

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ
Institut de génétique et de biologie moléculaire et cellulaire UMR 7104

THÈSE présentée par : **Abir Ben Bouzayene**

soutenue le : **27 Juin 2023**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : **Chimie**

ÉTUDE PAR RMN DES EFFETS DE L'INSERTION D'UN ATOME DE FLUOR SUR LES PROPRIÉTÉS DES PROLINES AU SEIN DES PROTÉINES

THÈSE dirigée par :

M. KIEFFER Bruno
M. TORBEEV Vladimir

Professeur, université de Strasbourg
Professeur, université de Strasbourg

RAPPORTEURS :

M. MICLET Emeric
Mme. SIZUN Christina

Maître de conférence, université Paris Sorbonne
Directrice de recherche, CNRS, université Paris Saclay

AUTRES MEMBRES DU JURY :

Mme. LANDRIEU Isabelle
Mme. ZANIER Katia

Directrice de recherche, CNRS, université de Lille
Chargée de recherche, CNRS, université de Strasbourg

INVITÉS :

M. SINNAEVE Davy Chargé de recherche, CNRS, université de Lille
Mme. DESAGE El-Murr Marine Professeur, université de Strasbourg

Acknowledgement

I warmly thank the jury members, who accepted to evaluate the scientific relevance of my thesis.

Major thanks to both my supervisors Prof. Bruno Kieffer and Prof. Vladimir Torbeev for believing in me and giving me the chance to work with them. I truly appreciate the confidence you showed in me and the assistance you have provided me with during this journey. Bruno, I very much appreciate your wise guidance, support and enriching ideas. Thank you for always taking the time to answer my questions and discuss with me. Words cannot express my gratitude for your invaluable patience and advices. Thank you for always pushing me beyond what I thought being my limits, and thank you for introducing me to the NMR community in many occasions. Valdimir, I am deeply grateful for the generous knowledge and expertise that you provided me with. I thank you for constantly reviewing my progress and guiding me through my PhD studies. I warmly thank you for the motivational speech that you gave me when needed. Thank you both for being as excited as I was with all my projects.

I would like to extend my sincere thanks to my team members Marc-André Delsuc, Célia Deville, Claude ling, Jocelyn Ceraline, Eva Erdmann, Mariam Sai, Philippe Dumas, Laura Duciel and Anne Briot-Dietsh. I am so glad to be part of this multidisciplinary team, I learned something from each and every one of you. Thank you all for your kindness. Marc-André Delsuc, thank you for your help and the numerous fruitful scientific discussion that we had. I am really lucky to be in the same lab as you. A special thank you to Claude ling with whom I shared the office for the last 4 years. I warmly thank you for patiently assisting me in front of the spectrometer when I started. I really appreciate the endless time you spent with me ameliorating my Python scripts. Thank you for your constant willingness to help in my professional and personal life. I will always cherish our moments and discussions. I also thank the student Fatiha Jouf Grondin who very much helped me during her internships in the team.

I would like to express my gratitude to Yves Nominé for his help with CD experiments, and especially for his valuable feedback and advices regarding my teaching experience. I also thank Davy Sinnaeve for the fruitful scientific discussions around NMR and the numerous

explanations he provided. I also thank Youssef El Khoury for the FTIR measurements and Nadia Messadeq for the TEM imaging.

I genuinely thank Roland Stote and Annick Dejaegere for their extreme kindness. Roland, I sincerely appreciate your constant willingness to help me. Thank you for your time, for the cookies and for inviting me to your nice garden parties.

I am also thankful to all the protein chemistry lab members with whom I had the opportunity to discuss on several occasions. I especially thank Mariam for her big efforts to synthesize my peptides, and Alexis for his help with ITC measurements. I wish you all good luck with your exciting projects.

I would like to thank my sister Marwa and my brother Fahmi. You've always been my source of happiness and my biggest supporters. Thank you for believing in me, I would have never come this far without you by my side. As long as I have you, I have everything. Mom and Dad, thank you for your endless and unconditional love. The proud look on your faces pushed me way beyond my limits.

My heartfelt thanks go to my friends Fadwa, Ana, Safouen and Yosra. Thank you for always being with me, for listening to me and for boosting my confidence. I also thank Aurora, Delphine, Victor, Mohamed, Gurudat, Masha and Louis for all the joyful moments we shared. I wish you good luck.

Last but not least, I would like to thank my life partner Iadh. No words are enough to appreciate what you do for me. Without your immense understanding and reassurance, it would be impossible for me to overcome the challenges. Thank you for always being there for me and for going through all hardships along with me.



Table of contents

General introduction.....	10
Proline amino acid	11
Research objectives	13
Thesis organization	13
Bibliography	14
I. General concepts of the relaxation	19
1. Introduction	19
2. Mechanisms of the relaxation.....	20
- Dipolar relaxation:	21
- Chemical shift anisotropy (CSA):	21
- Spin rotation mechanism:	21
- Quadrupolar relaxation mechanism:.....	21
- Scalar relaxation mechanism:.....	21
3. Phenomenological Bloch equations	22
4. Longitudinal relaxation rate explained by transitions between the energy levels of a single spin 24	
4.1- The case of a single spin $\frac{1}{2}$	24
4.2- Longitudinal relaxation in a two spins case: the Solomon equations.....	26
4.3- Expression of transition rates as a function of random molecular motions.....	29
4.3-1- Correlation and Spectral density functions.....	29
4.3-2- Relaxation through dipolar interaction	32
4.3-3- Relaxation through chemical shift anisotropy	35
4.3-3-1- Chemical shift tensor.....	35
4.3-3-2- Haeberlen notation.....	36
4.3-3-3- Maryland notation	37
4.3-3-4- Switching between Haeberlen and Maryland notation.....	38
4.3-3-5- Contribution of the CSA to the longitudinal relaxation rate	38
5. Transverse relaxation rate	38
5.1. Relaxation through dipolar interaction.....	39
5.2. Relaxation through chemical shift anisotropy	39
5.3. Chemical exchange	39
5.3-1 Bloch-McConnell equation.....	41
5.3-2 Dependance of the exchange rate constant on the strength of the magnetic field	43
6. Spectral densities and diffusion models	44
6.1- Rigid molecule with isotropic diffusion.....	44
6.2- Prolate Ellipsoid molecule:	45
6.3- Model free approach	45
Bibliography	47
II. Relaxation of fluorine nucleus	50
1. Fluorine relaxation properties to study biological molecules.....	50
2. Fluorine in 4-fluorinated prolines.....	53
3. Relaxation of fluorine in 4-fluorine substituted prolines	55
3.1. Fluorine longitudinal relaxation rate	56
3.2. Fluorine transverse relaxation rate:	61
3.2.1. ^1H - ^{19}F cross-relaxation rate constant.....	63
3.2.2. Heteronuclear ^1H - ^{19}F NOE.....	64
3.2.2.1. Steady state NOE.....	65
3.2.2.2. Transient NOE.....	67

4. Conclusion	69
Appendix 1.....	70
Bibliography	71
III. Dynamics of 4-fluoroprolines and their usage to evaluate overall peptide's motion.	76
Introduction:.....	76
1. Relaxation properties of 4-fluorinated prolines.....	77
1.1. Longitudinal relaxation rates of 4R and 4S fluoroprolines.....	78
1.2. Measuring Heteronuclear ^1H - ^{19}F NOE within 4S fluoroproline.....	83
1.3. Measurement of transverse relaxation rate in 4S fluoroproline	84
2. ^{19}F relaxation measurement of fluoroprolines within proline-rich model peptide	85
2.1. Heteronuclear ^1H - ^{19}F NOE in MpSR model peptide	85
2.2. Fluorine transverse relaxation rate in MpSR model peptide	86
3. Dynamics of a peptide encompassing a phosphorylated serine	87
4. Conclusions	89
5. Material and methods	90
Appendix 1.....	92
Appendix 2.....	93
Bibliography	94
IV. Using ^1H-^{19}F cross relaxation rates within fluoroprolines to characterize the rotational diffusion tensor in PPII helices	96
1. Proof of concept	97
2. Optimization of 2D ^1H - ^{19}F HOESY sequence	99
3. Measurements of ^1H - ^{19}F NOE build-up in PPII helix	101
3.1- ^1H - ^{19}F heteronuclear NOE build-up curves of MpSR model peptide	103
3.2- ^1H - ^{19}F heteronuclear NOE build-up curves of MpSRSer _{phospho} model peptide	104
3.3- ^1H - ^{19}F heteronuclear NOE build-up curves of MpRRSer and MpRRAla model peptides.....	105
4. Conclusion and future work.....	107
5. Experimental section	108
Bibliography	108
V. Fluorine NMR study of proline-rich sequences using fluoroprolines.....	111
1. Introduction	112
2. Results.....	116
2.1. Assignment and spectral analysis of model peptides.....	116
2.2. ^{19}F relaxation and dynamics.....	121
2.3. Impact of proline modifications on the binding of SH3	125
3. Discussion.....	130
4. Material and methods	133
5. Supplementary Material	152
VI. Fine-tuning molten-globule conformational dynamics by selective incorporation of fluoroprolines	158
1. Introduction	158

2. Results.....	160
2.1. Insertion of conformationally biased fluoroproline into NCBD protein domain.....	160
2.2. NCBD secondary structure is not altered by fluoroproline insertion	162
2.3. Fluoroproline insertion alters NCBD hydrophobic packing.....	162
2.4. Fluoroproline insertion impacts global and local dynamics.....	168
2.5. Insertion of 4-(<i>R</i>)-FPro in the NCBD α 2- α 3 loop induces self-assembly into amyloid fibers	170
3. Discussion.....	172
4. Conclusions	174
Experimental section	175
Bibliography	179
Supplementary materials.....	182
<i>VII. NMR study of binding of fluorinated NCBD variants to transactivation domain (TAD) of p53 tumor suppressor protein and activation domain of ACTR transcriptional co-activator.....</i>	<i>201</i>
1. Introduction	201
2. NCBD/TAD-p53 interaction	203
NCBD/TAD-p53 complex formation followed from TAD-p53 side	207
3. NCBD/AD-ACTR interaction	210
4. Conclusions	213
5. Experimental section	214
Bibliography	216
<i>General conclusions</i>	<i>218</i>
<i>Résumé de thèse.....</i>	<i>221</i>

Abbreviation list

NMR: Nuclear magnetic resonance

HOESY: Heteronuclear Overhauser enhancement spectroscopy HSQC Heteronuclear single quantum correlation

TOCSY: Total correlation spectroscopy

NOESY: Nuclear Overhauser enhancement spectroscopy

HSQC: Heteronuclear single quantum correlation

PPI: Polyproline-I

PPII: Polyproline-II

FPRO: fluoroproline

CSA: chemical shift anisotropy

DD: dipole-dipole

ppm: parts per million

DFT: Density functional theory

IDP: Intrinsically disordered protein

MG: molten globule

NCBD: Nuclear coactivator binding domain

CBP: CREB-binding protein

ACTR: Activator for thyroid hormone and retinoid receptors

p53: tumor suppressor protein 53

TAD p53: N-terminal transactivation domain of p53

ThT: Thioflavin t

CD: Circular dichroism

TEM: transmission electron microscopy

FT-IR: Fourier transform Infrared

SPPS: solid peptide phase synthesis

Fmoc: 9-fluorenylmethyloxycarbonyl

D₂O: deuterium oxide

°C: Degree Celsius

K: kelvin

s: second

Å: Angstrom

General introduction

General introduction

My PhD was performed under the supervision of professor Bruno Kieffer head of the NMR team at the IGBMC and professor Vladimir Torbeev head of the protein chemistry lab at the ESBS.

When I started this journey, I had very basic knowledge about NMR techniques. Every day and to this day my fascination about it grows bigger. I can only be proud to be part of the large NMR community that is open on too many interesting fields of research. My admiration to NMR is only rivaled by protein chemistry. Unfortunately, I didn't have the time to acquire deep knowledge of protein synthesis techniques, in fact, the four years went by so fast that I couldn't discover everything that I wanted to.

During my PhD I used fluorine NMR as a tool to study chemically synthesized peptides. The precious fluorine spin is very attractive because it is 83% as sensitive as the proton, has a very large chemical shift range (400 ppm) and is absent in biological molecules. In fact, Nature rarely involves fluorine in its products⁴⁴. In 1943, fluoroacetate was discovered in a toxic south African plant *Dichapetalum cymosum* that accumulates fluoroacetate in its leaves. Later, it was identified in different plants organs in Africa, Australia, Brazil and south America. The toxicity of the plants is a result of the modification of fluoroacetate into (2*R*,3*R*)-fluorocitrate. Other less common natural fluorine products were discovered and are the 4-Fluoro-L-threonine, the ω fluorooleic acid, the nucleocidin, the fluoroacetone and the 5-Fluorouracils, even though the last two products are most probably misidentified.^{1,2}

The analysis of structure and motion of biomolecules are usually assessed through ^1H , ^{15}N , or ^{13}C NMR spectroscopy. This is because these spins are the most abundant atoms in proteins. The main inconvenient of these studies is their need to use labelled proteins. Even though protein labelling became a routine task, it may still be challenging for engineered molecules. Additionally, the acquisition of NMR spectra of a labelled protein is very demanding in terms of product amount and time. Moreover, it is not always easy to detect the information of interest in a crowded spectrum. On the other side, working with fluorinated systems is advantageous because it requires the analysis of simple NMR spectra rapidly acquired even at low concentrations. Even though the insertion of a fluorine atom opens the window on numerous NMR experiments, a ^1D ^{19}F NMR spectrum is often very informative and no additional 2D experiments are needed. For this purpose, fluorinated amino acids and moieties became very popular and are commercially available. Some amino acids were excessively used

such as fluorinated tryptophan, phenylalanine and tyrosine. In this work, we are interested in prolines amino acid.

Proline amino acid

Proline is a particular amino acid with a cyclic structure. Its side chain bonds to the nitrogen backbone atom to create a five membered cycle. Inside this cycle, the possible combinations of the dihedral angles φ and ψ is limited compared to all the other amino acids (except for glycine) therefore the pyrrolidine ring is considered to be rigid.³ The absence of the hydrogen atom on its nitrogen results in the formation of a secondary amine that later creates tertiary amides. This particularity is incompatible with frequent secondary structures alpha helices and beta sheets where proline residue settles in the N-terminus or in the edge respectively.⁴

The pyrrolidine ring is not planar and is said to be puckered. Despite its rigidity, the five membered cycle is animated by a very rapid movement at the picoseconds time scale resulting from its interconversion between two preferable conformations c_γ *endo* and c_γ *exo*.^{5,6} The energies of c_γ *endo* and c_γ *exo* are very close and only differ by 0.5 kcal/mol.⁷ The differentiation between the two puckers is based on the direction of the c_γ by respect to the carbon of the carbonyl. When they point in the same direction, we identify the c_γ *endo* pucker, when they point in different directions, we recognize the c_γ *exo* pucker. Both dihedral angles φ and ψ have particular values inside each conformer. φ angle is typically around -60° and -75° within c_γ *exo* and c_γ *endo* respectively, whereas the ψ angle is about 140° and 170° when in c_γ *exo* and c_γ *endo* respectively.⁸

Inside proteins, the double bond character of the peptide bond leads to the formation of two isomers *cis* and *trans*. It is the third dihedral angle ω that directs this equilibrium. A 180° is a typical value for *trans* peptide bond, while 0° is typical for a *cis* peptide bond⁹. The *cis* isomer is less favored due to the steric repulsion between the C_α of the proline and the C_α of the previous amino acid.¹⁰⁻¹³ The energy barrier separating the two conformers is 20 kcal/mol.¹²⁻¹⁴ and the population of the *cis* isomer is estimated to be 0.1% or 1.5%.^{11,15} When proline amino acid is involved, the activation energy barrier is lowered to 13 kcal/mol^{16,17} consequently, the *cis* population is enhanced.¹¹ The early experimental studies on different model peptides containing prolines recorded the presence of 10 to 40% *cis* isomer, either by ^{13}C NMR spectroscopy or by X-ray cristallography.^{15,18-20}

The *cis/trans* isomerization is a slow event that occurs at 10^{-3} to 10^{-2} s^{-1} at room temperature, and that can be accelerated by peptidyl-prolyl *cis/trans* isomerases.^{21–24} The slow *cis/trans* isomerization of proline usually imposes the kinetics of the rate limiting steps within protein folding.^{25,26} Additionally, proline isomerization is a key phenomenon of several biological processes such the self-assembly of polypeptides into amyloid fibers^{27,28} the modulation of protein ligand affinities²⁹ and the control of the autoinhibition of protein signaling.³⁰

The *cis/trans* isomerization and the *endo/exo* ring puckering are two correlated events.⁶ The carbonyl groups in two consecutive peptide bonds witness an $n - \pi^*$ interaction between the lone pair of the oxygen of the first carbonyl and the non-bonding π^* orbital of the second. This interaction is very sensitive towards the angle separating the carbonyl groups and the distance $\text{O}_i - \text{C}_{i+1}$ and is accompanied by a release of a stabilizing energy.³¹ The proximity of the carbonyl in the *trans* conformer allows this weak interaction to occur. When proline is involved in a peptide bond, the orbitals overlap highly depends on the ring pucker, it is the c_γ *exo* that strengthens the interaction by favorable carbonyl closeness. On the other hand, when in a *cis* conformer, it is the *endo* pucker that is favored.^{32,33}

Proline can form different secondary structures from α helices and β sheets, named polyproline helices.³⁴ Consecutive proline residues connected by *cis* peptide bonds form a tightly packed right-handed helix known as PPI. They are formed by 3.3 residues/turn and have a 5.6 Å/turn helical pitch.³² When interconnected by *trans* peptide bonds, polyprolines form an expanded left-handed helix denoted PPII with 3 residues/turn and a pitch of 9.3 Å/turn. PPII helices are stabilized by the weak nonbonding interaction $n - \pi^*$ between the successive carbonyls. Since no intra-hydrogen bonds are established within polyproline helices, PPI and PPII are usually described in terms of their dihedral angles. PPI has $\varphi = -75^\circ$, $\psi = -160^\circ$ and $\omega = 0^\circ$, while PPII has $\varphi = -75^\circ$, $\psi = 145^\circ$ and $\omega = 180^\circ$. Interestingly, PPII helices are solvable in aqueous solutions due to their exposed carbonyl groups capable of forming hydrogen bonds with the solvent. Within PPI helices, the carbonyl groups are not exposed to the solvent and are favored in organic solvents. Consequently, only PPII helices were detected in biological contexts.^{32,35} PPII was first identified as one of the collagen's triple-helix structures, but was recognized later as a naturally occurring structure within polypeptides and proteins that is not necessarily dominated by prolines. PPII is less abundant than α helices and β sheets but is comparable to a 3(10)-helix. Due to its extended nature that is ideal for structural stability, PPII helices are frequently encountered in interaction interfaces and domain linker regions.³⁶ It is also a very known SH3 domain binder, therefore it is crucial for signaling pathways and was

associated to amyloids formation in Parkinson's disease.³⁷ Although the discovery of PPII helices goes back to 1950³⁷, its crystallographic structure was only recently published in 2014.³⁸ Proline is subject to posttranslational modifications (PTM). One common PTM is the hydroxylation of proline by prolyl-4 hydroxylases to form hydroxyproline that stabilizes the collagen.³⁹⁻⁴¹ Proline hydroxylation is mandatory for the folding of the collagen's polypeptide chain and is crucial for its triple-helix thermal stability. The lack of prolyl hydroxylase activity in collagen directly affects its interaction with its receptor integrin.^{42,43} The substitution at the position 4 of Proline led to a new variety of synthetic prolines with various properties

Research objectives

Fluorination of prolines shifts the conformational equilibria of proline residues. 4-fluorinated prolines are prolines with a substituted H_γ by a fluorine spin. *4R* fluoroproline favors the c_γ *exo* ring pucker and primarily populates the *trans* conformer. *4S* fluoroproline mainly adopts the c_γ *endo* pucker and enhances the *cis* conformer population. During my PhD, we intended to study the effect of conformationally biased fluoroprolines on proteins properties. To begin with, we intend to demonstrate that fluoroprolines are reliable fluorinated probes to reveal molecular motion, structural properties and binding events at the atomic level. For this purpose, 4-fluorinated prolines were incorporated in the context of polyproline rich sequences where fluorine signal properties namely its chemical shift and linewidth as well as its relaxation parameters were observed using fluorine NMR. Furthermore, we aim to investigate the functional relationship between proline's properties and protein function. In this regard, two proline residues were substituted by both *4R* and *4S* Fluoroprolines in the molten globule NCBD of CBP to create new variants of NCBD.

Thesis organization

Peptide dynamics were evaluated through NMR relaxation. The effect of the motion on the relaxation parameters is widely described in the literature. In the first chapter, I illustrated my understanding of the phenomenon with a focus on longitudinal relaxation. A detailed study of the transverse relaxation is beyond the scope of this thesis due to the participation of the structural and chemical exchange, although preliminary results are given.

The effect of the motion on fluorine relaxation parameters in fluoroprolines is presented in chapter two. Various simulations are given to illustrate the evolution of these parameters

with the molecular correlation time, under the effect of ^{19}F CSA. This chapter also illustrates the evolution of my python programming skills. With no prerequisite skills or experience with Python, I took it upon myself to learn this programming language. Eventually, I have accumulated enough knowledge to qualify for a full-time ATER position teaching Python at Strasbourg university.

A quantitative interpretation of the experimentally measured fluorine relaxation parameters in free 4-fluorinated prolines and in model peptides in terms of molecular motion is illustrated in chapter three and shows the potential of fluoroprolines to report overall motion of biological molecules.

In Chapter four we explore the possibility to get reliable relaxation data to orient molecular diffusion tensors. Preliminary results using ^1H - ^{19}F HOESY experiments on model peptides of proline homopolymers proving the potential of the approach are given.

Chapter five is an experimental study of the dynamics and the interactions of polyproline model peptides designed to recognize the SH3.3 domain of the vinexine β . Two fluorinated prolines were incorporated at two different positions in the peptide to report the behavior of each polyproline segment. This work was published in magnetic resonance journal of the Ampere group.

In chapter six we performed a study of the molten globule NCBD of the CBP protein. two proline residues were substituted by *4R* and *4S* fluoroprolines to reveal their effect on the dynamics and the structure of NCBD. This work will be very soon submitted to be published this summer.

Finally, chapter seven is a preliminary NMR study of NCBD's interactions with two of its important binding partners TADp53 and AD-ACTRp160.

Bibliography

- (1) Chan, K. K. J.; O'Hagan, D. The Rare Fluorinated Natural Products and Biotechnological Prospects for Fluorine Enzymology. In *Methods in Enzymology*; Elsevier, 2012; Vol. 516, pp 219–235. <https://doi.org/10.1016/B978-0-12-394291-3.00003-4>.
- (2) Harper, D. B.; O'Hagan, D. The Fluorinated Natural Products. *Nat. Prod. Rep.* **1994**, *11*, 123. <https://doi.org/10.1039/np9941100123>
- (3) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; De Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by $\text{C}\alpha$ Geometry: ϕ, ψ and $\text{C}\beta$ Deviation. *Proteins Struct. Funct. Bioinforma.* **2003**, *50* (3), 437–450. <https://doi.org/10.1002/prot.10286>.
- (4) Chou, P. Y.; Fasman, G. D. Prediction of Protein Conformation. *Biochemistry* **1974**, *13* (2), 222–245. <https://doi.org/10.1021/bi00699a002>.

- (5) Ramachandran, G. N.; Lakshminarayanan, A. V.; Balasubramanian, R.; Tegoni, G. Studies on the Conformation of Amino Acids XII. Energy Calculations on Prolyl Residue. *Biochim. Biophys. Acta BBA - Protein Struct.* **1970**, *221* (2), 165–181. [https://doi.org/10.1016/0005-2795\(70\)90257-6](https://doi.org/10.1016/0005-2795(70)90257-6).
- (6) Costantini, N. V.; Ganguly, H. K.; Martin, M. I.; Wenzell, N. A.; Yap, G. P. A.; Zondlo, N. J. The Distinct Conformational Landscapes of 4 *S* -Substituted Prolines That Promote an *Endo* Ring Pucker. *Chem. – Eur. J.* **2019**, *25* (48), 11356–11364. <https://doi.org/10.1002/chem.201902382>.
- (7) Newberry, R. W.; Raines, R. T. 4-Fluoroproline: Conformational Analysis and Effects on the Stability and Folding of Peptides and Proteins. In *Peptidomimetics I*; Lubell, W. D., Ed.; Topics in Heterocyclic Chemistry; Springer International Publishing: Cham, 2016; Vol. 48, pp 1–25. https://doi.org/10.1007/7081_2015_196.
- (8) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **2002**, *124* (11), 2497–2505. <https://doi.org/10.1021/ja0166904>.
- (9) Pal, D.; Chakrabarti, P. Cis Peptide Bonds in Proteins: Residues Involved, Their Conformations, Interactions and Locations 1 Edited by J. M. Thornton. *J. Mol. Biol.* **1999**, *294* (1), 271–288. <https://doi.org/10.1006/jmbi.1999.3217>.
- (10) Perricaudet, M.; Pullman, A. AN AB INITIO QUANTUM-MECHANICAL INVESTIGATION ON THE ROTATIONAL ISOMERISM IN AMIDES AND ESTERS. *Int. J. Pept. Protein Res.* **2009**, *5* (2), 99–107. <https://doi.org/10.1111/j.1399-3011.1973.tb02324.x>.
- (11) Ramachandran, G. N.; Mitra, A. K. An Explanation for the Rare Occurrence of Cis Peptide Units in Proteins and Polypeptides. *J. Mol. Biol.* **1976**, *107* (1), 85–92. [https://doi.org/10.1016/S0022-2836\(76\)80019-8](https://doi.org/10.1016/S0022-2836(76)80019-8).
- (12) Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. In *Advances in Protein Chemistry*; Elsevier, 1968; Vol. 23, pp 283–437. [https://doi.org/10.1016/S0065-3233\(08\)60402-7](https://doi.org/10.1016/S0065-3233(08)60402-7).
- (13) LaPlanche, L. A.; Rogers, M. T. *Cis* and *Trans* Configurations of the Peptide Bond in N-Monosubstituted Amides by Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1964**, *86* (3), 337–341. <https://doi.org/10.1021/ja01057a007>.
- (14) Weiss, M. S.; Jabs, A.; Hilgenfeld, R. Peptide Bonds Revisited. *Nat. Struct. Mol. Biol.* **1998**, *5* (8), 676–676. <https://doi.org/10.1038/1368>.
- (15) Stewart, D. E.; Sarkar, A.; Wampler, J. E. Occurrence and Role Of cis Peptide Bonds in Protein Structures. *J. Mol. Biol.* **1990**, *214* (1), 253–260. [https://doi.org/10.1016/0022-2836\(90\)90159-J](https://doi.org/10.1016/0022-2836(90)90159-J).
- (16) Pal, D.; Chakrabarti, P. Cis Peptide Bonds in Proteins: Residues Involved, Their Conformations, Interactions and Locations 1 Edited by J. M. Thornton. *J. Mol. Biol.* **1999**, *294* (1), 271–288. <https://doi.org/10.1006/jmbi.1999.3217>.
- (17) Schulz, G. E.; Schirmer, R. H. *Principles of Protein Structure*; Springer advanced texts in chemistry; Springer-Verlag: New York Berlin Heidelberg [etc.], 1990.
- (18) Grathwohl, C.; Wüthrich, K. The X-Pro Peptide Bond as an Nmr Probe for Conformational Studies of Flexible Linear Peptides: X-PRO PEPTIDE BOND. *Biopolymers* **1976**, *15* (10), 2025–2041. <https://doi.org/10.1002/bip.1976.360151012>.
- (19) Thomas, W. A.; Williams, M. K. ¹³C Nuclear Magnetic Resonance Spectroscopy and Cis/Trans Isomerism in Dipeptides Containing Proline. *J. Chem. Soc. Chem. Commun.* **1972**, No. 17, 994. <https://doi.org/10.1039/c39720000994>.

- (20) Brandts, J. F.; Halvorson, H. R.; Brennan, M. Consideration of the Possibility That the Slow Step in Protein Denaturation Reactions Is Due to Cis-Trans Isomerism of Proline Residues. *Biochemistry* **1975**, *14* (22), 4953–4963. <https://doi.org/10.1021/bi00693a026>.
- (21) Grathwohl, C.; Wüthrich, K. Nmr Studies of the Rates of Proline *Cis* - *Trans* Isomerization in Oligopeptides: RATES OF *CIS* - *TRANS* ISOMERIZATION. *Biopolymers* **1981**, *20* (12), 2623–2633. <https://doi.org/10.1002/bip.1981.360201209>.
- (22) Alderson, T. R.; Lee, J. H.; Charlier, C.; Ying, J.; Bax, A. Propensity for *Cis* -Proline Formation in Unfolded Proteins. *ChemBioChem* **2018**, *19* (1), 37–42. <https://doi.org/10.1002/cbic.201700548>.
- (23) London et al. - 1990 - Probing the Role of Proline in Peptide Hormones.Pdf.
- (24) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. Side-Chain Effects on Peptidyl-Prolyl *Cis*/*Trans* Isomerisation. *J. Mol. Biol.* **1998**, *279* (2), 449–460. <https://doi.org/10.1006/jmbi.1998.1770>.
- (25) Zosel, F.; Mercadante, D.; Nettels, D.; Schuler, B. A Proline Switch Explains Kinetic Heterogeneity in a Coupled Folding and Binding Reaction. *Nat. Commun.* **2018**, *9* (1), 3332. <https://doi.org/10.1038/s41467-018-05725-0>.
- (26) Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. Proline *Cis*–*Trans* Isomerization and Protein Folding. *Biochemistry* **2002**, *41* (50), 14637–14644. <https://doi.org/10.1021/bi020574b>.
- (27) Torbeev, V. Yu.; Hilvert, D. Both the *Cis* - *Trans* Equilibrium and Isomerization Dynamics of a Single Proline Amide Modulate B2-Microglobulin Amyloid Assembly. *Proc. Natl. Acad. Sci.* **2013**, *110* (50), 20051–20056. <https://doi.org/10.1073/pnas.1310414110>.
- (28) Eakin, C. M.; Berman, A. J.; Miranker, A. D. A Native to Amyloidogenic Transition Regulated by a Backbone Trigger. *Nat. Struct. Mol. Biol.* **2006**, *13* (3), 202–208. <https://doi.org/10.1038/nsmb1068>.
- (29) Sinnaeve, D.; Ben Bouzayene, A.; Ottoy, E.; Hofman, G.-J.; Erdmann, E.; Linclau, B.; Kuprov, I.; Martins, J. C.; Torbeev, V.; Kieffer, B. Fluorine NMR Study of Proline-Rich Sequences Using Fluoroproline. *Magn. Reson.* **2021**, *2* (2), 795–813. <https://doi.org/10.5194/mr-2-795-2021>.
- (30) Sarkar, P.; Reichman, C.; Saleh, T.; Birge, R. B.; Kalodimos, C. G. Proline *Cis*-*Trans* Isomerization Controls Autoinhibition of a Signaling Protein. *Mol. Cell* **2007**, *25* (3), 413–426. <https://doi.org/10.1016/j.molcel.2007.01.004>.
- (31) Newberry, R. W.; Bartlett, G. J.; VanVeller, B.; Woolfson, D. N.; Raines, R. T. Signatures of $N \rightarrow \pi^*$ Interactions in Proteins: Signatures of $N \rightarrow \pi^*$ Interactions in Proteins. *Protein Sci.* **2014**, *23* (3), 284–288. <https://doi.org/10.1002/pro.2413>.
- (32) Horng, J.-C. Stereoelectronic Effects on Polyproline Conformation. *Protein Sci.* **2006**, *15* (1), 74–83. <https://doi.org/10.1110/ps.051779806>.
- (33) Kim, W.; Hardcastle, K. I.; Conticello, V. P. Fluoroproline Flip-Flop: Regiochemical Reversal of a Stereoelectronic Effect on Peptide and Protein Structures. *Angew. Chem. Int. Ed.* **2006**, *45* (48), 8141–8145. <https://doi.org/10.1002/anie.200603227>.
- (34) Steinberg, I. Z.; Harrington, W. F.; Berger, A.; Sela, M.; Katchalski, E. The Configurational Changes of Poly-L-Proline in Solution. *J. Am. Chem. Soc.* **1960**, *82* (20), 5263–5279. <https://doi.org/10.1021/ja01505a001>.
- (35) Mutter, M.; Wüthrich, T.; Gioria, S.; Keller, M. Pseudo-Proline: Induction Of *cis*/*Trans*-Conformational Interconversion by Decreased Transition State Barriers. *Biopolymers* **1999**, *51* (2), 121–128. [https://doi.org/10.1002/\(SICI\)1097-0282\(1999\)51:2<121::AID-BIP2>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0282(1999)51:2<121::AID-BIP2>3.0.CO;2-O)

- (36) Adzhubei, A. A.; Sternberg, M. J. E.; Makarov, A. A. Polyproline-II Helix in Proteins: Structure and Function. *J. Mol. Biol.* **2013**, *425* (12), 2100–2132. <https://doi.org/10.1016/j.jmb.2013.03.018>.
- (37) Narwani, T. J.; Santuz, H.; Shinada, N.; Melarkode Vattekatte, A.; Ghouzam, Y.; Srinivasan, N.; Gelly, J.-C.; de Brevern, A. G. Recent Advances on Polyproline II. *Amino Acids* **2017**, *49* (4), 705–713. <https://doi.org/10.1007/s00726-017-2385-6>.
- (38) Wilhelm, P.; Lewandowski, B.; Trapp, N.; Wennemers, H. A Crystal Structure of an Oligoproline PPII-Helix, at Last. *J. Am. Chem. Soc.* **2014**, *136* (45), 15829–15832. <https://doi.org/10.1021/ja507405j>.
- (39) Ganguly, H. K.; Basu, G. Conformational Landscape of Substituted Prolines. *Biophys. Rev.* **2020**, *12* (1), 25–39. <https://doi.org/10.1007/s12551-020-00621-8>.
- (40) Berg, R. A.; Prockop, D. J. The Thermal Transition of a Non-Hydroxylated Form of Collagen. Evidence for a Role for Hydroxyproline in Stabilizing the Triple-Helix of Collagen. *Biochem. Biophys. Res. Commun.* **1973**, *52* (1), 115–120. [https://doi.org/10.1016/0006-291X\(73\)90961-3](https://doi.org/10.1016/0006-291X(73)90961-3).
- (41) Gorres, K. L.; Raines, R. T. Prolyl 4-Hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **2010**, *45* (2), 106–124. <https://doi.org/10.3109/10409231003627991>.
- (42) Sipilä, K. H.; Drushinin, K.; Rappu, P.; Jokinen, J.; Salminen, T. A.; Salo, A. M.; Kämpylä, J.; Myllyharju, J.; Heino, J. Proline Hydroxylation in Collagen Supports Integrin Binding by Two Distinct Mechanisms. *J. Biol. Chem.* **2018**, *293* (20), 7645–7658. <https://doi.org/10.1074/jbc.RA118.002200>.
- (43) Zhou, T.; Erber, L.; Liu, B.; Gao, Y.; Ruan, H.-B.; Chen, Y. Proteomic Analysis Reveals Diverse Proline Hydroxylation-Mediated Oxygen-Sensing Cellular Pathways in Cancer Cells. *Oncotarget* **2016**, *7* (48), 79154–79169. <https://doi.org/10.18632/oncotarget.12632>.
- (44) Gerig, J. T. Fluorine NMR. Departement of chemistry.1997. <https://www.biophysics.org/Portals/0/BPSAssets/Articles/gerig.pdf>

Chapter 1:

General concepts of the relaxation

I. General concepts of the relaxation

The process that drives a system of spins from a high energy state to the thermal equilibrium, after a perturbation is named relaxation and constitutes a rich source of information on the molecular dynamics. While this has been largely exploited to unveil dynamical properties of proteins using mostly the relaxation properties of the ^{15}N nucleus, studies of ^{19}F relaxation of fluorinated compounds remains confidential. This spin is characterized by peculiar relaxation properties due to large chemical shift anisotropy and specific interactions with neighboring spins.

1. Introduction

When a group of spins undergoes a short radiofrequency pulse at its resonance frequency, they absorb the energy and transitions between the energy levels are observed. These transitions modify the populations of the energy levels that are defined by the Boltzmann distribution at thermal equilibrium. The absorbed energy is exchanged between the different spins and from the spins to the surrounding (named "lattice") through different relaxation mechanisms. Thus, we define the relaxation process as the phenomenon by which the thermal equilibrium is regained after a perturbation. This process is directly related to molecular motions and to the spin's local environments. The capacity of NMR spectroscopy to provide insights on the structural and dynamical properties of molecules originates from the interpretation of the relaxation rates of different spin states. The relaxation rates are sensitive to a wide range of time scales from pico-seconds to seconds.¹ Whenever the time scale of a relaxation parameter matches the kinetics of the molecular movements, we say that the relaxation parameter is sensitive to this motion and can thus be used to interpret the molecular dynamics. Figure 1.

In this chapter, we introduce the general aspects of the relaxation pathways as described in literature.

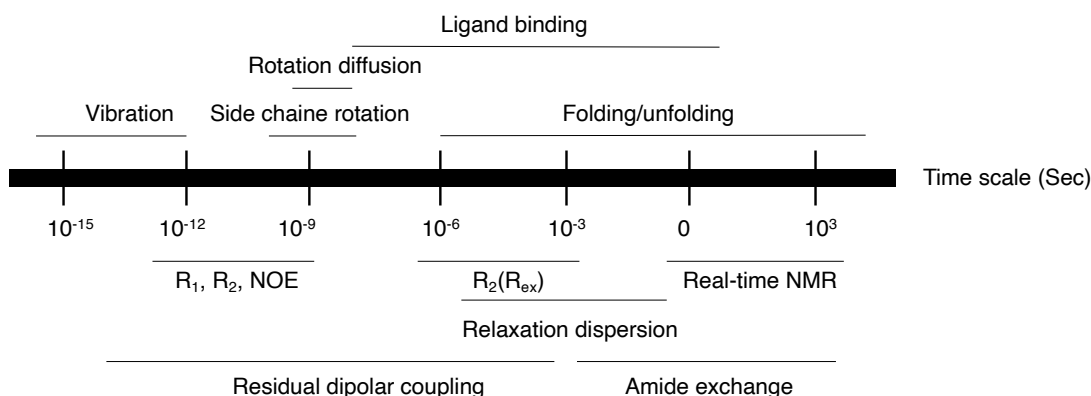


Figure 1: Motion time scales of biomolecules and their corresponding sensitive NMR parameters or methods. Inspired from ².

2. Mechanisms of the relaxation

The application of an external static magnetic field $\mathbf{B}_0$, leads to the formation of two energy levels named alpha and beta for a spin $\frac{1}{2}$. The energy difference is: $\Delta E = \frac{h\gamma B_0}{2\pi}$. The sum of the magnetic moments is non null and leads to a macroscopic magnetization that is represented by a vector $\mathbf{M}$ parallel to $\mathbf{B}_0$.

After a perturbation, the components that are aligned with $\mathbf{B}_0$ (M_z) or perpendicular to $\mathbf{B}_0$ (M_{xy}) may evolve at different rates and their evolutions define the longitudinal and the transverse relaxations, respectively. During the transverse relaxation, the M_{xy} components return to zero (their thermal equilibrium value) by exchanging the magnetization between the neighboring spins, a process named spin-spin relaxation. The thermal equilibrium value of M_z (the polarization) is reached by a transfer of the energy from the spin to its surrounding lattice, consequently it is termed spin lattice relaxation.³

An effective relaxation mechanism must involve magnetic interactions with the relaxing spin that fluctuate with time⁴. Within the NMR sample, each spin undergoes many local magnetic fields originating from different internal and external sources. These latter are modulated by molecular movements and therefore fluctuate with time. When the frequencies of the random fluctuations match the Larmor frequency of the spin, they contribute to an efficient relaxation mechanism.³ In high field solution NMR, five main relaxation mechanisms are considered:

- Dipolar relaxation:

This is the most important relaxation mechanism for nuclei with spin $\frac{1}{2}$. It originates from the interaction between the two magnetic moments of two spins. This interaction depends on the orientation of the vector linking the two spins relative to the direction defined by the external magnetic field as well as the distance between the spins (the module of the inter-spin vector). Both parameters are modulated by molecular movements and are consequently fluctuating with time.

- Chemical shift anisotropy (CSA):

Nuclear spins undergo slightly different magnetic field $\mathbf{B}_0$ upon the orientation of molecules relative to the direction of $\mathbf{B}_0$ due to the anisotropic arrangement of the molecular electronic orbitales. The molecular movements result in a reorientation the shielding factor that modulates the local magnetic field. The chemical shift anisotropy is weak for protons but could be more effective for spins with more electrons such as ^{13}C or ^{19}F .

- Spin rotation mechanism:

This mechanism results from the modulation of the molecular magnetic field produced by charged particles within the molecule. It is very effective for small molecules where the molecular rotation is fast.

- Quadrupolar relaxation mechanism:

This relaxation mechanism arises in presence of spins with spin quantum number higher than $\frac{1}{2}$ (such as ^{14}N) located in the vicinity of the observed nucleus. It is the result of the interaction between the local electric field gradient and the quadrupolar moment of the nucleus.

- Scalar relaxation mechanism:

A scalar coupling constant between two spins is independent of the molecular orientation, however, this interaction may be modulated by two distinct mechanisms:

- One of the spins is involved in a chemical exchange that can occasionally interrupt the coupling between the two spins, leading to a rapid fluctuation of the coupling constant.

- One of the spins relaxes very rapidly (with life-time shorter than $1/J$) introducing a random fluctuation of the interaction experienced by the first spin.

For a molecule in solution, the relaxation can result from several of these mechanisms with either common or distinct motion time scales. Usually, their effects add up for increased relaxation rates. In some cases, two interactions modulated by the same motion may cancel or reduce the amplitude of the fluctuating magnetic field, reducing the efficiency of the relaxation. This partial cancelation of the relaxation is known as the cross-correlated relaxation phenomenon.

3. Phenomenological Bloch equations

In 1946, Felix Bloch proposed a differential equation describing the time evolution of the magnetization vector $\mathbf{M}$ upon its interaction with external magnetic field $\mathbf{B}$. The equation describes both the precession of the magnetization around the magnetic field and its relaxation after a perturbation by the application of the oscillating $\mathbf{B}_1$ field.^{3,5}

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \wedge \mathbf{B} - R(\mathbf{M} - \mathbf{M}_0) \quad (\text{Eq. 1})$$

Where the magnetic field is described by the vector $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$ that is composed of the static

magnetic field $\mathbf{B}_0 = \begin{pmatrix} 0 \\ 0 \\ B_0 \end{pmatrix}$ and the time dependent electromagnetic component

$\mathbf{B}_1 = \begin{pmatrix} B_1(t) \\ 0 \\ 0 \end{pmatrix}$. The direction of $\mathbf{B}_0$ defines the orientation of the z axis in the laboratory frame

and the coordinates of the magnetization:

$$\mathbf{M} = M_x, M_y, M_z \quad (\text{Eq. 2})$$

R is the relaxation matrix expressed as:

$$\begin{pmatrix} R2 & 0 & 0 \\ 0 & R2 & 0 \\ 0 & 0 & R1 \end{pmatrix} \quad (\text{Eq. 3})$$

Where $R1 = \frac{1}{T1}$ and $R2 = \frac{1}{T2}$ are the transverse and the longitudinal relaxation rate constants, respectively.

At thermal equilibrium, the transverse components of the magnetization $\mathbf{M}$ are $M_x = M_y = 0$, and the longitudinal component M_z represents the polarization whose value is given by the Curie law:

$$M^0 = \frac{N\mu^2 B_0}{3kT} \quad (\text{Eq. 4})$$

where μ is the magnetic moment, N is the number of spins, B_0 is the static magnetic field, K is the Boltzmann constant and T is the temperature.

When the ensemble of the spins is perturbed by the application of an electromagnetic pulse whose magnetic component is described by $\mathbf{B}_1(t)$, the magnetization starts to precess around the effective field $\mathbf{B}_0 + \mathbf{B}_1(t)$. If the frequency of $\mathbf{B}_1$ matches the resonance condition, that is:

$$\mathbf{B}_1(t) = \begin{pmatrix} B_1 \cdot \cos(\omega_0 t) \\ 0 \\ 0 \end{pmatrix} \quad (\text{Eq. 5})$$

Where $\omega_0 = \gamma \mathbf{B}_0$ is the Larmor pulsation, the solution of the Bloch equation describes a precession of $\mathbf{M}$ around the x axis. The $\mathbf{B}_1$ radiofrequency pulse is usually applied during a time such as the magnetization undergoes a total rotation of 90° and gets aligned to the y axis. As soon as $\mathbf{M}$ has non-zero transverse components, the Bloch equation describes a precession around the $\mathbf{B}_0$ axis (z) and, after the $\mathbf{B}_1$ pulse, the Bloch equation may be written as:

$$\frac{dM_x}{dt} = \gamma(M_y B_0) - R_2(M_x) \quad (\text{Eq. 6.1})$$

$$\frac{dM_y}{dt} = -\gamma(M_x B_0) - R_2(M_y) \quad (\text{Eq. 6.2})$$

$$\frac{dM_z}{dt} = -R_1(M_z - M_0) \quad (\text{Eq. 6.3})$$

The solutions of the differential equations (6), given the initial state of a magnetization aligned with the y axis are:

$$M_x(t) = M^0 \sin(\omega t) e^{-R_2 t} \quad (\text{Eq. 7.1})$$

$$M_y(t) = -M^0 \cos(\omega t) e^{-R_2 t} \quad (\text{Eq. 7.2})$$

$$M_z(t) = M^0 (1 - e^{-R_1 t}) \quad (\text{Eq. 7.3})$$

While the first two equations describe the free precession of the transverse magnetization around the z axis with a pulsation ω (defined as the Larmor frequency γB_0 modulated by chemical shifts), the longitudinal component M_z returns to its equilibrium value with a first order kinetic.

The measure of R_1 and R_2 provides information about the dynamics of the spin system thanks to the semi-classical theory of relaxation proposed by Bloembergen and Redfield⁶. In the following section, we will introduce the underlying concepts allowing to relate relaxation parameters to molecular motion frequencies.

4. Longitudinal relaxation rate explained by transitions between the energy levels of a single spin

Describing how random motions affect the relaxation rates involves the calculation of transition probabilities between Zeeman states as shown below for a $\frac{1}{2}$ single spin.

4.1- The case of a single spin $\frac{1}{2}$

The component of the net magnetization along the magnetic field M_z is proportional to the population difference between the eigenstates α and β . During relaxation, the recovery of M_z requires random transitions between these energy levels to occur Figure 2. The rate at which equilibrium populations are established after a perturbation depends on the number of transitions per time unit experienced by the system that is both proportional to the transition rate denoted W and the deviation of the populations from the Boltzmann distribution. Thus, the transition rate from α to β is $W_{\alpha \rightarrow \beta} (n_\alpha - n_\alpha^{ea})$. Similarly, the transition rate from β to α is $W_{\beta \rightarrow \alpha} (n_\beta - n_\beta^{ea})$. Where n_α^{ea} and n_β^{ea} are the populations of the energy state α and β at thermal equilibrium, respectively.⁷⁻¹⁰¹¹

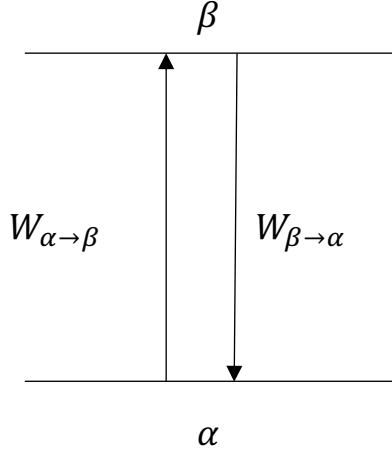


Figure 2: Transitions between the energy levels α and β responsible of the longitudinal relaxation at rate constants $W_{\alpha \rightarrow \beta}$ and $W_{\beta \rightarrow \alpha}$ in a single spin system.

The population changes are described by first order kinetic equations as follows:

$$\frac{d}{dt}n_{\alpha} = -W_{\alpha \rightarrow \beta}(n_{\alpha} - n_{\alpha}^{ea}) + W_{\beta \rightarrow \alpha}(n_{\beta} - n_{\beta}^{ea}) \quad (\text{Eq. 8})$$

$$\frac{d}{dt}n_{\beta} = W_{\alpha \rightarrow \beta}(n_{\alpha} - n_{\alpha}^{ea}) - W_{\beta \rightarrow \alpha}(n_{\beta} - n_{\beta}^{ea}) \quad (\text{Eq. 9})$$

Finally, we can deduce the evolution of the magnetization along z from equations (8) and (9) and assuming that $W_{\alpha \rightarrow \beta} = W_{\beta \rightarrow \alpha}$

$$\frac{d}{dt}M_Z = \frac{d}{dt}(n_{\alpha}) - \frac{d}{dt}(n_{\beta}) \quad (\text{Eq. 10})$$

$$\frac{d}{dt}M_Z = -W_{\alpha \rightarrow \beta}(n_{\alpha} - n_{\alpha}^{ea}) + W_{\beta \rightarrow \alpha}(n_{\beta} - n_{\beta}^{ea}) - W_{\alpha \rightarrow \beta}(n_{\alpha} - n_{\alpha}^{ea}) + W_{\beta \rightarrow \alpha}(n_{\beta} - n_{\beta}^{ea})$$

$$\frac{d}{dt}M_Z = -2W_{\alpha \rightarrow \beta}[(n_{\alpha} - n_{\alpha}^{ea}) - (n_{\beta} - n_{\beta}^{ea})]$$

$$\frac{d}{dt}M_Z = -2W_{\alpha \rightarrow \beta}[(n_{\alpha} - n_{\beta}) - (n_{\alpha}^{ea} - n_{\beta}^{ea})] \quad (\text{Eq. 11})$$

By analogy to the equation (6.3) we identify the longitudinal relaxation rate as:

$$R_1 = -2w_{\alpha \rightarrow \beta} \quad (\text{Eq. 12})$$

4.2- Longitudinal relaxation in a two spins case: the Solomon equations

While the relaxation rate due to the chemical shift anisotropy can be established from a single random transition rate defined above, the longitudinal relaxation due to dipole-dipole interaction requires at least two spins (denoted here after I and S) to be described. Two heteronuclear weakly coupled $\frac{1}{2}$ spins have 4 energy levels $\alpha\alpha$, $\beta\beta$, $\alpha\beta$ and $\beta\alpha$ and 12 possible transitions classified into eight single quantum transitions W_1 involving a single spin transition between its eigenstates, two double quantum transitions W_2 and two zero quantum transitions W_0 (Figure 3).^{7,11}

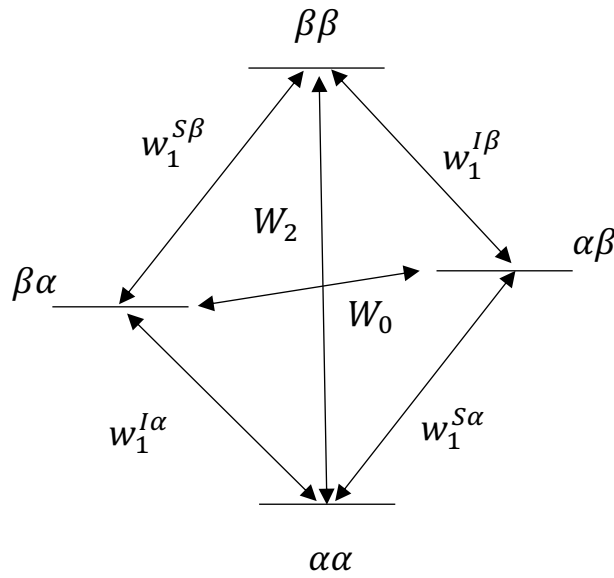


Figure 3: Transition probabilities in a weakly coupled AX spin system between the eigenstates. W is the transition rate constant, the under script is the order of the transition (zero, single or double quantum transitions noted 0,1 and 2 respectively). The upper script I and S indicate the active spin.

As shown above, the population changes between the four energy states will be described by first order kinetic differential equations as in (8) and (9).

Thus, for the energy level $\alpha\alpha$ the rate of the population change is:

$$\frac{d}{dt} n_{\alpha\alpha} = -w_1^{I\alpha}(n_{\alpha\alpha} - n_{\alpha\alpha}^{ea}) - w_1^{S\alpha}(n_{\alpha\alpha} - n_{\alpha\alpha}^{ea}) - W_2(n_{\alpha\alpha} - n_{\alpha\alpha}^{ea}) + w_1^{I\alpha}(n_{\beta\alpha} - n_{\beta\alpha}^{ea}) + w_1^{S\alpha}(n_{\alpha\beta} - n_{\alpha\beta}^{ea}) + W_2(n_{\beta\beta} - n_{\beta\beta}^{ea}) \quad (Eq. 13)$$

The positive and negative terms represent the transitions by which the population of the energy level increases or decreases, respectively. Similar expressions can be derived for the three other energy levels.

The net z magnetizations of the two spins I and S can be written as:

$$M_I^z = \mu_z((n_{\alpha\alpha} - n_{\beta\alpha}) + (n_{\alpha\beta} - n_{\beta\beta})) \quad (\text{Eq. 14})$$

$$M_S^z = \mu_z((n_{\alpha\alpha} - n_{\alpha\beta}) + (n_{\beta\alpha} - n_{\beta\beta})) \quad (\text{Eq. 15})$$

Where for a spin $\frac{1}{2}$ μ_z is the angular momentum along z, $\mu_z = \frac{1}{2}\gamma\hbar$. γ being the gyromagnetic ratio of the spin.

By definition, the longitudinal relaxation rate is the rate of change of the z magnetization thus, it is described by the evolution in time of M_I^z and M_S^z . To simplify the expressions, the angular momentum μ_z will no longer be included in the expressions of the net z magnetizations. Consequently, one should keep in mind that all the expressions we give need to be multiplied by μ_z .

$$\frac{d}{dt} M_I^z = \frac{d}{dt} (n_{\alpha\alpha} - n_{\beta\alpha}) + \frac{d}{dt} (n_{\alpha\beta} - n_{\beta\beta}) \quad (\text{Eq. 16})$$

$$\frac{d}{dt} M_I^z = -(w_0 + 2w_1 + w_2)(M_I^z - M_I^{zeq}) - (w_2 - w_0)(M_S^z - M_S^{zeq}) \quad (\text{Eq. 17})$$

Where M_I^{zeq} and M_S^{zeq} are the magnetization along the z axis at equilibrium of the spins I and S respectively.

By analogy to Bloch's equation (6.3) we identify the longitudinal relaxation rate R_1 also noted auto-relaxation rate ρ :

$$R_1 = \rho = w_0 + 2w_1 + w_2 \quad (\text{Eq. 18})$$

As well as an additional term, the cross-relaxation rate constant σ :

$$\sigma = w_2 - w_0 \quad (\text{Eq. 19})$$

The Solomon equation shown in equation (17) states that the longitudinal relaxation of a given spin, in dipolar interaction with neighboring spins depends on two relaxation rates (the auto- and cross relaxation rates) that differ by their transition rates and therefore by their dependencies from random molecular motions. One important conclusion of the Solomon

equations is that the resulting longitudinal relaxation rate R_1 depends on the magnetization of neighboring spins: a phenomenon that leads to the Nuclear Overhauser Enhancement (NOE). Another conclusion of these equations is that the effective longitudinal relaxation rate depends on the initial magnetization states of the interacting spins. After an inversion of both the populations of spins I and S, and assuming that the equilibrium values of I and S magnetizations are equal to M_I^{zeq} and M_S^{zeq} , with $M_I^{zeq} = M_S^{zeq}$, at $t=0$ equation (17) could be rewritten as:

$$\frac{d}{dt} M_I^z = -\rho(-2M_I^{zeq}) - \sigma(-2 M_I^{zeq}) \text{ at } t=0 \quad (\text{Eq. 20})$$

To which corresponds the solution:

$$M_I^z = M_I^{zeq}(1 - 2\exp^{-(\rho+\sigma)t}) \quad (\text{Eq. 21})$$

Equation (21) describes a return to the equilibrium population governed by a single exponential with an effective longitudinal relaxation rate: $R_1 = \rho + \sigma$

This can be generalized for N interacting spins as: $R_1 = \rho + \sum_1^N \sigma_{IS}$

A different behavior is obtained if the interacting spin S is left at equilibrium, after a selective perturbation of the observed I spin. In this case, $M_I^z = -M_I^{zeq}$ and $M_S^z = M_S^{zeq}$ at the origin of the evolution of populations and the return to equilibrium is governed by two exponential functions with two effective rates $(\rho + \sigma)$ and $(\rho - \sigma)$:

$$M_I^z = -M_I^{zeq}(1 - (\exp^{-(\rho+\sigma)t} - \exp^{-(\rho-\sigma)t})) \quad (\text{Eq. 22})$$

$$M_S^z = M_S^{zeq}(1 - (\exp^{-(\rho+\sigma)t} + \exp^{-(\rho-\sigma)t})) \quad (\text{Eq. 23})$$

The ability to set the magnetization of spins that are neighboring the observed spin allows modulating its effective longitudinal relaxation rate, a principle that has been exploited to speedup NMR acquisition time by shortening longitudinal recovery delays in SOFAST experiments of Band-selective Excitation Short Transient BEST variants of 3D triple resonance spectroscopies.¹²

4.3- Expression of transition rates as a function of random molecular motions

4.3-1- Correlation and Spectral density functions

The random nature of the magnetic fluctuation leading to relaxation requires the use of a statistical approach to describe its time dependence. This is accomplished using the ensemble average of the correlation function $\mathbb{G}(\mathcal{T})$ of the random fluctuating magnetic field $b(t)$. The auto-correlation function provides a statistical measure of the variations of the random magnetic field b between two times separated by a delay $\mathcal{T}$.^{13,14}

$$\mathbb{G}(\mathcal{T}) = \overline{b(t)b(t + \mathcal{T})} \quad (\text{Eq. 24})$$

$\overline{b(t)}$ expresses the average of the randomly fluctuating magnetic field. The magnitude of the fluctuating field is expressed through its non-null mean square $\overline{b^2}$ along the axis x,y and z.

$$\overline{b^2} = \overline{b_x^2(t)} = \overline{b_y^2(t)} = \overline{b_z^2(t)} \quad (\text{Eq. 25})$$

If we assume an isotropic rotational diffusive motion of the molecule, the correlation function is a single exponential function that decays as a function of time with a rate defined by the rotational diffusion constant D_r .¹⁵

$$\mathbb{G}(\mathcal{T}) = \overline{b^2} \exp \left(-\frac{\mathcal{T}}{\mathcal{T}_c} \right) \quad (\text{Eq. 26})$$

with $\mathcal{T}_c = \frac{1}{6D_r}$

τ_c is the correlation time that corresponds to the time needed for the molecule to rotate by $\sqrt{\frac{2}{3}}$ rad in average.¹⁰ Brownian diffusion of a vector in a sphere may be simulated by simple Monte-Carlo calculations allowing to visualize the exponential decay of the correlation function Figure 4. A short correlation time corresponds to a rapid exploration of all orientations in a sphere while a long correlation time corresponds to a slower motion.

George Gabriel Stokes derived an expression of the rotational correlation time for a sphere with a radius R_h tumbling within a medium of viscosity η and a temperature T :¹⁶

$$\mathcal{T}_c = \frac{\eta V}{K_B T}$$

V is the volume of the sphere with a hydrodynamic radius R_h that depends on the molecular size as well as on its hydration shell. More complex molecular motions lead to multiple exponentials decay of the correlation function described by additional correlation times. Several models have been designed to describe molecular motions, the most popular being the Woessner model that enables the description of anisotropic diffusive motion and the Lipari-Szabo "model-free" approach that allows a description of fast internal motions.^{17,18}

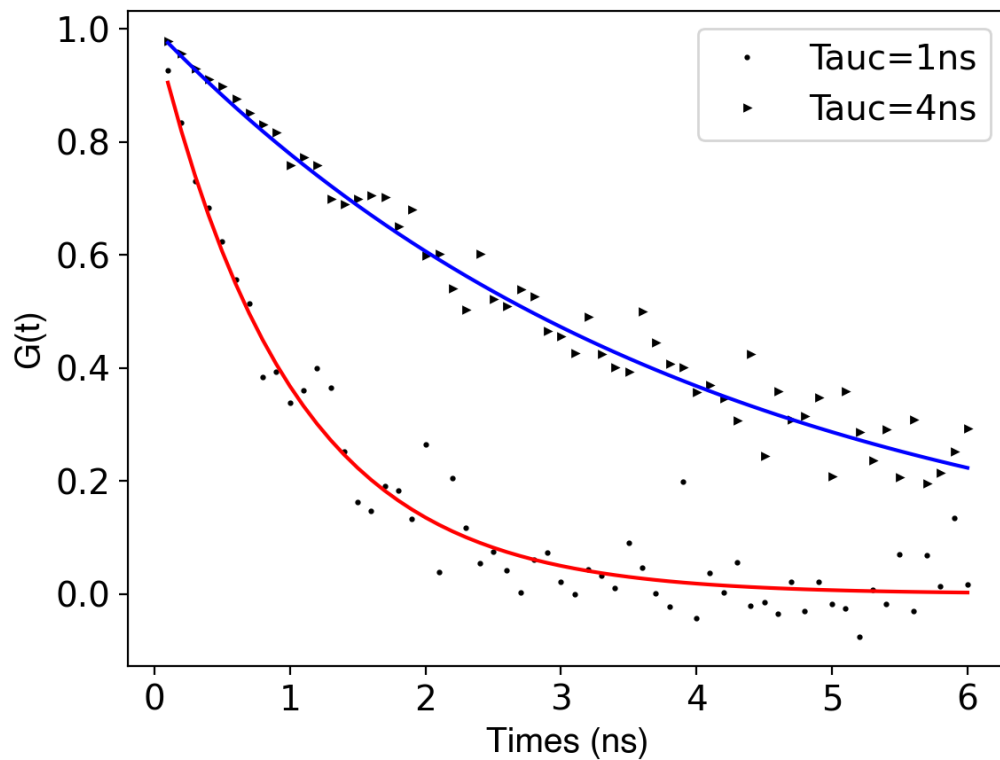


Figure 4: Simulation of the exponential decay of the average of the correlation time of two vectors under the effect of different correlation times 1 ns (dots) and 4 ns (triangles). The trajectory of each unit vector was determined through a random walk on a unit sphere during 6 ns. The random steps were separated by a delay of 0.02 ns thus the total number of steps was 300. The correlation function is the ensemble average over 100-unit vector. The red and blue curves represent the decay of an exponential function $\exp^{-\frac{t}{\mathcal{T}_c}}$ where $\mathcal{T}_c = 1 \text{ ns}$ and 4 ns respectively.

Since transitions between the spin states will have higher probabilities for random motions matching characteristic frequencies of the spin system (such as the Larmor frequency), descriptions of random motions are performed in the frequency domain. Consequently, a Fourier transform is applied on the autocorrelation function to obtain the spectral density function $J(\omega)$. The spectral density function quantifies the efficacy of the random fluctuations of the magnetic interactions $b(t)$ to the relaxation at a frequency ω .

$$J(\omega) = \int_{-\infty}^{\infty} \mathbb{G}(\mathcal{T}) \exp(-i \omega \mathcal{T}) d\mathcal{T} \quad (\text{Eq. 27})$$

Since the autocorrelation function for a time $\mathcal{T}$ and $-\mathcal{T}$ is equal, we can write:

$$J(\omega) = 2 \int_0^{\infty} \mathbb{G}(\mathcal{T}) \exp(-i \omega \mathcal{T}) d\mathcal{T} = 2 \int_0^{\infty} \overline{b^2} \exp^{-\frac{\mathcal{T}}{\mathcal{T}_c}} \exp(-i \omega \mathcal{T}) d\mathcal{T}$$

Which leads to:

$$J(\omega) = 2 \overline{b^2} \frac{\mathcal{T}_c}{1 + \omega^2 \mathcal{T}_c^2} \quad (\text{Eq. 28})$$

$\overline{b^2}$ depends on the magnitude of magnetic interactions that are modulated by random motions while the Lorentzian function $\frac{2\mathcal{T}_c}{1 + \omega^2 \mathcal{T}_c^2}$ describes the spectral components of these motions.¹⁰ This latter is denoted as the normalized spectral density $j(\omega)$ often directly introduced as the spectral density in the literature. For an isotropic tumbling with one representative correlation time, the normalized spectral density denoted is given by:

$$j(\omega) = \frac{2\mathcal{T}_c}{1 + \omega^2 \mathcal{T}_c^2} \quad (\text{Eq. 29})$$

In the following, we will use the term spectral density to denote the normalized spectral density. The spectral density is maximum when $\omega = 0$ and decays at higher frequencies with a rate that depends on the correlation time (figure 5). The area under $j(\omega)$ corresponds to the first point of the correlation function and is independent of the correlation time $\mathcal{T}_c$. Thus, the spectral density function describes how a given amount of energy available for relaxation is distributed over the frequency space.

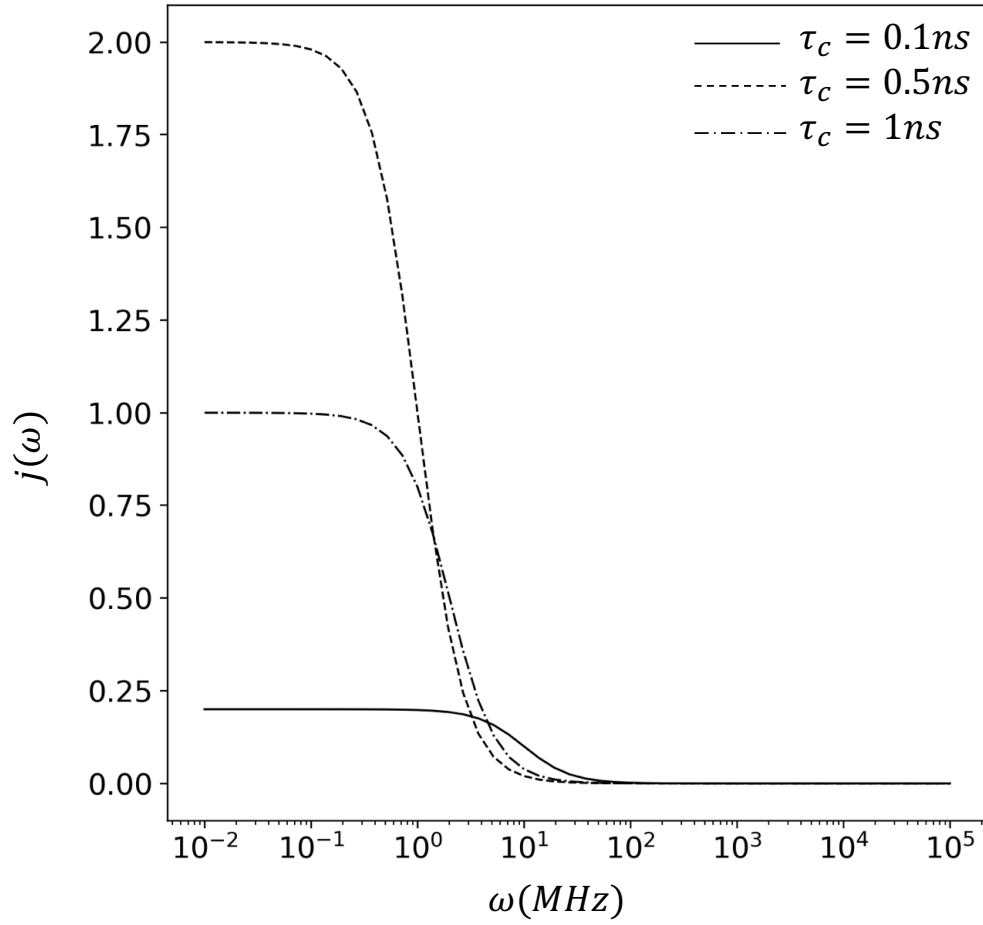


Figure 5: Simulation of the evolution of the spectral density as a function of the frequency at different correlation times.

4.3-2- Relaxation through dipolar interaction

The rates of transitions will depend on both the effective relaxation mechanisms (that will define the value of $\overline{b^2}$) and how the corresponding magnetic fields are modulated by random molecular motions as described by $j(\omega)$. This will be introduced below by describing how a dipolar interaction between two spins is modulated by the relative orientation of the two corresponding dipoles. The dipolar interaction energy depends on the distance separating the pair of spins as well as on the direction of the vector connecting them (Figure 6).

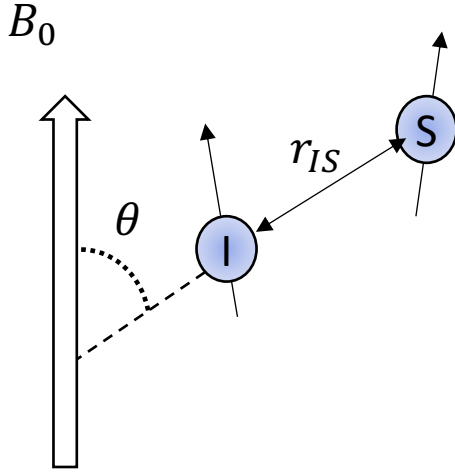


Figure 6: Dipolar interaction between two spins I and S separated by the vector r_{IS} under the effect of the magnetic field $\mathbf{B}_0$. θ is the angle formed between r_{IS} and $\mathbf{B}_0$.

The energy is proportional to the dipole-dipole coupling constant $b^2 = \frac{\mu_0}{4\pi} \left(\frac{\hbar \gamma_I \gamma_S}{r_{IS}^3} \right)^2$ and depends on the angle θ defined in figure 6:

$$E_{DD} \approx \frac{\mu_0}{4\pi} \left(\frac{\hbar \gamma_I \gamma_S}{r_{IS}^3} \right)^2 (3 \cos^2 \theta - 1) \quad (\text{Eq. 30})$$

Where γ_I and γ_S are the gyromagnetic ratios of the spins I and S respectively, μ_0 the magnetic permeability of the vacuum and r_{IS} is the distance between the interacting spins.

The dipolar energy changes its sign depending on the relative orientation of the two spins, with an unfavorable contribution when both spins are assuming a colinear orientation relative to the applied magnetic field and a favorable contribution when the inter-spin vector is perpendicular to the magnetic field. Noteworthy, the dipolar energy is null at the magic angle $\theta \cong 54.74^\circ$.

When the dipolar interaction is the only contributing mechanism to the longitudinal relaxation the transition probabilities between the energy levels can be written as follows:

$$w_0 = \frac{1}{20} b^2 j(\omega_I - \omega_S) \quad (\text{Eq. 31})$$

$$w_1 = \frac{3}{40} b^2 j(\omega_I) \quad (\text{Eq. 32})$$

$$w_2 = \frac{3}{10} b^2 j(\omega_I + \omega_S) \quad (\text{Eq. 33})$$

Where ω_I and ω_S are the Larmor frequencies of the observed spins I and S respectively.

The auto-relaxation rate constant ρ and the cross-relaxation rate constant σ are therefore:

$$\rho = \frac{1}{20} b^2 [j(\omega_I - \omega_S) + 3j(\omega_I) + 6j(\omega_I + \omega_S)] \quad (Eq. 34)$$

$$\sigma = \frac{1}{20} b^2 [6j(\omega_I + \omega_S) - j(\omega_I - \omega_S)] \quad (Eq. 35)$$

With $j(\omega)$ being the normalized spectral density function introduced in equation (29). The coefficients used in equations (31) to (33) arise from the averaging of the dipolar interaction over all possible orientations of the r_{IS} vector to compute a mean value of the dipolar interaction. The behaviors of the relaxation rate constants R_1 , ρ and the absolute magnitude of σ as well as the transverse relaxation rate constant R_2 (introduced in paragraph 5) as a function of the correlation time are displayed on figure 7. These constants were calculated for two protons separated by 1.79 Å at 600 MHz. For short correlation times ($\omega_0 \tau_c < 1$), all relaxation constants are increasing with τ_c . When $\omega_0 \tau_c \approx 1$, σ becomes negative explaining the singularity in the logarithmic representation. It is the contribution of $j(0)$ that dictates the evolution of R_2 , ρ and σ for $\omega_0 \tau_c > 1$. R_1 displays a maximum value when the correlation time matches the inverse of the Larmor frequency.

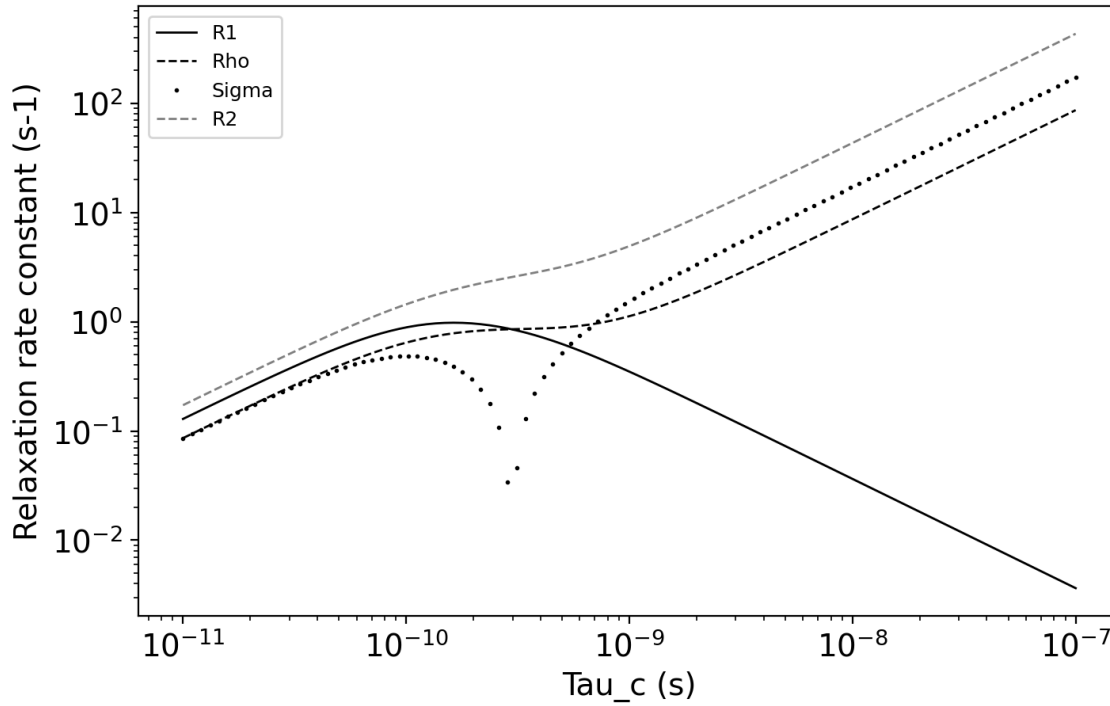


Figure 7: Evolution over correlation time of the longitudinal and transverse relaxation rate constants for a two-proton system at 600 MHz.

4.3-3- Relaxation through chemical shift anisotropy

The electronic cloud around a given NMR active nucleus is responsible of small variations from the external magnetic field B_0 due to the presence of an induced field (B_{induced}) leading to the chemical shifts. Depending on the chemical environment, the electronic molecular orbitals are not spherical and the effective field (B_{eff}) experienced by a given nucleus depends on the orientation of the molecule relatively to B_0 . The chemical shift anisotropy determines how much the nucleus's effective field strength depends on the molecule's orientation.^{7,19}

$$B_{\text{eff}} = B_0 + B_{\text{induced}} \quad (\text{Eq. 36})$$

4.3-3-1- Chemical shift tensor

The induced magnetic field is proportional to the applied magnetic field:

$$B_{\text{induced}} = \sigma B_0 \quad (\text{Eq. 37})$$

Where σ is the shielding tensor. To account for the dependency of the induced magnetic field to the molecular orientation, σ is expressed as a 3x3 matrix whose coefficients form the chemical shift tensor:

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \quad (Eq. 38)$$

The axes of the effective magnetic field are not necessarily parallel to those of the applied magnetic field B_0 , rather depend on the molecular orientation with respect to this latter and on the position of the nuclei within the molecule. The fluctuations of σ_{ij} coefficients due to the random molecular motions contribute to random fluctuations of $B_{induced}$ and subsequently to relaxation. We define the principal components as σ_{XX} , σ_{YY} and σ_{ZZ} values for a molecular orientation in which the tensor σ is diagonal. The defined axes are always orthogonal to each other. The average of the three principal values of the chemical shift tensor represents the isotropic chemical shift:

$$\sigma_{iso} = \frac{\sigma_{zz} + \sigma_{xx} + \sigma_{yy}}{3} \quad (Eq. 39)$$

The chemical shift tensor is considered as isotropic when the three principal components are equal, whereas it is said to be anisotropic when two of them or all of them are different. In literature, different notations are used to describe the chemical shift tensor leading to some confusion. In what follows, we present the two frequently used conventions and how to switch between them.^{7,20}

4.3-3-2- Haeberlen notation

To describe the tensor, this notation defines new combinations of the principal components relating them to the isotropic average such as²¹:

$$|\sigma_{zz} - \sigma_{iso}| \geq |\sigma_{xx} - \sigma_{iso}| \geq |\sigma_{yy} - \sigma_{iso}|$$

To describe the shielding tensor, three main parameters are defined:

- The **isotropic average** defined in equation (39)
- The **shielding anisotropy** that can take two different definitions:

$$\Delta\sigma = \sigma_{ZZ} - \frac{(\sigma_{XX} + \sigma_{YY})}{2} \quad (44) \text{ or } \zeta = \sigma_{ZZ} - \sigma_{iso} \quad (Eq. 40)$$

$$\text{Thus} \quad \Delta\sigma = \frac{3}{2}\zeta \quad (Eq. 41)$$

ζ is sometimes called the reduced anisotropy

- The **shielding asymmetry** η also referred to as bioaxiality term such as:

$$\eta = \frac{3}{2} \frac{\sigma_{YY} - \sigma_{XX}}{\Delta\sigma} = \frac{\sigma_{YY} - \sigma_{XX}}{\zeta} \quad (Eq. 42)$$

$$0 \leq \eta \leq 1$$

4.3-3-3- Maryland notation

The principal components are written using numerical subscripts and are given in the order of increasing shielding values:

$$\sigma_{11} \leq \sigma_{22} \leq \sigma_{33}$$

The anisotropy is replaced by the span Ω that is positive and used to describe the full extent of the spectrum:

$$\Omega = \sigma_{33} - \sigma_{11} \quad (Eq. 43)$$

The asymmetry is replaced by the skew k . The orientation of σ_{22} with respect to the isotropic chemical shift tensor determines the sign of k .

$$k = \frac{3(\sigma_{iso} - \sigma_{22})}{\Omega} \quad (Eq. 44)$$

The skew can be:

- Negative: ($\sigma_{iso} > \sigma_{22}$) thus the shielding tensor is axially symmetric and is represented by an oblate ellipsoid
- Positive: ($\sigma_{iso} < \sigma_{22}$) thus the shielding tensor is axially symmetric and is represented by a prolate ellipsoid
- Zero: when σ_{22} falls between σ_{11} and σ_{33} . It corresponds to $\eta = 1$

4.3-3-4- Switching between Haeberlen and Maryland notation

The ability to switch between these notations is crucial to compare CSA values. For this sake we have to consider these two situations:

When $\Delta\sigma > 0$ then:

$$\sigma_{11} = (\sigma_{iso} + \Delta\sigma) \quad (Eq. 45)$$

$$\sigma_{22} = \sigma_{iso} - \frac{\Delta\sigma(1-\eta)}{2} \quad (Eq. 46)$$

$$\sigma_{33} = \sigma_{iso} - \frac{\Delta\sigma(1+\eta)}{2} \quad (Eq. 47)$$

When $\Delta\sigma < 0$ then:

$$\sigma_{33} = (\sigma_{iso} + \Delta\sigma) \quad (Eq. 48)$$

$$\sigma_{22} = \sigma_{iso} - \frac{\Delta\sigma(1-\eta)}{2} \quad (Eq. 49)$$

$$\sigma_{11} = \sigma_{iso} - \frac{\Delta\sigma(1+\eta)}{2} \quad (Eq. 50)$$

4.3-3-5- Contribution of the CSA to the longitudinal relaxation rate

The CSA contribute to the relaxation when the shielding tensor is anisotropic ($\sigma_{XX} \neq \sigma_{YY} \neq \sigma_{ZZ}$) or axially symmetric ($\sigma_{XX} = \sigma_{YY}$).^{15,22}

The contribution of CSA to the longitudinal relaxation rate constants is expressed as follows:

$$R_{1\text{ CSA}} = \left(\frac{1}{15}\right) \gamma^2 B_0^2 \Delta\sigma^2 \left(1 + \frac{\eta^2}{3}\right) j(\omega_0) \quad (Eq. 51)$$

From equations 36 and 37 we notice that the contribution of the CSA to the relaxation is especially effective for spins with large chemical shift range such as: ^{13}C , ^{19}F and ^{15}N and increases with the square of the magnetic field.²³

5. Transverse relaxation rate

Similar to the longitudinal relaxation rate, the transverse relaxation rate R_2 is modulated through the random molecular motions and can be expressed in terms of transition probabilities. In contrast with the longitudinal relaxation rates, the transverse relaxation includes additional contributions from chemical exchange. Below, we give the expression of the transverse relaxation rate considering the dipolar and chemical shift anisotropy relaxation mechanisms as well as when it is hampered by an exchange contribution.

5.1. Relaxation through dipolar interaction

Using similar approach as in paragraph 4 to express the transverse relaxation for two spins in dipolar interaction:

$$R_2 = \frac{1}{20} b^2 \left[3j(\omega_I + \omega_S) + \frac{3}{2}j(\omega_I) + 3j(\omega_S) + \frac{1}{2}j(\omega_I - \omega_S) + 2j(0) \right] \quad (Eq. 52)$$

Equation 52 states that the transverse relaxation involves an additional component $j(0)$ known as adiabatic term. This latter is independent of the transitions between the energy levels and is a result to the relaxing local field at each spin.

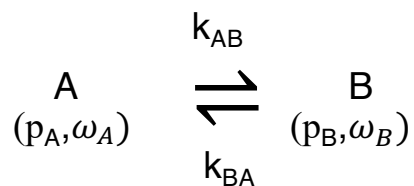
5.2. Relaxation through chemical shift anisotropy

The transverse relaxation rate constant originating from the CSA is expressed as follows: ^{15,22}

$$R_{2\text{ CSA}} = \left(\frac{1}{15} \right) \gamma^2 B_0^2 \Delta\sigma^2 \left(1 + \frac{\eta^2}{3} \right) \left[\frac{2}{3}j(0) + \frac{1}{2}j(\omega_0) \right] \quad (Eq. 53)$$

5.3. Chemical exchange

Chemical reactions and conformational changes occur on a large time scale ranging from μs to ms and lead to chemical shifts variations. These fluctuations with time scales greater than the rotational correlation time may contribute to the line widths and are represented by an additional contribution to the transverse relaxation rate constant R_2 : R_{ex} . This term can be significantly greater than R_2 and depends essentially on the chemical shift differences between the different states explored by the spin and the transition rate between these states k_{ex} . The impact of these parameters on the line width is usually introduced using a single resonance that interconverts between two states A and B. The resulting NMR spectrum depends on the rate constant k_{ex} of the interconversion between the two states ($k_{\text{ex}} = k_{AB} + k_{BA}$). Each state is characterized by the probability of its population p and its frequency ω .



We can distinguish three exchange regimes depending on the magnitude of the exchange constant rate k_{ex} by respect to $\Delta\omega = \omega_A - \omega_B$.²⁴

- A fast exchange regime when $k_{ex} \gg \Delta\omega$. The corresponding spectrum is composed of a unique resonance observed at an average frequency $p_A\omega_A + p_B\omega_B$.
- A slow exchange regime occurs when $k_{ex} \ll \Delta\omega$, two resonances are observed at ω_A and ω_B . The intensity of the resonance is proportional to its population probability, consequently, the detection of a very low populated state is sometimes challenging.
- An intermediate regime when $k_{ex} \sim \Delta\omega$ characterized by very broad single resonance (coalescence).

The interchanging states A and B might be equally populated, the exchange regimes and their corresponding NMR signals are displayed on figure 8. However, this is not always the case in biological molecules where excited states of the molecules are populated by few percent. In this case, the signal of the minor population gets too large and only the highly populated state's signal is observed. However, this signal is still broadened by the exchange with the excited state and the determination of R_{ex} can be used to derive useful information about the population, the chemical shifts and the transition rate of the excited state. In the situation of a slow exchange between skewed populations, the observation of a single broad peak may be misleading and recording NMR spectra at different fields may be used to resolve the ambiguity.

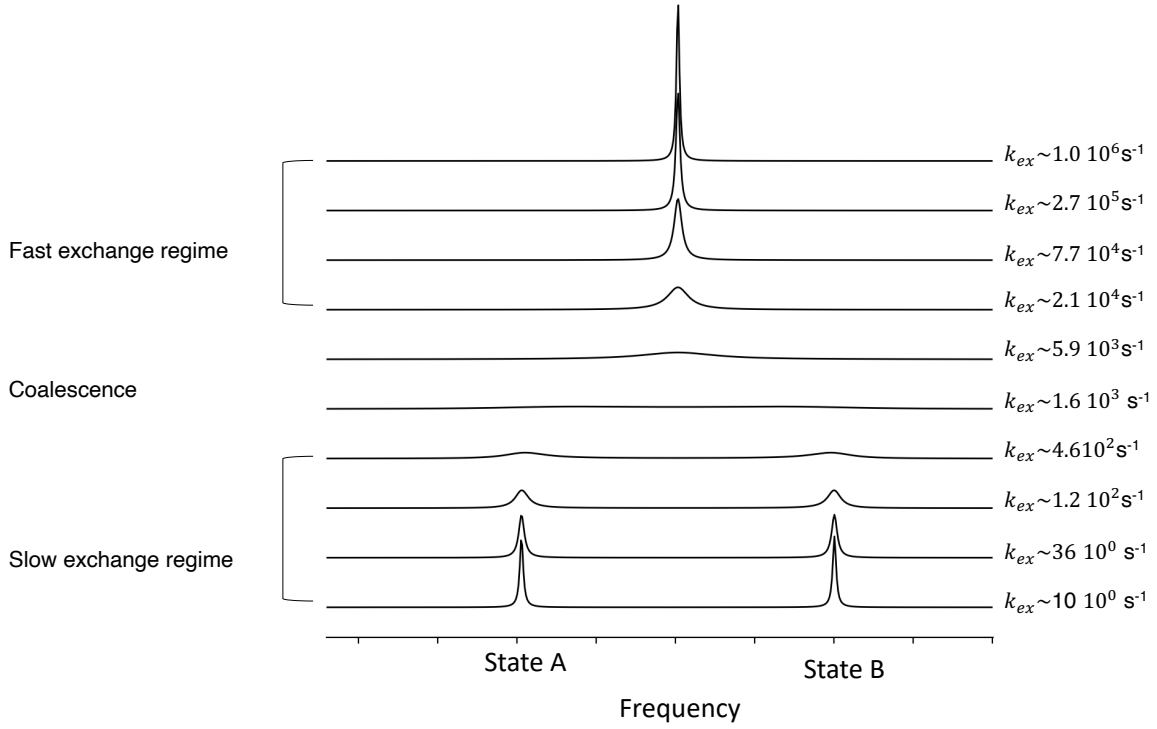


Figure 8: Simulated NMR spectra of chemical exchange between two equally populated states A and B as a function of the rate constant k_{ex} . $\Delta\omega = 1000 \text{ Hz}$.

5.3-1 Bloch-McConnell equation

The consequence of slow interconverting states on the NMR line shapes has to be addressed by resolving the coupled Bloch equations (known as Bloch-McConnell equations). Considering two interconverting states A and B with two kinetic rate constants k_{AB} and k_{BA} , the evolution of transverse magnetization can be described by:^{3,25,26}

$$\frac{d\mathbf{M}}{dt} = (i\omega + R + K) \mathbf{M} \quad (\text{Eq. 54})$$

Where:

- ω is the evolution matrix composed of the chemical shifts of the two states A and B

$$\omega = \begin{bmatrix} \omega_A & 0 \\ 0 & \omega_B \end{bmatrix}$$

- R is the relaxation matrix describing the relaxation of the transverse component originating from both states A and B

$$R = \begin{bmatrix} R_2 & 0 \\ 0 & R_2 \end{bmatrix}$$

- K is the rate matrix

$$k = \begin{pmatrix} -k_{AB} & k_{BA} \\ k_{AB} & -k_{BA} \end{pmatrix}$$

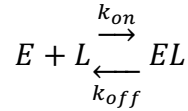
- $\mathbf{M}$ is the vector $\begin{bmatrix} M_{xy}^A \\ M_{xy}^B \end{bmatrix}$ where M_{xy}^A and M_{xy}^B are the transverse magnetization component for the A and B states, respectively. The transverse magnetization of one spin can be written as $M_{xy} = M_x + iM_y$

While an algebraic solution of equation 54 has been derived for two states, its expression is cumbersome and it is generally rather solved numerically. Simplified expressions of the solutions have been proposed. For instance, in the presence of a fast-exchanging regime, the contribution to the transverse relaxation rate is:²⁷

$$R_2^{ex} = \frac{\Delta\omega^2 P_A P_B \tau_{ex}}{1 + (\omega_{eff} \tau_{ex})^2} \quad (Eq. 55)$$

ω_{eff} being the frequency of the applied RF in the transverse plan.

When a molecule is binding in a non-covalent manner to a receptor, it undergoes a continuous shuttling between the free and bound states that can result into an alteration of line shapes depending on the exchange kinetic parameters. If we consider the following binding exchange:



The equilibrium involves an association rate $[E][L]k_{on}$ and a dissociation rate $[EL]k_{off}$. At equilibrium, these two flows are identical and the equilibrium concentrations are given by the equilibrium constant:

$$K_d = \frac{[E][L]}{[EL]} = \frac{k_{off}}{k_{on}}$$

If we observe a line of the receptor, E that is characterized by two distinct frequencies ω_E and ω_{EL} , the exchange regime will be defined by the association and dissociation rates k_{on} and k_{off} and the exchange matrix of the Bloch-McConnell equation will be written as:

$$K = \begin{pmatrix} -k_{on}[L] & k_{off} \\ k_{on}[L] & -k_{off} \end{pmatrix}$$

Solving the time evolution of the transverse magnetization $\mathbf{M} = \begin{bmatrix} M_{xy}^E \\ M_{xy}^{EL} \end{bmatrix}$ allows to investigate the effect of the binding kinetics on the line shapes as shown by the example below (Figure 9):

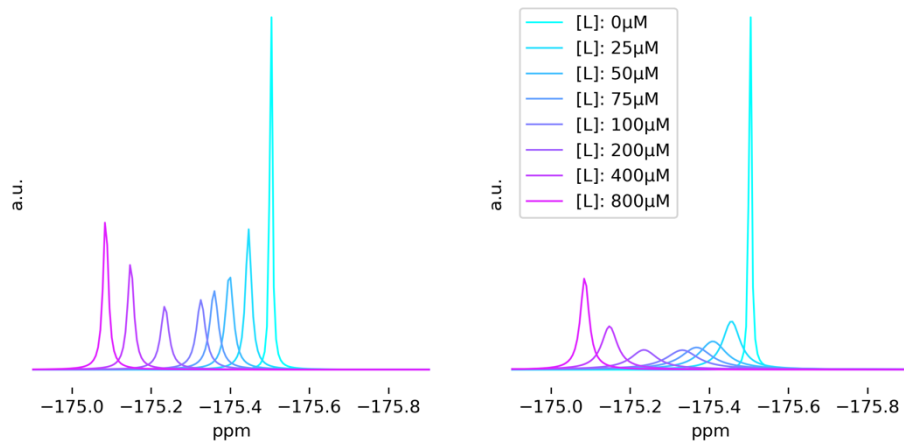


Figure 9: Simulated NMR spectra of chemical exchange during a titration experiment with $k_{on} = 10^8 \text{ Mol}^{-1}.\text{s}^{-1}$ (left) and $k_{on} = 2.10^7 \text{ Mol}^{-1}.\text{s}^{-1}$ (right). The titrated molecule contains a fluorine atom that resonates at -175.5 ppm in the free state and -175 ppm in the bound state. The dissociation equilibrium constant is set to 150 μM and the different concentrations of ligand (L) are indicated in the legend. The concentration of the observed molecule (E) is 50 μM .

5.3-2 Dependence of the exchange rate constant on the strength of the magnetic field

Millet et al suggested that in a fast exchange regime, R_{ex} varies with the quadratic magnetic field, whereas it is insensitive upon the variation of the magnetic field when the exchange regime is slow. To assess the time scale of the exchange regime one need to follow the variation of the line broadening under two different magnetic fields. A parameter α relating the strength of the magnetic field to R_{ex} is defined as following:

$$\alpha = \left(\frac{B_0^2 + B_1^1}{B_0^2 - B_1^1} \right) \left(\frac{R_{ex}^2 - R_{ex}^1}{R_{ex}^2 + R_{ex}^1} \right) \quad (Eq. 56)$$

Where B_0^2 and B_0^1 are the static magnetic fields to which corresponds the exchange rate constants R_{ex}^2 and R_{ex}^1 respectively.

The exchange regime is fast when $\alpha > 1$ and is slow when $\alpha < 1$.

6. Spectral densities and diffusion models

Molecular rotational diffusion is characterized by a rotational diffusion tensor that describes the rotations about the three orthogonal axes characterized by three principal values (D_x , D_y and D_z). The rotational diffusion tensor is represented by a diagonal 3x3 matrix when the principal axes of the diffusion tensor are colinear to those of the reference frame.

$$D_R = \begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix} \quad (Eq. 57)$$

Where,

$$D_x = \frac{1}{6\tau_x} \quad (Eq. 58)$$

$$D_y = \frac{1}{6\tau_y} \quad (Eq. 59)$$

$$D_z = \frac{1}{6\tau_z} \quad (Eq. 60)$$

Three particular cases may be defined, i) an isotropic molecular diffusion when $D_x=D_y=D_z$, ii) an Axially symmetric diffusion when $D_x=D_y$ and a fully anisotropic diffusion when $D_x \neq D_y \neq D_z$.

Here after, we give the spectral density expression for the three diffusion models.^{17,28-31}

6.1- Rigid molecule with isotropic diffusion

The simplest model consists in a rigid sphere whose isotropic tumbling is characterized by a single correlation time leading to a mono-exponential decay of the autocorrelation function. In this case, the different relaxation rates involve a single spectral density function and was above defined in equation (29).

6.2- Prolate Ellipsoid molecule:

When the molecular shape differs from a sphere, different correlation times are defined according the orientation of the geometric parameters of the relevant magnetic interaction within the molecular frame. For a prolate ellipsoid defined by an axially symmetric shape with a single long axis and two identical small axis two correlation times are defined to describe the diffusion around the long ($\mathcal{T}_L$) and the short axis ($\mathcal{T}_S$).

According to Woessner, the spectral density function is expressed as following:

$$j(\omega) = a_1 j(\omega, \mathcal{T}_1) + a_2 j(\omega, \mathcal{T}_2) + a_3 j(\omega, \mathcal{T}_3) \quad (\text{Eq. 61})$$

Where,

$$a_1 = 0.25(3\cos^2\beta - 1)^2 \quad (\text{Eq. 62})$$

$$a_2 = 3\cos^2\beta \sin^2\beta \quad (\text{Eq. 63})$$

$$a_3 = 0.75 \sin^4\beta \quad (\text{Eq. 64})$$

β describes the angle formed between the principal axis of the diffusion tensor and the vector that is relevant for relaxation (such as the inter-nuclei vector for the dipolar relaxation). The a_i coefficients are weighting factors with: $\sum_{i=1}^3 a_i = 1$

and $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$ are composite correlation times given by:

$$\mathcal{T}_1 = \mathcal{T}_L \quad (\text{Eq. 65})$$

$$\mathcal{T}_2 = \frac{6\mathcal{T}_L\mathcal{T}_S}{\mathcal{T}_L + 5\mathcal{T}_S} \quad (\text{Eq. 66})$$

$$\mathcal{T}_3 = \frac{3\mathcal{T}_L\mathcal{T}_S}{2\mathcal{T}_L + \mathcal{T}_S} \quad (\text{Eq. 67})$$

6.3- Model free approach

When the internal motions are considered, the same approach as above is applied with a spectral density function resulting from a weighted sum of terms describing different contribution of motions at different frequencies. Of most importance is that motions described by the different density functions have to be uncorrelated. Realizing that the number of NMR relaxation rates

is too reduced to provide a complete description of both internal and overall motions, Lipari and Szabo suggested a model where internal motions are described by a single effective correlation time, $\mathcal{T}_e$ that has to be small in front of the correlation time describing the overall molecular diffusion. The spectral density function is expressed as a sum of two components describing the overall rotational motion and internal motions weighted by an order parameter S^2 . This parameter provides a measure of the spatial restriction of fast internal motions.¹⁸

When the overall motion can be described by a single correlation time $\mathcal{T}_M$, the spectral density is expressed as follow:

$$j(\omega) = \left(\frac{S^2 \mathcal{T}_M}{1 + \omega^2 \mathcal{T}_M^2} + \frac{(1 - S^2) \mathcal{T}}{1 + \omega^2 \mathcal{T}^2} \right) \quad (Eq. 68)$$

With $\mathcal{T}^{-1} = \mathcal{T}_M^{-1} + \mathcal{T}_e^{-1}$

This formalism allows a description of local disordered regions by the analysis of the order parameter in different regions of a protein, provided that the global rotational diffusion of this protein can be determined from the relaxation parameters measured for rigid parts of the molecule.

Bibliography

- (1) Gurd, F. R. N.; Rothges, T. M. Motions in Proteins. In *Advances in Protein Chemistry*; Elsevier, 1979; Vol. 33, pp 73–165. [https://doi.org/10.1016/S0065-3233\(08\)60459-3](https://doi.org/10.1016/S0065-3233(08)60459-3).
- (2) Kumar, A.; Balbach, J. Real-Time Protein NMR Spectroscopy and Investigation of Assisted Protein Folding. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2015**, 1850 (10), 1965–1972. <https://doi.org/10.1016/j.bbagen.2014.12.003>.
- (3) Abragam, A. *The Principles of Nuclear Magnetism*, Repr.; International series of monographs on physics; Oxford Univ. Pr: Oxford, 2011.
- (4) Wüthrich, K. *NMR in Biological Research: Peptides and Proteins*; North-Holland Pub. Co.; sole distributors for the U.S.A. and Canada, American Elsevier Pub. Co: Amsterdam : New York, 1976.
- (5) Bloch, F.; Hansen, W. W.; Packard, M. The Nuclear Induction Experiment. *Phys. Rev.* **1946**, 70 (7–8), 474–485. <https://doi.org/10.1103/PhysRev.70.474>.
- (6) Bloembergen, N.; Purcell, E. M.; Pound, R. V. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys. Rev.* **1948**, 73 (7), 679–712. <https://doi.org/10.1103/PhysRev.73.679>.
- (7) Lectures by James Keeler, chapter 8 relaxation <http://www-keeler.ch.cam.ac.uk/lectures/>
- (8) Peng, J. W.; Wagner, G. Mapping of the Spectral Densities of Nitrogen-Hydrogen Bond Motions in Eglin c Using Heteronuclear Relaxation Experiments. *Biochemistry* **1992**, 31 (36), 8571–8586. <https://doi.org/10.1021/bi00151a027>.
- (9) Wittebort, R. J.; Szabo, A. Theory of NMR Relaxation in Macromolecules: Restricted Diffusion and Jump Models for Multiple Internal Rotations in Amino Acid Side Chains. *J. Chem. Phys.* **1978**, 69 (4), 1722–1736. <https://doi.org/10.1063/1.436748>.
- (10) Gupta, A.; Stait-Gardner, T.; Moghaddam, M. J.; Price, W. S. Dipolar Relaxation Revisited: A Complete Derivation for the Two Spin Case: DIPOLAR RELAXATION REVISITED. *Concepts Magn. Reson. Part A* **2015**, 44 (2), 74–113. <https://doi.org/10.1002/cmr.a.21334>.
- (11) Solomon, I. Relaxation Processes in a System of Two Spins. *Phys. Rev.* **1955**, 99 (2), 559–565. <https://doi.org/10.1103/PhysRev.99.559>.
- (12) Schanda, P.; Van Melckebeke, H.; Brutscher, B. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.* **2006**, 128 (28), 9042–9043. <https://doi.org/10.1021/ja062025p>.
- (13) Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd ed.; John Wiley & Sons: Chichester, England ; Hoboken, NJ, 2008.
- (14) Lectures by James Keeler, chapter 8 relaxation <http://www-keeler.ch.cam.ac.uk/lectures/>
- (15) Canet, D. On the Calculation of Spectral Density Functions for Spin Interactions without Axial Symmetry. *Concepts Magn. Reson.* **1998**, 10 (5), 291–297. [https://doi.org/10.1002/\(SICI\)1099-0534\(1998\)10:5<291::AID-CMR2>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1099-0534(1998)10:5<291::AID-CMR2>3.0.CO;2-S).
- (16) Lane, A. N.; Lefèvre, J.-F.; Jardetzky, O. A Method for Evaluating Correlation Times for Tumbling and Internal Motion in Macromolecules Using Cross-Relaxation Rate Constants from Proton NMR Spectra. *J. Magn. Reson.* 1969 **1986**, 66 (2), 201–218. [https://doi.org/10.1016/0022-2364\(86\)90024-7](https://doi.org/10.1016/0022-2364(86)90024-7).
- (17) Woessner, D. E. Nuclear Spin Relaxation in Ellipsoids Undergoing Rotational Brownian Motion. *J. Chem. Phys.* **1962**, 37 (3), 647–654. <https://doi.org/10.1063/1.1701390>.
- (18) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.*

1982, 104 (17), 4546–4559. <https://doi.org/10.1021/ja00381a009>.

(19) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd ed.; Methods in stereochemical analysis; Wiley: New York, 2000.

(20) James, T. L. Chapter 1 Fundamentals of NMR. **1998**.

(21) Harris, R. K.; Becker, E. D.; De Menezes, S. M. C.; Granger, P.; Hoffman, R. E.; Zilm, K. W.; Prepared for publication by. Further Conventions for NMR Shielding and Chemical Shifts (IUPAC Recommendations 2008): International Union of Pure and Applied Chemistry Physical and Biophysical Chemistry Division. *Magn. Reson. Chem.* **2008**, 46 (6), 582–598. <https://doi.org/10.1002/mrc.2225>.

(22) Kowalewski, J.; Werbelow, L. Evaluation of Spin Relaxation Induced by Chemical Shielding Anisotropy: A Comment on the Importance of the Antisymmetric Component. *J. Magn. Reson.* **1997**, 128 (2), 144–148. <https://doi.org/10.1006/jmre.1997.1238>.

(23) James - 1998 - Chapter 1 Fundamentals of NMR.Pdf. https://qudev.phys.ethz.ch/static/content/courses/phys4/studentspresentations/nmr/James_Fundamentals_of_NMR.pdf (accessed 2023-03-14).

(24) Andrew Atkinson, R.; Kieffer, B. The Role of Protein Motions in Molecular Recognition: Insights from Heteronuclear NMR Relaxation Measurements. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, 44 (3–4), 141–187. <https://doi.org/10.1016/j.pnmrs.2004.01.001>.

(25) McConnell, H. M. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **1958**, 28 (3), 430–431. <https://doi.org/10.1063/1.1744152>.

(26) Bain, A. D. Chemical Exchange in NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, 43 (3–4), 63–103. <https://doi.org/10.1016/j.pnmrs.2003.08.001>.

(27) Ishima, R.; Torchia, D. A. Estimating the Time Scale of Chemical Exchange of Proteins from Measurements of Transverse Relaxation Rates in Solution. *J. Biomol. NMR* **1999**, 14 (4), 369–372. <https://doi.org/10.1023/A:1008324025406>.

(28) Birchall, A. J.; Lane, A. N. Anisotropic Rotation in Nucleic Acid Fragments: Significance for Determination of Structures from NMR Data. *Eur. Biophys. J.* **1990**, 19 (2). <https://doi.org/10.1007/BF00185089>.

(29) Jarymowycz, V. A.; Stone, M. J. Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chem. Rev.* **2006**, 106 (5), 1624–1671. <https://doi.org/10.1021/cr040421p>.

(30) King, R.; Jardetzky, O. A General Formalism for the Analysis of NMR Relaxation Measurements on Systems with Multiple Degrees of Freedom. *Chem. Phys. Lett.* **1978**, 55 (1), 15–18. [https://doi.org/10.1016/0009-2614\(78\)85121-5](https://doi.org/10.1016/0009-2614(78)85121-5).

(31) Fushman, D.; Cowburn, D. The Effect of Noncollinearity of ^{15}N - ^1H Dipolar and ^{15}N CSA Tensors and Rotational Anisotropy on ^{15}N Relaxation, CSA/Dipolar Cross Correlation, and TROSY.

Chapter 2:

Relaxation of fluorine nucleus

II. Relaxation of fluorine nucleus

Fluorinated molecules in biology can be advantageously used to alter hydrophobicity, lipophilicity, acidity, basicity and conformational stability. They are also useful to recognize binding sites, to modify the response to metabolic transformations, and to modify chemical reactivity.^{1,2} In 1957, the 5-fluorouracil showed anticancer efficacy. Since then, fluorine chemistry emerged as an important tool in drug design with fluorinated small molecules comprising up to 33% of available drugs in 2020.³ Early studies of biological systems using ^{19}F NMR spectroscopy started around 1960s. Among them, we recall the work of Spotswood et al in 1967 to characterize the interaction of the digestive enzyme chymotrypsin with its substrates N-acetyl-DL-p-fluoro-phenylalanine and N-acetyl-DL-m-fluorophenylalanine.⁴

And the work of Griffiths et al in 1984 using fluorine NMR to characterize the *in vivo* metabolism of 5-fluorouracil in implanted tumors and livers in mice.⁵

In the early 2000, the technique became popular to assess protein and ligand binding for drug discovery, leading to various reviews covering its applications.⁶⁻¹⁰ These applications were favored by the unique properties of this nucleus: its sensitivity close to that of the proton, sparse spectra due to the large chemical shift dispersion, its absence in biological molecules and its extreme sensitivity to the chemical environment. Despite, the relaxation properties of fluorine were only marginally exploited in these studies. One possible reason, as explained in chapter 1, is the contribution of ^{19}F CSA to the transverse relaxation rates causing broad signals especially at high magnetic fields. Further, the interpretation of the fluorine relaxation parameters to get insights on molecular motions is hampered by the large contribution of the CSA to the relaxation mechanisms that contrasts with the weak body of reference data and studies on this subject.

In the following, we first enumerate few interesting studies where fluorine relaxation was put to use to characterize biological events. Later we will address some of these properties in the specific case of fluorinated prolines.

1. Fluorine relaxation properties to study biological molecules

A very popular and simple method to identify and characterize ligand protein binding is the Saturation Transfer Difference technique (STD). This method has been adapted for the

observation of fluorinated ligands. The fluorination might be site specific or global depending on the objectives of the study. Since we observe the ligand, no size restriction is applied on the studied protein. In a ^{19}F STD experiment, a weak RF field is selectively applied alternatively on the protein frequencies (usually using methyl resonances) and far from any resonances while the two resulting fluorine signals are compared. The protein on-resonance irradiation leads to a saturation of protein resonances that is transferred on the fluorinated ligand by magnetization exchange (NOE) when the ligand is bound resulting in intensity variations of the fluorine signal. Numerous applications could be found in the literature over the years. A recent interesting application of this approach was the characterization of the volatile anesthetic isoflurane binding to the bacterial voltage-gated sodium channel. Particular serine, leucine and threonine residues on the channel were predicted to interact with the isoflurane from MD simulations and were fluorinated using $-\text{CF}_2$ and $-\text{CF}_3$ moieties. The ^{19}F STD accurately identified those that actually binds to the anesthetic and their binding affinities.¹¹

In 2001, Peng et al developed a method for ranking the affinities of ligands to a particular protein. This method used the transverse cross-correlation rate between fluorine CSA and ^1H - ^{19}F dipolar relaxation mechanisms denoted η_{XY} . This rate depends on the overall molecular tumbling ($J(0)$) but not to the chemical exchange contribution. Thus, the quantification of the η_{XY}/R_2 ratio provides a measure of the exchange contribution to the transverse relaxation rate and subsequently a measure of the dissociation rate. The sensitivity of η_{XY} to the molecular tumbling is exploited to detect binding events.¹²

Large transverse relaxation of fluorine may be exploited in T_2 filter experiments based on the CPMG pulse sequence with common applications for protein-ligand binding screening studies. In presence of the protein, both fluorine CSA and chemical exchange contribute to the transverse relaxation and induce a remarkable intensity loss of the ligand's witnessing its binding. This relaxation effect is exploited by the FAXs methods (fluorine chemical shift anisotropy and exchange for screening) used to follow a binding competition between a weak fluorinated binder designed as a spy molecule and putative ligands of interest against the target. The transverse relaxation rate is well adapted for this kind of applications due to its extreme sensitivity to the molecule's motions. The FAXs experiment was developed by Claudio Dalvit in 2002 for the analysis of ligand binding to the p21-activated kinase. Simple 1D ^{19}F spectra containing the fluorinated ligand PHA-739917 and TFA serving as a control molecule were recorded while increasing amounts of kinase was added. A loss in the signal of ligand PHA-

739917 was observed indicating the complex formation while TFA signal remained invariant. The incorporation of a competing molecule capable of displacing the ligand induced a recovery of the original signal intensity.^{3,13,14}

Fragment based drug discovery FBDD is a widely used approach to determine new therapeutics molecules. It is faster than the traditional binding screening methods since it uses small high binding fragments. In 2011, Jordan et al suggested to implement fluorine NMR to FBDD. The idea was to use fluorinated fragment library for the screening. The detection of ^{19}F – ^{19}F NOE originating from two fragments that bind simultaneously to the target was suggested as a way to reveal the spatial proximity between the binders and the possibility to combine them. As a proof of concept, the group applied this methodology together with the original inter-ligand NOE (INOE) used for ^1H NMR, to screen fluorinated fragments against a drug target implicated in Alzheimer's disease, BACE-1 and were able to identify an efficient ligand.¹⁵

^{19}F paramagnetic relaxation enhancements is used to specifically shorten the longitudinal relaxation rate by the mean of a paramagnetic metal with minimal line broadening consequently enhancing the sensitivity by decreasing the experiment time. Structural information could be derived from ^{19}F PRE by determining long range distances in large proteins when NOEs are limited.¹⁶ A recent example used the dependence of the of longitudinal relaxation of fluorine in human glutamate transporter, on the distance to its neighboring Ni^{2+} ions to identify different interconverting states of the transporter and their interconversion rates. The transporter was labelled with trifluoroethanethiol and its corresponding 1D ^{19}F spectrum revealed the coexistence of three distinct conformations that were assigned based on the ^{19}F - Ni^{2+} distances estimated from R_1 PRE measurements.¹⁷

Even though the ligands affinities to the target molecule can be estimated from FAXs, it is sometimes challenging for weakly binding ligands. A new methodology called chemical shift anisotropy ranking (CSAR) is used to quantify the binding affinity. For this method, the CSA mechanism should be the major contributor to the transverse relaxation rate, hence both the dipolar and the chemical exchange are suppressed. For a given molecule, the contribution of CSA to the transverse relaxation is quantified through R_2 measurement at two different magnetic fields while the dipolar and exchange contributions are quantified through a spin lock at high magnetic fields. In this way, the measured relaxation rate is only dependent on the bound fraction of the ligand and estimation of this latter leads to the determination of the dissociation

constant. This method relies however on the challenges accompanying the calculation of the chemical shielding tensors to quantify the CSA contribution.¹⁸

Fluorine NMR is gaining attention as a tool to study bindings of biological molecules of interest. However, in literature, the description of fluorine relaxation is limited to few studies.¹³ In my work, we are interested to the use of 4-fluorinated prolines to report the motion of biological systems at the atomic level. Despite their great potential, fluoroprolines remained largely unstudied compared to other fluorinated amino acids. It is then crucial to characterize the relaxation of fluorine in free *4R* and *4S* fluoroprolines before interrogating more complex systems. In the following paragraph, we intend to follow the relaxation of the fluorine in the free *4R* and *4S* fluoroprolines through the simulation of its relaxation parameters R_1 , σ_{FH} and R_2 introduced in the first chapter.

2. Fluorine in 4-fluorinated prolines

4-fluoroprolines first appeared in 1965. At the early stages they gain their popularities because they could be incorporated into collagen protein and were then largely used to study collagen hydroxylation,^{19,20} and collagen structural stability.^{21,22} Incorporation of a fluorine atom on the $C\gamma$ results in the formation of a σ_{C-F}^* electron acceptor orbital. When the Fluorine atom and the carbon of the carbonyl group are on the opposite sides (in *4R* fluoroproline), the hyperconjugation interaction $\sigma_{C-H}-\sigma_{C-F}^*$ takes place when the ring is in its *exo* pucker, alternatively, when the a fluorine atom is on the same side of the carbon of the carbonyl group (in *4S* fluoroproline) the interaction favors the *endo* pucker (Figure1).²³ The ratio $C\gamma$ *endo*/ $C\gamma$ *exo* in the *4S* fluoroproline is 19:1 while it is 1:6 in *4R* fluoroproline.²⁴ Consequently, *4R* and *4S* fluoroprolines are said to bias the *endo/exo* equilibrium. As explained in the introduction, the ring pucker of the proline modulates the *cis/trans* isomerization. $n - \pi^*$ interaction takes place due to the proximity of the successive carbonyls within *4R* *exo*, which stabilizes the *trans* conformer. In *4S* *endo* ring pucker, the longer distance between the successive carbonyls weakens the $n - \pi^*$ interaction and stabilizes the *cis* conformer (figure 2).^{25,26}

The equilibrium constant of the *cis* to *trans* transition k_{ct} defined as $k_{ct} = \frac{[trans]}{[cis]}$ where $[trans]$ and $[cis]$ are the equilibrium concentrations of the *trans* and *cis* conformers respectively is altered when a proton of the five membered prolyl cycle is substituted by a fluorine atom. Compared to a non-modified proline residue, *4R* fluoroprolines favoring *exo* ring

pucker has higher k_{ct} whereas $4S$ fluoroproline favoring *endo* ring pucker has lower k_{ct} (Table1). Additionally, the incorporation of an electron withdrawing atom further influences the *cis trans* isomerization kinetics within proline amino acid. The double bond character of the amide bond is reduced and the rotation around the amide bond is enhanced. Compared to a non-modified proline, the *cis* to *trans* rotation velocity K_{ex} is higher within $4R$ fluoroproline, followed by that in $4S$ fluoroproline then the non-modified proline (Table1). Both thermodynamics equilibrium and kinetics of *cis trans* isomerization in fluoroproline correlate with conformational stability of proteins such as the collagen triple helix, the ubiquitin and the β 2-microglobulin as deduced from their melting temperatures and folding rates.^{27–30}

Table 1: Thermodynamic and kinetic data of N- acetylfluoroproline methyl ester derivatives from literature.

	k_{ct}	$K_{ex} (s^{-1})$
Ac-($4R$)-FPro-OMe	$6.7 \pm 5\%^{30}$	0.074
Ac-($4S$)-FPro-OMe	$2.5 \pm 5\%^{30}$	0.052
Ac-Pro-OMe	$4.6 \pm 5\%^{30}$	0.038

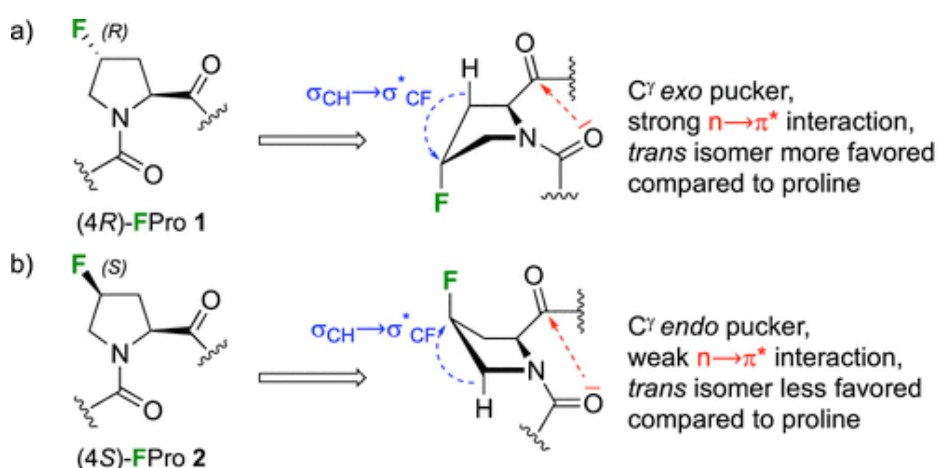


Figure 1: Illustration of the conformational bias in 4-fluorinated proline originating from different electronic interactions in the *endo* and the *exo* conformers in ($4R$) FPRO (a) and ($4S$) FPRO (b). σ_{CF}^* is the electron acceptor orbital and σ_{CH} is the electron donor orbital. The

interaction between these orbitals is favored in the *exo* pucker. $n-\pi^*$ interaction is stronger in the *4R* *exo* pucker and stabilizes the *trans* conformer. Reproduced from ³¹

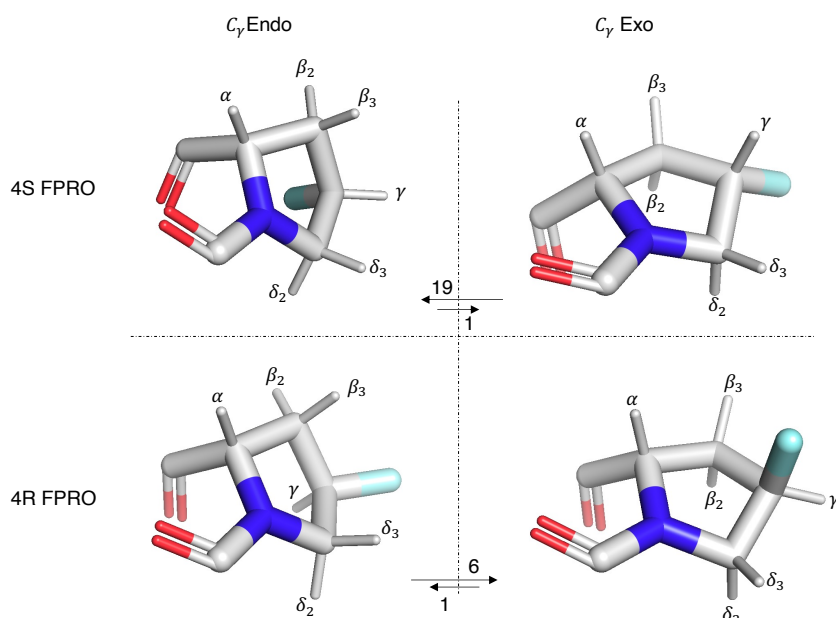


Figure 2: *4R* and *4S* fluoroprolines arranged into $c_\gamma endo$ and $c_\gamma exo$ puckers. $H_{\gamma 2}$ and $H_{\gamma 3}$ are substituted by a fluorine atom in *4R* and *4S* FPRO respectively. The fluorine atom is shown in cyan; it points towards the same direction of the carbonyl group in *4S* FPRO and towards the opposite direction in *4R* FPRO. The numbers represent the ratio $c_\gamma endo/c_\gamma exo$ the long arrows point towards the stabilized conformer of each fluoroproline. The models are from DFT calculations performed by Emile Ottoy.²⁴

3. Relaxation of fluorine in 4-fluorine substituted prolines

The fluoroproline models that we use are issued of DFT calculations.²⁴ For each fluoroproline we considered the two-ring pucker *endo* and *exo*. In the text we will refer to *4R* FPRO $c_\gamma endo$ and *4S* FPRO $c_\gamma exo$ as the minor conformers since they are less preferred. Compared to the models Ac-Pro-NHMe calculated by Kang and Choi, the ϕ dihedral angle of the fluorinated prolines are smaller than the average values of ϕ for local minima ($\sim 20^\circ$ difference for *4R* FPRO and $\sim 10^\circ$ difference for *4S* FPRO).⁴³

The internuclear distances between the fluorine and its neighboring protons in *4R* and *4S* FPRO conformers are reported on (Table 2). We notice that the F – H_γ distance is almost invariant in all conformers. H_δ and H_β protons are closer to the fluorine atom within the minor conformers

(except for F – H_{δ2} in *4S c_γexo*), whereas H_α is further away from the fluorine atom in the minor conformers.

Table 2: ¹⁹F-¹H internuclear distances within *4R* (*c_γendo* and *c_γexo*) and *4S* fluoroprolines (*c_γendo* and *c_γexo*) from Gaussian calculations in Ac-(*4R/4S*)-FPRO-NMe₂.

Distance (Å)	<i>4R</i> FPRO		<i>4S</i> FPRO	
	<i>c_γexo</i>	<i>c_γendo</i>	<i>c_γexo</i>	<i>c_γendo</i>
F – H _γ	2.02	2.02	2.02	2.01
F – H _{δ2}	3.31	2.96	2.52	2.45
F – H _{δ3}	2.53	2.44	2.88	3.27
F – H _{β2}	3.3	2.89	2.49	2.54
F – H _{β3}	2.59	2.5	2.88	3.3
F – H _α	2.86	4.39	4.28	3.98

3.1. Fluorine longitudinal relaxation rate

When the dipolar relaxation and the relaxation through the chemical shift anisotropies are the only relaxation mechanisms considered, the fluorine longitudinal relaxation rate in a two spins system composed of a fluorine and a proton is written as follows:

$$R_1^F = \frac{1}{20} b^2 (j(\omega_F - \omega_H) + 3j(\omega_F) + 6j(\omega_F + \omega_H)) + \left(\frac{1}{15}\right) \gamma^2 B_0^2 \Delta\sigma^2 \left(1 + \frac{\eta^2}{3}\right) j(\omega_F) \quad (\text{Eq. 1})$$

Where $b^2 = \frac{\mu_0}{4\pi} \frac{\hbar \gamma_F \gamma_H}{r_{FH}^3}$, r_{FH} is the internal distance between the fluorine and the proton, $\Delta\sigma$ is the fluorine shielding anisotropy, η is the asymmetry parameter, γ_F and γ_H are the fluorine and proton gyromagnetic ratios respectively and ω_F and ω_H represent their Larmor frequencies.

The evolution of the fluorine longitudinal relaxation rate with the correlation time, in a *4R exo* fluoroproline as given by equation (1) is shown on figure 3a. It is characterized by the presence of two maxima at 0.28 and 4 ns. The plot of each relaxing mechanism separately shows that the maximum at 4 ns corresponds to the sole contribution of the dipolar mechanism while the maximum at 0.28 ns corresponds to both the CSA and the dipolar interaction's mechanisms.

As stated by equation (1), the relaxation through chemical shift anisotropy is proportional to the square of the operating magnetic field. To visualize the effect of the CSA contribution to the longitudinal relaxation rate, we calculated the fluorine R_1 at 600 MHz and 400 MHz (Figure 3b). The large chemical shift anisotropy of the fluorine nucleus leads to a first maximum of the relaxation rate at the fluorine frequency that is mostly dominated by the CSA contribution at 600 MHz. At lower field (400 MHz), the dipolar relaxation becomes predominant for correlation times in the range of tens of nano-seconds due to the contribution of the spectral density at the frequency $(\omega_F - \omega_H)$.

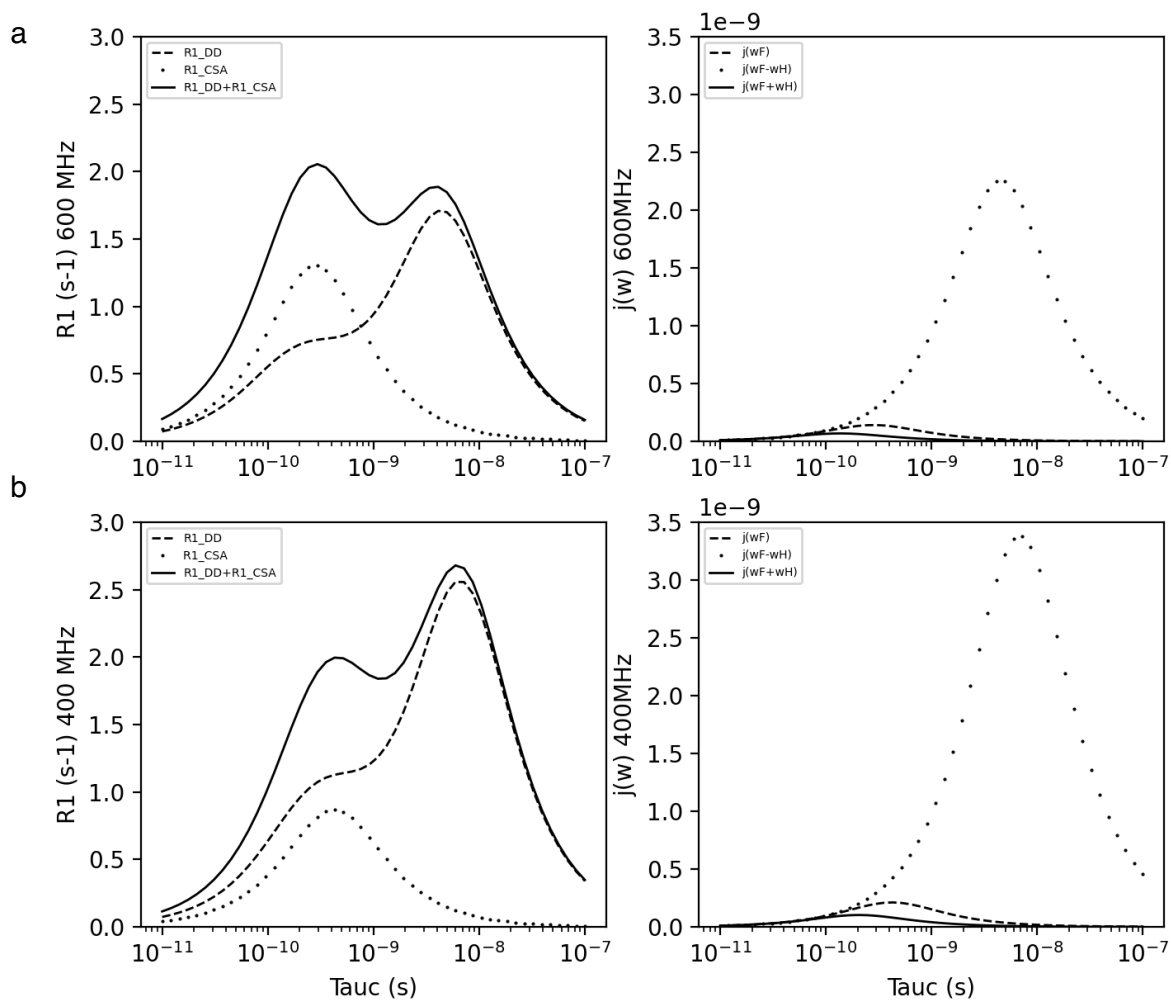


Figure 3: Left: Simulation of the fluorine longitudinal relaxation rate as expressed in Eq.1 in a fluoroproline *4R exo* at 600 MHz (a) and 400 MHz (b). r_{FH} distance is 2.02 Å. Right: Calculated spectral densities at each magnetic field as a function of the correlation time. $\Delta\sigma$ and η were set to -74.2 ppm and 0.12 respectively.

As stated by Table 2, the H_γ is the closest spin to the fluorine in a fluoroproline and provides a major contribution to the dipolar relaxation. However, one should not neglect the other neighboring spins when interpreting the relaxation parameters as shown by figure 4. Thus, the interpretation of the R_1 to extract information about the proline's ring dynamics should take into account both the global tumbling as well as internal ring motions that are susceptible to modulate proton-fluorine distances.

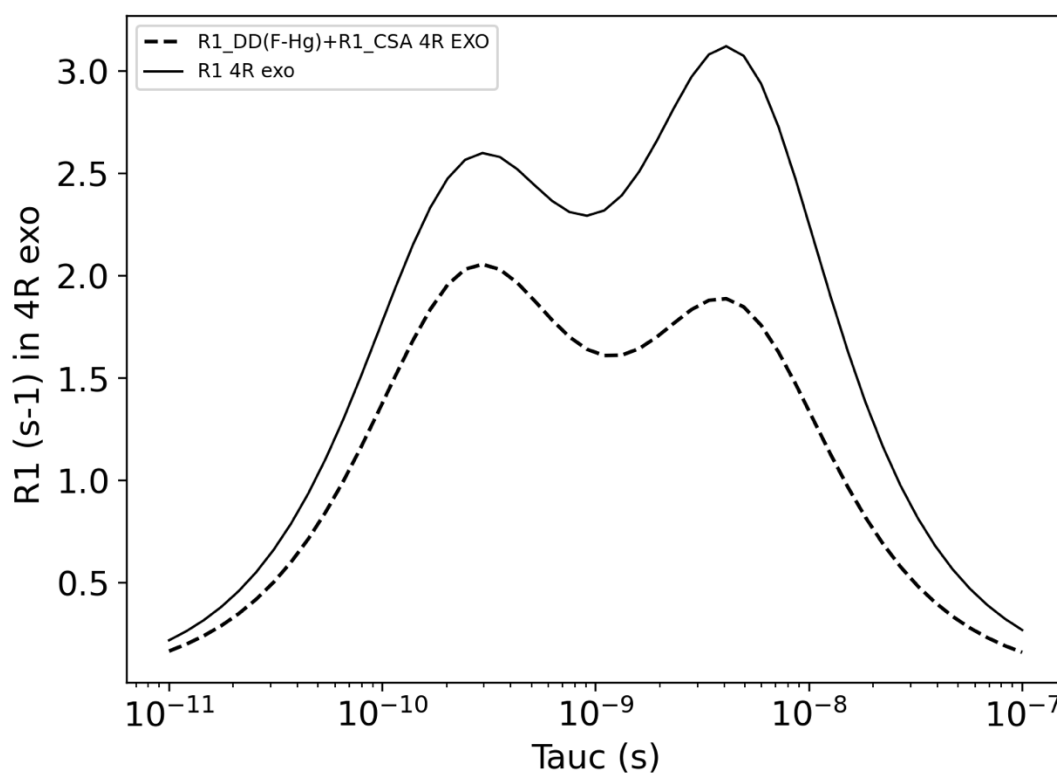


Figure 4: Fluorine longitudinal relaxation rate in a fluoroproline *4R exo* when the H_γ is the only relaxing spin (dashed line) and when the contribution of all neighboring protons is considered. Calculated from Eq.1 at 600 MHz. $\Delta\sigma$ and η were set to -74.2 ppm and 0.12 respectively.

In collaboration with Ilya Kuprov, the fluorine longitudinal relaxation rates in 4R and 4*S* fluoroprolines were calculated using spinach 2.6 software containing the relaxation theory derived from Bloch-Redfield-Wangsness theory.³⁶ Fluorine longitudinal relaxation rates were

calculated for correlation times varying from 10ps to 100 ns at 14.1 T. The chemical shift tensors of the major and the minor conformations were derived from DFT calculations performed by Emile Ottoy (Davy Sinnaeve group) and are displayed on table 3. We compared the calculated fluorine longitudinal relaxation rate using equation (1) (with all relaxing spins considered) to that derived from Spinach calculation (Figure 5). The perfectly identical curves within *4R* *exo* and *endo* conformers reveal that the fluorine relaxation in *4R* fluoroproline fully originates from the synergy of the dipolar and CSA relaxation mechanisms. Within *4S* *exo*, the simple model successfully predicts the fluorine longitudinal relaxation, on the other side, within *4S* *endo* conformer, this latter is ~5% underestimated.

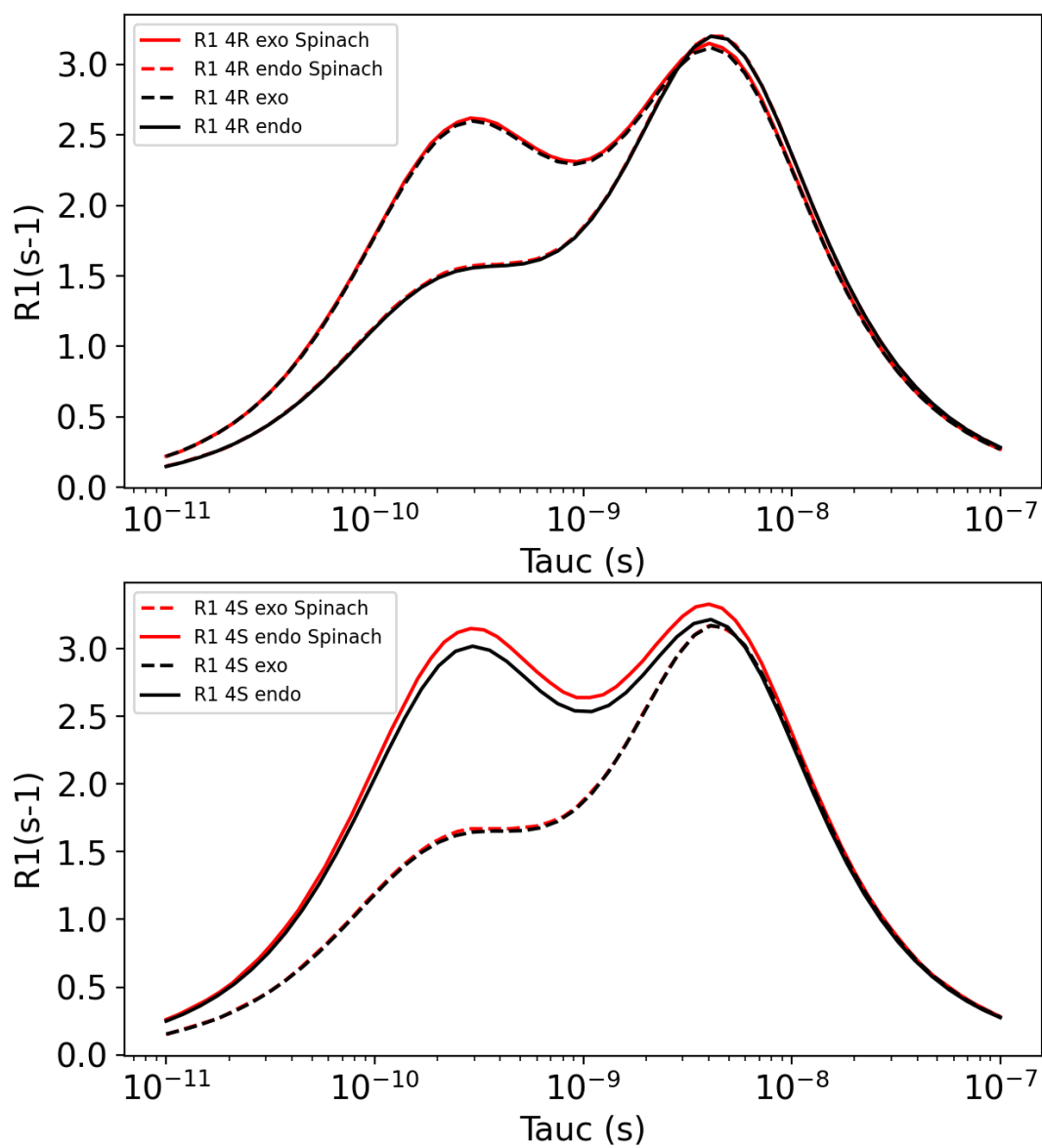


Figure 5: Fluorine longitudinal relaxation rate within *4R* and *4S* fluoroprolines calculated using Spinach software (red) and simplified equation Eq.1 (black). The fluorine relaxation in major and minor conformers is shown with solid and dashed lines respectively. Calculated at 600 MHz. The ^{19}F chemical shift tensor parameters are given in table 3.

Table 3: ^{19}F CSA tensor parameters for the major and minor conformers of *4R* and *4S* fluoroprolines.

	<i>4R</i> FPRO		<i>4S</i> FPRO	
	c_γ <i>exo</i>	c_γ <i>endo</i>	c_γ <i>exo</i>	c_γ <i>endo</i>
$\Delta\sigma$ (ppm)	-74.2	-25.6	-33.3	-84.9
η	0.12	0.396	0.483	0.392

NB: In literature several chemical shift tensors of fluorinated amino acids were published (Appendix 1) except for those of fluoroprolines. To our knowledge, our paper published in Magnetic resonance in collaboration with Davy Sinnaeve, is the first reference to bring up the topic. ³²

3.2. Fluorine transverse relaxation rate:

The fluorine transverse relaxation rate originating from the dipolar and the CSA relaxation mechanisms is given by equation (2):

$$R_2^F = \frac{1}{20} b^2 (3j(\omega_F + \omega_H) + \frac{3}{2}j(\omega_F) + 3j(\omega_H) + \frac{1}{2}j(\omega_F - \omega_H) + 2j(0)) + \left(\frac{1}{15}\right) \gamma^2 B_0^2 \Delta\sigma^2 \left(1 + \frac{\eta^2}{3}\right) \left(\frac{2}{3}j(0) + \frac{1}{2}j(\omega_F)\right) \quad (Eq. 2)$$

The calculated fluorine transverse relaxations rates within *4R* *exo* and *4S* *endo* fluoroprolines using Spinach software are shown on figure 6. Between 0.01ns and 100ns, the transverse relaxation displays the expected increase with the correlation time with a very small difference between the two fluoroprolines. The contribution of the CSA to the fluorine transverse relaxation in *4R* *exo* is exactly 2.25 more effective at 600 MHz than at 400 MHz, as stated by equation (2). The dipolar contribution to the transversal relaxation is invariant upon the strength of the magnetic field, yet a slight variation is observed from 0.12 ns to 1.2 ns where the dipolar relaxation at 400 MHz is higher than that at 600 MHz (Figure 7).

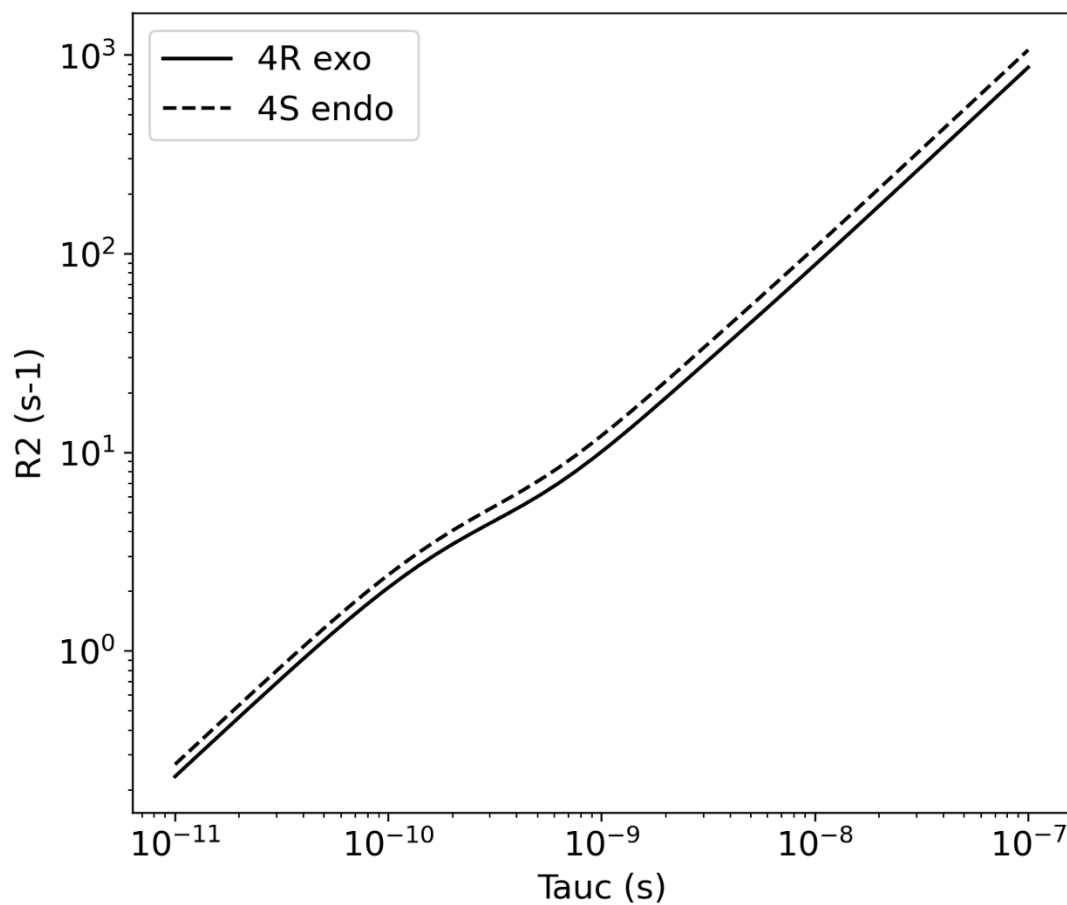


Figure 6: Variation of the Fluorine transverse relaxation rates with the correlation time within *4R exo* and *4S endo* fluoroprolines calculated from Eq.2 at 600 MHz. Chemical shift tensor parameters are given in table 3.

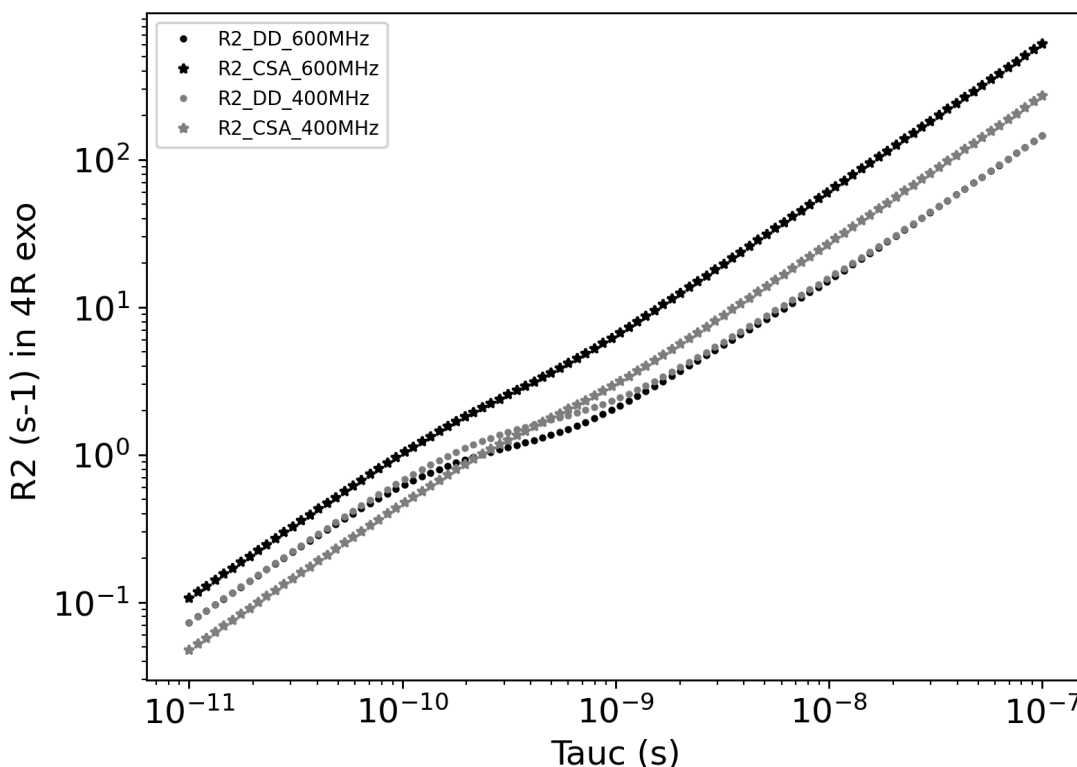


Figure 7: Visualization of the contribution of the CSA to the transverse relaxation rate at 600 MHz and 400 MHz within *4R exo* fluoroproline.

The interpretation of the transverse relaxation rate in terms of proteins motion is challenging. This is because the chemical shift anisotropy of the fluorine is not determined experimentally, consequently its contribution to the transverse relaxation rate cannot be accurately determined. The interpretation is even more challenging, when R_2 is hampered by chemical exchange. Additionally, when fluorine R_1 is chosen for a dynamical study, one should be very careful due to its particular behavior with τ_c . Therefore, an interesting relaxation parameter where nor the CSA nor the chemical exchange are problematic is the nuclear Overhauser effect originating from pure dipolar interactions. In the next paragraph we introduce the steady state and the transient NOE.

3.2.1. ^1H - ^{19}F cross-relaxation rate constant

The cross-relaxation rate is purely dipolar, invariant with the molecular orientation and only depends on the distance separating the interacting spins (Equation 3). In 4-fluoroproline, it is positive for small correlation times (up to 0.3 ns) and negative within slowly tumbling

molecules. Within *4S* and *4R* fluoroprolines, the σ_{FH_γ} is the most important cross-relaxation process (~ 3 times more effective than σ_{FH_β} and σ_{FH_δ}). Within both fluoroprolines, the magnetization is first transferred to the H_β and the H_δ pointing in the same direction as the fluorine atom and is later transferred to the remaining protons (Figure 8).

$$\sigma_{FH} = \frac{1}{20} b^2 [6j(\omega_F + \omega_H) - j(\omega_F - \omega_H)] \quad (Eq. 3)$$

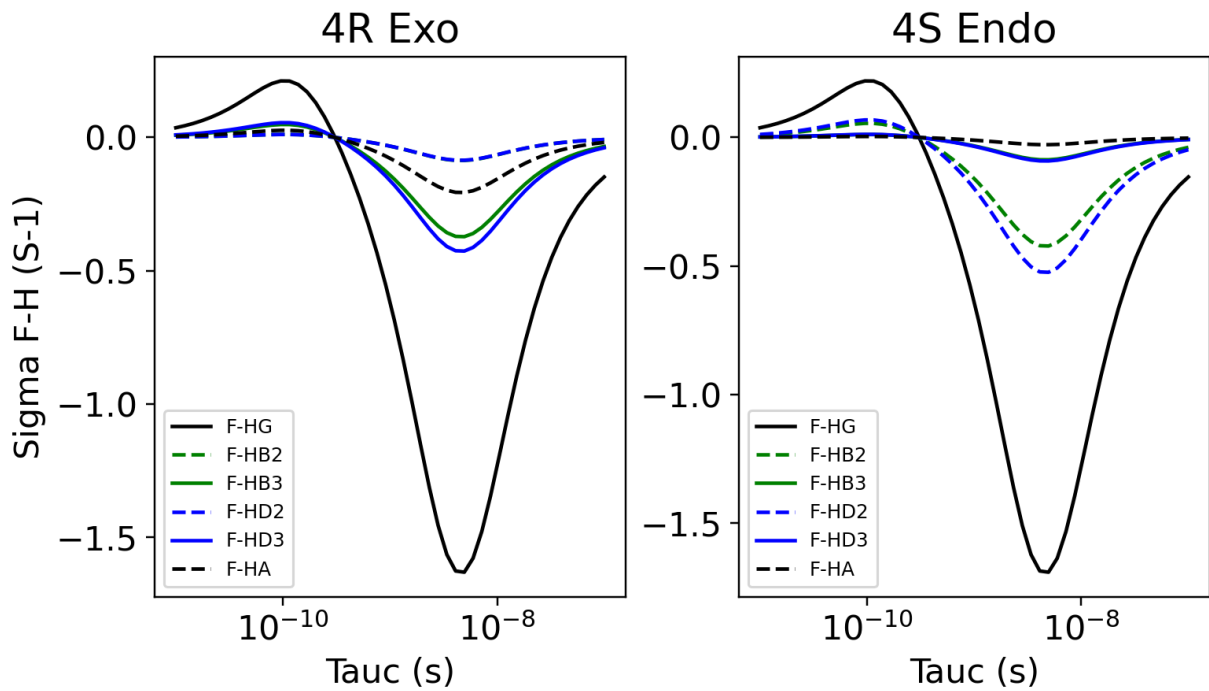


Figure 8: Calculated F-H cross relaxation rates in *4R exo* and *4S endo* fluoroprolines from Eq.3 at 600 MHz.

3.2.2. Heteronuclear ^1H - ^{19}F NOE

The nuclear Overhauser enhancement (NOE) describes the perturbation of the magnetization of the fluorine spin F when that of the proton H deviates from its thermal equilibrium. This perturbation depends on the magnitude of this deviation and on the cross-relaxation rate constant σ_{FH} , as stated by equation (3). The NOE is directly observable on NMR spectrum by a variation of the signal intensities upon perturbation of the magnetization of a neighboring nucleus. The NOE enhancement of the observed spin of interest, fluorine in our case, denoted η_F is then given by equation (4).

$$\eta_F = \frac{(M_F^Z - M_F^{Ze q})}{M_F^{Ze q}} \quad (\text{Eq. 4})$$

Where $M_F^{Ze q}$ is the magnetization of the fluorine at equilibrium.

The modification of the intensity can be both positive or negative depending on the sign of the cross-relaxation rate. In the case of a negative cross-relaxation, the magnetization of spin F drops upon saturation of spin H leading to a negative NOE. This is typically observed for large molecules in viscous solvents. When the cross-relaxation rate is positive the signal of spin F is enhanced and the NOE is said to be positive. This situation is witnessed for small molecules in non-viscous solvents.^{33 34 42}

3.2.2.1. Steady state NOE

A modification of the magnetization of spin F could be observed if the magnetization of spin H to which it is dipolarly coupled, is saturated for instance by the of a selective radiofrequency pulse during a time t. NOE build-up curve is established when the saturation time t is varied until it reaches a maximum value denoted $\text{NOE}_{\max}$.

The expression of the steady-state heteronuclear ^1H - ^{19}F NOE can be established from the Solomon equation:

$$\frac{d}{dt} M_F^Z = -\rho_F (M_F^Z - M_F^{Ze q}) - \sigma_{FH} (M_H^Z - M_H^{Ze q}) \quad (\text{Eq. 5})$$

Given that $M_H^Z = 0$ because it is saturated, and $\frac{d}{dt} M_F^Z = 0$

We can write:

$$0 = -\rho_F (M_F^Z - M_F^{Ze q}) + \sigma_{FH} M_H^{Ze q}$$

Where $M_H^{Ze q}$ is the magnetization of the proton at equilibrium, and $M_H^{Ze q} = M_F^{Ze q} \frac{\gamma_H}{\gamma_F}$

With γ_H and γ_F being the gyromagnetic ratios of the proton and the fluorine respectively.

Thus:

$$\text{NOE}_{\max} = \eta_F = \frac{(M_F^Z - M_F^{Ze q})}{M_F^{Ze q}} = \frac{\gamma_H \sigma_{FH}}{\gamma_F \rho_F} \quad (\text{Eq. 6})$$

The dependance of the NOE_{max} of the H-F spin pairs inside $4R$ and $4S$ fluoroproline conformers on the correlation time given in figure 9. $^1H_\gamma$ - ^{19}F heteronuclear NOE is the most effective pathway of the magnetization transfer due to the smaller distance between H_γ and ^{19}F . Except for $^1H_{\delta 2}$ - ^{19}F NOE, all the $^1H_\beta$ - ^{19}F and $^1H_\delta$ - ^{19}F NOE are higher within the minor conformers ($4R$ *endo* and $4S$ *exo*) due to the proximity of the interacting spins. Within the minor conformers, H_α is located further away from the fluorine spin, consequently, its corresponding NOEs are enhanced within the major conformers. $^1H_\alpha$ - ^{19}F NOE is only important within $4R$ *exo* conformer, and is close to zero in the other conformers. The comparison between the $^1H_\gamma$ - ^{19}F heteronuclear NOE within $4R$ *exo* and $4S$ *endo* (Figure 10) reveals that the two conformers are only distinguishable starting from $\tau_c = 0.25$ ns, at higher correlation times, they display nearly identical NOEs.

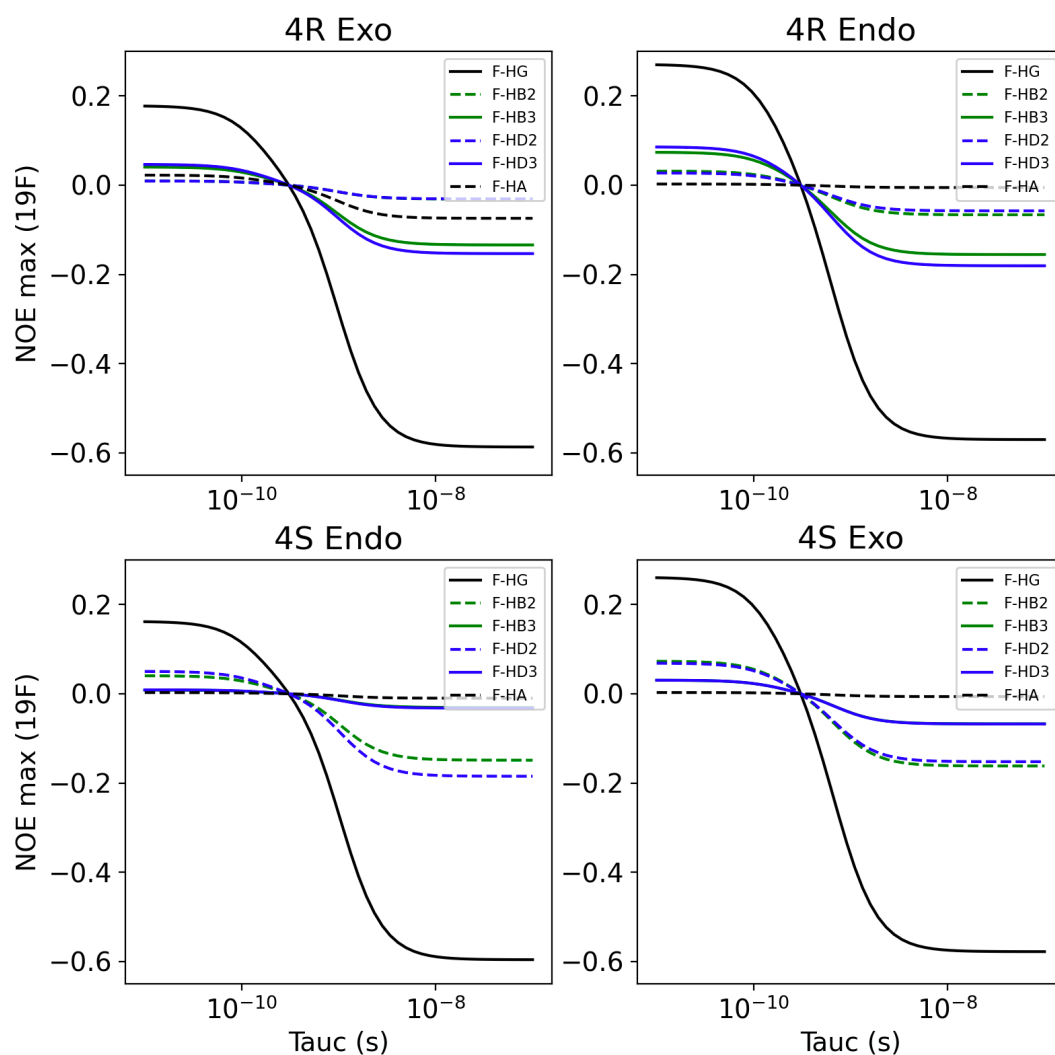


Figure 9: Calculated steady-state 1H - ^{19}F heteronuclear NOEs at 600 MHz within major and minor conformers of $4R$ and $4S$ fluoroprolines using (Eq.5).

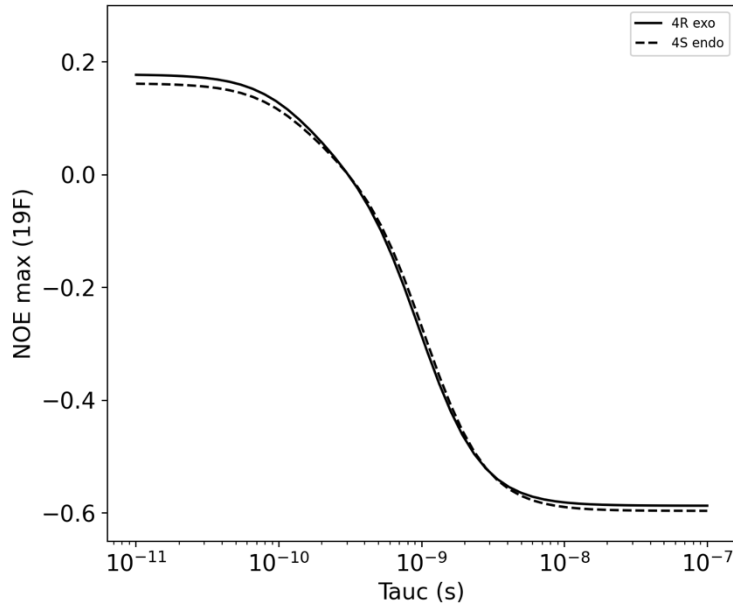


Figure 10: Maximum heteronuclear steady-state $^1H_\gamma$ - ^{19}F NOE within *4S endo* and *4R exo* at 600 MHz.

3.2.2.2. Transient NOE

In a transient NOE experiment, the perturbation of the thermal equilibrium of spin H is achieved by the mean of a short perturbation pulse such as a selective π pulse or the pair of $\frac{\pi}{2}$ pulses in a 2D HOESY experiment. During the mixing time τ_m , both H and F spins are exchanging magnetization through the cross-relaxation rate while relaxing to thermal equilibrium with their auto-relaxation rates. For instance, after a selective inversion of spin F, we have at time $\tau_m = 0$:

$$M_F^z(0) = M_F^{zeq} \text{ and } M_H^z(0) = -M_H^{zeq}$$

And equation 5 leads to

$$\frac{d}{dt} M_F^z(t) = 2\sigma_{FH} M_F^{zeq} \quad (\text{Eq. 7})$$

At short mixing times, equation 7 can be approximated to the first order:

$$M_F^z(\tau_m) = 2\sigma_{FH} M_H^{zeq} \tau_m + M_F^{zeq} \quad (\text{Eq. 8})$$

And:

$$\eta_F(\tau_m) = \frac{(M_F^z(\tau_m) - M_F^{zeq})}{M_F^{zeq}} = \frac{2\sigma_{FH} M_H^{zeq} \tau_m}{M_F^{zeq}} \quad (\text{Eq. 9})$$

Equation 8 implies that at short mixing time, the NOE enhancement varies linearly consequently, the cross-relaxation rate σ_{FH} can be extracted by plotting the NOE as a function of the mixing time.

However, the evolution at longer mixing time should take into account the deviations of M_F^z and M_H^z from equilibrium and their subsequent first order relaxation and the set of coupled differential equations should be solved:

$$\begin{aligned}\frac{d}{dt}M_F^z &= -\rho_F(M_F^z - M_F^{zeq}) - \sigma_{FH}(M_H^z - M_H^{zeq}) \\ \frac{d}{dt}M_H^z &= -\rho_H(M_H^z - M_H^{zeq}) - \sigma_{FH}(M_F^z - M_F^{zeq})\end{aligned}$$

The system can be rewritten using a matrix formalism as:

$$\frac{d}{dt}\vec{M} = -R \vec{\Delta M} \quad (Eq. 10)$$

$\vec{M}$ is the vector of magnetizations while $\vec{\Delta M}$ is the vector of deviations from equilibrium values and R is the relaxation matrix that corresponds to $\begin{pmatrix} \rho_F & \sigma_{FH} \\ \sigma_{FH} & \rho_H \end{pmatrix}$ for a two spins system composed of a proton and a fluorine. The solution of equation 10 involves the exponential of the relaxation matrix multiplied by the mixing time:

$$A(\tau_m) = e^{-(R\tau_m)} M_0 \quad (Eq. 11)$$

Where A is a (NxN) matrix of magnetizations (in the diagonal) and magnetization exchange out of the diagonal reflecting diagonal and cross-peak intensities measured in a NOESY or a HOESY experiment. Equation 11 is usually computed using the standard linear algebra toolbox provided in Matlab or Scipy toolboxes.

An experimental build-up NOE curve as a function of the mixing time extracted from 2D HOESY experiment recorded in 4R fluoroproline in a polyproline rich sequence context is displayed on figure 11.

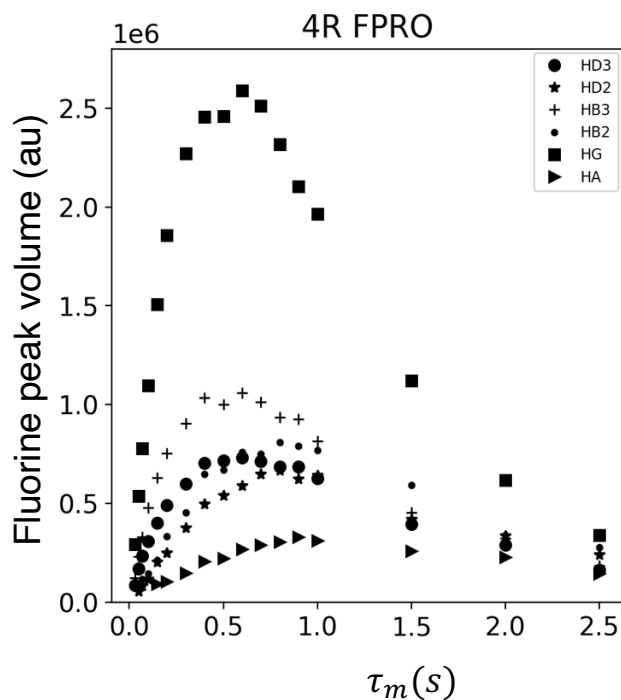


Figure 11: ^1H - ^{19}F heteronuclear NOE within *4R* fluoroproline in polypyrroline model peptide. Recorded at 288 K at 600 MHz in a phosphate buffer 20 mM pH 7.2.

4. Conclusion

In view of the agreement between the calculated relaxation parameters in Spinach software and using the derived simplified equations we conclude that, adding up the dipolar mechanism and ^{19}F CSA is sufficient to explain the behavior of the relaxation parameters in *4R* and *4S* fluoroprolines. There are a number of challenges involved in interpreting longitudinal and transverse relaxation rates in terms of motion due to the ^{19}F CSA contribution, especially when magnetic fields are high, and in the presence of exchange mechanisms that broaden the signals. In contrast, NOE solely depends on the distance separating the interacting spins and is exempt from the exchange and ^{19}F CSA contributions, therefore, it provides the most unambiguous interpretation of the motion in terms of correlation time. Nevertheless, this latter is deduced from $j(0)$, in our case, we can only provide a good estimation of τ_c since we dispose of $j(\omega_F - \omega_H)$ that is close to zero. In this chapter, we gave a close look to the relaxation of 4-fluorinated prolines through simulations of R_1 , R_2 , steady state and transient heteronuclear H-F NOE behavior with the correlation time. In the following chapters, we intend to determine the τ_c of free fluoroprolines and peptides by the interpretation of experimentally measured relaxation parameters.

Appendix 1

Published ^{19}F CSA tensor parameters

system	$^{19}\text{FCSA}$
CH_3F^{35}	$\Delta\sigma = -66 \text{ ppm} \pm 8$ $\Delta\sigma = \sigma \parallel -\sigma \perp$
$\text{C}_6\text{F}_6^{36}$	$\Delta\sigma = -154 \text{ ppm}$
CHFC12^{36}	$\Delta\sigma = -210 \text{ ppm} \pm 30$
1,2,4-trifluorobenzene, $\text{CFC1}_2\text{CFC1}_2$ and $\text{CF}_2\text{BrCF}_2\text{Br}^{37}$	$\Delta\sigma = -250 \text{ ppm} \pm 0.5$
Trifluoroacetic acid (CF_3COOH) 38	$\Delta\sigma = 235 \text{ ppm} \pm 3$
C_6F_6 in 4,4'-di-n-hexyloxyazoxybenzene 39	$\Delta\sigma = 158 \text{ ppm} \pm 4$ $\Delta\sigma = \sigma_{ZZ} - \frac{(\sigma_{XX} + \sigma_{YY})}{2}$
1,2-C ₆ H ₄ F ₂ in 4-hexanoyloxy-4'-ethoxy-azobenzene. 39	$\Delta\sigma = 117 \text{ ppm} \pm 5$
4F-Phe 40	$\zeta = \sigma_{ZZ} - \sigma_{iso}$ and $\eta = \frac{\sigma_{YY} - \sigma_{XX}}{\zeta}$ $\zeta = 58.9 \text{ ppm}$ $\eta = 0.74$
3F-Phe 40	$\zeta = 54.1 \text{ ppm}$ $\eta = 0.76$
3F-Tyr 40	$\zeta = -74.6 \text{ ppm}$ $\eta = 0.48$
5F-L-Trp 40	$\zeta = 51.2 \text{ ppm}$ $\eta = 0.04$
6F-L-Trp 40	$\zeta = 52.1 \text{ ppm}$ $\eta = 0.67$
3F-Ala 41	$\zeta = 33 \text{ ppm}$ $\eta = 0.84$
5F-Leu (2S,4S) 41	$\zeta = -19 \text{ ppm}$ $\eta = 0.68$
5F-Leu (2S,4R) 41	$\zeta = 10 \text{ ppm}$ $\eta = 0.72$
3F-Val 41	$\zeta = -34.6 \text{ ppm}$ $\eta = 0.4$

Bibliography

- (1) Buer, B. C.; Marsh, E. N. G. Fluorine: A New Element in Protein Design. *Protein Sci.* **2012**, *21* (4), 453–462. <https://doi.org/10.1002/pro.2030>.
- (2) Monkovic, J. M.; Gibson, H.; Sun, J. W.; Montclare, J. K. Fluorinated Protein and Peptide Materials for Biomedical Applications. *Pharmaceuticals* **2022**, *15* (10), 1201. <https://doi.org/10.3390/ph15101201>.
- (3) Buchholz, C. R.; Pomerantz, W. C. K. ¹⁹F NMR Viewed through Two Different Lenses: Ligand-Observed and Protein-Observed ¹⁹F NMR Applications for Fragment-Based Drug Discovery. *RSC Chem. Biol.* **2021**, *2* (5), 1312–1330. <https://doi.org/10.1039/D1CB00085C>.
- (4) Spotswood, T. M.; Evans, J. Mary.; Richards, J. Hall. Enzyme-Substrate Interaction by Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* **1967**, *89* (19), 5052–5054. <https://doi.org/10.1021/ja00995a047>.
- (5) Stevens, A. N.; Morris, P. G.; Iles, R. A.; Sheldon, P. W.; Griffiths, J. R. 5-Flourouracil Metabolism Monitored in Vivo by ¹⁹F NMR. *Br. J. Cancer* **1984**, *50* (1), 113–117. <https://doi.org/10.1038/bjc.1984.146>.
- (6) Gerig, J. T. Fluorine NMR of Proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26*, 293–370. [https://doi.org/10.1016/0079-6565\(94\)80009-X](https://doi.org/10.1016/0079-6565(94)80009-X).
- (7) Dalvit, C. Ligand- and Substrate-Based ¹⁹F NMR Screening: Principles and Applications to Drug Discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51* (4), 243–271. <https://doi.org/10.1016/j.pnmrs.2007.07.002>.
- (8) Gerig, J. T. [1] Fluorine Nuclear Magnetic Resonance of Fluorinated Ligands. In *Methods in Enzymology*; Elsevier, 1989; Vol. 177, pp 3–23. [https://doi.org/10.1016/0076-6879\(89\)77003-8](https://doi.org/10.1016/0076-6879(89)77003-8).
- (9) Gimenez, D.; Phelan, A.; Murphy, C. D.; Cobb, S. L. ¹⁹F NMR as a Tool in Chemical Biology. *Beilstein J. Org. Chem.* **2021**, *17*, 293–318. <https://doi.org/10.3762/bjoc.17.28>.
- (10) Arntson, K. E.; Pomerantz, W. C. K. Protein-Observed Fluorine NMR: A Bioorthogonal Approach for Small Molecule Discovery: Miniperspective. *J. Med. Chem.* **2016**, *59* (11), 5158–5171. <https://doi.org/10.1021/acs.jmedchem.5b01447>.
- (11) Kinde, M. N.; Bondarenko, V.; Granata, D.; Bu, W.; Grasty, K. C.; Loll, P. J.; Carnevale, V.; Klein, M. L.; Eckenhoff, R. G.; Tang, P.; Xu, Y. Fluorine-19 NMR and Computational Quantification of Isoflurane Binding to the Voltage-Gated Sodium Channel NaChBac. *Proc. Natl. Acad. Sci.* **2016**, *113* (48), 13762–13767. <https://doi.org/10.1073/pnas.1609939113>.
- (12) Peng, J. W. Cross-Correlated ¹⁹F Relaxation Measurements for the Study of Fluorinated Ligand–Receptor Interactions. *J. Magn. Reson.* **2001**, *153* (1), 32–47. <https://doi.org/10.1006/jmre.2001.2422>.
- (13) Dalvit, C.; Piotto, M. ¹⁹F NMR Transverse and Longitudinal Relaxation Filter Experiments for Screening: A Theoretical and Experimental Analysis: Fluorine NMR Relaxation for Screening: Theory and Experiments. *Magn. Reson. Chem.* **2017**, *55* (2), 106–114. <https://doi.org/10.1002/mrc.4500>.
- (14) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. Fluorine-NMR Competition Binding Experiments for High-Throughput Screening of Large Compound Mixtures. *Comb. Chem. High Throughput Screen.* **2002**, *5* (8), 605–611. <https://doi.org/10.2174/1386207023329923>.
- (15) Jordan, J. B.; Poppe, L.; Xia, X.; Cheng, A. C.; Sun, Y.; Michelsen, K.; Eastwood, H.; Schnier, P. D.; Nixey, T.; Zhong, W. Fragment Based Drug Discovery: Practical Implementation Based on ¹⁹F NMR Spectroscopy. *J. Med. Chem.* **2012**, *55* (2), 678–687. <https://doi.org/10.1021/jm201441k>.

- (16) Matei, E.; Gronenborn, A. M. ^{19}F Paramagnetic Relaxation Enhancement: A Valuable Tool for Distance Measurements in Proteins. *Angew. Chem. Int. Ed.* **2016**, *55* (1), 150–154. <https://doi.org/10.1002/anie.201508464>.
- (17) Huang, Y.; Wang, X.; Lv, G.; Razavi, A. M.; Huysmans, G. H. M.; Weinstein, H.; Bracken, C.; Eliezer, D.; Boudker, O. Use of Paramagnetic ^{19}F NMR to Monitor Domain Movement in a Glutamate Transporter Homolog. *Nat. Chem. Biol.* **2020**, *16* (9), 1006–1012. <https://doi.org/10.1038/s41589-020-0561-6>.
- (18) Rüdisser, S. H.; Goldberg, N.; Ebert, M.-O.; Kovacs, H.; Gossert, A. D. Efficient Affinity Ranking of Fluorinated Ligands by ^{19}F NMR: CSAR and FastCSAR. *J. Biomol. NMR* **2020**, *74* (10–11), 579–594. <https://doi.org/10.1007/s10858-020-00325-x>.
- (19) Hutton, J. J.; Marglin, A.; Witkop, B.; Kurtz, J.; Berger, A.; Udenfriend, S. Synthetic Polypeptides as Substrates and Inhibitors of Collagen Proline Hydroxylase. *Arch. Biochem. Biophys.* **1968**, *125* (3), 779–785. [https://doi.org/10.1016/0003-9861\(68\)90514-6](https://doi.org/10.1016/0003-9861(68)90514-6).
- (20) Kao, W. W.; Prockop, D. J.; Berg, R. A. Kinetics for the Secretion of Nonhelical Procollagen by Freshly Isolated Tendon Cells. *J. Biol. Chem.* **1979**, *254* (7), 2234–2243. [https://doi.org/10.1016/S0021-9258\(17\)30212-0](https://doi.org/10.1016/S0021-9258(17)30212-0).
- (21) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc.* **2001**, *123* (4), 777–778. <https://doi.org/10.1021/ja005542v>.
- (22) Malkar, N. B.; Lauer-Fields, J. L.; Borgia, J. A.; Fields, G. B. Modulation of Triple-Helical Stability and Subsequent Melanoma Cellular Responses by Single-Site Substitution of Fluoroproline Derivatives. *Biochemistry* **2002**, *41* (19), 6054–6064. <https://doi.org/10.1021/bi012071w>.
- (23) Costantini, N. V.; Ganguly, H. K.; Martin, M. I.; Wenzell, N. A.; Yap, G. P. A.; Zondlo, N. J. The Distinct Conformational Landscapes of 4 *S*-Substituted Prolines That Promote an *Endo* Ring Pucker. *Chem. – Eur. J.* **2019**, *25* (48), 11356–11364. <https://doi.org/10.1002/chem.201902382>.
- (24) Ottoy, E. (2022). *Combined NMR and ab initio analysis of fluorinated prolines for their application in peptides and proteins*. Ghent University. Faculty of Sciences, Ghent, Belgium.
- (25) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. Collagen Stability: Insights from NMR Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations. *J. Am. Chem. Soc.* **2002**, *124* (11), 2497–2505. <https://doi.org/10.1021/ja0166904>.
- (26) Newberry, R. W.; Raines, R. T. 4-Fluoroprolines: Conformational Analysis and Effects on the Stability and Folding of Peptides and Proteins. In *Peptidomimetics I*; Lubell, W. D., Ed.; Topics in Heterocyclic Chemistry; Springer International Publishing: Cham, 2016; Vol. 48, pp 1–25. https://doi.org/10.1007/7081_2015_196.
- (27) Crespo, M. D.; Rubini, M. Rational Design of Protein Stability: Effect of (2*S*,4*R*)-4-Fluoroproline on the Stability and Folding Pathway of Ubiquitin. *PLoS ONE* **2011**, *6* (5), e19425. <https://doi.org/10.1371/journal.pone.0019425>.
- (28) Torbeev, V. Yu.; Hilvert, D. Both the *Cis* - *Trans* Equilibrium and Isomerization Dynamics of a Single Proline Amide Modulate B2-Microglobulin Amyloid Assembly. *Proc. Natl. Acad. Sci.* **2013**, *110* (50), 20051–20056. <https://doi.org/10.1073/pnas.1310414110>.
- (29) Kim, W.; McMillan, R. A.; Snyder, J. P.; Conticello, V. P. A Stereoelectronic Effect on Turn Formation Due to Proline Substitution in Elastin-Mimetic Polypeptides. *J. Am. Chem. Soc.* **2005**, *127* (51), 18121–18132. <https://doi.org/10.1021/ja054105j>.

- (30) Siebler, C.; Maryasin, B.; Kuemin, M.; Erdmann, R. S.; Rigling, C.; Grünenfelder, C.; Ochsenfeld, C.; Wennemers, H. Importance of Dipole Moments and Ambient Polarity for the Conformation of Xaa-Pro Moieties – a Combined Experimental and Theoretical Study. *Chem. Sci.* **2015**, 6 (12), 6725–6730. <https://doi.org/10.1039/C5SC02211H>.
- (31) Hofman, G.-J.; Ottoy, E.; Light, M. E.; Kieffer, B.; Kuprov, I.; Martins, J. C.; Sinnaeve, D.; Linclau, B. Minimising Conformational Bias in Fluoroprolines through Vicinal Difluorination. *Chem. Commun.* **2018**, 54 (40), 5118–5121. <https://doi.org/10.1039/C8CC01493K>.
- (32) Sinnaeve, D.; Ben Bouzayene, A.; Ottoy, E.; Hofman, G.-J.; Erdmann, E.; Linclau, B.; Kuprov, I.; Martins, J. C.; Torbeev, V.; Kieffer, B. Fluorine NMR Study of Proline-Rich Sequences Using Fluoroprolines. *Magn. Reson.* **2021**, 2 (2), 795–813. <https://doi.org/10.5194/mr-2-795-2021>.
- (33) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd ed.; Methods in stereochemical analysis; Wiley: New York, **2000**.
- (34) Leeftang, B. R.; Kroon-Batenburg, L. M. J. CROSREL: Full Relaxation Matrix Analysis for NOESY and ROESY NMR Spectroscopy. *J. Biomol. NMR* **1992**, 2 (5), 495–518. <https://doi.org/10.1007/BF02192812>.
- (35) Hunt, E.; Meyer, H. Anisotropy of the ^{19}F Chemical Shift in Trapped CH_3F Molecules. *J. Chem. Phys.* **1964**, 41 (2), 353–357. <https://doi.org/10.1063/1.1725873>.
- (36) MacLean, C.; Mackor, E. L. The Anisotropy of the Chemical Shift of the Fluorine Nucleus in CHFCl_2 . *Proc. Phys. Soc.* **1966**, 88 (2), 341–349. <https://doi.org/10.1088/0370-1328/88/2/309>.
- (37) Andrew, E. R.; Tunstall, D. P. Anisotropy of Chemical Shift for Some Fluorine Compounds. *Proc. Phys. Soc.* **1963**, 81 (6), 986–995. <https://doi.org/10.1088/0370-1328/81/6/302>.
- (38) Dunn, M. B. Chemical Shift Anisotropy in Trifluoroacetic Acid. *Mol. Phys.* **1968**, 15 (4), 433–434. <https://doi.org/10.1080/00268976800101311>.
- (39) Nehring, J.; Saupe, A. Anisotropies of the ^{19}F Chemical Shifts in Fluorobenzene Compounds from NMR in Liquid Crystals. *J. Chem. Phys.* **1970**, 52 (3), 1307–1310. <https://doi.org/10.1063/1.1673130>.
- (40) Dürr, U. H. N.; Grage, S. L.; Witter, R.; Ulrich, A. S. Solid State ^{19}F NMR Parameters of Fluorine-Labeled Amino Acids. Part I: Aromatic Substituents. *J. Magn. Reson.* **2008**, 191 (1), 7–15. <https://doi.org/10.1016/j.jmr.2007.11.017>.
- (41) Grage, S. L.; Dürr, U. H. N.; Afonin, S.; Mikhailiuk, P. K.; Komarov, I. V.; Ulrich, A. S. Solid State ^{19}F NMR Parameters of Fluorine-Labeled Amino Acids. Part II: Aliphatic Substituents. *J. Magn. Reson.* **2008**, 191 (1), 16–23. <https://doi.org/10.1016/j.jmr.2007.11.016>.
- (42) Lectures by James Keeler, chapter 8 relaxation <http://www-keeler.ch.cam.ac.uk/lectures/>
- (43) Kang, Y.K. and Young Choi, H. Cis–trans isomerization and puckering of proline residue. *Biophys. Chem.* **2004**, 111(2), pp. 135–142. doi:10.1016/j.bpc.2004.05.006.

Chapter 3:
Dynamics of 4-fluoroprolines and
their usage to evaluate overall
peptide's motion

III. Dynamics of 4-fluoroprolines and their usage to evaluate overall peptide's motion.

Introduction:

The controlled stereo electronic effects together with the ability to modulate the peptide bond conformation led to various studies where fluoroprolines were substituted to natural prolines to investigate the resulting effect on the protein. For example, the substitution of proline 63 in mRFP1 (the monomer of the red fluorescent protein from *Discosoma* (DsRed)) by 4*R* fluoroproline enhances both its thermal and chemical stability and accelerates its maturation.¹ Similarly, the incorporation of 4*S* fluoroprolines was shown to accelerate the refolding rate upon thermal denaturation of the green fluorescent protein EGFP mainly due to the favorable *c_γendo* ring pucker². The effect of fluorination on protein can be complex and depends on both the studied protein and the used fluoroproline. For instance, while the incorporation of 4*R* fluoroproline enhances the thermal and chemical stability of the human ubiquitin as well as its refolding kinetics, the incorporation of 4*S* fluoroproline was not possible³. Additionally, some proteins may only tolerate the incorporation of one 4 fluoroproline conformer while the insertion of the other leads to unfolded proteins.⁴ One example is the Taq DNA polymerase that can be stabilized from a thermophilic organism, and maintains its enzymatic activity using a 4*R* fluoroproline, while its expression is not successful when a 4*S* fluoroproline is incorporated.⁴ On the contrary, The incorporation of 4*S* fluoroproline was shown to stabilize Pin1 WW domain and to strengthen its interactions with its ligands,⁵ and also to increase the affinity of the most immunogenic domain of anti-MUC1 antibodies for the MUC1 antigen.⁶ Finally, the modulation of Pro 32 *cis* and *trans* isomerization by both 4*R* and 4*S* fluoroprolines was shown to influence the amyloid assembly of the β 2-microglobulin.⁷ Remarkably, the presence of the fluorine spin was not often exploited to provide a description of the dynamical features of these systems using NMR relaxation.

In this chapter, we measured the fluorine relaxation parameters of 4*R* and 4*S* fluoroprolines either in isolation or within peptides. The dynamics are assessed through the interpretation of the three relaxation parameters R_1 , R_2 and ^1H - ^{19}F NOE.

1. Relaxation properties of 4-fluorinated prolines

Proton and fluorine spectra of isolated *4R* and *4S* fluoroprolines at 298 K are displayed on figure 1. Single fluorine populations are observed at -177.28 ppm and -175.22 ppm for *4R* and *4S* FPro respectively. The H_γ appears at a particular chemical shift 5.35 ppm within both fluoroprolines. *4R* Fpro is characterized by similar H_δ chemical shifts and distinguished H_β spins, while the opposite is true for *4S* FPro.

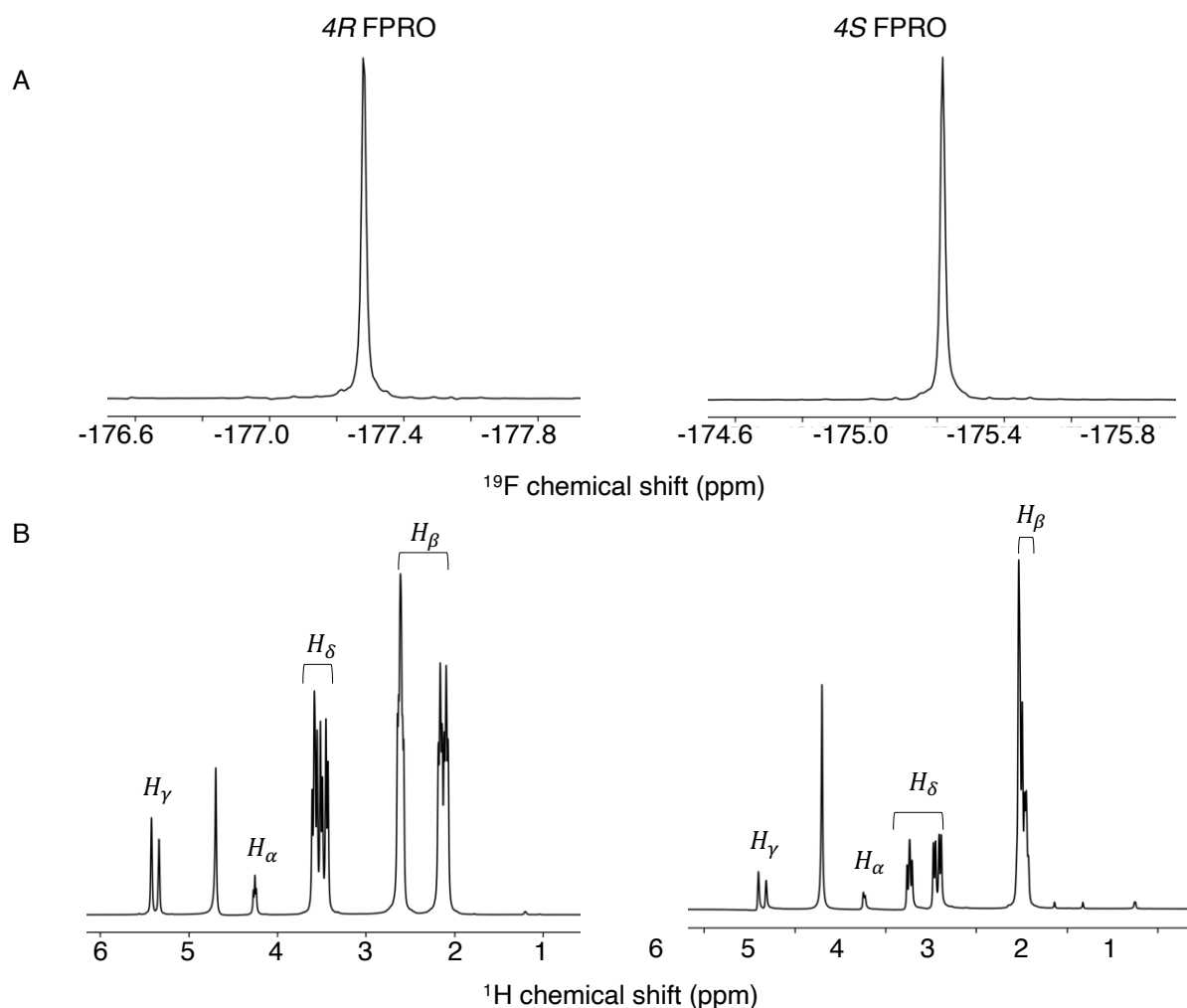


Figure 1: A)1D ^{19}F NMR spectra of free *4R* fluoroproline (left) and *4S* fluoroproline (right). B) 1D ^1H NMR spectra of *4R* and *4S* fluoroprolines shown on the left and right respectively. Both fluoroprolines were dissolved in phosphate buffer pH 7.2 20 mM in D_2O . Recorded at 600 MHz at 298 K.

1.1. Longitudinal relaxation rates of *4R* and *4S* fluoroprolines

The rotational diffusion of a molecule is determined by the temperature and the viscosity of the solvent, with the later that also depends on the temperature. Aiming at exploring the diffusion of fluoroprolines in a wide range of correlation times, we used both temperature and viscosity variations to probe their impact on longitudinal relaxation and compare the resulting values with theoretical ones. To modulate the viscosity of the solvent, various ratios of water/glycerol were used at different temperatures. These variations of temperatures and viscosity are used to modulate the value of the correlation time as follow:

$$\frac{\tau_c}{\tau_c^0} = \frac{\eta(T)T^0}{\eta^0(T^0)T}$$

where τ_c^0 is a reference correlation time at viscosity η^0 and temperature T^0 . The temperature dependence of the viscosity was approximated by those of the water published in the Handbook of Chemistry ⁸.

Temperature dependence of *4S* fluoroproline longitudinal relaxation in D₂O

Fluorine R₁ relaxation rate constants were measured for a single *4S* fluoroproline in phosphate buffer in D₂O solvent for temperatures ranging from 278 K to 320 K. The measured fluorine R₁ are shown on Appendix 1. Figure 2 displays the variation of the longitudinal relaxation rates using the value at 278 K as a reference point. For this point, the R₁ value was set to the theoretical value corresponding to a correlation time of 27 ps. The slope of the fluorine longitudinal rate is very close to the theoretical model and vary by 0.35 s⁻¹ from 278K to 320 K. The range of correlation time is found to be similar to those reported in previous studies and could be due to both the ring interconversion between the *endo* and the *exo* conformers and/or the rotational diffusion of the molecule. Indeed, Mikhailov *et al.* reported in 1995 that the correlation time of proline C γ -H in GPGG peptide measured by ¹³C longitudinal relaxation rate at 303 K is 27±1.5 ps, later in 2014 Aliev et al further confirmed this result.⁹

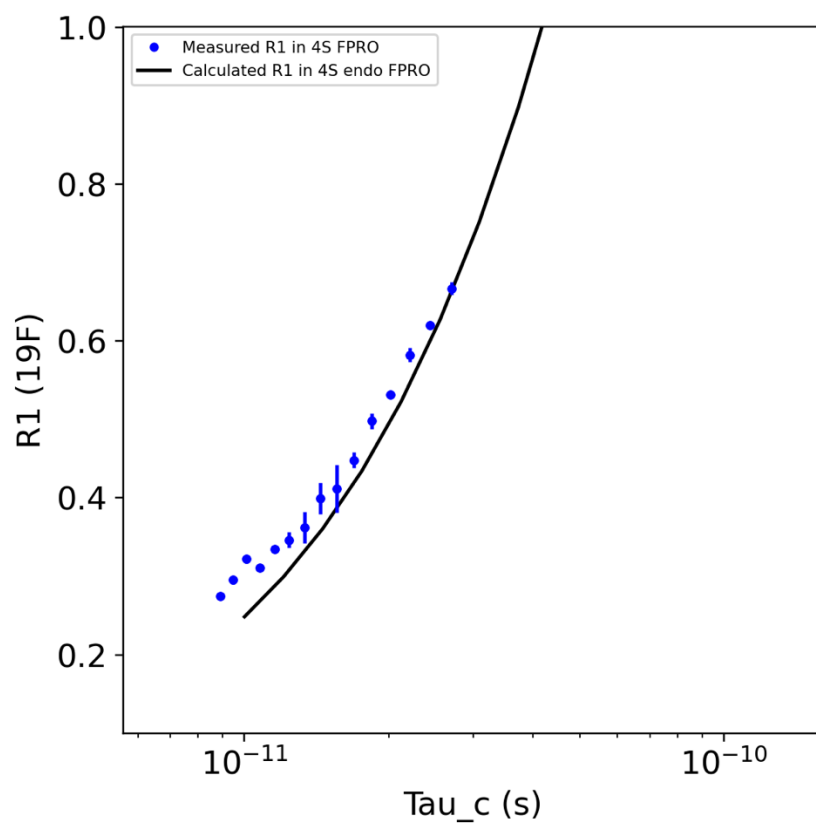


Figure 2: Zoomed section of the fit of the calculated fluorine longitudinal relaxation rate within 4S fluoroproline (solid black line) and the experimentally measured fluorine R_1 rates (blue) at varying temperature at 600 MHz using inversion recovery experiments.

Longitudinal relaxation of free *4R* and *4S* fluoroprolines in solvents of various viscosities

Aiming at experimentally probing the peculiar “camel hump” longitudinal relaxation rate profile of the fluorine as a function of the correlation time, a mixture of *4R* and *4S* fluoroprolines was dissolved in different H₂O/D₂O-Glycerol mixtures designed to provide solvents with varying viscosities over two degrees of magnitude. (See experimental section for details). Figure 3 shows an example of one inversion recovery experiment recorded for the mixture and the fit to extract the longitudinal relaxation rates while R_1 values are given in Appendix 2.

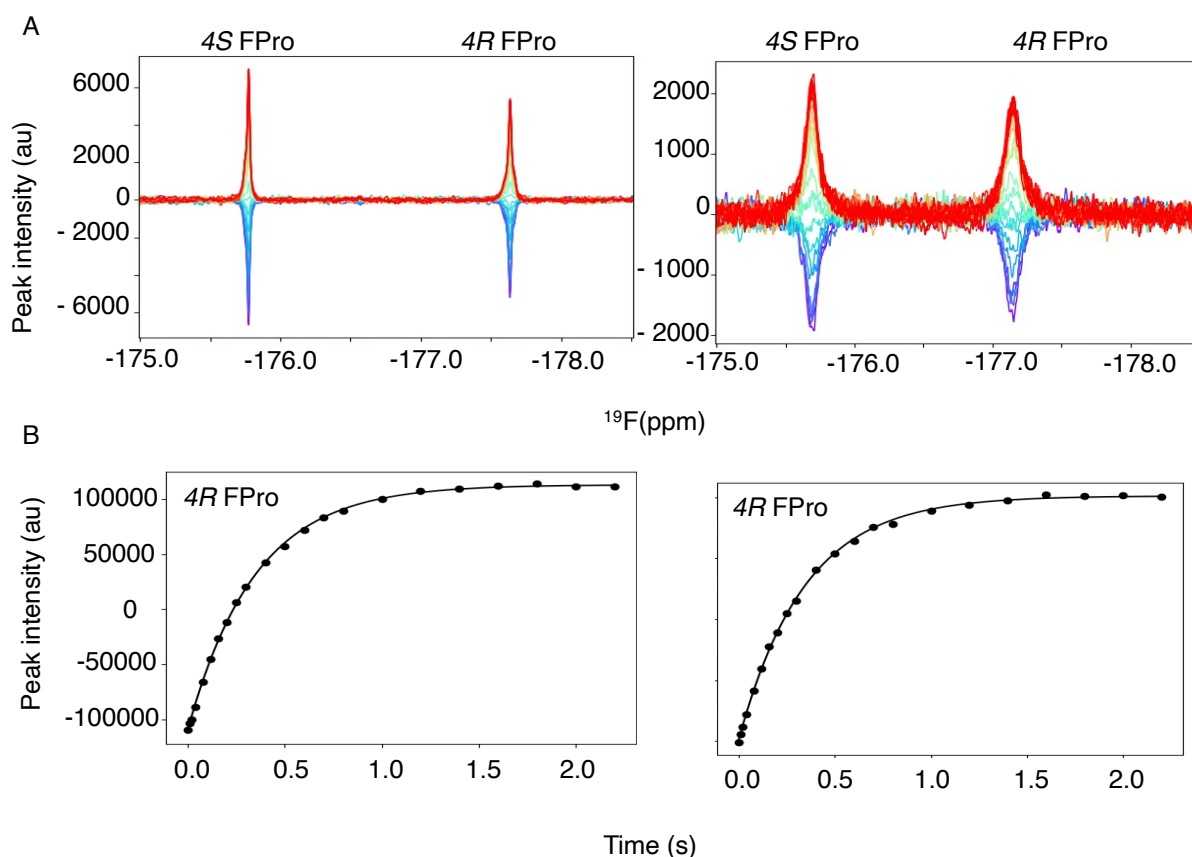


Figure 3: A: stack plot of extracted 1D ¹⁹F spectra during 2D inversion recovery experiment recorded for a mixture of free *4R* and *4S* fluoroprolines in solution at relative viscosity 13.63 (left) and 383.7 (right) at 278 K. The first spectrum is shown in blue and the last is in red. B: Example of the fit performed to extract fluorine R_1 . The dots are the experimental peak intensities of *4R* fluoroproline and the solid line is the corresponding fit to a monoexponential function.

The variation of the solution viscosity successfully sampled a large window of correlation times estimated to range from 0.02 ns to 11 ns. In contrast with theoretical relaxation rates constants computed from interaction parameters derived from Gaussian calculations, no obvious difference was observed between experimental fluorine longitudinal rates measured for *4R* and *4S* fluoroprolines (Figure 4). Both fluoroprolines display very similar R_1 values with a maximum value at 4ns of 3.4 s^{-1} , slightly higher than the theoretical value of 3.1 s^{-1} . In addition, experimental R_1 values do not exhibit a second maximum at 0.28 ns, that is due to CSA relaxation. The discrepancy between theoretical and experimental longitudinal relaxation may originate from several reasons: (i) the theoretical calculations were performed using static models originating from Gaussian calculations. They do not take into account *exo-endo* puckering fluctuations that can average the CSA contribution to the relaxation. This reason may also explain the global differences observed between the experimental and computed values since conformational fluctuations may bring protons closer to the fluorine atom leading to enhanced dipolar interactions. (ii) Another possible reason lies in the estimation of the CSA tensor parameters by Gaussian that may be overestimated. Further modeling studies in which the whole conformational space of the proline is taken into account are probably required to account for the experimental longitudinal relaxation profiles.

Nevertheless, the reasonable agreement between the experimental ^{19}F R_1 relaxation values and the theoretical values derived from a rigid model suggests that the relaxation is dominated by the slow rotational diffusion of the proline. This indicates that fluoroproline's ring may be considered as a rigid structure that is mostly sensing the global tumbling of the molecule. However, the use of R_1 relaxation to probe fluoroproline's dynamics faces two difficulties: (i) the profile of ^{19}F R_1 relaxation values as a function of the correlation time displays a flat profile around 3.1 s^{-1} for a wide range of correlation time values (from 0.5 to 10 ns) due to the similar Larmor frequencies of proton and fluorine nuclei, (ii) the proton-fluorine dipole-dipole interactions that contribute to R_1 relaxation are distributed in various directions which may complicate the interpretation of the relaxation in term of molecular motions if they are anisotropic.

In the following, we will take advantage of the peculiar frequency of the H_γ to measure NOE buildup curves using 1D saturation experiments.

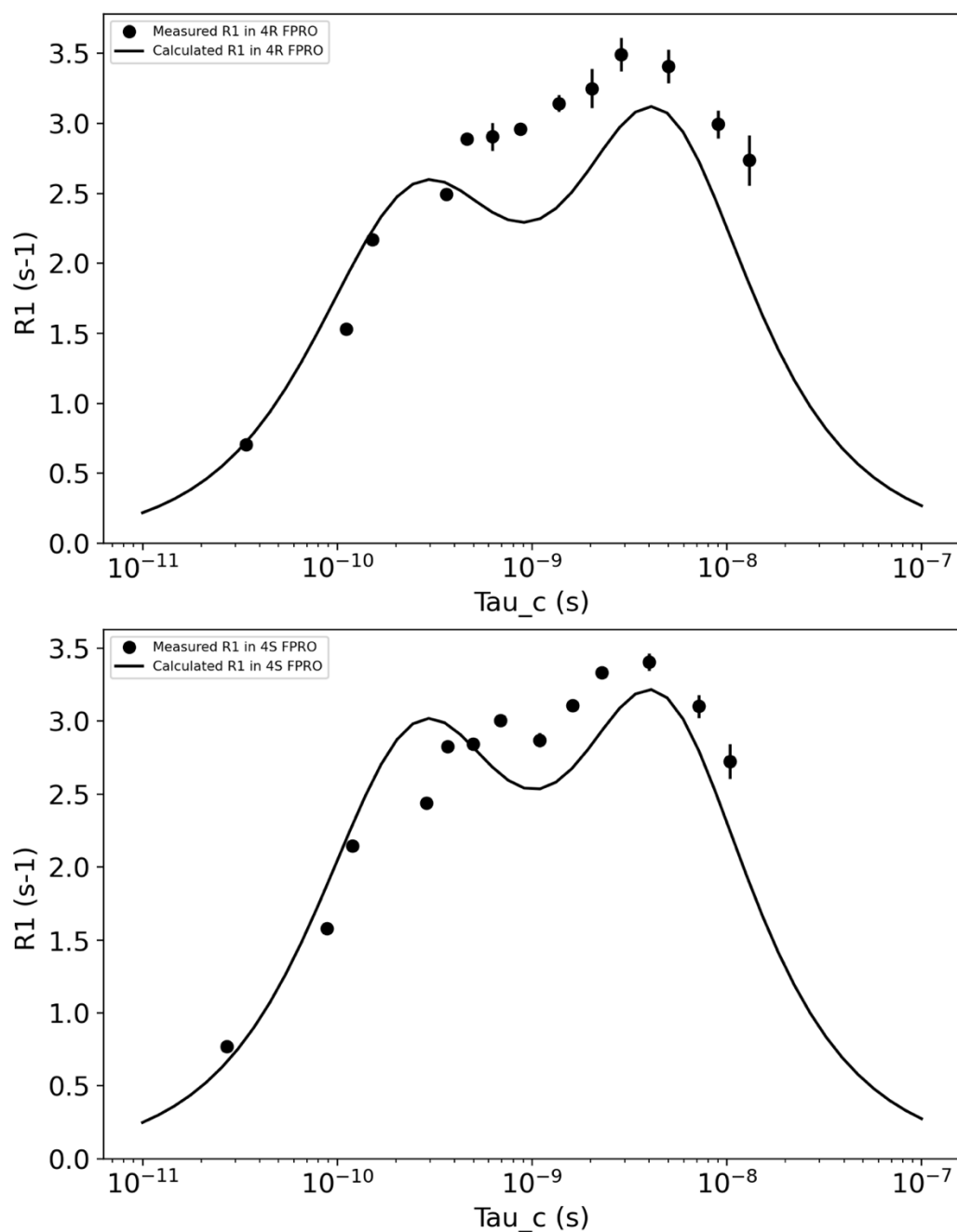


Figure 4: Dependence of the fluorine relaxation rates upon the correlation time within *4R exo* (upper) and *4S endo* (lower) fluoroprolines. The solid black line represents the calculated longitudinal relaxation rates as expressed from the equations derived in the first chapter. The dots represent the experimentally measured longitudinal relaxation rates in a mixture of free fluoroprolines dissolved in water-glycerol solvent with varying viscosity at 278 K and 600 MHz using an inversion recovery experiment.

1.2. Measuring Heteronuclear ^1H - ^{19}F NOE within 4*S* fluoroproline

The 1D ^1H NMR spectrum of 4-fluoroproline is characterized by a distinguishable chemical shift of the geminal H_γ . The observed signal is a doublet at 5.4 ppm separated by 55 Hz. Consequently, one can apply selective radio-frequency pulses to saturate the H_γ without disturbing the equilibrium of the remaining protons of the proline. This experiment has been designed using a single 4*S* fluoroproline dissolved in phosphate buffer (Figure 5). The build-up of the heteronuclear $^1\text{H}_\gamma$ - ^{19}F NOE, is established through a continuous saturation of the H_γ frequency at increasing delays while the fluorine signal's intensity is monitored on the ^{19}F channel. The heteronuclear NOE recorded is positive, as expected for small molecules and is estimated to be $12.6\% \pm 3$. Considering the variation of the NOE with the correlation time (presented in chapter 2 figure 10) such a value would correspond to a correlation time of 83 ± 20 ps, consistent albeit twice higher than the value deduced from R_1 relaxation rates ($\tau_c = 30$ ps). Several reasons may explain the observed discrepancy between these values: (i) the presence of spin diffusion that is neglected in the present analysis, (ii) an erroneous estimate of the distance between the fluorine and $^1\text{H}_\gamma$ (we used a distance of 2.01 Å from the Gaussian model). A variation of this distance as low as 5% may account for the observed discrepancy due to the r^{-6} dependence of the cross-relaxation rate to the inter-nuclei distance r .

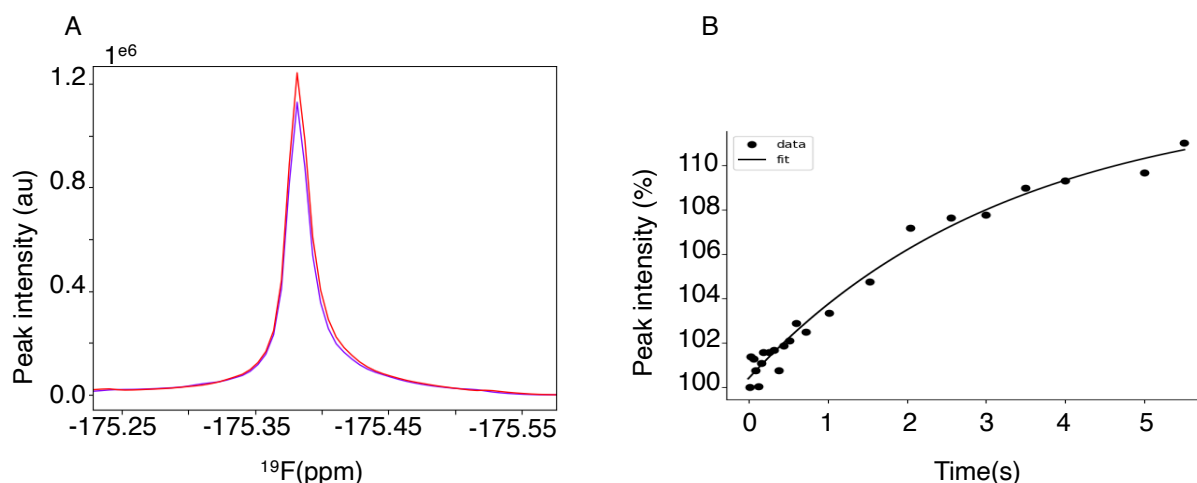


Figure 5: A: NOE enhancement observed on fluorine signal of free 4*S* fluoroproline upon selective saturation of H_γ proton at 300 K. The first spectrum (saturation time: 0.008 s) is shown in violet and the last (saturation time: 5.5 s) in red. The last recorded spectrum is shown in red. B: Corresponding experimental NOE (%) build-up at F_γ . The dots are the experimental

peak intensities and the solid line is the corresponding fit to a monoexponential function according to: $M_F^z(t) = M_F^{zeq} + \frac{\gamma_H \sigma_{FH}}{\gamma_F \rho_F} M_F^{zeq} (1 - e^{-\rho_F t})$

1.3. Measurement of transverse relaxation rate in 4S fluoroproline

In order to provide additional experimental data to validate the theoretical model we used to model fluorine relaxation; transverse relaxation rate was measured using a CPMG pulse sequence to remove the contribution of the chemical exchange (Figure 6). With the assumption of a single exponential decay, the measured R_2 was $1.27 \pm 0.5 \text{ s}^{-1}$ at 300 K, corresponding to a molecular correlation time of 48 ps. Noteworthy, this value lies between those measured by R_1 and R_2 relaxation which further suggests that the correlation time value derived from heteronuclear NOE may be overestimated.

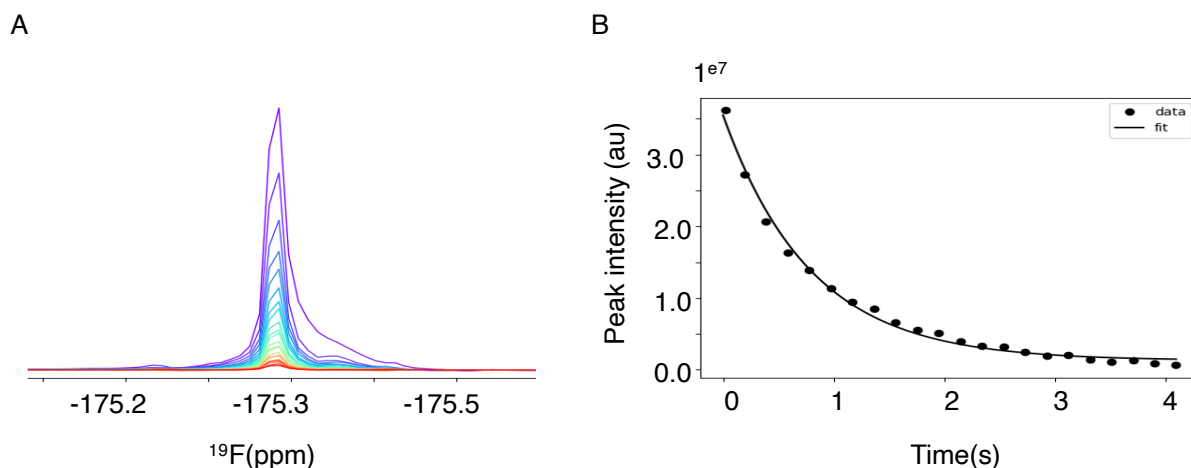


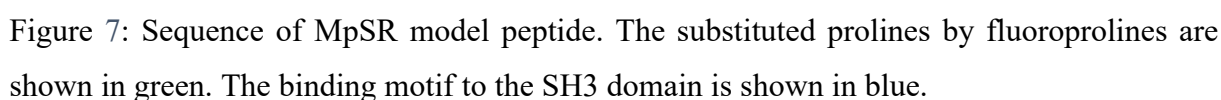
Figure 6: A: stack plot of fluorine R_2 CPMG. Delay (100us). The last spectrum is shown in red. B: Corresponding fit of the experimental peak intensity (dots) to a monoexponential function (solid line). Recorded at 300 K at 600 MHz.

The peculiar shape of the peak might be due to an effect of the CPMG pulse sequence. Even though its important width might suggest the presence of a minor peak, we strongly reject this hypothesis based on the good quality of both the ^{19}F and the ^1H spectra of this sample.

The interpretation of the relaxation data measured within 4S fluoroproline reveals a fast dynamic behavior on the ps time scale. The longitudinal relaxation rate is not a trustworthy parameter due to its peculiar “camel hump” like profile. Similarly, the transverse relaxation is not reliable to study fluoroproline dynamics due to the contribution of the chemical exchange.

To study relaxation parameters of *4R* and *4S* fluoroprolines in a peptide context, we measured fluorine relaxation rates of *4R* and *4S* fluoroprolines incorporated in a polyproline sequence. The results shown below provide complementary results to the study that was published in Magnetic resonance in 2021 (see chapter 5).

Fluoroprolines *4R* and *4S* were inserted in a model peptide composed by two segments of 5 successive prolines separated by a serine residue and followed by four amino acids Arg-Val-Tyr-Lys (RVYK) required for its interaction with the SH3.3 domain of the vinexine β . The two fluoroprolines are in positions 4 (*4S* fluoroproline) and 8 (*4R* fluoroproline). The model peptide is denoted MpSR. Figure 7.



The measured F- H_γ NOE of 4*R* and 4*S* fluoroproline within MpSR in water and deuterium oxide are displayed on table 1 and figure 8 together with the corresponding theoretical correlation time. We observe that all NOE values are negative, as expected for rotational correlation times above 0.3 ns. The NOE values are systematically larger in D₂O as expected for higher viscosity leading to correlation times that are between 20% and 25% larger in D₂O as compared to normal water. This result is in agreement with the D₂O viscosity of D₂O being

1.25 larger than normal water. Further, the measured NOEs are systematically larger for the FPRO at position 8 as compared to those measured for the FPRO at position 4.

This difference may reflect a distinct local dynamic in both polyproline segments. Indeed, the four C-terminal residues of the peptide (RVYK) may fold back on the second polyproline segment to form a globular structure with slower tumbling time while the FPRO at position 4 is sensing faster motion of the first polyproline segment.

Table 1: Correlation times extracted from experimental H_γ -F NOE in MpSR, at 298 K.

	4 <i>S</i> FPRO		4 <i>R</i> FPRO	
	D ₂ O	H ₂ O	D ₂ O	H ₂ O
H_γ - <i>F</i> (%)	-9.3	-6.2	-19	-14
τ_C (ns)	0.53	0.44	0.8	0.64

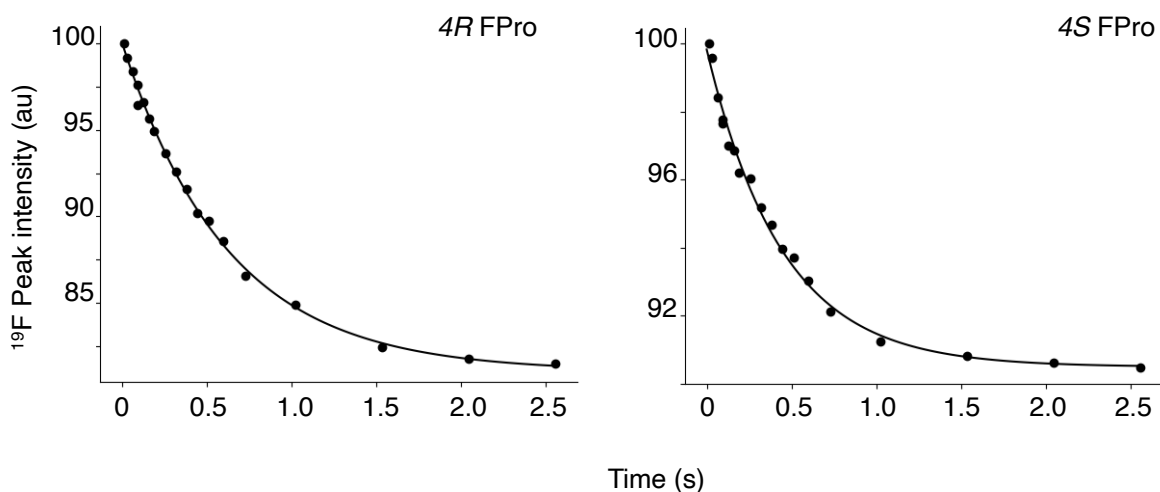


Figure 8: Experimental NOE (%) build-up at F_γ . In MpSR in D₂O at 298 K. The dots are the experimental peak intensities and the solid line is the corresponding fit to a monoexponential function

2.2. Fluorine transverse relaxation rate in MpSR model peptide

Fluorine transverse relaxation rate was measured at 600 MHz using both a spin echo and a CPMG pulse sequences to evaluate the possible contribution of exchange contribution. A qualitative analysis may be performed by approximating R_{ex}^* as $R_{ex}^* = R_{2\text{ spin echo}} - R_{2\text{ CPMG}}$.

R_{ex}^* denotes an apparent exchange rate constant because a residual exchange may still contribute to the relaxation even after the application of a CPMG pulse sequence.

To determine the time scale of the exchange regime, we measured the transverse relaxation rates at a second magnetic field 400 MHz (Table 2). We notice that the apparent chemical exchange contribution is different within the two fluoroprolines and is more important for the *4R* fluoroproline at position 8. This observation confirms our previous conclusion on the presence of μ s-ms dynamics within the second polyproline segment of the peptide. The parameter α relating the strength of the magnetic field to R_{ex} as defined by equation 56 of chapter 1 is smaller than the unit within both fluoroprolines (0.26 and 0.48 for *4R* and *4S* FPRO respectively) indicating a motion in the slow exchange regimes.

Table 2: Transverse relaxation rates of *4R* and *4S* fluoroprolines within the model peptide MpSR at two magnetic fields 400 MHz and 600 MHz at 298 K.

	<i>4R</i> FPRO			<i>4S</i> FPRO		
	Spin echo	CPMG delay(200 μ s)	R_{ex}^*	Spin echo	CPMG delay(200 μ s)	R_{ex}^*
$R_2(s^{-1})$ 400 MHz	15.94 ± 1.4	5.82 ± 0.4	10.12	10.57 ± 1.0	4.79 ± 0.3	5.78
$R_2(s^{-1})$ 600 MHz	18.52 ± 0.4	7.26 ± 0.4	11.26	12.36 ± 0.46	5.34 ± 0.2	7.02

3. Dynamics of a peptide encompassing a phosphorylated serine

We used the model peptide MpSR to investigate the effect of the serine phosphorylation on the peptide dynamics. Indeed, the phosphorylation of the serine linking the two polyproline segments might alter their relative dynamics, an effect that may be revealed by fluorine relaxation. To evaluate the validity of this hypothesis, we substituted the serine amino acid by a phosphorylated serine. The measured relaxation parameters within MpSRSer_{phospho} and MpSR are shown in table 3. The build-up of the heteronuclear H_γ -F NOE are displayed in figure 9.

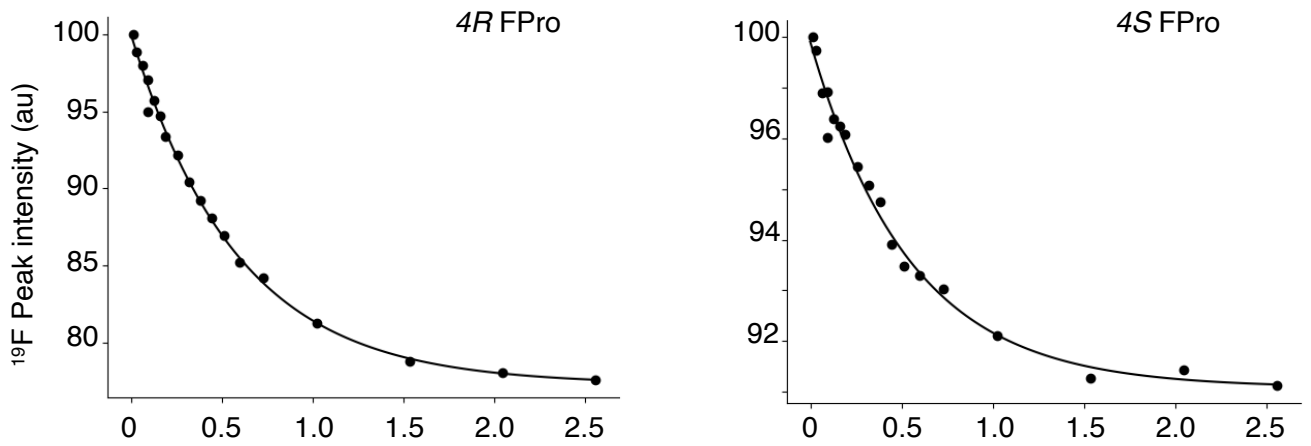


Figure 9: Experimental NOE (%) build-up at F_γ . In MpSRSer_{phospho} in D₂O at 298 K. The dots are the experimental peak intensities and the solid line is the corresponding fit to a monoexponential function

Table 3: Comparison of the relaxation of *4S* and *4R* fluoroprolines in MpSR and MpSRSer_{phospho} model peptides at 298 K. The τ_c were directly determined from the simulated profiles of σ_{FH} and R_2 as a function of the correlation time in chapter2.

	MpSRSer _{phospho}		MpSR	
	<i>4S</i> FPRO	<i>4R</i> FPRO	<i>4S</i> FPRO	<i>4R</i> FPRO
NOE _{max} H_γ -F (%)	-13.7	-22.8	-9.3	-19
ρ_F (s ⁻¹)	1.84	1.7	2.25	1.59
σ_{FH} (s ⁻¹)	-0.24	-0.36	-0.19	-0.28
τ_c (ns)	0.53	0.68	0.46	0.56
R ₁	2.08 ± 0.08	1.89 ± 0.09	2.31±0.01	2.13±0.01
R ₂ spin echo	13.16± 0.4	20.64 ± 0.33	12.36±0.46	18.52 ±0.39
R ₂ CPMG	6.53±0.16	10.3±0.046	5.4 ±0.34	9.67 ±0.46
τ_c (ns)	0.4	1	0.33	0.94

Both MpSR and MpSRSer_{phospho} peptides display similar longitudinal and transverse relaxation features. Both R_2 and heteronuclear NOE indicate faster tumbling for the first polyproline segment as compared to the second moiety of the sequence with a consistent ratio of 1.2 for the correlation times values derived from cross-relaxation rate σ_{FH} and 2.5-2.8 for the correlation times values derived from transverse relaxation.

Further, the apparent exchange contribution derived from the comparison of the R_2 measured from spin echo and CPMG experiments are very similar for the phosphorylated and non-phosphorylated peptides, indicating that phosphorylation does not affect μ s to ms motions. Noteworthy, the agreement between the correlation times measured from σ_{FH} and R_2 experiments for the 4*S* fluoroproline at position 4 provides some indications that exchange contributions are effectively quenched by the CPMG experiment. It is interesting to compare the experimental values derived from fluorine relaxation with expected values derived from the Stokes-Einstein expression. The expected correlation time for a globular peptide of 15 amino-acids tumbling in water at 298 K is 0.4 ns that perfectly agrees with the correlation times extracted from the cross-relaxation rate indicating the reliability of this parameter to report local motion. Several factors may explain the differences found between the correlation times of the two polyproline segments:

- 1) a faster internal dynamic of the 4*S* fluoroproline in position 4
- 2) a faster segmental motion of the first polyproline segment
- 3) the anisotropy of the rotational diffusion of the polyproline peptide

Distinguishing the genuine cause of the dynamical differences between the two peptide segments will require further experiments but the first hypothesis may be ruled out by exchanging the positions of the 4*R* and 4*S* fluoroprolines in the peptide and measure their relaxation. This experiment was performed and published in the MR paper and revealed similar differences between relaxation values measured at position 4 and 8, irrespective of the fluorine stereo-isomer.

4. Conclusions

In this chapter we conducted a quantitative study of the dynamic behavior of 4- fluoroprolines in both their free state and when incorporated into model peptides of polyproline rich sequences through an interpretation of the relaxation parameters R_1 , R_2 and ^1H - ^{19}F heteronuclear NOE. The fluorine longitudinal relaxation rate is not reliable for dynamical studies due to its small

contrast between 0.4ns and 4ns. Similarly, the transverse relaxation rate of fluoroprolines is hampered by an exchange contribution arising from conformational fluctuations occurring in the μ s to ms time scale. Ultimately, the purely dipolar ^1H - ^{19}F heteronuclear cross-relaxation rate is the most reliable parameter to evaluate local motion. Further studies of this parameter will be introduced in the next chapter.

5. Material and methods

Sample preparation:

The mixture of fluoroproline is prepared by dissolving commercially available *4R* and *4S* fluoroprolines in phosphate buffer 20 mM pH 7.2 for a final concentration of 5mM.

Glycerol solutions were prepared by adding glycerol 87%(wt) and 98%(wt) to the fluoroproline solution. The exact volumes of glycerol, water and fluoroprolines solution, was calculated depending on the target viscosity. The final volume contained 10% of D₂O. The prepared solutions are shown on appendix 2.

Commercially available *4S* fluoroproline was dissolved in phosphate buffer pH 7.2 20 mM.

Peptide synthesis

The model peptides were provided by Mraiam Sai using Fmoc solid peptide phase synthesis on an 431A peptide synthesizer from Applied Biosystems (Foster City, CA, USA) and Fmoc protected *4R* and *4S* fluoroprolines from Bachem SA. Purification of the peptide was performed on a reversed-phase HPLC and checked using mass spectrometry (ESI-TOF) and electrospray ionization.

NMR measurements

The ^{19}F R_1 data was measured at 600 MHz on a Bruker Avance I spectrometer using the standard inversion recovery pulse sequence. For free *4S* FPRO in D₂O 27 data points were acquired with varying inversion-recovery build-up delays ranging from 1ms to 5 s. 1 Data point was repeated to estimate the experimental error. The spectral width and the carrier were set to 8 ppm and -176ppm respectively. The interscan relaxation delay was 12 s. 3 mm tube was inserted in 5mm tube containing H₂O/D₂O for lock signal.

For the mixture of *4R* and *4S* FPRO in glycerol, 27 data points were acquired with varying inversion-recovery build-up delays ranging from 1 ms to 3 s. 1 Data point was repeated to estimate the experimental error. The spectral width and the carrier were set to 8 ppm and -176

ppm, respectively. The interscan relaxation delay was 4 s. 3 mm tube was inserted in 5mm tube containing H₂O/D₂O for better lock signal.

For the ¹H–¹⁹F NOE measurement, the H_γ (5.42 ppm) was irradiated by a selective soft sinc1.1000 shaped train of 180° pulses at 700 Hz. Each pulse has a duration of 2.8 ms and was applied each 4 ms preceding the acquisition on the fluorine channel. 15 data points were acquired at 15 different saturation times ranging from 10 ms to 2.6 s. 1 Data point was repeated to estimate the experimental error.

The R2 relaxation rates were measured using standard spin echo and CPMG experiments. The carrier frequency was set to –174 ppm for the model peptides and -175.3 ppm for free 4*S* FPro. The spectral width was 12 ppm. D1 was 4s and 12s for the model peptides and free FPro respectively. for the CPMG sequence, the half spin echo was 200 us for the model peptides and 100 us for the free 4*S* fluoroproline. 180° proton pulse was applied every 2.8 ms at the fluorine echo time to average cross-correlation effects and ensure a single exponential decay. 22 data points were recorded ranging from 1ms to 4.5 s with one repeated data point for error estimation.

The processing of 1D spectra was performed using “Spectrometry Processing Innovative Kernel” (SPIKE) that is an open-source Python package. An exponential line broadening of 8 Hz was applied prior to Fourier transform for signal apodization. 2D ¹⁹F spectra were also processed on SPIKE.

Appendix 1

Longitudinal relaxation rates measured within 4*S* fluoroproline sample at varying temperatures.

Temp(K)	R ₁ (s ⁻¹)
278	0.67±0.008
281	0.62±0.003
284	0.58±0.009
287	0.53±0.005
290	0.49±0.01
293	0.45±0.01
296	0.41±0.03
299	0.39±0.02
302	0.36±0.02
305	0.34±0.01
308	0.33±0.005
311	0.31±0.004
314	0.32±0.005
317	0.3±0.005
320	0.28±0.001

Appendix 2

Longitudinal relaxation rates measured in mixture of free fluoroprolines *4R* and *4S* at different viscosities at 278 K. The relative viscosity is determined from the chemistry handbook.⁸

Solution	R ₁ FPRO <i>4S</i> (s ⁻¹)	R ₁ FPRO <i>4R</i> (s ⁻¹)	$\frac{viscosity_{glycerol}}{viscosity_{water}}$	Volumes
S0	0.77±0.02	0.70±0.02	1.000	180μl FPro
S1	1.58±0.02	1.53±0.02	3.276	vG :180 μl,VH ₂ O :266.2 μl vFPRO :49.57 μl
S2	2.15±0.02	2.17±0.02	4.434	vG :180 μl,VH ₂ O :177 μl vFPRO :39.6 μl
S3	2.44±0.04	2.49±0.03	10.660	vG :180 μl,VH ₂ O :72 μl vFPRO :28 μl
S4	2.83± 0.03	2.89± 0.03	13.630	vG :180 μl,VH ₂ O :54 μl vFPRO :26 μl
S5	2.84± 0.03	2.90± 0.1	18.420	vG :180 μl,VH ₂ O :37 μl vFPRO :24.2 μl
S6	3.00± 0.03	2.96± 0.04	25.570	vG :180 μl,VH ₂ O :24.6 μl vFPRO :22.7 μl
S7	2.87± 0.05	3.15± 0.06	40.490	vG :180 μl, VH ₂ O :11 μl vFPRO :21.2 μl
S8	3.11± 0.03	3.25± 0.14	59.780	vG :180 μl,vFPRO :20 μl
S9	3.33± 0.04	3.49± 0.12	84.170	vG :180 μl,vFPRO :7.83 μl
S10	3.41± 0.06	3.41± 0.12	147.200	from sol 98%: vG :180 μl,VH ₂ O :5.13 μl vFPRO :20.57 μl
S11	3.10± 0.08	2.99± 0.1	265.450	from sol 98%: vG :180 μl,vFPRO :20 μl
S12	2.73± 0.12	2.74± 0.18	383.700	from sol 98%: vG :180 μl,vFPRO :19.5 μl

With:

VG: volume of the glycerol

VH₂O: volume of the water

VFPRO: volume of the fluoroproline mixture

When not indicated the glycerol solution is 87%(wt), otherwise it is a 98% (wt)

Bibliography

- (1) Deepankumar, K.; Nadarajan, S. P.; Ayyadurai, N.; Yun, H. Enhancing the Biophysical Properties of MRFP1 through Incorporation of Fluoroproline. *Biochem. Biophys. Res. Commun.* **2013**, *440* (4), 509–514. <https://doi.org/10.1016/j.bbrc.2013.09.062>.
- (2) Steiner, T.; Hess, P.; Bae, J. H.; Wiltshi, B.; Moroder, L.; Budisa, N. Synthetic Biology of Proteins: Tuning GFPs Folding and Stability with Fluoroproline. *PLoS ONE* **2008**, *3* (2), e1680. <https://doi.org/10.1371/journal.pone.0001680>.
- (3) Crespo, M. D.; Rubini, M. Rational Design of Protein Stability: Effect of (2S,4R)-4-Fluoroproline on the Stability and Folding Pathway of Ubiquitin. *PLoS ONE* **2011**, *6* (5), e19425. <https://doi.org/10.1371/journal.pone.0019425>.
- (4) Holzberger, B.; Obeid, S.; Welte, W.; Diederichs, K.; Marx, A. Structural Insights into the Potential of 4-Fluoroproline to Modulate Biophysical Properties of Proteins. *Chem. Sci.* **2012**, *3* (10), 2924. <https://doi.org/10.1039/c2sc20545a>.
- (5) Tang, H.-C.; Lin, Y.-J.; Horng, J.-C. Modulating the Folding Stability and Ligand Binding Affinity of Pin1 WW Domain by Proline Ring Puckering: Modulate WW Domain by Proline Ring Puckering. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (1), 67–76. <https://doi.org/10.1002/prot.24359>.
- (6) Somovilla, V. J.; Bermejo, I. A.; Albuquerque, I. S.; Martínez-Sáez, N.; Castro-López, J.; García-Martín, F.; Compañón, I.; Hinou, H.; Nishimura, S.-I.; Jiménez-Barbero, J.; Asensio, J. L.; Avenoza, A.; Busto, J. H.; Hurtado-Guerrero, R.; Peregrina, J. M.; Bernardes, G. J. L.; Corzana, F. The Use of Fluoroproline in MUC1 Antigen Enables Efficient Detection of Antibodies in Patients with Prostate Cancer. *J. Am. Chem. Soc.* **2017**, *139* (50), 18255–18261. <https://doi.org/10.1021/jacs.7b09447>.
- (7) Torbeev, V. Yu.; Hilvert, D. Both the *Cis* - *Trans* Equilibrium and Isomerization Dynamics of a Single Proline Amide Modulate B2-Microglobulin Amyloid Assembly. *Proc. Natl. Acad. Sci.* **2013**, *110* (50), 20051–20056. <https://doi.org/10.1073/pnas.1310414110>.
- (8) The Properties of Ordinary Water Substance. Compiled by N. E. Dorsey, New York (Reinhold), 1940 (Reprinted 1953). Pp. Xxiv. 673; 12 Figs., 289 Tables. 120s. *Q. J. R. Meteorol. Soc.* **1955**, *81* (347), 132–133. <https://doi.org/10.1002/qj.49708134736>.
- (9) Aliev, A. E.; Kulke, M.; Khaneja, H. S.; Chudasama, V.; Sheppard, T. D.; Lanigan, R. M. Motional Timescale Predictions by Molecular Dynamics Simulations: Case Study Using Proline and Hydroxyproline Sidechain Dynamics: Proline Force Field Parameters. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (2), 195–215. <https://doi.org/10.1002/prot.24350>.

Chapter 4:
Using ^1H - ^{19}F cross-relaxation rates
within fluoroprolines to characterize
the rotational diffusion tensor in PPII
helices

IV. Using ^1H - ^{19}F cross relaxation rates within fluoroprolines to characterize the rotational diffusion tensor in PPII helices

Relaxation involves the modulation by random molecular motions of interactions that are intrinsically anisotropic. Thus, the careful interpretation on relaxation parameters may reveal some information about the molecular shape if the magnetic interaction that is effective for the relaxation is oriented in the molecular frame. This is the case for the relaxation of ^{15}N in proteins that is dominated by the dipolar interaction with the bound amide proton. Analysis of relaxation rates of ^{15}N for many sites within a folded protein allows the fully asymmetric diffusion tensor to be determined^{1 2}. The anisotropy of the chemical shift may also be used if the corresponding tensor is well characterized as for the ^{13}C relaxation of the carbonyl carbon of proteins^{3 4}. The cross-relaxation rate provides an interesting tool to probe the anisotropic nature of molecular motions: it involves a pair of spins and the relaxation depends on the orientation of the vector linking their positions and it is an observable that is devoid of exchange contribution. These properties have been exploited in the past, for instance to study the anisotropic motion of DNA double strands⁵. The use of fluorine-proton heteronuclear cross relaxation rates as a probe for the molecular diffusion tensor is appealing for several reasons: (i) the 4*R* fluoroproline is strongly biased to a single conformation of the cycle ($c_\gamma\text{exo}$), thus the distances between the fluorine and the protons of the cycle as well as the relative orientations of the corresponding inter-nuclei vectors may be considered as constant, (ii) the geometry of the proline cycle is such that the fluorine-proton vectors do sample very different orientations (the $F_\gamma - H_\gamma$ vector is orthogonal to the $F_\gamma - H_\beta$ for instance), (iii) the fluorine induces a chemical shift dispersion of the protons that facilitates the measurement of the individual magnetization, (iv) a proline homopolymer is a highly anisotropic molecule, hence this method may be used to get relative orientations of polyproline stretches.

This chapter aims at exploring the possibility to get reliable relaxation data that would allow the orientation of a molecular diffusion tensor. We addressed two parts of the problem: the simulations of the magnetization transfer between the protons and the fluorine within a 4-fluoroproline in order to probe the effect of anisotropic diffusion and the measurement of these magnetization transfers using ^{19}F -H HOESY experiment. We present the first results on model peptides composed of proline homopolymers where either a serine or a phospho-serine residue is inserted.

1. Proof of concept

The coordinates of the *4R* fluoroprolines' protons and fluorine atoms were obtained from the set of models issued from DFT calculation. A relaxation matrix was then computed using the corresponding inter-nuclei distances and a correlation time whose value was dependent from the relative orientation of the considered proton-fluorine vector in a 3D frame where the diffusion tensor is diagonal according the Woessner diffusion model (see chapter 1, equation 29). These computations were performed using in-house Python script and the Numpy/Scipy library. To simulate the effect of anisotropy on ^{19}F - ^1H cross relaxation rates, the rotational diffusion tensor was rotated around an arbitrary axis before the computation of the relaxation matrix. Variations of the cross-relaxation rates constants are shown in figure 1 for a rotation axis that led to the maximal variations of the values and an axially symmetric diffusion tensor (ellipsoid shape with axis ratio of 0.2). The cross-relaxation rate is sensitive to the orientation of the internuclear vector within the diffusion tensor, albeit the relative amplitude of the variations is significantly weaker for heteronuclear interactions than for homonuclear ones. The effect is however pronounced for the $F_\gamma - H_\gamma$ vector whereas it is less apparent for the other vectors due to the scaling by the distance (r^{-6}). The distance scaling also explains the larger values found for homonuclear cross-relaxation rates. Indeed, the $F_\gamma - H_\gamma$ distance is 2 Å while it is 1.7 Å for the H_β geminal protons, leading to a factor of 0.35 between expected heteronuclear and homonuclear cross-relaxation rates.

The NOE matrix was then computed for a series of mixing times and for either an isotropic or for an axially symmetric molecular rotational diffusion of the *4R* fluoroproline using the linalg library of the Numpy library to compute the exponential of the relaxation matrix (See chapter 2, equation 11). (Figure 2) The effects of anisotropic diffusion are clearly visible on the NOE buildup curves albeit, the differences are subtle. The most important differences are found for NOEs between $F_\gamma\text{-H}\delta^2$ and $F_\gamma\text{-H}\delta^3$ for instance.

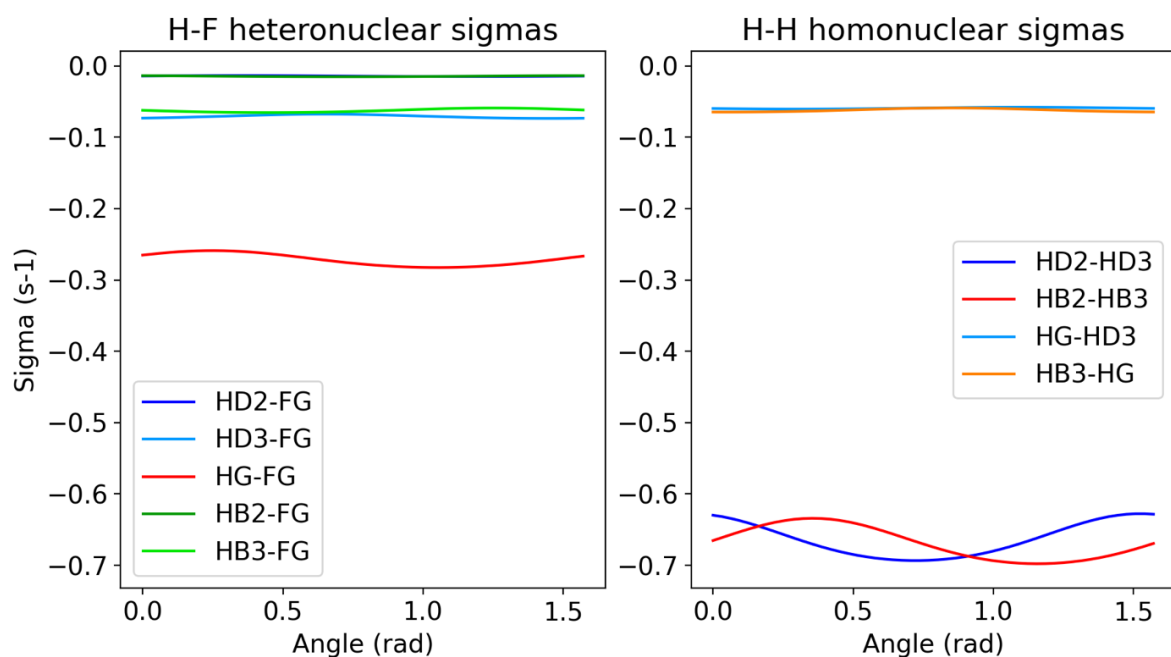


Figure 9 : Variation of the homonuclear and heteronuclear cross-relaxation rates σ_{HH} and σ_{FH} respectively upon evolution of the angle θ between the internuclear vector and the principal axis frame of an axially symmetric diffusion tensor of a *4R* fluoroproline (axis ratio=0.2). The correlation time was set to 1ns.

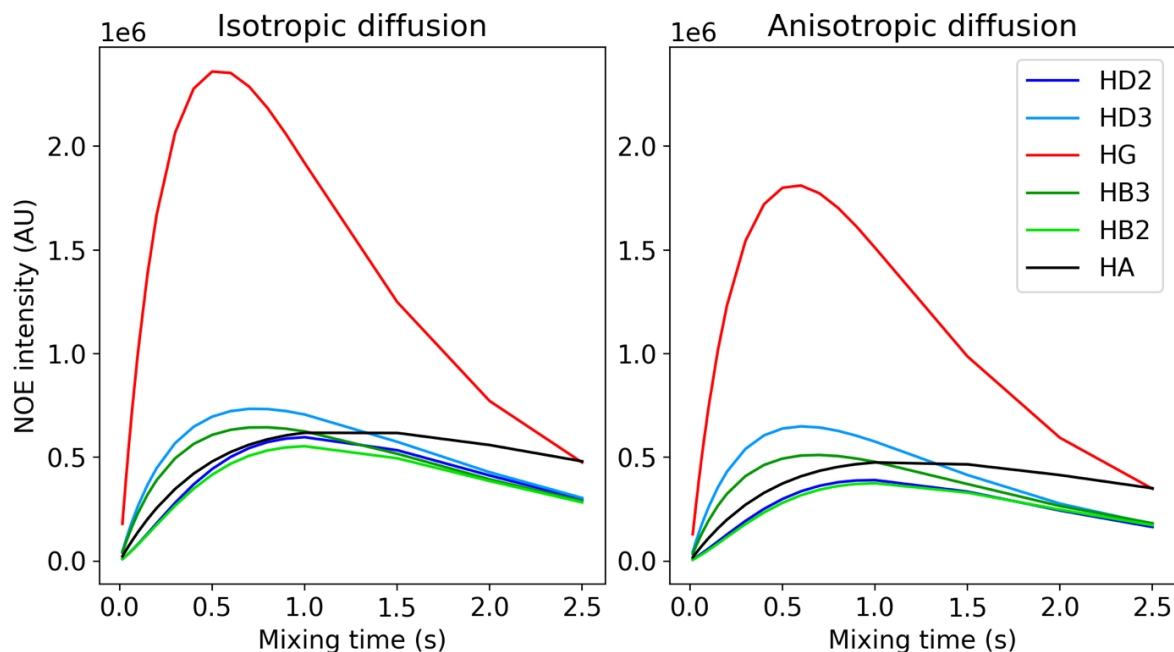


Figure 10 : Generated heteronuclear F-H NOE sets for *4R exo* fluoroproline in an isotopically diffusing molecule (left) and in an axially symmetric diffusion model (axis ratio=0.2) and a correlation time $\tau_c = 1 \text{ ns}$. Heteronuclear F_γ -H NOE are here denoted by the proton spin name. HD2 refers to F_γ - $H_{\delta 2}$ NOE.

We then investigated the ability to derive an orientation of a diffusion tensor from $F_\gamma - H$ heteronuclear NOE build-up curves. Different sets of synthetic NOE were computed with different initial parameters before applying a grid search procedure to compute the residual between the original synthetic dataset and each probed orientation (Figure 3). Residuals projected to a unit sphere display two polar regions corresponding to broad minima as expected for a symmetric system. The NOE buildups seem to be effective to discard orthogonal orientations shown as a purple belt on the figure.

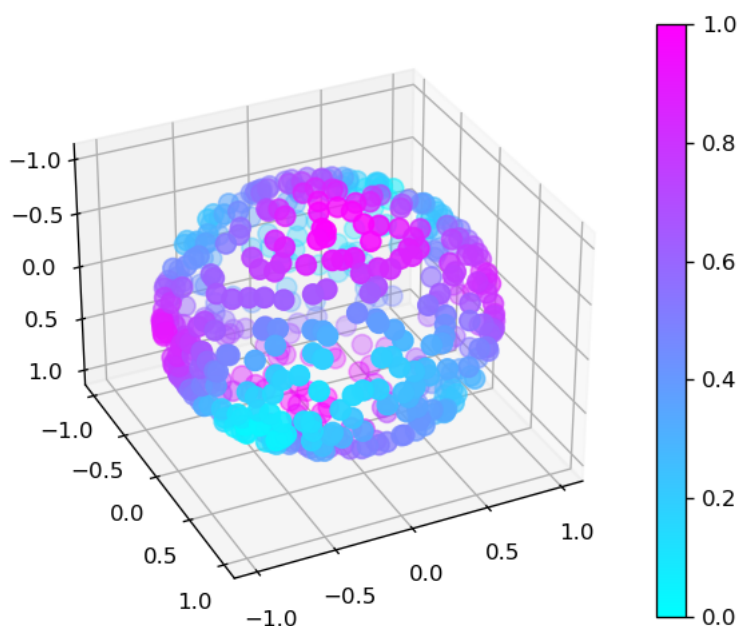


Figure 11 : Grid search to orient the diffusion tensor in a three-dimensional grid. The bar axis represents the normalized residual between the experimental and the synthetic $F_\gamma - H$ NOE.

2. Optimization of 2D ^1H - ^{19}F HOESY sequence

The original HOESY experiment was described by Yu and Levy in 1984.⁶ A starting 90° pulse applied on the proton channel transforms the longitudinal magnetization into a transverse magnetization. The proton magnetization is allowed to precess during the two $t_{\frac{1}{2}}$ delays for chemical shift labeling in the indirect F1 dimension. Applying a 180° on the fluorine channel in the middle of the t_1 delay refocuses the dephasing due to heteronuclear ^{19}F - ^1H couplings. Longitudinal proton magnetization is created by a 90° pulse and left for a mixing time τ_m during

which cross relaxation induces magnetization exchanges between protons and fluorine. At the end of the mixing time, a reading 90° pulse is applied on the fluorine channel. Initial trials of the standard HOESY pulse sequence showed that it was effective at relatively long mixing times ($\tau_m > 70$ ms), however for small mixing times of around 15 ms, artefactual correlation peaks were observed. Figure 4. These latter are believed to originate from the zero-quantum interferences. To avoid them, we implemented a zero quantum suppression element described by Thrifpleton and Keeler in 2003 composed of a gradient and a swept pulse as shown in figure 5.⁷ The application of the gradient relates the Larmor frequency of the spins to their positions in the NMR tube. The 180° swept-pulse will then inverse the spins that are at different positions in the tube at different times. The acquired zero quantum phases will thus vary with the spin positions in the sample and will then cancel out resulting in the suppression of zero quantum contribution.

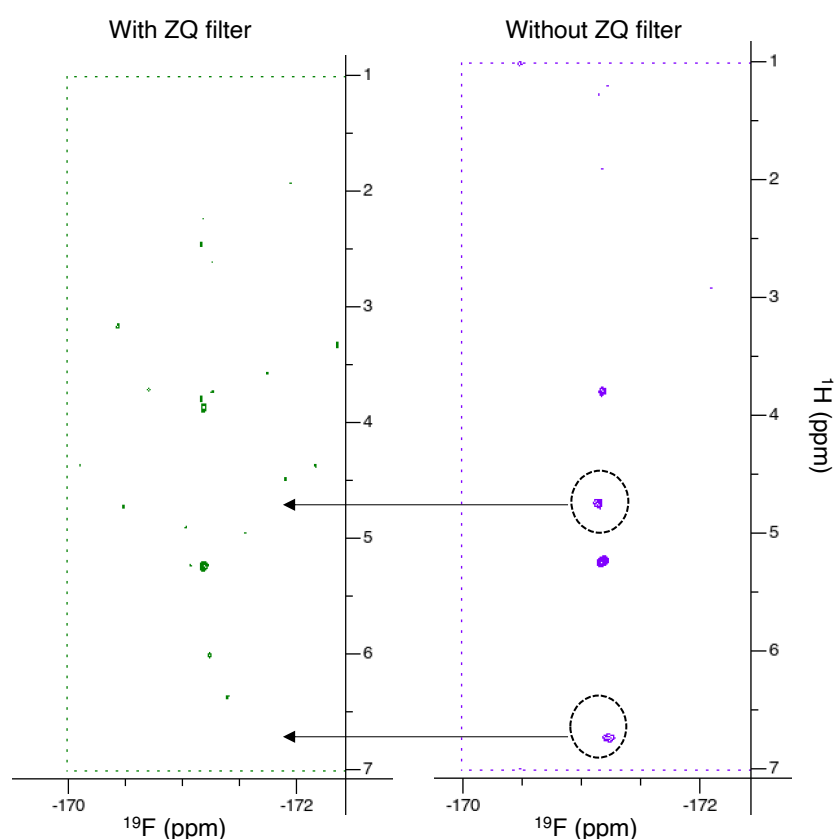


Figure 4: Recorded 2D ^1H - ^{19}F HOESY at 30 ms with (left) and without zero-quantum filter (right) in MpSR model peptide at 288 K at 600 MHz. The encircled signals are artefacts originating from zero-quantum (ZQ) interferences that disappear when the ZQ filter is applied.

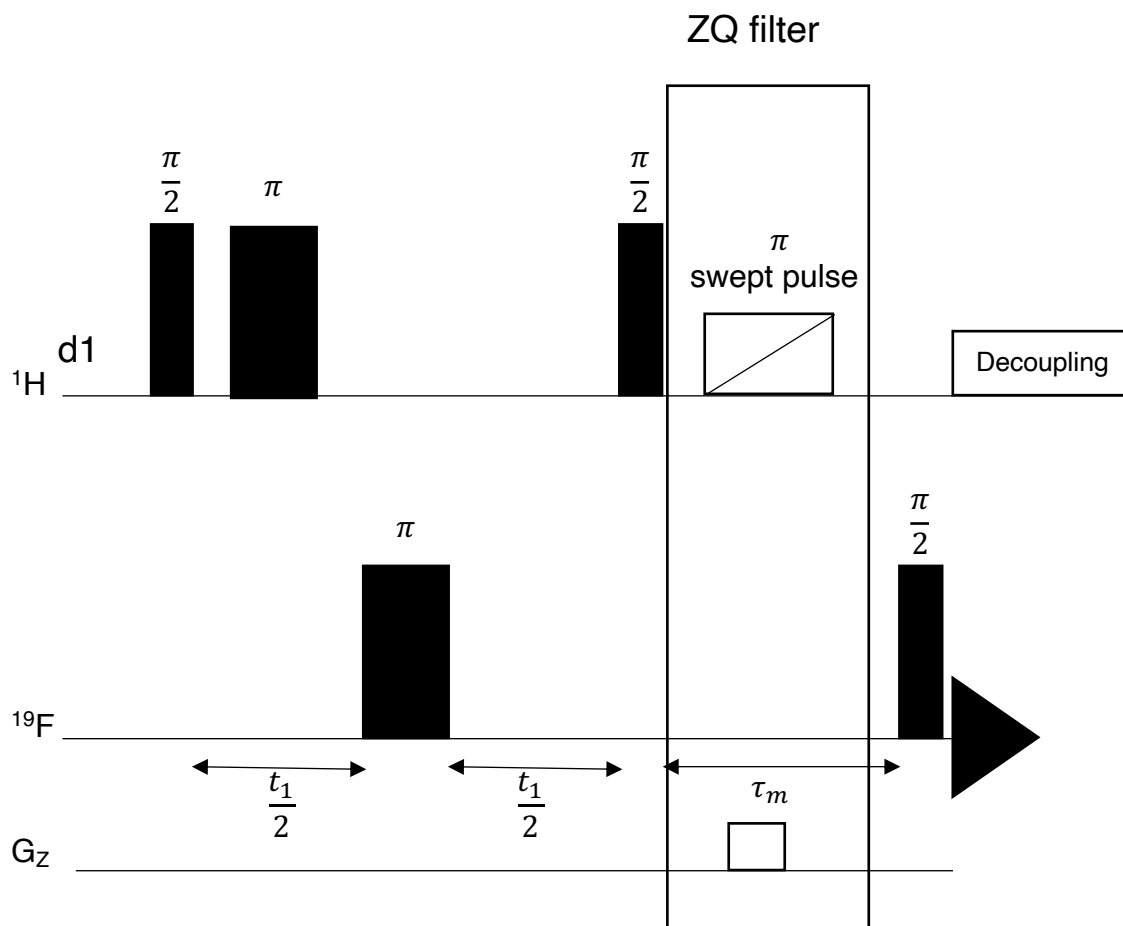


Figure 5: Pulse sequence of 2D ^1H - ^{19}F HOESY experiment with implemented zero-quantum filter shown in the box. The swept pulse is a crp20.10.20.1 applied during 10000 us (phase 0) and the gradient is a SINE.100 applied during 600 us.

3. Measurements of ^1H - ^{19}F NOE build-up in PPII helix

The modelling of heteronuclear cross-relaxation rates suggested the possibility to define the major axis of the rotational diffusion from the ^{19}F - ^1H NOE build-up curves. In the following, we measured these relaxation data for a set of model peptides composed of two segments of five prolines with an insertion of a fluoroproline in both segments (Figure 6). The aim was to establish whether some residues such as Alanine or Serine induce different relative orientations of the two polypyrroline segments, or if the phosphorylation of a serine leads to the same effect. In a first set of peptides, the first polypyrroline segment was incorporating a *4S* fluoroproline and

the second a *4R*. In order to investigate the possible effect of internal dynamics of the *4S* fluoroproline, we synthesized a second set of peptides with *4R* fluoroprolines at positions 4 and 8 (Figure 6).

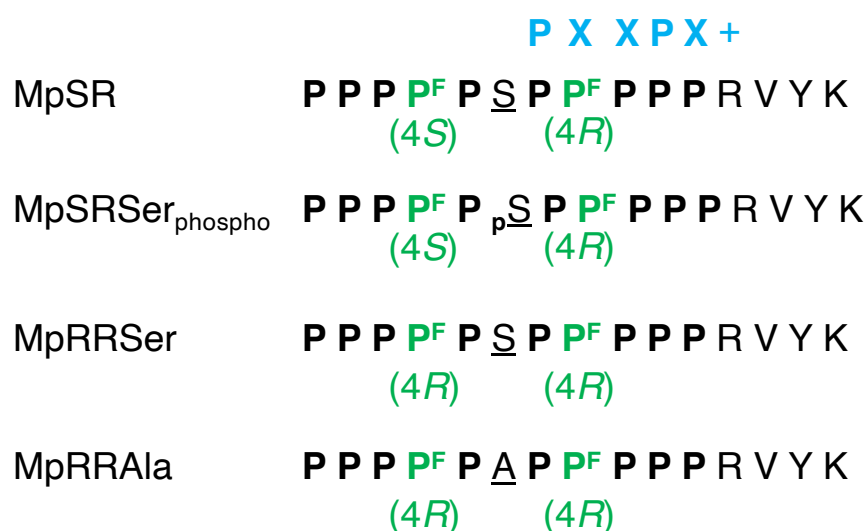


Figure 6: Designed model peptides sequences containing two fluoroprolines at two positions 4 and 8 highlighted in green. The muted amino acid separating the polyproline rich segments is underlined. The canonical SH3 domain binding motif is shown in blue.

Remarkably, despite a similar chemical (Pro) and structural environment (PPII helix), it was possible to distinguish the signals corresponding to the two *4R* fluoroproline in the fluorine spectrum enabling to monitor NOE build-ups at residue-specific locations in the peptide. (Figure 7)

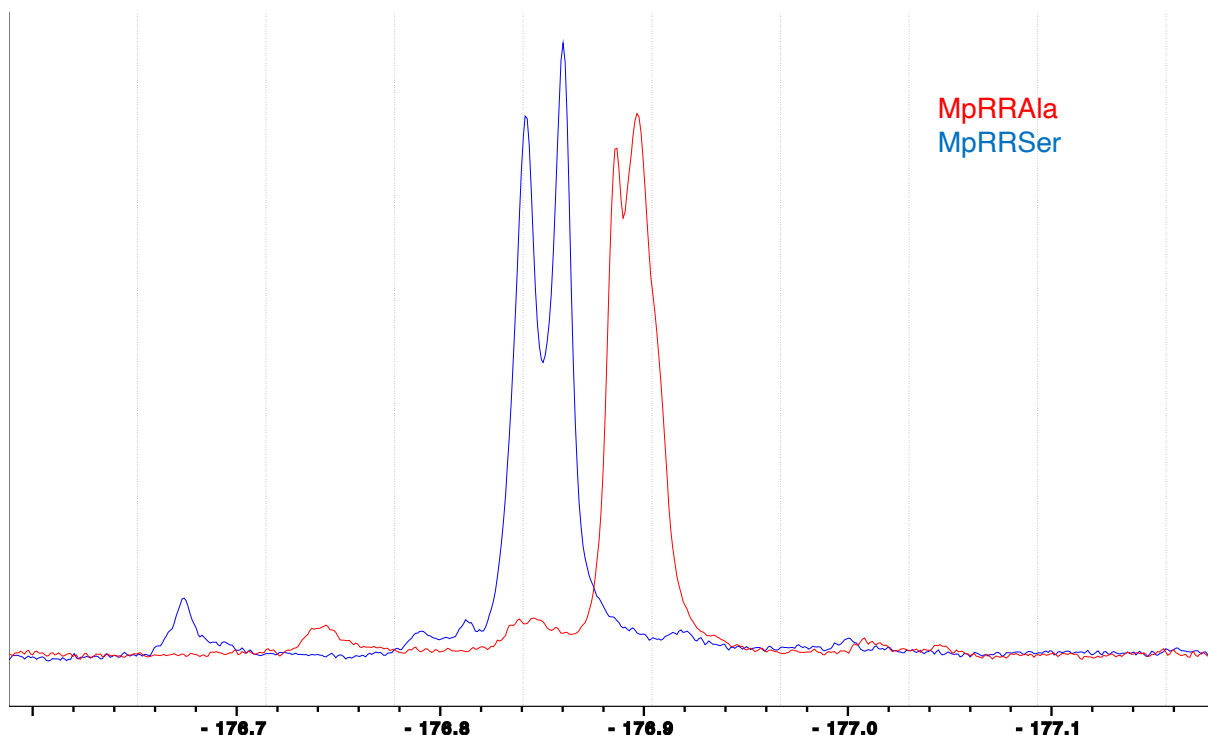


Figure 7: 1D ^{19}F spectra of model peptides MpRRAla and MpRRSer in phosphate buffer at 293 K. recorded at 600 MHz

3.1- ^1H - ^{19}F heteronuclear NOE build-up curves of MpSR model peptide

^1H - ^{19}F heteronuclear NOE build-up curves were established for MpSR At 288 K by recording a series of 2D ^1H - ^{19}F HOESY experiments at increasing mixing times (Figure 8). The assignment of the build-up curves was based on the known chemical shifts of the protons. To differentiate between geminal $H_{\delta 2}$ and $H_{\delta 3}$ as well as $H_{\beta 2}$ and $H_{\beta 3}$ we considered the F-H distances separating them from the fluorine atom. Thus, in *4R* fluoroproline $H_{\delta 3}$ being closer to the fluorine than $H_{\delta 2}$, will be the first to receive the transferred magnetization. Similarly, the equilibrium of $H_{\beta 3}$ is expected to be perturbed prior that of $H_{\beta 2}$.

Within MpSR peptide, we notice that the magnetization follows different routes in the five membered cycles *4R* and *4S* FPRO. This observation is consistent with the slightly different motion previously reported. Noteworthy, the distance matrices of the *4R* and *4S* fluoroprolines models that we used are not compatible with the observed magnetization pathways. For instance, it was not expected that H_{α} and both H_{β} in *4S* FPRO have similar build-up profiles, nor that the equilibrium of $H_{\delta 3}$ is perturbed prior both H_{β} . Similarly, $H_{\beta 3}$ being the first perturbed spin in *4R* FPRO even though our model indicated that $H_{\delta 3}$ is closer to F_{γ} is

intriguing. These observations suggest that internal motions may be effective in fluoroprolines and can't be discarded when interpreting these NOEs.

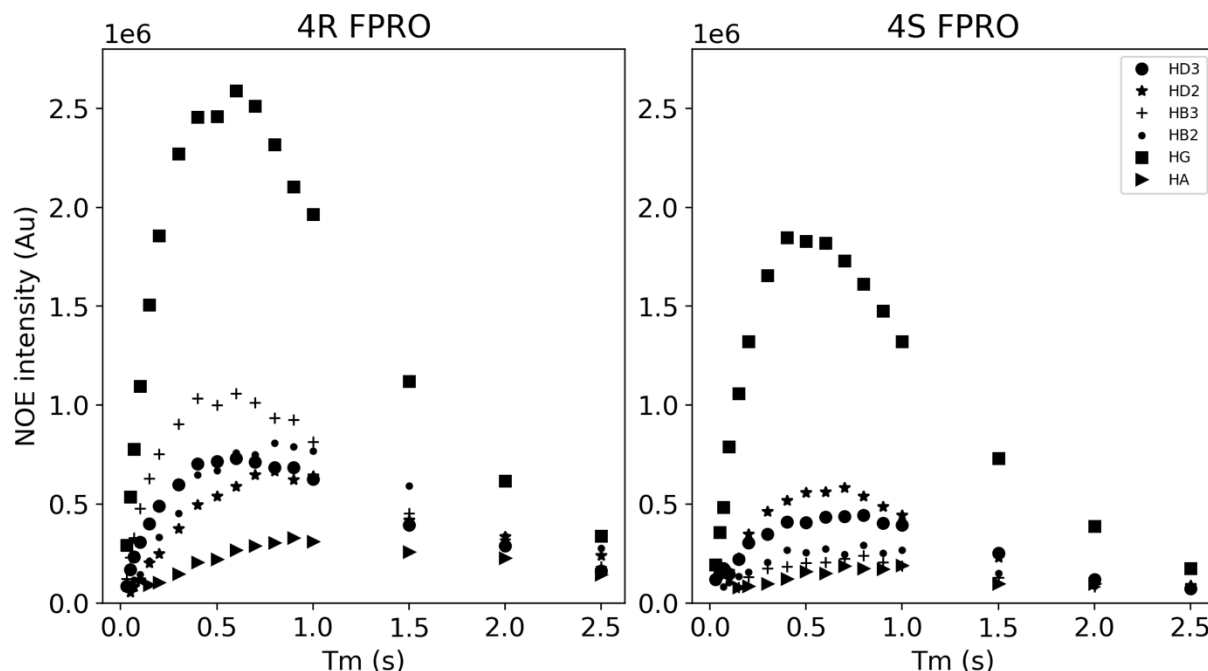


Figure 8: Build-up ^1H - ^{19}F NOE curves recorded in MpSR within *4R* (left) and *4S* (right) fluoroprolines in D₂O at 288 K. Recorded at 600 MHz. Heteronuclear F_{γ} -H NOE are here denoted by the proton spin name thus HD2 refers to F_{γ} - $H_{\delta 2}$ NOE.

3.2- ^1H - ^{19}F heteronuclear NOE build-up curves of MpSRSer_{phospho} model peptide

In chapter 3, we reported subtle differences in the relaxation properties for MpSR as compared to MpSRSer_{phospho}. Aiming at further understanding the origin of these differences, we recorded ^1H - ^{19}F heteronuclear NOE build-up curves for both peptides Figure 9. The intensities of the H-F NOE are very similar in both peptides suggesting that serine phosphorylation does not lead to a different orientation of the two polyproline segments.

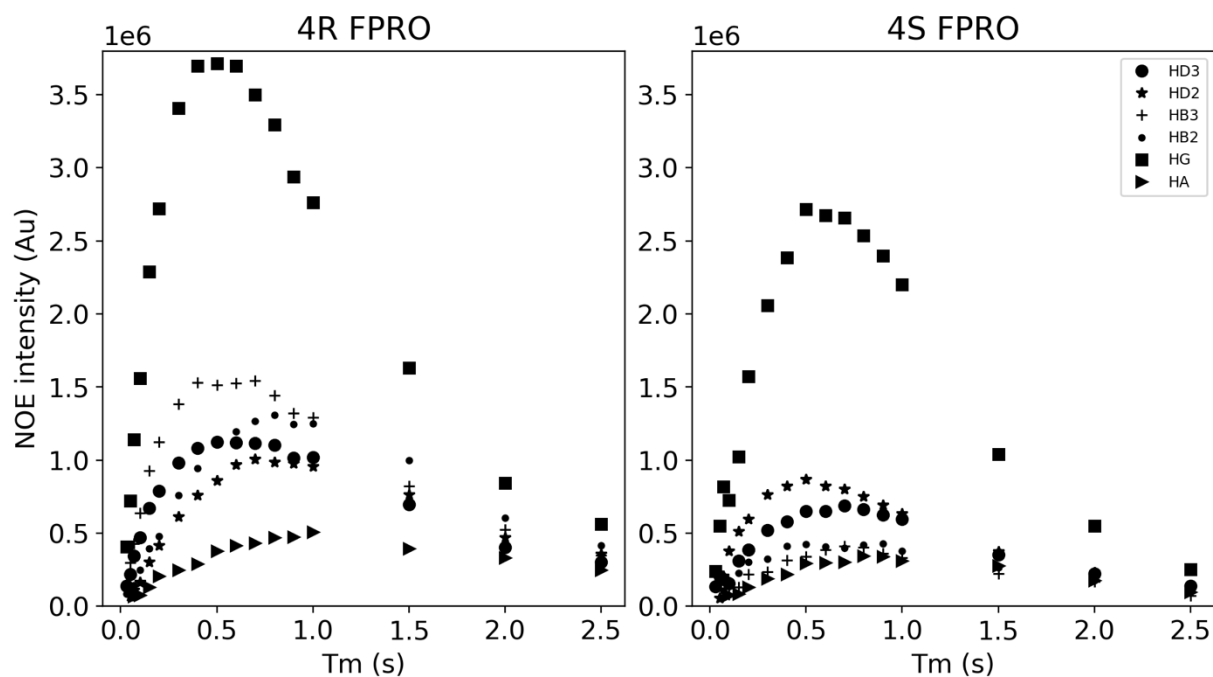


Figure 9: Build-up ^1H - ^{19}F NOE curves recorded in MpRRSer_{phospho} within *4R* (left) and *4S* (right) fluoroproline in D₂O at 288K. Recorded at 600 MHz.

3.3- ^1H - ^{19}F heteronuclear NOE build-up curves of MpRRSer and MpRRAla model peptides

We then investigated the effect of substituting a serine by an alanine within a PPII helix. Serine is known as a PPII breaker while Alanine is favorable for this secondary structure. We proceeded to also substitute the fluoroproline *4S* by a *4R* fluoroproline (this choice came after demonstrating that *4S* fluoroproline in the model peptide context doesn't alter the PPII helix, this will be discussed in chapter 5). While fluorine chemical shifts of the two *4R* fluoroproline within MpRRSer and MpRRAla peptides are separated, this was not the case for proton resonances making sequential assignment difficult to achieve (in the following, they will be labeled 1 and 2). Their corresponding build-up ^1H - ^{19}F NOE curves are displayed on figure 10 and 11.

The ^1H - ^{19}F NOE build-up curves are very similar for *4R* fluoroproline at positions 4 and 8 of the MpRRSer peptide suggesting a similar overall dynamic in both polyproline segments. This finding suggests that, in contrast to what was expected, the serine residue does not disconnect the motion of the polyproline segments in MpRRSer.

On the other hand, the pathways of the magnetization transfer in the two *4R* fluoroproline within MpRRAla are different. In *4R* FPRO 2, H_β spins are the first relaxing spins while only

$H_{\beta 3}$ display a behavior that differs from all other protons in *4R* FPRO 1. This latter behavior is similar to that recorded in MpSR and MpRRSer. Observations such as these may indicate that the alanine disconnects the relative motions of two polyproline segments. Another hypothesis would be that the Ser to Ala substitution affects the folding of the four C-terminal residues of the peptide leading to distinct diffusion behavior.

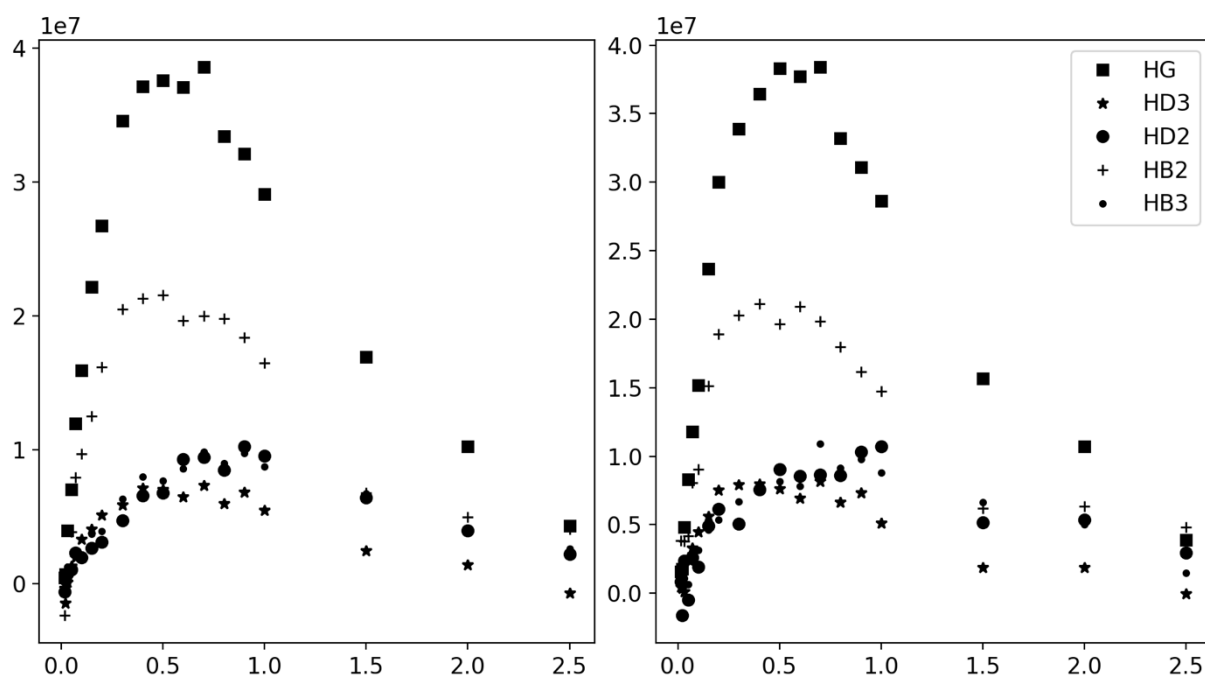


Figure 10: Build-up ^1H - ^{19}F NOE curves recorded in MpRRSer within *4R* FPRO recorded at 293 K at 600 MHz.

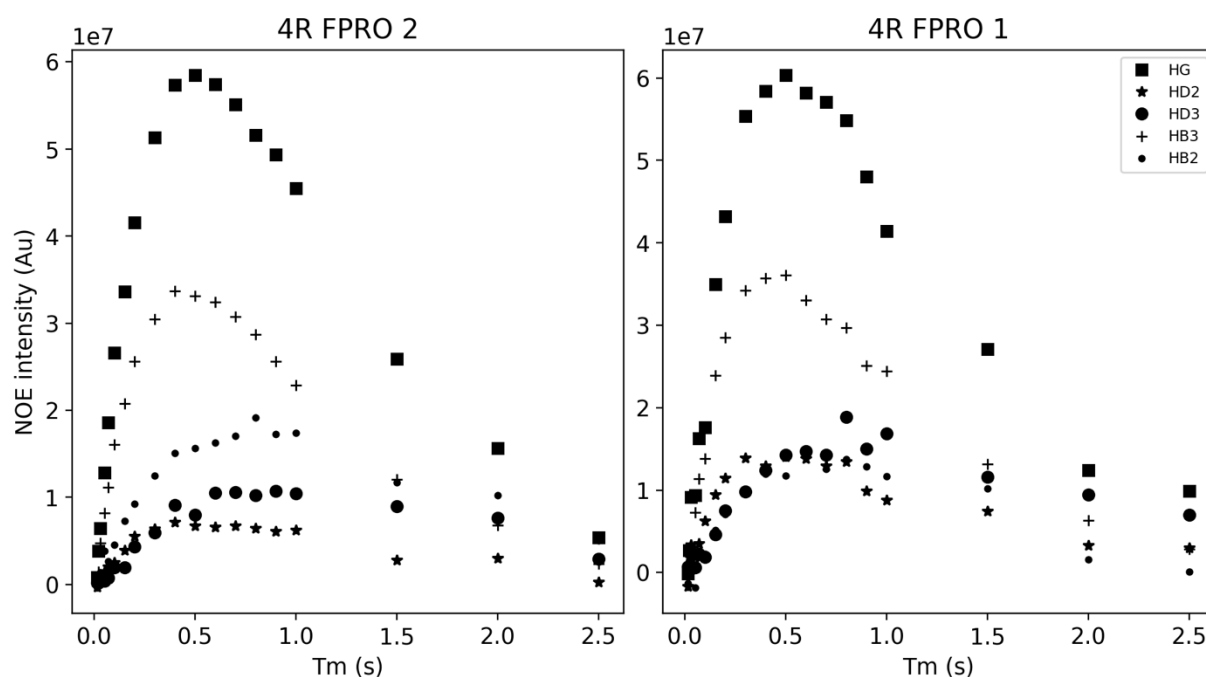


Figure 11: Build-up ^1H - ^{19}F NOE curves recorded in MpRRAla within *4R* FPRO recorded at 293 K at 600 MHz.

4. Conclusion and future work

In this chapter we suggested an approach based on the heteronuclear H-F NOE in 4-fluorinated prolines to characterize their diffusion tensors in peptides. For this we assume that proline is a rigid system. So far, we qualitatively showed that this approach is promising by simulating synthetic heteronuclear NOE considering an isotropic and an axially symmetric diffusion tensor and measuring a set of NOE buildup data. To further this preliminary work, several directions will have to be undertaken. The *4R* and *4S* fluoroproline may be inserted in simpler models composed of pure prolines, in order to assess the presence of internal motions and evaluate experimentally the effect of the global anisotropy (by using peptides of varying lengths). As suggested by our preliminary studies, completing the heteronuclear data by homonuclear NOE buildup data may significantly increase the precision of the procedure, provided that internal motions are handled properly. Finally, procedures to fit diffusion tensors to NOE buildup data will be elaborated and the expected precision in the orientation of the diffusion tensor will be quantitatively estimated.

5. Experimental section

All peptides were synthesized provided by Mariam Sai from the protein chemistry laboratory using solid phase peptide synthesis, and purified using HPLC and mass spectrometry to over 95%. NMR samples were prepared by dissolving the powder peptides in phosphate buffer 20 mM at pH 7.2.

2D ^1H - ^{19}F HOESY experiments were recorded at 288K (for MpSR and MpSRSer_{phospho}) and 293 K (for MpRRSer and MpRRAla) at 600 MHz using the pulse sequence described in the text. 1024 and 128 points were recorded in F2 and F1 domains respectively. The spectral width was set to 1.2 and 6 ppm. The resulting FID resolution was 0.66 Hz and 28.1 Hz in F2 and F1 domains respectively. The carrier OIP was fixed at -176.7 ppm. The shaped pulse was a crp20.10.20.1 applied during 10000 us. A SINE.100 gradient was applied during 600 us. The proton was decoupled during the acquisition using a GARP during 300 us. 16 scans were accumulated. 19 mixing times were sampled: 15, 20, 30, 50, 75, 100, 150, 200, 300, 500, 600, 700, 800, 900, 1000, 1500, 2000 and 25000 ms. The time experiment ranged from 1 h 47 minutes to 2 h 54 min for T_m=15 ms and 2.5 s respectively.

Spectra processing and peak peaking were performed using spike software on a Jupyter notebook. The peak volume was extracted using a box method drawn around the peak coordinates.

Bibliography

- (1) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin from ^{15}N NMR Relaxation. *J. Am. Chem. Soc.* **1995**, *117* (50), 12562–12566. <https://doi.org/10.1021/ja00155a020>.
- (2) Brüschweiler, R.; Liao, X.; Wright, P. E. Long-Range Motional Restrictions in a Multidomain Zinc-Finger Protein from Anisotropic Tumbling. *Science* **1995**, *268* (5212), 886–889. <https://doi.org/10.1126/science.7754375>.
- (3) Wylie, B. J.; Sperling, L. J.; Nieuwkoop, A. J.; Franks, W. T.; Oldfield, E.; Rienstra, C. M. Ultrahigh Resolution Protein Structures Using NMR Chemical Shift Tensors. *Proc. Natl. Acad. Sci.* **2011**, *108* (41), 16974–16979. <https://doi.org/10.1073/pnas.1103728108>.
- (4) Walker, O.; Mutzenhardt, P.; Haloui, E.; Boubel, J.-C.; Canet, D. Determination of the Rotation-Diffusion Tensor Orientation from NMR ^{13}C - ^1H Cross-Relaxation Rates. *Mol. Phys.* **2002**, *100* (17), 2755–2761. <https://doi.org/10.1080/00268970210141199>.
- (5) Birchall, A. J.; Lane, A. N. Anisotropic Rotation in Nucleic Acid Fragments: Significance for Determination of Structures from NMR Data. *Eur. Biophys. J.* **1990**, *19* (2). <https://doi.org/10.1007/BF00185089>.
- (6) Yu, C.; Levy, G. C. Two-Dimensional Heteronuclear NOE (HOESY) Experiments: Investigation of Dipolar Interactions between Heteronuclei and Nearby Protons. *J. Am. Chem. Soc.* **1984**, *106* (22), 6533–6537. <https://doi.org/10.1021/ja00334a013>.
- (7) Thriddleton, M. J.; Keeler, J. Elimination of Zero-Quantum Interference in Two-Dimensional NMR Spectra. *Angew. Chem. Int. Ed.* **2003**, *42* (33), 3938–3941. <https://doi.org/10.1002/anie.200351947>.

Chapter 5:
Fluorine NMR study of proline-rich
sequences using fluoroprolines
(Publication MR journal)

V. Fluorine NMR study of proline-rich sequences using fluoroprolines

Davy Sinnaeve^{1,2}, Abir Ben Bouzayene³, Emile Ottoy⁴, Gert-Jan Hofman^{4,5}, Eva Erdmann³, Bruno Linclau⁵, Ilya Kuprov⁵, José C. Martins⁴, Vladimir Torbeev⁶ and Bruno Kieffer³

¹Univ. Lille, Inserm, Institut Pasteur de Lille, CHU Lille, U1167 – Labex DISTALZ – RID-AGE – Risk Factors and Molecular Determinants of Aging-Related Diseases, F-59000 Lille, France

²CNRS, ERL9002 - Integrative Structural Biology, F-59000 Lille, France

³Département of Integrative Structural Biology, IGBMC, Université de Strasbourg, INSERM U1258, CNRS UMR7104, 1, rue Laurent Fries, F-67404 Illkirch, France

⁴Department of Organic and Macromolecular Chemistry, Ghent University, Campus Sterre, S4, Krijgslaan 281, B-9000 Gent, Belgium

⁵School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom

⁶Institut de Science et d'Ingénierie Supramoléculaires (ISIS), International Center for Frontier Research in Chemistry (icFRC), University of Strasbourg, CNRS UMR 7006, F-67000 Strasbourg, France

Correspondence to: B. Kieffer (kieffer@igbmc.fr) and D. Sinnaeve (davy.sinnaeve@univ-lille.fr)

Abstract. Proline homopolymer motifs are found in many proteins; their peculiar conformational and dynamic properties are often directly involved in those proteins' functions. However, the dynamics of proline homopolymers is hard to study by NMR due to lack of amide protons and small chemical shift dispersion. Exploiting the spectroscopic properties of fluorinated prolines opens interesting perspectives to address these issues. Fluorinated prolines are already widely used in protein structure engineering – they introduce conformational and dynamical biases – but their use as ¹⁹F NMR reporters of proline conformation has not yet been explored. In this work, we look at model peptides where C γ -fluorinated prolines with opposite configurations of the chiral C γ centre have been introduced at two positions in distinct polyproline segments. By looking at the effects of swapping these (4*R*)- and (4*S*)-4-fluoroprolines within the polyproline segments, we were able to separate the intrinsic conformational properties of the polyproline sequence from the conformational alterations instilled by fluorination. We assess the fluoroproline ¹⁹F relaxation properties, and exploit the latter in elucidating binding kinetics to the SH3 domain.

1. Introduction

The use of ^{19}F nuclei in medical and biological magnetic resonance is gaining popularity (Zhang et al., 2017). Since the pioneering incorporation of *p*-fluorophenylalanine (Chaiken et al., 1973) into ribonuclease-S' analogues, dozens of ^{19}F -labelled amino acid analogues have been evaluated (Odar et al., 2015; Mei et al., 2020; Muttenthaler et al., 2021, Salwiczek et al., 2012). Common ways to incorporate fluorinated amino acids in peptides or proteins are: (a) solid phase chemical synthesis (Behrendt et al., 2016); (b) post-translational addition of fluoroalkyl groups to reactive amino acid side chains (Liu et al., 2012); (c) addition of fluorinated precursors, such as fluoroindole, to bacterial culture media prior to protein overexpression (Crowley et al., 2012); (d) using recombinantly expressed orthogonal amber codon tRNA/tRNA synthetase pairs (Sharaf et al., 2015; Gimenez et al., 2021; Gee et al., 2016; Kitevski-LeBlanc et al., 2012). The advantages of ^{19}F nuclei in biological NMR are the absence of background signals, high gyromagnetic ratio, 100% natural abundance, and the sensitivity of ^{19}F chemical shift to the chemical environment (Rastinejad et al., 1995). Fluorine chemical shift range (~ 50 times wider than that of ^1H) makes it possible to study faster chemical exchange processes than those accessible to ^1H - and ^{13}C -based methods. This is useful in biomolecular interaction studies and examples include the deciphering of the signal transduction pathways through the $\beta 2$ -adrenergic trans-membrane receptor (Liu et al., 2012), the study of conformer interconversion and allostery that drive the catalytic process in the bacterial enzyme fluoroacetate dehalogenase (Kim et al., 2017), the monitoring of both kinetic and equilibrium thermodynamic binding parameters of a fluorine-labeled Src homology 3 (SH3) protein domain to peptides containing proline-rich motifs (PRM) (Stadmler et al., 2020), and the folding study of a small protein domain (Evanics et al., 2007). A downside of ^{19}F is high chemical shift anisotropy (CSA) – particularly in aromatic rings – resulting in rapid transverse relaxation and broad lines for large biomolecules at high magnetic fields (Kitevski-LeBlanc et al., 2012), although the recently proposed ^{19}F - ^{13}C aromatic TROSY experiment has alleviated this to some extent (Boeszoermenyi et al., 2019).

Fluorination is well known for its significant impact on the properties of organic molecules (Aufiero et al., 2018; Gillis et al., 2015, Berger et al., 2017). Apart from altering the interaction with the solvent (*cf.* hydrophobicity), replacing a hydrogen with fluorine can produce significant structural changes. Firstly, the volume of the moiety increases. Although fluorine is often considered isosteric to hydrogen based on its similar Van der Waals radius ($r_{\text{vdW}}(\text{F}) = 1.47 \text{ \AA}$ vs. $r_{\text{vdW}}(\text{H}) = 1.20 \text{ \AA}$) (Bondi et al., 1964) its covalent radius is significantly larger

($r_{\text{cov}}(\text{F}) = 0.57 \text{ \AA}$ vs. $r_{\text{cov}}(\text{H}) = 0.31 \text{ \AA}$) due to greater C–F bond length (Cordero et al., 2008; O’Hagan et al., 2008). As a result, fluorine may perturb the protein fold when the fluorinated side-chain is tightly packed within a protein structure. Secondly, the polar C–F bond brings in additional charge and polarisability effects (Salwiczek et al., 2012). In aromatic side chains, swapping a single hydrogen for fluorine does not normally (there are exceptions (Salwiczek et al., 2012; Boeszoermenyi et al., 2020; Yoshida et al., 1960)) alter the fold or the function of the protein (Welte et al., 2020). In contrast, fluorinating an aliphatic CH group can radically change local rotamer populations (O’Hagan et al., 2008; O’Hagan et al., 2012). This effect has been put to good use (Salwiczek et al., 2012; Berger et al., 2017), particularly in fluorinated prolines (Kubyskhin et al., 2021; Verhoork et al., 2018; Newberry et al., 2016).

Proline is the only proteinogenic amino acid with a secondary amino group, thus allowing for the *cis*-peptide bond isomer to be significantly populated. In addition, its pyrrolidine ring can adopt either a *C γ -endo* or *C γ -exo* conformation, with a slight preference for the former. Single or double fluorination at the β - and/or γ -positions shifts these conformational equilibria in a stereospecific way. For instance, (4*R*)-fluorination favours the *C γ -exo* ring conformer and enhances the *trans* isomer population, while (4*S*)-fluorination does the opposite (Fig. 1) (Eberhardt et al., 1996; Panasik et al., 1994).

This is caused by stabilizing $\text{C-H}\sigma_{(\text{HOMO})} \rightarrow \text{C-F}\sigma^*_{(\text{LUMO})}$ hyperconjugative delocalization, which is maximal when the C–H bond is antiperiplanar to the C–F bond, a phenomenon generally known as *gauche* effect (Thiehoff et al., 2017). The increased amount of *C γ -exo* conformer in (4*R*)-FPro in turn increases the *trans* isomer population since this is the most favourable configuration for further stabilizing hyperconjugative $\text{n} \rightarrow \pi^*$ delocalization between carbonyl groups in successive peptide bond (Newberry et al., 2016). Similarly, the reduced *C γ -exo* population as well as the steric impact from the longer C–F bond increases the population of the *cis*-isomer in (4*S*)-FPro. The increase or decrease of $\text{n} \rightarrow \pi^*$ hyperconjugation have been used to explain stabilization or destabilization of the polyproline-II (PPII) conformation in all-(4*R*)-fluorinated and all-(4*S*)-fluorinated oligoproline respectively (Horng et al., 2006). The ability to control the conformational preference of individual proline residues is central to elucidating the role of proline conformation on the stability, folding, and aggregation of various proteins, such as collagen (Holmgren et al., 1998; Shoulders et al., 2009), β_2 -microglobulin (Torbeev et al., 2015; Torbeev et al., 2013) and tau. (Jiji et al., 2016)

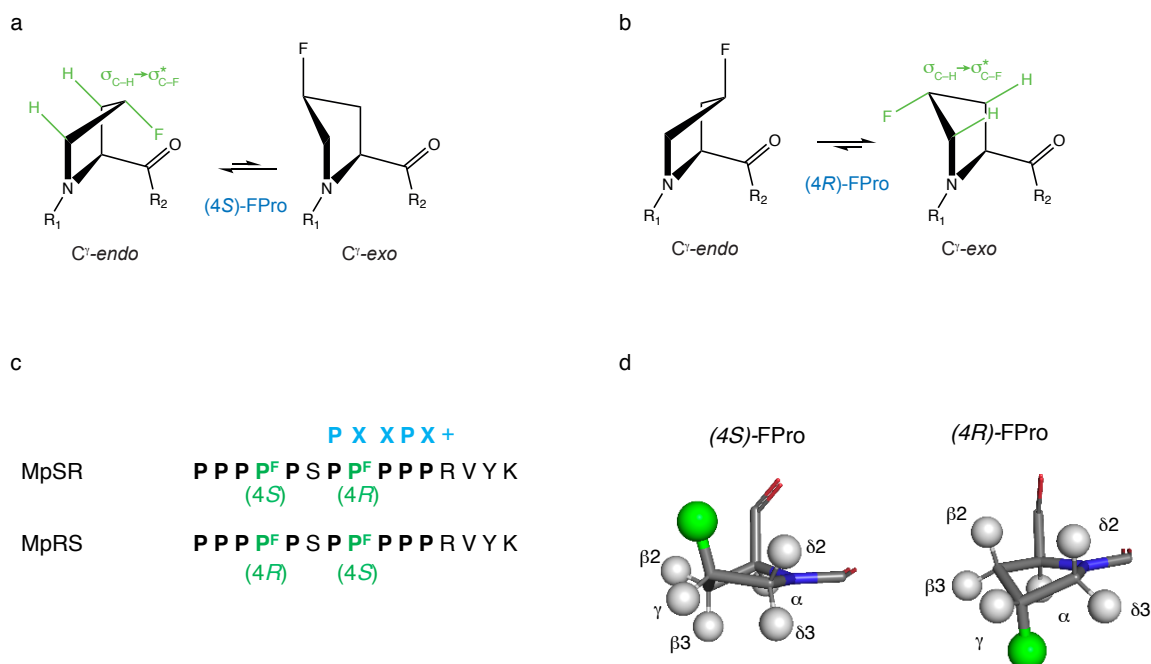


Figure 1: *C* γ -*endo* and *C* γ -*exo* puckering of the pyrrolidine ring in (4*S*)- and (4*R*)-fluoroprolines shown respectively in **a** and **b**. The *gauche* effect stabilizes the *C* γ -*endo* conformer of (4*S*)-fluoroproline, whereas the *C* γ -*exo* conformer is favoured in (4*R*)-fluoroproline. **c**: Fluoroprolines incorporated in a proline-rich sequence at two positions 4 and 8 highlighted in green. Two peptides are studied: in MpSR (4*S*)-fluoroproline is inserted at the fourth position and (4*R*)-fluoroproline is inserted at the eighth position. In MpRS, the positions of the (4*R*)- and (4*S*)-fluoroprolines are reversed, placing them in the position fourth and eighth, respectively. The canonical SH3 domain binding motif is shown in blue. **d, left side**: 3D model of (4*S*)-fluoroproline where H γ 2 is substituted by a fluorine atom. The carbonyl group and the fluorine atom point towards the same direction. **d, right side**: 3D model of (4*R*)-fluoroproline where H γ 3 is substituted by a fluorine atom. The carbonyl group and the fluorine atom point towards opposite sides.

Surprisingly, despite the well-established use of FPro residues in chemical biology, they have so far found very limited attention as ^{19}F NMR reporters in protein studies, in contrast to aromatic amino acids (Verhoork et al., 2018). In the limited protein or peptide studies that have used ^{19}F NMR, it was mainly used to confirm the local conformational state of the fluoroproline residue (Torbeev et al., 2013, Verhoork et al., 2018). To the best of our knowledge, only one study went further and exploited ^{19}F NMR of a *foldon* domain peptide containing (4*R*)-FPro and (4*S*)-FPro residues to monitor the folding/unfolding process as a function of temperature (Dietz et al., 2015). Yet the potential of FPro residues for advanced biomolecular ^{19}F NMR is clear given the abundance of proline in intrinsically disordered protein sequences, the prominent role of proline-rich regions as sites for protein-protein interaction and post-translational modification, the relatively small CSA of ^{19}F nuclei in prolines (thus, narrow lines), and the challenge of detecting minor *cis* isomers (Theillet et al., 2013). Possible explanations are the unknown ^{19}F NMR properties of these residues, as well as their undesirably strong conformational impact.

With the purpose of filling this gap, we have studied the impact of (4*R*)- and (4*S*)-FPro residues on the structure and dynamics of a polyproline peptide harbouring an SH3 binding motif, and used ^{19}F NMR to investigate the impact of fluorination on the binding affinity. We designed model peptides containing (4*R*)- and (4*S*)- fluorinated prolines with a sequence based on the motif located at the C-terminal part of the retinoic acid hormone nuclear receptor RAR γ that specifically binds to the third SH3 domain of the Vinexin β protein (Lalevee et al., 2010). First, we explored the impact of FPro introduction on the surrounding peptide sequence, and verified the preferred FPro ring pucker within the polyproline context. Next, we used ^{19}F relaxation analysis to gain insights into the local dynamics of the peptide. Finally, we monitored the interaction of the model peptides with the vinexin β SH3 domain using ^{19}F NMR, and demonstrated that FPro conformational bias can be used to modulate the kinetics of protein binding to proline-rich motifs. This work paves the way to using fluoroprolines as ^{19}F NMR reporters in protein interaction studies, where the conformational bias caused by fluorine is exploited to obtain information on binding kinetics.

2. Results

2.1. Assignment and spectral analysis of model peptides

The model peptide sequences shown in Fig. 1C contain two segments of five prolines separated by a single serine, and terminate with a four-residue sequence (RVYK) required for the SH3 class II binding specificity. FPro residues were inserted at positions 4 and 8, which are not directly involved at the protein-peptide interface according to homology models of PPII helices-SH3 complexes (Saksela et al., 2012). Position 4, located in the first polyproline segment, falls outside the expected PXXPX+ binding motif while proline 8, which is located within the canonical SH3-PPII binding motif, is expected to be solvent-exposed (Supplementary Fig.1). Thus, the fluorine atoms are not expected to contribute significantly to the protein-peptide binding interface. Two peptides were considered, with (4*R*)- and (4*S*)-FPro substitutions at positions 4 and 8 (hereafter named MpRS), or introduced at positions 8 and 4 (MpSR).

Full ^1H and ^{13}C chemical shift assignments of the non-proline and FPro residues in D_2O were achieved using standard ^1H - ^1H NOESY, ^1H - ^1H TOCSY and ^1H - ^{13}C HSQC experiments. The 8 non-fluorinated proline residues have very similar chemical shifts, but full assignment could still be achieved using a 2D ^1H - ^{13}C HSQC-NOESY experiment with very high ^{13}C digital resolution (*ca.* 4 Hz, see experimental section) (Fig. 2). For this, the spectral window was set to a narrow ^{13}C chemical shift region of 3 ppm containing the proline $\text{C}\delta$ resonances. To avoid interference from folded ^{13}C - $\text{H}\alpha$ autocorrelation peaks, a gradient-enhanced frequency-selective ^{13}C 180° refocusing pulse was applied in the HSQC experiment. At this spectral resolution, the minute $\text{C}\delta$ chemical shift dispersion (0.3 ppm for prolines 2 to 11 in MpSR) allowed resolving the sequential $\text{H}\delta(\text{i})$ to $\text{H}\alpha(\text{i}-1)$ NOE cross-peaks and thus completing ^1H and ^{13}C chemical shift assignment of both peptides (Table 1).

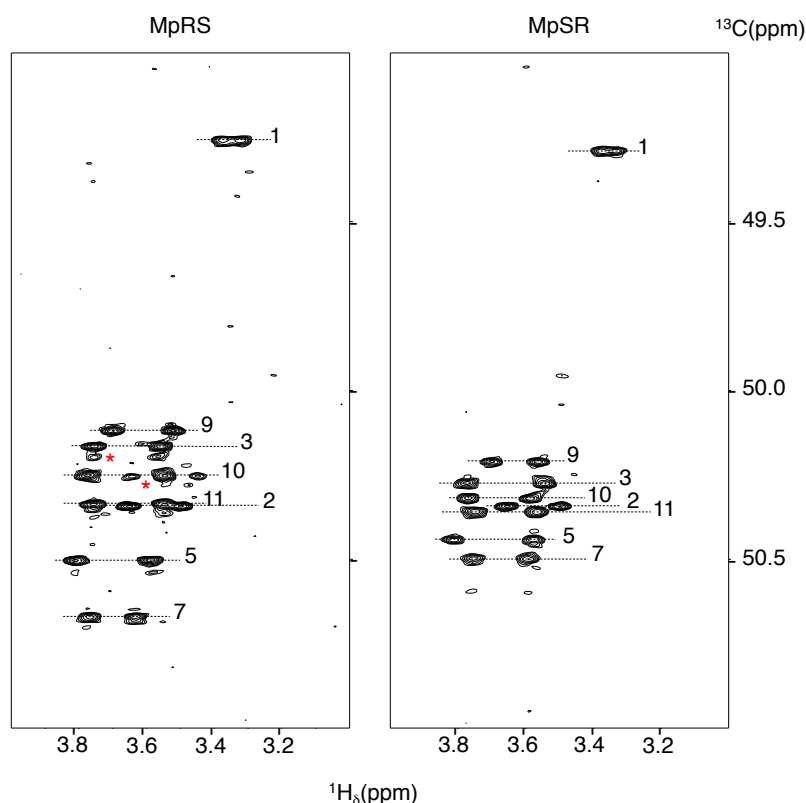


Figure 2: ^1H - ^{13}C HSQC-NOESY (mixing time: 80 ms) with a narrow ^{13}C window focussing on the $^{13}\text{C}\delta/\text{H}\delta$ correlations regions of both MpRS and MpSR peptides, recorded at 298 K and 700 MHz. The numbers indicate the position of the residue in the sequence. The red asterisks highlight minor forms of prolines.

When comparing the two peptides, ^1H and ^{13}C chemical shifts of each type of FPro (4*R* or 4*S*) turn out nearly identical, independently of their position in the sequence, suggesting that the local conformation of the pyrrolidine ring is not sensitive to the sequence context, but is dictated by the fluorination stereochemistry at the C_γ -center. To confirm this, we measured ^1H - ^{19}F and $^1\text{H}\alpha$ - $^1\text{H}\beta$ scalar couplings within the FPro residues (Table 2). Heteronuclear ^1H - ^{19}F couplings were measured in 2D ^1H - ^1H TOCSY spectra using the E.COSY cross-peak pattern $^1\text{H}\gamma$ - $^1\text{H}\beta$ and $^1\text{H}\gamma$ - $^1\text{H}\delta$ correlation peaks, while the homonuclear $^1\text{H}\alpha$ - $^1\text{H}\beta$ couplings were measured using SERF experiments (see experimental section) (Fig. 3). These latter couplings are diagnostic of the ring pucker, and a visual inspection of the coupling patterns observed for MpRS and MpSR peptides immediately indicates that both (4*R* or 4*S*) FPro retain their *exo* or *endo* ring pucker in the context of a polyproline segment. These scalar couplings are compared with the literature

values determined for the free amino acid (Table 2) (Gerig et al., 1973), and turn out to be very similar, except for the $^3J_{F-\delta 2}$ coupling in (4*S*)-FPro where a difference of about 5 Hz is seen. The reason for this is unclear, but could be due to the presence of either a neighbouring amide or amine group in the peptide or free amino acid, respectively. Thus, it can be concluded that the strong bias of the five-membered ring conformation introduced by 4-monofluorination (Gerig et al., 1973; DeRider et al., 2002) is fully preserved within the oligoproline context. Using density functional theory (M06/cc-pVDZ in SMD water), we previously calculated for Ac-FPro-OMe that the $C\gamma$ -*exo*: $C\gamma$ -*endo* population ratios are 93:7 for (4*R*)-FPro and 1:99 for (4*S*)-FPro (Hofman et al., 2018; Hofman et al., 2019). For the purpose of NMR conformation and dynamic analysis, it is thus fair to assume that only one ring conformer is present. It is known that proline normally interconverts between the $C\gamma$ -*exo* and $C\gamma$ -*endo* ring conformations within oligoprolines while adopting a PPII helix (Horng et al., 2006; Wilhelm et al., 2014). Similar to the concept of conformational frustration (Ferreiro et al., 2014), it can be stated that proline fluorination creates a form of “dynamic frustration” within the polyproline helix.

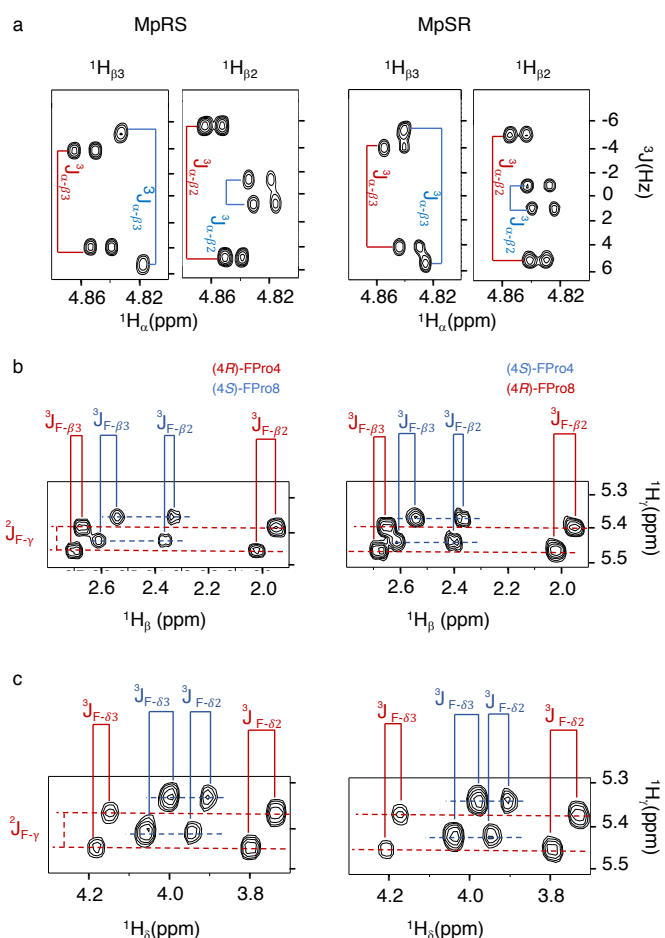


Figure 3: Puckering analysis of fluoroprolines. **a:** Homonuclear coupling constants $^3J_{H\alpha-H\beta 2}$ and $^3J_{H\alpha-H\beta 3}$ of the (4*R*) (in red) and (4*S*) (in blue) fluoroprolines in MpSR and MpRS peptides

measured from SERF experiments at 298 K. **b:** The $^3J_{F-H\beta}$ heteronuclear coupling constants extracted from 2D 1H - 1H TOCSY spectra using the E.COSY cross-peak pattern. **c:** The $^3J_{F-H\delta}$ heteronuclear coupling constants extracted from 2D 1H - 1H TOCSY spectra using the E.COSY cross-peak pattern.

For the non-fluorinated prolines, the ^{13}C chemical shifts are mostly similar in the MpRS and MpSR peptides (Table 1 and Supplementary Fig.2), suggesting that the overall conformational properties of the peptide are not greatly affected by the permutation of the two FPro residues. It is also observed that the insertion of the PPII destabilizing (4*S*)-FPro within both polyproline segments does not alter the intensities of the strong $H\alpha(i-1)$ to $H\delta(i)$ NOE cross-peaks observed between all prolines of the segment including the (4*S*)-FPro. In addition, the chemical shift differences between $^{13}C\beta$ and $^{13}C\gamma$ are found within 5 ppm for all eight natural prolines, indicating a *trans* conformation of the Xaa-Pro peptide bond (Table 1) (Schubert et al., 2002). The dynamics of the non-fluorinated prolines are also not impacted by the insertion of either (4*S*)- or (4*R*)-FPro, as measured from the difference between the diastereotopic $H\delta$ chemical shifts (Ahuja et al., 2016) (Supplementary Fig.3). All of this indicates that the overall PPII secondary structure of a polyproline segment is maintained, regardless of the conformational bias of the individual fluoroproline residue. Nevertheless, subtle ^{13}C chemical shift differences between both peptides are observed in the prolines neighboring the FPro residues (3, 5, 7 and 9) (Table 1 and Supplementary Fig.2), with the most pronounced differences seen in the $C\delta$ chemical shifts (Fig. 3). Indeed, it has been shown for Ac-FPro-OMe model compounds that the (4*R*)- and especially the (4*S*)-FPro residues change the preferred ψ dihedral angle (DeRider et al., 2002). The FPro residues thus appear to cause small local conformational equilibrium or dynamics changes in the local PPII helix backbone, and further detailed conformational analysis is ongoing to confirm and quantify this effect. Finally, a minor set of peaks for prolines 10 and 3 is observed in the MpRS peptide in the 2D 1H - ^{13}C HSQC-NOESY spectrum (Fig. 2); their origin could not be established.

The ^{19}F NMR spectra of each peptide are shown in Fig. 4. The assignment of the ^{19}F resonances can be made by comparing their chemical shifts with those of the Ac-FPro-OMe model compounds (*ca.* -178 ppm for (4*R*)-FPro, -173 ppm for (4*S*)-FPro) (Hofman et al., 2019). Just as for 1H and ^{13}C chemical shifts, ^{19}F chemical shifts of each type of FPro change only slightly between peptides. Several smaller peaks are found near the main peak of the (4*R*)-FPro at position 4 of the MpRS peptide, with one accounting for 35% of the total signal integral. By

analyzing this particular sample by analytical HPLC and mass-spectrometry we identified this species as an impurity (with mass increase of 14 Da that is localized to Pro1 residue based on tandem MS² experiment). Other minor peaks can correspond to minor forms of the peptide where a single proline or fluoroproline is in the *cis*-form. For oligoproline sequences, the *cis* form of internal prolines are known to be typically populated at a few percent, while the N- and C-terminal prolines can have populations above 10% (Best et al., 2007; Urbanek et al., 2020). This illustrates the remarkable sensitivity of fluorine to its chemical environment, as it is able to resolve not just local conformation of the FPro residue itself, but also chemical modification or conformations of nearby proline residues within the oligoproline.

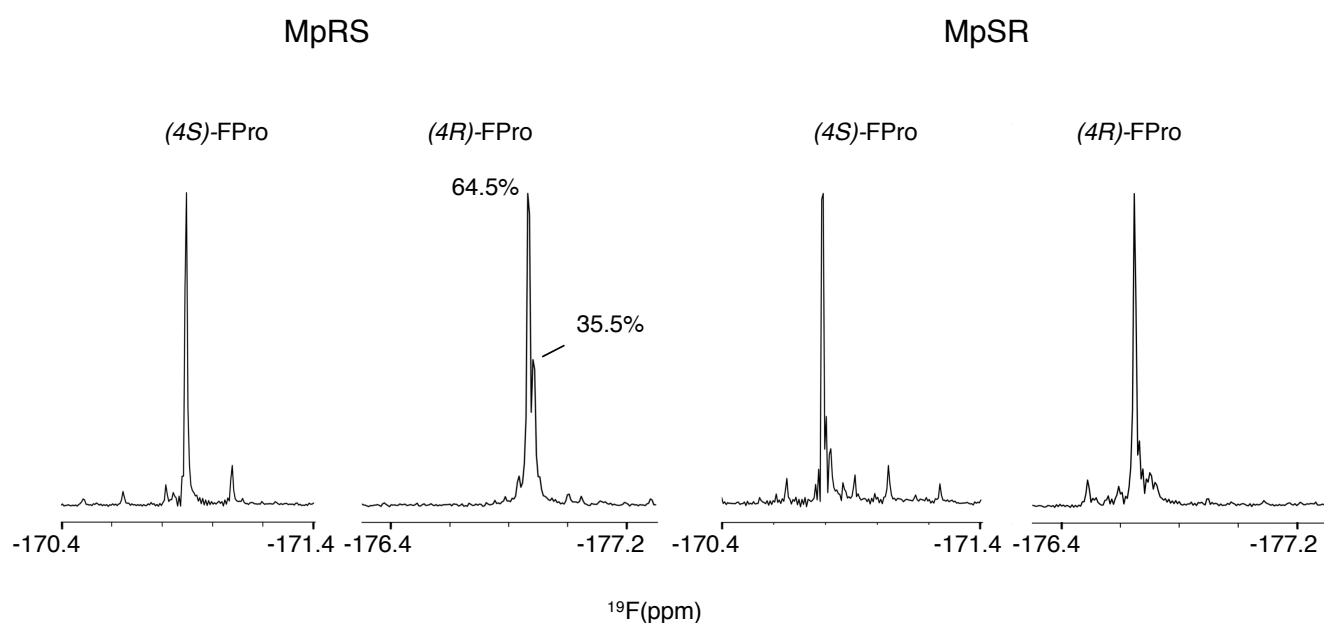


Figure 4: 1D ¹⁹F NMR spectra showing the fluoroprolines signals of the two model peptides MpRS on the left and MpSR on the right. The spectra were recorded at 298 K and 600 MHz in pure D₂O. The (4R)-FPro within MpRS resonance displays a second major species accounting for 35.5% of the total peak integral corresponding to a Hydrolytic and/or oxidative modification of the Pro1.

2.2. ^{19}F relaxation and dynamics

Spin relaxation rates are a useful source of information on molecular structure and dynamics. However, ^{19}F relaxation theory is rather complex, with multiple dipole-dipole (DD) interactions to neighboring protons, strong chemical shift anisotropy (CSA), and a multitude of cross-correlations (Dalvit et al., 2017; Lu et al., 2019). This stands in contrast with protein backbone ^{15}N relaxation where the dominant DD interaction with a single proton and ^{15}N CSA is well understood. A quantitative analysis of ^{19}F relaxation rates for both the (4*R*)– and (4*S*)–FPro residues in terms of dynamics thus requires knowledge of the various ^1H – ^{19}F distances within the fluoroproline structure, and also of the ^{19}F CSA tensor. These were obtained (Table 3) using density functional theory for the energy minimum structures of *C γ -exo* and *C γ -endo* ring conformations in the *trans* form of N-acetyl-FPro–NMe₂, where the capping groups were chosen to emulate the oligoproline peptide context. In each case, multiple protons had sufficiently small distances to the ^{19}F nucleus in order to significantly contribute to DD relaxation. While the distance to the H γ proton remains constant (2.0 Å), the distances to H β and H δ protons change with conformation (Table 3) (Gerig et al., 1973). Proline ring conformation also has a profound effect on the anisotropy parameter $\Delta\sigma$ of the chemical shift tensor: the major conformers (*C γ -exo* for 4*R*-FPro and *C γ -endo* for 4*S*-FPro) have $\Delta\sigma \approx -80$ ppm, while the minor conformers have $\Delta\sigma \approx -30$ ppm.

It has been shown that the proline ring pucker exchange occurs on a picosecond time scale (London et al., 1978; Sarkar et al., 1986; Kang et al., 2007), and thus faster than the overall tumbling of a peptide. Fluorine relaxation rates will be sensitive to this internal motion. However, because DD and CSA interactions vary in a correlated way with this motion, standard rotational diffusion autocorrelation functions cannot be used. Fortunately, as shown in the previous section, within the polyproline context both FPro residues adopt one dominant conformation. Although the work reported here can proceed, it would be beneficial in the long run to design case-specific relaxation models that involve a picosecond scale correlated switch in the spin Hamiltonian parameters.

Theoretical calculations of all relaxation rates were performed using the brute-force numerical implementation of Bloch-Redfield-Wangsness relaxation theory available in *Spinach* 2.6, which automatically accounts for all dipole-dipole interactions, all chemical shift anisotropies, and all of their cross-correlations (Hogben et al., 2011). Molecular geometries and the relevant magnetic parameters (chemical shielding tensors and *J*-couplings) were imported from DFT

calculations. The molecules in question are small enough that no spin system truncation is necessary. Longitudinal ^{19}F relaxation rates were calculated for rotational correlation times between 10 ps and 100 ns at 14.1 Tesla (Fig. 5A), for both *exo* and *endo* conformers to evaluate the impact of proline ring pucker. The correlation time dependence of longitudinal relaxation rates for the major conformers of all FPro residues shows a peculiar ‘camel hump’ shaped profile, with two maxima at 0.3 ns and at 4.4 ns. The same picture was reported earlier for fluorinated aromatic amino acids based on a simplified relaxation model (Dalvit et al., 2017). It is caused by CSA and dipolar ^1H - ^{19}F relaxation contributions being maximal at different frequencies, namely the fluorine Larmor frequency (ω_{F}) and the difference between proton and fluorine Larmor frequencies ($\omega_{\text{H}} - \omega_{\text{F}}$), respectively. For the minor conformers, the maximum at 0.3 ns is much lower than the one at 4.4 ns due to the lower CSA (Table 3). This demonstrates that, for longitudinal ^{19}F relaxation, the contribution of motions operating at time scales up to about 3 ns are strongly influenced by the ring pucker distribution. When assuming only the major ring pucker to be present – as in the polyproline context – R_1 shows very little contrast in the 0.1 ns to 10 ns range, implying it is not an appropriate parameter to unambiguously probe dynamics. The experimental relaxation rates measured for MpRS and MpSR peptides at 298K are reported in Table 4; they fall into a narrow range between 2.1 and 2.3 s^{-1} , in agreement with the calculated values within the aforementioned correlation time range.

Transverse ^{19}F relaxation rates (measured using the CPMG sequence with a half-echo delay of 200 μs) show the usual monotonic increase with the rotational correlation time (Fig. 5B). The difference in CSA between *exo* and *endo* puckers has a clear impact throughout, thereby complicating its interpretation in situations where puckers would be exchanging. To assess the contribution of slow motions, transverse relaxation rates were also measured using a spin echo (Table 4). This revealed about double values throughout, revealing exchange contributions on the ms time scale at both sites for both MpRS and MpSR peptides. As residual exchange contributions cannot be excluded in the CPMG experiment, an interpretation of transverse relaxation rates would also be unreliable. The origin of the exchange contribution is unclear, but possibly may arise from transient interactions between the polyproline segment and the flanking sequence (RVYK). Further studies will be required to investigate this unexpected finding.

In contrast to R_1 and R_2 , ^1H - ^{19}F cross-relaxation rates within the same carbon centre are purely dipolar and therefore likely to be easier to analyse. The $^1\text{H}_{\gamma}$ - ^{19}F NOE is ideal because H_{γ} has a distinct chemical shift at 5.6 ppm, allowing selective RF irradiation without perturbing the remaining protons of the proline ring; $^1\text{H}_{\gamma}$ - ^{19}F distance is independent on ring pucker. Fig. 5C

shows the calculated steady-state ^{19}F NOE upon $^1\text{H}\gamma$ saturation as a function of rotational correlation time. Just as for the R_1 curves, at long correlation times nearly identical curves for both the (4*R*)- and (4*S*)-FPro residues in each pucker are found, while at short correlation times a small difference is found between the puckers due to the dissimilar CSA. Importantly, the sigmoidal transition parts between fast-motion and slow-motion limits are similar in all four cases, making $^1\text{H}\gamma$ - ^{19}F NOE a reliable parameter sensitive to motions with correlation times between 0.1 and 4 ns.

Experimentally, $^{19}\text{F}\gamma$ signal intensities were measured for several $\text{H}\gamma$ selective irradiation times, leading to the observation of NOE build-up curves that were fitted with a single exponential function to extract the cross-relaxation rates (Fig. 5D-E). For both peptides, the steady-state NOE ranges from -6.8 % at position 4 to -19.9 % at position 8 (Table 4), indicating faster dynamics experienced by the first polyproline segment compared to the second. These values correspond to rotational correlation time estimates of 0.5 ns for the proline at position 4, and 0.8 ns for the proline at position 8. These correlation times suggest that local motions are different for the two polyproline segments, irrespective of the identity of the FPro residue. The reason for this difference between both polyproline segments is not obvious, and further relaxation or conformational studies will be required. We speculate that the distinct flanking sequences of each polyproline segment may determine their overall conformational and dynamical behavior, in a similar way as has recently been shown for other homopolymer sequences such as polyglutamine (Eftekharzadeh et al., 2016, Urbanek et al., 2020).

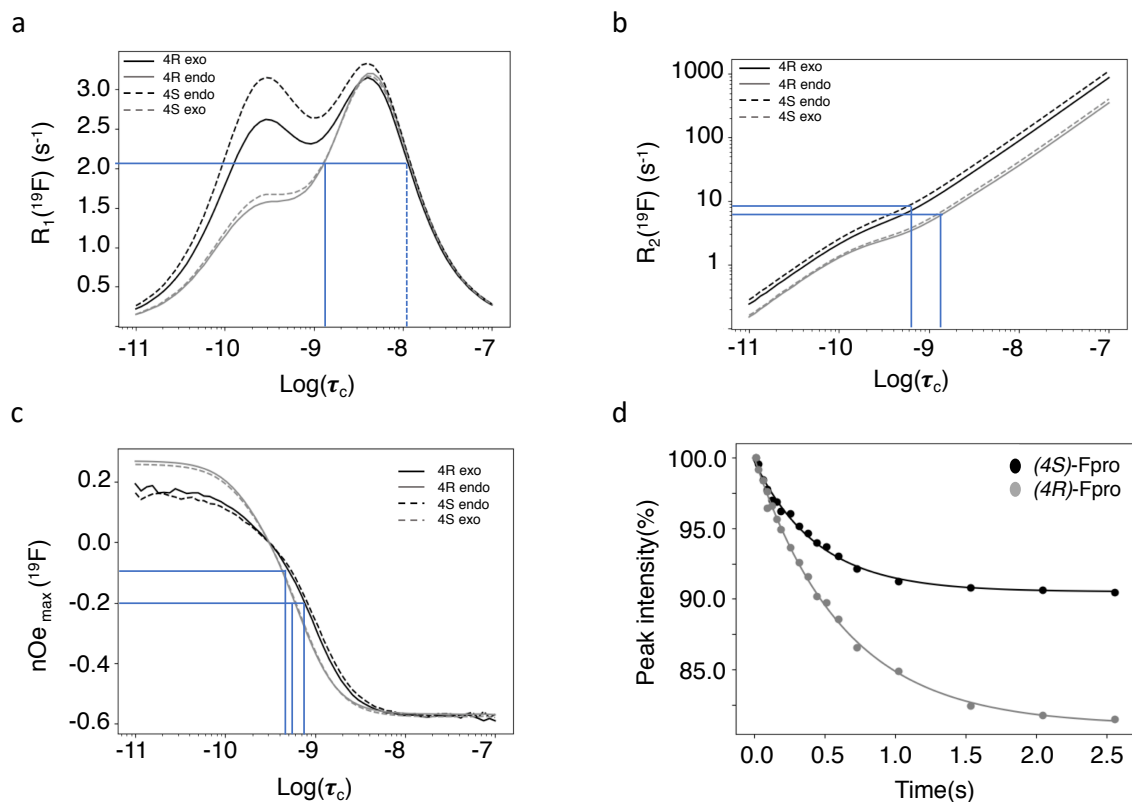


Figure 5: **a:** Calculated ^{19}F longitudinal relaxation rates as functions of rotational correlation time. The lines indicate the range of experimental values and their corresponding correlation times. The dotted line indicates a second, unrealistic correlation time for the observed R_1 . **b:** Calculated ^{19}F transverse relaxation rates as functions of rotational correlation time. The lines indicate the range of experimental values and their corresponding correlation times. **c:** Calculated steady state fluorine-proton heteronuclear NOEs. Relaxation data were calculated for (4R)-FPro in the major $\text{C}\gamma$ -exo (black plain line) and minor $\text{C}\gamma$ -endo (grey plain line) and for (4S)-FPro for the major $\text{C}\gamma$ -endo (black dashed line) and the minor $\text{C}\gamma$ -exo (grey dashed line) conformations. The lines indicate the range of experimental NOEs and their corresponding correlation times. **d:** Experimental NOE build-up at $\text{F}\gamma$ upon selective saturation of $\text{H}\gamma$ proton measured for the MpSR peptide at 298 K. The dots are the experimental peak intensities and the solid line is the corresponding fit to a monoexponential function.

2.3. Impact of proline modifications on the binding of SH3

SH3 domains are small modular protein domains of 50-70 amino acids that typically interact with proline rich motifs (PRM) and that are highly represented in the human genome (Saksela et al., 2012). Many experimental and theoretical studies have been conducted to decipher the molecular mechanisms underlying both binding affinity (in the 0.1-100 μ M range of dissociation equilibrium constant K_d) and specificity of SH3 domains that primarily recognize PXXP sequence motifs. This mechanism involves the aromatic indole ring of the tryptophan 37 residue exposed at the surface of the SH3 domain that mediates $CH\cdots\pi$ interaction with proline residues. Additional binding energy is provided by electrostatic interactions between the SH3 surface residues and those flanking the PXXP motif of the binding partner.

In order to measure the binding affinities between the Vinexin β SH3.3 domain and the model peptides, a titration experiment was performed where increasing amounts of peptide were added to a solution of 15 N labeled SH3 domain. Apart from MpRS and MpSR peptides, a titration was also performed with a non-fluorinated reference peptide. Just as for most SH3-PRM interaction studies, a gradual frequency shift of a subset of 1H - ^{15}N correlation peaks in the 1H - ^{15}N HSQC spectra was observed, indicative of fast exchange between bound and free states of the protein (Fig. 6A). Under this exchange regime, the chemical shifts provide an accurate measure of the bound protein fraction, enabling the determination of an equilibrium dissociation constant K_d (*vide infra*). Interestingly, a striking difference between the peptides is observed in the 1H - ^{15}N HSQC during the titration, where the trajectory of the tryptophan 37 $^{15}N_\epsilon$ - $^1H_\epsilon$ correlation appears different for the MpRS peptide compared to the MpSR peptide (insert Fig.6). Whether this reflects a direct interaction between the Trp37 aromatic ring with (4*S*)-FPro8 in the PXXP binding motif, or an alteration of the binding complex indirectly caused by the (4*S*)-FPro8 residue in MpRS will require further investigation.

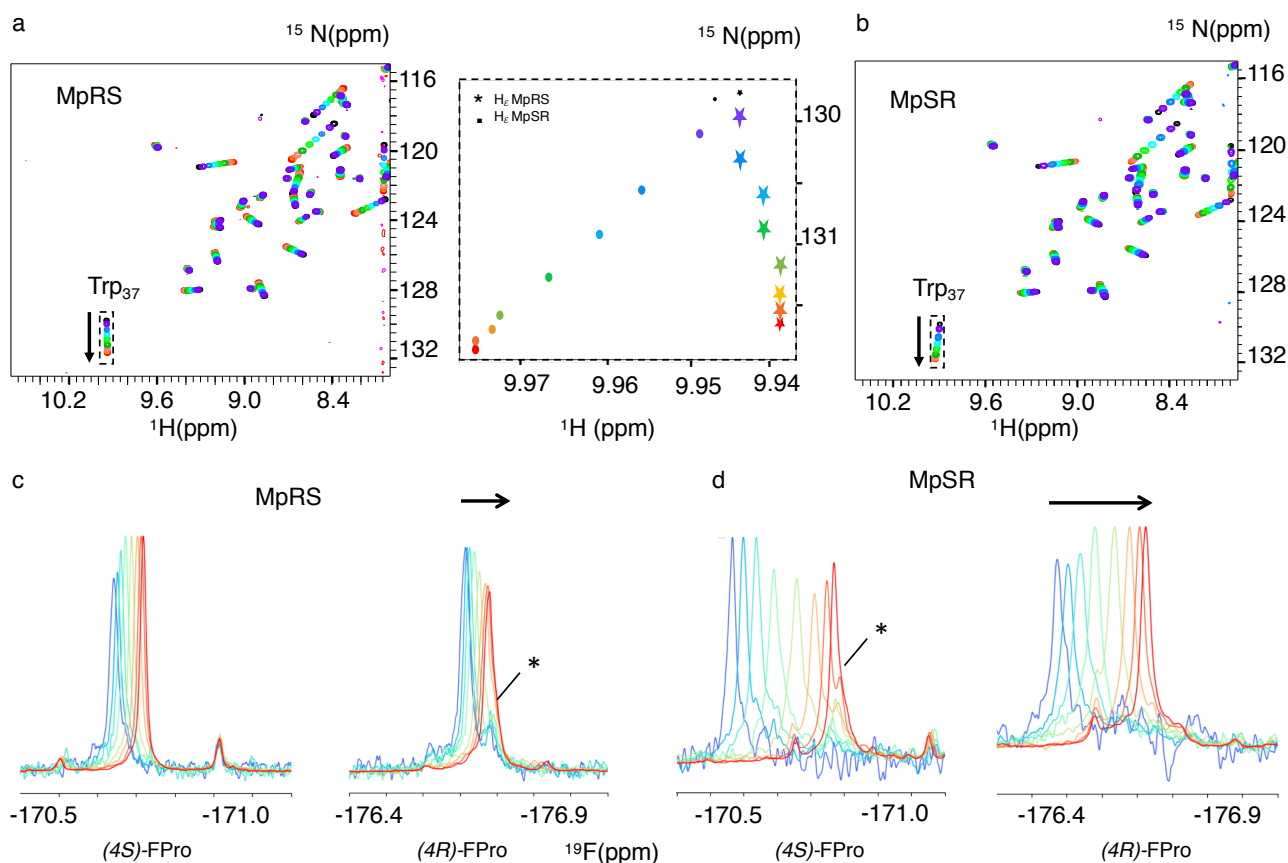


Figure 6: **a** and **b**: A series of ^1H - ^{15}N HSQC spectra of Vinexin β SH3.3 domain recorded upon successive addition of MpRS (A) and MpSR (B) peptides. The black spectrum corresponds to the first titration point where no peptide is added while the red spectrum corresponds to the last titration point. The arrows indicate the tryptophan 37 N_ϵ - H_ϵ cross-peak trajectories during the titration. The insert in the middle displays a zoom on these cross-peak trajectories shown as the position of the peak center for both MpRS (stars) and MpSR (disks) peptides. **c** and **d**: Series of 1D ^{19}F spectra recorded during the same titration experiment of Vinexin β SH3.3 domain by MpRS (C) and MpSR (D) peptides. Peak intensities were normalized to account for the difference in peptide concentrations and number of scans used to record the spectrum. For the (4R)-FPro4 in MpRS and (4S)-FPro4 in MpSR, a minor peak that overlaps with the main peak at high peptide concentrations is indicated by an asterisk. At low peptide concentrations (blue) the spectra are indicative of a mostly bound form while at high peptide concentrations (red), the spectra converge to those observed for the free peptides.

For the MpRS and MpSR peptides, the peptide-protein titration can also be observed using ^{19}F NMR, allowing to simultaneously monitor the binding event from the perspective of the protein

(receptor) and the peptide (the ligand) (Fig. 6C-D). Thanks to the availability of a cryogenic fluorine probehead, the ^{19}F signals could be detected even at the first titration point where the peptide concentration was just 50 μM and significant broadening was present. Just like for the ^1H - ^{15}N chemical shifts, increasing the peptide concentration resulted in a gradual shift of the ^{19}F resonances, indicative of a fast exchange regime. Interestingly, for both the signals from (4*R*)-FPro4 in MpRS and (4*S*)-FPro4 in MpSR, a minor peak is observed that does not shift during the titration (highlighted by a star in Fig. 6C and 6D). This minor peak thus appears to belong to a state that is not competent for SH3 binding. This proline is located in the first polyproline segment and this observation implies that at least two states of the complex are evidenced by the fluorine resonance at this position. At higher peptide concentrations, the peaks sharpen up with addition of peptide, which can be explained by the increasing fraction of the unbound peptide and thus lower amount of exchange broadening and faster tumbling correlation time. Visual inspection of the ^{19}F spectra reveals that the extent of chemical shift perturbation for both ^{19}F signals in each peptide appears similar, even though P8 falls within the binding motif and P4 outside (Supplementary Fig 1). These comparable CSP may be either due to a specific geometry of the two polyproline segments induced by the serine residue that may bend the PPII helix positioning P4 close to the SH3 surface and/or a dynamic averaging of CSP values due to one-dimensional diffusion of the SH3 domain on the peptide. When comparing the MpRS and MpSR peptides, it can be seen that the extent of chemical shift perturbations is the highest for the MpSR peptide, qualitatively already indicating the higher affinity of MpSR relative to MpRS.

Both the $^1\text{H}/^{15}\text{N}$ chemical shift perturbations of the SH3 domain as well as the ^{19}F chemical shift perturbations of the peptides can be used to assess the binding affinity. For this, the stoichiometry of the binding was first evaluated. Indeed, even though a single canonical PXXPX⁺ motif is present in the peptide sequence imposing binding specificity, a closer inspection shows that multiple non-specific PXXP motifs can be identified (Fig. 1), potentially leading to additional ways for the SH3 domain binding. For this, two binding models were used where the peptide and SH3 domain can bind either up to a 1:2 ratio or only in a 1:1 ratio. Both the ^{19}F and $^1\text{H}/^{15}\text{N}$ chemical shift data were fitted simultaneously using these models. Based on the goodness of fit reported as the reduced χ^2 , the ternary complex turned out to be unnecessary to explain the data, thus implying that only one SH3 binds to the peptide. The dissociation constants (K_d) found in this way were $96 \pm 30 \mu\text{M}$ for MpSR and $273 \pm 30 \mu\text{M}$ for MpRS. These values are slightly above the values found when only the ^1H - ^{15}N chemical shifts are considered (75 μM and 220 μM , supplementary table 1) but are within the reported uncertainties

that account for the uncertainty on protein and peptides concentration measurements that was estimated to be 15%. However, a strikingly good agreement was observed between the experimental and back-calculated ^{19}F chemical shifts, with a standard deviation of only 1.6 Hz despite the large peak widths of 10-20Hz (Fig. 7A). This excellent precision thanks to the sparsity of the 1D spectrum highlights one important feature of ^{19}F NMR spectroscopy to study molecular interactions. The fitted differences in bound and unbound ^{19}F frequencies is about twice as high for MpSR (265 ± 8 Hz for (4*S*)-FPro4 and 218 ± 8 Hz for (4*R*)-FPro8) than for MpRS (88 ± 17 Hz for (4*S*)-FPro4 and 100 ± 17 Hz for (4*R*)-FPro8).

In addition, the K_d value was also determined for the non-fluorinated peptide using the $^1\text{H}/^{15}\text{N}$ chemical shift perturbations alone using a binding model with 1:1 stoichiometry, which was found to be 74 ± 25 μM , which is similar to the MpSR peptide. It thus appears that the presence of (4*R*)-FPro within the binding motif has a negligible effect on the interaction with SH3, while (4*S*)-FPro significantly lowers the binding affinity despite our observations that suggest a preserved PPII conformation.

The exchange line broadening during the titration experiment also reports on the binding kinetic. Thus, the major ^{19}F peaks were fitted using a Lorentzian line shape and the line widths obtained in this way provide an estimate of the apparent transverse relaxation constant $R_2^\ddagger$ as a function of peptide concentration (Fig. 7B and C), that can be used to derive the binding kinetics. A simplified expression of the exchange contribution to $R_2^\ddagger$ as a function of peptide and SH3 concentration was used that is valid for the fast exchange approximation ($k_{exc} \gg \Delta\omega$) (Kovrigin et al., 2012):

$$R_2^\ddagger = p_f R_2^f + p_b R_2^b + p_f p_b \frac{\Delta\omega^2}{k_{exc}} \quad (1)$$

with:

$$k_{exc} = k_{on}([SH3]_{free} + K_d) \quad (2)$$

where p_b and p_f are the bound and free fractions of the peptide, R_2^b and R_2^f are the transverse relaxation rates of the bound and the free forms, and $\Delta\omega$ the frequency difference between the bound and free states multiplied by 2π . Taking the values of $\Delta\omega$, p_b , p_f , $[SH3]_{free}$ and K_d from the chemical shift perturbation fitting, the R_2^b , R_2^f and k_{on} values were subsequently fitted to the experimental $R_2^\ddagger$ values. For the MpSR peptide, the optimization was performed independently for the (4*S*)- and (4*R*)-FPro ^{19}F signals, leading to a fairly good agreement

between experimental and modeled values (Fig. 7B). This provided fitted association kinetic constants k_{on} of $0.9 \cdot 10^8 \pm 0.2 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ and $1.2 \cdot 10^8 \pm 0.2 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ for the (4*S*)-FPro4 and (4*R*)-FPro8 signals, respectively. These values are consistent with a simple one-to-one association mechanism driven by a free diffusion process of the two binding partners.

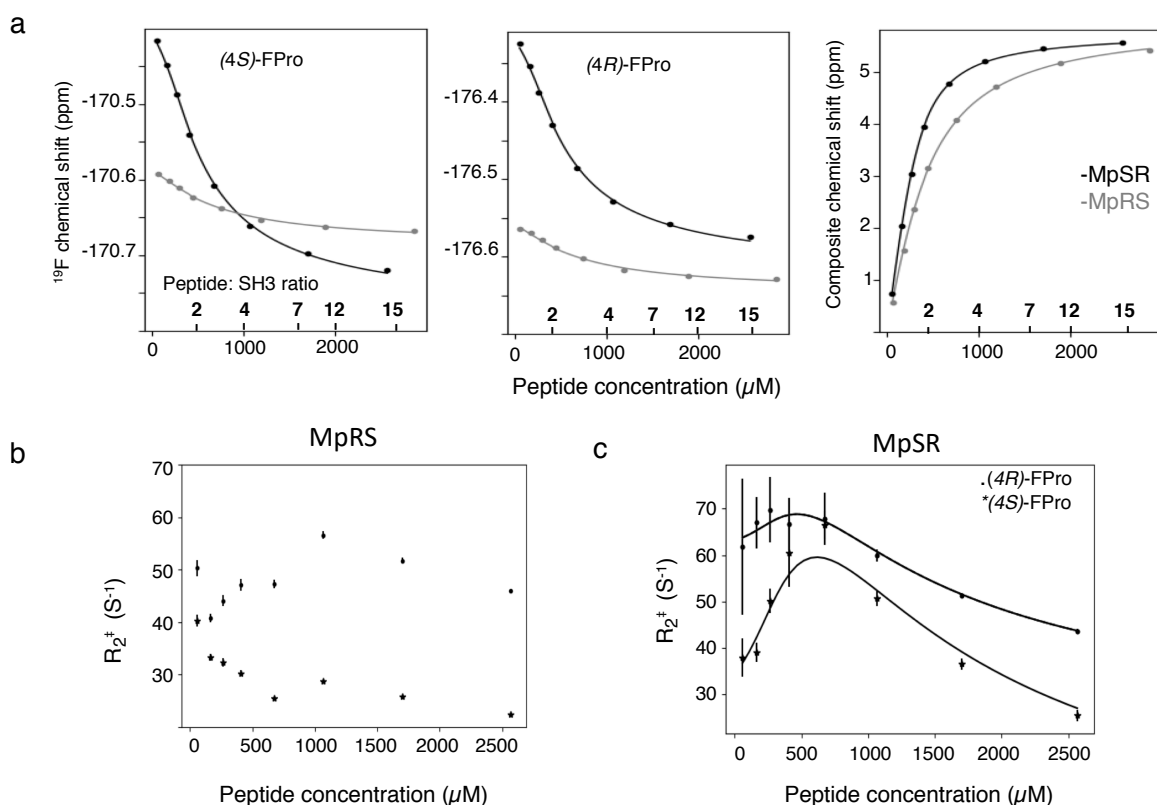


Figure 7: a: ^{19}F chemical shift variation of (4*S*)-FPro (left) and (4*R*)-FPro (middle) in MpSR (in black) and MpRS (in gray) peptides extracted from 1D ^{19}F spectra as a function of the total peptide concentrations. The peptide/protein ratios are indicated on the top of the axis. Left panel displays the ^1H , ^{15}N composite chemical shift from ^1H - ^{15}N HSQC. The experimental data (dots) were fitted simultaneously to derive the equilibrium dissociation constants for the two peptides (solid lines). **b:** Variation of the apparent ^{19}F transverse relaxation rates (R_2^+) as a function of MpRS peptide concentrations. **c:** Variation of the apparent ^{19}F transverse relaxation rates (R_2^+) derived from the ^{19}F line widths as a function of MpSR peptide concentrations. The black solid lines indicate the expected variation resulting from the fit of the experimental rates with

equation 1 and 2 using the transverse relaxation rates of the bound and free peptides (R_2^b , R_2^f) and the on-rate kinetic k_{on} as adjustable parameters.

For the MpRS peptide, the profile of $R_2^\ddagger$ as a function of peptide concentration showed a markedly different behavior. After an initial sharpening of about 10 Hz for both ^{19}F signals upon addition of the second peptide aliquot to the SH3 sample, a line broadening was observed for (4*R*)-FPro at position 4 while a continuous sharpening is experienced by the fluorine resonance of (4*S*)-FPro at position 8. This observation is peculiar, as in a simple one-binding site model one would expect a similar profile for both signals. This suggests a more complex binding mechanism involving at least one supplementary minor state. This is consistent with the observed significant reduction of the chemical shift differences between the bound and free forms of the MpRS peptide compared to MpSR as noted previously. For the MpRS peptide, the combined analysis of fluorine and proton spectral properties is insufficient to specify a specific binding model. However, together with the slight difference observed for the trajectory of the tryptophan $^{15}\text{N}_\epsilon$ - $^1\text{H}_\epsilon$ correlations in the ^1H - ^{15}N HSQC (Fig. 6A), this indicates that the structure or dynamics of the complex are altered by the insertion of (4*S*)-FPro within the canonical SH3 binding motif.

3. Discussion

(4*R*)- and (4*S*)-fluorinated prolines have, so far, been used in structural biology studies. This work demonstrates their hitherto neglected potential in biomolecular ^{19}F NMR investigations. In contrast to fluorination of most amino acids used in such studies, proline fluorination changes its conformational and dynamic properties, leading to modified protein-protein interactions. Although this may seem undesirable at first, this can be put to good use – as shown above – to modulate the interaction between a PRM and an SH3 domain. Using a model peptide containing two oligoproline sequences, permutations of two types FPro residues in conjunction with ^{19}F NMR analysis allowed studying the consequences of the conformational biases on the binding equilibrium with the SH3 domain. While the binding affinity appears unaltered by the introduction of (4*R*)-FPro at position 8 that lies within the SH3 binding motif, the insertion of (4*S*)-FPro at the same position leads to a substantial decrease of the binding affinity. Similar conclusions were drawn in studies involving SH3 domains of cortactin and human

hematopoietic-lineage cell-specific protein 1, where insertions of (4*R*)- or (4*S*)-FPro residues in the cognate PRM weakened the binding affinity (Ruzza et al., 2006; Borgogno et al., 2013). Interestingly, these and other (Horng et al., 2006) studies used circular dichroism spectroscopy to confirm that PPII conformational preference is stabilized by (4*R*)-FPro, meaning that the expected associated increase in SH3 binding affinity is negated by other effects introduced by the presence of the fluorine. Ruzza *et al* (Ruzza et al., 2006; Borgogno et al., 2013) suggested this could be due to a destabilization of the hydrogen bond formed by the proline's carbonyl group due to the inductive effect of fluorine, or by destabilization of proline's interaction with aromatic side-chains of the SH3. Further studies are required to disentangle these effects.

Apart from binding affinities, ^{19}F line shape analysis allowed kinetic information to be extracted. Thanks to the exquisite susceptibility of the ^{19}F signal line width to chemical exchange phenomena, it was found that the binding on-rate of the SH3 domain is fast and diffusion-limited. This result is consistent with a recent study reporting diffusion limited binding kinetics of an SH3 domain to a PRM peptide, where a fluorinated tryptophan inserted in the SH3 domain allowed simultaneous monitoring of ^{19}F and $^1\text{H}/^{15}\text{N}$ chemical shift perturbations measured from the SH3 domain (Stadmler et al., 2020). The difference with our study is that the ^{19}F chemical shift perturbations report on the binding event from the point of view of the binding peptide, providing complementary information with the $^1\text{H}/^{15}\text{N}$ chemical shift perturbation from the SH3 domain. Here, observation of fluorine resonances perturbations on the ligand evidenced a different dynamic of the SH3 domain on the polyproline peptide upon introduction of the conformationally biased (4*S*)-FPro in the cognate PRM.

Although the available numerical tools allow in principle to model spin relaxation processes in multi-spin systems very accurately, a major complication is the picosecond scale dynamics of the five-membered ring, mainly due to the strong dependence of ^{19}F CSA on ring pucker. This effect can be mitigated by a strong ring pucker bias, as is typically the case for (4*R*)- and (4*S*)-FPro residues. However, in general, and especially for other FPro variants without pucker bias (Hofman et al., 2018), a more advanced theoretical analysis will be required. Still, the measurement of the NOE between the geminal H_γ and F_γ provided an interesting way to probe local dynamics with correlation times between 0.3 and 4-5 ns.

The comparison of transverse relaxation at two different effective B_1 fields revealed the presence of motions occurring at the μs to ms time-scale. It should be noted that the presence of many ^1H - ^{19}F couplings within the FPro spin system implies that recently developed ^{19}F relaxation dispersion experiments cannot be applied (Overbeck et al., 2020). The dispersion of proton frequencies in the fluorinated prolines enable their selective excitation, a feature that

was exploited for the selective $H\gamma$ - $F\gamma$ NOE and can be further used to develop sequences adapted to FPro spin systems. The molecular origin of the difference in dynamics between the oligoproline segments remains unclear, and suggests that the flanking amino acid sequences can play a role in the conformational and dynamical preferences of polyproline segments. Importantly, given the absence of amide protons and the low $^1H\alpha$ and $^{13}C\alpha$ chemical shift dispersion, this information would be very difficult to obtain from 1H , ^{13}C or ^{15}N measurements. In conclusion, fluorinated prolines provide an attractive tool for biomolecular NMR studies, in addition to their well-established application of controlling proline conformation. Given the increasing capabilities of chemical biology techniques that allow introduction of unnatural amino acids in proteins, such as chemical ligation or genetic code expansion (Debelouchina et al., 2017), we foresee that ^{19}F NMR studies through FPro residues will find their way to larger protein constructs. Apart from (4*R*)- and (4*S*)-fluorinated prolines, many more mono- and difluorinated prolines have been described (Verhoork et al., 2018), providing a rich set of fluorine labelling options for PRMs that can be tuned to the specific needs in terms of conformational control and/or ^{19}F NMR properties. Further investigations in this respect are ongoing. Furthermore, this work demonstrates how the conformational changes caused by fluorination within a proline-rich SH3 binding motif subtly modulates the binding properties of an SH3 domain for its cognate binding site, despite it occurring at a position that is not directly involved in the binding – as defined by the canonical sequence binding motif – and the global PP-II conformation being preserved. We note that the change in binding affinity for the vinexin β SH3 upon (4*S*)-FPro insertion is comparable to what was observed upon serine phosphorylation in the proline-rich region of the RAR γ , upon which the MpRS and MpSR peptides were modelled upon (Lalevée et al. 2010). The introduction of fluoroproline in larger protein constructs thus provides an attractive tool to explore how small local conformational biases result into large biological effects within interaction networks, which is the basis for signalling mechanisms. Indeed, binding sites for SH3 – or other – domains are frequently clustered within larger proline rich regions where post-translational modifications lead to subtle changes in the weak binding affinities and thus a redistribution of the protein-protein interaction network. We are confident that fluorinated prolines and ^{19}F NMR provide an elegant way to shed light on these complex systems.

4. Material and methods

Sample preparation

The MpRS and MpSR peptides were produced by solid phase synthesis using Fmoc-amino-acids using a model 431A peptide synthesizer from Applied Biosystems (Foster City, CA, USA). Fmoc-protected (4*R*)- and (4*S*)-FPro amino acids were purchased from Bachem SA. Peptides were purified by reversed-phase HPLC, and checked by electrospray ionization-time of flight mass spectrometry (ESI-TOF). The Vinexin β SH3.3 was obtained using recombinant expression of a GST fusion protein in *E. coli* using pGEX plasmids as described (Lalevee et al., 2010). After thrombin cleavage, the protein was purified using size exclusion chromatography and eluted with phosphate buffer (40 mM Phosphate, NaCl 100 mM, DTT 2 mM, pH 7). Before titration experiments, a dialysis was performed using 1 kDa and 3 kDa cutoff membrane for the peptide and the protein, respectively, and a common dialysis bath containing the buffer used in interactions experiments. Protein concentrations were determined by measuring the OD at 280 nm (molar absorption coefficient 11460 M⁻¹.cm⁻¹). Peptide concentrations were measured by ¹H NMR by comparing the integrals of peptide resonances with those of tryptophan of known concentration in a sample containing small amounts (10-30 μ M) of both compounds in D₂O as described (Kohler et al., 2015). For assignments, lyophilized powder of MpRS and MpSR peptides were dissolved in 170 μ L of D₂O for a final concentration of 1 mM in 3 mm tubes.

NMR experiments

¹⁹F NMR spectra were recorded on a Bruker Avance I spectrometer operating at a ¹H frequency of 600 MHz and equipped with a cryogenic QCI-F probe. ¹H and ¹³C spectra were recorded using a Bruker Avance III spectrometer operating at a ¹H frequency of 700 MHz and equipped with a cryogenic TCI probe. Standard full-range ¹H-¹³C HSQC (10 ppm ¹H x 80 ppm ¹³C) were recorded on MpRS and MpSR peptides for the carbon assignment. The number of points in the time domain was 4096 in *F*₂ and 4096 in *F*₁. In addition, a high resolution 2D ¹H-¹³C HSQC-NOESY (600 ms mixing time) was recorded with the same ¹H spectral width, but with a narrow carbon bandwidth of 3 ppm, centered on the proline's C δ resonances (47.3 ppm). The number of points in the time domain was 1024 in *F*₂ and 256 in *F*₁. The resulting resolution in the ¹³C dimension was 4 Hz/pt. The usual ¹³C 180° pulse used to compensate for chemical shift evolution during the *echo-antiecho* encoding pulsed field gradient was replaced by a frequency-

selective ^{13}C 180° refocusing pulse (4 ms RSNOB) was applied on the $^{13}\text{C}\delta$ region in order to avoid interference from folded peaks from outside the spectral region. The NOESY mixing time was 80 ms. The relaxation time was set to 1 s and 300 transients were recorded for each t_1 point resulting into a total experiment time of 1 day and 4 hours.

The ^1H - ^{19}F heteronuclear coupling constants were measured from the ^1H - ^1H TOCSY spectra (MLEV spinlock 80 ms) recorded at 700 MHz. The spectral width was 10 ppm in both F_1 and F_2 , with the number of time domain points 4096 in F_2 and 512 in F_1 , resulting in resolutions in F_1 and F_2 of 23.4 and 2.9 Hz/pt, respectively. The ^1H - ^1H couplings were measured using SERF experiments (Fäcke et al., 1995) modified to use the Pell-Keeler method (Pell et al., 2007) to obtain 2D absorption mode line shapes, as recently proposed (Sinnaeve, 2021). The active spin refocussing selective 180° pulse was a RE-BURP pulse of 14.5 ms set to invert just the FPro $\text{H}\alpha$ signals, while the selective inversion 180° pulses were I-BURP pulses of 12.85 ms set to invert just one $\text{H}\beta$ proton per FPro residue at a time. The spectral width was set to 1.07 ppm in F_2 and 23.5 Hz in F_1 , with the number of time domain points 1024 in F_2 and 64 in F_1 , resulting in resolutions in F_1 and F_2 of 0.7 and 1.5 Hz/pt, respectively.

^{19}F R_1 and R_2 relaxation parameters were measured at 600 MHz (^1H frequency) and 298 K using standard inversion recovery and CPMG experiments respectively. The carrier frequency was set to -174 ppm with a spectral width of 12 ppm and an interscan relaxation delay of 4 s. The inversion recovery relaxation build-up delays ranged from 1 ms to 3 s with an exponential sampling with one point repeated for uncertainty estimation, resulting in 20 data points in total. The CPMG sequence was measured using a half-echo delay of 200 μs , or as a single spin echo with variable delays. A 180° proton pulse was applied every 2.8 ms at the fluorine echo time to average cross-correlation effects and ensure a single exponential decay (Farrow et al., 1995). Sampled relaxation delays ranged from 1 ms to 460 ms, with 16 data points in total.

^1H - ^{19}F nOe buildup experiments were measured by selectively saturating the $\text{H}\gamma$ proton, using a train of sinc shaped soft 180° pulses centered at 5.42 ppm. The pulse duration was 2.8 ms and was applied every 4 ms prior to fluorine acquisition. The saturation times ranged from 10 ms to 2.6 s, including one repeat for error estimation, resulting in 16 data points in total.

Processing of 1D ^{19}F spectra and quantification were performed using an open-source Python package dedicated to Fourier spectroscopies called “Spectrometry Processing Innovative Kernel” (SPIKE) (Chiron et al., 2016). An exponential line broadening of 8 Hz was applied

prior to Fourier Transform for signal apodization. Line fitting was done using the least-square minimizer of the Scipy optimize toolbox to find the optimal set of the signal parameters minimizing the squared differences between the experimental and calculated spectra. 2D spectra used for peptide assignments were processed using Topspin 2.6 (Bruker) and visualized in CcpNmr Analysis V2 (Vranken et al., 2005). Relaxation parameters were obtained by fitting relaxation data to a three parameters single exponential model using the least-square algorithm implemented in the Scipy optimize toolbox (Levenberg-Marquart).

The selective longitudinal relaxation rate constants ρ and the proton-fluorine cross-relaxation rate σ were obtained by identification of the three optimized parameters to the following equation:

$$I(t) = I_0 + \frac{\sigma \gamma_H}{\rho \gamma_F} I_0 (1 - e^{-\rho t}) \quad (3)$$

where I_0 is the equilibrium signal intensity and γ_H , γ_F are the proton and fluorine magnetogyric ratios, respectively.

Electronic structure theory and spin relaxation theory

All electronic structure theory calculations were performed using *Gaussian09* (Frisch et al., 2009). Molecular geometries of proline isomers and conformers were optimised for fluoroproline moieties (capped with an acetyl group on the NH side and a dimethylamino group on the COOH side) using density functional theory with the M06 exchange-correlation functional (Zhao et al., 2008) and cc-pVDZ basis set (Peterson et al., 2002) in SMD chloroform (Marenich et al., 2009). Hessians were checked for positive definiteness at convergence point, and magnetic property calculations (shielding tensors and J -couplings) then proceeded using gauge-independent atomic orbital method (London et al., 1937) with the basis set decontracted and augmented with tight Gaussian functions (Deng et al., 2006) for the calculation of isotropic J -couplings.

Spin relaxation theory calculations were performed using *Spinach 2.6* (Hogben et al., 2011). Cartesian coordinates, chemical shielding tensors, and J -couplings of all fluorine and hydrogen atoms were imported from *Gaussian09* logs, and a numerical evaluation (Goodwin et al., 2015) of Redfield's relaxation superoperator (Redfield et al., 1957) for the resulting 16-spin system

was carried out using the restricted state space approximation (Kuprov et al., 2007; Edwards et al., 2014), (IK-1(4,4) basis set) with a 5 Angstrom distance cut-off for dipolar interactions. Rigid-molecule isotropic rotational diffusion approximation was used. Longitudinal and transverse relaxation rates for the spins of interest were extracted as the matrix elements of the relaxation superoperator corresponding to L_z and L_+ states of those spins. The implementation of Bloch-Redfield-Wangsness theory in *Spinach* automatically accounts for all applicable cross-relaxation and cross-correlation effects (Kuprov et al., 2011).

Titration experiments

Titration experiments were performed by successive addition of stock peptide solutions to a sample of SH3 protein at 314 μM in a 3 mm NMR tube. In order to reduce the dilution of the initial protein solution and keep the aliquot volumes within values compatible with low pipetting errors (1 to 3 μL), initial aliquots were added using stock solutions diluted by a factor 2. The concentrations of the stock solutions were 5.1 mM and 5.7 mM for MpSR and MpRS, respectively. For every titration point, a 1D ^{19}F spectrum and a ^1H - ^{15}N -HSQC were recorded at 298K on the same spectrometer, taking advantage of the QCI-F probe. The ^1H - ^{15}N -HSQC was recorded using the standard pulse sequence and a 3-9-19 WATERGATE water suppression element. 200 points were recorded in the indirect dimension to achieve a final resolution of 7.0 and 18.8 Hz/pt in the acquisition and indirect dimensions, respectively. The relaxation delay was set to 1 s resulting in a total acquisition time of 16 min. The 1D ^{19}F spectra were recorded with a spectral width of 22522 kHz. The number of scans was adapted to achieve a sufficient signal/noise ratio for each concentration of peptide. For the first titration point, at a peptide concentration of 50 μM , 1600 scans were recorded for a total acquisition time of 1 hour and 6 min while 250 scans (10 min acquisition time) were used for the large peptide concentrations. The protein chemical shift perturbation was averaged over 9 ^1H - ^{15}N correlations that displayed a similar apparent titration profile (from the amino-acids Q14, N15 (side chain $\text{N}_{\delta 2}$), D17, L21, W37 (side chain N_{ϵ}), V39, G49, T50, V56).

The composite chemical shift was calculated using:

$$\Delta\delta = \sqrt{(\Delta\delta_N)^2 + (5\Delta\delta_H)^2} \quad (4)$$

where $\Delta\delta_N$ and $\Delta\delta_H$ are the ^{15}N and proton chemical shift difference measured between the free protein and the protein in presence of a given amount of peptide.

The modeling of the interaction was performed using an in-house Python script that solves the equilibrium concentrations of a set of interacting molecules by integrating the set of coupled differential equations until steady state is reached (<https://github.com/delsuc/SpinEq>). To fit the experimental data, we used the fluorine frequencies of free and bound states, and the frequency of the bounded SH3 as parameters. Depending on the model, one or more equilibrium constants were given. The goodness of fit was assessed using the reduced chi2:

$$\chi^2 = \frac{1}{N-NP} \sum_{i=1}^N (\delta_i^{exp} - \delta_i^{calc})^2 \quad (5)$$

Data availability

All NMR spectra used for this work have been deposited on Zenodo (DOI: doi.org/10.5281/zenodo.5548094)

Author contribution

DS and BK designed the experiments and ABB carried them out. GJH, BL and VT synthesized the peptides, EE produced the SH3 domain, EO performed quantum calculations under supervision of IK. JCM analyzed the data. DS and BK prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Agence Nationale de la Recherche (FLUOVIAL ANR-18-CE44-0009-01, FLUOPROLINE ANR-20-CE11-0025) the French Infrastructure for Integrated Structural Biology (FRISBI, ANR-10-INSB-05-01), Instruct-ERIC, the Centre National de la Recherche Scientifique and the University of Strasbourg. The Research Foundation – Flanders (FWO) is indebted for a research project to J.C.M. and D.S. (3G011015), a PhD fellowship to

E.O. and staff exchange funding (FWO-WOG Multimar). The EPSRC is thanked for a partial PhD grant to G.-J.H. (EPSRC-DTG EP/M50662X/1). A.B.B is supported by a fellowship from the région Grand Est and ANR. Claude Ling is warmly acknowledged for maintenance of the NMR spectrometers and Pascal Eberling for peptide synthesis.

Tables

Table 1: Chemical shift assignments of MpRS and MpSR peptides. $\Delta\delta$ is the chemical shift difference between $^{13}\text{C}\beta$ and $^{13}\text{C}\gamma$ resonances used as indicator for *cis* and *trans* conformations of the Xaa-Pro peptide bond. $\Delta\text{C}\alpha$ is the chemical shift difference between the measured $^{13}\text{C}\alpha$ and the corresponding random coil values. Chemical shifts were measured in D_2O at pH 7, 298 K and referenced to DSS- d_6 .

MpSR										
	H α	H β	H γ	H δ	C α	C β	C γ	C δ	$\Delta\delta$	$\Delta\text{C}\alpha$
Pro1	4.61	2.54 2.39	2.05	3.42 3.37	61.75	30.91	26.52	49.3	4.39	-1.59
Pro2	4.76	2.39 1.88	2.01	3.70 3.54	61.65	30.64	27.27	50.36	3.37	-1.69
Pro3	4.63	2.38 1.96	2.06	3.82 3.59	61.31	30.63	27.30	50.28	3.33	-2.03
(4S)-FPro	4.87	2.59 2.43	5.42	4.07 3.99	60.32	36.97	95.09	56.55	-58.12	-3.02
Pro5	4.43	2.27 1.88	2.02	3.85 3.62	63.06	31.89	27.27	50.44	4.62	-0.28
Ser6	4.71	3.84 3.72			56.24	63.04				-2.47
Pro7	4.68	2.31 1.89	2.02	3.79 3.64	61.51	30.66	27.26	50.50	3.40	-1.83
(4R)-FPro	4.88	2.72 2.06	5.45	4.22 3.84	59.77	37.20	95.10	56.58	-57.90	-3.57
Pro9	4.71	2.32 1.89	2.03	3.75 3.61	61.37	30.68	27.29	50.21	3.39	-1.97
Pro10	4.68	2.28 1.87	1.99	3.81 3.63	61.26	30.62	27.27	50.32	3.35	-2.08
Pro11	4.37	2.26 1.84	2.00	3.79 3.61	62.79	32.01	27.27	50.36	4.74	-0.55
Arg12	4.23	1.70	1.56 1.48	3.13	55.82	30.82	26.94	43.17		-0.96
Val13	4.06	1.93	0.86 0.83		61.82	33.04	20.41 20.98			-0.72
Tyr14	4.57	3.01 2.88		7.10 H ϵ 6.78	57.52	39.09		53.21 C ϵ 37.98		-0.66
Lys15	4.27	1.84	1.34	1.63 H ϵ 2.93	55.23	32.74	24.61	28.87 C ϵ 41.82		-1.73

MpRS										
	H α	H β	H γ	H δ	C α	C β	C γ	C δ	$\Delta\delta$	$\Delta C\alpha$
Pro1	4.61	2.54	2.04	3.42	61.76	30.86	26.51	49.28	4.35	-1.58
		2.04		3.37						
Pro2	4.76	2.39	2.01	3.69	61.62	30.57	27.26	50.37	3.31	-1.72
		1.88		3.54						
Pro3	4.71	2.29	2.05	3.80	61.24	30.49	27.32	50.20	3.17	-2.10
		1.84		3.60						
(4R)-FPro	4.89	2.73	5.44	4.20	59.72	37.22	95.66	56.52	-58.44	-3.62
		2.03		3.81						
Pro5	4.45	2.28	2.01	3.85	62.86	32.06	27.26	50.51	4.80	-0.48
		1.88		3.64						
Ser6	4.70	3.85			56.46	62.97				-2.25
		3.72								
Pro7	4.62	2.37	2.05	3.82	61.53	30.78	27.32	50.69	3.46	-1.81
		1.96		3.68						
(4S)-FPro	4.86	2.62	5.41	4.07	60.35	36.90	95.63	56.50	-58.73	-2.99
		2.38		3.96						
Pro9	4.69	2.3	2.02	3.74	61.22	30.52	27.25	50.13	3.27	-2.12
		1.87		3.56						
Pro10	4.68	2.31	2.02	3.82	61.19	30.73	27.25	50.25	3.48	-2.15
		1.87		3.59						
Pro11	4.37	2.26	2.00	3.8	62.82	32.01	27.26	50.33	4.75	-0.52
		1.83	1.98	3.59						
Arg12	4.24	1.68	1.54	3.13	55.80	30.80	26.96	43.17		-0.98
		1.67	1.47	3.13						
Val13	4.06	1.94	0.85		61.83	33.03	20.97			-0.71
			0.83				20.44			
Tyr14	4.57	3.01		7.10	57.51	39.09		53.21		-0.67
		2.87		H ϵ 6.78				C ϵ 37.98		
Lys15	4.28	1.84	1.34	1.65	55.16	32.72	24.63	28.84		-1.80
		1.72	1.33	1.62				C ϵ 41.82		
				H ϵ 2.92						

Table 2: Comparison of the scalar coupling constants (in Hz) of (4*R*)-FPro and (4*S*)-FPro measured in MpRS and MpSR peptides with those reported for the free fluoroproline residues (Gerig et al., 1973).

	MpRS		MpSR		Free amino acid	
	P4 (4 <i>R</i>)	P8 (4 <i>S</i>)	P4 (4 <i>S</i>)	P8 (4 <i>R</i>)	(4 <i>R</i>)-FPro	(4 <i>S</i>)-FPro
$^3J_{F-\beta_2}$	42.3	21.0	20.9	42.1	40.5	20.5
$^3J_{F-\beta_3}$	18.8	42.5	43.1	18.9	19.6	41.9
$^3J_{F-\delta_2}$	38.2	24.6	24.7	38.3	37.4	19.4
$^3J_{F-\delta_3}$	21.7	35.2	35.4	21.9	20.1	37.6
$^3J_{\alpha-\beta_2}$	10.1	3.0	3.1	10.2	10.4	2.8
$^3J_{\alpha-\beta_3}$	8.1	10.5	10.2	8.1	8.1	10.5

Table 3: Internuclear distances between ^{19}F atom and the neighboring protons for representative conformers of major and minor conformations of the (4*R*)- and (4*S*)- FPro and corresponding ^{19}F CSA tensor parameters derived from Gaussian calculations. $\Delta\delta$ is the chemical shift tensor anisotropy, η is the asymmetry parameter, and δ_{xy}^{anti} , δ_{xz}^{anti} , δ_{yz}^{anti} are the antisymmetric components to the full CSA tensor in the principal axes coordinate system of the symmetric part of the tensor.

	Distances (Å)					^{19}F CSA tensor				
	F-H γ	F-H β_2	F-H β_3	F-H δ_2	F-H δ_3	$\Delta\delta$ (ppm)	η	δ_{xy}^{anti} (ppm)	δ_{xz}^{anti} (ppm)	δ_{yz}^{anti} (ppm)
(4<i>R</i>)-exo major	2.03	3.29	2.56	3.3	2.5	-74.2	0.120	4.71	2.21	-3.45
(4<i>R</i>)-endo minor	2.02	2.89	2.50	2.97	2.44	-25.6	0.396	7.29	2.34	4.42
(4<i>S</i>)-endo major	2.01	2.49	3.29	2.4	3.25	-84.9	0.392	-3.26	-4.27	-6.01
(4<i>S</i>)-exo minor	2.02	2.49	2.89	2.52	2.88	-33.3	0.483	5.78	2.20	-2.32

Table 4: Experimental longitudinal and transverse relaxation rates together with the nuclear Overhauser effect measured for both peptides at 298 K on a 600 MHz spectrometer.

		MpRS		MpSR	
		P4(4 <i>R</i>)	P8(4 <i>S</i>)	P4(4 <i>S</i>)	P8(4 <i>R</i>)
$R_1(s^{-1})$		2.23± 0.04	2.2± 0.01	2.30± 0.01	2.13± 0.01
NOE max (%)		-6.8	-19.9	-9.3	-19.0
$\rho (s^{-1})$		1.85	1.76	2.25	1.59
$\sigma (s^{-1})$		-0.12	-0.33	-0.2	-0.28
$R_2(s^{-1})$	Spin	20.3 ± 0.5	24.96 ± 0.6	12.4 ± 0.5	18.5 ± 0.4
	Echo				
$R_2(s^{-1})$	CPMG	8.6 ± 0.5	8.2 ± 0.3	5.4 ± 0.3	9.7 ± 0.5

References

- Ahuja P., Cantrelle F.X., Huvent I., Hanouille X., Lopez J., Smet C., Wieruszeski J.M., Landrieu I., Lippens G.: Proline Conformation in a Functional Tau Fragment, *J. Mol. Biol.* 428 (2016) 79–91. <https://doi.org/10.1016/j.jmb.2015.11.023>
- Aufiero, M., Gilmour, R.: Informing Molecular Design by Stereoelectronic Theory: The Fluorine Gauche Effect in Catalysis, *Acc. Chem. Res.* 51 (2018) 1701–1710. <https://doi.org/10.1021/acs.accounts.8b00192>.
- Behrendt, R. P., White, P., Offer, J.: Advances in Fmoc solid-phase peptide synthesis, *J. Pept. Sci.* 22 (2016) 4–27. <https://doi.org/10.1002/psc.2836>.
- Berger, A. A., Völler, J.-S., Budisa, N., Kokschi, B.: Deciphering the Fluorine Code - The Many Hats Fluorine Wears in a Protein Environment, *Acc. Chem. Res.* 50 (2017) 2093–2103. <https://doi.org/10.1021/acs.accounts.7b00226>.
- Best, R. B., Merchant, K. A., Gopich, I. V., Schuler, B., Bax, A., Eaton W. A.: Effect of flexibility and cis residues in single-molecule FRET studies of polyproline, *Proc. Natl. Acad. Sci.* 104 (2007) 18964–18969. <https://doi.org/https://doi.org/10.1073/pnas.0709567104>.
- Boeszoermyeni, A., Chhabra, S., Dubey, A., Radeva, D. L., Burdzhiev, N. T., Chanev, C. D., Petrov, O. I., Gelev, V. M., Zhang, M., Anklin, C., Kovacs, H., Wagner, G., Kuprov, I., Takeuchi, K., Arthanari, H.: Aromatic ¹⁹F- ¹³C TROSY: a background-free approach to probe biomolecular structure, function, and dynamics, *Nat. Methods.* 16 (2019) 333–340. <https://doi.org/10.1038/s41592-019-0334-x>.
- Boeszoermyeni, A., Ogórek, B., Jain, A., Arthanari, H., Wagner, G.: The precious fluorine on the ring: fluorine NMR for biological systems, *J. Biomol. NMR.* 74 (2020) 365–379. <https://doi.org/10.1007/s10858-020-00331-z>.
- Bondi, A.: Van der waals volumes and radii, *J. Phys. Chem.* 68 (1964) 441–451. <https://doi.org/10.1021/j100785a001>.
- Borgogno, A., Ruzza, P.: The impact of either 4-R-hydroxyproline or 4-R-fluoroproline on the conformation and SH3m-cort binding of HPK1 proline-rich peptide, *Amino Acids.* 44 (2013) 607–614. <https://doi.org/10.1007/s00726-012-1383-y>.

Chaiken, I. M., Freedman, M. H., Lyster, J. R., Cohen, J. S.: Preparation and studies of ^{19}F -labeled and enriched ^{13}C -labeled semisynthetic ribonuclease-S' analogues., *J. Biol. Chem.* 248 (1973) 884–891. [https://doi.org/10.1016/s0021-9258\(19\)44350-0](https://doi.org/10.1016/s0021-9258(19)44350-0).

Chiron, L., Coutouly, M.-A., Starck, J.-P., Rolando, C., Delsuc, M.-A: SPIKE a Processing Software dedicated to Fourier Spectroscopies, (2016) 1–13. <http://arxiv.org/abs/1608.06777>.

Cordero, B., Gomez, V., Platero-Prats, A. E., Reves, M., Echeverria, J., Cremades, E., Barragan, F., Alvarez, S.: Covalent radii revisited, *J. Chem. Soc. Dalt. Trans.* (2008) 2832–2838. <https://doi.org/10.1039/b801115j>.

Crowley, P. B., Kyne, C., Monteith, W. B.: Simple and inexpensive incorporation of ^{19}F -Tryptophan for protein NMR spectroscopy, *Chem. Commun.* 48 (2012) 10681–10683. <https://doi.org/10.1039/c2cc35347d>.

Dalvit, C., Piotto, M.: ^{19}F NMR transverse and longitudinal relaxation filter experiments for screening: a theoretical and experimental analysis, *Magn. Reson. Chem.* 55 (2017) 106–114. <https://doi.org/10.1002/mrc.4500>.

Debelouchina, G. T., Muir, T. W.: A molecular engineering toolbox for the structural biologist, *Q. Rev. Biophys.* 50 (2017). <https://doi.org/10.1017/S0033583517000051>.

Deng, W., Cheeseman, J. R., Frisch, M. J.: Calculation of nuclear spin-spin coupling constants of molecules with first and second row atoms in study of basis set dependence, *J. Chem. Theory Comput.* 2 (2006) 1028–1037. <https://doi.org/10.1021/ct600110u>.

DeRider, M. L., Wilkens, S. J., Waddell, M. J., Bretscher, L. E., Weinhold, F., Raines, R. T., Markley, J. L.: Collagen stability: Insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations, *J. Am. Chem. Soc.* 124 (2002) 2497–2505. <https://doi.org/10.1021/ja0166904>.

Dietz D., Kubyshev V., Budisa N.: Applying gamma-Substituted Prolines in the Foldon Peptide: Polarity Contradicts Preorganization, *Chembiochem* 16 (2015) 403–406. <https://doi.org/10.1002/cbic.201402654>.

Eberhardt, E. S., Panasik, N., Raines, R. T.: Inductive effects on the energetics of prolyl peptide bond isomerization: Implications for collagen folding and stability, *J. Am. Chem. Soc.* 118 (1996) 12261–12266. <https://doi.org/10.1021/ja9623119>.

Edwards, L. J., Savostyanov, D. V., Welderufael, Z. T., Lee, D., Kuprov, I.: Quantum mechanical NMR simulation algorithm for protein-size spin systems, *J. Magn. Reson.* 243 (2014) 107–113. <https://doi.org/10.1016/j.jmr.2014.04.002>.

Eftekharzadeh, B., Piai, A., Chiesa, G., Mungianu, D., Garcia, J., Pierattelli, R., Felli, I.C., Salvatella X.: Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract, *Biophysical Journal* 110 (2016) 2361-2366.

<https://doi.org/10.1016/j.bpj.2016.04.022>

Evanics, F., Bezsonova, I., Marsh, J., Kitevski, J. L., Forman-Kay, J. D., Prosser, R.S.: Tryptophan Solvent Exposure in Folded and Unfolded States of an SH3 Domain by ¹⁹F and ¹H NMR, *Biochemistry*. 129 (2007) 1826–1835. <https://doi.org/10.1021/bi061389r>

Fäcke, T., Berger, S.: SERF, a New Method for ¹H, ¹H Spin-Coupling Measurement in Organic Chemistry, *J. Magn. Reson. Ser. A.* 113 (1995) 114–116. <https://doi.org/10.1006/jmra.1995.1063>.

Farrow, N. A., Zhang, O., Forman-Kay, J.D., Kay, L.E.: Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer, *Biochemistry*. 34 (1995) 868–878. <https://doi.org/10.1021/bi00003a021>.

Ferreiro, D. U., Komives, E. A., Wolynes, P. G.: Frustration in biomolecules, *Q. Rev. Biophys.* 47 (2014) 285–363. <https://doi.org/10.1017/S0033583514000092>.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A., Jr., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, N.J., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, Ö., Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox, D.J. Gaussian 09, in, Gaussian, Inc., Wallingford CT, USA, 2009.

Gee, C. T., Arntson, K. E., Urick, A. K., Mishra, N. K., Hawk, L. M. L., Wisniewski, A. J., Pomerantz, W. C. K.: Protein-observed ¹⁹F-NMR for fragment screening, affinity

quantification and druggability assessment, *Nat. Protoc.* 11 (2016) 1414–1427. <https://doi.org/10.1038/nprot.2016.079>

Gerig, J., McLeod, R.: Conformations of cis- and trans-4-Fluoroproline in Aqueous Solution, *J. Am. Chem. Soc.* 17 (1973) 4788–4795. <https://doi.org/10.1021/ja00798a046>.

Gillis, E. P., Eastman, K. J., Hill, M. D., Donnelly, D. J., Meanwell, N. A.: Applications of Fluorine in Medicinal Chemistry, *J. Med. Chem.* 58 (2015) 8315–8359. <https://doi.org/10.1021/acs.jmedchem.5b00258>.

Gimenez, D., Phelan, A., Murphy, C. D., Cobb S. L.: ^{19}F NMR as a tool in chemical biology, *J. Org. Chem.* 17 (2021) 293–318. <https://doi.org/10.3762/BJOC.17.28>.

Goodwin, D. L., Kuprov, I.: Auxiliary matrix formalism for interaction representation transformations, optimal control, and spin relaxation theories, *J. Chem. Phys.* 143 (2015). <https://doi.org/10.1063/1.4928978>.

Hofman, G.-J., Ottoy, E., Light, M., Kieffer, B., Kuprov, I., Martins, J. C., Sinnaeve, D., Linclau, B.: Minimising conformational bias in fluoroprolines through vicinal difluorination, *Chem. Commun.* 54 (2018) 5118–5121. <https://doi.org/10.1039/c8cc01493k>.

Hofman, G.-J., Ottoy, E., Light, M. E., Kieffer, B., Martins, J. C., Kuprov, I., Sinnaeve, D., Linclau, B.: Synthesis and Conformational Properties of 3,4-Difluoro-1-prolines, *J. Org. Chem.* 84 (2019) 3100–3120. <https://doi.org/10.1021/acs.joc.8b02920>.

Hogben, H. J., Krzystyniak, M., Charnock, G. T. P., Hore, P. J., Kuprov, I.: Spinach - A software library for simulation of spin dynamics in large spin systems, *J. Magn. Reson.* 208 (2011) 179–194. <https://doi.org/10.1016/j.jmr.2010.11.008>.

Holmgren, S. K., Taylor, K. M., Bretscher, L. E., Raines, R. T.: Code for Collagen's Stability Deciphered, *Nature*. 392 (1998) 666–667. <https://doi.org/10.1038/33573>.

Horng, J.-C. Raines, R. T.: Stereoelectronic effects on polyproline conformation, *Protein Sci.* 15 (2006) 74–83. <https://doi.org/10.1110/ps.051779806>.

Jiji, A. C., Shine, A., Vijayan, V.: Direct Observation of Aggregation-Induced Backbone Conformational Changes in Tau Peptides, *Angew. Chemie - Int. Ed.* 55 (2016) 11562–11566. <https://doi.org/10.1002/anie.201606544>.

Kang, Y. K.: Puckering transition of proline residue in water, *J. Phys. Chem. B.* 111 (2007) 10550–10556. <https://doi.org/10.1021/jp073411b>.

Kim, T. H., Mehrabi, P., Ren, Z., Sljoka, A., Ing, C., Bezginov, A., Ye, L., Pomès, R., Prosser, R. S., Pai, E. F.: The role of dimer asymmetry and protomer dynamics in enzyme catalysis, *Science*. 355 (2017) 6–11. <https://doi.org/10.1126/science.aag2355>.

Kitevski-LeBlanc, J. L., Prosser, R. S.: Current applications of ^{19}F NMR to studies of protein structure and dynamics, *Prog. Nucl. Magn. Reson. Spectrosc.* 62 (2012) 1–33. <https://doi.org/10.1016/j.pnmrs.2011.06.003>.

Kohler, C., Recht, R., Quinternet, M., de Lamotte, F., Delsuc, M.-A., Kieffer, B.: Accurate Protein–Peptide Titration Experiments by Nuclear Magnetic Resonance Using Low-Volume Samples, *Methods Mol. Biol.* 1286 (2015) 279–296. https://doi.org/10.1007/978-1-4939-2447-9_22.

Kovrigin, E. L.: NMR line shapes and multi-state binding equilibria, *J. Biomol. NMR.* 53 (2012) 257–270. <https://doi.org/10.1007/s10858-012-9636-3>.

Kubyshev, V., Davis, R., Budisa, N.: Biochemistry of fluoroprolines: The prospect of making fluorine a bioelement, *J. Org. Chem.* 17 (2021) 439–450. <https://doi.org/10.3762/BJOC.17.40>.

Kuprov, I., Wagner-Rundell, N., Hore, P.J.: Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction, *J. Magn. Reson.* 189 (2007) 241–250. <https://doi.org/10.1016/j.jmr.2007.09.014>.

Kuprov, I.: Diagonalization-free implementation of spin relaxation theory for large spin systems, *J. Magn. Reson.* 209 (2011) 31–38. <https://doi.org/10.1016/j.jmr.2010.12.004>.

Lalevée, S., Bour, G., Quinternet, M., Samarut, E., Kessler, P., Vitorino, M., Bruck, N., Delsuc, M.-A., Vonesch, J.-L., Kieffer, B., Rochette-Egly, C.: Vinexin β , an atypical “sensor” of retinoic acid receptor γ signaling: union and sequestration, separation, and phosphorylation, *FASEB J.* 24 (2010) 4523–4534. <https://doi.org/10.1096/fj.10-160572>.

Liu, J. J., Horst, R., Katritch, V., Stevens, R. C., Wüthrich, K.: Biased signaling pathways in β_2 -adrenergic receptor characterized by ^{19}F -NMR, *Science.* 335 (2012) 1106–1110. <https://doi.org/10.1126/science.1215802>.

London, F.: Théorie quantique des courants interatomiques dans les combinaisons aromatiques, *J. Phys. Le Radium.* 8 (1937) 397–409. <https://doi.org/10.1051/jphysrad:01937008010039700>.

London, R. E.: On the Interpretation of ^{13}C Spin-Lattice Relaxation Resulting from Ring Puckering in Proline, *J. Am. Chem. Soc.* 100 (1978) 2678–2685. <https://doi.org/10.1021/ja00477a018>.

Lu, M., Ishima, R., Polenova, T., Gronenborn, A.M.: ^{19}F NMR relaxation studies of fluorosubstituted tryptophans, *J. Biomol. NMR.* 73 (2019) 401–409. <https://doi.org/10.1007/s10858-019-00268-y>.

Marenich, A. V., Cramer, C. J., Truhlar, D. G.: Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant

and atomic surface tensions, *J. Phys. Chem. B.* 113 (2009) 6378–6396. <https://doi.org/10.1021/jp810292n>.

Mei, H., Han, J., Klika, K. D., Izawa, K., Sato, T., Meanwell, N. A., Soloshonok V. A.: Applications of fluorine-containing amino acids for drug design, *Eur. J. Med. Chem.* 186 (2020) 111826. <https://doi.org/10.1016/j.ejmech.2019.111826>.

Muttenthaler, M., King, G. F., Adams, D.J., Alewood, P. F.: Trends in peptide drug discovery, *Nat. Rev. Drug Discov.* 20 (2021) 309–325. <https://doi.org/10.1038/s41573-020-00135-8>.

Newberry, R. W., Raines, R. T.: 4-Fluoroproline: Conformational Analysis and Effects on the Stability and Folding of Peptides and Proteins, *Topics in Heterocyclic Chemistry: Peptidomimetics I.* (2016) 1–11. https://doi.org/10.1007/7081_2015_196.

Odar, C., Winkler, M., Wiltschi, B.: Fluoro amino acids: A rarity in nature, yet a prospect for protein engineering, *Biotechnol. J.* 10 (2015) 427–446. <https://doi.org/10.1002/biot.201400587>.

O'Hagan, D.: Understanding organofluorine chemistry. An introduction to the C–F bond, *Chem. Soc. Rev.* 37 (2008) 308–319. <https://doi.org/10.1039/b711844a>.

O'Hagan, D.: Organofluorine chemistry: Synthesis and conformation of vicinal fluoromethylene motifs, *J. Org. Chem.* 77 (2012) 3689–3699. <https://doi.org/10.1021/jo300044q>.

Overbeck, J.H., Kremer, W., Sprangers, R.: A suite of ^{19}F based relaxation dispersion experiments to assess biomolecular motions, *J. Biomol. NMR.* 74 (2020) 753–766. <https://doi.org/10.1007/s10858-020-00348-4>.

Panasik, N., Eberhardt, E. S., Edison, A. S., Powell, D. R., Raines, R. T.: Inductive effects on the structure of proline residues, *Int. J. Pept. Protein Res.* 44 (1994) 262–269. <https://doi.org/10.1111/j.1399-3011.1994.tb00169.x>.

Pell, A. J., Keeler, J.: Two-dimensional J-spectra with absorption-mode lineshapes, *J. Magn. Reson.* 189 (2007) 293–299. <https://doi.org/10.1016/j.jmr.2007.09.002>.

Peterson, K. A., Dunning, T. H.: Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al–Ar, and the first row atoms B–Ne revisited, *J. Chem. Phys.* 117 (2002) 10548–10560. <https://doi.org/10.1063/1.1520138>.

Rastinejad, F., Evilia, C., Lu, P.: Studies of Nucleic Acids and Their Protein Interactions by ^{19}F NMR, *Methods Enzymol.*, 261 (1995) 560–575. [https://doi.org/10.1016/s0076-6879\(95\)61025-1](https://doi.org/10.1016/s0076-6879(95)61025-1).

Redfield, A.G.: On the Theory of Relaxation Processes, *IBM J. Res. Dev.* (1957) 19–31. <https://doi.org/10.1147/rd.11.0019>.

Ruzza, P., Siligardi, G., Donella-Deana, A., Calderan, A., Hussain, R., Rubini, C., Cesaro, L., Osler, A., Guiotto, A., Pinna, L. A., Borin, G. 4-Fluoroproline derivative peptides: Effect on PPII conformation and SH3 affinity, *J. Pept. Sci.* 12 (2006) 462–471. <https://doi.org/10.1002/psc.750>.

Salwiczek, M., Nyakatura, E. K., Gerling, U. I. M., Ye, S., Koksche, B.: Fluorinated amino acids: Compatibility with native protein structures and effects on protein–protein interactions, *Chem. Soc. Rev.* 41 (2012) 2135–2171. <https://doi.org/10.1039/c1cs15241f>.

Saksela, K., Permi, P.: SH3 domain ligand binding: What’s the consensus and where’s the specificity?, *FEBS Lett.* 586 (2012) 2609–2614. <https://doi.org/10.1016/j.febslet.2012.04.042>.

Sarkar, S. K., Young, P. E., Torchia, D. A.: Ring Dynamics of DL-Proline and DL-Proline Hydrochloride in the Solid State: A ^2H Nuclear Magnetic Resonance Study, *J. Am. Chem. Soc.* 108 (1986) 6459–6464. <https://doi.org/10.1021/ja00281a002>.

Sharaf, N. G., Gronenborn, A. M.: ^{19}F -Modified Proteins and ^{19}F -Containing Ligands as Tools in Solution NMR Studies of Protein Interactions, *Methods Enzymol.* 565 (2015) 67–95. <https://doi.org/10.1016/bs.mie.2015.05.014>.

Sinnaeve, D. Selective Homonuclear 2D J-Resolved Spectroscopy. *eMagRes* 9 (2021) 267–282. <https://doi.org/10.1002/9780470034590.emrstm1544>

Schubert, M., Labudde, D., Oschkinat, H., Schmieder, P.: A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics, *J. Biomol. NMR.* 24 (2002) 149–154. <https://doi.org/10.1023/A:1020997118364>.

Shoulders, M.D., Raines, R.T.: Collagen structure and stability, *Annu. Rev. Biochem.* 78 (2009) 929–958. <https://doi.org/10.1146/annurev.biochem.77.032207.120833>.

Stadmler, S.S., Aguilar, J.S., Waudby, C.A., Pielak, G.J.: Rapid Quantification of Protein-Ligand Binding via ^{19}F NMR Lineshape Analysis, *Biophys. J.* 118 (2020) 2537–2548. <https://doi.org/10.1016/j.bpj.2020.03.031>.

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A. K., Daughdrill, G. W., Uversky, V. N.: The alphabet of intrinsic disorder, *Intrinsically Disord. Proteins.* 1 (2013) e24360. <https://doi.org/10.4161/idp.24360>.

Thiehoff, C., Rey, Y.P., Gilmour, R.: The Fluorine Gauche Effect: A Brief History, *Isr. J. Chem.* 57 (2017) 92–100. <https://doi.org/10.1002/ijch.201600038>.

Torbeev, V., Ebert, M. O., Dolenc, J., Hilvert, D.: Substitution of proline³² by α -methylproline preorganizes β 2-microglobulin for oligomerization but not for aggregation into amyloids, *J. Am. Chem. Soc.* 137 (2015) 2524–2535. <https://doi.org/10.1021/ja510109p>.

Torbeev, V.Y., Hilvert, D.: Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate β 2-microglobulin amyloid assembly, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 20051–20056. <https://doi.org/10.1073/pnas.1310414110>.

Urbanek, A., Popovic, M., Elena-Real, C. A., Morató, A., Estaña, A., Fournet, A., Allemand, F., Gil, A. M., Cativiela, C., Cortés, J., Jiménez, A. I., Sibille, N., Bernadó, P.: Evidence of the Reduced Abundance of Proline cis Conformation in Protein Poly Proline Tracts, *J. Am. Chem. Soc.* 142 (2020) 7976–7986. <https://doi.org/10.1021/jacs.0c02263>.

Urbanek, A., Popovic, M., Morató, A., Estaña, A., Elena-Real, C.A., Mier, P., Fournet, A., Allemand, F., Delbecq, S., Andrade-Navarro, M.A., Cortés, J., Sibille, N., Bernadó, P.: Flanking regions determine the structure of the poly-glutamine in huntingtin through mechanisms common among glutamine-rich human proteins, *Structure* 28 (2020) 733-746 <https://doi.org/10.1016/j.str.2020.04.008>

Verhoork, S. J. M., Killoran, P. M., Coxon, C. R.: Fluorinated Prolines as Conformational Tools and Reporters for Peptide and Protein Chemistry, *Biochemistry* 57 (2018) 6132–6143. <https://doi.org/10.1021/acs.biochem.8b00787>.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., Laue, E. D.: The CCPN data model for NMR spectroscopy: Development of a software pipeline, *Proteins Struct. Funct. Genet.* 59 (2005) 687–696. <https://doi.org/10.1002/prot.20449>.

Welte, H., Zhou, T., Mihajlenko, X., Mayans, O., Kovermann, M.: What does fluorine do to a protein? Thermodynamic, and highly-resolved structural insights into fluorine-labelled variants of the cold shock protein, *Sci. Rep.* 10 (2020) 1–12. <https://doi.org/10.1038/s41598-020-59446-w>.

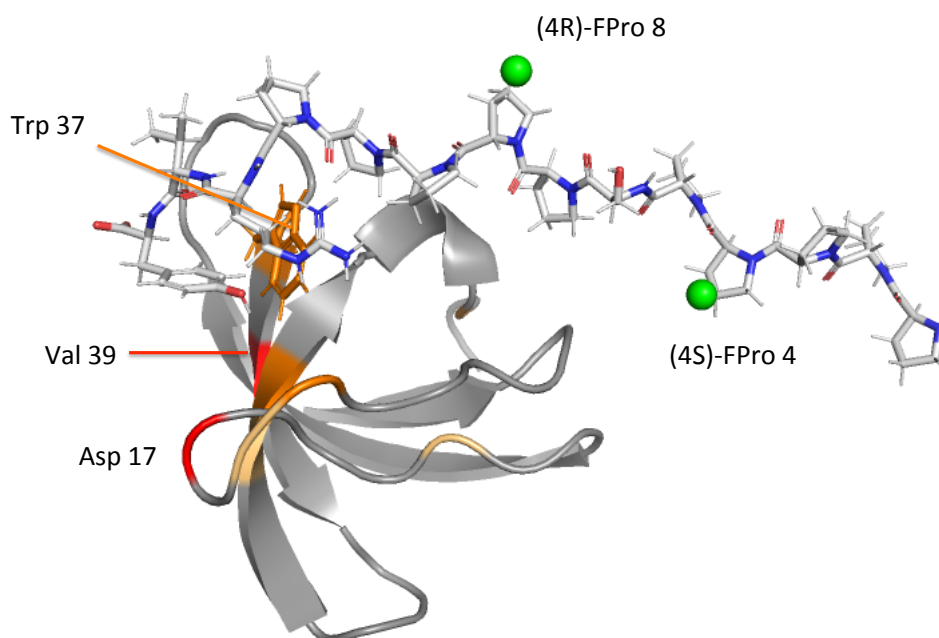
Wilhelm, P., Lewandowski, B., Trapp, N., Wennemers, H.: A crystal structure of an oligoproline PPII-Helix, at last, *J. Am. Chem. Soc.* 136 (2014) 15829–15832. <https://doi.org/10.1021/ja507405j>.

Yoshida, A.: Studies on the mechanism of protein synthesis, Incorporation of p-fluorophenylalanine into α -amylase of *Bacillus subtilis*, *BBA - Biochim. Biophys. Acta.* 41 (1960) 98–103. [https://doi.org/10.1016/0006-3002\(60\)90373-5](https://doi.org/10.1016/0006-3002(60)90373-5)

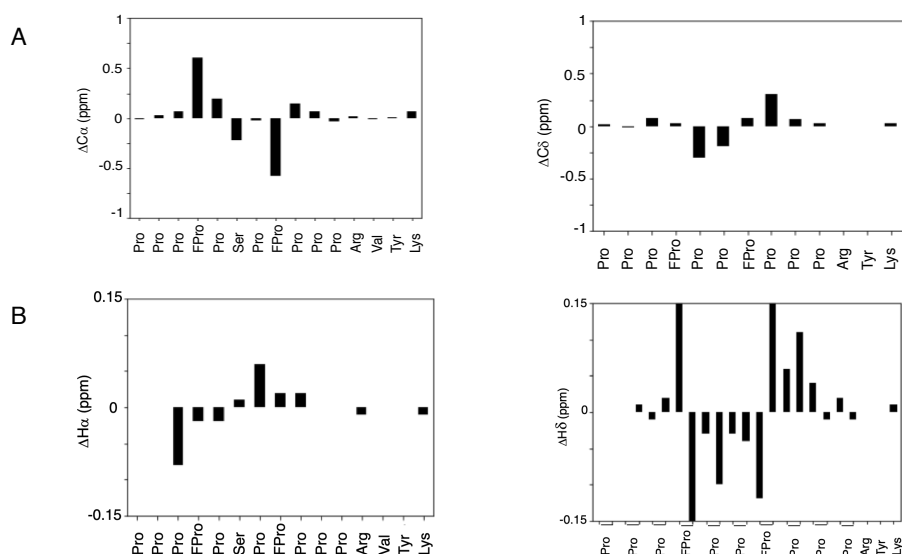
Zhao, Y., Truhlar, D.G.: The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.* 120 (2008) 215–241. <https://doi.org/10.1007/s00214-007-0310-x>.

Zhang, C., Moonshi, S. S., Han, Y., Puttick, S., Peng, H., Magoling, B. J. A., Reid, J. C., Bernardi, S., Searles, D. J., Král, P., Whittaker, A. K.: PFPE-Based Polymeric ^{19}F MRI Agents: A New Class of Contrast Agents with Outstanding Sensitivity, *Macromolecules*. 50 (2017) 5953–5963. <https://doi.org/10.1021/acs.macromol.7b01285>.

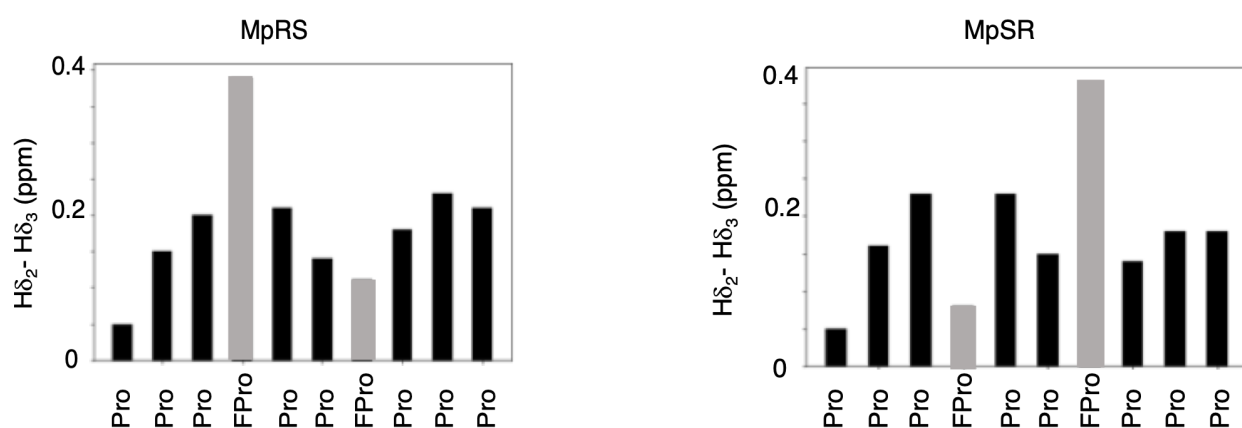
5. Supplementary Material



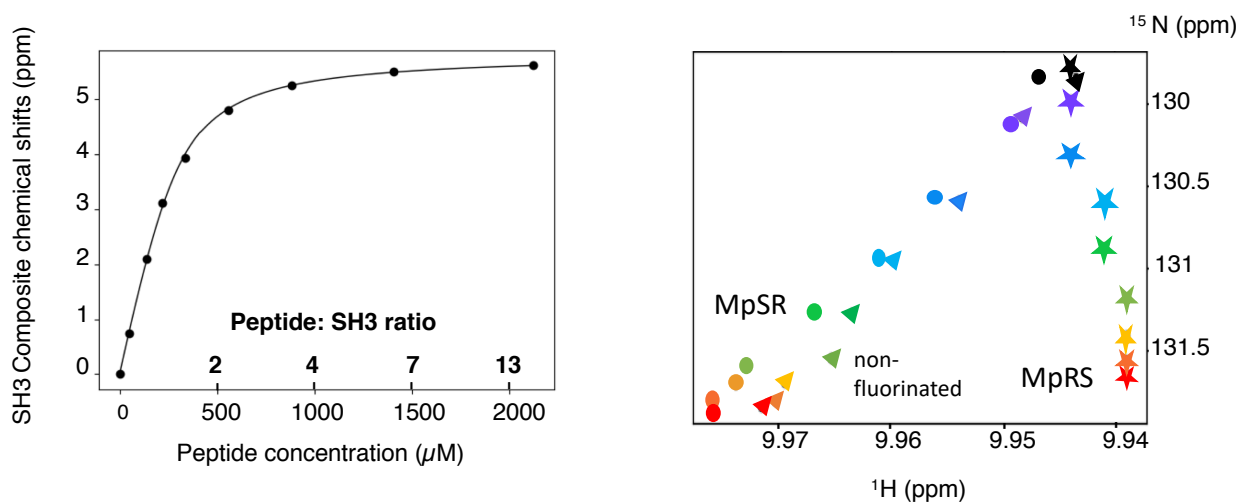
Supplementary Figure 1: 3D homology model of a complex between the MpSR model peptide and the SH3.3 domain of 25 Vinexin β . The structure of the SH3.3 domain was built using the structure of the homologous SH3.1 domain from the same protein (PDBID 2CT3) using MODELLER. The position of the ligand peptide was inferred from the structure 1PRL with sequence AFAPPLPRR. The polyproline segments were built assuming a PPII conformation with dihedral angles of -75° and 145° for phi and psi, respectively. The figure was made using Pymol. The residues displaying the largest chemical shift perturbations are colored from light orange to red according the importance of the perturbation



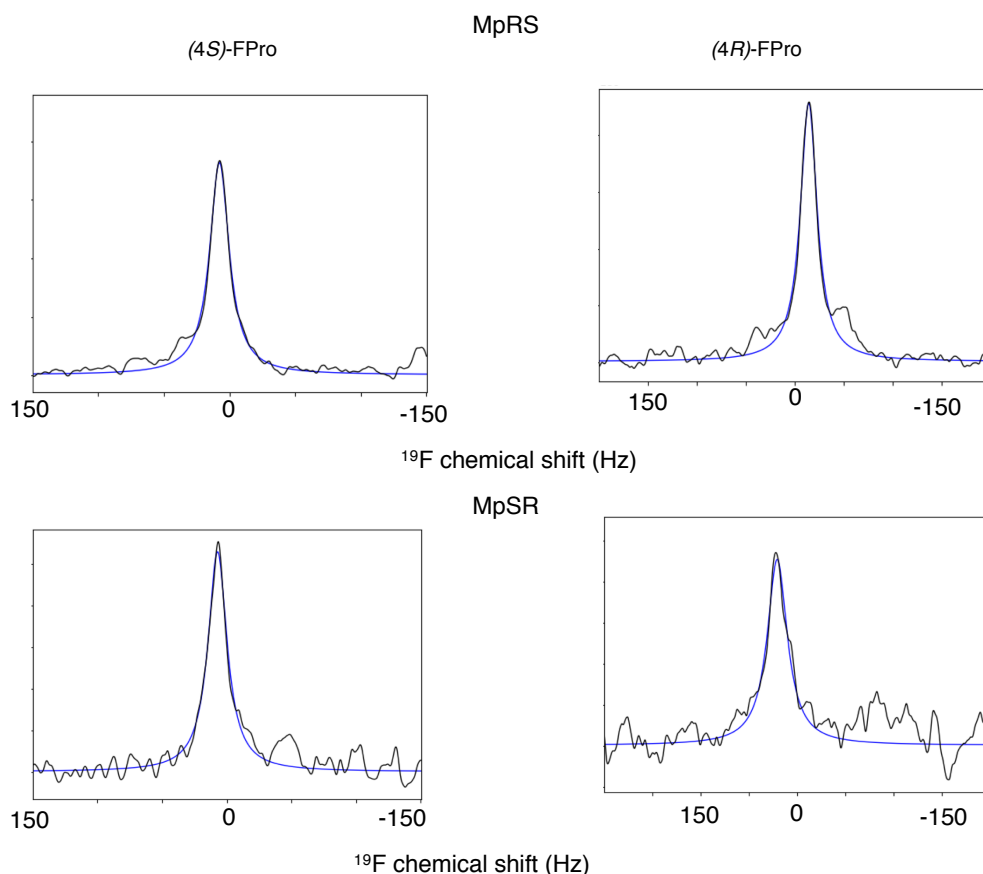
Supplementary Figure 2: Comparison of chemical shifts measured for MpRS and MpSR peptides in water at 298K at 600 MHz. A: ^{13}C Chemical shift difference between MpRS and MpSR at positions alpha (left) and delta (right). B: ^1H Chemical shift difference between MpRS and MpSR at two positions alpha (left) and delta (right).



Supplementary Figure 3: Comparison of delta geminal protons chemical shifts differences measured for the MpRS and MpSR peptides in water at 298 K at 600 MHz. Fluoroprolines at positions 4 and 8 are shown by gray bars.



Supplementary Figure 4: Left: ^1H , ^{15}N composite chemical shift measured in the ^1H - ^{15}N HSQC spectra during the titration of the Vinexin β SH3.3 domain with a non-fluorinated model peptide. The fit of the values with a Langmuir model is shown by a solid line and resulted into a dissociation equilibrium constant of $75 \pm 5 \mu\text{M}$. Right: Position of the Trp 37 H_n - N_ϵ cross 45 peak during the titration. The MpSR and MpRS fluorinated peptides are indicated by discs and stars, respectively while the cross peaks corresponding to the non-fluorinated peptide are show by triangles.



Supplementary Figure 5: Fit of the signals corresponding to (4S)-FPro (left) and (4R)-FPro (right) residues in the MpRS (top) and MpSR (bottom) peptides at the first titration point. Concentrations are 55.6 μ M and 50 μ M of MpRS and MpSR 50 respectively. The spectra were recorded at 298 K, on a 600 MHz in a 3 mm tube with 2048 scans. Measured signal to noise ratios are 71 for MpRS and 20 for MpSR, after applied apodization of 16 Hz. The fitted linewidths of (4S)-FPro are 12.95 Hz and 12.66 Hz in MpRS and MpSR respectively leading to R2 values of 40.7 s⁻¹ and 39.8 s⁻¹. For (4R)-FPro, the linewidths are 15.6 Hz and 20.24 Hz and the corresponding R2 are 49.3 s⁻¹ and 63.6 s⁻¹ in MpRS and MpSR respectively

	K_d (μ M)		
	MpRS	MpSR	Non-fluorinated peptide
Gln 14	223 $\pm$ 5	76 $\pm$ 2	119 $\pm$ 30
Asn 15	219 $\pm$ 2	80 $\pm$ 2	61 $\pm$ 2
Asp 17	219 $\pm$ 5	73 $\pm$ 2	71 $\pm$ 2
Leu 21	209 $\pm$ 7	73 $\pm$ 2	70 $\pm$ 4
Trp 37	216 $\pm$ 2	73 $\pm$ 2	70 $\pm$ 3
Val 39	219 $\pm$ 4	75 $\pm$ 2	69 $\pm$ 2
Gly 49	223 $\pm$ 8	70 $\pm$ 4	53 $\pm$ 3
Thr 50	227 $\pm$ 4	73 $\pm$ 2	72 $\pm$ 3
Val 56	218 $\pm$ 2	80 $\pm$ 5	56 $\pm$ 3

Supplementary Table 1: Dissociation equilibrium constants measured for the MpRS, MpSR and non-fluorinated peptide at 298 K in 40 mM phosphate buffer at pH 7. These constants have been obtained by fitting the variations of individual ^1H - ^{15}N cross-peaks positions upon addition of peptide with a 1:1 equilibrium model. The uncertainties are obtained from the covariance matrix of the fit and do not take in consideration uncertainties on peptide and protein concentrations.

	MpRS	MpSR	Non-fluorinated peptide
Gln 14	1.863	1.927	1.949
Asn 15	1.079	1.126	1.093
Asp 17	2.220	2.286	2.276
Leu 21	1.002	0.983	0.988
Trp 37	1.870	2.012	2.011
Val 39	3.094	3.198	3.214
Gly 49	0.800	0.788	0.812
Thr 50	1.841	1.807	1.881
Val 56	1.319	1.310	1.319

Supplementary Table 2: ^1H - ^{15}N Composite chemical shift perturbation measured for Vinexin β residues that the most affected by the addition of polyproline model peptides.

Chapter 6: Fine-tuning molten-globule conformational dynamics by selective incorporation of fluoroprolines

VI. Fine-tuning molten-globule conformational dynamics by selective incorporation of fluoroprolines

Abir Ben Bouzayene^{1,2}, co-authors, Vladimir Torbeev² and Bruno Kieffer¹

¹Departement of Integrative Structural Biology, IGBMC, Université de Strasbourg, INSERM U1258, CNRS UMR7104, 1, rue Laurent Fries, F-67404 Illkirch, France

²École Supérieure de Biotechnologie de Strasbourg (ESBS), UMR 7242 Biotechnology and Cellular Signalling, CNRS and University of Strasbourg, 67400 Illkirch, France

1. Introduction

The discovery of the ‘molten-globule state’ marked a turning point in the understanding of protein folding and function.^{1 2 3} Molten-globule, as the name suggests, is a fluctuating but compact protein structure effectively composed of multiple interconverting polypeptide conformations.⁴ In folding of many proteins, molten-globule links unfolded polypeptide and fully folded protein, serving as a key intermediate helping to acquire native intramolecular interactions and to converge to the correct tertiary structure.^{5 6 7} Furthermore, molten-globule is believed to act as a transitory state to misfolded and aggregated protein species, in particular observed in refolding experiments with isolated proteins in laboratory^{8 9 10} but also invoked in chaperon-assisted protein folding *in vivo*^{11 12}. Initially, to populate and characterize molten-globule state, conditions that favour partial unfolding were applied such as low pH, presence of moderate concentrations of denaturant or removal of disulfide bonds.^{1 2 13} Then, it has been invoked in various kinetic models of protein folding at physiological conditions.^{14 15 16 17 18} Later on, molten-globule was postulated to be a distinct species on the structure-function continuum in the transition from fully folded proteins to intrinsically disordered proteins, a class of natively unstructured proteins playing distinct and diverse biological functions.¹⁹

Molten-globules possess a high degree of native contacts and elements of secondary structure but little or no tertiary packing. In the so-called ‘wet’ molten globules the protein core is accessible to solvent^{3 4}, whereas ‘dry’ molten globules first observed in protein unfolding experiments constitute expanded forms of native proteins containing defects of close-packing while having their core protected from solvent.²⁰ Counterintuitively, despite the absence of tertiary structure and conformational fluctuations, the secondary structures are stabilized presumably due to transient intramolecular interactions. Such conformational plasticity of

molten-globules has some functional advantages by enabling an efficient allosteric regulation when binding to other biomolecules and, moreover, specific binding interactions with biomolecules that are structurally distinct.^{3 4}

Due to multiple conformations existing in dynamic equilibrium, the detailed structural characterization of molten-globules can be challenging. NMR spectroscopy has many advantages to probe the properties of such dynamic species being able to provide both structural information and to characterize the conformational dynamics. The protein of interest in this study is nuclear co-activator binding domain (NCBD) from transcriptional co-activator CREB-binding protein (CBP) which was previously shown to be molten-globule with a small cooperatively folded core and presence of the specific side chain packing.^{21 22 23} Remarkably, it can bind structurally unrelated proteins such as interferon regulatory factor 3 (IRF-3), activation domain of transcriptional co-activator ACTR (AD-ACTR) or transactivation domain of p53 protein (TAD-p53) adopting distinct conformations.²⁴ The two prolines at sites 2077 and 2094 in the sequence can regulate conformational sampling between these conformations in unbound form of NCBD and mediate binding interactions. Indeed, previously *cis-trans* amide isomerization of Pro2077 was invoked to play a role in the binding mechanism with AD-ACTR.²⁵

In this study we combine ¹⁹F NMR with the site-specific incorporation of stereoisomeric 4-(*R*)- and 4-(*S*)-fluoroprolines which can serve as local NMR probes but also perturb the conformational properties at substitution sites in a controlled way. For instance, 4-(*R*)-fluoroproline is known to enhance the propensity for *trans*-amide conformer, whereas 4-(*S*)-fluoroproline has tendency to overpopulate the *cis*-prolyl isomer.²⁶ In addition, the two proline analogues have distinct preferences to adopt either *exo*- or *endo*-conformations of five-membered pyrrolidine ring with 4-(*R*)-fluoroproline favouring the former and 4-(*S*)-fluoroproline preferring the latter.²⁶ Both 4-(*R*)- and 4-(*S*)-fluoroprolines have faster *cis-trans* amide isomerization kinetics.²⁷ Nevertheless, all these changes are rather subtle from structural and energetic point of view – single fluorine substituent is only slightly bigger than hydrogen and stereo electronic effects with resulting conformational alterations correspond to free energy changes below 0.5 kcal/mol. Here we present the detailed studies dissecting how the NCBD molten-globule responds to such changes and derive new insights into conformational transitions and their functional significance.

2. Results

2.1. Insertion of conformationally biased fluoroprolines into NCBD protein domain

The NCBD is a small domain composed of three α -helices and two loops. Both loops contain conserved proline residues albeit with different sequence contexts. The loop between helices α_1 and α_2 (loop L1) is five residues long and contains three serines and two prolines, whereas the loop between helices α_2 and α_3 (loop L2) contains a single proline preceded by an asparagine and followed up by glutamine (Figure 1A). Four modified NCBD 47-residue variants (corresponding to [2066–2112]-sequence fragment of CBP) were synthesized using solid-phase peptide synthesis by inserting either a 4-(*R*)-FPro or a 4-(*S*)-FPro at positions 2077 and 2094 hereafter named P13 and P30. The resulting proteins were purified and dissolved in a solution of 20 mM phosphate buffer at pH 7.2. The folding of the proteins was first investigated by circular dichroism (CD) spectroscopy. The four modified NCBD fragments display similar far-UV spectra with strong negative ellipticity peaks at 208 nm and 220 nm characteristic of alpha-helical secondary structure. These spectra are superimposable to the non-modified NCBD fragment suggesting that insertion of either 4-(*R*)-FPro or a 4-(*S*)-FPro within the two NCBD loops does not significantly affect the global helical secondary structure of the protein (Figure S1).

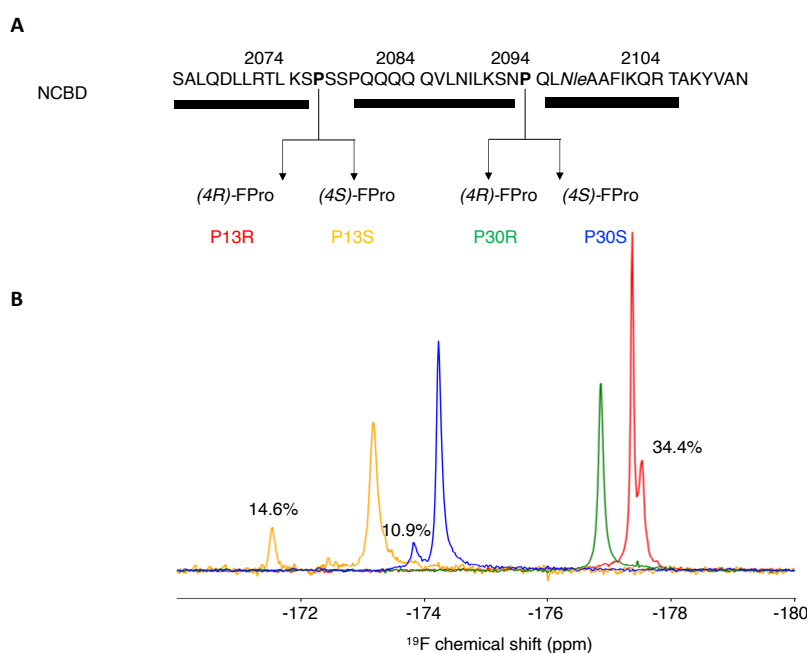


Figure 1: A) Sequence of NCBD protein and insertion positions of the fluoroprolines. B) 1D ¹⁹F NMR spectra of the four modified NCBD domains. Red and orange spectra correspond to

the insertion at position 2077 of 4-(*R*)-FPro or 4-(*S*)-FPro, respectively. Green and blue spectra denote the incorporation of 4-(*R*)-FPro or 4-(*S*)-FPro respectively at position 2094. *Nle* refers to norleucine amino acid replacing the methionine to avoid thioether oxidation. The peptides were dissolved in H₂O/D₂O buffered by 20 mM sodium phosphate. P13R, P13S, P30R and P30S refer to the four fluorinated NCBD variants. The spectra were recorded on a 600 MHz spectrometer at 298 K.

The comparison of the ¹⁹F NMR spectra recorded for the four modified NCBD variants reveals distinct features for prolines inserted within loops L1 (P13) and L2 (P30). The insertion of fluoroproline in L2 leads mostly to a single resonance at -174.2 and -176.86 ppm for the 4-(*S*)-FPro and 4-(*R*)-FPro residues, respectively (referenced as P30S and P30R in Figure 1B) displaying similar linewidths of 70 Hz. The presence of the second peak for the P30S variant (population 10.9 %) suggests the presence of *cis*-amide prolyl isomer enhanced in this protein analogue by a known propensity of 4-(*S*)-FPro to populate *cis*-conformer and suppressed by 4-(*R*)-FPro in the P30R NCBD variant due to its preference to populate *trans*-amide. A different behaviour was observed upon the insertion of fluoroproline in L1. In the ¹⁹F NMR spectra recorded for the (4*S*)-FPro13 and (4*R*)-FPro13 variants (labelled as P13S and P13R in Figure 1B) exhibited well-defined minor peaks accounting for 14.6% and 34.4% of the signal for P13S and P13R, respectively. Such percentages were not expected as typically minor *cis*-component is better enhanced by 4-(*S*)-FPro incorporation, whereas in this case the higher percentage of the second form was observed for 4-(*R*)-FPro-containing variant. As the major amide conformations of both Pro2077 and Pro2094 are *trans* based on the previous studies, the residue Pro2094 is likely to attain two conformations upon incorporation of 4-(*S*)-FPro for another reason than *cis-trans* amide isomerism. The linewidths measured for the two major peaks are significantly different with 90 Hz for P13S and 31 Hz for P13R. The chemical shift difference between minor and major peaks of P13S is 1.6 ppm, strikingly larger than the 0.15 ppm observed for P13R. For P13S, the selective irradiation of the major peak leads to a gradual decrease of the minor peak's intensity, indicating that the corresponding populations are exchanging in the ms time scale (Figure S2). Overall, these initial observations suggest that the insertion of fluorinated proline analogues with different conformational bias in loops L1 and L2 alters differentially the conformational properties and dynamics of the NCBD molten globule while preserving its global secondary structure.

2.2. NCBD secondary structure is not altered by fluoroproline insertion

The structure of the four modified NCBD was investigated by the analysis of ^1H - ^1H NOESY spectra recorded at 304 K and in each case at two different fields corresponding to proton frequencies of 600, 700 and 900 MHz. While the use of larger magnetic fields is generally expected to improve the quality of protein NMR spectra, we obtained contrasting results with modified NCBD samples. The highest quality spectra were obtained for the P30S protein while P13S led to lower quality spectra at 900 MHz. This was interpreted as an effect of the conformational fluctuations characterizing NCBD dynamics that is sensitive to the insertion site of the fluoroproline. Nevertheless, almost complete ^1H and ^{13}C chemical shift assignments of all NCBD variants in H_2O were obtained using standard ^1H - ^1H NOESY, ^1H - ^1H TOCSY and ^1H - ^{13}C HSQC experiments recorded at 304 K and 700 MHz proton frequency (Table S1). Noteworthy, the ^1H and ^{13}C frequency shifts introduced by the substitutions of proline P30 and P13 by fluoroprolines was helpful to get an unambiguous assignment of their spin systems.

The comparison of experimental $\text{H}\alpha$ and $\text{C}\alpha$ chemical shifts with tabulated random coil values provide a convenient probe of the local protein backbone structure. The profiles obtained for the four fluorinated NCBD variants display very little differences from the unmodified protein. Consistent with the CD spectra, this observation indicates that the three helices of NCBD spanning residues S₂₀₆₅-L₂₀₇₄, Q₂₀₈₁-K₂₀₉₁ and Q₂₀₉₅-K₂₁₀₂, respectively are insensitive to the modification of either P13 or P30 by a fluorine (Figure S3). Subtle but consistent differences between $\text{C}\alpha$ chemical shifts measured for 4-(*S*)-FPro and 4-(*R*)-FPro substituted NCBD are however observed (Figure 2A). These differences extend far beyond the position of the proline modification suggesting that the local conformational bias produced by the fluorination has a global impact on the NCBD domain.

2.3. Fluoroproline insertion alters NCBD hydrophobic packing

We then investigated the potential effects of the fluoroprolines on the global folding of NCBD molten globule by monitoring the hydrodynamic radii of the modified proteins as a function of temperature variation using DOSY experiments (Figure S4). The four modified proteins as well as the unmodified NCBD display very similar translational diffusion coefficient variations for temperatures ranging from 283 K to 309 K with a linear dependency of the diffusion coefficient as a function of the temperature over viscosity ratio. The hydrodynamic radii of the five proteins derived from the slopes of the least squares lines range from 14.8 Å to

15.5 Å for WT NCBD and P13R, respectively (Table 1). P13S and P30S show very similar hydrophobic radii as the WT NCBD while P13R and P30R display slightly larger radii.

Table 1: Hydrodynamic radii of fluorinated and non-fluorinated NCBD measured from 2D ^1H - ^1H DOSY. Uncertainties on the values were estimated to be 0.2 Å from the dispersion of measured diffusion coefficients from the linear model (see Figure S4).

	NCBD WT	P13S	P13R	P30S	P30R
$R_H(\text{\AA})$	14.8	14.7	15.5	14.8	15.3

The hydrodynamic radii measured for the NCBD variants correspond to a radius of gyration around 19 Å²⁸ in agreement with the value determined by SAXS measurements under native like conditions (R_g 18.8 Å).²² These data indicate that insertions of fluoroproline in L1 and L2 loops have only marginal effects on the globular shape of NCBD that is preserved in this range of temperatures. This was further confirmed by the analysis of the temperature dependence of CD ellipticities at 222 nm measured for the four variants. (Figure S5). The lack of cooperative unfolding, considered as one of the hallmarks of molten globules is preserved in NCBD harbouring fluoroproline within their loops.

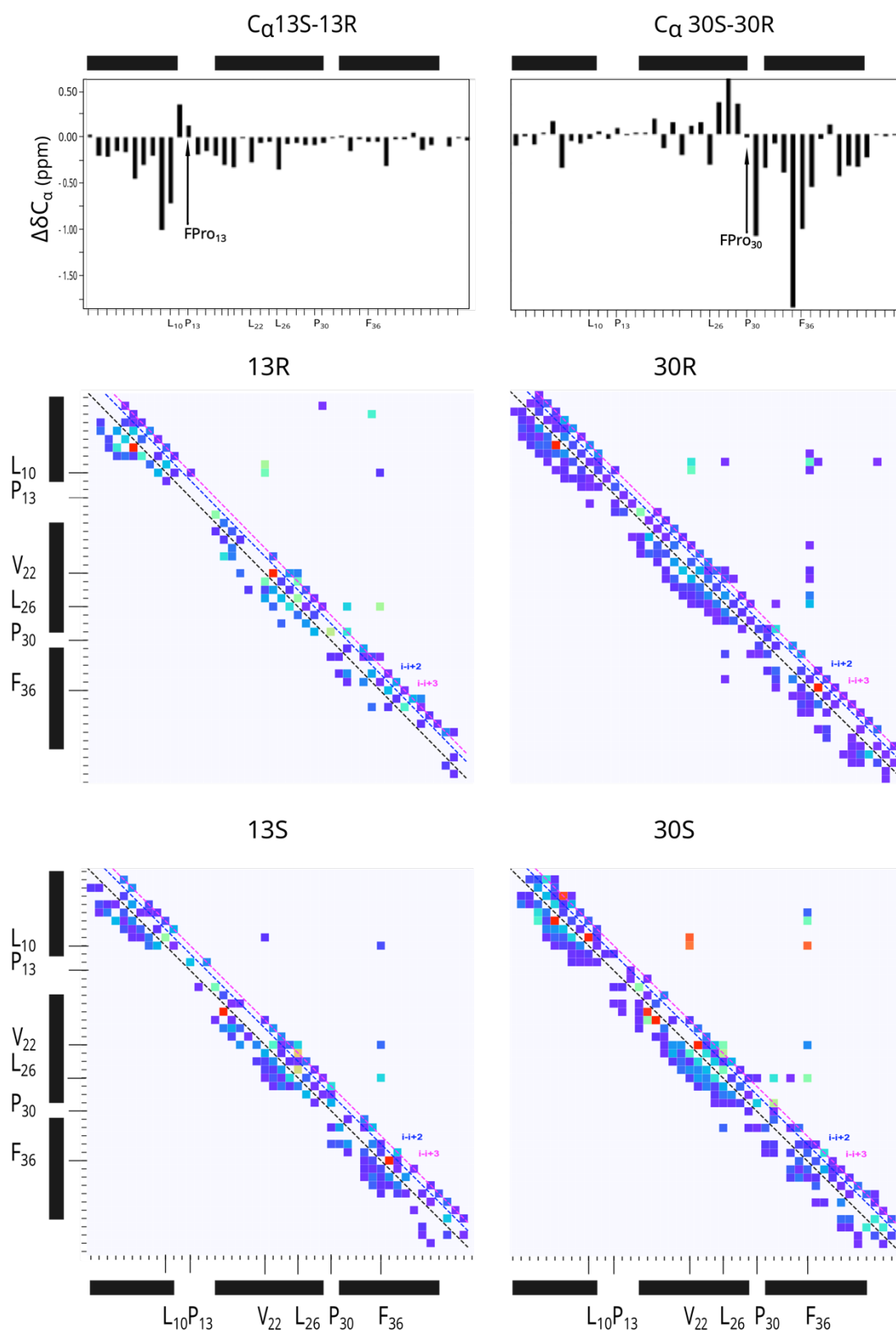


Figure 2: A) Differences between $C\alpha$ chemical shifts measured for NCBD incorporating either a 4-(*S*)-FPro or a 4-(*R*)-FPro at positions P13 (left panel) and P30 (right panel). The positions of the helices deduced from the comparison of $C\alpha$ chemical shifts with random coil values are

indicated by black rectangles. B) Contact maps displaying non-ambiguous NOEs in full NCBD identified by iterative interpretation of NOESY spectra using ARIA 2.0 structure calculation program. Upper triangles display residue-residue contacts from side chain atoms while lower triangles display the contacts between H α and HN backbone atoms. Intra residue NOEs are not represented and replaced by the diagonal dashed black line. The blue and magenta dashed lines correspond to NOE observed between a residue at position i and a residue at position $i+2$ and $i+3$, respectively. The colour of the squares reflects the intensity of the NOE compared to the total NOE of the corresponding category (backbone or side chain atoms). Red is the strongest NOE and blue is the lightest. Residues involved in long-range NOEs as well as prolines 13 and 30 are highlighted.

The fold of the five NCBD domains was further investigated by analysing the NOESY maps recorded at 700 and 900 MHz using an iterative structure calculation procedure implemented in ARIA 2.0. This procedure provides an unbiased interpretation of the five NOESY datasets based on prior manual assignments of the proton and carbon resonances, assuming a single conformation. The comparison of the NOE contact maps obtained for the four fluoroproline containing proteins (Figure 2B) shows a consistent set of hydrophobic residues whose side chain interactions contribute to the folding of NCBD. Similar to the unmodified NCBD, all fluorinated proteins display a number of NOE contacts between Phe36 (2101 in CBP) protons and side-chains protons of Leu26 and Val22 in helix α_2 and Leu10 in helix α_1 . NOE contacts are also consistently observed between Val22 and Leu10 side-chains characterizing the proximity between the first and the second helices. However, significant differences are observed in the number of NOE contacts and their intensities, reflecting the differences in the quality of the NOESY maps obtained for the five proteins. Noteworthy, insertion of a fluoroproline at position 30 resulted in sharper line-widths enabling the interpretation of medium range NOEs characterizing the helical secondary structures while the corresponding NOEs are depleted in datasets obtained for NCBD harbouring a fluoroproline at position 13. Several differences are observed in the NOE pattern of residues Leu26 and Phe36 when comparing the NOE contact maps of proteins incorporating 4-(*S*)-FPro or a 4-(*R*)-FPro at position 30, suggesting that the conformational bias of the fluorinated proline has a global impact on the packing of helices α_2 and α_3 .

The resulting sets of structures display different orientations of the three helices depending on the fluorination site and fluoroproline chirality (Figure 3). The largest angular fluctuations are observed between helices $\alpha 1$ and $\alpha 2$. Noteworthy, these fluctuations which are significant both within the same set of models and between the different proteins, are not directly related to the modification site. Thus, changing the chirality of the proline at position 30 leads to a significant change in the orientation of helix $\alpha 1$ relative to $\alpha 2$ while the orientation of helix $\alpha 2$ relative to $\alpha 3$ remains similar. Further, the substitution of Pro13 by a 4-(*S*)-FPro produced an ensemble of conformers with orientations of helices $\alpha 1$ and $\alpha 2$ that are similar to those observed for proteins harbouring a 4-(*R*)-FPro at position 30. It should be stressed that the use of a single conformer during the modelling is biasing the set of representative conformations. This is evidenced by the analysis of resulting dihedral angles that are often clustered to the edges the experimental values boundaries, an effect due to the different averaging properties of NOE and dihedral angle restraints (Figure S6). Nevertheless, the differences observed between the five structural sets suggest that insertions of fluoroprolines in the loops of NCBD modulate the ensemble of conformations sampled by the protein. Due to NOE averaging during the structure calculation procedure, this redistribution of conformers leads to slightly different hydrophobic packing and subsequently distinct orientations of the helices.

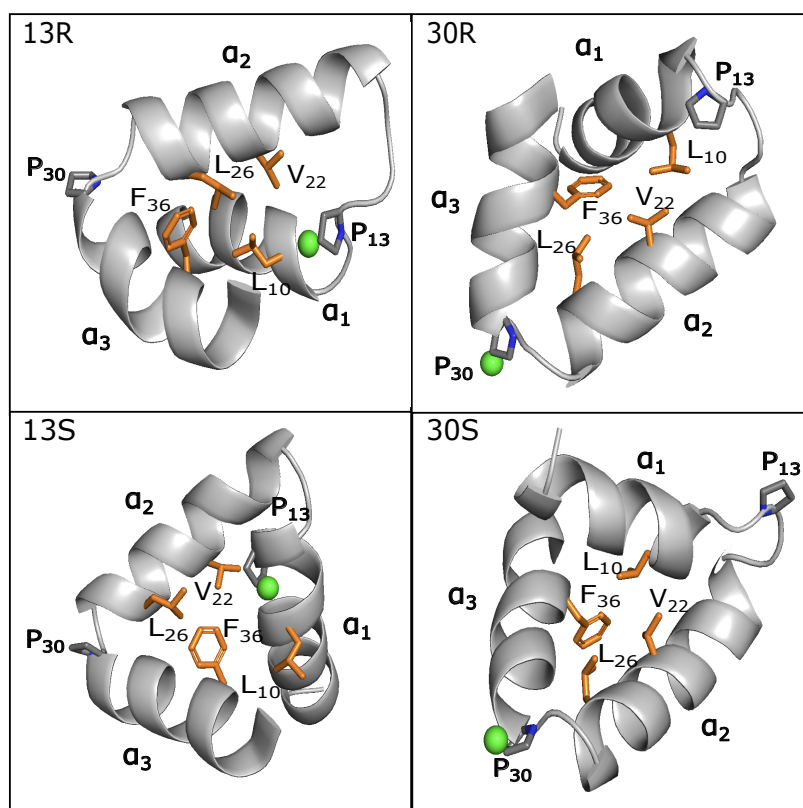
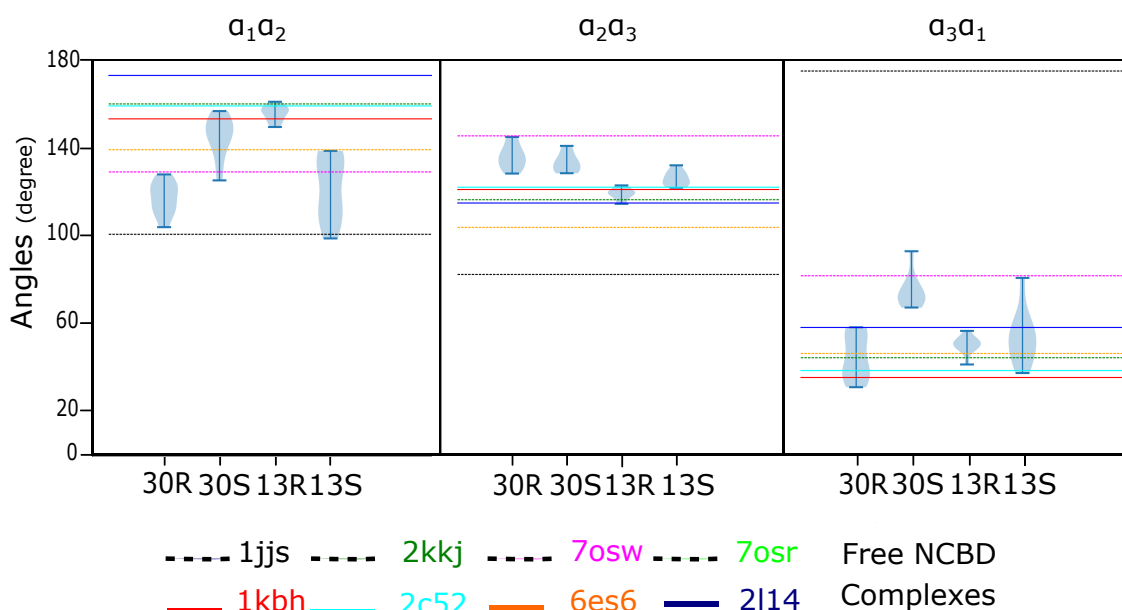


Figure 3: A) Orientation of the helices by respect to each other in the fluorinated NCBDs. The angles were measured for the 10 representative conformations of each NCBd obtained from ARIA structure calculations after refinement. The horizontal lines represent the orientation of the helices in published structures of ligand-free and ligand-bound NCBD. B) 3D model of fluorinated NCBDs showing the rearrangement of the hydrophobic core. (Statistics of the structure calculations are displayed on table S2)

2.4. Fluoroproline insertion impacts global and local dynamics

We previously noticed the sharp increase of NOESY spectral quality upon insertion of a 4-(*S*)-FPro or 4-(*R*)-FPro at position 30 contrasting with the non-modified protein or to the protein harbouring a fluoroproline at position 13. These spectral changes can be due to either a modulation of the transition rates between the accessible conformers or a reduction of the accessible conformational space. Taking advantage of the presence of a fluorine resonance on the four modified NCBD, we investigated the local dynamics at positions 13 and 30 by recording 1D ^{19}F spectra at temperatures ranging from 286 to 318 K. Fitting the resonances to Lorentzian shapes provided a good estimate of the transverse relaxation rate variations upon temperature changes (Figure 4). While fluorine signals display a general sharpening upon temperature increase, striking differences can be observed between the different samples. The signal of the 4-(*R*)-FPro at position 13 displays the sharpest profile, decreasing from 30 to 18 Hz. The temperature dependence of the linewidth follows the expected trend assuming that the transverse relaxation rate is dominated by the rotational diffusion and depends on the temperature and water viscosity changes. Incorporation of a 4-(*S*)-FPro at the same position leads to a marked difference. The major peak displays a broadening up to 288 K before sharpening at higher temperatures indicating the presence of chemical exchange. The interpretation of this behaviour is complex due to the presence of several exchanges at multiple time-scales. Indeed, the line shape could not be reproduced by a single exchange between the minor and major conformers whose exchange rate (0.4 s^{-1} at 300 K) was measured using a saturation transfer experiment (Figure S2). Instead, the fluorine signal is sensitive to a conformational exchange in the fast exchange regime leading to a line sharpening above 290 K. This multiple exchange is further suggested by the analysis of the temperature dependence of the chemical shift that displays a non-linear behaviour (Figure 4). The lack of information on the frequency of the minor conformation responsible for this fast exchange contribution prevents the extraction of the quantitative parameters of the exchange such as the exchange rate of the populations. Nevertheless, it is tempting to postulate that the position of the fluorine atom in the 4-(*S*)-FPro renders its resonance sensitive to the conformational exchange experience by the whole protein while the 4-(*R*) isomer is insensitive. In contrast, both 4-(*S*) and 4-(*R*) isomers of the FPro inserted at position 30 display line broadening at low temperature suggesting that fluorine atom experiences the conformational fluctuations of the NCBD. Again, the lack of information on the minor conformation's chemical shifts as well as the number of conformers prevents further quantitative interpretation. Contrasting with other three fluoroprolines, the

temperature dependence of fluorine chemical shift displays a singular non-linear behaviour with a ceiling value at -176.78 ppm.

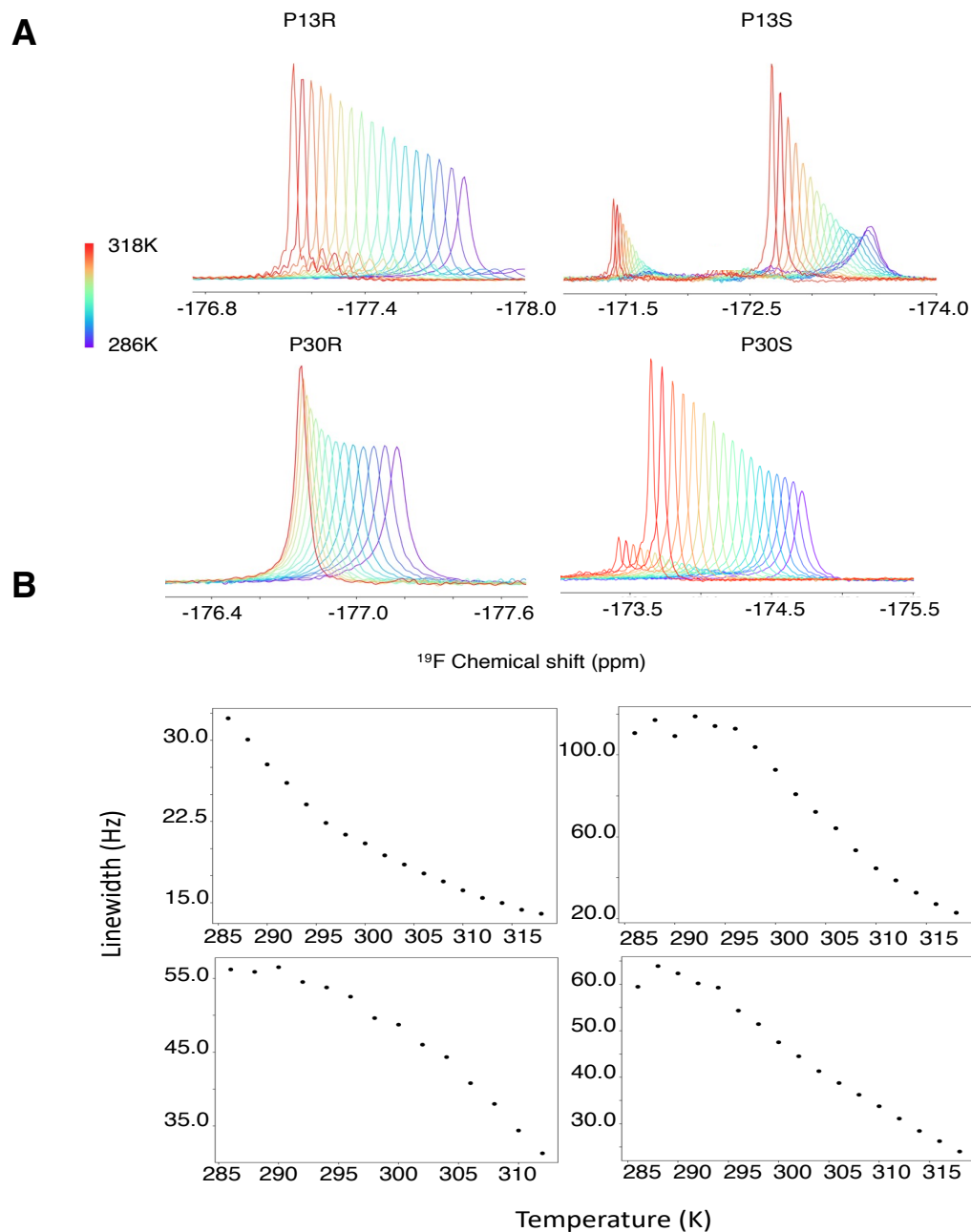


Figure 4: Temperature dependence of $^1\text{D } ^{19}\text{F}$ spectra of fluorinated NCBD recorded at 600 MHz (A), and their corresponding variation of linewidth (B). The linewidths were measured by fitting a Lorentzian model to the peaks using Spike.

2.5. Insertion of 4-(*R*)-FPro in the NCBD α 2- α 3 loop induces self-assembly into amyloid fibers

In the conditions used to study the fluorinated NCBD (1 mM of protein in 20 mM phosphate buffer, pH 7.2 at 304 K), we systematically observed a time evolution of the fluorine signal with a linear intensity decay during 90 min after a lag phase of few days suggesting a self-assembly process (Figure 5A). This phenomenon is very slow (few weeks depending on the concentration and the temperature of the storage) when the sample is not manipulated. The aggregated P30R-NCBD sample was further characterized using Thioflavin-T (ThT), a fluorescent dye sensitive to amyloids. The addition of aggregated P30R-NCBD to ThT induced a rapid fluorescence enhancement accompanied by a shift of both the maximal excitation and emission wavelengths from 340 nm to 448 nm and from 450 nm to 480 nm, respectively, revealing the presence of amyloid structures (Figure 5B). Indeed, transmission electron microscopy (TEM) allowed to visualize filamentous amyloid structures while the FT-IR spectra of the protein precipitate clearly indicated the presence of pure beta sheets while a mixture of alpha helices and beta sheets was observed in the soluble supernatant (Figure 5 C and D). Altogether, these results indicated that altering the conformational and dynamical properties of the short loop linking helices α 2 to α 3 of NCBD changes the conformational landscape of the domain favouring its self-assembly into amyloids.

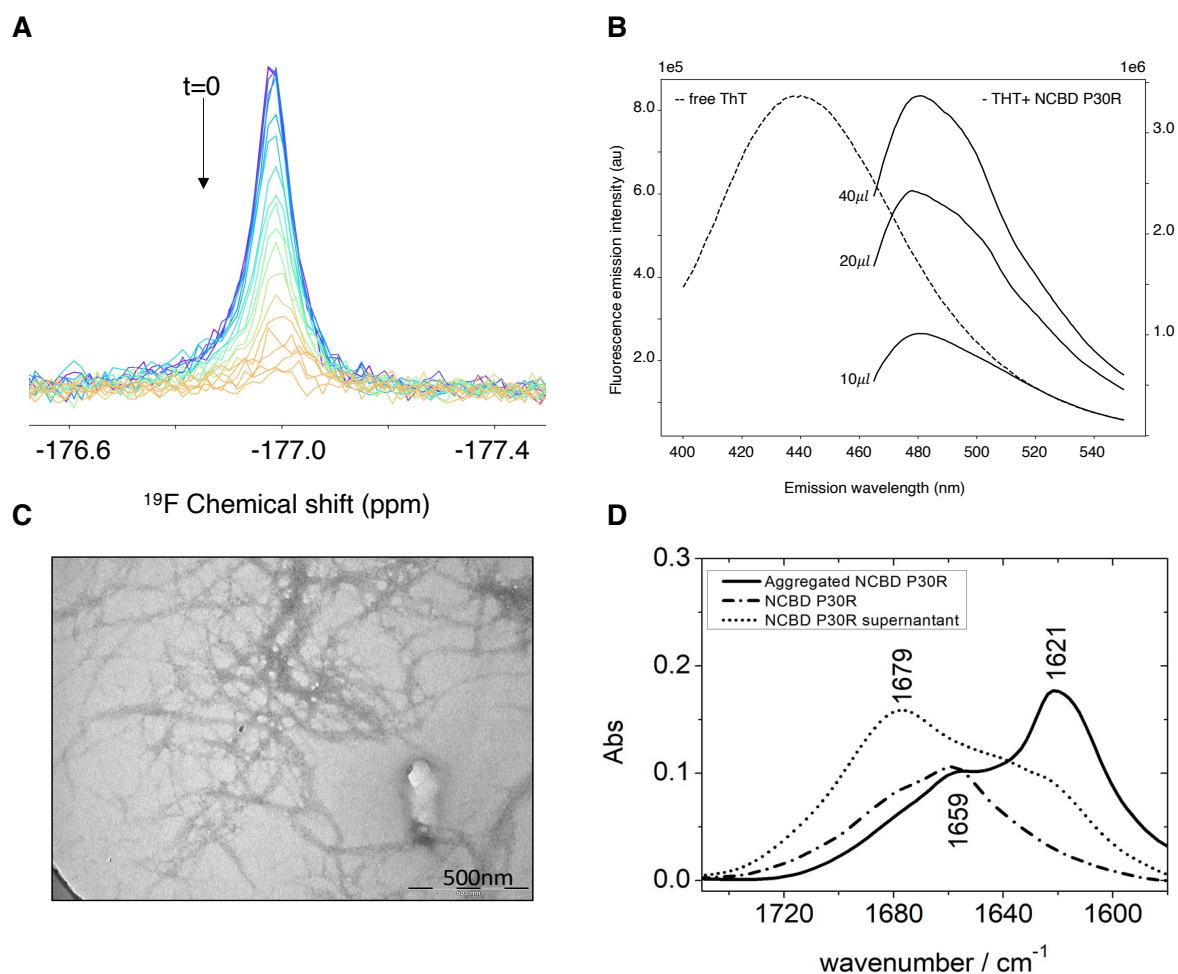


Figure 5: A) Self-assembly of P30R-NCBD during time followed by ^{19}F NMR at 304 K. The sample was dissolved in a phosphate buffer. After water evaporated at 327K the sample was dissolved in D_2O and incubated at 304 K. ^{19}F NMR resonance completely disappears after 90 min. B) Fluorescence emission spectrum of free ThT (18 μM , λ_{ex} 340 nm, λ_{em} 350-550nm) and fluorescence emission spectra of ThT upon addition of 10, 20 and 40 μl of aggregated P30R-NCBD. C) TEM micrographs of P30R-NCBD samples dissolved in phosphate buffer at pH 7.2 after an evaporation step at 327 K and an incubation step at 304 K during the night. D) FT-IR spectra of non-aggregated and aggregated P30R-NCBD. The protein precipitate is composed of pure beta sheets while a mixture of alpha helices and beta sheets is observed in the supernatant.

3. Discussion

The term ‘molten-globule’ was applied to various types of protein forms lacking well-defined tertiary structure. Obtaining high-resolution structural information for any kind of molten globule is essential to understanding their non-cooperative folding and conformational dynamics related to biological function. Previously, NCBD was reported as a molten globule with a partially folded core and known to interact with diverse binding partners resulting in induced folding and distinct structures. Conformational selection mechanism was proposed to play role in such interactions.

Incorporation of stereoisomeric 4-fluoroprolines enabled to shed light on distinct conformations that can be populated in NCBD molten-globule. Effectively, 4-(*R*)- and 4-(*S*)-fluoroprolines possessing distinct conformational preferences could bias the conformational landscape and overpopulate different subsets of conformations in the conformational ensemble.

The hydrodynamic radius of NCBD WT [2065-2111] of CBP, deduced from 2D DOSY is 14.6 Å that correspond to a gyration radius of 18.9 Å. This agrees well with the gyration radius of NCBD extracted from SAXS experiments (18.8 Å) performed by Poulsen’s group under native like conditions.²² Substituting proline 13 or proline 30 with 4-(*S*)-FPro does not alter the compactness of the molten globule. On the other side, the incorporation of 4-(*R*)-FPro induces a 5% expansion of the structure. Despite these differences, all fluorinated NCBD variants have large percentage of helical secondary structure similar to unmodified NCBD and are characterized by a non-cooperative unfolding transition, as previously reported.²²

The fluorination of the proline of the first loop reveals the coexistence of two unequally populated NCBD conformations exchanging in the ms time scale. Whereas, the perturbation of the proline of the second loop reveals a single conformation of NCBD. The published structure of free NCBD (2KKJ) is characterized by the proximity of helix $\alpha 1$ to $\alpha 2$ and $\alpha 2$ to $\alpha 3$. This conformer is considered as the major conformer of NCBD, the most stable and the one that resembles its adopted structure when in complex with AD-ACTR. NMR relaxation dispersion revealed an exchange between this conformer and a minor structure (8%, $k_{ex} = 3600 \text{ s}^{-1}$) characterized by a rearrangement of the three helices. The contact between $\alpha 2$ and $\alpha 3$ would be lost and $\alpha 1$ will get in the proximity of $\alpha 3$.²⁹ In the NOESY spectra of fluorinated NCBDs recorded at 304 K, we observed long range NOEs indicating contacts between helices $\alpha 1$ - $\alpha 2$ (L₁₀-V₂₂, T₉-V₂₂) characteristic of the major conformer, as well as interactions between helices $\alpha 1$ - $\alpha 3$ (L₁₀-F₃₆, T₉-F₃₆) characteristic of the minor conformer, and contacts between helices $\alpha 2$ - $\alpha 3$ (V₂₂-F₃₆, L₂₆-F₃₆). These observations suggest that fluorinated NCBD are in exchange with

one common minor conformation. The characteristic NOEs of the minor conformers are more intense and more numerous within P30R and P30S variants indicating that either their lifetime is longer or that it is more populated when proline 30 is modified.

The minor populations observed on 1D ^{19}F spectra of P13R and P13S NCBD variants are different from the minor conformers where the helices are rearranged. The two populations of P13S are separated by $\Delta\nu=900$ Hz. Thus, the chemical exchange constant rate previously extracted from the relaxation dispersion studies ($k_{ex} = 3600 \text{ s}^{-1}$) would result in a one resonance on the 1D ^{19}F spectrum of P13S and P13R. The exchange constant rate between the two P13S populations was determined by saturating the minor population resonance and reporting intensity variation of the major population at different temperatures. We report $k_{ex} = 0.05 \text{ s}^{-1}$ at 288 K and 0.4 s^{-1} at 300 K. This conformer might refer to a new helical rearrangement of P13S or to a *cis-trans* isomerization of the proline.

The obtained structure of NCBD WT is different from what Poulsen et al (2KKJ) found and resembles more the structure suggested by Fraenkel et al. (1JJS) characterized by parallel $\alpha 1$ and $\alpha 3$ helices. The sequence of NCBD (residues 2058-2116) in 2KKJ structure contains two disordered regions delimiting the three helices, while 1JJS structure has 4 different amino acids preceding alanine (2066) and stops at 2111 thus cutting short the disordered region succeeding $\alpha 3$. The NCBD that we synthesized correspond to (2065-2111) thus only containing the fully formed helices and a short-disordered region. This observation indicates that it is the 5 amino acids truncated from the disordered region that leads to one conformer or the other. This finding is important and highlights the effect of the tail of NCBD on the interactions between the helices. Similar result comes from the recent simulations discussing the pivotal role of NCBD's tail on the helices length and stability.³⁰ However, fluorinated NCBDs sample different conformations characterized by specific arrangement of the three helices that is very different from the U shape of the synthesized NCBD WT and rather resembles 2KKJ and the NCBD bounded structures to different extents. This means that ligand free NCBD samples ligand bound conformations. This result is in agreement with MD simulations.²⁴

NCBD P13R variant is the closest to the published structure of free NCBD 2KKJ and the structure of NCBD in complex with AD-ACTR (1KBH). On the other hand, NCBD P30S adopts a different structure. ITC titration data (see Chapter 7) reveal high affinity of NCBD-P13R to AD-ACTR and a lower affinity of NCBD P30S to AD-ACTR. This is in favor of a conformational selection mechanism explaining the different structures sampled by NCBD.

Previous study reported that NCBD's affinity to AD-ACTR is modulated by *cis-trans* isomerization of Pro13.²⁵ Here we expand this understanding and state that dynamics of both

prolines induce NCBD to rearrange itself to a particular structure, consequently affecting its binding affinities. This finding suggest that one can predict the affinity of NCBD towards the binder by comparing its structure to that of the complex. This can be valuable for screening approaches. Considering all the conformations that NCBD can sample implies the importance of the plasticity in NCBD's function.

The representative conformations of NCBD P30R variant obtained from ARIA software, indicate a particular orientation of the first helix that is rotated by 180° compared to its direction in the other fluorinated NCBDs to fit between $\alpha 2$ and $\alpha 3$. This predisposition is not simply adopted because of the available space between the helices, otherwise it would also be possible for $\alpha 1$ of P30S to acquire the same orientation. The particular conformation of NCBD P30R variant can explain its unique dynamics and especially its tendency to form amyloids. Interestingly, the orientation of the first helix is affected when the proline in the third helix is perturbed. This means that the effect of a local modification propagates in NCBD. Garcia et al, reported the aggregation of NCBD into amyloids but this required lowering the pH.³¹ They suggested that the early step of aggregation is due to the particular predisposition of the first helix that exposes the hydrophobic side chain to the solvent favoring its interaction with $\alpha 1$ of other NCBD molecules and leading to the formation of alpha helical oligomers – the initial step of aggregation followed by conversion to beta-sheet containing amyloid fibers. Our data experimentally supports this hypothesis. Furthermore, biasing the conformation of the Pro30 by incorporating 4-(*R*)-FPro induces the aggregation of NCBD at neutral pH where kinetics is faster at higher concentrations.

4. Conclusions

To conclude, fluorination of Proline 13 and 30 can dictate global NCBD structure, dynamics and binding affinities. Subtle stereo electronic effects of stereoisomeric 4-fluoroproline propagate depending on the site of substitution and sequence context. Using fluorinated prolines, we were able to expand the knowledge of NCBD's conformational landscape by reporting newly identified structures. All NCBD conformations undergo an exchange with a minor conformer previously reported and characterized by new helical contacts. When Pro13 is modified we identified two minor conformers that are accessible. We also showed that 5 amino acids in the disordered C-terminal of NCBD can influence its structure.

The aggregation of NCBD into amyloids can be enhanced by a particular predisposition of helix $\alpha 1$ when P30 is substituted by 4-(*R*)-FPro. The overall data also indicate that the multiple

structures in NCBD are encoded in NCBD's sequence and not a random process. This property is eventually enabling the conformational selection mechanism of NCBD in binding to its targets.

Experimental section

Protein synthesis

All synthesized NCBD variants correspond to the fragment [2065-2111] of the CREB binding protein. The Methionine of the native NCBD sequence was replaced by a nearly isosteric Norleucine to avoid oxidation of thioether moiety.

4-(*R*)- and 4-(*S*)-FPro amino acids were purchased from Bachem SA and Fmoc-protected in house using standard protocol.

The Synthesis was performed on a Liberty blue microwave assisted automatic synthesizer (CEM Corporation, USA). A preloaded Rink Amide resin (mesh 100-200, 0.14 mmol/g loading) was used. Cleavage and deprotection were performed using TFA/TIPS/H₂O (v/v/v) (19/0.5/0.5) mixture for 2.5 hours. Cold diethyl ether (-20 °C) was then added to induce precipitation. Afterwards, the mixture was centrifuged and the precipitate was dissolved in water/acetonitrile solution (1:1, v/v) containing 0.1% TFA to be lyophilized. The mass of the peptides was confirmed by LC-MS prior to purification by preparative HPLC on a C18 column using a gradient from 20-60 H₂O/MeCN, 0.1% TFA in 40 minutes. Purity of the samples was determined by analytical HPLC.

Evaporation: The samples were dissolved in a phosphate buffer, then frozen in liquid nitrogen. The water was evaporated at 327 K using a Savant SC110A Plus Speedvac . Once the solvent evaporated the peptides were dissolved in D₂O and incubated at 304 K.

NMR

All molten globules were dissolved in a freshly prepared phosphate buffer (20 mM) at a pH 7.2 in water. For NMR measurement 1 mM samples were used in 3 mm tubes.

¹⁹F NMR spectra were acquired at 565 MHz on a Bruker I spectrometer equipped with a cryogenic QCI-F probe. Proton and carbon spectra were recorded at 700 MHz on a Bruker Avance III spectrometer equipped with a cryogenic TCI probe. Unless mentioned otherwise, 2D spectra for the peptide assignment were recorded at 304 K.

Full ^1H , ^{13}C , and ^{15}N assignment of wild type and fluorinated variants of NCBD was performed by standard HSQC, TOCSY, NOESY and HSQC-TOCSY experiments. ^{13}C carbon assignment was obtained from standard full-range ^1H - ^{13}C HSQC (11 ppm ^1H x 70 ppm ^{13}C). The samples were in phosphate buffer (20 mM) at a pH 7.2 dissolved in D_2O . 4096 points were accumulated in the time domain F2 and 800 in F1. 32 scans and 4 demi scans were accumulated. The resulting resolutions are 1.6 Hz/pt and 13.2 Hz/pt in F2 and F1 respectively. The total experiment time was 12h14min. O1P was at 4.712 ppm and O2P at 40 ppm. At 900 MHz (10 ppm ^1H x 100 ppm ^{13}C). 2048 and 1024 points. 8 scans and 32 demi-scans. FID resolution 8.87Hz and 44.2Hz. on H_2O +1% D_2O samples. O2P at 50 ppm and O1P at 4.7 ppm.

^1H - ^1H NOESY spectra (11ppm ^1H x 11 ppm ^1H) were recorded at either 600 or 700 MHz at two mixing times 100 ms and 400 ms. The samples were dissolved in phosphate buffer (20 mM) at pH 7.2 in H_2O . The relaxation time was 2.5 s. 2048 points and 512 points were recorded in F2 and F1 domains respectively. The FID resolution is 3.25 Hz and 12.8 Hz. 16 scans and 8 demi-scans. The total experiment time is 7h and 6h20 for the long and the short NOESY respectively. Spectra recorded at 900 MHz (10 ppm ^1H x 10 ppm ^1H) were recorded on Bruker Avance NEO spectrometer. 4096 and 800 points. 16 scans and 16 demi-scans. FID resolution 4.4 Hz and 22.7 Hz. D1 was 1.5 s.

^1H - ^1H TOCSY spectra (11ppm ^1H x 11 ppm ^1H) were acquired for samples dissolved in phosphate buffer (20 mM) at pH 7.2 in H_2O at 600 or 700 MHz. Two mixing times were recorded 80 ms and 25 ms. The relaxation time was 2.5s. 4096 points and 800 points were recorded in F2 and F1 domains respectively. The FID resolution is 1.6Hz and 8.25 Hz. 16 scans and 8 demi-scans. The total experiment time is 10 h 26 min and 7 h 39 min for the long and the short TOCSY respectively. At 900 MHz (11ppm ^1H x 11 ppm ^1H) only TOCSY at 80 ms was recorded.: 4096 and 1024 points. 8 scans and 16 demi-scans. FID resolution 4.4 Hz and 17.75 Hz. D1 was set to 1s. HSQC-TOCSY (11 ppm ^1H x 70 ppm ^{13}C) at short mixing time (25 ms) were recorded on samples dissolved in phosphate buffer (20 mM) at pH 7.2 in D_2O . 4096 and 400 points were acquired in F2 and F1 domains respectively. A total of 256 scans and 8 demi-scans were accumulated. The FID resolution was 3.5Hz and 52.8 Hz in both domains respectively. the relaxation delay D1 was set to 1.35 s.

^{15}N resonances were assigned using 2D ^1H - ^{15}N HETSOFAST experiments recorded at 700 and 900 MHz. Spectrum for NCBD P30R was not recorded since it aggregated. NCBD P13R and P13S assignment was performed on spectra recorded at 700 MHz. Their spectra at 900 MHz showed less peaks. NCBD WT and NCBD 30S were both recorded at 700 MHz and 900 MHz. At 900 MHz the spectral width was (10 ppm x 35 ppm). 1024 and 128 points were

acquired in F2 and F1 domains respectively. 128 scans and 16 demi-scans were accumulated. The FID resolution was 17.7 Hz and 49.89 Hz in both domains respectively. O₁P and O₂P were set to 4.7 ppm and 117 ppm. The relaxation delay D1 was 0.1 s. At 700 MHz the spectral width was (14 ppm x 37 ppm). 1024 and 128 points recorded on F2 and F1 domains respectively. 2048 scans and 512 demi-scans were accumulated. The FID resolution was 19.14 Hz and 41 Hz. O₁P and O₂P were set to at 4.68 ppm and 117.5 ppm respectively. The relaxation delay D1 was 0.1s. The resulting experiment time is 17 h 33 min.

200 uL samples at 1mM concentration in 20 mM phosphate buffer at pH 7.2 dissolved in water were subject to a temperature rise from 276 K to 318 K with a 2-degree ramp. 1D ¹⁹F NMR spectra were recorded with 1024 scans, 8 DS and 2048 TD. D1 was set to 1.5 s. The spectral width is 30 ppm and the resolution is 8.2 Hz. The total experiment time was 26 min per spectrum.

All 1D spectra were processed using spike software. An exponential line broadening of 8 Hz was applied prior to Fourier transform for signal apodization. The peaks were fitted to a Lorentzian to extract their linewidth, chemical shift and intensity. 2D spectra were processed using Topspin 2.6 software. The assignment was performed using CcpNmr Analysis V2.

DOSY experiments were performed at 700 MHz at 14 different temperatures starting from 283 K up to 309 K with a 2-degree temperature ramp. For each experiment the FID size was 4096 points and 40 points in F2 and F1 dimensions, respectively. 32 scans were recorded with 16 demi-scans. The spectral width was 14 ppm and 10 ppm in F2 and F1 respectively with corresponding resolutions of 4.78 Hz and 361.3 Hz. D1 was set to 2.5 s. The DOSY spectra were processed using PALMA.³² To extract the diffusion constant, we considered a linear variation of the water viscosity and computed the viscosity at each measure temperature (T). The chemical shift and the diffusion constant (D) variation of one particularly isolated methyl peak were followed at each temperature. We then proceeded to plot $D=f(T/\text{viscosity})$ and extracted the hydrodynamic radius based on the Stock-Einstein formula: $R_h = \frac{kT}{6\pi\eta D}$, where k is the Boltzmann constant, T is the temperature and η is the viscosity of the sample.

Structure calculations

Chemical shifts of random coil non-modified NCBD sequence were generated from POTENCI random coil chemical-shift predictor from NCBD sequence. Secondary chemical shifts represent the measured chemical shifts of each NCBD - the random coil chemical shift.

Non-canonical amino acids are not recognized in POTENCI, thus norleucine was substituted by a methionine in the simulation.

Input data is composed of peak lists and measured chemical shifts tables issued from CCPNMR3.2. We applied constraints on hydrogen bonds and dihedral angles generated on CCPNMR 3.2. No constraints were applied on the loops.

The data is from ARIA file containing non ambiguous NOE between the residues and their corresponding volume. The NOE are classified in two groups depending on the interacting atoms being in the backbone or on the side chains. The backbone NOEs are represented in the lower part of the contact map while the sidechain NOE occupy the upper part. The color represents the intensity of the NOE compared to the total NOE of the corresponding category with the blue being the lowest and red being the strongest. This analysis was performed on a Jupiter notebook using Pandas library. Structure analysis was performed using the 10 representative conformations of each NCBD obtained from ARIA structure calculations after refinement. The orientation of the helices is measured from the pdb files on Pymol software using `angle_between_helices` widget. This latter can be added using an online available python module: `anglebetweenhelices.py`. Data of ligand-free and ligand-bound NCBD was generated using the same approach from published pdb files. The data analysis was performed on a Jupiter notebook using Pandas library.

Circular Dichroism

Far-UV CD experiments were acquired using a JASCO J-815 spectropolarimeter. All peptides were dissolved in 20 mM sodium phosphate buffer at pH 7.2 containing 10% of D₂O. The concentration of the samples was 250 μ M. The working temperature was set to 283 K and the observed wavelengths ranged from 185 nm to 270 nm. A spectrum was recorded every 1 nm after 6 scans. Response time was 1 s with a bandwidth of 1 nm. The continuous scan rate was 100 nm/min. Each spectrum was systematically corrected by subtracting the solvent spectrum recorded with the same experimental conditions.

Same samples were used for the thermal denaturation experiments using 1 mm cuvette. The concentration of the peptides was 80 μ M dissolved in 20 mM sodium phosphate buffer at pH 7.2 and containing 10% of D₂O. The temperature ranged from 10 to 90 °C with a temperature ramp fixed to 2 °C/min. Each data point resulted from a 4 s averaging integration. The working wavelength was 222 nm.

TEM analysis

Aliquots (15 μ L) of the aggregated NCBP P30R were placed on carbon-coated 300 mesh copper EM grids previously exposed to the UV-ozone cleaner for 5 minutes. TEM micrographs were acquired on a Morgagni 268D transmission electron microscope: equipped with Mega view III digital camera at room temperature.

ThT fluorescence

ThT solution was freshly prepared and filtered (200 mL at 18 mM in 20 mM sodium phosphate buffer at pH 7.2). The ThT concentration was determined by optical density using a 36 $\text{mM}^{-1} \text{cm}^{-1}$ extinction coefficient at 412 nm. The fluorescence intensity was recorded on a spectro-fluorometer PTI equipped with a lamp power supply, motor driver (MD-5020) and operating at room temperature. Analysis of the spectra was performed on Felix 32 software. The excitation and emission wavelengths of ThT alone are 340 nm/ 450nm. Different volumes of NCBP P30R were added to the ThT solution inducing shift of the emission and excitation wavelengths to 448 nm / 480 nm, respectively.

FT-IR spectroscopy

Experiments were recorded on Vertex 70 FTIR spectrometer (Bruker Optics, Karlsruhe, Germany). Residual TFA was removed from the samples by three cycles of lyophilization from 10 mM HCl. A drop of 2.5 μ L on the diamond surface was left to dry in order to record the spectra. The number of scans per spectrum was fixed to 64 ranging from 4000 cm^{-1} to 700 cm^{-1} with a 4 cm^{-1} resolution. 10 spectra were averaged and processed by OPUS 7.2 to correct the baseline.

Bibliography

- (1) Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S. Yu.; Ptitsyn, O. B. α -Lactalbumin: Compact State with Fluctuating Tertiary Structure? *FEBS Lett.* **1981**, *136* (2), 311–315. [https://doi.org/10.1016/0014-5793\(81\)80642-4](https://doi.org/10.1016/0014-5793(81)80642-4).
- (2) Ohgushi, M.; Wada, A. ‘Molten-Globule State’: A Compact Form of Globular Proteins with Mobile Side-Chains. *FEBS Lett.* **1983**, *164* (1), 21–24. [https://doi.org/10.1016/0014-5793\(83\)80010-6](https://doi.org/10.1016/0014-5793(83)80010-6).
- (3) Bhattacharyya, S.; Varadarajan, R. Packing in Molten Globules and Native States. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 11–21. <https://doi.org/10.1016/j.sbi.2012.10.010>.
- (4) Baldwin, R. L.; Rose, G. D. Molten Globules, Entropy-Driven Conformational Change and Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 4–10. <https://doi.org/10.1016/j.sbi.2012.11.004>.

- (5) Ptitsyn, O. B. Structures of Folding Intermediates. *Curr. Opin. Struct. Biol.* **1995**, 5 (1), 74–78. [https://doi.org/10.1016/0959-440X\(95\)80011-O](https://doi.org/10.1016/0959-440X(95)80011-O).
- (6) Kuwajima, K. The Molten Globule State as a Clue for Understanding the Folding and Cooperativity of Globular-Protein Structure. *Proteins Struct. Funct. Genet.* **1989**, 6 (2), 87–103. <https://doi.org/10.1002/prot.340060202>.
- (7) Ptitsyn, O. B.; Uversky, V. N. The Molten Globule Is a Third Thermodynamical State of Protein Molecules. *FEBS Lett.* **1994**, 341 (1), 15–18. [https://doi.org/10.1016/0014-5793\(94\)80231-9](https://doi.org/10.1016/0014-5793(94)80231-9).
- (8) Pillai, M.; Jha, S. K. The Folding and Aggregation Energy Landscapes of Tethered RRM Domains of Human TDP-43 Are Coupled via a Metastable Molten Globule-like Oligomer. *Biochemistry* **2019**, 58 (6), 608–620. <https://doi.org/10.1021/acs.biochem.8b01013>.
- (9) Pedrote, M. M.; De Oliveira, G. A. P.; Felix, A. L.; Mota, M. F.; Marques, M. D. A.; Soares, I. N.; Iqbal, A.; Norberto, D. R.; Gomes, A. M. O.; Gratton, E.; Cino, E. A.; Silva, J. L. Aggregation-Primed Molten Globule Conformers of the P53 Core Domain Provide Potential Tools for Studying P53C Aggregation in Cancer. *J. Biol. Chem.* **2018**, 293 (29), 11374–11387. <https://doi.org/10.1074/jbc.RA118.003285>.
- (10) Iram, A.; Alam, T.; Khan, J. M.; Khan, T. A.; Khan, R. H.; Naeem, A. Molten Globule of Hemoglobin Proceeds into Aggregates and Advanced Glycated End Products. *PLoS ONE* **2013**, 8 (8), e72075. <https://doi.org/10.1371/journal.pone.0072075>.
- (11) Martin, J.; Langer, T.; Boteva, R.; Schramel, A.; Horwich, A. L.; Hartl, F.-U. Chaperonin-Mediated Protein Folding at the Surface of GroEL through a 'Molten Globule'-like Intermediate. *Nature* **1991**, 352 (6330), 36–42. <https://doi.org/10.1038/352036a0>.
- (12) Rajaraman, K.; Raman, B.; Rao, Ch. M. Molten-Globule State of Carbonic Anhydrase Binds to the Chaperone-like α -Crystallin. *J. Biol. Chem.* **1996**, 271 (44), 27595–27600. <https://doi.org/10.1074/jbc.271.44.27595>.
- (13) Rösner, H. I.; Redfield, C. The Human α -Lactalbumin Molten Globule: Comparison of Structural Preferences at PH 2 and PH 7. *J. Mol. Biol.* **2009**, 394 (2), 351–362. <https://doi.org/10.1016/j.jmb.2009.09.025>.
- (14) Colón, W.; Roder, H. Kinetic Intermediates in the Formation of the Cytochrome c Molten Globule. *Nat. Struct. Biol.* **1996**, 3 (12), 1019–1025. <https://doi.org/10.1038/nsb1296-1019>.
- (15) Raschke, T. M.; Marqusee, S. The Kinetic Folding Intermediate of Ribonuclease H Resembles the Acid Molten Globule and Partially Unfolded Molecules Detected under Native Conditions. *Nat. Struct. Biol.* **1997**, 4 (4), 298–304. <https://doi.org/10.1038/nsb0497-298>.
- (16) Hoeltzli, S. D.; Frieden, C. Stopped-Flow NMR Spectroscopy: Real-Time Unfolding Studies of 6-19F-Tryptophan-Labeled Escherichia Coli Dihydrofolate Reductase. *Proc. Natl. Acad. Sci.* **1995**, 92 (20), 9318–9322. <https://doi.org/10.1073/pnas.92.20.9318>.
- (17) Jha, S. K.; Udgaonkar, J. B. Direct Evidence for a Dry Molten Globule Intermediate during the Unfolding of a Small Protein. *Proc. Natl. Acad. Sci.* **2009**, 106 (30), 12289–12294. <https://doi.org/10.1073/pnas.0905744106>.
- (18) Meinhold, D. W.; Wright, P. E. Measurement of Protein Unfolding/Refolding Kinetics and Structural Characterization of Hidden Intermediates by NMR Relaxation Dispersion. *Proc. Natl. Acad. Sci.* **2011**, 108 (22), 9078–9083. <https://doi.org/10.1073/pnas.1105682108>.
- (19) Uversky, V. N. Protein Intrinsic Disorder and Structure-Function Continuum. In *Progress in Molecular Biology and Translational Science*; Elsevier, 2019; Vol. 166, pp 1–17. <https://doi.org/10.1016/bs.pmbts.2019.05.003>.

- (20) Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry Molten Globule Intermediates and the Mechanism of Protein Unfolding: Mechanism of Protein Unfolding. *Proteins Struct. Funct. Bioinforma.* **2010**, 78 (13), 2725–2737. <https://doi.org/10.1002/prot.22803>.
- (21) Demarest, S. J. Packing, Specificity, and Mutability at the Binding Interface between the P160 Coactivator and CREB-Binding Protein. *Protein Sci.* **2004**, 13 (1), 203–210. <https://doi.org/10.1110/ps.03366504>.
- (22) Kjaergaard, M.; Teilum, K.; Poulsen, F. M. Conformational Selection in the Molten Globule State of the Nuclear Coactivator Binding Domain of CBP. *Proc. Natl. Acad. Sci.* **2010**, 107 (28), 12535–12540. <https://doi.org/10.1073/pnas.1001693107>.
- (23) Kjaergaard, M.; Poulsen, F. M.; Teilum, K. Is a Malleable Protein Necessarily Highly Dynamic? The Hydrophobic Core of the Nuclear Coactivator Binding Domain Is Well Ordered. *Biophys. J.* **2012**, 102 (7), 1627–1635. <https://doi.org/10.1016/j.bpj.2012.02.014>.
- (24) Burger, V. M.; Ramanathan, A.; Savol, A. J.; Stanley, C. B.; Agarwal, P. K.; Chennubhotla, C. S. Quasi-Anharmonic Analysis Reveals Intermediate States in the Nuclear Co-Activator Receptor Binding Domain Ensemble. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **2012**, 70–81.
- (25) Zosel, F.; Mercadante, D.; Nettels, D.; Schuler, B. A Proline Switch Explains Kinetic Heterogeneity in a Coupled Folding and Binding Reaction. *Nat. Commun.* **2018**, 9 (1), 3332. <https://doi.org/10.1038/s41467-018-05725-0>.
- (26) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc.* **2001**, 123 (4), 777–778. <https://doi.org/10.1021/ja005542v>.
- (27) Renner, C.; Alefelder, S.; Bae, J. H.; Budisa, N.; Huber, R.; Moroder, L. Fluoroprolines as Tools for Protein Design and Engineering. *Angew. Chem. Int. Ed.* **2001**, 40 (5), 923–925. [https://doi.org/10.1002/1521-3773\(20010302\)40:5<923::AID-ANIE923>3.0.CO;2-#](https://doi.org/10.1002/1521-3773(20010302)40:5<923::AID-ANIE923>3.0.CO;2-#).
- (28) Kok, C. M.; Rudin, A. Relationship between the Hydrodynamic Radius and the Radius of Gyration of a Polymer in Solution. *Makromol. Chem. Rapid Commun.* **1981**, 2 (11), 655–659. <https://doi.org/10.1002/marc.1981.030021102>.
- (29) Kjaergaard, M.; Andersen, L.; Nielsen, L. D.; Teilum, K. A Folded Excited State of Ligand-Free Nuclear Coactivator Binding Domain (NCBD) Underlies Plasticity in Ligand Recognition. *Biochemistry* **2013**, 52 (10), 1686–1693. <https://doi.org/10.1021/bi4001062>.
- (30) Luong, T. D. N.; Nagpal, S.; Sadqi, M.; Muñoz, V. A Modular Approach to Map out the Conformational Landscapes of Unbound Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci.* **2022**, 119 (23), e2113572119. <https://doi.org/10.1073/pnas.2113572119>.
- (31) Garcia, A. M.; Giorgiutti, C.; El Khoury, Y.; Bauer, V.; Spiegelhalter, C.; Leize-Wagner, E.; Hellwig, P.; Potier, N.; Torbeev, V. Aggregation and Amyloidogenicity of the Nuclear Coactivator Binding Domain of CREB-Binding Protein. *Chem. – Eur. J.* **2020**, 26 (44), 9889–9899. <https://doi.org/10.1002/chem.202001847>.
- (32) Afef Cherni, Emilie Chouzenoux, M-A Delsuc. PALMA, an improved algorithm for DOSY signal processing. [Research Report] LIGM - Laboratoire d’Informatique Gaspard-Monge. 2016. hal-01421737

Supplementary materials

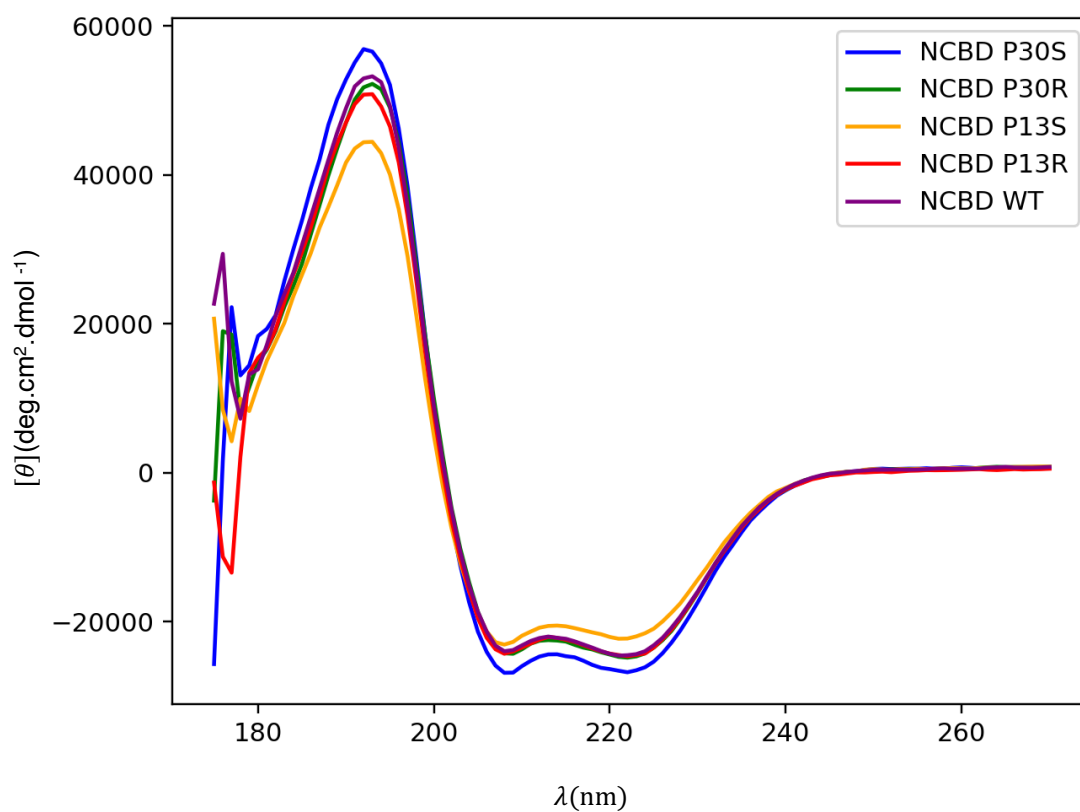


Figure S1: Far-UV CD spectra of fluorinated and wild type NCB D recorded at 283 K. The CD spectra were recorded using 10 mm quartz cuvettes. Protein samples were dissolved in Phosphate buffer 20 mM at concentrations of 250 μM .

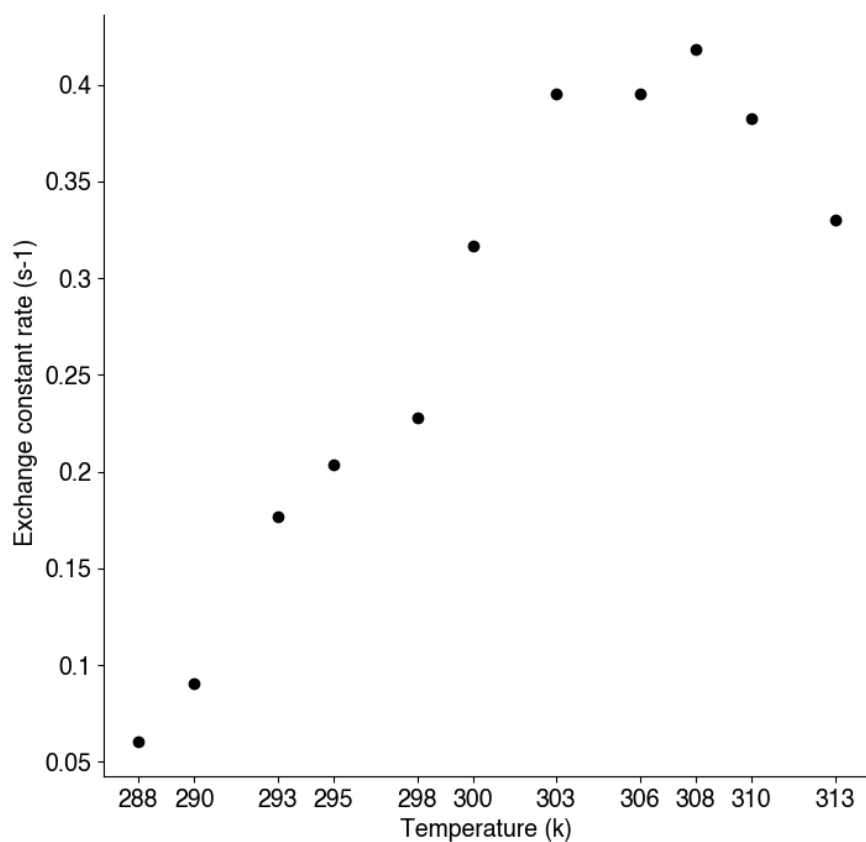


Figure S2: variation of the exchange constant rate between the two populated state in P13S with the temperature. The minor population of P13S was continuously saturated and the intensity changes of the major state was followed to extract the exchange rate constant. The intensity change of the major state follows the exponential model $MA(t) = MA0[1 + \frac{k'}{R1}(\exp(-R1t) - 1)]$. Where k' is the apparent exchange rate $k' = \frac{MB0}{MA0} * k_{AB}$. And $MB0$ and $MA0$ are the intensities of the major and the minor states respectively at $t=0$.

Table S1A: Assignments ^1H ^{13}C and ^{15}N resonances of a NCBD peptide harbouring a *4R* fluoroproline at position P13 at 304K, 20 mM phosphate buffer pH 7.2.

		H	N	H α	H β	H γ	H δ	H ϵ	C	C α	C β	C γ	C δ	C ϵ
1	Ser			3.75	3.62 3.54					74.78	65.22			
2	Ala	8.13		4.26	1.38					54.77	18.67			
3	Leu			4.00			0.89 0.85			58.15			25.44 23.18	
4	Gln	7.65	117.01	3.91	2.10	2.39				59.09	28.31	33.85		
5	Asp	8.13	120.49	4.36	2.82 2.62					57.31	39.99			
6	Leu	8.45	123.44	3.95	1.86 1.86	1.59	0.82 0.72			58.41	41.44	27.13	26.31 24.91	
7	Leu	8.43	118.71	3.93	1.52	1.86	0.80			58.25	41.09	26.87	23.78	
8	Arg	8.05	118.16	3.98	1.93	1.74 1.58	3.22			59.37	30.13	27.79	43.27	
9	Thr	7.89	116.43	3.98	4.23	1.22				66.45	68.56	22.84		
10	Leu	8.27	121.67	3.98	1.63 1.53	1.84	0.82 0.82			56.96	41.44	26.73	26.13 23.60	
11	Lys	7.41	116.95	4.27	1.95 1.81	1.58 1.47	1.64	2.92		56.58	32.78	25.22	29.43	41.83
12	Ser	7.48		4.67	3.98						62.84			
13	(4R)FPro			4.53	2.65 2.07	5.51	4.20 3.93			61.78				
14	Ser			4.34	3.87 3.81					58.94	63.38			
15	Ser	7.83	119.52	4.87	4.21 4.05					56.92	63.32			
16	Pro			4.35	2.43 2.18	2.02 1.92	3.89			65.59	31.84		50.59	
17	Gln	7.95		4.08	2.09 1.99	2.48 2.38				59.26	27.90	34.19		
18	Gln	7.98	123.27	4.01	2.10	2.44				59.20	27.97			
19	Gln	7.95		4.10						59.06				
20	Gln	7.95		3.93						58.46				
22	Val	8.04	119.45	3.34	2.17	0.95 0.65				67.28	31.44	23.47 21.43		
23	Leu	7.95	119.37	3.78	1.80 1.73	1.51	0.83 0.78			58.33	41.07	27.01	25.26 24.50	
24	Asn	8.28	117.44	4.39	2.91 2.83					55.97	37.92			
25	Ile	8.03	122.22	3.71	1.86 1.03 0.82	1.86 1.03 0.82	0.73			65.36	38.24	29.01 17.98	13.92	
26	Leu	8.11	119.22	3.88	1.82 1.45	1.21	0.54 0.48			57.70	41.04		25.70 22.16	
27	Lys	8.33		3.91	1.86	1.46	1.63	2.93		58.87	32.68	25.10	29.43	41.82
28	Ser	7.47		4.41	3.95					59.54	64.18			
29	Asn	7.49	116.66	5.25	2.71 2.60					51.51	40.38			
30	Pro			4.42	2.41 1.99	2.06	3.79 3.50			65.34	32.05	27.42	50.44	

31	Gln	8.82	118.54	4.23	2.12	2.51			58.57	27.46	33.88		
						2.37							
32	Leu	7.78	121.75	4.20	1.70	1.53	0.81		56.78	41.82	27.18	24.13	
					1.64		0.77					22.43	
33	Nle	7.94		3.88		1.30		0.82	59.97		31.06		16.77
34	Ala	7.93	119.28	4.01	1.45				54.99	18.23			
35	Ala	7.54	119.57	4.11	1.45				54.54	18.13			
36	Phe	8.23	118.36	4.23	3.11		7.04		60.26	39.79			
					3.04								
37	Ile	8.17		3.75	2.02	1.68	0.81		63.63	37.99	28.26	13.49	
						0.90					17.82		
38	Lys	7.84	121.49	4.10	1.86	1.44	1.66	2.92	58.07	32.71	25.10	29.19	41.84
39	Gln	7.94	118.01	4.16	2.04	2.39			56.55	29.01	33.90		
						2.33							
40	Arg			4.24	1.73		2.93		56.34	30.57		42.83	
					1.59								
41	Thr			4.25	4.23	1.16			62.55	69.68	21.77		
42	Ala			4.23	1.34				53.11	19.07			
43	Lys			4.16	1.64	1.28	1.59	2.90	56.95	32.84	24.67	28.99	41.85
						1.23							
44	Tyr	7.99	120.92	4.51	2.98		7.05	6.76	58.15	38.94		63.02	48.18
					2.93								
45	Val	7.82	122.38	3.99	1.98	0.87			61.93	33.05	21.19		
						0.87					20.63		
46	Ala	8.18	127.34	4.20	1.36				52.59	19.03			
47	Asn			4.61	2.77				52.97	38.96			
					2.74								

Table S1B: Assignments ^1H and ^{13}C resonances of a NCBD peptide harbouring a 4S fluoroproline at position P13 at 304 K, 20 mM phosphate buffer pH 7.2

		H	N	H α	H β	H γ	H δ	H ϵ	C	C α	C β	C γ	C δ	C ϵ
1	Ser			3.75	3.61 3.55					74.80	65.24			
2	Ala			4.27	1.40					54.61	18.65			
3	Leu	8.16		4.02		1.54	0.90			57.98		27.15	23.38	
4	Gln	7.73	117.21	3.94	2.11	2.39				58.97	28.34	33.89		
5	Asp	8.13	120.27	4.39	2.80 2.66					57.18	40.06			
6	Leu	8.29	122.88	4.01	1.64		0.84 0.76			58.05	41.92		25.96 24.92	
7	Leu	8.38	118.64	3.95						58.01				
8	Arg	7.99	118.13	4.03	1.94	1.75 1.61	3.23			59.21	30.15	27.76	43.26	
9	Thr	7.86	115.27	4.07	4.28	1.24				65.65	68.82	22.51		
10	Leu	8.01	121.31	4.08	1.82 1.55		0.82 0.81			56.39	41.49		25.97 23.37	
11	Lys	7.43		4.25	1.94	1.60	1.67	2.96		56.86	32.76	25.17	29.42	41.85
12	Ser	7.69	115.87	4.70	3.98						63.20			
13	(4S)FPro			4.62	2.61 2.39	5.48	4.09 4.03			61.88				
14	Ser	8.05		4.37	3.88 3.81					58.79	63.50			
15	Ser	7.82	118.64	4.84	4.16 4.04					56.80	63.18			
16	Pro			4.36	2.42 1.94	2.17 2.04	3.94 3.88			65.43	31.81	27.92	50.71	
17	Gln	7.94		4.07	2.17 2.14	2.44 2.04				59.02	28.32	34.14		
18	Gln	7.94	121.98	4.12	2.25 2.13	2.50 2.39				58.94	28.36	33.88		
19	Gln	8.21	118.38	3.80										
20	Gln			3.94						58.45				
20	Gln	8.02	117.98			2.40						34.13		
21	Gln	7.82	117.40	4.12						58.97				
22	Val	8.01	119.26	3.42	2.14	0.97 0.69				67.06	31.47	23.35 21.46		
23	Leu	8.00		3.82	1.80 1.72	1.53	0.80			58.28	41.87	27.15	24.31	
24	Asn	8.24	117.26	4.40	2.92 2.83					55.93	37.91			
25	Ile	7.98	122.05	3.75	1.89	1.82 1.07 0.83	0.75			65.08	38.19	28.98 17.86	13.79	
26	Leu	8.13	119.49	3.91	1.85 1.81		0.58 0.52			57.64	41.06		25.67 22.28	
27	Lys	8.25	116.81	3.94	1.88	1.50	1.65	2.94		58.82	32.59	25.11	29.42	41.82

28	Ser	7.51		4.43	3.96				59.47	64.15				
29	Asn	7.59	118.11	5.25	2.71				51.44					
					2.67									
30	Pro			4.43	2.40	2.07	3.81		65.29	32.04	27.40	50.51		
					2.02		3.54							
31	Gln	8.80	118.53	4.23	2.14	2.52			58.56	27.50	33.93	-		
						2.39								
32	Leu	7.79		4.21	1.71		0.83		56.79	41.99		24.21		
							0.83					22.58		
33	Nle	7.91		3.90		1.31		0.81	59.85		30.91		16.65	
34	Ala	7.94	119.25	4.03	1.46				54.97	18.13				
35	Ala	7.59	119.60	4.11	1.45				54.50	18.21				
36	Phe	8.18	118.39	4.25	3.15			7.07	60.22	39.73			60.93	
					3.05									
37	Ile	8.14		3.75	2.00	1.26	0.82		63.38	38.01	17.75	13.40		
						0.90								
38	Lys	7.89	121.45	4.10	1.86	1.46	1.66	2.95	58.05	32.65	25.11	29.14	41.83	
39	Gln	7.95	119.23	4.18	2.06	2.41			56.53	28.98	33.84			
						2.34								
40	Arg	7.57		4.25	1.66		2.97		56.38	30.55		42.87		
					1.61									
41	Thr			4.26	4.24	1.18			62.44	69.70	21.71			
42	Ala	7.67		4.24	1.36				53.04	19.06				
43	Lys			4.18	1.66	1.28	1.60	2.92	56.95	32.84	24.63	28.97	41.83	
						1.23								
44	Tyr	8.02	121.05	4.53	2.99		7.07	6.77	58.07	38.89				
					2.94									
45	Val	7.84	122.60	3.99	1.98	0.87			61.92	33.05	21.12			
						0.87					20.59			
46	Ala	8.21	127.52	4.21	1.36				52.56	19.02				
47	Asn			4.62	2.78				52.97	38.94				
					2.75									

Table S1C: Assignments ^1H and ^{13}C resonances of a NCBD peptide harbouring a *4R* fluoroproline at position P30 at 304 K, 20 mM phosphate buffer pH 7.2

		H	N	H α	H β	H γ	H δ	H ϵ	C α	C β	C γ	C δ	C ϵ
1	Ser			3.76	3.62 3.54				74.77	65.25			
2	Ala	8.12		4.27	1.40				54.80	18.63			
3	Leu	8.36		4.01	1.87 1.64	1.52	0.90		58.09	41.47	27.14	23.19	
4	Gln	7.65		3.93	2.14	2.40			59.10	28.33	33.78		
5	Asp	8.12		4.38	2.83 2.64				57.25	39.99			
6	Leu	8.44		3.97	1.85	1.62			58.21	41.43	27.09		
7	Leu	8.44		3.95	1.85 1.52		0.84		58.47	41.10		25.29 22.43	
8	Arg	8.02		4.01	1.94	1.75 1.59	3.23		59.44	30.15	27.76	43.23	
9	Thr	7.89		4.01	4.25	1.32 1.23			66.37	68.57	22.81		
9	Thr												
10	Leu	8.24		4.03	1.87 1.57	1.84	0.85 0.85		56.85	41.81	26.64	26.10 23.49	
11	Lys	7.44		4.28	1.96 1.83	1.59 1.48	1.66	2.95	56.66	32.78	25.12	29.37	41.84
12	Ser	7.52		4.71	3.99				57.56	62.87			
13	Pro			4.35	2.31 1.84	2.08 2.00	3.86 3.63		63.30	32.00	27.84	50.44	
14	Ser	8.36		4.35	3.88 3.82				58.90	63.43			
15	Ser	7.84		4.88	4.23 4.06				56.87	63.32			
16	Pro			4.38	2.44 1.95	2.19 2.03	3.90 3.86		65.55	31.79	27.98	50.36	
17	Gln	7.96		4.11	2.26 2.09	2.48 2.41			59.02	28.26	34.23		
18	Gln	7.99		4.03	2.11	2.43			59.17	28.30			
19	Gln	8.20		3.75	2.24 2.17	2.43							
20	Gln	7.99		3.95	2.15	2.42			58.30	28.42			
21	Gln	7.96		4.12	2.28 2.15	2.41			58.91	28.33	33.78		
22	Val	8.06		3.37	2.16 0.66	0.97 0.66				31.42	23.41 21.42		
23	Leu	7.94		3.79	1.82 1.53	1.74	0.80 0.85		58.43	41.17	27.01	24.43 23.79	
24	Asn	8.23		4.39	2.91 2.82		7.52 6.93		55.90	37.93			

25	Ile	8.04	3.71	1.85	1.84	0.73		65.28	38.25	28.96	13.89	
					1.04					17.95		
					0.83							
26	Leu	8.13	3.89	1.82	1.18	0.50		57.87	41.02		25.62	
				1.47		0.43					22.09	
27	Lys	8.34	3.90	1.86	1.47	1.64	2.94	58.73	32.50	25.12	29.39	42.46
28	Ser	7.43	4.42	3.97				59.31	64.21			
29	Asn	7.49	5.16	2.73				51.36	40.53			
30	(4R)FPro		4.59	2.81	5.55	4.19		63.95			50.61	
				2.30	5.47	3.90						
31	Gln	8.06	4.20	2.13	2.50			58.85	27.90	34.17		
					2.41							
32	Leu	7.67	4.23	1.69	1.52	0.85		56.92	41.81	27.01	25.40	
				1.66		0.77					24.11	
33	Nle	8.08	3.90	1.88	1.33		0.82	60.19	32.43		24.65	16.88
34	Ala	8.06	4.04	1.47				55.01	18.02			
35	Ala	7.57	4.14	1.52				54.74	18.23			
36	Phe	8.46	4.24	3.18			7.07	61.45	39.85			60.77
				3.10								
37	Ile	8.34	3.75	2.06	1.74	0.86		63.93	37.89	28.40	13.56	
					0.96					17.79		
38	Lys	7.84	4.10	1.89	1.44	1.68	2.96	58.35	32.63	25.08	29.19	41.88
39	Gln	7.92	4.16	2.06	2.42			56.57	28.98	33.83		
					2.35							
40	Arg		4.23	1.55	1.38	2.89		56.29	30.55	26.94	42.70	
41	Thr	7.97	4.25	4.26	1.19			62.78	69.67	21.76		
42	Ala	7.96	4.24	1.37				53.29	19.01			
43	Lys	8.01	4.16	1.65	1.25	1.58	2.88	57.14	32.76	24.62	28.96	41.82
						1.58						
44	Tyr	8.00	4.51	3.01		7.06	6.76	58.28	38.97		63.13	48.15
				2.93								
45	Val	7.83	4.00	2.00	0.88			61.95	33.02	21.20		
					0.88					20.60		
46	Ala	8.18	4.21	1.37				52.59	18.99			
47	Asn	8.23	4.62	2.78				52.99	38.94			
				2.75								

Table S1D: Assignments ^1H and ^{13}C resonances of a NCBD peptide harbouring a 4*S* fluoroproline at position P30 at 304 K, 20 mM phosphate buffer pH 7.2.

		H	N	H α	H β	H γ	H δ	H ϵ	C	C α	C β	C γ	C δ	C ϵ
1	Ser				3.60 3.52									
2	Ala	8.14		4.27	1.39					54.70	18.65			
3	Leu	8.35	118.22	4.01	1.74 1.64	1.53	0.90			58.07		27.16	23.23	
4	Gln	7.68	117.06	3.94	2.13	2.40				59.01	28.32	33.80		
5	Asp	8.14	120.49	4.39	2.83 2.64					57.26	40.01			
6	Leu	8.42	123.43	3.99		1.61	0.84 0.76			58.32		26.95	26.23 25.12	
7	Leu	8.41	118.56	3.94	1.79 1.47	1.86	0.82			58.18	41.09	26.87	22.48	
8	Arg	8.06	118.17	4.00	1.94	1.75 1.60	3.23			59.38	30.12	27.71	43.23	
9	Thr	7.90	116.19	4.01	4.26	1.24				66.29	68.68	22.73		
10	Leu	8.20	121.40	4.02	1.86 1.54	1.83	0.82			56.81	41.87	26.77	26.08	
11	Lys	7.45	117.19	4.27	1.96 1.83	1.59 1.47	1.65	2.95		56.68	32.78	25.20	29.41	41.83
12	Ser	7.52	116.94	4.72	4.00					57.52	62.91			
13	Pro			4.34	2.31	2.08 1.99	3.84 3.63			63.35		27.84	50.42	
14	Ser	8.34	116.38	4.34	3.88 3.82					58.89	63.43			
15	Ser	7.84	119.17	4.88	4.23 4.06					56.88	63.30			
16	Pro			4.34	2.31 1.84	2.20 2.04	3.91 3.82			65.56	31.97	27.96	50.61	
17	Gln	7.97	119.39	4.03	2.12	2.43				59.15	27.99			
18	Gln	7.96	119.47	4.12	2.28 2.11	2.41				59.05	28.31	33.74		
19	Gln	7.98		4.10	2.09 2.02					59.14	27.86	33.86		
20	Gln	7.96		3.95						58.40				
21	Gln	7.85		4.12										
22	Val	8.03	119.41	3.37	2.18	0.97 0.70				67.23	31.48	23.41 21.42		
23	Leu	7.93	119.39	3.82	1.81 1.55	1.73	0.84 0.80			58.25	41.09	26.98	25.25 23.79	
24	Asn	8.32	117.63	4.41	2.93 2.84		7.41 6.76			55.97	37.88			
25	Ile	8.05	122.32	3.73	1.87	1.06 1.05 0.85	0.75			65.38		17.92	13.88	

26	Leu	8.11	119.41	3.90	1.84	1.25	0.61		57.61	41.02		25.73
							0.54					22.44
27	Lys	8.33		3.90	1.89	1.50	1.65		59.00	32.51	25.07	29.37
28	Ser	7.54	111.61	4.41	3.96				59.79	64.05		
					3.94							
29	Asn	7.60	117.23	5.32	2.76				51.62	40.12		
					2.58							
30	(4S)FPro			4.67	2.70	5.52	4.13		63.92	31.82	70.91	
					2.43	5.46	3.74					
31	Gln	8.97	119.54	4.29	2.18	2.54			57.98	27.97	33.48	
					2.09	2.32						
32	Leu	7.78	120.95	4.22	1.75	1.41	0.82		56.63	42.00		24.15
					1.61							
33	Nle	7.79		3.90	1.87	1.32		0.81		32.27		16.57
34	Ala	7.86	119.22	4.03	1.46				54.93	18.14		
35	Ala	7.61	119.41	4.11	1.41				54.41	18.19		
36	Phe	8.04	117.97	4.27	3.13		7.07		59.97	39.69		
					3.03							
37	Ile	8.21	118.46	3.75	1.97	1.66	0.81		63.12	38.04	28.22	13.38
						1.23					17.72	
						0.88						
38	Lys	7.88	121.61	4.11	1.86	1.41	1.67	2.95	57.90	32.66	25.09	29.15
39	Gln	7.96	117.99	4.19	2.09	2.41			56.53	28.97	33.85	
					2.04	2.35						
40	Arg			4.26	1.78		2.99		56.37	30.52		42.88
					1.62							
41	Thr	7.91		4.26	4.24	1.18			62.42	69.69	21.72	
42	Ala			4.21	1.37				53.02	19.05		
43	Lys	8.04		4.18	1.66	1.30	1.60	2.93	56.86	32.86	24.65	28.98
44	Tyr	8.02	121.03	4.53	3.00		7.06	6.77	58.08	38.86		63.14
					2.94							48.22
45	Val	7.85	122.61	3.99	1.97	0.87			61.94	33.03	21.15	
						0.87					20.60	
46	Ala	8.21	127.44	4.21	1.37				52.57	19.00		
47	Asn			4.61	2.79				52.98	38.93		
					2.74							

Table S1E: Assignments ^1H and ^{13}C resonances of an unmodified NCBD peptide, 20 mM phosphate buffer pH 7.2.

		H	N	H α	H β	H γ	H δ	H ϵ	C	C α	C β	C γ	C δ	C ϵ
1	Ser			3.74	3.61 3.52					74.79	65.22			
2	Ala	8.12		4.26	1.38					54.80	18.63			
3	Leu	8.33	118.21	4.00	1.64	1.50	0.89 0.86			58.13	41.45	27.15	25.39 23.21	
4	Gln	7.67	117.00	3.92	2.11	2.38				59.09	28.32	33.81		
5	Asp	8.12	120.46	4.37	2.81 2.62					57.27	39.97			
6	Leu	8.41	123.31	3.97			0.73			58.36			26.17 24.84	
7	Leu	8.43	118.68	3.92	1.83	1.86	0.82 0.80			58.22	41.03	26.84	25.21 22.43	
8	Arg	8.03	118.16	4.00	1.93	1.73 1.58	3.22			59.34	30.10	27.80	43.24	
9	Thr	7.89	116.22	4.00	4.25	1.22				66.36	68.58	22.82		
10	Leu	8.19	121.42	4.01	1.53	1.83	0.82 0.81			56.90	41.44	26.77	26.05 23.50	
11	Lys	7.43	117.24	4.26	1.82 1.94	1.58 1.45	1.66 1.65	2.93		56.66	32.72	25.19	29.53	42.21
12	Ser	7.52			3.97						62.89			
13	Pro			4.34	2.30 1.82	1.98 1.98	3.84 3.62			63.33	32.02	27.84	50.47	
14	Ser	8.32	116.32	4.32	3.87 3.80					58.90	63.41			
15	Ser	7.81	119.16	4.86	4.20 4.05					56.89	63.31			
16	Pro			4.35	1.93 2.42	2.18 2.02	3.90			65.57	31.83	27.98	50.61	
17	Gln	7.97	119.19	4.10	2.08	2.41				59.25	27.87	34.18		
18	Gln	7.97		3.94										
19	Gln	8.01		4.11						59.07				
20	Gln	7.97		4.02	2.10	2.42				59.19	28.31			
21	Gln	7.95		4.11	2.10						27.92			
22	Val	8.03	119.37	3.36	2.15	0.96 0.66				67.22	31.43	23.47 21.46		
23	Leu	7.96	119.39	3.79	1.80 1.51	1.73	0.79			58.32	41.05	26.98	23.76	
24	Asn	8.27	117.44	4.39	2.92 2.82					55.94	37.93			
25	Ile	8.02	122.16	3.71	1.85	1.87 1.04 0.83	0.73			65.34	38.14	29.00 17.93	13.87	
26	Leu	8.12	119.27	3.89	1.44	1.80	0.54 0.48			57.68	41.00	25.96	25.62 22.09	
27	Lys	8.31		3.92	1.87	1.46	1.63	2.93		58.85	32.63	25.12	29.36	42.45

28	Ser	7.48		4.42	3.94				59.55	64.23		
29	Asn	7.51	116.98	5.25	2.72		7.16		51.51	40.42		
					2.62		7.16					
30	Pro			4.43	1.99	2.06	3.80		65.35	32.08	27.42	50.46
							3.51					
31	Gln	8.81	118.43	4.23	2.13	2.53			58.60	27.49	33.88	
						2.37						
32	Leu	7.77	121.70	4.21	1.69	1.54	0.78		56.78	41.84	27.11	24.11
					1.64							
33	Nle	7.94		3.88		1.30		0.81	60.01		31.08	16.77
34	Ala	7.94	119.39	4.02	1.45				55.01	18.13		
35	Ala	7.55	119.65	4.11	1.45				54.55	18.23		
36	Phe	8.23	118.40	4.24	3.13		7.05	7.05	60.32	39.78	133.12	131.75
					3.05							
37	Ile	8.19	117.02	3.75	2.00	1.67	0.81		63.49		17.82	13.47
						0.90						
38	Lys	7.84	121.52	4.10	1.85	1.49	1.66	2.95	58.05	32.72	25.09	29.15
						1.41						41.86
39	Gln	7.94	118.67	4.17	2.04	2.40			56.56	29.00	33.88	
						2.33						
40	Arg			4.24	1.73	1.42	2.94		56.33	30.59	26.98	42.82
					1.59							
41	Thr	7.88		4.25	4.25	1.16			62.55	69.71	21.76	
42	Ala	7.99	125.75	4.22	1.34				53.13	19.02		
43	Lys			4.16				2.91	56.97	32.83	24.63	29.00
43	Lys	8.03			1.65	1.22	1.59					41.81
						1.27						
44	Tyr	8.00	120.95	4.51	2.99		7.05	6.76	58.13	38.94	133.33	118.05
					2.93							
45	Val	7.82	122.43	3.98	1.97	0.86			61.94	33.05	21.14	
											20.59	
46	Ala	8.18	127.40	4.20	1.36				52.61	18.99		
47	Asn	8.27	117.33	4.61	2.78				52.99	38.93		
					2.74							

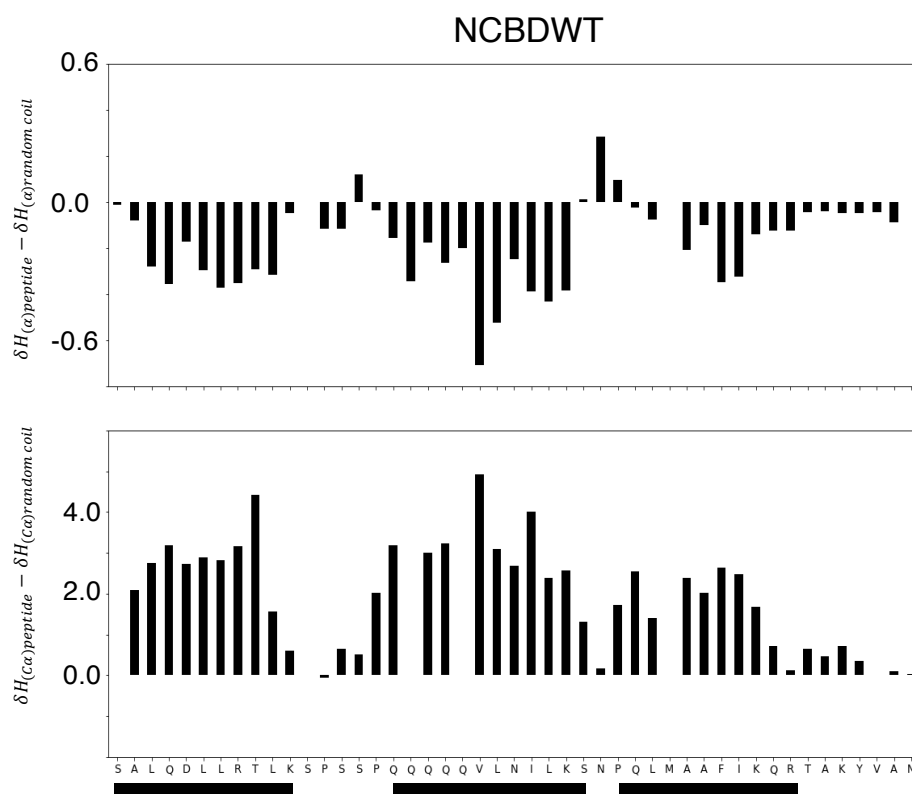


Figure S3A: Plots of the secondary chemical shift of unmodified NCBD. The random coil chemical shifts of native NCBD sequence are generated from POTENCI. Non canonical amino acids such as Norleucine are not recognized in POTENCI, thus we used the methionine amino acid.

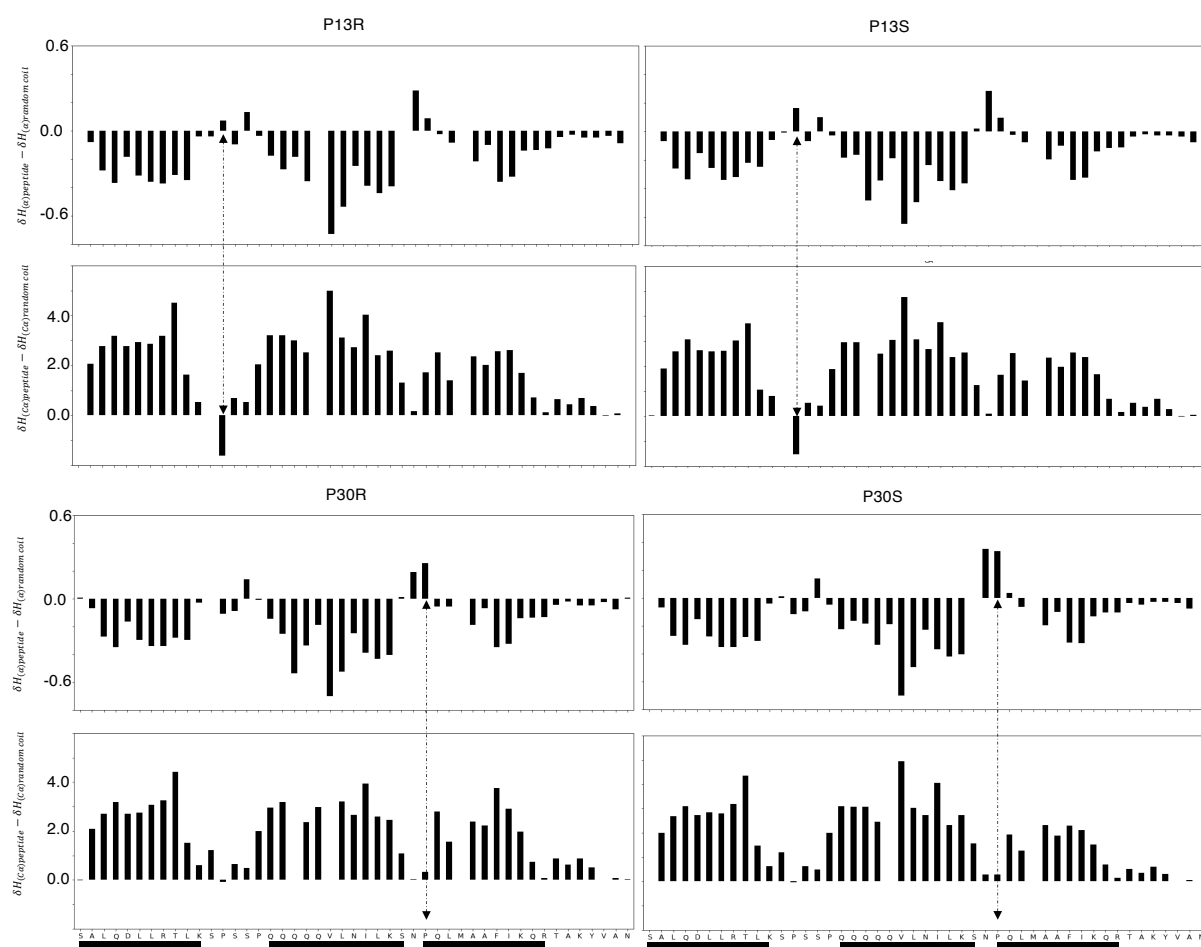


Figure S3B: Plots of the secondary chemical shift of fluorinated NCBs. The random coil chemical shifts of native NCB sequence are generated from POTENCI. Non canonical amino acids such as Norleucine are not recognized in POTENCI, thus we used the methionine amino acid.

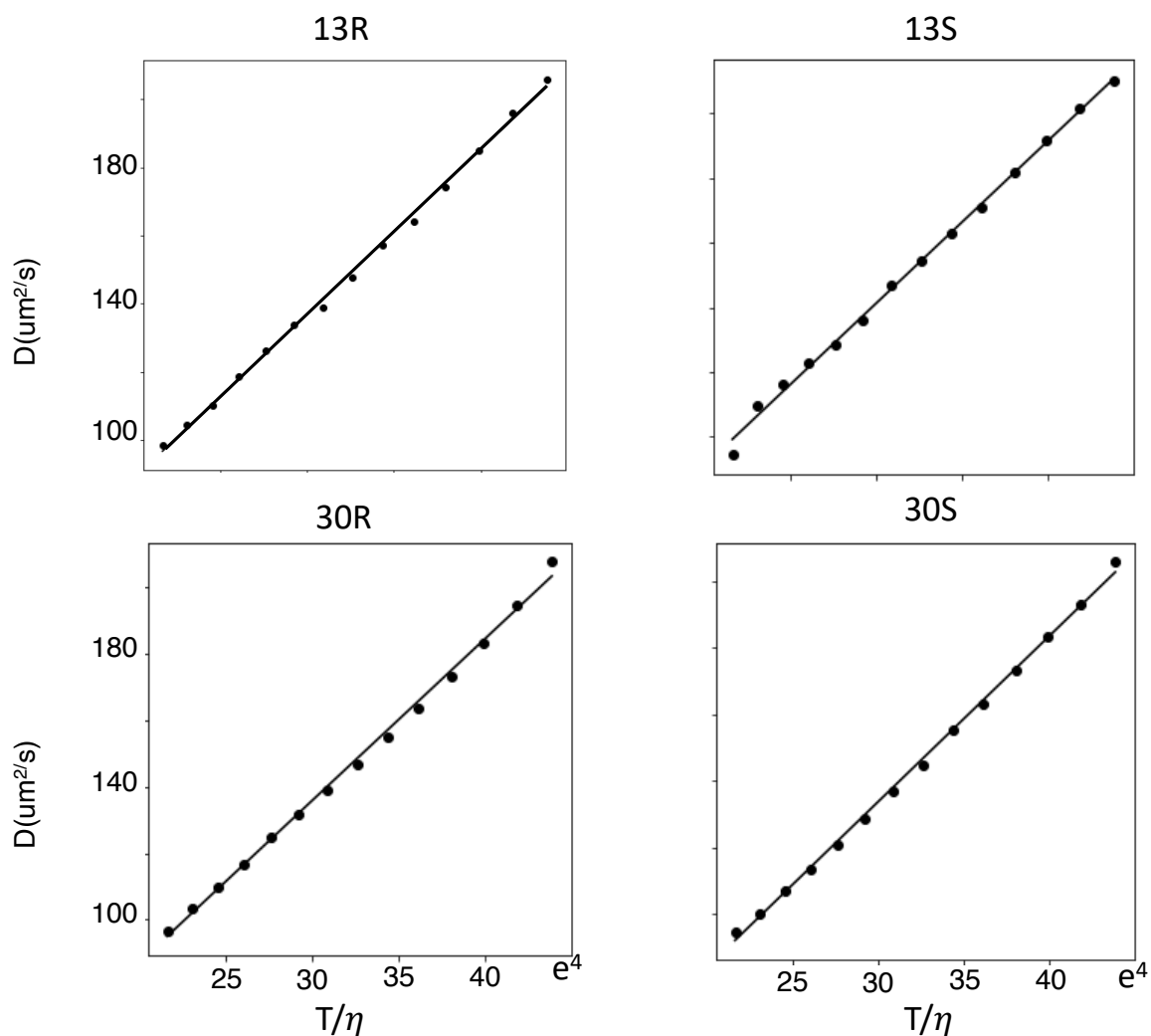


Figure S4: Linear variation of the rotational diffusion constant D of fluorinated NCBDs as a function of T/η (η being the viscosity of the sample). The diffusion constants are extracted from 2D ^1H DOSY experiments. The dots represent the measured diffusion coefficients of an isolated methyl signal at 0.6 ppm (at 283 K). the solid line represents a fit of the data to a polynomial of first degree. The temperature ranged from $T=283$ K to $T=309$ K. The viscosity of the water was considered to vary linearly in this temperature range and was computed using a home-made Python script. The hydrodynamic radius was then deduced from Stock-Einstein relation of the rotational diffusion.

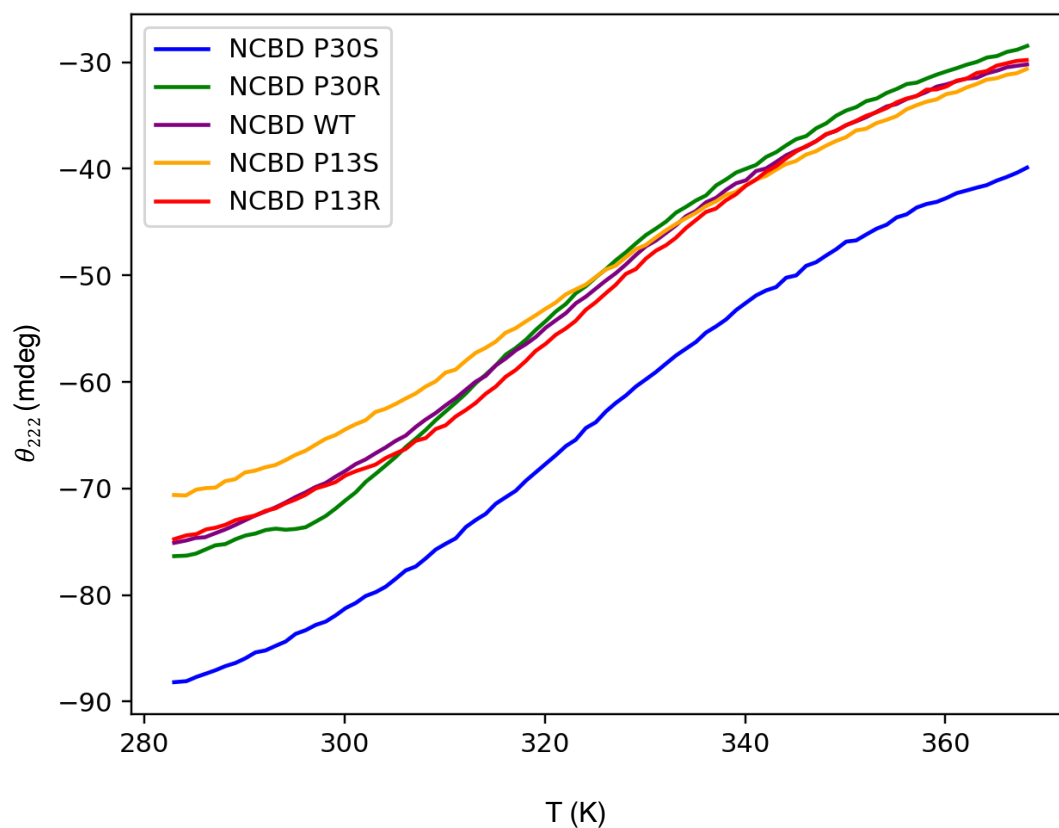


Figure S5: Thermal denaturation of NCBD variants harbouring a *4R* or a *4S* fluoroproline in the L1 loop (labelled P13R and P13S, respectively) or a *4R* or a *4S* fluoroproline in the L2 loop (labelled P30R and P30S, respectively). The denaturation profile of an unmodified NCBD is shown as WT. The protein concentrations were 80 μ M in phosphate buffer at pH 7.2. The heating's rate was set to 2 $^{\circ}$ /min.

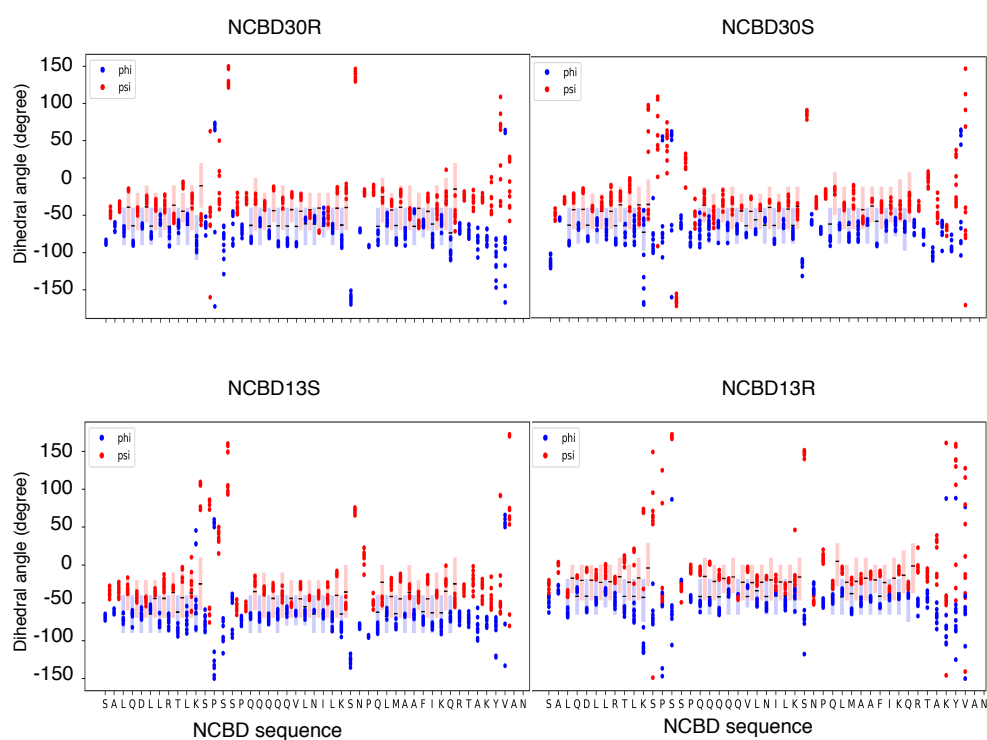


Figure S6: Analysis of dihedral angle restraints obtained from the NCBD chemical shifts measured for the four NCBD peptides modified by the insertion of a fluoroproline residue. The restraints were obtained from the DANGLE program installed in the CCPNmr package (version 2.1). The target values are indicated by a black line while the range of experimental values are indicated by a blue and red shading for phi and psi angles, respectively.

Table S2: statistics of structure calculations of non-fluorinated and fluorinated NCBD variants from ARIA

Distance Constraints					
	NCBD30R	NCBD30S	NCBD13R	NCBD13S	NCBDWT
Total	701	746	544	502	690
Intra _residual	368	328	329	252	330
Inter _residual	0	0	0	0	0
Short range (i - j = 1)	147	186	110	125	182
Middle range (i - j = 2 ,3,4)	147	192	79	111	129
Long range (i - j > 4)	34	37	24	13	40
Dihedral Angle Constraints					
Phi restraints	45	45	45	45	45
Psi restraints	45	45	45	45	45
NMR distance restraints					
RMS of violation(A°)	0.108 ± 0.011	0.071 ± 0.013	0.071 ± 0.005	0.051 ± 0.013	0.057 ± 0.010
Number of violation > 0.3 A°	15.9 ± 2.7	3.7 ± 2.2	0.9 ± 0.8	0.6 ± 0.7	1.0 ± 0.8
Number of violation > 0.1 A°	99.1 ± 3.1	60.9 ± 4.0	70.9 ± 2.6	33.0 ± 2.5	55.7 ± 4.7
Hydrogen bond restraints					
RMS of violation(A°)	0.045 ± 0.007	0.479 ± 0.332	0.034 ± 0.003	0.057 ± 0.012	0.110 ± 0.035
Number of violation > 0.3 A°	0.0 ± 0.0	1.6 ± 0.5	0.0	0.0	1.0 ± 0.5
Number of violation > 0.1 A°	2.2 ± 0.7	4.7 ± 1.0	1.3 ± 0.5	2.1 ± 0.6	4.6 ± 0.9
Dihedral angle restraints					
RMS of violation (°)	0.783 ± 0.242	0.638 ± 0.203	0.185 ± 0.066	0.497 ± 0.069	1.301 ± 0.182
Number of violation > 5°	0.3 ± 0.5	0.1 ± 0.3	0.0	0.0	2.1 ± 1.0
RMSD					
All residues _Backbone atoms (A°)	1.07 ± 0.22	1.06 ± 0.11	0.82 ± 0.20	1.22 ± 0.46	1.01 ± 0.28
All residues _Heavy atoms (A°)	1.57 ± 0.17	1.51 ± 0.15	1.29 ± 0.22	1.60 ± 0.48	1.39 ± 0.29
Ordered residues _Backbone atoms	0.74 ± 0.16	0.74 ± 0.16	0.39 ± 0.14	0.90 ± 0.38	0.52 ± 0.24
Ordered residues _Heavy atoms	1.28 ± 0.10	1.28 ± 0.21	0.89 ± 0.19	1.31 ± 0.37	1.00 ± 0.25
Ramachandran Plot					
residues in the most favoured region (%)	87.20 ± 3.77	78.60 ± 3.58	84.30 ± 3.02	85.40 ± 2.16	79.60 ± 3.01
residues in the additionally allowed region (%)	11.20 ± 4.05	18.40 ± 4.27	13.30 ± 3.02	11.60 ± 3.23	17.00 ± 2.87
residues in the disallowed region (%)	0.72 ± 1.16	0.69 ± 1.17	0.00	0.69 ± 1.17	0.69 ± 1.17

Chapter 7:
NMR study of binding of fluorinated
NCBD variants to transactivation
domain (TAD) of p53 tumor suppressor
protein and activation domain of ACTR
transcriptional co-activator

VII. NMR study of binding of fluorinated NCBD variants to transactivation domain (TAD) of p53 tumor suppressor protein and activation domain of ACTR transcriptional co-activator

1. Introduction

In its free form, NCBD is a molten globule that fluctuates between unfolded and folded conformations with a high helical content.^{1,2} Previously, two structures were resolved using NMR spectroscopy (PDB IDs: 1JJS and 2KKJ) in which the three helices were found to be arranged differently (Figure 1, A and B).

NCBD has several binding partners that are fully folded such as the interferon regulatory factor IRF-3 as well as intrinsically disordered binders such as the transactivation domain of tumor suppressor p53 protein (TAD-p53) and the activation domain of nuclear receptor coactivator ACTR (alternative name SRC-3) from p160 family (AD-ACTR). The structure of NCBD in complex with TAD-p53 and AD-ACTR resembles the free structure 2KKJ. However, the induced structure of NCBD upon binding to IRF-3 is characterized by a different arrangement of helices (Figure 1 D).

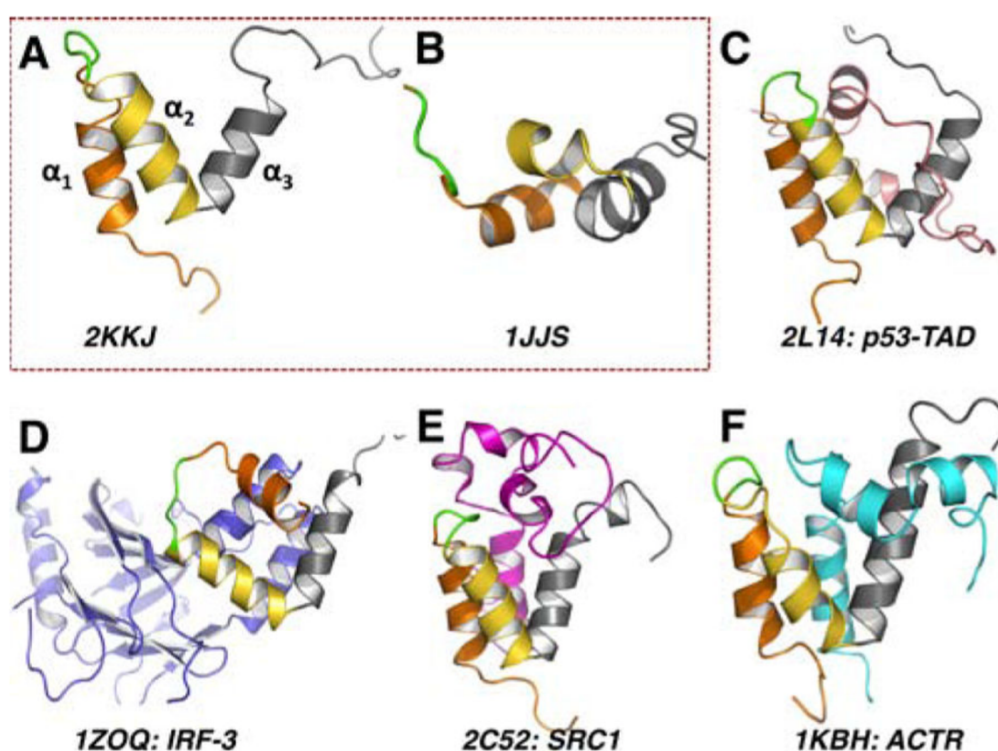


Figure 12: Free and bound NCBD structures. (A, B) NMR structures of free NCBD. (C) NCBD in complex with TAD-p53. (D) NCBD in complex with IRF-3. (E) NCBD in complex with SRC-1. (F) NCBD in complex with the activation domain (AD) of ACTR transcriptional co-

activator (SRC-3 from p160 family). The first, second and third helices of NCBD are shown in orange, yellow and gray respectively. Figure reproduced from ref.³.

Transcription factor p53 is known to be a tumor suppressor whose malfunction is related to cancers. It interacts with a large number of binding partners to accomplish various functions such as fertility and metabolism regulation.^{4,5} The transcriptional activity of p53 is mediated by a disordered domain at its N-terminus called the transactivation domain TAD-p53. This domain is composed of 61 amino acids and contains the two subdomains AD1 (1-42) and AD2 (43-61).⁶ TAD-p53 is responsible for the interactions of p53 with the transcriptional machinery components as well as to numerous components involved in DNA recombination and repair. Interestingly, the two subdomains can either bind to the same partner or to two different partners simultaneously⁶. The interaction with CREB-binding protein (CBP) mostly involves AD2 that can bind to nuclear coactivator binding domain (NCBD) of CBP.⁶ The structure of the NCBD/TAD-p53 complex (2L14) was first determined by NMR in 2010⁶. The NCBD molten globule and the intrinsically disordered TAD-p53 undergo binding induced folding to form five helices, three from NCBD and two from TAD-p53.⁷ This interaction is mainly mediated through hydrophobic interactions. The full TAD-p53 domain has a high affinity towards NCBD ($K_d = 3 \mu\text{M}$). The kinetic k_{off} and k_{on} constants were determined using stopped-flow experiments and are estimated to be around $200\text{-}500 \text{ s}^{-1}$, and $2\text{-}5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, respectively, at ionic strength 0.2 M ⁸ and 277 K .

The activation domain (AD) from the p160 transcriptional coactivator for thyroid hormone and retinoid receptors (ACTR, also called SRC-3 – steroid receptor coactivator 3 from p160 family) is unstructured in its free form with a partial alpha helical structure in its N-terminus.⁹ NCBD/AD-ACTR complex was first determined by NMR in 2002 and displayed a high helical content in the form of six helices, three from each protein. NCBD/AD-ACTR interaction was among the first examples describing a mutual synergetic folding of two IDPs.¹⁰ Binding was shown to be driven by electrostatic interactions between positively charged NCBD and negatively charged AD-ACTR resulting in a protein complex characterized by a high hydrophobic interface.¹¹ NCBD binds to AD-ACTR with a high affinity $0.034 \mu\text{M}$ ¹⁰. Kinetic constants k_{on} and k_{off} were estimated to be $3 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ and 2.7 s^{-1} , respectively, based on stopped-flow methods at 0.2 M ionic strength.⁸ Later, both the affinity of NCBD towards AD-ACTR and the interaction kinetics were shown to be modulated by the *cis/trans* isomerization of the proline lying in the vicinity of the first loop between helix 1 and 2 of NCBD.^{12,13}

Additionally, the binding between NCBD and AD-ACTR could be modulated through an enhancement of the helical content of AD-ACTR when alpha methylated amino acids were incorporated in the AD-ACTR sequence.¹⁴

Although NCBD is known to be a molten globule and characterized by structural plasticity, its mechanisms of binding to protein partners are not well understood. Its interaction with AD-ACTR was stated to follow a conformational selection mechanism where AD-ACTR recognizes a pre-folded NCBD structure among the conformations in the NCBD conformational ensemble.^{1,15} An “extended conformational mechanism” where the selection step is followed by an adjustment step of the structure seems to be taking place.¹⁶ However, an induced fit mechanism was mentioned to be possible as well.¹⁷ Finally, an induced fit mechanism based on the introduction of alpha methylated amino acids and constraining ϕ and ψ dihedral angles in the unbound AD-ACTR was reported.^{9,18}

When in interaction with p53, the binding mechanism of NCBD is also controversial, and so far the interaction is believed to adopt a local induced fit and global conformational selection binding mechanisms.⁷

Hereafter, we decided to study both complexes NCBD/TAD-p53 and NCBD/AD-ACTR by NMR spectroscopy to uncover the recognition principles at the atomic level through the use of fluorinated proline probes. Two proline residues lying in the two loops of NCBD were substituted by either 4-(*R*)- or 4-(*S*)-fluoroprolines favoring *trans*- and *cis*-amide isomers of proline, respectively. The complex formation was followed by titration experiments where the binding partner was gradually added to the fluorinated NCBD analogue.

2. NCBD/TAD-p53 interaction

Full length wild type TAD-p53 was gradually added to each of the four fluorinated NCBD variants. For each titration point, we recorded 1D ¹⁹F spectrum (Figure 2).

The addition of TAD-p53 to fluorinated NCBDs induces a small shift of fluorine resonances. A shift towards higher ppm values is recorded for P13R, P30R variants and the major peak of P13S analogue. However, a shift towards smaller ppm values is recorded for P30S variant and the minor state of P13S analogue. The largest shift was observed for the major peak of P13S variant ($\Delta = 0.15$ ppm) and the smallest one for P13R variant ($\Delta = 0.015$ ppm). The observed unique fluorine peak of NCBD during binding with TAD-p53 results from an averaging between the bound and the free fractions of NCBD. The results suggest an exchange

rate constant faster than 85 s^{-1} ($\Delta\text{ppm (0.15)} * 565\text{ MHz}$). This result agrees with previously reported k_{off} 200-500 s^{-1} obtained by using stopped flow method.⁸

The complex formation induces a variation of the linewidth of fluorine resonance. A quantitative evaluation of linewidth perturbation upon TAD-p53 addition was performed by fitting the peaks with a Lorentzian function (Figure3). For P13S, P30R and P30S NCBD variants fluorine resonances undergo a line sharpening. The P13R NCBD analogue, on the other hand, displays a line broadening. Interestingly, the complex P13R NCBD/TAD-p53 displays a linewidth around 23.5 Hz, thus almost 3 times smaller than P13S NCBD/TAD-p53 and 2 times smaller than P30R NCBD/TAD-p53 and P30S NCBD/TAD-p53.

These observations emphasize the sensitivity of chiral configuration of fluorine substituent at the Pro13 site compared to Pro30. The broadening of P13R major peak from 16 Hz to 23.5 Hz can be associated to slower rotational diffusion upon complex formation. The sharpening of fluorine signal observed for all other variants may indicate a suppression of the conformational averaging. Both minor states of P13R and P13S experience less pronounced chemical shift and line width perturbations. This difference might indicate that the dominant populations have more affinities towards TAD-p53.

Furthermore, with the exception of NCBD P13S, all titrations led to a plateau value of the perturbation for a TAD-p53/NCBD ratio between 0.6 and 0.8. Attempting to fit the data to a one-to-one equilibrium is successful for P13R and indicates a dissociation constant $K_d = 3 \pm 5\text{ }\mu\text{M}$ (Figure4). Similar dissociation constant value was reported for wild type NCBD/TAD-p53 complex.^{6,8} Consequently, the substitution of P13 by 4-(*R*)-fluoroproline does not affect the affinity of NCBD towards TAD-p53. On the other hand, tempting to fit the NCBD P13S linewidth variation to a one-to-one model was not satisfactory indicating differently sampled dynamics.

Together with the linewidth analysis, this might indicate the presence of bigger complexes such as 1 NCBD / *n* TAD-p53 where *n* is the number of TAD-p53 attached to NCBD molecule. Further analysis of the formed complexes is crucial to derive accurate conclusions.

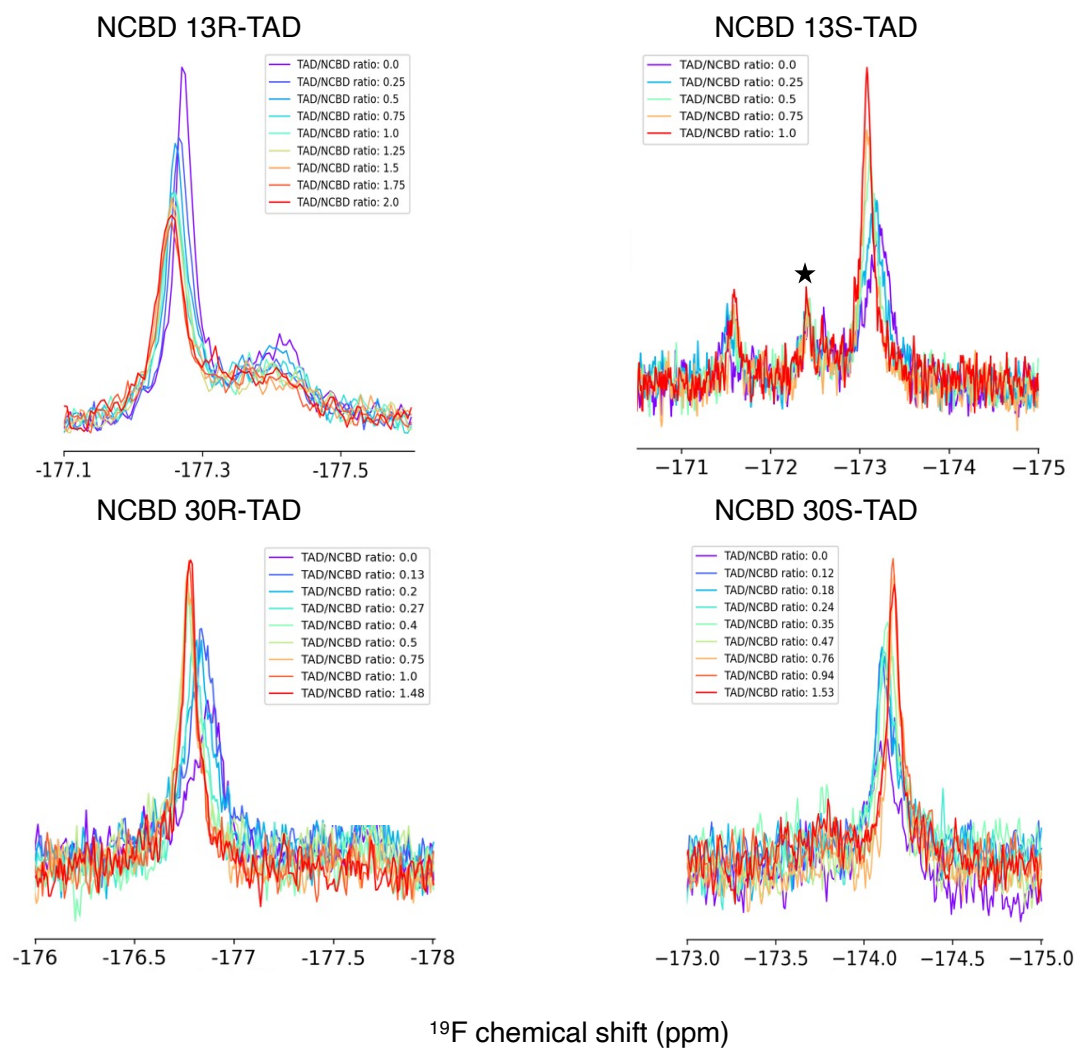


Figure 2: 565 MHz ^{19}F NMR spectra upon addition of TAD-p53 observed for the four modified NCBD proteins with fluoroprolines at position 13 (upper panels) and position 30 (lower panels) with carbon-4 in 4-(*R*)- (left panels) or 4-(*S*)- (right panels) absolute configurations. The star denotes an impurity.

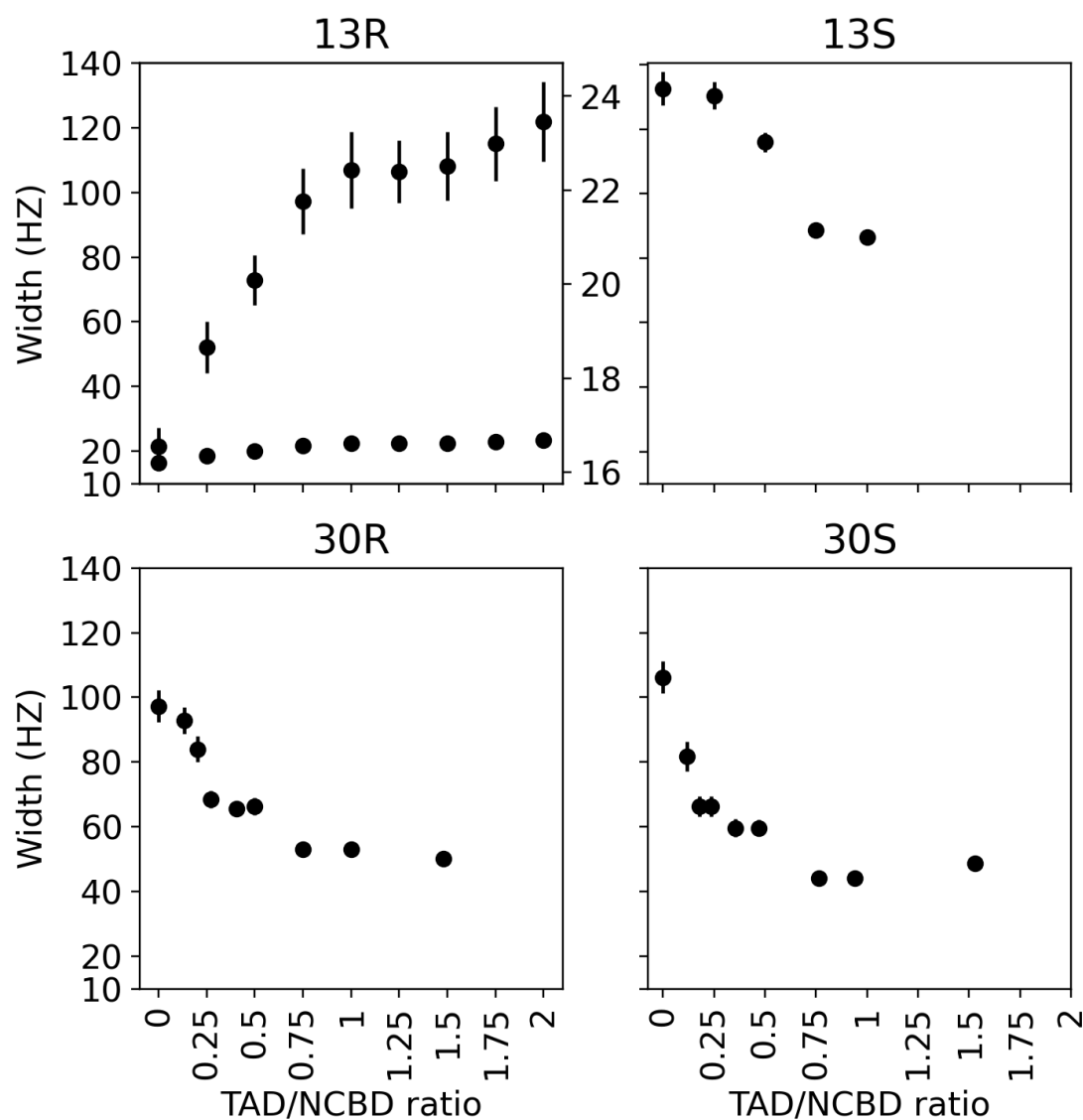


Figure 3: ^{19}F NMR linewidth variation upon addition of TAD-p53 observed for the four modified NCBD protein variants with a fluoroproline at position 13 (upper panels) or position 30 (lower panels) with a fluorine atom in 4-(*R*)- (left panels) or 4-(*S*)- (right panels) absolute configurations. For P13R NCBD variant, the same dataset is shown at two scales for the sake of comparison with the three other proteins. Magnified scale (16 to 24 Hz) is shown on the right side.

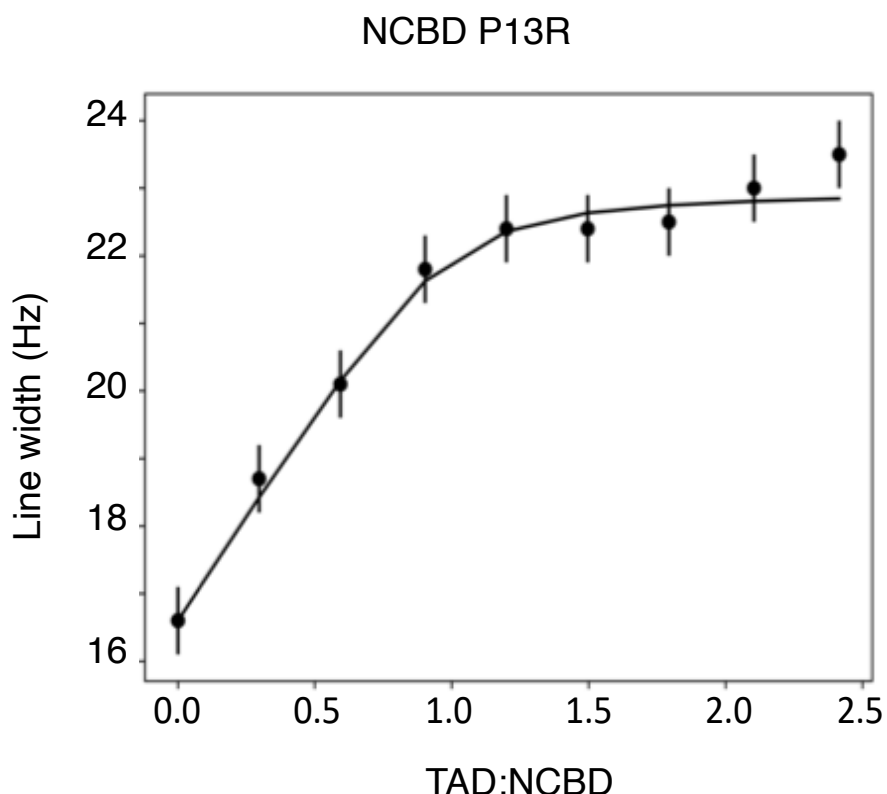


Figure 4: Fitting of the linewidth variation of ^{19}F NMR signal for P13R NCBD variant upon addition of TAD-p53 at 304 K. The circles are the experimental linewidths extracted by fitting the ^{19}F NMR peaks to a Lorentzian function. The solid line is the calculated linewidth during the interaction of P13R NCBD analogue to TAD-p53 obeying to a one-to-one stoichiometry with a dissociation constant $K_d = 3 \mu\text{M}$.

NCBD/TAD-p53 complex formation followed from TAD-p53 side

TAD-p53 has two tryptophans (W23 and W53) situated in both subdomains AD1 and AD2 respectively. The NCBD's sequence does not contain any tryptophan. Consequently, NCBD/TAD-p53 interaction can be followed through the analysis of the N-H indole ring signal of the tryptophan in the 1D ^1H spectra (Figure 5). NCBD/TAD-p53 binding can be further investigated using another probe that is Leu26 residue of NCBD. This latter has distinctive methyl chemical shifts that do not overlap with other signals from either NCBD or TAD-p53 and might bring more understanding on the rearrangement of the hydrophobic core of NCBD upon binding and is yet to be applied.

Hereafter, we followed the binding from TAD-p53's perception. The resonance assignment of the two tryptophans is adapted from the previously solved structure of p53.¹⁹

Upon addition of TAD-p53, the tryptophan signals slightly shift and broaden for P13R and P13S NCBD variants. On the other hand, a more pronounced variation of both the frequency and the linewidths is observed when complexation is carried out with P30R and P30S NCBD variants (Figure 6).

The difference in linewidths between W23 and W53 in the free form is consistent in all experiments with the exception of P30R NCBD variant. The signal sharpening observed for all experiments upon increasing TAD-p53 concentrations is consistent with the disordered nature of TAD that folds upon binding. The different behaviour of fluorinated NCBDs in contact with TAD-p53 is consistent with their different dynamics (presented in Chapter 6).

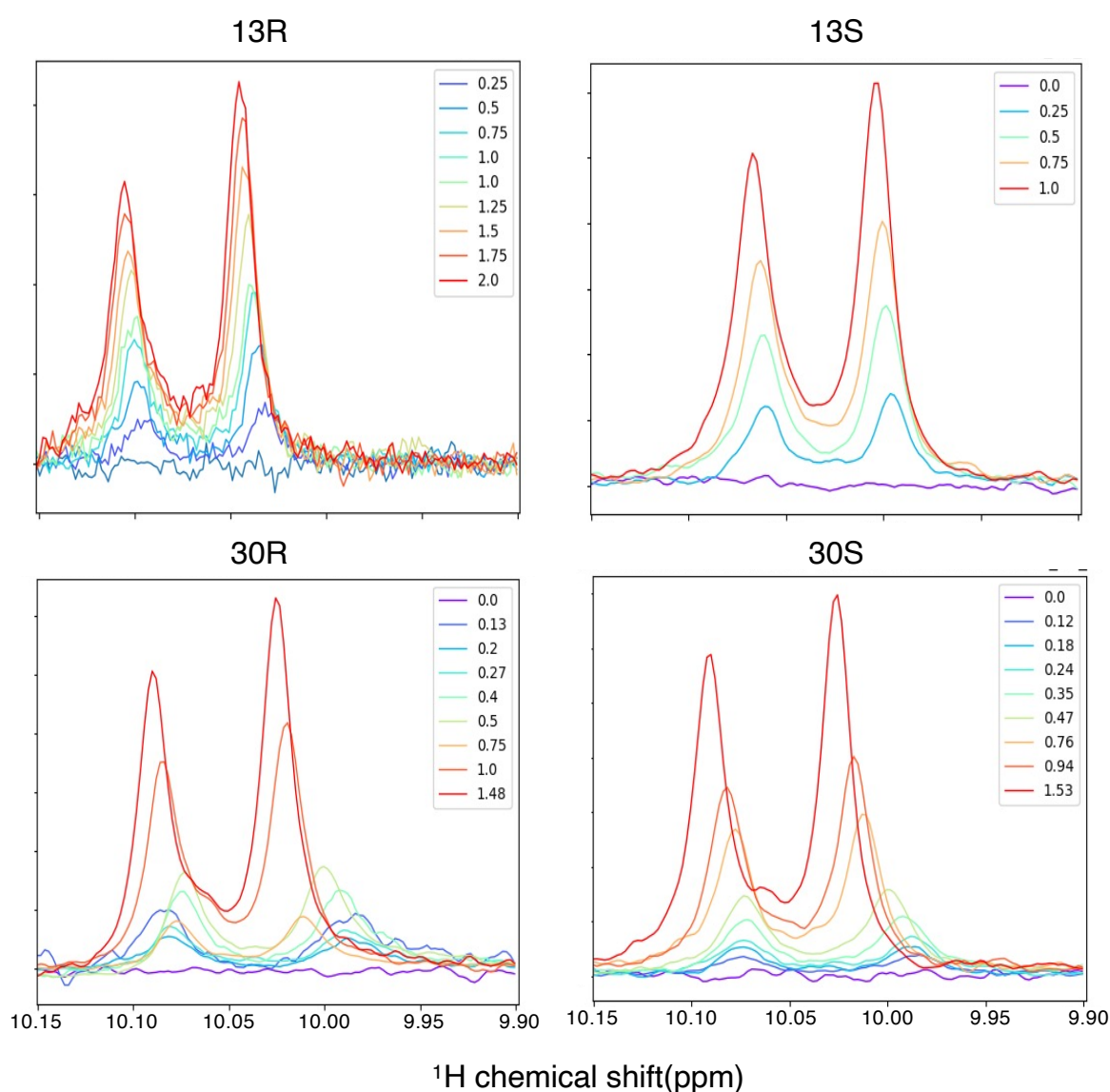


Figure 5: ^1H NMR spectra of TAD-p53/NCBD complexes (tryptophan indole region) for different TAD-p53/NCBD concentration ratios during titration experiments.

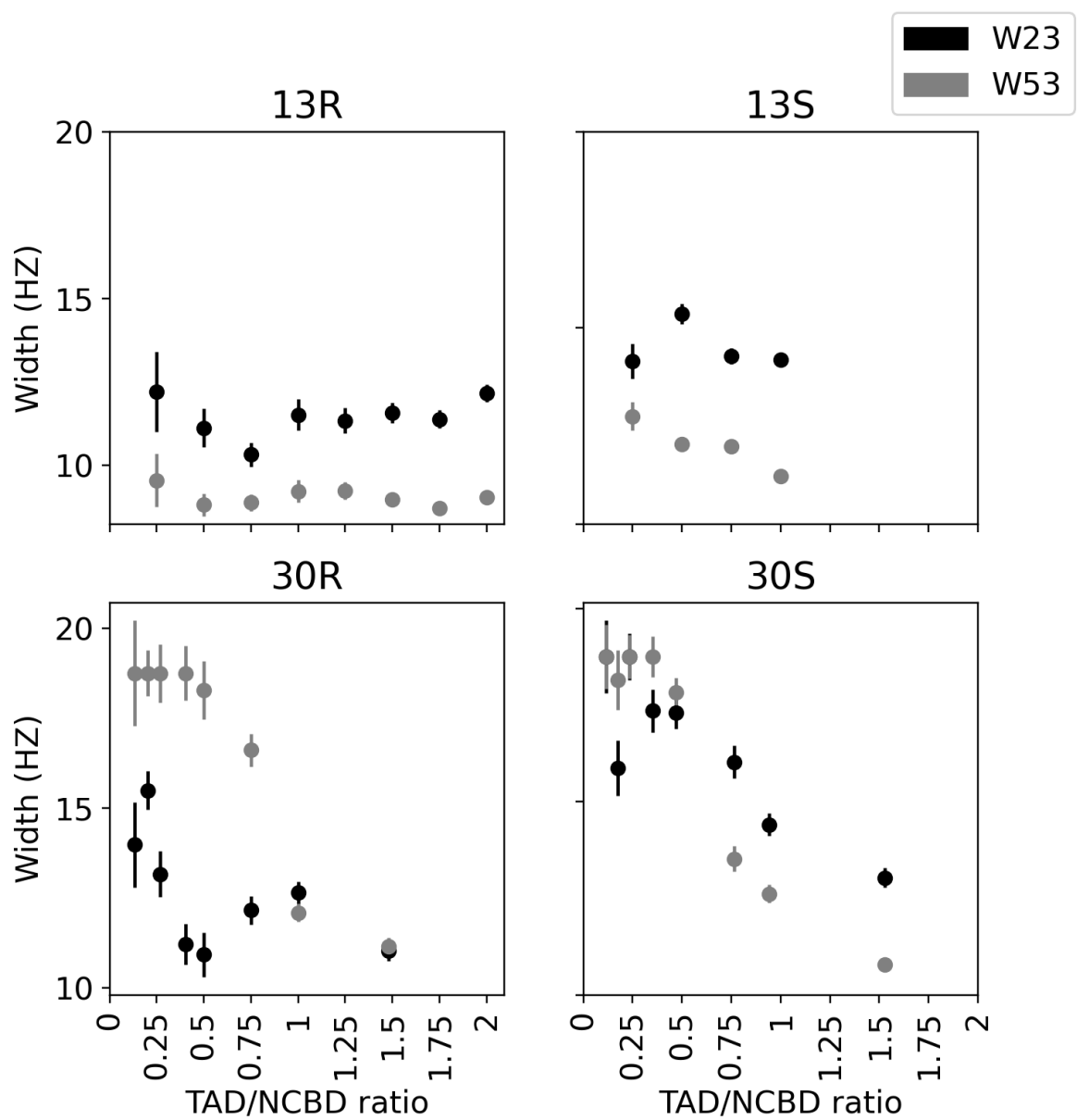


Figure 6: Variation of tryptophan indole HNε linewidths upon TAD-p53 titration. Tryptophans were identified based on their chemical shift previously published in solved structure of p53.¹⁹

3. NCBD/AD-ACTR interaction

NMR titration experiments were performed to get insights into NCBD/AD-ACTR complex formation at the atomic level. Except for P30R NCBD variant, the addition of AD-ACTR to the fluorinated NCBDs induces broadening of the fluorine resonance in comparison to the free NCBD and the appearance of a new peak (Figure 7). The largest shift is $\Delta = 0.64$ ppm in the P13S NCBD analogue. The smallest shift is $\Delta = 0.06$ ppm in the P13R NCBD variant. The two unequally populated peaks on the fluorine spectra are characteristic of a slow exchange regime with an exchange rate constant slower than 34 s^{-1} ($\Delta\text{ppm} (0.06) * 565 \text{ MHz}$). The rise of the new titrating population for P13R, P13S and P30S analogues, together with the disappearance of the minor states of P13R and P13S upon addition of AD-ACTR are in favour of a conformational selection mechanism as such behaviour corresponds to a shift in chemical equilibrium.

The small chemical shift drift in P13R NCBD analogue might be explained by the minor rearrangement of NCBD in the complex. Thus, the helical packing topology in P13R NCBD variant is the closest to that in the NCBD/AD-ACTR complex. This is consistent with the fact that P13R NCBD analogue resembles the previously resolved structure of free NCBD in solution (PDB ID: 2KKJ). (Chapter 6)

Upon addition of AD-ACTR, the fluorine resonance in P13R NCBD variant broadens, whereas the other fluorinated NCBDs witness a peak sharpening. This behavior is very similar to what we observed for NCBD binding to TAD-p53. Again, the broadening of P13R can be associated to slower rotational diffusion upon complex formation. The sharpening of fluorine signal observed for all other proteins may indicate a suppression of the conformational averaging (Figure 8). Interestingly, the linewidths of the complexes P13R NCBD/AD-ACTR, P13S NCBD/AD-ACTR and P30R NCBD/AD-ACTR are very close (25 Hz), however, for the P30S NCBD/AD-ACTR complex it is slightly broader.

To get further insights into the NCBD/AD-ACTR interaction, we performed ITC experiments to measure the thermodynamic parameters of the complexes (Table 1). The high affinity of P13R NCBD to AD-ACTR is in favour of its predisposition to a complex-like structure. On the other hand, P30S NCBD analogue is the only NCBD variant reporting a significant affinity loss towards AD-ACTR. Together with its slightly higher complex linewidth, it might indicate that the complex undergoes a conformational averaging. The P13S and P30R report similar affinities to ACTR as the NCBD WT.

Previous single-molecule FRET studies of NCBD reported $K_d = 80$ nM of NCBD/AD-ACTR complex when proline 13 is in its *trans* isomer,¹² which agrees with $K_d = 89$ nM for the P13R-NCBD/AD-ACTR complex. The same study reported $K_d = 650$ nM, when proline 13 is in the *cis*-amide configuration. This is more than two times higher than $K_d = 297$ nM of P13S-NCBD/AD--ACTR favouring the *cis* isomer.

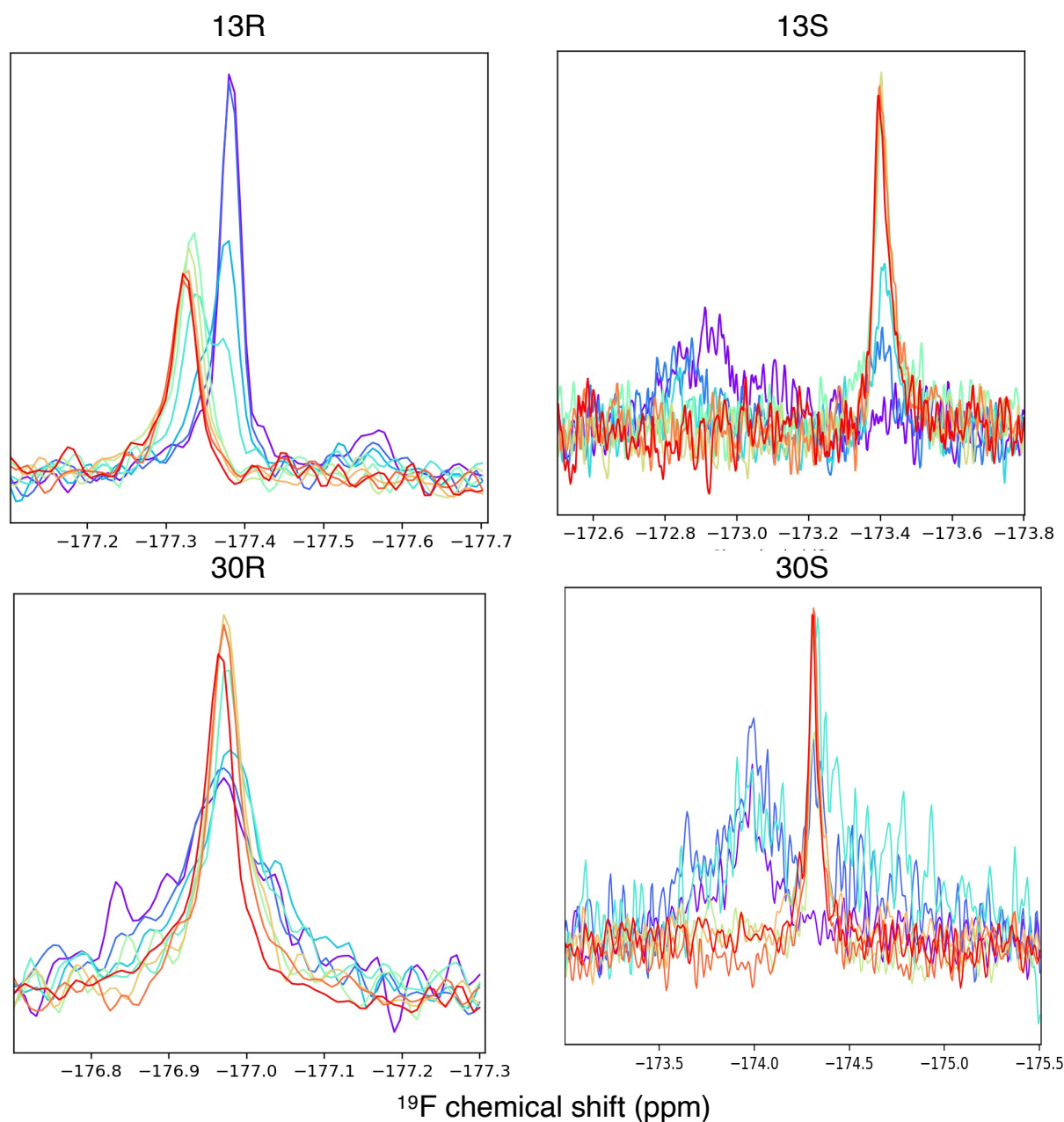


Figure 7: 565 MHz ^{19}F NMR spectra observed for the four fluorinated NCBDs P13R, P13S (upper panels) and P30R, P30S (lower panels) upon addition of AD-ACTR at 304 K.

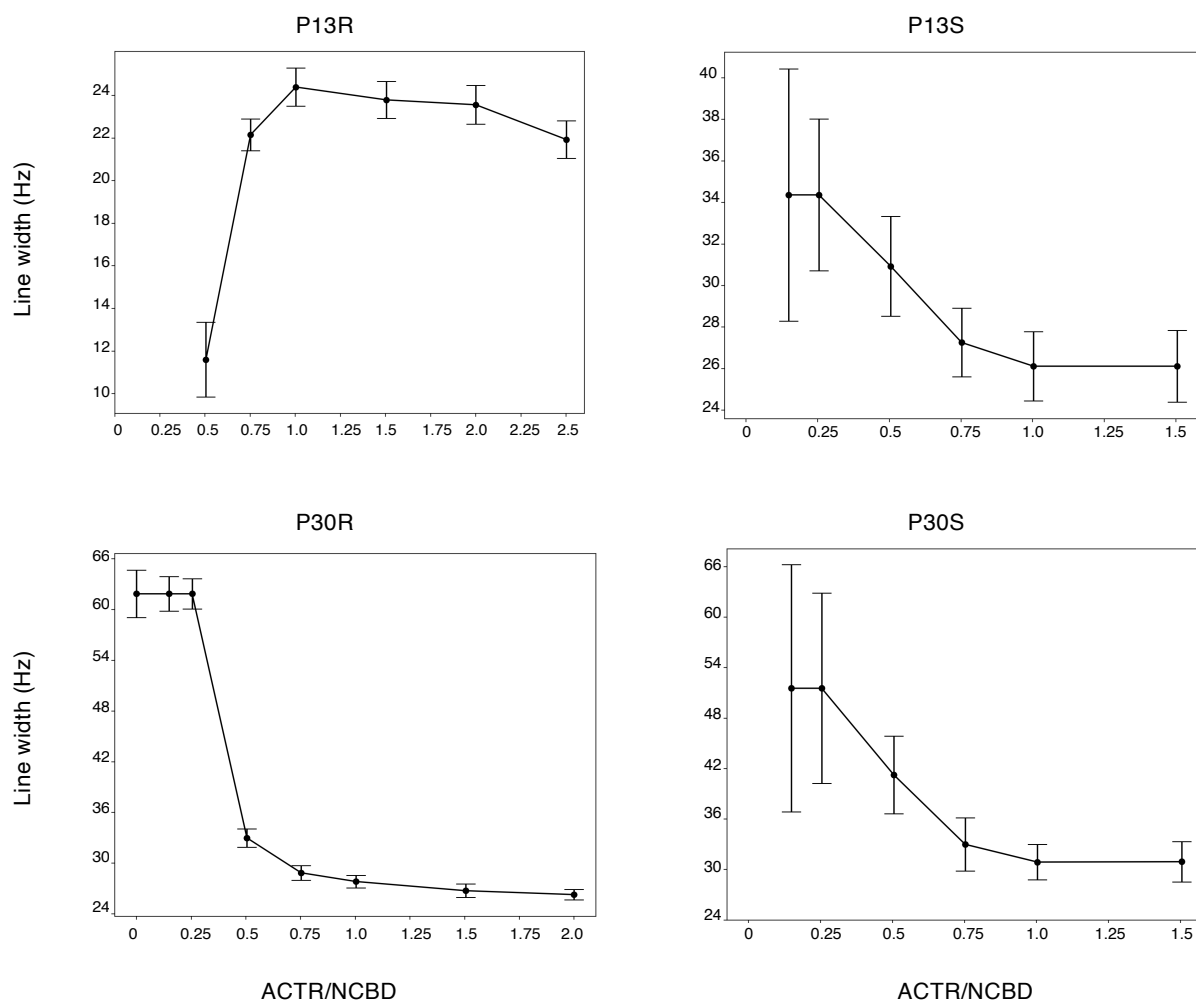


Figure 8: Variation of the ^{19}F NMR peak linewidths in the titration experiments of fluorinated NCBDs upon interaction with AD-ACTR. Recorded at 565 MHz ^{19}F NMR frequency at 304 K.

Table 1: Thermodynamic parameters of NCBD/AD-ACTR complex formation from isothermal titration calorimetry experiments.²⁰

	K_d [μM]	ΔG [kcal mol⁻¹]	ΔH [kcal mol⁻¹]	-T ΔS [kcal mol⁻¹]
NCBD WT-ACTR	0.207	-9.3	-12.65	3.35
P13R-ACTR	0.089	-9.81	-9.37	-0.45
P13S-ACTR	0.297	-9.08	-9.86	0.77
P30R-ACTR	0.243	-9.23	-8.65	0.57
P30S-ACTR	0.433	-8.91	-13.55	4.63

4. Conclusions

In this work we studied the interaction between fluorinated NCBDs and two of its important binding partners TAD-p53 and AD-ACTR by NMR titration experiments. Fluoroproline incorporated in NCBD served as local probes reporting on changes at the residue level upon binding by modifications of chemical shifts, linewidths and intensities of fluorine signals. We demonstrated that the NCBD interactions are affected by structural heterogeneity originating from proline modifications. Modifying Pro13 has a significant impact on NCBD's dynamics and affinities towards both TAD-p53 and AD-ACTR. When NCBD contains 4-(*R*)-fluoroproline at residue 13 (P13R-NCBD) structure it forms the most stable complex with TAD-p53 ($K_d = 3 \mu\text{M}$) determined through fitting the linewidth of the fluorine signal to a one-to-one interaction model. This result is in agreement with the previously determined affinities of NCBD towards TAD-p53.⁸ The data extracted from other fluorinated NCBDs (i.e., P13S, P30R and P30S) could not be fitted to this model. The preliminary data suggest the formation of bigger complexes resulting from the interaction of 1 NCBD and 2 or more TAD-p53 molecules. Further experiments are needed to investigate this hypothesis. Finally, we estimated that the exchange rate constant of NCBD/TAD-p53 is faster than 85 s^{-1} . The presence of tryptophan residues on TAD-p53 and particularly isolated methyl groups on NCBD is very

advantageous as the interaction can be followed using ^1H NMR from both NCBD and TAD-p53 perspectives.

When AD-ACTR is the binding partner, the P13R NCBD variant has the highest affinity among the fluorinated NCBDs ($K_d = 89 \text{ mM}$). The binding kinetics of fluorinated NCBD and AD-ACTR occurs on a timescale slower than 34 s^{-1} . Upon addition of AD-ACTR, fluorine signal of free NCBD gradually disappears and a new fluorine population appears in the 1D ^{19}F spectrum. This observation is in favour of a conformational selection binding mechanism between NCBD and AD-ACTR from the NCBD point of view.

Upon addition of both TAD-p53 and AD-ACTR, the fluorine signal of P13R NCBD analogue undergoes a broadening that could refer to the formation of the complexes. In contrast, the binding of P13S, P30R and P30S NCBD analogues is accompanied by a peak sharpening. This peculiar behaviour could be explained by different dynamics sensed by these complexes that are yet to be understood.

In conclusion, this work was performed to reveal the effect of proline residues of NCBD on its binding affinities and kinetics. Future work will include modelling of the data to understand the stoichiometry of NCBD/TAD-p53 complexes. Other techniques such as native mass spectrometry could be useful for this matter. Further understanding of the complex formation could be derived from titration experiments performed between fluorinated NCBDs and both subdomains AD1 and AD2 of TAD-p53 separately.

5. Experimental section

Peptides with purity above 95% were synthesized in the Laboratory of Biosystems Chemistry (UMR 7242 Biotechnology and Cellular Signalling) using SPPS. Their purity was determined by reverse-phase HPLC analysis and by electrospray ionization mass-spectrometry. TAD-p53 was provided by Aromal Asokan, AD-ACTR by Alexis Jouin and wild type and fluoroproline-containing variants of NCBD by Mariam Sai.

NMR titrations

NMR measurements were performed at 600 MHz (^1H frequency) and 565 MHz (^{19}F frequency) using NMR spectrometers from Bruker equipped with a cryogenic QCI probe. All experiments were performed at 304 K.

NMR samples were prepared by dissolving dry powders of NCBD and its variants in 160 μ L 20 mM phosphate buffer at pH 7.2. 3 mm tubes were inserted into 5 mm tube of D₂O to provide an external lock signal. The titration experiments were performed by adding increasing amounts of TAD-p53 protein (1 mM) to a 100 μ M sample of NCBD. TAD/NCBD ratio ranged from 0 to 2. The number of recorded titration points was adapted for each sample. For NCBD-ACTR titration experiments, NMR samples were prepared by dissolving NCBD dry powder in 200 μ L 20 mM phosphate buffer at pH 7.2 containing 10% of D₂O. The final NCBD concentration was 50 μ M. The titration experiments were performed by adding increasing amounts of AD-ACTR (330 mM) protein to NCBD solution in a 3 mm NMR tube. ACTR/NCBD ratio ranged from 0 to 2.5 and the number of recorded titration points was adapted for each sample.

A total of eight titrations were performed:

- P13R – TAD-p53
- P13S - TAD-p53
- P30R - TAD-p53
- P30S - TAD-p53
- P13R – AD-ACTR
- P13S – AD-ACTR
- P30R – AD-ACTR
- P30S – AD-ACTR

For each titration point, a ¹⁹F NMR spectrum was recorded in addition to a 1D ¹H NMR spectrum. The number of scans was adapted to achieve a reasonable signal over noise ratio.

Preliminary Isothermal titration calorimetry (ITC) experiments were performed by Valentin Bauer (PhD thesis defended in September 2020).²⁰

Bibliography

- (1) Kjaergaard, M.; Teilum, K.; Poulsen, F. M. Conformational Selection in the Molten Globule State of the Nuclear Coactivator Binding Domain of CBP. *Proc. Natl. Acad. Sci.* **2010**, *107* (28), 12535–12540. <https://doi.org/10.1073/pnas.1001693107>.
- (2) Kjaergaard, M.; Andersen, L.; Nielsen, L. D.; Teilum, K. A Folded Excited State of Ligand-Free Nuclear Coactivator Binding Domain (NCBD) Underlies Plasticity in Ligand Recognition. *Biochemistry* **2013**, *52* (10), 1686–1693. <https://doi.org/10.1021/bi4001062>.
- (3) Burger, V. M.; Ramanathan, A.; Savol, A. J.; Stanley, C. B.; Agarwal, P. K.; Chennubhotla, C. S. Quasi-Anharmonic Analysis Reveals Intermediate States in the Nuclear Co-Activator Receptor Binding Domain Ensemble. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **2012**, 70–81.
- (4) Soussi, T. The P53 Tumor Suppressor Gene: From Molecular Biology to Clinical Investigation. *Ann. N. Y. Acad. Sci.* **2006**, *910* (1), 121–139. <https://doi.org/10.1111/j.1749-6632.2000.tb06705.x>.
- (5) Vousden, K. H.; Prives, C. Blinded by the Light: The Growing Complexity of P53. *Cell* **2009**, *137* (3), 413–431. <https://doi.org/10.1016/j.cell.2009.04.037>.
- (6) Lee, C. W.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. Structure of the P53 Transactivation Domain in Complex with the Nuclear Receptor Coactivator Binding Domain of CREB Binding Protein. *Biochemistry* **2010**, *49* (46), 9964–9971. <https://doi.org/10.1021/bi1012996>.
- (7) Yu, Q.; Ye, W.; Wang, W.; Chen, H.-F. Global Conformational Selection and Local Induced Fit for the Recognition between Intrinsic Disordered P53 and CBP. *PLoS ONE* **2013**, *8* (3), e59627. <https://doi.org/10.1371/journal.pone.0059627>.
- (8) Dogan, J.; Jonasson, J.; Andersson, E.; Jemth, P. Binding Rate Constants Reveal Distinct Features of Disordered Protein Domains. *Biochemistry* **2015**, *54* (30), 4741–4750. <https://doi.org/10.1021/acs.biochem.5b00520>.
- (9) Schmidtgal, B.; Chaloin, O.; Bauer, V.; Sumyk, M.; Birck, C.; Torbeev, V. Dissecting Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein by Chemical Synthesis of Conformationally Constrained Analogues. *Chem. Commun.* **2017**, *53* (53), 7369–7372. <https://doi.org/10.1039/C7CC02276J>.
- (10) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. Mutual Synergistic Folding in Recruitment of CBP/P300 by P160 Nuclear Receptor Coactivators. *Nature* **2002**, *415* (6871), 549–553. <https://doi.org/10.1038/415549a>.
- (11) Demarest, S. J. Packing, Specificity, and Mutability at the Binding Interface between the P160 Coactivator and CREB-Binding Protein. *Protein Sci.* **2004**, *13* (1), 203–210. <https://doi.org/10.1110/ps.03366504>.
- (12) Zosel, F.; Mercadante, D.; Nettels, D.; Schuler, B. A Proline Switch Explains Kinetic Heterogeneity in a Coupled Folding and Binding Reaction. *Nat. Commun.* **2018**, *9* (1), 3332. <https://doi.org/10.1038/s41467-018-05725-0>.
- (13) Buholzer, K. J.; McIvor, J.; Zosel, F.; Teppich, C.; Nettels, D.; Mercadante, D.; Schuler, B. Multilayered Allosteric Modulation of Coupled Folding and Binding by Phosphorylation, Peptidyl-Prolyl *Cis* / *Trans* Isomerization, and Diversity of Interaction Partners. *J. Chem. Phys.* **2022**, *157* (23), 235102. <https://doi.org/10.1063/5.0128273>.
- (14) Bauer, V.; Schmidtgal, B.; Gógl, G.; Dolenc, J.; Osz, J.; Nominé, Y.; Kostmann, C.;

- Cousido-Siah, A.; Mitschler, A.; Rochel, N.; Travé, G.; Kieffer, B.; Torbeev, V. Conformational Editing of Intrinsically Disordered Protein by α -Methylation. *Chem. Sci.* **2021**, *12* (3), 1080–1089. <https://doi.org/10.1039/D0SC04482B>.
- (15) Naganathan, A. N.; Orozco, M. The Native Ensemble and Folding of a Protein Molten-Globule: Functional Consequence of Downhill Folding. *J. Am. Chem. Soc.* **2011**, *133* (31), 12154–12161. <https://doi.org/10.1021/ja204053n>.
- (16) Zhang, W.; Ganguly, D.; Chen, J. Residual Structures, Conformational Fluctuations, and Electrostatic Interactions in the Synergistic Folding of Two Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2012**, *8* (1), e1002353. <https://doi.org/10.1371/journal.pcbi.1002353>.
- (17) Knott, M.; Best, R. B. Discriminating Binding Mechanisms of an Intrinsically Disordered Protein via a Multi-State Coarse-Grained Model. *J. Chem. Phys.* **2014**, *140* (17), 175102. <https://doi.org/10.1063/1.4873710>.
- (18) Dogan, J.; Mu, X.; Engström, Å.; Jemth, P. The Transition State Structure for Coupled Binding and Folding of Disordered Protein Domains. *Sci. Rep.* **2013**, *3* (1), 2076. <https://doi.org/10.1038/srep02076>.
- (19) Lee, H.; Mok, K. H.; Muhandiram, R.; Park, K.-H.; Suk, J.-E.; Kim, D.-H.; Chang, J.; Sung, Y. C.; Choi, K. Y.; Han, K.-H. Local Structural Elements in the Mostly Unstructured Transcriptional Activation Domain of Human P53. *J. Biol. Chem.* **2000**, *275* (38), 29426–29432. <https://doi.org/10.1074/jbc.M003107200>.
- (20) Bauer, V. (2020). Doctoral thesis. “*Chemical synthesis of proteins to study the structures and functions of a complex formed by enriched disordered proteins.*” University of Strasbourg, France.

General conclusions

Fluorination is advantageous in biological studies due to the absence of fluorine in the naturally occurring molecules. The use of fluoroprolines in advanced biological studies is justified by many reasons: first prolines are largely present in disordered proteins and polyproline regions are involved in many protein-protein interactions. Second, compared to all other amino acids, proline enhances the population of the *cis* conformer of the peptide bond. Finally, the challenging ^{19}F CSA responsible of NMR signal broadening is relatively small in fluoroprolines. However, the use of fluorinated prolines in advanced biological studies is very limited compared to other aromatic amino acids namely tryptophan. Therefore, in this work, we highlight the potential of 4-fluoroprolines as a fluorine probe to report molecular motion and binding events at the atomic level in polyproline rich peptides, and as structural and dynamical modulator of proline residues properties to uncover their role in a molten globule's function.

On fluoroproline models *4R exo*, *4R endo*, *4S endo* and *4S exo* issued from DFT calculations, a qualitative study of the relaxation was performed through simulations of R_1 , R_2 , heteronuclear H-F NOE and cross relaxation rate σ_{FH} . The effect of the molecular motion of each relaxation parameter was predicted based on derived simple equations from relaxation theory. The agreement of the predictions with those from SPINACH software is very satisfying and reveals that the synergy of the dipolar relaxation and the fluorine chemical shift anisotropy mechanisms is enough to explain the relaxation of 4-fluorinated prolines. A quantitative relaxation study of 4-fluorinated prolines in polyproline sequence context was performed through experimental measurements of the relaxation parameters. The purely dipolar ^1H - ^{19}F heteronuclear cross-relaxation rate is the most reliable parameter to evaluate molecular motion.

4R and *4S* FPro were incorporated in polyproline peptide harboring an SH3 binding motif to determine their effect on the structure and binding affinity. Two model peptides were studied MpSR and MpRS. Fluoroprolines inserted in the fourth and eighth positions are not believed to affect the protein-peptide binding interface. The dynamics of polyproline segments were determined from steady state heteronuclear NOE. Within both peptides the first polyproline segment tumbles faster 0.5ns compared to 0.8ns for the second regardless of the nature of the fluoroproline. The comparison of the heteronuclear coupling constants the two in fluoroprolines in the polyproline sequence and the free amino acids revealed that the

conformational bias is fully preserved. Even though *4S* fluoroproline destabilizes the PPII helix, strong $H\alpha(i-1)$ to $H\delta(i)$ NOE cross-peaks were reported between all prolines and the chemical shift differences between $^{13}C\beta$ and $^{13}C\gamma$ all lied within 5 ppm witnessing the formation of *trans* conformation of the Xaa–Pro peptide bond and the maintain of the overall PPII helix. Binding affinities were determined through NMR titrations. 1H – ^{15}N HSQC spectra and 1D ^{19}F spectra were recorded at each titration point allowing the binding event to be observed from both SH3 and peptides perspectives. Within both peptides, fluorine and 1H – ^{15}N signals went through a gradual shift at each titration point indicating fast exchange regime. The chemical shifts were simultaneously fitted and revealed a 1-to-1 complex stoichiometry with MpSR being the strongest binder to SH3. Compared to the non-fluorinated peptide, the incorporation of *4S* FPro in the binding motif induces a drastic loss of affinity, whereas *4R*FPro maintains it. Finally, the binding kinetics were estimated for MpSR peptide from the line broadening of the fluorine signal during the titration. K_{on} was $0.9 \times 10^8 \pm 0.2 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ and $1.2 \times 10^8 \pm 0.2 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ for *4S* and *4R* respectively. Particular line broadening was recorded for MpRS indicating more complex binding mechanism probably due to the *4S* FPro in the binding motif.

4R and *4S* fluoroprolines were used to elucidate the role of the conformational and dynamical properties of two prolines (P13 and P30) to modulate the structural landscape and the motion of the molten globule NCBD of CBP. In this work, we biophysically characterized the dynamics and the structure of non-modified NCBD and four synthesized fluorinated NCBD variants: NCBD 13R, NCBD 13S, NCBD 30R and NCBD 30S. The fluorination did not alter the global helical secondary structure of the molten globule as observed from CD spectra and the secondary chemical shifts. The determination of the hydrodynamic radii of NCBD from DOSY experiments showed that the globular structure of NCBD is not affected by the insertion of fluoroprolines in its loops. On the residue level, several common side-chain interactions of hydrophobic residues were visible on the NOESY spectra, however the number of these NOE and their intensities varies between the fluorinated NCBDs. The difference of the NOE pattern in NCBD indicates that a site-specific modification of proline's conformation has a global effect on the helix arrangement of NCBD. For each fluorinated NCBD, structure calculations revealed different orientation of the three helices, thus, introducing a conformational bias in NCBD's loop directly affects its conformational landscape. When NCBDs are subject to a heating, their transverse relaxation rate variation displayed unique complex profiles indicator of multiple exchanges at several timescales, making it challenging to extract quantitative exchange parameters. The perturbation of the second loop of NCBD through *4R* FPro incorporation,

systematically leads to its self-assembly into amyloids at neutral pH. The kinetics of this process highly depended on the concentration and the working temperature.

Addressing biological challenges requires a detailed understanding of the structural and dynamical properties of the system. We hope that the present work will encourage more fluorine NMR studies with fluorinated prolines, especially that beside the 4-fluoroprolines here presented, a rich library of synthesized, c_β and c_γ mono or double fluorinated proline analogs are described in literature. The study of polyproline rich sequences interaction to the SH3 domain is very interesting and can be used as a reference to further designed model peptides with tuned properties and within more complex systems. Characterizing the conformational landscape of a NCBD is very important to understand their significant function. Our work will be further completed by deeper understanding of NCBD's interactions with its binding partners. Additional designed variants of fluorinated NCBD can be envisaged either to bring more understanding of its properties or to create molten globules with predictable behavior for specific applications. In continuity of this study, it would be interesting to understand the effect of fluorinated NCBDs on full length CBP properties.

Résumé de thèse

La proline est un acide aminé particulier de nature cyclique. Son cycle peut adopter deux conformères non planaires c_γ *endo* et c_γ *exo* qui s'échangent à une échelle de temps de picoseconde.^{1,2} Lorsque la proline est impliquée dans une liaison peptidique, l'isomère *cis* est significativement peuplé.³ La vitesse d'interconversion *cis-trans* est intrinsèquement lente et se produit à l'échelle de secondes et peut être accélérée par les peptidyl-prolyl *cis-trans* isomérases.⁴⁻⁷ La conformation du cycle, les populations *cis* et *trans* ainsi que la cinétique de ces événements peuvent être biaisées en remplaçant un atome d'hydrogène par un atome de fluor dans la proline. *4R* et *4S* fluoroprolines favorisent la formation des conformères c_γ *exo* et c_γ *endo* et enrichissent les isomères *trans* et *cis* respectivement.^{8,9} Ce travail a deux objectifs principales, le premier est de montrer le potentiel de ces sondes fluorées à reporter d'une façon fiable les propriétés structurales, les mouvements moléculaires ainsi que les événements d'interactions des peptides modèles à l'échelle atomique. Le deuxième est d'exploiter le potentiel de modification des propriétés conformationnelles et/ou dynamiques de la proline pour explorer le lien fonctionnel entre les propriétés de la proline et la fonction protéique. Le noyau ^{19}F est actif en RMN et a une sensibilité intrinsèque proche de celle du ^1H ¹¹, son absence dans les systèmes biologiques rend les spectres facilement interprétables sans interférence de bruit de fond. Au cours de ma thèse, deux systèmes ont été étudiés : le domaine de liaison du coactivateur nucléaire (NCBD) du coactivateur transcriptionnel CBP. Le NCBD est un globule fondu, qui s'agrège en fibres amyloïdes. Le second système est tout un répertoire de peptides modelés de poly proline qui se lient au domaine SH3. Figure 1.

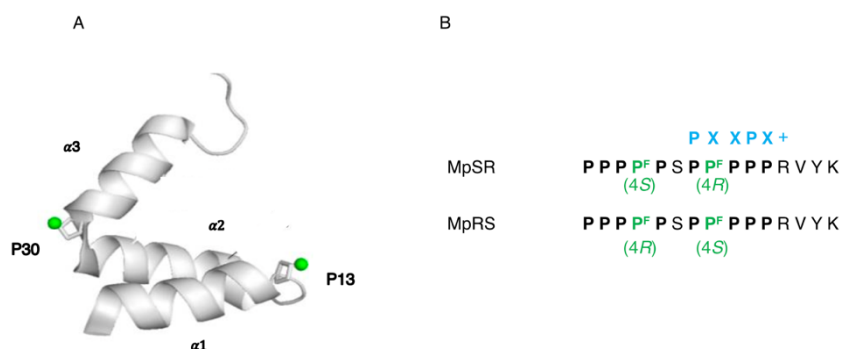


Figure 1 : A : Modèle 3D de NCBD mettant en évidence deux résidus de proline étudiés où Hy est remplacé par un atome de fluor (sphères vertes). B : Fluoroprolines incorporées dans deux séquences riches en proline à deux positions 4 et 8 surlignées en vert. Le motif canonique de liaison au domaine SH3 est représenté en bleu.

Résultats

1- Étude RMN de séquences riches en prolines à l'aide de fluoroprolines (publication)

Dans ce travail, les biais de structure et de dynamique contrôlés au sein des fluoroprolines *4R* et *4S*, ont été appliqués pour étudier l'interaction entre le domaine SH3 et les peptides riches en polyprolines conçus. Les fluoroprolines ont été insérées à deux positions différentes dans le peptide figure 1.B. La dynamique des peptides modèles a été décrite en termes de paramètres de relaxation R_1 , R_2 et de NOE hétéronucléaire. Nous montrons que les vitesses de relaxation longitudinale et transversale ne sont pas appropriées pour sonder sans ambiguïté la dynamique en raison du très faible contraste et de l'échange chimique résiduel. Cependant, l'interprétation du NOE ^1H - ^{19}F hétéronucléaire purement dipolaire a révélé que les deux prolines du peptide subissent une dynamique légèrement différente et que cette dynamique est dictée par le contexte de la séquence plutôt que par la nature de la fluoroproline. Nous rapportons les temps de corrélation de 0.5 ns pour la proline en position 4 et de 0.8 ns pour la proline en position 8. Les affinités de liaison entre le domaine SH3 et les peptides modèles ont été déterminées par des expériences de titrage. Le K_d a été déterminé à partir de la variation du déplacement chimique à la fois du fluor et des $^1\text{H}/^{15}\text{N}$. Cette méthodologie permet de suivre l'interaction du point de vue peptidique et protéique en même temps. Nos données suggèrent que l'incorporation de fluoroproline *4R* dans le site de liaison du domaine SH3 ne modifie pas les affinités, mais une fluoroproline *4S* réduit considérablement l'affinité de liaison. Enfin, la constante cinétique d'association a été déterminée par une analyse de la forme des raies fluor et a suggéré un mécanisme d'association un à un. Au cours de ce travail, Nous avons optimisé plusieurs méthodes standards pour élucider les spécificités dynamiques et conformationnelles des peptides modèles de polyproline. Pour l'attribution sans ambiguïté des différents résidus de proline, les techniques standard telles que ^1H - ^{13}C HSQC et NOESY ne sont pas fiables. Pour surmonter ce défi, nous avons conçu une 2D ^1H - ^{13}C HSQC -NOESY résolue avec une fenêtre carbone étroite de 3 ppm, centrée sur les résonances C δ de la proline. Cela a permis de résoudre les pics croisés séquentiels H δ (i) à H α (i-1) NOE, d'où une attribution complète et réussie du déplacement chimique ^1H et ^{13}C du peptide. L'enquête conformationnelle locale pour distinguer les fluoroprolines *4S* et *4R* a été réalisée en mesurant les constantes de couplage homo et hétéronucléaire $^3\text{JF-H}\beta$, $^3\text{JF-H}\delta$ et $^3\text{JH}\alpha\text{-H}\beta$ dans chaque cycle pyrrolidine par des expériences 2D ^1H - ^1H TOCSY et SERF à haute résolution.

2- Moduler les propriétés d'un « molten globule » avec les fluoroprolines *4R* et *4S*.

Dans ce projet, nous étudions le rôle des résidus de proline pour régler les propriétés du NCBD. Deux prolines se trouvant dans deux environnements chimiques différents sont les sujets d'intérêt de ce travail. Figure 1A. L'isomérisation de la proline 13 est connue pour moduler la cinétique de liaison du NCBD ainsi que son affinité avec son partenaire ACTR de la famille p160.¹⁰ La proline 30 est un résidu initiateur d'hélice et à notre connaissance, n'a pas été étudiée auparavant. Les prolines 13 et 30 ont été remplacées par des fluoroprolines *4R* et *4S*. Par conséquent, une bibliothèque de 5 variants de NCBD est considérée : NCBDwt, NCBD13R, NCBD13S, NCBD30R et NCBD30S. La compacité de la protéine n'est pas altérée par l'incorporation de l'atome de fluor comme révélé par le rayon hydrodynamique inchangé (15 Å) estimé par les expériences 2D ¹H DOSY. De même, la structure globale des NCBD fluorés et leur comportement de dépliement ne sont pas modifiés. Les trois hélices sont conservées au sein des NCBD fluorés comme le révèle d'abord l'examen des déplacements chimiques secondaires et plus tard en effectuant un calcul de structure. Figure 2. La perturbation de la proline 13 induit la formation d'au moins deux variants de NCBD inégalement peuplés et qui coexistent et s'échangent à l'échelle de la milliseconde, cependant la perturbation de la proline 30 révèle une seule population de NCBD. Tous les NCBD fluorés sont en échange avec une forme minoritaire commune caractérisée par des contacts inter hélices modifiés. Cette conclusion est tirée à partir des NOE spécifiques repérés sur les spectres NOESY de chaque NCBD. Parmi les structures calculées de NCBD fluoré, NCBD P13R est la structure la plus proche de la forme majoritaire de NCBD publiée (2KKJ). Malgré les similitudes structurelles globales, les différentes cartes de NOE des NCBD fluorés suggèrent que le noyau hydrophobe se réorganise différemment au sein de chaque variant. Cette observation contredit la description commune de "liquide like" des noyaux hydrophobes de globules fondus. Tous les NCBD fluorés subissent des événements dynamiques uniques, en effet la variation des largeurs de raies des résonances du fluor lorsqu'ils sont soumis à la chaleur est différente et demande encore de travail pour être modélisée. La substitution de la proline 30 par une fluoroproline *4R* induit son réarrangement en fibres amyloïdes. Figure 3. Ce phénomène est dû à la disposition particulière de la première hélice $\alpha 1$ qui expose sa chaîne latérale hydrophobe au solvant favorisant ainsi son interaction avec d'autres $\alpha 1$ d'autres molécules NCBD et conduisant à la formation d'oligomères.

L'interaction de NCBD avec deux de ses principales partenaires d'interaction le domaine TAD du suppresseur du tumeur p53 ainsi que le coactivateur de transcription AD-ACTR de la famille p160, a été suivie en RMN par des expériences de titration et en ITC. Nous montrons que la modification de la proline 13 principalement modifie les affinités de NCBD à ses partenaires et que sa substitution par une fluoroproline *4R* est accompagnée par une amélioration de l'affinité.

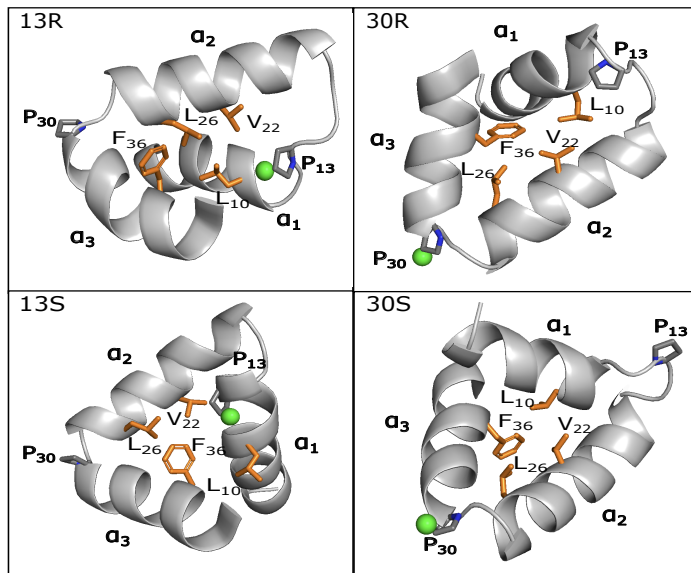


Figure 2 : Modèles 3D de structures de NCBD fluorés calculées à l'aide du logiciel ARIA montrant le réarrangement de leurs noyaux hydrophobes.

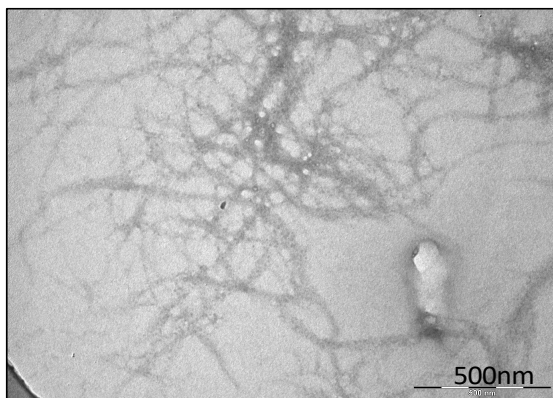


Figure 3 : Micrographies TEM d'échantillons de NCBD 30R dissous dans du tampon phosphate à pH 7.2 après une étape d'évaporation à 327 K et une étape d'incubation à 304 K pendant la nuit.

Conclusions

4*R* et 4*S* fluoroprolines ont un grand potentiel de rapporter les propriétés structurales, et dynamiques des peptides ainsi que leurs événements d'interaction. Du fait qu'elles introduisent un biais dynamique et structurel sur les prolines, elles servent aussi à dévoiler leur rôle joué au sein des systèmes biologiques. Dans ce travail, les fluoroprolines ont été utilisées à la fois comme sonde fluorée pour l'étude d'un répertoire de peptides modèles conçus qui se lient au domaine SH3 et comme un fort modulateur des propriétés des prolines au sein d'un globule fondu atypique NCBD de la protéine CBP. Les paramètres de relaxation de fluor ont été interprétés en termes de mouvement moléculaires. Les déplacements chimiques et largeurs de raies des signaux fluor sont interprétés pour déterminer les constantes de dissociation ainsi que la cinétique des événements d'interaction. Au cours de ces travaux, nous décrivons en détails la relaxation de fluor dans les fluoroprolines 4*R* et 4*S*, les paramètres de son tenseur d'anisotropie de déplacement chimique ainsi que les matrices de distance hétéronucléaire F-H dans les conformères *exo* et *endo* de ces prolines fluorés. Nous espérons que ces données encourageront l'implication des fluoroprolines dans de nombreuses études avancées.

Références

- (1) Ramachandran, G. N.; Lakshminarayanan, A. V.; Balasubramanian, R.; Tegoni, G. Studies on the Conformation of Amino Acids XII. Energy Calculations on Prolyl Residue. *Biochim. Biophys. Acta BBA - Protein Struct.* **1970**, 221 (2), 165–181. [https://doi.org/10.1016/0005-2795\(70\)90257-6](https://doi.org/10.1016/0005-2795(70)90257-6).
- (2) Costantini, N. V.; Ganguly, H. K.; Martin, M. I.; Wenzell, N. A.; Yap, G. P. A.; Zondlo, N. J. The Distinct Conformational Landscapes of 4 *S*-Substituted Prolines That Promote an *Endo* Ring Pucker. *Chem. – Eur. J.* **2019**, 25 (48), 11356–11364. <https://doi.org/10.1002/chem.201902382>.
- (3) Ramachandran, G. N.; Mitra, A. K. An Explanation for the Rare Occurrence of *Cis* Peptide Units in Proteins and Polypeptides. *J. Mol. Biol.* **1976**, 107 (1), 85–92. [https://doi.org/10.1016/S0022-2836\(76\)80019-8](https://doi.org/10.1016/S0022-2836(76)80019-8).
- (4) Grathwohl, C.; Wüthrich, K. Nmr Studies of the Rates of Proline *Cis* - *Trans* Isomerization in Oligopeptides: RATES OF *CIS* - *TRANS* ISOMERIZATION. *Biopolymers* **1981**, 20 (12), 2623–2633. <https://doi.org/10.1002/bip.1981.360201209>.
- (5) Alderson, T. R.; Lee, J. H.; Charlier, C.; Ying, J.; Bax, A. Propensity for *Cis* -Proline Formation in Unfolded Proteins. *ChemBioChem* **2018**, 19 (1), 37–42. <https://doi.org/10.1002/cbic.201700548>.
- (6) London et al. - 1990 - Probing the Role of Proline in Peptide Hormones.Pdf.
- (7) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. Side-Chain Effects on Peptidyl-Prolyl *Cis/Trans* Isomerisation. *J. Mol. Biol.* **1998**, 279 (2), 449–460. <https://doi.org/10.1006/jmbi.1998.1770>.

- (8) Horng, J.-C. Stereoelectronic Effects on Polyproline Conformation. *Protein Sci.* **2006**, *15* (1), 74–83. <https://doi.org/10.1110/ps.051779806>.
- (9) Kim, W.; Hardcastle, K. I.; Conticello, V. P. Fluoroproline Flip-Flop: Regiochemical Reversal of a Stereoelectronic Effect on Peptide and Protein Structures. *Angew. Chem. Int. Ed.* **2006**, *45* (48), 8141–8145. <https://doi.org/10.1002/anie.200603227>.
- (10) Zosel, F.; Mercadante, D.; Nettels, D.; Schuler, B. A Proline Switch Explains Kinetic Heterogeneity in a Coupled Folding and Binding Reaction. *Nat. Commun.* **2018**, *9* (1), 3332. <https://doi.org/10.1038/s41467-018-05725-0>.
- (11) Gerig, J. T. Fluorine NMR. Departement of chemistry.1997. <https://www.biophysics.org/Portals/0/BPSAssets/Articles/gerig.pdf>

Abir Ben Bouzayene

**ÉTUDE PAR RMN DES EFFETS DE
L'INSERTION D'UN ATOME DE
FLUOR SUR LES PROPRIÉTÉS DES
PROLINES AU SEIN DES PROTÉINES**

Résumé

La proline est un acide aminé particulier qui peut adopter deux conformères non planaires *cy endo* et *cy exo* et qui peuple de manière significative l'isomère *cis* comparé à tous les autres acides aminés naturels. La conformation du cycle, les populations *cis* et *trans* ainsi que la cinétique de ces événements peuvent être biaisées en remplaçant un atome d'hydrogène par un atome de fluor dans la proline. *4R* et *4S* fluoroprolines sont utilisées d'abord comme sondes atomiques de la structure, la dynamique et les interactions des peptides suivies par la RMN de fluor et ensuite pour explorer le lien fonctionnel entre les propriétés de la proline et la fonction protéique. Deux systèmes biologiques ont été étudiés, le premier est un peptide modèle riche en poly proline qui se lit au domaine SH3 et le deuxième est un globule fondu NCBD de la protéine CBP. Les variants fluorés de chaque système ont été étudiés en détail dans leurs formes isolées ainsi qu'en complexe avec leurs partenaires de liaison.

Mots clés : Proline, RMN, Fluor, Fluoroproline, NCBD, Globule fondu, Poly proline

Résumé en anglais

Proline is a particular amino acid that can adopt two non-planar conformers *cy endo* and *cy exo*, and that significantly populates the *cis* isomer compared to all other natural amino acids. The ring puckering, *cis* and *trans* populations as well as the kinetics of these events can be biased by replacing a hydrogen atom with a fluorine atom in proline. *4R* and *4S* fluoroprolines are used first as atomic probes of peptide structure, dynamics and interactions followed by fluorine NMR and then to explore the functional link between proline properties and protein function. Two biological systems were studied, the first is a poly proline-rich model peptide that binds to the SH3 domain, the second is the molten globule NCBD of the CBP protein. The fluorinated variants of each system have been studied in detail in their isolated forms as well as in complex with their binding partners.

Keywords : Proline, NMR, Fluorine, Fluoroproline, NCBD, Molten globule, Polypoline