, the inverse Fourier transform of $F(\mathrm{h\,k\,l})^2$, and thus of the intensity of the diffraction spots, is the Patterson function. Therefore, defining the vector $\boldsymbol{u}$ as $\boldsymbol{u} = \boldsymbol{a}\cdot u + \boldsymbol{b}\cdot v + \boldsymbol{c}\cdot w$ allows to write

$$P(\mathbf{u}) = \frac{1}{\mathrm{V}}\sum_h\sum_k\sum_l F(\mathrm{h\,k\,l})^2 \exp[-2\pi\mathrm{i}(\mathrm{hu} + \mathrm{kv} + \mathrm{lw})]\ .$$

However, the electron density $\rho(\boldsymbol{r})$ of the atoms that scatters the X-Rays is a periodic function of the position $\boldsymbol{r}$ due to the periodic spatial locations of these atoms within the

lattice and therefore the Patterson function can be more conveniently written as a Fourier series with the squared structure factor magnitude $F(\mathrm{h\,k\,l})^2$, representing the detected intensity of the h k l, as the coefficient of the series and with the phase angle fixed to 0, yielding[471–473]

$$P(\mathbf{u}) = \frac{1}{\mathrm{V}} \sum_h \sum_k \sum_l F(\mathrm{h\,k\,l})^2 \cos[2\pi(\mathrm{hu} + \mathrm{kv} + \mathrm{lw})] \,.$$

The Patterson function can therefore be calculated from the sole intensity of the reflections without any phase information. Nevertheless, it still provides precious information. Indeed, the obtained Patterson map shows peaks at the endpoint of vectors **u** equals to real space distances between atoms in the unit cell, while it is otherwise null. Of course, the highest peak in the centrosymmetric map is always the origin as it corresponds to the vector between every atom and itself. As for the other peaks, their height is directly proportional to the product of the atomic numbers of the two atoms of that particular vector **u**, thereby allowing to locate the heaviest atoms, determine the distances with neighbouring atoms and therefore, to approximate the phase angle. However, as the number of atomic positions is unfortunately much smaller than the number of interatomic vectors, the Patterson map is typically highly overlapped and it is therefore not straightforward to interpret. It is therefore convenient to cut the Patterson map close to the origin in order to observe only intramolecular vectors **u** and not the ones describing distances between different molecules. Such map is called self-Patterson map and these intramolecular vectors, self-Patterson vectors while the intermolecular are called cross-Patterson vectors[471–473].

1.5.3.3.2 Isomorphous replacement and anomalous dispersion methods

In order to retrieve this lost phase information, two strategies are typically used: experimental phasing and molecular replacement.

Experimental phasing can be performed using two different methods. The first one is the isomorphous replacement method which is realised by adding a single (SIR) or multiple (MIR) heavy atoms to the crystallised protein and comparing the obtained dataset to the native one. Due to their high diffraction power, these heavy atoms can be localised on the Patterson map and these positional information can in turn be used to obtain the phase

of the heavy-atom derived crystal diffraction dataset, which can finally be used to derive the phase of the native crystal dataset by comparison with the native Patterson map[474,475].

The second experimental phasing method make use of anomalous dispersion, whether at a single (SAD) or at multiple (MAD) wavelengths. Indeed, when the wavelength of the incident X-rays closely matches the wavelength of the K absorption edge of the atom, i.e the energy levels of the X-Ray beam is just above the binding energy of the K-shell electron of the atom, it promotes transitions of these electrons to higher orbitals. Thus, the electron distribution is not centrosymmetric around the nucleus anymore and therefore, so is the scattering, which is thus said to be anomalous. Consequently, Friedel's law doesn't hold anymore and therefore, the h,k,l reflections and -h,-k,-l reflections have different intensities and phase which can therefore be used to estimate it. In practice, this method is performed via the soaking of the protein's crystals in halides solutions or via the crystallisation of proteins that covalently derivatised with selenium atom, for example expressed using seleno-methionine instead or regular methionine. This is due to bromide, iodide or selenium having a strong anomalous signal and K-edge at wavelength accessible for synchrotron X-rays[471–473].

1.5.3.3.3 Molecular replacement

If the structure of a protein or complex is already known from another crystal or thought to be similar to an already known structure, it is possible to transfer it to the unknown arrangement by the *molecular replacement* method. Indeed, the placement of the known, or search, structure in the target unit cell can be achieved first by a rotation to determine the angular relationship of the known and unknown structure with respect to each other, and then by a translation to superpose them.

This is possible as the self-Patterson maps of these structures ought to be similar and therefore, calculating them and rotating one should permit the overlap onto the other as these self-Patterson vectors depends on the molecule structure and not on their relative positions within the unit cell. Then, both self- and cross-Patterson vectors can be offset so

the known structure model fits in the crystal lattice, allowing to calculate the structure factor **F**(h k l), including its phase, for the positioned model, which is in turn iteratively refined in order to minimise the difference between the observed and calculated reflections and therefore, allowing to solve the structure of the target crystal[471–473].

1.5.3.4 Non-crystallographic symmetries and crystal defects

Apart from the group space symmetry of the lattice, particular symmetries may also be found within each asymmetric unit but not throughout the whole crystal, in which case it is accordingly deemed non-crystallographic symmetry (NCS). Indeed, NCS occurs when more than one identical molecule or subunit crystallises in the asymmetric unit and are related by local translational or rotational symmetry operations. As such, the presence of NCS can typically be seen in the Patterson map as a strong peak, unless it is a translation along a crystallographic screw axis. NCS-related proteins or subunits have therefore similar electron density, which can be used as constraints on the structure factor and thus, on the phase angles. Moreover, the identification of NCS can aid in the initial placement of protein in the molecular replacement methods. Indeed, the known symmetry relationships can be used to predict the positions of all NCS-related copies once one has been accurately placed. Finally, during the refinement of the model, the presence of NCS permit the imposition of constraints or restraints that ensure NCS-related molecules maintain similar conformations, improving the overall model accuracy[471–473].

Moreover, crystals can present several defects, amongst them twinning and mosaicity. Twinning occurs when two or more adjacent crystals share some of the same lattice points in a symmetrical manner. These crystals are therefore oriented differently, creating a composite structure that can complicate the interpretation of diffraction patterns. Twinning can be detected by analysing the diffraction data for overlapping or anomalous reflections, and it often requires particular care in order to deconvolute the data from the twinned crystals. However, the presence of translational NCS may mask the presence of twinning as both can result in similar effect on the diffraction pattern[471–473].

In turn, mosaicity, refers to the presence of slightly misaligned unit cells within a single crystal, which can be described as "mosaic-like". These domains are almost but not

perfectly aligned, resulting in the of the diffraction spots, which can reduce the resolution and quality of the collected data.

1.5.3.5 The B-factor

In the previous sections, the atoms in the unit cell of the crystal have been considered as occupying fixed positions in the lattice. However, this is not correct as atoms undergo thermally induced vibrational and librational motions even at cryogenic temperature which therefore induce a mean square displacement noted $\langle u^2 \rangle$ of the scattering centres in a direction normal to the diffraction planes that increases with the temperature. Because of these motions, the intensity of the diffracted spots is reduced which can be taken into account on the form, or atomic scattering, factor f defined in eq. 0.70 as

$$f = f_0 \exp\left[-B\left(\frac{\sin(\theta)}{\lambda}\right)^2\right]$$

with

$$B = 8\pi^2\langle u^2 \rangle \,.$$

This coefficient B is called the Debye-Waller factor, the temperature factor due to the temperature dependence of $\langle u^2 \rangle$ or simply, the B-factor. However, being an atomic mean-square displacement, the B factor also reports on protein flexibility and disorder. Indeed, high B-factors indicate regions of the protein that are more flexible or disordered, as these regions exhibit greater atomic displacements while lower B-factors suggest more rigid and well-ordered regions of the protein structure[476–478].

1.5.4 Dynamic Light Scattering

Dynamic light scattering (DLS) or photon correlation spectroscopy (PCS) is a light scattering technique with many applications. It is easily accessible as it does not require a synchrotron light source[460,479].

1.5.4.1 Scattering considerations

If the size of the scattering particles is significantly smaller than the wavelength of the incident light, the particle scatters the light elastically, which means that the scattered wave has the same wavelength as the incident wave, a phenomenon known as *Rayleigh scattering*[460,479,480].
However, if the size of the particle is bigger than a threshold commonly approximated to a tenth of the wavelength, then the anisotropic *Mie scattering* phenomenon occurs instead, which is angle dependent. As a solution to Maxwell's equation, the Mie scattering theory describes the scattering as the superposition of light radiated by all the oscillating electric and magnetic moments of the particle. When the particle is small enough with respect to the light wavelength, the scattering is dominated by the electric dipole moment, which is the only one taken into account in Rayleigh scattering[460,460,479,481].

As modern DLS lasers have a typically wavelength on the order of 600 nm, when the hydrodynamic radius R_h of scattered particles passes the approximate limit of 60 nm, which can typically happen for large assembly of proteins like fibrils or polymers, Mie scattering can therefore occur. However, it is usually neglected despite its anisotropic properties, possibly making the study of big species unreliable.

1.5.4.2 Fluctuations and autocorrelation function

Unlike SAXS, DLS is not measured as a function of the wavevector q, but the intensity of the scattered light is detected at a fixed q and is therefore dependent of the position of the scattering centres at a given time. Here, q is defined as

$$q = \frac{4\pi\eta}{\lambda}\sin\left(\frac{\theta}{2}\right)$$

which, compared to eq. 0.59, is also function of the refractive index of the solvent η and with θ being divided by two to represent the angle at which the detector is located[460].

However, collisions with solvent molecules result in constant rotational and translational diffusion of the particles in solution (cf. section 1.4.5.3.1.1 page 80). These Brownian motions therefore cause a Doppler broadening of the monochromatic incident light, i.e. the individual scattered waves from each scattering centre either mutually cancel each other due to destructive interference, or, conversely, their mutually constructive phases give rise to the detected signal. Therefore, as the particles diffuse, the scattered light intensity randomly fluctuates over time, which is measured by the autocorrelator to give rise to the time-averaged autocorrelation function of the sample. This is the $G_2(\tau)$ second-order intensity correlation function expressed as the integral over the product of time-dependent intensities as[460,479,482,483]

$$G^{(2)}(\tau) = \langle I(t)I(t+\tau)\rangle$$

which is normalised as

$$g^{(2)}(\tau) = \frac{G^{(2)}(\tau)}{\langle I(t)\rangle^2} = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2}$$

with I being the measured intensity, t the time, τ the time interval between time-points and $g^{(2)}(\tau)$ the normalised intensity correlation function, while the brackets denote the averaging over time.

Thus, although it is not possible to obtain the positions of the scattering centres at all measured times in DLS, the signal detected is a function of the coherence of their motion and can therefore be correlated as mean of correlation functions. However, the temporal intensity correlation function $G^{(2)}(\tau)$ describes the second-order coherence of the scattered waves through the detection of their interference signal. The first-order correlation function, or electric field correlation function $G^{(1)}(\tau)$ and its normalised expression $g^{(1)}(\tau)$ are therefore necessary and defined as[460,479,482,483]

$$g^{(1)}(\tau) = \frac{G^{(1)}(\tau)}{\langle I(t)\rangle} = \frac{G^{(1)}(\tau)}{\langle E^*(t)E(t)\rangle} = \frac{\langle E^*(t)E(t+\tau)\rangle}{\langle E^*(t)E(t)\rangle} \qquad 0.76$$

where E represents the actual electric field of the scattered light and with the light intensity I associated with the electric field E being $I(t) = E^*(t)E(t)$. The above equation is a consequence of the Wiener-Khintchine theorem expressing the light electric field power spectrum $S(\omega)$ as a function of a first order correlation function for a stationary random process as $S(\omega) = \int g^{(1)}(\tau)\, exp^{i\omega\tau} d\tau$. As such, it links the spatial correlations to the detected coherence, i.e. to the phase correlations[482–485].

A complete description of the coherence of the scattered waves would require the correlation functions $G_n(\tau)$ at all n orders[486]. However, in the case of a random scattered field arising from a process with a Gaussian probability distribution in an ergodic medium, the correlation functions at all orders can be related to each other. In the case of DLS, assuming that the photodetector is homodyne, meaning that only the scattered light signal is detected, which is typically enforced by measurement at sufficiently large angles θ to detect outside the incident beam, the Siegert relation linking the two first-order correlation functions arising from a spatially coherent scattered polarised light applies, as[460,482,483,487]

$$g^{(2)}(\tau) = B + \beta\left|g^{(1)}(\tau)\right|^2 \qquad 0.77$$

with B representing the baseline, typically equal to 1 – which will be used for the following calculations – and β the coherence factor representing the overall optical detection efficiency of the experimental setup. Therefore, directly measuring the autocorrelation of fluctuations of the scattered light over time $G^{(2)}(\tau)$ gives a satisfactory description of the scattering process.

It can be shown from eq. 0.76 and **0.51** that the electric field correlation functions of a single species undergoing Brownian motion exponentially decays as[482]

$$g^{(1)}(\tau) = \exp(-D_T q^2 \tau) \qquad 0.78$$

with D_T the translational diffusion coefficient of the species, q the wave vector defined in eq. 0.59 and τ the time interval of the measurement. Thus, by putting together equations 0.77 and 0.78, the measured $g^{(2)}(\tau)$ function is[482]

$$g^{(2)}(\tau) = B + \beta\exp(-2D_T q^2 \tau)\,. \qquad 0.79$$

Thus, this equation states that the fluctuations of the scattered intensities measured as the $g^{(2)}(\tau)$ autocorrelation function at a fixed wave vector q are directly a function of the particles' motions represented by their translational diffusion coefficient D_T. From this

equation, it is straightforward to see that the faster the translational diffusion, i.e. the larger the diffusion coefficient D_T, the faster the decay. This is illustrated in Figure 2.4 on the left where the baseline subtracted function $g^{(2)}(\tau) - B$ decays faster for the green curve calculated with arbitrarily chosen $D_T = 5\text{x}10^{-6}\ \text{m}^2\text{s}^{-1}$ than for the red one calculated with $D_T = 5\text{x}10^{-8}\ \text{m}^2\text{s}^{-1}$.

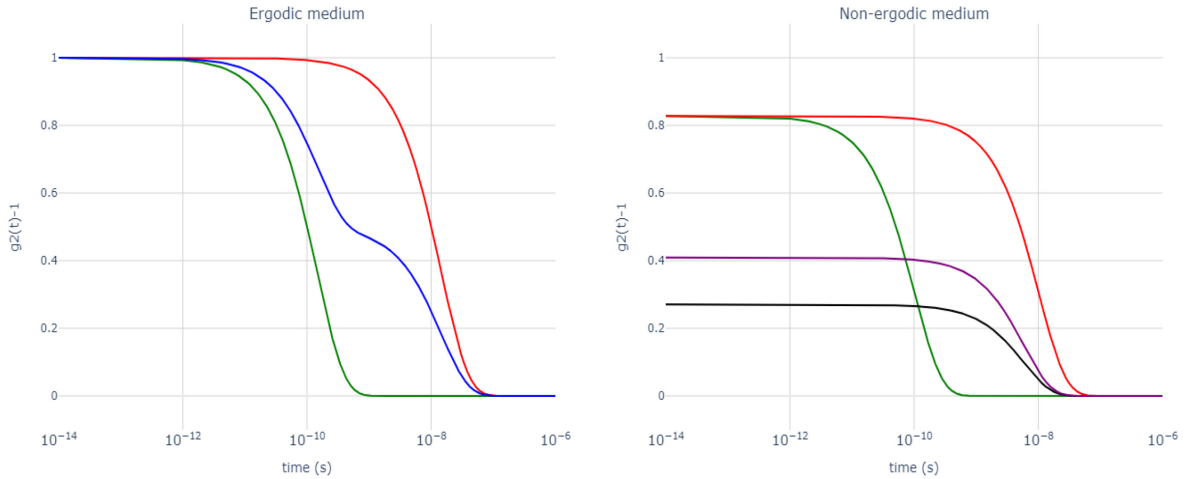


Figure 2.4: DLS autocorrelation functions for ergodic and non-ergodic media.

Baseline-subtracted normalised DLS intensity correlation function $g^{(2)}(t) - 1$ for the classical, ergodic case (left), calculated using eq. 2.79, and the non-ergodic case (right), calculated using eq 2.80. The red and green curves in the two plots correspond to species with respectively $D_T = 5 \times\ m^2 s^{-1}$ and $D_T = 5 \times 10^{-6}\ m^2 s^{-1}$ while the blue curve in the left plot correspond to the presence of both species in equal proportions. The non-ergodic correlation functions (right) were calculated for 10^6 scattering centres, with the non-ergodicity parameter Y=1.3 and the non-diffusing fraction $g^{(1)}(q, \infty) = 0.3$ for the red and green curves. The black and purples profiles were calculated using $D_T = 5 \times 10^{-8}\ m^2 s^{-1}$and with Y=1.3, $g^{(1)}(q, \infty) = 0.6$ (black) or Y=1.6, $g^{(1)}(q, \infty) = 0.3$ (purple). All functions were calculated using a wavelength of 632.8 nm, a refractive angle of 1.33 corresponding to water and a detection angle θ of 173°.

1.5.4.3 DLS in non-ergodic media

The ergodic character of a system means that, over a sufficiently long time, any individual component of the system can fully sample all its possible spatial configurations, and therefore that the measurement of a property of that system over the course of an experiment is a representative ensemble average of that property. In the case of DLS, this means that the time-averaged intensity correlation function measured is the same as the ensemble-averaged correlation, that is $\langle G^{(2)}(\tau) \rangle = \overline{G^{(2)}(\tau)}$.

However, some media like gels or glasses are non-ergodic systems. Indeed, the scattering particles in such samples can only experience localised and physically limited

translational and rotational diffusion arising from Brownian motions or mutual diffusion due to concentration inhomogeneities. Consequently, this means that the scattering particles are only able to sample a sub-ensemble of positions which is defined by the time-averaged positions of these scattering centres as well as the magnitude of their displacements. As non-ergodic samples are typically non homogeneous, different samples or even different measurements at different positions of the very same sample will give different results because they describe different sub-ensembles.

Therefore, the time-averaged intensity correlation function $G^{(2)}(\tau)$ obtained from such a DLS experiment is not equal to the ensemble-averaged correlation function but only to one of its particular sub-ensemble-averaged correlation functions, which results in the Siegert relation (eq. 0.77) and its subsequent development not being valid any more, as pointed out by Pusey and Van Megen who developed a model for such cases[488]. Assuming that the sub-ensemble view described above holds; that the field scattered by such a sub-ensemble is the sum of a fluctuating, ergodic component and of a constant, non-ergodic, component representing the gel network; and that the observed collective diffusion coefficient is dependent on q[488–490], the time-averaged normalised intensity correlation function $g^{(2)}(q,\tau)$ can be expressed as

$$g^{(2)}(q,\tau) = \sigma_l^2 \exp\left(\frac{-Y D_T q^2 \tau}{\sigma_l}\right)^2 \qquad 0.80$$

with Y being a parameter characterising the ergodicity and expressed being the ratio of the ensemble and time averaged intensities, i.e. $Y = \langle I(q)\rangle / \overline{I(q)}$, while σ_l^2 is the mean-square intensity fluctuation and is defined as

$$\sigma_l^2 = Y^2\left(B - g^{(1)}(q,\infty)\right) = g^{(2)}(q,0) - B$$

which depends on the field correlation function value at infinite time $g^{(1)}(q,\infty)$ that effectively represents the fraction of the scattering species that are fixed and not fluctuating. Consequently, in ergodic media, $Y = 1$ and $g^{(1)}(q,\infty) = 0$, which effectively reduces eq. 0.80 to eq. 0.79.

Because of the difference between the time- and ensemble-averaged intensities, a non-ergodic intensity correlation function does not begin at 1 when $\tau = 0$ but at σ_l^2 which can be seen in Figure 2.4. The figure also varies Y and $g^{(1)}(q,\infty)$ in order to highlight their

influence on the shape of the correlation function. Altogether, the observation of a suppression of the initial $g^{(2)}(0)$ value is therefore a straightforward way to recognise the apparition of non-ergodicity which has been successfully used to characterise the gelation point of colloids and polymer gels[490,490,491,491].

1.5.4.4 Hydrodynamic radius

As the translational diffusion coefficient D_T is linked to the hydrodynamic radius R_h using relations like the Stokes-Einstein equation or versions that take into account the shape of the molecule (cf. eq. **0.52** and section **1.4.5.3.1**.1 page 80), fitting the measured intensity correlation function allows to obtain these parameters if the dilution regime does not render the equation invalid. For a monodisperse sample of particles whose shape is known and can be approximated as a sphere or a spheroid, this is easily done using a cumulant fit based on the Taylor expansion of the logarithm of the correlation function[492].[492]. However, in the case of a polydisperse sample, the ill-posed nature of the inversion renders the obtention of a reliable Rh difficult. Indeed, an infinite number of possible combinations of particles exists that results in the same autocorrelation function. A regularisation of the solution is therefore necessary. For example, the most widely used method, known as CONTIN, performs a numerical Laplace transform inversion using a Tikhonov regularisation to constrain the solution[493,493,494,494], a procedure that has also found uses in diffusion NMR[495]. The regularisation results in the most conservative solution to be chosen, which therefore often yields significantly spread peaks instead of multiple narrow ones for species of close radii.

As the obtention of R_h also depends on viscosity, any particular nano-viscosity experienced by the observed species may yield inaccurate estimations if the shear viscosity is used instead, which is particularly true for crowded and viscous environments like condensates and gels. Moreover, the R_h of proteins may vary depending on how extended the crowding allows them to be due do the excluded volume effect[496], while reaching the semidilute crowding regime or entanglement causes a departure from the Stokes-Einstein equation. For this reason, no DLS-derived hydrodynamic radii are presented in this thesis.

1.6 Atomic Force Microscopy

Atomic Force Microscopy (AFM) is a microscopy technique that probes the surface of a sample from the detected forces between a solid tip and the sample. As such, it doesn't produce an image, but determine an array of points that corresponds to a topographical image by encoding the measured height as the intensities of pixel points[497–499]. It is, however, a common practice to visualize AFM topography using a false-color scale that resembles a classical picture.

However, these heights are measured indirectly. Indeed, the tip is mounted on a spring-like cantilever and as the tip approaches the surface of the sample, the cantilever bends vertically or horizontally due to the effect of the different atomic forces. In turn, the extent of this bending is quantified by measuring the reflection of an incident laser beam on the cantilever onto a photodetector, yielding a detectable deflection that can be translated into vertical height values using a feedback loop connected to a sensitive piezoelectric scanner linked to the cantilever that does not only moves the sample laterally in x and y for the tip to probe but also adjusts its setpoint vertically in z in order to not exceed the limit of the cantilever bending[498,499].

1.6.1 Tip-sample interaction

Depending on the mode (vide supra), the experimental conditions, the shape and material of the tip as well as the properties of the surface of the sample, different forces of varying magnitudes can act on the tip, thus bending the cantilever upward if repulsive or downward if the forces are attractive[500].

Indeed, as the AFM tip apex approaches the sample, the atomic distance between them gets gradually small enough for the fluctuating electrons distribution of one atom to repel the nearest electrons of the other atom. This Coulombic repulsion eventually induces an instantaneous dipole moment in both atoms, resulting in the attractive long-range London dispersion force which increases as the 6th power of the inverse of the interatomic distance[500–502]. Additionally, Keesom and Debye forces can also contribute to the long-

range attractions if respectively permanent dipoles or dipole-induced dipole interactions between atoms or molecules are present[500,502].
These van der Waals forces are therefore typically seen as the dominant ones driving tip-sample adherence. However, if the tip continues to draw closer to the surface and their distances reaches approximately 2 Å, electron orbitals start to overlap and the Pauli exclusion principle causes a repulsive force that grows with the 12th power of the interatomic distance inverse[500,502]. This is why the tip-sample interaction is often modelled as the "6-12" Lennard-Jones potential for van der Waals interactions[500].

1.6.2 Different modes, different experiments

AFM is a polyvalent method. Indeed, measurement scan be realised not only in air or in liquid but also in different modes. In the classic, contact mode, the tip is close enough to the sample to be considered in contact with its surface height which causes the feedback. This mode is therefore best suited for rigid and flat surfaces due to the frictions and shear force generated which may limit the lateral resolution if the sample is soft. However, due to the constant contact, the z-resolution is often really high[498,499].

The most used mode nowadays is the tapping mode, in which the cantilever oscillates at a particular resonant frequency as a driven harmonic oscillator and therefore has the tip lightly taps the sample's surface as it scans over it. Variations on the surface height of the sample cause a damping of the oscillation, therefore triggering an amplitude and phase change that results in the feedback needed to probe the surface. As such, tapping mode limits the strain due to lateral shear forces on the sample and tip compared to the contact mode which in turn improves the lateral resolution, making it well suited for soft or deformable samples. While the topological height of the surface causes the amplitude change, the phase change in turn results from the interactions between the tip and the surface. Indeed, while a surface can be flat and as such, won't change the amplitude of the tip's oscillation, a softer or more adhesive region than the rest have stronger attractive or repulsive properties, which results in a phase change due to the introduction different harmonic frequencies in the cantilever oscillation resulting from the tip-sample interactions. Phase images and Fourier harmonics analysis are therefore a great complement to the topological height images and provide significant insights in the

material properties and heterogeneity of the sample with a high lateral resolution due to their sensitivity to boundary changes[498,499,503–506].

Additionally, the non-contact mode of AFM has the tip oscillates in the vicinity of the sample at a fixed height without adhering to the sample, which amounts to be close enough for the tip to be within the attractive forces range without reaching the repulsive regime. As such, the feedback is triggered by the adjustment of the amplitude or phase necessary to maintain the constant cantilever oscillation frequency. This effectively reduces both normal and lateral forces strain and represent as well an advantage for soft samples[498,499].

Finally, AFM can also perform force spectroscopy by making single point indentation measurements of forces between the tip and the sample, recording the deflection of the cantilever. The obtained force curve can in turn be fitted to different model to obtain mechanical and rheological information about the sample such as its stiffness or viscoelastic properties through Young's modulus, how attractive or repulsive its surface is or even its depths[507].

1.6.3 Tip-sample convolution

As AFM measurements are done in response to the surface height, this information is the most reliable and as it is probed by the tip, the sharper it is, the higher the resolution. In ideal conditions, AFM z-resolution can measure surface height down to the angstrom, allowing the observation of even hydrogen bonds in chemical compounds[508].

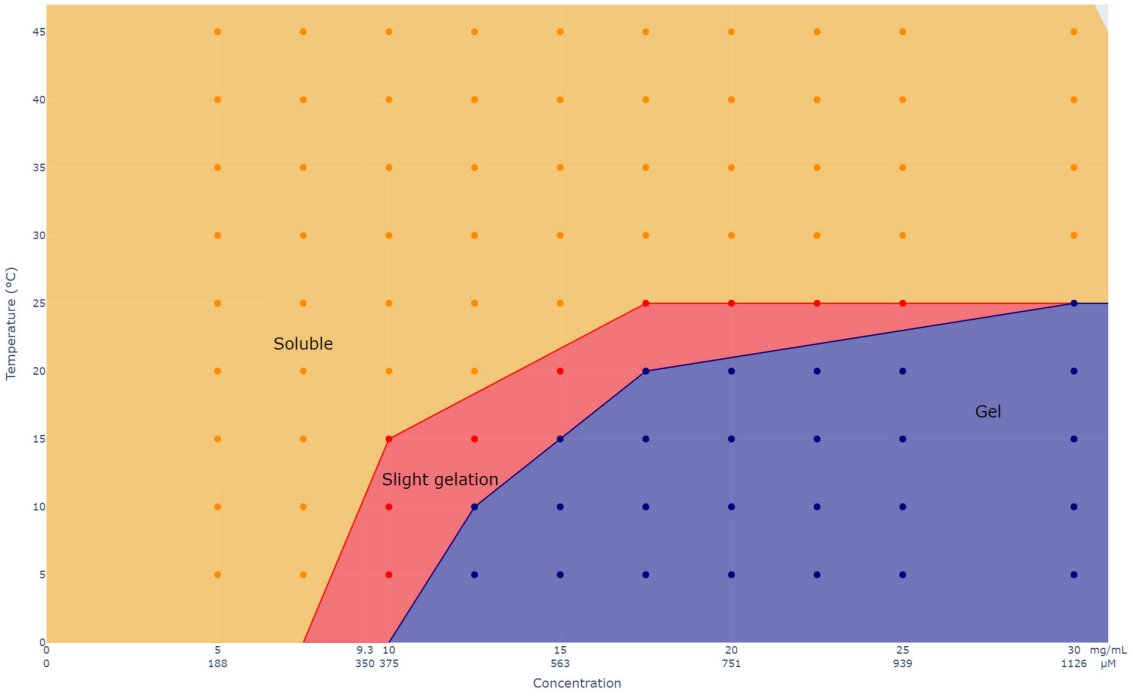
However, as the tip has a finite size as well as a defined shape and scans over the sample laterally, it encounters the surface sideways which causes a convolution effect with the object. Indeed, a lateral contact with the tip in its basis rather than its apex with a protruding, convex feature of the object causes an artificial elevation from the cantilever due to its geometrical exclusion, making this feature appear wider and its elevation slope smaller than what they really are. Conversely, a hole or concave feature cannot be accurately probed if the tip is wider as the cantilever adjusts the setpoint only once and only as much as the tip can physically go into the hole, thus making it appear narrower

and possibly shallower than what it really is. This is called the tip convolution effect and limits the lateral resolution of the measurement as well as the accuracy of the z-measurement for deep features. The extent of this effect therefore depends on the shape and size of the tip, with broad tips, in comparison with narrow ones, resulting in greater broadening effects for protruding objects and greater narrowing effects for holes[498,504,509,510].

Multiple strategies have been developed for limiting this effect experimentally like adjusting the measurement angle of the tip, or to compensate for it in the data processing in order to reconstruct the real sample width using the known shape and dimensions of the tip[511–513] or using correcting algorithms on the image directly without a priori information on the tip[514–516]. One of them, the Laplacian weight filter method, calculates the value of the Laplacian function for each pixel of the image and uses it as a weighting factor for the smoothening of their intensity value by two-dimensional Gaussian window. Indeed, as a second derivative operator, the Laplacian function is well suited to detect edges and therefore, this method allows a correction of the tip effect and greatly enhances the lateral resolution of AFM images by sharpening the boundaries of the objects[516,517].

Chapter 3: Overview of CAHS-8 and its different forms

Before presenting the results obtained during this thesis, it is necessary to introduce briefly CAHS-8 and its phase diagram. At the very beginning of this thesis, during the first purifications of the protein, it became quickly obvious that CAHS-8 would form an apparent gel at high concentrations. It was subsequently noticed that the low temperature favoured the formation of such putative gel. However, it was also observed on occasion that the samples would become highly viscous, but without forming an apparent gel. This phase was later revealed to correspond to the formation of fibrillar assemblies which would in turn entangle to form a gel. The conditions of fibrillation are termed “slight gelation” in the following phase diagram of CAHS-8, which was obtained by DLS as detailed in the chapter 7 which is dedicated to the study of the gels. The fibrillar assemblies are mentioned and first shown in the chapter 5, while the chapter 6 is dedicated to their study.



Chapter 4: Published article

An article was published during this thesis in late October 2021 and is reproduced here under terms of the CC-BY license. Unlike the following result chapters which follow a hierarchical organisation, the article presents the findings of the project as a whole. All the data in it are nevertheless presented and discussed in the result chapters.

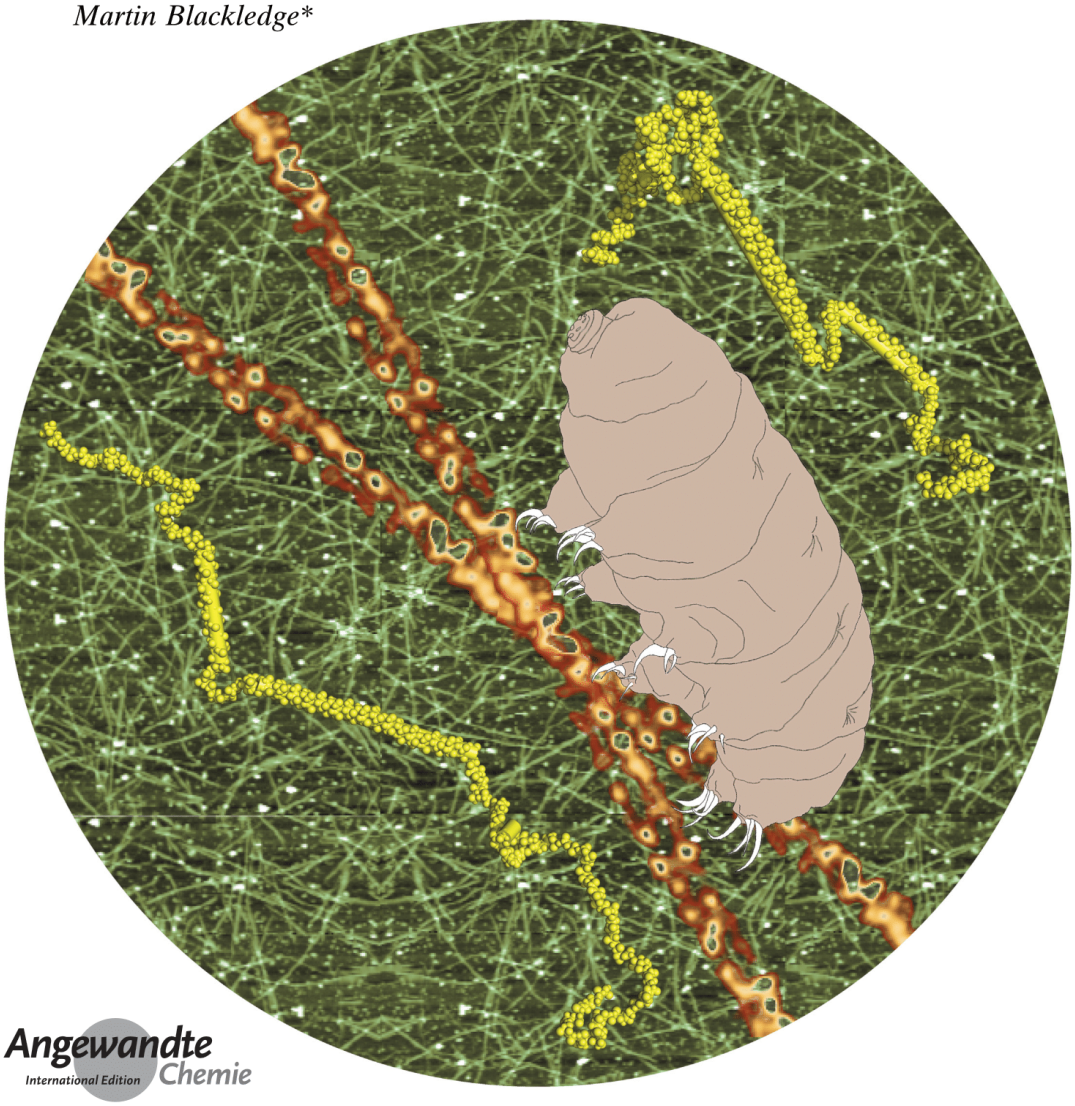
GDCh Communications Angewandte International Edition Chemie www.angewandte.org
Check for updates

Intrinsically Disordered Proteins | Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2022**, *61*, e202109961
International Edition: doi.org/10.1002/anie.202109961
German Edition: doi.org/10.1002/ange.202109961

Intrinsically Disordered Tardigrade Proteins Self-Assemble into Fibrous Gels in Response to Environmental Stress

*Anas Malki, Jean-Marie Teulon, Aldo R. Camacho-Zarco, Shu-wen W. Chen, Wiktor Adamski, Damien Maurin, Nicola Salvi, Jean-Luc Pellequer, and Martin Blackledge**



© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202109961, Wiley Online Library on [22/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Abstract: *Tardigrades are remarkable for their ability to survive harsh stress conditions as diverse as extreme temperature and desiccation. The molecular mechanisms that confer this unusual resistance to physical stress remain unknown. Recently, tardigrade-unique intrinsically disordered proteins have been shown to play an essential role in tardigrade anhydrobiosis. Here, we characterize the conformational and physical behaviour of CAHS-8 from Hypsibius exemplaris. NMR spectroscopy reveals that the protein comprises an extended central helical domain flanked by disordered termini. Upon concentration, the protein is shown to successively form oligomers, long fibres, and finally gels constituted of fibres in a strongly temperature-dependent manner. The helical domain forms the core of the fibrillar structure, with the disordered termini remaining highly dynamic within the gel. Soluble proteins can be encapsulated within cavities in the gel, maintaining their functional form. The ability to reversibly form fibrous gels may be associated with the enhanced protective properties of these proteins.*

Tardigrades are microscopic aquatic animals[1] that exhibit remarkable resistance to environmental stresses such as extreme temperature, radiation, high pressure, or absence of oxygen or water, a characteristic that has remained mysterious for centuries.[2–4] In response to such stress, tardigrades transform into a characteristic *tun* state that can survive for long periods of time.[5] Although the molecular mechanisms that confer this protection against extreme physiological stress remains poorly understood,[6] recent studies have identified the importance of two families of tardigrade-unique intrinsically disordered proteins (cytosolic abundant heat-soluble (CAHS) proteins and secreted abundant heat-soluble (SAHS) proteins).[7–10] These proteins are collectively known as tardigrade-specific disordered proteins (TDPs).[9] RNA interference demonstrated that TDPs were required for desiccation tolerance in tardigrades, while heterologous expression of certain TDPs demonstrated their ability to increase desiccation tolerance in both yeast and *E. coli*.[9]

[*] A. Malki, J.-M. Teulon, A. R. Camacho-Zarco, W. Adamski, D. Maurin, N. Salvi, J.-L. Pellequer, M. Blackledge
Univ. Grenoble Alpes, CNRS, CEA
Institut de Biologie Structurale
Grenoble (France)
E-mail: martin.blackledge@ibs.fr
S.-w. W. Chen
niChe Lab for Stem Cell and Regenerative Medicine
Department of Biochemical Science and Technology
National (Taiwan) University
Taipei 10617 (Taiwan)
Supporting information and the ORCID identification number for some of the authors of this article can be found under: https://doi.org/10.1002/anie.202109961.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

The protective role of TDPs remains under discussion in the literature.[11–13] The observation of the glass transition temperature of TDPs by differential scanning calorimetry was interpreted as evidence that TDPs can reversibly vitrify the cellular environment.[9] Similar phenomena were observed for the intrinsically disordered late embryogenesis abundant (LEA) proteins,[14] through the stabilization and apparent vitrification of co-solutes such as trehalose. This sugar is known to play an important role in anhydrobiosis in a wide range of organisms,[15] although trehalose is not generally abundant in tardigrades.[16]

To better understand how TDPs alone can confer protective properties in response to desiccation and other environmental stresses, we investigated the conformational behaviour of a CAHS from *Hypsibius exemplaris* at atomic resolution. By using high-field NMR spectroscopy, we reveal that the protein comprises an extended central helical domain, with apparent coiled-coil propensity, flanked by long, highly disordered N- and C-terminal domains. Small-angle X-ray scattering (SAXS), atomic force microscopy (AFM), and dynamic light scattering (DLS) were used to follow the successive formation of oligomers, long fibres, and finally gels constituted of fibres. We were able to probe the behaviour of soluble proteins that are encapsulated within the gel matrix, and propose possible mechanisms by which the formation of such fibrous gels can enable anhydrobiosis.

The ^{15}N-^{1}H HSQC spectrum of CAHS-8 (Figure 1 A) is typical of an intrinsically disordered protein, with low dispersion in the $^{1}H^{N}$ dimension. Nevertheless, >98% of the backbone resonances could be assigned (see Methods in the Supporting Information), revealing the presence of a helical domain between residues 95 and 195 (Figure 1 B).

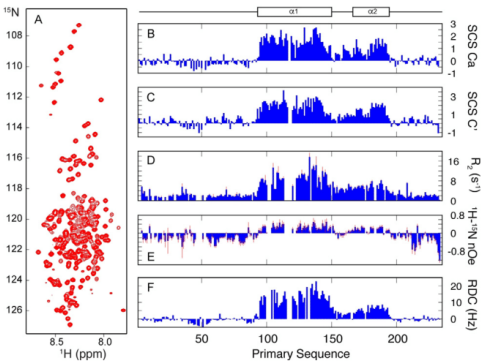


Figure 1. Characterisation of the structural and dynamic properties of CAHS-8 in solution. A) The ^{15}N-^{1}H correlation spectrum is characteristic of an intrinsically disordered protein. B, C) Secondary ^{13}Cα and ^{13}C′ chemical shifts showing strong α-helical propensities from 95–195, divided into two distinct helices, α_1 (E95-F150) and α_2 (V167-L194). ^{15}N relaxation (D; 850 MHz, E; 600 MHz), ^{15}N-^{1}H RDCs (F), and PREs (see Figure S1) suggest the protein acts as a disordered chain comprising two highly populated α-helices (303 K) and no persistent tertiary structure. The NMR spectrum was recorded at 350 μM (9.3 mg mL^{-1}). The NMR assignments are deposited in BMRB (accession code 51115).

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202109961, Wiley Online Library on [22/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This region is divided into two segments, α_1 (E95–F150) and α_2 (V167–L194), with α_2 showing a lower level of helical propensity (Figure 1B,C; estimated from ^{13}C secondary shifts to be around 60 and 30%, respectively). The linker between α_1 and α_2 (12 consecutive sites from 152–163, with ^{1}H-^{15}N nOe effects <0.2 and ^{13}C shifts showing a secondary structure population of <0.25) exhibits reduced helical propensity (Figure 1B,C), and enhanced flexibility (Figure 1D,E).

The tertiary folding of CAHS-8 was probed using paramagnetic relaxation enhancements, with TEMPO-maleimide labelling of positions A134C and A185C. The absence of significant enhancements beyond immediate neighbours indicates an absence of persistent long-range structure (see Figure S1).

The presence of two helices connected by a more flexible linker is supported by the measurement of ^{15}N-^{1}H residual dipolar couplings (RDCs; see Methods in the Supporting Information). The RDCs show a profile typical of IDPs with helical elements that do not specifically interact with each other, as predicted by simulation and experimentally validated in numerous systems. The RDCs exhibit positive values in the helical regions and lower or negative values in the unstructured parts of the chain (Figure 1f).[17,18] A model of CAHS-8, comprising a flexible chain with two consecutive helices connected by a short linking peptide and flanked by intrinsically disordered termini was calculated using the disordered protein ensemble generator *flexible-meccano*[19,20] and compared to SAXS measurements. Ensemble selection using the ASTEROIDS approach suggests that the protein samples a largely extended conformation in the monomeric state with an average radius of gyration of 58 ± 7 Å (Figure 2A). The average radius of gyration $R_{g,av}$ of an ensemble of fully disordered conformers is 47 ± 10 Å, while an ensemble of conformers containing the helical elements ($R_{g,av}=59\pm11$ Å; see Figure S2) coincides with the experimental radius of gyration, thereby supporting the nature of the ensemble of unfolded chains comprising significant populations of helices α_1 and α_2.

CAHS-8 shows a strongly concentration-dependent propensity to oligomerize, as illustrated by SAXS, where scattering curves that are characteristic of fibril formation are observed at higher concentrations (Figure 2A). The ^{13}C chemical shifts show strong temperature dependence, revealing a significantly higher helical content at 278 K for helix α_2 (see Figure S3). The ^{15}N relaxation is also temperature dependent (Figure 2C and see also Figure S4), notably higher than that normally observed in intrinsically disordered proteins containing helical elements[21,22] (see Figure S2), thus supporting the formation of higher order oligomers at lower temperature. Dynamic light scattering (DLS) measurements were used to map the phase diagram of CAHS-8 by observing the dependence of the decay in the autocorrelation of scattered intensity as a function of temperature and concentration (Figure 2B). We observe the formation of an apparent physical hydrogel phase at low temperatures and higher concentration, in qualitative agreement with observations from the NMR and SAXS studies.

We used atomic force microscopy (AFM) in liquid and air modes to observe the apparent formation of fibrils that eventually form a dense mesh that resembles a gel and to investigate the nature of these gels. Extension often follows a twisted helix structure comprising paired filaments, exhibiting a pitch of approximately 600 nm (Figure 3A). AFM was used to follow gel formation, and revealed the generation of

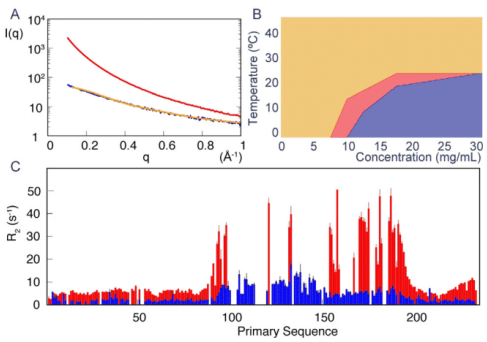


Figure 2. Monomeric CAHS-8 forms higher order oligomers as a function of concentration and temperature. A) SAXS measured at 2 mg mL^{-1} (blue) and 5 mg mL^{-1} (red). At the higher concentration, the curve shows characteristics associated with fibril formation. The orange curve shows the predicted SAXS curve from an ASTEROIDS-selected ensemble of 7 disordered conformers calculated using *flexible-meccano* (average R_g of best-fitting ensembles $=58\pm7$ Å). B) Phase diagram derived from DLS. Orange represents the isotropic phase, red represents increased oligomerization associated with slight gel formation, and blue represents gel formation. C) Transverse ^{15}N relaxation (R_2) measured at 303 K (blue) and 278 K (red) at 850 MHz, showing a very large increase in molecular tumbling at the lower temperature.

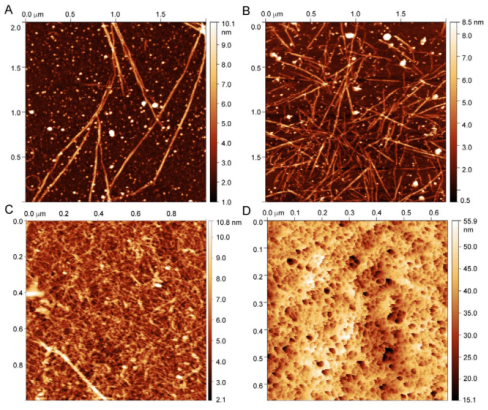


Figure 3. Different states of CAHS-8 imaged by AFM in the liquid (A,B) and air mode (C,D). A) Dissolving the gel on mica revealed fibrils and allowed a kinetic series to be recorded (see Figure 3). B) Allowing fibrillation to continue whilst the sample is drying shows an increasingly dense matrix of fibrils. C) After the kinetic series from (A) was recorded and the sample fully dehydrated, air imaging displays the entanglement of the fibrils forming a gel. D) Placing a 30 mg mL^{-1} gel on glass reveals porous structures, comprising cavities with diameters of tens of nanometers.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202109961, Wiley Online Library on [22/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

fibrillar structures as a function of time (Figure 3B, see also Figure S3), eventually leading to a dense mesh of fibres (Figure 3C). Circular fibrillar structures with a radius of about 150 nm were also formed (see Figure S5). Application of a Laplacian-weight method[23,24] highlights the twist, and again suggests the presence of paired strands forming single fibrils (see Figure S6). Placing concentrated CAHS-8 on glass and cooling resulted in the formation of a porous structure comprising large cavities with diameters of 10–30 nm (Figure 3D).

NMR spectroscopic analysis of the gels formed from ^{15}N-labelled CAHS-8 demonstrates that the intrinsically disordered domains (excluding helix α_1) remain flexible in the context of the fibrillar gel (Figure 4A). No signals from helix α_1 were detected in the ^{15}N-^{1}H HSQC spectrum of a gel formed from ^{15}N-labelled CAHS-8, and ^{15}N spin relaxation rates reveal an overall similar dynamic behaviour of the N-terminal intrinsically disordered region at 303 K as the protein in free solution. The 62–72 strand, comprising both hydrophobic and acidic residues, shows elevated R_2 values, which suggests increased correlation times, possibly arising from interaction with more rigid elements of the gel. The C-terminal domain shows a similar dynamic behaviour to the monomeric phase, with the exception of the terminus, where the non-native TEV-cleavage site (228–233) also shows enhanced R_2 values. Signals from helix α_2 are visible in the spectrum, but show considerably higher relaxation rates than those in free solution (25 s^{-1} compared to 10 s^{-1}), which suggests that this domain is tethered to a much larger, or immobile, object. These observations are compatible with the involvement of α_1 in the formation of fibrils that constitute the gel, with α_2 and the disordered domains remaining flexible within the fibril cavities.

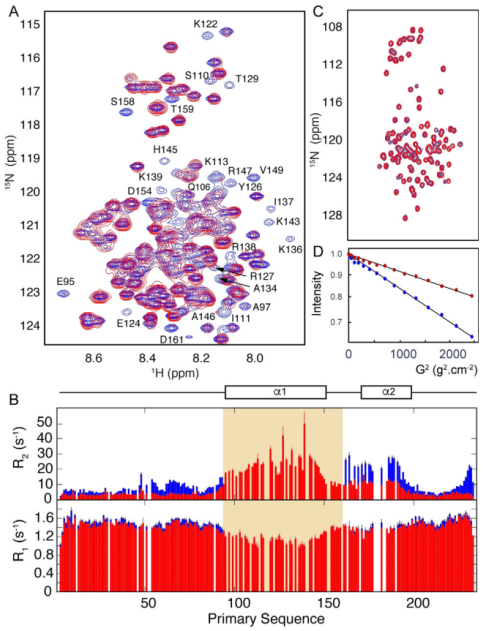


Figure 4. CAHS-8 disordered domains remain flexible in gels that confine soluble proteins. A) Comparison of the ^{15}N-^{1}H correlation spectra (293 K, 850 MHz) of CAHS-8 in monomeric (blue) and gel (red) forms. Residues in helix α1 that disappear in the gel spectrum are annotated. B) Transverse (R_2) and longitudinal (R_1) ^{15}N relaxation (293 K, 850 MHz) of monomeric (red) and gel (blue) CAHS-8. Signals from the shaded region were not visible in the spectrum. C) ^{15}N-^{1}H correlation spectra of the disordered domain of transcription factor ANP32a in buffer (blue) and gel (red) reveal that conformational sampling of ANP32a is unperturbed in gel. D) DOSY measurements of ANP32a in buffer (blue) and gel (red) reveal that translational diffusion of ANP32a is significantly reduced (4.8 compared to 9.6 10^{-11} m^2s^{-1}) in the gel.

To investigate the possible role of gel formation in the establishment of protective properties of CAHS-8, gels were formed in mixtures of the TDP and soluble proteins that were isotopically labelled to allow observation by NMR spectroscopy. The results (illustrated in Figure 4C) demonstrate that the ^{15}N-^{1}H HSQC spectrum of the intrinsically disordered domain of ANP32a,[25] a transcription factor, is essentially unchanged in the gel. The translational diffusion of the protein is reduced by approximately a factor of two compared to standard buffer conditions (Figure 4D). This level of restriction is similar to that measured for a small, folded protein in 6% polyacrylamide gels.[26] Despite this slow diffusion, the conformational sampling of the sequestered protein is maintained in the CAHS-8 gel cavity. Similar results were measured on the intrinsically disordered C-terminal domain of Sendai virus nucleoprotein (N_{TAIL}).[27] Similarly, a folded SH3 domain[28] exhibited an essentially identical HSQC spectrum in gel and isotropic solution (see Figure S7). The intensity of the spectrum is very weak, which suggests a significant fraction is partially immobilised in the gel.

Comparison of the sequence of α_1 with related CAHS sequences reveals highest conservation in the helical domain, which exhibits an irregular coiled coil propensity comprising both canonical heptads and stutters,[29,30] as illustrated for the segment A133–R152 (Figure 5A,B, see also Figure S8). Formation of a coiled-coil conformation has also been associated with desiccation-protectant LEA proteins upon dehydration.[14,31,32]

On the basis of these observations, we propose that the long helical domains of the partially disordered protein associate through coiled-coil-like interactions under certain environmental conditions (low temperature, high concentration) to form oligomers, then fibrils (Figure 5C), which eventually form a gel-like matrix. Both charged and hydrophobic amino acids in this region are strongly conserved over CAHS proteins from different species. NMR relaxation data also show strong evidence that this region is central to the intermolecular interaction, with signals in this region disappearing at lower temperatures, where higher order oligomerization is favoured, and relaxation rates increasing more in this region than elsewhere. Although the molecular structure of the twisted fibril is not known, helix α_1 likely contributes to the structured core, possibly through a super-helical[33] or alternatively a cross-α[34] configuration, while the remainder

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

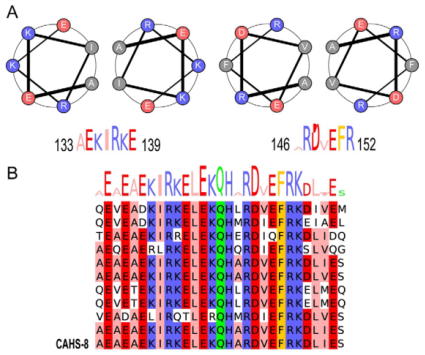


Figure 5. A) CAHS-8 comprises coiled-coil-like sequences located specifically in helix α_1. Two sequences are shown: A133–E139 and A146–R152, which exhibit a typical heptad repeat. B) This propensity is conserved in related sequences (see Figure S8).

of the molecule remains flexible, as shown by NMR spectroscopy.

The exact mechanism by which the formation of such gels protects the organism remains unknown, although the total water volume required to maintain biomolecules in their functional state is clearly reduced in such a matrix. The disordered termini of CAHS-8 remain flexible in the gel and may contribute to the solvation of sequestered biomolecules. In this respect, it is interesting to note that the C-terminal IDR, in particular, is rich in polar amino acids. Cryoprotection may also be facilitated by a reduction in the volume of bulk water, thereby reducing the formation of ice crystals.[35] Under our conditions, we see no evidence of vitrification during the formation of fibrillar hydrogels, as has been observed for trehalose, sucrose, and LEA proteins and proposed as a mechanism of TDP protection.[12,13]

In conclusion, we present a detailed characterization of the transformation of the physical properties of CAHS-8 as a function of environmental conditions. We describe the monomeric behaviour of the protein as partially disordered, with a long two-helical domain flanked by long flexible tails. We propose that this domain mediates homo-dimerization through a coiled-coil interaction, and have experimentally observed the successive formation of oligomers, fibrils, and eventually gels in response to concentration and temperature change. Importantly the gels can sequester soluble proteins in their native state, maintaining their functional form. We propose that the observed transformation of this important protein from soluble monomer to intracellular matrix may be responsible for its protective role in the phenomenal response of tardigrades to environmental stress.

Acknowledgements

This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). We thank Caroline Mas for assistance and access to the Biophysical platform. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). This work acknowledges the AFM platform at the IBS and the ESRF. M.B. acknowledges support from ERC advanced grant DynamicAssemblies.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: atomic force microscopy · gels · intrinsically disordered proteins · NMR spectroscopy · tardigrades

[1] H. Greven, in *Water Bears: The Biology of Tardigrades* (Ed.: R. O. Schill), Springer International Publishing, **2018**, pp. 1–55.
[2] L. Spallanzani, J. G. Dalyell, M. Baillie, *Tracts on the Nature of Animals and Vegetables.*, William Creech, Edinburgh, **1804**.
[3] W. Welnicz, M. A. Grohme, L. Kaczmarek, R. O. Schill, M. Frohme, *J. Insect Physiol.* **2011**, *57*, 577–583.
[4] L. Somme, *EJE* **2013**, *93*, 349–357.
[5] R. Bertolani, R. Guidetti, I. K. Jönsson, T. Altiero, D. Boschini, L. Rebecchi, *J. Limnol.* **2004**, *63*, 16–25.
[6] R. O. Schill, B. Mali, T. Dandekar, M. Schnölzer, D. Reuter, M. Frohme, *Biotechnol. Adv.* **2009**, *27*, 348–352.
[7] A. Yamaguchi, S. Tanaka, S. Yamaguchi, H. Kuwahara, C. Takamura, S. Imajoh-Ohmi, D. D. Horikawa, A. Toyoda, T. Katayama, K. Arakawa, A. Fujiyama, T. Kubo, T. Kunieda, *PLoS One* **2012**, *7*, e44209.
[8] T. Hashimoto, D. D. Horikawa, Y. Saito, H. Kuwahara, H. Kozuka-Hata, T. Shin-I, Y. Minakuchi, K. Ohishi, A. Motoyama, T. Aizu, A. Enomoto, K. Kondo, S. Tanaka, Y. Hara, S. Koshikawa, H. Sagara, T. Miura, S.-I. Yokobori, K. Miyagawa, Y. Suzuki, T. Kubo, M. Oyama, Y. Kohara, A. Fujiyama, K. Arakawa, T. Katayama, A. Toyoda, T. Kunieda, *Nat. Commun.* **2016**, *7*, 12808.
[9] T. C. Boothby, H. Tapia, A. H. Brozena, S. Piszkiewicz, A. E. Smith, I. Giovannini, L. Rebecchi, G. J. Pielak, D. Koshland, B. Goldstein, *Mol. Cell* **2017**, *65*, 975–984.
[10] S. Piszkiewicz, K. H. Gunn, O. Warmuth, A. Propst, A. Mehta, K. H. Nguyen, E. Kuhlman, A. J. Guseman, S. S. Stadmiller, T. C. Boothby, S. B. Neher, G. J. Pielak, *Protein Sci.* **2019**, *28*, 941–951.
[11] T. C. Boothby, G. J. Pielak, *BioEssays* **2017**, *39*, 1700119.
[12] K. Arakawa, K. Numata, *Mol. Cell* **2021**, *81*, 409–410.
[13] T. C. Boothby, *Mol. Cell* **2021**, *81*, 411–413.
[14] J. Ingram, D. Bartels, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 377–403.
[15] L. M. Crowe, *Comp. Biochem. Physiol. Part A* **2002**, *131*, 505–513.
[16] S. Hengherr, A. G. Heyer, H.-R. Köhler, R. O. Schill, *FEBS J.* **2008**, *275*, 281–288.
[17] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.
[18] M. R. Jensen, M. Zweckstetter, J. Huang, M. Backledge, *Chem. Rev.* **2014**, *114*, 6632–6660.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202109961, Wiley Online Library on [22/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[19] P. Bernado, L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, M. Blackledge, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17002 – 17007.
[20] V. Ozenne, F. Bauer, L. Salmon, J.-R. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Bioinformatics* **2012**, *28*, 1463 – 1470.
[21] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2016**, *138*, 6240 – 6251.
[22] W. Adamski, N. Salvi, D. Maurin, J. Magnat, S. Milles, M. R. Jensen, A. Abyzov, C. J. Moreau, M. Blackledge, *J. Am. Chem. Soc.* **2019**, *141*, 17817 – 17829.
[23] S.-W. W. Chen, J.-M. Teulon, C. Godon, J.-L. Pellequer, *J. Mol. Recognit.* **2016**, *29*, 51 – 55.
[24] S.-W. W. Chen, A.-S. Banneville, J.-M. Teulon, J. Timmins, J.-L. Pellequer, *Nanoscale* **2020**, *12*, 22628 – 22638.
[25] A. R. Camacho-Zarco, S. Kalayil, D. Maurin, N. Salvi, E. Delaforge, S. Milles, M. R. Jensen, D. J. Hart, S. Cusack, M. Blackledge, *Nat. Commun.* **2020**, *11*, 3656.
[26] H.-J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, *J. Biomol. NMR* **2000**, *18*, 303 – 309.
[27] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, M. Blackledge, *J. Am. Chem. Soc.* **2008**, *130*, 8055 – 8061.
[28] J. L. O. Roldan, M. Blackledge, N. A. J. van Nuland, A. I. Azuaga, *J. Biomol. NMR* **2011**, *50*, 103 – 117.
[29] J. H. Brown, C. Cohen, D. A. Parry, *Proteins* **1996**, *26*, 134 – 145.
[30] A. N. Lupas, M. Gruber, *Adv. Protein Chem.* **2005**, *70*, 37 – 38.
[31] J. Browne, A. Tunnacliffe, A. Burnell, *Nature* **2002**, *416*, 38.
[32] K. Goyal, L. Tisi, A. Basran, J. Browne, A. Burnell, J. Zurdo, A. Tunnacliffe, *J. Biol. Chem.* **2003**, *278*, 12977 – 12984.
[33] S. Mondal, L. Adler-Abramovich, A. Lampel, Y. Bram, S. Lipstman, E. Gazit, *Nat. Commun.* **2015**, *6*, 8615.
[34] E. Tayeb-Fligelman, O. Tabachnikov, A. Moshe, O. Goldshmidt-Tran, M. R. Sawaya, N. Coquelle, J.-P. Colletier, M. Landau, *Science* **2017**, *355*, 831 – 833.
[35] P. Mazur, *Am. J. Physiol. Cell Physiol.* **1984**, *247*, C125 – C142.

Manuscript received: July 26, 2021
Revised manuscript received: November 3, 2021
Accepted manuscript online: November 9, 2021
Version of record online: November 25, 2021

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202109961, Wiley Online Library on [22/03/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Intrinsically Disordered Tardigrade Proteins Self-Assemble into Fibrous Gels in Response to Environmental Stress

*Anas Malki, Jean-Marie Teulon, Aldo R. Camacho-Zarco, Shu-wen W. Chen, Wiktor Adamski, Damien Maurin, Nicola Salvi, Jean-Luc Pellequer, and Martin Blackledge**

anie_202109961_sm_miscellaneous_information.pdf

Material and methods

Protein expression, purification, and mutagenesis

CAHS-8 was cloned into pET-28b(+) between NcoI and XhoI cleavage sites with a C-terminal 8His-tag preceded by a TEV cleavage site (ENLYFQG). They were expressed in E. coli BL21(DE33) cells for 5 hours after induction at an optical density of 0.6 with 1 mM of isopropyl-β-d-thiogalactopyranoside. The unlabeled protein expression was performed in LB medium whilst isotopically labelled protein expression for NMR (^{15}N and ^{13}C) was performed in M9 minimal media prepared in H_2O or in D_2O to produce ^{2}H-labelled proteins.
The cysteine mutations (A134C and A185C) were made by site directed mutagenesis using the quick change method [1].
The proteins were sonicated in 50 mM Tris, pH = 8.0 and 150 mM NaCl (lysis buffer) and then boiled at 95°C for 20 min for standard nickel beads purification. The elution was performed in the same buffer with 300 mM Imidazole, then the protein was incubated overnight with TEV protease while dialyzing against the lysis buffer, subjected to a second nickel beads affinity column and the flow-through was loaded into a Superdex 75/200 column for size exclusion chromatography (SEC) in 30 mM NH4HCO3 and lyophilized. The obtained proteins were weighted and resuspended in 50 mM phosphate buffer at pH = 6.5 and 100 mM NaCl (AFM buffer) or 500 mM (NMR buffer) at the desired concentration.
Paramagnetic labelling was done using maleimide chemistry. In short, the protein was incubated overnight with 10 mM DTT, then dialyzed against 50 mM NaP (pH = 7.0) and 150 mM NaCl. A 5 times molar excess of maleimide reactant (4-maleimido-TEMPO dissolved in DMSO) was added for overnight incubation. The excess of label was then removed through in the NMR buffer.
ANP32a IDR was expressed and purified as described.[2] Concentrations were verified using a Thermo Fischer NanoDrop 2000.
Sequence of the IDR of ANP32a :
DRDDKEAPDSDAEGYVEGLDDEEEDEDEEEYDEDAQVVEDEEDEDEEEEGEEEDVSGEEEEDEEGYNDGEVDDEE
DEEELGEEERGQKRKREPEDEGEDDD
N_{TAIL} IDR was expressed and purified as described.[3]
Sequence of the IDR of N_{TAIL} :
MSGGDGAYHKPTGGGAIEVALDNADIDLETEAHADQDARGWGGESGERWARQVSGGHFVTLHGAERLEEETNDED
VSDIERRIAMRLAERRQEDSATHGDEGRNNGVDHDEDDDAAAVAGIGGI
SH3A from CD2AP was expressed and purified as described.[4]

NMR Spectroscopy
NMR experiments were measured on 600, 700, 850 and 950 MHz Bruker spectrometers in the NMR buffer using 3 mm NMR tubes with 10% D_2O.
Backbone resonance assignment was performed using triple resonance experiments for Ca, CB and CO chemical shift correlation. Secondary Structure Propensity (SSP) and secondary chemical shift were done using the software SSP [5].
The relaxation rate experiments on the CAHS-8 free form were measured at 350 μM protein concentration whilst the experiments with the gel were measured at a total CAHS-8 concentration of 1.12 mM.
^{15}N R1ρ rates were measured as described using a spin lock of 1.5 kHz [6] with a set of delays of 1, 20, 50, 70, 90, 120, 160 and 200 ms with the repetition of two delays. A recycle delay of 1.5s was used.
^{15}N R1 and ^{1}H-^{15}N heteronuclear NOE rates were measured as described [7] with a set of delays for R1 of 0, 0.1552, 0.2716, 0.388, 0.5044, 0.6208, 0.7372, 0.9312 and 1.164 s with the repetition of one delay. A recycle delay of 1.5s was used.
^{15}N–^{1}H CSA/DD transverse cross correlated rates (ηxy) were measured as described [8,9] with a cross-relaxation delay of 50 ms.
The rates were extracted using an in-house software with Monte Carlo simulation-based noise-estimation. R2 rates were determined from R1 and R1ρ as described [10].
NMR Pipe [11] was used to process all spectra and CCPNMR Analysis [12] to analyze them. PRE effects were determined from the peak intensity ratios in a 15N HSQC between the paramagnetic state and the diamagnetic state obtained by addition of 5 mM ascorbic acid. A recycle delay of 1.575s was used.
For residual dipolar couplings (RDCs) measurement, the protein was aligned using ASLA Biotech pf1 phages at 15 mg/mL concentration. The scalar couplings were measured in CCPN and the RDCs

values were obtained by subtracting the couplings in the isotropic and anisotropic (phage aligned) measurements.
DOSY were measured with gradient strength of 1.016, 4.269, 7.522, 10.775, 14.028, 17.280, 20.533, 23.786, 27.039, 30.292, 33.544, 36.797, 40.050, 43.303, 46.556, 49.808 G/cm. ANP32a total concentration was 150 μM, NTAIL 300 μM and SH3A 150 μM.

SAXS
SAXS experiments were performed at BM29 BioSAXS in the ESRF Grenoble in the NMR buffer. The measurement used an incident beam wavelength of 0.99 Å and with a distance between the sample and the Pilatus3 2M detector of 2.867 m. This gives a scattering momentum transfer comprised between 0.0039 $Å^{-1}$ and 0.49 $Å^{-1}$. Every measurement comprised 10 frames which were analyzed and rejected in case of radiation damage. The solvent was measured before and after each sample and its scattering was subtracted from the sample's. The data were processed and analyzed using ScÅtter.
10000 conformers were calculated using *flexible-meccano*[13] assuming the helical domains α_1 (E95-F150) and α_2 (V167-L194) were either populated or not. Small angle scattering curves were calculated for each conformer using the program Crysol.[14] The ensemble selection algorithm ASTEROIDS[15] was used to calculate combinations of equally populated conformers that reproduced experimental data. Ensembles were calculated with 2 to 20 members. The optimal number was estimated to be 7. Radii of gyration were calculated for each conformer and the average over the best fitting ensemble is quoted in the main manuscript.

Dynamic Light Scattering (DLS)
DLS measurement were performed on a Malvern Zetasizer Nano S at an angle of 173° in a Malvern ZEN0040 cuvette. A temperature trend from 318K to 278K was used with 3 measurements (comprising 10 runs of 60s) performed every 5K after 300s of equilibration.
The protein phase was determined from the correlation function.[16]

AFM
Images were recorded on a Bruker multimode 8 microscope with a Nanoscope V controller using the PeakForce Tapping mode in liquid and soft tapping in air. For liquid imaging, a silicon nitride cantilevers (ScanAsyst Fluid, Bruker AFM probes, Camarillo, CA, USA) with the following nominal characteristics: k = 0.7 N/m, Fq = 150 kHz, r = 20 nm were used. A semi-automated ScanAsyst mode was used by switching off the autogain. A scan rate of 1 Hz was commonly used with a setpoint of about 0.1 V. For imaging in air, etched silicon cantilevers (RFESP-75, Bruker AFM probes) with the following nominal characteristics: k = 3 N/m, Fq = 75 kHz, r = 8 nm were used. A target amplitude of 300 mV (~13 nm) with a setpoint of about 185 mV was commonly used with a scan rate of 1Hz.
For air imaging, 2 μL of CAHS8 at 30 mg/mL, kept at 42°C, were deposed on mica or glass and dried for 5 min in air. For liquid imaging, 2 μL of CAHS8 at 30 mg/mL, kept at 42°C, was deposited on mica with 50 μL of AFM buffer, also kept at 42°C.
Images were processed using Gwyddion [17] : raw images were leveled by mean-plane subtraction, rows were then aligned by the median method, strokes were corrected and the minimal intensity was set to 0. For further analysis, images were denoised using DeStripe [18] and improved visibility was performed using the Laplacian weight method [19,20].

REFERENCES

[1] J. Braman, C. Papworth, A. Greener, in *In Vitro Mutagenesis Protocols* (Ed.: M.K. Trower), Humana Press, Totowa, NJ, **1996**, pp. 31–44.
[2] A. R. Camacho-Zarco, S. Kalayil, D. Maurin, N. Salvi, E. Delaforge, S. Milles, M. R. Jensen, D. J. Hart, S. Cusack, M. Blackledge, *Nat Commun* **2020**, *11*, 3656.
[3] K. Houben, D. Marion, N. Tarbouriech, R. W. H. Ruigrok, L. Blanchard, *J. Virol.* **2007**, *81*, 6807–6816.
[4] J. L. O. Roldan, M. Blackledge, N. A. J. van Nuland, A. I. Azuaga, *J. Biomol. NMR* **2011**, *50*, 103–117.
[5] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795–2804.
[6] N.-A. Lakomek, J. Ying, A. Bax, *Journal of Biomolecular NMR* **2012**, *53*, 209–221.
[7] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2016**, *138*, 6240–6251.

[8] P. Pelupessy, G. M. Espallargas, G. Bodenhausen, *Journal of Magnetic Resonance* **2003**, *161*, 258–264.
[9] P. Pelupessy, F. Ferrage, G. Bodenhausen, *J. Chem. Phys.* **2007**, *126*, 134508.
[10] M. Akke, A. G. Palmer, *J. Am. Chem. Soc.* **1996**, *118*, 911–912.
[11] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *Journal of Biomolecular NMR* **1995**, *6*, 277–293.
[12] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687—696.
[13] V. Ozenne, F. Bauer, L. Salmon, J.-R. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Bioinformatics* **2012**, *28*, 1463–1470.
[14] D. Svergun, C. Barberato, M. Koch, *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
[15] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *Journal of the American Chemical Society* **2010**, *132*, 8407–8418.
[16] M. Shibayama, T. Norisuye, *BCSJ* **2002**, *75*, 641–659.
[17] D. Nečas, P. Klapetek, *Open Physics* **2012**, *10*, 181–188.
[18] S. W. Chen, J.-L. Pellequer, *BMC Struct Biol* **2011**, *11*, 7.
[19] S. W. Chen, J.-M. Teulon, C. Godon, J.-L. Pellequer, *J Mol Recognit* **2016**, *29*, 51–55.
[20] S.-W. W. Chen, A.-S. Banneville, J.-M. Teulon, J. Timmins, J.-L. Pellequer, *Nanoscale* **2020**, *12*, 22628–22638.
[21] M. R. Jensen, M. Blackledge, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E1557-1558.
[22] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, M. Blackledge, *J. Am. Chem. Soc* **2008**, *130*, 8055–8061.
[23] J. H. Brown, C. Cohen, D. A. Parry, *Proteins* **1996**, *26*, 134–145.
[24] A. N. Lupas, M. Gruber, in *Advances in Protein Chemistry*, Elsevier, **2005**, pp. 37–38.

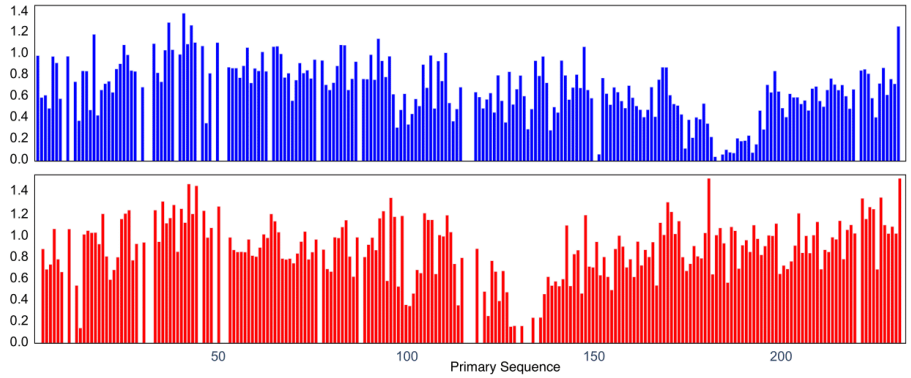


Figure S1. Paramagnetic Relaxation Enhancement (PRE) experiment with the spin label on positions 185 (top) 134 (bottom). The absence of significant enhancement beyond the immediate neighboring residue indicates that the protein is extended and is lacking a tertiary structure. Values that are greater than 1 are non-physical, and are due to experimental noise, also in part due to potential differences in pH between oxidized and reduced samples.

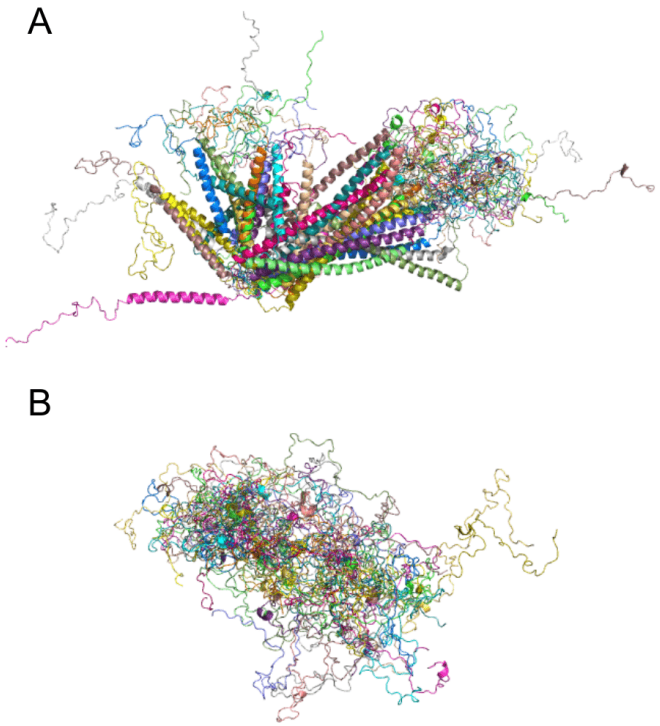


Figure S2. Representations of the dimensions of CAHS-8 in the presence of the helical domains (A), and assuming that these regions sample random coil statistics (B). The two ensembles have radii of gyration of 59±11 and 47±10 Å respectively.

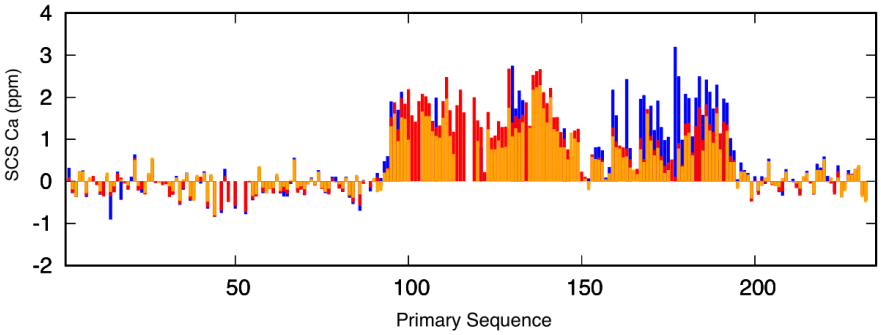


Figure S3. Temperature dependent conformational sampling. Temperature compensated secondary chemical shift for each backbone $^{13}C^{\alpha}$ are shown for 278K (blue), 293K (red) and 303K (orange). A clear increase in helical propensity in the α_2 helix is visible at 278K. Many peaks from the α_1 helix are not visible at this temperature.

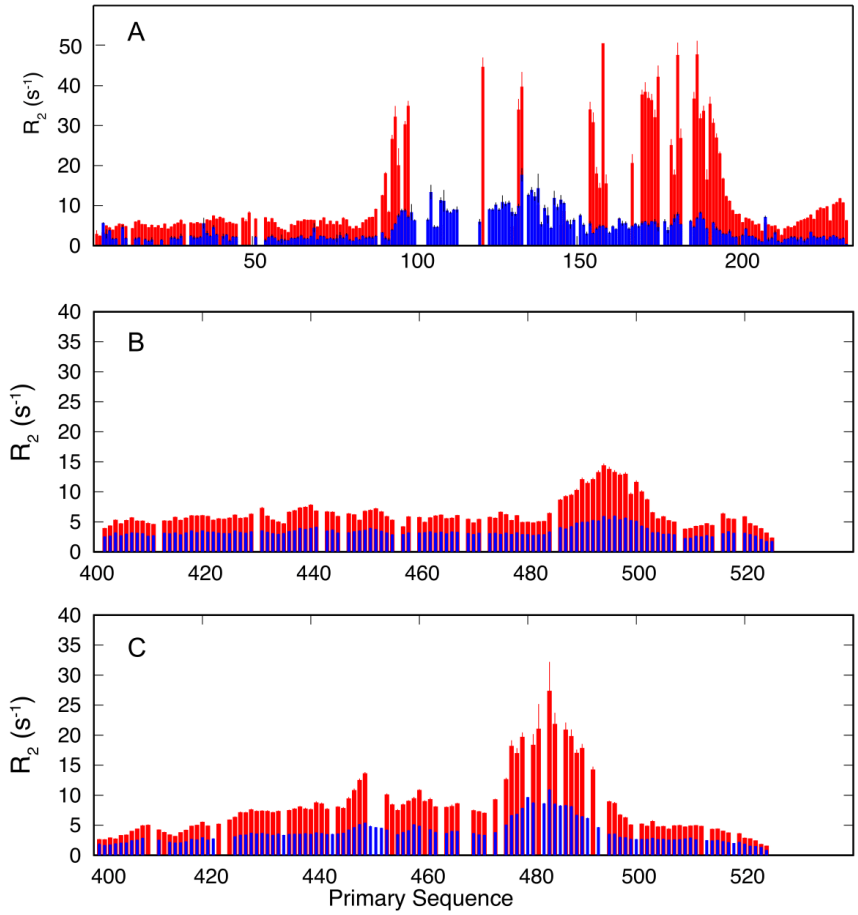


Figure S4. A Transverse ^{15}N relaxation (R_2) measured at 303K (blue) and 278K (red) at 850 MHz, showing very large increase in molecular tumbling at the lower temperature. B Transverse ^{15}N relaxation (R_2) of Measles virus nucleoprotein C-terminal disordered domain measured at 298K (blue) and 278K (red) at 850 MHz . The helical element (positions 490-500) is populated to approximately 90% in the central region.[21] C. Transverse ^{15}N relaxation (R_2) of Sendai virus nucleoprotein C-terminal disordered domain measured at 298K (blue) and 278K (red) at 850 MHz.[7] The helical element (positions 478-491) is populated to approximately 80% in the central region.[22] The increase in relaxation rates (and disappearance of signal) in the helical region of CAHS-8 is significantly greater than for the two other viral proteins that do not show signs of oligomerization at 278K.

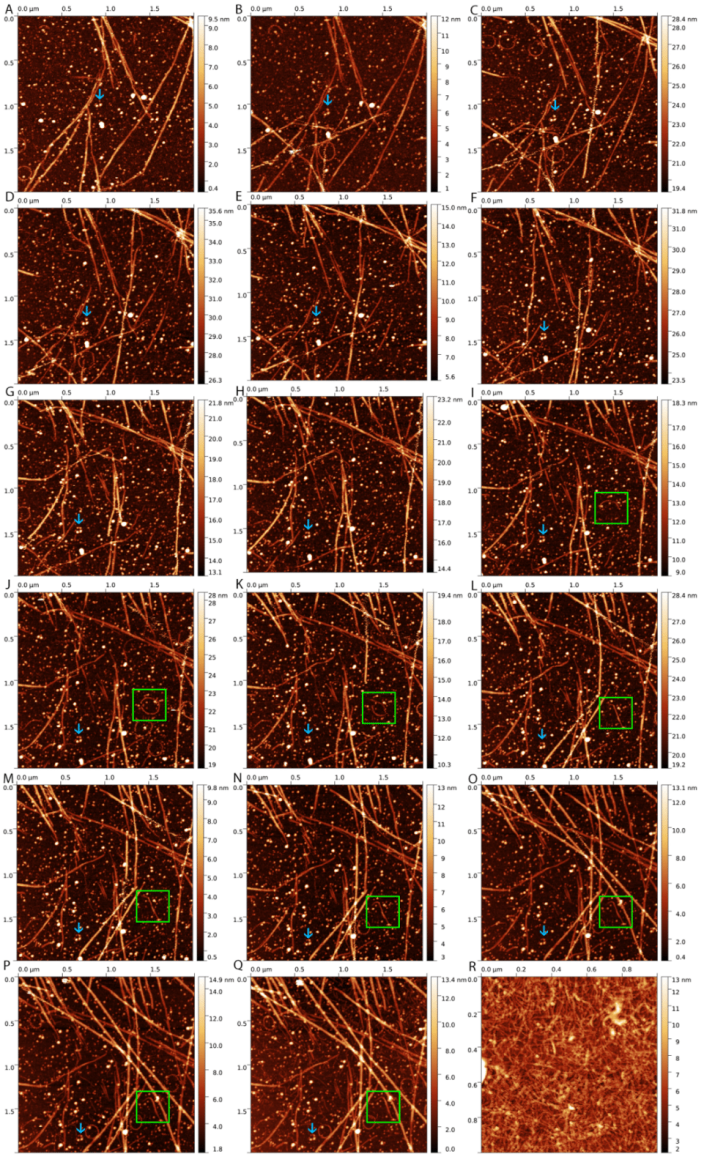


Figure S5. (A-Q) Liquid mode AFM images of the kinetic of fibrils formation. The 17 frames of the movies are separated by 8 min. The images show the fibrillation starting from a few already fully formed fibres to a mesh of longer and and smaller fibrils. The blue arrow denotes oligomers thought to be either aborted fibrils (possibly "off pathway oligomers") or prefibrillar oligomers. The green circle highlights small nascent circular fibrils that later strengthen to an early fibril.
(R). Air mode AFM image of the gel obtained by letting the sample imaged in (A-Q) dry displaying a network of interwined fibres.

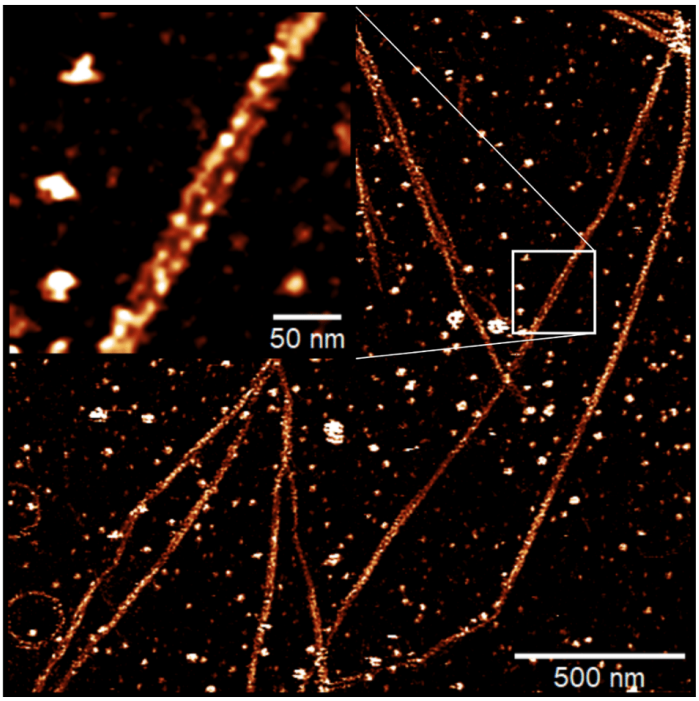
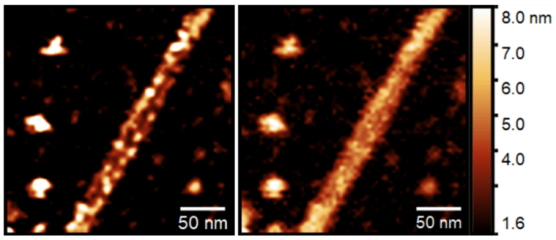


Figure S6a. Image processing of Fig. 3A using the triple L-weight filter.[19,20] The L-weight filter suggests a twisted pair of proteins stacking onto each other with their local long axis perpendicular to the fiber extension. The measured width of a single protein stacking is about 9 nm; which could correspond to the full length of the central folded α-helix. The original image has a physical size of 2 x2 μm² with 512 x 512 pixel². The inset is made using a computer magnification (X4) using the Schaum interpolation method with Gwyddion.[17] The size of the white square is 64 x 64 pixel² representing a physical size of 250 x 250 nm². Because the intensity values of processed images using the L-weight filter do not represent height values, there is no corresponding colour scale bar present on this filtered image. The colour ramp of pixel intensities has been manually adjusted to improve visibility.

Fibres are continuously growing and shrinking during AFM imaging and consequently, the thickness of fibers changes over time. In addition, some fibres lie on top of others, not directly in contact with the mica surface, and thus the tip-effect in such fibers is expected to be larger than for fibres directly adsorbed on the mica surface. The apparent twisting, evidenced by the change in height along a single fibre, is not expected to be an artefact of the tip-effect.
We have tested the processing on several images and obtained a similar result, suggesting a fibre made of pairs of proteins.

Figure S6b



Left, the same L-weight filter processed image as in Fig. 6a and on the right the corresponding original height image. The original image has been cropped from Fig. 3a with a physical size of 250 x250 nm² with 64 x 64 pixel², then the imaged was made using a computer magnification (X4) using the Schaum interpolation method with Gwyddion.[17] Because the intensity values of processed images using the L-weight filter do not represent height values, there is no scale bar present on the left image whereas the color bar on the right represents true height values. It can be seen that the middle of the cropped fibre is flatter (about 3 nm) than the two twisted extremities (about 6 nm).

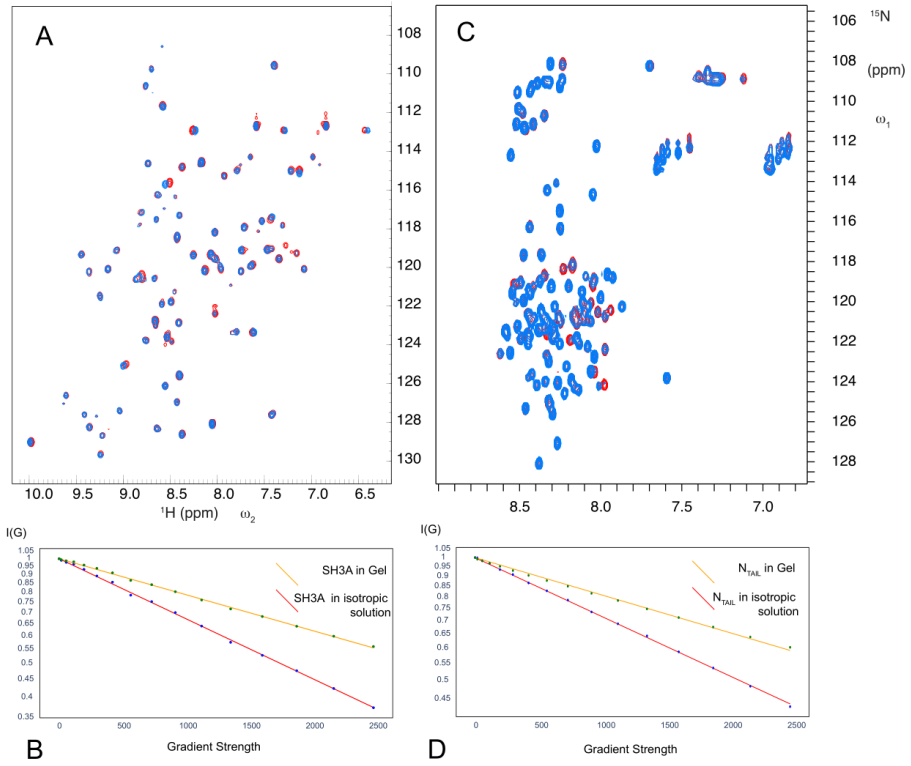


Figure S7.
Comparison of the properties of folded and unfolded proteins in the isotropic phase and in the gel.
A – SH3A domain from CD2AP. Red - ^{15}N-^{1}H HSQC spectrum in free solution, blue - spectrum in CAHS-8 gel. The intensity of the peaks is severely diminished in the gel, suggesting that a significant fraction of the protein becomes invisible. The visible fraction has the same fold, and (B) has a diffusion constant that is significantly slower than in free solution (Free : 2.18808203 10^{-10} m^2s^{-1} compared to 1.31157081 10^{-10} m^2s^{-1} in the gel).
The conditions are 150 μM concentration of protein, 298K, 600 MHz in NaP 50 mM, NaCl 500 mM, pH 6.5. The gel is at 20 mg/mL.

C – The disordered domain of Sendai virus nucleoprotein (N_{TAIL}) (red - ^{15}N-^{1}H HSQC spectrum in free solution, blue - spectrum in CAHS-8 gel). D - The protein has a diffusion constant that is again significantly slower than in free solution (Free : 1.83816502 10^{-10} m^2s^{-1} compared to 1.16630247 10^{-10} m^2s^{-1} in the gel).
Experimental conditions were 150 μM concentration of protein, 298K, 600 MHz in NaP 50 mM, NaCl 500 mM, pH 6.5. The gel is at 30 mg/mL.

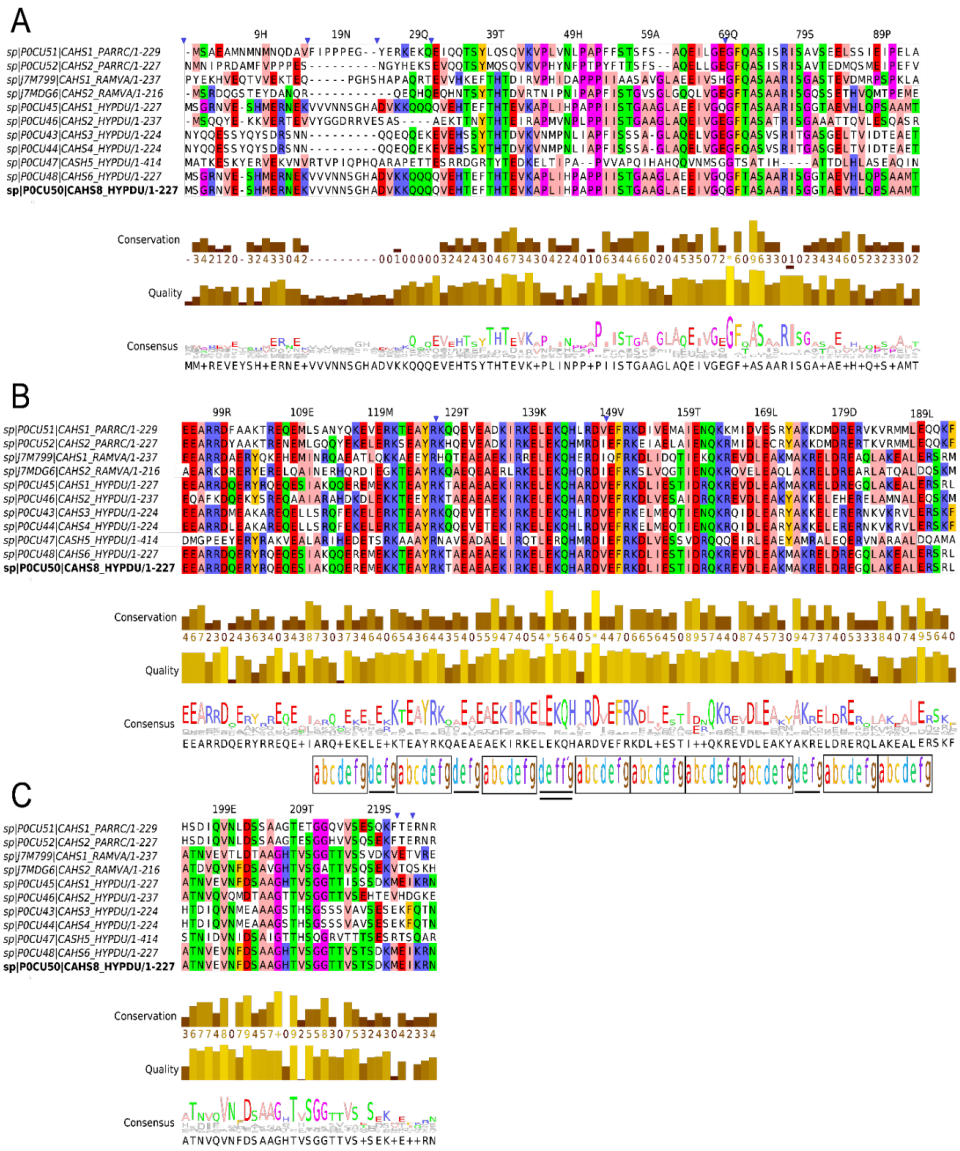


Figure S8. Sequence alignment of CAHS-8 with other CAHS from *Hypsibius exemplaris* (HYPDU), the closely related tardigrade *Ramazzottius varieornatus* (RAMVA) and the more distant *Paramacrobiotus richtersi* (PARRC). A, B and C correspond respectively to CAHS-8 disordered N-terminus (M1 – T94), its helical domain (E95 – F194) and its disordered C-terminus (A195 – N227). Blue arrows indicate hidden residues from the aligned proteins in order to visualize alignment on CAHS-8 only. The sequence alignment shows a significant conservation of T39 onward.
(B) The helical domain displays the highest amount of conservation and the analysis of its sequence reveals an irregular coiled coil [23,24] comprising 9 *a-b-c-d-e-f-g* canonicals heptads, 3 *d-e-f-g* stutters

(M119-E120-K121-K122, A130-E131-A132-E133 and A173-K174-R175-E176) and one *d-f-f'-g* five residues insertion (L142-E143-K144-Q145-H146). The position *a* and *d* being preferentially occupied by hydrophobic amino acids while the *e* and *g* by charged residues of opposites charges.
The irregularity of this coiled coil would make it unstable whose affinity may be modulated by conditions. In agreement with NMR measurements on the gel (fig.4), the helical domain of CAHS-8 would then be the fibrillation core with the coiled coil formation being its first step toward the formation of paired helical filament as already seen. The high conservation of the sequence between these CAHS indicate that they most probably have a similar structure (alpha helix that would assemble into coiled-coil), fibrillation and gelation properties.

Part 1: CAHS-8 – Monomer, coiled-coils and fibres

Chapter 5: CAHS-8 and its helices

« he is prone of old habit to believe that after all it is something more than a spiral »

D'Arcy Wentworth Thompson

On Growth and Form, 1917

Introduction

α-helices

First described in 1951 by Pauling, Corey and Branson[518], this type of secondary structure occurs when the protein backbone forms a helix with a turn – or linear translation – of 3.63 residues per turn and with a rise – or orthogonal circular rotation – of 5.4 Å per turn. This helical structure arises from hydrogen bonding between the carbonyl (CO) of a residue and the amide proton (NH) of the residues located 4 positions after (i+4), which therefore pulls the protein backbone into a curled shape to accommodate those repetitive hydrogen bonds that lie almost parallel to the helix axis. As for the sidechains, they physically stick outward from the helix thus stabilising it by limiting steric hindrances and giving them the freedom to interact, whether with other proteins, other domains of the same protein or with neighbouring residues like the physically close i, i+3 and i+4 sidechains due to the turn being 3.63 residues per turn[519].

Other kind of helices exist as well: if instead of occurring with the i+4 residue, the hydrogen bonding occurs with the i+3 residue, the structure is the less common 3_{10}-helix, which is mostly observed at extremities of α-helices and are rarely more than 4 residues long, while when it occurs with the i+5 residues, the structure is the even less common π-helix which is typically found within a longer α-helix and appear as a "bulge" in the helix axis[519,520].

The α-helices are overwhelmingly right-handed, with optimal backbone ϕ and ψ torsion angles often corresponding to the centre of a region of minimum energy in the Ramachandran plot while the preferred torsion angles of the other types of right-handed helices are less thermodynamically favourable, thus explaining the higher abundance of α-helices in regard to the others. The left-handed α-helix is also less favourable than the right one, due to the steric hindrance between the backbone carbonyl and the sidechains of L-amino-acids, is located in a region of the Ramachandra diagram far from the deep minimum corresponding to the different kind of β-sheets and the right-handed α-helix and thus making the left-handed α-helix exceedingly rare[520].

Because all the amides group of the amino-acids are oriented in the same axis, an α-helix has a dipole moment along this axis with the positive end at the N-terminus due to the three non-hydrogen-bonded NH groups and the negative end at the C-terminus end of the helix due to the three non-hydrogen-bonded CO groups[521,522]. Moreover, if the α-helix is long and folded or highly populated, it is not rare for long α-helices to display an anisotropic rotational diffusion as the overall shape of the helix would be that of a prolate spheroid[523].

Coiled-coils

Canonical double stranded coiled-coils

The important tertiary structure or quaternary structure for this thesis is the coiled-coil arrangement. Described for the first time in 1952 by Crick[524] and 1953 by Pauling and Corey[518] to explain the X-ray diffraction pattern of the α-keratin type of fibrous proteins that display a strong reflection at 5.1 Å instead of the 5.4 Å reflection expected for a canonical α-helix, a coiled-coil was described as a type of structure in which at least two adjacent α-helices coiling around each other to form a "super-helix"[518,524,525]. Indeed, Crick realised that if one α-helix would experience a left-handed axis rotation of ~20°, the sidechain from this helix could intercalate into the surface cavities between the sidechains of the facing residues from another other helix. Thus, he deemed this close interlocking arrangement *knobs-into-holes*[524,525] and noted that this it cannot happen if the monomers stay straight. Therefore, the formation of a canonical coiled-coil requires the bending of the individual α-helices, also called supercoiling, with the interface side being concave and the solvent-exposed side being convex[526,527]. Although somewhat similar as another interhelical mode of interactions, coiled-coils should not be confused with the more irregular "ridges-into-grooves" packing found in globular proteins where the sidechains of a residue packs either above or underneath the sidechains of the facing residue and does not imply a perturbation of the helix periodicity[528–531].

Coiled-coils form when an α-helix is amphipathic, meaning that one side presents predominantly polar amino-acids while the other side presents predominantly hydrophobic amino-acids. As hydrophobic groups tend to cluster together due to the hydrophobic effect, this particular nonpolar face of the helix forms a surface shielded from the water with another hydrophobic face of an α-helices, whether it is with the hydrophobic face of another α-helix of the same monomer, in which case it is a tertiary structure element, or with another protein, that can be either a monomer of the same protein if it is a homo-oligomeric coiled-coil or an altogether different protein in the case of a hetero-oligomeric coiled-coil[525–527,532].

Coiled-coils have an ideal periodicity of seven residues, with the particular position of this heptad repeat being noted **abcdefg**, and comprise at least 2 of such repeats. As the helices face each other in a parallel fashion, with the second denoted prime, and thus with a central symmetry point between them, the residues **a** and **d'** of the heptads are in close vicinity and therefore have to be hydrophobic to form the core of the superhelix through the hydrophobic effect. It is therefore possible to predict a coiled-coil structure by looking for a frequency of hydrophobic acids repeats of $i\text{-}i_{+4}\text{-}i_{+7}$ in the primary sequence of a protein. In turn, the **e** and **g** of both helices are solvent-exposed amino-acids that face each other a bit more distantly on the side of the core, and thus are typically pairs of charged residues which will further stabilize the coiled-coil through salt-bridges. As such, these **e** and **g** amino-acids of the repeats are important and give the coiled-coil its pairing specificity, as the repulsion between these residues can outweigh the hydrophobic effect[525,526,533]. Finally, the remaining **b**, **c** and **f** residues of the two helices are ideally hydrophilic amino-acids as they are exposed to the solvent. In turn, If the coiled-coil is antiparallel, the symmetry is axial instead of central and thus, the residues at the same position of the heptad face each other, with the amino-acid in position **a** facing the **a'** residue of the other helix or helices instead of **d'** and therefore **d** facing **d'** instead of **a'** and so on. This ideal periodicity of seven over two turns of the helix has therefore the effect of changing the periodicity of the individual α-helices from 3.63 to $7/2 = 3.5$ residues per turn instead, which can be seen for example in the well-studied canonical 7/2 coiled-coil leucine-zipper GCN4-pLI (PDB: 1GCM)[526,527,531–534].

However, canonical coiled-coils often present imperfections in their distribution of amino-acids in the heptad repeats, in particular the presence of polar residues in position **a** or **d**, which are therefore buried in the hydrophobic core and not only constitute on average about a quarter of the interface residues but are often highly conserved. This typically destabilises the coiled-coil but also increases the specificity of the interaction and can change the oligomeric state of the assembly, which can be attributed to the reduction of the hydrophobicity of its interface, particular electrostatic attraction or repulsion as well as the sidechains of polar residue favouring conformations in which their polar group can be solvated, which therefore favours dimeric coiled due to their smaller hydrophobic core than higher order coiled-coils[526,527,533,535,536]. This effect was notably demonstrated for the GCN4 coiled-coil where an asparagine residue in position **a** of the heptad, and therefore buried, controls the specificity and stability of the interaction as mutating it for hydrophobic amino-acids results in a different number of helices involved in the coiled-coil but also a greater stability[536–538] while mutating it to a glutamine also result into a lower stability but doesn't change the stoichiometry[539]. A similar behaviour was also demonstrated for the trimeric autotransporter adhesins where the introduction of asparagines at position **d** of the heptad progressively destabilised the structure while making it more specific but still trimeric, thereby increasing the minimum concentration at which the coiled-coil formed[540].

Discontinuities and non-canonical repeats

The heptad repeats of coiled-coils often present discontinuities. Typically, residues are inserted between the heptads which perturb the coiled-coil stability and periodicity but not significantly enough to abrogate it. The insertion of three residues between heptads is called a *stammer* and have the effect of overwinding the helices by reducing the helices local periodicity to $(7 + 3)/3 = 10/3 \approx 3.33$ residues per turn as this insertion of three residue is smaller than the 3.5 residue per turn of a canonical coiled-coil. In turn, the insertion of four residues is called a *stutter* and unwinds the coiled-coil by raising its periodicity to $(7 + 4)/3 = 11/3 \approx 3.66$ residues per turn as four residues is more than one normal α-helix turn. Although sometimes also noted **defg**, these extra stutter residues are commonly noted as **hijk** which will be used in this thesis. Consequently, because of

this change of periodicity, the core and its packing are also altered. For instance, because of the unwinding of the supercoil, the neighbouring repeats have to compensate the residue positions and therefore, the **h** position of a stutter is projected towards the centre of the core in-between residue **e** of the previous repeat and **a** of the next. Because this residue faces directly the other helix, the knobs-into-holes packing is not possible anymore and is locally changed into the *knobs-to-knobs* packing in which the particular sidechain of this **h** residue directly faces the sidechain of the facing **h'** position instead of packing into a hole formed the sidechains of the two residues facing it like in the canonical knobs-into-hole packing. Thus, only small amino-acids such as alanines or glycines may fit without clashes, which are otherwise avoided by the coiled-coil adopting an antiparallel conformation. Indeed, such symmetry not only allows the sidechains to point towards different positions and therefore prevents the clashes, but also as is not forced to pair a particular repeat with a mirrored version of itself but can be instead paired with another distant repeat and thus can induce the pairing an **a** residue with an **a'** one that ought to be smaller, or alternatively, with a polar amino-acids with its aliphatic chain buried in the core and its polar groups being solvent-exposed[526,531,541–544].

Those discontinuities are not rare, in particular stutters, and their repetition over several heptads can durably disrupt the periodicity of the coiled-coil. However, it is incorrect to consider them solely as perturbing element as the presence of stutter or stammers allow to accommodate a nearby discontinuities, for instance the insertion of one residue called a *skip*. Moreover, stammers and stutters may not be present only as discontinuities but be instead regularly repeated along the sequence, which durably perturbs the periodicity of the coiled-coil which therefore departs from the canonical model. While Pauling and Corey already noted in 1953 that other periodicities than heptads may also be possible[518], the discovery of repeated motifs of 11 residues in several proteins such as LEA proteins[545] or surface protein tetrabrachion of archaea[546] lead to the realisation that the regular repetition of an heptad (7) and a stutter (4) forms an hendecad repeats (11, also called undecatad, noted **abcdefghijk**) and that these non-canonical repeats can also form coiled-coils. Eventually, evidences of pentadecads (15) which consists in a hendecad (11) and another stutter (4) and therefore with a $(11 + 4)/4 = 15/4 = 3.75$ residue per turn accumulated as well as the discovery of common additional 15/4, 18/5,

20/6 or 25/7 periodicities that all consist in diverse combinations of heptads, stammers and stutters[526,527,531,532,541,543,547]. It is noteworthy that while canonical heptad-based coiled-coils are left-handed, the hendecad-based coil-coils are essentially straight due to the relaxation of the supercoiling provided by the stutter while pentadecad-based coiled-coil are right-handed due to the further unwinding of the supercoil provided by the second stutter resulting in a supercoiling in the other direction[527,531,533,541,547]. This can be seen by comparing the structure of the aforementioned canonical 7/2 coiled-coil GCN4-pLi (1GCM) with the 11/3 tetrabrachion which consists of 4 essentially straight helices (1FE6)[548], while a notable example of a 25/7 periodicity can be observed in the coiled-coil domain of the Sendai virus phosphoprotein (1EZJ) which consists of 4 helices being initially straight in their N-terminal but showing a slight left-handed supercoil afterward[531,549].

Geometrical analysis of coiled-coils

The orientation positional angle, or Crick angle, of the residues corresponds to the angle they form with respect to the centre of the coiled-coil on their plane and therefore represents the phase of the coiled-coil, with consecutive residues separated by 360°/P with P being the periodicity of the coil. In a canonical coiled-coil, $P = 7/2 = 3.5$ and thus, each residue of the heptad is separated by $\sim 102.8°$[550,551]. As such, they show a clear clustering by position in the heptad when these Crick angles are plotted versus the sequence. For instance, the buried residues in **a** and **d** positions are oriented toward the core of the coiled-coil and have an average orientation of respectively 17° and -34°, which indicate that **a** is typically twice as close to the core than **d** and explain why buried polar residues have different effects depending on their position[544,550,552]. In turn residues at position c and f typically points to the outside and are therefore close to ±180°. As such, Crick angles can be used to determine the heptad position of the residues[550]. Observation of the Crick angles can therefore also used to detect discontinuities, as for instance, in the case of the insertion of a stutter which corresponds to approximately 3.67 residues per turn, the residues are separated by $\sim 98.2°$ instead and thus, they cluster differently. It is however noteworthy that these phase shifts are often modulated by a constant, in

particular if the coiled-coil oligomer is of a higher order than two or antiparallel, and therefore, it is not rare to see variations of their distribution[544,551].

The difference between the expected Crick angles and the experimental ones is called the axial rotation of the helices and is therefore also useful to detect deviations. Indeed, using the Crick angles for a canonical 7/2 coiled-coil, the closer the values are to a deviation of 0°, the closer they are to an ideal geometry. Experimentally, canonical heptad-based coiled-coils tend to have a variation of ± 26° and therefore, bigger discrepancies indicate departure from this geometry[531,544,553].

Finally, another metric is the axial shift of the helices which is the offset along the axis of the helix between a residue and the equivalent residue on the other helix. If perfectly aligned, their helical axial shift is null, while if two facing residues are shifted by x Å, the shift will therefore be equal to respectively +x Å and -x Å. As such, helical axial shift can be used to detect local variations of the helices that modulate the offset between the helices. This axial helical shift is typically close to 0 for parallel coiled-coils but has a broader distribution for antiparallel ones, as steric clashes can arise between the **a** and **a'** residues of the helices if one is not small, which is therefore accommodated by the shifting of one helix relative to the other and therefore of one **a** residue towards a **d** one, which probably explains why antiparallel coiled-coils have particular preferences for the amino-acids in **a** and **a'** positions as well as **a** and **d**[542,544,550,553–555].

Coiled-coils based fibrils

As discussed previously, coiled-coils have been first theorised to explain the fibre diffraction patterns of keratin-type of fibrous proteins. Decades of subsequent development have shown that coiled-coils are found in numerous fibre-forming proteins as they allow the formation of elongated and repetitive structures from a simple distribution of residues in α-helices, through the polymerisation of coiled-coils base-

units[526,533]. Unless specified otherwise, all structures discussed in this section were resolved via X-Ray diffraction crystallography.

The most relevant type of fibrous proteins for this thesis are the intermediate filament (IF) proteins which form, along microtubules and actin, one of the filament type of the cytoskeleton of animals[556]. The previously mentioned keratins are one of the types of IF proteins which also include the nuclear lamins as well as vimentin, which is the most common one. These IF proteins share an organisation in domains, historically called the N-terminal head, the so-called rod-domain made of the α-helical domains 1A, 1B, 2A and 2B separated by three linkers called L1, L12 and L2, as well as a final by a C-terminal tail. Indications that they form coiled-coils which would form the base unit for their polymerisation into filaments have accumulated since the fibre diffraction experiments of Crick on keratins in the 1950's, although the resolution of their structure has taken decades and is not finished to this day[524,557–559].

Indeed, although a crystalised fragment of the short domain 1A of vimentin was shown to be a lone amphipathic α-helix in 2002[560], the predicted canonical two-stranded parallels coiled-coil of this entire domain was only resolved in 2009 for vimentin using mutations of the **a** and **d** positions in the repeats[561] and in 2020 for lamins using fragments fused with capping motives[562]. However, no structure of domain 1A is available yet for keratins.

As per the longer domain 1B, it was first resolved in 2011 for vimentin by a combination of X-Ray crystallography and Electron Paramagnetic Resonance (EPR) for a construct containing the linker L1 and the domain 1B, revealing two parallel coiled-coil pairs assembling antiparallel into a tetrameric protofibril (PDB: 3UF1)[563], which was also shown in 2019 for lamin for the longest IF protein construct resolved yet which comprises part of domain 1A, linker L1, the entire helical domain 1B, linker L12 and a short segment of helix 2A (6JLB)[564]; as well as in 2019 also for keratins (6E2J)[565]. These structures demonstrated that the linkers L1 and L12 actually adopt an α-helical conformation and are crucial for interaction of the coiled-coil pairs to form the fibrils, being not only part of the interaction sites but also showing bending necessary to accommodate it[559,564].

In turn, the only complete structure of the fragment 2A of any IF protein were resolved for vimentin in 2010, revealing the expected parallel dimeric coiled-coil, and include the linker L2 as well as a part of domain 2B (3KLT)[566], of which a segment fused in a chimeric construct was resolved first in 2002 (1GK4)[560]; as well as in 2004 for lamins (1X8Y)[567] and in 2012 for keratins (3TNU)[568]. Additionally, structures of the almost-complete domain 2B of keratins were resolved in 2014 (4ZRY)[569] and 2020 (6UUI)[570]. The 2002 domain 2B coiled-coil vimentin structure revealed the presence of a stutter after eight heptads that unwind the coiled-coil to accommodate the π-π stacking of two phenylalanines[560]. This stutter sequence appears to be highly conserved amongst all IF proteins despite variability around it[560,571], appearing also in the domain 2B coiled-coils of the lamins[567] and keratins[568,569] and has been hypothesised to allow the bending and deformation of the structure[559,572]. This was followed by the realisation in 2006 that the domains 2A and linker L2 actually followed a hendecad periodicity and that this linker actually also adopted an α-helical conformation like L1 and L12[573] which was observed in the 2010 vimentin structure[566] while the hendecad periodicity was subsequently studied in greater detail via mutations and observation of the Crick angles[574].

It therefore appears that coiled-coil presenting a particular complementary sequence on its outer faces, that the considered coiled-coil can't interact with but another can, allows the polymerisation of many coiled-coils units into fibres. Thus, peptides can be designed to form coiled-coils that also present these "sticky-ends". Several such designed systems have been presented, for instance using canonical dimeric parallel coiled-coils that form fibres via electrostatic interactions either parallel[575,576] or antiparallel[577], trimeric coiled-coils with one antiparallel strand that polymerise through exposed tryptophans π-π stacking[578] , or modulating the stoichiometry of the coiled-coil assembly to nanotube-like fibres[579].

Additionally, it is noteworthy for the second part of the results of the thesis that such designed coiled-coils can also form reversible hydrogels at low temperature, either through the formation of fibrils[580] or without if the designed sequences have not only coiled-coils forming regions that interact together but also disordered chains[581,582].

4.1 Results

4.1.1 CAHS-8, an IDP with a long helical domain

Expression and purification protocols were developed during the first months of the thesis in order to obtain a high quantity of pure CAHS-8, which was initially challenging as the protein gelation was not understood and would clog chromatography columns or concentrators. This protocol allowed to isotopically label the protein in order to measure heteronuclear ^{15}N-^{1}H HSQC spectra (fig. 1 A) which displays the narrow proton chemical shift dispersion associated with intrinsically disordered proteins (IDPs). The protein was assigned using triple resonance experiments using two samples, requiring deuteration as approximately one third of the peaks of the ^{15}N-^{1}H HSQC spectrum of ^{13}C,^{15}N labelled CAHS-8 had no detectable transfer in triple resonance assignment experiments of the protonated sample, probably because of dipole-dipole induced proton relaxation.

The protein backbone resonances were assigned to more than 95% and the experimental chemical shifts were compared to the random-coil chemical shits of the sequence of the protein corrected for the experimental conditions of pH, temperature and salt concentration[583]. The obtained secondary chemical shifts Δδ (fig. 2) shows that CAHS-8 comprises two fully disordered termini (residues 1-94 and 195-233) flanking two helical regions (residue 95-151 and 160-194). Indeed, the secondary chemical shifts of the alpha carbon (Δδ $^{13}C^{\alpha}$, fig. 1 B) and carbonyl carbon (Δδ $^{13}C'$, fig. 1 C) vary around 0 while the two central regions show consistently positive values, which is associated with an α-helical secondary structure. This is also confirmed by the negative secondary chemical shifts for the amide proton (Δδ $^{1}H^{N}$, fig. 1 D) of this central region which are also associated with the presence of α-helices.

In order to detect a possible tertiary structure, backbone amide ^{15}N-^{1}H Residual Dipolar Couplings (RDCs or $^{1}D_{NH}$) were measured at 293 K at 350 µM. RDCs require an alignment with respect to the static magnetic field axis and report on the ensemble averaged orientational properties of the internuclear bond-vector. They are commonly measured to obtain information both on local and long-range order in the protein.

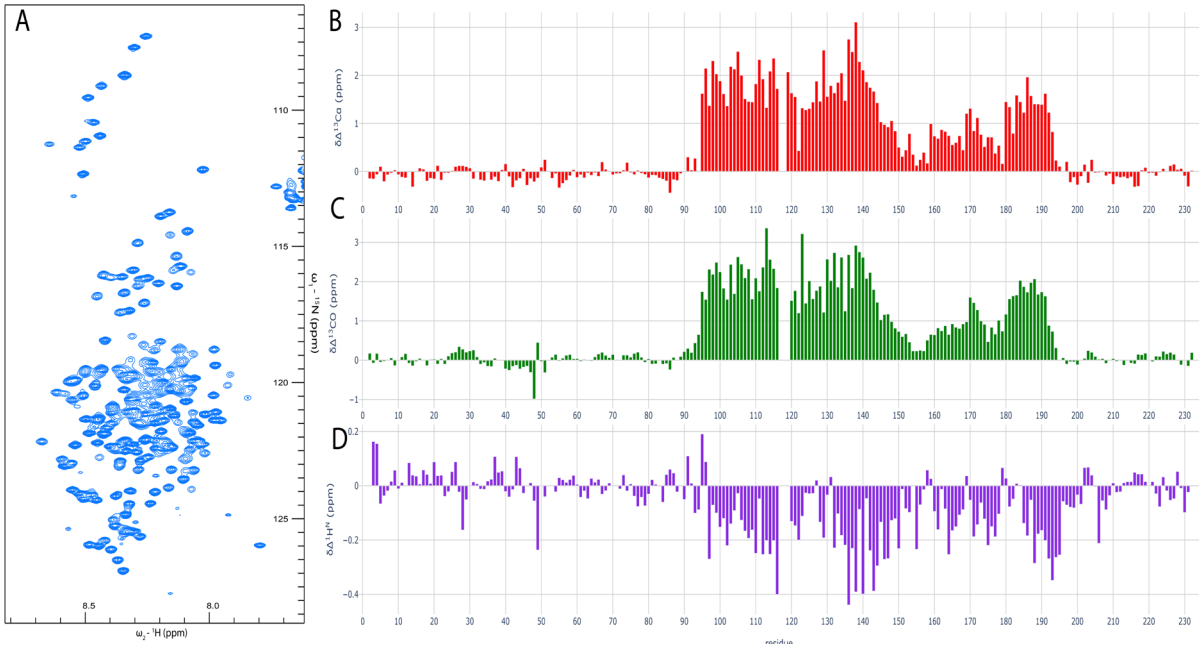


Figure 5.1: CAHS-8 is an intrinsically disordered protein with a central helical domain.
A: ^{15}N-^{1}H HSQC spectrum of a ^{2}D ^{13}C ^{15}N labelled CAHS-8 recorded at 293 K and 850 MHz.
B: Secondary chemical shifts of the α-carbons Δδ ^{13}C$^{\alpha}$ of CAHS-8 at 293 K.
C: Secondary chemical shifts of the carbonyl carbon Δδ ^{13}C′ of CAHS-8 at 293 K.
D: Secondary chemical shifts of the amide proton Δδ ^{1}H^{N} of CAHS-8 at 293 K.
Positive values for the secondary chemical shifts of the C$^{\alpha}$ and C′ and negative values for H^{N} are associated with α-helical secondary structure. The experimental chemical shifts were corrected for the deuterium isotopic and the reference chemical shifts were calculated using POTENCI[1] taking into account not only the protein's sequence but also the temperature, pH and ionic strength of the buffer.

^{15}N-^{1}H RDCs obtained for CAHS-8 (fig. 3 A) show negative values for the disordered regions, and positive values for the helical domains identified from the ^{13}C chemical shifts, in agreement with numerous studies of IDPs containing helical elements[584–588]. This is due to RDCs reflecting the average orientation of internuclear bond vectors. In disordered chains, backbone amide N-H bonds are on average perpendicular to the main axis of the chain, which in the case of steric alignment, tends to corresponds to the direction of the static magnetic field B_0. However, in helical regions, backbone amide NH- bond vectors are rather expected to align with the direction of the chain and therefore have an inverted sign with respect to the non-structured regions. The presence of negative $^{1}D_{NH}$ values at 152, 154 and 166 concur with the smaller Δδ ^{13}C$^{\alpha}$ values (fig. 1 B) supporting the observation that the helical domain can be separated in two α-helices, with the first one (helix 1, 95-151) showing an apparent higher helical propensity, i.e. carbon Δδ values, than the second one (helix 2, 167-194). Secondary-structure propensity calculation algorithm indicates that the first helix is populated to approximately 70%, the linker to

approximately 15% and the second helix to 50%[589]. The magnitude of the RDCs depends on the length of a helix, which may explain the difference between the two helices of CAHS-8 as well as the difference in the helical population.

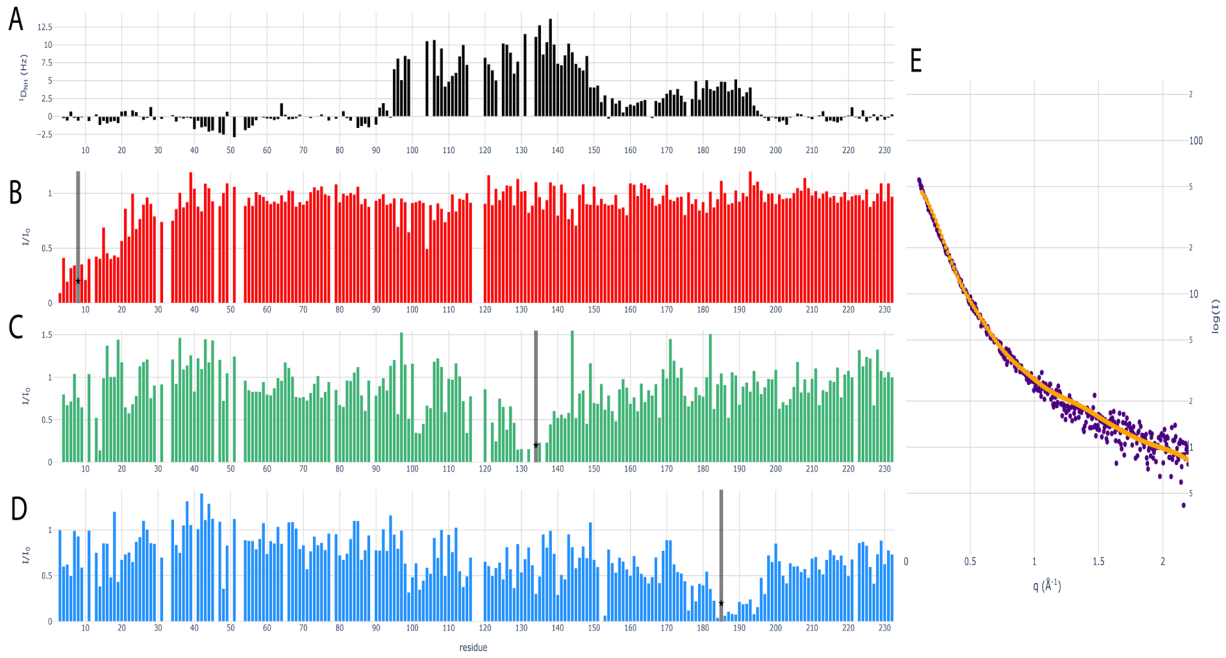


Figure 5.2: Absence of tertiary structure elements in CAHS-8.
A: Residual Dipolar Couplings (RDCs, $^1D_{NH}$, in black) and Paramagnetic Relaxation Enhancement (PREs) experiments of CAHS-8, with the paramagnetic spin-label position highlighted in grey and with a black star, being respectively in position **B**: 8 (red), **C**: 134 (green) and **D**: 185 (blue). **E**: Experimental SAXS profile of CAHS-8 at 75 µM i.e. 2 mg/mL (violet) and calculated SAXS profile (yellow) from the conformational ensemble selected using ASTEROIDS.

This absence of tertiary structure is confirmed by Paramagnetic Relaxation Enhancement (PREs) experiments performed with the paramagnetic spin-label on position 8, 134 and 185 (fig. 2 B, C and D) measured at 293 K and 150 µM to avoid intermolecular PREs effects. Although impaired by experimental noise, the results show no clear broadening beyond the neighbouring residues, supporting the absence of tertiary structure.

Moreover, SEC-MALLS experiments confirmed that CAHS-8 is monomeric in solution as no higher order oligomers could be detected at room temperature. Using these chemical shifts and RDCs information, models of CAHS-8 as a monomeric protein with disordered termini and two central helices separated by a linker was generated using the IDP ensemble generator flexible-meccano[590] and an ensemble of 7 conformers was selected using ASTEROIDS[591]. The obtained conformational ensemble comprises monomeric

CAHS-8 proteins with variably populated helices compatible with the secondary chemical shifts and helical propensity. The average radius of gyration of this ensemble $R_{g,av}$ = 59 ± 11 Å reproduced the experimental SAXS-derived radius of gyration obtained using the Guinier method $R_{g,exp}$ = 58 ± 7 Å (fig. 2 E).

4.1.1.1 Oscillating patterns

In order to understand the dynamics of CAHS-8, backbone ^{15}N amide nitrogen longitudinal (R_1) and transverse (R_2) autorelaxation as well as {^1H}-^{15}N heteronuclear nOe and cross-correlated transverse (η_{xy}) relaxation rates were measured at 600 and 850 MHz at 350 µM and 293 K. The obtained data (fig. 3) are consistent with the secondary structure derived from the secondary chemical shifts, PREs and RDCs. Indeed, R_1 shows a dip in the central helical domain while R_2 and η_{xy} show higher values, both indicating the presence of smaller amplitude fast motions and more substantial slower motions. The linker appears clearly in R_2 at 600 MHz, showing smaller average values (~6.5 s^{-1}) than helix 1 (~17 s^{-1}) and helix 2 (~10 s^{-1}), indicating faster motions on the nanosecond and above timescale. As per nOe, the helical region appears rigid, as seen from the high values even at 600 MHz, while the disordered regions appear flexible, that is with low nOe rates, with the exception of the 220-230 region the C-terminus that appears more rigid than the rest, possibly due to the local amino acid composition slowing-down the fast motions in this region as this region does not appear unusual in neither R_1 nor R_2.

The relaxation profile displays a significant difference between the R_2 values at 600 and 850 MHz. Indeed, at 600 MHz, the average R_2 rates in the first helix is approximately 17 s^{-1} at 600 MHz while it is closer to 28 s^{-1} at 850 MHz. This very large difference is unlikely to be due to CSA-induced transverse relaxation alone, as the dipolar contribution to R_2 is

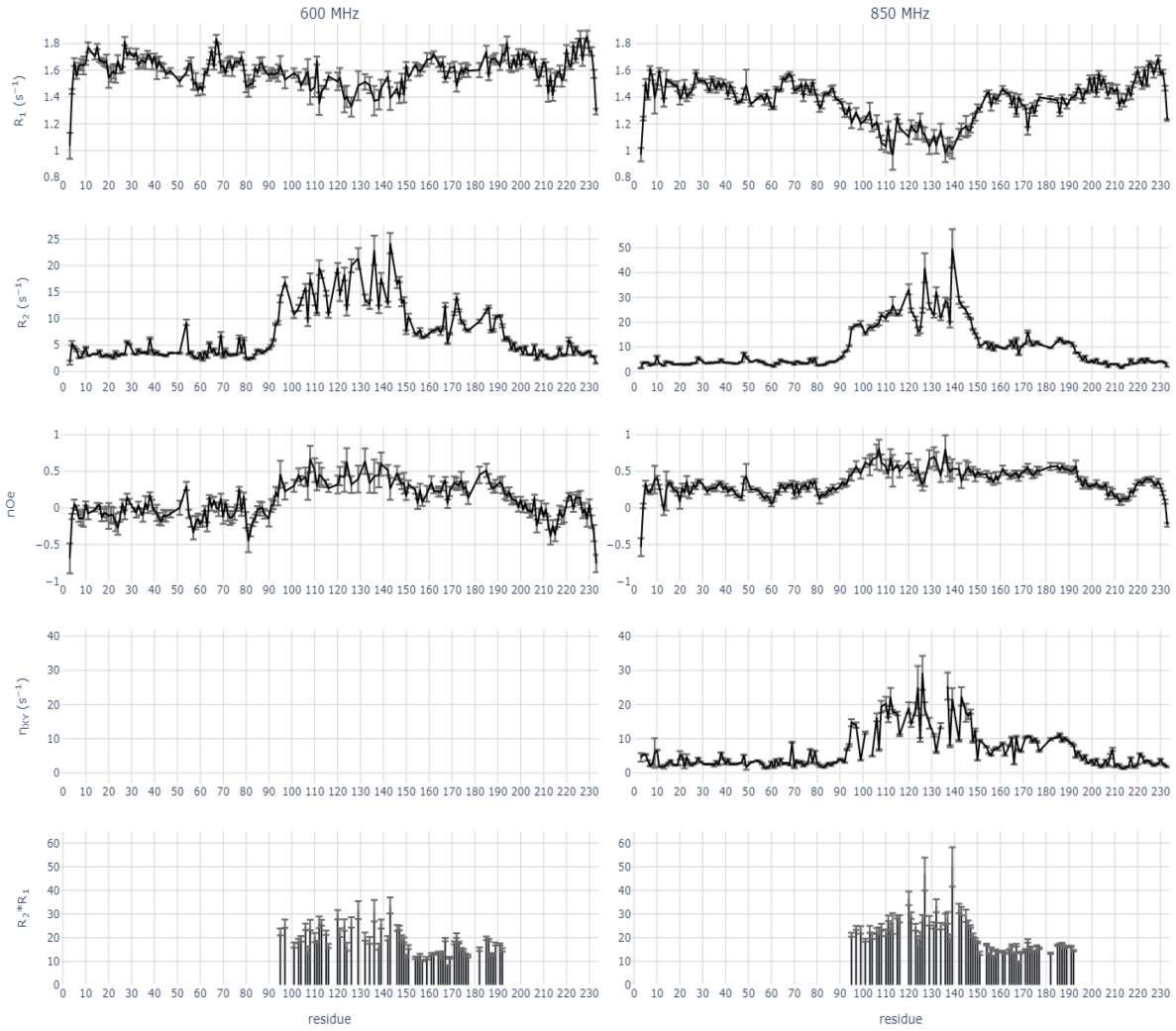


Figure 5.3: Relaxation of CAHS-8 at 293K
From top to bottom: R_1, R_2, nOe and η_{xy} relaxation as well as the R_2R_1, at 600 Mhz (left) and 850 (MHz)

independent of magnetic field strength, and is therefore more likely to be indicative of a chemical exchange. As the R_2 rates were calculated using $R_{1\rho}$ rates measured with a spin-lock of 1500 Hz, any chemical exchange process slower than this should be quenched, which indicates that this putative exchange phenomenon should be faster than 1500 Hz.

Periodic oscillations are visible in the relaxation profile in the first helix, in particular in R_1 and R_2 at 600 MHz and in η_{xy} at 850 MHz (fig. 3). In anisotropically tumbling molecules, it is necessary to consider a rotational diffusion tensor, characterised by different reorientation about the three orthogonal axes which is important to note as the unusually long length of the first helix (55 residues) may give rise to an anisotropy of the diffusion

tensor. Due to the orientational properties of the dominant relaxation-active interactions, which are the N-H dipole-dipole interactions in this case, it is possible to detect an orientational dependence of measured ^{15}N relaxation properties along a helix[592–594]. In this case the observed periodicity should match that of the helix, that is every 3.6 residues, however it is clear that while a periodicity in CAHS-8 is present, the period is rather every 7th or 8th residue. While the presence of a putative fast-exchange contribution to R_2 could mask or alias such effects, the periodicity still appears in R_1 and thus, the origin of this periodicity is not simple orientation of the helices relative to the diffusion tensor. It should also be noted that in the presence of a dynamic equilibrium, where 30% of the signal is apparently coming from non-helical conformers, such an analysis is probably beyond the signal to noise of these datasets.

Using the R_2R_1 product in the helical domain can be used to highlight chemical exchange contributions to R_2 which manifests as larger values when exchange is present (fig. 3, bottom row). Indeed, if $\omega\tau_c \gg 1$ and in the absence of internal motions, the R_2R_1 product is essentially constant at a given field while the contributions of the rotational anisotropy tend to cancel. As the Larmor frequency ω depends on B_0, using the highest available magnetic field strength ensure a better fulfilment of this slow tumbling limit condition[595,596].

A wave like pattern is also observed in the secondary chemical shifts of the amide proton $\Delta\delta\ ^1H^N$ of helix 1 in the regions 95-98, 127-133 and 140-145 and in 180-185 in helix 2 (fig. 4). Oscillations of the $\Delta\delta\ ^1H^N$ values around 0 in folded helices indicate a bending of the helix. Indeed, as the hydrogen bonds in the concave side of the helix are shorter than in the convex side, they experience an increased deshielding effect from the carbonyl with which they form the hydrogen bond and therefore show a bigger $\Delta\delta\ ^1H^N$ than the convex side[597–599]. However, the helices of the CAHS-8 are not fully populated as seen in the $\Delta\delta\ ^{13}C^\alpha$ values (fig. 4), resulting in the contribution of disordered conformations mitigating the deshielding effect to the ensemble-averaged measured chemical shifts resulting in smaller shifts. Localised curvatures of helix 1 nevertheless appears highly probable due some residues having positive secondary chemical in the 95-98 region, with a bend centred around A97 as this residue presents a clear break of continuity with neighbouring

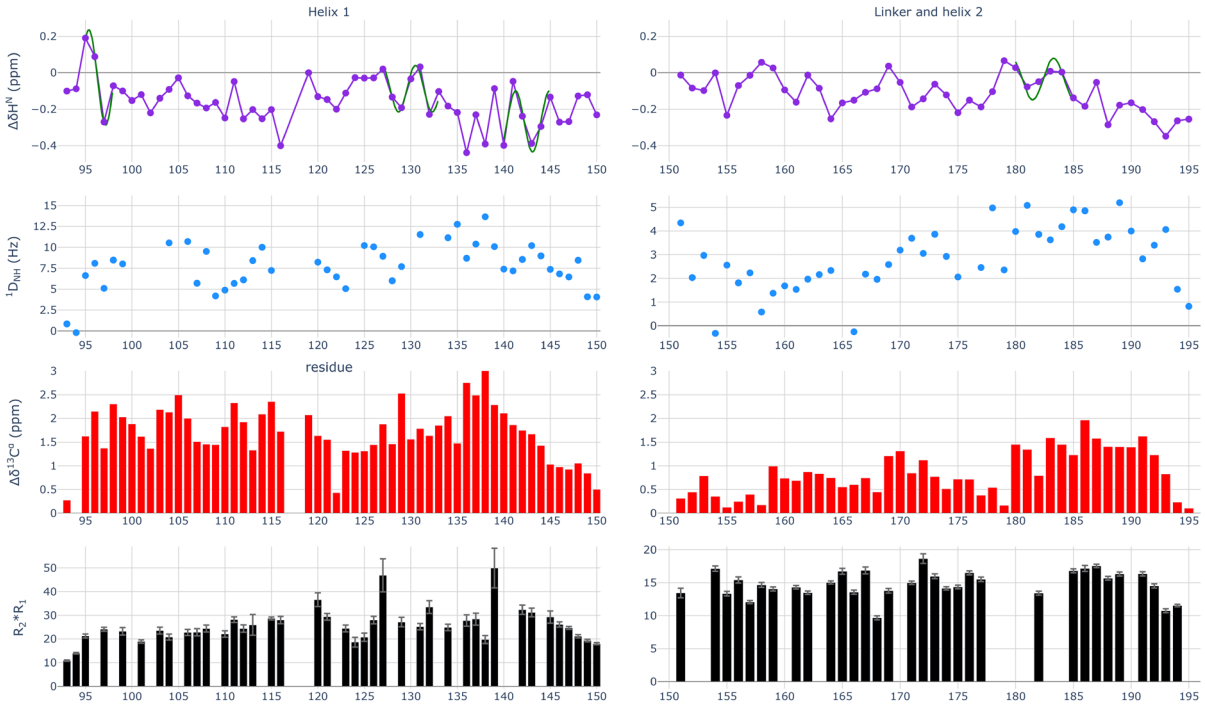


Figure 5.4: The helices of CAHS-8 present local variations.
A: Secondary chemical shifts of the amide protons Δδ $^{1}H^{N}$ (purple); for (left) the first helix on the (95-150) and (right) the linker (150-165) as well as the second helix (166-195). The green curve corresponds to a sine fit with a period of 3.63, which is the periodicity of a normal α-helix, and indicate a local bending of the helix.
B: Residual Dipolar Couplings as presented in fig. 2.
C: Secondary chemical shifts of the alpha carbon Δδ $^{13}C^{\alpha}$ (red), which highlights the difference in population of the helical conformations.
D: Secondary R_2R_1 product (black) at 850 MHz which permit to detect the R_{ex} chemical exchange contribution while limiting the contribution of a possible rotational diffusion anisotropy.
The bending of the helix seen in the Δδ $^{1}H^{N}$ values is correlated with N-H bond vectors pointing in unexpected directions in the RDCs and are highlighted in orange for A97, T129, K143 and Q183.

residues in RDCs (fig. 4) as well as in the 127-133 region with the break of continuity in the RDCs appearing at K128. Moreover, the presence of Y126 may induce a ring-current shifts, resulting in a downfield shift of neighbouring residues which includes R127 which may damp this oscillation as well as possibly affecting the preceding residues E124 and A125 which have a very similar Δδ $^{1}H^{N}$ to Y126 and thus hiding a longer oscillation. A similar damping of oscillations may appear round Y104 as the Δδ $^{1}H^{N}$ of R105 is suspiciously high. As per the oscillation in the 140-145 portion of helix 1, all the secondary chemical shifts are below 0 which is surprising given the high helical population of this region (fig. 4). However, this region presents a bell-shaped pattern in the RDCs centred

around K143 which may nevertheless indicate a local bending of the helix while another process may be responsible for the downfield shift of the chemical shifts.

Additionally, S158 is also suspiciously high and present a clear discontinuity point in the RDCs, however the small helical population of this linker region make it difficult to conclude if an additional bend of the helix occurs at this position. Moreover, the R_2R_1 product of this whole linker region (152-166) is also surprisingly elevated, presenting a similar baseline to that of the following helix 2, despite its lower helical content. This may indicate that the linker is also involved in an exchange process.

In principle, RDCs measured in disordered chains containing helical elements also contain valuable information concerning the ensemble of states sampled up to the millisecond timescale. Indeed similarly to relaxation profiles, the orientational properties of a helical element report on the average orientation of the helix with respect to the magnetic field, or direction of alignment. If this average orientation is tilted with respect to B_0, this can give rise to a *dipolar wave* along the helix, again with periodicity 3.6. However, no such dipolar wave could be detected in the RDCs of CAHS-8. This absence may result from experimental noise, absence of tilt with respect to the principal direction of alignment, the presence of numerous helices of different length in rapid exchange which interferes with observed periodicity or curvature of the helix. Moreover, such absence can also result from a high level of local internal mobility and the presence disordered chains, which for instance appear to account to 30 % of the population of the first helix, as they give rise to more random RDCs that eventually mask the phenomenon. In turn, it is also possible that CAHS-8 interacts weakly with the alignment medium.

4.1.1.2 Sequence analysis

The comparison of the sequence of the helical domain of CAHS-8 with 84 others identified CAHS protein from the same tardigrade species, *Hypsibius exemplaris*, as well as known

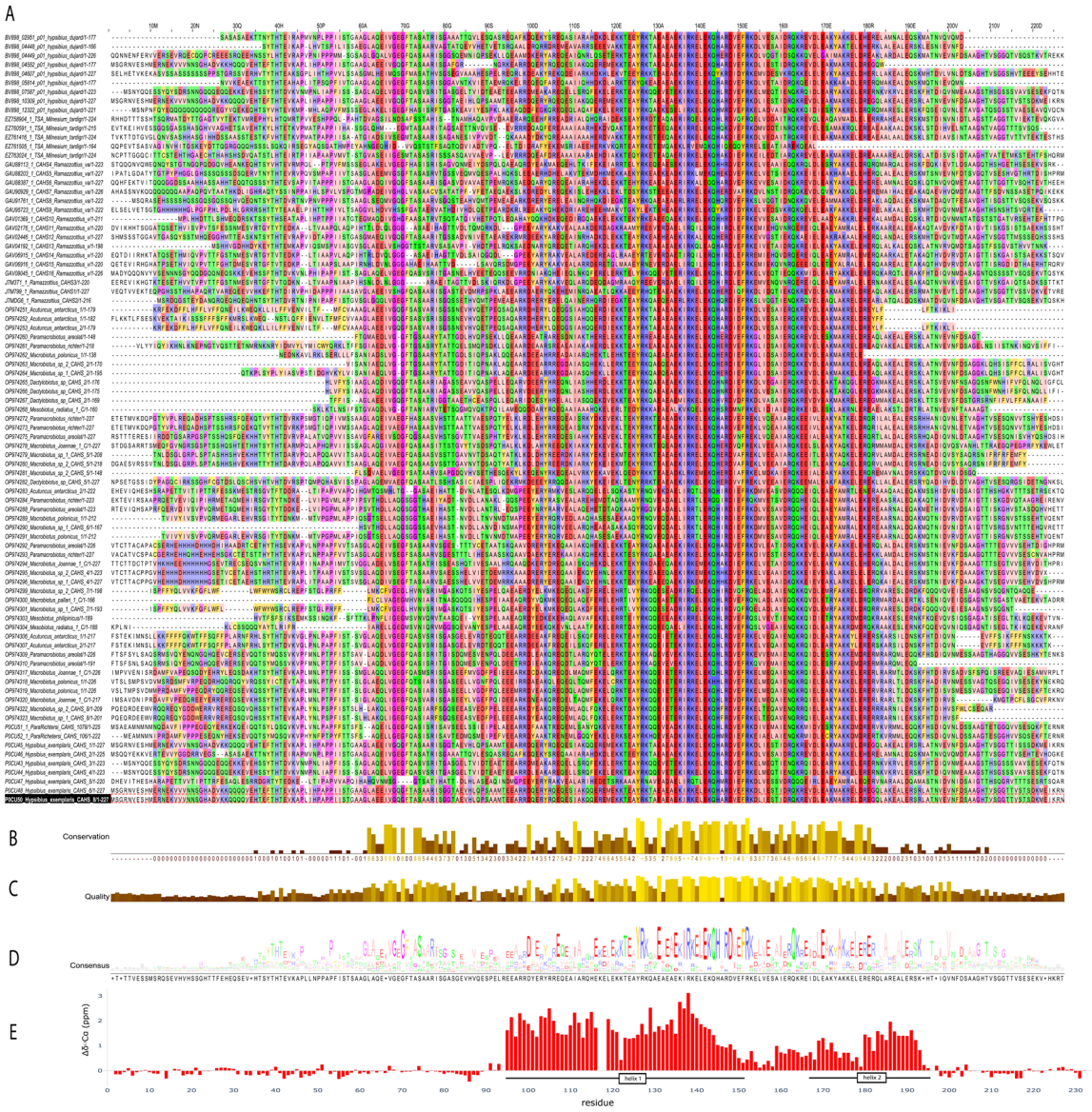


Figure 5.5: The helical domain sequence is highly conserved amongst CAHS proteins.
A: Sequence alignment of CAHS-8 with related known and putative CAHS. The red bars in the sequence name separate the different tardigrade species. Gaps in the alignments relative to the studied *H.exemplaris* CAHS-8 where hidden and indicated by a blue arrow on the top and **B**: conservation and, **C**: quality score.
D: Consensus logo of the sequence alignment.
E: Secondary chemical shifts of the alpha carbons $\Delta\delta^{13}C\alpha$ (red) aligned with the sequence, with the helices indicated.

and putative CAHS proteins identified through BLAST analysis from more or less closely related other tardigrade species shows a considerable sequence conservation in the helical domain of the CAHS-8 (fig. 5). The second half of helix 1 and first residue of the linker (Y126-R152) is almost entirely conserved amongst most species and corresponds to the region previously highlighted. In particular, the residues Y126 and F151 that flanks

this region are conserved in all 85 sequences as well as E131 and R152 are constant in all the 85 aligned sequences.

The end of the linker and first half of the helix 2 (165-177) displays a lower, but still high residue conservation but nevertheless a strong quality conservation of the residue type, in particular L169, E170 and A174 that are highly conserved. The second half of helix 2 (178-194) that repeats every two to three residues up to the V198 in the C-terminal disordered region. Some hydrophobic residues, in particular, L178, A185 and L189 are also highly conserved. Finally, the region 62-84 in the N-terminal IDR also shows a high amount of conservation with 9 amino acids having a conservation score of 8 or higher.

The careful examination of the sequence in the helical domain reveals a periodic distribution of hydrophobic, charged and hydrophilic residues that is typical of coiled-coil forming helices. Indeed, charged and hydrophobic residues are found with a frequency of 3 or 4 residues, which effectively results in the helices being amphiphilic, i.e. with the hydrophobic residues being located on the same face of the helix and thus, are prone to cluster together as a result of the hydrophobic effect. Thus, specialised prediction algorithms predict the occurrence of a coiled-coil on the helical domain of CAHS-8 (fig. 7 A) which is also supported by the high conservation of these hydrophobic residues amongst the many CAHS sequences.

As previously highlighted, the helices of CAHS-8 are bent which is typical of coiled-coil forming helices as a curvature is necessary to accommodate the “knobs-into-holes” packing of another helix, or often, in the case of “knobs-into-knobs”, to accommodate heptad discontinuities or imperfections as discussed in the introduction. As such, perfects heptads, that is with hydrophobic residue in position **a** and **d** and with charged residues of opposing signs in position **e** and **g**, can be found throughout the helix. In particular, the heptad 134-**A**EK**IR**K**E**-140 correspond to the maximum of sequence conservation but also R_2 rates, chemical exchange. Another heptad in the second helix, 171-**A**KM**AK**R**E**-177 also correspond to a highly conserved sequence amongst tardigrade species K172 has the highest R_2 rates and seemingly exchange contribution of helix 2.

The periodic distribution of the residue types is not perfect in CAHS-8, which indicate the presence of coiled coil irregularities such as stutters. Nevertheless, even accounting for this, the hydrophobic face of CAHS-8 helices would comprise polar residues while some hydrophobic ones have to be solvent exposed, in particular in helix 2. Moreover, charged residues of the same sign can also be found in probable positions **e** and **g**, leading to ionic repulsion. This imperfect heptad distribution is probably the reason why CAHS-8 is monomeric in solution instead of being in a coiled-coil oligomeric state. In turn, the presence of chemical exchange faster than 1500 s^{-1} (see above), in these regions can be plausibly attributed to the transient oligomerisation of CAHS-8 at 293 K as an oligomeric coiled-coil.

4.1.1.3 Evidence of oligomerisation

Concentration dependent ^{15}N-^{1}H HMQCs spectra were measured at 220, 250, 350, 450 and 550 µM at 293 K. The results show that the intensities I increases between 220 and 250 µM as the ratio of I_{250}/I_{220} is approximately 1.5 on average throughout the sequence (fig. 7 B). However, the intensities drop by ~20% between 250 and 350 µM and continue to decrease further at 450 and 550 µM (fig. 7 C). This shows that CAHS-8 oligomerises in a concentration-dependent way in qualitative agreement with the phase diagram presented in the previous chapter (fig. 7 D), where the higher the concentration, the more fibrillation and gelation. As the 350 µM and 450 µM concentrations are still in the soluble region, it is likely that any oligomerisation is still transient as no fibres could be detected by DLS at these concentrations, while 550 µM is already in the slight gelation region, at which a significant fibre scattering could be detected.

Tentative characterisation of this exchange using linewidth analysis could not be performed due to the high overlap of the peaks while efforts to characterise this exchange process in the helical domain of CAHS-8 by off-resonance $R_{1\rho}$ relaxation dispersion unfortunately did not succeed either due to the low signal to noise ratio in this region.

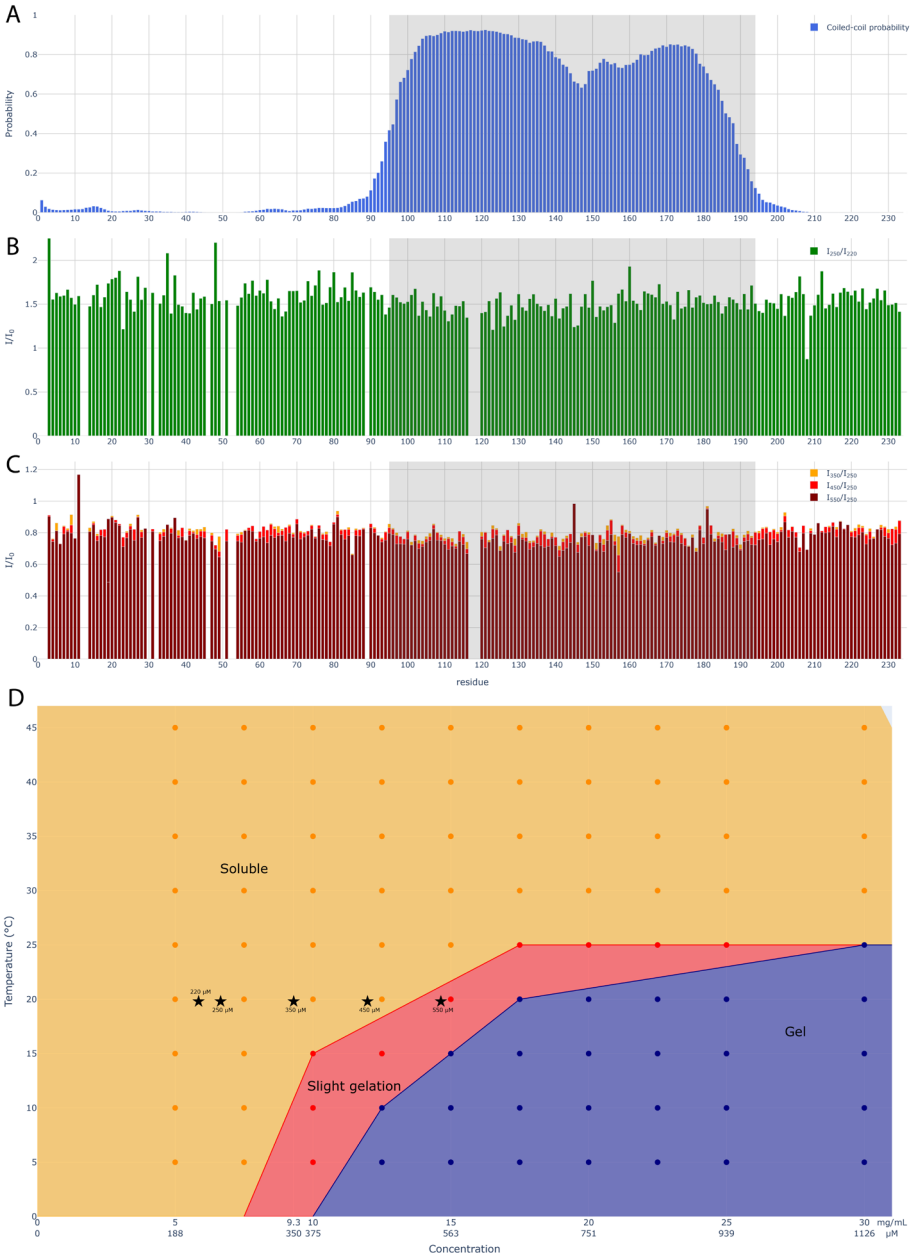


Figure 5.6: Prediction and evidence of concentration-dependent coiled-coil oligomerisation of CAHS-8.
A: Predictions of coiled-coil formation using DeepCoil[2] show a high probability of approximately 70 % over the whole helical domain of CAHS-8 (95-194), which is highlighted in grey.
B: The ratio of the intensities in ^{15}N-^{1}H HMQCs spectra at 250 µM over that at 220 µM shows an increase of the intensity associated with a higher concentration.
C: However, the ratios of the intensities at 350 µM (**bottom**, gold), 450 µM (red) and 550 µM (brown) over that at 250 µM show a decrease of the intensities as the concentration increases, with the helical domain being more affected, supporting a concentration-dependent oligomerisation via this domain.
D: DLS-derived phase diagram of CAHS-8 in function of the concentration and the temperature. The 220, 250, 350, 450 and 550 µM concentration are indicated by black stars.

4.1.1.4 A high temperature dependence

This phase diagram also shows that CAHS-8 has a temperature-dependent behaviour, with fibrillation and gelation at low temperatures, and thus, the protein was also assigned at 278 and 303 K. The comparison of the HSQCs spectra at 303, 293 and 278 K (fig. 7 A) shows a significant decrease of the signal to noise ratio of many peaks at low temperature. Indeed, many resonances in the first helix are broadened beyond detection at 278 K. The comparison of secondary chemical shifts for the alpha carbons $\Delta\delta\ ^{13}C\alpha$ (fig. 7 B) also show a temperature-dependent behaviour in the helical region: accounting for calibrated temperature dependent random coil shifts, the $\Delta\delta\ ^{13}C\alpha$ values still increase as the temperature decreases with significantly larger shifts in the second helix at 278 K than at 293 K which in turn are consistently larger than at 303 K. Due to the broadening beyond detection, secondary chemical shifts values are not available at 278 K for most of the helix but the trend between 303 and 293 K and the behaviour of the second helix, it is justified to infer a similar effect on the $\Delta\delta\ ^{13}C\alpha$, indicating that the helical domain of CAHS-8 is increasingly populated as the temperature decreases.

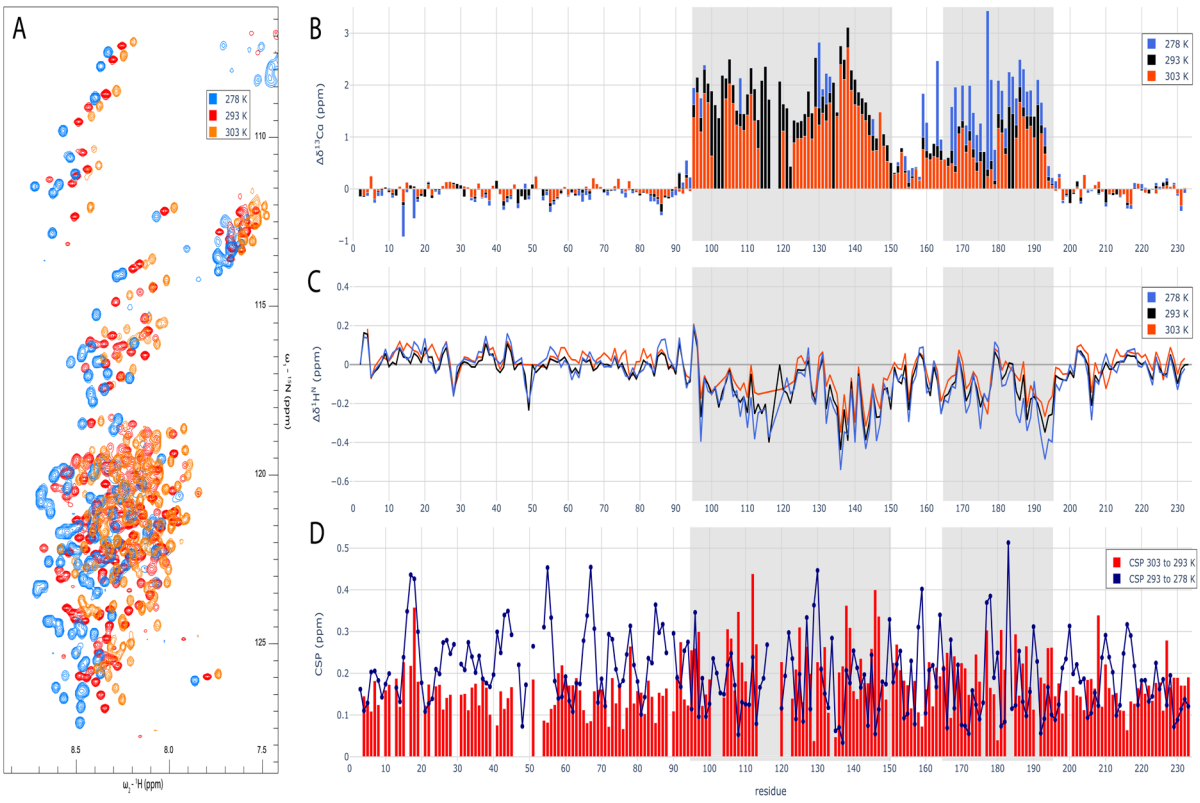


Figure 5.7: Temperature-dependent behaviour of CAHS-8.
A: ^{15}N-^{1}H HSQCs spectra of a ^{2}D ^{13}C ^{15}N CAHS-8 at 850 MHz at 303 K (orange), 293 K (red) and 278 K (light blue).
Associated secondary chemical shifts of **B:** the alpha carbon $\Delta\delta$ $^{13}C^{\alpha}$ and **C**: amide protons $\Delta\delta$ $^{1}H^{N}$ at 303 K (orange), 293 K (black) and 278 K (light blue). **D:** Chemical Shifts Perturbations (CSP) from 303 to 293 K (red) and 293 K to 278 K (dark blue).
The random coils values were calculated using POTENCI for each temperature while the experimental values were corrected for deuterium isotopic shift. The two helices are shaded in grey.

This is supported by the amide protons secondary chemical shifts $\Delta\delta$ $^{1}H^{N}$ profile obtained from ^{15}N-^{1}H HSQCs which permit to observe helix 1 (fig. 7 C) in which the values are smaller at 278 K than at higher temperature, which suggest an increased helical propensity at lower temperature. In turn, no particular temperature effect on the chemical shifts is apparent in the flexible termini.

The ^{15}N-^{1}H chemical shift perturbations (CSP) between 303 and 293 K (fig. 7 D, in red) are overall consistently about 0.1 to 0.2 ppm in the termini but show an oscillating pattern in the helical domain with amplitude ranging from less than 0.05 to more than 0.4 ppm. However, the CSP between 293 and 278 K (fig. 7 D, in dark blue) show a complex, seemingly oscillating pattern in both the disordered termini and helices.

Spin relaxation rates measured at 303 K (fig. 8) show clearly in both R_2 and NOE rates the difference between the central helical domain of CAHS-8 and the flanking disordered termini. The increase of R_2 between 600 and 850 MHz magnetic field strength is modest in comparison to the relaxation profile at 293 K (fig. 3, fig. 10), increasing from an average of approximately 7.5 to 10 s^{-1} in the first helix and from 5 to 6.5 s^{-1} in the second helix which reflects a smaller exchange contribution. This is also apparent from the modest deviation of the exchanging residues from the baselines in the R_2R_1 product (fig. 10, bottom). However, relaxation measurement at 278 K (fig. 9) show significant differences in the helical domain in comparison with 293 K (fig. 10). In particular, the R_2 rates are much faster, going from approximately 10 s^{-1} on average at 293K in the second helix to more than 30 s^{-1} or the first helix going from approximately 17 s^{-1} at 600 MHz to most of



Figure 5.8: Relaxation of CAHS-8 at 303K.
From top to bottom: R_1, R_2,, nOe relaxation as well as the R_2R_1 product, at 600 MHz (left) and 850 MHz (right)

the signal broadened beyond detection. Moreover, the nOe shows a significant rigidity of the whole helical domain in the few accessible peaks, with positives values even at 600 MHz.

Although the secondary chemical shifts show an increase of the helical content as the temperature decreases (fig. 7 D) which would translate into an increase of both the

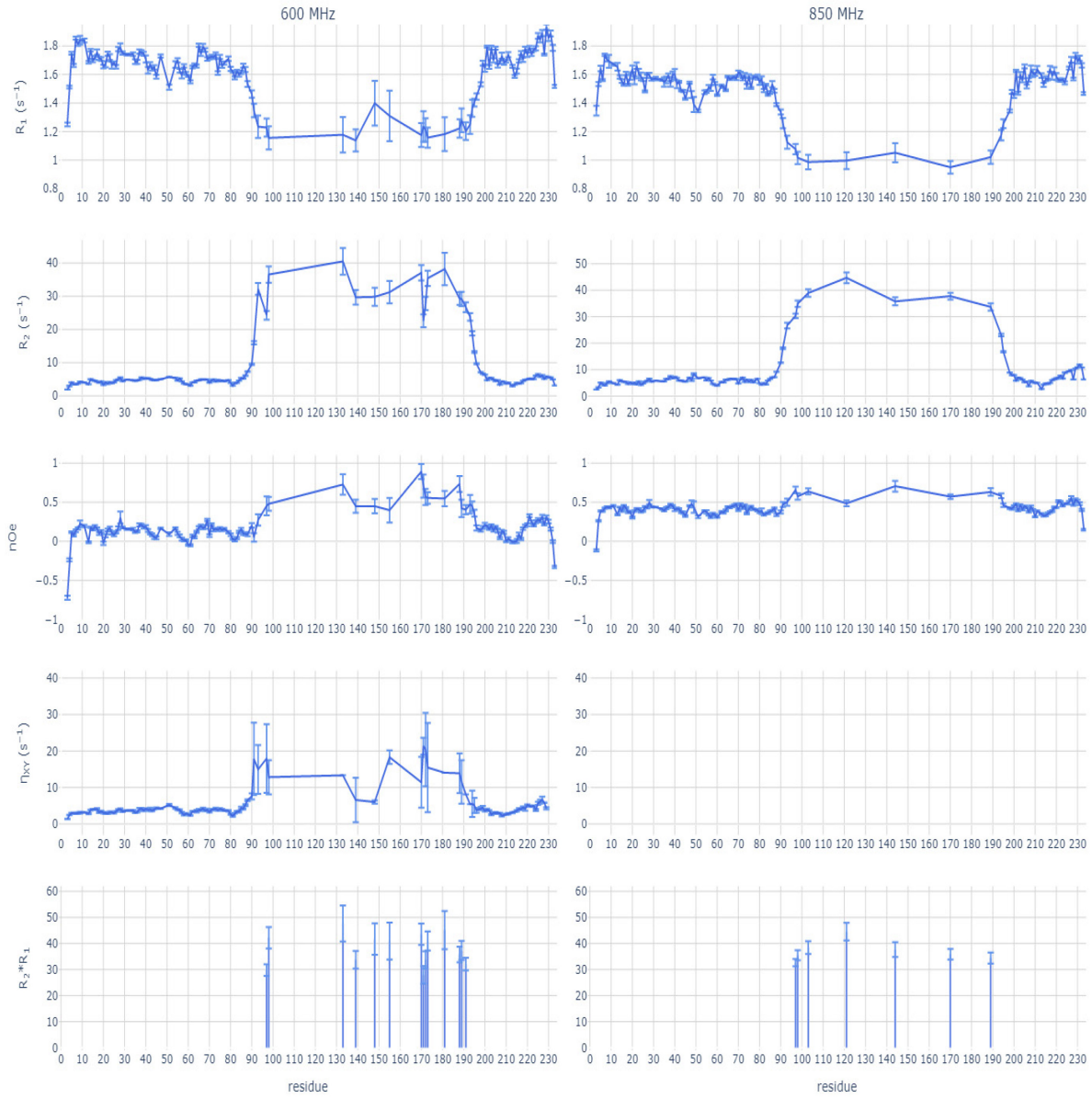


Figure 5.9: Relaxation of CAHS-8 at 278 K.
From top to bottom: R_1, R_2, nOe relaxation as well as the R_2R_1 product, at 600 MHz (left) and 850 MHz (right).

rigidity, as seen from the increase in nOe, and the transverse relaxation rates, it is unlikely that this would result in such a dramatic increase toward R_2 rates. Indeed, even folded proteins the size of this helix can usually be studied at this temperature.

Altogether, this indicates that the transient putative coiled-coil oligomerisation of CAHS-8 via its helical domain is highly temperature dependent with unequal fractional

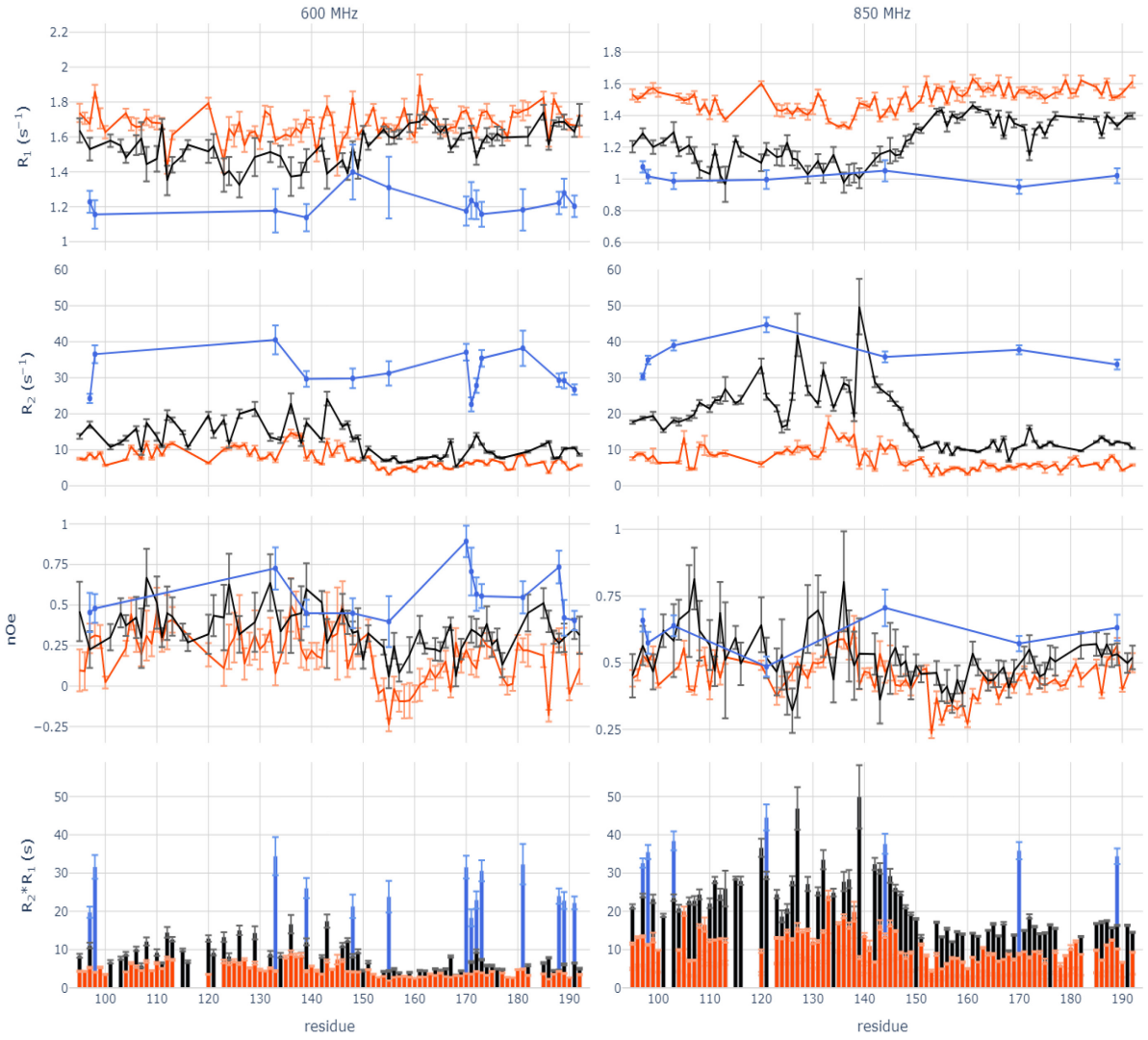


Figure 5.10: Temperature variation of the relaxation of the helical domain of CAHS-8. Relaxation rates at 600 MHz (left) or 850 MHz (right) at 303 K (orange), 293 K (black) and 278 K (blue). From top to bottom: R_1, R_2, nOe and the R_2R_1 product

populations in favour of the monomeric species at 293 K while the fractional population of oligomers increases as the temperature decreases. Given the overall phase-diagram behaviour of CAHS-8, with the fibrillation of the protein happening at low temperature, this strongly suggests that the putative coiled-coil formation is necessary for the formation of fibrils and, like several of the fibrous proteins mentioned in the introduction, can be inferred to be the base unit of the fibrils. Additionally, the absence of changes in peak intensity over the course of several days, if not weeks of measurements at 293 K indicate that the sample is at a thermodynamic equilibrium or at least, in a steady-state over the course of a couple of weeks.

4.1.2 Study of the helical domain

In order to study the behaviour of the helical region of CAHS-8 on its own, a construct of it hereafter called HD (Helical Domain, 88-198, with mutations P88Q and G90A for N-terminal TEV cleavage convenience) was produced.

4.1.2.1 Similarities and differences in spin relaxation

The ^{15}N-^{1}H HSQC spectrum of HD presents a great similarity with the corresponding peaks in the full-length CAHS-8 – which will be simply referred to as CAHS-8 hereafter – (fig. 11 A). Nevertheless, HD was assigned at 293 K in order to ensure that its secondary structure was not perturbed. Comparison of the secondary chemical shifts $\Delta\delta\ ^{13}C^{\alpha}$ and $\Delta\delta\ ^{13}C'$ of HD with the corresponding region of CAHS-8 shows an almost unperturbed secondary structure propensity in comparison with the full-length protein for the alpha carbons and in particular for the carbonyl carbons (fig. 11 B and C), with slightly higher $\Delta\delta\ ^{13}C^{\alpha}$ values indicating that HD has a slightly more populated helices than the full-length protein.

The amide protons $\Delta\delta\ ^{1}H^{N}$ profile at 293 K shows similar oscillations as CAHS-8 (fig. 11 D), indicating an overall similar curvature of the helices. RDCs of HD were also acquired (fig. 11 E) and like the full-length protein, the two helices of HD have the same $^{1}D_{NH}$ sign, indicating that they have the same relative orientation with respect to the magnetic field and therefore that HD doesn't have any particular tertiary structure element either. The RDC profile of HD shows slight differences with CAHS-8 in the first helix, indicating a different alignment tensor.

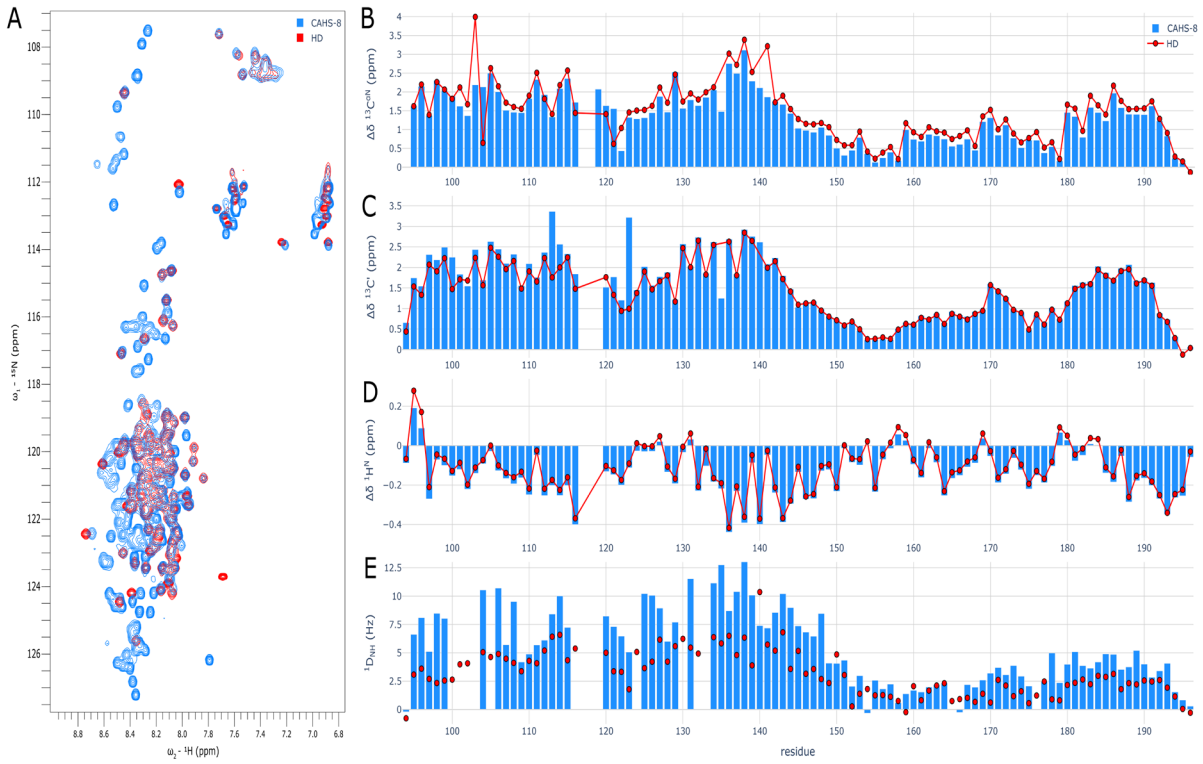


Figure 5.11: Comparison of the HD construct with the helical domain of the full length CAHS-8. **A**: ^{15}N-^{1}H HSQC spectra, secondary chemical shifts of the **B**: alpha carbons Δδ $^{13}C^{\alpha}$, **C**: carbonyl carbons Δδ $^{13}C'$ and **D**: amide protons $\Delta\delta H^{N}$, as well as **E**: residual dipolar couplings $^{1}D_{NH}$ of the backbone amide N-H bond vectors. CAHS-8 is in blue and HD in red.

Additionally, although the region around S158 appears to have a higher helical population in HD than in CAHS-8 as per the slightly higher Δδ $^{13}C^{\alpha}$, it has higher amide proton secondary chemical shifts instead of a lower one associates with a higher helical propensity (fig. 11 B and D). Like in CAHS-8, a discontinuity is present in the RDCs but at T159 instead of S158. This supports the possibility of a bend of the helix at this position evoked for CAHS-8, with a delocalisation of the curvature a residue further in HD.

Relaxation of HD at 600 MHz at 293 K shows an overall similar R_1 and nOe baseline to the WT CAHS-8, indicating similar dynamic properties of their helices (fig. 12). However, R_2 rates are significantly smaller for HD in helix 1, with an average of ~17 s^{-1} for the former and ~10 s^{-1} for the latter, and show an oscillating profile for helix 2 instead of a reverse bell-shaped profile like in CAHS-8. This similarity in R_1 but different R_2 result in a different R_2R_1 profile, with not only a smaller product but also different outlier peaks, which

strongly suggests that the exchange contribution to R_2 is smaller in HD and localised slightly differently.

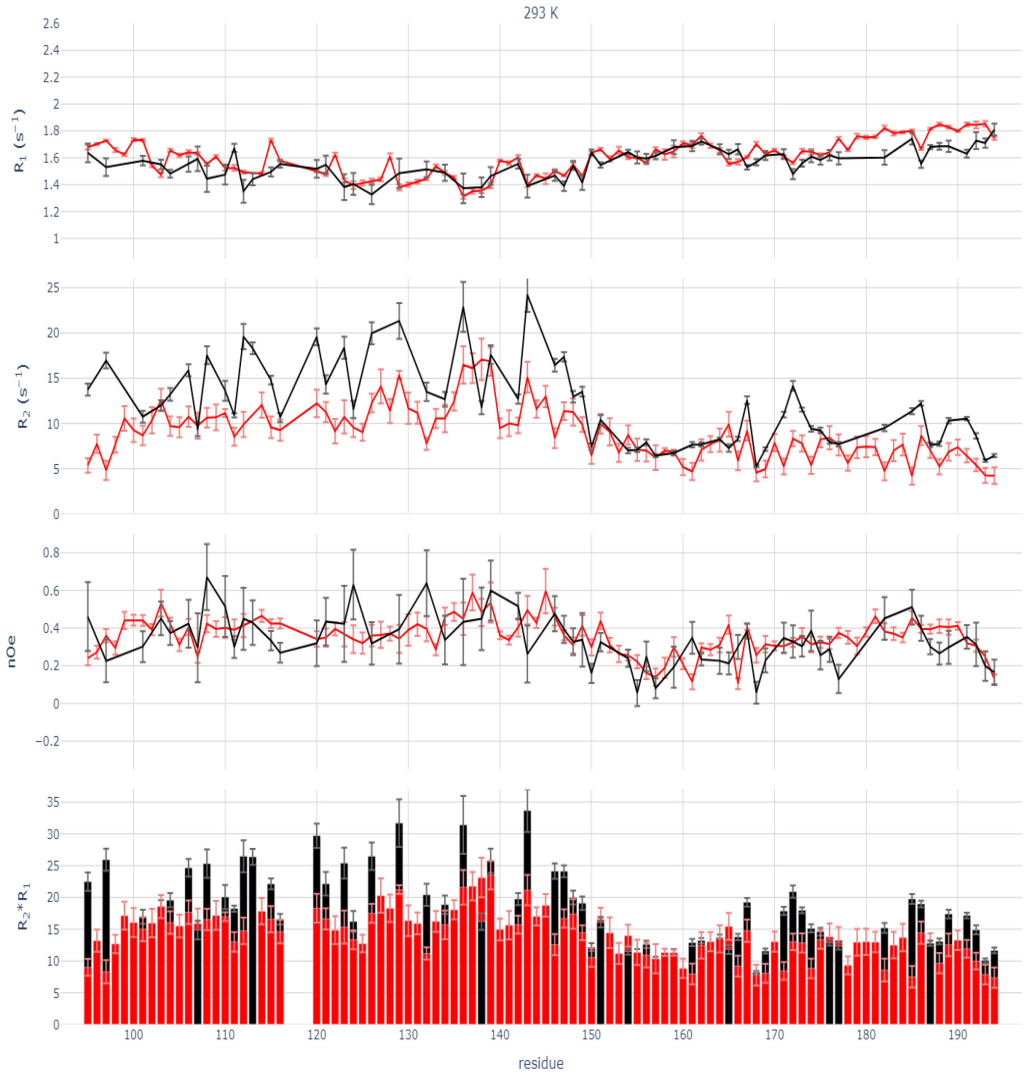


Figure 5.12: Comparison of the relaxation of HD with the corresponding helical domain of CAHS-8 at 293 K and 600 MHz.
CAHS-8 is in black and HD in red.
From top to bottom: R_1, R_2 and nOe relaxation as well as the R_2R_1 product.

Finally, comparisons of the relaxation rates of HD at 293 K measured at 600 MHz and 950 MHz (fig. 13) does show signs of chemical exchange as observed by the increase in the R_2R_1 product baseline with the field which is similar to what was previously observed in CAHS-8 between 600 and 850 MHz (fig. 3), supporting a similar occurrence of the exchange process attributed to a transient coiled-coil oligomerisation.

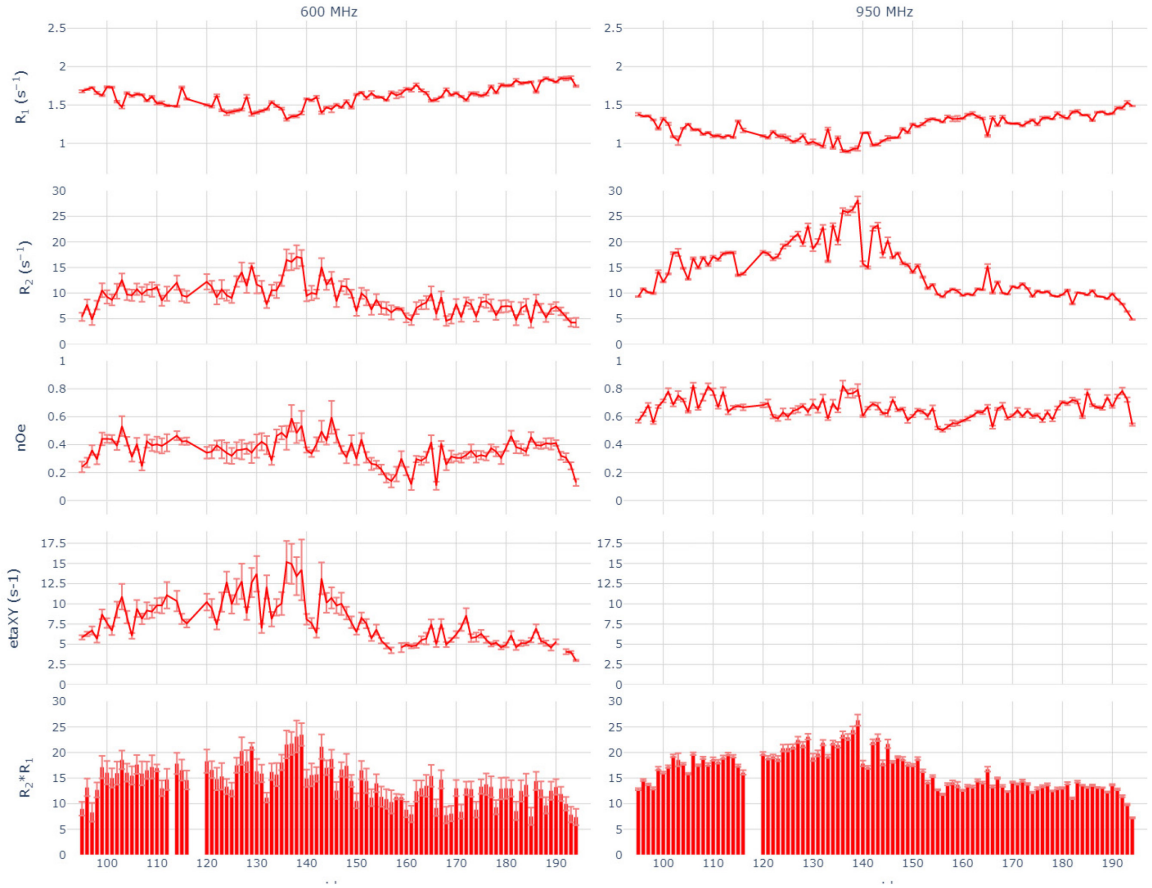


Figure 5.13: Comparison of the relaxation of HD with the corresponding helical domain of CAHS-8 at 293 K and 600 MHz.
CAHS-8 is in black and HD in red. From top to bottom: R_1, R_2 and nOe relaxation as well as the R_2R_1 product.

4.1.2.2 Crystals of the coiled-coil dimer

Numerous experiments using NMR, SAXS, AFM, SEC-MALLS, DLS and mass photometry were performed in order to observe the hypothesised coiled-coil oligomer. However, it could never be observed directly. Altogether, these experiments show that depending on the conditions, CAHS-8 is either overwhelmingly monomeric, with its an oligomeric form being too transient and too lowly populated to be detected, or, at higher concentration, forms fibrils and eventually a gel, masking the oligomers which appears to be a transient intermediate species.

Nevertheless, despite the helices being only transiently populated in solution, an intermolecular coiled-coil of CAHS-8 requires them to be formed enough for the oligomerisation to happen, in a probable conformation selection mechanism and therefore, crystallisation was attempted in order to isolate the oligomeric form as only the conformations with a highly populated helix are likely to crystallise.

After months of optimisation, conditions could be found where HD would form crystals which were only more or less big microcrystal, often needle shaped. Increasing the protein concentration only resulted in smaller microcrystals and urchins-like clusters, while decreasing it only yields marginally bigger but fewer microcrystals. Depending on the conditions, the microcrystals appeared physically quite irregular, with often display of obvious physical twinning and even apparent multiple birefringence under polarized light

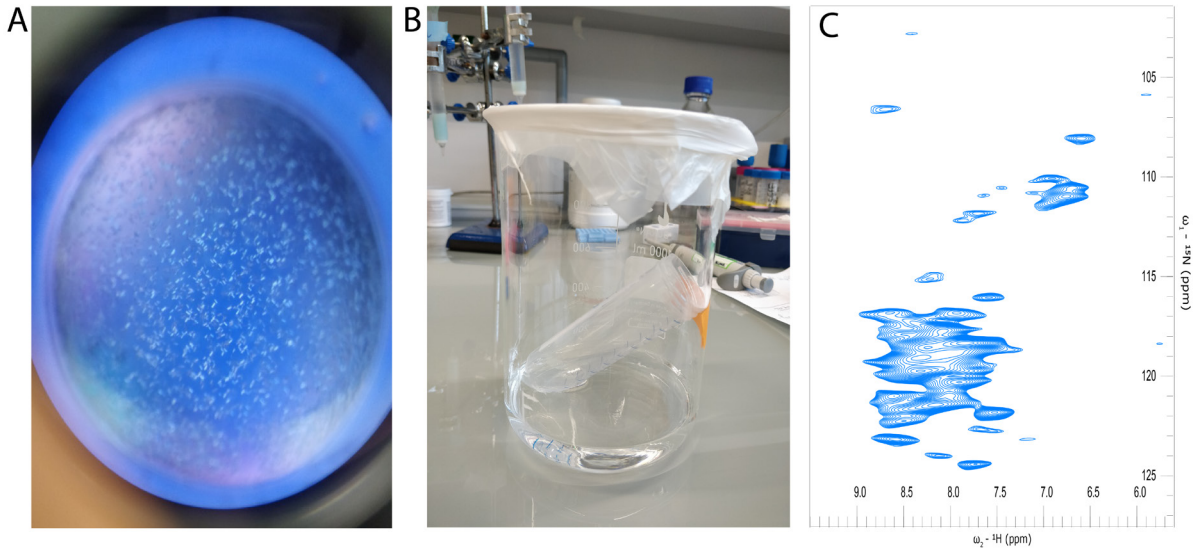


Figure 5.14: Tentative ssNMR study of HD microcrystals.
A: Microcrystals of HD that were scaled up in **B**: a homemade crystallisation device to produce enough ^{13}C ^{15}N microcrystals for **C**: proton-detected ssNMR hNH spectrum at 850 MHz and 55 kHz MAS. The spectrum reveals a high level of inhomogeneity despite the crystalline form.

One of the conditions that led to microcrystals that were regular to the eye (fig. 14 A) was successfully scaled-up in a homemade giant crystallization well made from a sealed beaker in order to produce several milligrams of ^{13}C ^{15}N labelled microcrystals to pack a 1.3 mm rotor for proton detected Magic Angle Spinning (MAS) solid-state NMR (ssNMR) study (fig. 14 B), permitting to measure hNH and hCaNH spectra at 50 kHz MAS, 293K and 800 MHz. Unfortunately, these experiments ^{1}H-^{15}N hNH and ^{1}H-^{15}N-^{13}Ca hCANH

experiment revealed significant degree of inhomogeneous line broadening despite the crystalline form (fig. 14 C) and as a consequence, no analysis of these data could be performed. Deuterated crystals couldn't be made as the use of deuterated protein significantly hampered the crystal growth, which was hypothesised to results from a lower ability of the deuterated protein to oligomerise.

Crystals were regularly diffracted for 1 year and a half at the beamline ID23-2 of the European Synchrotron Radiation Facility (ESRF). Most crystals were too anisotropic to obtain a high-quality dataset, even after several rounds of optimization including the use of additives to improve the best conditions. Eventually, the merged diffractions of 51 microcrystals originating from the same drop, mounted together on a mesh and diffracted individually resulted in a dataset with high isomorphism going up to 2.6 Å. Initially, molecular replacement was attempted. As the helix 1 has significantly higher helical propensity than the rest of the helical domain, it was hypothesised to be the only one crystallising, or, at least, the domain which drives the oligomerisation, and therefore thousands of conformers of parallel coiled-coil dimers of helix 1, were generated by flexible-meccano and used as seeds for molecular replacement solutions. However, despite the promising molecular replacement scores of some conformers, none were ultimately correct. Molecular replacement using AlphaFold models, deposited structures of others coiled-coils or cross-alpha amyloids was also unsuccessful.

As no other structure of any other protein from the CAHS family was ever resolved, experimental phasing was pursed using seleno-methionine labelling, in order to take advantage of the anomalous scattering of selenium and the presence of 3 methionines per monomers of HD. However, the inclusion of selenium atoms instead of sulphur in the methionines changes their properties, which prevented the crystallisation from happening in the same conditions, likely indicating that the methionines locate at the interface between the helices. Optimisation, including through the use of additives and screening in other conditions resulted in very few microcrystals that poorly diffracted.

Eventually, while attempting molecular replacement with other coiled-coil sequences, it was noticed that the log-likelihood gain score, which quantifies how well the data can be

predicted from the model compared to a random distribution of the same atoms, was better with search models that showed little supercoiling of the helices. From there, manual trial-and-error modifications of the best-scoring conformers previously generated coupled with optimisation using ARCIMBOLDO[600,601] finally resulted in a satisfactory molecular replacement solution.

4.1.2.2.1 Structure of the dimer

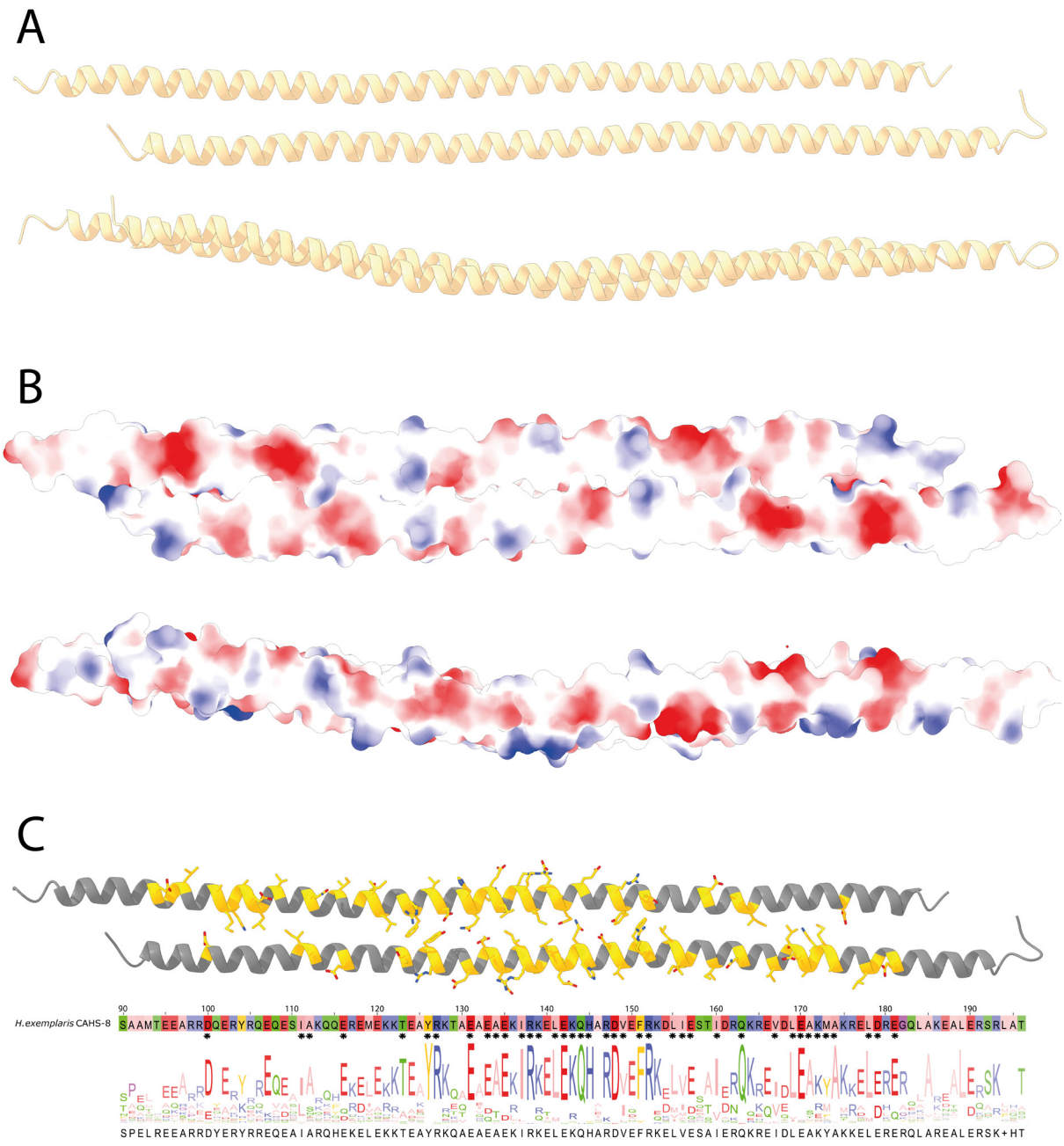


Figure 5.15: Crystallographic structure of the antiparallel coiled-coil dimer of the HD construct of CAHS-8.
A and **B:** Front and lateral view of the coiled-coil with the two antiparallel chains in cartoon representation (**A**) and in electrostatic surface representation (**B**). The positive patches are highlighted in blue and the negative ones in red, with the neutral ones appear in white.
C: Highly conserved residues (conservation score >7) highlighted in yellow and marked with a star below the CAHS-8 sequence, followed by the logo from the sequence alignment (fig. 5). The region between the aromatic stacking of Y126 and F151 stands out by its conservation, as well as the 167-181 region. The sequences are aligned with the bottom chain of the above it.

The structure of the final model reveals an antiparallel two-stranded coiled-coil comprising the full helical domain of HD (fig. 15), therefore including not only helix 1 but also the linker and helix 2. The statistics of the dataset and the model are given in the following table:

Table 1. Data collection and refinement statistics.

	CAHS-8
Wavelength	0.873
Resolution range	46.79 - 2.6 (2.693 - 2.6)
Space group	C 1 2 1
Unit cell	61.83 55.45 96.38 90 103.85 90
Total reflections	54745 (5223)
Unique reflections	17232 (914)
Multiplicity	3.2 (3.2)
Completeness (%)	96.47 (95.40)
Mean I/sigma(I)	2.82 (0.88)
Wilson B-factor	34.34
R-merge	0.5392 (2.801)
R-meas	0.6254 (3.275)
R-pim	0.3054 (1.641)

CC1/2	0.922 (0.185)
CC*	0.979 (0.558)
Reflections used in refinement	9589 (913)
Reflections used for R-free	440 (46)
R-work	0.2610 (0.3396)
R-free	0.2945 (0.3588)
CC(work)	0.891 (0.480)
CC(free)	0.830 (0.493)
Number of non-hydrogen atoms	1883
macromolecules	1785
ligands	0
solvent	98
Protein residues	211
RMS(bonds)	0.010
RMS(angles)	1.33
Ramachandran favored (%)	99.52
Ramachandran allowed (%)	0.48
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	10.33
Clashscore	4.46
Average B-factor	52.33
macromolecules	53.08
solvent	38.71

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

The geometrical analysis of the structure using manual inspection and the coiled-coil analysis library samCC turbo[551,553] show that this antiparallel coiled-coil presents non-canonical characteristics. Plot of the Crick angles against the sequence permit to visualise the different phase relationship between consecutive residues and therefore, allows the assignment of the repeats from the structure and the clear clustering of the positions (fig. 16 B) as explained in the introduction of this part. This highlights the sequence repeats starts with a stutter (4 residues, noted **hijk**) in position 93-96 and then alternate heptads (7 residues, noted **abcdefg**) and stutters three times namely with the stutters in position 104-107, 115-118 and 126-129. After these, a three heptads consecutive repeats at 130-136, 137-143 and 144-150 are followed by another stutter in 151-154, two alternate heptads-stutters repeats with the stutters in 162-165 and 173-176 and a final heptad in position 177-183.

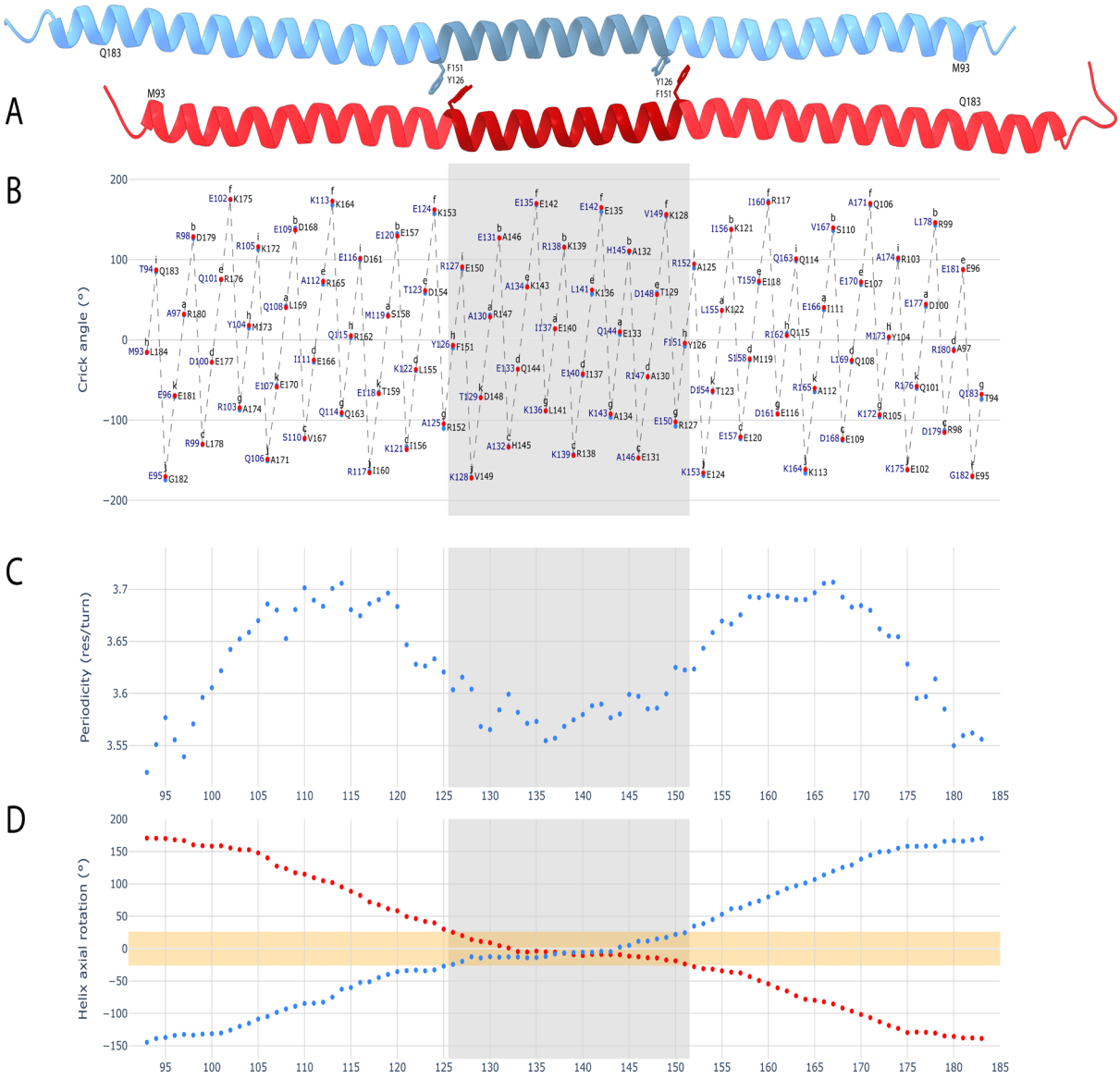


Figure 5.16: Coiled-coil metrics of the structure.
A: The antiparallel coiled-coil of the HD construct of CAHS-8, with the chain A in red and chain B in blue. This colour code is used in the whole figure. The limits of the coiled-coil M93 and Q183 are denoted. In turn, Y126 and F151 are highlighted and their sidechain displayed while the region between them is highlighted in the entire figure in grey.
B: Crick angles of the two chains, with the chain A being a red rot and its residue being written in its left while the one in chain B is blue and the residue is on its right. For most of the residues, these values are almost identical and only the red can be distinguished. The position in the coiled-coil repeat is written on top of each point.
C: Periodicity of the coiled-coil per residue. It is identical for both chains.
D: Axial rotations of the helices with the usual expected canonical values (-26° to 26°) highlighted in orange. The x value corresponds to the residues in the chain A only while the chain B is inverted to follow it, as it is an antiparallel coiled-coil. Consequently, the values of the two chains mirror each other.

As discussed in the introduction of this chapter, the insertion of a stutter, or presence of hendecads, relaxes locally the supercoil as four residues form more than a turn of an α-helix but their position is still constrained by the necessity of forming a hydrophobic core, in which case the residues in position **a**, **d**, **e** and **h** that forms it, with **h** projected directly towards core and flanked by the previous **e** residue and the following **a** residue of the same α-helix. This can be seen in the distribution of the Crick angles (fig. 16 B) with the residues in position **h** of the stutters having the phase the closest to 0 of the whole structure, indicating that it faces the interface. This particularity results into a knobs-to-knobs type of packing for the stutter, where the sidechain of the core residue does not pack into a hole in-between the sidechains of the two facing residues as in the classic knobs-into-hole packing, but instead packs directly against the sidechain of the facing residue from the other α-helix. As a consequence, this particular relaxation of the supercoil induced by stutters or hendecads lead to essentially straight helices instead of the left-handed supercoiling of the canonical heptads.

Thus, this effect on the periodicity is seen when plotting the local periodicity of the helix (fig. 16 C), with the 105-120 and 154-176 segments showing an approximate 3.67 residues per turn periodicity, corresponding to three turns of an α-helix over a heptad and a stutter (forming together an hendecad repeat), that is $(7 + 4)/3 = 11/3 \approx 3.67$, while the central 130-150 segment that corresponds to the three consecutive heptads has a periodicity of approximately 3.55 residues per turn. A canonical heptad corresponds to two turns of the α-helix and therefore has a $7/2 = 3.5$ residues per turn periodicity, which

is not the case here. It appears instead that the periodicity in this region is better described by a stutter and three heptads over 7 turns of the helix, that is $(4 + 7 + 7 + 7)/7 = 25/7 \approx 3.57$. In turn, initial 97-104 heptad and the 151-154 stutter display an in-between periodicity, corresponding respectively to the unwinding of the coiled-coil from respectively a classic α-helix periodicity relaxed by the initial stutter for the latter and the 25/7 periodicity of the central region for the former, to the looser 11/3 hendecad periodicity. Conversely, the 121-129 residues also display an in-between periodicity that correspond to a winding of helices toward the tighter one of the 25/7 periodicity of the 130-150 segment.

As per the lack of supercoiling of the helices, it is visible in the structure from the two coiled-coil extremities in the segments 94 to 125 and residue 152-183, which are essentially parallel to each other. However, the central region delimited by the π-π stacking of Y126 and F151 on both sides of the antiparallel coiled-coil does show a supercoiling of the helices due to a closer following of the canonical heptad periodicity as seen in the plot of the axial rotation. As such, this can also be observed in the plot of the axial rotation of the helices (fig. 16 D), which corresponds to the difference between the experimental Crick angles with the ones expected for an ideal canonical coiled-coil, in which case the experimental axial rotation takes values between -26° and 26° (orange box) with the closer to 0°, the closer to an ideal packing[544]. In this CAHS-8 coiled-coil, this is fulfilled in the central region 126-F151 which is therefore highlighted in all plot by the grey box. In term of repeats, this region therefore corresponds to the 126-129 stutter and the three consecutive heptads from residue 130 to 150.

Putting all together, this coiled-coil structure is therefore more accurately described as an initial stutter at 93-96 followed by two consecutive hendecads (one heptad followed by one stutter, i.e. 11 residues noted **abcdefghijk**), in position 97-107 and 108-118, a heptad in position 119-125 followed by a stutter in 126-129, three heptads at 130-136, 137-143 and 144-150, followed by another stutter at 151-154, two hendecads in position 155-165 and 166-176 and a final heptad at 177-183.

Nevertheless, despite these particular periodicities occurring to accommodate for the distribution of the amino-acids in the sequence, polar residues are also present at the interface of the two monomers (hereafter called A and B as per the PDB file numbering) and participate to the stabilisation of the dimer through the making of salt-bridges. Examples includes the salt-bridge network at the extremities of the dimer composed of E96B-R180A-D100B-R176A, which also exist on the other side of the dimer where it includes coordinated waters as well. Other examples include E107A-K172B, K136A-E140B, R147A-E133B and D154A-K122B. Additionally, multiple hydrogen bond also appear to stabilise the interaction between the two monomers on both sides of the dimer.

On the other hand, although most hydrophobic residues are located at the interface of the coiled-coil as they need to be, the outer side of the second half of the helical domain of both monomers is remarkably hydrophobic (fig. 16 B). This corresponds to part of the linker and to the helix 2 and is discussed in details in the next chapter.

4.1.2.2.2 Deformations and comparison with the solution conformations

The axial shift of the helices corresponds to the offset between a residue at a given position of the repeats and the corresponding residue in the facing helix. As such, it should remain more or less constant if the coiled-coil is unperturbed, but shows local variations in case of perturbation such as the bending of the helix, whether this curvature is vertical or lateral. Thus, the plot of the axial shift permits to localise precisely the residue where the α-helix bends, being the residue showing a clear discontinuity with the neighbouring residues and after which the trend changes.

Analysis of the structure and repeats permit to explain these surprisingly numerous local curvatures as induced by the stutters, seemingly often to avoid the burial of polar groups within the hydrophobic core.

Indeed, the bend at A97 allows the accommodation of the salt-bridge and hydrogen-bond network of E96-R180-E177/D100-R176 previously mentioned, which in turn allow the

solvation of D100 and E177. Indeed, D100 is in **d** position of the heptad and the salt-bridge with its polar carboxyl group solvated by pointing outside of the core, allowing its oxygens δ to be at respectively between 2.6 and 3 Å of the guanidino groups of R180 and R176 of the other helix and R103 of the same protein. Moreover, the bending allows its facing residue, E177 in position **a** of the heptad, to point entirely outside which can be seen by its Crick angles of 41°, far from the average 17° expected for this position, thereby locating its oxygen 2.5 Å away from the R180 guanidino group of the same protein, for the B chain only, to form a hydrogen bond with a water molecule 3.5 Å away. This network of hydrogen-bonds is probably very important to the function of CAHS proteins, likely to hold the coiled-coil together at its extremities, as all the residues involved show either a high level of conservation or mutations that also allow hydrogen-bonding, but not necessarily salt-bridges. Indeed, E96 is a glutamic acid, a glutamine or a lysine in respectively 45 %, 19 % and 14 % of the sequences; D100 is an aspartic acid in 80 % of them and a glutamic acid in 18 %; R176 is an arginine in CAHS-8 and overall, 42 % of the sequences but is a lysine in 48 %; E177 is a glutamic or an aspartic acid in 67 % and 8%; and finally, R180 is an arginine in 66 % of the sequence and a histidine in 15%.

On the other protein or at the C-terminal extremity of the same helix, these interactions are rendered possible by two deviations visible in the axial shift plot at R176 and L178, although this region appears quite straight.

As per Y104, it is in position **h** of the stutter 104-107 but has its sidechain exposed to the solvent and the deformation of the α-helix at Y104 both laterally away from the other helix and vertically appears necessary to accommodate not only the presence of the aforementioned large salt-bridge network in the core, but also to allow the two facing bulky apolar amino-acids, M173 and L169 respectively in position **h** and **d** of the hendecad, to be fully buried away from the solvent at the interface of the two helices.

The location of the aromatic ring of Y104 outside of the core is surprising, however it is located sideway and approximately 6.3 Å from the sulphur atom of M173, which is in a favourited range for a sulphur-π interaction[602] which, coupled with the solvation of the hydroxyl group of the tyrosine, likely makes up for the energy penalty of exposing the solvent-exposure of the aromatic ring.

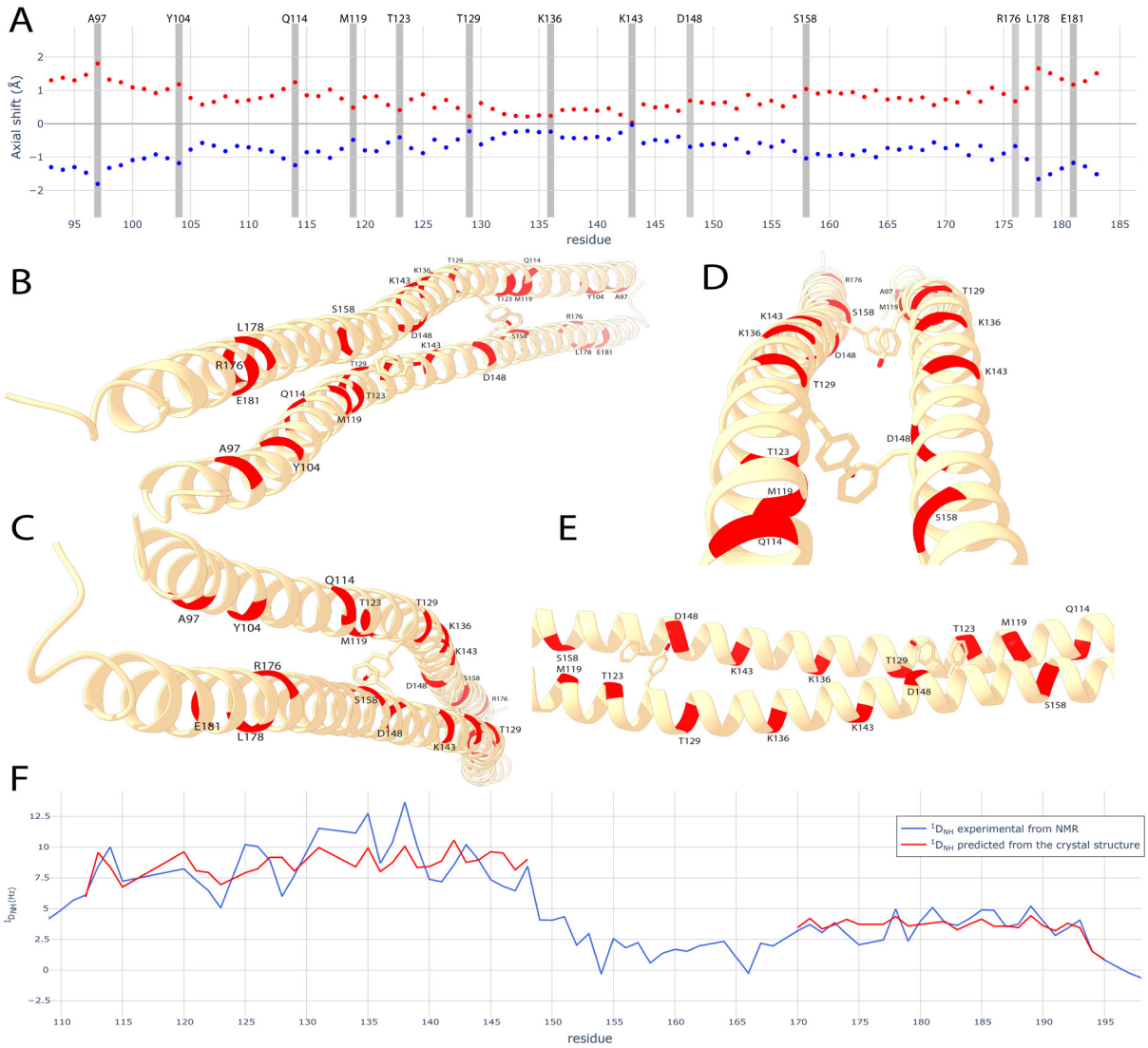


Figure 5.17: The coiled-coil crystal structure presents particular deformations and is present in solution.
A: Axial shift plot, which allows to clearly identify deformations of the helices. Like in the previous figure, the red dots correspond to the A chain of the PDB structure file, while the blue correspond to the B chain.
B, **C**, **D**, **E**: Views of the coiled-coil structure with the key residues involved in deformations highlighted in red and with the Y126-F151 stacking visible.
F: Experimental residual dipolar couplings of the backbone N-H bond vectors $^1D_{NH}$ of CAHS-8 at 293 K (blue) and fitted dipolar coupling from a single helix of the crystal structure (red).

This bend at Y104 is compensated by a change in the direction of the α-helix at Q114 visible in the axial shift plot and the structure (fig. 17) which brings the two helices closer and likely results from the pair of polar residues Q115-R162 having their sidechains pointing outside the core to be solvated, despite being in position **h** of their stutter thus being projected directly towards the core, hence their Crick angles being comprised

between 2.8° and 8.2°. Although this change of direction of the helix doesn't appear on the amide proton secondary chemical shift plot, either because it is not present in solution or because it is not a curvature of the helix per se, the RDC profile of CAHS-8 does present a clear discontinuity at Q114 which supports the later hypothesis.

Further helix deviations at M119, T123 and T129 as well as at K143, D148 and S158 are visible in the helical shift plot (fig. 17) and probably arise to accommodate the π-π interaction between the large Y126 and F151 as well as to allow the formation of salt-bridges and the solvation of polar residues in core positions. This π-π stacking is probably crucial for the function of the CAHS protein as Y126 is a tyrosine in 100 % of the aligned sequences while F151 is phenylalanine in 96 % and a tyrosine in the 4 remaining

In turn, the effects of these bends on the helix are numerous in the regions 119-129 and 143-151. First, M119 is in position **a** of the heptad which is followed by a stutter to make a hendecad, and thus is projected more towards the outside than normal so the following **h** residue, Y126, can face directly the core. As such, their respective Crick angles are respectively 29.7°, more than twice the expected average 14°, and -6.7°. A similar situation happens with L155, in position **a** which is flanked by two stutters with F151 and R162 in position **h**. As these two positions face directly the core, with Crick angles of -4.0° and 5.5°, L155 faces more the outside and therefore has a Crick angle of 37°. This results from the presence of two stutters and allow the accommodation of F151 within the core, allowing it to face Y126. In turn, this creates an alignment of the hydrophobic sidechains of M119, L155 and F151 which may limit their exposure penalty.

As per the **h** residue between M119 and Y126 and thus counterpart of L155, K122, its sidechains points towards the outside where its α-amino group locate on both sides of the antiparallel coiled coil in a very favourable region for the formation of hydrogen-bonds and salt-bridges with residues of the other helix. Indeed, the nitrogen of the α-amino group of K122A and K122B locate at respectively 2.5 and 2.9 Å of the hydroxyl oxygen of S158, 5.2 and 3.7 Å of D154 furthest carboxyl oxygen, 5 and 5.2 Å of E157's as well as 7.2 and 6 Å of D161's. Finally, it also locates at 5 and 5.6 Å of E118 of the same residue on the same helix. The location of K122's long sidechain outside the core thus allows a closer

proximity between the aromatic rings of Y126 and F151. In turn, the hydrogen-bonding of K122 is possible with S158, which is the residue facing M119, as its sidechain points outside, hence the deviation of the helix axis observed at this position, which allows the solvation of its hydroxyl group and this interaction with K122.

The T123 deviation widens the distance between the helices in order to permit the following core residue Y126 to face F151. As discussed, this π-π stacking appears crucial for the function of CAHS proteins given its extreme conservation, and thus particular deviation appears crucial to accommodate it. This position requires a small amino-acid to not sterically hinder the Y126-F151 interaction and indeed, T123 is a threonine in 81% of the sequences, while in respectively 15% and 2% of them, it is an alanine and a serine.

As per the deviations at T129 and D148, they happen concurrently as these residues are each other symmetrical counterpart, being respectively in positions **k** and **e** and as such, are projected in opposite direction with respect to the core. Thus, while bending at T129 pulls the helices closer after the one at T123 moved them apart to permit the Y126-F151 interaction, the one at D148 widen the distances again so F151 and its facing Y126 can interact.

In turn, in between these deviations, the region 132-139 is indeed mostly straight with a small change of direction at K136 in the structure. This particular region has the highest sequence conservation of any region of the sequence as well as the fast transverse relaxation rates and apparent exchange contribution, in particular K136 and I137, which are respectively in position **g** and **a** of their heptads. The two K136 of the antiparallel coiled-coil are not equivalent, with K136A having its sidechain coming back towards the other helix, allowing its nitrogen to locate at 3 Å from the carboxyl oxygen of the facing E140 of the other monomer, but also at 3.6 Å and 6.5 Å of E133's and E140's of the same helix, while K136B points more toward the outer face of its helix, locating therefore at more than 8 Å from the facing E140 but at 5.7 Å of the closest E133 carboxyl oxygen of the same protein. As per this particular E133 residue, its closest carboxyl oxygen is situated at less than 4.3 Å on the two sides of the facing R147's closest nitrogen, just before the aforementioned D148 deviation.

Finally for these deviations of the centre, as previously mentioned, the central region 130-150 follows a 25/7 periodicity, which arises from one stutter and three consecutive heptads. This periodicity typically results in helices being initially straights and then eventually having a slight-supercoil, which can be seen from E140 to E150, and triggers the bending around K143 observed in solution as well.

The fit of the RDC measured on CAHS-8 at 293 K using a single helix of the crystal structure shows a good agreement (fig. 17 F) not only for the 112-148 region in helix 1 (reduced $\chi^2 = 3.24$) but also for the 170-195 region of helix 2 (reduced $\chi^2 = 0.55$). This is remarkable considering the fact that these regions are partially populated and are overwhelmingly monomeric at the 350 µM and 293 K conditions in which they were measured while the crystal structure is that of a dimer with the numerous singular deformations previously mentioned. This suggests that the helices of CAHS-8 sample conformations similar to the DCC crystal structure in solution.

4.1.2.3 HD has a reduced ability to fibrillate

As discussed in chapter 3, CAHS-8 forms fibres at room-temperature at a high enough concentration, with these fibres being composed of fibrillar units and with a height of 2-3 nm (fig. 18 A, B). In turn, HD only forms exceedingly rare and unstable fibres which quickly dissolve (fig. 18 C, D). Moreover, room-temperature SAXS measurements of HD at 1.14 mM, a high concentration at which CAHS-8 is already in a gel state, does not report on the presence of extended objects such as fibres which would dominate the scattering at small angles, which can be seen in CAHS-8 at 190 µM already (fig. 18 E, F). In particular, while the profile of CAHS-8 at 190 µM on the Kratky plot reports clearly on the presence of a highly structured domain (fig. 18 F, green), that of the 1.14 mM HD (red) is similar to the profile of CAHS-8 at 76 µM (violet) which is chat is expected of an intrinsically disordered protein[470].

However, at 278 K, a temperature at which CAHS-8 forms thermally reversible fibres if concentrated enough, HD forms increasingly numerous amorphous aggregates at a rate

dependent on the concentration, eventually giving a milk colour to the sample. These aggregates are also reversible when the sample is heated, but forms again when the temperature decreases again.

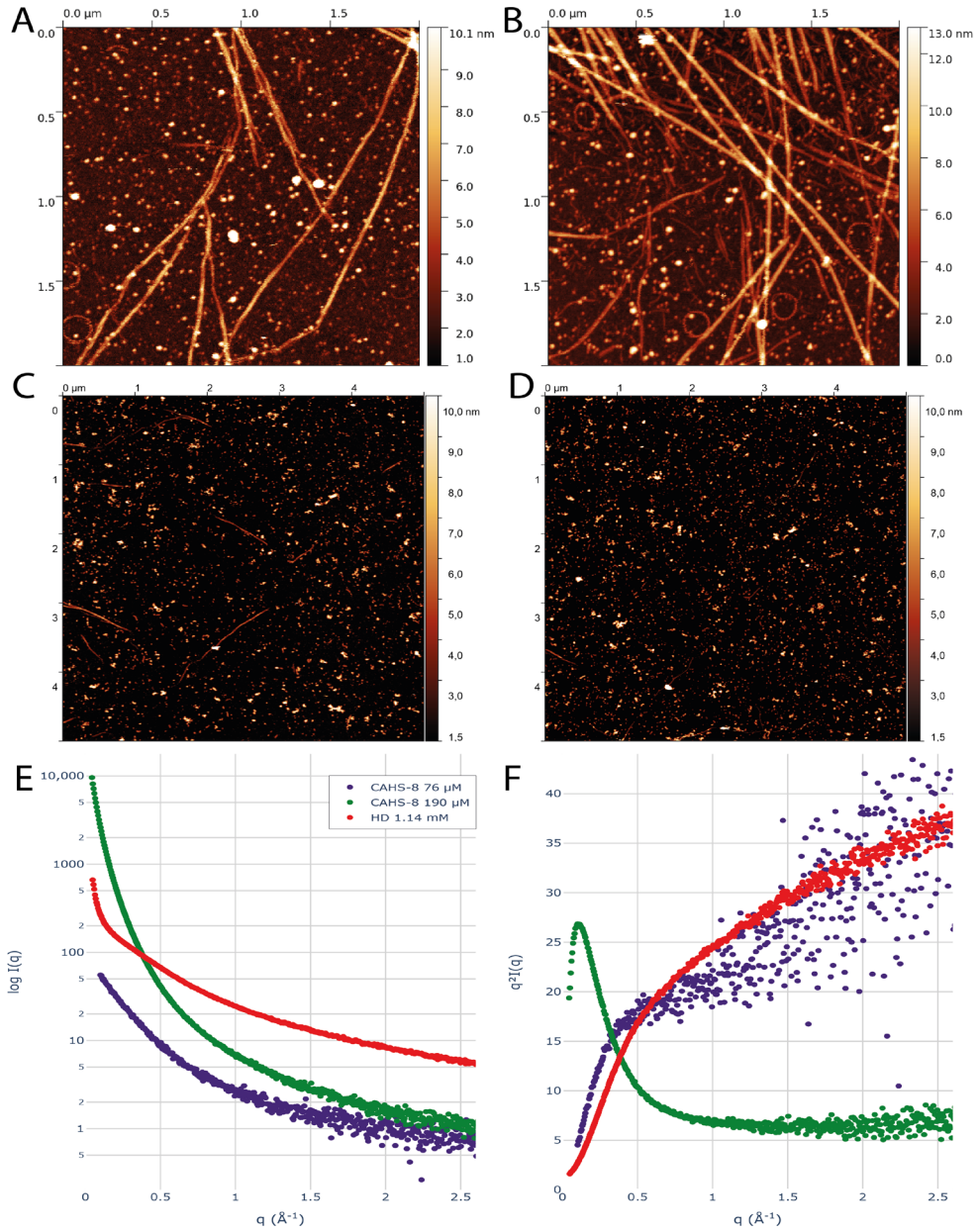


Figure 5.18: HD has a reduced ability to form fibrillar assemblies compared to CAHS-8.
A and **B**: AFM images of the fibres of CAHS-8 (**A**) and the same sample approximately 40 min later (**B**), showing a continuous fibrillation.
C and **D**: AFM images of the fibrillar assemblies of HD (**A**) and the same sample approximately 35 min later (**B**), showing their dissolving.
E and **F**: SAXS profile (**E**) and Kratky plot (**F**) of CAHS-8 at 76 µM (2 mg/mL, violet), CAHS-8 at 190 µM (5 mg/mL, green) and HD at 1.14 mM (15 mg/mL, red). HD shows signs of aggregation but not of fibres, unlike CAHS-8 at 1/6th of the concentration of HD.
The Kratky profile of CAHS-8 at 76 µM was scaled by the square of its ratio with that of HD's for easiness of comparison.

This behaviour suggests that although HD can form coiled-coil dimer, it has a greatly reduced ability to form fibrils. When the temperature decreases and therefore starts to favour the forms with the lowest potential energy, HD appears to chaotically self-assemble into aggregates, forming only rare fibrils in the process. Consequently, HD was never observed to form a gel in any temperature or concentration at which CAHS-8 does.

In order to avoid the aggregation of HD at 278 K, the protein was deuterated as it was observed in the crystal making process that deuteration limits its oligomerisation. This effectively prevented a visible change of colour of the sample even after days of measurements and permitted to assign the protein.

The secondary chemical shifts for the alpha carbons ($\Delta\delta\ ^{13}C^{\alpha}$, fig. 19 A) and amide protons ($\Delta\delta\ ^{1}H^{N}$, fig. 19 B) show that HD has a greater helical propensity at 278 K than at 293 K, including in helix 1. This was hypothesised for CAHS-8 but couldn't be observed due to the broadening beyond detection of helix 1. Indeed, the $\Delta\delta\ ^{13}C^{\alpha}$ profile shows a consistent increase of the secondary chemical shifts from 293 to 278 K while the $\Delta\delta\ ^{1}H^{N}$ profile shows a consistent decrease of the secondary chemical shift as the temperature decreases from 303 K to 293 K and then to 278 K, with both of these trends indicating an increase of the helical populations in the conformational sampling of HD.

The comparison of the $R_{1\rho}$ relaxation of HD at 278 and 293 K at 950 MHz (fig. 19 C) shows that the $R_{1\rho}$ rates constants of helix 1 passes from ~18 s^{-1} on average at 293 K to ~40 s^{-1} at 278K, with many outliers displaying relaxation rates above 50s^{-1} such as S110, R113, T123, R127, T129, A130, E131, K136, I137 and H145. In the linker and helix 2, the rates constants increase from an almost uniform baseline of ~10 s^{-1} at 293 K to ~25 s^{-1} on average at 278 K. Although the rates cannot be directly compared per se due to the $R_{1\rho}$ relaxation rates at 278 K including the contribution of the non-experimentally decoupled deuterium, it is highly unlikely that the increase viscosity associated with a lower temperature as well as more populated helical conformations would increase the $R_{1\rho}$ rates to such extent.

This behaviour is better explained by the transient oligomerisation of HD at 278 K, with this increase in the relaxation rates resulting from the contribution of the global rotational diffusion time of the dimer as well as the exchange between the monomer and dimer form.



Figure 5.19: The deuterated HD at 278 K shows signs of the coiled-coil formation.
A: Secondary chemical shifts for the alpha carbons Δδ $^{13}C^{\alpha}$ shows a consistently higher helical population at 278 K (blue) than at 293 K (black) throughout the whole protein.
B: Secondary chemical shifts for the amide protons Δδ $^{1}H^{N}$ also shows a higher helical population at 278 K compared with 293 K or 303 K (orange) as well as bigger oscillations.
C: $R_{1\rho}$ relaxation of the deuterated HD at 950 MHz and at 278 K and 1.14 mM (blue) and 293 K and 350 µM (black) shows significantly faster relaxation at 278 K. Residues T123, T129, K136 and S158 are highlighted as they show sign of chemical exchange, having faster relaxation than the neighbouring residues baseline, are associated with the deformation of the helices to accommodate the coiled-coil oligomerisation.
D: CPMG relaxation dispersion curves of the deuterated HD at 293 K for residues in the linker R152, E157 and S158 with the experimental point in grey and the fitted two-state profile in blue.

Nevertheless, this exchange would have to still be faster than 1500 s^{-1} as per the spin-lock frequency used during the measurement of the $R_{1\rho}$ relaxation.

However, CPMG relaxation dispersion experiments on the deuterated HD at 293 K and 600 MHz reveals the presence of an exchange process in the linker (152-166) which could be fitted to a two-state model using residues 151-153, 156-161 and 165 together, yielding $k_{\mathrm{ex}} = 749 \pm 105\ s^{-1}$ (fig. 19 D, reduced $\chi^2 = 1.97$). As dispersion was significant only in the linker and in residue 133 and as the k_{ex} of this exchange process is smaller than 1500 s^{-1} and thus cannot be the one process attributed to the coiled-coil oligomerisation, this exchange in the linker can be assigned to a conformational exchange between a disordered and a helical conformation, thus explaining how the linker can be helical in the crystal structure. Unfortunately, due to a very low signal to noise ratio in the helical domain of CAHS-8, no CPMG relaxation dispersion characterisation could be performed.

Thus, HD appears to retains its ability to form a coiled-coil dimer but has a significantly reduced ability to form fibrils as this higher-order fibrillation happens chaotically without the disordered termini. In order to understand why, a model-free analysis of the spin relaxation rates was performed on CAHS-8 and HD.

4.1.2.4 Model-free analysis of the helices

Model-free analysis treatment of the relaxation rates of the helices and neighbouring residues was attempted for CAHS-8 at 278 and 293 K as well at HD at 293 K. In each of these 3 conditions, the 7 measured rates were fitted together (R_1, R_2 and nOe at two fields as well as η_{xy} at one field, and previously shown in fig. 3, 9, 10 and 13). The shape of the spectral density functions was modelled using either two (fitting therefore 5 parameters, i.e. τ_3, τ_2, A_2, $\mathrm{R_{ex}}$ and the CSA/DD angle θ), or three dynamic modes (fitting therefore an additional sixth parameters corresponding to the amplitude of a fast motional mode, i.e. A_1, with its timescale being fixed to 50 ps as described[439,452]). The model free analysis is not presented for CAHS-8 at 303 K as only 6 rates were measured, rendering the three motional mode model fit non robust, while two motional modes is poorly fitting, which is agreement with the observed trend of less helical population and increased flexibility at 303 K.

The resulting fits were discriminated as per χ^2 and Akaike information criteria. Namely, the 3 dynamic modes fits were chosen for residues 151-200, which correspond to the very end of helix 1, the linker, helix 2 and the first 5 residues of the C-terminal disordered chain, as well as the 87-94 residues of the N-terminal IDR. The reproduction of the rates, CSA/DD angle, and fit metrics are given in the annex.

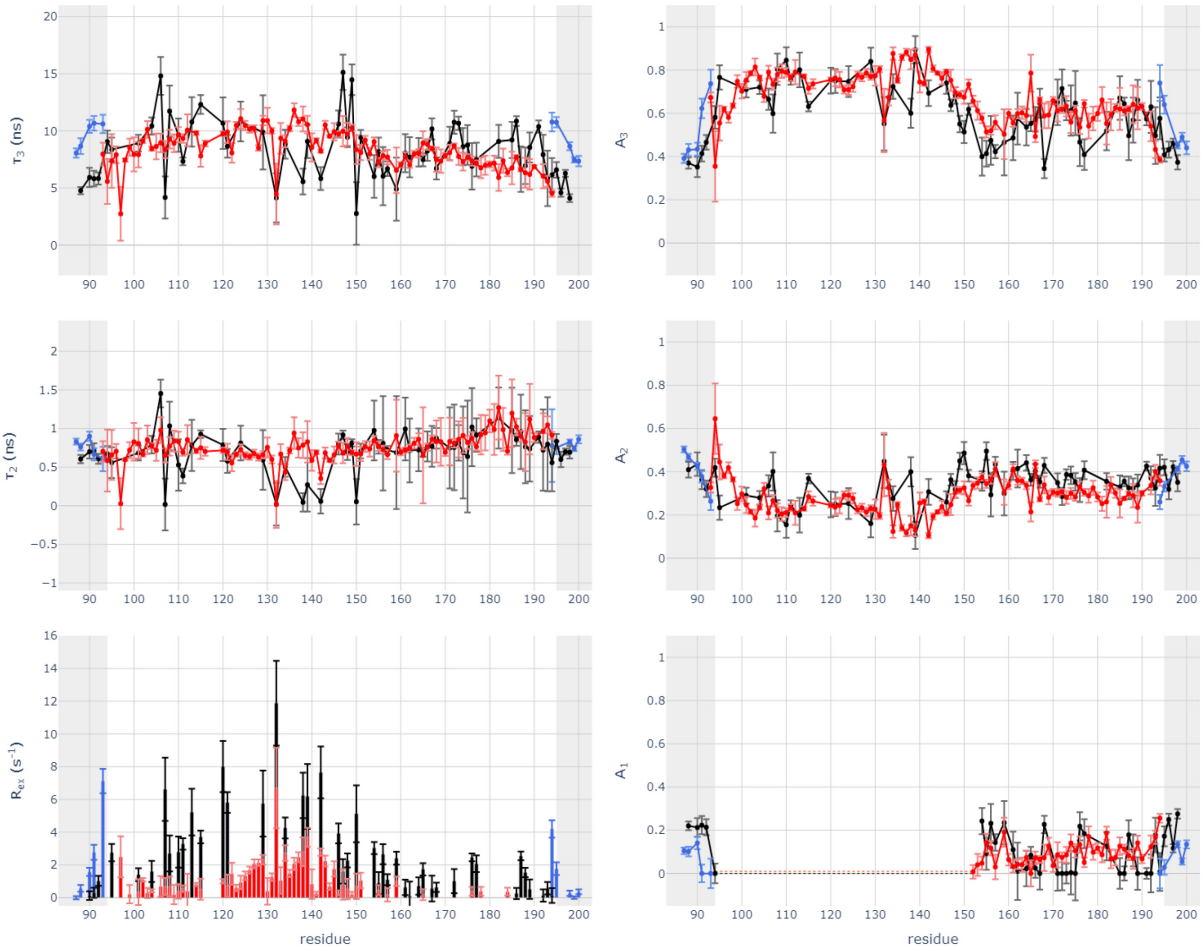
Given the signal to noise ratio and number of rates compared to the parameters, the fits are not expected to be extremely precise but nevertheless, constitute a useful first estimation of these parameters.

4.1.2.4.1 Temperature dependence of CAHS-8

The comparison of the CAHS-8 parameters at 278 K (fig. 20, blue) and 293 K (black) for the few residues available at 278 K show a massive increase in the timescale τ_3 of the slow-motions which appears to be of more than 10 s^{-1}, while their associated amplitudes A_3 are consistently higher than at 293 K.

Moreover, a higher exchange contribution is also detected at 278 K in the disordered residues flanking the helices, which is absent at 293 K. This is in agreement with the elevated R_2 rates obtained at these temperatures.

Figure 5.20: The model-free analysis treatment of the spin relaxation rates of CAHS-8 and HD shows different exchange profile.
CAHS-8 at 278 K is in blue, CAHS-8 at 293 K in black and HD at 293 K in red.
Top: Correlation time of the slow motional mode τ_3 (left) and its associated amplitude A_3 (right).
Middle: Correlation time of the intermediate motional mode τ_2 (left) and its associated amplitude A_2 (right).
Bottom: Chemical exchange contribution to the transverse relaxation R_{ex} (left) and amplitude of the fast motional mode A_1 (left, its correlation time is fixed to 50 ps).
Several neighbouring residues in the flanking disordered regions in N-terminal (87-94) and C-terminal (195-200) are included in the figure for CAHS-8 and are highlighted in grey. These residues, as well as the last residue of helix (151), the linker (152-166) as well as helix 2 (166-194) were fitted using a shape of the spectral density function comprising three motional modes, while the rest of the residues of helix 1 (95-150) uses two as their α-helical conformations are significantly populated, hence their absence of fast motional mode amplitude A_1. The choice of three or two motional mode was done as per χ^2 and Akaike criterion metrics which are shown in the annex, as well as the reproduction of the rates.



4.1.2.4.2 HD and CAHS-8

The comparison of HD (fig. 20, red) results with CAHS-8 (black) at 293 K reveals that the timescale of the slow components τ_3 and its associated amplitude A_3 are broadly similar, with τ_3 appearing only slightly more bell-shaped in HD. In turn, the timescales and amplitudes of the intermediate motional mode τ_2 and A_2 in HD are also similar to in the full-length CAHS-8 (fig. 20, middle row). However, the amplitudes of the fast timescale motions A_1 (fig. 20, bottom right) in helix 2 are higher in HD than in CAHS-8, indicating a greater flexibility.

Finally, the exchange contribution R_{ex} have a very different profile, showing the stairs-like profile of HD already observed in the R_2R_1 product between residues 120-129 and 133-139, and an overall much smaller exchange contribution than in CAHS-8 which shows localised and higher exchange peaks (fig. 20, bottom left).

It has been demonstrated that the presence of disordered flanking chains on an α-helix have a length-dependent effect, with the longer the chain, the higher the increase in the timescale of occurrence of the slowest motions τ_3 due to an effective increase of the rotational diffusion time in a viscosity-dominated drag effect[452]. However, their associated amplitudes A_3 and the amplitudes or timescales of the intermediate (A_2 and τ_2) and fast motions (A_1 and τ_1) are barely affected by the presence or length of such flanking disordered regions[452].

Thus, the similarity in the τ_3 values of CAHS-8 and HD is surprising because they indicate that the helices of HD have a tumbling time relatively similar to CAHS-8's despite the presence of 94 disordered residues in N-terminus and 39 disordered ones in C-terminus. Although HD appears to sample slightly more populated helical conformation than CAHS-8 at 293 K (fig. 11), it doesn't appear more rigid as per their nOe relaxation (fig. 12) and therefore, this increased amplitude of the slow reorientational motions cannot be attributed the tumbling of a stiffer helix.

An explanation to these results can be that the helices of HD, in particular the first one, present a rotational diffusion anisotropy. Indeed, such anisotropy of the diffusion tensor manifests as fictitious slow motions in an isotropically modelled spectral density function, and would therefore explain why HD appears to tumble so slowly, which in turn would bias its associated amplitude and thus the other amplitudes as well because their sum is by definition equal to 1.

Using either of the monomers in the crystal structure and excluding the residues that display obvious sign of exchange, the fit of the diffusion tensor of helix 1 using NMRFit or Tensor2 result in a very good reproduction of the R_2/R_1 ratio. However, the eigenvalues of the prolate tensor do not appear physical and the angular distribution of the N-H bond vectors peaks at π/2 instead of being minimum there, while a second unexpected minimum is present between π/2 and π which indicate a probable second order effect. The oblate tensor presents a more coherent angular distribution, but as helix 1 is 55 residues long and therefore highly prolate, such a tensor shape appears improbable.

NMRFit was used to fit the diffusion tensor to thousands of conformers generated by flexible-meccano on the basis of the secondary chemical shift data. This allowed to identify several conformers that reproduces well the relaxation. It is notable that these well-fitting conformers are bent helices in one or several of the positions previously highlighted in the crystal structure, which supports their correctness. However, these conformers present the same counter-intuitive angular distribution for the prolate tensor and have an oblate one being more coherent.

An oblate diffusion tensor would require many parallel helices to be aligned, which is not supported by the shape of the SAXS curve, wasn't observed in AFM and it would be nevertheless highly dubious that this could be observed by solution NMR. In turn, the more probable prolate tensor suggests that HD is indeed in a monomer-dimer equilibrium exchange process in solution that affects one side of the helix and not the other, as in the coiled-coil crystal structure. Unfortunately, this exchanging side of the helix would have to be excluded from the fit to obtain the diffusion tensor of the monomer, which in turn would bias the angular distribution of the N-H bond vectors. Therefore, the rotational diffusion tensor could not be obtained and no model free analysis using an expression of the spectral density function taking this anisotropy into account could be performed.

Nevertheless, although HD appears to exchange with its coiled-coil dimer form in solution and was crystallised in this conformation, its greatly reduced ability to form fibrils in comparison with CAHS-8 as well as the likely correlated differences in their exchange profile and apparent increased flexibility in helix 2 highlights the role of the intrinsically disordered chains flanking this helical domain. The study of the fibrils and therefore of the role of the disordered termini is the object of the next chapter.

Chapter 6: The fibres and the disordered chains

« As above, so below. »

The Emerald Tablet, 9th century.

5.1 Results

5.1.1 Structure of the fibrillar assemblies of CAHS-8

5.1.1.1 Fibrils in the lattice

The analysis of the structure of the antiparallel dimeric coiled-coil of CAHS-8 presented in the previous chapter shows that the outermost face of the C-terminal half of the coiled-coil, that is the **b**, **c** and **f** positions of the repeats, are enriched in hydrophobic residues (fig. 1 A). This is highly unusual for a coiled-coil which typically arise from amphipathic α-helices, as having one apolar side and the rest being polar result in the apolar faces clustering together due to the hydrophobic effect. The representation of the hydrophobic surface of the crystal structure shows that this particular high distribution of apolar residues is indeed distributed on the C-terminal outer face of the α-helices, with V149, I156, I160, V167, A171, A174, A185 and L189 all facing outwards (fig. 1 B). As such, this other hydrophobic face extends beyond the residues involved in the coiled-coil which ends at Q183.

Moreover, this particular solvent-exposed face of helix 2 is also where the B-factor of the structure is the lowest, indicating that this is the most stable and least-variable region of the whole structure (fig. 1 C). In turn, the sequence-alignment of the 85 known and putative CAHS proteins presented in the previous chapters shows that these hydrophobic residues are also very highly conserved (fig. 1 D, marked by a star), which indicates that these residues have an important role in the function of the desiccation-protection CAHS proteins.

The expansion of the symmetry of the asymmetric unit within the crystal lattice reveals the presence of a fibrillar network throughout the crystal (fig. 2 A), with the fibrils having two directions due to the *C*121 symmetry of the crystals. As such, while the two-stranded coiled-coil presented in the previous chapter is the asymmetric unit of the unit cell which comprise two copies of this coiled coil, these fibrils are most likely the real biological assembly of the protein.

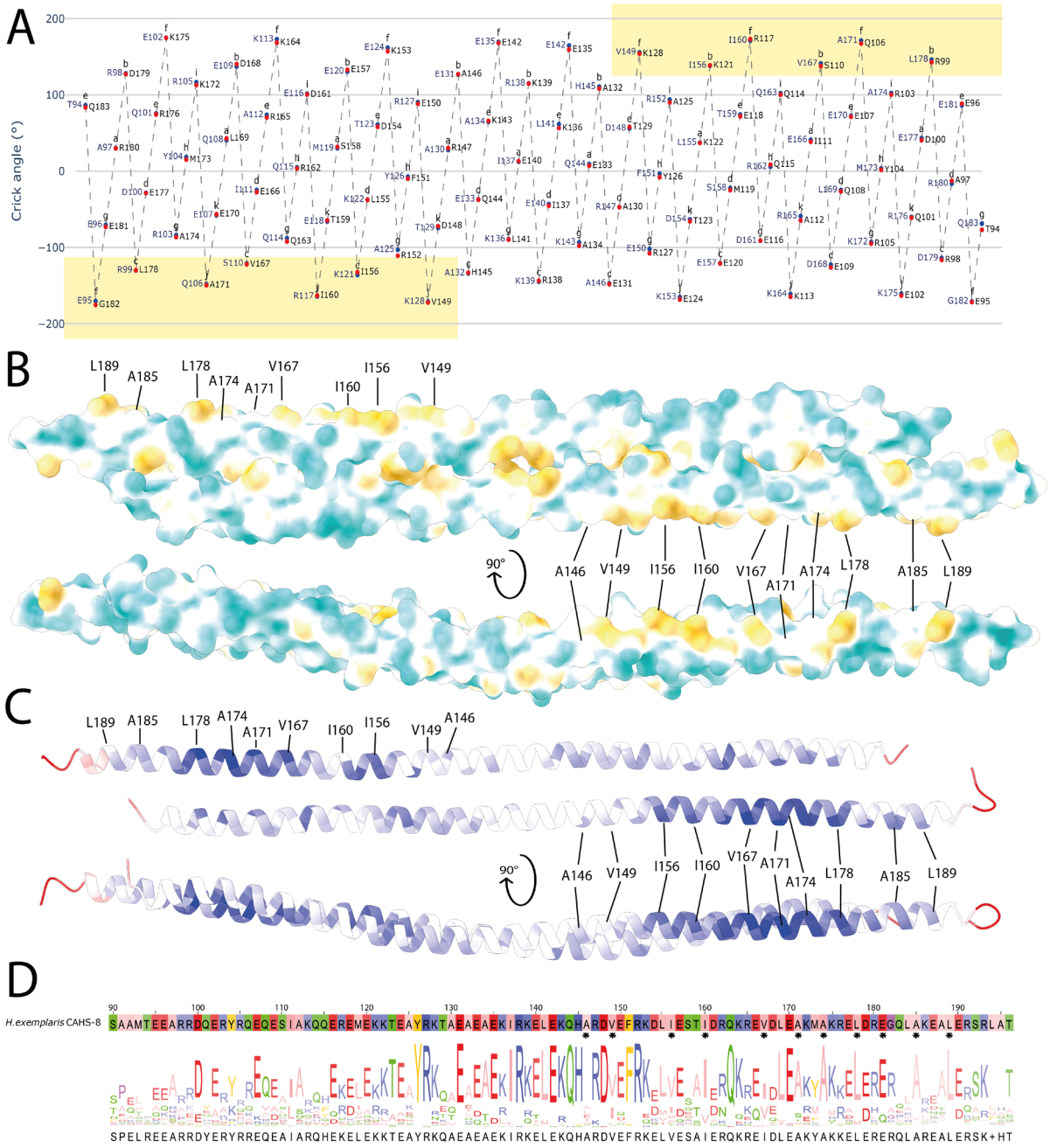


Figure 6.1: The outer face of the C-terminal half of the helical domain of CAHS-8 is apolar.

A: Plot of the Crick angles of the coiled-coil crystal structure shows that the outer face of the C-terminal half of the coiled-coil is enriched in hydrophobic residues instead of polar ones (yellow box).

B: Hydrophobic surface representation of the crystal structure (polar in cyan, apolar in yellow) shows this hydrophobic side, with the relevant residues being marked and extending beyond the coiled-coil dimer interface.

C: B-factor of the crystal structure, with the lowest B-factor in blue and highest in red. The C-terminal outer face has the lowest B-factor of the whole structure.

D: These hydrophobic residues on the C-terminal outer face of the helical domain of *H.exemplaris* CAHS-8 (marked by a star) are highly conserved amongst CAHS proteins.

Isolating two of these fibrils (fig. 2 B) shows that these fibrils are composed of the dimeric coiled-coils of CAHS-8 as a unit that assemble with one another via the outer face of the C-terminal half of the α-helices highlighted previously. These fibrils have a total thickness of 2.9 nm, which is bigger than the thickness of a regular α-helix due do the topological deformations along the helix discussed in the previous chapter, and have a width of about 4 nm at their widest, where the coiled-coil dimers interact with each other (comprising 4 helices), and 2 nm at their thinnest (comprising two helices). In turn, two paired fibrils together form a fibre of approximately 9 nm of width.

As discussed in chapter 4, CAHS-8 was observed to form an increasing number of fibres as the concentration of the protein increased due to dehydration of the sample (fig. 2 C), with the fibres having a depth (estimated from AFM) of approximately 2 to 3 nm. As mentioned in the dedicated section in the introductory chapter 2, although the dimensions in the transvers (x-y) plane of AFM images cannot be interpreted in a simple manner, the Laplacian-weight filtering method allows to somewhat mitigate the effects of the tip convolution on the object and therefore enhances the lateral resolution of AFM imaging[516,517]. Using this method on the fibres image (fig. 2 D) allows to distinguish that the CAHS-8 fibres are composed of two paired fibrillar assemblies (fig. 2 E, F). Moreover, Negative-Staining Transmission Electron Microscopy (NS-TEM) of the CAHS-8 fibres shows that these fibres are approximately 9.5 nm large and that they are composed of two thinner fibrils (fig. 2 G, H).

This organisation of a fibre being composed of two fibrils is in agreements with the AFM images and with the crystal structure of the fibrils (fig. 2 B). Indeed, the individual fibres observed in the processed AFM image and in NS-TEM (fig. 2 D-H) are 9.5 nm large as per NS-TEM. However, they appear larger in AFM, with apparent width of 20 to 27 nm, which is most probably due to their instability, as they were forming and dissolving in that image, persistent convolution effects as well as the presence of the disordered chains that may trigger a change of phase or the amplitude of the AFM tip, and as such, may not be entirely corrected by the Laplacian-weight filtering method. Nevertheless, the 9.5 nm width observed in NS-TEM appear to correspond to the 9 nm width of two adjacent fibrils in the crystal lattice (fig. 2 B). Moreover, these individual fibrils appear large of

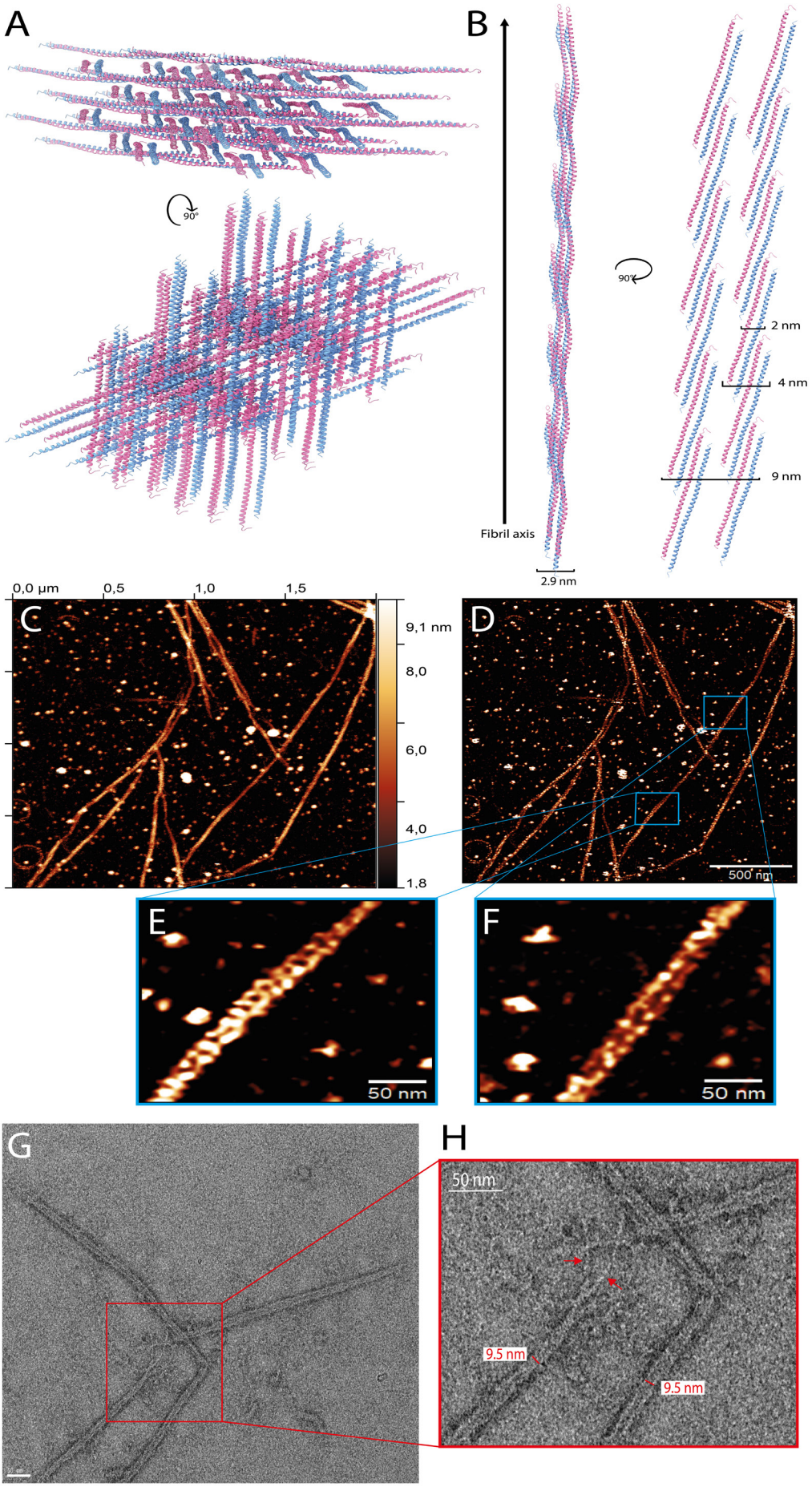

A
90°
B
Fibril axis
2.9 nm
90°
2 nm
4 nm
9 nm
0,0 µm
0,5
1,0
1,5
C
9,1 nm
8,0
6,0
4,0
1,8
D
500 nm
E
50 nm
F
50 nm
G
H
50 nm
9.5 nm
9.5 nm

Figure 6.2: Structure of the fibrils and fibres of CAHS-8.

A: The expansion of the crystal symmetry reveals a fibril network spanning the lattice. The A chains of the coiled-coil dimer are coloured in pink and the B chains in blue.

B: The crystal structure of these fibrils shows that the antiparallel coiled-coils dimer of CAHS-8 assemble in an antiparallel arrangement via the outer face of the C-terminal half of the helical domain, forming fibrils approximately 4 nm large and 2.9 nm thick, with two adjacent fibrils together forming a 9 nm large fibre.

C: AFM images of the fibres of CAHS-8, showing a height of approximately 3 nm.

D, E and **F**: Laplacian-weight filtered version of the images of the panel **C** (**D**), and two excepts (**E**, **F**) permitting to distinguish that a fibre is made of two fibrils.

E and **H**: Negative-staining transmission electron microscopy of the fibres of CAHS-8 shows an approximately width of 9.5 nm and that they are constituted of two fibrils large of approximately 9.5 nm (**H**).

approximately 2.5 to 4 nm (fig. 2 H), which is also the width of the individual fibrils observed in the crystal structure (fig. 2 B).

Additionally, the fibres being approximately 2 to 3 nm thick as per the original AFM image (fig. 2 C) also correspond to the thickness of the crystal structure of the fibrils (fig. 2 B). As the most precise AFM measurement, this slight depth variation can probably be attributed to the instability of the fibrils or is due to the imaging setpoint, i.e. the amplitude of the tip oscillation, which therefore corresponds to how strongly the tip presses on the sample.

The imaging of the fibres of CAHS-8 by cryo Electron Microscopy (cryoEM, fig. 3 A, B) and the obtained two-dimensional classes confirm that the fibres are indeed 9.5 nm wide and composed of two fibrils (fig. 3 C). In turn, the width of the individual fibrils in the 2D classes varies from 2.5 nm at their thinnest to 4 nm at their widest which is in agreement with the crystal structure (fig. 2 B).

However, unfortunately, the 2D classes features are not well defined on close inspection (fig. 3 C). This possibly due to the inhomogeneity of the particles, contribution of the disordered regions as well as the small difference between the dimensions of the individual fibrils, i.e. 3 nm thick and large of 2.5 nm wide at their thinnest to 4 nm on the region where the coiled-coil pairs assemble.

Moreover, as the fibres were often entangled, careful manual picking had to be performed which was limited by their small, 2.5-4 nm individual width of the fibrils and overall difficulty to see them due to their 3 nm thickness greatly limiting the scattering of the electron beam. As such, not enough particles could be picked to obtain better quality classes or a three-dimensional model.

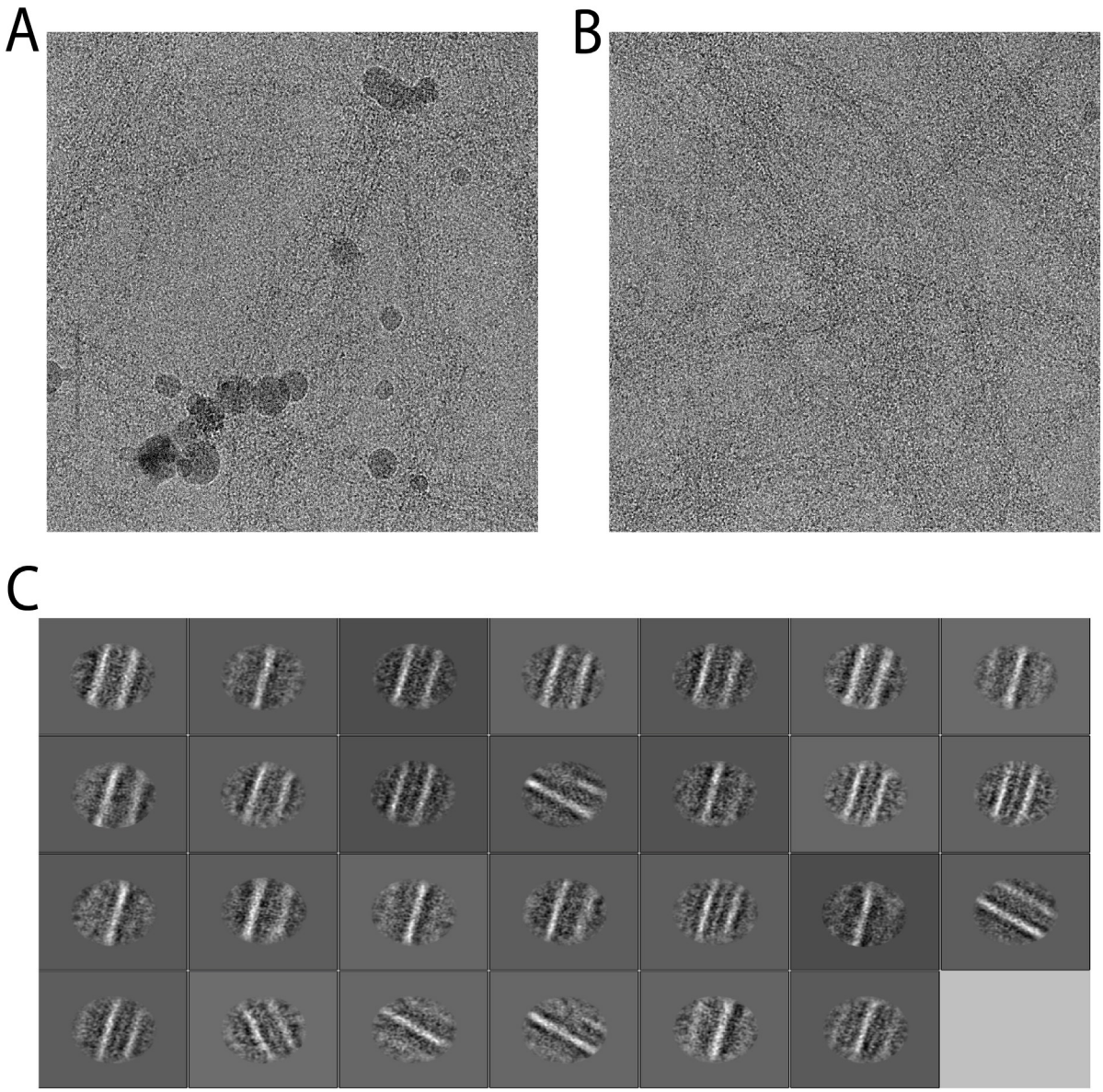


Figure 6.3: Cryo Electron Microscopy study of the fibrillar assemblies of CAHS-8

A and **B:** CryoEM images of the cross-linked fibrillar assemblies of CAHS-8. The fibrils and fibres are relatively straight but often entangled and are inhomogeneous.

C: Two dimensional classes of the particles manually picked showing either the two fibrils together composing one fibre, or single fibrils. The width of a fibre made of two fibrils is consistently 9 to 10 nm while the width of the individual fibrils varies from 2.5 nm at their thinnest to 3.5 to 4 nm at their widest.

5.1.1.2 A second coiled-coil drives the fibrillation

The analysis of the fibril crystal structure shows that the coiled-coils dimers assemble with each other by forming a second antiparallel coiled-coiled which will be henceforth called the fibrillation-driving coiled-coil (FCC), spanning the 149-190 region (fig. 4 A, black rectangle). As per the first coiled-coil presented in the previous chapter, it will be called the dimerisation-driving coiled-coil (DCC) to distinguish it. As the coiled-coil pair assemble via the outer-side of the C-terminal half of the helices, to form a second coiled coil, all the A chains have the same directions in the fibrils, and, mutatis mutandis, the B chains as well.

The measurements of the metrics of this second, fibrillation-driving coiled-coil permit to see that like the dimerisation-driving coiled-coil, the FCC also follows a hendecad periodicity. Indeed, beginning with a stutter (149-152), the FCC has three consecutives hendecads, respectively 153-163, 164-174 and 175-185 as well as an incomplete, five-residue long heptad 186-190 (fig. 4 B). The angles show a clear clustering of the residues depending on the positions, with for instance the **h** positions, namely V149, I160, A171 and G182 showing the closer orientational angle to 0° of any positions, as expected from stutters **h** positions as previously described. The analysis of the core positions of the repeats of the FCC shows four lysines in position **a**, namely K153, K164, K175 and K186. This is surprising considering that **a** position is a core position forming the hydrophobic interface between the helices, although the presence of the four stutters in the repeats changes slightly the core packing with **a** facing more outward to accommodate the stutter **h** position (fig. 4 B). In turn, the periodicity of the coiled is approximately 3.67 over the FCC which corresponds to the hendecad periodicity, i.e. 11 residues for three turns (fig. 4 C). This is also visible from the axial rotation of the helices as this metric evaluates the deviation of the Crick angles compared with expected values for a canonical coiled-coil, and indicates that the FCC consistently deviates from the canonicity, which correspond to values of 0° (fig. 4 D). Additionally, the axial shift of the helices shows a shift between the facing positions of the two helices smaller than 1 Å with slight deviations (fig. 4 E) which are studied in greater detail in the next paragraphs.

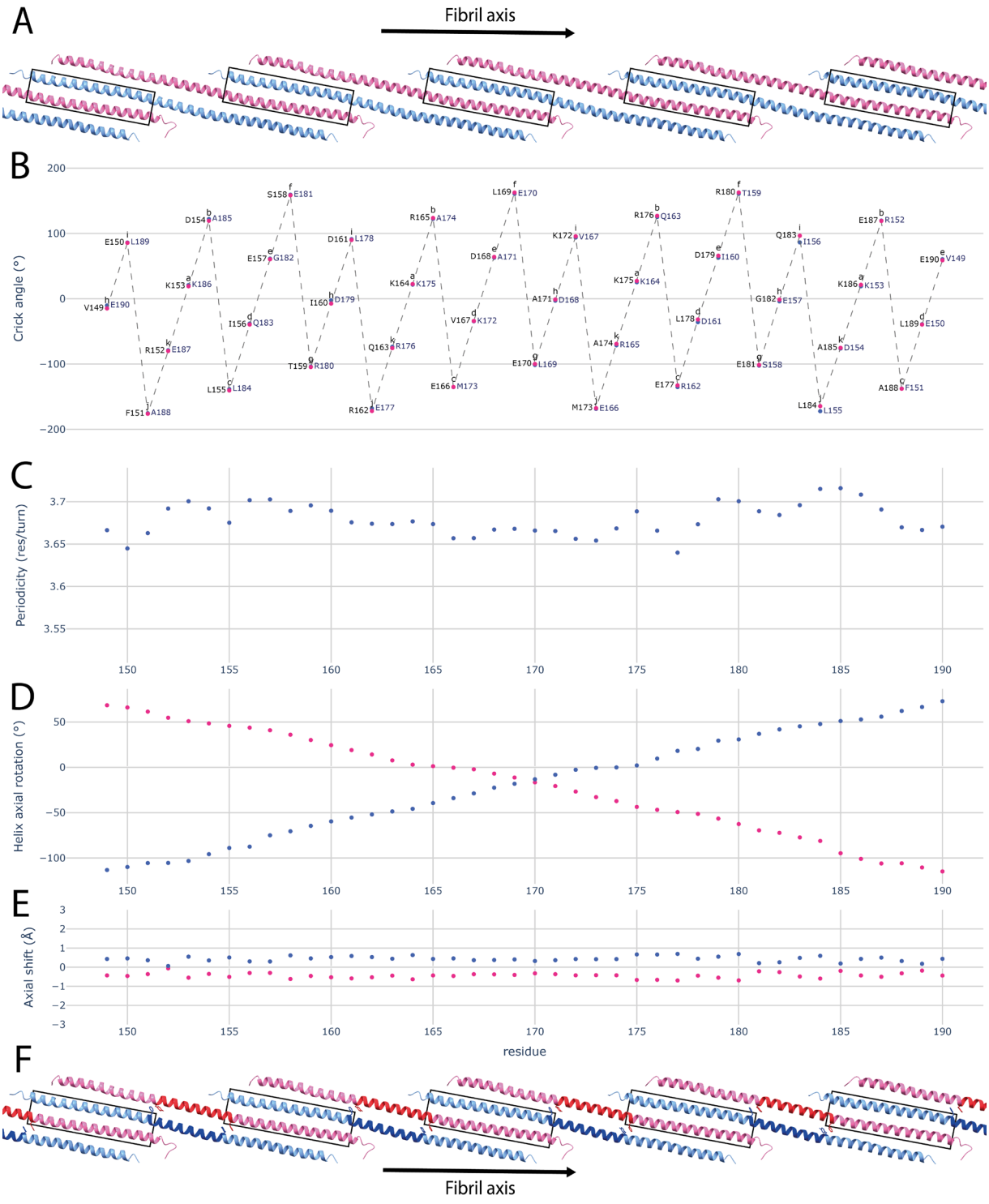


Figure 6.4: The coiled-coil dimers assemble with each other via a second antiparallel coiled-coil to form fibrils.

A: Crystal structure of a fibril composed of six coiled-coil dimer units. The FCC is indicated by a black rectangle. The A chains are in pink, the B chain in blue and the axis of the fibril is indicated by an arrow, with this colour being conserved throughout the figure.

B: Cricks angle of the FCC showing the different repeats. The residues in the A chains are indicated on the left of the pink dots and their position on top while the residues of the B chains are indicated on the right of the blue dots. As the two chains are mostly similar, the B dots may be hidden.

C: Per-residue periodicity of the FCC, showing the average of 3.67 residue per turn periodicity characteristics of hendecads

D: Axial rotations of the helices in the FCC, which corresponds to the deviation from the expected canonical Crick angles.

E: Axial shift of the FCC helices, revealing local deformation of the helices (cf. fig. 7).

F: Crystal structure of the fibrils with the remarkable central region of the dimerisation-driving coiled-coil (DCC) located between Y126-F151 coloured in red for chain A and blue for the B chain and the sidechains of these two aromatic residues are shown. The black rectangles delimit the FCC site (149-190) which almost perfectly follows it.

For **C**, **D** and **E** the residues numbering in the x axis correspond to the A chain (pink), but as the FCC is antiparallel, the facing residue of the B chain (blue) is represented and therefore follows an inverted axis. As the periodicity is the same for the A and B chains, the pink and blue markers have the same value in **C**.

Finally, it is noteworthy that the limits of the remarkable central region of the first DCC located between the two antiparallel π-π stackings of Y126-F151 is almost perfectly followed by the FCC site, which spans V149-E190, although it is located on the other face of the helix (fig. 4 F).

The visual inspection of the C-terminal region of the helical domain where the FCC takes place (fig. 5 A) shows that the core residues of the coiled-coil are indeed mostly hydrophobic as expected of a coiled-coil, but with the four lysine, K153, K164, K175 and K186, having their charged sidechains pointing outward instead toward the interface, thus avoiding the desolvation energetic penalty (fig. 5 B and C). In turn, these four lysines engage in salt-bridges with their counterparts, namely K153-E190, K164-D179, K175-D168 and K186-E157, thus stabilising the FCC with eight salt-bridges as the FCC is also antiparallel (fig. 5 D and E).

However, the accommodations of the orientation of the sidechains of these lysines towards the solvent despite being in the **a** interfacial positions result in deviations of the α-helices geometry which can be detected in their FCC axial shift. (fig. 5 F).

Indeed, to prevent a steric hindrance between the sidechains of K153 and of its salt-bridge partner E190 as well as for their involved groups to be in an optimal distance, the α-helix

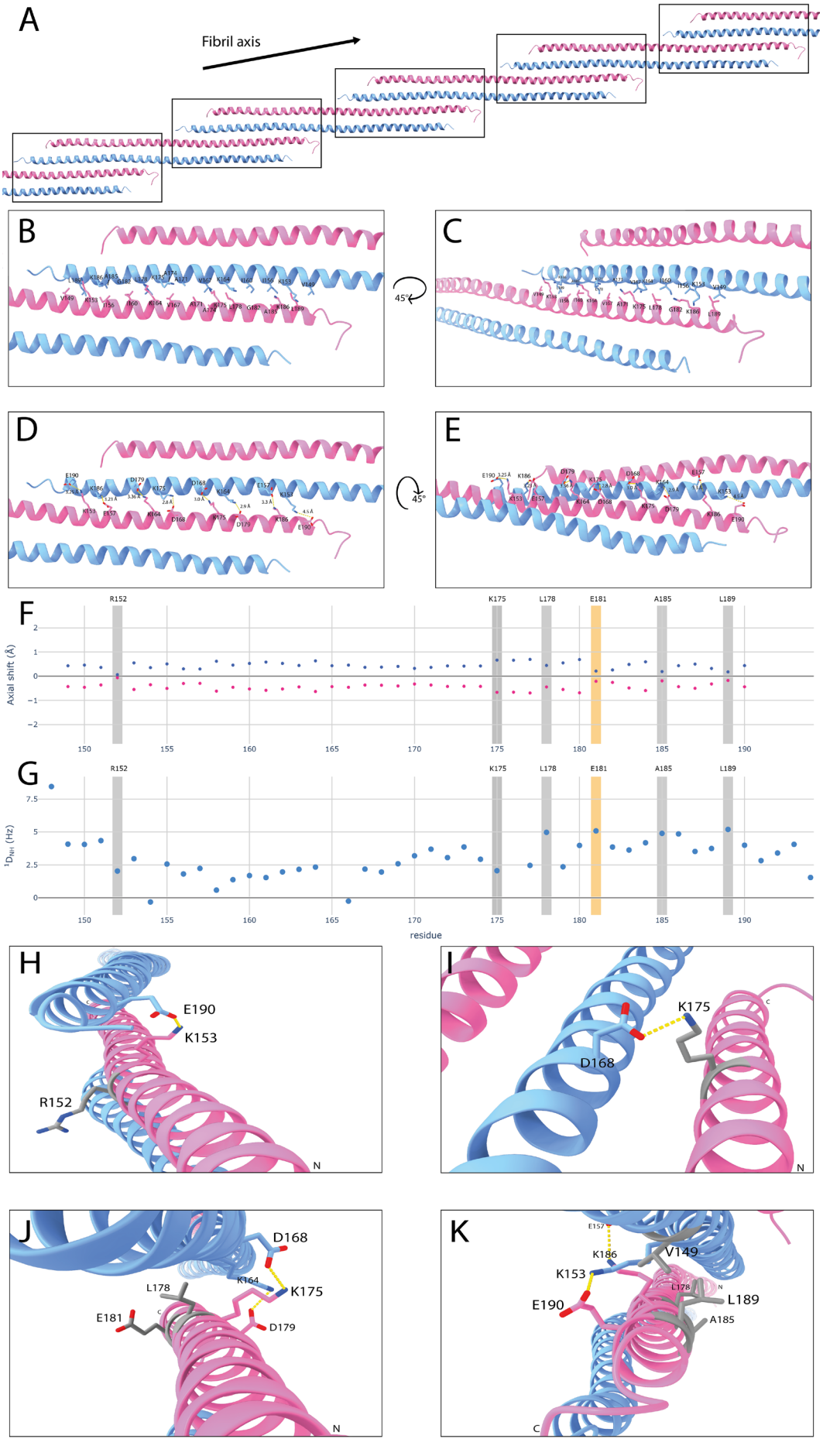

A
Fibril axis
B
C
45°
D
E
45°
F
Axial shift (Å)
R152
K175
L178
E181
A185
L189
G
$^1D_{NH}$ (Hz)
residue
H
E190
K153
R152
N
I
K175
D168
N
J
D168
K164
K175
L178
E181
D179
N
K
E157
K186
V149
K153
L178
N
L189
E190
A185
C

Figure 6.5: The fibrillation-driving coiled-coil is stabilised by salt-bridges mediated by four lysines in core positions.

A: Crystal structure of a fibril composed of six coiled-coil dimer units. The A chains are in pink, the B chain in blue and the axis of the fibril is indicated by an arrow, with this colour code being conserved in the whole figure. The black boxes correspond to the region highlighted in the next panels.

B and **C**: Views of the residues at the interface of the fibrillation-driving coiled-coil (FCC) showing the apolar residues at the hydrophobic interface. Additionally, A174 and A185 are also shown despite being in **k** position of hendecads as both are followed by lysines with sidechains pointing outside of the core and thus, the alanines participate to the apolarity of this side of the α-helices.

D and **E**: Views of the charged residues engaging in salt-bridges in the FCC, that is K153-E190, K164-D179, K175-D168 and K186-E157.

F: Axial shift of the FCC (149-190) showing signs of geometrical at R152, K175, L178, E181, A185 and L189. These residues are highlighted in grey for deviations and in orange for a curvature of the helix in the case of E181, which was described in the previous chapter.

G: Residual dipolar couplings of the backbone N-H bond vectors $^{1}D_{NH}$ of CAHS-8 at 293 K (blue).

H: View of the deviation induced by K153 to favour its salt-bridge with E190. The A chain (pink) presents an alteration of the direction of the α-helix at R152 (grey). The B chains are in blue, with the FCC partner in top and the DCC partner in the bottom. The N and C terminal ends of the A chain are indicated as well.

I: View of the K175-D168 salt-bridge. As K175 (grey) is in **a** position of the FCC hendecad, its polar sidechain points away from the core, triggering a slight deviation in the otherwise mostly straight 172-180 segment.

J: View of the deviations induced by the K175-D168 salt-bridge with L178 indicated in grey, as well as E181 where the helix starts bending.

K: View of the deviations at A185 (grey) to permit the K186-E157 salt-bridge. This deviation is corrected at the next helix turn position, L189 (grey), permitting the positioning of its sidechain toward V149 (grey). The FCC B chain relative to the studied A chain is on the top, while the bottom B chain is the DCC helix.

undergoes a lateral change of direction at R152 (fig. 5 H), hence the smaller axial shift at this position than in the neighbouring residues (fig. 5 F). A clear discontinuity is visible at R152 in the NMR backbone N-H bond vector Residual Dipolar Couplings (RDCs or $^{1}D_{NH}$, fig. 5 G) in solution. However, as R152 marks the beginning of the linker region in solution, this discontinuity is likely to indicate the end of helix 1 rather than the helix deviation visible in the crystal structure.

As per K175, it is located in the quite straight 172-180 segment of the crystal structure and induces a slight deviation of the helix direction due to it being in the **a** position of the repeat, permitting its sidechain to escape the hydrophobic interface between the two helices and to form a salt-bridge with D168 (fig. 5 F and I). The effect of this change of

direction can also be seen at R176 in the axial shift of the DCC as discussed in the previous chapter.

In turn, this deviation of the helix is corrected at L178. Indeed, as K175 induces a slight deviation as well as due to the B chain also being bent and L178 being in **d** position in a hendecad, a slight deviation at L178 is necessary for its apolar chain to face toward the interface with the other helix (fig. 5 F and J). This also results in a significant change in the axial shift of the DCC as previously discussed.

Moreover, as no significant exchange contribution could be seen in helix 2 from NMR relaxation or CPMG relaxation dispersion compared to the DCC, it is likely that the fibrillation and thus the FCC are insignificant at this concentration and temperature. This is in agreement with behaviour derived from DLS, at 350 µM and 293 K has the protein in the soluble region still were no important fibrillar assembly could be detected nor does the sample eventually gelate, unlike the slight fibrillation region.

Finally, in order to permit the K186-E157 salt-bridge, a slight deviation of the helix direction is present at A185 which is corrected at the next helix turn position, i.e. L189, allowing the positioning of its hydrophobic sidechain toward the facing FCC residue, V149 (fig. 5 F and K).

5.1.1.3 The fibrillation can be repressed by perturbating the FCC

One of the cysteine mutants presented in the previous chapter and used for paramagnetic relaxation enhancement experiments is A185C. As observed in the crystal structure, A185 is involved in the FCC (fig. 4 B and fig. 5 K) but not in the DCC, making its mutation to a polar cysteine residue optimal to perturb the fibrillar assemblies but not the dimerisation.

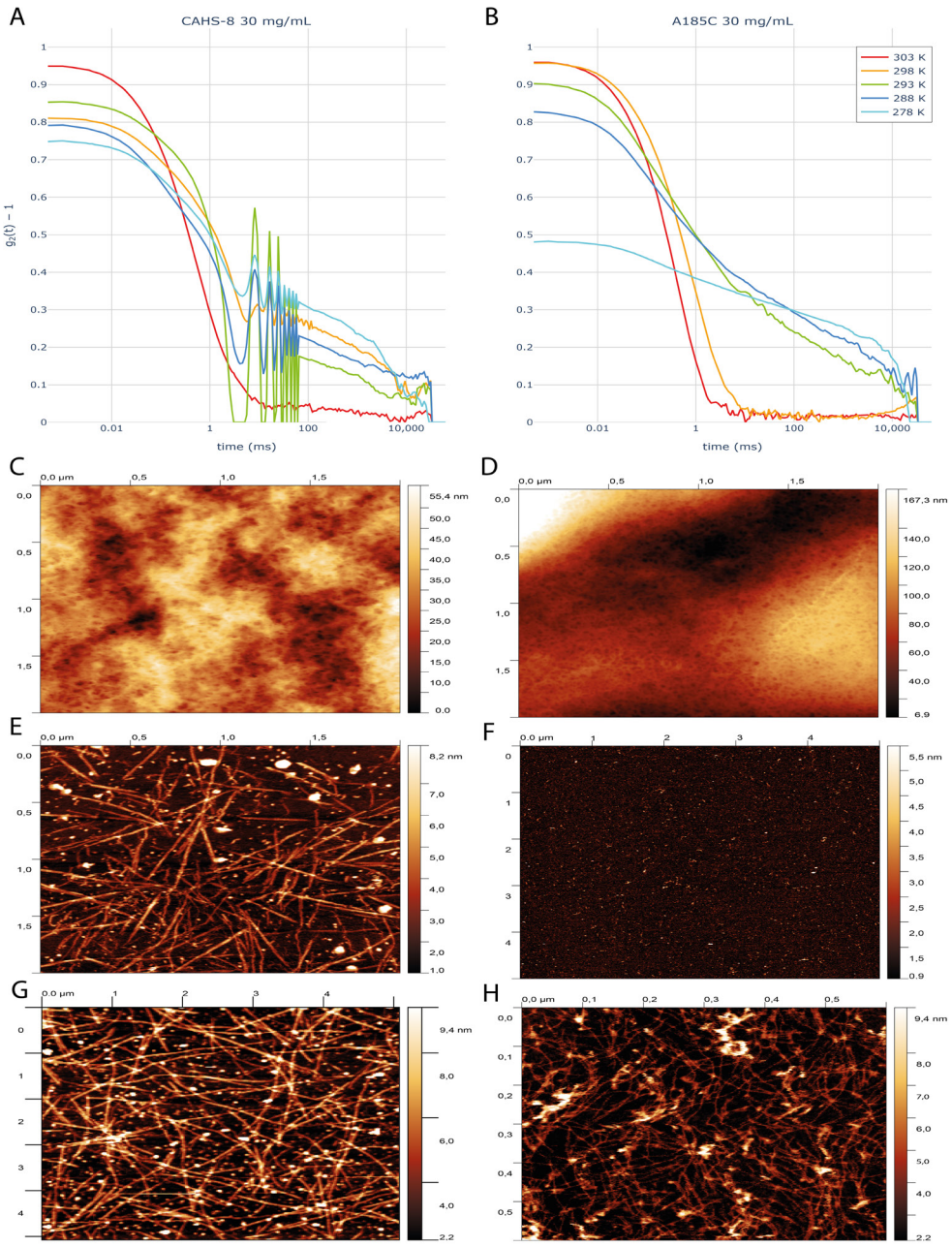


Figure 6.6: A single mutation in the fibrillation-driving coiled-coil is enough to repress the fibrillation.

A and **B**: Temperature-dependent normalised Dynamic Light Scattering intensity autocorrelation function of CAHS-8 (**A**) and A185C (**B**) at 30 mg/mL. CAHS-8 gelates between the 303 and 298 K points while A185C does between 298 and 293 K. The spikes associated with the scattering of fibres are not detected in A185C.

C and **D**: AFM imaging in air of desiccated drops at 30 mg/mL of CAHS-8 (**C**) and A185C (**D**), revealing a reticular network, i.e. a dried gel. **E** and **F**: AFM imagining in liquid mode of the resuspension of such desiccated drops, revealing numerous long fibres and fibres in CAHS-8 (**E**) but only a couple of small ones are visible for A185C (**F**), hence the bigger field of view to see some.

G and **H**: Letting these previous sample dry results in a significant number of very long fibres of CAHS-8 (**G**) but only smaller fibres for A185C (**H**), hence the much smaller field of view.

As discussed in the chapter 3, CAHS-8 forms fibres and a gel, both of which could be detected by looking at the decay and pattern of the Dynamic Light Scattering (DLS) normalised intensity autocorrelation function $g^{(2)}(t) - 1$. Indeed, fibres manifests by important spikes while the gelation manifests as an important drop of the initial autocorrelation value, which is explained in more details in the chapter 7. At 30 mg/mL, i.e. 1.14 mM, CAHS-8 forms a gel at a temperature between 303 and 298 K (fig. 6 A) and fibres can be clearly detected. However, as per the same temperature-dependent DLS measurement, the A185C mutant at the same concentration appears to start a gel only at a point between 298 and 293 K (fig. 6 B). Additionally, at 278 K, the initial $g^{(2)}(t) - 1$ value is much lower than any measure of CAHS-8, indicating a higher degree of non-ergodicity of the A185C gel. Finally, the characteristic spikes in the autocorrelation function resulting from the scattering of long fibres are only visible in CAHS-8 and could not be detected at any temperature in A185C.

The desiccation of drops at 30 mg/mL results in a dry gel in both CAHS-8 and A185C (fig. 6 C and D). However, resuspending these dry gels in buffer reveals a significant number of long fibres in CAHS-8 (fig. 6 E) but only a low number of short ones in A185C (fig. 6 F). In turn, if these samples are let to dry again at room-temperature, the polymerisation of CAHS-8 into fibres continues and these fibres gets longer and more numerous (fig. 6 G), while for A185C, only much shorter fibrils are visible (fig. 6 H). The CAHS-8 fibres have an average apparent width of 20-27 nm, which corresponds to 9.5 nm in NS-TEM as discussed earlier, and a depth of 2.8 nm while the A185C fibrillar assemblies are approximately 8 nm wide but 1.5 to 3 nm deep, which most likely correspond to the individual fibrils with thus a real width of 2 to 4 nm.

Altogether, these results indicate that the mutation of A185 into a cysteine does indeed destabilise the fibrillation of CAHS-8. Indeed, its gelation temperature appears to have been shifted to several degrees below that of CAHS-8 while A185C does not form fibres but only short fibrils. It is likely that this mutation causes the FCC to be less stable, possibly resulting in the fibrils constantly assembling and disassembling at a higher rate than in CAHS-8, thereby preventing them from polymerising to long fibres. This instability is visible in their poorly defined contour in the AFM image (fig. 6 H). In turn, the 1.5 nm depth could correspond to disassembling fibrils with the helices losing the deformation

responsible for their 2.9 nm width and necessaries for the FCC formation or simply to an AFM setpoint too high for individual fibrils.

5.1.2 Role of the disordered termini in the fibre assembly

The crystal structure of the dimerisation-driving coiled-coil and of the fibrils explains their formation, while AFM and EM evidences shows that fibres are composed of two fibrils. However, the structure does not explain how these two fibrils form fibres together nor why the HD construct introduced in the previous chapter has a significantly decreased ability to fibrillate. This suggests, especially for the latter, a role for the intrinsically disordered chains in N-terminal (1-94) and C-terminal (195-233) that flank the central helical domain of CAHS-8 (95-194).

5.1.2.1 Model-free analysis of the disordered termini

The comparison of the relaxation of the disordered termini of CAHS-8 at 278, 293 and 303 K as well as the model-free analysis treatment of the 278 and 293 K rate permit to obtain insights of their role, especially as unlike the helical domain, the disordered chains are visible at 278 K. Performing the model-free analysis using 7 rates (R_1, R_2 and nOe at 600 and 850 MHz, as well as η_{xy} at one field) to fit 6 parameters (A_1, A_2, τ_2, τ_3, R_{ex} and θ) is not expected to yield results of utmost precision, but nevertheless represent an estimation of theses parameters as well as a useful way to detect the presence of R_{ex}. For this reason, model-free analysis at 303 K was not included as only 6 rates were measured. The relaxation results of the N-terminal disordered chain (fig. 7 A-F) show a behaviour similar to that of other IDPs in function of the temperature, with an increased rigidity and a general increase of R_1 and R_2 at lower temperature[603]. This, alongside, the absence of regions presenting an elevated R_2, indicate that the N-terminal disorder chain remains disordered at 278 K and does not appear to engage in interactions. In turn, the model-free analysis of the N-terminal relaxation (fig. 7 G-L) show that while no significant exchange contribution to the transverse relaxation R_{ex} is detected at 293 K, an R_{ex} of approximately 0.5 to 1.5 s^{-1} is observed throughout the chain at 278 K in particular in the segments 35-43 and 54-79 (fig. 7 K).

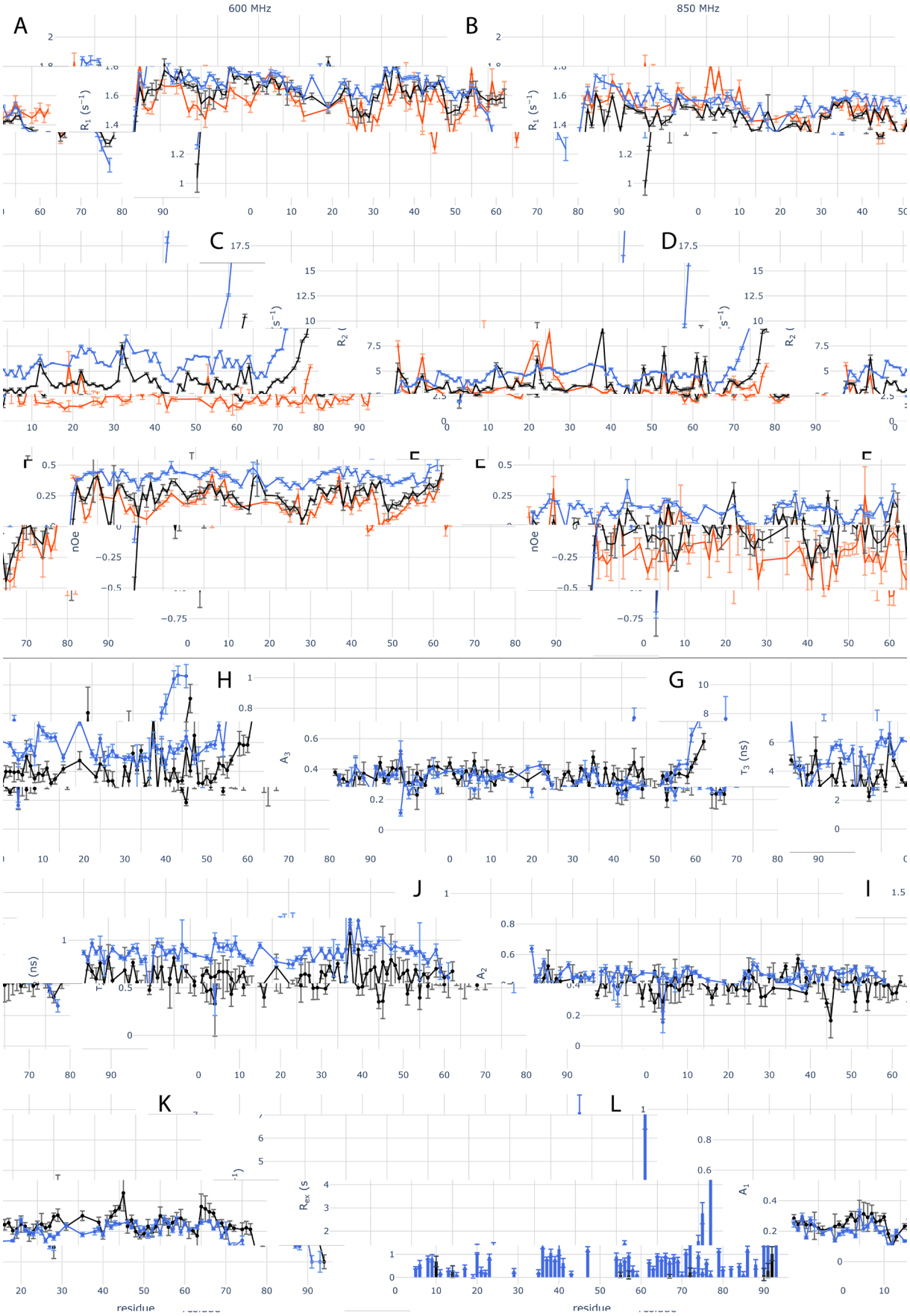

600 MHz
850 MHz
A
B
C
D
E
F
G
H
I
J
K
L
residue

Figure 6.7: Relaxation of the N-terminal disordered chain of CAHS-8 and model-free analysis.

A-**F**: Longitudinal relaxation rate R_1 (**A**, **B**), transverse relaxation rate R_2 (**C**, **D**) and heteronuclear $^{15}\mathrm{N} - \{^1\mathrm{H}\}$ nOe relaxation rate (**E**, **F**) at 600 MHz (**A**, **C**, **E**) and 850 MHz (**B, D, F**) for CAHS-8 at 278 K (blue), 293 K (black) and 303 K (orange).

G-**L**: Model-free analysis of the relaxation of CAHS-8 at 278 K (blue) and 293 K (black): Correlation time of the slow motional mode τ_3 (**G**) and its associated amplitude A_3 (**H**), correlation time of the intermediate motional mode τ_2 (**I**) and its associated amplitude A_2 (**J**), chemical exchange contribution to the transverse relaxation R_{ex} (**K**) and amplitude of the fast motional mode A_1 (**L**, with its correlation time τ_1 being fixed to 50 ps[1,2]).

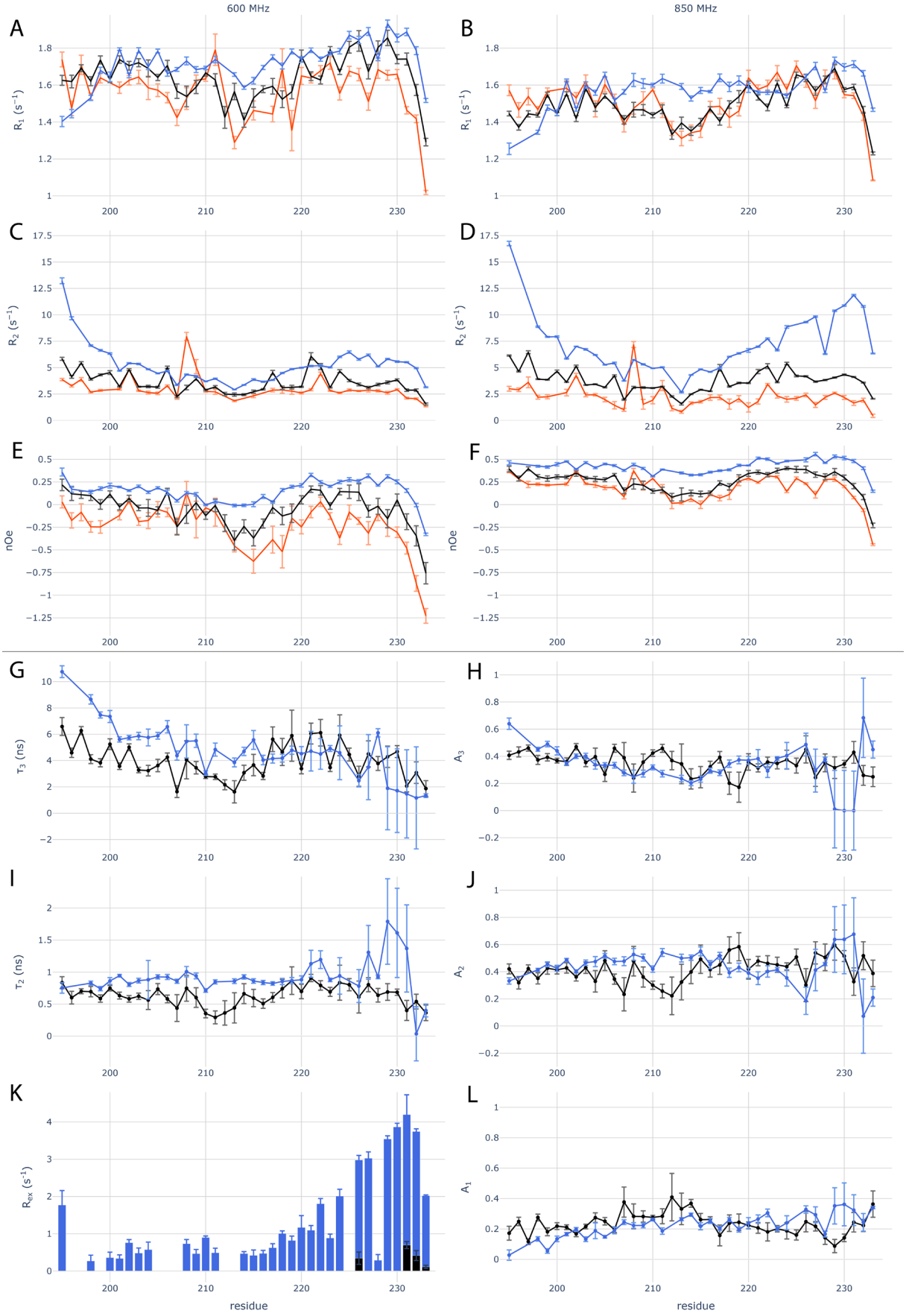
Moreover, as the temperature decreases, the correlation times of the slow and intermediate motions τ_3 and τ_2 (fig. 7 G, I) increase as expected due do the increase of the solvent-mediated nano-viscosity they experience, as well as due to the influence of the helical domain in C-terminal[603]. However, the values of the motional amplitude of the segmental motions A_3 (fig. 7 H) at 278 K closely reproduce the 293 K profile up to residue 87 after which they increase noticeably higher than at 293 K.

This effectively indicate despite the viscosity-modulated increase of the slow-motion correlation time τ_3, a maximal amplitude of the chain motions is already reached at 293 K and does not further increase at 278 K up to residue 87. In turn, because the slow and intermediate motional timescales τ_3 and τ_2 increase between 293 and 278 K but A_3 does not, this translates into an effective increase of the amplitude of the local backbone sampling motions A_2 between 278 and 293 K instead of their expected decrease[603].

Figure 6.8: Relaxation of the C-terminal disordered chain of CAHS-8 and model-free analysis.

A-**F**: Longitudinal relaxation rate R_1 (**A**, **B**), transverse relaxation rate R_2 (**C**, **D**) and heteronuclear $^{15}\mathrm{N} - \{^1\mathrm{H}\}$ nOe relaxation rate (**E**, **F**) at 600 MHz (**A**, **C**, **E**) and 850 MHz (**B, D, F**) for CAHS-8 at 278 K (blue), 293 K (black) and 303 K (orange).

G-**L**: Model-free analysis of the relaxation of CAHS-8 at 278 K (blue) and 293 K (black): Correlation time of the slow motional mode τ_3 (**G**) and its associated amplitude A_3 (**H**), correlation time of the intermediate motional mode τ_2 (**I**) and its associated amplitude A_2 (**J**), chemical exchange contribution to the transverse relaxation R_{ex} (**K**) and amplitude of the fast motional mode A_1 (**L**, with its correlation time τ_1 being fixed to 50 ps [1,2]).


600 MHz
850 MHz
A
B
C
D
E
F
G
H
I
J
K
L
R1 (s−1)
R2 (s−1)
nOe
τ3 (ns)
A3
τ2 (ns)
A2
Rex (s−1)
A1
residue
residue

The relaxation of the C-terminal disordered region (fig. 8) shows a clear increase of the R_2 rates at 278 K in the very terminal residues compared to the neighbouring ones which is contrary to what is typically seen at the end of disordered chains[439,603,604] and which this phenomenon being more marked at 850 MHz compared to 600 MHz (fig. 8 C, D). This indicates a chemical exchange contribution, possibly due to an interaction or to conformational exchange. This is supported by model-free analysis applied to these C-terminal relaxation rates which shows the transverse exchange contribution R_{ex} consistently increases along the sequence from residue 217 to 231, reaching more than 2 s^{-1} in the 224-233 region (fig. 8 K). Moreover, like in the N-terminal chain, the parameters also appear to shows a conspicuous behaviour at 278 K. Indeed, while the 195-216 region has slower correlations times for the segmental and backbone motions τ_3 and τ_2 at 278 K than at 293 K as expected (fig. 8 G, I), they have a respectively decreasing and increase associated amplitudes A_3 and A_2 (fig.8 H, J), with the former being smaller at 278 K than at 293 K for the 201-216 segments and the latter being higher from V198 already. In turn, this region also has a modest but consistent R_{ex} contribution of approximately 0.5 s^{-1} (fig. 8 K). As per the 217-233 segment of the chain, it follows a more expected behaviours at 278 K with apparent deviations, especially in the last few terminal residues, that can be attributed to the large error bars.

Altogether, the behaviour of the 195-216 region may be possibly be explained by a motional restriction at 278K, hence the smaller chain-like motional amplitudes A_3 than at 293 K, while the 217-233 terminal region shows a more classic disordered chain behaviour. The presence of a significant exchange in the last residues is surprising but as he 227-233 residues are exogeneous amino-acids resulting from a TEV-cleavage site put in C-terminal in the original plasmid, this behaviour is probably not physiologically relevant.

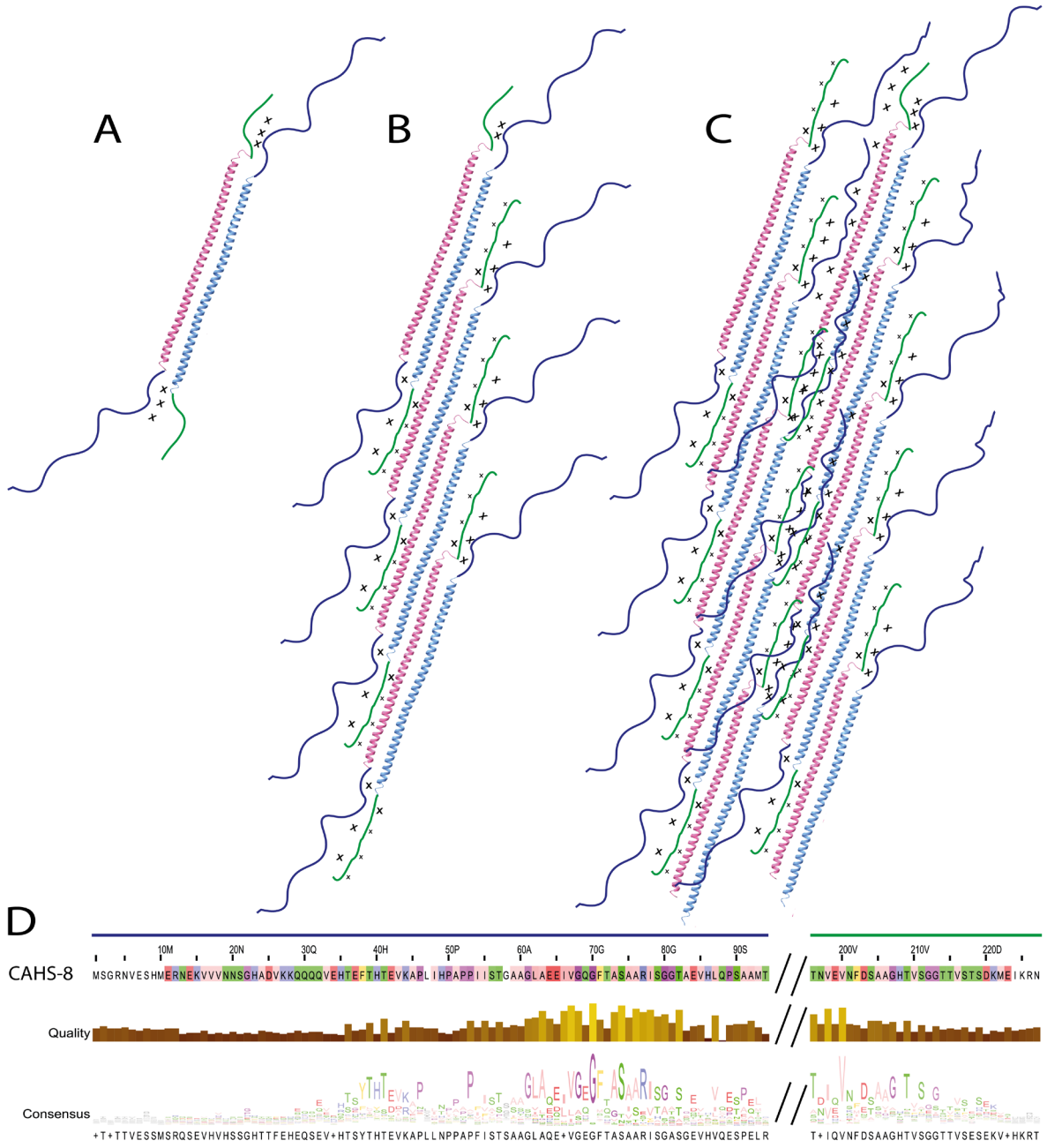


Figure 6.9: The different assemblies of CAHS-8 result in physical restrictions of the disordered termini which show various level of conservation.

A: The formation of the antiparallel dimerisation-driving coiled-coil result in the N-terminal disordered chain (dark blue) of a monomer and the C-terminal chain (green) of the other limiting each other lateral motions, with the hindrance being marked by a cross. The termini are not to the scale.

B: The fibrils of CAHS-8 result from the formation of the fibrillation-driving coiled-coil (FCC) which results in the C-terminal disordered chain being also physically restricted on the other side by the part of the helical domain of its partner not engaged in the FCC.

C: The fibres of CAHS-8 result in the additional mutual physical restrictions of the disordered chains of the two fibrils in the space between them.

D: Sequence conservation of the disordered terminal chain resulting from the sequence alignment presented in the previous chapter. The helical domain is hidden and denoted by the two slashes.

The analysis of the fibril crystal structure shows that the helical domain of CAHS-8 dimerises via an antiparallel coiled-coil, the DCC, and polymerise to form fibrils via a second antiparallel coiled-coil of the dimeric coiled-coils, the FCC, with these mechanisms being favoured at 278 K compared to 293 K and 303 K. As per the phase diagram of CAHS-8, the protein at 350 μM and 303 or 293 K is in the soluble region, while at 278 K, it is in the slight gelation region which is characterised by a significant number of fibres being detectable by DLS, but not enough for the whole sample to gelate during the course of the measurements. Thus, this indicate that the DCC and FCC are significantly more present at 278 K as they are required for the fibre formation.

The crystal structure may suggest a putative physical restriction and intermolecular contacts. First, in the case of the dimerisation via the DCC, as two CAHS-8 proteins form an antiparallel coiled-coil, their disordered chains may encounter some physical limitations due to the presence of the other, i.e. the N-terminal IDR of one protein may be restricted by the C-terminal IDR of the other protein and so on (fig. 9 A). However, in the case of the helices polymerising via the FCC to form fibrils, which takes place at roughly half of the helical-domain length, the C-terminal disordered chain of each helix is directly restricted in its lateral motions by the presence of the remaining half of the other FCC helix on one side and by the presence of the N-terminal chain of the DCC helix on the other side (fig. 9 B).

Moreover, additionally to these two restrictions, in the case of the presence of another fibril chain to form a fibre, the N-terminal chain of one of DCC helix of each fibril would be projected towards the other fibril, thus physically restricting each other while being restricted on the one remaining empty side by the presence of helices of the other fibrils (fig. 9 C).

The model-free analysis shows a detectable exchange contribution to the transverse relaxation R_{ex} at 278 K in the regions 35-41 and 54-84 of the N-terminal disordered chain (fig. 7, bottom left) while in turn, the C-terminal flexible chain shows small but consistent exchange contribution in 198-216 region (fig. 8, bottom left) as well as the residues in the vicinity of the helical domain. The analysis of the sequence conservation of the disordered termini shows a significant sequence conservation in the N-terminal regions 35-45 and

60-80, as well as 198-213 for the C-terminal (fig. 9 D). This suggest that the termini engage in interactions to stabilise, or destabilise, the different fibrillar assemblies.

5.1.2.2 One-armed constructs

In order to understand to understand the role of the two flexible terminal, two more constructs were made comprising either the N-terminal disordered chain and the helical domain (1-198, hereafter called the NTDHD construct) or the helical domain and the C-terminal disordered chain (88-227, hereafter called the HDCTD construct). The two constructs have had their TEV-cleavage site put in N-terminal and as such, do not carry exogeneous leftover residues in C-terminal, unlike CAHS-8.

Thanks to their very similar ^{15}N-^{1}H HSQC to HD and to CAHS-8, neither the NTDHD nor the HDCTD constructs were assigned but the already known assignment was transferred and both constructs were assumed to still be helical.

Comparison of the relaxation rates at 600 MHz at 303 K of the full-length CAHS-8 (black) with the constructs NTDHD (blue), HD (red) and HDCTD (green) show multiple effects (fig. 10).

First, at 303 K, the N-terminal chain of the NTDHD construct behave slightly differently than CAHS-8's. Indeed, its longitudinal and transverse autorelaxation rates R_1 and R_2 as well as nOe rates are systematically smaller than CAHS-8's indicating an increased flexibility as well as faster correlation times. Due to the presence of a flexible linker, especially at this temperature, between the two helices and indicated in orange, the global motions of helix 1 should be reproduced, which is not the case either. As the R_2R_1 product is also consistently smaller than in CAHS-8, this likely indicates a reduced exchange process in the first helix, i.e. less dimerisation-driving coiled-coil (DCC) at 303 K.

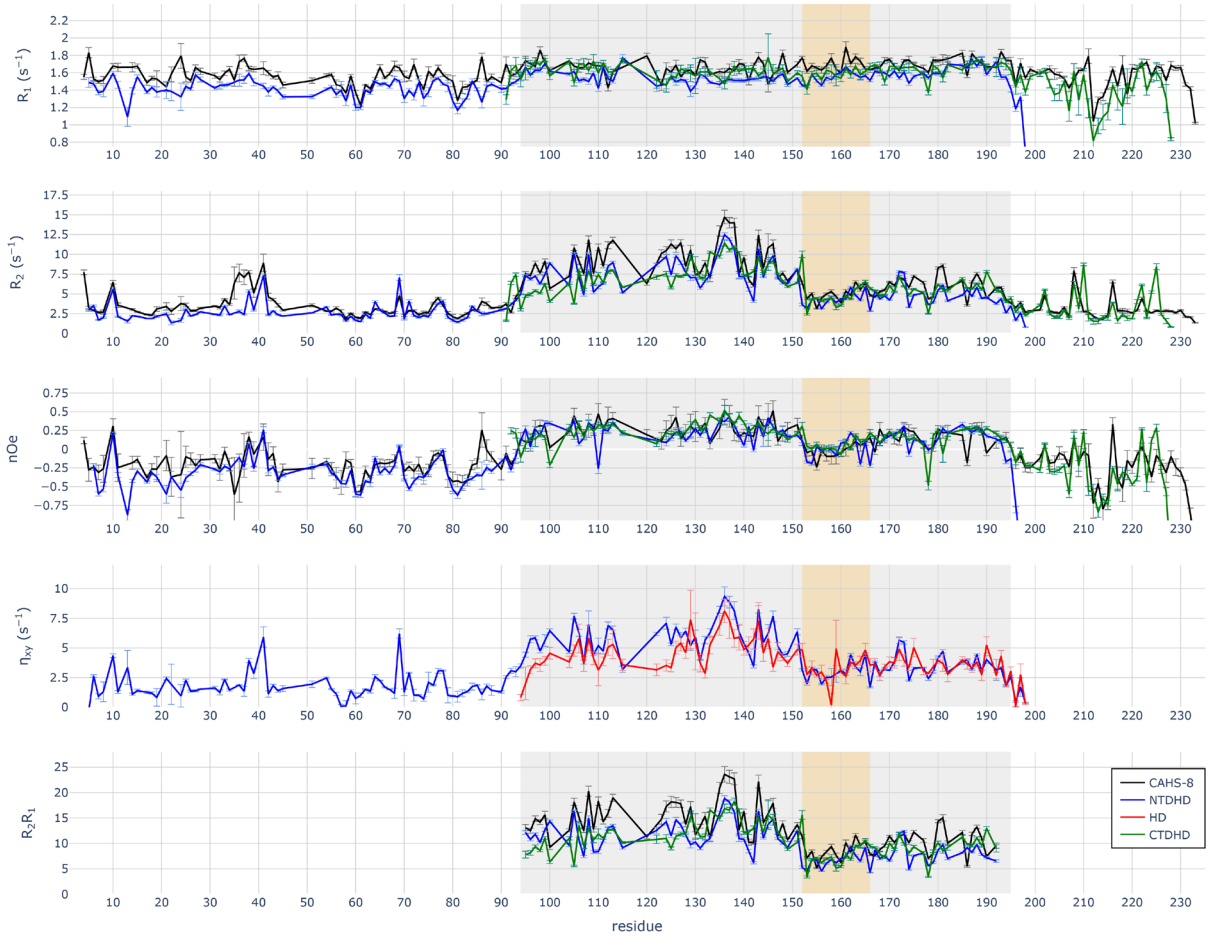


Figure 6.10: Comparison of the relaxation of the different constructs of CAHS-8 at 600 MHz, 350 µM and 303 K.

From top to bottom: Backbone amide longitudinal autorelaxation R_1, transversal autorelaxation R_2, {^{1}H}-^{15}N heteronuclear nOe relaxation, cross-correlated transverse relaxation η_{xy} as well as the R_2R_1 production for the wild-type CAHS-8 (black), the NTDHD construct (blue), the HD construct (red) and the CTDHD construct (green).

The helical domain (95-194) is indicated by the grey shading and the linker (152-166) between the two helices is highlighted in orange.

In turn, the HDCTD relaxation at 303 K closely reproduces the longitudinal and nOe relaxation of CAHS-8 over the whole helical region as well as the C-terminal IDR, with smaller R_1 rates around G213. The difference in the R_2 rates in helix 1 is expected as the absence of the long N-terminal disordered chain reduces the rotational diffusion time of the helix, hence the almost identical helix1 R_2 between HDCTD and HD in helix 1[603]. In turn, the behaviour of helix 2 is similar to that of CAHS-8 due to the presence of the C-terminal disordered chain.

At 293 K (fig. 11), as per NTDHD, the significantly slower R_1 rates in the disordered chain but marginally slower R_2 compared to CAHS-8 are surprising compared to similar studies of other IDPs comprising helical elements, where the peripheral disordered domains show similar changes between different length constructs[603]. As per the HDCTD construct, the C-terminal disordered chain shows similar autorelaxation rates constants to CAHS-8 but appears slightly more flexible around residue V210. The F202 residue shows a higher R_2 rate constant, a slower R_1 rate and appears more rigid in nOe in the HDCTD than in CAHS-8.

In turn, helix 1 in the NTDHD construct shows elevated R_2 and lower R_1 values than CAHS-8 but a similar R_2R_1 product. In turn, its helix 2 behaves similarly to HD. The origin of this anomaly is not clear and will require the measurement of a larger range of relaxation rates at different magnetic field strengths, as well as a detailed analysis of the dependence on sample variation.

At 278 K (fig. 12), as stated in the previous chapter, HD eventually aggregates, turning the sample white and its relaxation could not be measured as it couldn't be considered to be in a steady-state. However, the measured T_1 and T_2 decay of the HDCTD construct at 278 K present a clear bi-exponential profile and therefore the rates could not be extracted. This bi-exponential profile likely results from separate population with different solvent exchange rate and therefore accessibility, which strongly supports the HDCTD construct being in a stable oligomeric state at 278 K. This bi-exponential profile of the HDTCD construct suggest a slow exchange process. Unfortunately, the signal to noise ratio CPMG relaxation dispersion was too low to distinguish a dispersion while no cross-peaks could be distinguished in ZZ-exchange over a wide range of mixing times. Nevertheless, two populations would be in qualitative agreement with the DLS autocorrelation function which clearly shows at least two well separated and populated species, although the measurements were performed at 1.5 mM while NMR is at 350 µM (fig. 13 C).

This is not the case of the NTDHD construct whose helical domain could be observed unlike CAHS-8's, while their N-terminal disordered chain presents very similar R_2 as well

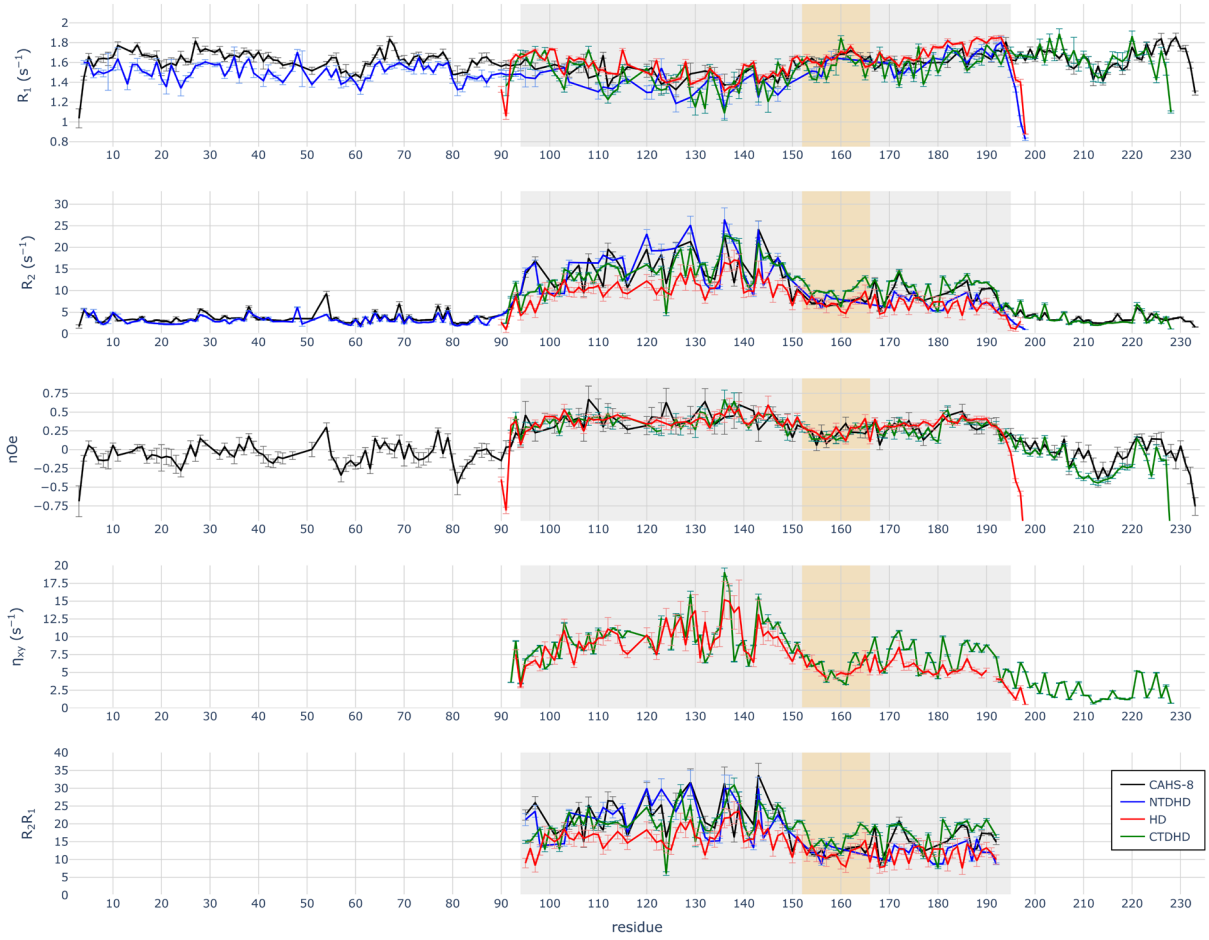


Figure 6.11: Comparison of the relaxation of the different constructs of CAHS-8 at 600 MHz, 350 µM and 293 K.

From top to bottom: Backbone amide longitudinal autorelaxation R_1, transversal autorelaxation R_2, {^{1}H}-^{15}N heteronuclear nOe relaxation, cross-correlated transverse relaxation η_{xy} as well as the R_2R_1 production for the wild-type CAHS-8 (black), the NTDHD construct (blue), the HD construct (red) and the CTDHD construct (green).

The helical domain (95-194) is indicated by the grey shading and the linker (152-166) between the two helices is highlighted in orange.

similar, a slightly lower nOe rates but a consistently smaller R_1 profile. As per the helical domain, it has a similar R_1 values as the few peaks available in CAHS-8 but much smaller R_2 rate constants, being approximately 20 s^{-1} on average while most of CAHS-8's are broadened beyond detection.

Altogether, these observations allow to identify the role of the different domains of CAHS-8. The helical domain self-associates into the DCC. However, due to the defects in the heptad repeats, the association is transient at room-temperature but is increasingly favoured as the temperature decreases.

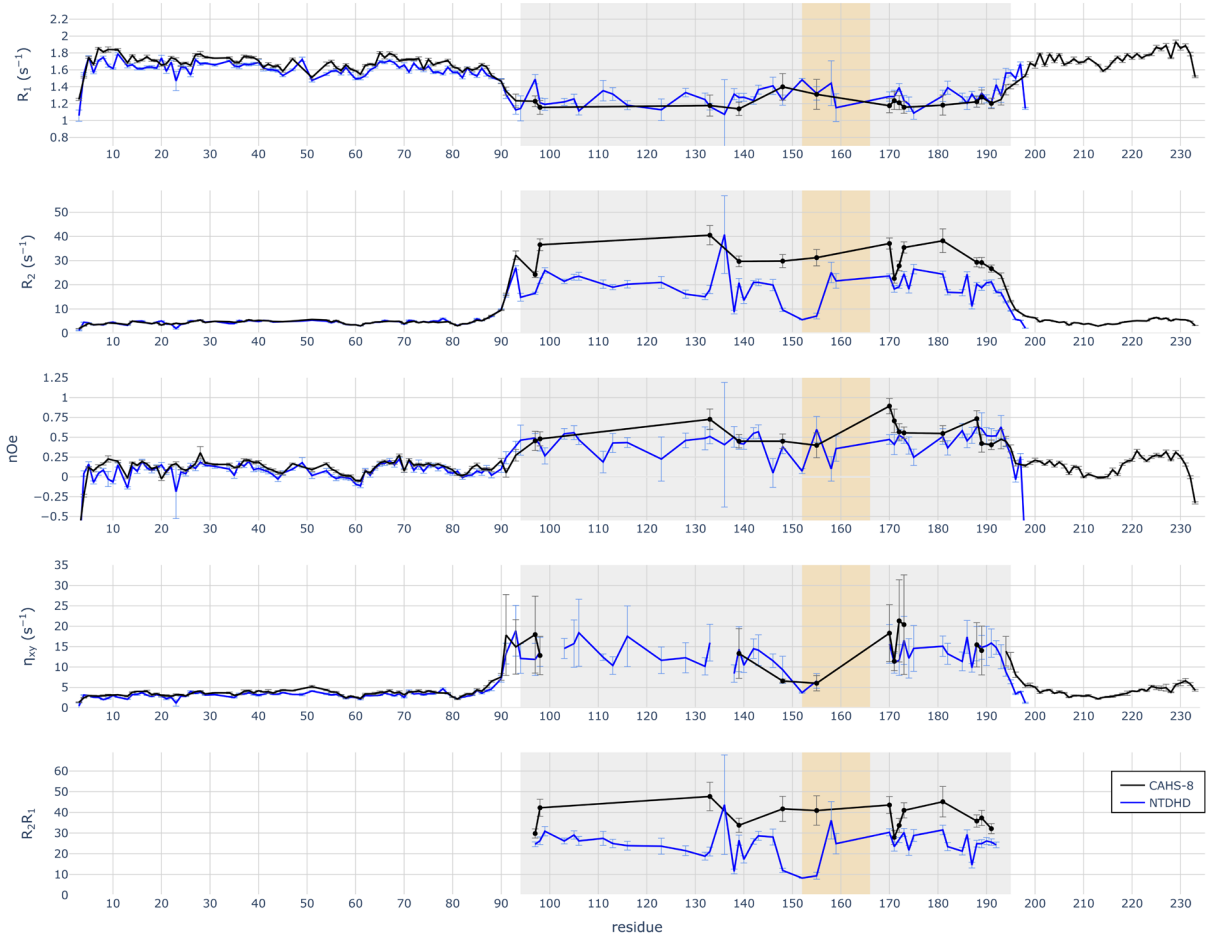


Figure 6.12: Comparison of the relaxation of the different constructs of CAHS-8 at 600 MHz, 350 µM and 278 K.

From top to bottom: Backbone amide longitudinal autorelaxation R_1, transversal autorelaxation R_2, {^{1}H}-^{15}N heteronuclear nOe relaxation, cross-correlated transverse relaxation η_{xy} as well as the R_2R_1 production for the wild-type CAHS-8 (black), the NTDHD construct (blue), the HD construct (red) and the CTDHD construct (green).

The helical domain (95-194) is indicated by the grey shading and the linker (152-166) between the two helices is highlighted in orange.

At 278 K and 350 µM, the helical region alone of CAHS-8, i.e. the HD construct, is inferred to chaotically self-associate which results in aggregates as discussed in the previous chapter. However, when the C-terminus is present, as in the HDCTD construct, this

association may be better guided, resulting in stable oligomeric population but no aggregates, nor fibrils or gels at this concentration. The increase in transverse relaxation rates at the C-terminus of the protein suggests that the C-terminal disordered chain is involved in an interaction at this temperature, possibly with the helices and in particular, as drawn in fig. 9 B, with the N-terminal half of the helical domain of the FCC partner.

As mentioned in the introduction of this part, the specificities of coiled-coil assemblies come from the charged residues flanking the apolar interface, as their attraction or repulsion determine which protein, face, or site interacts with which. As such, in the case of the FCC, it appears probable that the specificity of this coiled-coil is also determined by the C-terminal disordered chain that guides proper assembly. Indeed, in such mechanism, HD, which does not have this disordered chain, may interact on the FCC face with parts of the DCC phase and vice versa, forming chaotic unspecific coiled-coils aggregates at 278 K. Direct evidences for this are so far lacking but it is supported by the fact that, as discussed in the previous chapter, deuterating it appear to repress the FCC but not the DCC, resulting in the deuterated HD not forming aggregates at 278 K.

In turn, when only the N-terminal disordered chain is present alongside the helical region, i.e. the NTDHD construct, oligomerisation still appears to happen but with a lower efficiency than any other construct. It is therefore possible that the NTDHD constructs retains the ability to forms the DCC but represses the formation of the FCC and therefore, fibrillation. Indeed, the helical domain of the NTDHD is visible by NMR at 278 K which may indicate a lower oligomeric state or substantial drag effects due to the disordered tails. However, it does not form a great number of chaotic aggregates like HD but, missing the C-terminal IDR, NTDHD aggregates more than CAHS-8 and HDCTD while forming small fibrils in the process (fig. 13 E vs D and F). In turn, as per the DLS autocorrelation function, NTDHD appears not to form fibres and appears to be barely able to gelate at 278 K, while CAHS-8 does it at a temperature between 303 and 298 K (in fig. 13 B). Thus, as the FCC is apparently repressed but not inhibited, the formation of fibrils and fibres would still be possible, although reduced compared to CAHS-8 and HDCTD. Finally, as the HDCTD shows a stable population of fibrils but not CAHS-8, this supports an effect of the N-terminal disordered chain on the FCC, causing its disassembly.

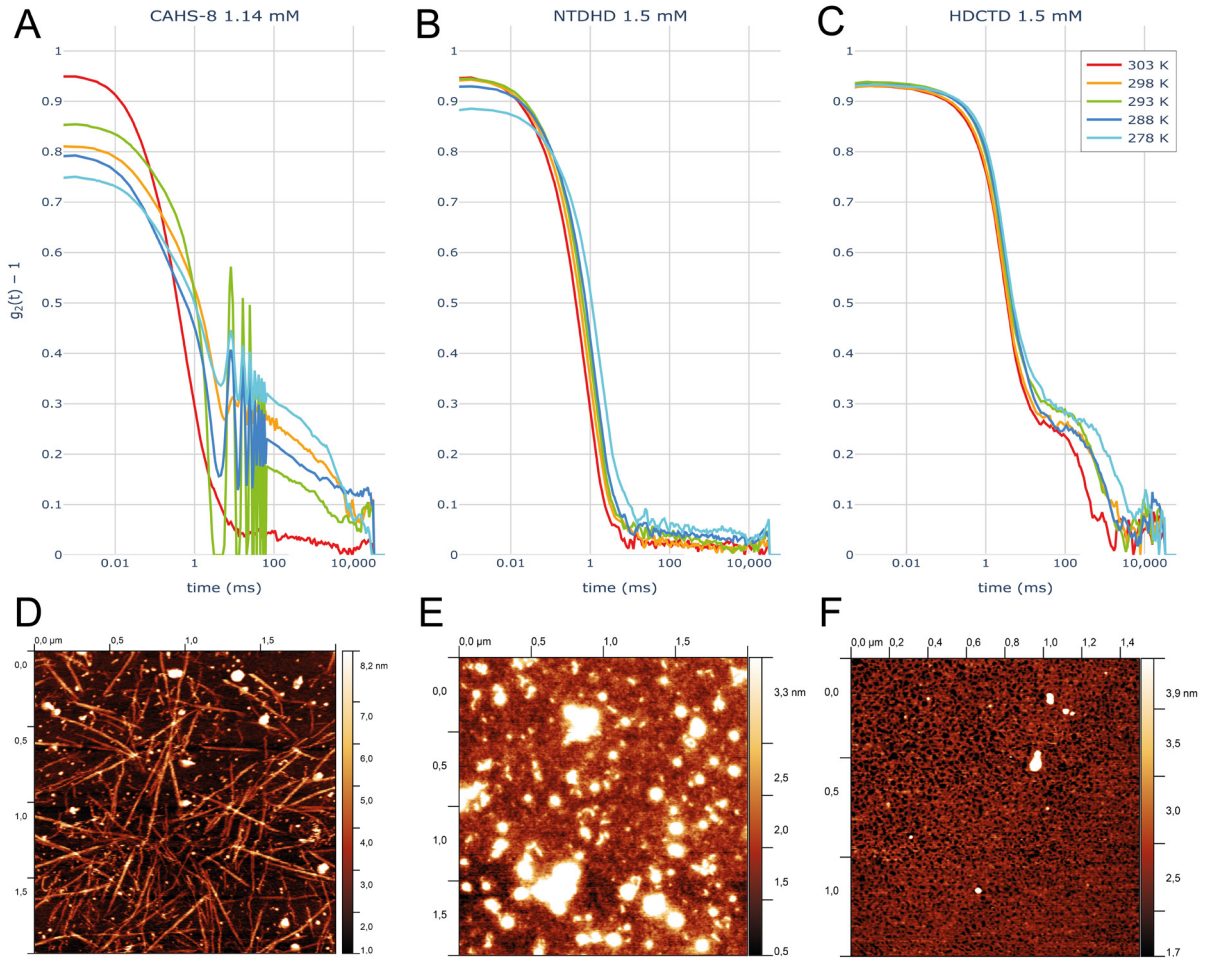


Figure 6.13: The disordered termini are important for the gel formation and the morphology of the fibrillar assemblies of CAHS-8.

A and **D**: Temperature-dependent normalised Dynamic Light Scattering (DLS) intensity autocorrelation function of CAHS-8 at 1.14 mM (**A**, corresponding to 30 mg/mL) and AFM imagining in liquid mode of a drying sample of CAHS-8 (**D**). The DLS data shows the presence of fibres and the formation of a gel.

B and **E**: Similar DLS intensity correlation function of the NTDHD construct at 1.5 mM (**B**) and AFM imagining in liquid mode of a drying sample of NTDHD (**E**). The DLS data does not detect the presence of fibres nor a significant number of fibres, but a slight gelation at 278 K is visible due to the decrease of the initial autocorrelation value. In turn, AFM imaging reveals numerous aggregates in the foreground (clearer species) but also fibrils in the background when the sample is let to dry.

C and **F**: Similar DLS intensity correlation function of the HDCTD construct at 1.5 mM (**C**) and AFM imagining in liquid mode of a drying sample of HDCTD (**F**). The DLS autocorrelation function does not show signs of gelation but a stable population of a second species is visible by the clearly bi-modal decay of the function. In turn, upon drying, AFM shows a high number of small fibrils appearing to percolate the sample.

5.1.3 CAHS-8 can form fibres-like crystals of the fibrils

5.1.3.1 The dehydration rate controls the assembly of CAHS-8

During the crystal screening and optimisation process of the helical domain construct HD, some remarkable crystal morphologies were observed.

First, polycrystalline spherulites were observed in dried or almost dried drops, showing clearly the so-called “Maltese cross pattern” under polarised light (fig. 14 A). These spherulites typically were either small and numerous, or big enough for one or two to span most of the dried drop. Then, very small needle-like crystals were also observed appearing to grow out of near-dried drops, giving the interface a spiky appearance and often coexisting with the spherulites. Spherulites are typically associated with fibrous protein or polymer crystallisation and crystal growing on the outside of the drop is uncommon, however, as the obtaining of diffracting crystals of the, at that time only hypothesised, coiled-coil of CAHS-8 was the priority, none of these conspicuous morphologies were investigated further.

Nonetheless, a drop in a poorly sealed crystallisation well eventually showed several millimetres-long spikes growing out of it and its observation under polarised light revealed a highly birefringent spherulite spanning the entire as well as a significant birefringence of these spikes (fig. 14 B). These spikes, which have a mushy texture and are deformable, were collected and subjected to X-Ray diffraction and proved to diffract to high-resolution in a fibre-diffraction like pattern but with separated enough diffraction spots (fig. 14 C), demonstrating that these spikes were poly-crystalline assemblies. However, such diffraction spots are too spread to be indexed, which likely arise from a significant mosaicity of these crystalline spikes.

In order to obtain better diffracting crystalline spikes, reproducing them under a controlled environment was necessary which required understanding their origin. However, no similar behaviour could be found in the literature to provide a starting hypothesis.

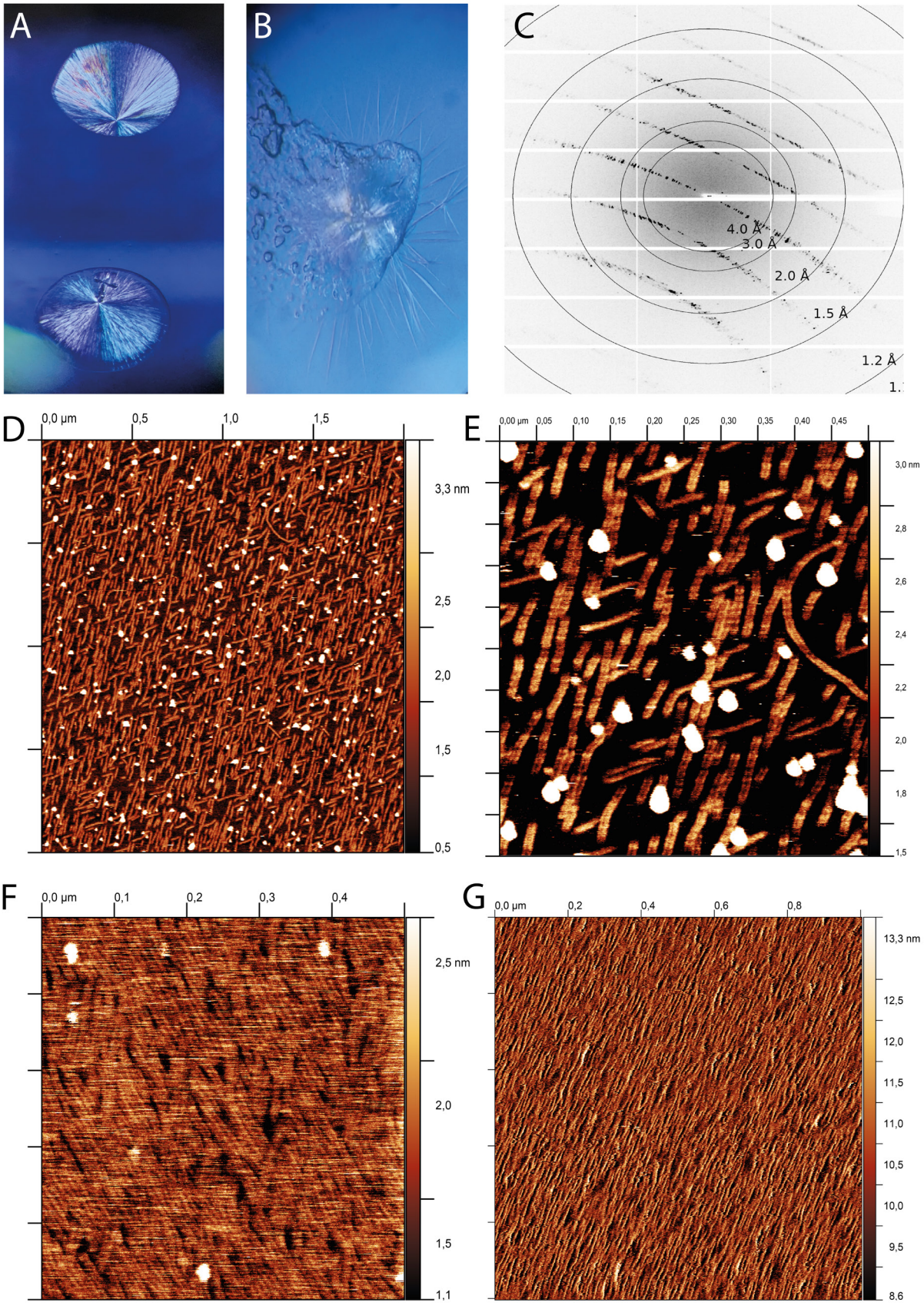

A
B
C
4.0 Å
3.0 Å
2.0 Å
1.5 Å
1.2 Å
D
0,0 µm
0,5
1,0
1,5
3,3 nm
2,5
2,0
1,5
0,5
E
0,00 µm
0,05
0,10
0,15
0,20
0,25
0,30
0,35
0,40
0,45
3,0 nm
2,8
2,6
2,4
2,2
2,0
1,8
1,5
F
0,0 µm
0,1
0,2
0,3
0,4
2,5 nm
2,0
1,5
1,1
G
0,0 µm
0,2
0,4
0,6
0,8
13,3 nm
12,5
12,0
11,5
11,0
10,5
10,0
9,5
8,6

Figure 6.14: CAHS-8 assembles into crystalline fibrillar morphologies.

A: Two dried drop of HD in a crystallisation well observed under polarised light reveals spherulites showing the distinctive "Maltese cross" and associated with polymer or fibrous crystallisation.
B and **C**: A dried drop of HD observed under polarised light shows birefringent spikes growing out of the drop (**B**). The X-Ray diffraction pattern of one of these spikes yields a polycrystalline fibre diffraction pattern with imperfectly resolved diffraction spots (**C**).
D and **E**: AFM imaging in liquid of a dissolved CAHS-8 gel at 30 mg/mL showing periodic ordering of fibrillar assemblies of CAHS-8 reminiscent of a liquid crystal, with two clear orientations (**C**). The focus on a zone (**D**) shows that these linear assemblies of CAHS-8 proteins have an average apparent width of 15 to 17 nm, are composed of two smaller assemblies of approximately 4 nm width, composed of regular units separated by 9 to 10 nm and have a height 1.2 to 1.6 nm. Although somewhat similar to the regular fibfres of CAHS-8, these assemblies appear slightly larger and half as thin.
F: AFM imaging in liquid of another sample of CAHS-8 prepared from the same stock and imaged on the same day as **C** and **D** showing an even clearer crystalline arrangement of these fibrillar assemblies. These fibrillar assemblies appear to also have an apparent width of 17 nm but a height of 0.7 to 1 nm.
G: AFM imaging in liquid of a different samples of CAHS-8 at 500 µg/mL with 5 M of TMAO also showing a clear orientation of the fibres, being 8 to 10 nm wide and 3 to 3.3 nm high.

However, these crystalline fibres weren't the only hard to explain phenomenon observed in CAHS-8. Indeed, over the many AFM measurements in liquid throughout the span of this thesis, the fibrillar assemblies of CAHS-8 were observed several times to spontaneously assume a periodic arrangement. This was observed for CAHS-8 both without (fig. 14 D, E, F) and with (fig. 14 G) the influence of the fibrillation-triggering chemical TMAO (cf. the dedicated section 5.1.4.2 page 236). However, this behaviour could never be reproduced in a controlled way despite several attempts and would remain only randomly observed.

Putting these two phenomena together led to the hypothesis that CAHS-8, and by extension HD, could aggregate not only as the fibrillar assemblies discussed previously but also as crystalline assemblies such as liquid crystals or crystalline fibres under the right dehydration rate. As such, the observed crystalline spikes could have resulted from the improper sealing of the crystallisation well, yielding a suitable rate for the phenomenon to happen, while the spontaneous self-orientation seen in AFM would result from an uncontrolled ideal conjunction of conditions.

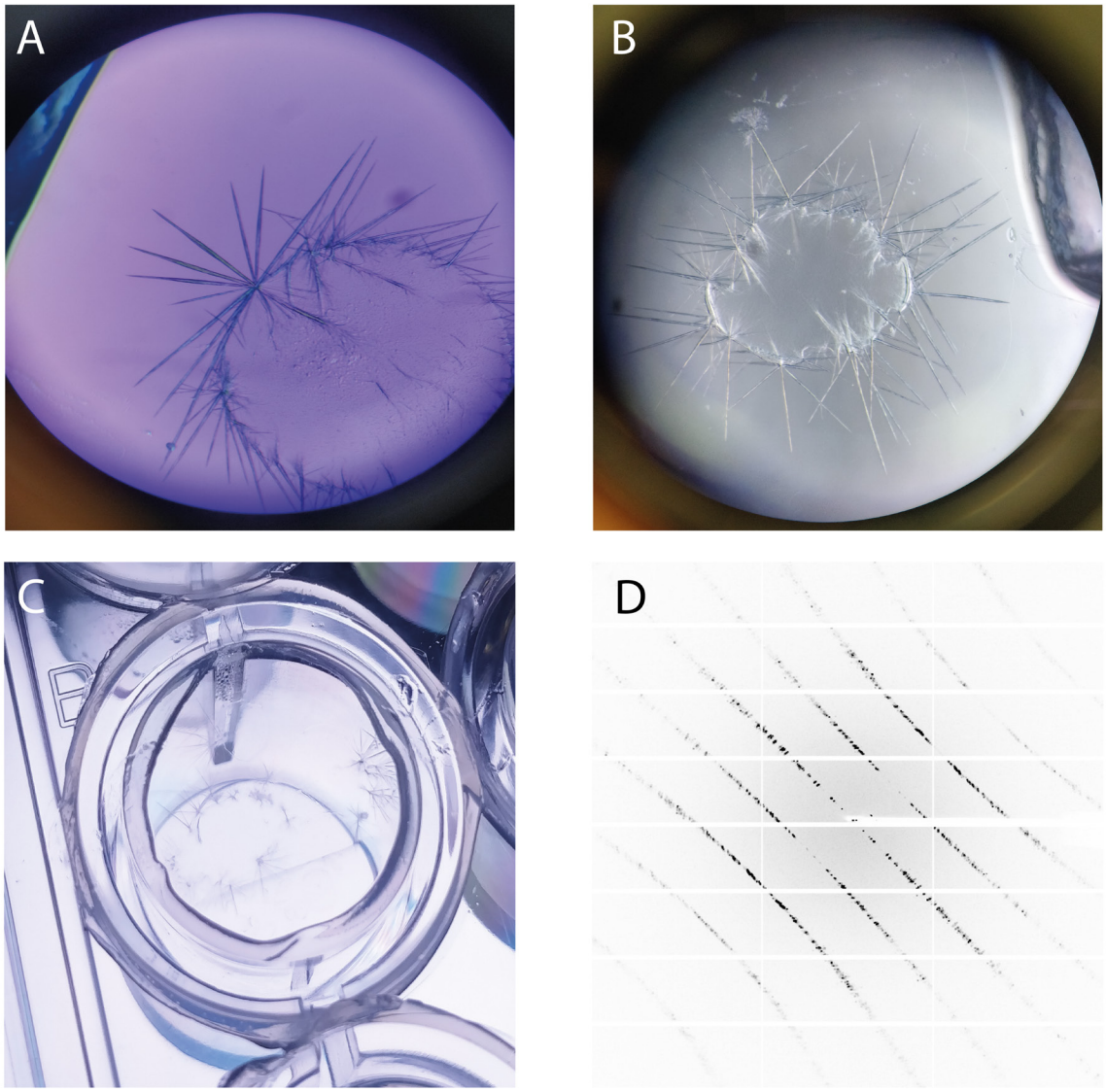


Figure 6.15: The assembly of HD into crystalline fibres depends on the RH and concentration.
A and **B**: Images of the crystalline fibres growing out of the dehydrated drop from HD at 3 mg/mL (**A**) and 7.5 mg/mL (**B**) grown at RH = 43%.
C: Picture of one of the hanging drop settings showing that these crystalline fibres are long enough to be visible to the naked eye.
D: X-Ray diffraction pattern of one of the crystalline fibres from **A** showing a high-resolution, fibre-diffraction like pattern. The resolution at the limit of the image is approximately 1.1 Å.

In order to verify this hypothesis, the relative humidity (RH) of crystallisation drops were systematically varied, as the RH is directly linked to the dehydration rate, as well as the precipitants.

The results shows that dehydration of HD at the probed RH of 43%, 57% and 67% reproduced the crystalline fibres growing out of the dehydrated drop, with the lower the RH or the concentration, the thinner or fewer the spikes, and, eventually, that precipitants

were not required for this crystallisation to happen (fig. 15 A and B). These crystalline fibres could reach a maximum length of several centimetres and were clearly visible to the naked eye (fig. 15 C).

At lower RH, such as 32 % or 23 %, only small, grainy aggregates that looked like a dehydrated gel were observed, while at higher RH such at 75 % and higher, bigger amorphous aggregates were observed which shows that the dehydration rate does control the morphology of the helical domain of CAHS-8 as observed.

Moreover, the volume of the drop was also shown to be a critical parameter. If too small, the crystalline fibres either were small and thick or missing, with the whole drop looking precipitated, and if too large, there were no spikes growth. Depending on the RH, the optimal drop size was found to vary between 2 and 4 µL.

The diffraction of these crystalline fibres of HD showed the same pattern as the one found serendipitously which indicate a similar arrangement (fig. 15 D). Although they diffracted up to a higher resolution, the spots were unfortunately still too broad to be indexed.

5.1.3.2 An assembly of crystalline fibrils

Surprisingly, not only the HD construct crystallised in such fashion but so could the wild type CAHS-8, despite its the 94 disordered residues in its N-terminal and the 38 residues in its C-terminal. Compared to HD, the crystalline fibres of CAHS-8 grow faster and at higher RH than HD (fig. 16 A and C). Additionally, CAHS-8 readily shows dendritic, or seaweed-like crystals over a wide range of RH, i.e. tree-like branching crystals of various size and fractal degree, with the higher the RH, the bigger the crystals composing these dendrites. These dendritic crystals grow later than the crystalline fibres (fig. 16 A and B) and unlike them, grow towards the inside of the drop rather than towards the outside (fig. 16 B, C, D) and, occasionally, appear to share a common nucleation point with the fibres (fig. 16 B and C). If this crystallisation follows Ostwald's rule, then this would indicate that the crystalline fibres are species less stable then the dendrites.

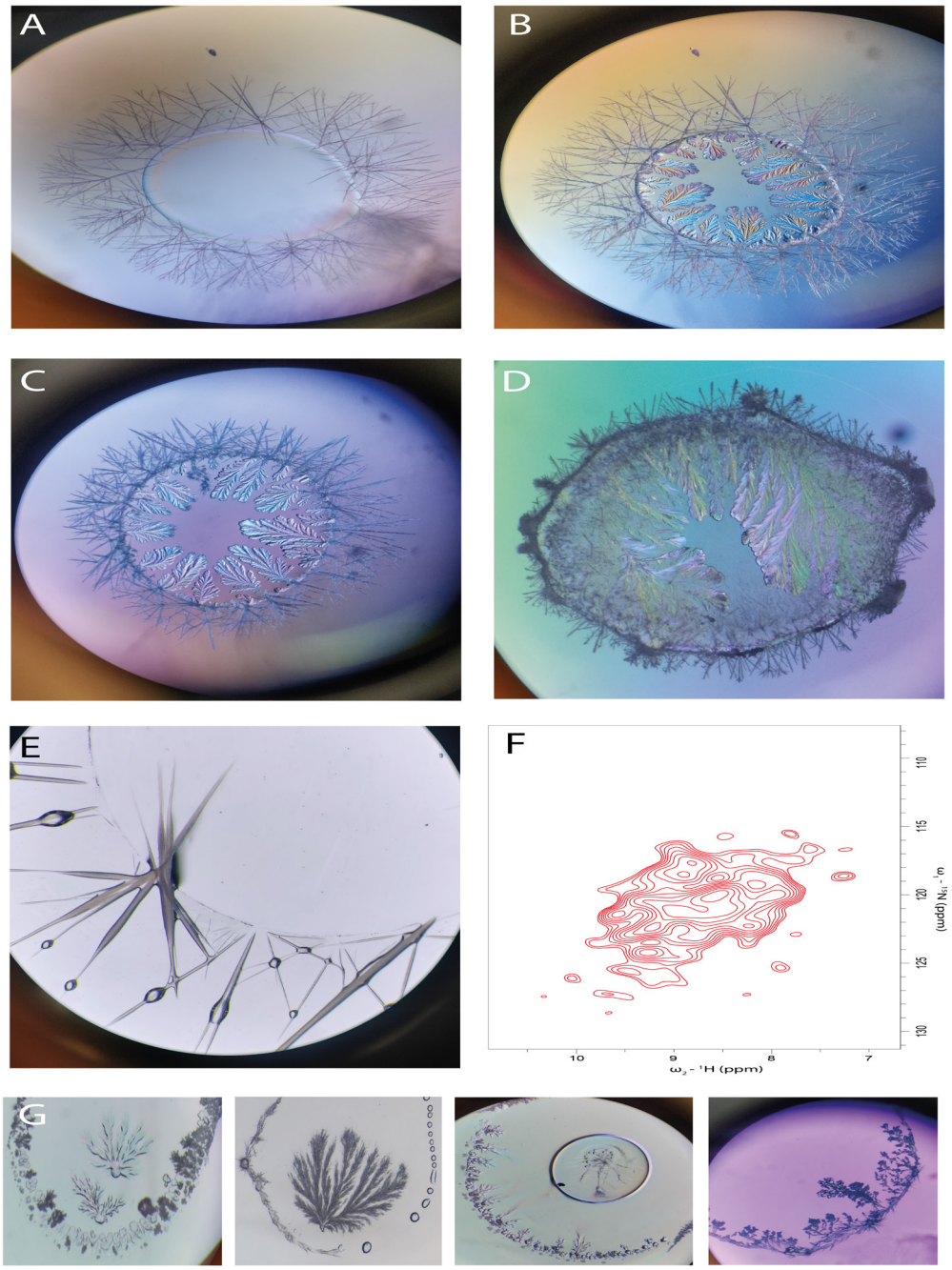


Figure 6.16: The full-length CAHS-8 can assemble into crystalline fibres and dendritic crystals.

A and **B**: Pictures of a drop of CAHS-8 at RH = 57 % and 5 mg/mL showing the crystalline fibres (**A**). The same drop 5 weeks later reveals a thickening of the crystalline fibres as well as dendritic crystals sharing an apparent nucleation point with the fibres, but growing on the inside (**B**).

C and **D**: Pictures of a drop of CAHS-8 at RH = 67 % at 5 mg/mL (**C**) and 10 mg/mL (**D**) showing thicker crystalline fibres than at 67 % (**B**).

E: The fibres can be rehydrated up to RH = 87 % without dissolving.

F: Proton detected MAS ssNMR hNH spectrum of crystalline fibres grown as in **A** and rehydrated to as in **E**, revealing broad linewidths.

G: Collection of dendritic crystals of CAHS-8. Although hardly suitable for crystallography, they have some aesthetic merits.

As the diffraction of these crystalline fibres showed the same spreading of the diffraction spots, ssNMR was therefore attempted to study them. In order to limit the broadening due to the chemical shift anisotropy and to reintroduce a degree of internal dynamics, the rehydration of the crystalline fibres was studied. The results show that once formed, those crystalline fibrils dissolve if put at a RH of 94% or above but not at 87%, indicating that the critical RH of dissolution is in-between these values (fig. 16 E). It is notable that the surface of these crystallised fibrils appears to be hydrophobic, with the water drop condensing circularly on them while they show an affinity for apolar solvents.

Many of drops of CAHS-8 were thus manually grown, the spikes individually collected in acetonitrile and rehydrated in order to fill a 0.7 mm MAS ssNMR rotor but unfortunately, they proved to be too long and solid to pack enter the rotor, forming instead entangled structures. They were therefore fitted into a 1.3 mm rotor, but as not enough material was produced, the rotor could not be fully packed. Nevertheless, the crystalline fibres were rehydrated up to at least 87% in order to permit enough internal dynamic to obtain thinner linewidth rather instead of a possible powder spectrum. The resulting proton-detected hNH spectra at 60 kHz MAS had a low sensitivity and showed broad linewidth still despite the crystallinity of these fibres (fig. 16 F). Unfortunately, as observed previously, the deuteration of the proteins inhibited the assembly of the crystalline fibres as well and thus, this strategy could not be used to limit the homogeneous line-broadening induced by proton-proton dipolar couplings.

Nevertheless, despite the dimer coiled-coil crystal structure not being resolved at that time and its existence unproven, these crystalline fibres were hypothesised to result from a periodic assembly of the helical domain of CAHS-8. As α-helices have a slightly anisotropic magnetic susceptibility and these crystalline fibres could reach the centimetre range, indicating a long-range order, a growth under a static magnetic field was therefore attempted in order to increase their homogeneity. After weeks of growth in crystal trays placed under a 950 MHz NMR spectrometer at different RH, drops of CAHS-8, HD as well as the NTDHD and HDCTD constructs were retrieved, showing a physically oriented growth in the opposite direction to B_0 (fig. 17 A, D, G, J).

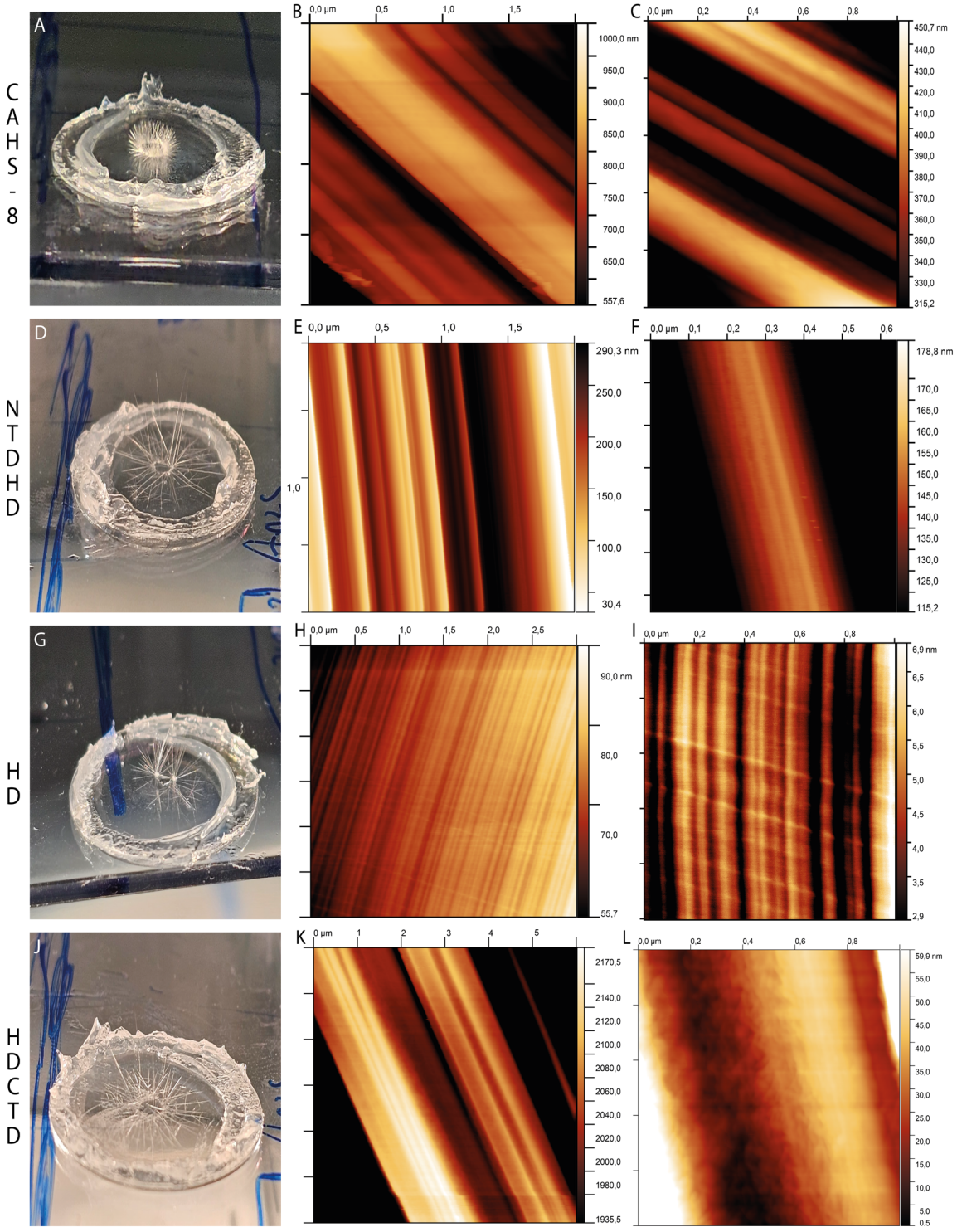


Figure 6.17: The different constructs of CAHS-8 form crystalline fibres that orient under magnetic field and are composed of c-axis grown crystals.

A, **D, G** and **J**: Picture of a flipped glass lamella used in a hanging drop set-up to form the crystalline fibres grown under a 950 MHz spectrometer at RH = 67 %, for respectively CAHS-8 (**A**), NTDHD (**D**), HD (**G**) and HDCTD (**J**). The crystals show a clear growth in the opposite direction to B_0 which corresponds to the up to down direction as the glass lids are flipped down compared to their regular sitting drop setting.

AFM imaging of the crystallised fibres of CAHS-8 (**B**) and focus on the crystalline arrangements that compose it (**C**), as well as for NTHD (**E** and **F**), HD (**H** and **J**) and HDCTD (**K** and **L**)

AFM imaging of these oriented crystalline fibres of CAHS-8 reveals that they are composed of smaller crystalline arrangement, oriented almost parallelly and with gaps between them (fig. 17 B). The subsequent focus on these smaller units reveals a very plane surface, but showing slight signs of being composed of even smaller crystalline unit themselves (fig. 17 C).

The crystalline fibres the NTDHD construct formed were longer than CAHS-8's (fig. 17 D vs A), but not all of them oriented under the magnet, with the remaining flat needles showing variable degree of curvature. The AFM imaging of the NTDHD fibres show a similar arrangement, with the smaller crystalline arrangement being smaller than in CAHS-8 (fig. 17 E). The imaging of one of these arrangement shows that they are composed of parallel, c-axis grown crystalline units stacked together and forming a tubular arrangement (fig. 17 F).

As per HD, the crystalline fibres orient under the magnet as well but appear to grow from a limited amount of nucleation points, showing an urchin-like pattern (fig. 17 G). They appear to be composed directly of parallel crystalline units which are not bundled together into the crystalline arrangement observed for the other constructs, thus showing limited space between them (fig. 17 H and I). Occasional glides are visible and shared amongst the many crystalline c-axis grown units forming the lattice of these crystalline fibres.

Finally, the crystalline fibres of HDCTD appear macroscopically similar to those of NTDHD (fig. 17 J vs D) but are organised similarly to those of CAHS-8 (fig. 17 K vs fig. 17 B). However, the focus on of these arrangement shows that they are the less regularly arranged than the other constructs, looking significantly more amorphous (fig. 17 L).

The X-Ray diffraction of these crystalline fibres of CAHS-8 (fig. 18 A) and HD (fig. 18 B) grown under the magnetic field shows a well resolved poly-crystalline fibre diffraction pattern for CAHS-8, with some of the images that could be indexed to P_1 symmetry. However, those of HD appear more amorphous, then previously, an averaging of the meridional reflections and an amorphous halo or alternatively, ice signal.

Nevertheless, the azimuthal integration of these two images (fig. 18 C) show a very similar pattern, indicating a similar structure despite the halo.

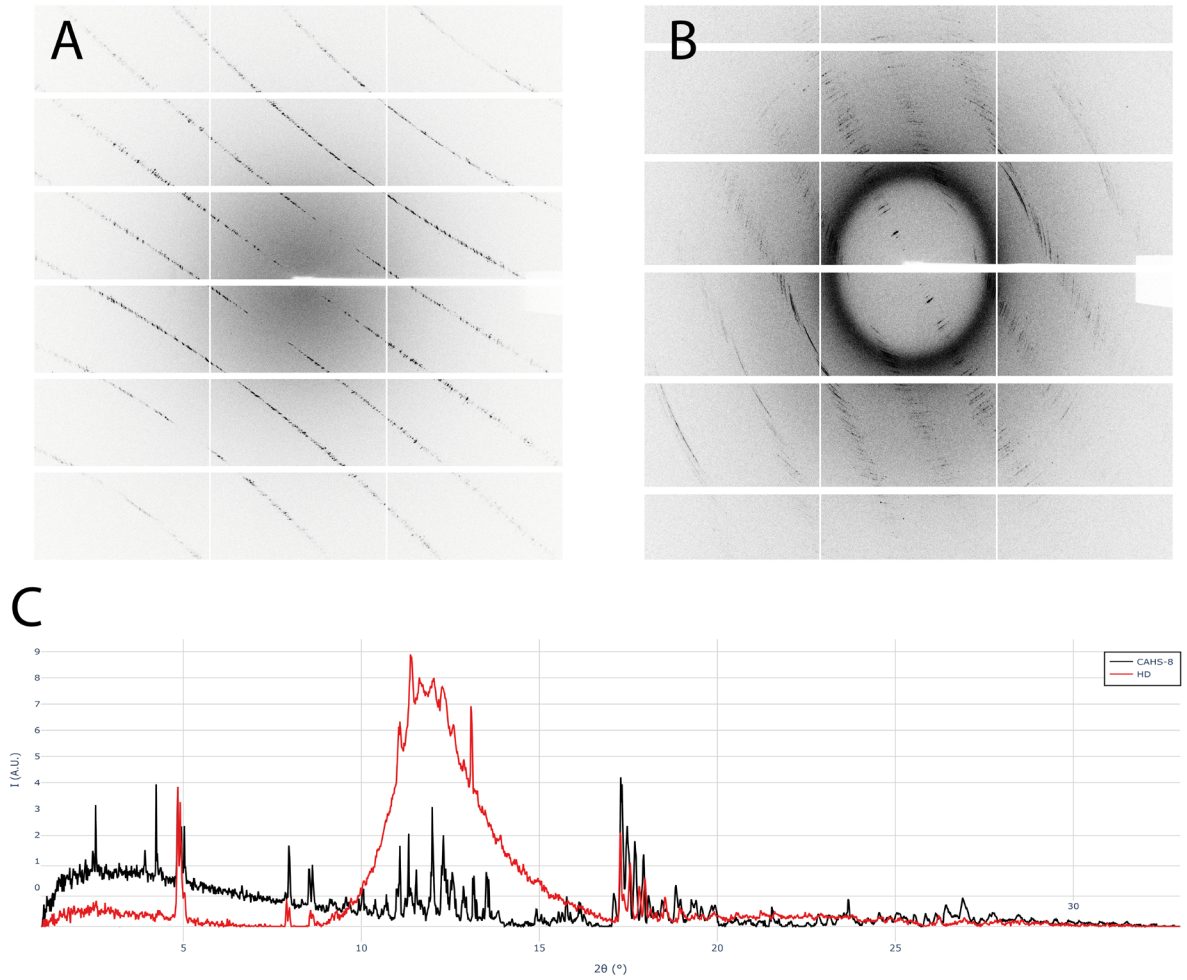


Figure 6.18: Fibre diffraction of the crystalline fibres grown under the magnet of CAHS-8 and HD.

A, and **B**: X-Ray diffraction of the crystalline fibre grown under the spectrometer of respectively CAHS-8 (**A**) and HD (**B**). The crystalline fibres of HD appear less crystalline than CAHS-8's, showing a clear amorphous halo and an averaging of the equatorial reflections.

C: Baseline-corrected azimuthal integration of **A** (CAHS-8, black) and **B** (HD, red), showing similar reflections despite the clear amorphous halo or ice signal.

As the crystalline fibres of CAHS-8 appeared to have the higher crystallinity of all the constructs, despite its disordered residues, they were imaged in order to resolve the crystal lattice parameters. High-resolution AFM over one of the crystalline arrangements composing the crystalline fibres of CAHS-8 reveals that these were indeed composed of fibrils of CAHS-8 arranged periodically and in parallel (fig. 19 A). This is even more visible in the Laplacian-weight treatment of the image, allowing to recognise that the fibrils are c-axis grown crystals (fig 19 B). This parallel arrangement of these fibrils shows some slight variations and defects, hence the spread of the diffraction spots, but renders difficult to study them individually as they are closely spaced and show an apparent twist.

The focus on a region where a fibril appears isolated enough and plane shows dimensions compatible with α-helices and the coiled-coils and fibril structure obtained from the microcrystals (fig. 19 C and D), which permit an overlay of the image and the fibril structure (fig. 19). As for the microcrystals, the coiled-coil is therefore the unit-cell of these crystalline fibrils.

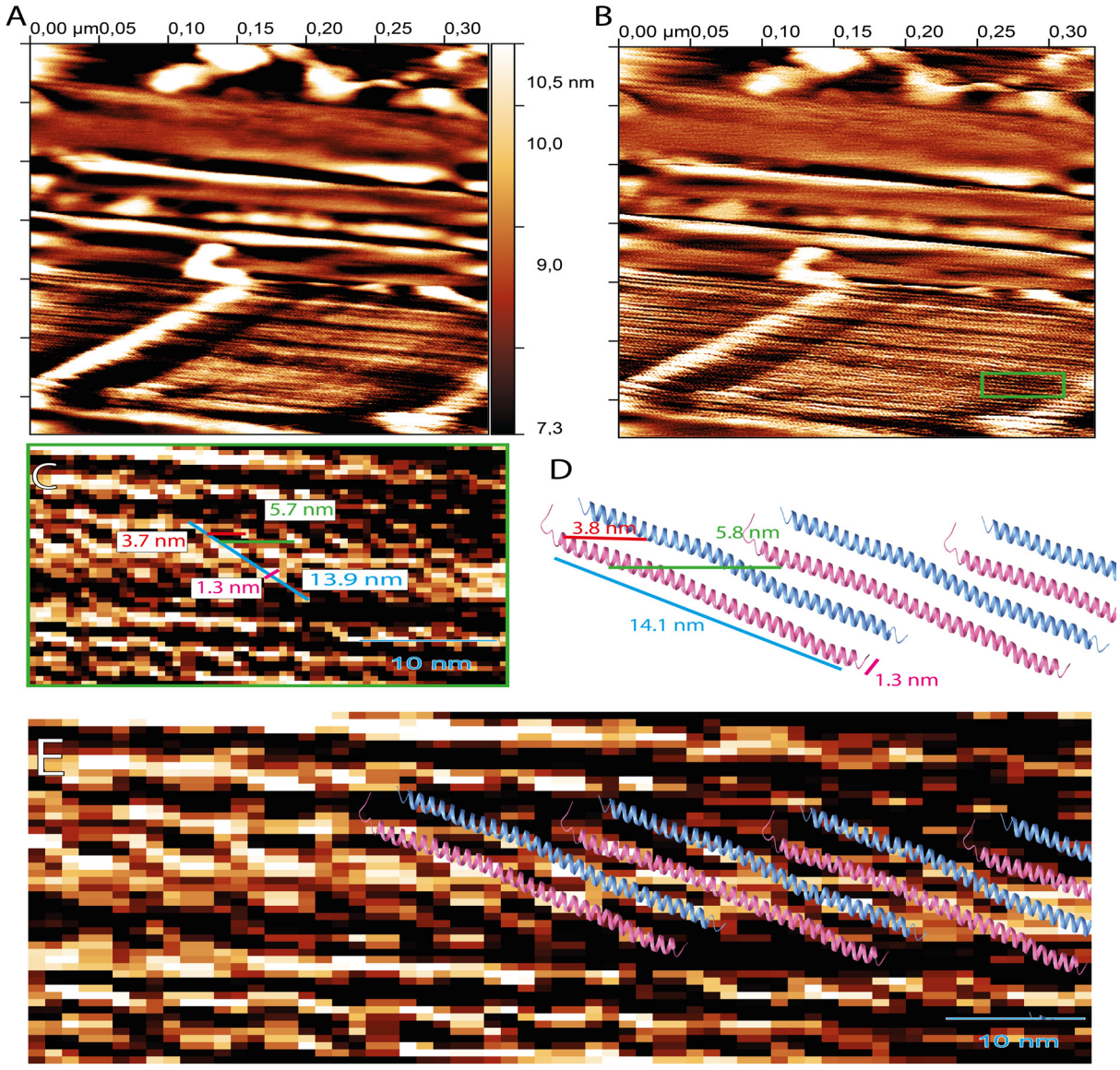


Figure 6.19: The magnetically oriented crystalline fibres of CAHS-8 are composed of a lattice of crystallised twisting fibrils.

High-resolution air-mode AFM imaging of one of the crystalline fibres of CAHS-8 grown under the magnetic field (**A**) and its processing using the Laplacian-weight method **(B)**, revealing that the crystalline arrangements are composed of closely spaced, parallels fibrils. A focus on a region of **B** where individual fibrils can be distinguished shows regular units of dimensions (**C**) compatible with the crystal structure of the coiled-coils and fibrils (**D**), permitting an overlay of the two (**E**). However, the fibrils show a twisting of the coiled-coils not present in the microcrystal structure as can be seen from the smaller dimensions on the left of the studied fibrils or on the others on top or bottom.

5.1.4 Manipulation of the fibrillation

As the helices of CAHS-8 are only transiently populated, the coiled-coil oligomerisation requires at least one initial step of conformational selection or induced fit. As such, the fibrillation of the protein would be limited by the probability of encounter of monomers in the right sub-conformation of their sampled ensemble for both the DCC and the FCC. Thus, by manipulating their helical conformational or their probability of encounter, the minimum concentration of fibrillation can be lowered.

5.1.4.1 TFE

The first compound tried to this aim is 2,2,2-Trifluoroethanol or TFE. Indeed, during the crystal making process, alcohols such as ethanol, isopropanol and TFE were observed to make the crystallisation of HD faster, although it was later shown to result in more crystal defects. Moreover, TFE has long been known to favour the formation of α-helix through a probable dehydration effect, solvating the protein and therefore preventing it from hydrogen-bonding with the solvent, which causes the protein to form hydrogen-bonds with itself and thus, to aggregate or to form α-helices depending on the concentration[605–607].

AFM imaging show that indeed, TFE appears at first to collapse CAHS 8 and HD into a gel-like state at a 100th of the minimum gel concentration of CAHS-8. This gel likely made of aggregated protein coated with TFE, as the resulting TFE-induced gel was observed to be hydrophobic unlike the native gel and thus required to be imaged on a highly oriented pyrolytic graphite (HOPG) surface instead of the common, hydrophilic mica. Indeed, on mica, the sample remains as a highly elastic drop, rendering imaging challenging.

However, TFE is volatile and once evaporated enough, fibres start to appear. As TFE is thought to aggregate around the protein, a high concentration of it could locate on the DCC or FCC faces of the helical domain of CAHS-8 which would prevent it from forming coiled-coil assemblies[605]. The eventual formation of fibres therefore would correspond to a

concentration low enough to allow the intra CAHS 8 coiled coil formation, but high enough to limit the hydrogen bonding with the solvent. Unfortunately, no adjustment of the protein concentration or percentage of TFE could prevent the initial gelation. Due to this, as well as the instability of the sample resulting from the volatility of TFE, apparent different properties of the TFE-induced gel state and its corrosive effect on the AFM apparatus, TFE studies were not pursued further.

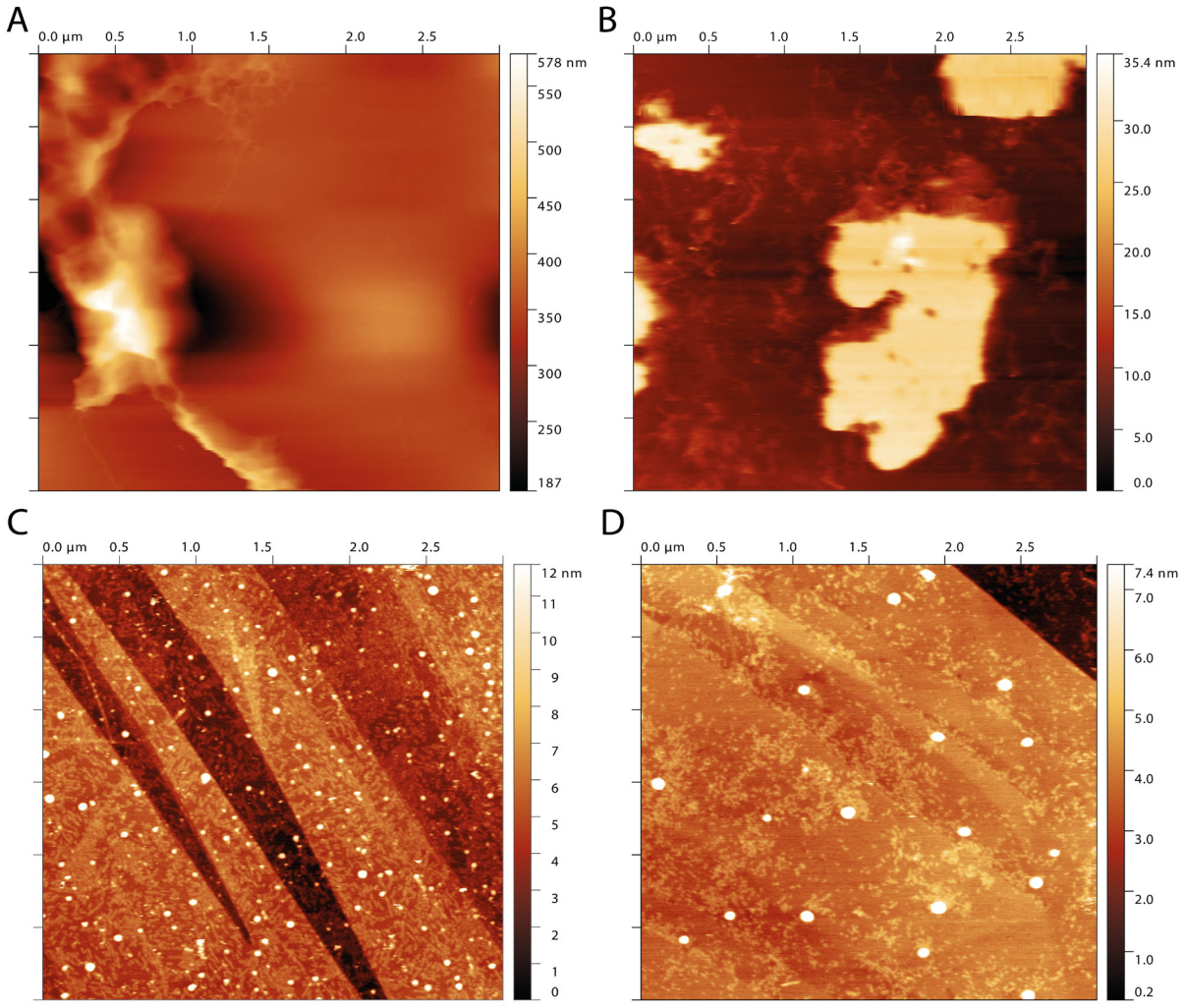


Figure 6.20: TFE induce a gelation and a particular fibrillation of CAHS-8 and HD.

A and **B**: Liquid-mode AFM imaging of CAHS-8 (**A**) and HD (**B**) at 1 mg/mL after addition of 5 % v:v of TFE, showing the immediate formation of a gel.

C and **D**: As TFE is volatile, after approximately 30 min, the gel dissolves and reveals small fibrillar assemblies for CAHS-8 (**C**) and HD (**D**). Both fibrillar assemblies have a height of 1.2 nm and are large of approximately 22 nm which does not correspond to the dimensions of natural fibrils or fibres. The one fibre of CAHS-8 visible on the upper left corner of **C** is 27 nm large and 3 nm high which corresponds to the regular dimensions.

As this TFE-induced gel shows hydrophobic properties, it was imaged on HOPG instead of mica. The surface of HOPG is responsible for the patterns visible in C and D.

5.1.4.2 TMAO

Trimethylamine N-oxide (TMAO) is an amine oxide compound with chemical chaperoning effects found in marine fish and crustaceans, allowing the stabilisation of the secondary structure of their proteins against water pressure. This chemical chaperoning ability of TMAO opposes the denaturation effect of urea to almost a 1:1 ratio. As it is a stable compound and not a volatile alcohol like TFE, TMAO was used in order to increase the population of the α-helices of CAHS-8, thus allowing to bypass the initial conformational selection step.

The results show the potency of TMAO to trigger the fibrillation of CAHS-8, even at a minimum protein concentration, and with TMAO-concentration dependent effects.

Indeed, at the very low concentration of 5.7 µM of CAHS-8 in 1 M of TMAO triggers the formation of fibrils (fig. 21 A). These fibrils appear poorly defined and their AFM regular dimensions of approximately 3 nm depth and 11 nm of apparent width which most likely corresponds to the 2-4 nm width of the fibrils visible in EM and the crystal structure.

At 11.4 µM of CAHS-8 and 2.5 M of TMAO, only a few small assemblies are visible (fig. 21 B) but at 3.5 M of TMAO, are much more visible (fig. 21 C). These small fibrils, which appear to display a certain degree of self-orientation, have a width of approximately 10 nm of average but an apparent depth of 2.2 nm, which is smaller than the native ones. Raising the concentration to 4 M of TMAO yields fibrils of CAHS-8 (fig. 21 D), which also appear to show a certain level of self-orientation and have a depth of 2 to 3 nm and an apparent width of approximately 14 nm, which is slightly wider than the native fibres. Numerous small species with mostly the same orientation are visible and the background and are 1.5 nm heigh and approximately 15 nm wide, which corresponds to the length of the helical domain of CAHS-8.

However, adding 3.5 M of TMAO to a solution of CAHS-8 at 11.4 µM, which corresponds to the same conditions as fig. 21 C but with the reverse preparation order, results in a very fibrous gel spanning the whole sample (fig. 21 E).

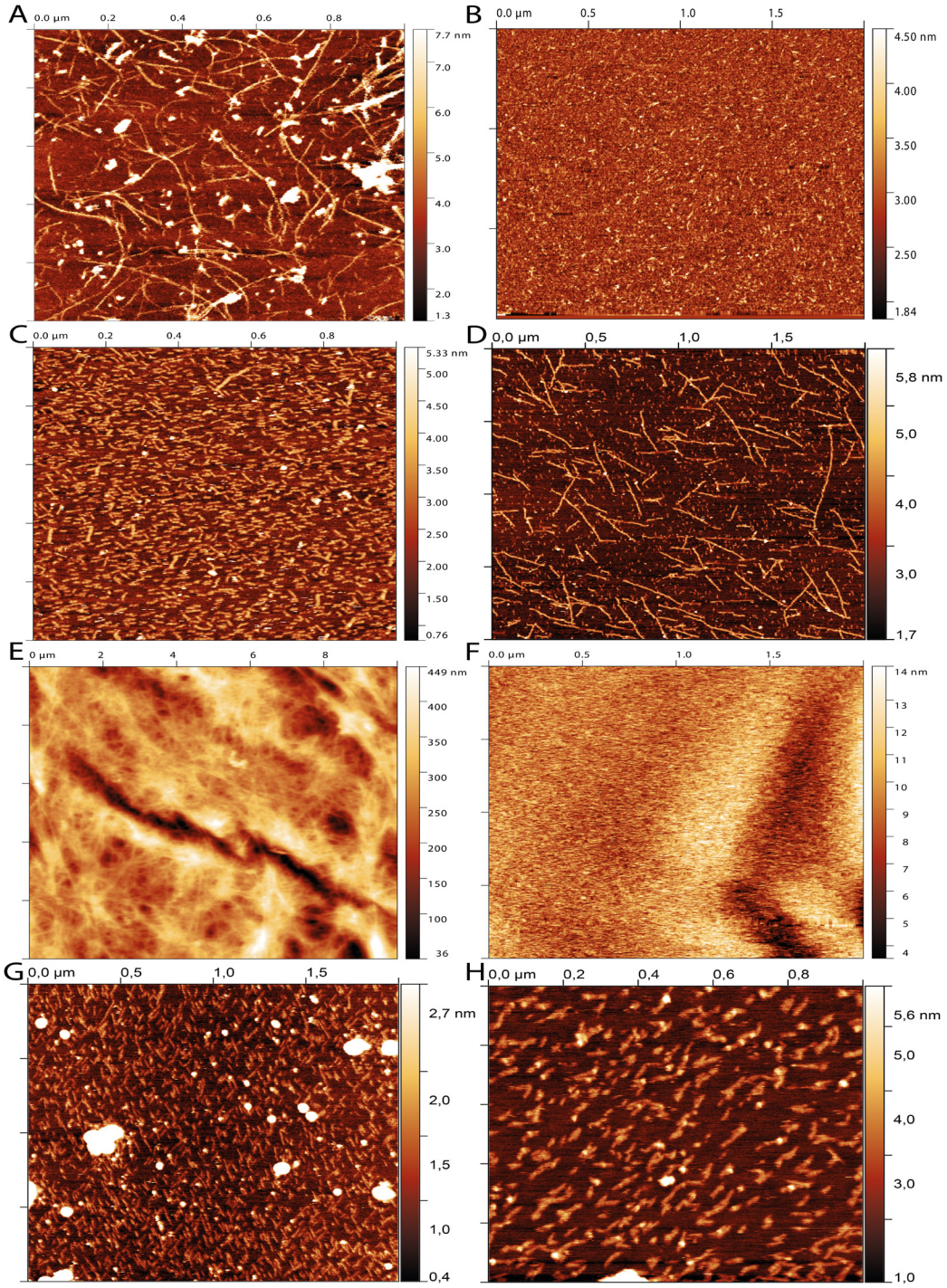


Figure 6.21: TFE induces a gelation and a particular fibrillation of CAHS-8 and HD.

A and **B**: Liquid-mode AFM imaging of CAHS-8 (**A**) and HD (**B**) at 1 mg/mL after addition of 5 % v:v of TFE, showing the immediate formation of a gel.

C and **D**: As TFE is volatile, after approximately 30 min, the gel dissolves and reveals small fibrillar assemblies for CAHS-8 (**C**) and HD (**D**). Both fibrillar assemblies have a height of 1.2 nm and are large of approximately 22 nm which does not correspond to the dimensions of natural fibrils or fibres. The one fibre of CAHS-8 visible on the upper left corner of **C** is 27 nm large and 3 nm high which corresponds to the regular dimensions.

As this TFE-induced gel shows hydrophobic properties, it was imaged on HOPG instead of mica. The surface of HOPG is responsible for the patterns visible in **C** and **D**.

As per HD, a concentration of 11.4 µM of protein in 5 M of TMAO yields a dense coat of fibrillar assemblies with hydrophobic properties, which thus requires imaging on HOPG instead of mica (fig. 21 F). Finally, 6.5 µM of NTHD in 5 M of TMAO yields aligned fibres of 27 nm of apparent width but a depth smaller than 1 nm (fig. 21 G) while 9.3 µM of CTHD in 5 M of TMAO also results in fibres but of 22 nm of apparent width and 1.6 nm deep.

Altogether, these results demonstrate that TMAO is indeed successful in triggering the fibrillation of the different constructs of CAHS-8 at concentration considerably lower than in native conditions, which would be agree with the possibility that fibrillation of CAHS-8 is limited by an initial step of conformational selection. However, while the fibres of CAHS-8 obtained in fig. 21 A and D have dimensions and morphologies similar to their native form, the other fibrillar assemblies of CAHS-8 or of the other constructs show distorted dimensions.

5.1.4.3 Gadolinium

Finally, solvent Paramagnetic Relaxation Enhancement (sPRE) experiments were attempted in order to find a proof of the existence of the then-hypothesised coiled-coil of CAHS-8. These experiments used Gd(DTPA-BMA), a chelate of the paramagnetic lanthanide gadolinium also known as Gadodiamide and were performed at 350 µM and 293 K. However, the samples of CAHS-8 always ended up gelating after a few hours while normally, a concentration of approximately 665 µM is necessary for the protein to gelate at 293 K.After verification, it became clear that Gd(DTPA-BMA) was responsible for this gelation. Unbeknown when these sPRE experiments were performed, lanthanides ions have been demonstrated to stabilise coiled-coils thanks to their 3 positive charge interacting with anionic sidechains, thereby resulting in a non-covalent cross-linking of the α-helices[608–610]. In turn, the use of reduced Gd(DTPA-BMA) on HD resulted in molten-globule like NMR spectra, indicating a collapse of the protein further confirmed by both AFM and SAXS. This further supports that that absence of the C-terminal disordered region results in non-specific interactions.

5.2 Discussion

Through a combination of NMR secondary chemical shifts, PREs, RDCs, relaxation as well as SAXS, it is now clear that CAHS-8 is an intrinsically disordered protein with a central helical domain made of two transiently populated α-helices and flanked by fully disordered chains, with the helices remaining flexible with respect to one another due to the flexible interhelical linker and their helical populations being increasingly more populated as the temperature decreases (ch. 5 fig. 1, 2, 7).

Although CAHS-8 is mainly monomeric, it appears to transiently exchange with a higher order oligomer, termed the antiparallel dimerisation-driving coiled-coil (DCC), of which the crystal structure of the DCC could be obtained by X-Ray diffraction crystallography (ch. 5 fig. 15). This coiled-coil structure reveals that the linker between the two helices undergoes a conformational change to form a helix, so that the helical domain becomes a single long helical domain comprising both helices present in the monomeric form. However, it is unknown whether the dimeric form in solution presents this feature, but this seems probable considering the slight helical propensity of the linker (ch. 5 fig. 2) and its conformational exchange (ch. 5 fig. 20) which may be between disordered and helical conformation. The crystal structure presents a similar local topology to its ensemble averaged solution conformation (ch. 5 fig. 17) and displays an unusual periodicity, with hendecad repeats in its extremities associated with the helices showing virtually no supercoiling, and flanking a central region showing a regular heptad periodicity and delimited by two π-π interactions (ch. 5 fig. 16). In turn, as the concentration increases or the temperature decreases, the equilibrium shifts and the fractional population of the coiled-coil dimer increases which eventually leads to fibrillation (ch. 5 fig. 6, 10).

This fibrillation is driven by a second stabilisation mechanism, involving an apparent, antiparallel fibrillation-driving coiled-coil (FCC) taking place concurrently on the outer face of the C-terminal half of the helical domain of CAHS-8 which was also resolved in the crystal structure (ch. 6 fig. 4), with the fibrils spanning the entire lattice and representing the biological assembly of the protein (ch. 6 fig. 2). The 4 nm wide fibrils could be observed

by electron microscopy with dimensions agreeing with the crystal structure, and also display a hierarchical assembly with two fibrils forming together a 9 nm wide fibre. These fibres could be observed in AFM and in electron microscopy and display dimensions compatible with the model derived from the crystal structure and actually present in the crystal lattice (ch. 6 fig. 2).

The structure of the fibrils could also be observed by AFM of a crystallised fibril form obtained by the dehydration of CAHS-8 at a particular rate and representing a novel kind of crystal morphology (ch. 6 fig. 19, 15, 16). Finally, the fibrillation via the FCC was shown to be guided and favoured by the C-terminal disordered termini of CAHS-8, and repressed by the N-terminal disordered termini (ch. 6 fig. 11, 12, 13). Without any of these termini, the DCC could still form but the FCC would happen in a non-specific way, resulting in aggregates.

5.2.1 A first characterisation of CAHS proteins

In the founding paper published in 2012 announcing the discovery of the CAHS family of proteins, Yamaguchi and co-workers from the Kunieda group already noted on the basis of Circular Dichroism (CD) experiments and sequence analysis that these newly discovered CAHS proteins in the tardigrade *R.varieornatus* were largely "intrinsically unstructured" and may be able to form amphipathic α-helices like LEA proteins discussed in the introduction of this thesis[611]. In 2017, Boothby and co-workers used deuterium-exchange experiment on CAHS proteins to conclude that they were indeed intrinsically disordered. However, until our article published in 2021 which includes some of these results on the CAHS-8 of *H.exemplaris*, no characterisation at residue-resolution of the secondary or tertiary structure of any CAHS was published[396]. To this day, the only other characterisation of the secondary structure of a CAHS protein at residue resolution was performed by NMR on a 62 residue-fragment of the CAHS-1 protein of *P.richtersi* corresponding to portion of the helical domain of CAHS-8 and published in 2024[612]. Thus, the characterisation of CAHS-8 presented in these chapter constitute the first per-residue characterisation of a CAHS protein, or of any Tardigrade exclusive Disordered Protein (TDP). As such, it represents a significant advance for the field and is, at the moment of

the writing of this thesis, the only complete one. It shows that indeed, the protein exhibits extensive disorder, with over 100 disordered amino acids, but also shows significant level of secondary structure, with two helices that are well populated in the monomeric state.

Moreover, the comparison of 85 attested and putative CAHS protein sequence amongst many tardigrade species showed a significant amount of sequence conservation, in particular in what corresponds to the helical domain of the studied *H.exemplaris* CAHS-8. Thus, the findings presented in these chapters, and more generally in this thesis, are probably largely exportable to these other CAHS proteins. This is supported by the use of secondary-structure prediction algorithms as well as AlphaFold predictions on these other CAHS proteins which all predict a central helical domain flanked by more or less disordered termini.

Nevertheless, the study of the behaviour of CAHS-8 by NMR was made difficult by the multitude of phenomenon happening at the same time. Indeed, the transient DCC over the helical domain, in particular the region between Y126-F151, and its exchange contribution appear to coexist with an intramolecular conformational exchange in the linker region as well as with the FCC oligomerisation (ch. 5 fig. 10), with all these species and conformation probably having specific, and very different anisotropic rotational diffusion tensors. Thus, it is difficult to spectroscopically distinguish these phenomena as they have similar effect on the relaxation. As such, the different constructs, namely HD, NTDHD and HDCTD, as well as the crystal structure of the fibrils proved essential to deconvolute the different effects and their location.

5.2.2 Stutters, hendecads and an antiparallel orientation

The particular geometry and features of the dimerisation-driving coiled-coil (DCC) of CAHS-8 result from the non-canonical distribution of polar and apolar residues in its helical-domain, resulting in the presence of multiple stutters amongst heptads, which in turns results in the alternating of 11/3, 25/7 and 11/3 again periodicities in respectively

its M93-A125, Y126-F151 and R152-Q183 (ch. 5 fig. 16). Consequently, the stutters render the coiled-coil antiparallel in order to accommodate the knobs-to-knobs packing mode of the stutter **h** residue facing directly the interface[526,531,544]. In light of CAHS-8's ability to form fibrils and fibres, it is also noteworthy that the non-canonical knobs-to-knobs packing displayed in the crystal structure has been inferred for decades to favour higher-order oligomerisation of two-stranded coiled-coils[526].

This coexistence of knobs-to-knobs and knobs-into-holes packing as well as the topological deformations of the α-helices explain the difficulties to identify a satisfying search model for molecular replacement phasing. Indeed, none of the tried search models of other coiled-coils or cross-α amyloids had this peculiar arrangement, nor any of the conformers generated nor any of the AlphaFold generated models, most likely resulting in Patterson maps too different for a satisfying phasing solution. Moreover, the fibrous lattice resulting from the lateral FCC between asymmetric DCC units results in a tight packing with barely any water molecules around the C-terminal half of the α-helices, which likely complicates the separation between the self- and cross- Patterson vectors performed in the molecular replacement[613]. This was clearly visible in the models generated with ARCIMBOLDO which repeatedly mistook the C-terminal half of the α-helices for belonging to another asymmetric unit. Nevertheless, the overall lack of supercoiling over more than two third of the DCC coiled-coil also validates the hypothesis made during the molecular replacement step, where the log-likelihood gain score was better with search models displaying little supercoiling, with this observation allowing to eventually find a satisfying molecular replacement solution.

5.2.3 A transient coiled-coil and possible sequence of events

The structure of the dimerisation-driving coiled-coil (DCC) may also rationalise why CAHS-8 is mainly monomeric up to a certain concentration or temperature. The presence of the many polar residues discussed in the core positions **a**, **d** or **h** of the repeats as well as residues facing each other having the same charge destabilises the coiled coil

respectively due to desolvation penalty and possible burial of ions for the former and charge repulsion for the latter (ch. 5 fig. 16).

Although the structure shows the features mentioned above to limit their destabilising effects, the behaviour of CAHS-8 and HD in solution indicates that the free-energy minimum of the coiled-coil conformation is not low enough for the proteins to not dissociate due to thermal energy. As such, it is likely that the dissociation k_{off} of the dimerisation is significantly bigger than the association constant k_{on}, providing a possible explanation of why the coiled-coil dimer is only transient.

However, the formation of the FCC may stabilise the assembly, thereby triggering fibrillation. As such, the isolated DCC is a transient species which either rapidly dissociates, or form fibrils. This explains why the isolated DCC could never be directly observed until the obtaining of the crystal structure, with the different experiments showing monomers or higher-order assemblies.

The helical domain of CAHS-8 is delimited in one first helix (E95-F151), a linker (R152-E166) and a second helix (V167-L194), with respective α-helical populations of approximately 70, 15 and 50 % (ch. 5 fig. 1). As in the crystal structure, the DCC spans the entire helical domain (ch. 5 fig. 15, 16) while the FCC spans the linker and helix 2 (ch. 6 fig. 4), the formation of any of these coiled-coils, and thus of the fibrils, is inherently limited by the α-populations of these domain and therefore involve at least an initial step of conformational selection or may alternatively arise from an induced fit triggered by transient interactions.

The model of conformational selection is supported by the use of TFE or TMAO which both increase the helical populations of CAHS-8 and thus trigger the fibrillation at concentration as low as 5.7 µM at room-temperature (ch. 6 fig. 20, 21) instead of approximately 500 µM without (ch. 5 fig. 6). However, compared to CAHS-8, the helical domain construct HD has a reduced ability to fibrillate, yet it has a slightly higher helical

propensity which indicate that the helical population is not the only important parameter (ch. 5 fig. 11, 18).

As the Y126-F151 central region of the DCC has a more canonical, heptad-based organisation due to its more regular distribution of hydrophobic and polar residues, as well as being stabilised by two π-π stackings on its extremities (ch. 5 fig. 16) and shows the fastest transverse relaxation rates of in all construct of CAHS-8 (ch. 5 fig. 10, ch. 6 fig. 10-12), it is likely that this region is the first to dimerise. This is also supported by this region being the only one to show a significant exchange contribution in CAHS-8 at 303 K (ch. 5 fig. 20).

Moreover, HD appears to form the DCC at a lower efficiency than CAHS-8 (ch. 5 fig. 12) and shows a slightly different residual dipolar coupling profile (ch. 5 fig. 11). Yet, both CAHS-8 and HD do show a highly similar RDC profile in the 131-139 segment, which indicate a similar average orientation over the ensemble and may support the initial formation of the DCC in this central region first. This is further supported by HD showing a significant, and periodic, apparent exchange contribution to the transverse relaxation only in this region Y126-F151 at 293 K (ch. 5 fig 21).

It is then unclear whether the formation of the FCC over the linker and helix 2 requires the DCC being already formed over these domains too or can happen independently of the DCC. As helix 1 faces helix 2 in an antiparallel manner in the DCC and helix 1 has a higher helical propensity than helix 2 (ch. 5 fig. 1), it is likely that the DCC forms first as helix 1 would be on average helical already, with helix 2 being then stabilised by the formation of the FCC its other side with another helix 2 (ch. 6 fig. 4). This is supported by the observation of a more significant exchange contributions in the transverse relaxation in the whole helix 1 than in helix 2 at 293 K in CAHS-8 (ch.5 fig. 21), which can indicate, as per the DLS-derived phase diagram (ch. 5 fig. 6), that the formation of the FCC is either marginal and quenched by the 1500 Hz $R_{1\rho}$ spin-lock and thus is a slower exchange phenomenon, is too minor to be detected or does not happen at this temperature.

However, the relative lack of exchange contribution in HD in helix 1 before Y126 (ch. 5 fig. 21) and HD's formation of the FCC being impaired by its lack of the C-terminal disordered chain, suggests that the formation of the DCC on E95-A125, which faces R152-G182 in an antiparallel direction, and the formation of FCC are correlated phenomena. This is strongly implied by the faster R_2 rates and higher R_2R_1 product of the HDCTD constructs in the E95-A125 compared to HD's, being even comparable to CAHS-8's (ch. 6 fig. 11). As the HDCTD construct lacks the N-terminal disordered chain, these faster R_2 rates in helix 1 cannot result from a drag-effect slowing down its rotational correlation time and thus can only results come from an exchange phenomenon. As HDCTD was shown to have a strong propensity to form the FCC (ch. 6 fig. 12, 13), this indicates altogether that the E95-A125 DCC formation with its partner R152-G182 is favoured by the formation of the FCC, probably due to the C-terminal disordered chain stabilising E95-A125 on the other side (ch. 6 fig. 9).

Finally, R152 is precisely the first residue of the linker region (152-166), and highlights its importance. As shown in the deuterated HD at 278 K, the linker appears to undergo an intermediate exchange between helical and disordered conformations (ch. 5 fig. 19) which, as per its $k_{ex} = 749 \pm 105\ s^{-1}$, cannot be the exchange visible in the relaxation as this would have been quenched already by the spin-lock. As this region is helical in the crystal structure, it is unclear whether the formation of the DCC or the FCC in solution can happen with the disordered conformations of the linker but helix 1 and 2 being helical. As the linker represents almost half of the coiled-coil interface of R152-G182 with E95-A125 (ch. 5 fig. 16), its helical conformation may be required and would suggest a conformational selection step which may act as a further limitation on the fibrillation. However, the helical conformation of the linker may not be required for the initial the formation of the DCC but may happen upon formation of the FCC which may represent an induced fit mechanism and may be one of the mechanisms preventing the dissociation of the DCC upon formation of the FCC as discussed previously. Finally, it is also possible that the linker remains exchanging between helical and not helical conformations in the fibrils, conferring them an inherent flexibility, and thus adaptability to physical constraints.

These possible conformational selection steps and other inhibitory effects on the oligomerisation of CAHS-8 may be functionally useful to the tardigrade. Indeed, although

the studied *Hypsibius exemplaris* or species like *Paramacrobiotus richtersi* express CAHS proteins in response to desiccation, other tardigrade species like *Milnesium tardigradum* expresses them constantly[614]. As such, it appears necessary for the tardigrade to have the protein forming fibres and gels in its cells only in the event of a desiccation stress and not at all time, which would likely be detrimental to the tardigrade. As such, a latent ability to form the DCC, fibrils and fibres, being prevented in normal conditions but unlocked once dehydration happens as, even without surexpression of CAHS, it automatically translates to an increase of the protein concentration due to the loss of water, thus favouring contacts between proteins and the probability of association and thus, the realisation of the protective function of the proteins. The possible mechanisms of this protection are studied in the second part of this thesis.

5.2.4 Relevance for CAHS and LEA proteins

Additionally, in the aforementioned paper published in 2012 reporting for the first time the tardigrade CAHS family of proteins, Yamaguchi and co-workers from the Kunieda suggested that these new CAHS proteins may have, like LEA proteins have been hypothesised to, a possible similar “molecular shield” role but without specifying how exactly[611].

In our paper published in 2021, we advanced the hypothesis that CAHS-8 could form a coiled-coil oligomer, which we also hypothesised to serve as the base unit of the fibril form of the protein. The obtained crystal structure is therefore the first structure of a CAHS protein and provides a molecular description supporting this hypothesis which was indirectly stated for the first time 12 years ago.

Additionally, the very high sequence conservation of the amino acid sequence in the helical domain in the region corresponding to the helical domain of the CAHS-8 protein of *H.exemplaris* studied here amongst many other CAHS proteins of other tardigrade species of more or less distantly related families indicate strongly suggests that these findings are

relevant for most of these proteins too (ch. 5, fig. 6 and fig. 16 C). Namely, these other CAHS proteins are highly likely to also form α-helices and coiled-coil oligomers, being predicted to do so by prediction algorithms and models.

Indeed, one of the defining particularities of the CAHS-8 coiled coil is the π-π stacking of Y126 and F151 delimiting a region characterised by a different, canonical coiled-coil packing from the rest of the coiled-coil and which also presents the fastest transverse relaxation in its monomeric solution form, making it the probable anchoring point of the dimerisation (ch. 5 fig. 16, 10). Both Y126 and F151 are present in all 85 aligned CAHS protein sequences, while the residues in-between display the highest average conservation of any region (ch. 5 fig. 5).

Additional evidences of the presence a coiled-coil amongst these other CAHS are also visible in the mutation patterns of the different CAHS proteins, assuming that the structural adjacency seen in CAHS-8 is reproduced for these sequences. For instance as previously mentioned, in the studied *H.exemplaris* CAHS-8 antiparallel DCC, Y104 faces M173, probably forming a stabilising sulphur-π interaction. However, this particular tyrosine-methionine pairing is found only in 14 sequences out of the 85 sequences, in 22 of the sequences the tyrosine would be paired with a lysine thereby permitting the formation of a cation-π interactions, while in respectively 25 and 3 of them, the tyrosine would face another tyrosine or a phenylalanine, which can introduce two supplementary π-π interactions at the extremities of the coiled-coil (ch. 5 fig. 5).

Moreover, this structure is also significant for LEA proteins. These proteins, which were discussed in greater details in the first chapter of this thesis, were discovered in the 1980's and share a similar desiccation-protection role in plants and some small animals to the tardigrade CAHS proteins. The discovery an "11-mer" amino-acid motif in their sequence led to hypothesise in 1993 that these proteins form an amphipathic α-helix which would promote their oligomerisation as coiled-coils when dehydrated and that these coiled-coils would not be based on the heptads known at the time but on repeats of a proposed novel 11 residue motif[545,615]. This 11-residue coiled-coil repeat basis would later be recognised as common motifs and referred to as hendecads[547] or undecatads[616].

Although supported by sequence analysis, ambiguous Fourier-Transform Infrared (FT-IR) measurements[617] or Molecular Dynamics simulations[618], the possibility for LEA protein to form coiled-coil has not been decisively proven experimentally to this day, with no structure having been resolved either. Thus, our tardigrade CAHS coiled-coil crystal structure is also the first of this class of the desiccation-protection protein and strikingly presents the remarkable hendecad periodicity hypothesised for the first time for LEA proteins[545]. As noted above, the very high sequence conservation of the coiled-coil forming motifs strongly suggest that its formation is important for the desiccation-protection function of the CAHS proteins. As such, given the structural homology and similar function of the α-helices forming intrinsically disordered LEA proteins families[619], the tardigrade crystal structure supports their long-hypothesised ability to form coiled-coils as well.

5.2.5 Crystallisation of CAHS-8

The crystallisation of the helical domain of CAHS-8 was attempted to force the dimerisation and entrap the protein in this form without it fibrillating, thereby allowing to directly observe its hypothesised intermediary coiled-coil form. However, obtaining of crystals, and in particular crystals diffracting to high-resolution, was made difficult by the fact that the helical domain is not a folded domain, with the helices being only transiently populated, in particular in the linker and helix 2.

It was eventually noticed that HD aggregated in various forms, including crystalline, in a few hours to a few days in conditions involving dehydrating agents such as acetone or alcohols, in particular isopropanol and TFE. This suggests that a partial desolvation of CAHS-8 favours the formation of the coiled-coil dimer, probably through the increase of the helical content of CAHS-8 as the protein is forced to form helix-forming hydrogen-bonds with itself to compensate the loss of hydrogen-bonds with the solvent. This is consistent with the fact that the protein protects tardigrades from desiccation and therefore, dehydration triggers its function.

Yet, the formation of the DCC also triggers the self-assembly of the protein into fibrils via the FCC, which represented a serious challenge in the obtaining of high-resolution diffraction crystals. Indeed, the propensity of CAHS-8 to aggregate as fibrils even when crystallised, or alternatively to crystallise as fibrils, likely manifests in their overwhelming tendency to form needle-shaped microcrystals, either isolated or in urchin-like clusters. As such, no conditions were found to permit to obtain big crystals as decreasing the protein concentration or the saturation conditions in the drop only led to either fewer crystals or marginally bigger microcrystals. This is most likely due to a slight but increasing shift of the crystal lattice resulting from fibrils running over the entire crystal length as were finally obtained (ch. 6 fig. 2 A). However, the crystal lattice is typically straight, and thus, a slight coiling pitch from the fibrils results in a shear stress on the crystal, thereby breaking the crystal once it has reached a critical size, or rather, the critical strain it can withstand has been reached. This may also explain the widespread crystal twinning commonly observed during the crystallisation of HD, resulting from a dislocation in the crystal lattice that could occur once this shear mechanical stress is too significant to accommodate.

Thus, although HD was crystallised in order to observe the coiled-coil dimer and avoid fibril formation it was the fibrils that eventually ended up being crystallised.

This is not an isolated case as fibrous proteins are known to be difficult to crystallise due to their strong tendency to aggregates as fibres or amyloids instead of crystals, which is why fibre diffraction has historically been the diffraction technique of choice in these cases rather than crystallography[526]. In the relatively few cases where the structure of fibrous proteins has been resolved through X-Ray diffraction crystallography, it often took decades of efforts, the resolution is low or the crystals are typically obtained on truncated forms of the protein.

For example, the well-known coiled-coil of the fibre-forming tropomyosin protein has had the structure of its complete form resolved to a 15 Å resolution in 1986 (PDB: 2TMA)[620], to 7 Å resolution in 2000 (1C1G)[621], and to this day, only fragments of the coiled-coil forming domain have been resolved to a higher resolution. Moreover, as discussed in more details in the introduction in the section “Coiled-coils based fibrils” page 148,

intermediate filament proteins, which also form coiled-coil based fibrils and fibres, have also been exceedingly difficult to crystallise, which could only be achieved so far for selected segments of some of the intermediate filament proteins[524,557–559].

Similarly, amyloid-forming segments, for example of Sup35, Tau, Aβ or PSMα3, crystallise as fibrils or proto-fibrils that span the entire lattice, and have thus been reported to only ever yield needle-shaped microcrystals[622–624]. Additionally, amyloid fibrils have been observed to grow out of the extremities of microcrystals, which highlight the similarity between the structures in the microcrystals and the fibrils[623,625].

This is an important observation as it strongly suggests that the crystallised fibrils form of CAHS-8 is the native one, minus the possible twist observed in the crystallised fibres form. Indeed, the DCC was shown to bear similarities to the ensemble conformation of the protein in solution, there are precedents in these other fibrous proteins, the crystallised fibrils of CAHS-8 agree with the experimental dimensions (ch. 6 fig. 2) and the α-helices organisation could be directly observed in the crystallised fibre form (ch. 6 fig. 19). Moreover, the entire 99 residue-long helical domain of CAHS-8 has been crystallised. This is notable because fibrous protein are typically only crystallised in truncated segments due to the inherent difficulties posed by the formation of fibrils, and while these segments may form crystallised into aberrant and non-physiological states due to the truncations, this is less likely with the complete domain[626].

5.2.6 The fibril structure and assembly mode

In our article published in late 2021, we reported the formation of fibrils by *H.ex* CAHS-8 and speculated at the time that the fibrillation could result from an assembly of coiled-coils of helix 1 transversal to the fibril axis as the dimension of the fibrils, that is 9 nm in width and 3 nm in height, corresponded respectively to the approximate length of helix 1 and to the approximate height of a canonical coiled-coil[627].

At a few days difference from our article, Yagi-Utsumi, co-workers from the Kato team and collaborators reported that *R.va* CAHS-1 could form fibrils via homo-oligomerisation of the supposed α-helical region, but without proposing a model[628].

In an article published in early 2024, Sanchez-Martinez and co-workers from the Boothby team proposed on the basis of measurements of dimensions and mutants that instead, fibres of *H.ex* CAHS-8 result from bundles of 22 dimers which would assemble parallel to the fibril axis and be stabilised via putative β-β interactions between the termini, a hypothesis which is further discussed in the next section[629].

The obtained crystal structure of the fibrils and its agreement with experimental measurements therefore allows to decisively end the speculation around the structure of these fibrils.

The assembly of the CAHS-8 fibrils resembles the so-called A_{11} or A_{22} **a**ssembly modes of tetramers of intermediate filaments (IF) proteins in which coiled-coil dimer units assemble laterally in an antiparallel manner with another coiled-coil dimer unit, on respectively their coil 1 domain or on the coil 2, eventually yielding the so-called unit-length filament[626,630,631]. Although these assembly modes have been modelled for many years[630,632,633], only relatively recent crystal structure of fragment of vimentin[563], lamins[564,634], keratins[565] or the glial fibrillary acidic protein[635] allow the experimental observation of the structural basis of these assembly types to form filaments.

Moreover, despite the highlighted deformation of the α-helices of CAHS-8 in the crystal structure, some of the crystal structures of these IF proteins have a remarkable similarity to CAHS-8's. For example, the RMSD of the chain B of CAHS-8 with the keratin 5 chain (chain A) of the heterodimer coiled-coil crystal structure of the second half of coil 2 between humans keratin 5 and keratin 14 (PDB: 3TNU)[568] is 0.917 Å over 65 atoms, i.e. almost 65 % of the 99-residue helical domain of CAHS-8, with the structures overlaying precisely on the C-terminal part of the helical domain of CAHS-8 where the fibrillation-driving coiled-coil happens.

However, the CAHS-8 coiled-coil dimer units do not assemble transversally with respect to the fibril axis but in a staggered parallelogram way, which has been resolved in 2024 by cryo-EM to be also the case for the protofilament of vimentin, with 5 of these vimentin protofilaments assembling as the mature tubular intermediate filament[636,637]. Although the assembly mode of these vimentin filaments is very similar to CAHS-8's, it is important to note that CAHS-8 was not observed to form tubular fibres in solution as the measured heights of fibrils and fibres mostly the same, with only their width changing. Moreover, the FCC assembly of CAHS-8 appears to be a real and concomitant coiled-coil and not simply some complementary interaction site or "sticky-end" as in recent synthetically designed fibre-forming coiled-coils.

The similarity of CAHS-8's helices and fibril assembly mode with intermediate filament proteins is especially intriguing as tardigrades, and *Panarthropoda* in general, have lost the canonical cytoplasmic IF proteins like keratins or vimentin, and only possess a unique IF protein called cytotardin which is apparently derived from the nuclear IF protein lamins and forms a filamentous network around the nucleus and the cell border[638]. As such, although tardigrades do possess the other types of cytoskeleton structures, namely actin and microtubules, the particular role of cytoplasmic intermediate filaments would have to be fulfilled by other proteins, and include the protection of the cells from soft mechanical deformation[639,640]. In early 2021, we injected fluorescently-labelled CAHS-8 to cells and noticed formation of the fibres reported in this part of the thesis which would appear to form cytoskeleton-like structures spanning the entire cell when they were subjected to desiccation-mimicking osmotic shock. A more detailed, systematic study of this behaviour was published in 2022 by the Kunieda team using three *Ramazzottius varieornatus* CAHS proteins and proposed that "CAHS proteins are novel cytoskeleton-like proteins that form filamentous networks"[641].

Although the precise origin of CAHS proteins is still unclear, genes for these proteins have only been found in *Eutardigrada* and not in the other tardigrade class, *Heterotardigrada*, which suggest they appeared after this split[410]. As such, they cannot simply originate from IF proteins as the whole *Panarthropoda* clade to which tardigrades but also arthropods or velvet worms belong, lacks this family of proteins[638]. Thus, on top of their already highly similar desiccation-similar role and organisation into hendecads coiled-

coils to LEA proteins, the CAHS proteins, or at least CAHS-8, also present unique similarities cytoplasmic IF proteins in the structural arrangement of their fibrillar assemblies. Altogether, CAHS proteins may therefore be a remarkable case of convergent evolution to both LEA proteins and cytoplasmic IF proteins.

5.2.7 Assembly of the fibres

The precise mechanism of assembly of two CAHS-8 4 nm wide and 3 nm deep fibrils, into one 9 nm wide and 3 nm deep fibres remains unclear. HD appears able to form 15-20 nm wide fibres (ch.5 fig. 18) but with a much lower proficiency than CAHS-8, while some of these HD fibres have a 3 nm height, other showed a consistent 4.3 nm depth which likely indicate a different morphology. As the HDCTD construct was observed to form fibrils, but not fibres, of an apparent 10 nm width and 2.1 nm depth (ch. 6 fig. 13 F), which most likely correspond to 4 nm actual width as observed in Electron Microscopy (EM), while the NTDHD construct only formed fibrils of 9 nm width, likely corresponding to the 2-4 nm wide fibrils with the broadening effect of the AFM tip, but with an apparent height of less than 1 nm (ch. 6 fig. 13 E). If not an effect of the AFM settings, then these data indicate different morphologies compared to the native CAHS-8 fibrillar assemblies which therefore appear to require both disordered termini to be formed.

This is supported by the exchange contribution visible at 278 K in the N-terminal localised in the high sequence conservation 35-40 and 60-80 regions. In those two regions, F38, H40 and F71 stands out, with CAHS-8's F38 being a tyrosine in 55% of the 85 aligned CAHS sequences, and a phenylalanine in 13%; H40 is a histidine, a tyrosine or a phenylalanine in respectively 56%, 11% and 2% of the sequences; F71 is a phenylalanine in 72% of the sequences (ch. 5 fig. 5, ch. 6 fig. 9). This strongly suggest that these aromatic residues are important to a particular function of the N-terminal disordered chain.

The potential partner in the helical domain is likely be the conspicuous H145, which is one of the most conserved residues in the protein, being a histidine in 96% of the sequences and a tyrosine in 2%, despite being involved neither in the DCC nor in the FCC as its sidechain point perpendicularly with respect to these lateral coiled-coils. As such,

H145 may serve as an anchor point to one of these aromatic residues of the N-terminal disordered arm, holding together the two fibrils via a π-π interaction to form a fibre which are the configuration that would permit it (ch. 6 fig. 9). This is further supported by the fact that Q144, which precedes it, is a glutamine in 99% of the sequences and a histidine in the remaining 1%.

Indeed, the sequence F38-T39-H40-T41 include these two threonines, with T39 and T41 being a threonine in respectively 62% and 68% of the sequences, or a serine in respectively 2% and 11% of the sequences. As per F71, it is preceded by a highly conserved G70, which is a glycine in 94% of the sequences and a serine in 11%, or G68 which is a glycine, an alanine or a serine in 69, 13 and 11%, which indicate the need of a small amino-acid, and followed by T72 which is respectively a threonine, glutamine serine or histidine in respectively 34, 27, 11 and 9% of the sequences (ch. 5 fig. 5, ch. 6 fig. 9). This suggests an additional hydrogen-bond stabilisation of this putative fibre-holding interaction as all these polar residues are able to engage in such interactions.

Eicher and co-workers from the Pielak group have proposed in late 2022[642], as well as Sanchez-Martinez and co-workers from the Boothby group in late 2023[629], that the extremities of CAHS-8 would form β-strands that would interact with each other. These proposed mechanism come after our published findings in late 2021[627] which show that the secondary chemical shifts of the disordered terminal chains do not display a β-strand propensity at any temperature, although if present they should engage in such putative interactions at 278 K given the formation of fibres at this temperature and concentration (ch. 5 fig. 7). Moreover, we could not observe any particular change of relaxation that would be associated with the formation of interacting β-sheets at 278 K (ch. 5 fig. 9) nor signs of the presence of β-strands in the gel state relaxation as presented in the next chapter. However, the 40-60 region does appear slightly extended in the residual dipolar couplings of CAHS-8 (ch. 5 fig. 2) but as no secondary-structural propensity could be detected, and neither any predicting algorithm nor AlphaFold predict any β-strands, suggesting that this RDC tendency likely reports only on the amino-acid composition.

Finally, the presence of the sidechain of H145 pointing toward the transverse direction to the DCC and FCC plane and not being involved in these coiled-coil may render an alternate interaction mode of the fibrils which would result from the π-π interactions of the H145 of the different helices of two fibrils. If the fibrillar assemblies of HD (ch. 5 fig. 18 C) are indeed fibres and not broadened fibrils, then this tighter, histidine zipper mode may explain how they could form and why some of these HD fibrillar assemblies have a depth of 4.3 nm, which would corresponding to stacking of two fibrils. Such zippering of coiled-coils mediated by π-π stacking of accessible aromatic residues can be designed in synthetic coiled-coil or coiled-coil-like peptide to trigger their assembly into fibres, demonstrating its feasibility[578,643].

5.2.8 The crystalline fibres: a saturation-induced physiological state?

The relative humidity (RH)-dependent crystallisation of the helical-domain of the CAHS-8 protein without precipitants is not a commonly seen phenomenon. In fact, the only similar phenomenon that could be found in the literature is an article showing that one of the desiccation-associated LEA proteins, *af*LEA1, could "undergo[…] crystallization when dried more slowly for hours rather than minutes"[644] in pure water which, remarkably, was published at the same time the fibre crystallisation of CAHS-8 and its dependence on the RH were discovered during this thesis. However, the relative humidity (RH) at which of these *af*LEA1 crystal does not appear to have been controlled, they have a branching, dendrite-like morphology and are less than 100 µm long, which is hardly comparable with the centimetre-long crystallised fibres of CAHS-8 (ch. 6 fig. 16).

Nonetheless, the behaviour of CAHS-8 and the morphology of its dehydration-induced crystals show a high similarity to polymer crystals. Indeed, polymer crystals typically show spherulite, dendrite, seaweed or the so called "shish-kebab" crystal morphologies, as regions of polymers can form local crystalline zones while other zones remain amorphous. As such, considering the helical domain of CAHS-8 as the crystalline region and the disordered termini as flexible, amorphous regions of the polymers, the crystal growth could result from an initial nucleation point caused by the periodical coiled-coil

arrangement of the helical domain of CAHS-8. From this seed, depending on the conditions and in particular, the saturation, the surface tension and diffusion, the crystal would adopt more or less regular morphologies, that is, from the most regular to the least, dendrites, seaweed dendrites or lamellar spherulites (ch. 6 fig. 14 A-B, fig. 16)[645–648].

However, although the internal structure of polymer crystals resembles that of the crystalline fibres of CAHS-8, no similar crystals growing out of the drop could be found in the literature. These crystalline fibres therefore appear to be a newly reported kind of crystal form. As such, several points need to be considered in order to explain them.

First, the units aggregating on a crystal and driving its growth require diffusion to reach it. However, as the growth of these CAHS-8 crystalline fibres happens outside the drop (ch. 6 fig. 15, 16) the monomers cannot diffuse in the air outside of the drop to partake in the crystal lattice.

A first possibility would be that these crystalline fibres are covered by a thin water film which would allow the CAHS-8 monomers to diffuse along the outside of the fibre. However, the elongate morphology of the fibres ought to cause a faster evaporation than in the drop, which would quickly dry these fibrils unless a constant influx of water is maintained and on experimentation, the surface of the crystallised fibres of CAHS-8 appears highly hydrophobic, as revealed by the condensation of water into circular drops on their surface when they are put at a higher RH (ch. 6 fig. 16 E).

Alternatively, a growth of the c-axis grown individual fibril crystals from their base, which would be the drop-air interface, may circumvent this diffusion problem, but the force needed to push them would be considerable.

However, as seen in the AFM images, large spaces are present between the different fibrillar assemblies, i.e. the bundles of crystallised fibrils, which may serve as channels that could retain bulk water and in which the monomers would diffuse (ch. 6 fig. 17). In turn, the hydrophobic surface would prevent evaporation, essentially making these crystallised fibres water-storage containers which could thus keep other proteins hydrated, and protected from the nefarious effects of dehydration. As per the comparison

of the crystallised fibres of the different constructs, the space between the fibrils bundle appears directly correlated with the presence of the disordered termini which would therefore be free to interact with entrapped proteins (ch. 6 fig. 17). This hypothesis is attractive as the protein was shown to participate in the protection of tardigrades to desiccation, however it is unknown at this point if these fibres represent any physiological relevance or are simply a remarkable effect of the biophysical properties of CAHS-8.

The question of what determines whether CAHS-8 aggregates into the regular fibrils, fibres and gel, or into these crystallised fibres is worth considering, in particular if these crystallised fibres are an actual physiological mechanism. The RH was shown to be critical but other parameters are also important such as the volume of the drop. These two parameters point towards the supersaturation, as determinant in the selection of the aggregation type. Indeed, as the RH controls the dehydration rate, the dehydration of the drop would eventually cause it to become supersaturated, at which point the crystallisation of CAHS-8 may occur. However, crystallisation of a protein provides a pathway to drive the system to thermodynamic equilibrium, therefore decreasing the saturation. Nonetheless, if a balance between the dehydration rate, controlled by the RH as well as the exposed drop surface which increase the saturation of the solution, and the crystallisation, which decreases the saturation, this may keep the system in the right supersaturation range necessary for their growth despite their length, as it would be thermodynamically more favourable for diffusing monomers to join the crystal, also rationalising the importance of the RH and the volume of the drop. The fact that these crystallised fibres represent a favoured thermodynamic state is also visible from the fact that these fibres are preferentially grown up to a certain point, after which only dendritic crystals forms grow in the inside of the drop (ch. 6 fig. 16 A, B). As per the surface tension anisotropy of the drop, it may also contribute as it has also been shown to be an important parameters in the morphology of the polymer dendritic crystals[646,647].

Intriguingly, the amyloid formation, which has often been compared to crystallisation as both show a seeding effect, or amorphous aggregation were favoured was also reported to depend on the supersaturation[649–651]. Moreover, amyloid fibrils of a peptide were

shown to order into a 2D crystalline phase at the air-water interface[652] and formation of fibrils by the liquid-liquid phase separation FUS protein was also recently reported upon maturation at the interface between the dense, droplet phase and the dilute phase[653].

Additionally, low molecular weight peptides or compounds forming fibres entangling into gels have been reported to preferentially arrange into crystals instead of gels upon slight chemical changes[654], to transition from a fibrous gel to microcrystals upon addition of ethanol[655,656], DMSO[656] or sonication[657], or to spontaneously transition from a fibrous gel to microcrystals via fibre-fibre interactions[658] or via spontaneous crystal formation[659]. In turn, a theoretical model argues that in the case of these low-molecular weight molecules forming fibrous gels, the gel may represent a metastable equilibrium which, depending on the condition, may be overcome to reach a true thermodynamic equilibrium which can be crystals[660]. Although not applicable to CAHS-8 per se, which is much bigger than any of these species, these examples are reminiscent of the rapid formation of microcrystals of HD upon addition of alcohols such as isopropanol (ch. 5 fig. 14) and highlight the counter-intuitive interplay between fibres, fibrous gels and crystals, which would support the view that these crystalline fibres represent a particular region of the phase diagram of CAHS-8, in which the fibrils would pack into parallel bundles instead of into a gel.

In fact, this may have been observed many times already as the spontaneous alignments of the fibrillar assemblies of the protein (ch. 6 fig. 14 D, E, F) which, in particular, appear to be favoured by high concentration of TMAO, that is 3.5 M to 5 M, on CAHS-8 (ch. 6 fig. 14 G, fig. 21 C, D) or its constructs (ch. 6 fig. 21 G, H). Being an osmolyte, high concentration of TMAO increases the saturation of the system and may therefore favour this crystalline arrangement of CAHS-8. As AFM images the species adsorbed to the surface, it is not known whether the soluble species or bigger, sedimenting fibrillar assemblies order into a liquid-crystalline arrangement, which is a likely possibility and would explain the mechanism of formation of these crystals.

Indeed, lyotropic liquid crystals can readily be formed from sizeable molecules with a clear distribution of polar and apolar groups, particularly if they are rod-shaped. Due to their excluded volume, the competition between orientational and translational entropy,

which respectively favour isotropically distributed or parallelly aligned distributions of the rod-shaped objects, has long been established to permit the spontaneous formation of a liquid crystal phase at volume fraction depending on the conditions and dimensions of the rod-shaped species[661,662]. In particular, the Onsager rod-model as well as other models predicting this disorder-to-order transitions highlight the necessity for the rod-shaped objects to be significantly longer than they are wide, while the flexibility of the rod-shaped species or the heterogeneity of their length being predicted to increase the critical volume fraction of the transition[661–667]. Consequently, fibrillar assemblies of proteins or peptides, including amyloids or even of globular protein such as actin, have been shown to form well characterised, mainly nematic, liquid crystal phases[668–673].

As such, it is probable, that when CAHS-8 forms fibrils long enough and at a density high enough, whether because of the dehydration rate or TMAO, a liquid crystalline phase is formed and which may therefore result in a solid crystal when dehydrated.

Finally, it is noteworthy that as tardigrades undergo dehydration, the saturation of their cytosol mechanically increases due to the loss of water which may, at least locally, permit the formation of such crystallised fibres, if possible in vivo, as the effects of the crowding and of others proteins on these assemblies are unknown. Crystallisation of protein in vivo is a well-reported phenomenon[674], for instance, the crystals of protein naturally crystalising in vivo have been reported to be able to passively entrap enzymes and to protect them to thermal denaturation and proteolysis[675]. The crystallisation of CAHS-8 into these crystalline fibres in vivo would therefore serve as anti-desiccation shelters in which other proteins would be trapped and protected, may therefore be a reality and represent an alternative protection mechanism. In turn, as the tardigrade is rehydrated and water returns to the cytosol, these fibres would dissolve as observed (ch. 6 fig. 16 E).

In such a model, a coexistence of the gel state of CAHS-8 and of a crystal state would be theoretically possible. This was indeed observed and is presented in the following second part of the results of the thesis, which is dedicated to the study of the gel state of CAHS-8 and its properties.

5.3 Material and methods

Unless specified otherwise, the experiments were performed in a buffer containing 50 mM NaP, pH = 6.5, 500 mM NaCl.

5.3.1 Sample expression and purification

The gene sequence of CAHS-8, HD, NTDHD and CTDHD were optimized for *E.coli* BL21(DE3) and cloned into pET-28b(+) between the cleavage sites NcoI and XhoI with a purification tag made of 8 histidines in C-terminal for CAHS-8 and N-terminal for the others, respectively preceded and followed by a TEV cleavage site (ENLYFQG). The proteins were expressed in *E.coli* BL21(DE3) cells for 5 hours after induction at an optical density of 0.6 with 1 mM of isopropyl-β-d-thiogalactopyranoside. The unlabelled protein expression was performed in LB medium while isotopically labelled protein expression for NMR (^{15}N and ^{13}C) was performed in M9 minimal media prepared in H_2O or in D_2O to produce ^{2}H-labelled proteins.

The proteins were sonicated in 50 mM Tris, pH = 8.0 and 150 mM NaCl and then boiled at 95°C for 20 min for standard nickel beads purification. The elution was performed in the same buffer with 300 mM Imidazole, then the protein was incubated overnight with TEV protease while dialysing against the lysis buffer, subjected to a second nickel beads affinity column and the flow-through was loaded into a Superdex 75/200 column for size exclusion chromatography (SEC) in 30 mM NH_4HCO_3. The proteins were subsequently freeze-dried.

All concentrations were verified using a Thermo Fischer NanoDrop 2000.

The final sequences of the protein used after TEV cleavages are:

CAHS-8:

MSGRNVESHMERNEKVVVNNSGHADVKKQQQQVEHTEFTHTEVKAPLIHPAPPIISTGAAGL
AEEIVGQGFTASAARISGGTAEVHLQPSAAMTEEARRDQERYRQEQESIAKQQEREMEKKTEA

YRKTAEAEAEKIRKELEKQHARDVEFRKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEA
LERSRLATNVEVNFDSAAGHTVSGGTTVSTSDKMEIKRNENLYFQ

HD:

QSGAMTEEARRDQERYRQEQESIAKQQEREMEKKTEAYRKTAEAEAEKIRKELEKQHARDVE
FRKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEALERSRLATNV

NTDHD:

SGRNVESHMERNEKVVVNNSGHADVKKQQQQVEHTEFTHTEVKAPLIHPAPPIISTGAAGLAE
EIVGQGFTASAARISGGTAEVHLQPSAAMTEEARRDQERYRQEQESIAKQQEREMEKKTEAYR
KTAEAEAEKIRKELEKQHARDVEFRKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEALE
RSRLATNV

HDCTD:

CAAMTEEARRDQERYRQEQESIAKQQEREMEKKTEAYRKTAEAEAEKIRKELEKQHARDVEF
RKDLIESTIDRQKREVDLEAKMAKRELDREGQLAKEALERSRLATNVEVNFDSAAGHTVSGGT
TVSTSDKMEIKRN

5.3.2 Sequence alignment and coiled-coil prediction

The sequences of the protein were aligned using MUSCLE[676] via the EMBL-EBI website[677].

The sequences were taken from the supplementary materials of Fleming et al[678]. Out of the original 114 sequences of the alignment, 29 were taken out as they present significant deletions in the helical domain or in both termini.

The neural-network DeepCoil[679] was used to predict the probability of coiled-coil formation along the sequence.

5.3.3 Solution-state NMR

NMR experiments were measured on 600, 700, 850 and 950 MHz Bruker spectrometers in the buffer using 3 mm NMR tubes supplemented with 10% D_2O. The experiments with the HDCTD included 2 mM DTT.

5.3.3.1 Assignment

Backbone resonance assignment was performed using the following BEST-based triple resonance experiments[680,681].

CAHS-8 at 293 K: HNCA and HNCOCA, HNCO and HNCACO

CAHS-8 at 278 K: HNCA

CAHS-8 at 303 K: HNCA

HD at 293 K: HNCA and HNCOCA, HNCO and HNCACO

HD at 278 K: HNCA and HNCOCA

NMRPipe[682] was used to process the spectra and the assignment was manually performed using CCPNMR Analysis 2[683].

5.3.3.2 Secondary chemical shifts

The secondary chemical shifts were obtained by subtracting the experimental chemical shifts with the random-coil one obtained using the POTENCI method in order to take into account the temperature, pH and ionic force of the buffer[583,598,684,685].

When necessary, the chemical shifts were corrected for deuterium isotopic shift.

5.3.3.3 Relaxation measurement

The relaxation rate experiments on the CAHS-8 free form were measured at 350 µM protein concentration. The ^{15}N R_1 and {^{1}H}-^{15}N heteronuclear nOe rates were measured as described[686]. The R_1 experiments were measured with a set of delays of 0, 0.1552, 0.2716, 0.388, 0.5044, 0.6208, 0.7372, 0.9312 and 1.164 s with one delay being repeated for experimental error estimation and using a recycle delay of 1.5 s, while the saturation of the nOe was 8 s.

The ^{15}N $R_1\rho$ rates were measured as described using a spin lock of 1500 Hz[687] with a set of delays of 1, 20, 50, 70, 90, 120, 160 and 200 ms with, two delay delays being repeated for estimation of the experimental error and using a recycle delay of 1.5 s. The R_2 rates were determined from R_1 and $R_1\rho$ as described[688].

The ^{15}N–^{1}H CSA/DD transverse cross correlated rates (η_{xy}) were measured as described[689,690], using a cross-relaxation delay of 50 ms. All rates were extracted using in-house scripts with Monte Carlo simulation-based noise-estimation.

NMRPipe[682] was used to process all spectra and the software CCPNMR Analysis 2[683] was used to analyse the spectra.

The errors of the R_2R_1 product was calculated using the propagation of uncertainty formulas:

$$\sigma_{R_2R_1} = R_2R_1\sqrt{\left(\frac{\sigma_{R_1}}{R_1}\right)^2 + \left(\frac{\sigma_{R_2}}{R_2}\right)^2}$$

with σ denoting the standard deviation.

5.3.3.4 Chemical Shifts Perturbation

The Chemical Shifts Perturbation (CSP) was calculated for each residue using the formula

$$CSP = \sqrt{(\Delta\delta_{H^N})^2 + \left(\frac{\Delta\delta_N}{6.51}\right)^2}$$

with $\Delta\delta_{H^N}$ and $\Delta\delta_N$ referring to the difference between respectively the chemical shifts for the amide proton and amide nitrogen for the datasets to compare.

5.3.3.5 Paramagnetic Relaxation Enhancement

Cysteine mutations (S8C, A134C and A185C) were made by site directed mutagenesis using the quick change method[691]. Paramagnetic labelling was performed by attaching the 4-maleimido-TEMPO spin-label on the cysteine sidechain using maleimide chemistry. The protein was incubated overnight with 10 mM DTT, then dialysed against 50 mM NaP ,pH = 7.0, 150 mM NaCl. A 5 times molar excess of 4-maleimido-TEMPO dissolved in DMSO was added for overnight incubation and the excess of label was then removed through SEC.

Paramagnetic relaxation enhancement effects were determined from the peak intensity ratios of ^{15}N HSQC spectra of the paramagnetic state and the diamagnetic state, which was obtained by the addition of 5 mM ascorbic acid to the former. A recycle delay of 1.575s was used for both states.

5.3.3.6 Residual Dipolar Couplings

For residual dipolar couplings (RDCs) measurement, the protein was aligned using ASLA Biotech pf1 phages at 15 mg/mL concentration for CAHS-8 and 10 mg/mL for HD. The alignment was verified by measuring the quadrupolar splitting of D_2O. BEST HNCO-JNH

experiments[692] were measured in the isotropic and the phage aligned anisotropic states and the N-HN scalar couplings were obtained by measuring the line-splitting in the ^{13}C dimension using CCPN. The amide ^{15}N-^{1}H RDCs values were obtained by subtracting the couplings in the measurements. The spectra were zero-filled to 8192 points in the ^{13}C dimension to obtain a satisfactory digital resolution.

The prediction of alignment tensors and thus of RDCs from the crystal structure and their fit to the experimental RDC was performed using PALES[693].

5.3.3.7 Analysis of the relaxation rates

To perform model free analyses as well as spectral density mapping and fit of the rotational diffusion tensor, a python package called NMRFit was written.

The model-free analysis was performed using a Levenberg-Marquardt algorithms to fit the relaxation rates using a nonlinear least-squares approach for each residue i, minimizing the equation

$$\chi_i^2 = \sum_{n=1}^{k} \left[\frac{\left(R_{n,\mathrm{exp}} - R_{n,\mathrm{calc}}\right)}{\sigma_{n,\mathrm{exp}}} \right]^2$$

with n identifying the different residue-specific experimental rates available $R_{n,exp}$ (that is R_1, R_2, nOe and η_{xy}) at every field available and $\sigma_{n,exp}$ their associated experimental error. The calculated rates $R_{n,calc}$ were calculated using equations given in the introduction[694] and with a form of the spectral density function modelled with 2 motional modes for the helical domain (thus fitting τ_1, τ_2 and A_2) and 3 modes for the disordered regions (thus fitting $\tau_1, \tau_2, \tau_3, A_2$ and A_3)[686]. The fast motions timescale $\tau_1$was fixed to 50 ps and as $\sum_k A_k = 1$, then $A_1 = 1 - A_2$ for 2 modes and $A_1 = 1 - A_2 - A_3$ for 3 modes. Additionally, R_{ex} was also fitted using the definition $R_{2,eff} = R_{ex} + R_2$ with R_{ex} scaled to the field strength. Finally, the angle θ was fitted as well for the rates at 278 and

293 K but was fixed at 22.5° at 303 K due to the absence of experimental cross-correlated relaxation rate constants.

The errors of the fitted parameters were estimated using a noise-based Monte Carlo simulation with 50 iterations.

5.3.3.8 CPMG relaxation dispersion

^{15}N Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion were measured using a constant time relaxation delay of 32 ms and 14 CPMG frequencies (νCPMG) ranging from 31.25 to 1 kHz, with one being repeated to permit the estimation of the experimental intensity error. The proton decoupling field was 11 kHz. The error on the R_{2eff} were estimated using a Monte Carlo estimation.

The results were analysed using the program ChemEx[695] using a two-state exchange model, with, when necessary, the deuterium flag in the parametrisation.

5.3.4 Solid-state NMR

All solid-state experiments were performed at the Centre de Résonance Magnétique Nucléaire à Très Hauts Champs (CRMN, Lyon, France) with the collaboration of Dr. Tanguy Lamarchand and Dr. Guido Pintacuda.

5.3.4.1 Microcrystals

Approximately 3 mg of microcrystals of ^{13}C ^{15}N labelled HD were obtained by mixing HD at 7.5 mg/mL in a 1:1 ratio with a solution of 0.1 M Tris, 25 % isopropanol, 0.2 M ammonium acetate at pH = 8.5 in 50 mL Falcon tubes in a 1 L Becher sealed with parafilm, resulting in a pseudo sitting-drop set up. The microcrystals were grown for three days after which they were used to fill a 1.3 mm ssNMR rotor by ultra-centrifugation.

The proton-detected magic-angle spinning ssNMR ^{1}H-^{15}N hNH and ^{1}H-^{15}N-^{13}Ca hCANH experiments were measured on a 800 MHz spectrometer at 55 KHz MAS and at a sample temperature of approximately 20 °C, taking into account the rotor frictions.

5.3.4.2 Crystallised fibres

The crystallised fibres were prepared by dehydrating a 2 µL drop of ^{13}C ^{15}N CAHS-8 at concentration of 3.75 mg/mL in a relative humidity of 43 % using three 24-well crystallisation trays. After 5 weeks of growth, the spikes growing out of each drop were manually collected one by one in acetonitrile and packed into a 1.3 mm MAS rotor using ultracentrifugation.

The cross-polarisation based proton detected hNH spectra was measured on a 800 MHz Bruker spectrometer at 60 KHz MAS and at a sample temperature of approximately 293 K.

5.3.5 Dynamic Light Scattering

DLS measurements were performed using a Malvern Zetasizer Nano S using a back-scattering. The samples at the desired concentration were centrifugated at high temperature for 5 min and 150 µL were pipetted into a Malvern ZEN0040 cuvette. A temperature program ranging from 318 K to 278 K with three measurements every 5 K after 300 s of equilibration was used. Each measurement is composed of 10 runs of 60 s.

5.3.6 Small-angle X-Ray Scattering

SAXS experiments were performed at the BM29 BioSAXS beamline in the ESRF Grenoble in the buffer. The measurement used an incident beam wavelength of 0.99 Å and with a distance between the sample and the Pilatus3 2M detector of 2.867 m, giving a range of

the magnitude of the scattering vector q comprised between 0.0039 $Å^{-1}$ and 0.49 $Å^{-1}$. Every measurement comprised 10 frames which were analysed and rejected in case of radiation damage. The scattering of the solvent was measured before and after each sample and was subtracted from the sample's.

The data were processed and analysed using ScÅtter and PRIMUS. 10000 conformers were calculated using flexible-meccano[696] assuming the helix 1 (E95-F151) and helix 2 (V167-L194) were either populated or not. SAXS curves were calculated for each conformer using Crysol[697] and the ensemble selection algorithm ASTEROIDS[698] was used to calculate combinations of equally populated conformers that reproduced the SAXS experimental data. The ensembles were calculated with members ranging from 2 to 20 with 7 being estimated to be the best. The individual radius of gyration was calculated for each conformer and their average over the final best-fitting ensemble is the radius of gyration given in the results.

5.3.7 X-Ray diffraction crystallography

The microcrystals of HD were obtained by mixing the protein at 5 mg/mL in a 1:1 ratio to a solution of 0.1 M MES/Imidazole at pH = 6.5, 0.02M Sodium formate, 0.02M Ammonium acetate, 0.02M Sodium citrate tribasic dihydrate, 0.02M Potassium sodium tartrate tetrahydrate, 0.02M Sodium oxamate, 20% v/v Glycerol, 10% w/v PEG 4000. These conditions were initially identified using a Morpheus MD screen via the EMBL's High-throughput crystallisation (HTX) platform in Grenoble. The needle-shaped microcrystals were visible after 5 days and stopped growing after 3 weeks after which they were collected. They were supplemented with glycerol for cryoprotection, mounted on a mesh, flash-frozen in liquid nitrogen and their diffraction data were individually collected using the mesh collection mode on the beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) in collaboration with Dr. Max H. Nanao[699], using a wavelength of 0.873 Å, a temperature of 100 K and using a Pilatus3 2M detector.

The data of 51 microcrystals were integrated, scaled and merged using XDS and XSCALE[700]. Molecular replacement was performed using ARCIMBOLDO_LITE in coiled-coil mode[600,601] which uses PHASER[701] and SHELXE[702] as part of the CCP4 suite[703,704]. The model was built using Coot[705,706] and refined using PHENIX[707] at first and then REFMAC5[708]. The statistics were generated using MolProbity[709].

The structure was analysed using ChimeraX[710], PISA[711] and the coiled-coil analysis package samCC turbo[551,553].

5.3.8 Fibre diffraction

The crystallised fibres were prepared by dehydrating 2 µL to 4 µL drops of the proteins at a final protein concentration comprised between 1 and 10 mg/mL in hanging-drop crystallisation trays at varying relative humidity. The crystallised fibrils start to appear after 5 days to 1 weeks and take approximately 1 months to 6 weeks to fully grow.

For the fibres grown under the magnetic field, they were prepared using the optimal identified conditions which are 2 µL drops at RH = 43 % and a final protein concentration of 3.75 mg/mL. The trays were then placed under a 950 MHz Bruker NMR spectrometer for four weeks, at the level of one of its anti-vibration pods.

The crystallised fibres were manually collected using a large mounted crystal loop, soaked in glycerol and diffracted in helical collection mode at ID23-2 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) in collaboration with Dr. Max H. Nanao. A wavelength of 0.873 Å, a temperature of 100 K and, for the diffraction data presented in ch.6 fig. 13 and 14, a Pilatus3 2M detector while for those in ch.6 fig. 18, using a Dectris EIGER X 9M detector.

5.3.9 Atomic Force Microscopy

2 µL of CAHS-8 or HD at 1.2 mM were deposited on a mica plate, let dehydrate for 5 minutes and covered with 50 µL of 50 mM phosphate buffer and 500 mM of NaCl at pH = 6.5. AFM images were recorded in liquid mode on a Bruker multimode 8 microscope equipped with a Nanoscope V controller and a silicon cantilever nitride with characteristics k = 0.7 N/m, Fq = 150 kHz, r = 20 nm. The ScanAsyst mode was used, using a scan rate of 1 Hz and a setpoint of approximately 0.1 V.

The experiments were performed at the Institut de Biologie Structurale, Grenoble, France, in collaboration with Jean-Marie Teulon and Dr. Jean-Luc Pellequer.

The images were processed using Gwyddion[712] using a mean-plane subtraction levelling, a median method alignment of the rows, correction of the strokes and with the minimal height set to 0. The Laplacian-weight method treatment[516,517] of some of the CAHS-8 images was performed by Dr. Jean-Luc Pellequer and Dr. Shu-wen W. Chen.

5.3.10 Electron Microscopy

CAHS-8 was incubated 2 days in the fridge at 20 mg/mL. On the day of the imaging, the protein was cross-linked using 0.18 % v:v Glutaraldehyde (Sigma) for 2 min. The reaction was then stopped using 100 mM Tris at pH = 6.5.

5.3.10.1 Negative-stain Transmission Electron Microscopy

The cross-linked protein sample were applied to the clean side of a carbon grid on mica and then stained using 2% v:v of sodium silicotungstate (SST) at pH = 7. The electron micrographs were taken using a FEI Tecnai T12 microcscope operated at a tension of 120 kV and equipped with a GATAN Orium 1000 camera. The magnification was set to 23 000

times. The experiments were performed at the Institut de Biologie Structurale in collaboration with Emmi Mikkola.

5.3.10.2 Cryo Transmission Electron Microscopy

The cross-linked sample were pipetted on a grid and vitrified using a Vitrobot. The microcrraphs were acquired using a Thermofischer Glacios microscope with a 200 kV operating tension and equipped with a K2 direct electron detector at the Institut de Biologie Structurale, Grenoble, France by Dr. Guy Schoehn.

The micrographs were aligned using MotionCor2[713] and their contrast transfer function corrected using Gctf[714]. The picking and two dimensional classification was attempted in both Relion[715,716] and cryoSPARC[717]. The presented 2D classes were obtained using Relion and particles picked by Dr. Guy Schoehn.

5.3.11 Manipulation of the fibrillation

5.3.11.1 TFE

A volume of 2,2,2-Trifluoroethanol (TFE, Sigma) was pipetted into the 1 mg/mL protein sample in order to reach the final desired v:v ratio, here 5%.

5.3.11.2 TMAO

Trimethylamine N-oxide (TMAO, Sigma) was weighted and resuspended in the buffer used at the desired concentration. The protein samples were deposed on the AFM imaging surface and diluted to the final desired concentration using the buffer containing the TMAO concentration.

In the case of the fibrous gel, the liquid AFM imaging cell was filled with the buffer containing 3.5 M of TMAO and the concentrated protein was pipetted in, resulting in a dilution to the final 300 μg/mL protein concentration.

5.3.11.3 Solvent PREs

Gadodiamide (Gd(DTPA-BMA), Sigma) was weighted and resuspended in the buffer at a high concentration and pipetted to the sample for a final concentration of 3 mM.

Part 2: CAHS-8 – Its gels and their properties

Chapter 7: The gels of CAHS-8

« Marco Polo describes a bridge, stone by stone.

“But which is the stone that supports the bridge?” Kublai Khan asks.

“The bridge is not supported by one stone or another,” Marco answers, “but by the line of the arch that they form.”

Kublai Khan remains silent, reflecting. Then he adds: “Why do you speak to me of the stones? It is only the arch that matters to me.”

Polo answers: “Without stones there is no arch.” »

Italo Calvino

Invisible Cities, 1972

6.1 Introduction

6.1.1 Gels

6.1.1.1 Gels and gelation

A gel is a soft solid defined as a macroscopic three-dimensional network of individual chains connected either by covalent bonds or by non-covalent interactions in nodes[718]. The first case is called a chemical gel and typically results from the branching polymerisation of a chemical or the crosslinking of the individual units using reaction such as vulcanisation, and is therefore irreversible[662,718]. The second case is called a physical gel, typically results from non-covalent interactions such as the hydrophobic effect, π-π or electrostatics interactions or simply physical entanglement, and is therefore reversible as these non-covalent interactions can be easily disrupted, typically by heating[718].

In turn, gelation is the constitution of this three-dimensional gel network from a solution and is therefore a phase transition from a liquid to a solid state. Historically, chemical gelation via polymerisation has been described by the Flory-Stockmayer model, permitting to predict the degree of polymerisation necessary for a macroscopic network with observable viscoelastic properties to gelate[662,719,720]. The model was subsequently generalised as a percolation phenomenon, taking into account the presence of solvent molecules as well as the interaction between monomers in adjacent sites, with the gelation point being the critical threshold of bond occupancy of the sites above which the system is percolated, i.e. where it is a gel[440,718,721,722]. As such, below the probability occupancy of this gel point and although the system is in a liquid state, local occupied junctions are still present because the gelation is a gradual process[440]. Nonetheless, the Flory-Stockmayer model and the percolation model are not qualitatively different, with the former being a mean-field picture of the percolation model on a Bethe lattice tree and presenting the advantage of being simple and somewhat accurate[440,718,721].

Because of the nature of the gelation process, that is the random formation of junctions by chains which may themselves result from aggregation, gels are often inhomogeneous which is particularly the case of physical gels as they do not arise from the interaction of

unique groups but from multiple different possible junctions, or from simply result from sheer entanglement. As such, gel meshes present spatial inhomogeneities due to variations of the local density of the junctions which result in pores of different dimensions, connectivity inhomogeneities of the pores which may be either connected to each other or isolated, and topological inhomogeneities of the freedom of chains, their entanglement and distribution[723–725].

6.1.1.1.1 Characterisation of the gelation

Gels are non-ergodic media due to their porous nature. Indeed, molecules physically confined in these cavities can only diffuse in a restricted part of the spatial configuration of the system. As such, the gel point can be characterised by the loss of ergodicity using techniques sensitive to this property like Dynamic Light Scattering (DLS). Indeed, as pointed out in chapter 2, in a non-ergodic medium the normalised time-averaged intensity correlation function $g^{(2)}(\tau)$ measured by DLS is not equal to the ensemble-average of the same correlation function but only to particular sub-ensembles due to the above-mentioned spatial inhomogeneities, which translates into a suppression of the initial $g^{(2)}(0)$ value[723,726,727].

In term of scattering, this is due to the oscillations of these static scattering centres remaining within the probed length scale by the DLS experiment, that is the inverse of the scattering vector i.e. q^{-1}, and therefore do not trigger detectable changes in the scattering pattern, thus does not resulting in a loss of correlation. Thus, considering gels, this means that the extent of this suppression effect of the non-ergodicity on the scattering depends mainly on the stiffness of the chains rather than the amount of inhomogeneity itself, with gels composed of rigid fibres having been noted to exhibit marked non-ergodic effects[726,728,729]. As a consequence, the observation of a significant loss of ergodicity by DLS has therefore been used successfully to characterise the gel point of chemical gels or thermoreversible physical gels[723,730,731].

Besides this initial suppression of the normalised DLS intensity correlation function $g^{(2)}(\tau)$, the actual shape of its decay as the system approaches the gel point has been

observed to progressively diverge from a simple exponential decay, gaining first a stretched exponential tail which transitions to a power-law at the gel point, yielding[730,732]

$$g^{(2)}(\tau) - 1 = \sigma_I^2 \left[A \exp(-D_{\text{fast}} q^2 \tau) + (1 - A) \left(1 + \frac{\tau}{\tau^*}\right)^{\left(\frac{n-1}{2}\right)} \right]^2 \qquad 6.1$$

with σ_I^2 being the mean-square intensity fluctuation term defined in chapter 2, which corresponds to the asymptotic value of the field correlation function at infinite time $g^{(1)}(q, \infty)$ and thus, is the contribution of the non-ergodicity to the scattering which is therefore the initial suppressed value measured, i.e. $g^{(2)}(0) - 1$; A and D_{fast} respectively the fraction of the sample engaged in the fast apparent translational diffusion mode and its diffusion coefficient, q the magnitude of the scattering vector, τ^* the time at which the power-law manifests in the decay and n the fractal dimension of the scattered photons which is a rheological coefficient related to its viscoelastic properties although the exact relationship between the two appears to depend on the nature of the gel[730,732].

This power-law tail is sensitive to the difference in timescales of translational diffusion of the molecules engaged in the gel network and the ones which are not. Indeed, the latter cause a normal exponential decay of the intensity correlation function, while the former becomes increasingly slower, as more joints are formed near the chains which yields a stretched exponential first and then, once the system is percolated, that is when the junction occupancy reaches the critical point gelation point, the diffusion of such species is so slow that it is essentially independent of the time interval probed by DLS, which is the definition of a power law[732].

As such, the apparition of such power-law behaviour in the DLS correlograms is commonly used to characterise the gel point of different gelating systems[730,733–737]. However, beyond the gel point, the contrast between the initial exponential decay caused by the diffusion of the fast mode and the power-law tail is reduced due to a potential retardation of the diffusion in the gel pores as well as the inhomogeneities of the pores causing irregularities in the scattering as well as loss of ergodicity resulting in the obtained correlograms only describing a sub-ensemble of the system[726,732].

Another method commonly used to characterise gelation is Differential Scanning Calorimetry (DSC), a technique which measures the difference in either heat flux or power compensation between a sample and a reference in response to variation of the temperature due to differences in heat capacity[738,739]. As such, when the crucibles containing the sample and the reference are heated or cooled, the difference is constant if the sample in in a steady-state, yielding the baseline of the measurement. However, if a thermodynamic transition that changes the heat capacity happens in the sample in response to the variation of the temperature, the equilibrium breaks down and the difference in signal with the reference varies, yielding a peak that can be either endothermic or exothermic depending on the transition, with the integration of the peak corresponding to the total change in enthalpy[739]. As both the gelation and the melting of thermoreversible hydrogels are first-order transitions, i.e. they involve an absorption or a release of heat unlike second-order transition which do not, they yield peaks in DSC and because, by definition, the solvent is the major component of the sample, the baseline after the transition remains essentially the same as before[738,739].

6.1.1.2 Hydrogels, water and dry gels

Hydrogels are defined as a hydrophilic three-dimensional mesh, which implies the presence of pores that are permeable to water and in which the *hydrogel* can retain a relatively large volume of water with respect to the actual dry weight of gel-forming material. As such, depending on the properties of the chains, hydrogels have been demonstrated to be able to hold up to a thousand times as much water as their dry weight[740]. This manifests physically as a swelling whose extent depends on the interactions between the hydrophilic groups of the chains and the solvent, its properties, the number of nodes and their strengths as well as the temperature and other related parameters[718]. As such, gels display a viscoelastic behaviour, and in particular a rubber elasticity, with this swelling being counterbalanced by the retractive force of the hydrogel network[662,718,740]. As the reversible junctions of physical gels are weaker than the permanent ones of chemical gels, the former typically have lower viscoelastic properties than the latter and may even rupture if the swelling breaks their bonds[740,741].

Several populations of water therefore coexist in a hydrogel, namely the bound water in the primary and secondary hydration shells of the hydrophilic chains and the bulk interstitial water present in the pores of the hydrogel without being directly bound to it, as well as possible hydrophobic solvation water in the case where the chains also have accessible apolar groups[502,740,742]. The diffusion in such an environment is discussed in the section 6.1.2 hereafter.

The hydrogel can dry and therefore unswell, resulting in what is called a *xerogel* (from ancient Greek ξηρός, that is *xērós*, meaning dry), or alternatively the water can be extracted under supercritical or freeze-drying conditions and replaced by a gas, which prevents the shrinkage and results in an *aerogel* (also sometimes referred to as a *cryogel* if obtained via freeze-drying). As a consequence, aerogels are more porous, less dense and physically more bulky than xerogels[743,744].

6.1.1.3 Gels, fibres, crystals and glasses

Due to the relatively simple requirements of chains and junction occupancy, gels can be formed by many species such as polymers, colloidal clusters or protein. Common examples of respectively a polymer and a protein gel are the ubiquitous agarose gels found in most wet labs and gelatin, which is derived from collagen. The two gels present some similarities, with both agarose and gelatin being disordered and flexible at high temperature but transition respectively to paired-helical fibres of aggregated molecules and into characteristic collagen triple-helical fibrils around 30 °C, which eventually yield physical gels due to interactions or bridging of fibres as well as entanglement of the disordered parts. These gels can be melted if heated, as the heat breaks and denatures the helical structures, but nevertheless reform if cooled again[718,745,746].

As briefly discussed in the previous chapter, polymers can crystallise, in which case local segments of different individual polymers form aligned and ordered layers which therefore coexist with disordered, amorphous regions[648,747]. As such, the gelation of

certain polymers can be triggered by their crystallisation, with these crystalline segments representing the junctions of the gels while the flexible part of the polymers are the chains constituting the network[718,748]. This is for instance the case of the widely used Poly(vinylchloride) (PVC) plastic where the crystallisation of portions of its chain is induced by cycles of heating and cooling, yielding a gel[749].

Finally, gels can present some similarities with glasses. First, gels with static inhomogeneities are non-ergodic and so are glasses. Secondly, the correlation functions of supercooled liquids near the glass transition temperature are characterised by two dynamic modes, the so-called β fast regime and the α slow regime, which can be described together using a power-law form as per the mode-coupling theory (MCT)[750–752]. This is reminiscent of the shape of DLS intensity correlation function at the gel point previously described with the similarities between the two phenomena and interplay between the MCT and percolation theory having been noted and discussed in many articles[730,751–757].

6.1.2 Dynamics in complex environments

As hydrogels are a mesh with more or less large and connected cavities which retain solvent, the pores of hydrogels are viscous environments that can physically confine the molecules entrapped in them and possibly induce a molecular crowding if the concentration is high enough within these cavities. These different, yet related, phenomena have different effects on these encapsulated molecules.

6.1.2.1 Shear and bulk viscosities

Liquids have an intrinsic internal resistance to any flow or shear, meaning that they experience a drag force termed viscosity due their frictional properties. The shear viscosity η (or μ), also called absolute or dynamic viscosity, is therefore the tangential force per area unit necessary to slide one layer of molecules over another and thus directly depends on the resistance to this sliding due to frictions between the layers.

Consequently, as a force per time, its S.I. unit are $\frac{\mathrm{N}}{\mathrm{m}^2} \cdot \mathrm{s}$, which is equivalent to $\mathrm{Pa} \cdot \mathrm{s}$. As liquids are nearly incompressible due to the close packing of the molecules, i.e. that their freedom of motion is limited by the many neighbouring molecules, any transport of momentum in a liquid is the result of the influence of individual molecules on each other which thus explains why the viscosity of liquids decreases as the temperature increases.

This shear viscosity is to be differentiated from the bulk viscosity which is the relaxation time of the molecular arrangement of the fluid after a perturbation and is therefore often assumed to be null and neglected for typical solvent as they reach their equilibrium very fast[758]. However, it is particularly high in hydrogels due to their viscoelastic rheological properties, in particular in semi-rigid hydrogels which are able to sustain significant deformation before relaxing[759–761].

As discussed in the chapter 2, the constant collision of spherical objects of hydrodynamic radius R_h with the molecules of the solvent cause them to experience a viscous drag opposing both transport and reorientation driven by the thermal energy. As such, the viscosity is an important parameter of both translational and rotational self-diffusion, being the solvent contribution to the Stokes friction coefficients $\xi_T = 6\pi\eta R_\mathrm{h}$ for the former and $\xi_R = 8\pi\eta {R_\mathrm{h}}^3$ for the latter which appear in the different treatment of diffusion. Thus, the translational D_T and rotational D_R diffusion coefficients are simply the available thermal energy over this friction coefficients, being respectively $D_T = \frac{k_B T}{6\pi\eta R_\mathrm{h}}$ which is the Stokes-Einstein equation, and $D_\mathrm{R} = \frac{k_B T}{8\pi\eta {R_\mathrm{h}}^3}$ which is the Stokes-Einstein-Debye equation[434,435,438]. The term ξ_T also appears explicitly in the Langevin equation[437] and indirectly through D_T in the Smoluchowski equation[436]. However, all these different equations describe a normal diffusion, i.e. that the probability of displacement or reorientation follows a Gaussian distribution which implies that diffusion is a Markovian process, i.e. that the probability of a diffusion motion is independent from previous motions, and are valid as long as the objects do not perturb each other's hydrodynamic radii R_h, both conditions not being always the case as discussed in the following section[435,438,440,441,762,763].

6.1.2.2 Crowding and confinement

Macromolecular *crowding* is generally defined as the state of a solution in which macromolecules constitute a non-negligible percentage of the volume fraction of the solution.

Before discussing crowding *per se*, it is important to note that crowding increases the shear viscosity of the solution as well, which is why crowders are also referred to as *viscogens*. Indeed, any translation of the chain of crowder polymer requires cooperative motions of all the connected segments, creating a resistance to any flow. This is the definition of shear viscosity and therefore explains its fast increase with the concentration of polymer. As this dynamic viscosity increases, the bulk viscosity of the solution also becomes progressively less negligible, but at a much slower rate than the dynamic viscosity because those cooperative movements are locally limited near the polymer chain, unlike the necessary cooperative motions of a gel network[719,759,764].

Additionally, when observing the dynamics of proteins in crowded solutions, it is necessary to distinguish the experienced viscosity, termed *nano-viscosity*, from the overall shear viscosity of the solution. Indeed, the viscosity felt by a probe depends on its radius of gyration $R_{g,probe}$ and that of the crowder $R_{g,crowder}$, which implies several cases. In the extreme limit where $R_{g,probe} \ll R_{g,crowder}$, the probe experiences a viscosity close to that of the pure solvent without any crowder, while in the other extreme limit $R_{g,probe} \gg R_{g,crowder}$, the nano-viscosity affecting the probe is approaches the macroscopic shear viscosity of the solution. This can be intuitively pictured as a probe much smaller than the crowder may diffuse almost unimpacted by the much larger crowder species that leaves plenty of free space due their mutual excluded volume while on the other case, the probe may be effectively surrounded by crowders that must be therefore displaced to permit a diffusion and thus causes significant frictions, which is therefore similar to the definition of the shear viscosity of the solution. In turn, when the two radii of gyration are close, the nano-viscosity experienced by the probe depends on their exact ratio as well as on the correlation-length of the crowder polymer and on the interactions between the species as

expressed in an equation developed by the Hołyst group[440,765–768]. This has important consequences as crowders can be small molecules like glycerol, large proteins like BSA or rod-shaped fibres long of several micrometres, which therefore have different effects on the dynamics and conformational exchange of proteins, depending on the ratio of their size[769].

As per crowding, two cases should be considered separately: the case where crowder species are mobile, for instance soluble polymers such as PEG or dextran added to a solution of protein and which therefore diffuse alongside them, and the case were the crowder species can be considered spatially fixed, for example the network constituting a hydrogel.

6.1.2.2.1 Mobile crowders

In the case of mobile polymer crowders, three regimes of crowding can be separated, namely the dilute, the semi-dilute and the concentrated regime[440,764].

In the dilute regime, the polymers are considered coiled and to behave as hard sphere, freely diffusing without impinging on each others hydrodynamic radii, there isn't significant interactions between them and the behaviour of individual chain can be described independently. However, upon increase of their concentration, the excluded volume of the polymers progressively overlaps, yielding local entanglement of chains which is defined as the semi-dilute regime. Yet, despite this partial overlap, the concentration of the polymer is still considered low as bulk solvent is still significantly present between the chains but upon further increase of the concentration, the system enters the concentrated regime in which the different individual polymers are densely packed and extensively entangled with each other[440,764].

In turn, the increase of the concentration of polymer crowders, which may also be described in term of volume fraction, results in an increase of the shear viscosity as described previously, with the transition to semi-dilute regime and the concentrated

regime resulting in significant increases due to the entangled chains providing even further resistance to shear flow than isolated chains[440]. Consequently, this increase of the viscosity results in slow down of the diffusion of the crowded species following the length-scale effects previously described. Besides this slowdown of the diffusion, the probability of displacement of other species is also affected by the crowding. This is particularly visible when observing the translational mean-square displacement $\langle r^2(t)\rangle$ which does not increase linearly with the time t anymore as in normal diffusion as per $\langle r^2(t)\rangle = 6D_T\tau$ in three-dimensions, but show progressively smaller displacement over time than expected, which is called sub-diffusion, and may arise due to the presence of the crowder inducing either a geometrical restriction on the diffusion or altering the probability of displacement which becomes non-Gaussian, that is anomalous[770–772].

For example, the diffusion of fluorescent beads was shown by X-ray Photon- and Fluorescence- Correlation Spectroscopy (XPCS and FCS) as well as Single Particle Tracking (SPT) to be slowed down and sub-diffusive in the semi-dilute regime of PEG if the mesh size induced by the polymer entanglement if smaller than the bead diameter[772]. However, besides the volume fraction, the shape of the crowder is also a critical parameter. For instance, the translational diffusion of a globular protein was shown using FCS and Brownian dynamics simulations to be reduced by 60% by a volume fraction of 5% of dsDNA, which has a cylindrical shape, while the reduction of the diffusion is only 10% at the same volume fraction for Ficoll70 which has a spherical shape, highlighting the importance of the actual excluded volume in the crowding[773].

Crowders can also perturb the conformational sampling of IDPs but the effects depend on the IDP as well as on the crowder properties, size and volume fraction. Indeed, some IDPs have been show to experience a compaction of their conformational sampling upon crowding, thereby minimising their excluded volume and may collapse at high concentration of crowder as shown using single-molecule Förster resonance energy transfer spectroscopy (smFRET) and predicted from simulation[774,775]. Alternatively, the D-Rec1 IDP has been demonstrated using Small-Angle X-ray Scattering (SAXS) as well as small- and ultra-small angle Neutron- Scattering (SANS and USANS) to experience

different conformational changes depending on the nature of the polymer crowder and even its volume fraction, sampling more extended conformational ensemble, apparently due to soft attractive interactions with polymer crowder at low volume fraction, but at higher concentration of the same crowder, the protein was shown to contract[776].

Adamski and co-workers from the Blackledge group performed an extensive study of the effects of crowding on the backbone dynamics of two IDPs using NMR relaxation, therefore probing the reorientational rotational diffusion dynamics of the proteins, as a function of the concentration of dextran40, a polymer crowder, by quantifying the increase in nano-viscosity via measurement of the longitudinal relaxation rate of water protons[439]. Using the $k = 3$ dynamic modes spectral density function described in chapter 2, they defined their associated timescales as a function of the frictions with the solvent as

$$\tau_k(C) = \tau_{k,0}(\varepsilon_k \rho(C) + 1)$$

for each component k, with $\tau_{k,0}$ being the initial correlation time without any crowder, ε_k the experienced friction coefficient by the spins and $\rho(C)$ the solvent friction coefficient depending on the concentration C of crowder. This last term was determined beforehand as $\rho(C) = (\eta_C - \eta_0)/\eta_0 = \left(R_{1,C} - R_{1,0}\right)/R_{1,0}$, with η the viscosity, R_1 the longitudinal relaxation rate of water protons and the indices C and 0 denoting respectively these rates were measured at a concentration C of crowder and without any crowder[439].

They demonstrated that in the dilute regime of dextran40, that is up to 145 mg/mL, the nanoseconds timescales of segmental chain motions and backbone sampling motions of the IDPs increased linearly with the friction coefficients ε quantified from the nano-viscosity, with ε of the former being approximately 3 times bigger than for the latter which indicate a stronger dependence on the viscosity, while picosecond timescale vibrational and librational motions didn't show a detectable dependence on the nano-viscosity. In turn, no significant change of their amplitude was observed which indicate altogether that this crowding does not affect much the conformational sampling of the backbone of these IDPs. Moreover, these results indicate that fundamental length-scale of nano-viscosity effects as detected by NMR relaxation depend on the dimensions of the segment of the molecule. In the case of single peptide motions, which are on the order of the size of the

solvent molecule, the coupling to the water is lower than multi-peptide segments, which closely follows what the model developed by the Hołyst group predicts. In turn, at 195 mg/mL of dextran40, the linear dependence of these motions on the viscosity breaks down, indicating the transition to the semi-dilute regime in which the protein hydrodynamic radius R_h is perturbed by the polymer crowder[439,766,768].

Moreover, in the dilute and semi-dilute regime, polymer crowders exert a particular effect called depletion interaction on crowded species, causing an attractive force between neighbouring monomers[777–780]. Polymers cannot freely diffuse in the vicinity of these crowded species due to their excluded volume, resulting in a zone around these particles referred to as the *depletion layer* in which the local concentration of polymer is lower than in the bulk to minimise losses of configurational entropy from the polymers. In turn, if several of these crowded species are spatially close to each other, their depletion layers overlap and polymers concentrate around them which eventually results in an attractive force between these crowded particles due to osmotic pressure[777–780].

Interestingly, it was noted already in the first model of the depletion interaction by Asakura and Oosawa that rod-shaped objects, such as fibres, are more efficient depletants than spherical ones of the same volume[777,778].

Additionally, this depletion interaction results in a somewhat disconcerting effect: interactions can be favoured by crowding, despite the increase in viscosity. Indeed, it was demonstrated using FCS and fluorescence anisotropy that the k_{on} of the interaction between two globular proteins was slowed down by small crowders such as glycerol and sucrose as well as PEG in the dilute regime but was increased in the semi-dilute of PEG of different sizes which was attributed to the depletion interaction, before significantly slowing down in the concentrated PEG regime[781].

Similarly, Zosel and co-worker from the Schuler team showed using smFRET and FCS that the affinity of the interaction of two folding-upon-binding IDPs increased with the size of the PEG crowder, which was explained by the depletion interaction as the bigger the crowder, the bigger the induced depletion layer which essentially represent cavities larger than the proteins in which they experience a nano-viscosity close to that of the pure

solvent and thus translationally diffuse accordingly, with the attractive depletion interaction driving an increase of the affinity[782]. Additionally, after the initial increase of the k_{on} with the concentration of PEG, they also observed a decrease of the k_{on} at high concentration of large PEG species, which was attributed to the viscosity-driven slow-down of the diffusion overcoming the accelerating effect of the depletion interaction as per the recent Berezhkovskii-Szabo theory which combines these two effects[782,783]. In this model describing the interaction of proteins in crowded environment, the retardation of the diffusion is explained by collision of the proteins with the crowder and nonspecific reversible binding between them while the short-range attractive depletion interaction is modelled as a simple distance dependent square-well potential[783].

6.1.2.2.1.1 Self-crowding

First, for folded proteins, increasingly solution of self-crowded globular protein of different sizes and interaction properties shows a scaling of their translational diffusion measured on the millisecond timescale with the macro-viscosity as observed by NMR, SAXS and FCS[784]. However their rotational diffusion was shown to approach that of the pure solvent if the proteins were weakly interacting and globular, while stronger interactions or non-globular shapes resulted in a correlated tumbling with neighbouring proteins, indicating an experienced nano-viscosity coupled to the macroscopic shear viscosity of the solution[784].

The self-crowding of proteins at high monodisperse concentrations is conceptually similar to a polymer-induced crowding, especially for IDPs which behave similarly to polymers, which allowed several concepts and models from polymer physics to be applied to IDPs, although they show more complex effects than homo-polymers due to the variations of their amino-acid sequence[439,16,603,604,785].

In particular, the polymer-like liquid-liquid phase separating (LLPS) behaviour of IDPs results in the formation of highly concentrated condensates, in which the concentration

can reach several hundreds of mg/mL and therefore represent extreme but particular cases of self-crowding. Indeed, Brady and co-workers from the Kay and Forman-Kay groups studied the diffusion and backbone dynamics of an IDP forming LLPS condensate by NMR and observed a translational diffusion 100 time slower than in its dilute phase, as well as ^{15}N transverse relaxation rates approximately 4 times faster on average[786]. They further compared a non-separating mutant of the IDP concentrated to similar concentration to that in the condensate, which showed approximately 1.5-times slower relaxation rates than in the condensate, demonstrating that the slowdown was not only viscosity driven but also had a contribution attributed to intermolecular contacts[786].

In turn, Guseva, Schnapka and co-workers from the Blackledge group compared experimental NMR relaxation data of another IDP in LLPS condensates or with different concentration of PEG crowder as well as with MD simulation of boxes with different concentration of the IDP to derive insights on the dynamics in the droplets[604]. They demonstrated that like in the crowder, the different dynamic modes sampled by the protein backbone underwent a significant increase of their associated timescales of motions however unlike in the crowder, their associated amplitude of motions were significantly altered, with a restriction of the librational motions suggesting an partial entanglement of the IDPs chains[604].

6.1.2.2.2 Static crowders and confinement

In this case, the crowders are considered fixed compared to the timescale of diffusion of the considered species.

Such static crowders, like the network of a hydrogel or anchored membrane protein in lipids bilayers, act as barriers that prevent the diffusion of other molecules due to their excluded volume, effectively entrapping them in delimited pores. As such, these encapsulated species experience a confined diffusion in these cavities. However, depending on the considered timescale and the actual crowding within the pores, the translational diffusion of these entrapped molecules may or may not appear anomalous. Indeed, if the pores are large enough and somewhat not significantly crowded, the

molecules may be observed to diffuse normally at short timescales, but long enough to not observe ballistic motions, with the anomalous effect of confinement becoming visible if the timescales of diffusion are long enough for the molecules to encounter the uncrossable walls of the pores[787–790].

In turn, the pores can be crowded due to remaining gel-forming species not engaged in the network or due to the self-crowding of the observed species resulting from its own concentration or the pore dimensions, which therefore constitute mobile crowders as described previously. As such, due to the possible relatively low crowding in the pores or the size difference between the entrapped species and the mobile crowder present in the pores, the apparent diffusion of entrapped species in hydrogel pores is slowed-down but not as much as what would be expected considering the increased bulk viscosity of a gel, highlighting the difference between its macro-viscosity and the actual experienced nano-viscosity in its pore[761,768,791]. However, the walls of the pore may not be inert and interactions between them and the entrapped species can further perturb their rotational and translational diffusion.

Hydrogels by definition, are permeable to water but are able to retain significant amount of it thanks to their accessible hydrophilic groups and as such, perturb the solvent dynamics[761,792]. Molecular Dynamics (MD) on PEG hydrogels at different densities predicted that the diffusion of water is slowed-down as a function of increasing density of the gel junctions, which resulted from a gradual decrease of the number of hydrogen-bond per water molecule as well as an increase of their lifetime, with effectively no bulk water at the highest gel density[793]. This was also demonstrated using as two dimensional Fourier-Transform Infrared (2D FT-IR) as well as polarization-selective pump-probe (PSPP) ultrafast spectroscopies in polyacrylamide hydrogel[794], however combination of MD, PPSP and 2D FT-IR on solutions of acrylamide or polyacrylamide as well as on polyacrylamide gels shows that these effects arise from disruptive effects on the hydrogen-bond network on the femtosecond timescale caused by the presence of soluble acrylamide monomers due to a reduction of the available hydrogen-bonding water molecule partner[795]. This indicates that the retardation of water dynamics may be due to interactions with the available hydrophilic groups present in the hydrogel rather than effect of confinement which can manifest similarly[795,796].

The translational and rotational diffusion of 400 nm large spindle-shaped nanoparticles has been shown by Mössbauer spectroscopy to be mechanically restricted in dense polyacrylamide hydrogels as a function of the mesh size, with even pores larger than the particles significantly restricting it already[797]. Similar slow-down of diffusion of different µm-sized particles has been shown by SPT to be correlated to the density of mucin hydrogels but also to the electrostatic interactions between the probe and the gel network, which can act either as a strong or permeable barrier depending on their charge[798]. In turn, the use of variable-length-scale FCS to study the diffusion of large fluorescent beads in agarose gel reveal an anomalous diffusion due to confinement in the pores[771,799].

Interestingly, there is another overlap between these two categories of crowders, that is mobile and fixed. Indeed, in the semi-dilute regime of polymers, the local entanglement may cause the formation of local transient mesh networks that can trap proteins inside and act as static barrier, even if only transiently. In turn, the increase of the concentration reduces the sizes of these cavities until the concentrated regime is reached[764].

6.2 Results

6.2.1 The fibrous hydrogel and xerogel

6.2.1.1 The entangled fibrils form a gel

As discussed in the previous chapters, CAHS-8 self-associates into coiled-coils to form fibrils and fibres in solution. The intertwinement of the fibres forms a reticular network that can be observed by negative-stain transmission electron microscopy (fig. 1 A, B) and by AFM (fig. 1 C), resulting in an apparent hydrogel where the fibres crossing each other form nodes while they are themselves the chains as per the percolation picture of gelation[718,800]. The drying at room-temperature of this hydrogel results in a xerogel[743] which can also be observed by AFM (fig. 1 D). All these images show that the entanglement of the fibres results in empty regions between fibre bundles which form the pores of the gel that are particularly visible in the xerogel (fig. 1 D).

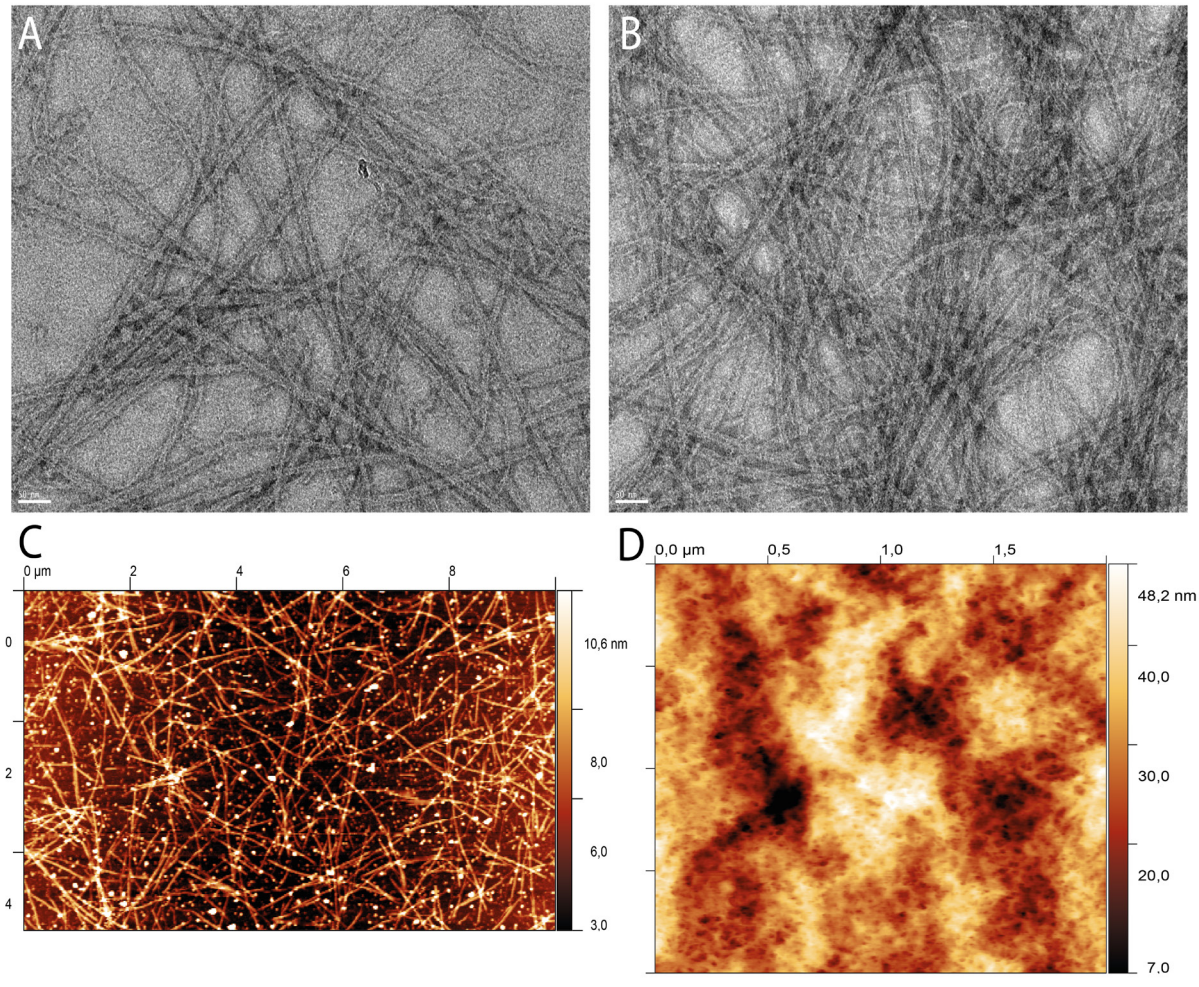


Figure 7.1: The fibrous gels of CAHS-8 observed by EM and AFM.
A and **B**: Negative-stain transmission electron microscopy micrographs of the entangled fibres of CAHS-8 forming a gel.
C: Atomic Force Microscopy image in solution of the entangled fibres of CAHS-8 forming a hydrogel.
D: Atomic Force Microscopy image in air of the dried fibres of CAHS-8 forming a xerogel.

As the fibrillation of CAHS-8 was observed to depend on the concentration and temperature, systematic measurements of Dynamic Light Scattering (DLS) correlograms were acquired at 10 different concentrations ranging from 5 mg/mL to 30 mg/mL, each from 45 °C to 5°C every 5 °C (fig. 2 A). Far from the gel point, the normalised time-dependent intensity correlation function $g_2(t)$ shows a smooth exponential decay, however when CAHS-8 approaches the gel point, sizeable fibres or local gel patches they form manifest by spikes visible in the tens of millisecond timescale as such species are inhomogeneous and different from the ensemble-average of the system, as well as by a stretched exponential tail in the decay of the function[726,730,732]. In turn, when the system reaches the gel point and is percolated, i.e. when CAHS-8 gelates, the loss of ergodicity due to the inhomogeneities and physical confinement in the pores manifests as a suppression of the $g_2(0)$ value, while the oscillations of the fibres constituting the gel network become so slow that they do not result in a loss of correlation in the time interval probed by DLS, which manifests as the apparition of a power-law tail in the correlogram which is, as discussed in the introduction of this part, the strongest at the gel point and immediately after it[726,730,732]. The measurements at 30 mg/mL and 298 K clearly show this features, which permitted its fitting of the correlogram to the equation of Martin, Wilcoxon and Odinek to characterise the gel point[730,732] (fig. 2 B, reduced $\chi^2 = 6.76 \times 10^{-5}$), yielding a fast diffusion coefficient of the free species of 2.41×10^{-10} $m^2 s^{-1}$, with these free species accounting for 11 % of the scattering which has a fractal exponent of 0.83.

These features were therefore used to draw a phase diagram of CAHS-8 (fig. 2 B), with the three shapes of the intensity correlation function described being termed respectively the soluble region, the slight gelation region and the gel region.

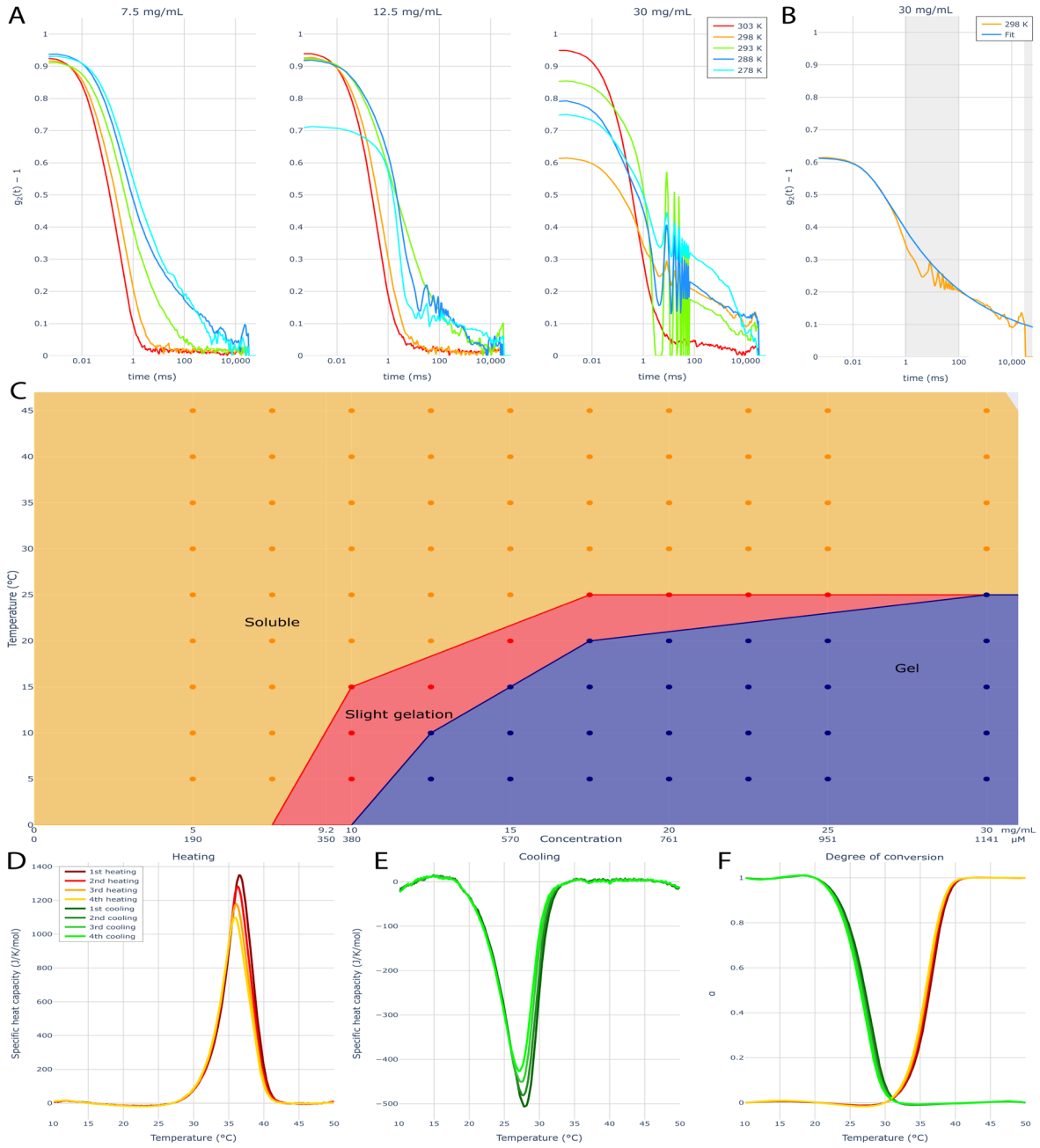


Figure 7.2: Phase diagram of CAHS-8 and characterisation of its gelation and melting.

A and **B**: Dynamic Light Scattering (DLS) correlograms showing the time-dependent decay of the normalised intensity autocorrelation function $g_2(t)$ at different temperatures for samples of CAHS-8 at 7.5, 12.5 and 30 mg/mL (**A**). The correlograms were measured from the highest temperature to the lowest at each concentration. As the system approaches the gel point, fibres or local, static inhomogeneous gel patches manifest by the noisy spikes visible at intermediate time, while the gelation manifests as a significative suppression of the $g_2(0)$ value due to the loss of ergodicity as well as the manifestation of a power-law behaviour, with these effects being the strongest around the gel point as discussed in the introduction of this part. This is clearly visible at 298 K at 30 mg/mL and permit the fitting of this decay to the equation given in the introduction for characterisation of the gel point (**B**, reduced $\chi^2 = 6.76 \times 10^{-5}$), yielding a fast diffusion coefficient of $2.41 \times 10^{-9}\ \mathrm{m}^2 s^{-1}$ accounting for 11 % of the scattering species and a fractal exponent of 0.83. The intermediate part of the function highlighted in grey was excluded from the fit as it results from spatial inhomogeneities.

C: Phase diagram of CAHS-8 drawn from these DLS correlograms, with each point representing a measurement. The yellow region corresponds to the smooth exponential decay of the $g_2(t)$ function while the red, "slight gelation" region corresponds to the observation of inhomogeneities spikes without suppression of $g_2(0)$, as it means a significant number of fibres are present to yield local gel patches but not enough for the percolation of the whole sample. Finally, the blue, gel region corresponds to the clear suppression of $g_2(0)$ and the presence of an apparent power-law decay. As the measurements were performed every 5°C, the gelation was not observed at 30 °C at 30 mg/mL of CAHS-8 but was apparent at 25 °C, indicating that precise gelation temperature is in-between these two values.
D, **E** and **F**: Measurements of Differential Scanning Calorimetry (DSC) thermograms on CAHS-8 at 30 mg/mL shows a transition at 36.2 °C on average when heating the sample (**D**) which corresponds to the melting of the gel, and a transition at 27.5 °C on average when cooling the sample (**E**), which corresponds to its gelation. This gelation temperature is in agreement with the DLS phase diagram. The calculation of the degree of conversion of the species α (**F**) highlights that the gelation and melting happen at distinct temperatures, indicating that CAHS-8 displays a thermal hysteresis, with the temperature at which the two curves meet being approximately 31.1 °C, that is 304 K.
The DSC thermograms follow the "exo down" convention. The gelation is therefore an exothermic reaction and the melting of the gel an endothermic reaction. The baseline has been subtracted from the specific heat capacity.
The precise transition temperatures changes in enthalpy and specific heat capacity are given in the following table.

At 30 mg/mL, the gelation was not observed at 30 °C of CAHS-8 but was clear at 25 °C (fig. 1 A, right), indicating that precise gelation temperature is in-between these two values. The system was therefore more precisely characterised by Differential Scanning Calorimetry (DSC). The measurement of four heating and cooling thermograms on the hydrogel of CAHS-8 at 30 mg/mL with a scan rate of 0.5 °C/min show a first-order endothermic transition at 36.2 °C on average in the heating runs (fig. 2 D), which corresponds to the melting of the gel, and a first-order exothermic transition at 27.5 °C on average on the cooling runs, which corresponds to the gelation (fig. 2 E).

This precise temperature of gelation T_{gel} determined by DSC is in agreement with the range derived from the DLS correlograms (fig. 2 C). The average changes of enthalpy associated with the gelation and its melting are respectively -2.72 ± 0.05 kJ/mol and 5.96 ± 0.05 kJ/mol and are presented in the following table.

Table 1: Differential Scanning Calorimetry parameters of CAHS-8 at 30 mg/mL.

Sample and run	Transition temperature (°C)	Change in enthalpy ΔH (kJ/mol)	Change in specific heat capacity (J/K/mol)
CAHS-8 30 mg/mL 1st heating	36.5	6.58 ± 0.04	$2.6 \times 10^2 \pm 3$
CAHS-8 30 mg/mL 1st cooling	27.9	-2.96 ± 0.05	-71 ± 10
CAHS-8 30 mg/mL 2nd heating	36.3	6.26 ± 0.06	$2.0 \times 10^2 \pm 5$
CAHS-8 30 mg/mL 2nd cooling	27.7	-2.82 ± 0.04	-73 ± 10
CAHS-8 40 mg/mL 3rd heating	36.1	5.81 ± 0.04	$2.0 \times 10^2 \pm 5$
CAHS-8 30 mg/mL 3rd cooling	27.4	-2.63 ± 0.04	-61 ± 10
CAHS-8 30 mg/mL 4th heating	35.9	5.5 ± 0.04	$1.8 \times 10^2 \pm 6$
CAHS-8 30 mg/mL 4th cooling	27.1	-2.47 ± 0.04	-52 ± 10

Although already apparent from the difference of the temperatures of gelation and melting, the calculation of the degree of conversion of the species α highlights the fact that the gelation process of CAHS-8 displays a thermal hysteresis (fig. 2 F), i.e. that its state at a particular temperature depends on its history. Indeed, the difference of the midpoint of the gelation and melting curves results in an intermediate temperature range where the sample state can be either soluble or a hydrogel depending on which transition it last experienced. The gelation and melting curves meet at respectively their end and onset at approximately 31.1 °C or 304 K. Thus, as the gel melts at $T_{\mathrm{melt}} = 36.2\ °C$ but forms only at $T_{\mathrm{gel}} = 27.5\ °C$, the melting of a gel sample at a temperate above T_{melt} and its subsequent cooling at a temperature higher T_{gel}, for instance 31 °C, yields a soluble sample. However, if a sample was subjected to the reverse process, i.e. its gelation was triggered by being subjected to a temperature lower than T_{gel} and then the gel was heated up to 31 °C, which is lower than T_{melt}, then the sample remains a hydrogel.

This possibility of studying either the hydrogel state or a soluble state at the same concentration and temperature thanks to this hysteresis of the protein is used in the next chapter.

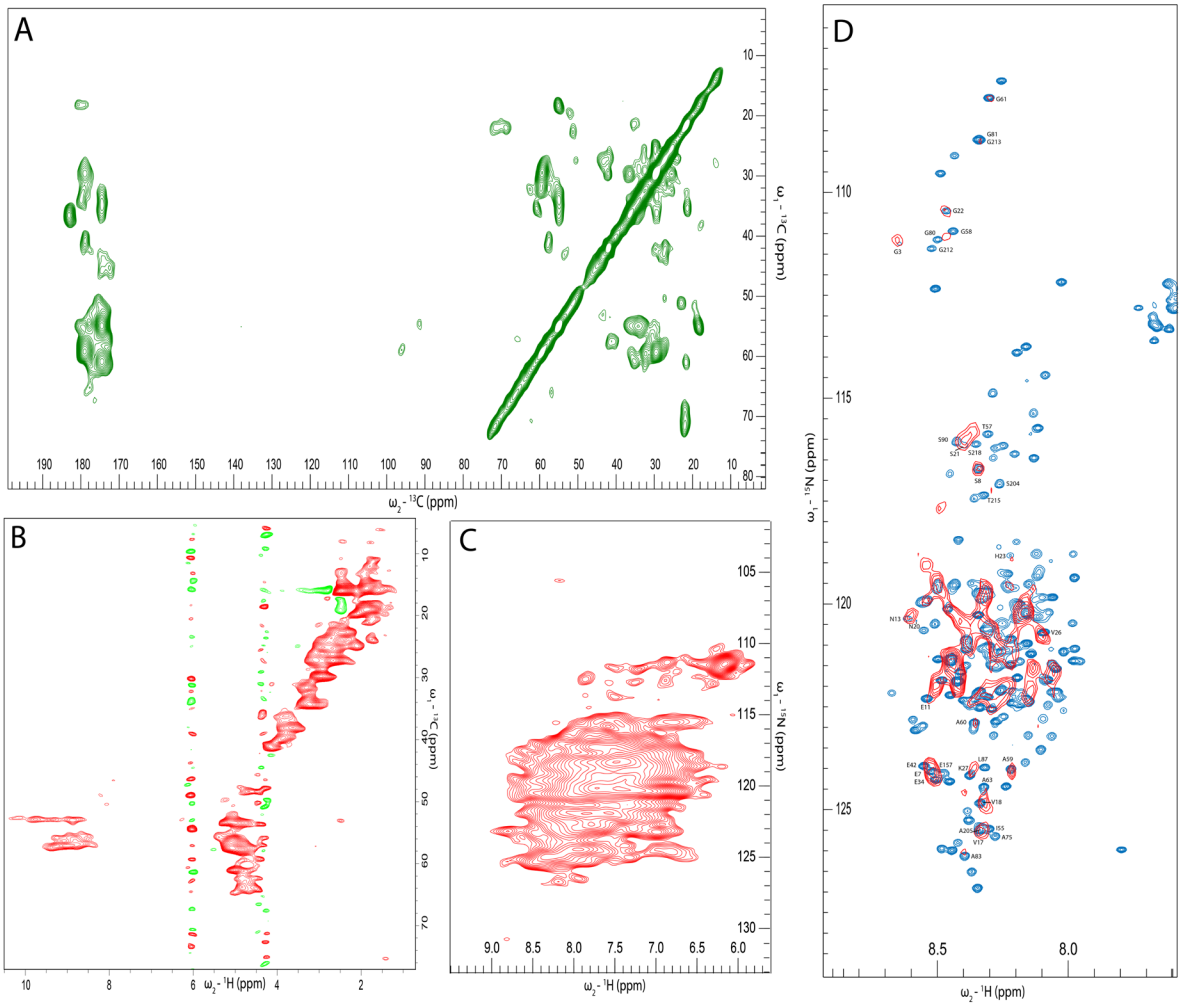


Figure 7.3: Solid-state NMR spectra of the CAHS-8 hydrogel.

A: Carbon-detected 13C-13C DARR (Dipolar Assisted Rotational Resonance) spectrum of TMAO-induced fibres of CAHS-8 measured in a 3.2 mm rotor at 800 MHz, 16.67 kHz magic-angle spinning (MAS) speed, 293 K and 20 ms of mixing time.

B, C and D: Proton-detected cross-polarisation based 13C-1H hCH (B), cross-polarisation based 15N-1H hNH (C) and INEPT-based 15N-1H hNH (D) ssNMR spectra of a 30 mg/mL deuterated gel of CAHS-8 (red). These spectra were measured in a 1.3 mm rotor at 1 GHz, 55 kHz MAS and 293 K.

6.2.1.2 The disordered terminal chains of the fibres are free in the gel state

Characterisation of the CAHS-8 hydrogel and fibres constituting it was initially attempted by solid-state NMR. Several ^{13}C-^{13}C carbon-detected DARR (Dipolar Assisted Rotational Resonance) spectrum of TMAO-induced fibres of CAHS-8 were measured but the linewidths were too broad to permit a subsequent assignment (fig. 3 A). Proton-detected

cross-polarisation (CP) based ^{13}C-^{1}H hCH (fig. 3 **B**) and ^{15}N-^{1}H hNH (fig. 3 **C**) as well as INEPT-based hNH (fig. 3 D) were measured on a deuterated gel of CAHS-8 and also show linewidths too broad to permit an assignment, despite the deuteration which limits the magnitude of the proton-proton dipolar coupling interaction. These broad linewidths therefore likely arise from inhomogeneous line broadening which can be attributed to the multitude of species coexisting in the gel: fibres, fibrils, possible isolated monomers as well as DCC or FCC oligomers etc. Nevertheless, comparing the CP and INEPT based hNH spectra (fig. 3 C and D) shows that while most of the protein is rigid enough in the hydrogel to permit a significant magnetisation transfer via CP, the presence of strong resonances in the INEPT-based spectrum indicates that some species remain flexible. The comparison of this ssNMR INEPT-based hNH with the solution-state HSQC of CAHS-8 at 350 µM and 293 K presented in the previous chapters reveals similarities in the peak position (fig. 3 D), all belonging to the disordered termini. As suspected due to the structure of the fibril described in the previous chapter, this suggests that the disordered termini remain flexible in the gel.

Using the temperature dependence of CAHS-8, a ^{15}N labelled hydrogel at 30 mg/mL, i.e. 1.14 mM, could be introduced in a solution NMR tube. The obtained ^{15}N-^{1}H HSQC spectrum spectra at 293 K and its comparison with that of CAHS-8 at 350 µM reveals that the peaks of the disordered termini are visible in the gel state and have indistinguishable chemical shifts (fig. 4 A).

The measurement of a stimulated-echo diffusion-NMR experiment with a long diffusion delay $\Delta = 300$ ms can be used to distinguish species of significantly different translational diffusion coefficient D.

The decay of the magnetisation intensity of the CAHS-8 hydrogel at 278 K as a function of the gradient field strength reveals a bi-exponential profile (fig. 4 B). The subsequent fit of the intensity decay to a bi-exponential Stejskal-Tanner ($\chi^2 = 6.49 \times 10^{-3}$) yields two species with apparent translational diffusion coefficients $D_{\text{free}} = 1.53 \times 10^{-9} \pm 4.0 \times 10^{-11}$ and $D_{\text{static}} = 2.78 \times 10^{-11} \pm 1.85 \times 10^{-12}$, and with respective populations of 58% and 42% $\pm$ 0.6%. This indicates that two species coexist in the gel: a static species, which

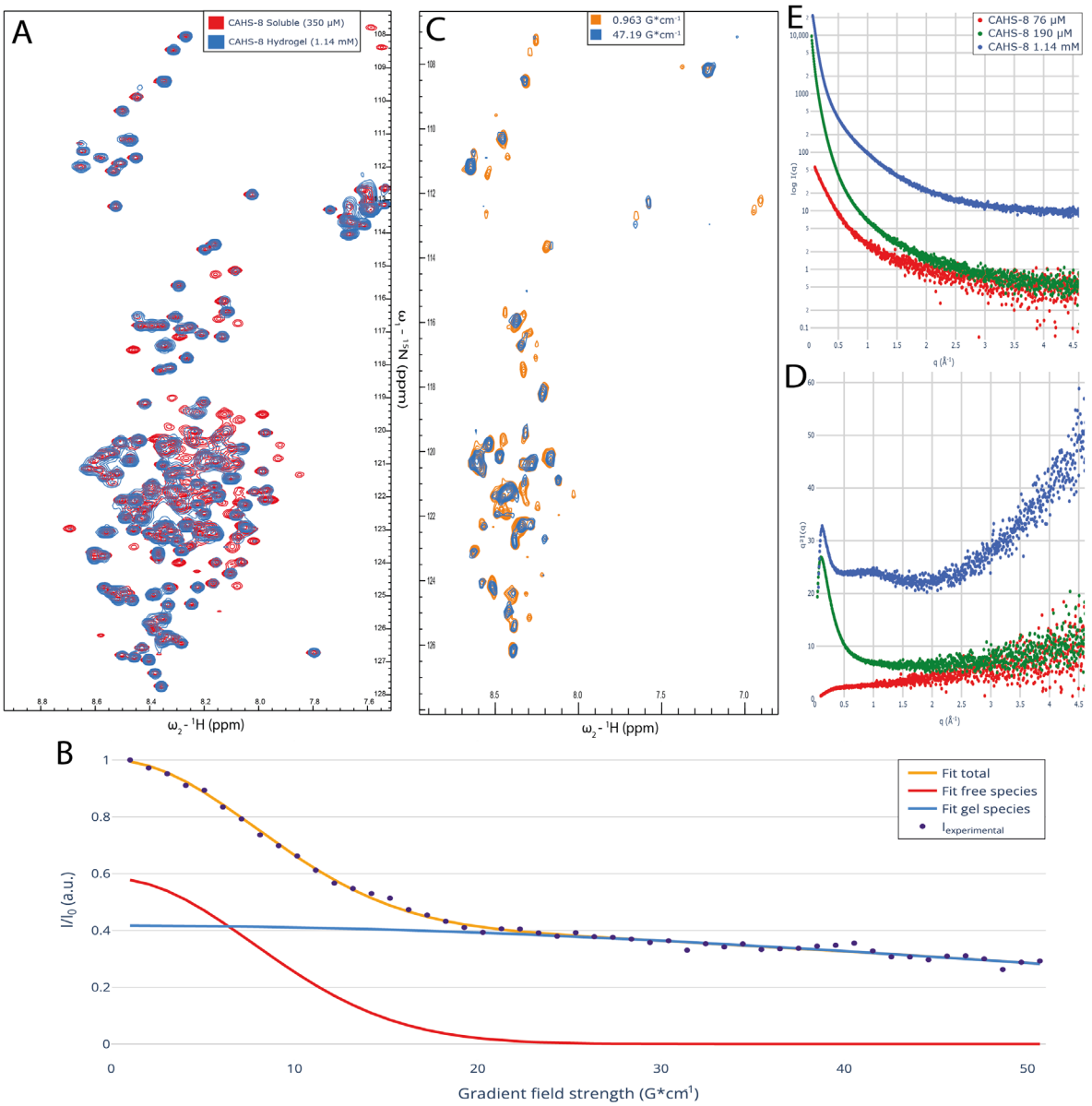


Figure 7.4: The gel-forming fibres of CAHS-8 have free disordered termini and coexist with free CAHS-8 species.

A: ^{15}N-^{1}H HSQC spectrum of CAHS-8 at 293 K and at 350 µM in its soluble form (red), and at 1.14 mM (30 mg/mL) in its hydrogel form (blue). The chemical shifts of the disordered termini are unchanged.

B: The decay of the magnetisation intensity in a diffusion-NMR experiment with a long diffusion delay $\Delta = 300$ ms on the CAHS-8 hydrogel at 1.14 mM and 278 K shows a clear bi-exponential profile (purple) which was fitted to a bi-exponential Stejskal-Tanner equation (yellow, $\chi^2 = 6.49 \times 10^{-3}$). This yield one normal diffusing species with a translational diffusion coefficient $D_{\text{free}} = 1.53 \times 10^{-9} \pm 4.0 \times 10^{-11}$ (red, with an apparent population of 58% ± 0.6%) and one static species with a translational diffusion coefficient $D_{\text{static}} = 2.78 \times 10^{-11} \pm 1.85 \times 10^{-12}$ (blue, with an apparent population of 42% ± 0.6%).

C: Diffusion-edited HSQC with a diffusion delay $\Delta = 300$ ms of the CAHS-8 hydrogel at 1.14 mM (30 mg/mL), 600 MHz and 278 K at low (orange) and strong (blue) field gradient strength showing respectively both the free tumbling and static species for the former and only the static species in the second. The chemical shifts are almost identical, indicating similar conformational sampling between both species.

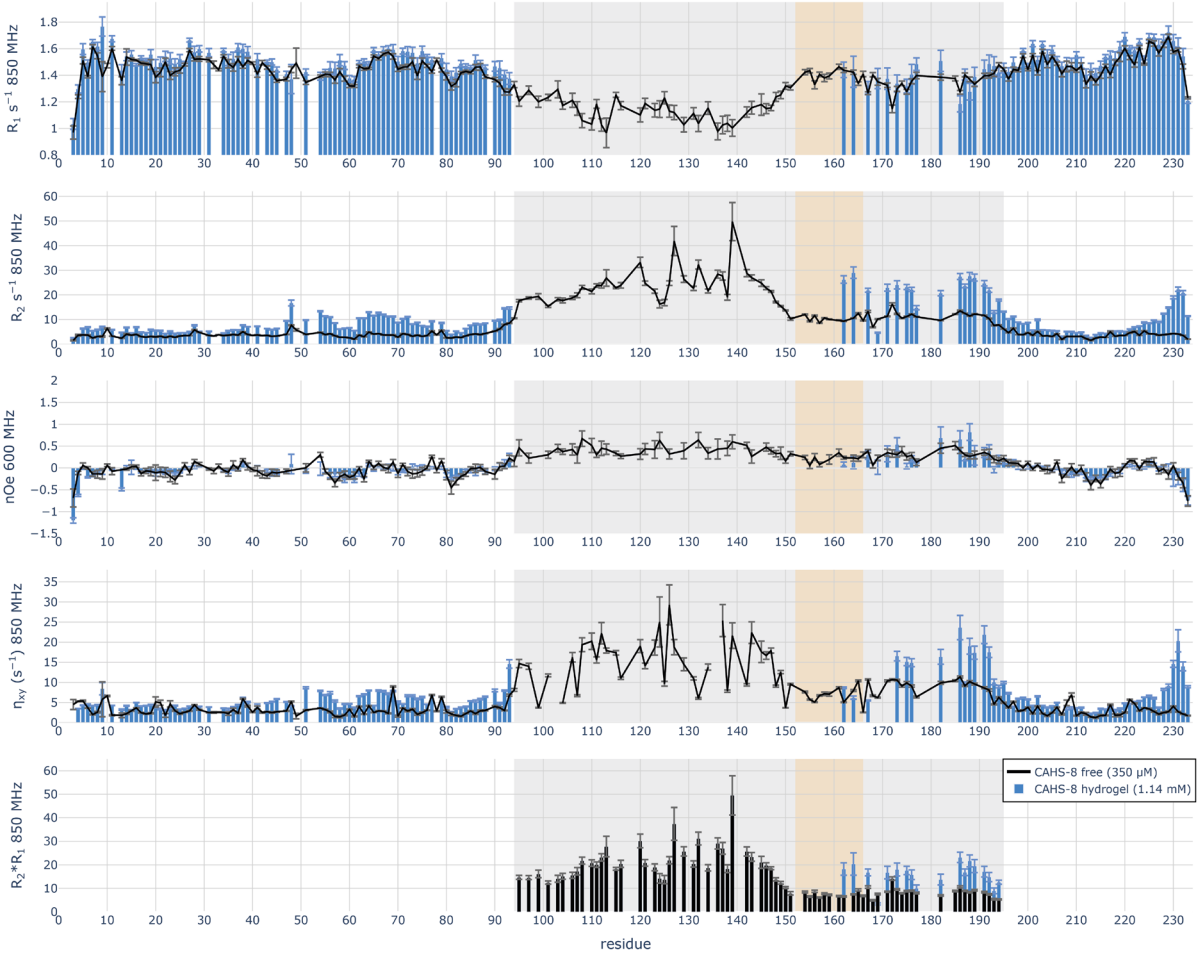
D and **E**: Comparison of the SAXS scattering profile (**C**) and Kratky profile (**D**) of CAHS-8 in its soluble form at 76 µM (red), in its soluble form but with the scattering dominated by fibrils at 190 µM (green) and in its hydrogel form at 1.14 mM (blue). In the Kratky plot, the presence of a peak at small magnitude of the scattering vector q indicate the presence of rigid domain, the presence of an almost horizontal line indicates a partial flexibility and the rise at high q indicates disorder. The hydrogel profile in the Kratky plot was divided by a factor of four to permit a visual comparison with the lower concentrated forms.
Altogether, these data indicate that the disordered termini of the CAHS-8 fibres are static, as the fibres form the gel, and visible by solution NMR while free species of CAHS-8 are tumbling in the cavities of the gel and are also visible by solution NMR.

can be attributed to the free disordered termini of the CAHS-8 monomers forming the fibres, and a freely tumbling species, which can be attributed to CAHS-8 species evolving in the cavities of the gel. The value of D_{free} at 278 K is in agreement with the value of D_{fast} obtained by DLS at 298 K (fig. 2 B) but not the populations, which is likely due to the static population being underestimated in NMR as their transverse relaxation time must be considerably slower than the free diffusing CAHS-8, which therefore dominate the obtained signal.

In turn, the measurement of a diffusion-edition HSQC with the same diffusion delay $\Delta =$ 300 ms permit to observe these two species (fig. 4 C), with the lower gradient strength (orange) displaying the resonances of both species but being dominated by the free species due to likely difference in transverse relaxation rates, while the strongest gradient (blue) corresponds to the static species as per fig. 4 B. The two species display almost indistinguishable chemical shifts belonging to the disordered terminal chains of CAHS-8, which indicate that indeed, the static species are static termini which appear to have similar conformational sampling.

This is further supported by SAXS measurements (fig. 4 D and E). Indeed, the Kratky profile of the CAHS-8 hydrogel at 30 mg/mL increases significantly at high q values, which is characteristic of the presence of disordered species within the gel.

Figure 7.5: Comparison of the relaxation of CAHS-8 in its soluble and hydrogel forms at 293 K.
From top to bottom: R_1, R_2, nOe and η_{xy} relaxation as well as the R_2R_1 product for the soluble form of CAHS-8 (350 µM i.e. 9.2 mg/mL, black) and hydrogel form (1.14 mM i.e. 30 mg/mL, blue). All the rates were measured at 850 MHz except the heteronuclear nOe relaxation which was measured at 600 MH.
The helical domain of CAHS-8 is highlighted in grey, with the linker between the helices being highlighted in orange.



The intensities of the CAHS-8 hydrogel in ^{15}N-^{1}H HSQCs show virtually no change over a period of at least three weeks, indicating that the gel is in a steady-state over the course of this period of time. The measurement of the relaxation of the CAHS-8 30 mg/mL, that is 1.14 mM, hydrogel at 293 K and its comparison with the relaxation of CAHS-8 at 350 µM at this temperature (fig. 5) confirms that the disordered termini are indeed entirely visible as well as peaks belonging to the C-terminal part of the linker region and to helix 2 (highlighted respectively in orange and the following grey box).

The comparison of the relaxation rate shows similar profiles, with a consistently higher R_2 rates in the hydrogel compared to the soluble protein, slightly lower R_1 rates and similar nOe rates. This suggests similar contribution of the fast motions to the relaxation in the low-concentrated, free CAHS-8 and the CAHS-8 species visible in the hydrogel state, but greater quenching of the spectral density function at $J(\omega_N)$ and before. In turn, this is consistent with effects from a slight increase in viscosity and from possible exchange contributions. In particular, the 62-72 region in the N-terminal disordered chain displays

fasters R_2 rate constants than the neighbouring residues which may possibly result from interactions. This region includes the conserved 71F and hydrogen bond-capable residues highlighted in the previous chapter and suspected of being involved in π-π interactions with other aromatic residues in the disordered chain or with H145 as well as hydrogen bonding. In turn, the exogeneous C-terminal ENLYFQ sequence also shows a similar behaviour which may arise from analogous causes due to the presence of two aromatic residues, Y231 and F232 as well as the possibilities of E228, N229, Y231 and Q233 to engage in hydrogen bonds. However, a similar behaviour was also observed at 278K at the dilute 350 µM concentration, which indicate that this is not a gel-specific phenomenon and may arise from a change in secondary structure propensity but, as ^{13}C secondary chemical shifts were not measured in the gel state, no definitive explanation for this behaviour can be provided.

The residues belonging to the linker and helix 2 shows significantly higher R_2 rates in the hydrogel than in the soluble CAHS-8, increasing from approximately 10 s^{-1} to approximately 25 s^{-1}. In turn, the difference between the transverse autorelaxation R_2 and the transverse cross-correlated relaxation η_{xy} in the linker region highlights the presence of exchange.

As these regions of the protein are involved in the formation of the fibrils, as seen in the crystal structure, it is therefore highly probable that these peaks belong exclusively to the CAHS-8 protein tumbling in the cavities of the gel, with these peaks being visible thanks to their faster correlation time than helix 1.

6.2.1.3 No exchange of the free CAHS-8 with the fibres could be characterised

As the fibres of CAHS-8 are reversible by changing conditions, free neighbouring monomers or coiled-coil dimers of CAHS-8 are likely to exchange with the fibrils or to be tethered to them. Dark-state exchange saturation transfer (DEST) experiments[801,802] were measured using a 150 Hz and a 350 Hz B_1 field in order to characterise this putative exchange and only residues belonging to the helical domain were used, being considered to belong exclusively to the free diffusing species in the cavities. Although peaks showed

a DEST effect, not enough residues had signal to noise ratio high enough to permit a satisfactory fit.

6.2.2 The fibrous aerogel

The last step of the purification of CAHS-8 after expression is its lyophilisation, also called freeze-drying, in a volatile buffer which eventually allow resuspension of the protein to the desired concentration. However, while most freeze-dried proteins yield a more or less fluffy powder, the lyophilised CAHS-8 is a white, compact solid which invariably shows a helical shape when freeze-dried in falcon tubes (fig. 6 A, B). Although it can reach 12 cm, this CAHS-8 lyophilisate is not dense and as such, rarely contain weight more than 10 mg. Yet, these solid displays a mechanical resistance to being torn apart, which nevertheless reveal it is composed of long fibres under binocular lenses (fig. 6 C).

In order to understand the nature of this additional state of CAHS-8, the protein was freeze-dried *in situ* on AFM supports and imaged, revealing a highly porous structure which can reach several hundred of nanometres (fig. 6 D, E). In turn, the subsequent focus on the denser regions (fig. 6 F, G) shows that this solid is composed of intertwined fibres which show their characteristic 3 nm height (fig. 6 G) and organised similarly to the hydrogel and xerogel form of CAHS-8 (fig. 1 C and D).

This fibrous organisation with high porosity resulting in a dry solid with a large surface yet with a low density is the definition of an aerogel, which are typically prepared by either supercritical CO_2 drying or, precisely, freeze-drying[743,744,803,804].

On the edges of the *in situ* lyophilisate on AFM plates, CAHS-8 shows an apparent crystalline arrangement (fig. 7 A) and when freeze-dried on steel plates, the aerogel appear constituted of ordered, more or less parallel fibres or layers (fig. 7 B) which is reminiscent of the arrangement of the crystallised fibres described in the previous chapter.

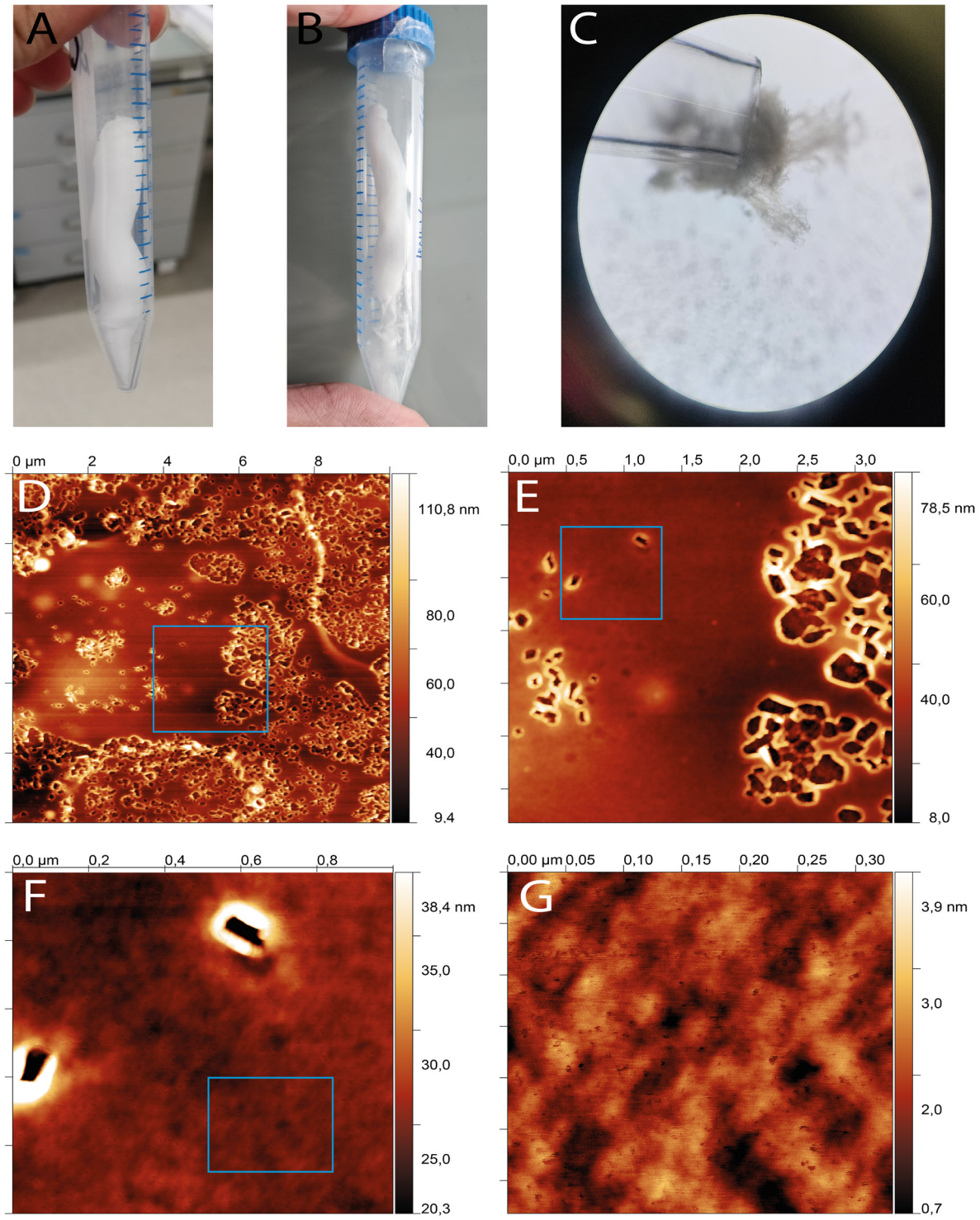


Figure 7.6: The aerogel of CAHS-8 is highly porous and constituted of fibres.
A, **B** and **C**: Freeze-dried CAHS-8 showing a macroscopical helicity (A and B) and revealing a fibrous composition under a binocular lens (**C**).
D, **E**, **F** and **G**: Air-mode AFM imaging of in-situ freeze-dried CAHS-8 on an AFM glass support. The blue rectangles denote the zone imaged in the next panels. Broad field of view images show a highly porous structure (**D**) which differs from xerogel (fig. 1 **D**). However, the focus on the denser region reveals a similar organisation to the hydro- and xerogels (**G**) with the 3 nm high fibres of CAHS-8 being recognisable.

In turn, the X-Ray diffraction of these aerogels of CAHS-8 as well as HD both unlabelled or seleno-methionine labelled reveals the presence of small crystalline regions (fig. 7 C), although it is unclear whether these are microcrystals embedded in the aerogel or if they are regions of the fibres with a highly periodical crystalline arrangement as in the crystallised fibres presented in the previous chapter. However, the diffraction profile is clearly poly-crystalline (fig. 7 D) and as such, despite their diffraction to high resolution, the data couldn't be processed.

Finally, as the trehalose desiccation-protection mechanism of forming a glass has been put forward for CAHS proteins and an apparent glass transition has been observed by Boothby and co-workers for dried CAHS-8 protein around 60 °C and 140 °C[614], the CAHS-8 aerogel were tested by solid-state DSC to see whether a glass transition could be observed. The obtained thermograms (fig. 7 E) do not show transitions around 60 °C but does show a slight peak on the first scan at 145 °C indicating an apparent first-order transition which does not appear to perturbate the baseline. As such, it cannot correspond to a glass transition which is a second-order transition, which therefore does not show a peak but triggers a change in the baseline as the phase transition modify durably the heat capacity of the sample[739]. Moreover, this is unlike the well reproducible gel and melting transitions previously described for the hydrogel (fig. 2). Consequently, as this peak is only visible on the first scan, it could be is either artefactual, possibly arising from residual water trapped within the aerogel or may correspond to some transition of the structure of the aerogel, possibly becoming more crystalline. Additionally, Arakawa and Numata published a letter in which they measured Thermal Gravimetric Analysis (TGA) simultaneously to DSC on another CAHS proteins and noted a weight-loss upon these apparent transitions. They therefore proposed that the apparent glass transition observed for CAHS-8 by Boothby and co-workers were artefacts due to residual water, which is compatible with our data[805].

Thus, if the formation of a protective glass is not the mechanism of protection of CAHS proteins, other mechanisms are necessary to explain their effect. This is the object of the next chapter of this thesis.

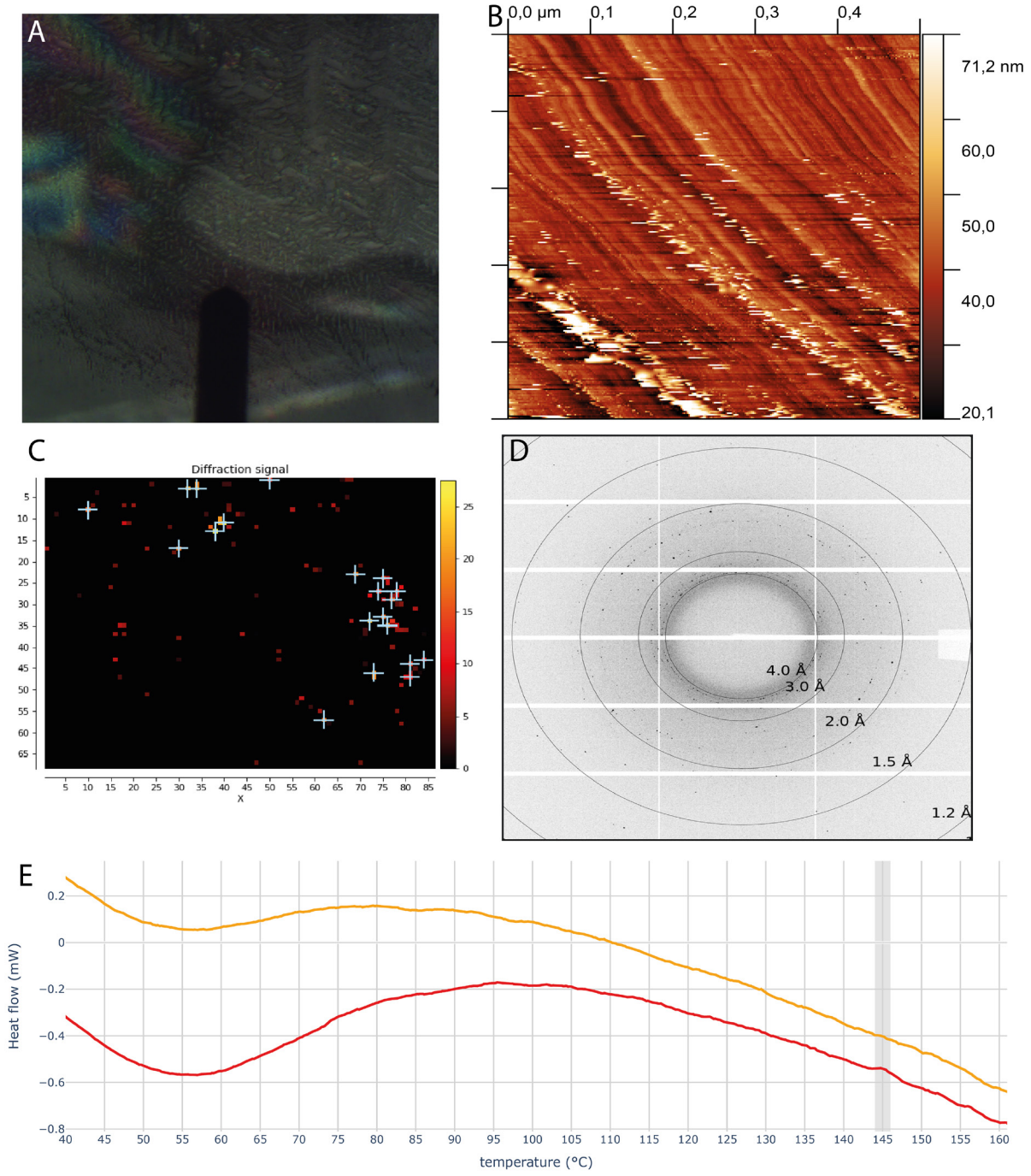


Figure 7.7: The aerogel of CAHS-8 displays crystallinity but dubious glass properties.
A: Freeze-drying patterns of CAHS-8 on AFM glass plates showing crystalline arrangement. The black rod on the bottom is the AFM tip.
B: AFM imaging of CAHS-8 freeze-dried on a steel plate showing what appears to be more or less parallels fibres or layers which bears resemblance to the crystallised fibres.
C and **D**: Mounting the aerogel on cryoloop and performing a mesh-scan reveals multiple crystalline regions (**C**) which show a poly-crystalline X-ray diffraction pattern (**D**).
E: Solid-state DSC thermograms of two heating scans of the aerogels of CAHS-8 showing non-reproducible transition, with the first scan in red and the second in yellow. The scan rate is 0.5 °C per minute.

Chapter 8: Effects of the CAHS-8 hydrogel on entrapped proteins

« In the case of all things which have several parts and in which the totality is not, as it were, a mere heap, but the whole is something beside the parts, there is a cause;

for even in bodies contact is the cause of unity in some cases, and in others viscosity or some other such quality. »

Aristotle

Metaphysics, book XIII, 4th century B.C.E.

7.1 Results

CAHS-8 was demonstrated in the previous chapters to self-associate into a coiled-coil dimer which leads to the formation of fibrils and fibres that entangle to form a hydrogel at high concentration. During dehydration, the concentration increases as water is removed and thus, the CAHS-8 hydrogel may form *in vivo* in tardigrades and fill their cell during desiccation. As such, the study of the effects of the CAHS-8 hydrogel on protein entrapped in it may reveal the mechanism by which CAHS proteins protect tardigrade from desiccation[614].

7.1.1 Proteins and macromolecular assemblies can be encapsulated in the gels

In order to visualise such entrapment of exogeneous proteins CAHS-8, gels were formed in the presence of other proteins and imaged by AFM (fig. 1 A). Unfortunately for small proteins of appropriate size for subsequent NMR studies, little effect was seen at the resolution available to AFM. However using macromolecular assemblies such as the 18 x 300 nm rod-shaped Tobacco Mosaic Virus (TMV, fig. 1 B)[806,807] or large proteins such as the approximately 14.5 x 8.5 x 4 nm Immunoglobulin G (IgG, fig. 1 C, D)[808,809] permit to see that these objects localise in the gel, either sequestered in the cavities or tethered to the gel via the CAHS-8 disordered chain. In particular, IgG proteins appear to closely follow the wall of the gel. Moreover, both TMV and IgGs can be observed entering the pores of the hydrogel, suggesting more are trapped below, within the gel.

As demonstrated in the previous chapter, CAHS-8 gelates below approximately 27.5 °C if concentrated beyond the gel-point concentration. This temperature dependence can therefore be used to form an unlabelled gel around other ^{15}N-labelled protein in a solution NMR tube. Unlike phenomena like liquid-liquid phase separation, a gelation spans the entire sample which is therefore not separated between a macroscopical dense and dilute phase. As such, this allows observation of only these encapsulated and labelled species.

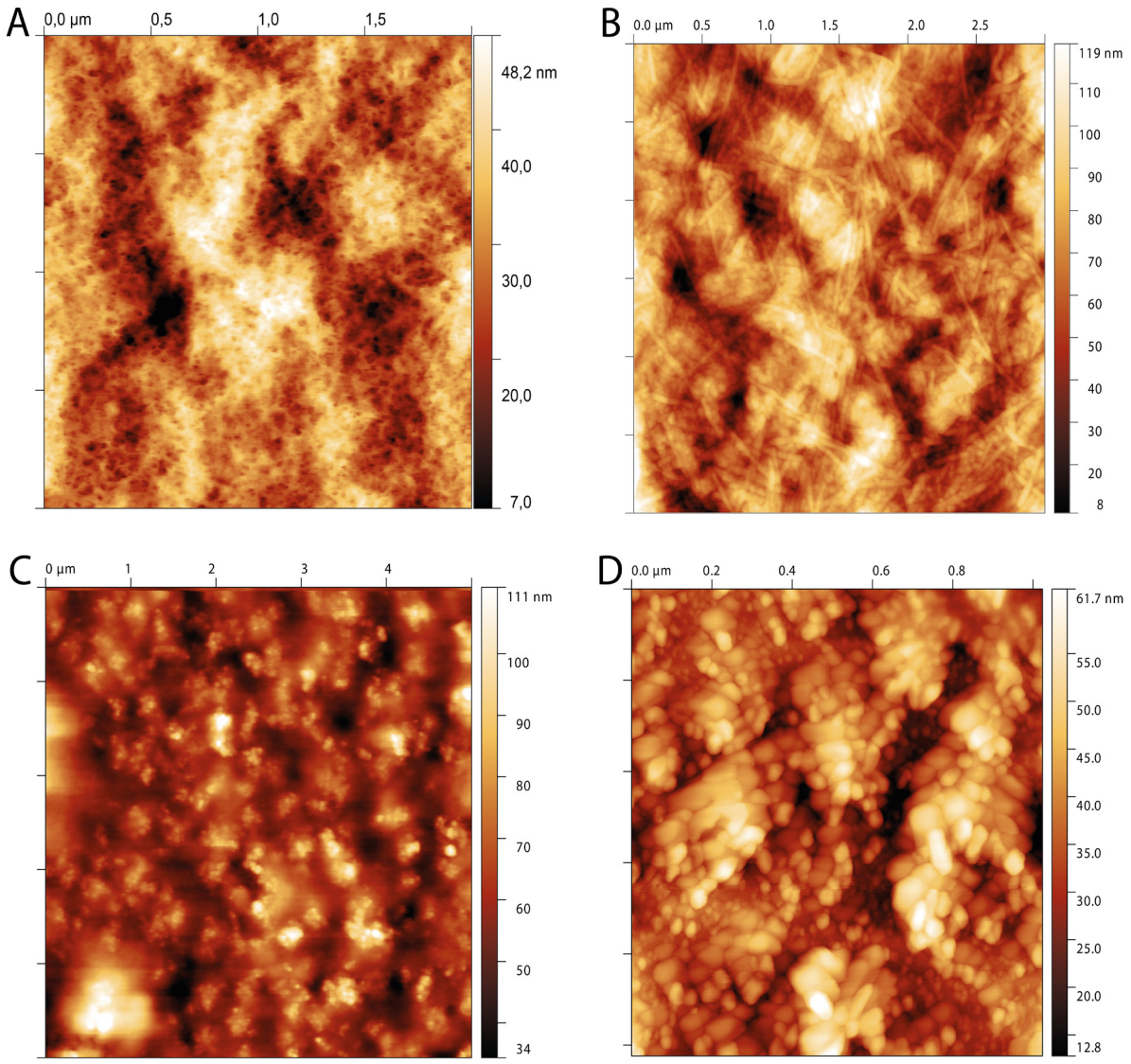


Figure 8.1: AFM imaging of proteins and macromolecular assemblies entrapped in the CAHS-8 hydrogel.

A: AFM imaging of the CAHS-8 hydrogel form in air.

B: AFM imaging in solution of the CAHS-8 hydrogel with rod-shaped Tobacco Mosaic Viruses (TMV) entrapped in it. Some TMV can be seen going into the pores of the gel.

C and **D**: AFM imaging in solution of the CAHS-8 hydrogel with Immunoglobulin G (IgG) entrapped in it. The IgG can be seen covering the whole gel including the walls of the cavities of the gel.

Indeed, the SH3A domain of the CD2AP protein shows almost unperturbed chemical shifts in a 20 mg/mL CAHS-8 hydrogel (fig. 2 A), as well as the intrinsically disordered domain of the transcription factor ANP32a (fig. 2 B) or the C-terminal disordered domain of the Sendai virus nucleoprotein (N_{tail}, fig. 2 C) in 30 mg/mL CAHS-8 hydrogel.

As the hydrogel presents pores between the fibres as can be seen in AFM (fig. 1 A), these encapsulated proteins most probably diffuse in these cavities.

7.1.2 The gel is a viscous environment in which diffusion is slowed and restricted

The measurement of the translational diffusion of these SH3A, ANP32a and Sendai N_{tail} proteins entrapped in the gel via pulse diffusion-NMR show that their apparent translational diffusion coefficient is slowed down by approximately a factor of two (fig. 2 D-I). Indeed, the apparent translational diffusion coefficient of SH3A goes from $2.18 \times 10^{-10} \pm 3.52 \times 10^{-11}$ to $1.31 \times 10^{-10} \pm 3.38 \times 10^{-11}$ m^2s^{-1} in the gel (fig. 2 D, G), that of ANP32a from passes from $9.61 \times 10^{-11} \pm 3.97 \times 10^{-11}$ in the buffer to $4.89 \times 10^{-11} \pm 2.08 \times 10^{-11}$ m^2s^{-1} (fig. 2 E, H) and that of Sendai N_{tail} slows down from $1.83 \times 10^{-10} \pm 1.94 \times 10^{-11}$ in the buffer to $1.16 \times 10^{-10} \pm 3.63 \times 10^{-11}$ m^2s^{-1} in the CAHS-8. As their fold or conformational sampling are not perturbed, this suggests a viscosity effect or the impact of either partial or complete sequestration of the client proteins within the cavities of the gel, with sufficient unrestrained reorientational freedom to permit acquisition of high-resolution NMR spectra.

In turn, to probe the dynamics of proteins in the CAHS-8 hydrogel, the relaxation of the intrinsically disordered protein Sendai N_{Tail} was measured in the gel at 298 K at 600 MHz and 700 MHz (fig. 3). The comparison with the relaxation of Sendai N_{Tail} in buffer[439,603] reveals that the region 441-463 (grey) and the helical segment 476-492 (orange) have higher R_2 values in the gel than compared to the buffer, with respectively increases from on average ~4 s^{-1} to ~9 s^{-1} in the 441-463 region at 600 MHz and ~7 s^{-1} in the buffer to ~10 s^{-1} in the gel in the helix. The heteronuclear nOe rates are highly similar in the buffer and in CAHS-8 hydrogel, which is similar to observations recently made in the same protein as a function of inert crowding[439], and as the longitudinal autorelaxation rates in these regions R_1 is comparable to the buffer at 600 MHz but smaller in the gel at 700 MHz, this indicates slower correlation times over the nanosecond and above timescales in the gel.

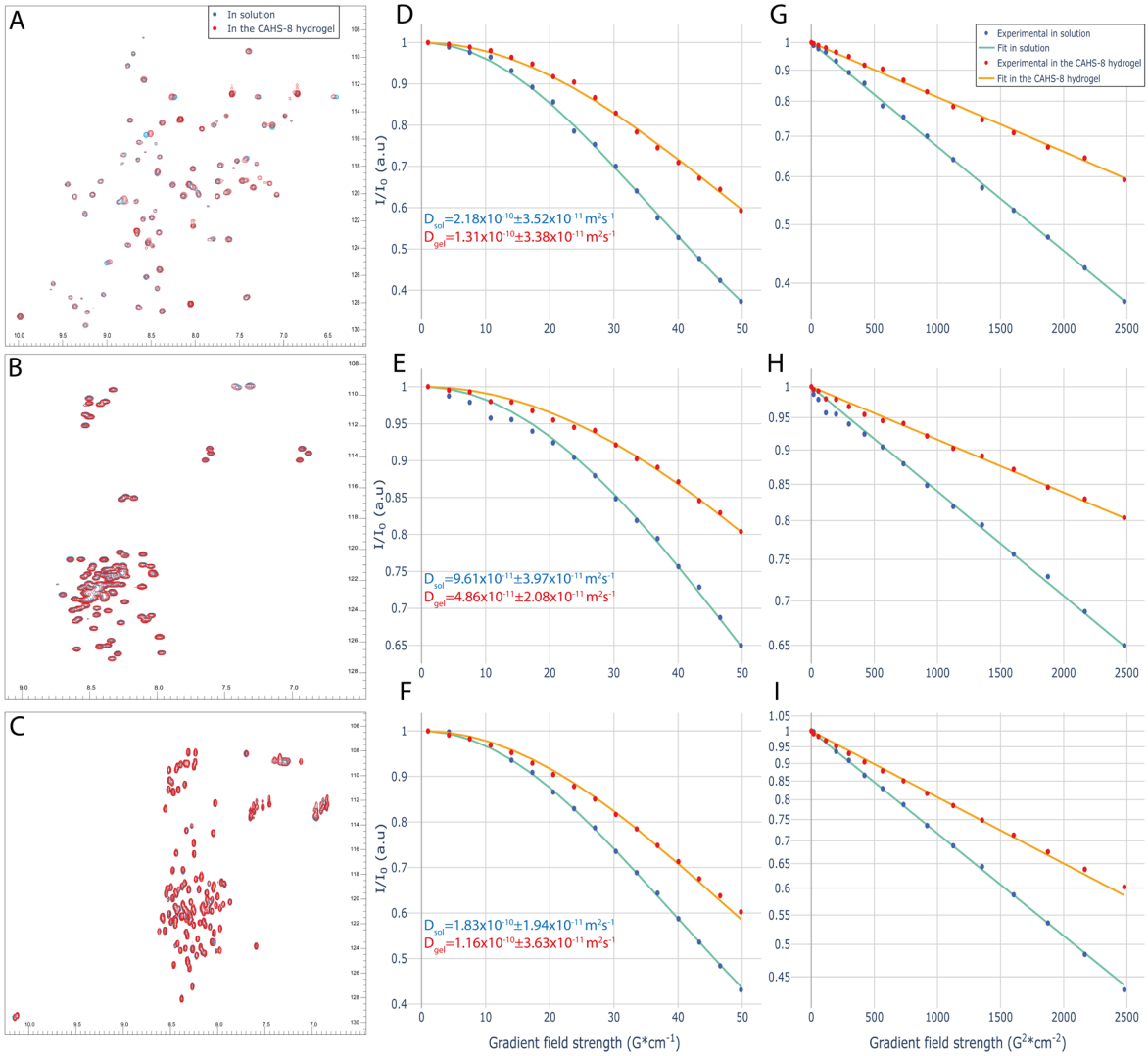


Figure 8.2: Proteins can tumble in the CAHS-8 hydrogel without significant change of their fold or conformational sampling but their translational diffusion is slowed down.

A, B and C: Comparison of the 15N-1H HSQC spectra in the buffer (red) or inside an unlabelled CAHS-8 hydrogel (blue) of the SH3A domain of CD2AP (A), the intrinsically disorder domain of the transcription ANP32a (B) and the C-terminal intrinsically disordered of the Sendai virus nucleoprotein (Ntail, C). The chemical shifts show next to no perturbation upon entrapment of the proteins in the gel.

D, E and F: Diffusion-NMR normalised intensities against field gradient strength of SH3A (D), ANP32a (E) and Sendai Ntail (F) in solution (blue circles) and in the CAHS-8 hydrogel (red circles), with their fit permit to access the apparent translational diffusion coefficient D denoted respectively Dsol and Dgel. The diffusion of the proteins is halved in the gel.

G, H and I: Logarithmic representation of D, E and F against the squared gradient strength.

The reduced χ^2 of the fits are respectively 4.0×10^{-5} and 4.74×10^{-5} for SH3A, 7.0×10^{-5} and 2.26×10^{-5} for ANP32a and finally 1.32×10^{-5} and 5.54×10^{-5} for Sendai Ntail.

To understand the diffusion and viscosity in the CAHS-8 hydrogel cavities environment, first, the diffusion delay Δ was varied during the measurement of the diffusion

experiments on Sendai N_{Tail} in the CAHS-8 hydrogel in order to probe whether the translational diffusion of the proteins is restricted (fig. 4 A, B). Indeed, in the case of an entirely free diffusion, the decay of the intensity remains strictly exponential as in fig. 2 D-F and therefore appearing linear if plotted on a logarithmic scale against the square of the strength of the gradients as in fig. 2 G-I. As such, the apparent diffusion coefficients changes and the entire decay cannot be fitted to the Stejskal-Tanner diffusion equation anymore.

The results show the onset of an anomalous diffusion, i.e. a loss of linearity, between $\Delta = 300$ ms and $\Delta = 600$ ms which increases the longer the diffusion delay (fig. 4 B). This indicates that the diffusion of Sendai N_{tail} is confined within the gel with the protein experiencing free translational diffusion up to at least 300 ms, after which it is probably limited by the size of the pores and cannot diffuse freely anymore, probably bouncing against the walls. As no sharp disruptions of the intensity could be observed, this suggests that the proteins are confined in these CAHS-8 pores and cannot exchange between them, being probably too large to get into the channels between the CAHS-8 fibres forming the gel. As the mean squared displacement (MSD) travel distance of a species experiencing a Brownian self-diffusion random walk in three dimension is equal to $\mathrm{MSD} = 6D\tau$. Thus using the apparent diffusion coefficient resolved in fig. 2 F, that is $D_{app} = 1.16 \times 10^{-10}$ m^2s^{-1} (fig. 2 F) and $\tau = 300$ to 600 ms yields an MSD of 14.5 to 20.4 µm.

This therefore constitutes a rough approximation of the diameter of the CAHS-8 cavities but assumes, a free Brownian self-diffusion until the pores wall are encountered, which is not likely given the estimated and imaged pores of the gel compared to the size of proteins. However, mutual diffusion due to concentration inhomogeneities may also be present and may therefore bias this estimation, although it is probably unlikely to contribute significantly to the apparent diffusion. Unfortunately, despite an extensive search, no clear diffusion-diffraction could be observed and as such, the sizes of the pores of the CAHS-8 hydrogel as it is in the NMR tube with proteins encapsulated in it could not be characterised more precisely by this method.

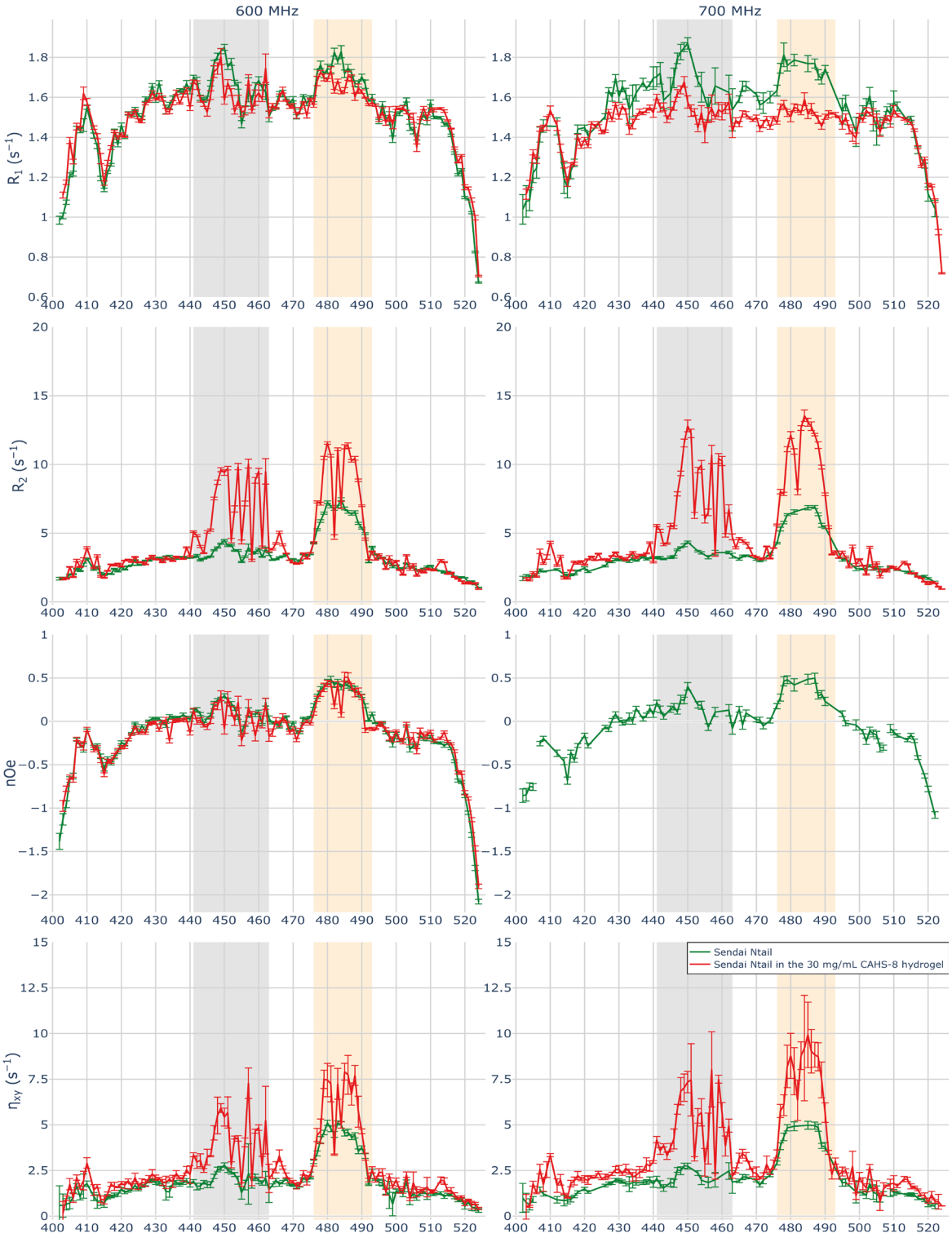


Figure 8.3: The relaxation of Sendai NTail is modified in the CAHS-8 hydrogel.

From top to bottom: Longitudinal and transverse autorelaxation rates R_1 and R_2, heteronuclear nOe and transverse cross-correlated relaxation rates η_{xy} at 600 (left) and 700 MHz (right) of the intrinsically disordered protein Sendai Ntail at 298 K in the sole buffer (green) and in the CAHS-8 hydrogel (red) at 30 mg/mL.

The region 441-463, which comprises W443 and W451, is highlighted in grey and the helical domain 476-492 is highlighted in orange.

As per the translational diffusion of water in the CAHS-8 hydrogel (fig. 3 C, D), the result that the diffusion of water appears slightly biexponential at 278 K as it does not fit well the mono-exponential decay used, suggesting two distinct populations. As the whole visible sample in the NMR tube is a gel, without a separate dilute phase like in liquid-liquid phase separation, it is likely that these two possible water populations are inside the gel and may therefore report on the water molecules encounter limiting boundaries during the diffusion delay.

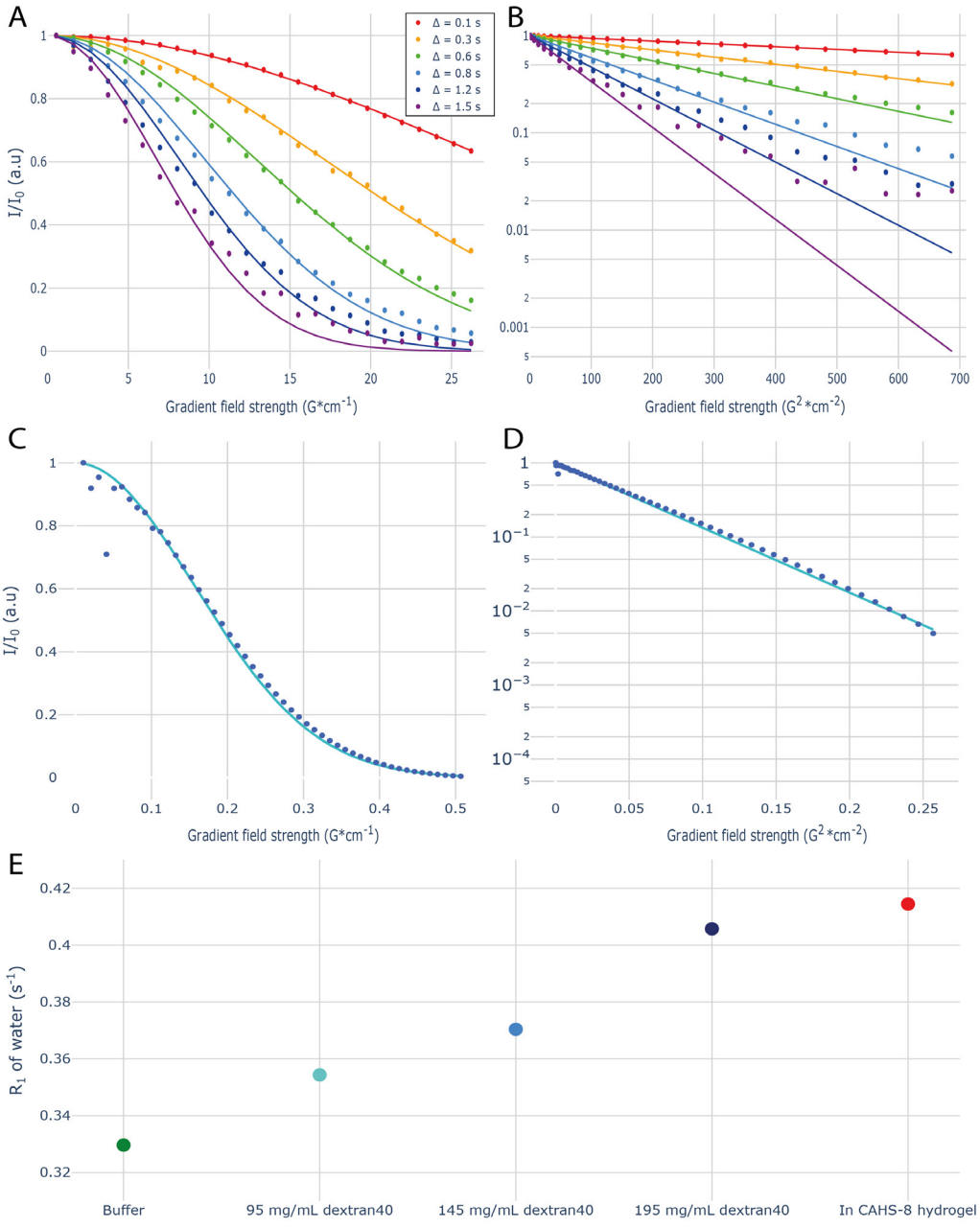


Figure 8.4: The cavities of the CASH-8 hydrogel are viscous environments in which the diffusion is confined.

A and B: Diffusion-NMR normalised intensity decay probing the translational diffusion of for increasing diffusion delays Δ (A). The circles are the experimental values and the lines are the fits. As the diffusion delay Δ increases, the decays are not strictly exponential anymore and as such do not fit the Stejskal-Tanner equation anymore. This translates by a loss of the linearity of the values on the logarithmic plot of the values against the squared strength of the gradient field (B).

C and D: Diffusion-NMR normalised intensity decay probing the translational diffusion of water protons at 278 K in the CAHS-8 hydrogel (C) and logarithmic representation (D).

E: Transverse relaxation rates of the water protons R_1 in presence of Sendai NTail either in the sole buffer, with various concentration of dextran40 viscogen or in the CAHS-8 gel. This rate is expected to be dominated by the rotational diffusion of the water molecules.

Finally, in order to characterise the nanoviscosity in the cavities of the CAHS-8 hydrogel, the longitudinal relaxation of water protons was measured. The obtained water R_1 obtained in the presence of Sendai N_{Tail} at 298 K is higher than the value measured by Adamski et. al in the same buffer with Sendai N_{Tail} in the presence of 195 mg/mL of dextran40 viscogen (fig. 4 E), a concentration in which the crowding is high enough to transition from the dilute to the semidilute regime[439].

The exchange of bulk and bound water to substantially large species as well as surface effect are known to give rise to water relaxation dispersion and are visible at low magnetic field, however, at high magnetic field where the extreme narrowing limit of water spectral density function is reached, this rate R_1 is expected to be dominated by τ_c, that is rotational diffusion of water molecules and therefore constitute a good probe of the nanoscale solvent friction and should be independent of B_0[439,810–812]. Nonetheless, hydrogels typically have a fractal dimension and therefore a large surface[813–815], which is augmented in the case of the CAHS-8 hydrogel by the presence of its terminal disordered chains, as fibrils and fibres not involved in the gel architecture may also be present in the cavities which have been shown to be able to confine the water's translational diffusion (fig. 4 C, D), it is plausible that this R_1 rate of water in the CAHS-8 hydrogel presents an exchange contribution and therefore appears higher than what it really is, which is supported by the similarity between of the in-gel nOe profile to the buffer one. Nonetheless, although it is reasonable to doubt its extent, it is clear from the diffusion coefficients that the pores of the gel exhibit a high nanoviscosity on top of the obviously high shear and bulk viscosities associated with a gel state.

7.1.3 The disordered termini of CAHS-8 are in contact with entrapped proteins

The comparison of the ^{15}N relaxation of Sendai N_{tail} at 298 K and 600 MHz encapsulated in the CAHS-8 hydrogel with its relaxation in buffer or with the 145 mg/mL and 195 mg/mL concentrations of dextran40 viscogen previously mentioned as per the R_1 of water shows several phenomena specifics to the CAHS-8 hydrogel (fig.5).

First, in the α-helical region 476-492 highlighted in orange, Sendai N_{tail} in the CAHS-8 gel shows elevated R_2 and η_{xy} rates comparable with the values measured by Adamski *et al*. 2019 with 145 mg/mL of dextran40[439]. However, the R_1 rates in this helical component when the protein is in the gel are higher than in the fully disordered residues, appearing intermediary between that of Sendai N_{tail} in buffer and with 145 mg/mL of dextran40, being actually closer to the former on the edges of the helix. In turn, the heteronuclear nOe rates also appear closer to the values measured for Sendai N_{tail} without any viscogen than to any other dextran40 admixture.

Remarkably, the nOe rates of the 441-463 segment (grey) in the 30 mg/mL CAHS-8 gel are similar to those in the buffer while R_1 rate constants are similar to those in 145 mg/mL of dextran40 but the R_2 rates are higher than in 195 mg/mL of dextran40 (∼6 s^{-1} on average to ∼9 s^{-1} in the hydrogel) while their η_{xy} rates are similar, which suggest the presence of chemical exchange or different dynamics in the gel compared to in the presence of this crowder. This particular segment comprises two tryptophans, namely W443 and W451, presents a weak predicted helical propensity and a slightly positive N-H bond vector residual dipolar coupling (RDC) profile around residue 510[586] as well as an oscillating RDC profile of the alpha carbon and proton bond vector[586]. Moreover, this region displays a significant increase of its R_2 and η_{xy} rates with the concentration of dextran40, indicating slower correlations times (fig. 5). This may indicate the presence of a transient and weakly populated secondary structure, possibly a kind of helical structure or a π-π stacking of the two tryptophans W443 and W451 which would be stabilised in the CAHS-8 gel.

Finally, the rest of the protein, being fully disordered, shows relaxation rates intermediary to that in the buffer and at 145 mg/mL of dextran40.

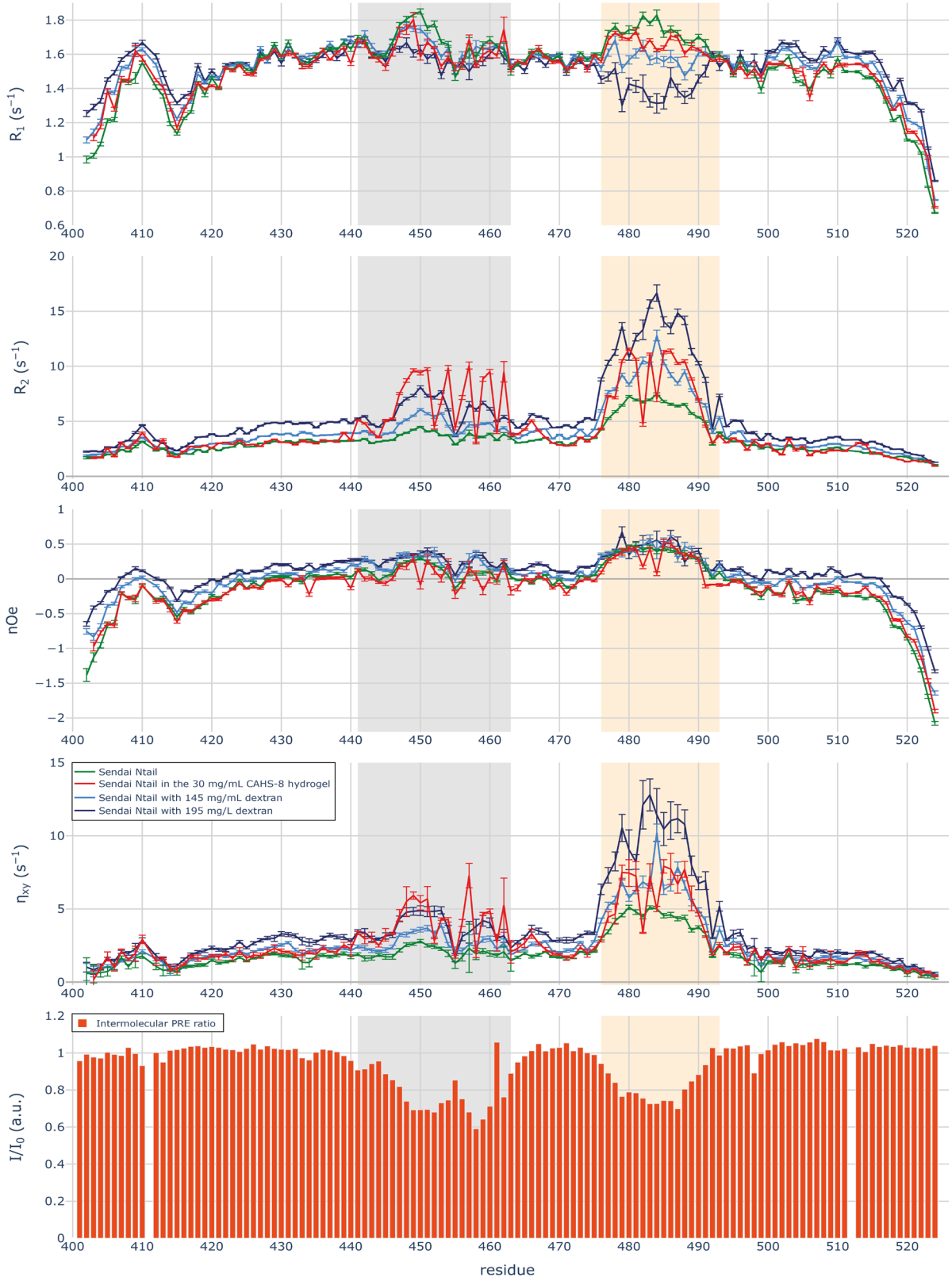

R1 (s−1)
R2 (s−1)
nOe
ηxy (s−1)
I/I0 (a.u.)
residue
Sendai Ntail
Sendai Ntail in the 30 mg/mL CAHS-8 hydrogel
Sendai Ntail with 145 mg/mL dextran
Sendai Ntail with 195 mg/L dextran
Intermolecular PRE ratio

Figure 8.5 : The relaxation of Sendai NTail in the CAHS-8 is not only an effect of viscosity.

From top to bottom: Longitudinal and transverse autorelaxation rates R_1 and R_2, heteronuclear nOe and transverse cross-correlated relaxation rates η_{xy} at 600 MHz of the intrinsically disordered protein Sendai Ntail at 298 K in the sole buffer, in the CAHS-8 hydrogel (red) at 30 mg/mL and with 145 mg/mL and 195 mg/mL of dextran40 viscogen.

The region 441-463, which comprises W443 and W451, is highlighted in grey and the helical domain 476-492 is highlighted in orange.

The last row is the intermolecular Paramagnetic Relaxation Enhancement (PRE) ratio obtained from the ratio of the intensities of the 15N-1H HSQC peaks of Sendai NTail in the CAHS-8 hydrogel in the para- and diamagnetic state of the spin-label attached on position 8 of CAHS-8, that is on its disordered N-terminal chain. A PRE-broadening can be observed in the two zones highlighted which have their relaxation change the most in the CAHS-8 gel.

A paramagnetic TEMPO spin-label was attached to the residue 8 of the N-terminal disordered chain of the ^{14}N CAHS-8 protein and used to form a hydrogel at 30 mg/mL in which the ^{15}N Sendai N_{tail} was observed. Given the high concentration of CAHS-8 in the gel (1.15 mM) and its crowded nature, the percentage of paramagnetic CAHS-8 was optimised to 10 % of the concentration in order to damp the paramagnetic relaxation enhancement (PRE) effect on the entrapped ^{15}N Sendai N_{tail}. The ratio of the intensities of the ^{15}N-^{1}H HSQC peaks of Sendai N_{Tail} in the para- and diamagnetic state of CAHS-8 (fig. 5 bottom) shows a PRE broadening precisely in the 441-463 (grey) and 476-492 (orange) regions, indicating an ensemble averaged spatial proximity of the positions of Sendai N_{tail} with the N-terminal disordered chain of CAHS-8 in the hydrogel. This is consistent with the structure of the fibrils and fibres, formed by the helical domain and in which the disordered chains projected toward the exterior of the fibres form a lateral fuzzy-coat, as well as the general architecture of the CAHS-8 gels, being composed of entangled fibres, discussed in the previous chapter.

In order to understand whether these effects are particular to the hydrogel state or result from the viscous crowding, the hysteresis of CAHS-8 was exploited to study Sendai N_{Tail} at 304 K in both states of CAHS-8. Indeed, as seen in the previous chapter, at 304 K, the state of CAHS-8 beyond the gel point concentration depends on its history. Thus, the experiments were first recorded at 298 K and then at 304 K in which case CAHS-8 remained a hydrogel, as it is below the gel melting temperature of 309.8 K (36.2°C). Then, the temperature was raised to 313 K to melt the gel and then decreased to 304 K again in

which case CAHS-8 remained dilute and thereby forming a dense solution, as this temperature is above the gelation temperature of 300.6 K (27.5 °C), permitting to study both states with the same sample and at the same temperature. The melting of the gel and subsequent absence of gelation was visually confirmed in the NMR tube before and after the experiments, as well as being clear in the spectrometer due to the difference in the signal intensity.

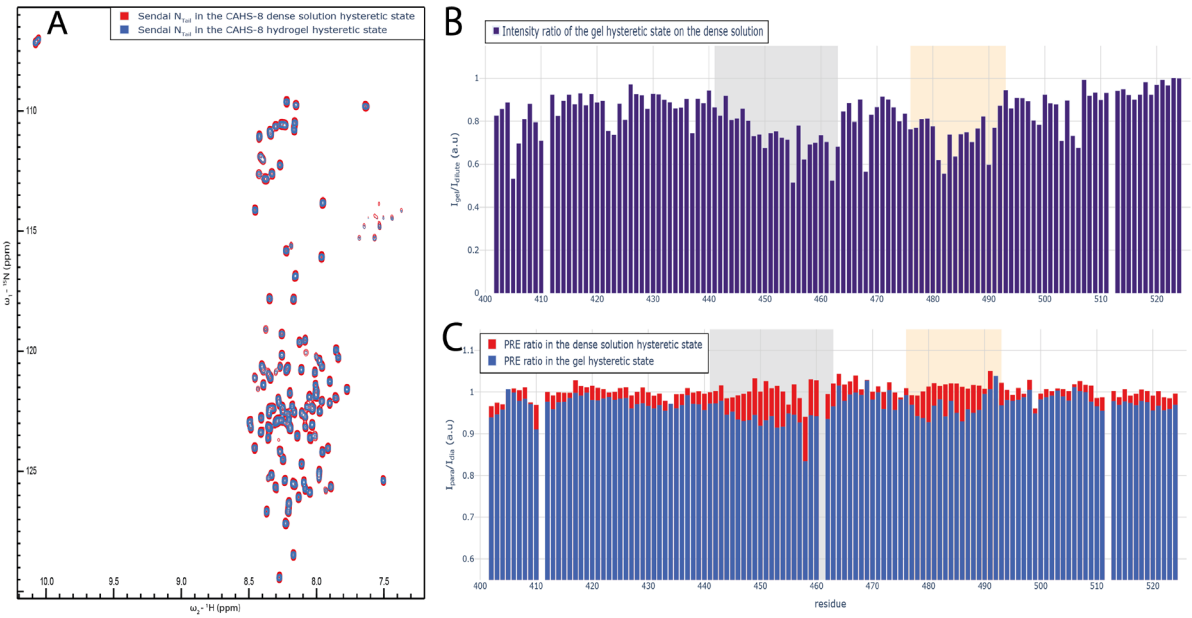


Figure 8.6: The hysteresis of CAHS-8 shows effects on Sendai NTail specific to the gel state.

A and **B**: ^{15}N-^{1}H HSQC spectra of Sendai N_{Tail} at 850 MHz and 304 K (**A**) in the hydrogel hysteretic state of CAHS-8 (blue) and its dense solution hysteretic state (red) as well as their ratio of their normalised intensities (**B**) showing lower intensities in the 441-463 region and the helical domain 476-492 respectively highlighted in grey and orange.

C: Intermolecular PRE ratio observed on Sendai N_{Tail} with the spin-label on position 8 of CAHS-8 at 304 K for the dense solution hysteretic state of CAHS-8 at 30 mg/mL (red) and the hydrogel hysteretic state (blue). Although the PRE effect is more modest than at 298 K (fig. 5 bottom), both highlighted regions have different values than the rest of the proteins in the CAHS-8 gel state.

The comparison of ^{15}N-^{1}H HSQC spectra of Sendai N_{Tail} at 304 K in both CAHS-8 hysteretic states, that is measured in the same sample, at the same temperature, with the same protein concentrations, shows indistinguishable chemical shifts (fig. 6 A), indicating that the conformational sampling of the protein is not perturbed in the hydrogel compared to the dense solution phase. The ratio of their normalised intensities indicates a specific broadening in the hydrogel state of CAHS-8 in the two previously highlighted regions, that

is the 441-463 segment and the helical domain 476-492 (fig. 6 B). In turn, the intermolecular PRE ratios induced by a spin-labelling of position 8 of the N-terminal disordered chain of CAHS-8 on Sendai N_{Tail} in the dense solution hysteretic state at 304 K (fig. 6 C, red) is consistently 1 while in the hydrogel state, it shows a slight but consistent PRE effect over the whole protein with a stronger broadening in these two regions (fig. 6 B, blue).

Altogether, the PRE experiments at 298 and 304 K suggest that the N-terminal disordered chain of CAHS-8 is in contact with the entrapped protein (fig. 5 bottom) and that this effect is specific to the hydrogel state compared to the dense solution (fig. 6 C). Resident time effect, as well as exchange, may contribute to the obtained ratio, with the regions displaying a slower correlation time appearing more affected, which may be supported by the modest PRE broadening observed in the gel at 304 K compared to 298 K. However, the very similar apparent translational diffusion coefficient in both hysteretic states (fig. 6 D, E) suggests that this PRE effect is dominated by an effect of specific spatial proximity, although the ratio of the intensities of Sendai N_{Tail} in the two CAHS-8 states (fig. 6 B) suggest a specific slowing down of the rotational correlation times of the protein in the hydrogel state, which may possibly contribute to the observed PRE effect. Additionally, it is unknown whether the pores of the gel retain the same dimensions at 298 K and 304 K and if the fraction of CAHS-8 protein forming the fibres and gel is the same as well, with such variations possibly affecting the extent of the PRE broadening and explaining the difference at these two temperatures if the cavities are larger at 304 K than at 298 K.

The relaxation experiments (fig. 5) suggests that the pores of the CAHS-8 hydrogel in which encapsulated proteins are likely to tumble are an environment in which slow motions occurring over several nanoseconds are particularly slowed down, as R_2 and η_{xy} rates show a significant increase, while motions occurring over the nanosecond timescale are less affected, as seen in the modest increase of the R_1 rates, and while finally the picosecond timescale motions are not much affected, as the nOe values are close to that in the buffer (fig. 5). This suggests that the spatial confinement imposed by the CAHS-8 pores (fig. 4 A) yields a significant crowding in these cavities, with this crowding resulting from not only the self-crowding of the encapsulated protein Sendai N_{tail}, but also due to the walls of the pores being covered by at least the CAHS-8 N-terminal disordered chain, and

probably the C-terminal one too as per the proposed structure of the CAHS-8 fibres, as well as probable free CAHS-8 species evolving inside the lumen of the hydrogel pores. It is worth reminder that the N-terminal disordered chain of CAHS-8 is 94 residue long, which is almost the size of the whole Sendai N_{tail} protein used here, and therefore likely extended several nanometer into the pores.

As such, the encapsulated proteins would evolve in a confined, dense and crowded pores, with the proteins on the edges of the pores being in frequent contact with the disordered termini of the CAHS-8 lining the walls, all hindering these proteins from freely experiencing large reorientational motions, hence the increase in R_2 and η_{xy} rates, perturbing smaller amplitude motions occurring on the nanosecond timescale, hence the modest increase in R_1, but not hindering their inherent flexibility or librational or internal motions hence the only slightly affected nOe rates. Moreover, the chemical shifts are unperturbed which indicates a similar conformational sampling as in the buffer and therefore, and thus, an absence of crowding-induced conformational collapse. In turn, the R_1 of water values probably present a significant contribution of exchange between the different water species which gives the impression that the gel is more viscous than what it really is, otherwise the nOe rates would be higher.

Finally, the contacts with the disordered termini of CAHS-8 can give rise to exchange with particular segments of the entrapped proteins (fig. 5), as observed in the 441-463 region of Sendai N_{tail}, and may stabilise transiently populated conformational state as this region displays a weak helical propensity[586]. Moreover, the very periodical profile of the 451-462 segment of this region, appearing to have a period of 3 to 4 residue, suggest a helix interacting on one of its sides, possibly with the disordered termini of CAHS-8. Nonetheless, the measurement of CPMG relaxation dispersion and CEST did not yield any dispersion or clear effect, which suggests that there is no detectable exchange at least in the millisecond to second timescales probed by these two experiments.

7.1.4 Interactions are perturbed in the hydrogel

The interaction of the Sendai N_{Tail} disordered domain with the C-terminal domain of the Sendai phosphoprotein (PX) has been demonstrated via CPMG relaxation dispersion by Schneider *et al.* to funnel the conformational state of Sendai N_{Tail} via the stabilisation of one of its helical conformers[33]. This interaction can clearly be observed by its induction of ^{15}N dispersion to an extent dependent on molar ratio of Sendai N_{Tail} and PX at 298 K, with the strongest effect having been observed at 8 % of PX[33]. The dissolution of the previously studied CAHS-8 hydrogel with ^{15}N-labelled Sendai N_{Tail} in it, subsequent addition of 8 % of PX, gelation of CAHS-8 and measurement of CPMG relaxation dispersion at 298 K and 600 MHz reveals an absence of dispersion in the expected positions, namely 476, 477, 484, 486, 488 and 489[33] (fig. 7 A). This indicates that the Sendai N_{Tail}-PX reaction has been repressed in the CAHS-8 hydrogel or has been shifted over timescales to which ^{15}N CPMG relaxation dispersion is not sensitive.

Another well studied interaction is the interaction of the intrinsically disordered C-terminal domain of the Nucleoprotein of the Measles virus (Measles N_{Tail}) with the C-terminal three-helix bundle domain of the phosphoprotein of the Measles virus (P_{XD}), which has been characterised by CPMG relaxation dispersion by Guseva, Milles *et al.*[816]. ^{15}N-labelled Measles N_{Tail} was therefore encapsulated in a 30 mg/mL CAHS-8 hydrogel with 10 % of its partner P_{XD} and CPMG relaxation dispersion was measured at 298 K and at both 600 and 850 MHz with CPMG frequency ranging from 31 to 1000 Hz. The global fit of the two-field data for the residues 483, 484, 486, 491 and 507 yields $k_{\mathrm{ex}} = 329 \pm 69\ \mathrm{s}^{-1}$ with $p_B = 1.52 \pm 0.26$ % in the gel (reduced $\chi^2 = 1.13$, fig. 7 B) as compared to $k_{\mathrm{ex}} = 276 \pm 50\ s^{-1}$ with $p_{\mathrm{B}} = 7.3$ % outside[816], which shows a broadly similar reaction rate and a strong decrease of the population of the complex. However, the same experiment performed at 600 MHz in a CAHS-8 hydrogel at 22 mg/mL for residues 483, 484, 487, 489, 490, 505, 506 and 507 yields $k_{\mathrm{ex}} = 1023 \pm 300\ \mathrm{s}^{-1}$ with $p_B = 2.81 \pm 5.21$ % (reduced $\chi^2 = 0.9$). As the experiment was performed at only 1 field at this CAHS-8 concentration, the p_B estimation may not be reliable and is given as an indication to be compared to the more reliable, 2-fields derived p_B at 30 mg/mL CAHS-8. Moreover, the exchange process appears to be in the fast intermediate regime, which further hinder the precision of the estimation of these parameters.

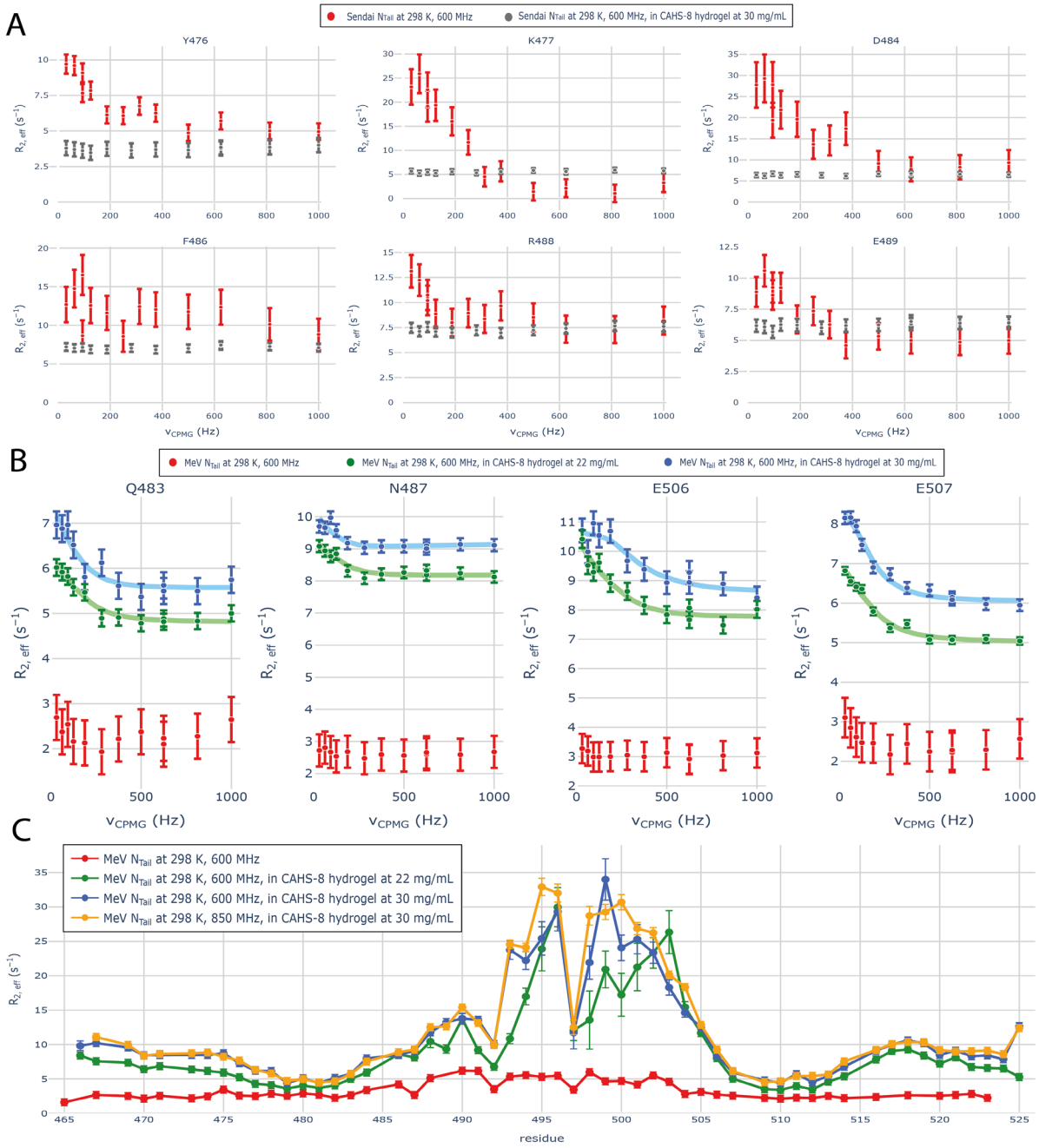


Figure 8.7: Interactions are perturbed in the CAHS-8 hydrogel.

A: 15N CPMG relaxation dispersion plots of Sendai NTail with 8 % of its partner PX at 298 K and 600 MHz in the CAHS-8 hydrogel at 30 mg/mL (grey) compared to the dispersion in the same conditions with CAHS-8 from Schneider et al. 2015 (red).

B: 15N CPMG relaxation dispersion plots of Measles NTail with 10 % of its partner PXD at 298 K at 600 MHz and in the buffer (red), in the CAHS-8 hydrogel at 30 mg/mL at 600 MHz (green) and 850 MHz (blue). While $k_{ex} = 276 \pm 50\ s^{-1}$ and $p_B = 7.3$ % in the buffer as per Guseva, Milles et al. 2020, the fit of the interaction in the CAHS-8 hydrogel yields $k_{ex} = 1023 \pm 300\ s^{-1}$ and $p_B = 2.81 \pm 5.21$ % (reduced $\chi^2 = 0.9$) in the 22 mg/mL gel and $k_{ex} = 330 \pm 69\ s^{-1}$ and $p_B = 1.52 \pm 0.26$ % (reduced $\chi^2 = 1.13$) in the 30 mg/mL gel.

C: Comparison of the $R_{2,0}$ at 1000 MHz CPMG frequency of Measles NTail at 298 K and 600 MHz in the buffer (red), in a CAHS-8 hydrogel at 22 mg/mL (green) or 30 mg/mL (blue) and at 850 MHz in 30 mg/mL CAHS-8 gel (orange).

Accordingly, the comparison of the $R_{2,0}$ rates obtained at the 1000 Hz CPMG frequency in these different experiments (fig. 7 C) shows an increase in the transverse relaxation dependent on the concentration of the CAHS-8, with the rates of Measles N_{Tail} being higher at 30 mg/mL CAHS-8 than at 22 mg/mL as expected due to the increased experienced nanoviscosity affecting in particular the segmental, several nanoseconds long motions which dominate R_2[439]. Interestingly, unlike in the sole buffer, the obtained $R_{2,\mathrm{eff}}$ rates of Measles N_{Tail} in both the 22 mg/mL and 30 mg/mL CAHS-8 hydrogels do not show a bell-shaped profile along the sequence of the protein, with rates approximately twice as fast in the N- or C-terminal residues than around respectively residue 480 and 510. Such behaviour has also been observed for a longer construct of the same protein in liquid-liquid phase separation droplets and in molecular dynamics simulation in highly self-crowded environments, indicating a restriction of the high-amplitude fast motions normally observed in the flexible termini[604]. This is supported by the C-terminal amino-acid showing an increase of the $R_{2,0}$ rates compared to the previous residues at 30 mg/L CAHS-8 but not at 22 mg/mL, indicating an even stronger damping of the fast and intermediate motions and translating by an increased sensitivity of its relaxation to slower motions.

In turn, the increase of k_{ex} of the Measles N_{Tail}-P_{XD} in the 22 mg/mL hydrogel of CAHS-8 compared to the buffer and its subsequent decrease at 30 mg/mL to similar level to what it is in the buffer suggest a similar trend to what has been reported upon addition of large polymer crowders, where k_{on} initially increases, supposedly due to attractive depletion interactions despite the increase in the nanoviscosity, before eventually decreasing at higher crowder concentration[817,781,782]. The hydrogel of CAHS-8 therefore appears to have a neutral or increasing effect on k_{ex} depending on the concentration, however, the hydrogel also appears to limit the interaction as the population of the complex p_{B} consistently decrease, probably due to physical segregation of the partners.

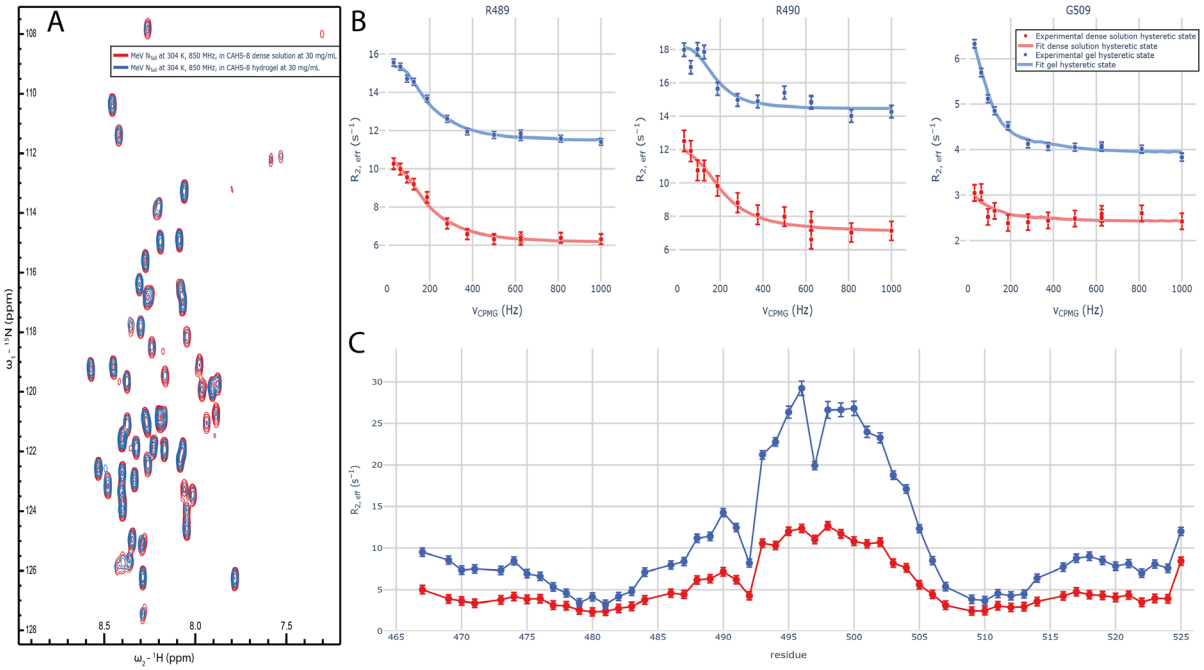


Figure 8.8: The hysteresis of CAHS-8 demonstrates that the hydrogel phase has a specific effect on the interactions.

A: 15N-1H HSQC of Measles NTail with 10 % of its partner PXD at 304 K encapsulated in the 30 mg/mL CAHS-8 (blue) or in the dense solution phase of CAHS-8 at 30 mg/mL (red).

B: Associated 15N CPMG relaxation dispersion plots of Measles NTail with 10 % of its partner PXD at 304 K and 850 MHz in the gel (blue) and dense solution phase (red) of CAHS-8 at 30 mg/mL for R489, R490 and G509. The dispersion profiles show clearly the slow-down of the exchange process in the gel and the subsequent fitting of these profile (lines), yields $k_{ex} = 267 \pm 42\ s^{-1}$ in the hydrogel and $k_{ex} = 564 \pm 113\ s^{-1}$ in the dense solution phase of CAHS-8 (reduced $\chi^2 = 1.64$ and 0.76).

C: Comparison of the $R_{2,0}$ at 1000 MHz CPMG frequency of Measles NTail with 10 % PXD at 304 K and 850 MHz in the gel (blue) and dense solution (red) phase showing a clear increase of the correlation times of the protein in the gel phase.

In order to decorrelate the effect of CAHS-8 hydrogel itself from the effect of the viscosity, the hysteretic properties of CAHS-8 were used to probe this interaction between Measles N_{Tail} and P_{XD} by CPMG relaxation dispersion at 850 MHz and 304 K in consecutive experiments with the same sample in the hydrogel and then the dense solution hysteretic states of CAHS-8 at 30 mg/mL, i.e 1.14 mM. The comparison of ^{15}N-^{1}H HSQC spectra of Measles N_{Tail} obtained entrapped in the CAHS-8 gel or tumbling in the dense solution state show almost undistinguishable chemical shifts, indicating once again that the conformational sampling of the protein is not perturbed in the gel (fig. 8 A). The global fit of residue 489, 490, 494, 495 and 509 yields $k_{ex} = 267 \pm 42\ s^{-1}$ and $p_B = 1.54 \pm 0.19\%$ (reduced $\chi^2 = 1.64$) in the hydrogel phase and $k_{ex} = 564 \pm 113\ s^{-1}$ and $p_B = 1.2 \pm 0.1\%$

(reduced $\chi^2 = 0.76$) in the dense solution of CAHS-8 (fig. 8 C). This demonstrate that the CAHS-8 hydrogel itself, not simply the viscosity, slows down this reaction by a factor of approximately two. These dispersion experiments were measured only at one field and therefore, the p_B estimation may not be reliable, although they are similar to the one obtained in the same sample at 298 K from experiments at two fields.

Moreover, the comparison of the $R_{2,\text{eff}}$ rates obtained at the 1000 Hz CPMG frequency shows that the transverse relaxation of Measles N_{Tail} is approximately 2 to 2.5 times higher in the CAHS-8 hydrogel (fig. 8 C) which demonstrates that the hydrogel, and not only the induced increase of the nanoviscosity, also results in a specific retardation of the rotational correlation times of the encapsulated proteins.

7.2 Discussion

Using DLS and DSC, the gelation of CAHS-8 into a thermoreversible physical hydrogel and its melting were characterised, permitting to draw a phase diagram and revealing a thermal hysteresis between these two-phase transitions (ch. 7 fig. 2). The entanglement of the fibres forming the hydrogel network was observed by AFM and EM (ch. 7 fig. 1) and was studied by SAXS as well as both solid state and solution NMR (ch. 7 fig. 3-5). In agreement with the model of the fibres derived from the crystal structure of the fibrils present in the previous part (ch. 6 fig. 9), these experiments showed that the disordered terminal chains of CAHS-8 can tumble freely in the gel despite being anchored as part of the gel structure and coexist with free species of CAHS-8 which diffuse supposedly in the pores of the gel.

These pores were imaged by AFM showing their ability to accommodate large proteins and assemblies like Immunoglobulin G and Tobacco Mosaic Viruses (ch. 8 fig. 1). They were also demonstrated by solution NMR to allow the tumbling of folded and disordered protein without changing their fold and conformational sampling but reducing their translational diffusion by a factor of two (ch. 8 fig. 2) which was shown for Sendai N_{Tail} to be confined (ch.8 fig. 4). In turn, the relaxation of this IDP was shown to be modified in the hydrogel, showing two regions with fast R_2 but almost unperturbed nOe rate

constants which were established by intermolecular PRE to be in contact with the N-termini of CAHS-8 (ch. 8 fig. 5). The thermal hysteretic property of CAHS-8 was used to study the dynamics of Sendai N_{Tail} in both the hydrogel and a dense solution of CAHS-8 at the same concentration and temperature in successive experiments with the same sample, revealing similar translational diffusion constants in both CAHS-8 states but specific line-broadening and PRE contact in the hydrogel state (ch. 8 fig. 6).

Interactions between proteins were also studied in the hydrogel of CAHS-8 using CPMG relaxation dispersion, in particular the C-terminal domain of Sendai virus nucleoprotein, N_{Tail} with its partner PX (the C-terminal domain of the phosphoprotein of the same virus) being shown to be abrogated in the hydrogel of CAHS-8 (ch. 8 fig. 7). The interaction of Measles N_{Tail} with its partner P_{XD} in the hydrogel was shown to be perturbed in a dense gel as well, showing in particular a population of the complex being significantly reduced, while its R_2 rates were shown to increase in function of the concentration of CAHS-8 used to form the hydrogel (ch. 8 fig. 7). In turn, the hysteresis of CAHS-8 was used to demonstrated that at the same temperature, same concentration of proteins and in the same sample, the interactions Measles N_{Tail} was not slowed down in the hydrogel but shown a significantly reduced population of the interaction complex, despite the transverse relaxation of its interacting region being more than twice as fast as in the dense solution of CAHS-8 (ch. 8 fig. 8).

Finally, the freeze-drying of solutions of CAHS-8 was shown to result in a highly porous fibrous aerogel which has a macroscopic helical shape and was shown by AFM to have a coexistence of amorphous gel regions (ch. 7 fig. 7) and regions of ordered mostly parallel fibres resembling the crystallised fibres, with the diffraction of the aerogel showing polycrystalline pattern X-Ray diffraction pattern and an absence of glass transition in DSC (ch. 7 fig. 8).

7.2.1 The gelation of CAHS-8

As discussed in the previous part of this thesis, at low concentration and room temperature, CAHS-8 forms a transient and low-populated coiled-coil dimer, the DCC. When the temperature decrease, the thermodynamic equilibrium shifts and favour DCC, which increase its population. In turn, the second coiled-coil, the FCC, is inferred to form after the DCC and to stabilise it. As such, the probability of encounter and thus of formation of these coiled-coils increase with concentration of protein, which results in an increased formation of both the DCC and FCC and with them, fibrils and fibres.

This allows to explain the temperature and concentration dependence of the phase diagram of CAHS-8, as well as its shape (ch. 7 fig. 2 C). In the vocabulary of the percolation theory of gelation discussed in the introduction of this second part of the thesis, the fibres would be the chain of the hydrogel while its junctions are their physical crossing of each other and entangling (ch. 7 fig. 1 A-C). As such, low temperature or high concentration results in increasing fibrillation and when a significant number of fibres are present, the junction occupancy is high and regions of the sample may be described as local gel patch, yielding inhomogeneities that manifest as oscillation in the measured DLS intensity autocorrelation function[726,732]. This corresponds to the slight gelation region of the phase diagram (ch. 7 fig. 2 C), which is also characterised by a high macroscopic viscosity due to the presence of fibres, and is the region of coexistence between local gel phase and soluble phase predicted by the percolation theory[721,800].
In turn, once the percolation threshold has been reached because enough fibres are intricated, CAHS-8 gelates, which corresponds to the gel part of the phase diagram. This gelation, and the gel state, is apparent by the loss of ergodicity manifesting as a significant suppression of the initial value of the autocorrelation function, the change of its shape as well as the apparition of strong inhomogeneities due to the difference in pores and the non-ergodicity (ch. 7 fig. 2 A)[726,730,732].

In the DLS correlogram, the temperature was measured every 5 °C starting from 45 °C to 5 °C, with every measurement being characterised as soluble, slight-gelation or gel, which therefore didn't permit to obtain the precise temperature of gelation at each concentration. This was obtained precisely via DSC for CAHS-8 at 30 mg/mL, with an

average gelation temperature of 27.5 °C and melting of the gel of 36.2 °C. Moreover, the changes in enthalpy ΔH associated with the gelation and its melting could be characterised at respectively -2.72 ± 0.05 kJ/mol and 5.96 ± 0.05 kJ/mol on average, indicating respectively enthalpy and entropy driven reactions. The occurrence of thermal hysteresis in the hydrogel of CAHS-8 is not unique, as other protein physical hydrogels such as gelatin gels[718,818] or even polysaccharides ones such as κ-carrageenan gels[819,820] also display similar hysteresis, which corresponds in both case to the difference between the helix-coil transition temperature and its thermal denaturation temperature. However, the process is probably different in CAHS-8. Indeed, the changes in enthalpy ΔH associated with the gelation and the melting are respectively on the order of -280 kJ/mol[821] and 125 kJ/mol[822], indicating significant changes for gelatin while the ΔH for CAHS-8 are much smaller, which suggest that the melting of the CAHS-8 hydrogel does not mean the complete denaturation of all the CAHS-8 fibres but instead, a disruption of their entangled junctions high enough to disrupt the macroscopic gel network.

Unfortunately, the hydrogel and the fibres couldn't be studied neither by carbon nor proton detection MAS ssNMR due do their very broad linewidths, despite sample optimisation and deuteration. This likely arise from inhomogeneous line-broadening as not only fibres are typically heterogeneous with hydrogen-bonds of different lengths[823], but also due to the different form of CAHS-8 coexisting in the hydrogel: monomers, dimer, fibril or fibres may all have different chemical shifts which may be exacerbated by the various degree of entanglement, interactions or crowding in the pores between these different species.

7.2.2 Formation and crystallinity of the aerogel

The formation of an aerogel by freeze-drying of soluble solutions of CAHS-8 at low concentration and in buffer of low ionic strength contained in 15 mL falcon tubes is surprising but can be explained by ice-templating during the freezing. Ice-templating, also called directional freezing, occurs when water freezes along a defined axis, forming elongating needle-shaped ice crystals which separate and locally concentrate the particle between them in the remaining liquid-water and can drive their association. As such, ice-

templating can result in the formation of fibres or highly porous fibrous aerogels if the material has the ability to form these structures, which made this method commonly used for the formation of aerogels[824–826].

The obtained aerogel of CAHS-8 and the coexistence of gel-like amorphous regions (ch. 7 fig. 7) and more crystalline regions (ch. 7 fig. 8) as well as the macroscopic long-range range order of the material, giving it a helical shape visible to the eye, indicate that this aerogel is semi-crystalline and highlights once again that CAHS-8 can either aggregate into randomly arranged fibres, yielding gels, or into periodically arranged fibrillar structures, yielding crystals. Considering the results of the previous part of the thesis where the rate of dehydration was shown to control the formation and morphology of crystallised fibres, it is possible that the different dehydration rates during the freezing are also responsible for these different morphologies of CAHS-8. Indeed, during the directional freezing, as needle-shaped ice crystals grow by drawing in the nearby supercooled water, the CAHS-8 monomers may be concentrated and dehydrated at different rates in different regions, thus resulting in either fibrous gels, crystallites or periodically arranged fibrillar assemblies, therefore resulting in a semi-crystalline aerogel.

7.2.3 Organisation, viscosity, crowding and confinement in the hydrogel pores

In the model of the fibres derived from the crystal structure of the fibrils presented in the previous part (ch. 6 fig. 9), disordered termini of certain CAHS-8 monomers would remain free, forming essentially a fuzzy coat around the fibrils and therefore would be able to tumble. This was demonstrated by the diffusion-NMR experiments which show clearly a static species of the disordered termini coexisting with a free diffusion species (ch. 7 fig. 4 B, C). As the fibres constitute the frame of the hydrogel, this means that its pores are lined with these disordered termini which, although spatially anchored due do the helical domain being engaged in the fibres and therefore not translating, are continuously experiencing rotational diffusion and disordered chain motions, therefore permitting their detection by solution-state NMR.

The determination of the crowding and viscosity in the hydrogel is unfortunately not straightforward. Indeed, a significant fraction of CAHS-8 is engaged in the fibres network of the gel and as such, constitute a static crowder that does not contribute to the crowding and the nanoviscosity in the centre of the pore. However, free species of CAHS-8 tumbling in the pores were observed (ch. 7 fig. 4 B, C) and would contribute to both crowding and nanoviscosity, in particular if not only monomers but also dimers or fibrils are present in the cavities.

In turn, if the individual fibres constituting the fibres network of the hydrogel are static and therefore perfectly insensitivity to water flow, they would not contribute to the shear viscosity either. Nevertheless, such a strong immobility is unlikely, especially as the fibres were regularly imaged to display some local curvature instead of being perfectly straight and stiff (ch. 7 fig. 1) and as such, they may show some limited freedom of movement that can cause resistance to the flow and thus contribute to the shear viscosity of the interstitial water.

Finally, the disordered termini of CAHS-8, both from the monomers engaged in the hydrogel fibres network and the free species, are most probably strong contributor to the increased viscosity in the hydrogel. Indeed, hydrogel retain water thanks to their polar groups thereby perturbating water dynamics by decreasing the number of available hydrogen-bond per water molecule and extending their lifetime, with dense gel perturbating their interstitial network to the extent that no water showed bulk properties[793,792,794,795]. The disordered termini of CAHS-8 are rich in polar groups, in particular the C-terminus and therefore are expected to perturb the water hydrogen-bond network. Additionally, confinement show similar effects on water dynamics but are unlikely to happen in the CAHS-8 pores as they appear consistently wide of several hundred of nanometres (ch. 7 fig. 1, ch. 8 fig. 1), which is large enough to accommodate diffusing proteins much bigger than water molecules without perturbing their fold or conformational sampling (ch. 8 fig. 2). In turn, hydrogels are permeable and as such, are not expected to forcibly keep water sealed in unlike for instance liposomes[795,796].

7.2.4 Effects of the pore environment on encapsulated proteins

Immunoglobulin G (IgG) proteins are 14.5 x 8.5 x 4 nm in size[808,809] and thus should have enough space to tumble within the pores. Yet, the imaging in solution of a hydrogel of CAHS-8 to which IgG proteins were added show the gel being coated with IgG (ch. 8 fig. 1 C, D). This is surprising as the IgG are soluble proteins and AFM images adsorbed objects, which means that these IgG clusters are deposited onto the gel and weren't washed off. This suggests that the hydrogel of CAHS-8 can catch and interact with passing proteins, possibly via its disordered termini, while freely diffusing IgG are in the solution and are therefore invisible to AFM.

The intermolecular PRE between the spin-labelled disordered termini of the ^{14}N gel and the entrapped ^{15}N Sendai N_{tail} as well as its relaxation in the gel supports this model. Indeed, the R_2 relaxation rates of the protein show a strong increase in two regions (441-463 and 476-492) in the gel while the rest of the chain doesn't change significantly (ch. 8 fig. 3), and these regions display a specific PRE effect but not the rest of the protein (ch. 8 fig. 5). In turn, the comparison of the R_2 and η_{xy} relaxation rates measured at 600 Mhz and 700 MHz supports an exchange contribution in these two regions. Altogether, these NMR data indicate that these two regions of Sendai N_{tail} are in contact with the disordered termini of CAHS-8 in the hydrogel as per the PRE, thus increasing their correlation times and triggering a presumably intermolecular exchange process faster than 1500 s^{-1} as per the $R_{1\rho}$ spin lock frequency used. However, the lack of chemical shift difference between the protein in the buffer and in the hydrogel (ch. 8 fig. 1 C) argues against a fast exchange, but no sign of dispersion could be observed either using CPMG relaxation dispersion.

As such, these dangling disordered termini of CAHS-8 lining the hydrogel interior appear to hinder the large reorientational motions of the proteins they interact with but not to affect smaller scale motion as per the relaxation, which is strikingly different from the relaxation observed with 145 or 195 mg/mL of dextran40 crowder (ch. 8 fig. 5), which highlights that the actual crowding in the pores of the hydrogel is lower than in these

dextran crowder experiments, as the total 30 mg/mL concentration of CAHS-8 is not only lower but a significant part of the protein is immobilised in the fibres network of the gel.

An alternative explanation can be that the fibrils or fibres species diffusing in the hydrogel pores would have a large rod-shaped excluded volume and as such, may similarly perturb the large reorientation motions of neighbouring proteins but not the faster motions. However, the use of the hysteretic properties of CAHS-8 permitted to study the effect on proteins either in the hydrogel or in a dense solution of CAHS-8, at same temperature and using the same sample so with the same total concentration of proteins, which confirm gel specific effects compared to the dense solution, with these two same regions showing a specific line broadening in the gel compared to the dense solution as well as slight PRE effect (ch.8 fig. 6). In such a dense solution where the junctions between the fibres have been broken enough to melt the gel, but as previously discussed, would not break all the fibres, the crowding is expected to be higher than in the hydrogel state but the specific effects seen in the hydrogel state support an effect of the dangling disordered termini more than a crowding-based effect.

7.2.5 Slow-down of the interactions in the hydrogel

The $R_{2,0}$ rates of measles N_{Tail} in the presence of 10 % P_{XD} obtained from the CPMG relaxation dispersion experiments at 298 K show a dependence on the concentration of CAHS-8 protein used to form the hydrogel, with the rates appearing slightly faster in the termini at 30 mg/mL CAHS-8 than at 22 mg/mL However, while the k_{ex} appears faster at in 22 mg/mL CAHS-8 hydrogel than at 30 mg/mL as expected, it also is faster than in the free solution. This is similar to the increase of k_{on} of interactions observed in the semi-dilute regime of PEG before decreasing at higher concentration by Kozer *et al.*[781] as well as Zosel *et al.*[782], which was attributed to the attractive depletion interaction and explained as per the Berezhkovskii-Szabo model[783] as discussed in the introduction of this part. In this model of protein interaction in crowded media, collisions and nonspecific reversible binding with the crowder slows-down the diffusion and thus represses the reaction and competes with the attractive short-range depletion interaction which

favours the reaction, with the balance of the two, depending on the properties and size of the crowder, determining the effect on the k_{on}.

Using this model, in the CAHS-8 hydrogel the crowder would be the disordered chain lining the pores of the hydrogel as well as the free CAHS-8 species diffusing in them, which may include fibrils or fibres which would cause a significant depletion layer, as well as the self-crowding of measles N_{Tail} and P_{XD}. While at 22 mg/mL, the balance between the increase in viscosity and depletion interaction would accelerate their interaction, at 30 mg/mL the balance may result in next to no effect on the observable k_{ex}. In turn, the decrease of the population can also be explained as physical separation of the species due to confinement within pores or the disordered chain of CAHS-8 as well as possible preferential interaction with one of the proteins, thus effectively distancing the two partners. This is further supported by the only slight difference between the $R_{2,0}$ of measles N_{Tail} in the CAHS-8 hydrogels at 22 mg/mL and 30 mg/mL and the absence of difference between their chemical shits, which indicate the rotational diffusion of measles N_{Tail} is not modified but slowed-down due to increased nano-viscosity, and thus points towards perturbating effects on the self-diffusion process instead.

Such model also permits to explain the fact that k_{ex} of the interaction is twice faster in the dense solution than in hydrogel at 304 K with the same total concentration of proteins but the apparent translational diffusion coefficient staying being identical (ch.8 fig. 8): the amount and nature of the crowding is different in the hydrogel and in the dense solution.

Indeed, as previously discussed, while in the hydrogel most of CAHS-8 is engaged in the fibres gel network with a fraction diffusing freely in the pores (ch. 7 fig. 4), in the dense solution the gel network is broken and CAHS-8 is diffusing in the medium, either as monomer, dimer, fibrils or fibres, yielding a more crowded solution.

As such, the large depletion layers induced by the CAHS-8 fibres diffusing in the dense solution results in the attractive depletion interactions crowding together measles N_{Tail} and P_{XD} due to overlap of their depletion layers and osmotic pressure created by the exclusion of the CAHS-8 fibres from these local depleted local zones. Such model is all the more likely as rod-shaped objects, such as fibres, are well known to be more efficient

depletants than spherical ones of the same volume[777,778]. As such, in these fibre-depleted local regions, called cavities in the Berezhkovskii-Szabo model, the following effects would occur compared to the hydrogel pores[777–780,782,783]:

- The interaction between measles N_{Tail} and P_{XD} would not be hindered anymore because of the presence of CAHS-8 fibres species and are favoured by the attractive potential of the depletion interaction, resulting in an increase of k_{on}.
- The different reorientational motions of measles N_{Tail} would not be restricted and no interaction would occur with the CAHS-8 disordered chains besides at the boundary between the regions, thus explaining the $R_{2,0}$ rates being two to three times smaller in the dense solution than in the hydrogel.
- Measles N_{Tail} would translationally diffuse at two different speeds: a diffusion in the cavities depleted of CAHS-8 fibres which would be close in term of speed to what it would be in a CAHS-8 free environment and a slower diffusion outside of the cavities due to viscosity arising from the CAHS-8 crowding. As only one part of the decay curved was probed, the apparent diffusion coefficient has two contributions. In turn, in the hydrogel pores, the first mode is unlikely but in the second, measles N_{Tail} would be less crowded as discussed and diffuse faster, yielding a similar apparent diffusion coefficient.

As a final note, an exchange contribution could also theoretically contribute to this observed difference in $R_{2,0}$ between the hydrogel and dense solution hysteretic states but is unlikely due to the $R_{2,0}$ at 298 K being similar at 600 MHz and 850 MHz apart from a few residues in the helix. Moreover, compared to the relaxation in buffer, the $R_{2,0}$ relaxation rates of measles N_{Tail} do not show regions with local increase of relaxation unlike Sendai N_{Tail}, which indicate that CAHS-8 does not interact with the former like it does with the latter. Thus, these effects can be satisfactorily explained by the different nature of the crowding between the hydrogel pore and the dense solution

7.2.6 Model of the molecular basis of the desiccation resistance of tardigrades

In tardigrades, during dehydration, the concentration of CAHS-8 and other CAHS protein would automatically increase, therefore triggering the formation of fibres and possibly of

a hydrogel if the concentration is high enough, which would therefore dry as a xerogel in the dry state of the tardigrade. Although the mechanisms of protection *in vivo* in the dry tardigrade may be different from the effect of the hydrogel presented here, evaluating these results in light of the different proposed mechanism of dehydration protection (ch. 1, section 2.3.1.1) suggests that these findings are relevant.

First, the glass transition hypothesis proposes that molecules, in particular trehalose would vitrify upon dehydration, effectively filling the cytosol with a glassy matrix which would physically immobilise and separate the biomolecules, thus suspending the metabolism and preventing their aggregation and denaturation[239,246–248]. This hypothesis was also proposed for CAHS proteins on the basis of an apparent glass transition[614] but has been disputed as artefactual[805] which is in agreement by the lack of glass transition observed in the DSC experiments on the CAHS-8 aerogel (ch. 7 fig. 8 E). Nonetheless, as the pores of the CAHS-8 hydrogel have been established to be lined with disordered termini that can interact with proteins (ch. 8 fig. 1, 5), the drying of the hydrogel and with it of protein intercalated between the disordered termini would constitute a glass-like material in its structural sense, especially as xerogels experience a significant shrinkage that could effectively tightly crowd entrapped biomolecules in a matrix composed of the disordered termini of CAHS-8[718,827,828]. As such, these molecules would effectively be protected from aggregation and possibly denaturation arising from desolvation due to desiccation. This is similar to the molecular shield mechanism proposed for the LEA desiccation-protective proteins in which they would act as crowders upon dehydration, intercalating between the biomolecules and preventing their aggregation and denaturation[295,297,305,306], which appear to already be happening in the 30 mg/mL gel as interactions were shown to be restricted from the significant decrease of the population of the interacting complex, indicating a decrease of the physical contact between the partners (ch. 8 fig. 7, 8).

Secondly, the water replacement hypothesis proposes that the protectant against desiccation forms hydrogen-bonds with biomolecules upon dehydration, thereby allowing them to maintain their fold and prevent their aggregation by replacing these lost

bonds with water[252,253]. While no evidences have been found suggesting this model in CAHS-8, the disordered termini, proven to be in contact with entrapped proteins in the hydrogel, are rich in polar amino-acids, in particular C-terminal chain, and may therefore be able to solvate the biomolecules in the dry tardigrade cells.

Finally, the preferential hydration theory proposes that the protectant preserves or even actively favours the water hydration layer needed by the biomolecules during dehydration, thereby indirectly protecting them[257–259]. No direct evidences have been found to corroborate this mechanism for the CAHS-8 mediated protection of tardigrade. However, the protein was demonstrated to form a hydrogel which are by definition able to retain water using the abundant polar groups in the disordered chains of CAHS-8, and therefore, disrupt the water hydrogen-bonds network via the preferential hydration of these groups and the viscosity-mediated increase of their lifetime, with dense PEG hydrogels for instance affecting retained water molecules, effectively amounting to an absence of bulk water[761,792–795]. Dried tardigrades in their tun state have only have 1 to 3 % remaining water and the porous nature of gels do not make for a good water-retaining material upon dehydration[367,368]. However, the bulk water molecules evaporate before the hydration layer molecules due to their hydrogen-bonding which, along the increased intracellular viscosity due to the presence of the fibres, may slow-down the dehydration of the gel and permit the tardigrade additional time to carry the cellular process necessary for the survival to dehydration.

7.3 Material and methods

Unless specified otherwise, the experiments were performed in a buffer containing 50 mM NaP, pH = 6.5, 500 mM NaCl.

7.3.1 Sample expression and purification

CAHS-8 was expressed and purified as described in the previous part. The gels were formed by weighing the freeze-dried protein and resuspending at 42 °C in the corresponding volume of buffer to obtain the desired concentration. The gelation was triggered by letting the sample cool to room-temperature.

Sendai N_{tail} corresponds to the residues 401-524 of the nucleoprotein of the Harris strain of Sendai virus and PX corresponds to the residues 474-568 of the phosphoprotein of the same strain with both being expressed and purified as described[33,829].

Measles Ntail corresponds to the residue 465-525 of the nucleoprotein of the Edmonston strain of the measles virus and XD to the residues 457-507 of the phosphoprotein of the same strain with both being expressed and purified as described[816,830]

The IDR domain of ANP32a was expressed and purified as described[831].

The SH3A domain of CD2AP was expressed and purified as described[832].

All the concentration were verified using a Thermo Fischer NanoDrop 2000. For the gel samples of CAHS-8, this was performed by taking a volume of the heated sample and diluting it 10 times, yielding concentrations below the gelation threshold.

7.3.2 Negative-stain Transmission Electron Microscopy

As in the previous chapter, the samples were applied to the clean side of a carbon grid on mica and then stained using 2% v:v of sodium silicotungstate (SST) at pH = 7. The electron micrographs were taken using a FEI Tecnai T12 microcscope operated at a tension of 120 kV and equipped with a GATAN Orium 1000 camera. The magnification was set to 23 000 times. The experiments were performed at the Institut de Biologie Structurale in collaboration with Emmi Mikkola.

7.3.3 Atomic Force Microscopy

The experiments were performed at the Institut de Biologie Structurale, Grenoble, France, in collaboration with Jean-Marie Teulon and Dr. Jean-Luc Pellequer.

AFM images were recorded in liquid mode on a Bruker multimode 8 microscope equipped with a Nanoscope V controller.

For imaging in liquid mode, a silicon cantilever nitride with characteristics k = 0.7 N/m, Fq = 150 kHz, r = 20 nm. The ScanAsyst mode was used, using a scan rate of 1 Hz and a setpoint of approximately 0.1 V.

For imaging in air mode, etched silicon cantilevers were used with characteristics k = 3 N/m, Fq = 75 kHz, r = 8 nm. The measurements were typically performed using a target amplitude of 300 mV and a setpoint of approximately 185 mV, with a scan rate of 1 Hz.

The fibres in solution were prepared by depositing 2 µL of CAHS-8 at 30 mg/mL on a mica plate, covered with 50 µL of 50 mM of buffer and imaged as the sample was dehydrating.

The aerogels were prepared by depositing 5 µL of CAHS-8 at approximately 5 mg/mL in 30 mM ammonium bicarbonate on either mica, glass or steel plates which were subsequently freeze-dried individually and imaged.

The dried hydrogel (xerogel) was imaged by depositing 5 µL of CAHS-8 at 30 mg/mL on a mica plate and letting the sample dry at room temperature for 15 to 30 min.

The hydrogel in solution were prepared by depositing 5 µL of CAHS-8 at 30 mg/mL on a mica plate and adding solution of TMV or IgG to the sample and letting the sample dry, after which the sample was rinsed with buffer once, covered with 50 µL of 50 mM of buffer and imaged as the sample was dehydrating.

The images were processed using Gwyddion[712] using a mean-plane subtraction levelling, a median method alignment of the rows, correction of the strokes and with the minimal height set to 0.

7.3.4 Dynamic Light Scattering

The measurements of the DLS correlograms were performed using a Malvern Zetasizer Nano S instrument using a back-scattering angle of 173°. The samples at the desired concentration were centrifugated at high temperature for 5 min and 150 µL were pipetted into a Malvern ZEN0040 cuvette. A temperature program ranging from 318 K to 278 K with three measurements every 5 K after 300 s of equilibration was used. Each measurement is composed of 10 runs of 60 s.

The correlogram at 30 mg/mL at 298 K was fitted using the gel-point equation[730,732]

$$g^{(2)}(\tau) - 1 = \sigma_l^2 \left[A \exp(-D_{\text{fast}} q^2 \tau) + (1 - A) \left(1 + \frac{\tau}{\tau^*}\right)^{\left(\frac{n-1}{2}\right)} \right]^2$$

with the oscillations at intermediate times being excluded from the fit. As previously describe in the introduction of this part, σ_l^2 corresponds to the asymptotic value of the field correlation function at infinite time $g^{(1)}(q, \infty)$ and is therefore the contribution of the non-ergodicity to the scattering which is therefore the initial suppressed value measured which was fixed in the fit. A and D_{fast} are the fraction of the sample engaged in the fast apparent translational diffusion mode for the former and its diffusion coefficient for the later, q the magnitude of the scattering vector and was fixed to 173 °, τ^* the time at which the power-law manifests in the decay and n the fractal dimension of the scattered photons.

7.3.5 Differential Scanning Calorimetry

The hydrogel thermograms were measured on a Texas Instruments MC-DSC calorimeter using 238.7 mg of a hydrogel sample of CAHS-8 at 30 mg/mL, yielding 7.16 mg of sample, using a scan rate of 0.5 °C per minute, from 5 °C to 80 °C. The reference thermograms of the buffer were measured at the same time using the same parameter.

The data were processed using pyDSC[833,834].

Solid-state DSC thermograms were measured in a SETARAM DSC-131 calorimeter using 2.5 mg of the CAHS-8 xerogel resulting from freeze-drying. The thermograms were measured from 20 °C to 173 °C with a heating rate of 0.5 °C per minute.

Both instruments were located in the Institut Laue-Langevin (ILL), Grenoble, France.

7.3.6 Nuclear Magnetic Resonance

7.3.6.1 Solid-state NMR

All solid-state experiments were performed at the Centre de Résonance Magnétique Nucléaire à Très Hauts Champs (CRMN, Lyon, France) with the collaboration of Dr. Tanguy Lemarchand and Dr. Guido Pintacuda.

Carbon-detected ^{13}C-^{13}C DARR (Dipolar Assisted Rotational Resonance) spectrum of TMAO-induced fibres of ^{13}C ^{15}N CAHS-8 at 15 mg/mL measured in a 3.2 mm rotor at 800 MHz, 16.67 kHz magic-angle spinning (MAS) speed, 293 K and 20 ms of mixing time.

The proton-detected ^{15}N-^{1}H hNH (CP and INEPT) as well as the ^{13}C-^{1}H hCH experiments were measured as described using a ^{2}D ^{13}C ^{15}N hydrogel of CAHS-8 at 30 mg/mL packed in a 1.3 mm rotor, at 1 GHz, 55 kHz MAS and 293 K.

7.3.6.2 Solution-state NMR

7.3.6.2.1 Sample preparation

The experiments on the CAHS-8 hydrogel were performed at 30 mg/mL using ^{15}N labelled protein.

The experiments using encapsulated proteins were performed by mixing solutions of pre-heated CAHS-8 and the proteins to obtain the desired final concentrations. The gelation was realised by letting the sample cool to room-temperature.

The concentration of the proteins in the hydrogels were:

ANP2a – 150 µM

SH3A – 150 µM

Sendai NTail – 300 µM

PX – 24 µM (8% of Sendai N_{Tail})

Measles N_{Tail} – 260 µM

XD – 26 µM (10% of Measles N_{Tail}).

7.3.6.2.2 Relaxation

The ^{15}N R_1 rates and $\{^1H\}$-^{15}N heteronuclear nOe were measured as described[687] with a set of delays for R_1 of 0, 0.1552, 0.2716, 0.388, 0.5044, 0.6208, 0.7372, 0.9312 and 1.164 s with one delay being repeated for experimental error estimation and also using a recycle delay of 1.5 s, while the nOe used a saturation delay of 8 s.

The ^{15}N $R_1\rho$ rates were measured as described using a spin lock of 1500 Hz[687] with a set of delays of 1, 20, 50, 70, 90, 120, 160 and 200 ms with, two delay delays being repeated for estimation of the experimental error and using a recycle delay of 1.5 s. The R_2 rates were determined from R_1 and $R_1\rho$ as described[688].

The $^{15}N–^{1}H$ CSA/DD transverse cross correlated rates (η_{xy}) were measured as described[689,690], using a cross-relaxation delay of 50 ms. All rates were extracted using in-house scripts with Monte Carlo simulation-based noise-estimation.

NMRPipe[682] was used to process all spectra and the software CCPNMR Analysis 2[683] was used to analyse the spectra.

7.3.6.2.3 Paramagnetic Relaxation Enhancement

The S8C cysteine mutation of CAHS-8 used in the previous part for the intermolecular PRE was used for the in-gel intermolecular PRE experiments. The paramagnetic labelling was performed by attaching the 4-maleimido-TEMPO spin-label on the cysteine sidechain using maleimide chemistry. The protein was incubated overnight with 10 mM DTT, then dialysed against 50 mM NaP, pH = 7.0, 150 mM NaCl. A 5 times molar excess of 4-maleimido-TEMPO dissolved in DMSO was added for overnight incubation and the excess of label was then removed through SEC.

A ^{14}N Hydrogel sample was prepared at 30 mg/mL of CAHS-8 with 300 µM of ^{15}N Sendai N_{Tail} encapsulated in it. The ratio of ^{14}N TEMPO-labelled CAHS-8 was optimised at 10 % while maintaining the total CAHS-8 concentration and the Sendai N_{Tail} concentration.

Paramagnetic relaxation enhancement effects were determined from the peak intensity ratios of ^{15}N-^{1}H HSQC spectra of the paramagnetic state and the diamagnetic state, which was obtained by the addition of 5 mM ascorbic acid to the former. The same recycle delay were used in the para- and diamagnetic experiments.

7.3.6.2.4 CPMG relaxation diffusion

^{15}N Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion were measured using a constant time relaxation delay of 32 ms and 14 CPMG frequencies (νCPMG) ranging from 31.25 to 1 kHz, with one repeated frequency to permit the estimation of the experimental intensity error. The proton decoupling field was 11 kHz. The error on the R_{2eff} were estimated using a Monte Carlo estimation.

The results were fitted using the program ChemEx[695] using a two-state exchange model.

7.3.6.2.5 Diffusion NMR

The ^{15}N-edited pulse field gradient experiments were measured with pulse field gradient strength up to 50.68 G/cm. The volumes of the amide region were integrated and fitted to the Stejskal-Tanner equation[835].

7.3.6.2.6 Diffusion-edited HSQC

The diffusion-edited ^{15}N-^{1}H HSQC was measured as described[836], using a diffusion delay $\Delta = 300$ ms and with pulsed field gradients strength of 0.963 G/cm and 47.19 G/cm.

7.3.7 X-Ray Diffraction

Small parts of the aerogels were torn using a tweezer and the visible fibrous assemblies were mounted on cryoloop, plunged in glycerol for cryoprotection and flash frozen in liquid N_2. Their diffraction data were individually collected using the mesh collection mode on the beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) in collaboration with Dr. Max H. Nanao[699], using a wavelength of 0.873 Å, a temperature of 100 K and using a Pilatus3 2M detector.

7.3.8 Small-Angle X-Ray Scattering

The SAXS measurements of the gel were acquired in glass capillary tube manually mounted in the BM29 beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The sample of CAHS-8 was heated to dissolve the gel and pipetted in a glass capillary sealed on one end with beeswax, with the other end being sealed afterwards. The sample was left to gelate at room-temperature.

The measurement used an incident beam wavelength of 0.99 Å and with a distance between the sample and the Pilatus3 2M detector of 2.867 m, which yields a range of the magnitude of the scattering vector q comprised between 0.0039 $Å^{-1}$ and 0.49 $Å^{-1}$. Every measurement comprised 10 frames which were analysed and rejected in case of radiation damage. The scattering of the solvent was measured before and after each sample and was subtracted from the scattering of the sample.

Chapter 9: Conclusion and perspectives

This thesis aimed to study CAHS-8 and its remarkable properties in order to understand how they could explain the resistance to desiccation it confers to organisms in which it is expressed[614].

In the chapter 5 and 6, CAHS-8 was demonstrated to be an intrinsically disordered protein with a central helical domain composed of two highly populated α-helices, which transiently exchange with an antiparallel dimeric coiled-coil, which can further self-associate into fibrils by forming a second antiparallel coiled-coil on its outer side. The crystal structure of these coiled-coils as well as the fibrils could be obtained for the first time and is in agreement with atomic-force microscopy and electron microscopy imaging, with two fibrils forming a fibre. Finally, it was demonstrated that the presence of the disordered termini of CAHS-8 was necessary to form the fibres, with the C-terminal chain favouring the formation of the coiled-coils necessary for the fibrillation and the N-terminal chain appearing necessary for the formation of the fibres.

In turn, in the chapter 7, the gels resulting from the entangling of these fibres were studied, the reversible gelation was characterised and shown to display a thermal hysteresis and the hydrogel of CAHS-8 was characterised as having the disordered termini of the proteins forming the fibres, and therefore the gel, being free to tumble and coexisting with freely diffusion CAHS-8 species. Finally, the chapter 8 showed that the hydrogel is able to entrap other proteins in its pores which were shown to be a porous and crowded complex environment, which therefore restricts the diffusion of these proteins and perturbate their interactions, with a decrease in the population of the complex probably arising from a physical separation of the interacting partners and with the hydrogel showing a specific retardation effect on the experiment compared to a dense solution at the same concentration of CAHS-8. This is reminiscent of the molecular shield mechanism proposed decades ago for the dehydration-protective LEA proteins, which suggest that these protein may act as crowders upon dehydration that would physically intercalate in-between the biomolecules to prevent their aggregation and

denaturation[295,297,305,306]. Moreover, a model of the intrinsically disordered LEA proteins forming α-helices and coiled-coils upon desiccation which would assemble into "filaments" that would fill the cytosol has been proposed in 2004[306], which bears a remarkable resemblance to what has been demonstrated in this thesis with only the formation of a gel missing.

However, if CAHS proteins have been demonstrated by Tanaka and co-workers from the Kunieda group to form a strong fibre network in vivo in the cells upon dehydration-mimicking osmotic shock[641], it is yet unknown whether they also form a gel in the cytosol. Indeed, after all, these gels were shown to be an entangled network of fibres which may arise in vitro due to their sheer density. However, it appears likely that these proteins do form a gel in vivo during the dehydration, as such a dense fibrillar network in a cell that is shrinking due to water loss, appear to present the characteristic morphology of a gel, whose formation is rendered likelier by automatic increase of the concentration of the CAHS proteins due to the evaporation of water.

As such, the study of CAHS-8 in vivo by fluorescence microscopy and FRAP to see whether it forms a gel or not in the cells was started during the course of this thesis but could not be finished, and therefore present an important next step in the study of this protein.

However, as evident by its number, CAHS-8 is not the only CAHS protein of *Hypsibius exemplaris*, nor of tardigrades in general. Other CAHS, which are more or less close as seen in chapter 5, present different sequences and thus properties. As such, the study of their interplay and possible synergistic effects also appears as a further step in this field. In fact, the CAHS proteins from *Ramazzottius varieornatus (*R.va*)* studied by Tanaka and co-workers from the Kunieda group, with R.vaCAHS-3 showing a strong filament forming ability while R.vaCAHS-8 (which is different from the protein studied in this thesis) forming condensates that may arise from liquid-liquid phase separation (LLPS)[641]. It is notable that these behaviours may be rationalised using the crystal structure of the coiled-coils presented in this thesis, as these protein present mutations that can be explained to strengthen or perturbate the coiled-coil. Nevertheless, the study of such LLPS condensates formed by CAHS proteins would constitute an interesting development, in

particular if these condensates mature into gels and are able to further trap other proteins, which can constitute a complementary mechanism of protection against dehydration stress.

Although the study of CAHS-8 by ssNMR was unsuccessful during this thesis, ssNMR study of labelled proteins trapped in CAHS-8 gels similar to the chapter 8 of this thesis may provide unvaluable insights into the mechanism of protection against dehydration. Indeed, the comparison of dehydrated proteins with or without the CAHS-8 gel matrix may prove different denaturation or aggregation level, in which case they are likely to reveal interaction with the disordered termini of CAHS-8. Moreover, the variation of the relative humidity level may permit to follow the different step of this mechanism and would constitute a significant step forward in the understanding of the dehydration protection ability of these fascinating proteins and the remarkable tardigrades.

References

[1] P. J. Kennelly, K. M. Botham, O. McGuinness, V. W. Rodwell, P. A. Weil, *Harper's Illustrated Biochemistry, Thirty-Second Edition*, McGraw Hill LLC, **2022**.
[2] C. B. Anfinsen, *Science* **1973**, *181*, 223–230.
[3] A. K. Dunker, P. Romero, Z. Obradovic, E. C. Garner, C. J. Brown, **2000**, DOI 10.11234/gi1990.11.161.
[4] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, Z. Obradovic, *J. Mol. Graph. Model.* **2001**, *19*, 26–59.
[5] P. Tompa, *Trends Biochem. Sci.* **2002**, *27*, 527–533.
[6] H. J. Dyson, P. E. Wright, *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
[7] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, Z. Obradović, *Biochemistry* **2002**, *41*, 6573–6582.
[8] H. J. Dyson, P. E. Wright, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
[9] C. J. Oldfield, A. K. Dunker, *Annu. Rev. Biochem.* **2014**, *83*, 553–584.
[10] A. K. Dunker, M. M. Babu, E. Barbar, M. Blackledge, S. E. Bondos, Z. Dosztányi, H. J. Dyson, J. Forman-Kay, M. Fuxreiter, J. Gsponer, K.-H. Han, D. T. Jones, S. Longhi, S. J. Metallo, K. Nishikawa, R. Nussinov, Z. Obradovic, R. V. Pappu, B. Rost, P. Selenko, V. Subramaniam, J. L. Sussman, P. Tompa, V. N. Uversky, *Intrinsically Disord. Proteins* **2013**, *1*, e24157.
[11] K. A. Dill, *Biochemistry* **1990**, *29*, 7133–7155.
[12] K. A. Dill, J. L. MacCallum, *Science* **2012**, *338*, 1042–1046.
[13] H. Frauenfelder, S. G. Sligar, P. G. Wolynes, *Science* **1991**, *254*, 1598–1603.
[14] L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205–211.
[15] W. Kauzmann, in *Adv. Protein Chem.*, Elsevier, **1959**, pp. 1–63.
[16] M. R. Jensen, M. Zweckstetter, J. Huang, M. Blackledge, *Chem. Rev.* **2014**, *114*, 6632–6660.
[17] C. J. Oldfield, V. N. Uversky, A. K. Dunker, L. Kurgan, in *Intrinsically Disord. Proteins*, Elsevier, **2019**, pp. 1–34.
[18] N. Salvi, in *Intrinsically Disord. Proteins*, Elsevier, **2019**, pp. 37–64.
[19] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, V. N. Uversky, *J. Mol. Biol.* **2006**, *362*, 1043–1059.
[20] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, M. M. Babu, *Chem. Rev.* **2014**, *114*, 6589–6631.
[21] K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, N. E. Davey, *Chem. Rev.* **2014**, *114*, 6733–6778.
[22] L. Mollica, L. M. Bessa, X. Hanoulle, M. R. Jensen, M. Blackledge, R. Schneider, *Front. Mol. Biosci.* **2016**, *3*, DOI 10.3389/fmolb.2016.00052.
[23] P. E. Wright, H. J. Dyson, *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.
[24] P. Csermely, R. Palotai, R. Nussinov, *Trends Biochem. Sci.* **2010**, *35*, 539–546.
[25] D. D. Boehr, R. Nussinov, P. E. Wright, *Nat. Chem. Biol.* **2009**, *5*, 789–796.
[26] M. Arai, *Biophys. Rev.* **2018**, *10*, 163–181.
[27] C.-J. Tsai, B. Ma, R. Nussinov, *Proc. Natl. Acad. Sci.* **1999**, *96*, 9970–9972.

[28] C. Tsai, S. Kumar, B. Ma, R. Nussinov, *Protein Sci.* **1999**, *8*, 1181–1190.
[29] B. Ma, S. Kumar, C.-J. Tsai, R. Nussinov, *Protein Eng. Des. Sel.* **1999**, *12*, 713–720.
[30] D. E. Koshland, *Proc. Natl. Acad. Sci.* **1958**, *44*, 98–104.
[31] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, *BMC Genomics* **2008**, *9*, S1.
[32] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, M. Blackledge, *J. Am. Chem. Soc.* **2008**, *130*, 8055–8061.
[33] R. Schneider, D. Maurin, G. Communie, J. Kragelj, D. F. Hansen, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2015**, *137*, 1220–1229.
[34] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2–8.
[35] E. Delaforge, J. Kragelj, L. Tengo, A. Palencia, S. Milles, G. Bouvignies, N. Salvi, M. Blackledge, M. R. Jensen, *J. Am. Chem. Soc.* **2018**, *140*, 1148–1158.
[36] A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best, B. Schuler, *Nature* **2018**, *555*, 61–66.
[37] S. Milles, D. Mercadante, I. V. Aramburu, M. R. Jensen, N. Banterle, C. Koehler, S. Tyagi, J. Clarke, S. L. Shammas, M. Blackledge, F. Gräter, E. A. Lemke, *Cell* **2015**, *163*, 734–745.
[38] S. L. Shammas, A. J. Travis, J. Clarke, *J. Phys. Chem. B* **2013**, *117*, 13346–13356.
[39] J. M. Rogers, A. Steward, J. Clarke, *J. Am. Chem. Soc.* **2013**, *135*, 1415–1422.
[40] M. M. Misiura, A. B. Kolomeisky, *J. Phys. Chem. B* **2020**, *124*, 20–27.
[41] K. Namba, *Genes Cells* **2001**, *6*, 1–12.
[42] H.-X. Zhou, X. Pang, C. Lu, *Phys. Chem. Chem. Phys.* **2012**, *14*, 10466.
[43] V. N. Uversky, *Protein J.* **2009**, *28*, 305–325.
[44] U. Jakob, R. Kriwacki, V. N. Uversky, *Chem. Rev.* **2014**, *114*, 6779–6805.
[45] N. S. Latysheva, T. Flock, R. J. Weatheritt, S. Chavali, M. M. Babu, *Protein Sci.* **2015**, *24*, 909–922.
[46] M. Marín, T. Ott, *Chem. Rev.* **2014**, *114*, 6912–6932.
[47] C. L. Cuevas-Velazquez, J. R. Dinneny, *Curr. Opin. Plant Biol.* **2018**, *45*, 68–74.
[48] J.-H. Jung, A. D. Barbosa, S. Hutin, J. R. Kumita, M. Gao, D. Derwort, C. S. Silva, X. Lai, E. Pierre, F. Geng, S.-B. Kim, S. Baek, C. Zubieta, K. E. Jaeger, P. A. Wigge, *Nature* **2020**, *585*, 256–260.
[49] S. Hutin, J. R. Kumita, V. I. Strotmann, A. Dolata, W. L. Ling, N. Louafi, A. Popov, P.-E. Milhiet, M. Blackledge, M. H. Nanao, P. A. Wigge, Y. Stahl, L. Costa, M. D. Tully, C. Zubieta, *Proc. Natl. Acad. Sci.* **2023**, *120*, e2304714120.
[50] Y. Dorone, S. Boeynaems, E. Flores, B. Jin, S. Hateley, F. Bossi, E. Lazarus, J. G. Pennington, E. Michiels, M. De Decker, K. Vints, P. Baatsen, G. W. Bassel, M. S. Otegui, A. S. Holehouse, M. Exposito-Alonso, S. Sukenik, A. D. Gitler, S. Y. Rhee, *Cell* **2021**, *184*, 4284-4298.e27.
[51] T. M. Franzmann, M. Jahnel, A. Pozniakovsky, J. Mahamid, A. S. Holehouse, E. Nüske, D. Richter, W. Baumeister, S. W. Grill, R. V. Pappu, A. A. Hyman, S. Alberti, *Science* **2018**, *359*, eaao5654.
[52] J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, D. A. Drummond, *Cell* **2017**, *168*, 1028-1040.e19.
[53] M. Lichtheim, *Ancient Egyptian Literature: A Book of Readings. Vol. 3, Vol. 3,* University Of California Press, Berkeley, CA, **2006**.
[54] P. Barguet, *La Stèle de La Famine, à Séhel*, L'Institut Français D'archéologie Orientale, **1953**.
[55] B. Weiss, *Clim. Change* **1982**, *4*, 173–198.

[56] E. H. Cline, in *1177 BC Year Civiliz. Collapsed*, Princeton University Press, **2014**.
[57] R. L. Heathcote, *Drought and the Human Story: Braving the Bull of Heaven*, Routledge, **2016**.
[58] S. W. Manning, C. Kocik, B. Lorentzen, J. P. Sparks, *Nature* **2023**, *614*, 719–724.
[59] Gilgamesh, N. K. Sandars, *The Epic of Gilgamesh: An English Version with an Introd.*, Penguin Books, Harmondsworth, Middlesex, **1972**.
[60] S. Dalley, *Myths from Mesopotamia: Creation, the Flood, Gilgamesh, and Others*, Oxford University Press, **1998**.
[61] K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. W. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barrett, G. Blanco, W. W. L. Cheung, S. Connors, F. Denton, A. Diongue-Niang, D. Dodman, M. Garschagen, O. Geden, B. Hayward, C. Jones, F. Jotzo, T. Krug, R. Lasco, Y.-Y. Lee, V. Masson-Delmotte, M. Meinshausen, K. Mintenbeck, A. Mokssit, F. E. L. Otto, M. Pathak, A. Pirani, E. Poloczanska, H.-O. Pörtner, A. Revi, D. C. Roberts, J. Roy, A. C. Ruane, J. Skea, P. R. Shukla, R. Slade, A. Slangen, Y. Sokona, A. A. Sörensson, M. Tignor, D. Van Vuuren, Y.-M. Wei, H. Winkler, P. Zhai, Z. Zommers, J.-C. Hourcade, F. X. Johnson, S. Pachauri, N. P. Simpson, C. Singh, A. Thomas, E. Totin, P. Arias, M. Bustamante, I. Elgizouli, G. Flato, M. Howden, C. Méndez-Vallejo, J. J. Pereira, R. Pichs-Madruga, S. K. Rose, Y. Saheb, R. Sánchez Rodríguez, D. Ürge-Vorsatz, C. Xiao, N. Yassaa, A. Alegría, K. Armour, B. Bednar-Friedl, K. Blok, G. Cissé, F. Dentener, S. Eriksen, E. Fischer, G. Garner, C. Guivarch, M. Haasnoot, G. Hansen, M. Hauser, E. Hawkins, T. Hermans, R. Kopp, N. Leprince-Ringuet, J. Lewis, D. Ley, C. Ludden, L. Niamir, Z. Nicholls, S. Some, S. Szopa, B. Trewin, K.-I. Van Der Wijst, G. Winter, M. Witting, A. Birt, M. Ha, J. Romero, J. Kim, E. F. Haites, Y. Jung, R. Stavins, A. Birt, M. Ha, D. J. A. Orendain, L. Ignon, S. Park, Y. Park, A. Reisinger, D. Cammaramo, A. Fischlin, J. S. Fuglestvedt, G. Hansen, C. Ludden, V. Masson-Delmotte, J. B. R. Matthews, K. Mintenbeck, A. Pirani, E. Poloczanska, N. Leprince-Ringuet, C. Péan, *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)]. IPCC, Geneva, Switzerland.*, Intergovernmental Panel On Climate Change (IPCC), **2023**.
[62] G. Cooper, K. Adams, *The Cell: A Molecular Approach*, Oxford University Press, **2022**.
[63] T. C. Hsiao, *Annu. Rev. Plant Biol.* **1973**, *24*, 519–570.
[64] H. Mitchell, T. Hamilton, F. Steggerda, H. Bean, *J. Biol. Chem.* **1945**, *158*, 625–637.
[65] G. E. Walrafen, *J. Chem. Phys.* **1964**, *40*, 3249–3256.
[66] R. Ludwig, *Angew. Chem. Int. Ed.* **2001**, *40*, 1808–1827.
[67] A. Geiger, *J. Mol. Liq.* **2003**, *106*, 131–146.
[68] D. Laage, G. Stirnemann, F. Sterpone, R. Rey, J. T. Hynes, *Annu. Rev. Phys. Chem.* **2011**, *62*, 395–416.
[69] D. Laage, J. T. Hynes, *Science* **2006**, *311*, 832–835.
[70] D. Laage, J. T. Hynes, *J. Phys. Chem. B* **2008**, *112*, 14230–14242.
[71] D. Laage, *J. Phys. Chem. B* **2009**, *113*, 2684–2687.
[72] M. Ji, M. Odelius, K. J. Gaffney, *Science* **2010**, *328*, 1003–1005.
[73] R. Schulz, Y. Von Hansen, J. O. Daldrop, J. Kappler, F. Noé, R. R. Netz, *J. Chem. Phys.* **2018**, *149*, 244504.
[74] K. Amann-Winkel, M.-C. Bellissent-Funel, L. E. Bove, T. Loerting, A. Nilsson, A. Paciaroni, D. Schlesinger, L. Skinner, *Chem. Rev.* **2016**, *116*, 7570–7589.
[75] D. Laage, G. Stirnemann, F. Sterpone, J. T. Hynes, *Acc. Chem. Res.* **2012**, *45*, 53–62.
[76] P. J. W. Debye, *Polar Molecules*, Chemical Catalog Company, **1929**.
[77] A. Rahman, F. H. Stillinger, *J. Chem. Phys.* **1971**, *55*, 3336–3359.

[78] A. Rahman, F. H. Stillinger, *J. Am. Chem. Soc.* **1973**, *95*, 7943–7948.
[79] A. Geiger, F. H. Stillinger, A. Rahman, *J. Chem. Phys.* **1979**, *70*, 4185–4193.
[80] H. E. Stanley, J. Teixeira, *J. Chem. Phys.* **1980**, *73*, 3404–3422.
[81] A. Luzar, D. Chandler, *Nature* **1996**, *379*, 55–57.
[82] J. Dore, *Chem. Phys.* **2000**.
[83] P. Ball, *Chem. Rev.* **2008**, *108*, 74–108.
[84] D. E. Moilanen, D. Wong, D. E. Rosenfeld, E. E. Fenn, M. D. Fayer, *Proc. Natl. Acad. Sci.* **2009**, *106*, 375–380.
[85] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. Van Der Spoel, Y. Xu, A. E. Garcia, *Chem. Rev.* **2016**, *116*, 7673–7697.
[86] J. D. Eaves, J. J. Loparo, C. J. Fecko, S. T. Roberts, A. Tokmakoff, P. L. Geissler, *Proc. Natl. Acad. Sci.* **2005**, *102*, 13019–13022.
[87] F. Sciortino, A. Geiger, H. E. Stanley, *Nature* **1991**, *354*, 218–221.
[88] B. Halle, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359*, 1207–1224.
[89] I. Popov, P. B. Ishai, A. Khamzin, Y. Feldman, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13941–13953.
[90] D. Laage, T. Elsaesser, J. T. Hynes, *Chem. Rev.* **2017**, *117*, 10694–10725.
[91] G. E. Walrafen, *J. Chem. Phys.* **1966**, *44*, 1546–1558.
[92] D. C. Elton, *Phys. Chem. Chem. Phys.* **2017**, *19*, 18739–18749.
[93] I. D. Kuntz, W. Kauzmann, in *Adv. Protein Chem.*, Elsevier, **1974**, pp. 239–345.
[94] G. Otting, E. Liepinsh, K. Wüthrich, *Science* **1991**, *254*, 974–980.
[95] M. Tarek, D. J. Tobias, *Phys. Rev. Lett.* **2002**, *88*, 138101.
[96] C. J. Fecko, J. D. Eaves, J. J. Loparo, A. Tokmakoff, P. L. Geissler, *Science* **2003**, *301*, 1698–1702.
[97] C. Mattea, J. Qvist, B. Halle, *Biophys. J.* **2008**, *95*, 2951–2963.
[98] M. Heyden, M. Havenith, *Methods* **2010**, *52*, 74–83.
[99] L. Nilsson, B. Halle, *Proc. Natl. Acad. Sci.* **2005**, *102*, 13867–13872.
[100] S.-H. Chen, L. Liu, E. Fratini, P. Baglioni, A. Faraone, E. Mamontov, *Proc. Natl. Acad. Sci.* **2006**, *103*, 9012–9016.
[101] F. Persson, P. Söderhjelm, B. Halle, *J. Chem. Phys.* **2018**, *148*, 215104.
[102] M. S. Cheung, A. E. García, J. N. Onuchic, *Proc. Natl. Acad. Sci.* **2002**, *99*, 685–690.
[103] Y. Levy, J. N. Onuchic, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.
[104] M. A. Williams, J. M. Goodfellow, J. M. Thornton, *Protein Sci.* **1994**, *3*, 1224–1235.
[105] B. Halle, in *Hydration Process. Biol. Theor. Exp. Approaches* (Ed.: M.C. Bellissent-Funel), IOS Press, **1999**, pp. 233–249.
[106] V. P. Denisov, J. Peters, H. D. Hörlein, B. Halle, *Nat. Struct. Mol. Biol.* **1996**, *3*, 505–509.
[107] S. Kaieda, B. Halle, *J. Phys. Chem. B* **2013**, *117*, 14676–14687.
[108] F. Persson, B. Halle, *J. Am. Chem. Soc.* **2013**, *135*, 8735–8748.
[109] B. D. Armstrong, J. Choi, C. López, D. A. Wesener, W. Hubbell, S. Cavagnero, S. Han, *J. Am. Chem. Soc.* **2011**, *133*, 5987–5995.
[110] J. Qvist, G. Ortega, X. Tadeo, O. Millet, B. Halle, *J. Phys. Chem. B* **2012**, *116*, 3436–3444.
[111] F.-X. Gallat, A. Laganowsky, K. Wood, F. Gabel, L. van Eijck, J. Wuttke, M. Moulin, M. Härtlein, D. Eisenberg, J.-P. Colletier, G. Zaccai, M. Weik, *Biophys. J.* **2012**, *103*, 129–136.
[112] O. Rahaman, S. Melchionna, D. Laage, F. Sterpone, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3570.
[113] P. Rani, P. Biswas, *J. Phys. Chem. B* **2015**, *119*, 13262–13270.

[114] C. A. Angell, *Science* **1995**, *267*, 1924–1935.
[115] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach, A. H. Xie, R. D. Young, *Phys. Rev. Lett.* **1989**, *62*, 1916–1919.
[116] F. Parak, E. W. Knapp, D. Kucheida, *J. Mol. Biol.* **1982**, *161*, 177–194.
[117] W. Doster, S. Cusack, W. Petry, *Nature* **1989**, *337*, 754–756.
[118] M. Ferrand, A. J. Dianoux, W. Petry, G. Zaccaï, *Proc. Natl. Acad. Sci.* **1993**, *90*, 9668–9672.
[119] J. Fitter, *Biophys. J.* **1999**, *76*.
[120] K. Wood, A. Frölich, A. Paciaroni, M. Moulin, M. Härtlein, G. Zaccai, D. J. Tobias, M. Weik, *J. Am. Chem. Soc.* **2008**, *130*, 4586–4587.
[121] J. R. Knab, J.-Y. Chen, Y. He, A. G. Markelz, *Proc. IEEE* **2007**, *95*, 1605–1610.
[122] H. Jansson, R. Bergman, J. Swenson, *J. Phys. Chem. B* **2011**, *115*, 4099–4109.
[123] J. R. Lewandowski, M. E. Halse, M. Blackledge, L. Emsley, *Science* **2015**, *348*, 578–581.
[124] M. Bin, M. Reiser, M. Filianina, S. Berkowicz, S. Das, S. Timmermann, W. Roseker, R. Bauer, J. Öström, A. Karina, K. Amann-Winkel, M. Ladd-Parada, F. Westermeier, M. Sprung, J. Möller, F. Lehmkühler, C. Gutt, F. Perakis, *J. Phys. Chem. B* **2023**, *127*, 4922–4930.
[125] D. Vitkup, D. Ringe, G. A. Petsko, M. Karplus, *Nat. Struct. Biol.* **2000**, *7*, 34–38.
[126] C. Arcangeli, A. R. Bizzarri, S. Cannistraro, *Chem. Phys. Lett.* **1998**, *291*, 7–14.
[127] A. Paciaroni, A. R. Bizzarri, S. Cannistraro, *Phys. Rev. E* **1999**, *60*, R2476–R2479.
[128] G. Caliskan, A. Kisliuk, A. P. Sokolov, *J. Non-Cryst. Solids* **2002**, *307–310*, 868–873.
[129] B. F. Rasmussen, A. M. Stock, D. Ringe, G. A. Petsko, *Nature* **1992**, *357*, 423–424.
[130] W. Doster, M. Settles, *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2005**, *1749*, 173–186.
[131] A. P. Sokolov, H. Grimm, A. Kisliuk, A. J. Dianoux, *J. Biol. Phys.* **2001**, *27*, 313–327.
[132] G. Caliskan, R. M. Briber, D. Thirumalai, V. Garcia-Sakai, S. A. Woodson, A. P. Sokolov, *J. Am. Chem. Soc.* **2006**, *128*, 32–33.
[133] V. De Michele, M. Levantino, A. Cupane, *J. Chem. Phys.* **2019**, *150*, 224509.
[134] A. M. Tsai, D. A. Neumann, L. N. Bell, *Biophys. J.* **2000**, *79*, 2728–2732.
[135] A. Paciaroni, S. Cinelli, G. Onori, *Biophys. J.* **2002**, *83*, 1157–1164.
[136] H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, R. D. Young, *Proc. Natl. Acad. Sci.* **2009**, *106*, 5129–5134.
[137] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, F. G. Parak, *Proc. Natl. Acad. Sci.* **2002**, *99*, 16047–16051.
[138] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, R. D. Young, *Proc. Natl. Acad. Sci.* **2004**, *101*, 14408–14413.
[139] K. L. Ngai, S. Capaccioli, N. Shinyashiki, *J. Phys. Chem. B* **2008**, *112*, 3826–3832.
[140] A. L. Lee, A. J. Wand, *Nature* **2001**, *411*, 501–504.
[141] B. Busi, J. R. Yarava, A. Hofstetter, N. Salvi, D. Cala-De Paepe, J. R. Lewandowski, M. Blackledge, L. Emsley, *J. Phys. Chem. B* **2018**, *122*, 9697–9702.
[142] Y. He, P. I. Ku, J. R. Knab, J. Y. Chen, A. G. Markelz, *Phys. Rev. Lett.* **2008**, *101*, 178103.
[143] W. Adamski, N. Salvi, D. Maurin, J. Magnat, S. Milles, M. R. Jensen, A. Abyzov, C. J. Moreau, M. Blackledge, *J. Am. Chem. Soc.* **2019**, *141*, 17817–17829.
[144] A. D. Stephens, J. Kölbel, R. Moons, C. W. Chung, M. T. Ruggiero, N. Mahmoudi, T. A. Shmool, T. M. McCoy, D. Nietlispach, A. F. Routh, F. Sobott, J. A. Zeitler, G. S. Kaminski Schierle, *Angew. Chem. Int. Ed.* **2023**, *62*, e202212063.
[145] R. B. Best, W. Zheng, J. Mittal, *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

[146] S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
[147] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2016**, *138*, 6240–6251.
[148] N. Salvi, A. Abyzov, M. Blackledge, *Sci. Adv.* **2019**, *5*, eaax2348.
[149] P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus, G. N. Somero, *Science* **1982**, *217*, 1214–1222.
[150] R. G. W. Jones, J. Gorham, in *Physiol. Plant Ecol. III Responses Chem. Biol. Environ.* (Eds.: O.L. Lange, P.S. Nobel, C.B. Osmond, H. Ziegler), Springer Berlin Heidelberg, Berlin, Heidelberg, **1983**, pp. 35–58.
[151] Potts M, *Microbiol. Rev.* **1994**, *58*, 755–805.
[152] Z. Huang, A. Tunnacliffe, *J. Physiol.* **2004**, *558*, 181–191.
[153] M. J. Oliver, J. M. Farrant, H. W. M. Hilhorst, S. Mundree, B. Williams, J. D. Bewley, *Annu. Rev. Plant Biol.* **2020**, *71*, 435–460.
[154] H. T. Meryman, *Science* **1956**, *124*, 515–521.
[155] J. H. Crowe, J. F. Carpenter, L. M. Crowe, T. J. Anchordoguy, *Cryobiology* **1990**, *27*, 219–231.
[156] R. Pearce, *Ann. Bot.* **2001**, *87*, 417–424.
[157] A. Blomberg, L. Adler, in *Adv. Microb. Physiol.*, Elsevier, **1992**, pp. 145–212.
[158] H. Simonin, L. Beney, P. Gervais, *Biochim. Biophys. Acta BBA - Biomembr.* **2007**, *1768*, 1600–1610.
[159] A. K. Adya, E. Canetta, G. M. Walker, *FEMS Yeast Res.* **2006**, *6*, 120–128.
[160] P. S. Romero-Perez, Y. Dorone, E. Flores, S. Sukenik, S. Boeynaems, *Chem. Rev.* **2023**, *123*, 9010–9035.
[161] C. Womersley, *Comp. Biochem. Physiol. Part B Comp. Biochem.* **1981**, *70*, 669–678.
[162] C. Walters, J. Farrant, N. Pammenter, P. Berjak, in *Desiccation Surviv. Plants Dry. Dying* (Eds.: M. Black, H.W. Pritchard), CABI Pub., **2002**, pp. 263–291.
[163] J. Edsall, *Adv. Biophys.* **1983**, *16*, v–vi.
[164] H. Frauenfelder, P. W. Fenimore, G. Chen, B. H. McMahon, *Proc. Natl. Acad. Sci.* **2006**, *103*, 15469–15472.
[165] S. J. Prestrelski, N. Tedeschi, T. Arakawa, J. F. Carpenter, *Biophys. J.* **1993**, *65*, 661–671.
[166] F. A. Hoekstra, E. A. Golovina, J. Buitink, *Trends Plant Sci.* **2001**, *6*, 431–438.
[167] V. P. Denisov, B.-H. Jonsson, B. Halle, *Nat. Struct. Biol.* **1999**, *6*.
[168] D. A. Stein, Ed. , *Protein Aggregation*, Nova Biomedical Books, New York, NY, **2011**.
[169] A. Dong, S. J. Prestrelski, S. D. Allison, J. F. Carpenter, *J. Pharm. Sci.* **1995**, *84*, 415–424.
[170] K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hiroaki, M. Shirakawa, *Nature* **2009**, *458*, 106–109.
[171] A. C. Miklos, M. Sarkar, Y. Wang, G. J. Pielak, *J. Am. Chem. Soc.* **2011**, *133*, 7116–7120.
[172] Y. Wang, M. Sarkar, A. E. Smith, A. S. Krois, G. J. Pielak, *J. Am. Chem. Soc.* **2012**, *134*, 16614–16618.
[173] T. Arakawa, S. N. Timasheff, *Biochemistry* **1984**, *23*, 5912–5923.
[174] C. J. van Oss, R. J. Good, M. K. Chaudhury, *J. Protein Chem.* **1986**, *5*, 385–405.
[175] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, M. Stefani, *Nature* **2002**, *416*, 507–511.
[176] M. Stefani, C. M. Dobson, *J. Mol. Med.* **2003**, *81*, 678–699.
[177] W. M. Holmes, C. L. Klaips, T. R. Serio, *Crit. Rev. Biochem. Mol. Biol.* **2014**, *49*, 294–303.

[178] M. A. Mourão, J. B. Hakim, S. Schnell, *Biophys. J.* **2014**, *107*, 2761–2766.
[179] A. P. Minton, *J. Biol. Chem.* **2001**, *276*, 10577–10580.
[180] S. Schnell, T. E. Turner, *Prog. Biophys. Mol. Biol.* **2004**, *85*, 235–260.
[181] K. Luby-Phelps, *Int. Rev. Cytol.* **1999**, 189–221.
[182] J. A. Dix, A. S. Verkman, *Annu. Rev. Biophys.* **2008**, *37*, 247–263.
[183] R. Kopelman, *Science* **1988**, *241*, 1620–1626.
[184] H. Berry, *Biophys. J.* **2002**, *83*, 1891–1901.
[185] J. Shin, A. G. Cherstvy, R. Metzler, *New J. Phys.* **2014**, *16*, 053047.
[186] Y. Shibata, J. Hu, M. M. Kozlov, T. A. Rapoport, *Annu. Rev. Cell Dev. Biol.* **2009**, *25*, 329–354.
[187] D. Chapman, *Q. Rev. Biophys.* **1975**, *8*, 185–235.
[188] P. Yeagle, *The Membranes of Cells*, Elsevier/AP, Amsterdam ; Boston, **2016**.
[189] G. Bryant, J. M. Pope, J. Wolfe, *Eur. Biophys. J.* **1992**, *21*, 223–232.
[190] P. L. Steponkus, M. Uemura, M. S. Webb, in *Permeability Stab. Lipid Bilayers* (Eds.: E.A. Disalvo, S.A. Simon), CRC Press, USA, **1995**, pp. 77–104.
[191] J. H. Crowe, L. M. Crowe, D. Chapman, *Science* **1984**, *223*, 701–703.
[192] J. H. Crowe, F. A. Hoekstra, L. M. Crowe, *Annu. Rev. Physiol.* **1992**, *54*, 579–599.
[193] H. Binder, K. Gawrisch, *J. Phys. Chem. B* **2001**, *105*, 12378–12390.
[194] J. H. Crowe, F. A. Hoekstra, L. M. Crowe, *Proc. Natl. Acad. Sci.* **1989**, *86*, 520–523.
[195] W. J. Gordon-Kamm, P. L. Steponkus, *Proc. Natl. Acad. Sci.* **1984**, *81*, 6373–6377.
[196] G. Bryant, J. Wolfe, *Eur. Biophys. J.* **1989**, *16*, DOI 10.1007/BF00257886.
[197] L. Yang, H. W. Huang, *Science* **2002**, *297*, 1877–1879.
[198] S.-J. Marrink, A. E. Mark, *Biophys. J.* **2004**, *87*, 3894–3900.
[199] M. R. Becwar, P. C. Stanwood, E. E. Roos, *Plant Physiol.* **1982**, *69*, 1132–1135.
[200] F. A. Hoekstra, J. H. Crowe, L. M. Crowe, in *Recent Adv. Dev. Germination Seeds*, Springer, **1989**, pp. 71–88.
[201] M. Uemura, R. A. Joseph, P. L. Steponkus, *Plant Physiol.* **1995**, *109*, 15–30.
[202] S. Dupont, L. Beney, J.-F. Ritt, J. Lherminier, P. Gervais, *Biochim. Biophys. Acta BBA - Biomembr.* **2010**, *1798*, 975–985.
[203] K. L. Koster, R. A. Balsamo, C. Espinoza, M. J. Oliver, *Plant Growth Regul.* **2010**, *62*, 293–302.
[204] R. Cruz De Carvalho, M. Catalá, C. Branquinho, J. Marques Da Silva, E. Barreno, *Physiol. Plant.* **2017**, *159*, 277–289.
[205] S. Pukacka, M. Malec, E. Ratajczak, *Acta Physiol. Plant.* **2011**, *33*, 2219–2227.
[206] S. Parkhey, S. C. Naithani, S. Keshavkant, *Plant Physiol. Biochem.* **2012**, *57*, 261–267.
[207] J. Chandra, S. Keshavkant, *Physiol. Mol. Biol. Plants* **2018**, *24*, 75–87.
[208] B. D. Mckersie, J. H. Crowe, L. M. Crowe, *Biochim. Biophys. Acta BBA-Biomembr.* **1989**, *982*, 156–160.
[209] S. Dussert, Mark. W. Davey, A. Laffargue, S. Doulbeau, R. Swennen, H. Etienne, *Physiol. Plant.* **2006**, *127*, 192–204.
[210] E. Cabiscol Catalā, J. Tamarit Sumalla, J. Ros Salvador, **2000**.
[211] A. M. Winger, N. L. Taylor, J. L. Heazlewood, D. A. Day, A. H. Millar, *J. Biol. Chem.* **2007**, *282*, 37436–37447.
[212] M. Schieber, N. S. Chandel, *Curr. Biol.* **2014**, *24*, R453–R462.
[213] M. U. H. Joardder, M. H. Masud, in *Food Preserv. Dev. Ctries. Chall. Solut.*, Springer International Publishing, Cham, **2019**, pp. 57–66.
[214] M. Perchonok, C. Bourland, **2002**, *18*.
[215] N. Lynnerup, *Am. J. Phys. Anthropol.* **2007**, *134*, 162–190.

[216] I. U. Kang, J. Han, J. H. Hong, J. Kim, D. H. Shin, V. H. Mair, in *Handb. Mummy Stud. New Front. Sci. Cult. Perspect.* (Eds.: D.H. Shin, R. Bianucci), Springer Singapore, Singapore, **2021**, pp. 1033–1048.
[217] Herodotus, *Herodotus : The Histories*, New Edition. London, Eng. ; New York : Penguin Books, 1996., **1996**.
[218] A. Lucas, *J. Egypt. Archaeol.* **1914**, *1*, 119–123.
[219] B. BRIER, R. S. WADE, *Z. Für Ägyptische Sprache Altertumskunde* **1997**, *124*, 89–100.
[220] K. Jeremiah, *Living Buddhas: The Self-Mummified Monks of Yamagata, Japan*, McFarland, Incorporated, Publishers, **2010**.
[221] P. Alpert, M. Oliver, in *Desiccation Surviv. Plants Dry. Dying* (Eds.: M. Black, H.W. Pritchard), CABI Pub., **2002**, pp. 3–43.
[222] A. Wood, M. Jenks, in *Plant Desiccation Toler.*, **2008**, pp. 1–10.
[223] A. van Leeuwenhoek, in *Sel. Works Antony Van Leeuwenhoek*, Whittingham And Arliss, London, **1702**, pp. 207–213.
[224] A. Tunnacliffe, J. Lapinski, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2003**, *358*, 1755–1771.
[225] D. Keilin, *Proc. R. Soc. Lond. B Biol. Sci.* **1959**, *150*, 149–191.
[226] J. S. Clegg, **2001**.
[227] J. C. Wright, *Zool. Anz. - J. Comp. Zool.* **2001**, *240*, 563–582.
[228] R. Guidetti, K. I. Jönsson, *J. Zool.* **2002**, *257*, 181–187.
[229] J. Shen-Miller, M. B. Mudgett, J. W. Schopf, S. Clarke, R. Berger, *Am. J. Bot.* **1995**, *82*, 1367–1380.
[230] S. Sallon, E. Solowey, Y. Cohen, R. Korchinsky, M. Egli, I. Woodhatch, O. Simchoni, M. Kislev, *Science* **2008**, *320*, 1464–1464.
[231] T. C. Hawes, J. S. Bale, *J. Exp. Biol.* **2007**, *210*, 2585–2592.
[232] J. Serwe-Rodriguez, K. Sonnenberg, B. Appleman, S. Bornstein-Forst, *J. Invertebr. Pathol.* **2004**, *85*, 175–181.
[233] O. Gusev, Y. Nakahara, V. Vanyagina, L. Malutina, R. Cornette, T. Sakashita, N. Hamada, T. Kikawada, Y. Kobayashi, T. Okuda, *PLOS ONE* **2010**, *5*, e14008.
[234] O. Gusev, T. Okuda, in *Life Earth Planet. Bodies* (Eds.: A. Hanslmeier, S. Kempe, J. Seckbach), Springer Netherlands, Dordrecht, **2012**, pp. 121–138.
[235] A. Ryabova, K. Mukae, A. Cherkasov, R. Cornette, E. Shagimardanova, T. Sakashita, T. Okuda, T. Kikawada, O. Gusev, *Extremophiles* **2017**, *21*, 109–120.
[236] T. Sakamoto, T. Yoshida, H. Arima, Y. Hatanaka, Y. Takani, Y. Tamaru, *Phycol. Res.* **2009**, *57*, 66–73.
[237] J. M. Thevelein, *Microbiol. Rev.* **1984**, *48*, 42–59.
[238] L. M. Crowe, J. H. Crowe, *Adv. Space Res.* **1992**, *12*, 239–247.
[239] L. M. Crowe, *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2002**, *131*, 505–513.
[240] M. Iordachescu, R. Imai, in *Abiotic Stress Plants* (Eds.: A. Shanker, B. Venkateswarlu), IntechOpen, Rijeka, **2011**.
[241] S. H. Lillie, J. R. Pringle, *J. Bacteriol.* **1980**, *143*, 1384–1394.
[242] T. Hottiger, T. Boller, A. Wiemken, *FEBS Lett.* **1987**, *220*, 113–115.
[243] J. S. Clegg, M. F. Filosa, *Nature* **1961**, *192*, 1077–1078.
[244] L. M. Higa, C. Z. Womersley, *J. Exp. Zool.* **1993**, *267*, 120–129.
[245] C. Erkut, S. Penkov, H. Khesbak, D. Vorkel, J.-M. Verbavatz, K. Fahmy, T. V. Kurzchalia, *Curr. Biol.* **2011**, *21*, 1331–1336.
[246] J. H. Crowe, J. F. Carpenter, L. M. Crowe, *Annu. Rev. Physiol.* **1998**, *60*, 73–103.
[247] J. H. Crowe, J. F. Carpenter, L. M. Crowe, *Annu. Rev. Physiol.* **1998**, *60*, 73–103.

[248] J. K. Kaushik, R. Bhat, *J. Biol. Chem.* **2003**, *278*, 26458–26465.
[249] R. J. Williams, A. C. Leopold, *Plant Physiol.* **1989**, *89*, 977–981.
[250] S. Hengherr, M. R. Worland, A. Reuner, F. Brümmer, R. O. Schill, *Physiol. Biochem. Zool.* **2009**, *82*, 749–755.
[251] M. Sakurai, T. Furuki, K. Akao, D. Tanaka, Y. Nakahara, T. Kikawada, M. Watanabe, T. Okuda, *Proc. Natl. Acad. Sci.* **2008**, *105*, 5093–5098.
[252] J. F. Carpenter, J. H. Crowe, *Biochemistry* **1989**, *28*, 3916–3922.
[253] J. H. Crowe, J. S. Clegg, L. M. Crowe, in *Prop. Water Foods ISOPOW 6* (Ed.: D.S. Reid), Springer US, Boston, MA, **1998**, pp. 440–455.
[254] S. D. Allison, B. Chang, T. W. Randolph, J. F. Carpenter, *Arch. Biochem. Biophys.* **1999**, *365*, 289–298.
[255] C. Olsson, H. Jansson, J. Swenson, *J. Phys. Chem. B* **2016**, *120*, 4723–4731.
[256] C. Olsson, S. Genheden, V. García Sakai, J. Swenson, *J. Phys. Chem. B* **2019**, *123*, 3679–3687.
[257] S. N. Timasheff, *Biochemistry* **1992**, *31*, 9857–9864.
[258] S. N. Timasheff, *Biochemistry* **2002**, *41*, 13473–13482.
[259] N. K. Jain, I. Roy, *Protein Sci.* **2009**, *18*, 24–36.
[260] P. S. Belton, A. M. Gil, *Biopolymers* **1994**, *34*, 957–961.
[261] R. D. Lins, C. S. Pereira, P. H. Hünenberger, *Proteins Struct. Funct. Bioinforma.* **2004**, *55*, 177–186.
[262] F.-F. Liu, L. Ji, L. Zhang, X.-Y. Dong, Y. Sun, *J. Chem. Phys.* **2010**, *132*, 225103.
[263] D. Corradini, E. G. Strekalova, H. E. Stanley, P. Gallo, *Sci. Rep.* **2013**, *3*, 1218.
[264] H. Kawai, M. Sakurai, Y. Inoue, R. Chûjô, S. Kobayashi, *Cryobiology* **1992**, *29*, 599–606.
[265] M. Sola-Penna, J. R. Meyer-Fernandes, *Arch. Biochem. Biophys.* **1998**, *360*, 10–14.
[266] M. Heyden, E. Bründermann, U. Heugen, G. Niehues, D. M. Leitner, M. Havenith, *J. Am. Chem. Soc.* **2008**, *130*, 5773–5779.
[267] M. E. Gallina, L. Comez, A. Morresi, M. Paolantoni, S. Perticaroli, P. Sassi, D. Fioretto, *J. Chem. Phys.* **2010**, *132*, 214508.
[268] L. R. Winther, J. Qvist, B. Halle, *J. Phys. Chem. B* **2012**, *116*, 9196–9207.
[269] M. Heyden, G. Schwaab, M. Havenith, *J. Phys. Chem. B* **2014**, *118*, 10802–10805.
[270] B. Halle, *J. Phys. Chem. B* **2014**, *118*, 10806–10812.
[271] N. Katyal, S. Deep, *Phys Chem Chem Phys* **2014**, *16*, 26746–26761.
[272] K. Ahlgren, C. Olsson, I. Ermilova, J. Swenson, *Phys. Chem. Chem. Phys.* **2023**, *25*, 21215–21226.
[273] A. S. Rudolph, J. H. Crowe, *Cryobiology* **1985**, *22*, 367–377.
[274] A. E. Oliver, L. M. Crowe, J. H. Crowe, *Seed Sci. Res.* **1998**, *8*, 211–221.
[275] K. L. Koster, Y. P. Lei, M. Anderson, S. Martin, G. Bryant, *Biophys. J.* **2000**, *78*, 1932–1946.
[276] C. J. Garvey, T. Lenné, K. L. Koster, B. Kent, G. Bryant, *Int. J. Mol. Sci.* **2013**, *14*, 8148–8163.
[277] J. Liu, C. Chen, C. Lu, W. Li, *J. Mol. Liq.* **2019**, *275*, 839–848.
[278] L. Dure III, C. Chlan, *Plant Physiol.* **1981**, *68*, 180–186.
[279] I. Dure Leon, G. A. Galau, *Plant Physiol.* **1981**, *68*, 187–194.
[280] A. C. Cuming, in *Seed Proteins* (Eds.: P.R. Shewry, R. Casey), Springer Netherlands, Dordrecht, **1999**, pp. 753–780.
[281] J. R. Battista, M.-J. Park, A. E. McLemore, *Cryobiology* **2001**, *43*, 133–139.
[282] J. Rodriguez-Salazar, S. Moreno, G. Espín, *Cell Stress Chaperones* **2017**, *22*, 397–408.
[283] S. Abba', S. Ghignone, P. Bonfante, *BMC Genomics* **2006**, *7*, 39.

[284] A. Tunnacliffe, J. Lapinski, B. McGee, *Hydrobiologia* **2005**, *546*, 315–321.
[285] N. Y. Denekamp, R. Reinhardt, M. Kube, E. Lubzens, *Biol. Reprod.* **2010**, *82*, 714–724.
[286] J. Browne, A. Tunnacliffe, A. Burnell, *Nature* **2002**, *416*, 38–38.
[287] J. A. Browne, K. M. Dolan, T. Tyson, K. Goyal, A. Tunnacliffe, A. M. Burnell, *Eukaryot. Cell* **2004**, *3*, 966–975.
[288] J. S. Clegg, *Comp. Biochem. Physiol.* **1967**, *20*, 801–809.
[289] B. Janis, V. N. Uversky, M. A. Menze, *J. Biomol. Struct. Dyn.* **2018**, *36*, 3291–3309.
[290] O. Gusev, Y. Suetsugu, R. Cornette, T. Kawashima, M. D. Logacheva, A. S. Kondrashov, A. A. Penin, R. Hatanaka, S. Kikuta, S. Shimura, H. Kanamori, Y. Katayose, T. Matsumoto, E. Shagimardanova, D. Alexeev, V. Govorun, J. Wisecaver, A. Mikheyev, R. Koyanagi, M. Fujie, T. Nishiyama, S. Shigenobu, T. F. Shibata, V. Golygina, M. Hasebe, T. Okuda, N. Satoh, T. Kikawada, *Nat. Commun.* **2014**, *5*, 4784.
[291] T. Kikawada, Y. Nakahara, Y. Kanamori, K. Iwata, M. Watanabe, B. McGee, A. Tunnacliffe, T. Okuda, *Biochem. Biophys. Res. Commun.* **2006**, *348*, 56–61.
[292] T. Kazuoka, K. Oeda, *Plant Cell Physiol.* **1994**, *35*, 601–611.
[293] T. Maula, N. Vahvelainen, H. Tossavainen, T. Koivunen, M. T. Pöllänen, A. Johansson, P. Permi, R. Ihalin, *Virulence* **2021**, *12*, 1239–1257.
[294] Y. Liu, L. Wang, X. Xing, L. Sun, J. Pan, X. Kong, M. Zhang, D. Li, *Plant Cell Physiol.* **2013**, *54*, 944–959.
[295] R. Hatanaka, Y. Hagiwara-Komoda, T. Furuki, Y. Kanamori, M. Fujita, R. Cornette, M. Sakurai, T. Okuda, T. Kikawada, *Insect Biochem. Mol. Biol.* **2013**, *43*, 1055–1067.
[296] B. M. LeBlanc, S. C. Hand, *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2021**, *1869*, 140642.
[297] K. Goyal, L. J. Walton, A. Tunnacliffe, *Biochem. J.* **2005**, *388*, 151–157.
[298] J. Grelet, A. Benamar, E. Teyssier, M.-H. Avelange-Macherel, D. Grunwald, D. Macherel, *Plant Physiol.* **2005**, *137*, 157–167.
[299] L. C. Boswell, M. A. Menze, S. C. Hand, *Physiol. Biochem. Zool.* **2014**, *87*, 640–651.
[300] A. V. Popova, S. Rausch, M. Hundertmark, Y. Gibon, D. K. Hincha, *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2015**, *1854*, 1517–1525.
[301] S. Chakrabortee, C. Boschetti, L. J. Walton, S. Sarkar, D. C. Rubinsztein, A. Tunnacliffe, *Proc. Natl. Acad. Sci.* **2007**, *104*, 18073–18078.
[302] V. Boucher, J. Buitink, X. Lin, J. Boudet, F. A. Hoekstra, M. Hundertmark, D. Renard, O. Leprince, *Plant Cell Environ.* **2010**, *33*, 418–430.
[303] T. Furuki, T. Shimizu, S. Chakrabortee, K. Yamakawa, R. Hatanaka, T. Takahashi, T. Kikawada, T. Okuda, H. Mihara, A. Tunnacliffe, M. Sakurai, *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2012**, *1824*, 891–897.
[304] S. Koubaa, A. Bremer, D. K. Hincha, F. Brini, *Sci. Rep.* **2019**, *9*, 3720.
[305] A. Tunnacliffe, D. K. Hincha, O. Leprince, D. Macherel, in *Dormancy Resist. Harsh Environ.* (Eds.: E. Lubzens, J. Cerda, M. Clark), Springer Berlin Heidelberg, Berlin, Heidelberg, **2010**, pp. 91–108.
[306] M. J. Wise, A. Tunnacliffe, *Trends Plant Sci.* **2004**, *9*, 13–17.
[307] R. I. Feigin, D. H. Napper, *J. Colloid Interface Sci.* **1980**, *75*, 525–541.
[308] A. N. Semenov, *Macromolecules* **2008**, *41*, 2243–2249.
[309] C. Navarro-Retamal, A. Bremer, H. I. Ingólfsson, J. Alzate-Morales, J. Caballero, A. Thalhammer, W. González, D. K. Hincha, *Biophys. J.* **2018**, *115*, 968–980.
[310] P. L. Steponkus, M. Uemura, R. A. Joseph, S. J. Gilmour, M. F. Thomashow, *Proc. Natl. Acad. Sci.* **1998**, *95*, 14570–14575.

[311] A. Thalhammer, G. Bryant, R. Sulpice, D. K. Hincha, *Plant Physiol.* **2014**, *166*, 190–201.
[312] D. Tolleter, M. Jaquinod, C. Mangavel, C. Passirani, P. Saulnier, S. Manon, E. Teyssier, N. Payet, M.-H. Avelange-Macherel, D. Macherel, *Plant Cell* **2007**, *19*, 1580–1589.
[313] D. Tolleter, D. K. Hincha, D. Macherel, *Biochim. Biophys. Acta BBA - Biomembr.* **2010**, *1798*, 1926–1933.
[314] D. S. Moore, R. Hansen, S. C. Hand, *Biochim. Biophys. Acta BBA - Biomembr.* **2016**, *1858*, 104–115.
[315] D. S. Moore, S. C. Hand, *Cryobiology* **2016**, *73*, 240–247.
[316] W. F. Wolkers, S. McCready, W. F. Brandt, G. G. Lindsey, F. A. Hoekstra, *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **2001**, *1544*, 196–206.
[317] G. Hunault, E. Jaspard, *BMC Genomics* **2010**, *11*, 1–9.
[318] L. Dure, M. Crouch, J. Harada, T.-H. D. Ho, J. Mundy, R. Quatrano, T. Thomas, Z. R. Sung, *Plant Mol. Biol.* **1989**, *12*, 475–486.
[319] A. Garay-Arroyo, J. M. Colmenero-Flores, A. Garciarrubio, A. A. Covarrubias, *J. Biol. Chem.* **2000**, *275*, 5668–5674.
[320] M. Battaglia, Y. Olvera-Carrillo, A. Garciarrubio, F. Campos, A. A. Covarrubias, *Plant Physiol.* **2008**, *148*, 6–24.
[321] L. Dure III, *Plant J.* **1993**, *3*, 363–369.
[322] W. D. McCubbin, C. M. Kay, B. G. Lane, *Can. J. Biochem. Cell Biol.* **1985**, *63*, 803–811.
[323] J. Eom, W. R. Baker, A. Kintanar, E. S. Wurtele, *Plant Sci.* **1996**, *115*, 17–24.
[324] K. Goyal, L. Tisi, A. Basran, J. Browne, A. Burnell, J. Zurdo, A. Tunnacliffe, *J. Biol. Chem.* **2003**, *278*, 12977–12984.
[325] D. Li, X. He, *Biomacromolecules* **2009**, *10*, 1469–1477.
[326] T. Furuki, T. Shimizu, T. Kikawada, T. Okuda, M. Sakurai, *Biochemistry* **2011**, *50*, 7093–7103.
[327] K. Sasaki, N. K. Christov, S. Tsuda, R. Imai, *Plant Cell Physiol.* **2014**, *55*, 136–147.
[328] T. Ahlstrand, H. Tuominen, A. Beklen, A. Torittu, J. Oscarsson, R. Sormunen, M. T. Pöllänen, P. Permi, R. Ihalin, *Virulence* **2017**, *8*, 115–134.
[329] B. M. LeBlanc, M. T. Le, B. Janis, M. A. Menze, S. C. Hand, *Cell Stress Chaperones* **2019**, *24*, 979–990.
[330] M. Shih, S. Lin, J. Hsieh, C. Tsou, T. Chow, T. Lin, Y. C. Hsing, *Plant Mol. Biol.* **2004**, *56*, 689–703.
[331] M.-D. Shih, F. A. Hoekstra, Y.-I. C. Hsing, in *Adv. Bot. Res.*, Elsevier, **2008**, pp. 211–255.
[332] J. Lapinski, A. Tunnacliffe, *FEBS Lett.* **2003**, *553*, 387–390.
[333] B. Hespeels, X. Li, J.-F. Flot, L.-M. Pigneur, J. Malaisse, C. D. Silva, K. V. Doninck, *PLOS ONE* **2015**, *10*, e0131313.
[334] D. R. Nelson, R. Guidetti, L. Rebecchi, in *Ecol. Classif. North Am. Freshw. Invertebr.*, Elsevier, **2010**, pp. 455–484.
[335] N. Møbjerg, A. Jørgensen, R. M. Kristensen, R. C. Neves, in *Water Bears Biol. Tardigrades* (Ed.: R.O. Schill), Springer International Publishing, Cham, **2018**, pp. 57–94.
[336] C. Bonnet, J. A. E. Goeze, *Herrn Karl Bonnets Abhandlungen Aus Der Insektologie*, Bey J.J. Gebauers Wittwe Und Joh. Jac. Gebauer, Halle :, **1773**.
[337] H. Greven, *Acta Biolocica Benrodis* **2015**.
[338] L. Spallanzani 1729-1799, *Opuscoli Di Fisica Animale e Vegetabile [Microform] / Dell 'Abate Spallanzani ; Aggiuntevi Alcune Lettere Relative Ad Essi Opuscoli Dal Celebre*

Signor Bonnet Di Ginevra, e Da Altri Scritte All ' Autore, Presso La Societa' Tipografica, Modena, **1776**.
[339] P. Degma, R. Bertolani, R. Guidetti, others, **2019**.
[340] A. Jørgensen, R. M. Kristensen, N. Møbjerg, in *Water Bears Biol. Tardigrades* (Ed.: R.O. Schill), Springer International Publishing, Cham, **2018**, pp. 95–114.
[341] R. O. Schill, K. I. Jönsson, M. Pfannkuchen, F. Brümmer, *J. Zool. Syst. Evol. Res.* **2011**, *49*, 66–70.
[342] G. Rahm, *Zool Anz* **1937**, *120*, 65–71.
[343] G. T. Grothman, C. Johansson, G. Chilton, H. Kagoshima, M. Tsujimoto, A. C. Suzuki, *Zoolog. Sci.* **2017**, *34*, 5–10.
[344] K. I. Jönsson, I. Holm, H. Tassidis, in *Evo-Devo Non-Model Species Cell Dev. Biol.* (Eds.: W. Tworzydlo, S.M. Bilinski), Springer International Publishing, Cham, **2019**, pp. 231–249.
[345] N. Møbjerg, A. Jørgensen, R. M. Kristensen, R. C. Neves, in *Water Bears Biol. Tardigrades* (Ed.: R.O. Schill), Springer International Publishing, Cham, **2018**, pp. 57–94.
[346] J. R. Garey, S. J. McInnes, P. B. Nichols, in *Freshw. Anim. Divers. Assess.* (Eds.: E.V. Balian, C. Lévêque, H. Segers, K. Martens), Springer Netherlands, Dordrecht, **2008**, pp. 101–106.
[347] R. Bertolani, R. Guidetti, T. Altiero, D. R. Nelson, L. Rebecchi, in *Dormancy Aquat. Org. Theory Hum. Use Model.* (Eds.: V.R. Alekseev, B. Pinel-Alloul), Springer International Publishing, Cham, **2019**, pp. 43–51.
[348] D. R. Nelson, P. J. Bartels, N. Guil, in *Water Bears Biol. Tardigrades* (Ed.: R.O. Schill), Springer International Publishing, Cham, **2018**, pp. 163–210.
[349] L. Rebecchi, T. Altiero, R. Guidetti, *Invertebr. Surviv. J.* **2007**, *4*, 65–81.
[350] J. Renaud-Mornant, *Meteor Forschungsergebnisse Reihe Biol.* **1975**, *21*, 54–61.
[351] C. Bussau, *Zool. Scr.* **1992**, *21*, 79–92.
[352] R. O. Schill, F. Huhn, H.-R. Köhler, *Zool. Middle East* **2007**, *42*, 83–88.
[353] D. A. Neher, S. A. Lewins, T. R. Weicht, B. J. Darby, *J. Arid Environ.* **2009**, *73*, 672–677.
[354] K. Zawierucha, M. Ostrowska, T. R. Vonnahme, M. Devetter, A. Nawrot, S. Lehmann, M. Kolicka, *J. Limnol.* **2016**, *75*, 545–559.
[355] H. Dastych, *Mitt Hamb Zool Mus Inst* **2004**, *101*, 169–183.
[356] W. R. MILLER, J. D. MILLER, H. HEATWOLE, *Zool. J. Linn. Soc.* **1996**, *116*, 175–184.
[357] M. Tsujimoto, S. Imura, H. Kanda, *Cryobiology* **2016**, *72*, 78–81.
[358] Y. Yoshida, G. Koutsovoulos, D. R. Laetsch, L. Stevens, S. Kumar, D. D. Horikawa, K. Ishino, S. Komine, T. Kunieda, M. Tomita, M. Blaxter, K. Arakawa, *PLOS Biol.* **2017**, *15*, e2002266.
[359] L. I. Campbell, O. Rota-Stabelli, G. D. Edgecombe, T. Marchioro, S. J. Longhorn, M. J. Telford, H. Philippe, L. Rebecchi, K. J. Peterson, D. Pisani, *Proc. Natl. Acad. Sci.* **2011**, *108*, 15920–15924.
[360] O. Rota-Stabelli, E. Kayal, D. Gleeson, J. Daub, J. L. Boore, M. J. Telford, D. Pisani, M. Blaxter, D. V. Lavrov, *Genome Biol. Evol.* **2010**, *2*, 425–440.
[361] K. W. Cooper, *Psyche J. Entomol.* **1964**, *71*, 41–48.
[362] M. A. Mapalo, N. Robin, B. E. Boudinot, J. Ortega-Hernández, P. Barden, *Proc. R. Soc. B Biol. Sci.* **2021**, *288*, 20211760.
[363] D. R. Nelson, *Integr. Comp. Biol.* **2002**, *42*, 652–659.
[364] R. M. Kristensen, T. E. Hallas, *Zool. Scr.* **1980**, *9*, 113–127.
[365] K. A. Halberg, A. Jørgensen, N. Møbjerg, *PLoS One* **2013**, *8*, e85091.

[366] J. C. Wright, *J. Exp. Biol.* **1989**, *142*, 267–292.
[367] J. H. Crowe, *Biol. Bull.* **1972**, *142*, 407–416.
[368] D. D. Horikawa, T. Sakashita, C. Katagiri, M. Watanabe, T. Kikawada, Y. Nakahara, N. Hamada, S. Wada, T. Funayama, S. Higashi, Y. Kobayashi, T. Okuda, M. Kuwabara, *Int. J. Radiat. Biol.* **2006**, *82*, 843–848.
[369] A. Jørgensen, N. Møbjerg, *Mar. Biol. Res.* **2015**, *11*, 214–217.
[370] T. L. Hygum, L. K. B. Clausen, K. A. Halberg, A. Jørgensen, N. Møbjerg, *Zool. J. Linn. Soc.* **2016**, *178*, 907–911.
[371] A. Jørgensen, N. Møbjerg, R. M. Kristensen, *J. Limnol.* **2007**, *66*, 77.
[372] K. K. Newsham, N. R. Maslen, S. J. McInnes, *Cryoletters* **2006**, *27*, 291–294.
[373] D. Persson, K. A. Halberg, A. Jørgensen, C. Ricci, N. Møbjerg, R. M. Kristensen, *J. Zool. Syst. Evol. Res.* **2011**, *49*, 90–97.
[374] R. C. Neves, A. Møbjerg, M. Kodama, J. Ramos-Madrigal, M. T. P. Gilbert, N. Møbjerg, *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2022**, *267*, 111169.
[375] S. Hengherr, M. R. Worland, A. Reuner, F. Brümmer, R. O. Schill, *Physiol. Biochem. Zool.* **2009**, *82*, 749–755.
[376] K. Seki, M. Toyoshima, *Nature* **1998**, *395*, 853–854.
[377] F. Ono, M. Saigusa, T. Uozumi, Y. Matsushima, H. Ikeda, N. L. Saini, M. Yamashita, *J. Phys. Chem. Solids* **2008**, *69*, 2297–2300.
[378] K. I. Jönsson, E. Rabbow, R. O. Schill, M. Harms-Ringdahl, P. Rettberg, *Curr. Biol.* **2008**, *18*, R729–R731.
[379] A. Traspas, M. J. Burchell, *Astrobiology* **2021**, *21*, 845–852.
[380] J. C. Wright, *J. Ultrastruct. Mol. Struct. Res.* **1988**, *101*, 23–39.
[381] L. J. Rothschild, R. L. Mancinelli, **2001**, *409*.
[382] L. Rebecchi, T. Altiero, R. Guidetti, M. Cesari, R. Bertolani, M. Negroni, A. M. Rizzo, *Astrobiology* **2009**, *9*, 581–591.
[383] Kh. O. Kharkevych, N. G. Sergeeva, *Vestn. Zool.* **2013**, *47*, 17–27.
[384] P. Westh, H. Ramløv, *J. Exp. Zool.* **1991**, *258*, 303–311.
[385] R. Guidetti, T. Altiero, L. Rebecchi, *J. Insect Physiol.* **2011**, *57*, 567–576.
[386] S. Hengherr, A. G. Heyer, H.-R. Köhler, R. O. Schill, *FEBS J.* **2008**, *275*, 281–288.
[387] K. Ingemar Jönsson, O. Persson, *Open Zool. J.* **2010**, *3*.
[388] T. C. Boothby, H. Tapia, A. H. Brozena, S. Piszkiewicz, A. E. Smith, I. Giovannini, L. Rebecchi, G. J. Pielak, D. Koshland, B. Goldstein, *Mol. Cell* **2017**, *65*, 975-984.e5.
[389] K. Arakawa, K. Numata, *Mol. Cell* **2021**, *81*, 409–410.
[390] E. Schokraie, A. Hotz-Wagenblatt, U. Warnken, B. Mali, M. Frohme, F. Förster, T. Dandekar, S. Hengherr, R. O. Schill, M. Schnölzer, *PLOS ONE* **2010**, *5*, e9502.
[391] S. Tanaka, J. Tanaka, Y. Miwa, D. D. Horikawa, T. Katayama, K. Arakawa, A. Toyoda, T. Kubo, T. Kunieda, *PLOS ONE* **2015**, *10*, e0118272.
[392] A. M. Rizzo, M. Negroni, T. Altiero, G. Montorfano, P. Corsetto, P. Berselli, B. Berra, R. Guidetti, L. Rebecchi, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2010**, *156*, 115–121.
[393] R. O. Schill, G. H. B. Steinbrück, H.-R. Köhler, *J. Exp. Biol.* **2004**, *207*, 1607–1613.
[394] K. I. Jönsson, R. O. Schill, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2007**, *146*, 456–460.
[395] A. Yamaguchi, S. Tanaka, S. Yamaguchi, H. Kuwahara, C. Takamura, S. Imajoh-Ohmi, D. D. Horikawa, A. Toyoda, T. Katayama, K. Arakawa, A. Fujiyama, T. Kubo, T. Kunieda, *PLOS ONE* **2012**, *7*, e44209.
[396] A. Malki, J. Teulon, A. R. Camacho-Zarco, S. W. Chen, W. Adamski, D. Maurin, N. Salvi, J. Pellequer, M. Blackledge, *Angew. Chem. Int. Ed.* **2022**, *61*, e202109961.

[397] P. Gąsiorek, D. Stec, W. Morek, Ł. Michalczyk, *Zootaxa* **2018**, *4415*, DOI 10.11646/zootaxa.4415.1.2.
[398] M. Richaud, E. Le Goff, C. Cazevielle, F. Ono, Y. Mori, N. L. Saini, P. Cuq, S. Baghdiguian, N. Godefroy, S. Galas, *Sci. Rep.* **2020**, *10*, 4324.
[399] Y. Fukuda, Y. Miura, E. Mizohata, T. Inoue, *FEBS Lett.* **2017**, *591*, 2458–2469.
[400] Y. Fukuda, T. Inoue, *Protein Sci.* **2018**, *27*, 993–999.
[401] K. Miyazawa, S. G. Itoh, H. Watanabe, T. Uchihashi, S. Yanaka, M. Yagi-Utsumi, K. Kato, K. Arakawa, H. Okumura, *J. Phys. Chem. B* **2021**, *125*, 9145–9154.
[402] K. Miyazawa, S. G. Itoh, Y. Yoshida, K. Arakawa, H. Okumura, *J. Phys. Chem. B* **2022**, *126*, 2361–2368.
[403] T. Hashimoto, D. D. Horikawa, Y. Saito, H. Kuwahara, H. Kozuka-Hata, T. Shin-I, Y. Minakuchi, K. Ohishi, A. Motoyama, T. Aizu, A. Enomoto, K. Kondo, S. Tanaka, Y. Hara, S. Koshikawa, H. Sagara, T. Miura, S. Yokobori, K. Miyagawa, Y. Suzuki, T. Kubo, M. Oyama, Y. Kohara, A. Fujiyama, K. Arakawa, T. Katayama, A. Toyoda, T. Kunieda, *Nat. Commun.* **2016**, *7*, 12808.
[404] T. Hashimoto, T. Kunieda, *Life* **2017**, *7*, 26.
[405] C. Chavez, G. Cruz-Becerra, J. Fei, G. A. Kassavetis, J. T. Kadonaga, *eLife* **2019**, *8*, e47682.
[406] E. Burgio, P. Piscitelli, L. Migliore, *Int. J. Environ. Res. Public. Health* **2018**, *15*, 1971.
[407] C. Ricci, G. Riolo, C. Marzocchi, J. Brunetti, A. Pini, S. Cantara, *Biology* **2021**, *10*, DOI 10.3390/biology10100970.
[408] M. Zarubin, T. Azorskaya, O. Kuldoshina, S. Alekseev, S. Mitrofanov, E. Kravchenko, *iScience* **2023**, *26*, 106998.
[409] X.-H. Li, C. W. H. Yu, N. Gomez-Navarro, V. Stancheva, H. Zhu, A. Murthy, M. Wozny, K. Malhotra, C. M. Johnson, M. Blackledge, B. Santhanam, W. Liu, J. Huang, S. M. V. Freund, E. A. Miller, M. M. Babu, *PNAS Nexus* **2024**, *3*, pgae006.
[410] M. Kamilari, A. Jørgensen, M. Schiøtt, N. Møbjerg, *BMC Genomics* **2019**, *20*, 607.
[411] H. B. Lübsen, *Ausführliches Lehrbuch Der Analytischen Oder Höhern Geometrie Zum Selbstunterricht.*, Leipzig, **1885**.
[412] J. Cavanagh, N. J. Skelton, W. J. Fairbrother, M. Rance, A. G. Palmer, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, Amsterdam ; Boston, **2007**.
[413] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, Wiley, **2008**.
[414] E. Merzbacher, *Quantum Mechanics*, Wiley, **1998**.
[415] A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Press, **1961**.
[416] P. Alken, E. Thébault, C. D. Beggan, H. Amit, J. Aubert, J. Baerenzung, T. N. Bondar, W. J. Brown, S. Califf, A. Chambodut, A. Chulliat, G. A. Cox, C. C. Finlay, A. Fournier, N. Gillet, A. Grayver, M. D. Hammer, M. Holschneider, L. Huder, G. Hulot, T. Jager, C. Kloss, M. Korte, W. Kuang, A. Kuvshinov, B. Langlais, J.-M. Léger, V. Lesur, P. W. Livermore, F. J. Lowes, S. Macmillan, W. Magnes, M. Mandea, S. Marsal, J. Matzka, M. C. Metman, T. Minami, A. Morschhauser, J. E. Mound, M. Nair, S. Nakano, N. Olsen, F. J. Pavón-Carrasco, V. G. Petrov, G. Ropp, M. Rother, T. J. Sabaka, S. Sanchez, D. Saturnino, N. R. Schnepf, X. Shen, C. Stolle, A. Tangborn, L. Tøffner-Clausen, H. Toh, J. M. Torta, J. Varner, F. Vervelidou, P. Vigneron, I. Wardinski, J. Wicht, A. Woods, Y. Yang, Z. Zeren, B. Zhou, *Earth Planets Space* **2021**, *73*, 49.
[417] F. Bloch, *Phys. Rev.* **1946**, *70*, 460–474.
[418] H. M. McConnell, *J. Chem. Phys.* **1958**, *28*, 430–431.
[419] H. C. Torrey, *Phys. Rev.* **1956**, *104*, 563–565.
[420] I. Solomon, *Phys. Rev.* **1955**, *99*, 559–565.

[421] M. Goldman, *Quantum Description of High-Resolution NMR in Liquids*, Clarendon Press, **1988**.
[422] F. Bloch, *Phys. Rev.* **1957**, *105*, 1206–1222.
[423] R. K. Wangsness, F. Bloch, *Phys. Rev.* **1953**, *89*, 728–739.
[424] A. G. Redfield, *IBM J. Res. Dev.* **1957**, *1*, 19–31.
[425] A. G. Redfield, in *Adv. Magn. Opt. Reson.*, Elsevier, **1965**, pp. 1–32.
[426] H.-P. Breuer, F. Petruccione, *The Theory of Open Quantum Systems*, Clarendon Press, Oxford, **2009**.
[427] P. S. Hubbard, *Phys. Rev.* **1969**, *180*, 319–326.
[428] D. M. Brink, G. R. Satchler, *Angular Momentum*, Clarendon P., **1968**.
[429] L. G. Werbelow, D. M. Grant, in *Adv. Magn. Opt. Reson.*, Elsevier, **1977**, pp. 189–299.
[430] T. Yuwen, J. P. Brady, L. E. Kay, *J. Am. Chem. Soc.* **2018**, *140*, 2115–2126.
[431] N. Tjandra, A. Bax, *Science* **1997**, *278*, 1111–1114.
[432] M. Blackledge, *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23–61.
[433] I. Bertini, C. Luchinat, G. Parigi, *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**.
[434] G. G. Stokes, *Trans. Camb. Philos. Soc.* **1851**, *9*, 8.
[435] A. Einstein, *Ann. Phys.* **1905**, *4*.
[436] M. Von Smoluchowski, *Ann. Phys.* **1906**, *326*, 756–780.
[437] P. Langevin, *CR Acad Sci Paris* **1908**, *146*, 530.
[438] P. J. W. Debye, *Polar Molecules*, Chemical Catalog Company, **1929**.
[439] W. Adamski, N. Salvi, D. Maurin, J. Magnat, S. Milles, M. R. Jensen, A. Abyzov, C. J. Moreau, M. Blackledge, *J. Am. Chem. Soc.* **2019**, *141*, 17817–17829.
[440] P. De Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, **1979**.
[441] G. D. J. Phillies, G. S. Ullmann, K. Ullmann, T.-H. Lin, *J. Chem. Phys.* **1985**, *82*, 5242–5246.
[442] F. Perrin, *J Phys Radium* **1934**, *5*, 497–511.
[443] D. E. Woessner, *J. Chem. Phys.* **1962**, *37*, 647–654.
[444] M. Blackledge, F. Cordier, P. Dosset, D. Marion, *J. Am. Chem. Soc.* **1998**, *120*, 4538–4539.
[445] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
[446] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.
[447] B. Halle, *J. Chem. Phys.* **2009**, *131*, 224507.
[448] A. G. Palmer, *Chem. Rev.* **2004**, *104*, 3623–3640.
[449] R. Brueschweiler, P. E. Wright, *J. Am. Chem. Soc.* **1994**, *116*, 8426–8427.
[450] F. Ochsenbein, J.-M. Neumann, E. Guittet, C. Van Heijenoort, *Protein Sci.* **2002**, *11*, 957–964.
[451] K. Modig, F. M. Poulsen, *J. Biomol. NMR* **2008**, *42*, 163–177.
[452] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2016**, *138*, 6240–6251.
[453] H. A. Kramers, *Physica* **1940**, *7*, 284–304.
[454] M. Bier, I. Derényi, M. Kostur, R. D. Astumian, *Phys. Rev. E* **1999**, *59*, 6422–6432.
[455] L. Rondin, J. Gieseler, F. Ricci, R. Quidant, C. Dellago, L. Novotny, *Nat. Nanotechnol.* **2017**, *12*, 1130–1133.
[456] A. D. Bain, *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *43*, 63–103.
[457] A. G. Palmer, C. D. Kroenke, J. Patrick Loria, in *Methods Enzymol.*, Elsevier, **2001**, pp. 204–238.
[458] A. Sekhar, P. Vallurupalli, in *Two-Dimensional 2D NMR Methods*, John Wiley & Sons, Ltd, **2023**, pp. 435–460.

[459] I. W. Hamley, *Small-Angle Scattering: Theory, Instrumentation, Data and Applications*, Wiley, Hoboken, NJ, **2021**.
[460] B. J. Berne, R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*, Courier Corporation, **2000**.
[461] O. Glatter, O. Kratky, *Small Angle X-Ray Scattering*, Academic Press, **1982**.
[462] D. I. Svergun, M. H. J. Koch, *Rep. Prog. Phys.* **2003**, *66*, 1735–1782.
[463] I. Pilz, O. Glatter, O. Kratky, in *Enzyme Struct. Part H*, Academic Press, **1979**, pp. 148–249.
[464] A. N. Tikhonov, V. I. (Vasiliĭ I. Arsenin, *Solutions of Ill-Posed Problems*, Winston And Distributed Solely By Halsted Press, **1977**.
[465] O. Glatter, *J. Appl. Crystallogr.* **1977**, *10*, 415–421.
[466] A. Guinier, in *Ann. Phys.*, EDP Sciences, **1939**, pp. 161–237.
[467] D. I. Svergun, M. H. J. Koch, *Rep. Prog. Phys.* **2003**, *66*, 1735–1782.
[468] G. Porod, *Kolloid-Z.* **1951**, *124*, 83–114.
[469] P. Debye, H. R. Anderson, H. Brumberger, *J. Appl. Phys.* **1957**, *28*, 679–683.
[470] A. G. Kikhney, D. I. Svergun, *FEBS Lett.* **2015**, *589*, 2570–2577.
[471] M. M. Woolfson, *An Introduction to X-Ray Crystallography*, Cambridge Univ. Press, Cambridge, **1997**.
[472] J. Drenth, J. Mesters, *Principles of Protein X-Ray Crystallography*, Springer, New York, **2007**.
[473] Y. Waseda, E. Matsubara, K. Shinoda, *X-Ray Diffraction Crystallography: Introduction, Examples and Solved Problems*, Springer Berlin Heidelberg, Berlin, Heidelberg, **2011**.
[474] D. W. Green, V. M. Ingram, M. F. Perutz, W. L. Bragg, *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **1954**, *225*, 287–307.
[475] D. M. Blow, F. H. C. Crick, *Acta Crystallogr.* **1959**, *12*, 794–802.
[476] P. Karplus, G. Schulz, *Naturwissenschaften* **1985**, *72*, 212–213.
[477] P. Radivojac, Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, A. K. Dunker, *Protein Sci.* **2004**, *13*, 71–80.
[478] Z. Sun, Q. Liu, G. Qu, Y. Feng, M. T. Reetz, *Chem. Rev.* **2019**, *119*, 1626–1665.
[479] K. S. Schmitz, G. D. J. Phillies, *Phys. Today* **1991**, *44*, 66–66.
[480] J. W. Strutt, *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1871**, *41*, 447–454.
[481] G. Mie, *Ann. Phys.* **1908**, *330*, 377–445.
[482] K. Schätzel, *Appl. Phys. B Photophysics Laser Chem.* **1987**, *42*, 193–213.
[483] J. Stetefeld, S. A. McKenna, T. R. Patel, *Biophys. Rev.* **2016**, *8*, 409–427.
[484] N. Wiener, *Acta Math.* **1930**, *55*, 117–258.
[485] A. Khintchine, *Math. Ann.* **1934**, *109*, 604–615.
[486] R. J. Glauber, *Phys. Rev.* **1963**, *130*, 2529–2539.
[487] A. J. F. Siegert, M. I. of T. R. Laboratory, *On the Fluctuations in Signals Returned by Many Independently Moving Scatterers*, Radiation Laboratory, Massachusetts Institute Of Technology, **1943**.
[488] P. N. Pusey, W. Van Megen, *Phys. Stat. Mech. Its Appl.* **1989**, *157*, 705–741.
[489] J. G. H. Joosten, in *Appl. Scatt. Methods Dyn. Polym. Syst.* (Eds.: B. Ewen, E.W. Fischer, G. Fytas), Steinkopff, Darmstadt, **1993**, pp. 149–152.
[490] H. B. Bohidar, S. Ghosh, *Eur. Polym. J.* **2000**.
[491] M. Okamoto, T. Norisuye, M. Shibayama, *Macromolecules* **2001**, *34*, 8496–8502.
[492] D. E. Koppel, *J. Chem. Phys.* **1972**, *57*, 4814–4820.
[493] S. W. Provencher, *Comput. Phys. Commun.* **1982**, *27*, 229–242.
[494] S. W. Provencher, *Comput. Phys. Commun.* **1982**, *27*, 213–227.

[495] K. F. Morris, C. S. Johnson, *J. Am. Chem. Soc.* **1993**, *115*, 4291–4299.
[496] N. Das, P. Sen, *Biochemistry* **2018**, *57*, 6078–6089.
[497] G. Binnig, C. F. Quate, Ch. Gerber, *Phys. Rev. Lett.* **1986**, *56*, 930–933.
[498] P. J. Eaton, P. West, *Atomic Force Microscopy*, Oxford University Press, Oxford ; New York, **2010**.
[499] G. Haugstad, in *At. Force Microsc.*, John Wiley & Sons, Ltd, **2012**, pp. 1–32.
[500] G. Haugstad, in *At. Force Microsc.*, John Wiley & Sons, Ltd, **2012**, pp. 33–90.
[501] S. Buhmann, D. Welsch, *Prog. Quantum Electron.* **2007**, *31*, 51–130.
[502] J. N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, Burlington, MA, **2011**.
[503] S. N. Magonov, V. Elings, M.-H. Whangbo, *Surf. Sci.* **1997**, *375*, L385–L391.
[504] G. Haugstad, in *At. Force Microsc.*, John Wiley & Sons, Ltd, **2012**, pp. 137–186.
[505] G. Haugstad, in *At. Force Microsc.*, John Wiley & Sons, Ltd, **2012**, pp. 187–257.
[506] M. Marrese, V. Guarino, L. Ambrosio, *J. Funct. Biomater.* **2017**, *8*, 7.
[507] G. Haugstad, in *At. Force Microsc.*, John Wiley & Sons, Ltd, **2012**, pp. 91–136.
[508] J. Zhang, P. Chen, B. Yuan, W. Ji, Z. Cheng, X. Qiu, *Science* **2013**, *342*, 611–614.
[509] U. D. Schwarz, H. Haefke, P. Reimann, H. -J. Güntherodt, *J. Microsc.* **1994**, *173*, 183–197.
[510] F. Gołek, P. Mazur, Z. Ryszka, S. Zuber, *Appl. Surf. Sci.* **2014**, *304*, 11–19.
[511] D. Keller, *Surf. Sci.* **1991**, *253*, 353–364.
[512] J. S. Villarrubia, *Surf. Sci.* **1994**, *321*, 287–300.
[513] P. Markiewicz, M. C. Goh, *Langmuir* **1994**, *10*, 5–7.
[514] F. J. Giessibl, *Appl. Phys. Lett.* **2001**, *78*, 123–125.
[515] J. Canet-Ferrer, E. Coronado, A. Forment-Aliaga, E. Pinilla-Cienfuegos, *Nanotechnology* **2014**, *25*, 395703.
[516] S. W. Chen, J.-M. Teulon, C. Godon, J.-L. Pellequer, *J. Mol. Recognit.* **2016**, *29*, 51–55.
[517] S. W. Chen, A.-S. Banneville, J.-M. Teulon, J. Timmins, J.-L. Pellequer, *Nanoscale* **2020**, *12*, 22628–22638.
[518] L. Pauling, R. B. Corey, *Nature* **1953**, *171*, 59–61.
[519] *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.* **1950**, *203*, 321–357.
[520] R. Armen, D. O. V. Alonso, V. Daggett, *Protein Sci.* **2003**, *12*, 1145–1157.
[521] A. Wada, *Adv. Biophys.* **1976**, 1–63.
[522] W. G. J. Hol, P. T. van Duijnen, H. J. C. Berendsen, *Nature* **1978**, *273*, 443–446.
[523] M. L. Tillett, M. J. Blackledge, J. P. Derrick, L.-Y. Lian, T. J. Norwood, *Protein Sci.* **2000**, *9*, 1210–1216.
[524] F. H. C. Crick, *Nature* **1952**, *170*, 882–883.
[525] F. H. C. Crick, IUCr, "The packing of -helices: simple coiled-coils," **1953**.
[526] A. N. Lupas, M. Gruber, *Adv. Protein Chem.* **2005**, *70*, 37–38.
[527] A. N. Lupas, J. Bassler, *Trends Biochem. Sci.* **2017**, *42*, 130–140.
[528] C. Chothia, M. Levitt, D. Richardson, *Proc. Natl. Acad. Sci.* **1977**, *74*, 4130–4134.
[529] C. Chothia, M. Levitt, D. Richardson, *J. Mol. Biol.* **1981**, *145*, 215–250.
[530] D. Walther, F. Eisenhaber, P. Argos, *J. Mol. Biol.* **1996**, *255*, 536–553.
[531] A. N. Lupas, J. Bassler, S. Dunin-Horkawicz, *Fibrous Proteins Struct. Mech.* **2017**, *82*, 95–129.
[532] D. A. D. Parry, R. D. B. Fraser, J. M. Squire, *J. Struct. Biol.* **2008**, *163*, 258–269.
[533] M. D. Hartmann, in *Fibrous Proteins Struct. Mech.* (Eds.: D.A.D. Parry, J.M. Squire), Springer International Publishing, Cham, **2017**, pp. 63–93.
[534] P. B. Harbury, P. S. Kim, T. Alber, *Nature* **1994**, *371*, 80–83.
[535] D. N. Woolfson, in *Adv. Protein Chem.*, Elsevier, **2005**, pp. 79–112.

[536] J. M. Fletcher, G. J. Bartlett, A. L. Boyle, J. J. Danon, L. E. Rush, A. N. Lupas, D. N. Woolfson, *ACS Chem. Biol.* **2017**, *12*, 528–538.
[537] P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **1993**, *262*, 1401–1407.
[538] K. J. Lumb, **n.d.**
[539] D. M. Eckert, V. N. Malashkevich, P. S. Kim, *J. Mol. Biol.* **1998**, *284*, 859–865.
[540] M. D. Hartmann, O. Ridderbusch, K. Zeth, R. Albrecht, O. Testa, D. N. Woolfson, G. Sauer, S. Dunin-Horkawicz, A. N. Lupas, B. H. Alvarez, *Proc. Natl. Acad. Sci.* **2009**, *106*, 16950–16955.
[541] J. H. Brown, C. Cohen, D. A. D. Parry, *Proteins Struct. Funct. Genet.* **1996**, *26*, 134–145.
[542] K. Szczepaniak, G. Lach, J. M. Bujnicki, S. Dunin-Horkawicz, *J. Struct. Biol.* **2014**, *188*, 123–133.
[543] M. Gruber, *Trends Biochem. Sci.* **2003**, *28*, 679–685.
[544] K. Szczepaniak, J. Ludwiczak, A. Winski, S. Dunin-Horkawicz, *J. Struct. Biol.* **2018**, *204*, 117–124.
[545] L. Dure III, *Plant J.* **1993**, *3*, 363–369.
[546] J. Peters, W. Baumeister, A. Lupas, *J. Mol. Biol.* **1996**, *257*, 1031–1041.
[547] M. R. Hicks, D. V. Holberton, C. Kowalczyk, D. N. Woolfson, *Fold. Des.* **1997**, *2*, 149–158.
[548] J. Stetefeld, M. Jenny, T. Schulthess, R. Landwehr, J. Engel, R. A. Kammerer, *Nat. Struct. Biol.* **2000**, *7*, 772–776.
[549] N. Tarbouriech, J. Curran, R. W. H. Ruigrok, W. P. Burmeister, *Nat. Struct. Biol.* **2000**, *7*, 777–781.
[550] G. Grigoryan, W. F. DeGrado, *J. Mol. Biol.* **2011**, *405*, 1079–1100.
[551] K. Szczepaniak, A. Bukala, A. M. da Silva Neto, J. Ludwiczak, S. Dunin-Horkawicz, *Bioinformatics* **2021**, *36*, 5368–5376.
[552] D. L. Akey, V. N. Malashkevich, P. S. Kim, *Biochemistry* **2001**, *40*, 6352–6360.
[553] S. Dunin-Horkawicz, A. N. Lupas, *J. Struct. Biol.* **2010**, *170*, 226–235.
[554] E. B. Hadley, O. D. Testa, D. N. Woolfson, S. H. Gellman, *Proc. Natl. Acad. Sci.* **2008**, *105*, 530–535.
[555] E. B. Hadley, S. H. Gellman, *J. Am. Chem. Soc.* **2006**, *128*, 16444–16445.
[556] H. Herrmann, H. Bär, L. Kreplak, S. V. Strelkov, U. Aebi, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 562–573.
[557] N. Geisler, E. Kaufmann, K. Weber, *J. Mol. Biol.* **1985**, *182*, 173–177.
[558] A. A. Chernyatina, D. Guzenko, S. V. Strelkov, *Curr. Opin. Cell Biol.* **2015**, *32*, 65–72.
[559] S. A. Eldirany, I. B. Lomakin, M. Ho, C. G. Bunick, *Curr. Opin. Cell Biol.* **2021**, *68*, 132–143.
[560] S. V. Strelkov, H. Herrmann, N. Geisler, T. Wedig, R. Zimbelmann, U. Aebi, P. Burkhard, *EMBO J.* **2002**, *21*, 1255–1266.
[561] M. Meier, G. P. Padilla, H. Herrmann, T. Wedig, M. Hergt, T. R. Patel, J. Stetefeld, U. Aebi, P. Burkhard, *J. Mol. Biol.* **2009**, *390*, 245–261.
[562] G. Stalmans, A. V. Lilina, P.-J. Vermeire, J. Fiala, P. Novák, S. V. Strelkov, *Cells* **2020**, *9*, 1633.
[563] A. Aziz, J. F. Hess, M. S. Budamagunta, J. C. Voss, A. P. Kuzin, Y. J. Huang, R. Xiao, G. T. Montelione, P. G. FitzGerald, J. F. Hunt, *J. Biol. Chem.* **2012**, *287*, 28349–28361.
[564] J. Ahn, I. Jo, S. Kang, S. Hong, S. Kim, S. Jeong, Y.-H. Kim, B.-J. Park, N.-C. Ha, *Nat. Commun.* **2019**, *10*, 3757.
[565] S. A. Eldirany, M. Ho, A. J. Hinbest, I. B. Lomakin, C. G. Bunick, *EMBO J.* **2019**, *38*, e100741.

[566] S. Nicolet, H. Herrmann, U. Aebi, S. V. Strelkov, *J. Struct. Biol.* **2010**, *170*, 369–376.
[567] S. V. Strelkov, J. Schumacher, P. Burkhard, U. Aebi, H. Herrmann, *J. Mol. Biol.* **2004**, *343*, 1067–1080.
[568] C.-H. Lee, M.-S. Kim, B. M. Chung, D. J. Leahy, P. A. Coulombe, *Nat. Struct. Mol. Biol.* **2012**, *19*, 707–715.
[569] C. G. Bunick, L. M. Milstone, *J. Invest. Dermatol.* **2017**, *137*, 142–150.
[570] I. B. Lomakin, A. J. Hinbest, M. Ho, S. A. Eldirany, C. G. Bunick, *Yale J. Biol. Med.* **2020**, *93*, 3–17.
[571] H. Herrmann, S. V. Strelkov, *BMC Biol.* **2011**, *9*, 16.
[572] M. Arslan, Z. Qin, M. J. Buehler, *Comput. Methods Biomech. Biomed. Engin.* **2011**, *14*, 483–489.
[573] D. A. D. Parry, *J. Struct. Biol.* **2006**, *155*, 370–374.
[574] A. A. Chernyatina, S. V. Strelkov, *J. Struct. Biol.* **2012**, *177*, 46–53.
[575] M. J. Pandya, G. M. Spooner, M. Sunde, J. R. Thorpe, A. Rodger, D. N. Woolfson, *Biochemistry* **2000**, *39*, 8728–8734.
[576] D. Papapostolou, A. M. Smith, E. D. T. Atkins, S. J. Oliver, M. G. Ryadnov, L. C. Serpell, D. N. Woolfson, *Proc. Natl. Acad. Sci.* **2007**, *104*, 10853–10858.
[577] H. Dong, S. E. Paramonov, J. D. Hartgerink, *J. Am. Chem. Soc.* **2008**, *130*, 13691–13695.
[578] E. D. Santis, V. Castelletto, M. G. Ryadnov, *Phys. Chem. Chem. Phys.* **2015**, *17*, 31055–31060.
[579] N. C. Burgess, T. H. Sharp, F. Thomas, C. W. Wood, A. R. Thomson, N. R. Zaccai, R. L. Brady, L. C. Serpell, D. N. Woolfson, *J. Am. Chem. Soc.* **2015**, *137*, 10554–10562.
[580] E. F. Banwell, E. S. Abelardo, D. J. Adams, M. A. Birchall, A. Corrigan, A. M. Donald, M. Kirkland, L. C. Serpell, M. F. Butler, D. N. Woolfson, *Nat. Mater.* **2009**, *8*, 596–600.
[581] W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, D. A. Tirrell, *Science* **1998**, *281*, 389–392.
[582] C. Xu, V. Breedveld, J. Kopeček, *Biomacromolecules* **2005**, *6*, 1739–1749.
[583] J. T. Nielsen, F. A. A. Mulder, *J. Biomol. NMR* **2018**, *70*, 141–165.
[584] R. Mohana-Borges, N. K. Goto, G. J. A. Kroon, H. J. Dyson, P. E. Wright, *J. Mol. Biol.* **2004**, *340*, 1131–1142.
[585] P. Bernadó, C. W. Bertoncini, C. Griesinger, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969.
[586] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, M. Blackledge, *J. Am. Chem. Soc.* **2008**, *130*, 8055–8061.
[587] M. R. Jensen, P. R. L. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernadó, M. Blackledge, *Structure* **2009**, *17*, 1169–1185.
[588] V. Iešmantavičius, M. R. Jensen, V. Ozenne, M. Blackledge, F. M. Poulsen, M. Kjaergaard, *J. Am. Chem. Soc.* **2013**, *135*, 10155–10163.
[589] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795–2804.
[590] V. Ozenne, F. Bauer, L. Salmon, J. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Bioinformatics* **2012**, *28*, 1463–1470.
[591] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.
[592] P. Tsan, J.-C. Hus, M. Caffrey, D. Marion, M. Blackledge, *J. Am. Chem. Soc.* **2000**, *122*, 5603–5612.
[593] L. K. Lee, M. Rance, W. J. Chazin, A. G. Palmer, *J. Biomol. NMR* **1997**, *9*, 287–298.
[594] J. B. Hall, D. Fushman, *J. Biomol. NMR* **2003**, *27*, 261–275.
[595] J. M. Kneller, M. Lu, C. Bracken, *J. Am. Chem. Soc.* **2002**, *124*, 1852–1853.

[596] L. Frey, N. Lakomek, R. Riek, S. Bibow, *Angew. Chem. Int. Ed.* **2017**, *56*, 380–383.
[597] N. E. Zhou, B. Y. Zhu, B. D. Sykes, R. S. Hodges, *J. Am. Chem. Soc.* **1992**, *114*, 4320–4326.
[598] T. Asakura, K. Taoka, M. Demura, MichaelP. Williamson, *J. Biomol. NMR* **1995**, *6*, DOI 10.1007/BF00197804.
[599] S. Bibow, Y. Polyhach, C. Eichmann, C. N. Chi, J. Kowal, S. Albiez, R. A. McLeod, H. Stahlberg, G. Jeschke, P. Güntert, R. Riek, *Nat. Struct. Mol. Biol.* **2017**, *24*, 187–193.
[600] I. Caballero, M. Sammito, C. Millán, A. Lebedev, N. Soler, I. Usón, *Acta Crystallogr. Sect. Struct. Biol.* **2018**, *74*, 194–204.
[601] M. Sammito, C. Millán, D. Frieske, E. Rodríguez-Freire, R. J. Borges, I. Usón, *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71*, 1921–1930.
[602] A. L. Ringer, A. Senenko, C. D. Sherrill, *Protein Sci. Publ. Protein Soc.* **2007**, *16*, 2216–2223.
[603] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2016**, *138*, 6240–6251.
[604] S. Guseva, V. Schnapka, W. Adamski, D. Maurin, R. W. H. Ruigrok, N. Salvi, M. Blackledge, *J. Am. Chem. Soc.* **2023**, *145*, 10548–10563.
[605] A. Kentsis, T. R. Sosnick, *Biochemistry* **1998**, *37*, 14613–14622.
[606] M. Goodman, I. G. Rosen, *Biopolymers* **1964**, *2*, 537–559.
[607] V. L. Anderson, W. W. Webb, *Biophys. J.* **2012**, *102*, 897–906.
[608] M. R. Berwick, D. J. Lewis, A. W. Jones, R. A. Parslow, T. R. Dafforn, H. J. Cooper, J. Wilkie, Z. Pikramenou, M. M. Britton, A. F. A. Peacock, *J. Am. Chem. Soc.* **2014**, *136*, 1166–1169.
[609] L. N. Slope, O. J. Daubney, H. Campbell, S. A. White, A. F. A. Peacock, *Angew. Chem.* **2021**, *133*, 24678–24682.
[610] A. M. Webster, A. F. A. Peacock, *Chem. Commun.* **2021**, *57*, 6851–6862.
[611] A. Yamaguchi, S. Tanaka, S. Yamaguchi, H. Kuwahara, C. Takamura, S. Imajoh-Ohmi, D. D. Horikawa, A. Toyoda, T. Katayama, K. Arakawa, A. Fujiyama, T. Kubo, T. Kunieda, *PLOS ONE* **2012**, *7*, e44209.
[612] D. Kang, M. J. Yang, H. Kim, C. Park, *Protein Sci.* **2024**, *33*, e4913.
[613] J. M. H. Thomas, R. M. Keegan, J. Bibby, M. D. Winn, O. Mayans, D. J. Rigden, *IUCrJ* **2015**, *2*, 198–206.
[614] T. C. Boothby, H. Tapia, A. H. Brozena, S. Piszkiewicz, A. E. Smith, I. Giovannini, L. Rebecchi, G. J. Pielak, D. Koshland, B. Goldstein, *Mol. Cell* **2017**, *65*, 975-984.e5.
[615] L. Dure, M. Crouch, J. Harada, T.-H. D. Ho, J. Mundy, R. Quatrano, T. Thomas, Z. R. Sung, *Plant Mol. Biol.* **1989**, *12*, 475–486.
[616] P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **1998**, *282*, 1462–1467.
[617] K. Goyal, L. Tisi, A. Basran, J. Browne, A. Burnell, J. Zurdo, A. Tunnacliffe, *J. Biol. Chem.* **2003**, *278*, 12977–12984.
[618] D. Li, X. He, *Biomacromolecules* **2009**, *10*, 1469–1477.
[619] C. Hesgrove, T. C. Boothby, *Cell Commun. Signal.* **2020**, *18*, 178.
[620] G. N. Phillips, *J. Mol. Biol.* **1986**, *192*, 128–131.
[621] F. G. Whitby, G. N. Phillips, *Proteins* **2000**, *38*, 49–59.
[622] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe, D. Eisenberg, *Nature* **2005**, *435*, 773–778.
[623] M. R. Sawaya, D. Eisenberg, *Amyloid Fibrils Prefibrillar Aggreg. Mol. Biol. Prop.* **2013**, 21–38.
[624] E. Tayeb-Fligelman, O. Tabachnikov, A. Moshe, O. Goldshmidt-Tran, M. R. Sawaya, N. Coquelle, J.-P. Colletier, M. Landau, *Science* **2017**, *355*, 831–833.

[625] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, D. Eisenberg, *Nature* **2007**, *447*, 453–457.
[626] D. Guzenko, A. A. Chernyatina, S. V. Strelkov, in *Fibrous Proteins Struct. Mech.* (Eds.: D.A.D. Parry, J.M. Squire), Springer International Publishing, Cham, **2017**, pp. 151–170.
[627] A. Malki, J. Teulon, A. R. Camacho-Zarco, S. W. Chen, W. Adamski, D. Maurin, N. Salvi, J. Pellequer, M. Blackledge, *Angew. Chem. Int. Ed.* **2022**, *61*, e202109961.
[628] M. Yagi-Utsumi, K. Aoki, H. Watanabe, C. Song, S. Nishimura, T. Satoh, S. Yanaka, C. Ganser, S. Tanaka, V. Schnapka, E. W. Goh, Y. Furutani, K. Murata, T. Uchihashi, K. Arakawa, K. Kato, *Sci. Rep.* **2021**, *11*, 21328.
[629] S. Sanchez-Martinez, K. Nguyen, S. Biswas, V. Nicholson, A. V. Romanyuk, J. Ramirez, S. Kc, A. Akter, C. Childs, E. K. Meese, E. T. Usher, G. M. Ginell, F. Yu, E. Gollub, M. Malferrari, F. Francia, G. Venturoli, E. W. Martin, F. Caporaletti, G. Giubertoni, S. Woutersen, S. Sukenik, D. N. Woolfson, A. S. Holehouse, T. C. Boothby, *Protein Sci.* **2024**, *33*, e4941.
[630] A. Premchandar, N. Mücke, J. Poznański, T. Wedig, M. Kaus-Drobek, H. Herrmann, M. Dadlez, *J. Biol. Chem.* **2016**, *291*, 24931–24950.
[631] P.-J. Vermeire, G. Stalmans, A. V. Lilina, J. Fiala, P. Novak, H. Herrmann, S. V. Strelkov, *Cells* **2021**, *10*, 2457.
[632] P. M. Steinert, L. N. Marekov, D. A. Parry, *J. Biol. Chem.* **1993**, *268*, 24916–24925.
[633] A. V. Sokolova, L. Kreplak, T. Wedig, N. Mücke, D. I. Svergun, H. Herrmann, U. Aebi, S. V. Strelkov, *Proc. Natl. Acad. Sci.* **2006**, *103*, 16206–16211.
[634] A. V. Lilina, A. A. Chernyatina, D. Guzenko, S. V. Strelkov, *J. Struct. Biol.* **2020**, *209*, 107404.
[635] B. Kim, S. Kim, M. S. Jin, *Biochem. Biophys. Res. Commun.* **2018**, *503*, 2899–2905.
[636] M. Eibauer, M. S. Weber, R. Kronenberg-Tenga, C. T. Beales, R. Boujemaa-Paterski, Y. Turgay, S. Sivagurunathan, J. Kraxner, S. Köster, R. D. Goldman, O. Medalia, *Nat. Struct. Mol. Biol.* **2024**, *31*, 939–949.
[637] *Mol. Cells* **2024**, *47*, 100080.
[638] L. Hering, J.-E. Bouameur, J. Reichelt, T. M. Magin, G. Mayer, *eLife* **2016**, *5*, e11117.
[639] D. A. Fletcher, R. D. Mullins, *Nature* **2010**, *463*, 485–492.
[640] A. F. Pegoraro, P. Janmey, D. A. Weitz, *Cold Spring Harb. Perspect. Biol.* **2017**, *9*, a022038.
[641] A. Tanaka, T. Nakano, K. Watanabe, K. Masuda, G. Honda, S. Kamata, R. Yasui, H. Kozuka-Hata, C. Watanabe, T. Chinen, D. Kitagawa, S. Sawai, M. Oyama, M. Yanagisawa, T. Kunieda, *PLOS Biol.* **2022**, *20*, e3001780.
[642] J. E. Eicher, J. A. Brom, S. Wang, S. S. Sheiko, J. M. Atkin, G. J. Pielak, *Protein Sci.* **2022**, *31*, e4495.
[643] Y. Jiang, W. Zhang, F. Yang, C. Wan, X. Cai, J. Liu, Q. Zhang, Z. Li, W. Han, *Sci. Adv.* **2021**, *7*, eabd0492.
[644] B. Janis, C. Belott, T. Brockman, M. A. Menze, *Biomolecules* **2022**, *12*, 425.
[645] V. Ferreiro, J. F. Douglas, J. A. Warren, A. Karim, *Phys. Rev. E* **2002**, *65*, 042802.
[646] L. Gránásy, T. Pusztai, T. Börzsönyi, J. A. Warren, J. F. Douglas, *Nat. Mater.* **2004**, *3*, 645–650.
[647] L. Gránásy, T. Pusztai, J. F. Douglas, in *Encycl. Polym. Compos.* (Ed.: S. Palsule), Springer Berlin Heidelberg, Berlin, Heidelberg, **2013**, pp. 1–35.
[648] J. Xu, G. Reiter, R. Alamo, *Crystals* **2021**, *11*, 304.

[649] Y. Yoshimura, Y. Lin, H. Yagi, Y.-H. Lee, H. Kitayama, K. Sakurai, M. So, H. Ogi, H. Naiki, Y. Goto, *Proc. Natl. Acad. Sci.* **2012**, *109*, 14446–14451.
[650] M. Adachi, M. So, K. Sakurai, J. Kardos, Y. Goto, *J. Biol. Chem.* **2015**, *290*, 18134–18145.
[651] M. So, D. Hall, Y. Goto, *Curr. Opin. Struct. Biol.* **2016**, *36*, 32–39.
[652] M. Lepère, C. Chevallard, G. Brezesinski, M. Goldmann, P. Guenoun, *Angew. Chem.* **2009**, *121*, 5105–5109.
[653] L. Emmanouilidis, E. Bartalucci, Y. Kan, M. Ijavi, M. E. Pérez, P. Afanasyev, D. Boehringer, J. Zehnder, S. H. Parekh, M. Bonn, T. C. T. Michaels, T. Wiegand, F. H.-T. Allain, *Nat. Chem. Biol.* **2024**, *20*, 1044–1052.
[654] D. J. Adams, K. Morris, L. Chen, L. C. Serpell, J. Bacsa, G. M. Day, *Soft Matter* **2010**, *6*, 4144.
[655] P. Zhu, X. Yan, Y. Su, Y. Yang, J. Li, *Chem. – Eur. J.* **2010**, *16*, 3176–3183.
[656] P. Terech, N. M. Sangeetha, U. Maitra, *J. Phys. Chem. B* **2006**, *110*, 15224–15233.
[657] B. Roy, P. Bairi, A. K. Nandi, *Soft Matter* **2012**, *8*, 2366.
[658] J. R. Moffat, D. K. Smith, *Chem. Commun.* **2008**, 2248.
[659] Y. Wang, L. Tang, J. Yu, *Cryst. Growth Des.* **2008**, *8*, 884–889.
[660] I. Ramos Sasselli, P. J. Halling, R. V. Ulijn, T. Tuttle, *ACS Nano* **2016**, *10*, 2661–2668.
[661] L. Onsager, *Ann. N. Y. Acad. Sci.* **1949**, *51*, 627–659.
[662] P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, **1953**.
[663] P. J. Flory, A. Abe, *Macromolecules* **1978**, *11*, 1119–1122.
[664] M. Warner, P. J. Flory, *J. Chem. Phys.* **1980**, *73*, 6327–6332.
[665] A. R. Khokhlov, A. N. Semenov, *Phys. Stat. Mech. Its Appl.* **1982**, *112*, 605–614.
[666] Z. Y. Chen, *Macromolecules* **1993**, *26*, 3419–3423.
[667] M. Franco-Melgar, A. J. Haslam, G. Jackson, *Mol. Phys.* **2008**, *106*, 649–678.
[668] R. Furukawa, R. Kundra, M. Fechheimer, *Biochemistry* **1993**, *32*, 12346–12352.
[669] J.-M. Jung, R. Mezzenga, *Langmuir* **2010**, *26*, 504–514.
[670] I. W. Hamley, *Soft Matter* **2010**, *6*, 1863.
[671] D. Cannon, A. M. Donald, *Soft Matter* **2013**, *9*, 2852.
[672] M. Bagnani, G. Nyström, C. De Michele, R. Mezzenga, *ACS Nano* **2019**, *13*, 591–600.
[673] Z. Wang, P. Servio, A. D. Rey, *Front. Soft Matter* **2022**, *2*, DOI 10.3389/frsfm.2022.904069.
[674] J. Doye, W. Poon, *Curr. Opin. Colloid Interface Sci.* **2006**, *11*, 40–46.
[675] B. S. Heater, Z. Yang, M. M. Lee, M. K. Chan, *J. Am. Chem. Soc.* **2020**, *142*, 9879–9883.
[676] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792–1797.
[677] F. Madeira, N. Madhusoodanan, J. Lee, A. Eusebi, A. Niewielska, A. R. N. Tivey, R. Lopez, S. Butcher, *Nucleic Acids Res.* **2024**, *52*, W521–W525.
[678] J. F. Fleming, D. Pisani, K. Arakawa, *Genome Biol. Evol.* **2024**, *16*, evad217.
[679] J. Ludwiczak, A. Winski, K. Szczepaniak, V. Alva, S. Dunin-Horkawicz, *Bioinforma. Oxf. Engl.* **2019**, *35*, 2790–2795.
[680] P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, *128*, 9042–9043.
[681] E. Lescop, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2007**, *187*, 163–169.
[682] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.
[683] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687—696.
[684] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *5*, 67–81.

[685] K. Tamiola, B. Acar, F. A. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 18000–18003.
[686] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R. W. H. Ruigrok, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2016**, *138*, 6240–6251.
[687] N.-A. Lakomek, J. Ying, A. Bax, *J. Biomol. NMR* **2012**, *53*, 209–221.
[688] M. Akke, A. G. Palmer, *J. Am. Chem. Soc.* **1996**, *118*, 911–912.
[689] P. Pelupessy, G. M. Espallargas, G. Bodenhausen, *J. Magn. Reson.* **2003**, *161*, 258–264.
[690] P. Pelupessy, F. Ferrage, G. Bodenhausen, *J. Chem. Phys.* **2007**, *126*, 134508.
[691] J. Braman, C. Papworth, A. Greener, in *Vitro Mutagen. Protoc.* (Ed.: M.K. Trower), Humana Press, Totowa, NJ, **1996**, pp. 31–44.
[692] R. M. Rasia, E. Lescop, J. F. Palatnik, J. Boisbouvier, B. Brutscher, *J. Biomol. NMR* **2011**, *51*, 369–378.
[693] M. Zweckstetter, *Nat. Protoc.* **2008**, *3*, 679–690.
[694] A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Pr, Oxford, **1986**.
[695] P. Vallurupalli, G. Bouvignies, L. E. Kay, *J. Am. Chem. Soc.* **2012**, *134*, 8148–8161.
[696] V. Ozenne, F. Bauer, L. Salmon, J.-R. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Bioinforma. Oxf. Engl.* **2012**, *28*, 1463–1470.
[697] D. Svergun, C. Barberato, M. H. J. Koch, *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
[698] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.
[699] M. Nanao, S. Basu, U. Zander, T. Giraud, J. Surr, M. Guijarro, M. Lentini, F. Felisaz, J. Sinoir, C. Morawe, A. Vivo, A. Beteva, M. Oscarsson, H. Caserotto, F. Dobias, D. Flot, D. Nurizzo, J. Gigmes, N. Foos, R. Siebrecht, T. Roth, P. Theveneau, O. Svensson, G. Papp, B. Lavault, F. Cipriani, R. Barrett, C. Clavel, G. Leonard, *J. Synchrotron Radiat.* **2022**, *29*, 581–590.
[700] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.
[701] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
[702] I. Usón, G. M. Sheldrick, *Acta Crystallogr. Sect. Struct. Biol.* **2018**, *74*, 106–116.
[703] Collaborative Computational Project, Number 4, *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760–763.
[704] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242.
[705] P. Emsley, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.
[706] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501.
[707] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221.
[708] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367.
[709] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12–21.

[710] T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, *Protein Sci. Publ. Protein Soc.* **2018**, *27*, 14–25.
[711] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774–797.
[712] D. Nečas, P. Klapetek, *Open Phys.* **2012**, *10*, 181–188.
[713] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, *Nat. Methods* **2017**, *14*, 331–332.
[714] K. Zhang, *J. Struct. Biol.* **2016**, *193*, 1–12.
[715] S. H. W. Scheres, *J. Struct. Biol.* **2012**, *180*, 519–530.
[716] J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, S. H. Scheres, *eLife* **2018**, *7*, e42166.
[717] A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, *Nat. Methods* **2017**, *14*, 290–296.
[718] M. Djabourov, K. Nishinari, S. B. Ross-Murphy, *Physical Gels from Biological and Synthetic Polymers*, Cambridge University Press, **2013**.
[719] P. J. Flory, *J. Am. Chem. Soc.* **1941**, *63*, 3083–3090.
[720] W. H. Stockmayer, *J. Chem. Phys.* **1944**, *12*, 125–131.
[721] A. Coniglio, H. E. Stanley, W. Klein, *Phys. Rev. Lett.* **1979**, *42*, 518–522.
[722] D. Stauffer, A. Aharony, *Introduction To Percolation Theory*, Taylor & Francis, **2018**.
[723] M. Okamoto, T. Norisuye, M. Shibayama, *Macromolecules* **2001**, *34*, 8496–8502.
[724] M. Asai, T. Katashima, U. Chung, T. Sakai, M. Shibayama, *Macromolecules* **2013**, *46*, 9772–9781.
[725] S. Seiffert, *Polym. Chem.* **2017**, *8*, 4472–4487.
[726] P. N. Pusey, W. Van Megen, *Phys. Stat. Mech. Its Appl.* **1989**, *157*, 705–741.
[727] J. G. H. Joosten, in *Appl. Scatt. Methods Dyn. Polym. Syst.* (Eds.: B. Ewen, E.W. Fischer, G. Fytas), Steinkopff, Darmstadt, **1993**, pp. 149–152.
[728] A. Moussaïd, P. N. Pusey, J. J. M. Slot, J. G. H. Joosten, *Macromolecules* **1999**, *32*, 3774–3787.
[729] S. Seiffert, *Prog. Polym. Sci.* **2017**, *66*, 1–21.
[730] M. Shibayama, T. Norisuye, *Bull. Chem. Soc. Jpn.* **2002**, *75*, 641–659.
[731] X. Li, S. Nakagawa, Y. Tsuji, N. Watanabe, M. Shibayama, *Sci. Adv.* **2019**, *5*, eaax8647.
[732] J. E. Martin, J. Wilcoxon, J. Odinek, *Phys. Rev. A* **1991**, *43*, 858–872.
[733] T. Matsunaga, M. Shibayama, *Phys. Rev. E* **2007**, *76*, 030401.
[734] A. V. Teixeira, E. Geissler, P. Licinio, *J. Phys. Chem. B* **2007**, *111*, 340–344.
[735] Y. Piñeiro, M. A. López-Quintela, J. Rivas, D. Leisner, *Phys. Rev. E* **2009**, *79*, 041409.
[736] N. Quennouz, S. M. Hashmi, H. Sung Choi, J. Woong Kim, C. O. Osuji, *Soft Matter* **2016**, *12*, 157–164.
[737] I. Dueramae, S. Fukuzawa, N. Shinyashiki, S. Yagihara, R. Kita, *J. Biorheol.* **2017**, *31*, 21–28.
[738] M. Djabourov, K. Nishinari, S. B. Ross-Murphy, in *Phys. Gels Biol. Synth. Polym.*, Cambridge University Press, **2013**, pp. 18–63.
[739] G. W. H. Höhne, W. F. Hemminger, H.-J. Flammersheim, *Differential Scanning Calorimetry*, Springer Berlin Heidelberg, Berlin, Heidelberg, **2003**.
[740] A. S. Hoffman, *Adv. Drug Deliv. Rev.* **2002**.
[741] Y. Zhang, Y. Huang, *Front. Chem.* **2021**, *8*, 615665.
[742] M. Djabourov, K. Nishinari, S. B. Ross-Murphy, in *Phys. Gels Biol. Synth. Polym.*, Cambridge University Press, **2013**, pp. 156–181.
[743] J. V. Alemán, A. V. Chadwick, J. He, M. Hess, K. Horie, R. G. Jones, P. Kratochvíl, I. Meisel, I. Mita, G. Moad, S. Penczek, R. F. T. Stepto, *Pure Appl. Chem.* **2007**, *79*, 1801–1829.

[744] A. C. Pierre, G. M. Pajonk, *Chem. Rev.* **2002**, *102*, 4243–4266.
[745] I. J. Haug, K. I. Draget, O. Smidsrød, *Food Hydrocoll.* **2004**, *18*, 203–213.
[746] M. Ghebremedhin, S. Seiffert, T. A. Vilgis, *Curr. Res. Food Sci.* **2021**, *4*, 436–448.
[747] A. Keller, *Rep. Prog. Phys.* **1968**, *31*, 623–704.
[748] T.-C. Ho, C.-C. Chang, H.-P. Chan, T.-W. Chung, C.-W. Shu, K.-P. Chuang, T.-H. Duh, M.-H. Yang, Y.-C. Tyan, *Molecules* **2022**, *27*, 2902.
[749] S. J. Guerrero, A. Keller, *J. Macromol. Sci. Part B* **1981**, *20*, 167–184.
[750] S. P. Das, *Rev. Mod. Phys.* **2004**, *76*, 785–851.
[751] W. Götze, *Complex Dynamics of Glass-Forming Liquids: A Mode-Coupling Theory*, Oxford Univ. Press, Oxford, **2009**.
[752] F. Mallamace, G. Mensitieri, M. Salzano de Luna, P. Lanzafame, G. Papanikolaou, D. Mallamace, *Int. J. Mol. Sci.* **2022**, *23*, 5316.
[753] S. Z. Ren, C. M. Sorensen, *Phys. Rev. Lett.* **1993**, *70*, 1727–1730.
[754] M. Kroon, G. H. Wegdam, R. Sprik, *Phys. Rev. E* **1996**, *54*, 6541–6550.
[755] P. Segrè, V. Prasad, A. Schofield, D. Weitz, *Phys. Rev. Lett.* **2001**, *86*, 6042–6045.
[756] M. Shibayama, S. Ozeki, T. Norisuye, *Polymer* **2005**, *46*, 2381–2388.
[757] H. H. Winter, *Macromolecules* **2013**, *46*, 2425–2432.
[758] R. E. Graves, B. M. Argrow, *J. Thermophys. Heat Transf.* **1999**, *13*, 337–342.
[759] N. Hirai, H. Eyring, *J. Polym. Sci.* **1959**, *37*, 51–70.
[760] K.-Y. Huang, C. N. Kingsley, R. Sheil, C.-Y. Cheng, J. C. Bierma, K. W. Roskamp, D. Khago, R. W. Martin, S. Han, *J. Am. Chem. Soc.* **2016**, *138*, 5392–5402.
[761] N. Nandi, K. Bhattacharyya, B. Bagchi, *Chem. Rev.* **2000**, *100*, 2013–2046.
[762] R. Metzler, J. Klafter, *Phys. Rep.* **2000**, *339*, 1–77.
[763] J. A. Dix, A. S. Verkman, *Annu. Rev. Biophys.* **2008**, *37*, 247–263.
[764] H. Fujita, *Polymer Solutions*, Elsevier ; Distributors For The U.S. And Canada, Elsevier Science Pub, Amsterdam ; New York : New York, NY, U.S.A, **1990**.
[765] J. Szymański, A. Patkowski, A. Wilk, P. Garstecki, R. Holyst, *J. Phys. Chem. B* **2006**, *110*, 25593–25597.
[766] A. Wisniewska, K. Sozanski, T. Kalwarczyk, K. Kedra-Krolik, R. Holyst, *Macromolecules* **2017**, *50*, 4555–4561.
[767] N. Ziębacz, S. A. Wieczorek, T. Kalwarczyk, M. Fiałkowski, R. Hołyst, *Soft Matter* **2011**, *7*, 7181.
[768] T. Kalwarczyk, K. Sozanski, A. Ochab-Marcinek, J. Szymanski, M. Tabaka, S. Hou, R. Holyst, *Adv. Colloid Interface Sci.* **2015**, *223*, 55–63.
[769] A. Sekhar, M. P. Latham, P. Vallurupalli, L. E. Kay, *J. Phys. Chem. B* **2014**, *118*, 4546–4551.
[770] A. Ochab-Marcinek, R. Hołyst, *Soft Matter* **2011**, *7*, 7366.
[771] D. S. Banks, C. Tressler, R. D. Peters, F. Höfling, C. Fradin, *Soft Matter* **2016**, *12*, 4190–4203.
[772] P. Nath, R. Mangal, F. Kohle, S. Choudhury, S. Narayanan, U. Wiesner, L. A. Archer, *Langmuir* **2018**, *34*, 241–249.
[773] T. Skóra, F. Vaghefikia, J. Fitter, S. Kondrat, *J. Phys. Chem. B* **2020**, *124*, 7537–7543.
[774] A. Soranno, I. Koenig, M. B. Borgia, H. Hofmann, F. Zosel, D. Nettels, B. Schuler, *Proc. Natl. Acad. Sci.* **2014**, *111*, 4874–4879.
[775] H. Kang, P. A. Pincus, C. Hyeon, D. Thirumalai, *Phys. Rev. Lett.* **2015**, *114*, 068303.
[776] R. Balu, N. Wanasingha, J. P. Mata, A. Rekas, S. Barrett, G. Dumsday, A. W. Thornton, A. J. Hill, N. Roy Choudhury, N. K. Dutta, *Sci. Adv.* **2022**, *8*, eabq2202.
[777] F. Oosawa, S. Asakura, *J. Chem. Phys.* **1954**, *22*, 1255–1255.
[778] S. Asakura, F. Oosawa, *J. Polym. Sci.* **1958**, *33*, 183–192.

[779] A. Vrij, *Pure Appl. Chem.* **1976**, *48*, 471–483.
[780] H. N. W. Lekkerkerker, R. Tuinier, M. Vis, *Colloids and the Depletion Interaction*, Springer International Publishing, Cham, **2024**.
[781] N. Kozer, Y. Y. Kuttner, G. Haran, G. Schreiber, *Biophys. J.* **2007**, *92*, 2139–2149.
[782] F. Zosel, A. Soranno, K. J. Buholzer, D. Nettels, B. Schuler, *Proc. Natl. Acad. Sci.* **2020**, *117*, 13480–13489.
[783] A. M. Berezhkovskii, A. Szabo, *J. Phys. Chem. B* **2016**, *120*, 5998–6002.
[784] M. Roos, M. Ott, M. Hofmann, S. Link, E. Rössler, J. Balbach, A. Krushelnitsky, K. Saalwächter, *J. Am. Chem. Soc.* **2016**, *138*, 10365–10372.
[785] B. Schuler, A. Soranno, H. Hofmann, D. Nettels, *Annu. Rev. Biophys.* **2016**, *45*, 207–231.
[786] J. P. Brady, P. J. Farber, A. Sekhar, Y.-H. Lin, R. Huang, A. Bah, T. J. Nott, H. S. Chan, A. J. Baldwin, J. D. Forman-Kay, L. E. Kay, *Proc. Natl. Acad. Sci.* **2017**, *114*, E8194–E8203.
[787] M. J. Saxton, *Biophys. J.* **1994**, *66*, 394–401.
[788] P. A. Netz, T. Dorfmüller, *J. Chem. Phys.* **1997**, *107*, 9221–9233.
[789] S. K. Ghosh, A. G. Cherstvy, R. Metzler, *Phys. Chem. Chem. Phys.* **2015**, *17*, 1847–1858.
[790] N. Kamerlin, C. Elvingson, *J. Phys. Condens. Matter* **2016**, *28*, 475101.
[791] S. K. Lai, D. E. O'Hanlon, S. Harrold, S. T. Man, Y.-Y. Wang, R. Cone, J. Hanes, *Proc. Natl. Acad. Sci.* **2007**, *104*, 1482–1487.
[792] V. M. Gun'ko, I. N. Savina, S. V. Mikhalovsky, *Gels* **2017**, *3*, 37.
[793] Y. Wu, S. Joseph, N. R. Aluru, *J. Phys. Chem. B* **2009**, *113*, 3512–3520.
[794] C. Yan, P. L. Kramer, R. Yuan, M. D. Fayer, *J. Am. Chem. Soc.* **2018**, *140*, 9466–9477.
[795] S. A. Roget, Z. A. Piskulich, W. H. Thompson, M. D. Fayer, *J. Am. Chem. Soc.* **2021**, *143*, 14855–14868.
[796] S. Mondal, B. Bagchi, *J. Chem. Phys.* **2020**, *152*, 224707.
[797] J. Landers, L. Roeder, S. Salamon, A. M. Schmidt, H. Wende, *J. Phys. Chem. C* **2015**, *119*, 20642–20648.
[798] O. Lieleg, I. Vladescu, K. Ribbeck, *Biophys. J.* **2010**, *98*, 1782–1789.
[799] J. Labille, N. Fatin-Rouge, J. Buffle, *Langmuir* **2007**, *23*, 2083–2090.
[800] A. Aharony, D. Stauffer, *Introduction To Percolation Theory*, Taylor & Francis, **2003**.
[801] N. L. Fawzi, J. Ying, R. Ghirlando, D. A. Torchia, G. M. Clore, *Nature* **2011**, *480*, 268–272.
[802] N. L. Fawzi, J. Ying, D. A. Torchia, G. M. Clore, *Nat. Protoc.* **2012**, *7*, 1523–1533.
[803] C. Jiménez-Saelices, B. Seantier, B. Cathala, Y. Grohens, *Carbohydr. Polym.* **2017**, *157*, 105–113.
[804] L. Manzocco, S. Plazzotta, J. Powell, A. De Vries, D. Rousseau, S. Calligaris, *Food Hydrocoll.* **2022**, *122*, 107117.
[805] K. Arakawa, K. Numata, *Mol. Cell* **2021**, *81*, 409–410.
[806] R. Pattanayek, G. Stubbs, *J. Mol. Biol.* **1992**, *228*, 516–528.
[807] M. Trinh, M. Odorico, L. Bellanger, M. Jacquemond, P. Parot, J. Pellequer, *J. Mol. Recognit.* **2011**, *24*, 503–510.
[808] Y. H. Tan, M. Liu, B. Nolting, J. G. Go, J. Gervay-Hague, G. Liu, *ACS Nano* **2008**, *2*, 2374–2384.
[809] R. C. Chaves, J. Teulon, M. Odorico, P. Parot, S. W. Chen, J. Pellequer, *J. Mol. Recognit.* **2013**, *26*, 596–604.
[810] B. Halle, V. P. Denisov, K. Venu, in *Biol. Magn. Reson.* (Eds.: N.R. Krishna, L.J. Berliner), Kluwer Academic Publishers, Boston, **2002**, pp. 419–484.
[811] E. Persson, B. Halle, *Proc. Natl. Acad. Sci.* **2008**, *105*, 6266–6271.

[812] R. G. Bryant, in *New Dev. NMR* (Ed.: R. Kimmich), Royal Society Of Chemistry, Cambridge, **2018**, pp. 207–228.
[813] W. Liao, Y. Zhang, Y. Guan, X. X. Zhu, *Langmuir* **2012**, *28*, 10873–10880.
[814] E. M. Saffer, M. A. Lackey, D. M. Griffin, S. Kishore, G. N. Tew, S. R. Bhatia, *Soft Matter* **2014**, *10*, 1905.
[815] K.-C. Hung, U.-S. Jeng, S. Hsu, *ACS Macro Lett.* **2015**, *4*, 1056–1061.
[816] S. Guseva, S. Milles, M. R. Jensen, N. Salvi, J.-P. Kleman, D. Maurin, R. W. H. Ruigrok, M. Blackledge, *Sci. Adv.* **2020**, *6*, eaaz7095.
[817] N. Kozer, G. Schreiber, *J. Mol. Biol.* **2004**, *336*, 763–774.
[818] M. Djabourov, P. Papon, *Polymer* **1983**, *24*, 537–542.
[819] C. Rochas, M. Rinaudo, *Biopolymers* **1984**, *23*, 735–745.
[820] M. R. Mangione, D. Giacomazza, D. Bulone, V. Martorana, P. L. S. Biagio, *Biophys. Chem.* **2003**.
[821] B. Jeong, Y. H. Bae, S. W. Kim, *Macromolecules* **1999**, *32*, 7064–7069.
[822] H. B. Bohidar, S. S. Jena, *J. Chem. Phys.* **1993**, *98*, 8970–8977.
[823] R. Linser, *Solid State Nucl. Magn. Reson.* **2017**, *87*, 45–53.
[824] W. Mahler, M. F. Bechtold, *Nature* **1980**, *285*, 27–28.
[825] S. Deville, *J. Mater. Res.* **2013**, *28*, 2202–2219.
[826] G. Nyström, W.-K. Fong, R. Mezzenga, *Biomacromolecules* **2017**, *18*, 2858–2865.
[827] C. A. Angell, *Science* **1995**, *267*, 1924–1935.
[828] N. R. Richbourg, N. A. Peppas, *Prog. Polym. Sci.* **2020**, *105*, 101243.
[829] K. Houben, D. Marion, N. Tarbouriech, R. W. H. Ruigrok, L. Blanchard, *J. Virol.* **2007**, *81*, 6807–6816.
[830] S. Milles, M. R. Jensen, G. Communie, D. Maurin, G. Schoehn, R. W. H. Ruigrok, M. Blackledge, *Angew. Chem. Int. Ed.* **2016**, *55*, 9356–9360.
[831] A. R. Camacho-Zarco, S. Kalayil, D. Maurin, N. Salvi, E. Delaforge, S. Milles, M. R. Jensen, D. J. Hart, S. Cusack, M. Blackledge, *Nat. Commun.* **2020**, *11*, 3656.
[832] J. L. O. Roldan, M. Blackledge, N. A. J. van Nuland, A. I. Azuaga, *J. Biomol. NMR* **2011**, *50*, 103–117.
[833] A. Cisse, J. Peters, G. Lazzara, L. Chiappisi, *J. Therm. Anal. Calorim.* **2021**, *145*, 403–409.
[834] leonardo-chiappisi, "PyDSC," can be found under https://github.com/leonardo-chiappisi/pyDSC, **n.d.**
[835] E. O. Stejskal, J. E. Tanner, *J. Chem. Phys.* **1965**, *42*, 288–292.
[836] C. A. Waudby, M. D. Mantle, L. D. Cabrita, L. F. Gladden, C. M. Dobson, J. Christodoulou, *J. Am. Chem. Soc.* **2012**, *134*, 11312–11315.

Annex

Model free analysis reproduction of the experimental rates

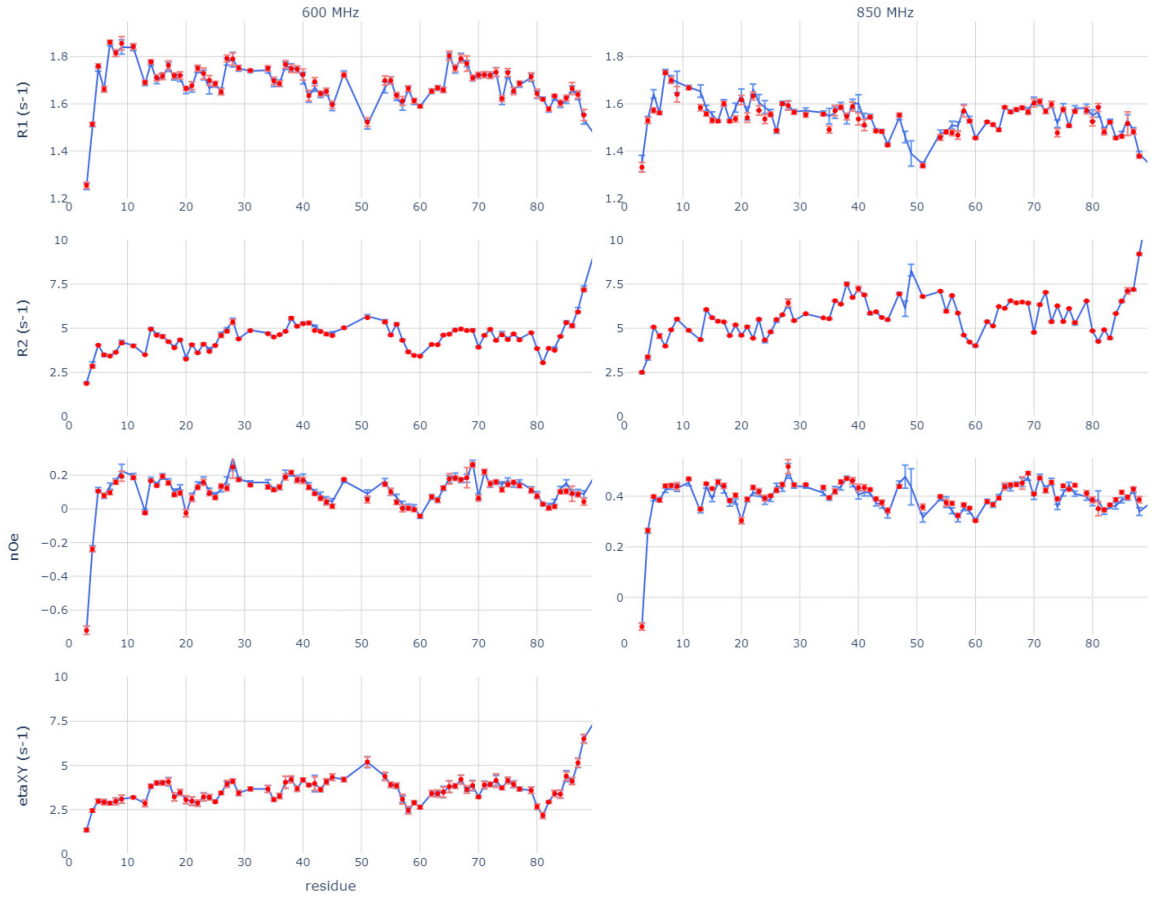
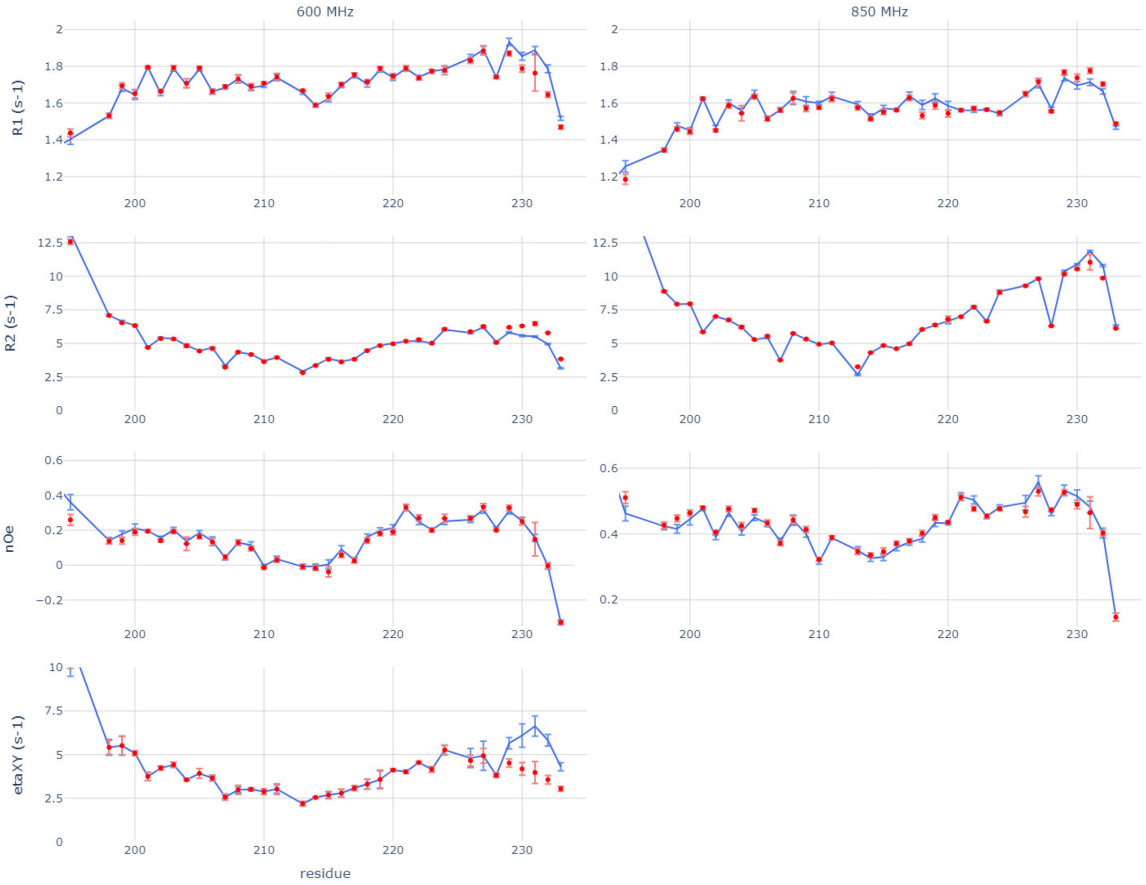


Figure S1: Reproduction of the experimental spin relaxation rates calculated from the model free analysis parameters for the N-terminal disordered chain of CAHS-8 at 278 K.

The experimental rates are in blue and the calculated rates are in red. The spectral density function was modelled using three dynamic modes.

Figure S2: Reproduction of the experimental spin relaxation rates calculated from the model free analysis parameters for the C-terminal disordered chain of CAHS-8 at 278 K.

The experimental rates are in blue and the calculated rates are in red. The spectral density function was modelled using three dynamic modes.


600 MHz
850 MHz
R1 (s-1)
R2 (s-1)
nOe
etaXY (s-1)
residue

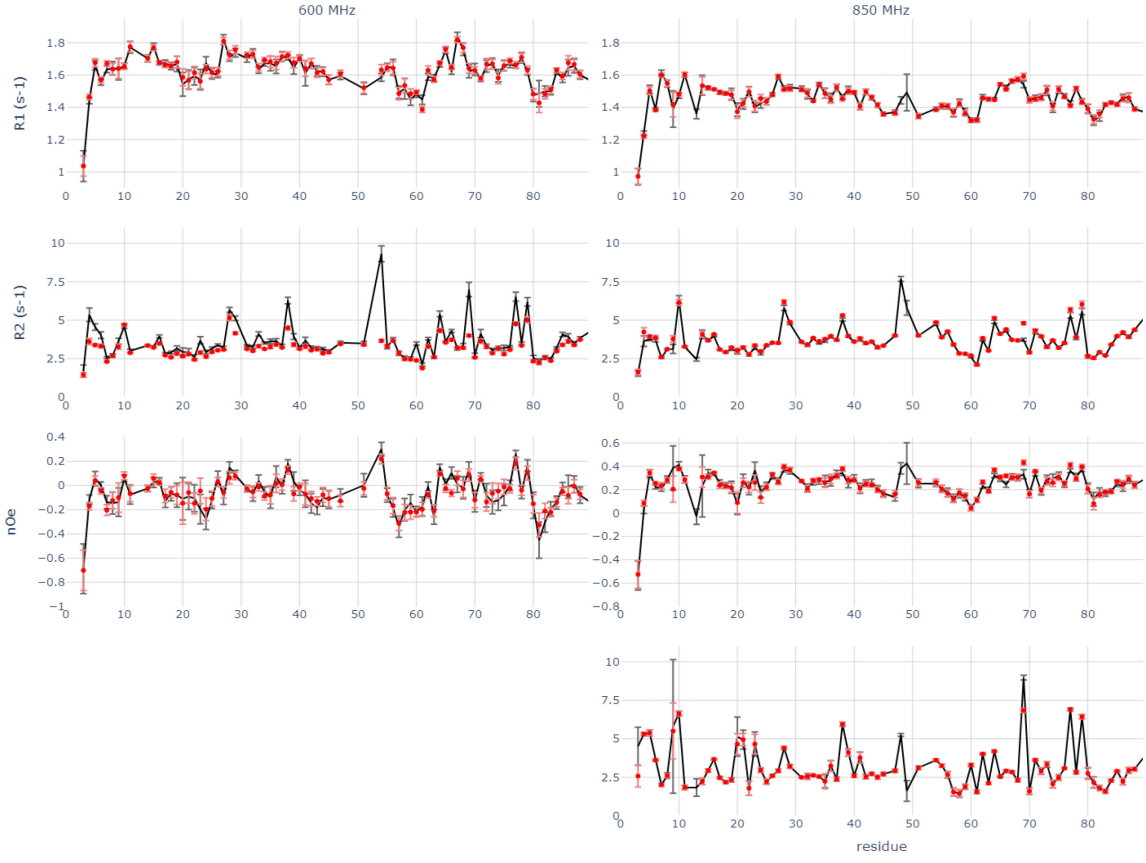

600 MHz
850 MHz
R1 (s-1)
R2 (s-1)
nOe
residue

Figure S3: Reproduction of the experimental spin relaxation rates calculated from the model free analysis parameters for the N-terminal disordered chain of CAHS-8 at 293 K.

The experimental rates are in black and the calculated rates are in red. The spectral density function was modelled using three dynamic modes.

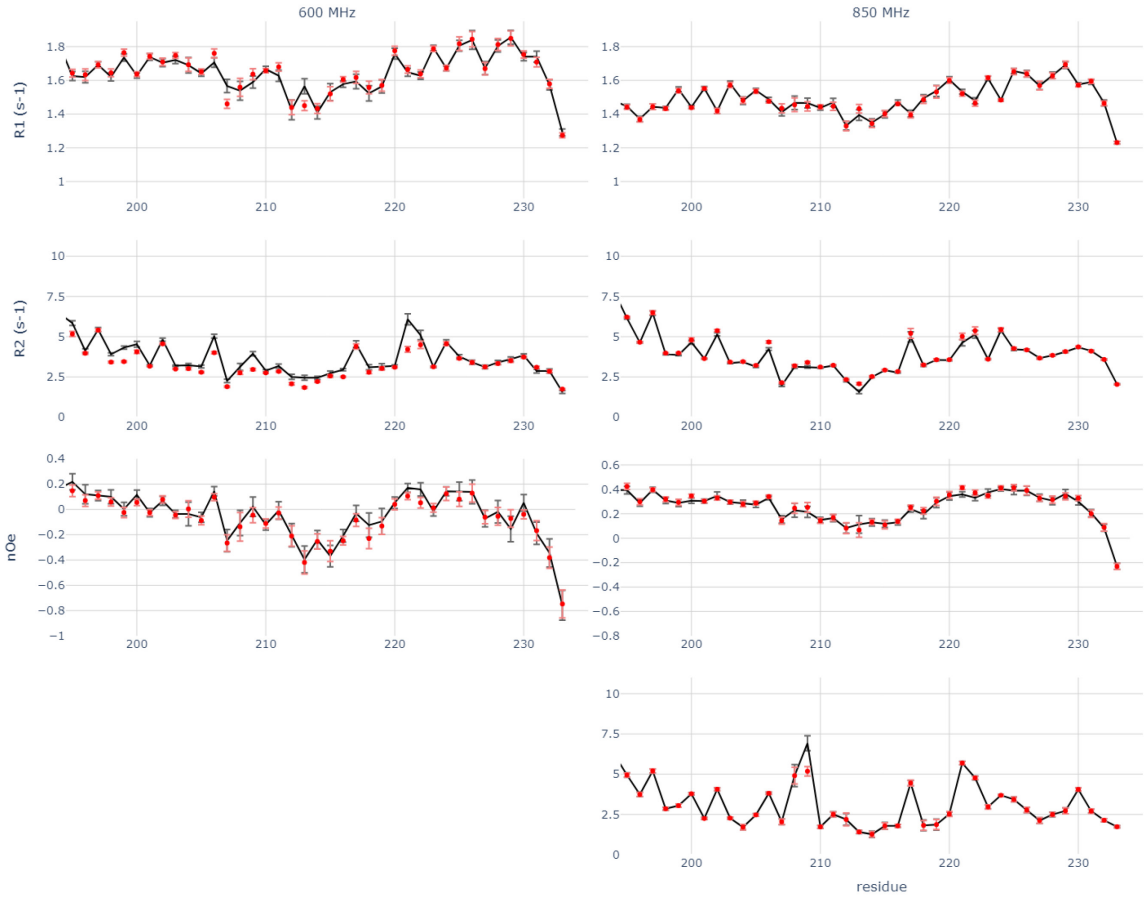


Figure S4: Reproduction of the experimental spin relaxation rates calculated from the model free analysis parameters for the helical domain of CAHS-8 at 293 K.

The experimental rates are in black and the calculated rates are in red. The spectral density function was modelled using two dynamic modes for residues 95-150, corresponding to helix 1, and three dynamic modes were used for residues 151-192, corresponding to the linker and helix 2.

Figure S5: Reproduction of the experimental spin relaxation rates calculated from the model free analysis parameters for the C-terminal disordered chain of CAHS-8 at 293 K.

The experimental rates are in black and the calculated rates are in red. The spectral density function was modelled using three dynamic modes.


600 MHz
850 MHz
R1 (s-1)
R2 (s-1)
nOe
residue

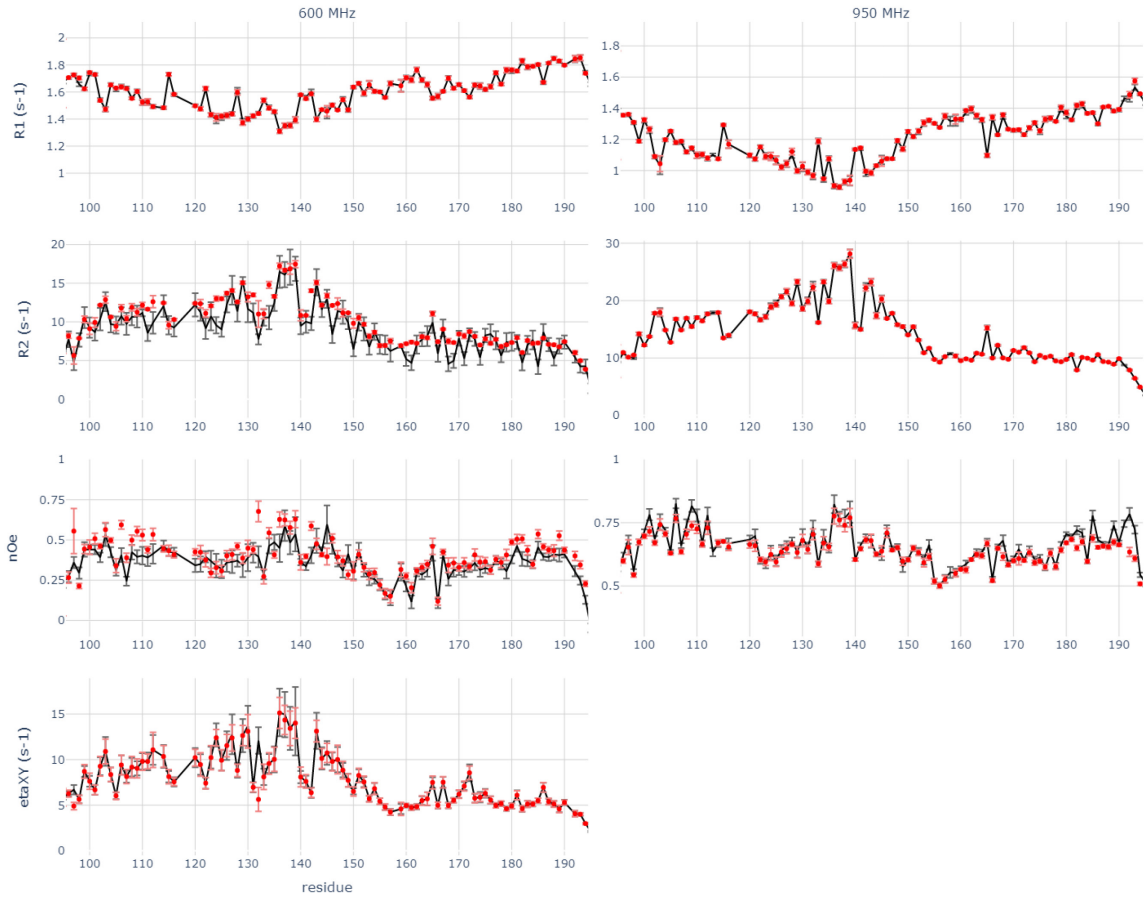

600 MHz
950 MHz
R1 (s-1)
R2 (s-1)
nOe
etaXY (s-1)
residue

Figure S6: Reproduction of the experimental spin relaxation rates calculated from the model free analysis parameters for the C-terminal disordered chain of CAHS-8 at 293 K.

The experimental rates are in black and the calculated rates are in red. The spectral density function was modelled using three dynamic modes.

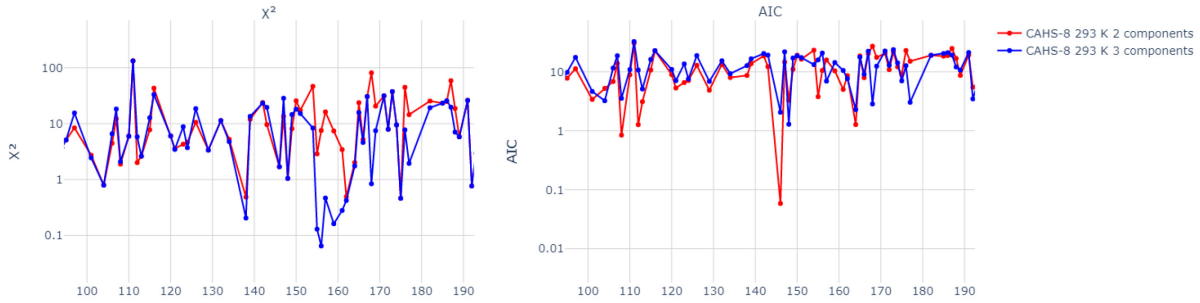


Figure S7: χ^2 and Aikaike information criterion statistics of the model free analysis fits of the relaxation rates of the helical domain of CAHS-8 at 293 K.

The χ^2 are similar for the two and three components models for helix 1 and therefore the Aikaike information criterion (AIC) is better for the former. The three dynamic model clearly outperforms the other in the linker and helix 2. This behaviour was expected considering that helix 1 is more populated than helix 2 which is still significantly disordered, as is the linker, hence the need for a more complex shape of the spectral density function to model them.

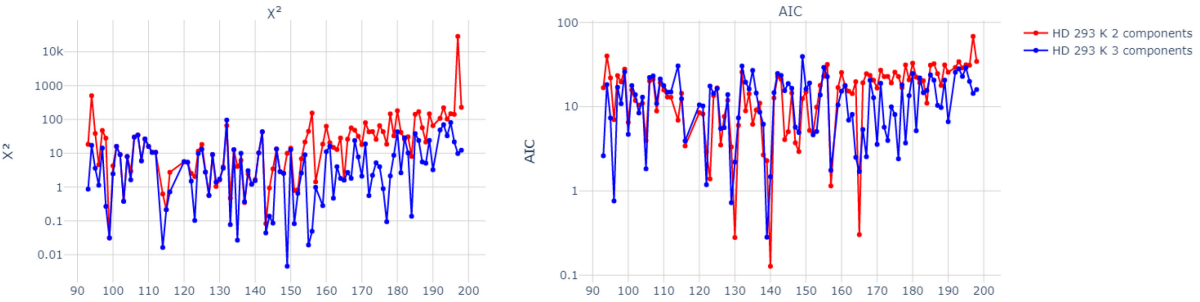


Figure S8: χ^2 and Aikaike information criterion statistics of the model free analysis fits of the relaxation rates of HD at 293 K.

The same results and effects are observed than in CAHS-8, with an even clearer difference in the helix 2.

These model free analyses were performed using a python program written for this and named NMRFit, which also performs fits of anisotropic diffusion tensor and spectral density mapping. It will eventually be available at https://github.com/anasmalki/NMRFit