fakultät für chemie
und chemische biologie

technische universität
dortmund

Characterisation of Protein Structure and Dynamics by NMR Spectroscopy and Computational Methods

Dissertation zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

der Fakultät für Chemie und Chemischen Biologie
der Technischen Universität Dortmund

vorgelegt von

Benedikt Söldner

geboren in Rosenheim

November 2025

Diese Dissertation wurde nach erfolgreich bestandener mündlicher Prüfung im Januar 2026 von der Fakultät für Chemie und Chemische Biologie der Technischen Universität Dortmund veröffentlicht.

Tag der Einreichung:	01.10.2025
Erstgutachter:	Prof. Dr. Rasmus Linser
Zweitgutachter:	Prof. Dr. Lars Schäfer
Prüfungsvorsitz:	Prof. Dr. Susanne Brakmann
Tag der mündlichen Prüfung:	24.11.2025

Acknowledgements

After all of the many years as a Ph.D. student – a special stage of life that I will never forget – I want to thank all the people, who guided, supported and accompanied me throughout this journey.

First and foremost, I would like to thank my supervisor Prof. Dr. Rasmus Linser. Your constant support, guidance, and encouragement has been invaluable throughout the entire period. I especially want to thank you for your trust in me, the freedom you gave me in my work and the pleasant working atmosphere. I also want to express my sincere gratitude for allowing me to visit all of the many workshops and conferences worldwide giving me the opportunity to deepen my understanding of NMR spectroscopy and to become acquainted with the scientific community.

My second thanks go to Prof. Dr. Lars Schäfer, not only for being my second reviewer, but also for the fruitful collaboration during the project on the determination of accurate distances from solid-state NMR and his precious advice for the MD simulations for the project on the hCAII–SBR complex.

I want to thank Dr. Kristof Grohe and Peter Neidig for their collaboration with the project on the determination of accurate distances from solid-state NMR. Furthermore, I want to thank Dr. Hosein Geraili Daronkola and Dr. Ana Vila Verde for their collaboration on the project about the influence of salt concentration on protein dynamics. A special thank goes to my former colleague Dr. Himanshu Singh for his support and help with the NMR measurements in the course of the aforementioned project and his collaboration on the PstA project.

I also want to sincerely thank my former colleague Dr. Julia Kotschy for her collaboration on the hCAII–SBR complex. Moreover, I want to thank all of my current and former colleagues Dr. Himanshu Singh, Dr. Suresh Vasa Kumar, Dr. Alexander Klein, Dr. Sara Medina Gómez, Dr. Ekaterina Burakova, Dr. Julia Kotschy, Dr. Laura Kukuk, Hanna Kavaleuskaya, Suchandra Roy Acharyya, Dr. Elnaz Banijamali, Dr. Dmitrii Blokhin, Dr. Florian Lindemann, Dorothea Dirks, Tye Gonzalez, Vanessa Slota, Christine Wehbrink and Nils Schloemp for the nice working atmosphere, all the scientific help and advices as well as for the personal exchange and activities.

Finally, I want to thank my parents, my brother and all other family members for their continuous support and encouragement throughout the challenging time of my study and Ph.D.

Summary

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for studying the structure and dynamics of proteins. In contrast to almost all other experimental techniques, NMR spectroscopy facilitates the elucidation of site-specific protein dynamics on various timescales, making it an indispensable tool for structural biology of proteins. In the first chapter, the theory of NMR spectroscopy is introduced and an overview of frequently used NMR spectroscopic methods for studying protein structure and dynamics is given. In addition, an introduction about molecular dynamics (MD) simulations, a technique for studying protein dynamics on an atomic scale used for explaining the dynamics detected by NMR spectroscopy as well as constituting a technique for structure determination of proteins based on observables from NMR spectroscopy, is given.

In chapter 2 – 5, the four major projects investigated for my Ph.D. are presented. In chapter 2, a newly developed method for determining accurate distances from ^{1}H-detected solid-state NMR spectroscopy is presented and demonstrated by structure determination and restrained MD simulations of the chicken α-spectrin SH3 domain. In the project presented in chapter 3, microsecond-timescale dynamics of a small-molecule ligand bound to the active site of the human carbonic anhydrase II (hCAII) was revealed with solid-state NMR spectroscopy, where my contribution consisted in determining the origin of the dynamics on an atomic level using MD simulations. In the project shown in chapter 4, the influence of salt concentration on the protein dynamics was investigated by NMR spectroscopy and MD simulations. In the project presented in chapter 5, the secondary-messenger-induced allosteric modulation of conformational loop dynamics in the P_{II}-like protein A (PstA) is investigated. In addition to the modulation of the spatial properties of the 30-residue long loops, in absence of the ligand, also slow µs-ms timescale dynamics in the core of PstA are revealed.

Zusammenfassung

Kernresonanzspektroskopie (NMR Spektroskopie) ist eine leistungsstarke Technik zur Untersuchung der Struktur und Dynamik von Proteinen. Im Gegensatz zu fast allen anderen experimentellen Techniken ermöglicht NMR-Spektroskopie die Aufklärung von ortsspezifischer Proteindynamik auf verschiedenen Zeitskalen und ist damit ein unverzichtbares Werkzeug für die Strukturbiologie von Proteinen. Im ersten Kapitel wird die Theorie der NMR-Spektroskopie vorgestellt und ein Überblick über häufig verwendete NMR-spektroskopische Methoden zur Untersuchung der Proteinstruktur und -dynamik gegeben. Darüber hinaus wird eine Einführung in Molekulardynamik-Simulationen (MD Simulationen) gegeben, eine Technik zur Untersuchung der Proteindynamik auf atomarer Ebene, die zur Erklärung der durch NMR-Spektroskopie erfassten Dynamik sowie zur Strukturbestimmung von Proteinen auf der Grundlage von Beobachtungen aus der NMR-Spektroskopie verwendet wird.

In den Kapiteln 2 bis 5 werden die vier Hauptprojekte vorgestellt, die ich für meine Promotion untersucht habe. In Kapitel 2 wird eine neu entwickelte Methode zur Bestimmung akkurater Distanzen aus ^{1}H-detektierter Festkörper-NMR-Spektroskopie vorgestellt und anhand der Strukturbestimmung und durch experimentelle Daten eingeschränkte MD-Simulationen der SH3-Domäne von Hühner-α-Spektrin demonstriert. Im Projekt, das in Kapitel 3 vorgestellt wird, wurde für ein kleines Molekül, das an die aktive Tasche der humanen Carboanhydrase II (hCAII) gebunden ist, mit Festkörper-NMR-Spektroskopie eine Dynamik im Mikrosekundenbereich aufgedeckt. Mein Beitrag an dem Projekt bestand darin, den Ursprung der Dynamik auf atomarer Ebene mit MD-Simulationen zu erklären. Das in Kapitel 4 vorgestellte Projekt handelt von der Untersuchung des Einflusses der Salzkonzentration auf die Proteindynamik mit NMR-Spektroskopie und MD-Simulationen. Im Projekt, das in Kapitel 5 behandelt wird, wurde die durch einen sekundären Botenstoff induzierte allosterische Modulation der Loop-Konformationsdynamik des P_{II}-ähnlichen Signalproteins A (PstA) untersucht. Zusätzlich zur Veränderung der räumlichen Eigenschaften des 30 Reste langen Loops wurde unter Abwesenheit des Liganden eine Dynamik auf der µs-ms Zeitskala im Kern von PstA aufgedeckt.

Abbreviations

AmtB	amphotericin B
ADP	adenosine diphosphate
aMD	accelerated molecular dynamics
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BMRB	Biological Magnetic Resonance Bank
BMRD	Bloch-McConnell relaxation dispersion
BSH-CP	band-selective homonuclear cross polarisation
CCRR	cross-correlated cross-relaxation rate
c-di-AMP	cyclic diadenylate monophosphate
CEST	chemical-exchange saturation transfer
cMD	classical MD
CPMG	Carr–Purcell–Meiboom–Gill
CP	cross polarisation
cryo-EM	cryogenic electron microscopy
CSA	chemical-shift anisotropy
CSI	chemical-shift index
CSP	chemical-shift perturbation
DAC	diadenylyl cyclase
DARR	dipolar-assisted rotational resonance
DEER	electron-electron double resonance
DQ	double-quantum
DRAG	dinitrogenase reductase-activating glycohydrolase
DREAM	dipolar recoupling enhanced by amplitude modulation
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
EXSY	exchange spectroscopy
FID	free induction decay
FRET	Förster resonance energy transfer
hCAII	human carbonic anhydrase II
hetNOE	heteronuclear steady-state NOE
HORROR	double-quantum homonuclear rotary resonance
GaMD	Gaussian-accelerated molecular dynamics
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
INEPT	insensitive nuclei enhancement by polarisation transfer
ITC	isothermal titration calorimetry
lab	laboratory frame

LJ	Lennard-Jones
MAS	magic-angle spinning
MBP	maltose binding protein
MD	molecular dynamics
MDSA	molecular dynamics simulated annealing
ML	maximum likelihood
NAGK	N-acetylglucosamine kinase
NERRD	near-rotary resonance relaxation dispersion
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
NOE	nuclear Overhauser effect
PAS	principal axis system
PDB	protein data bank
PDSD	proton-driven spin diffusion
PME	particle mesh Ewald
PRE	paramagnetic relaxation enhancement
PstA	P_{II}-like signal transduction protein A
RDC	residual dipolar coupling
REDOR	rotational-echo, double-resonance
RFDR	RF-driven dipolar recoupling
RF	radio frequency
RMSD	root mean square deviation
RMSF	root mean square fluctuation
rot	rotor frame
SASA	solvent-accessible surface area
SAXS	small angle X-ray scattering
SH3	α-spectrin SRC homology 3 domain
SPECIFIC-CP	spectrally induced filtering in combination with cross polarisation
ssNMR	solid-state NMR
TCF	time-correlation function
TMS	tetramethylsilane
vdW	van-der-Waals
ZQ	zero-quantum

Contents

1

Introduction

In this chapter, I will give an introduction about the theory of NMR spectroscopy and will give an overview of different methods for studying the structure and dynamics of proteins, which were in part also used for the projects presented in this thesis. Furthermore, I will give a short introduction about molecular-dynamics simulations, a technique for studying the dynamics of macromolecules on an atomistic scale, which was used for explaining the dynamics detected by NMR spectroscopy on an atomic level in this thesis.

1.1 Fundamentals of NMR Spectroscopy[1,2]

Nuclear magnetic resonance (NMR) spectroscopy is an analytical method for observing the effect of electromagnetic pulses on the magnetic moment $\vec{\mu} = \gamma\vec{I}$ of nuclei placed within an external magnetic field. The magnetic moment $\vec{\mu}$ is proportional to the nuclear spin angular momentum $\vec{I}$ with the nucleus-specific gyromagnetic ratio γ as proportionality factor. The spin angular momentum $\vec{I}$ and its z-component depend on the nuclear spin quantum number $I \in \{0, \frac{1}{2}, 1, \frac{3}{2}, ...\}$, which depends on the composition of the atom core, but is not theoretically predictable:

$$|\vec{I}| = \hbar\sqrt{I\,(I+1)} \tag{1.1}$$

$$I_z = \hbar m_I \tag{1.2}$$

Therefore, all nuclei with a spin quantum number $I \neq 0$ possess an angular momentum and a magnetic moment. If such a nucleus is exposed to a static magnetic field with the field strength B_0, the magnetic quantum number $m_I \in \{-I, -I+1, ..., I-1, I\}$ leads to a splitting into $2I+1$ energy levels

$$E = -\mu_z B_0 = -\gamma P_z B_0 = -\gamma\hbar m_I B_0 \tag{1.3}$$

This effect is called Zeeman splitting and the most fundamental NMR spectroscopic interaction. Only transitions between adjacent energy levels are allowed ($\Delta m_I = \pm 1$). Therefore, the resulting energy difference corresponds to

$$\Delta E = \gamma\hbar B_0 \tag{1.4}$$

Nuclei with $I = \frac{1}{2}$ are the most relevant for NMR spectroscopy (especially ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P), because nuclei with higher spin quantum number show very unfavourable effects called quadrupolar interactions leading to strong distortions in the resulting spectra. For spin-$\frac{1}{2}$-nuclei, two energy levels, called α and β, exist. When exposed to an external magnetic field, the two energy levels are unequally populated according to the Boltzmann distribution, which corresponds for a two-state system to

$$\frac{N_\beta}{N_\alpha} = \exp\left(-\frac{\Delta E}{k_B T}\right) = \exp\left(-\frac{\gamma\hbar B_0}{k_B T}\right) \tag{1.5}$$

Since the energy difference is very small, α and β states are nearly equally populated with only a tiny excess of α spins (e.g., 0.010 % for 600 MHz, 0.013 % for 900 MHz, and 0.019 % for 1.2 GHz). Although the difference in populations is very small, it still causes an observable net magnetisation $\vec{M} = \sum_i \vec{\mu}_i$ in z-direction.

According to the Planck-Einstein relation $E = h\nu = \hbar\omega$, the frequency necessary for inducing a transition between adjacent states corresponds to

$$\nu_0 = \frac{\omega_0}{2\pi} = \frac{\gamma}{2\pi} \cdot B_0 \tag{1.6}$$

This frequency is termed Larmor frequency and corresponds to the precession frequency of the nuclear magnetic moment around the z-axis defined along the direction of the magnetic field. An radio frequency (RF) pulse with a field strength in the range of the Larmor frequency applied orthogonally to the B_0, can force the net magnetisation, initially pointing along the B_0-field, to precesses around the additionally irradiated magnetic field and can bring the net magnetisation thereby into the transversal plane. In the transversal plane, different effects occur to the magnetisation, especially, chemical shift dependent evolution of the magnetisation in time, the development of scalar coupling (also called indirect spin-spin coupling or J-coupling) with other nuclei, which can finally be measured.

1.1.1 Chemical Shift

In addition to the magnetic field strength B_0, the Larmor frequency depends on the shielding of the core by the electron density and surrounding atoms. Hence, the effective magnetic field B_{eff} prevailing at the core and the observed frequency ν_{sample} correspond to

$$B_{\text{eff}} = (1 - \sigma) \cdot B_0 \tag{1.7}$$

$$\nu_{\text{sample}} = \frac{\gamma}{2\pi} \cdot B_{\text{eff}} = \frac{\gamma}{2\pi} \cdot (1 - \sigma) \cdot B_0 \tag{1.8}$$

with the dimensionless second-rank shielding tensor σ, which reduces in solutions due to the fast and unrestricted molecular tumbling to a single value (see section 1.3). Practically, the chemical shift δ is commonly used as field-independent alternative to frequencies and measured in *parts per million* (ppm) with respect to a reference substance (e.g., tetramethylsilane (TMS) or 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) for ^{1}H and ^{13}C):

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{ref}}}{\nu_{\text{ref}}} \cdot 10^6 \text{ ppm} \tag{1.9}$$

After referencing of ^{1}H chemical shifts, heteronuclear chemical shifts can also be referenced indirectly.[3] Although chemical shifts are extremely sensitive to the local chemical environment and molecular conformations, an interpretation on an atomic level is barely feasible due the complexity associated with the high sensitivity.

For proteins, the dependency of backbone chemical shifts on the secondary structure is well understood, such that the comparison with protein backbone chemical shifts deposited in databases facilitates a reliable prediction of secondary structures and backbone dihedral angles. The chemical-shift index (CSI)[4] is one of the oldest methods using an easy scoring based on chemical-shift differences with respect to random coil chemical shifts and can be even performed by hand. Much more elaborate programs used for the prediction of protein secondary structure and backbone dihedral angles based on chemical shifts are TALOS[5] and CheSPI.[6] Among others, these programs for example also take chemical shifts of neighbouring residues into account.

1.1.2 Scalar Coupling Interactions[2,7,8]

Scalar coupling (also called indirect dipole-dipole coupling or J-coupling) is a coupling mechanism mediated through covalent bonds and caused by the Fermi contact, an interaction between nuclear spins and the magnetic moments of paired binding electrons, generating a hyperfine splitting in energetic states. Scalar couplings have a predominantly isotropic nature and are characterised by the coupling constant J observed for example as splitting between protons in ^{1}H-spectra. The strength of the scalar coupling is determined by the number of bonds the between coupling nuclei (indicated by a superscript before the coupling constant) and other factors like the type of involved nuclei, valence and torsion angles, bond lengths and the chemical environment caused by neighbouring atoms. J-couplings involving proton and carbon atoms are between 30 - 250 Hz for ^1J, < 20 Hz for ^2J, < 10 Hz for ^3J, and < 1 Hz for J couplings involving more bonds.[7]
^3J(H-H) coupling constants are physically related to dihedral angles according to the Karplus equation[9,10] and enable for example the determination of torsion angles in proteins.[11,12] Furthermore, ^1J couplings are of outstanding importance in pulse sequence elements, like INEPT[13] (insensitive nuclei enhancement by polarisation transfer), because the development of anti-phase magnetisation via ^1J couplings facilitates a magnetisation transfer between covalently bonded nuclei and additionally, due to the high gyromagnetic ratio of ^{1}H, polarisation transfers from ^{1}H to insensitive heteronuclei like ^{13}C or ^{15}N even enhances the NMR signal by a factor of $\gamma_{\mathrm{H}}/\gamma_{\mathrm{C/N}}$. Moreover, relaxation delays between successive experiments depend now on T_1 relaxation times of ^{1}H, which are much shorter than those for ^{13}C and ^{15}N, such that acquisition can be accomplished in shorter times.

1.1.3 Dipolar Coupling Interactions

Dipolar or direct dipole-dipole coupling interactions result from through-space interactions between the small local magnetic fields generated by nuclear spins, similar to the interaction between two bar magnets adjacent to each other. The dipolar coupling interaction strength depends on the angle between the inter-spin vector and the external magnetic field (see section 1.2.3) as well as the dipolar coupling constant

$$d_{IS} = -\frac{\mu_0}{4\pi}\frac{\gamma_I \gamma_S \hbar}{r_{IS}^3} \tag{1.10}$$

where μ_0 is the vacuum permeability, $\gamma_{I/S}$ the gyromagnetic ratio of spin I and S, $\hbar$ the reduced Planck constant and r_{IS} the inter-spin distance. The nuclear Overhauser effect (NOE), a relaxation mechanism via direct dipole-dipole interactions, facilitates a through-space magnetisation transfer between adjacent spins and can be detected using nuclear Overhauser effect spectroscopy (NOESY) experiments. The distance dependence of the dipolar coupling strength can be exploited for the determination of internuclear distances, which can be in return used for structure determination.

1.2 Theoretical Background[1,2,14–19]

1.2.1 Quantum Mechanical Representation of Spins

The Hamiltonians describing all kinds of interaction in NMR spectroscopy, e.g., the Zeeman interaction $\hat{H}_z = -\gamma \hat{I}_z B_0$ or the scalar coupling interaction $\hat{H}_J = 2\pi J_{IS} \hat{I}_z \hat{S}_z$, all depend on the spin angular momentum operators $\hat{I}$ and $\hat{S}$ for spins I and S. The eigenvalues of the spin angular momentum operators $\hat{I}_{x/y/z}$ acting on spin-$\frac{1}{2}$-nuclei with spin states α and β along all three axes are

$$\begin{aligned} \hat{I}_x |\alpha\rangle &= +\frac{1}{2} |\beta\rangle & \hat{I}_y |\alpha\rangle &= +\frac{1}{2} i |\beta\rangle & \hat{I}_z |\alpha\rangle &= +\frac{1}{2} |\alpha\rangle \\ \hat{I}_x |\beta\rangle &= +\frac{1}{2} |\alpha\rangle & \hat{I}_y |\beta\rangle &= -\frac{1}{2} i |\alpha\rangle & \hat{I}_z |\beta\rangle &= -\frac{1}{2} |\beta\rangle \end{aligned} \tag{1.11}$$

In quantum mechanics, wave functions can be represented as vectors and operators as matrices. The overall wavefunction of a system can be either described as linear combination $|\Psi\rangle = \sum_i c_i |\psi_i\rangle$ of its eigenstates $|\psi_i\rangle$ constituting an orthonormal set of eigenfunctions ψ_i, or alternatively, as a vector defined by the coefficients c_i as vector components. For spin-$\frac{1}{2}$-nuclei, the overall wavefunction with eigenstates $|\alpha\rangle$ and $|\beta\rangle$ corresponds to

$$|\Psi\rangle = \sum_i c_i |\psi_i\rangle = c_1 |\alpha\rangle + c_2 |\beta\rangle = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \langle \alpha \,|\, \Psi \rangle \\ \langle \beta \,|\, \Psi \rangle \end{pmatrix} \Rightarrow |\alpha\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \,\&\, |\beta\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.12}$$

and its hermitian to

$$\begin{gathered} \langle \Psi | = \sum_i c_i^* \langle \psi_i | = c_1^* \langle \alpha | + c_2^* \langle \beta | = \begin{pmatrix} c_1 & c_2 \end{pmatrix} = \begin{pmatrix} \langle \alpha \,|\, \Psi \rangle \\ \langle \beta \,|\, \Psi \rangle \end{pmatrix}^\dagger = \begin{pmatrix} \langle \Psi \,|\, \alpha \rangle & \langle \Psi \,|\, \beta \rangle \end{pmatrix} \\ \Rightarrow \langle \alpha | = \begin{pmatrix} 1 & 0 \end{pmatrix} \,\&\, \langle \beta | = \begin{pmatrix} 0 & 1 \end{pmatrix} \end{gathered} \tag{1.13}$$

The matrix representation of an operator $\hat{A}$ is constructed from the matrix components $A_{ij} = \left\langle \psi_i \middle| \hat{A} \middle| \psi_j \right\rangle$. Therefore, the spin angular momentum operators $\hat{I}_{x/y/z}$ forming (together with the identity operator $\hat{E}$) an orthogonal basis set for a single spin-$\frac{1}{2}$-spin system with the eigenstates $\psi_i \in \{|\alpha\rangle, |\beta\rangle\}$ are represented by the matrices

$$\hat{I}_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \hat{I}_y = \frac{1}{2} i \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad \hat{I}_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.14}$$

These matrices are called Pauli matrices and obey the cyclic commutation relations

$$\left[\hat{I}_x, \hat{I}_y\right] = i\hat{I}_z \qquad \left[\hat{I}_y, \hat{I}_z\right] = i\hat{I}_x \qquad \left[\hat{I}_z, \hat{I}_x\right] = i\hat{I}_y \tag{1.15}$$

Eigenstates and operators describing multi-spin systems are constructed from those of spin-$\frac{1}{2}$ systems using the tensor product $\otimes$ (also called Kronecker product). The multi-spin system constructed of two operators $\hat{A}$ and $\hat{B}$ is therefore defined by

$$\hat{A} \otimes \hat{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \otimes \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}\hat{B} & a_{12}\hat{B} \\ a_{21}\hat{B} & a_{22}\hat{B} \end{pmatrix} \tag{1.16}$$

1.2.2 Density Matrix Theory and Product Operator Formalism

The matrix representation of an operator defined as $\hat{\sigma} = |\Psi\rangle\langle\Psi|$ (also often described as $\hat{\rho}$) consists, according to the definition in the previous section, of the matrix elements

$$\begin{aligned}\sigma_{ji} &= \langle\psi_j \,|\, \Psi\rangle\langle\Psi \,|\, \psi_i\rangle = \left\langle \psi_j \middle| \sum\nolimits_l c_l \middle| \psi_l \right\rangle \sum\nolimits_m c_m^* \langle\psi_m \,|\, \psi_i\rangle \\ &= \sum\nolimits_l c_l \delta_{jl} \sum\nolimits_m c_m^* \delta_{mi} = c_j c_i^*\end{aligned} \tag{1.17}$$

The operator $\hat{\sigma}$ is termed density operator and has the special property that the expectation value of an operator $\hat{A}$ can be extracted from it:

$$\begin{aligned}\left\langle \hat{A} \right\rangle &= \left\langle \psi \middle| \hat{A} \middle| \psi \right\rangle = \sum\nolimits_i c_i^* \left\langle \psi_i \middle| \hat{A} \sum\nolimits_j c_j \middle| \psi_j \right\rangle \\ &= \sum\nolimits_i \sum\nolimits_j c_i^* c_j A_{ij} = \sum\nolimits_i \sum\nolimits_j \sigma_{ji} A_{ij} = \mathrm{Tr}(\hat{\sigma}\hat{A})\end{aligned} \tag{1.18}$$

This holds not only true for a single spin-$\frac{1}{2}$ system (called *pure state*) as in the equations above, but also for an ensemble of spins (called *mixed state*), for which the density operator is defined as

$$\hat{\sigma} = N^{-1}\left(|\Psi_1\rangle\langle\Psi_1| + |\Psi_2\rangle\langle\Psi_2| + ... + |\Psi_N\rangle\langle\Psi_N|\right) = \overline{|\Psi\rangle\langle\Psi|} \tag{1.19}$$

where N is the number of ensemble members. For phase-sensitive signal detection in NMR spectroscopy (quadrature detection), the signal (called free induction decay, FID), is measured in both x- and y-dimensions using two channels, so that the measurement operator (denoted as $\hat{A}$ in equation 1.18) corresponds to the raising operator $\hat{I}^+$, which is, analogous to the lowering operator $\hat{I}^-$, constructed from the spin angular momentum operators $\hat{I}_{x/y}$ in the transversal plane as $\hat{I}^\pm = \hat{I}_x \pm i\hat{I}_y$. The final spectrum in frequency domain is obtained by Fourier transformation of the FID in time domain

$$S(\omega) = \mathrm{Re}\left\{\int_{-\infty}^{+\infty} S(t)e^{-i\omega t}dt\right\} = \mathrm{Re}\left\{2\int_0^{+\infty} S(t)e^{-i\omega t}dt\right\} \tag{1.20}$$

with signal S. The evolution of the density operator $\hat{\sigma}$ in time is derived from the time-dependent Schrödinger equation and described by the Liouville-von Neumann equation:

$$\frac{\partial}{\partial t}\hat{\sigma}(t) = \frac{\partial\,|\Psi(t)\rangle}{\partial t}\langle\Psi(t)| + |\Psi(t)\rangle\frac{\partial\,\langle\Psi(t)|}{\partial t} = i\left[\hat{H}, \hat{\sigma}(t)\right] \tag{1.21}$$

where $\frac{\partial}{\partial t}|\Psi(t)\rangle = -i\hat{H}\,|\Psi(t)\rangle$ corresponds to the time-dependent Schrödinger equation and $\frac{\partial}{\partial t}\langle\Psi(t)| = i\,\langle\Psi(t)|\,\hat{H}$ to its hermitian. If $\hat{H}$ is time-independent, the Liouville-van Neumann equation can be solved as

$$\hat{\sigma}(t) = e^{-i\hat{H}t}\hat{\sigma}(0)e^{i\hat{H}t} \tag{1.22}$$

The density operator $\hat{\sigma}$ can be expressed as any linear combination of an orthogonal set of basis operators (e.g., $\hat{E}$, $\hat{I}_x$, $\hat{I}_y$, and $\hat{I}_z$ for a spin-$\frac{1}{2}$ system). If the Hamiltonian $\hat{H}$ can be defined as $\hat{H} \equiv \phi\hat{A}$ and $\hat{\sigma}$ as $\hat{\sigma} \equiv \hat{B}$ such that an operator $\hat{C}$ exists, for which these operators obey the cyclic commutation relations $\left[\hat{A}, \hat{B}\right] = i\hat{C}$ (and its counterparts), the following transformation for an operator $\hat{A}$ acting on operator $\hat{B}$ exists:

$$\hat{B} \xrightarrow{\phi\hat{A}} e^{-i\phi\hat{A}}\hat{B}e^{+i\phi\hat{A}} = \hat{B}\cos\phi - i[\hat{A},\hat{B}]\sin\phi = \hat{B}\cos\phi + \hat{C}\sin\phi \tag{1.23}$$

i.e., operators $\hat{B}$ and $\hat{C}$ change over time with phase $\phi = \omega t$. In NMR spectroscopy, such transformations are frequently used for the description of magnetisation evolution in pulse sequences. This simplified description is termed product operator formalism. Examples for typical transformation are the application of a y-pulse on z-magnetisation about angle $\theta = \omega_{\mathrm{rf}} t$, chemical-shift evolution of $\hat{I}_x$ or the evolution of scalar coupling of $\hat{I}_x$ with a second spin S:

$$\hat{I}_z \xrightarrow{\theta\hat{I}_y} \hat{I}_z\cos\theta + \hat{I}_x\sin\theta \tag{1.24}$$

$$\hat{I}_x \xrightarrow{(\Omega_I t)\hat{I}_z} \hat{I}_x\cos(\Omega_I t) + \hat{I}_y\sin(\Omega_I t) \tag{1.25}$$

$$\hat{I}_x \xrightarrow{(\pi J_{IS} t)2\hat{I}_z\hat{S}_z} \hat{I}_x\cos(\pi J_{IS} t) + 2\hat{I}_y\hat{S}_z\sin(\pi J_{IS} t) \tag{1.26}$$

1.2.3 Hamiltonians in NMR Spectroscopy

In NMR spectroscopy, several components contribute to the total Hamiltonian. First of all, the total Hamiltonian can be divided into internal and external spin interactions

$$\hat{H} = \hat{H}_{\mathrm{ext}} + \hat{H}_{\mathrm{int}} \tag{1.27}$$

On the left side of the equation,

$$\hat{H}_{\mathrm{ext}} = \hat{H}_{\mathrm{Z}} + \hat{H}_{\mathrm{rf}} = \omega_0\hat{I}_z + 2\omega_{\mathrm{rf}}\hat{I}_{x/y}\cos(\omega_{\mathrm{rf}} t) \tag{1.28}$$

is the Hamiltonian describing spin interactions with the static external field aligned along the z-axis (the so-called Zeeman interactions denoted by $\hat{H}_{\mathrm{Z}}$) and spin interaction with radio-frequency pulses along the transversal plane (x- or y-axis). On the right side

$$\hat{H}_{\mathrm{int}} = \hat{H}_{\mathrm{J}} + \hat{H}_{\mathrm{CS}} + \hat{H}_{\mathrm{D}} + \hat{H}_{\mathrm{Q}} \tag{1.29}$$

is the internal Hamiltonian comprising the Hamiltonians of scalar coupling $\hat{H}_{\mathrm{J}}$, chemical shift $\hat{H}_{\mathrm{CS}}$, dipolar coupling $\hat{H}_{\mathrm{D}}$, and quadrupolar couplings $\hat{H}_{\mathrm{Q}}$. For spin-$\frac{1}{2}$ nuclei, all relevant Hamiltonians can be either represented by a 3×3 matrices of the form

$$\hat{H} = C_\mu \cdot \vec{I} \cdot \hat{A} \cdot \vec{K} = C_\mu \begin{pmatrix} \hat{I}_x & \hat{I}_y & \hat{I}_z \end{pmatrix} \begin{pmatrix} a_{xx} & a_{xy} & a_{xz} \\ a_{yx} & a_{yy} & a_{yz} \\ a_{zx} & a_{zy} & a_{zz} \end{pmatrix} \begin{pmatrix} \hat{K}_x \\ \hat{K}_y \\ \hat{K}_z \end{pmatrix} \tag{1.30}$$

where C_μ is a constant depending on the type of interaction μ ($C_{\mathrm{Z}} = -\gamma_I$, $C_{\mathrm{CS}} = \gamma_I$, $C_{\mathrm{D}} = -2 \cdot \hbar\mu_0\gamma_I\gamma_S/4\pi$), $\vec{I}$ is the vector of spin operators, $\hat{A}$ is the interaction tensor, and $\vec{K}$ contains the operators of another spin or the magnetic field. Alternatively, Hamiltonians can be written in terms of spherical harmonics

$$\hat{H} = C_\mu \sum_{l=0}^{2} \sum_{m=-l}^{+l} (-1)^m A_{l,m}\hat{T}_{l,-m} \tag{1.31}$$

where l is the rank, m the component, $A_{l,m}$ is the spacial part, and $\hat{T}_{l,-m}$ is the spin operator (listed in Table 1.1). The isotropic components correspond to $l = 0$, the anti-symmetric anisotropic components to $l = 1$ and the symmetric anisotropic components to $l = 2$. Anti-symmetric components do not influence NMR spectra directly but affect relaxation.

Table 1.1: Spherical tensor representation of spin operators $T_{l,-m}$.[15]

	Shielding	Spin-Spin
$T_{0,0}$	$-\frac{1}{\sqrt{3}}\hat{I}_z B_0$	$-\frac{1}{\sqrt{3}}\vec{I}\cdot\vec{S}$
$T_{1,0}$	0	$-\frac{1}{2\sqrt{2}}\left(\hat{I}^+\hat{S}^- - \hat{I}^-\hat{S}^+\right)$
$T_{1,\pm1}$	$-\frac{1}{2}\hat{I}^\pm B_0$	$-\frac{1}{2}\left(\hat{I}_z\hat{S}^\pm - \hat{I}^\pm\hat{S}_z\right)$
$T_{2,0}$	$\sqrt{\frac{2}{3}}\hat{I}_z B_0$	$-\frac{1}{\sqrt{6}}\left(3\hat{I}_z\hat{S}_z - \vec{I}\cdot\vec{S}\right)$
$T_{2,\pm1}$	$\mp\frac{1}{2}\hat{I}^\pm B_0$	$\mp\frac{1}{2}\left(\hat{I}_z\hat{S}^\pm + \hat{I}^\pm\hat{S}_z\right)$
$T_{2,\pm2}$	0	$\frac{1}{2}\left(\hat{I}^\pm\hat{S}^\pm\right)$

If the interaction tensor $\hat{A}$ is given in the principal axis system (PAS), i.e., if the symmetric component of the matrix representation of $\hat{A}$ is diagonal, the isotropic part a_{iso}, the anisotropy δ and the asymmetry η can be calculated as

$$a_{\text{iso}} = \frac{1}{3}\left(a_{xx} + a_{yy} + a_{zz}\right) \tag{1.32}$$

$$\delta = a_{zz} - a_{iso} \tag{1.33}$$

$$\eta = \frac{a_{yy} - a_{xx}}{\delta} \tag{1.34}$$

Depending on the type of interaction, $\delta_{\text{CS}} = \frac{2}{3}\Delta\sigma$ for the chemical shift tensor with the chemical-shift anisotropy (CSA) $\Delta\sigma$ or $\delta_{\text{D}} = r_{IS}^{-3}$ in case of dipolar coupling interactions. Based on these values, the spherical components $A_{l,m}^{\text{PAS}}$ can be calculated as

$$A_{0,0}^{\text{PAS}} = -\sqrt{3}a_{iso}$$

$$A_{1,0}^{\text{PAS}} = -i\sqrt{2}a_{xy} \qquad A_{1,\pm1}^{\text{PAS}} = a_{xz} \pm i a_{yz} \tag{1.35}$$

$$A_{2,0}^{\text{PAS}} = \sqrt{\frac{3}{2}}\delta \qquad A_{2,\pm1}^{\text{PAS}} = 0 \qquad A_{2,\pm2}^{\text{PAS}} = -\frac{1}{2}\delta\eta$$

Back-transformation of $A_{l,m}^{\text{PAS}}$ into the laboratory frame (lab) (or in the case of magic-angle spinning solid-state NMR (ssNMR) first into the rotor frame and progressively into the laboratory frame with $D^l_{m',m}(\omega_{\text{R}}t, \beta_{\text{R}}, 0)$, where β_{R} is the angle between the external field and the rotor axis) is possible by active rotation around the three Euler angles α, β, and γ using either the rotation matrix $R(\alpha\beta\gamma) = R_z(\alpha)R_y(\beta)R_z(\gamma)$ for the matrix representation of $A_{l,m}$ or in terms of spherical harmonics using the Wigner rotation matrix $D^l_{m',m}$ with the

reduced rotation matrices $d^l_{m',m}(\beta)$ where m' and m denote the old and new components:

$$\begin{aligned} A^{\text{lab}}_{l,m} &= R(\alpha\beta\gamma) A^{\text{PAS}}_{l,m'} R^{-1}(\alpha\beta\gamma) \\ &= \sum_{m'=-l}^{+l} D^l_{m',m}(\alpha,\beta,\gamma)\, A^{\text{PAS}}_{l,m'} = \sum_{m'=-l}^{+l} e^{i\alpha m'} d^l_{m',m}(\beta) e^{-i\gamma m} A^{\text{PAS}}_{l,m'} \end{aligned} \tag{1.36}$$

The Hamiltonian $\hat{H}$ in spherical tensor operator notation contains time-dependent contributions $\exp(-im\omega_0 t)$ (with the angular Larmor frequency ω_0). At high magnetic fields, all terms with $m \neq 0$ (i.e., time-dependent terms) average to 0 over time due to the fast spin rotation and can therefore be neglected. Only time-independent contributions of the internal Hamiltonian (called secular terms with $m = 0$) are preserved. This approximation is called secular or high-field approximation.
Under consideration of the secular approximation and neglecting antisymmetric anisotropic components ($l = 1$), the chemical-shift Hamiltonian derives from spherical harmonics to

$$\begin{aligned} \hat{H}^{\text{lab}}_{\text{CS}} &= C_{\text{CS}} \cdot \left(A^{\text{lab}}_{0,0} \hat{T}_{0,0} + A^{\text{lab}}_{2,0} \hat{T}_{2,0} \right) \\ &= C_{\text{CS}} \cdot \left[A^{\text{lab}}_{0,0} \hat{T}_{0,0} + \sum_{m'=-2}^{2} e^{i\alpha m'} d^2_{m',0}(\beta) A^{\text{PAS}}_{2,m'} \hat{T}_{2,0} \right] \\ &= \omega_0 \hat{I}_z \cdot \left[\sigma_{\text{iso}} + \frac{\delta_{\text{CSA}}}{2} \left(3\cos^2\theta - 1 - \eta \sin^2\theta \cos 2\phi \right) \right] \end{aligned} \tag{1.37}$$

where $C_{\text{CS}} = \gamma$, $\delta_{\text{CSA}} = \frac{2}{3}\Delta\sigma$ with the chemical shift anisotropy $\Delta\sigma$, $\theta = -\beta$ is the angle between the z-axis of the chemical-shift tensor and the B_0 field, and $\phi = -\alpha$ is the azimuth angle. The change in sign is a result of the polar angles θ and ϕ being defined by passive instead of active rotations. In contrast to the chemical shift, the dipolar coupling tensor is symmetric ($\eta = 0$) and purely anisotropic (i.e., traceless).

$$\begin{aligned} H^{\text{lab}}_{\text{DD}} &= C_{\text{D}} \cdot \sum_{m=-2}^{+2} (-1)^m A^{\text{lab}}_{2,m} \hat{T}_{2,-m} \\ &= C_{\text{D}} \cdot \left[A_{2,0}\hat{T}_{2,0} - A_{2,-1}\hat{T}_{2,1} - A_{2,1}\hat{T}_{2,-1} + A_{2,-2}\hat{T}_{2,2} + A_{2,2}\hat{T}_{2,-2} \right] \\ &= -\frac{\hbar\mu_0\gamma_I\gamma_S}{4\pi r^3_{IS}} \cdot \left[(\hat{A} + \hat{B}) + \hat{C} + \hat{D} + \hat{E} + \hat{F} \right] \end{aligned} \tag{1.38}$$

with $C_{\text{D}} = -2\hbar\mu_0\gamma_I\gamma_S/4\pi$, $\delta_{\text{DD}} = r^{-3}_{IS}$, the polar angles $\theta = -\beta$ and $\phi = -\gamma$ for passive rotation, and the components of the dipolar alphabet

$$\begin{aligned} \hat{A} &= \left(3\cos^2\theta - 1\right) \cdot \hat{I}_z \hat{S}_z & \hat{B} &= -\frac{1}{4}\left(3\cos^2\theta - 1\right) \cdot \left[\hat{I}^+\hat{S}^- + \hat{I}^-\hat{S}^+\right] \\ \hat{C} &= \frac{3}{2}\sin\theta\cos\theta e^{-i\phi} \cdot \left[\hat{I}^z\hat{S}^+ + \hat{I}^+\hat{S}^z\right] & \hat{D} &= \frac{3}{2}\sin\theta\cos\theta e^{+i\phi} \cdot \left[\hat{I}_z\hat{S}^- + \hat{I}^-\hat{S}_z\right] \\ \hat{E} &= \frac{3}{4}\sin^2\theta e^{-2i\phi} \cdot \hat{I}^+\hat{S}^+ & \hat{F} &= \frac{3}{4}\sin^2\theta e^{+2i\phi} \cdot \hat{I}^-\hat{S}^- \end{aligned} \tag{1.39}$$

According to the high-field approximation, only the terms $\hat{A}$ and $\hat{B}$ need to be considered for homonuclear dipolar couplings and only term $\hat{A}$ for heteronuclear dipolar couplings.

1.3 Solid-State NMR Spectroscopy[14,15,19–21]

1.3.1 Magic-Angle Spinning (MAS)

Solid-state NMR spectroscopy facilitates the measurement of various materials from crystalline solids to disordered materials like amorphous materials, gels, glasses or polymers. In the following, only micro-crystalline solids, which are typically used for studying proteins, are considered. In contrast to isotropic liquids, anisotropic (i.e., orientation-dependent) interactions are not averaged out by fast molecular tumbling leading to strong distortions with very broad lines in the resulting spectra, called *powder pattern*. Linewidths are on the order of several kHz for ^{13}C and ^{15}N CSA and dipolar interactions up to several tens of kHz for dipolar-coupling interactions involving ^{1}H nuclei.

Magic-angle spinning (MAS) is a routinely used method for alleviating the effects of anisotropic interactions, where the sample is rapidly spun about a specific angle β_R with respect to the magnetic field, which suppresses components of zeroth order of the second rank Hamiltonians, as these components cover the strongest anisotropic interactions at high magnetic field strengths. Higher order components (containing e.g., the usually negligibly small asymmetry η) are also affected by MAS, but not averaged out. Under MAS conditions, the transformation of the dipolar or chemical-shift Hamiltonian from principal axis system (PAS) to the laboratory frame (lab) can be separated in two different transformations, where the PAS is first transformed to the rotor frame (rot) and second from rotor frame to the laboratory frame:

$$\begin{aligned} D^{l=2}_{m=0,m'=0}(\alpha_{\mathrm{lab}},\beta_{\mathrm{lab}},\gamma_{\mathrm{lab}}) &= D^{2}_{0,0}(\alpha_{\mathrm{rot}},\beta_{\mathrm{rot}},\gamma_{\mathrm{rot}})\cdot D^{2}_{0,0}(\omega_{\mathrm{R}}t,\beta_{\mathrm{R}},0) \\ &= \frac{1}{2}\left(3\cos^2\beta_{\mathrm{rot}}-1\right)\cdot\frac{1}{2}\left(3\cos^2\beta_{\mathrm{R}}-1\right) \end{aligned} \tag{1.40}$$

where ω_{R} denotes the MAS spinning speed. If β_{R} is set to $\beta_{\mathrm{R}} = \arccos\left(\sqrt{1/3}\right) \approx 54.74°$, the so-called *magic angle*, and the spinning rate is sufficiently fast compared to the strength of the anisotropic interactions, the time-average of the orientation dependence vanishes. If MAS rates are not fast enough, spinning sidebands separated by multiples of the spinning frequency from each other and the isotropic chemical shift appear and linewidths are still slightly broadened. While spinning rates of 5 kHz are sufficient for averaging ^{13}C and ^{15}N CSA and dipolar interactions, for obtaining narrow linewidths in proton-detected ssNMR, high spinning rates, optimally in combination with partial deuteration (e.g., ^{2}H-labelling and ^{1}H back-exchange,[22,23] methyl labelling schemes[24–26]), are required. While for fully protonated samples MAS rates beyond 300 kHz would be theoretically be required to obtain solution-like lineshapes,[27] at the moment, spinning rates up to 200 kHz are experimentally achievable,[21] while routinely used MAS rates for proton-detected NMR go up to 160 kHz.[28–30]

1.3.2 Magnetisation Transfer in Solid-State NMR[31]

In solution state NMR, magnetisation transfers are usually either accomplished via covalent bonds using insensitive nuclei enhancement by polarisation transfer (INEPT)-based experiments or through space using NOESY experiments. In solid-state NMR, however, ^{1}H transverse relaxation times are extremely fast for rigid solids due to the strong ^{1}H-^{1}H dipolar couplings, such that the signal would fade away before scalar couplings can evolve sufficiently.[32–35] Thus, INEPT is usually either only used as a complementary method facilitating the detection of flexible regions or for magnetisation transfers between weakly coupled spins, like ^{13}C.[23,36] Cross polarisation (CP)[37,38] is an alternative solid-state NMR technique, where simultaneous RF irradiation on two nuclei can provoke zero-quantum (ZQ) and double-quantum (DQ) transitions if an energy matching is achieved by adjusting the RF field strengths B_1^I and B_1^S correctly. Without sample spinning, ZQ transitions can be accomplished according to the Hartmann-Hahn condition

$$\gamma_I B_1^I = \gamma_S B_1^S \tag{1.41}$$

Under MAS conditions, additional sidebands at $\pm n\nu_\mathrm{r}$ with $n \in \mathbb{N}$ appear. In addition, DQ transitions can occur for

$$\gamma_I B_1^I = -\gamma_S B_1^S \pm n\nu_\mathrm{r} \tag{1.42}$$

Optimal transfer efficiency can be achieved using shaped RF pulses, where the amplitude of at least one of both RF pulses is modulated.[39,40] Like INEPT, CP also provides a signal enhancement for insensitive spins (like ^{13}C or ^{15}N) if magnetisation is initially derived from ^{1}H and faster proton T_1 times shorten measurement times of successive acquisitions.[41] Although dipolar couplings are averaged out by magic-angle spinning, they can be selectively reintroduced (e.g., using CP-based experiments with simultaneous RF irradiation or by application of rotor-synchronised RF pulses) and used for magnetisation transfer in solid-state NMR. This process is referred to as *dipolar recoupling*. A variety of dipolar recoupling sequences has been developed in the last decades,[42] where among many others PDSD[43] (proton-driven spin diffusion), REDOR[44,45] (rotational-echo, double-resonance), RFDR[46,47] (RF-driven dipolar recoupling), HORROR[48] (double-quantum homonuclear rotary resonance), DREAM[49] (dipolar recoupling enhanced by amplitude modulation), DARR[50,51] (dipolar-assisted rotational resonance), SPECIFIC-CP[52] (spectrally induced filtering in combination with cross polarisation) and BSH-CP[53,54] (band-selective homonuclear cross polarisation) are commonly used examples.

1.4 Relaxation and Nuclear Overhauser Effect[1,55,56]

1.4.1 Relaxation for the Study of Protein Dynamics

The majority of all currently known protein structures deposited in the protein data bank (PDB) were determined via X-ray crystallography or cryogenic electron microscopy (cryo-EM). Although, these methods allow an accurate determination of the three-dimensional structure of proteins and other biomolecules, they are unable to capture flexible regions, protein dynamics or the underlying conformational ensemble. The elucidation and quantification of protein dynamics is, however, of great biological interest, because the dynamics of proteins is directly coupled to their biological function. NMR spectroscopy allows on the one hand the determination such conformational ensembles of small to medium sized proteins (up to $\approx 50\,\mathrm{kDa}$) using ensemble-averaged structure calculation based on NMR restraints and on the other hand – much more important and less cumbersome – even a quantitative assessment of protein dynamics on a wide range of timescales capturing various different biological processes. The latter is facilitated by relaxation studies using a set of complementary NMR experiments enabling, among others, the determination of exchange rates, populations, and the extent of order and disorder.

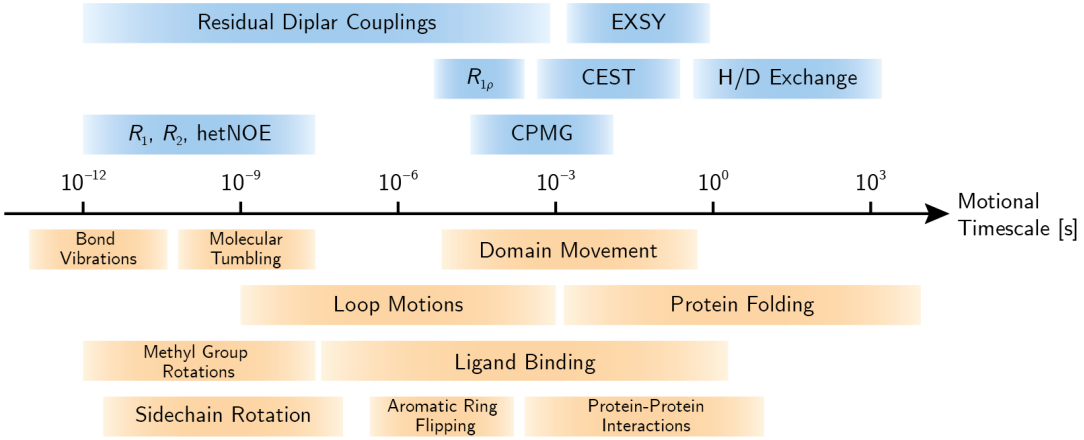

Figure 1.1: Timescales for different protein motions and NMR experiments for observing protein dynamics in respective time regimes.[57]

1.4.2 Spin-Lattice and Spin-Spin Relaxation

Relaxation describes the process of bulk magnetisation excited by RF irradiation returning back to the equilibrium and is primarily caused by dipolar interactions for ^{1}H and additionally by CSA interactions for ^{13}C and ^{15}N. For dynamics studies in biomolecular NMR, usually relaxation of heteronuclei like ^{13}C and ^{15}N, is investigated, because relaxation is mostly determined by dipolar interactions with the directly bound proton. For ^{1}H, in contrast, interactions to the numerous surrounding protons all contribute to the relaxation process, which hampers the interpretation of observed relaxation not directly reflecting a specific, locally-restricted type of dynamics. The two most fundamental types of relaxation

are spin-lattice and spin-spin relaxation. Spin-lattice relaxation (also called longitudinal relaxation) is an enthalpy driven process describing the relaxation of bulk magnetisation along the direction of the magnetic field (defined as z axis) and is quantified by the longitudinal relaxation rate R_1 (or time T_1). In contrast, spin-spin relaxation is an entropy driven process describing the dephasing of magnetisation coherently precessing in the transversal plane and is quantified by the transversal relaxation rate R_2 (or time T_2). The time-dependent decline of magnetisation due to longitudinal and transversal relaxation was already described in 1946 by Felix Bloch:[58]

$$\frac{dM_z(t)}{dt} = -R_1 \cdot [M_z(t) - M_z^{\mathrm{eq}}] \quad \Rightarrow \quad M_z(t) = [M_z(0) - M_z^{\mathrm{eq}}] \cdot e^{-R_1 t} + M_z^{\mathrm{eq}} \tag{1.43}$$

$$\frac{dM_{x/y}(t)}{dt} = -R_2 M_{x/y}(t) \quad \Rightarrow \quad M_{x/y}(t) = M_{x/y}(0) e^{-R_2 t} \tag{1.44}$$

Longitudinal relaxation is most sensitive to the ns timescale, because the R_1 rate depends motions on the timescale of the Larmor frequency, while transversal relaxation is sensitive to slower timescales.

1.4.3 Nuclear Overhauser Effect[1,55,56]

^{15}N heteronuclear steady-state NOE (hetNOE) experiments are a commonly used relaxation experiments for capturing fast timescale dynamics in the ps to fast ns regime and nuclear Overhauser effect spectroscopy (NOESY) experiments facilitate the determination of internuclear distances (especially proton-proton distances). Both experiments base on the nuclear Overhauser effect (NOE), a dipolar relaxation mechanism acting on spins close in space with respect to each other. Figure 1.2 depicts the four energy levels with associated transition rates W resulting for a two-spin system with spins I and S.

Solomon Equations

The relaxation of longitudinal magnetisation in a two-spin system can be classically described by the Solomon equations[59]

$$\frac{d}{dt}\begin{pmatrix} \langle \hat{I}_z \rangle \\ \langle \hat{S}_z \rangle \end{pmatrix} = -\begin{pmatrix} \rho_I & \sigma_{IS} \\ \sigma_{IS} & \rho_S \end{pmatrix}\begin{pmatrix} \langle \hat{I}_z \rangle - I_z^{\mathrm{eq}} \\ \langle \hat{S}_z \rangle - S_z^{\mathrm{eq}} \end{pmatrix} \tag{1.45}$$

where I_z^{eq} and S_z^{eq} correspond to the equilibrium longitudinal magnetisation for spin $\hat{I}$ and $\hat{S}$, ρ_I and ρ_S represent the auto-correlated auto-relaxation rate constants and σ_{IS} is the auto-correlated cross-relaxation rate constant for magnetisation exchange. Auto- and cross-relaxation rate constants depend on the transition rate constants between the energy levels depicted in Figure 1.2:

$$\rho_I = W_0 + 2W_{1I} + W_2 \tag{1.46}$$

$$\rho_S = W_0 + 2W_{1S} + W_2 \tag{1.47}$$

$$\sigma_{IS} = W_2 - W_0 \tag{1.48}$$

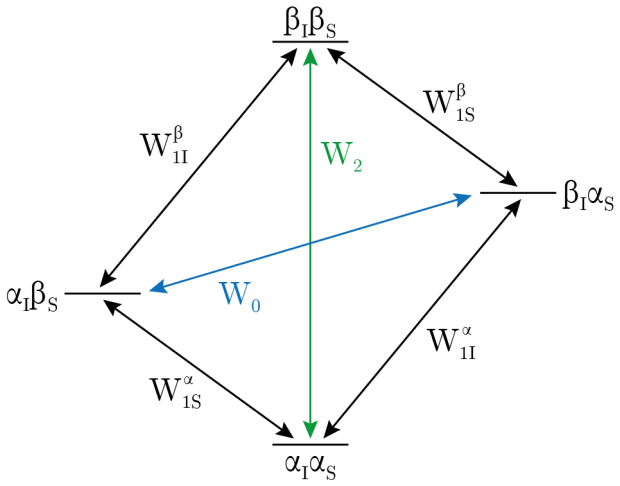

Figure 1.2: Energy level diagram of a two-spin system with spins I and S. Rate constants $W^{\alpha/\beta}_{1I/1S}$ denote single-quantum transitions, where the flipping spin is indicated as second subscript and the spin state of the second spin is indicated as superscript. In the classical Solomon equations, $W_{1I} \equiv W^{\alpha}_{1I} = W^{\beta}_{1I}$ and $W_{1S} \equiv W^{\alpha}_{1S} = W^{\beta}_{1S}$ are equal, which is however not the case if cross-correlated relaxation is taken into account. Rate constants W_0 and W_2 denote zero- and double-quantum transitions.

The Solomon equations show, that if longitudinal magnetisation is not at equilibrium, transfer of z-magnetisation between both spins occurs via cross-relaxation. This type of magnetisation transfer is called nuclear Overhauser effect (NOE). In NOESY experiments, auto-relaxation gives rise to diagonal peaks and cross-relaxation leads to cross-peaks in the corresponding spectra. Cross-relaxation rates are proportional to the square of the dipolar coupling constant $d_{IS} \propto r^{-3}$ (compare equation 1.10) and therefore depend on the internuclear distance according to $\sigma_{IS} \propto r^{-6}$, which can be utilised for distance determination. For multi-spin systems, magnetisation transfer via relayed nuclei (called spin-diffusion) is however a main error source restricting the accuracy of determined distances. Fortunately, the Solomon equations can be extended to multi-spin systems and based thereon, methods for correction of spin-diffusion and determination of highly accurate distances even for large multi-spin systems like proteins were developed.[60–63]

Heteronuclear steady-state NOE[56,64]

^{1}H-^{15}N heteronuclear steady-state NOE (hetNOE) experiments are frequently used for studying fast-timescale dynamics in proteins. In hetNOE measurements, one of both spins, for example ^{1}H in ^{1}H-^{15}N hetNOE experiments, is continuously irradiated by a weak RF field in the beginning of the experiment, such that this spin (represented in the following as spin I) gets saturated, i.e., the net magnetisation of spin I gets zero ($\langle \hat{I}_z \rangle(t) = 0$). If RF irradiation on spin I is long enough, it can be furthermore assumed, that the magnetisation of the other spin S (e.g., ^{15}N in ^{1}H-^{15}N hetNOE experiments) gets into a *steady state* $\langle \hat{S}_z \rangle(t) = S_z^{\mathrm{ss}}$, i.e., the time-evolution gets zero

$$\frac{d}{dt}\langle\hat{S}_z\rangle(t) = -\sigma_{IS}\cdot(-I_z^{\mathrm{eq}}) - \rho_S\cdot(S_z^{\mathrm{ss}} - S_z^{\mathrm{eq}}) = 0 \tag{1.49}$$

$$\Rightarrow S_z^{\mathrm{ss}} = \frac{\sigma_{IS}}{\rho_S} I_z^{\mathrm{eq}} + S_z^{\mathrm{eq}} \tag{1.50}$$

The equilibrium magnetisation $I_z^{\mathrm{eq}}/S_z^{\mathrm{eq}}$ depends on the equilibrium population difference, given by the Boltzmann distribution. It can be shown, that this population difference is proportional to the gyromagnetic ratio, such that

$$\frac{I_z^{\mathrm{eq}}}{S_z^{\mathrm{eq}}} = \frac{\gamma_I}{\gamma_S} \Leftrightarrow I_z^{\mathrm{eq}} = \frac{\gamma_I}{\gamma_S} S_z^{\mathrm{eq}} \tag{1.51}$$

Substituting I_z^{eq} in equation 1.50 by the expression in equation 1.51 results in

$$S_z^{\mathrm{ss}} = \frac{\gamma_I}{\gamma_S}\frac{\sigma_{IS}}{\rho_S} S_z^{\mathrm{eq}} + S_z^{\mathrm{eq}} = \left(\frac{\gamma_I}{\gamma_S}\frac{\sigma_{IS}}{\rho_S} + 1\right)\cdot S_z^{\mathrm{eq}} = (\eta_{\mathrm{ss}} + 1)\cdot S_z^{\mathrm{eq}} = \varepsilon_{\mathrm{ss}}\cdot S_z^{\mathrm{eq}} \tag{1.52}$$

where η_{ss} is the NOE enhancement and $\varepsilon_{\mathrm{ss}}$ is the nuclear Overhauser effect. By measuring the heteronuclear steady-state NOE experiment with and without initial irradiation on ^{1}H, the NOE can be determined from the intensity ratio:

$$\varepsilon_{\mathrm{ss}} = \frac{I^{\mathrm{sat}}}{I^{\mathrm{unsat}}} = \frac{\varepsilon_{\mathrm{ss}}\cdot S_z^{\mathrm{eq}}}{S_z^{\mathrm{eq}}} \tag{1.53}$$

Extended Solomon Equations

As mentioned in the last chapter, cross-correlation between different dipolar coupling interactions or a dipolar coupling interaction and a CSA interaction tensor can also occur.[65] These effects are reflected in the extended Solomon equations:

$$\frac{d}{dt}\begin{pmatrix}\langle\hat{I}_z\rangle\\ \langle\hat{S}_z\rangle\\ \langle 2\hat{I}_z\hat{S}_z\rangle\end{pmatrix} = -\begin{pmatrix}\rho_I & \sigma_{IS} & \Delta^I\\ \sigma_{IS} & \rho_S & \Delta^S\\ \Delta^I & \Delta^S & \rho_{IS}\end{pmatrix}\begin{pmatrix}\langle\hat{I}_z\rangle - I_z^{\mathrm{eq}}\\ \langle\hat{S}_z\rangle - S_z^{\mathrm{eq}}\\ \langle 2\hat{I}_z\hat{S}_z\rangle\end{pmatrix} \tag{1.54}$$

Auto-relaxation rates are on the diagonal of the relaxation matrix and cross-relaxation rates are off-diagonal. ρ_{IS} is the cross-correlated auto-relaxation rate. Δ_I and Δ_S are the cross-correlated cross-relaxation rates (CCRRs) reflecting the interconversion of one-spin order, $\hat{I}_z$ or $\hat{S}_z$, into two-spin order, $2\hat{I}_z\hat{S}_z$. For the two-spin magnetisation operator $2\hat{I}_z\hat{S}_z$, the equilibrium magnetisation is 0. For All rate constants depend on the transition rate constants depicted in 1.2 as follows:

$$\rho_I = W_0 + W_{1I}^{\alpha} + W_{1I}^{\beta} + W_2 \tag{1.55}$$

$$\rho_S = W_0 + W_{1S}^{\alpha} + W_{1S}^{\beta} + W_2 \tag{1.56}$$

$$\sigma_{IS} = W_2 - W_0 \tag{1.57}$$

$$\Delta_I = W_{1I}^{\alpha} - W_{1I}^{\beta} \tag{1.58}$$

$$\Delta_S = W_{1S}^{\alpha} - W_{1S}^{\beta} \tag{1.59}$$

$$\rho_{IS} = W_{1I}^{\alpha} + W_{1I}^{\beta} + W_{1S}^{\alpha} + W_{1S}^{\beta} \tag{1.60}$$

1.4.4 Redfield Theory and the Spectral Density Function[66–72]

Redfield relaxation theory is a semi-classical approach for the description of spin relaxation, where the time-dependent Hamiltonian

$$\hat{H}(t) = \hat{H}_0(t) + \hat{H}_1(t) \tag{1.61}$$

is decomposed into a deterministic part $\hat{H}_0(t)$ (containing only isotropic components like the Zeeman interaction, isotropic chemical shifts, J-couplings) and a time-dependent, stochastic part $\hat{H}_1$, causing relaxation. The deterministic part $\hat{H}_0$ in the Liouville-van Neumann equation can be eliminated by transformation of all operators $\hat{Q}$ into an interaction-frame with operators $\tilde{Q} = e^{i\hat{H}_0 t}\hat{Q}e^{-i\hat{H}_0 t}$. Mathematical transformation, ensemble averaging (indicated by an overbar) and replacement of $\tilde{\sigma}(t) \rightarrow \tilde{\sigma}(t) - \tilde{\sigma}_{\mathrm{eq}}$, where σ_{eq} is the thermal equilibrium density matrix, yields the modified Liouville-van Neumann equation

$$\frac{d}{dt}\overline{\tilde{\sigma}(t)} = -\int_0^t \overline{\left[\tilde{H}_1(t), \left[\tilde{H}_1(t-\tau), \tilde{\sigma}(t-\tau) - \tilde{\sigma}_{\mathrm{eq}}\right]\right]} d\tau \tag{1.62}$$

where τ reflects the timescale of the stochastic process and t the evolution time of the density operator. If the Redfield or weak collision limit with $\tau \ll t$ is fulfilled, which is typically the case for solution NMR due to the fast molecular tumbling but not necessarily solid-state NMR, the approximation $\tilde{\sigma}(t-\tau) \approx \tilde{\sigma}(t)$ can be applied and the upper boundary of the integral can be set to infinity. Back-transformation of equation 1.62 into the laboratory frame leads to the master equation

$$\begin{aligned}\frac{d}{dt}\overline{\hat{\sigma}(t)} &= -i\left[\hat{H}_0(t), \hat{\sigma}(t)\right] - \int_0^\infty \overline{\left[\hat{H}_1(t), \left[e^{-i\hat{H}_0\tau}\hat{H}_1(t-\tau)e^{i\hat{H}_0\tau}, \hat{\sigma}(t) - \hat{\sigma}_{\mathrm{eq}}\right]\right]} d\tau \\ &= -i\left[\hat{H}_0(t), \hat{\sigma}(t)\right] - \hat{\hat{\Gamma}}\{\hat{\sigma}(t) - \hat{\sigma}_{\mathrm{eq}}\}\end{aligned} \tag{1.63}$$

where $\hat{\hat{\Gamma}}$ is the relaxation superoperator. After writing the Hamiltonian $\hat{H}_1(t)$ in spherical-tensor notation according to equation 1.31, and additional mathematical transformations, the real part of the relaxation superoperator $\hat{\hat{\Gamma}}$ derives to

$$\hat{\hat{\Gamma}} = \frac{1}{2}\sum_{\mu,\mu'}\sum_{l,l'}\sum_{m,m'}(-1)^{m+m'}\sum_{p'}\left[\hat{T}^{\mu}_{l,m}, \left[\left(\hat{V}^{\mu'}_{p'}\right)^{\dagger}, \right]\right] j^{\mu,\mu'}_{l,l',-m,-m'}(\omega_{p'}) \tag{1.64}$$

where μ and μ' describe the types of interaction (i.e., $\mu = \mu'$ for auto-correlated relaxation, $\mu \neq \mu'$ for cross-correlated relaxation) and $j^{\mu,\mu'}_{l,l',-m,-m'}$ is the spectral density function

$$j^{\mu,\mu'}_{l,l',-m,-m'}(\omega_{p'}) = \int_0^\infty \overline{A^{\mu}_{l,-m}(t)\left(A^{\mu'}_{l',-m'}\right)^*(t-\tau)} \cdot e^{i\omega_{p'}\tau} d\tau \tag{1.65}$$

The time-averaged expression corresponds to the stochastic correlation function, which is only non-vanishing if $l = l'$ and $m = m'$. $\hat{V}^{\mu'}_{p'}$ and $\omega_{p'}$ are eigenfunctions and eigenfrequencies of the Hamiltoninan commutation superoperator $\hat{\hat{H}}_0$:

$$\hat{\hat{H}}_0\hat{V}^{\mu'}_{p'} = \left[\hat{H}_0, \hat{V}^{\mu'}_{p'}\right] = \omega_{p'}\hat{V}^{\mu'}_{p'} \tag{1.66}$$

where operators $\hat{V}^{\mu'}_{p'}$ are basis operators of $\hat{T}^{\mu'}_{l',m'} = \sum_{p'}\hat{V}^{\mu'}_{p'}$ and $\hat{H}_0 = \omega_I\hat{I}_z + \omega_S\hat{S}_z$ is for example in case of uncoupled 2-spin systems.

Redfield theory[66–69] shows that relaxation depends on the spectral density function

$$\begin{aligned} j_{l,m}(\omega) &= \mathrm{Re}\left\{2\int_0^\infty \langle A_{l,m}(t)A_{l,m}(t-\tau)\rangle_t \cdot e^{-i\omega\tau}d\tau\right\} \\ &= \mathrm{Re}\left\{2\int_0^\infty \langle A_{l,m}(t)A_{l,m}(t+\tau)\rangle_t \cdot e^{-i\omega\tau}d\tau\right\} \\ &= \mathrm{Re}\left\{2\int_0^\infty C_{l,m}^{\mathrm{stoch}}(\tau)\cdot e^{-i\omega\tau}d\tau\right\} \end{aligned} \tag{1.67}$$

where $C_{l,m}^{\mathrm{stoch}}(\tau)$ is the stochastic correlation function, τ is the time delay, and the angle brackets symbolise the average over time t. For isotropic liquids at high magnetic fields, the spatial components $A_{l,m}$ with rank $l = 2$ and component $m = 0$ are the most relevant for studying relaxation induced by the dipolar coupling and chemical-shift anisotropy. Therefore, the stochastic correlation function can be rewritten as

$$C_{2,0}^{\mathrm{stoch}}(\tau) = \left\langle c_0(t)\, c_0(t+\tau)\, Y_2^0(\theta(t))\, Y_2^0(\theta(t+\tau))\right\rangle \tag{1.68}$$

where θ is the angle between the z-axis of the spherical tensor and the B_0 field, Y_l^m with $Y_2^0 = \sqrt{\frac{5}{4\pi}}\cdot d_{0,0}^2 = \sqrt{\frac{5}{4\pi}}\cdot\frac{1}{2}(3\cos^2\theta - 1)$ is a normalised spherical harmonic function describing the orientation dependence of the spatial component $A_{l,m}$ and c_0 is a function of physical constants (e.g., for dipolar couplings depending on $\delta_{\mathrm{D}} = r_{IS}^{-3}(t)$, compare section 1.2.3). Because relaxation of ^{15}N spins in amide groups in the protein backbone is mainly determined by dipolar couplings to the (via a fixed distance/bond length) covalently bound ^{1}H atom, c_0 is often assumed to be roughly constant in time. Therefore, an orientational correlation function can be defined as

$$C(\tau) = \left\langle Y_2^0(\theta(t))\, Y_2^0(\theta(t+\tau))\right\rangle = C^{\mathrm{stoch}}(\tau)/c_0^2 \tag{1.69}$$

and the corresponding orientational spectral density function can be defined as

$$J(\omega) = \mathrm{Re}\left\{2\int_0^\infty C(\tau)\cdot e^{-i\omega\tau}\right\} = j(\omega)/c_0^2 \tag{1.70}$$

According to Redfield theory, relaxation rates can be expressed as function of the spectral density function at different frequencies, e.g., ^{15}N longitudinal and transversal relaxation rates and ^{15}N cross-relaxation rates from ^{1}H-^{15}N hetNOE experiments correspond to

$$\begin{aligned} R_1 &= \Gamma_{S_z,S_z} = \mathrm{Tr}\{\langle\hat{S}_z|\hat{\hat{\Gamma}}|\hat{S}_z\rangle\}/\,\mathrm{Tr}\{\langle\hat{S}_z|\hat{S}_z\rangle\} \\ &= \frac{d_{IS}^2}{4}\cdot[J(\omega_{\mathrm{H}}-\omega_{\mathrm{N}}) + 3J(\omega_{\mathrm{N}}) + 6J(\omega_{\mathrm{H}}+\omega_{\mathrm{N}})] + c_{\mathrm{CSA}}^2 J(\omega_{\mathrm{N}}) \end{aligned} \tag{1.71}$$

$$\begin{aligned} R_2 &= \Gamma_{S^+,S^+} = \mathrm{Tr}\{\langle\hat{S}^+|\hat{\hat{\Gamma}}|\hat{S}^+\rangle\}/\,\mathrm{Tr}\{\langle\hat{S}^+|\hat{S}^+\rangle\} \\ &= \frac{d_{IS}^2}{8}\cdot[4J(0) + J(\omega_{\mathrm{H}}-\omega_{\mathrm{N}}) + 3J(\omega_{\mathrm{N}}) + 6J(\omega_{\mathrm{H}}) + 6J(\omega_{\mathrm{H}}+\omega_{\mathrm{N}})] \\ &\quad + \frac{c_{\mathrm{CSA}}^2}{6}[4J(0) + 3J(\omega_{\mathrm{N}})] \end{aligned} \tag{1.72}$$

$$\begin{aligned} \sigma_{\mathrm{NH}} &= \Gamma_{I_z,S_z} = \mathrm{Tr}\{\langle\hat{I}_z|\hat{\hat{\Gamma}}|\hat{S}_z\rangle\}/\,\mathrm{Tr}\{\langle\hat{I}_z|\hat{I}_z\rangle\} \\ &= \frac{d_{IS}^2}{4}\cdot[6J(\omega_{\mathrm{H}}+\omega_{\mathrm{N}}) - J(\omega_{\mathrm{H}}-\omega_{\mathrm{N}}))] \end{aligned} \tag{1.73}$$

where $d_{IS} = -(\mu_0 \gamma_I \gamma_S \hbar)/(4\pi r_{IS}^3)$ is the dipolar coupling constant (compare equation 1.10), $c_{\mathrm{CSA}} = \Delta\sigma\omega_{\mathrm{N}}/\sqrt{3}$ is the CSA coupling constant with the amide CSA ($-172\,\mathrm{ppm}$) and spins I and S represent ^{1}H and ^{15}N, respectively.[73]

In most cases, the correlation of an interaction tensor (i.e., a N-H bond vector) with itself over time is studied. This type of rotational correlation is termed auto-correlation. However, also different interaction tensors can be correlated over time, e.g., the two ^{13}C-^{1}H dipolar coupling tensors or one ^{13}C-^{1}H dipolar coupling tensor and the ^{13}C CSA tensor within one methyl group can show a correlation over time and lead to observable relaxation. This type of relaxation is termed cross-correlated relaxation and is much weaker than auto-correlated relaxation. In the following only auto-correlated relaxation is considered.

In the Lipari-Szabo model-free analysis,[74] overall and internal protein motion are considered to be independent, such that the total correlation function can be written as

$$C(t) = C_{\mathrm{O}}(t) C_{\mathrm{I}}(t) \tag{1.74}$$

where the correlation function for the overall molecular tumbling in isotropic solutions is

$$C_{\mathrm{O}}(t) = \frac{1}{5} e^{-t/\tau_c} \tag{1.75}$$

and τ_c is the molecular correlation time, approximately corresponding to the average time for a molecule to rotate about 1 radian. In anisotropic solutions, C_{O} has to be determined in a slightly more complicated manner based on the axial components $D_\perp$ and $D_\parallel$ of the rotational diffusion tensor D_{rot}.[72]

The internal correlation function C_{I} can be approximated in the simplest case by a single internal correlation time τ_i and the order parameter S^2:

$$C_{\mathrm{I}}(t) = S^2 + (1 - S^2) e^{-t/\tau_i} \tag{1.76}$$

According to the addition theorem for spherical harmonics, C_{I} can also be calculated by

$$C_{\mathrm{I}}(\tau) = \langle P_2 \left(\cos\theta(t, t+\tau)\right) \rangle_t \tag{1.77}$$

where P_2 is the second Legendre polynomial ($P_2(\cos\theta) = d_{00}^2(\theta) = (3\cos^2\theta - 1)/2$) and $\theta(t, t+\tau)$ is the angle between z-axis of the spherical tensor at time t and $t+\tau$, which coincides with the direction of the interaction, i.e., the N-H amide bond vector in the protein backbone in case of ^{15}N R_1, R_2 or hetNOE experiments. Using this equation, orientational correlation functions and NMR relaxation parameters can be back-calculated from molecular dynamics (MD) simulations.

By Fourier transformation of the orientational correlation function, the corresponding spectral density function can be obtained as

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + (\omega\tau_c)^2} + \frac{(1 - S^2)\tau_{\mathrm{eff}}}{1 + (\omega\tau_{\mathrm{eff}})^2} \right) \tag{1.78}$$

where $\tau_{\mathrm{eff}}^{-1} = \tau_c^{-1} + \tau_i^{-1}$ is the effective correlation time.

Using a set of R_1, R_2, and hetNOE experiments, for which each corresponds to a different linear combination of the spectral density function, all three parameters from the Lipari-Szabo model-free analysis (τ_c, τ_i, and S^2) can be estimated using spectral density mapping. In the extended model-free analysis, internal motion is described by two different timescales, a fast and slow internal correlation time and associated fast and slow order parameters.[75,76] Using the same set of relaxation experiments as above measured at different magnetic field strengths, protein backbone dynamics can be quantitatively assessed even with two different time scales of internal motion.[73]

1.4.5 Experiments Sensitive to Slow Timescale Dynamics[57,77]

In the slow µs-ms timescale, various types of important biological process like loop and domain motions, enzymatic reactions, ligand binding, protein folding, and protein–protein interactions occur. In this slow time regime, exchange rates are much smaller than the difference in Larmor frequencies ($k_{\text{ex}} \ll \Delta\omega$, called slow exchange) leading to separate peaks for the low-populated excited and the high-populated ground state. In contrast, for fast exchange, where $k_{\text{ex}} \gg \Delta\omega$, a single peak at the population-weighted average resonance frequency is observed. Intermediate exchange induces strong line broadening especially affecting the excited state, making it difficult to observe directly. Because excited states are usually only sparsely populated and therefore difficult to detect, two-site exchange models are primarily used for analysis of relaxation experiments.

A variety of different relaxation experiments were developed, for probing different portions within the µs-ms regime. Exchange spectroscopy (EXSY)[78] and chemical-exchange saturation transfer (CEST)[79,80] experiments are at the slow end capturing dynamics between tens of ms to s. Chemical shifts of ground and excited states can be directly observed in these experiments. Complementary, relaxation dispersion experiments facilitate the detection of dynamics ranging from few µs to hundreds of ms.[81]

In $R_{1\rho}$ experiments, an RF field with field strength ω_1 is applied in the transverse plane, which spin-locks the magnetisation in the rotating frame along the effective field strength $\vec{\omega}_e$ constructed from the components ω_1 in the transversal plane and the frequency offset Ω from the RF carrier aligned along the z-axis. Depending on the tilt angle $\theta = \arctan(\omega_1/\Omega)$ of the spin-locked magnetisation with respect to the B_0-field, $R_{1\rho}$ relaxation rates contains contributions from longitudinal and transversal relaxation rates, which can in the fast-exchange limit be expressed as

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta = R_1 \cos^2\theta + (R_2^0 + R_{\text{ex}}) \sin^2\theta \tag{1.79}$$

where R_2^0 and R_2 are the intrinsic and the effectively observable transverse relaxation rate constant, respectively, and

$$R_{\text{ex}} = \frac{k_{\text{ex}}\Phi_{\text{ex}}}{k_{\text{ex}}^2 + \omega_e^2} = \frac{k_{\text{ex}} p_{\text{A}} p_{\text{B}} \Delta\omega^2}{k_{\text{ex}}^2 + \omega_e^2} \tag{1.80}$$

is the contribution to the transverse relaxation arising from chemical exchange between states A and B with populations $p_{A/B}$. Depending on the field strength ω_1 of the applied RF field, $R_{1\rho}$ experiments are sensitive to different time regimes (fast µs regime for strong spin-lock fields and slow ms regime for low-power $R_{1\rho}$).[82,83] Depending on the positioning of the applied spin-lock field with respect to the resonance frequency of the spins of interest, $R_{1\rho}$ experiments can be performed on-resonance (also referred to as near-resonance) and off-resonance. Furthermore, $R_{1\rho}$ experiments cannot only be performed in solution, but also in solid-state NMR. Depending on the strength of the effective irradiation field strength ω_e with respect to the spinning frequency ω_r, two different types of relaxation can be differentiated: (i) Bloch-McConnell relaxation dispersion (BMRD) occurs for $\omega_e \ll \omega_r$ and results from slow timescale fluctuations of the isotropic chemical shielding tensor, while (ii) near-rotary resonance relaxation dispersion (NERRD) dominates at $\omega_e \approx \omega_r$ and results from time-dependent stochastic fluctuations of the anisotropic contributions of spin-interaction tensors.[84] In contrast to BMRD, NERRD can only be measured in solids but not in solution.

Carr–Purcell–Meiboom–Gill (CPMG)[85,86] relaxation dispersion is a technique sensitive to dynamics ranging from hundreds of µs to the slow ms timescale and is based on consecutive 180° pulses (called CPMG pulse trains) at a frequency ν_{CPMG} for probing the decay of transverse relaxation rates $R_2(\nu_{CPMG}) = R_2^0 + R_{ex}(\nu_{CPMG})$ with increasing frequencies ν_{CPMG}. CPMG experiments are nowadays usually performed as constant-time version, where the CPMG relaxation period T_{CPMG} is kept constant.[87,88] Hence, effective transverse relaxation rates can be calculated by

$$R_2(\nu_{CPMG}) = -\frac{1}{T_{CPMG}} \cdot \ln\left(\frac{I(\nu_{CPMG})}{I_0}\right) \tag{1.81}$$

where I_0 is the transverse magnetisation measured at $T_{CPMG} = 0$.[57,89] For fast timescale exchange between states A and B, the analytic expression for the transverse relaxation rate is given by

$$R_2(\nu_{CPMG}) = p_A R_{2A}^0 + p_B R_{2B}^0 + \frac{\Phi_{ex}}{k_{ex}}\left[1 - \frac{4\nu_{CPMG}}{k_{ex}} \cdot \tanh\left(\frac{k_{ex}}{4\nu_{CPMG}}\right)\right] \tag{1.82}$$

with $\Phi = p_A p_B \Delta\omega_{AB}^2$, the populations $p_{A/B}$, the intrinsic transverse relaxation rates $R_{2A/B}^0$, the resonance frequency difference ω_{AB} between states A and B and the exchange rate k_{ex}.[90,91] For exchange in the intermediate timescale regime ($k_{ex} \approx \langle\Delta\omega_{AB}\rangle$), transverse relaxation rates can be expressed by the Carver–Richards equation.[91–93] Instead of fitting the decay of transverse relaxation to corresponding analytical equations, which can get relatively complicated, e.g., for intermediate exchange or three-site exchange, hampering the determination of quantitative fitting parameters, numerical modelling according to the Bloch–McConnell equations[94] can be performed using programs like ChemEx.[80]

1.5 Basics of Molecular Dynamics Simulations[95,96]

Molecular dynamics (MD) simulations have become an indispensable tool for studying the dynamics of biomacromolecules on an atomic scale. In force-field based molecular dynamics simulations, atoms are treated as classical particles, such that their motion and resulting forces can be described by Newton's second law

$$\vec{F}_i = -\vec{\nabla}_i U(\vec{r}) = -\left(\frac{\partial}{\partial r_{i,x}}, \frac{\partial}{\partial r_{i,y}}, \frac{\partial}{\partial r_{i,z}}\right) U(\vec{r}) = m_i \vec{a}_i = m_i \frac{\partial^2}{\partial t^2} \vec{r}_i \tag{1.83}$$

where the force $\vec{F}_i$ acting on particle i with mass m_i, coordinates $\vec{r}_i$, and acceleration $\vec{a}_i$ is derived from the negative gradient of the coordinates of particle i acting on the total potential energy $U(\vec{r})$ of the system. Force fields like AMBER,[97–99] CHARMM,[100] GROMOS,[101–104] and OPLS[105,106] provide a set of empirical energy functions and parameters for accurately representing all kinds of interatomic forces (e.g., bond stretching, bending, torsions, electrostatic interactions, van der Waals forces, and hydrogen bonding interactions) used for the calculation of the potential energy U. Aside from huge systems with millions of atoms, usually all-atom force fields are used instead of united-atom force fields, where for the purpose of reducing the complexity of calculations aliphatic hydrogens are merged with the bound carbon into a single particle instead of explicitly parametrising each atom as individual particle like in all-atom force fields.

In MD simulations, Newton's equation of motion is numerically integrated over small time steps via algorithms like the Verlet,[107] velocity Verlet[108] or Leap Frog[109] algorithm. For accurate description of motion, the time step has to be several times smaller than the fastest motion present (i.e., hydrogen bond vibrations) and is usually set to 1 fs. If hydrogen bond lengths and angles between them are constrained, the time step can be increased to 2 fs.

MD simulations can be performed in different thermodynamic ensembles like the micro-canonical (NVE), canonical (NVT) or isothermal-isobaric (NpT) ensemble, where the variables indicated in brackets are kept constant. The NVT and NpT ensembles are practically much more relevant, as they better reflect the natural and experimental conditions, in contrast to the NVE ensemble, they however require a thermostat[110,111] and barostat[112–114] for temperature and pressure control.

The standard procedure for the setup of classical MD (cMD) simulations of solvated proteins consists of multiple steps. At first, a periodic box is constructed to account for boundary effects. The box is filled with water atoms neutralised with counter ions and energy minimised. Before starting the production run in the desired thermodynamic ensemble, usually, an initial equilibration in the NVT ensemble is performed for getting the system to the desired temperature and subsequently, an equilibration in NpT ensemble ensures that the equilibrium density of the system is reached at the given pressure and temperature. During all equilibration steps position restraints are applied to all non-hydrogen

atoms of the solute for preventing potential artifacts arising from random initial velocities or unrealistic displacements as a result of changes in the box size. Optimally, these position restraints are gradually released prior to the production run.
Despite the extensive insights cMD simulations provide, reaching time scales in the slow µs or even ms regime, where a variety of important biological processes like protein folding, loop and domain motions, and protein–protein interactions occur, as well as incomplete exploration of the conformational space by not overcoming local energy minima are still major bottlenecks. To overcome these limitations, different enhanced sampling methods[115] were developed, e.g., accelerated MD[116–118] (aMD), replica-exchange MD,[119–121] and metadynamics.[122,123]

1.6 NMR Observables for Structure Determination and Refinement of MD Simulations

NMR spectroscopy provides a variety of observables such as short-range NOESY-based distances (up to 6 Å), chemical shifts,[124,125] torsion angles from scalar couplings,[11,12,126] order parameters[127,128] from fast-timescale relaxation experiments, orientational restraints from residual dipolar couplings (RDCs) measurements,[129–131] and long-range distances from paramagnetic relaxation enhancement (PRE) experiments,[132–134] which can be exploited for structure determination as well as for perturbing molecular dynamics simulations. All of these experimental observables represent averages of the conformational ensemble on different timescales. While NOE-derived distances and order parameters from ^{15}N R_1, R_2, and hetNOE measurements are averaged on the fast, picosecond to nanosecond timescale, PRE are sensitive to the intermediate microsecond timescale and RDCs are affected from conformational dynamics ranging from the fast ps-ns timescale up to microseconds.
NMR-based structure calculation typically mainly relies on distances determined from NOESY cross-peak intensities ($r \propto V^{-1/6}$). The ambiguity in the assignment of cross-peaks to the corresponding pair of protons is a major hurdle and tackled by iterative protocols, which comprise cross-peak assignment, calculation of a bundle of structures and assignment validation based on the calculated temporary structure models, and are implemented in structure determination programs like ARIA,[135,136] CYANA[137,138] and Xplor-NIH.[139,140] The structure determination itself is based on a molecular dynamics simulated annealing (MDSA) protocol guided by restraints constituting a bias potential. Since NOE-based distances are r^{-6}-averages of the conformational ensemble, restraints can be enforced in an ensemble-averaged manner.[141,142]
Accurate determination of structural ensembles of proteins using molecular dynamics simulations still faces two major problems: Limitations in conformational sampling due to the high computational costs associated with reaching e.g., the slow µs-ms time scale as well as force-field inaccuracies, which are in particular for highly dynamic regions and systems

like intrinsically disordered regions and proteins (IDRs/IDPs) still significant.[143–145] Therefore, experimental observables from NMR and other techniques like small angle X-ray scattering (SAXS),[146–148] electron-electron double resonance (DEER)[149] or Förster resonance energy transfer (FRET)[150,151] are often used for either *a priori* restraining MD simulations or *a posteriori* reweighting unbiased MD simulations. A variety of methods have been developed for obtaining accurate trajectories from MD simulations guided by ensemble-averaged experimental data. All of these different ensemble determination methods can be grouped into two opposing categories, the maximum-entropy principle and the maximum-parsimony principle.[152–154] According to the maximum-entropy principle, the perturbation inflicted on the structural ensemble by *a priori* restraining or *a posteriori* reweighing should be the least possible to conform experimental data, while the maximum-parsimony principle aims for a smallest possible ensemble size compatible to the experimental observables. Maximum parsimony methods are well suited for rigid systems with few relevant states, whereas maximum entropy methods are more versatile and also appropriate for describing flexible systems possessing a broad energy landscape with many local minima arising from the high conformational freedom.[154]

2

Integrated Assessment of Structure and Dynamics of Solid Proteins

The work presented in this chapter was published as B. Söldner, K. Grohe, P. Neidig, J. Auch, S. Blach, A. Klein, S. K. Vasa, L. V. Schäfer, R. Linser, *J. Phys. Chem. Lett.* **2023**, *14*, 1725–1731. An open-access preprint of this work is available at bioRχiv.

My contributions to this work:

- Together with Kristof Grohe, I developed the method for determining accurate distances from solid-state NMR ("eRFDR").
- Kristof Grohe measured all RFDR spectra and adopted the correction method for spin-diffusion from solution to solid-state.
- I performed the side-chain and RFDR assignment.
- I performed all SIMPSON simulations and thereon, I developed the correction method for dipolar truncation. Suresh Kumar Vasa helped with the initial setup of the SIMPSON simulations.
- I performed all structure calculations except for those from partially truncated datasets.

2.1 Abstract

Understanding macromolecular function, interactions, and stability hinges on detailed assessment of conformational ensembles. For solid proteins, accurate elucidation of the spatial aspects of dynamics at physiological temperatures is limited by the qualitative character or low abundance of solid-state nuclear magnetic resonance (ssNMR) internuclear distance information. Here, we demonstrate access to abundant proton–proton internuclear distances for integrated structural biology and chemistry with unprecedented accuracy. Apart from highest-resolution single-state structures, the exact distances enable molecular dynamics (MD) ensemble simulations orchestrated by a dense network of experimental interproton distance boundaries gathered in the context of their physical lattices. This direct embedding of experimental ensemble distances into MD will provide access to representative, atomic-level spatial details of conformational dynamics in supramolecular assemblies, crystalline and lipid-embedded proteins, and beyond.

2.2 Introduction

Access to site-resolved structural data is usually a precondition for a detailed structural understanding data is of usually biomacromolecular function. Elucidation of key biological and biophysical properties ranging from intermolecular interactions and recognition to enzymatic activity, regulation, and allostery, however, also requires insights into different details of atomic motion. NMR spectroscopy provides access to various aspects of structural dynamics under close-to-physiological conditions. Much in contrast to many other structural-biology techniques providing only ground-state structures, NMR parameters are shaped by additional conformations that are populated at room (or body) temperature and thus can also contribute to the molecular properties via the underlying dynamic conformational ensemble. In solution, MD simulations are regularly augmented by experimental NMR restraints, which are translated into (system-specific) pseudopotentials that are added to the regular protein force field. These experimentally derived bias potentials can be implemented in the MD simulations via ensemble-averaging approaches that obey the maximum-entropy principle.[125,128,156–158] In addition to chemical shifts, accurate structural restraints (e.g., exact NOEs) are available that permit reconstruction of dynamic ensembles with great detail.[142,159–161] Complementing the class of soluble or solubilized proteins and increasingly aided by cryo-electron microscopy (cryo-EM) tomography for information on the nanometer to micrometer scale,[162–164] solid-state NMR (ssNMR) grants access to structural and motional features of proteins in a solid lattice, including supramolecular assemblies and fibrillar or membrane proteins. With the onset of direct proton detection, because of fast magic-angle spinning (MAS) and/or partial deuteration, ssNMR has seen a revolution with respect to its associated technical possibilities and feasibility for biological studies, with a particular prospect for high-molecular weight targets.[23,165,166]

Structure elucidation from proton-detected ssNMR has initially been hinging on amide and methyl groups, whose contacts are sparse for perdeuterated proteins and have limited structural assessment to a reasonably defined backbone fold.[167–170] Fully protonated proteins at MAS rates exceeding 100 kHz, by contrast, now open the door to a dense network of internuclear distances. The restraints are an order of magnitude more numerous than for deuterated proteins but are associated, as of now, with severe shortcomings regarding a quantitative interpretation, which compromises the value of established structure elucidation approaches.[171–173] In the solid state, for detailed elucidation of dynamic ensembles within crystalline proteins, supramolecular assemblies, fibrils, and membrane proteins, a similar impact of concatenation of experiment and simulation via the maximum-entropy principle can be expected as for solution NMR. However, due to the sparseness (e.g., isolated REDOR[44] spin pairs) or rather qualitative character (e.g., current homonuclear recoupling techniques[47,167,170,171,173–176]) of distance restraints in MAS solid-state NMR, prospects have been limited to chemical-shift anisotropy and dipolar coupling-based implementations in oriented membrane protein samples.[177,178] Opposing the limited perspectives prevailing for the large range of solid proteins, here we demonstrate integrated assessment of structure and dynamics via experimental data from fast-MAS NMR, leveraged by a dense network of accurate proton–proton distances made attainable.

2.3 Methods and Results

In solid-state NMR, the build-up of coherence-driven magnetization transfer usually behaves like a linear combination of Bessel functions of the first kind.[179] Here, to revisit the potential of such data for detailed assessment of structure and dynamics, we recorded and assigned three-dimensional (3D) time-shared ^{15}N/^{13}C-edited RFDR magnetization build-ups[167] at 110 kHz MAS on a microcrystalline, fully protonated (^{13}C/^{15}N) sample of the chicken α-spectrin SH3 domain in a time-dependent fashion (Figure 2.1A and Figure 2.5; see further details in section 2.4). Using the automated assignment procedure of ARIA, in total, 1033 unambiguously assigned peaks were obtained, including 274 cross-peaks from bidirectional transfers, 485 cross-peaks from unidirectional transfers, and 274 diagonal peaks. For the generation of exact solid-state distance restraints, we adapted the eNORA framework[60,180] to the specific behaviour of ssNMR magnetization transfers, in particular tailored fitting of magnetization build-up and correction for dipolar truncation and spin diffusion. Bearing advantages over numerical simulations, an analytical function for fitting recoupling curves in a powder sample (i.e., with a distribution of angles β between the B_0 field and interspin vector according to random molecular orientations) was derived by Mueller using a linear combination of Bessel functions of the first kind (J_α with order α),[179] whose series expansion can be truncated practically to $k \leq 5$ elements

$$I(\lambda) = A_1 \cdot \left\{ 1 - \left[J_0\left(\sqrt{2}\lambda\right) \right]^2 + \sum_k \frac{2}{16k^2 - 1} \left[J_k\left(\sqrt{2}\lambda\right) \right]^2 \right\} + C \tag{2.1}$$

where I is the intensity as a function of mixing time, A_1 is a vertical scaling factor (amplitude), λ is the argument of the Bessel function, and C is an optional offset factor (not used in this study). The effective behaviour of homonuclear dipolar recoupling and coherence-driven transfers is reasonably similar to the REDOR case (compare Figure 2.1B).[46,181] Hence, the argument λ can be constructed as $\sqrt{2}\lambda = D\tau_{\mathrm{mix}}$ from build-up time τ_{mix}, modulated by the interaction strength D, which in an ideal two-spin case is directly proportional to the dipolar coupling constant. The distance r is then obtained via

$$r = A_2 \cdot (D \cdot F)^{-1/3} + B \tag{2.2}$$

with correction factor F combining spin diffusion and dipolar truncation and user-defined values for slope A_2 and y-intercept B, both calibrated with respect to a preliminary starting structure or known distances, e.g., between geminal protons. The scaling factor of the apparent interaction strength is not known beforehand but merges into a global fit parameter A_2. Even though the procedures for fitting and extraction of distances from the fitted interaction strengths as well as the developed methods for correction of spin diffusion and dipolar truncation (see below and section 2.4 for details) are heuristic solutions, the close correspondence of the distances determined from the corrected interaction strengths and the true distances are apparent in Figure 2.1C (blue curve).

Direct interpretation of magnetization build-ups is compromised by the fact that the magnetization transfer from spin I to spin S is modulated by nearby third spins (e.g., dipolar truncation) as well as spin diffusion. The latter is heuristically taken into account by assuming a merely additive effect on the observed magnetization flow, as has been done in the framework of NOE build-up in solution.[63] To obtain correction factors for dipolar truncation, a simulation-based method hinging on nearest-neighbour interactions was developed (see details in section 2.4). In brief, 17,500 theoretical three-spin scenarios of RFDR transfers from I to S with a third spin K with different geometries, in addition to a simple two-spin build-up for each I – S distance, were simulated in SIMPSON.[182] All synthetic build-ups were fitted in analogy to the experimental data, and for each geometry, a correction factor was derived as the ratio between the apparent interaction strength D obtained in the two-spin simulation and the fit value D obtained when fitting the output of the three-spin simulation. These correction factors, applied on the basis of similarity between simulation geometry and spin environment of the site in a template structure (see below) as well as the strength of this modulating effect comparing different third spins, enabled association of experimental build-ups with relevant two-spin distance values.

As correct cross-peak assignment can be efficiently obtained from iterative assignment procedures,[136] distance restraints were benchmarked here for a series of 15 3D time-shared, ^{15}N/^{13}C-edited RFDR spectra of different mixing times. With the increased dispersion of four-dimensional spectra,[167,169,170] the benefits of full protonation could be exploited even further, implying fewer overlap and hence further improvements in data quality but necessitating more measurement time. As template structure for correction of spin

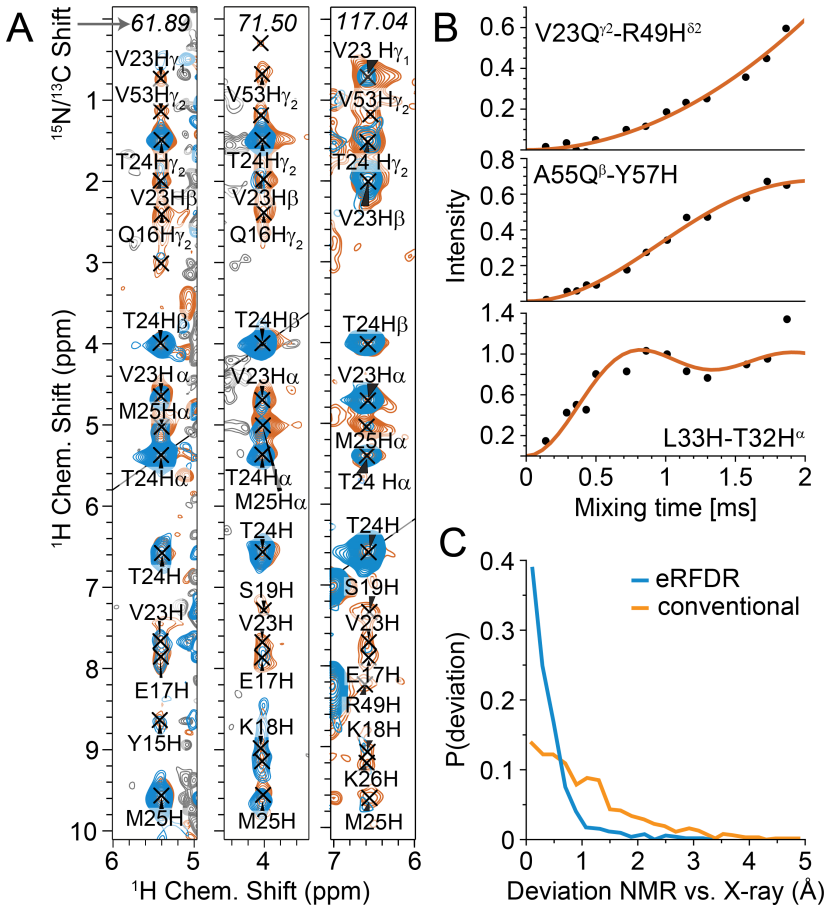

Figure 2.1: Generation of accurate solid-state NMR distance restraints. (A) Exemplary strips of time-shared RFDR experiments (diagonal at T24 H^{α}, H^{β} and H^{N}, respectively) for a fully protonated sample of the SH3 domain of chicken α-spectrin at 110 kHz MAS (blue, 0.288 ms mixing time; orange, 1.008 ms mixing time). Instead of 3D data, increasing peak dispersion by a fourth dimension is possible.[167,169,170] (B) Build-up curves (slow, intermediate, and fast) and their fits for exemplary transfers (with distances of 6.4 Å, 4.8 Å, and 2.1 Å from top to bottom). (C) Statistics of the accuracies obtained for the determined distances, shown as deviation to corresponding distances found by crystallography. The strips in panel A are shown with a larger spectral width in Figure 2.5.

diffusion and dipolar truncation, we used either (1) the X-ray structure 2NUZ (as proof of principle and for demonstrating the achievable distance restraint accuracies), (2) a structure obtained from conventional ssNMR structure calculation (see section 2.4) based on cross-peak volumes of a *single* RFDR spectrum (for restraint accuracies that can be obtained when no X-ray, cryo-EM, or AlphaFold[183] structure is available), and (3) an incorrectly folded starting structure (for assessing the error tolerance of the method for correction of spin diffusion and dipolar truncation toward a worst-case scenario, see Table 2.2). Table 2.2 also shows that spin diffusion correction is the correction of choice for the deuterated sample, whereas the combination of both corrections yields best results for fully protonated samples. In addition, Table 2.2 and Figure 2.9 compare results for calculation based on only five (instead of the excessive 15) RFDR spectra, in which case quantitative restraints are still obtained with an only minute decrease of the structural statistics.

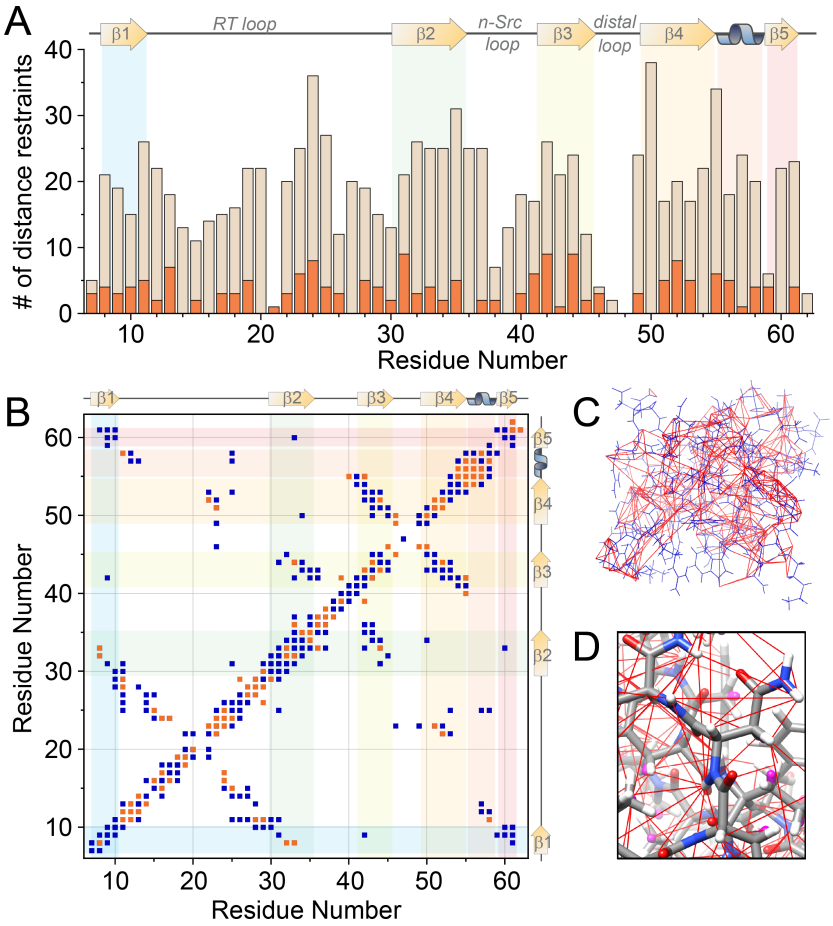

Figure 2.2: Quantity of the obtained restraints and their connectivities. (A) Number of distance restraints per residue (ochre, fully protonated sample; orange, deuterated and proton-back-exchanged sample). (B) Contact map with unidirectional (blue) and bidirectional (orange) distance restraints. (Only one dot is shown per pair of residues even for multiple contacts via different protons.) (C and D) Depiction of the coverage of obtained distances, forming a dense network of restraints.

For the protonated sample, given the high number of protons available in the side chains, 718 of 759 build-ups could be fitted, or 25.6 per residue on average (only including assignable residues, 7–62). Among these, 98 peaks were discarded due to unreliable fitting (see the Supporting Information for details), 204 peaks could be included as bidirectional restraints, with magnetization transfer found and quantified for both $i \rightarrow j$ and $j \rightarrow i$, and 416 peaks were used as unidirectional restraints (Figure 2.2 and Table 2.1). In contrast, for the deuterated and back-exchanged sample, only 52 peaks could be included as bidirectional restraints and 67 peaks as unidirectional ones, which is only 18 % of the restraints of the protonated case.

At first, to assess the extent to which the utilization of accurate distance restraints can improve regular structure determination in solid-state NMR using either protonated or deuterated and $^{1}H^{N}$ back-exchanged samples, structure calculations (Figure 2.3) were performed with ARIA,[136] as laid out in more detail in section 2.4. Indeed, structure calculation resulting from the high abundance of exact distance restraints yields very

high precision with root-mean-square deviations (RMSDs) of 0.43 and 1.02 Å (backbone and all atoms, respectively), determined over all non-terminal residues from position 7 to 61 (or even 0.30 and 0.83 Å, respectively, determined over the secondary structural elements only (see Table 2.1)). The closeup in Figure 2.3B and the residue-wise comparison in Figure 2.3C demonstrate the improvement in structural precision with regard to a deuterated/back-exchanged sample under otherwise comparable conditions (backbone and all-atom precision values of 0.86 and 1.56 Å, respectively). Similarly, comparison of the structural ensemble with X-ray structure 2NUZ yields backbone RMSDs of 0.45 and 1.01 Å for protonated and deuterated/back-exchanged structures, respectively, for the secondary structural elements (compare Table 2.1).

Table 2.1: Structural statistics are shown for the entire protein and only structured regions, i.e., excluding residues 17–23 (RT loop), 36–41 (n-SRC loop), and 47–48 (distal loop) and termini, using the combined (spin diffusion and dipolar truncation) correction method with PDB 2NUZ as a template. The manifold higher number of distance restraints for the protonated sample strongly improves the precision and accuracy of the NMR structure calculations. Compared to the conventional restraints, exact solid-state restraints enable a further increase in both precision and accuracy.

number of restraints		**protonated**	**deuterated**
intraresidual	$\|i-j\| = 0$	180	0
sequential	$\|i-j\| = 1$	137	26
short-range	$\|i-j\| =$ 2-3	46	18
medium-range	$\|i-j\| =$ 4-5	7	5
long-range	$\|i-j\| > 5$	148	44
total		518	93

structure	**protonated**		**deuterated**	
	conventional	eRFDR	conventional	eRFDR
precision of backbone of the entire protein (Å)	0.75	0.43	1.10	0.86
precision of heavy atoms of the entire protein (Å)	1.34	1.02	1.80	1.56
precision of backbone for structured regions (Å)	0.52	0.31	0.85	0.66
precision of heavy atoms for structured regions (Å)	1.07	0.83	1.59	1.33
accuracy of backbone of the entire protein (Å)	1.38	0.61	2.13	1.36
accuracy of heavy atoms of the entire protein (Å)	2.26	1.45	3.04	2.37
accuracy of backbone for structured regions (Å)	0.93	0.45	1.59	1.01
accuracy of heavy atoms for structured regions (Å)	1.78	1.09	2.41	1.85

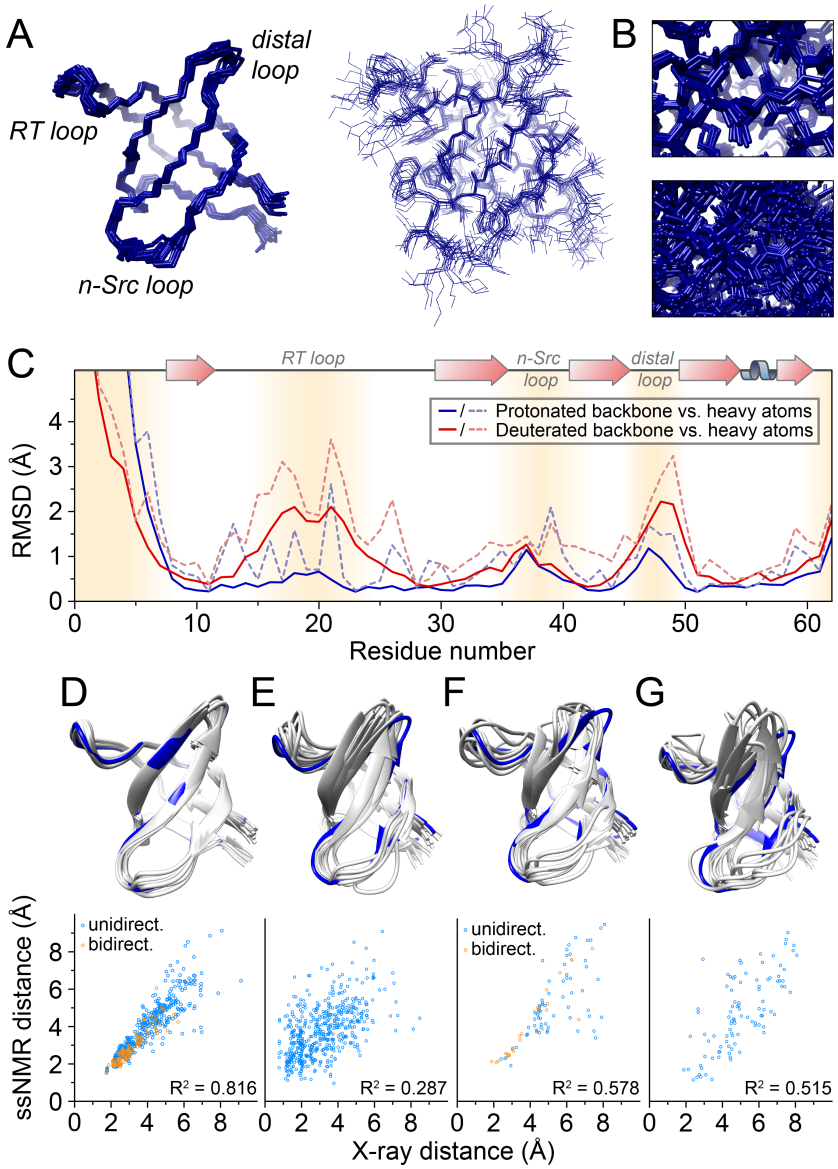

Figure 2.3: Single-state structure calculation from fully protonated samples via accurate distance restraints. (A) Backbone structural ensemble and overlay of these structures showing all heavy atoms. (B) Visual comparison of structural definition of this structure elucidation (closeup at the top) with the ensemble resulting from a deuterated/back-exchanged sample, assessed using the same procedure (bottom). (C) Structural precision as a function of residue for this structure elucidation compared to the calculation derived from a deuterated/back-exchanged sample. (D – G) Structure calculations from fully protonated (D and E) or deuterated and proton back-exchanged samples (F and G), via eRFDR restraints (D and F) or conventional restraints (E and G), with their corresponding distance correlations to X-ray data (bottom row). eRFDR restraints were determined using the combined (spin diffusion and dipolar truncation) correction method using X-ray structure 2NUZ as a template. The eRFDR structure for the fully protonated case, in conjunction with those obtained using only dipolar-truncation correction or only spin diffusion correction for comparison, has been deposited into the PDB under accession codes 8CF4, 8CHG, and 8CHH, respectively.

More importantly, as even modern state-of-the-art MD simulations are limited in accuracy for assessing dynamic structural ensembles via MD as a stand-alone technique due to inaccuracies of the force fields used and incomplete conformational sampling, combining MD simulations with experimental NMR data (either by *a posteriori* reweighting methods or by restrained-ensemble techniques) has become a gold standard of integrated structural biology in solution.[156–158,161] The availability of a dense network of exact solid-state NMR distances now enables restrained-ensemble MD simulations for solid protein preparations. Such an integrated process has important prospects for detailed assessment of conformational ensembles of the various targets whose overall molecular weight or solubility excludes solution NMR. The maximum-entropy principle is implemented in the MD ensembles by considering only the ensemble average for the comparison with the experimentally derived restraints (which are also ensemble averages). Importantly, the lattice and intermolecular contacts in ssNMR samples are often unknown and challenging to implement in the MD. In the presence of the experimental data, however, restrained-ensemble MD simulations can be carried out for the monomeric protein in solution, whereas the ssNMR restraints now guide the conformational sampling to regions of the energy landscape that are in line with the experimental conditions.

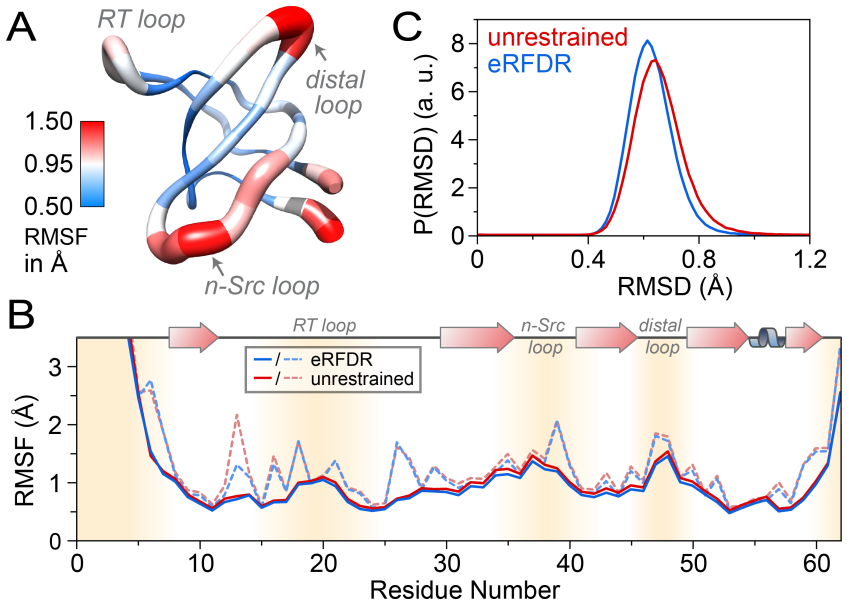

Figure 2.4: Ensemble-restrained MD simulations based on experimental eRFDR restraints. (A) Structure of SH3 with C^{α} RMSF indicated as ribbon thickness. (B) Comparison of the ensemble-averaged sequential C^{α} and heavy-atom RMSFs (solid and dashed lines, respectively) of eRFDR-restrained MD simulations (blue) to those from unrestrained MD simulations (red). (C) Normalised probability distributions of C^{α} RMSDs for residues 7–61 with respect to X-ray structure 2nuz after energy minimization and equilibration, comparing unrestrained MD simulations (red) and an ensemble of MD simulations restrained by exact solid-state restraints (blue). All restraints accord to eRFDR data from the fully protonated sample. Final frames reached within the restrained MD simulations are deposited in the PDBDev under accession code PDBDEV_00000199.

The MD ensemble comprised 16 replicas, which were subjected to bias potentials on the proton–proton distances according to the experimental eRFDR restraints from the protonated sample. MD simulations were performed over a total of 2.56 µs ($16 \cdot 160$ ns) with the Amber ff99SB*-ILDN protein force field[97,184] together with the TIP3P water model[185] (see the section 2.4 for more details). The number of distance restraint violations is very small (15 of 518 restraints are violated on average, with an average absolute violation (over all 518 restraints) of < 0.001 nm; sum of all violations, 0.388 nm). In Figure 2.4, the root-mean-square fluctuations (RMSFs) of the atoms around their mean position from the hybrid NMR/MD assessment are plotted, reflecting atomic-resolution dynamics in the backbone (solid lines) and the entire protein including side chains (dashed lines). In fact, the eRFDR-restrained MD ensemble shows sequential RMSFs that are highly similar but slightly lower than those of the corresponding unbiased MD simulations, with average decreases of 6.2 % and 6.4 % for backbone and all-heavy-atom fluctuations, respectively. The same applies for RMSDs with respect to the X-ray structure (Figure 2.4C). Despite the tight restraints, the differential behaviour of the individual atoms and the relative mobility of different structural elements are highly consistent (correlation coefficients of 0.999 and 0.996 for backbone and all-heavy-atom RMSFs, respectively, between the restrained and unrestrained ensemble (see also Figure 2.13)), demonstrating that an artificial "over-restraining" to a static structure is avoided. Most interestingly, the data confirm the high degree of similarity of dynamics for this protein domain in a crystal and in solution, complementing previous experimental studies[161,186,187] with atomic-resolution spatial details. By contrast, larger and more flexible proteins are expected to be more sensitively impacted by their lattices, which can be brought to light by the integrative modelling approach outlined in this work.

Here we have shown that high-accuracy ssNMR distance restraints for solid proteins for integrated structural biology are afforded by combining fully protonated proteins, facilitated at MAS rates in the 100 kHz regime, and time- and context-dependent assessment of ^{1}H–^{1}H magnetization transfer. In addition to the small amount of the sample required (a single 0.5 mg nondeuterated sample), the procedure, which has also been made available via the Bruker "Dynamics Center" in a semiautomated fashion, enables very high numbers of accurate restraints, outcompeting the advantages of partially deuterated samples, and allows for ssNMR-restrained ensemble MD to elucidate experiment-driven spatial dynamics in solid protein samples in atomic detail. For the small globular protein used as a case example in this proof of concept study, the protein dynamics are confirmed to be highly similar between solution and crystal. Conversely, for manifold future targets, including supramolecular assemblies, fibrils, and membrane proteins, whose plasticity is functionally modulated by the complexation/aggregation state, ssNMR-restrained ensemble MD simulations will be key to elucidating diverse biological questions, possibly also addressing the challenge of in silico reconstruction of the explicit crystal lattices.[188–190]

2.4 Supporting Information

2.4.1 NMR Acquisition and Processing Parameters

Experiments were performed using an 800 MHz Bruker Ascend Avance NEO spectrometer equipped with a CPMAS 0.7 mm H/C/N/D standard bore probe at 110 kHz MAS frequency. Assignments of the protein backbone were used from previous studies.[191] Sidechain resonances were initially assigned using a 3D S2B pulse sequence[192] with MOCCA-mixing[193] between carbons, and later completed using a 5D S2B combined with a simultaneous 4D HCCH in a 700 MHz Bruker Ascend Neo spectrometer with an H/C/N standard bore probe.[194] In total, 72.1 % of all side chain carbons and protons were assigned. Not counting aromatic and oxygen-bound protons, which cannot be accessed with the assignment spectra employed, this translated into a fraction of 82.3 % of the sidechain carbons and protons. In addition, 36.8 % of all sidechain amides could be assigned. In order to decrease the recycling delay d1 to 0.5 s, $[CuEDTA]^{2-}$ was added for paramagnetic relaxation enhancement as described before.[195] NMR chemical shifts have been deposited in the Biological Magnetic Resonance Bank (BMRB) under accession code 34785.
RFDR experiments were recorded as a series of 15 3D-experiments with ^{1}H-^{1}H homonuclear mixing times from 150 µs to 1.9 ms, using a time-shared ^{15}N/^{13}C-edited H-RFDR-h(N/C)H RFDR pulse sequence.[167] These experiments were recorded with 160 points in each indirect dimension with an experimental time of approx. 38 hours for each spectrum and two weeks in total. (Note that the high number of mixing time points was used for method development purpose and would probably be cut down to 4-5 in future applications. In order to run the integrated data analysis as performed by the Bruker "Dynamics Center", a minimal number of three spectra are required, and very similar results to the 15 points were obtained using 5 mixing times, see Table 2.2 below.) The indirect ^{1}H dimension was covering a spectral width of 11.5 ppm. In the time-shared ^{15}N/^{13}C-dimension, the ^{15}N resonances were shifted by 1500 Hz relative to ^{13}C, the aromatic ^{13}C resonances were folded, such that in total 80 ppm spectral width was sufficient. Equivalently, ^{15}N and ^{13}C encoded pathways can be separated via phase differences for heteronuclear pulses in the framework of an additional loop. Even for short and intermediate mixing times, high quality spectra with minimal peak overlap were obtained (compare Fig. 2.5).

2.4.2 Automated Cross-Peak Assignment

Cross-peak assignments are a general prerequisite for structure calculations and need to be determined by methodology outside the scope of this work. In the case of ambiguity in overlapped spectra, higher-dimensionality options have been proposed.[167,169,170] Here we fast-tracked this step by automated cross-peak assignment in ARIA[136] aided by the known structure of the protein. For the deuterated SH3 sample the RFDR spectrum with the longest mixing time and for the fully protonated SH3 sample an intermediate mixing time (1.008 ms) were used. Peak intensities were squared for compensating the

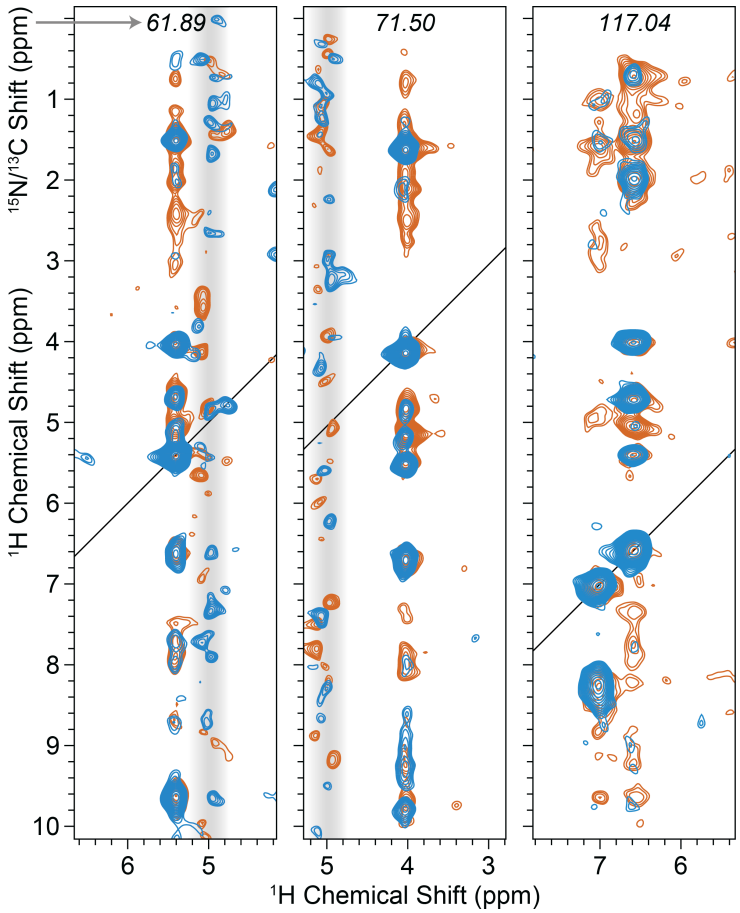

Figure 2.5: Strips of a time-shared RFDR experiment (diagonal at T24 H^{α}, H^{β} and H^{N}) for a fully protonated sample of the SH3 domain of chicken α-spectrin at 110 kHz MAS (blue: 0.288 ms mixing time, orange: 1.008 ms) as shown in main text Fig. 2.1A but cut out with a larger proton spectral width. The noise at around 5 ppm in the direct dimension (gray shaded areas) is t_1 noise from water, which could not be completely suppressed.

calculation of distances from the peak volume to the power of $-1/3$ instead of $-1/6$ from solution state NMR. This does not compensate the averaging of ambiguous restraints. However, ambiguous restraints are not used for following investigations. Also, for unambiguous distance restraints involving methyl groups, the errors in distances due to incorrect averaging of methyl protons are expected to incur a negligible degree of incorrect assignments due to the applied distance tolerances. Hence, the incorrect averaging can be neglected, especially when considering that effects like dipolar truncation and spin diffusion have a much stronger effect on the resulting distances. The automated assignment with ARIA was performed only based on dihedral angle restraints from TALOS-N[5] and distance restraints from automated assignment in an iterative fashion for improving the spectral assignment (3 successive runs for the fully protonated sample and 2 successive runs for the deuterated sample). The first run was performed with the last 6 iterations of the standard protocol plus water refinement using 2NUZ as template for calibration for the first iteration. The following ARIA runs based on the dihedral-angle restraints from TALOS-N, the preliminary assigned RFDR spectra, and the structure ensemble

from the previous run for calibration. They were only performed for the last iteration plus water refinement for retaining the reached level of ambiguity. In all iterations, 100 structures were calculated, and the best 25 structures were used as input for the next iteration. The 10 lowest-energy structures were finally refined in water. The calculations were performed with 20,000 high-temperature steps, 60,000 and 50,000 steps for the first and second cooling procedure, respectively, and 60,000 refinement steps, torsion angle dynamics during the high-temperature step, a log-harmonic potential, and – because the maximum range of magnetization transfer reached with RFDR experiments is higher than for NOESY experiments in solution state – the distance cut-off was set to 8 Å for the deuterated sample (for the protonated sample the default distance cut-off of 6 Å was used). The spin diffusion correction was deactivated. For the automated assignment of the protonated/deuterated sample, chemical-shift tolerances of 0.05 ppm/0.03 ppm and 0.08 ppm/0.04 ppm were used for the direct and indirect proton dimension, respectively, and 0.5 ppm/0.4 ppm for the heteronuclear dimension.
The peak lists obtained from the final ARIA run were filtered (see below) and only peaks that led to unambiguous distance restraints, their corresponding symmetry peaks designated to be merged during the ARIA runs, and diagonal peaks, which were mostly manually assigned, were retained. The automated assignments were inspected, and additional manual assignments were added. Intermolecular distances were not taken into account as usually these can only be identified for mixed-labelled samples.

2.4.3 General Workflow for Exact Solid-State NMR Distances

The workflow consists of the following elements: (1) Compensation of differential polarization transfer efficiencies during the heteronuclear transfers as well as pulse imperfections during mixing via normalization of cross peak volumes to their diagonal-peak volume in the same spectrum, (2) fitting of build-ups with Bessel functions, (3) averaging of the fitted interaction strength in those cases where cross peaks are present from both magnetization transfers $i \rightarrow j$ and $j \rightarrow i$, and (4) correction for dipolar truncation and spin diffusion based on an (iteratively obtained) starting structure, as detailed below.

Elimination of Unreliable Fitting Results from the Dynamics Center

In order to eliminate results from unreliable fitting, an acceptable range for the apparent interaction strength D was defined. For the analysis described here, a minimum value for D of 200 s^{-1} , a maximum value of 6900 s^{-1}, and a maximum percental error for D of 150 % was obtained. Very slow build-ups (low apparent interaction strength D) result in high fitting errors and thus, as the bounds of the distance restraints are calculated based on the fitting error, huge boundaries result, so that the distance restraint is of no use. Conversely, for very high D values, build-ups are likely to be overfitted, also resulting in false distances. It is therefore justified to set a lower and an upper cut-off for the value D, which is possible in the “Distances tab” of the tool “Exact SolidState Distances”.

Correction for Dipolar Truncation and Spin Diffusion

Spin diffusion correction was implemented in a simplified fashion derived from the eNOE framework in eNORA[62] (three-spin approach), where the observed peak intensities are scaled down before fitting by the factor F using a starting structure obtained from a conventional approach (see below) with

$$F_{\mathrm{corr}}(t) = \frac{I_{\text{2-spin}}(t)}{I_{\text{2-spin}}(t) + \sum_i I^i_{\text{3-spin}}(t)} \tag{2.3}$$

In contrast to NOE magnetization transfer, no comparably elaborated time-*dependent* theory for solid-state NMR exists so far, so that a time-*independent* approximation for correction of the apparent interaction strength was developed. The approach relies on the heuristic assumption of an additive contribution of transferred magnetization from indirect pathways (I $\rightarrow$ K $\rightarrow$ S) via third spins to the build-up on spin S with an effective indirect contribution approximated by the weaker one of both contributions (I $\rightarrow$ K or K $\rightarrow$ S), as the weaker interaction constitutes the limiting factor:

$$F_{\mathrm{corr}} = \frac{D_{\mathrm{IS}}}{D_{\mathrm{IS}} + \sum_k \min(D_{\mathrm{IK}}, D_{\mathrm{KS}})} \tag{2.4}$$

This simplification, which merely yields tendencies of spin diffusion contributions for different spin densities, turns out to be sufficiently robust to allow conversion of the procedure. Similarly as shown for solution[62,196,197] as well as solid-state NMR previously,[180] such iterative calculation still converges to the correct structure even with a wrong fold of the starting structure, demonstrated here again in Fig. 2.10 below.

The distance cut-off for spins taken into account should be kept to appropriate boundaries of magnetization transfer (around 8 Å for deuterated and 6 Å of protonated samples). For values strongly exceeding this, the assumed contribution from an increasing number of theoretical transfer pathways may artificially diminish the final transfer contribution, thus inflating the determined distances and potentially causing structural distortions.

In addition, in solid-state NMR, nearby third spins and especially the nearest neighbour spin of a spin I strongly modulate a two-spin build-up in a geometry-dependent way that at current poses significant analytical and numerical obstacles. The specific features (e. g., oscillations) of the build-up are modulated, in addition to a reduced overall amount of magnetization being transferred to spin S due to a (geometry-dependent) dipolar truncation. To pursue correction for these distortions, we establish a numerical-simulation-based method including I and S spins as a source and sink of magnetization, in addition to one third (main effector) spin in a multitude of possible geometries. In total, around 17,500 theoretical 3-spin-scenarios of RFDR transfers from I to S with a third-spin K were simulated for distances r_{IS} and r_{IK} ranging from 1.5 Å to 8.0 Å in steps of 0.25 Å and angles α_{SIK} ranging from 0° to 180° in 7.5° steps, respectively. In addition, two-spin scenarios for all I, S distances covered were simulated. The RFDR simulations were performed via

SIMPSON,[182] using the crystal file rep2000, a magnetic field strength of 800 MHz and an MAS rate of 111 kHz. All simulated build-ups of intensity I were again fitted (analogously to the experimental data) with a linear combination of Bessel functions of first kind for determining an apparent interaction strength D:

$$I(D \cdot \tau_{\mathrm{mix}})) = A_1 \cdot \left\{ 1 - \left[J_0 \left(D \cdot \tau_{\mathrm{mix}} \right) \right]^2 + \sum_k \frac{2}{16k^2 - 1} \left[J_k \left(D \cdot \tau_{\mathrm{mix}} \right) \right]^2 \right\} \tag{2.5}$$

The correction factors are derived as a ratio between the apparent interaction strength derived via SIMPSON from a 2-spin simulation and that obtained from fitting the 3-spin build-up. For the four nearest neighbours found in the template structure for a given spin, the four correction factors are selected for which the Euclidean distance determined by

$$\sqrt{\left(r_{\mathrm{IS}}^{\mathrm{template}} - r_{\mathrm{IS}}^{\mathrm{lookuptable}} \right)^2 + \left(r_{\mathrm{IK}}^{\mathrm{template}} - r_{\mathrm{IK}}^{\mathrm{lookuptable}} \right)^2 + \left(r_{\mathrm{KS}}^{\mathrm{template}} - r_{\mathrm{KS}}^{\mathrm{lookuptable}} \right)^2} \tag{2.6}$$

is minimal, i.e., the simulated spatial arrangement coincides best with the architecture in the template structure. Of the four selected correction factors only the lowest one, i. e., the one where the dipolar truncation has the strongest effect, is used (compare S6 and S7). Both methods (spin diffusion and dipolar truncation correction) can also be used together. Note that the Bessel function, strictly appropriate only for two-spin fits, is used for both, the experimental build-ups (in reality comprising multiple third-spin modulations) and the simulated (three-spin) build-ups. In the framework that one strongest third-spin has a main influence on the build-up, a reasonable similarity of the multi-spin and the three-spin scenario is expected. Fitting these curves consistently with the above function allows the match between experimental (multi-spin) and simulated three-spin build-ups. Finally, as within the simulations, both two-spin and three-spin build-ups are again fitted consistently using Bessel fits, selection of an appropriate correction factor derived from the simulation becomes possible, eventually correlating the complex experimental fit with the associated I – S proximity in the matching simulation. The correlation between obtained distances and the crystallographic distances leaves some room for improvement, which in part reflects the added third-spin contributions (within fittable build-ups) not addressed in the routine. By contrast, however, without the above corrections, the correlation between determined and crystallographic distances remains extremely poor. In addition, whereas this procedure certainly is simplistic, neither more complex fitting of the experimental data nor more complex simulations of the different geometric constellations within the protein are currently realistic. Hence, the approach constitutes a pragmatic but reasonably accurate match between experimental data and geometry correction.

As the proportionality of the apparent interaction strength to the fitted build-up is corrected subsequently by a calibration as explained in the next section, only the proportion of the fitted D values to each other matter. Note that for a different field strength or MAS regime, the same table can be used, because on the one hand, the field strength does not change build-up behaviour and, on the other hand, MAS rates

alter the apparent interaction strength in a proportional manner, and the corresponding proportionality constant would cancel out in numerator and denominator of the correction factor. The lookup table containing the angle α_{SIK}, distances r_{IK}, r_{IS}, and r_{KS}, and the correction factor (in this order, as tab-separated columns) is stored in Bruker/DynamicsCenter/data/DipolarTruncation_LookUpTable.txt. It can potentially be replaced with a corresponding lookup table based on other simulations, e. g. of DREAM, MIRROR, BASS-SD, MODIST etc.,[49,173,174,198] but a high similarity in the build-ups and thus good performance of the RFDR build-ups even for other homonuclear mixing schemes is expected. The method "Exact Solid State Distances" is therefore also adaptable to other NMR correlations than RFDR. Both correction methods, for dipolar truncation and spin diffusion, can be used either individually or together.

Calibration of Determined Distances

Since in terms of polarization transfer via dipolar recoupling in solid-state NMR, no detailed analytical function for derivation of internuclear distances is known, determined distances have to be calibrated with respect to distances from a template model. This is performed by linearly fitting the measured distance r and adjusting the calibration factors A_2 and B towards an optimal correlation with respect to the reference distances used.

$$r = A \cdot (D \cdot F)^{-\frac{1}{3}} + B \tag{2.7}$$

The distances used for calibration should not be strongly affected by the tertiary structure of the protein. Therefore, the tool "Exact Solid State Distances" enables interactive calibration using only distances between protons of the same and neighbouring residues. These distances are influenced mostly by the primary and secondary structure, which are defined by the sequence and by dihedral restraints generated from chemical shifts, but less by the tertiary structure, which is to be determined.
Additionally, also a distance error Δr is calculated from the fitting error ΔD:

$$\Delta r = \frac{1}{3} \cdot A \cdot (D \cdot F)^{-\frac{4}{3}} \cdot \Delta D \cdot F \tag{2.8}$$

Due to the $D^{-4/3}$ dependence, this error is usually relatively small, so that in addition, a user-defined absolute or percental error was specified. From Main Text Fig. 2.2H, containing the distance-over-distance plot, it becomes obvious that restraints with longer distances deviate more strongly from the expected distance, so that the usage of percental errors is highly recommended. Upper and lower distance bounds are calculated via

$$U = r + \max(\Delta r, \Delta r_{\mathrm{user}}) \tag{2.9}$$

$$U = r - \max(\Delta r, \Delta r_{\mathrm{user}}) \tag{2.10}$$

To avoid over-restriction of the MD simulation, the minimum absolute and percental bound with respect to the determined distance was set to 0.75 Å and 40 %, respectively.

Comparison of Simulated Build-ups of 3-Spin Systems with Those of Corresponding 2-Spin Systems

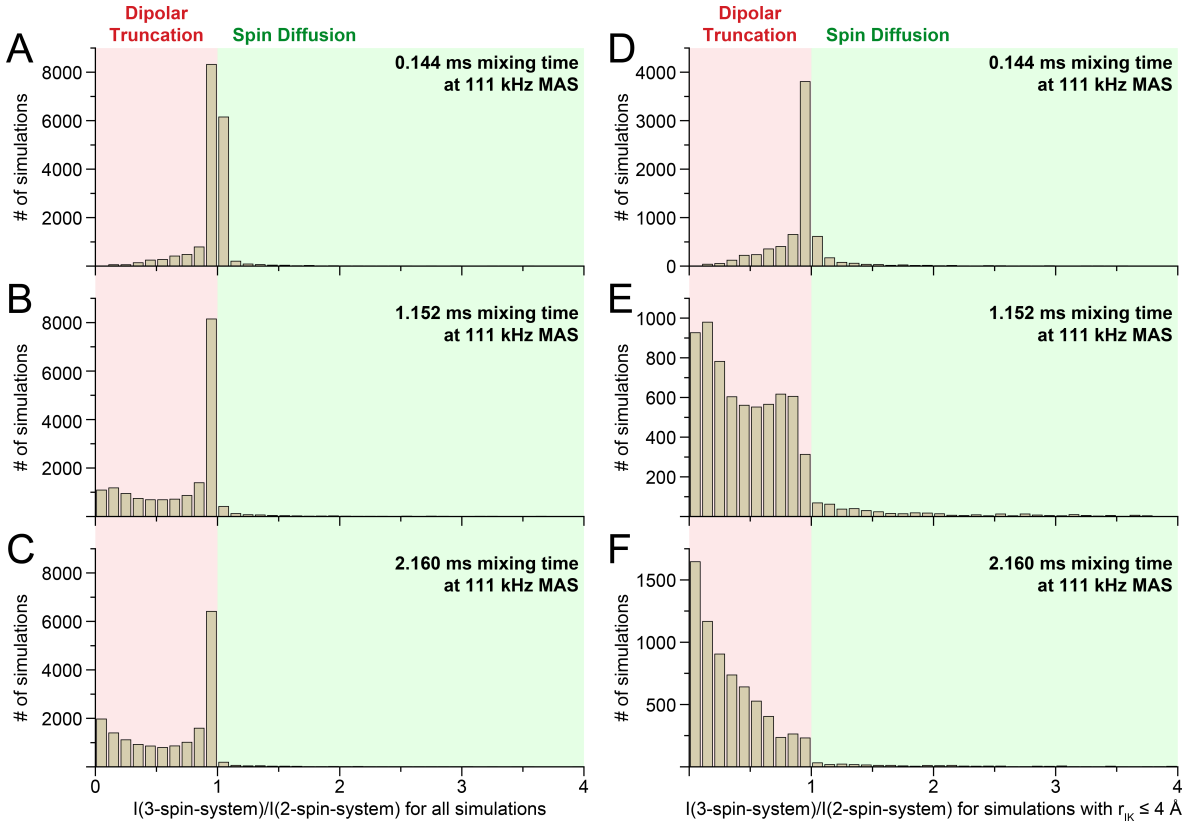

Figure 2.6: Histogram showing the intensity ratio of 3-spin and respective 2-spin systems at different RFDR mixing times [in A) and D) for 0.144 ms, B) and E) 1.152 ms, C) and F) 2.160 ms]. A-C) for all 3-spin simulations and D-F) only for close nearest neighbour distances ($\leq$ 4 Å) as expected for fully protonated proteins. Intensity ratios below 1 indicate that dipolar truncation predominates, while for integral ratios above 1, spin diffusion predominates. Intensity ratios range from 0 to 6.3 for the simulated scenarios.

For assessing the interplay of dipolar truncation and spin diffusion in this framework, intensities of the simulated 3-spin build-ups and analogously simulated 2-spin build-ups were extracted for three different mixing times (0.144 ms, 1.152 ms and 2.160 ms at 111 kHz MAS) and compared for obtaining a measure of the amount of magnetization transferred. In Fig. 2.6 histograms of the intensity ratios of 3-spin systems and the respective 2-spin systems are shown. Especially for near neighbouring spins K with distances $r_{IK} \leq 4$ Å, dipolar truncation seems to appear strongly, while for far neighbouring spin K, nearly no effect is seen. Interestingly, with increasing mixing time, the difference between 3-spin and 2-spin intensities becomes more pronounced, meaning that dipolar truncation is not overcome. In total, for none of the six histograms, spin diffusion predominates, but either dipolar truncation predominates or no predominating effect is observed. Although distances below 4 Å are typical for distances to nearest neighbouring atoms, the proton density in the simulated 3-spin scenarios is much less than it is typical for fully protonated samples, where 5 to 6 neighbour spins are present. In principle, this would amplify the

effect of both dipolar truncation and spin diffusion in different ways but overall, one would expect a maintained trend, i.e., if for the 3-spin case, one effect is already predominant, this effect is likely to be predominant also for multi-spin scenarios. Therefore, the 3-spin simulations already give a good impression about the predominating effects.

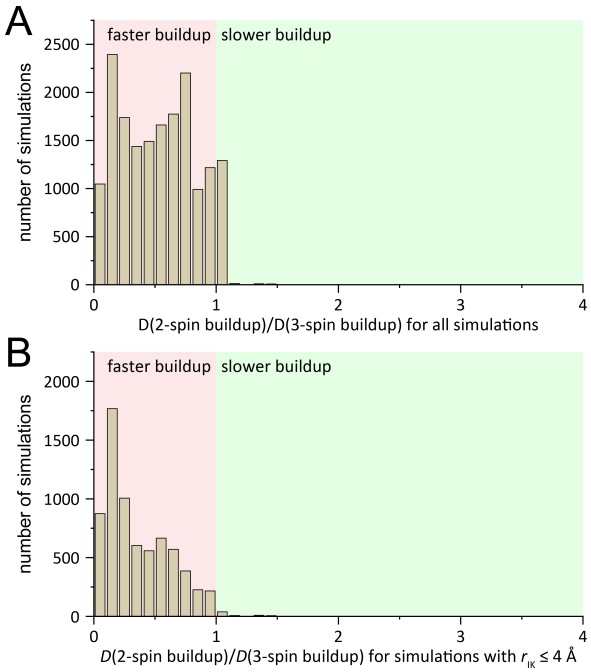

Figure 2.7: Histogram showing the ratio of the apparent interaction strengths of 3-spin and the respective 2-spin systems. A) for all 3-spin simulations, and B) only for close nearest-neighbour distances (≤ 4 Å), as expected for fully protonated proteins. Ratios below 1 indicate that the obtained fit parameter D in 3-spin systems is increased, while for ratios above 1, the obtained D values are lower. Note that the fit parameters obtained from consistent fitting with Bessel functions allow to establish pairwise correction factors for the experimental fits that are obtained likewise, but a seemingly higher interaction strength obtained does not have the physical meaning of a more effective build-up (compare Fig. 2.6 for the amount of magnetization eventually obtained on spin S).

In addition to the question whether dipolar truncation or spin diffusion predominates, it is also interesting to analyse the difference in apparent interactions strengths (represented by the fit parameter D) in the magnetization transfers (as determined by fitting with equation 2.5) in 3- and 2-spin systems (in contrast to the total amount of magnetization, as analysed above). Fig. 2.7 shows histograms of the ratio of this fit parameter for 2-spin divided by 3-spin systems, which also corresponds to the correction factor obtained in the routine for dipolar truncation (see above). It turns out that the fit parameter for the apparent interaction strength between I and S is generally higher in the presence of a third spin. This can be explained by a fast initial build-up of transferred magnetization on S (with a stronger network of interconnecting dipolar interactions) which is, however, abrogated sooner, reaching a lower "plateau" as spin S captures part of the magnetization (compare Fig. 2.6).

2.4.4 Structure Calculation

Structure Calculation Based on Conventional and eRFDR Restraints

All structure calculations were performed with ARIA using 50,000 high-temperature steps, 150,000 and 120,000 steps for the first and second cooling procedure, respectively, and 120,000 refinement steps. In addition to the distance restraints, dihedral-angle restraints obtained from TALOS-N prediction were used. For generating restraints using the conventional approach, the same unambiguously assigned cross-peaks were used as for the eRFDR approach. For the deuterated sample, the RFDR spectrum with the longest mixing time (4.03 ms) at 55 kHz MAS was used and for the fully protonated sample, an RFDR spectrum with an intermediate mixing time (1.01 ms) at 111 kHz was used. In case of bidirectional restraints, cross-peak volumes were averaged and in addition, cross-peak volumes were corrected by a factor of 3 or 9 if one or two methyl groups were involved in the cross-peak, respectively. For determining distances from cross-peak volumes, the cross-peak volumes to the power of $-1/3$ for the fully protonated sample were calibrated to an average distance of 4.00 Å.

Comparison of eRFDR Structure Calculation Results Using Different Correction Methods and Template Structures

In Table S1, a comparison of different structure calculations using the eRFDR and conventional approach are shown, using different correction methods and template structures for the correction of spin diffusion and dipolar truncation. For the deuterated sample, the eRFDR approach with the spin diffusion correction method delivers in general the most accurate structures. Due to the lower proton density, the dipolar truncation effect, which is mostly caused by nearby spins, is strongly reduced and the spin diffusion effect predominates. Compared to the standard approach, structure accuracies are consistently much improved for the eRFDR approach, especially when applying spin diffusion correction. The distance over distance plots only show an improved correlation (R^2 value) for the eRFDR approach when applying the spin diffusion correction or both, whereas the dipolar truncation correction on its own performs poorly.

For the fully protonated sample, the correction methods for spin diffusion and dipolar truncation lead to similarly good results when a reasonable, overall correctly folded start structure is used. The combined correction approach, compared to spin diffusion or dipolar truncation as individual corrections, shows further improvements both, when using the X-ray structure or a conventional NMR structure as the template structure for the correction terms. (Atomic coordinates of the structures based on combined correction, dipolar-truncation correction, and spin diffusion correction – based on 2NUZ as a template – have been deposited in the Protein Data Bank under accession codes 8CF4, 8CHG, and 8CHH, respectively.) When starting from an incorrectly folded starting structure for correction of spin diffusion and dipolar truncation (assuming the correct assignment is

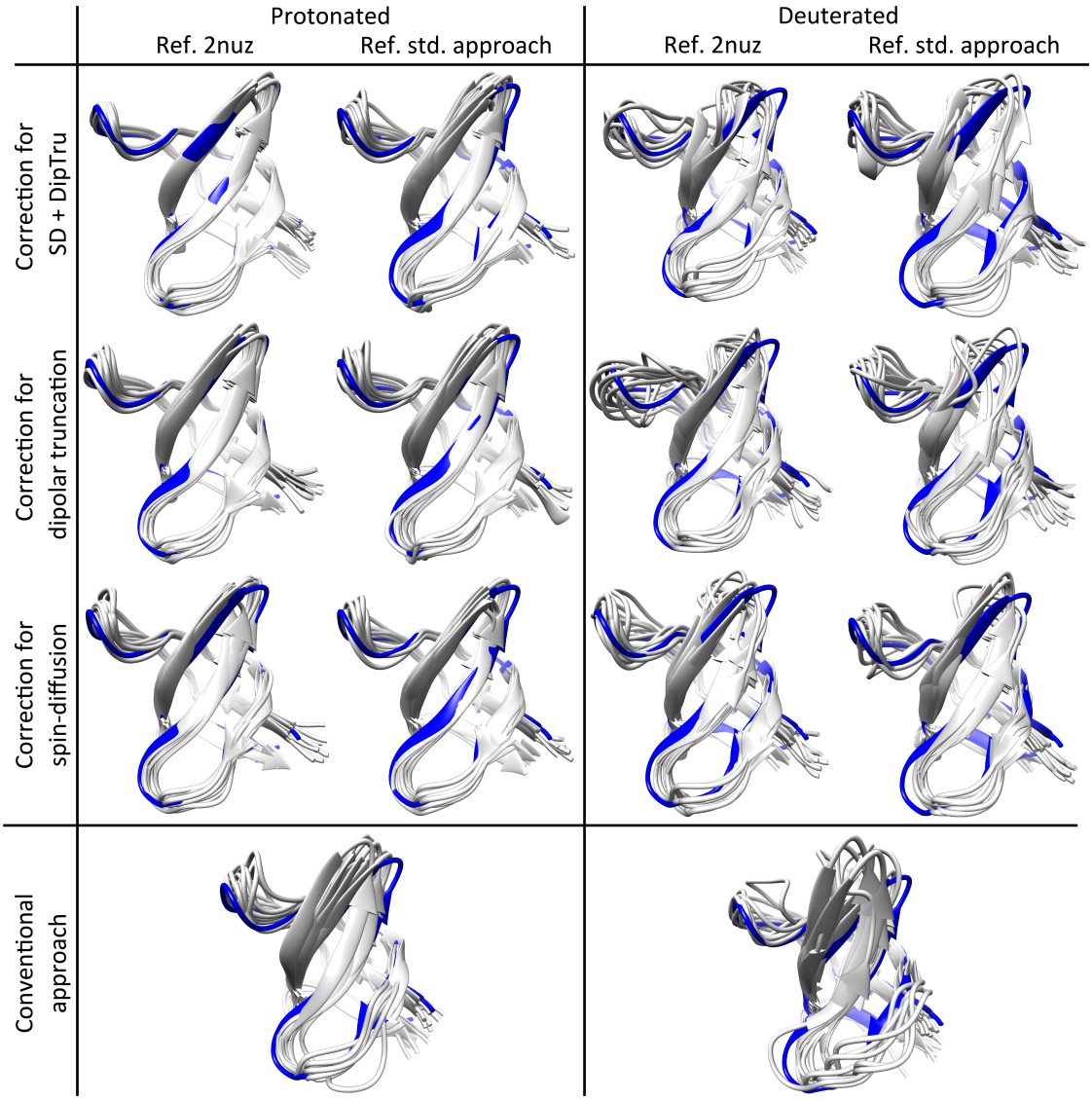

Figure 2.8: Comparison of structure ensembles calculated with the eRFDR and conventional approach using different correction methods (correction for spin diffusion, dipolar truncation, or both methods in a combined fashion) and template structures for correction (X-ray structure or NMR structure from conventional approach). Compare quantitative measures of structural quality given in Table S1.

known), the correction method for spin diffusion works better (Fig. 2.10). Adding a second iteration after this first one (using the lowest-energy structure of the structure ensemble of the first run as template structure for correction of the second run), it provides similarly good results as when using a template structure with a correct fold. This demonstrates the error tolerance of the spin diffusion correction method towards incorrectly folded template structures. In total, full protonation delivers much better precisions and accuracies (RMSD with respect to the X-ray structure) due to the manifold higher number of distance restraints and denser network of distance restraints. This becomes understandable when comparing the contact map of the fully protonated sample (also shown in main text Fig. 2.2B) with the contacts from the deuterated sample (Fig. 2.11).

Table 2.2: Precision and accuracy of the various structure calculations using different correction methods and their performance with different template structures. For elucidating whether less points in the RFDR build-up provide restraints with comparable quality, one structure calculation (marked with an asterisk) was only based on 5 instead of all 15 RFDR spectra, with mixing times equally spaced from 360 µs to 1.73 ms. Structures obtained from calculations in rows marked with indices *a*, *b*, and *c* have been deposited in the BMRB under accession codes 8CF4, 8CHG, and 8CHH, respectively. Estimation of stereochemical quality of the calculated structures was performed with CCP4 Rampage.[199]

	Approach	Correct. Method	Template Structure	Iteration	R^2	Accuracy BB3/HA		Precision BB3/HA		Rampage allowed/outlier	
Deuterated	eRFDR	combined	2NUZ	1	0.58	1.36	2.37	0.86	1.56	99.5 %	0.5 %
	eRFDR	DipTru	2NUZ	1	0.39	1.63	2.89	1.01	1.74	98.8 %	1.2 %
	eRFDR	SpinDiff	2NUZ	1	0.71	1.45	2.46	0.78	1.49	99.0 %	1.0 %
	eRFDR	combined	std. appr.	1	0.56	1.79	2.93	0.91	1.70	99.2 %	0.8 %
	eRFDR	DipTru	std. appr.	1	0.41	1.71	2.81	0.87	1.63	98.3 %	1.7 %
	eRFDR	SpinDiff	std. appr.	1	0.67	1.51	2.53	0.84	1.63	98.2 %	1.9 %
	conv.	-	-	1	0.53	2.13	3.04	1.10	1.80	98.7 %	1.3 %
Fully Protonated	eRFDR[a]	combined	2NUZ	1	0.82	0.61	1.45	0.43	1.02	99.7 %	0.3 %
	eRFDR*	combined	2NUZ	1	0.72	0.69	1.60	0.44	1.05	99.0 %	1.0 %
	eRFDR[b]	DipTru	2NUZ	1	0.71	0.75	1.73	0.47	1.14	98.6 %	1.4 %
	eRFDR[c]	SpinDiff	2NUZ	1	0.60	0.86	1.69	0.54	1.17	99.0 %	1.0 %
	eRFDR	DipTru	bad	1	0.22	1.39	2.42	0.80	1.47	98.5 %	1.5 %
	eRFDR	DipTru	result it1	2	0.43	1.27	2.21	0.62	1.21	98.6 %	1.4 %
	eRFDR	SpinDiff	bad	1	0.09	1.18	2.21	0.67	1.26	99.3 %	0.7 %
	eRFDR	SpinDiff	result it1	2	0.43	0.94	1.90	0.47	1.13	99.4 %	0.7 %
	eRFDR	combined	std. appr.	1	0.50	1.09	1.93	0.55	1.20	98.3 %	1.7 %
	eRFDR	DipTru	std. appr.	1	0.43	0.98	1.90	0.58	1.24	98.6 %	1.4 %
	eRFDR	SpinDiff	std. appr.	1	0.44	0.94	1.89	0.50	1.16	98.8 %	1.2 %
	conv.	-	-	1	0.32	1.38	2.26	0.75	1.34	98.9 %	1.2 %

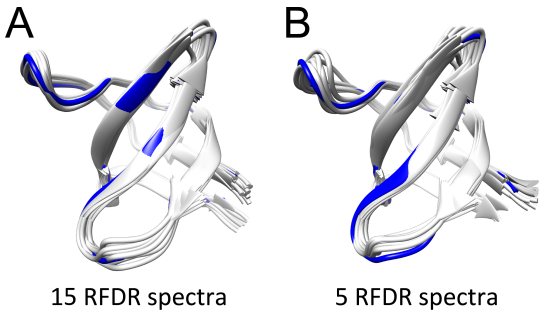

Figure 2.9: Comparison of structure ensembles calculated with the eRFDR approach using a different number of RFDR spectra (A) 15 spectra, B) 5 spectra). For both structures, the X-ray structure 2NUZ was used as reference for correction of spin diffusion and dipolar truncation. Even though fitting of build-ups is expected to be more reliable with more data points of the build-up, similar precisions (0.43 Å vs 0.44 Å BB3) and accuracies (0.61 Å vs 0.69 Å BB3) were reached (compare Table 2.2).

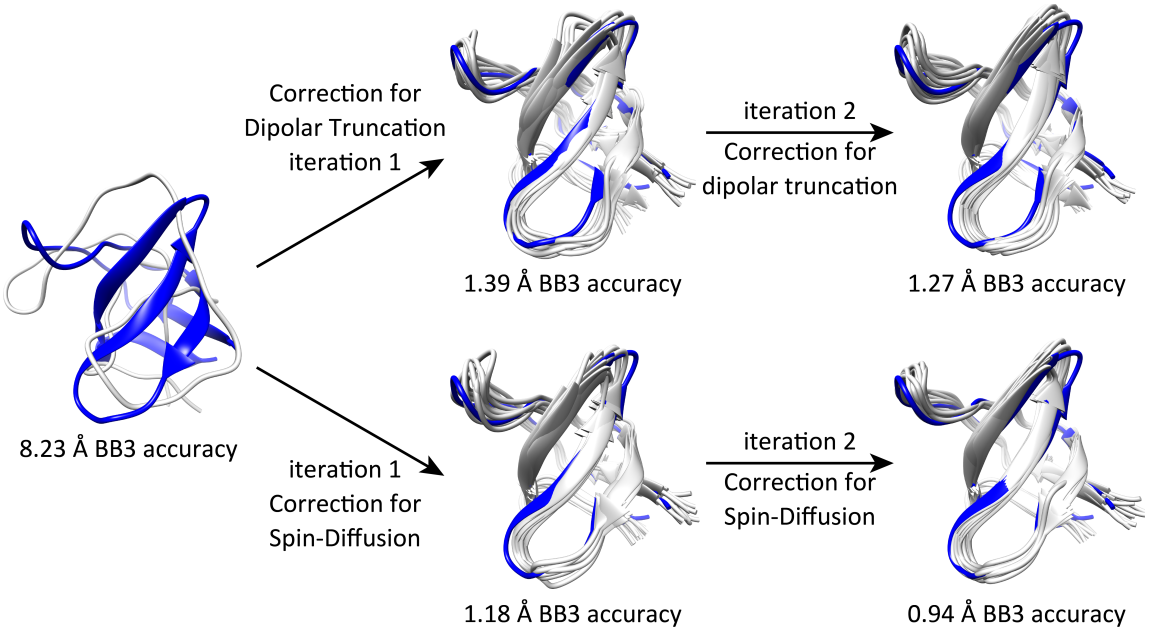

Figure 2.10: Iterative improvement of an incorrectly folded structure for correction of spin diffusion and dipolar truncation, respectively, under the assumption that the correct cross peak assignments are known. The huge improvement in accuracy, which after two iterations is in the range of the accuracy of structure ensembles calculated using a conventional NMR structure as template, demonstrates the tolerance for incorrectly folded template structures and proves that an iterative improvement is possible. Here, the correction for spin diffusion leads to a better accuracy and precision than the correction dipolar truncation. This might, however, also behave differently in other, bigger protein samples. Note that, during iterative improvement, also different correction methods can be used in each step, e.g., a first iteration using the correction for spin-diffusion and a second iteration using the combined correction method. Such combinations might additionally improve the resulting structure ensembles.

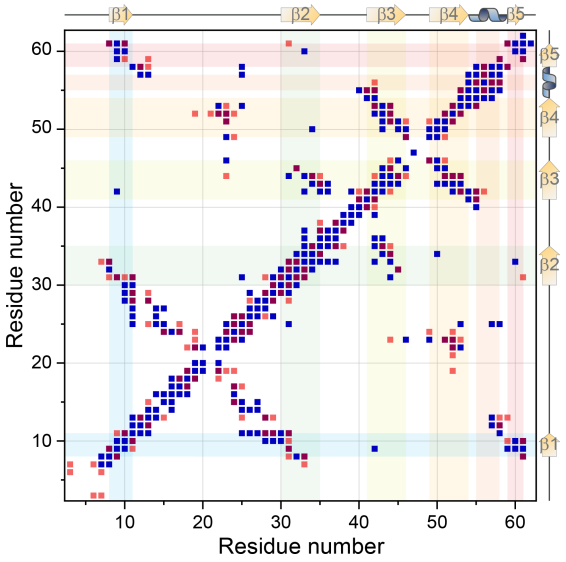

Figure 2.11: Comparison of distances covered in experiments on fully protonated (blue) vs. deuterated and amide back-exchanged samples (orange). Such contacts obtained in both approaches (overlaps) appear purple. Note that per residue pair only one dot is shown in this plot, however, multiple connections from different protons of each residue are usually obtained in the protonated samples.

Conventional Structure Determination and Artificially Truncated Data Sets

For added insights regarding specific benefits of full protonation compared to other labelling schemes bearing partial deuteration, the shift list and the completely assigned RFDR peak lists obtained from previous ARIA runs of the fully protonated sample were truncated in two ways. First, all amide and ILV methyl protons (and the corresponding magnetization transfers between them) were kept ("ILV labeling"), and in a second set, only the amide protons and the transfers between those were kept. The peculiarity is the reduction of restraints without the reduction of dipolar truncation that applies for a physically deuterated sample, hence allowing a direct assessment of the changes in data quality associated with linewidths and dipolar truncation, in particular with the real deuterated/$^{1}H^{N}$-back-exchanged one. The obtained shift and peak lists were used as input for a conventional structure calculation with ARIA: During each iteration, 100 structures were calculated of which 30 were kept for the next iteration. 50000 steps were applied in the high-temperature phase (given as number of Cartesian MD steps), 120000 in the first and 96000 in the second cooling stage. The refinement was performed in 80000 steps. All other settings were kept as described above in the chapter about the automated cross peak assignment.

The data truncated to only amides does not converge to a reasonable fold (18 distance restraints only), the data set including amides and methyls (74 restraints) does converge, but to a structure of considerably poorer quality than the fully protonated structure calculation (530 restraints). By contrast, the actual (physically) deuterated and amide-back-exchanged sample resulted in 108 restraints and useful quality even for the traditional

procedure of structure calculation (using a single mixing time) and with similar precision and accuracy as for the data set artificially truncated to amide and methyl protons only (ILV labelling). These data are shown in Fig. 2.12 and Table 2.3. This clearly demonstrates the drawback of significant peak overlap and broader lines associated with higher proton density at 110 kHz MAS. Indeed, in the fully protonated sample, only 8,0 % of the 6592 potential restraints become utilizable for the eventual structure calculation, whereas in the deuterated and back-exchanged sample at 60 kHz MAS, 35 % (108 restraints) of all potential restraints are obtained. In spite of the relative decrease of restraints, the absolute number of restraints, however, which is almost five time higher than for the latter, creates the benefits for full protonation that are described in more detail in the main text.

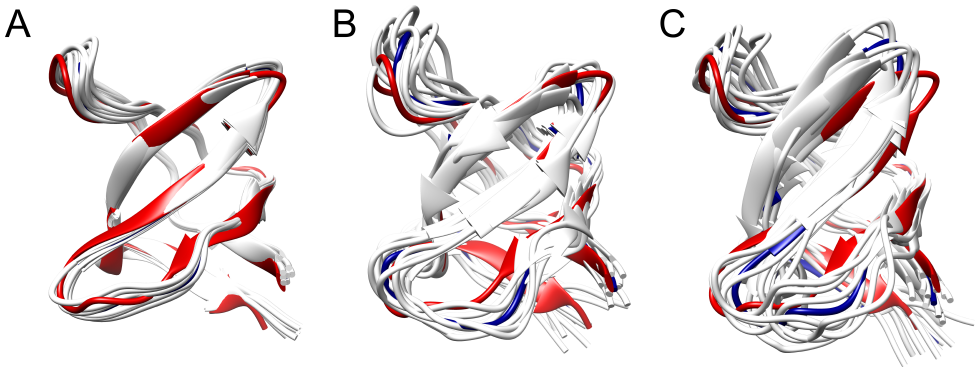

Figure 2.12: Comparison of ARIA structure calculation results originating from data of the fully protonated sample without truncation (A), with artificial truncation to amides and the methyl groups corresponding to ILV labelling only (B), and originating from untruncated data from the deuterated and amide back-exchanged sample (C) (no methyls, but sharper lines and less dipolar truncation). The average structure is overlaid in blue and the X-ray structure 2NUZ in red. The structure calculation based on the fully protonated sample with truncation to only amides, which would correspond to a deuterated and amide back-exchanged sample, did not converge and is therefore not shown.

2.4.5 MD Simulations

MD simulations were carried out with GROMACS 2019.2[200] using the X-ray crystal structure (PDB 2NUZ) as a starting structure. The N-terminal residues Met1, Asp2, Glu3, Thr4, Gly5, and Lys6 and the C-terminal residue Asp62, which were not resolved in the X-ray structure, were added using the coordinates from the solution structure (PDB 6SCW). After adding hydrogens to the protein and crystal water heavy atoms, the system was solvated with water molecules in a periodic dodecahedron simulation box. The overall charge was neutralized by adding sodium ions. The Amber ff99SB*-ILDN protein force field[97,184] was used together with the TIP3P water model.[185] Lennard-Jones 6,12 interactions were smoothly shifted to zero at a cut-off distance of 1.0 nm. This distance was also used for switching between short-range and long-range Coulomb interactions, which were treated with the particle mesh Ewald method[201] with default settings in GROMACS. The systems were energy-minimized (with steepest descent) and equilibrated for 500 ps in

Table 2.3: Comparison between conventional ARIA structure calculation results based on nontruncated data sets from the fully protonated sample, (physically) deuterated and amide back-exchanged sample, and partially truncated data sets originating from data of the fully protonated sample by sparing out part of the contacts. Accuracy and precision were determined for residues 7-61 as RMSD of all backbone atoms and all heavy atoms (in brackets). Accuracies were determined with respect to the X-ray structure 2NUZ.

	Fully protonated	**Truncated to ILV/back-ex.**	**Truncated to deut./back-ex.**	**Phys. deut./ back-ex.**
Accuracy [Å]	1.04 ± 0.09 (1.89 ± 0.10)	2.34 ± 0.23 (3.32 ± 0.23)	18.86 ± 3.58 (19.08 ± 3.45)	2.32 ± 0.43 (3.38 ± 0.37)
Precision [Å]	0.57 ± 0.06 (1.36 ± 0.05)	1.77 ± 0.12 (2.71 ± 0.12)	11.53 ± 1.03 (12.09 ± 0.98)	2.01 ± 0.16 (2.99 ± 0.13)
E [kcal mol^{-1}]	-2417 ± 91	-2164 ± 68	-2058 ± 69	-2279 ± 107
# unambiguous distance restraints	530	74	18	108
CCP4 Rampage favoured + allowed	98.0 %	99.2 %	98.2 %	99.5 %
CCP4 Rampage outliers	2.0 %	0.2 %	1.8 %	0.5 %

the NpT ensemble with harmonic position restraints on all protein heavy atoms (force constants of 1000 kJ mol^{-1} nm^{-2}). Temperature and pressure were kept constant at 300 K and 1 bar, respectively, using the thermostat of Bussi and coworkers[111] and the Berendsen barostat, respectively. All bond lengths in the protein were constrained with LINCS, and water molecules were kept rigid with SETTLE, allowing to integrate the equations of motion with 2 fs time steps.

Two sets of MD simulations were carried out, unrestrained MD and restrained-ensemble MD. In the unrestrained simulations, 16 individual simulations of 160 ns each were initiated using different random seeds for generating the initial atomic velocities from a Maxwell-Boltzmann distribution at 300 K. To ensure comparability, the ensembles in the restrained-ensemble MD simulations comprised of 16 replicas, each of which was simulated for 160 ns at constant 300 K temperature and 1 bar pressure (as described above). The eRFDR-derived upper distance bounds were converted into distance restraints and included in the GROMACS protein topology. In total, 518 distance restraints were used. Flat-bottom harmonic distance restraints were used as biasing potentials. Back-calculation of the proton-proton distances included r^{-6} averaging over the ensemble, and the restraints were only applied if the averages calculated from the MD ensemble deviated from the experimentally derived values. The force constants of the flat-bottom harmonic restraining potentials were set to 1000 kJ mol^{-1} nm^{-2}. In principle, one can argue that for the ssNMR-derived restraints, r^{-3} averaging would be more appropriate. However, this was not done because it is not yet implemented in GROMACS, and the difference between r^{-3} and r^{-6} averaged distances is very small (for example, summed over all 518 restraints,

the violations are 0.387 nm and 0.323 nm for r^{-3} and r^{-6} averaging, respectively – for comparison, linear averaging would yield 0.536 nm). For the analyses (but not for the averaging in the ensemble simulations themselves), r^{-3} distance averaging was applied. The extent of dynamics witnessed in all MD simulations correctly reflect those from dedicated relaxation/recoupling experiments (Fig. 2.13). Final frames reached within the restrained MD simulations are deposited in the PDBDev under accession code PDBDEV_00000199.

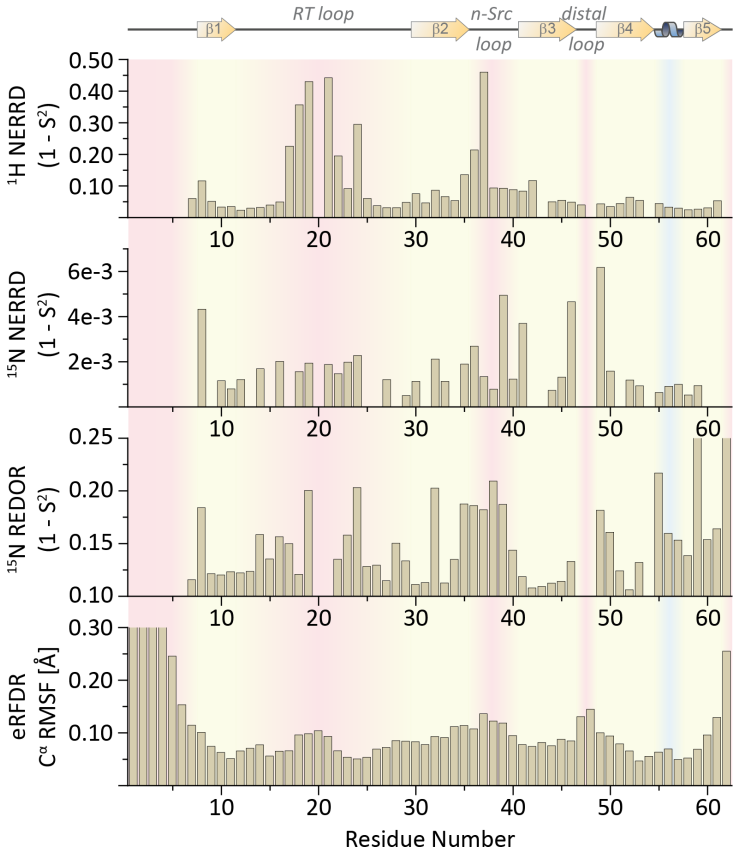

Figure 2.13: Comparison of experimental information about protein dynamics from NMR relaxation studies (NERRD and REDOR data)[186,202] with the ensemble-averaged sequential C^{α}-RMSF of the MD simulation restrained with eRFDR restraints (bottom). Whereas strong quantitative correlations are compromised by differential, timescale-dependent responses of the NMR assessments to protein motion, qualitative agreement is observed. Residues generally escaping solid-state NMR analyses like the N-terminus and residues 47-48 are too dynamic for cross polarization transfer of magnetization. Residues 20 and 54 are prolines.

3

Microsecond Timescale Dynamics of a Small-Molecule Ligand within the Active Site of a Protein

The work presented in this chapter was published as J. Kotschy[†], B. Söldner[†], H. Singh, S. K. Vasa, R. Linser, *Angew. Chem. Int. Ed.* **2024**, *63*, e202313947 and J. Kotschy[†], B. Söldner[†], H. Singh, S. K. Vasa, R. Linser, *Angew. Chem.* **2024**, *136*, e202313947.
[†]: These authors contributed equally to this work.

My contributions to this work consisted in performing and analysing MD simulations of hCAII in complex with the ligand SBR crucial for gaining mechanistic insights in the µs dynamics observed by ^{1}H NERRD experiments.

3.1 Abstract

The possible internal dynamics of non-isotope-labelled small-molecule ligands inside a target protein is inherently difficult to capture. Whereas high crystallographic temperature factors can denote either static disorder or motion, even moieties with very low B-factors can be subject to vivid motion between symmetry-related sites. Here we report the experimental identification of internal µs timescale dynamics of a high-affinity, natural-abundance ligand tightly bound to the enzyme human carbonic anhydrase II (hCAII) even within a crystalline lattice. The rotamer jumps of the ligand's benzene group manifest themselves both, in solution and fast magic-angle spinning solid-state NMR ^{1}H $R_{1\rho}$ relaxation dispersion, for which we obtain further mechanistic insights from MD simulations. The experimental confirmation of rotameric jumps in bound ligands within proteins in solution or the crystalline state may improve understanding of host-guest interactions in biology and supra-molecular chemistry and may facilitate medicinal chemistry for future drug campaigns.

3.2 Introduction

Protein-ligand complexes are usually explored via crystallographic studies, which in case of structurally rigid ligands can accurately report on the underlying site-specific interactions. Conversely, for structural elements with enthalpically less restricted interactions to the target, X-ray crystallographic electron density maps often indicate heterogeneity and/or flexibility of atoms via increased B-factors.[205,206] However, motion cannot confidently be distinguished from static disorder with a single complex structure, and in particular exchange between like conformations remains elusive.[207] With cryo-electron microscopy, molecules can be trapped at different points in time to assess larger structural differences (non-equivalent conformations),[208] but timescale information is unavailable. In particular, motion in the µs regime, a hallmark of enzymatic catalysis and recognition at physiological temperatures, which for *proteins* has constituted a prime focus of structural biology, has been difficult to observe for bound *ligands*, including substrates or small-molecule messengers. Compared to their entropic losses through rigid-body motion of around 15–20 kJ/mol,[209,210] configurational entropy losses upon binding have been estimated to reach up to 100 kJ/mol.[211] Finetuning of residual ligand mobility in the bound state thus has important consequences for the overall entropic contributions ($-T\Delta S$, normally a few tens of kJ/mol) to the binding affinity.[212] Although in silico predictions have become increasingly reliable,[213] the difficulty of site-specific assessment of residual mobility by experiments has set limitations to the understanding of the dynamic properties of host:guest complexes in supramolecular chemistry and catalysis, of substrate affinity and selectivity in natural and synthetic enzymes, for predicting the thermodynamic properties of pharmacological inhibitors, and for the rational optimization of a diverse

range of compounds through targeted synthetic and medicinal chemistry.[206,212,214] While in solution, a plastic behaviour and in particular the possibility of protein breathing is known to facilitate internal motion associated with steric restrictions,[215–217] it has remained unclear whether such conformational exchange, e.g., aromatic ring flips in tightly bound ligands, is maintained within proteins restrained in a crystal lattice. Here we show that µs timescale rotameric jumps are experimentally observed for a tightly packed, crystallographically well-defined element of a high-affinity pharmaceutical inhibitor in hCAII. Using proton $R_{1\rho}$ relaxation dispersion in solution and in micro-crystals, combined with MD simulations, µs timescale dynamics of a ligand aromatic ring are revealed both in the expectedly more plastic protein in solution and within the crystalline enzyme.

3.3 Methods and Results

We prepared natural-abundance, $^{13}C/^{15}N$- and $^{2}H/^{13}C/^{15}N$-labeled hCAII (Figure 3.1A), a ubiquitous 29 kDa enzyme and well-studied, FDA-approved drug target interconverting bicarbonate and carbon dioxide, using procedures similar to those established[172,218] (see section 3.5 for details). For crystalline samples of hCAII in complex with (*R*)-N-(3-indol-1-yl-2-methyl-propyl)-4-sulfamoyl-benzamide (denoted as "SBR" in the following, according to the PDBeChem ligand dictionary, Figure 3.1B), a high-affinity binder with a K_D of ≈ 30 pM,[219] the ligand was either soaked into preformed hCAII crystals or added before crystallization (co-crystallization). As a reference to NMR-spectroscopic studies, both types of crystals were subjected to X-ray diffraction, the co-crystallized structure with a resolution of 1.46 Å revealing better-defined electron densities for SBR than the soaked sample (see Figures 3.4 and 3.5 as well as Tables 3.1 and 3.2 in section 3.5). B-factors within the SBR ligand, which estimate the atomic displacement from its average position, imply strong conformational heterogeneity or flexibility near the indole group (B-factors up to 32 even for the co-crystallized complex). By contrast, the structural elements close to the Zn^{2+}-bound warhead (≈ 14) and the benzene (≈ 17) appear highly defined. Figure 3.1C and D show the electron densities as well as the temperature B-factors, respectively, for the SBR ligand bound to hCAII upon co-crystallization. Figure 3.1E represents the dense packing of its benzene ring in the pocket.

We wondered whether there is any "hidden" motion of SBR in the hCAII complex crystals for the crystallographically well-defined aromatic ring, similar to observations of aromatic ring flips in amino acid side chains recently identified by ^{13}C relaxation dispersion.[190,221] First, solution NMR characterization of the complex with a $^{13}C,^{15}N$-labelled protein was pursued after concentration of the sample in 50 mM phosphate buffer to ≈ 8.5 mg/ml and titration of an equimolar amount of SBR into the solution. Figures 3.6 and 3.7 show the chemical-shift perturbations (CSPs) of hCAII *amide peaks* upon addition of SBR. Regions, in which the highest CSPs for backbone spins cluster, comprise amino acids 125–142 and 198–205, both close to the binding pocket. In agreement with the X-ray data, the titration

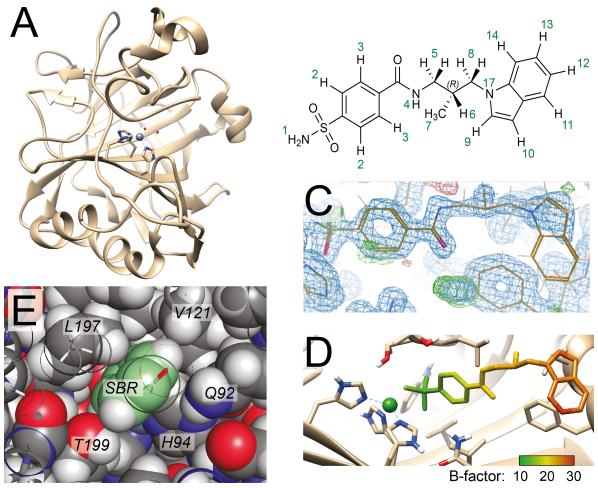

Figure 3.1: The hCAII-SBR complex. (A) hCAII crystal structure in apo form (PDB 3GZ0).[220] (B) Ligand SBR, depicting sequential numbering colored in green as used throughout this study. Protons 2 and 3 are called benzylic protons and protons 9–14 indole protons in the following. (C) Ligand electron densities in X-ray crystallographic data of the co-crystalline sample at cryogenic temperatures, deposited under PDB ID 8R1I. (D) Representation of ligand B-factors of this sample (color code below). (E) Packing of the benzene moiety of SBR (green) deep inside the active-site pocket.

data confirm a single binding site of the ligand. Importantly, the data also confirm a residence time longer than the NMR timescale,[222] as expected from the dative bond between warhead and Zn^{2+}, which avoids interference of ligand association/dissociation with the relaxation dispersion studies in the following. To further rule out contributions from potential backbone motion in such data, we performed ^{15}N relaxation dispersion, reporting on (local) shift fluctuations for the amides (Figure 3.8A). In consistency with previous work employing two other sulfonamide inhibitors,[172,218,223,224] the data confirm that any intrinsic µs timescale motion in the pocket of *apo* hCAII is quenched upon binding of the ligand. ^{1}H $R_{1\rho}$ relaxation dispersion experiments of the apo ligand on its own (Figure 3.8B) do not show dispersion either.

Access to site-specific ligand resonances and their properties in the bound state in solution without interference from protein nuclei was obtained by a combination of ^{13}C/^{15}N-filtered[225] 2D ^{1}H/^{1}H correlations (with a relay transfer) with water suppression modules and an $R_{1\rho}$ relaxation block (see Figure 3.9 for the pulse sequence) on ^{2}H/^{13}C/^{15}N-labelled hCAII. Peak assignments were transferable from spectra of the apo ligand in D_2O (Figures 3.10, 3.11, and 3.12). We recorded ^{1}H $R_{1\rho}$ relaxation dispersion profiles (Figures 3.2C and 3.13) to probe the possible µs timescale conformational-exchange dynamics of the ligand in the active site. To combat proton-proton cross relaxation, ^{1}H $R_{1\rho}$ data were recorded in a constant-time fashion using a spin lock at 35.3°, at which angle NOESY and ROESY contributions cancel.[226] In contrast to the absolute values, which are susceptible to overall tumbling, the existence of Bloch-McConnell relaxation dispersion (BMRD) found for varying spin lock field strengths can report on slow-motional internal contributions

(Figure 3.2A).[227] Figures 3.2C and Figure 3.13 show that a dispersive behaviour is indeed observed for cross peaks representing the benzylic ring, whereas expectedly, neither of the indole protons shows a clear dispersion profile. Even though the signal-to-noise ratio in the filtered experiments as well as chemical-shift differences for chemically equivalent protons (in the apo state) in different rotameric states (in the complex) and hence dispersion effects are expectedly small, the dispersion curves for part of the peaks meet the statistical criteria ($p < 0.05$) to suggest a flipping motion of the benzylic ring on a tens of µs timescale while bound to hCAII in solution. By contrast, the flat profiles of the indole protons at rather high baseline $R_{1\rho}$ rates (around $7\,s^{-1}$) denote faster motion, exceeding the timescale accessible by the available spin lock field strengths, however, still significantly slower than the dynamics of the free ligand (see apo vs. bound-ligand linewidths in Table 3.3).
Motion has been reported to be highly similar between proteins in solution and those in a crystalline setting.[66,186] Still, the lattice might reduce breathing motion and thus abrogate dynamics of tightly packed ligand moieties. To assess whether the apparent conformational exchange of the bound ligand is maintained even in the context of the crystal, we composed a proton-detected $^{13}C/^{15}N$ double-filtered ^{1}H $R_{1\rho}$-edited sequence with tailored water suppression modules for solid-state NMR at fast MAS. (see the H/N spectrum of micro-crystalline, co-crystallized hCAII/SBR complex in Figure 3.14; the ^{1}H $R_{1\rho}$-edited pulse sequence and a filtered spectrum of the ligand only are shown in Figures 3.15 and 3.16, respectively). In the solid state, assessing fast-timescale ligand motion is again compromised by the unavailability of NMR-active heteronuclei. Redfield-type auto relaxation of protons as well as assessment of slow conformational exchange by BMRD are limited by strong coherent contributions to the effective rates and small chemical-shift differences between states. However, proton near-rotary resonance relaxation dispersion (^{1}H NERRD), introduced in theory[202] and for two model proteins[187,228] recently, is responsive to µs-ms timescale changes in the regional homonuclear dipolar networks and hence relative bond orientation, rather than to changes in isotropic chemical shifts. Figure 3.2B shows a simulation of the ^{1}H NERRD effect for a range of timescales. The measurement benefits from a very high amplitude of dispersion for a natural-abundance nucleus in case of µs-timescale architectural changes in the dipolar network close to the half-rotary-resonance (HORROR) condition. On the flipside, amplitude information is blurred as the effective strength of the interaction cannot be accurately determined.[187,202]
Peak assignments of hCAII-bound SBR in the solid state were transferred from the complex in solution, tolerating deviations of up to 0.2 ppm due to likely temperature inaccuracies. Owing to partial peak overlap, four signals were assessed that in part entangle multiple protons each, namely 3/11/14 and 12/13, in addition to proton 2 and 10. Profiles of the latter two peaks, representatives of one of the aromatic systems each, are depicted in Figure 3.2D. (See the others, as well as dispersion curves of protein amides from previous work for comparison, in Figure 3.18.) Note that in contrast to the *local* (chemical-shift)

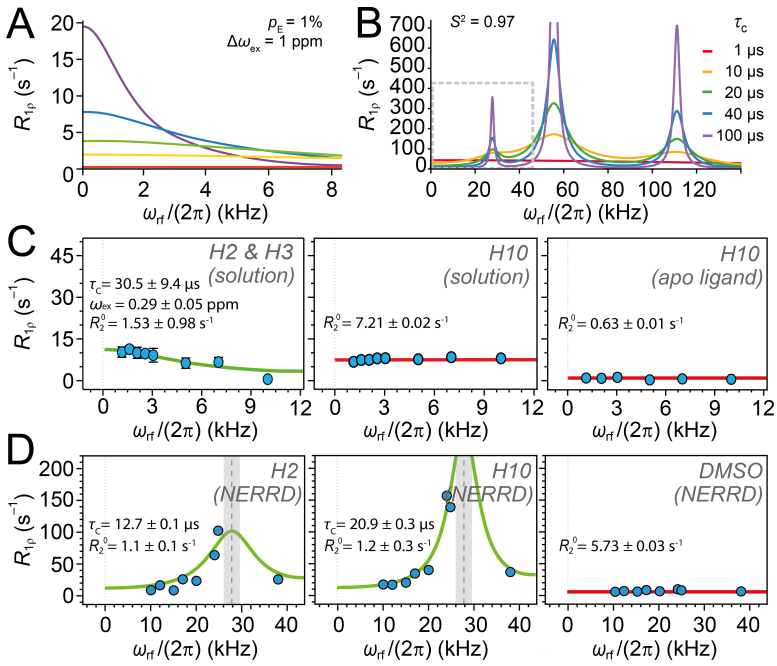

Figure 3.2: Slow-motional ligand dynamics assessed via ^{1}H ligand relaxation dispersion in solution and in the solid state. (A) ^{1}H $R_{1\rho}$ BMRD relaxation dispersion curves simulated for populations p_E of 1 % and a chemical-shift difference $\Delta\omega_{ex}$ of 1 ppm for different timescales of exchange. (B) Simulated profiles of proton NERRD at 55.555 kHz MAS for different time regimes of motion (considering homonuclear interactions with 3 Å effective distance and a CSA of 10 ppm at 700 MHz ^{1}H Larmor frequency at an S^2 of 0.97). The dashed box denotes the regime (around the HORROR condition) investigated here. (C) Ligand µs timescale motion in the complex, seen only for benzylic protons (left), opposed to motion faster than refocussed by the spin lock for indole protons (center) and much faster motion for apo ligand in general (right). See more profiles in Figures 3.8B and 3.13. (D) NERRD behavior observed in the co-crystallized hCAII/SBR sample for all discernable ligand proton signals, including the benzyl and the indole ring (left and center, respectively), but not for residual DMSO as a reference case for isotropic motion outside the µs regime (right). (See more profiles in Figure 3.18.) Spins are numbered (gray/italic) as shown in Figure 3.1B. The dashed gray lines in D denote the HORROR condition at 27.777 kHz. Note that the amplitude of the ^{1}H NERRD profiles depends on the strength of the effective proton–proton dipolar network at the site, which in the presence of motion is a time- dependent property that as of now is hard to predict (see section 3.5). This blurs an assessment of motional amplitudes and at this point reduces the quantitative assessment to a timescale information.

changes probed in BMRD, changes in a tightly coupled proton-proton dipolar network in case of µs timescale contributions for one of the rings would be felt by protons from both rings. (See a more detailed assessment below.) This is reflected in the distinct NERRD profiles, where fast µs timescale contributions, with timescales between 5 and 20 µs, variable amplitudes, and maximal $R_{1\rho}$ rates up to $\approx 200\,s^{-1}$, show for both rings. The DMSO solvent peak, by contrast, with an isotropic motion much faster than the µs regime, displays a flat profile of the fast-motional case with constantly low $R_{1\rho}$ rates of around $6\,s^{-1}$ only (bottom-right profile in Figure 3.2D).

To understand the slow motional components of the ligand mechanistically, we conducted MD simulations in GROMACS with the Amber ff19SB force field. For this purpose, the system was prepared with the AmberTools22 package, including parametrization of the zinc cofactor with the metal center parameter builder.[229] (See details in section 3.5.) Three production runs were performed for the SBR-bound form for 4 µs each in addition to 2 µs for the apo-form. In contrast to most of the protein backbone, being rather rigid over the course of the simulation (apart from the very N-terminus and residues around sites G234, A133, L84, and P21, see Figure 3.20 for root mean square fluctuations of the protein in apo and SBR-bound states), the ligand itself again displays a large spectrum of dynamics. Most of these motions are expectedly fast (ns timescale dynamics, see Figures 3.3A–C) outside the sensitivity of NERRD experiments and in line with the flat BMRD profiles for the ligand tail.

Intriguingly, by contrast, the benzylic ring, the central element between the sulfonamide warhead and the attached indole moiety, displays ≈ 180° rotameric jumps on the µs timescale, determined by its tight π-interaction with the side chain of L197[230] (Figures 3.3B, red trace, 3D, and 3.21/3.22). Consistent rotameric jumps and a short period with a 45° twist in the third trajectory are shown in Figure 3.21E. Both, the intrinsic insufficiencies of the MD force fields as well as insufficient sampling of the event in the 4 µs simulations prevent accurate determination of the timescale of the flips. However, qualitatively, their sparsity (five 180° flips in 12 µs) is well in agreement with the µs timescale observed experimentally. Both is in congruence with the spectrum of timescales observed for ring flips in aromatic sidechains of protein amino acids.[190,215,217,221]

Figure 3.3E displays three representative time-correlation functions (TCFs) for different (intra- and inter-ring) proton-proton internuclear vectors, obtained from the MD trajectory as described by Hoffmann et al.[231] (More TCFs are shown in Figures 3.23–3.25, for more details see section 3.5.) The TCFs represent time-dependent changes of the individual contributions to the local magnetic field at a given site, i.e., the fluctuating dipolar interaction strength dependent on relative movement of one proton to the other, equally affecting both protons. Due to the point symmetry of the nuclear magnetic field, after a 180° ring flip, two out of three (intra-ring) dipolar interactions for the benzylic protons impose local magnetic fields of the exact same magnitude and direction as before the flip (blue coupling partners in Fig. 3.3F; second TCF in Fig. 3.3E). Only one proton in the ring (red, with a ≈ 5 Å distance) causes changes to the local magnetic field (first TCF in Fig. 3.3E). The fast-timescale dynamics of the indole moiety average the *intra*-indole dipolar interactions on a timescale inaccessible for NERRD (third TCF in Fig. 3.3E). However, the benzylic ring flips further modulate the fast-timescale average of the local magnetic field at the indole protons by additional *inter*-ring interactions (fourth TCF in Fig. 3.3E) on the µs timescale. (This is the case when α spins in the benzyl ring (e.g., H2 or H3) interconvert with β spins (when H2' or H3', respectively, have opposite spin state)

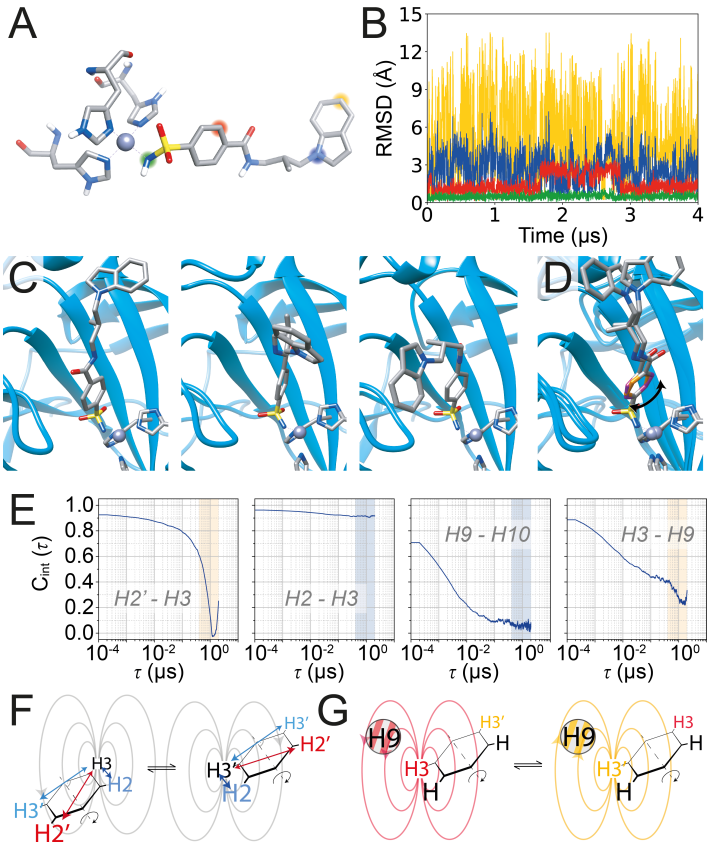

Figure 3.3: Different modes of ligand dynamics, including aromatic ring flips on the µs timescale, as seen in MD simulations. (A) Constitution of the ligand within the protein, showing colors employed in B. (B) Root-mean-square deviations (RMSDs) of individual ligand protons as a function of simulation time, red reflecting the benzylic ring flip on the µs timescale. (C) Excerpts from a 4 µs trajectory giving representative positions of the fast but non-isotropic motion of the indole ring (three snapshots) and D) the ≈ 180° ring flip of the benzylic ring (two snapshots overlayed). For visualization, the ring has different colors on different sides. (E) Time correlation functions (TCFs, shown up to half the simulation time) of different intra-ring (H2'–H3, left; H2–H3, second from the left; H9–H10, third from the left) and inter-ring vectors (H3–H9, right), the fast µs timescale regime marked in orange (present) or blue (absent). In contrast to the interaction between protons on the same side (e.g., H3 and H2, vector parallel to the rotation axis) and that between the C_2-symmetry-related protons (e.g., H3 and H3', 180° flip of the inter-spin vector), only the diagonal protons in the benzylic ring (e.g., H3 and H2') cause a flip-dependent modulation of the local fields. Whereas intra-indole vectors like between H9 and H10 represent purely fast-timescale modulations, inter-ring interactions bear fast and slow-timescale contributions modulating the local magnetic fields. (F) Visualization of the reconstitution of the (intra-ring) dipolar network after a ≈ 180° benzene flip (exemplarily shown for H3) both, with regard to the symmetry-related partner (H3') as well as for the directly adjacent proton (H2) (see text for details). (G) µs timescale inter-ring modulation of the effective dipolar network around an indole proton (top left) due to a benzene flip. Benzene flips further modulate the fast-timescale average of the local magnetic fields in the indole imposed by intra-indole dipolar interactions (and vice versa). A modulation applies whenever interchanging protons 3 and 3' (or 2 and 2' or both) have opposite spin state (here visualized by red vs. orange color).

and vice versa, i.e., in 50 % of the cases (Fig. 3.3G). By contrast, for exchange between like spin states in the benzyl ring, the effective inter-ring dipolar network is reconstituted. Given the high plasticity of the overall ligand structure (Fig. 3.3C), the effective distances applicable for inter-ring modulations of the local magnetic fields vary strongly (see Fig. 3.26). As such, the inter-ring proton-proton distances do reach below 5 Å, which distance also determines the *intra*-benzylic (H3/H2') contribution. The experimental NERRD profiles are hence well in line with the mechanistic assessment by MD.

While each of the above assessments has its technical limitations, in conjunction, the above data for the SBR ligand strongly prompt that moieties with highly defined electron densities can comprise hard-to-detect µs timescale conformational exchange motion even in the crystalline state. In structure-based drug design and medicinal chemistry, the understanding of internal dynamics for high-affinity ligands on different timescales has been limited, and its demonstration for hCAII and a sulfonamide ligand, a prototypical pair of protein and inhibitor, via NMR may open several new avenues. The quantitative information content conveyed by the individual experiments exploiting natural-abundance proton relaxation dispersion is limited by several factors mentioned along the above findings. Nevertheless, even qualitative *identification* of ligand dynamics, its temperature dependence, or its abrogation by mutations or upon changes of external conditions, may help better understand host-guest interactions. This has the mentioned consequences for the evaluation and optimization of affinity and entropy in synthetic small-molecule ligands, e.g., pharmacological binders, for supramolecular chemistry and structural biology of protein-ligand complexes but equally concerns natural or biotechnological processes (inhibitors, substrates, or products), as well as studies involving nucleotides or small DNA or RNA molecules, which pose similar limitations to isotope labelling. We expect that in particular for solid samples, with new prospects for NMR-restrained MD simulations,[155] higher magnetic fields, and increasing MAS rates,[28,166] an exploitation of ^{1}H relaxation dispersion will bring a greenfield of hitherto inaccessible applications.

3.4 Conclusion

In conclusion, we have demonstrated the experimental detection of fast µs timescale dynamics in a sterically restrained part of a high-affinity ligand while bound to its target protein, even in the context of the crystal lattice. Although a quantitative level as known for heteronuclear relaxation is yet unavailable for ^{1}H relaxation dispersion due to the difficulty of quantifying the interaction strength, the obtained profiles for the first time identify and distinguish motion on the µs timescale for a non-isotopically labelled target compound in solution and in the crystalline state, which brings about new perspectives for understanding and tailoring host-guest chemistry.

Chemical-shift assignments and X-ray data were deposited into the BMRB (ID 52196) and the PDB (ID 8R1I), respectively.

3.5 Supporting Information

3.5.1 Experimental

hCAII Expression and Purification

To express fully deuterated protein, the cells needed to be adapted to growth in D_2O. An LB starter culture of 50 ml volume was prepared and incubated overnight. On the next day, 50 ml M9* medium (1 l of sterile medium included 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 2 g ^{2}H-,^{13}C-glucose, 0.5 g $^{15}NH_4Cl$, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1 mg thiamine, 1 mg biotin, 60 mM $ZnCl_2$ and 1 µM ampicillin) was inoculated to an $OD_{660\,nm}$ of 0.05–0.10 using the starter culture. The same starting OD was always used subsequently upon inoculation. When the cell density reached an $OD_{660\,nm}$ of 0.6, a cell aliquot was transferred to 100 ml of M9* medium with 30 % D_2O. The culture was incubated at 37 °C while shaking at 200 rpm until an $OD_{660\,nm}$ of 0.4 – 0.5. The same OD value needed to be reached during all subsequent steps until the cells were transferred to fresh medium. Next, 100 ml M9* medium with 60 % D_2O were prepared and inoculated. At sufficient density, the required number of cells was used for the next incubation step at 37 °C and 200 rpm in 300 ml M9* medium with 90 % D_2O overnight.

On the next day, the final M9* medium with 100 % D_2O was prepared. An initial OD_{660nm} of 0.05 – 0.10 was prepared using an aliquot from the overnight culture with 90 % D_2O content. After inoculation, the culture was incubated at 37 °C while shaking at 220 rpm until $OD_{660nm} = 0.6 - 0.8$ was reached. Subsequently, the induction was performed by adding 1 mM IPTG and 1 mM $ZnCl_2$. From then on, the culture was incubated at 30 °C while shaking at 200 rpm for 16 h. The cells were harvested by centrifuging the suspension at 6.24 g at 4 °C for 20 min. The supernatant was discarded, and the pellet was flash-frozen in liquid nitrogen for further storage at −80 °C.

The purification of hCAII included affinity and size exclusion chromatography as well as proteolysis using thrombin. The lysis was performed with the Avestin Emulsiflex C5 using a cooled PBS buffer (139.3 mM NaCl, 2.7 mM KCl, 10.0 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 1 mM DTT) at pH = 7.5. The lysate was centrifuged at 6.24 g for 30 min at 4 °C. Next, an affinity chromatography was performed using a gravity-flow column filled with glutathione agarose beads with a binding capacity of 10 mg ml^{-1}. First, the resin was washed five times with two column volumes of distilled water and equilibrated another five times with two column volumes of PBS supplemented with 1 mM DTT at a pH of 7.5. The supernatant resulting from centrifuging the lysate was added to the beads and incubated while rotating slowly at 4 °C for one hour. Next, the resin was filled into a small manual column and washed with PBS buffer until the 280 nm absorption of the fraction reached a baseline. The protein was eluted with three times one column volume of 50 mM TRIS, 10 mM glutathione, and 1 mM DTT at pH = 8.0. Again, this step was continued until the 280 nm absorption of the fractions reached the baseline. The elution fractions

which contained most of the protein were pooled and incubated with a few milligrams of lyophilized thrombin at room temperature overnight. The sample was subsequently purified by size exclusion chromatography. The column was equilibrated with 50 mM ammonium bicarbonate at a pH of 7.7. The fractions were pooled and lyophilized.

Crystallization of hCAII

The ligand SBR was kindly provided by Prof. Dr. Klebe from Marburg University. Instruments and chemicals for crystallization were kindly shared by the crystallization facility of the MPI for Molecular Physiology in Dortmund, Germany, run by Dr. Raphael Gasper-Schönenbrücher.
The lyophilized hCAII powder was dissolved in 50 mM TRIS buffer at pH = 8.0. When a co-crystallization approach was used, SBR was added to a $\approx 10\,\mathrm{mg\,ml^{-1}}$ hCAII solution in a 1:1 molar ratio and incubated on ice for a few minutes. For soaking, the ligand SBR was only added in a 1:1 molar ratio after the crystals were fully grown. The solution was subsequently prepared for crystallization in a sitting-drop format using the Dragonfly instrument for pipetting. The precipitant solution contained 50 mM deuterated TRIS buffer and 2.7 – 3.0 M ammonium sulfate at a pH of 7.8 – 8.0. The protein solution was mixed with the precipitant solution in a 1:1 ratio of volumes resulting in a final volume of 6 µl. In addition, 400 µl precipitant solution was kept below the crystallization drop well. The sitting drop plate was sealed and kept at 4 °C for one week. For solid-state NMR, deuterated DMSO and buffers were used.

X-ray Data Acquisition and Processing

The crystals were harvested with 0.2 mm loops, cryoprotected with precipitant solution containing 10 % glycerol, and then frozen with liquid nitrogen. Data sets were collected at the Swiss Light Source, Villigen, Switzerland. All data sets were processed and refined with several software packages like XDS including xscale[232] as well as CCP4 including pointless.[233] The structures were then refined and constructed using Phenix[234] and Coot.[235]
The accuracy of B-factors usually amounts to $\approx 15\,\%$.[236] Looking at the hCAII – SBR complex, lower occupancy of the ligand likely also causes the B-factors of ligand atoms to rise compared to protein atoms. During X-ray structure refinements with Coot,[235] the relative occupancy of Zn^{2+} and SBR was estimated based on the provided electron densities. For the dataset for the soaked crystals, the occupancies of SBR and Zn^{2+} were manually set to 68 % and 40 %, respectively, and refined with the software Phenix, estimating an overall occupancy for all atoms of the molecule. The B-factors of the dataset of the co-crystallized sample were estimated for each SBR ligand atom by Phenix during the refinement.

Table 3.1: Experimental details on acquisition and processing parameters of X-ray diffraction data sets for the soaked and co-crystallized hCAII – SBR complexes. Both datasets were collected at an excitation voltage of 12.4 keV, with waves of 1 Å and 100 K with an EIGER2 16M detector at a PXII X10SA at the Swiss Light Source. For each dataset, one crystal was used, which had a space group of P 1 2_1 1 in each case.

Data collection parameters	**Co-crystallized hCAII – SBR complex**	**hCAII crystals soaked with SBR**
Unit cell dimensions	42.352 41.595 72.243 90 104.343 90	42.339 41.615 72.286 90 104.364 90
Resolution	1.459 Å – 32.098 Å	1.481 Å – 41.015 Å
Number of reflections	42546	39468
Completeness	99.77 %	96.77 %
R_{work}, R_{free}	0.1680, 0.1917	0.1779, 0.1996
RMS (bonds, angles)	0.009, 1.00	0.0046, 0.75
Ramachandran (favoured, outliers)	95.70 %, 0.39 %	95.69 %, 0.00 %
Clashscore	1.67	4.32

Table 3.2: (Averaged) B-factors and occupancies derived from the software Phenix. For histidine residues H94, H96 and H119, which complex the Zn^{2+} ion, the B-factors and occupancies of both sidechain nitrogen atoms were averaged. The averaged B-factor and occupancy of the SBR warhead originates from the two oxygen and the nitrogen atoms of the warhead, which are in direct contact with the Zn^{2+} ion. For the SBR benzene and indole groups, the B-factor and occupancies of all atoms were averaged per group. Top number: B-factors, lower numbers: occupancies.

Sample	**Property**	**Zn^{2+} ion**	**H94, H96 and H119**	**SBR warhead**	**SBR benzene ring**	**SBR indole group**
Co-crystallised	B-factor	10.29	14.53	14.02	17.01	28.00
	Occupancy	0.36	1.0	0.75	0.76	0.68
Soaked with SBR	B-factor	19.88	20.43	23.80	29.63	51.00
	Occupancy	0.44	1.0	0.77	0.77	0.77

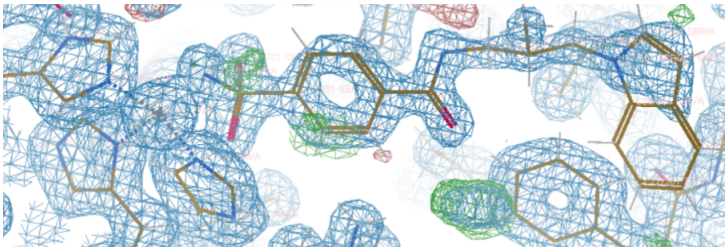

Figure 3.4: Ligand electron density in the X-ray structure of co-crystallized hCAII – SBR.

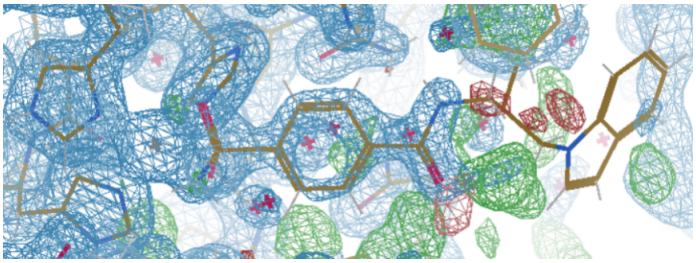

Figure 3.5: Ligand electron density in the X-ray structure of hCAII crystals soaked with SBR.

Solution NMR protein experiments

All solution NMR experiments were performed using an ASCEND 800 NMR spectrometer with an AVANCE NEO console and a cryo-probe if not indicated differently. For protein-observed experiments, uniformly (u-^{13}C/^{15}N) labelled hCAII was prepared in a mixed solvent of 90 % H_2O and 10 % ^2H_2O (in 80 mm phosphate buffer and 20 mm NaCl at pH 7.4). All protein-containing experiments were carried out at 310 K with protein concentrations of around 0.5 mM.

SBR was titrated in up to a 1:1 ratio (see Figs. 3.6 and 3.7) using a 10 mm DMSO stock solution, resulting in ≈ 13 % final DMSO content. The DMSO was shown not to cause any significant chemical-shift perturbations using a DMSO-only titration (data not shown), such that CSPs observed in the ligand titration do derive from the bound ligand as expected.

Bloch-McConnell relaxation dispersion experiments (see Fig. 3.8) of the ligand-bound protein were performed using the CPMG sequence "hsqcrexetf3gpsi3d" from the Bruker library, with a sample containing ^{13}C,^{15}N labelled hCAII in 80 mm phosphate buffer and 50 mm NaCl at pH 7.4 with 1:1 molar ratio of the ligand SBR at 310 K. The data were fitted using NESSY software.[237] As expected from previous studies,[223] the protein in the complex does not show any significant relaxation dispersion profiles for the pocket.

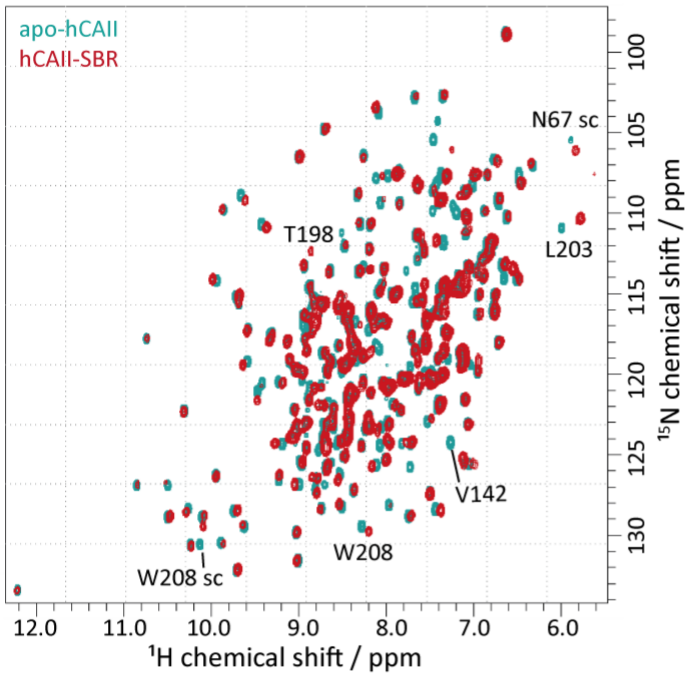

Figure 3.6: Solution HSQC protein spectra of ^{13}C,^{15}N-labelled apo-hCAII (cyan) and the hCAII-SBR complex at 1:1 molar ratio (red) in 80 mM phosphate buffer and 20 mM NaCl at pH 7.4 and 310 K.

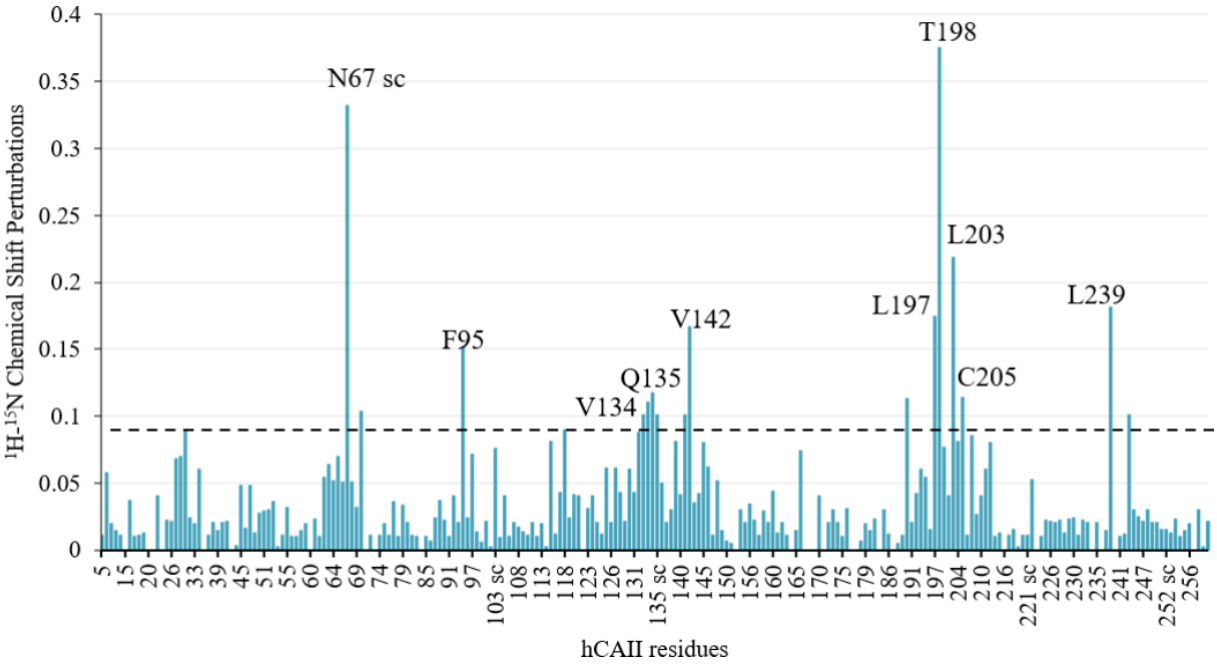

Figure 3.7: Chemical-shift perturbations of hCAII amide peaks upon binding of the ligand SBR (observed via solution NMR) as a function of sequence, highlighting the residues with the largest perturbations.

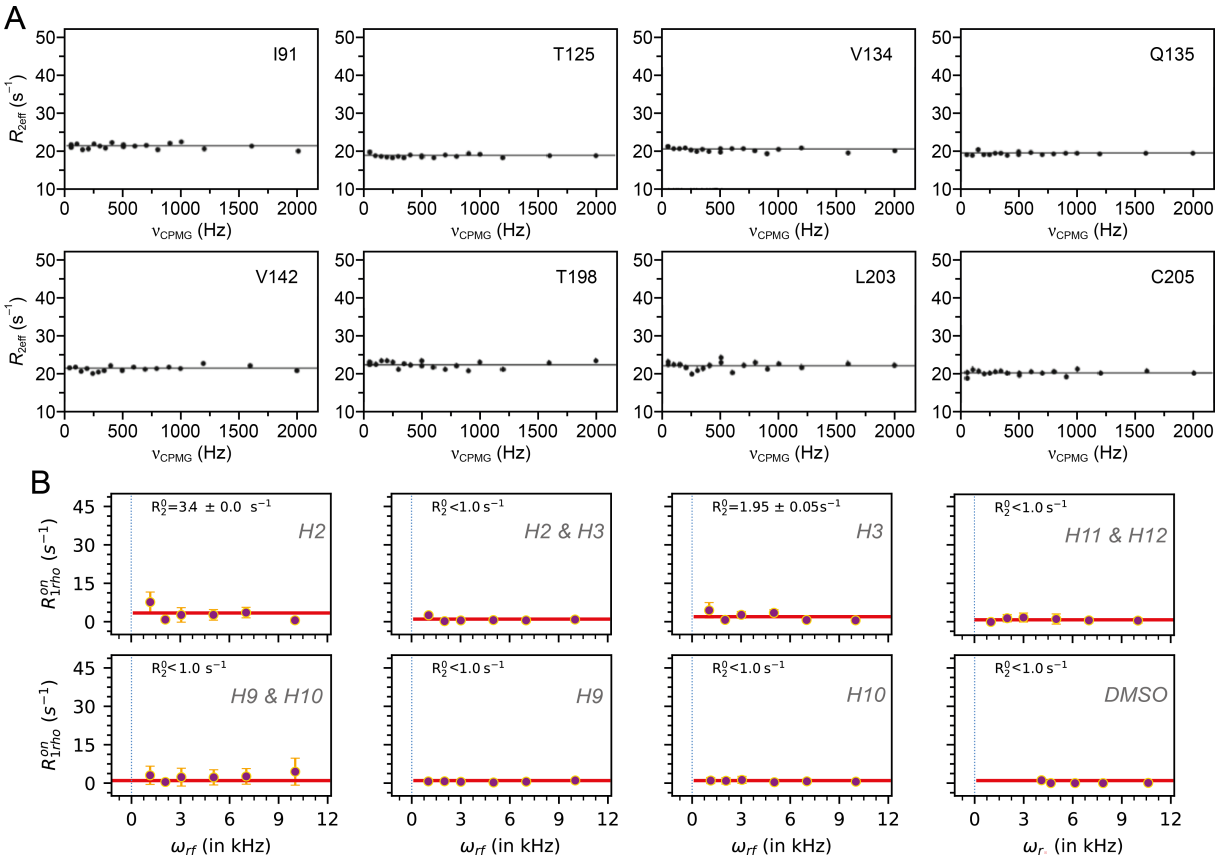

Figure 3.8: Selection of CPMG solution NMR plots of apo protein and apo ligand. (A) hCAII backbone dynamics while bound to SBR. Each plot represents the ^{15}N-backbone dynamics of one amino acid as specified. A sample of ^{13}C,^{15}N-labelled hCAII at 1:1 molar ratio of the ligand SBR was used. (B) Bloch McConnell relaxation dispersion profiles of the apo ligand, recorded using a ^{1}H-^{1}H 2D correlation with a refocussed INEPT in D_2O. Cross peaks are denoted as "x & y". The relaxation behaviour that is encoded in the peak heights refers to the first number mentioned (direct dimension).

Solution NMR ligand experiments

Proton resonances of the apo ligand were assigned using standard experiments HSQC, HMBC, and TOCSY spectra (see Fig. 3.10), recorded in either d_6-DMSO or D_2O on a 700 MHz or 800 MHz spectrometer, respectively.

Ligand spectra of bound ligand (Fig. 3.11) were obtained using heteronuclear double-filtered 2D experiments. For ligand assignments and relaxation dispersion measurements, a 480 µM u-^{2}H/^{13}C/^{15}N-labeled hCAII sample with ligand in a 1:1 molar ratio in 80 mM phosphate and 20 mM NaCl at pH 8 in D_2O were used. The $R_{1\rho}$ relaxation dispersion experiments (see the pulse sequence in Fig. 3.9) consisted of a refocused INEPT[223] element to observe cross peaks between the ligand protons via scalar couplings as well as a constant-time, on-resonance spin relaxation element, two double-^{13}C,^{15}N-filters, water suppression with WATERGATE, and heat compensation. It was derived from the 2D ^{13}C/^{15}N-doubly filtered TOCSY experiments in the Bruker pulse sequence library.[225,238–240]

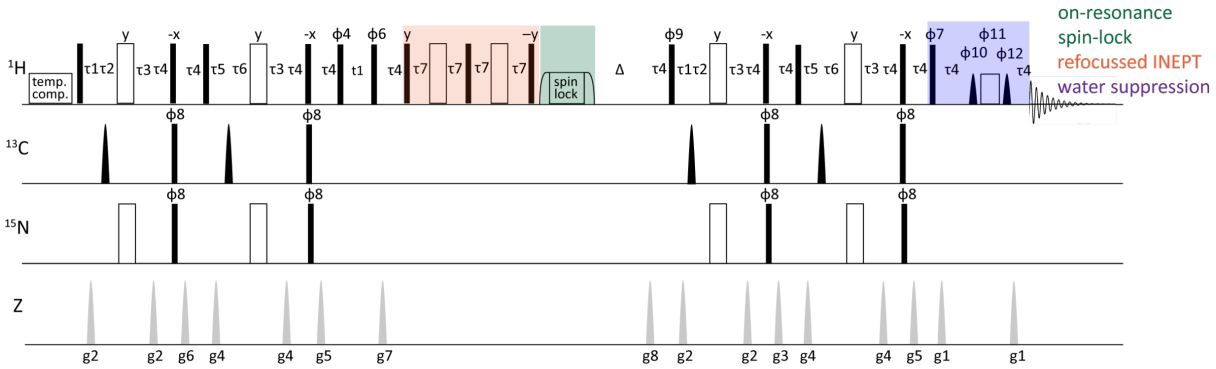

Figure 3.9: $R_{1\rho}$ relaxation dispersion solution pulse sequence, adapted from "refineptgpphwgxf" and "dipsi2gpphwgxf" from the Bruker library,[238] with the evolution of the indirect dimension during t_1. The Z channel displays all gradient pulses implemented. Apart from ^{13}C,^{15}N-double filters, the ^{1}H-BMRD experiment comprises temperature compensation, refocused INEPT (apricot), on-resonance spin lock (green) with a constant time delay, Δ, and a water suppression module (purple). The delays were defined as $\tau_1 = 476\,\mu s$, $\tau_2 = 204\,\mu s$, $\tau_3 = 1.977\,ms$, $\tau_4 = 200\,\mu s$, $\tau_5 = 375\,\mu s$, $\tau_6 = 305\,\mu s$, $\tau_7 = 24.995\,ms$. Δ is adjusted to 100 ms minus the spin lock length applied. Spin lock fields were applied at 1.0, 1.5, 2.0, 2.5, 3.0, 5.0, 7.0 and 10.0 kHz. Spin lock times were chosen as 10, 25, 50, 75 and 100 ms. Phase cycling was performed for ϕ_4 = 0 0 0 0 2 2 2 2, ϕ_6 = 0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2, ϕ_7 = 0 2, ϕ_8 = 0 1 2 3, ϕ_9 = 0 0 2 2, ϕ_10 = 2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1, ϕ_11 = 0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3, and ϕ_12 = 2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1; ϕ_{rec} = 0 2 2 0 0 2 2 0 2 0 0 2 2 0 0 2.

The effective relaxation rate at each spin lock field strength was calculated according to

$$R_{1\rho} = -\frac{1}{T} \cdot \ln\left(\frac{I}{I_0}\right) \tag{3.1}$$

where I is the peak intensity, I_0 is the corresponding intensity in a reference spectrum recorded without the spin lock, and T is the spin lock length. Data were analysed in Python according to the below models of no-exchange and fast-exchange behaviour. The statistically justified model for each residue was determined based on the reduced χ^2 using a p-value of 5 %.

$$\text{Model 1, no exchange:} \quad R_{1\rho} = R_{1\rho}^0 \tag{3.2}$$

$$\text{Model 2, two state fast exchange:} \quad R_{1\rho} = R_{1\rho}^0 + \frac{\omega_{ex}\tau_c}{\omega_{rf}^2\tau_c^2 + 1} \text{ with } \omega_{ex} = p_1 p_2 \Delta\omega^2 \tag{3.3}$$

τ_c, ω_{rf}, p_i, $R_{1\rho}^0$, and $\Delta\omega$ being the correlation time of the exchange, the spin lock field strength, the associated populations, the expected relaxation rate assuming no µs timescale dynamics, and the chemical-shift difference between the states.

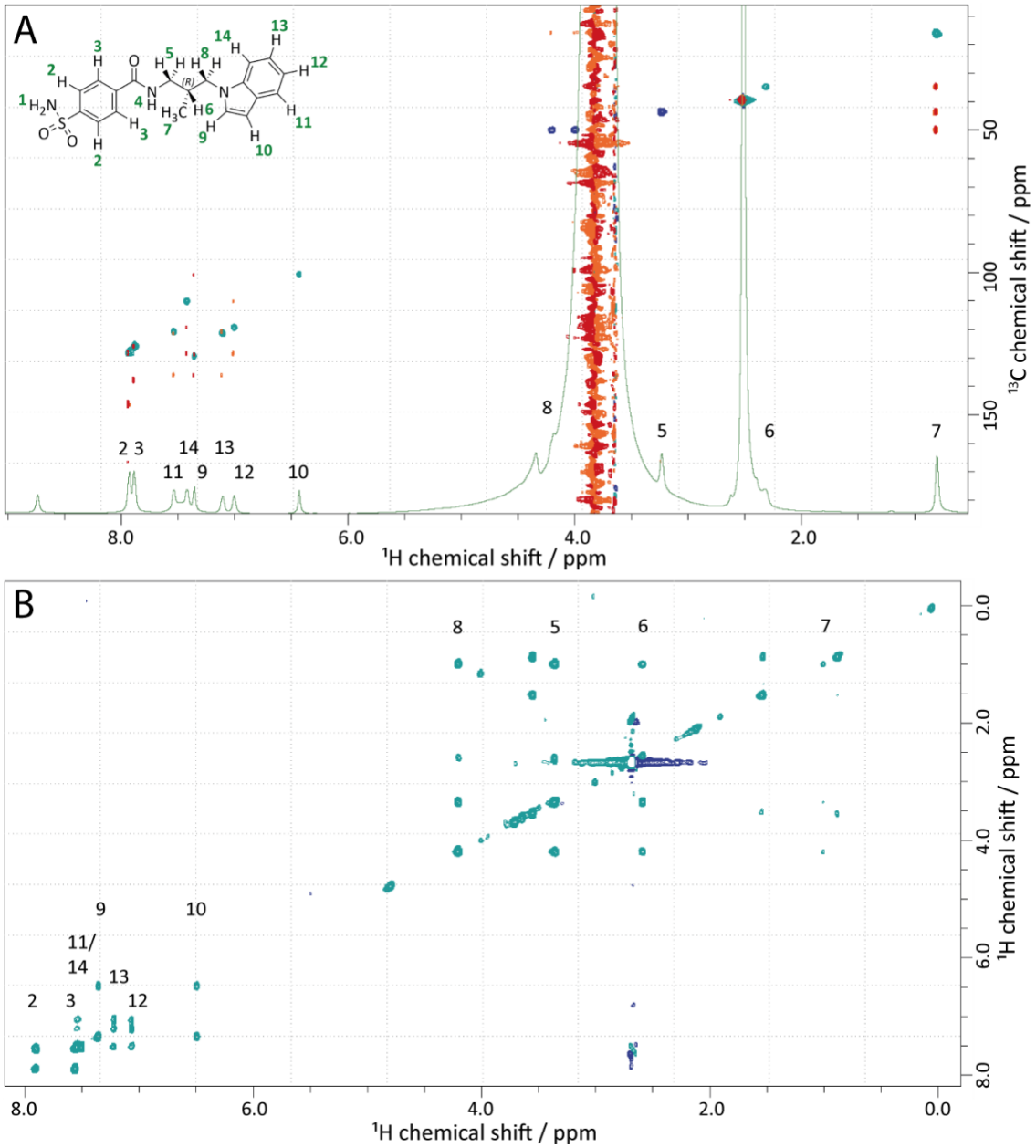

Figure 3.10: SBR ^{1}H chemical-shift assignments. (A) 2D ^{1}H–^{13}C HMBC spectrum (red for positive, orange for negative peaks), 1D ^{1}H spectrum (light green) and ^{1}H–^{13}C HSQC spectrum (cyan for positive and dark blue for negative peaks, corresponding to methylene groups) of ligand 9.7 mM SBR in d_6-DMSO, recorded on a 700 MHz spectrometer. The peak at ≈ 2.5 ppm corresponds to a signal from non-deuterated DMSO. (B) ^{1}H–^{1}H TOCSY spectrum (cyan for positive and dark blue for negative peaks, pulse sequence "dipsi2gpph19" from the Bruker library, in contrast to the ^{1}H–^{1}H relay correlation in Fig. 3.11) of a 200 µM SBR in D_2O, recorded at 298 K on an 800 MHz spectrometer. In addition to the expected peaks, assigned via standard 1D and 2D small-molecule experiments,[219] additional peaks were obtained in the aliphatic region that likely belong to other solvents and impurities from the SBR synthesis. SBR is only soluble up to ≈ 200 µM concentration in aqueous solution.

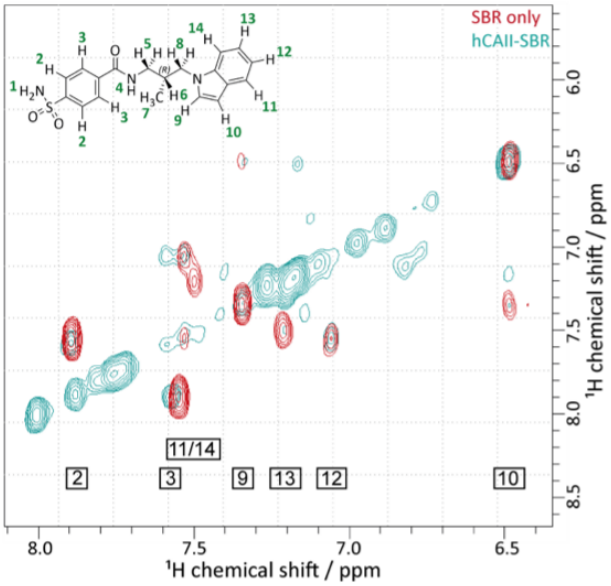

Figure 3.11: ^{13}C- and ^{15}N-filtered ^{1}H–^{13}C relay correlation of SBR bound to hCAII (cyan), overlaid with apo ligand (red) at 298 K. A 200 µM ligand sample in D_2O and a 480 µM hCAII sample with ligand in a 1:1 molar ratio in 80 mM phosphate and 20 mM NaCl at pH 8 in D_2O were used, respectively.

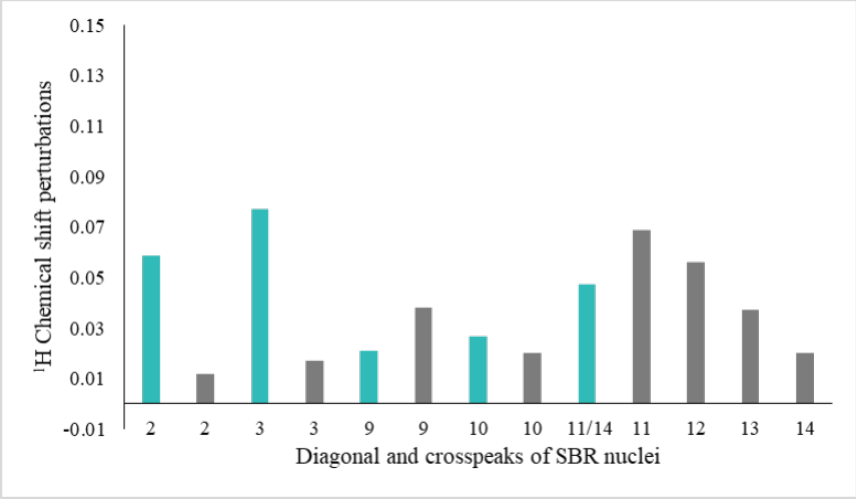

Figure 3.12: Chemical-shift perturbations of aromatic ligand protons upon binding. Given the flexibility of the aromatic parts of the ligand, CSPs are small. Differences between diagonal and cross peaks are due to the limited accuracy of the measurement of peak maxima in the bound state. These peaks are broadened due to their larger natural linewidth and apodization needed for better signal to noise ratios.

Table 3.3: Proton linewidths of well-separated ligand signals. Whereas the general increase in linewidth upon binding reflects the expected change in τ_c (< 1 kDa to 29 kDa MW), the maintained large linewidths of the indole protons suggest that their internal dynamics is still significantly slower than the dynamics of the free ligand (τ_c of ≈ 1 ns).

Sample	**Peak Attribute**	**Linewidth for H2**	**Linewidth for H10**
SBR in D_2O (298 K)	Diagonal peak	-	2.8 Hz
	Cross peak with H3/H9	10.1 Hz	2.5 Hz
hCAII – SBR in D_2O (298 K)	Diagonal peak	19.8 Hz	19.1 Hz
	Cross peak with H3/H9	16.7 Hz	20.4 Hz
hCAII – SBR in D_2O (283 K)	Diagonal peak	29.4 Hz	13.2 Hz
	Cross peak with H3/H9	29.7 Hz	16.0 Hz

Solid-State NMR Ligand Experiments

The triple-labelled crystals were harvested from the sitting drop plates using a 1 – 20 µl pipette, carefully detaching them from the well surface. For facilitated 1.3 mm rotor filling, large crystals were crushed into smaller parts with a pipette tip.

All solid-state NMR experiments were performed on an Ascend 700 NMR spectrometer with an AVANCE NEO console and a 1.3 mm triple-resonance HNC probe at 55.555 kHz MAS and roughly 298 K. The spectra shown in Fig. 3.14 were recorded at 100 kHz MAS in a 0.7 mm HNC probe at roughly the same temperature.

The assignments of SBR were transferred from solution to solid-state NMR spectra, referenced by the water and DMSO peaks. Direct detection of SBR dynamics was pursued using a ^{13}C- and ^{15}N-filtered sequence to remove all contributions from residual protein protons, which are bound to ^{13}C or ^{15}N atoms. The pulse sequence was adapted from the sequence designed for solution NMR measurements (see below).

Solid-State $R_{1\rho}$ Relaxation Dispersion

$R_{1\rho}$ relaxation dispersion under MAS or "NEar Rotary Resonance relaxation Dispersion" for residues moving on the timescale of the rotor frequency is characterized by increased effective relaxation rates not only at but also around the various recoupling conditions.[187,228,241] The effect is observed whenever an interference occurs between internal motion and MAS (in the case of similar timescales of internal motion and the rotor frequency). Autorelaxation in ^{15}N NERRD[242,243] (partial recoupling between the amide proton and nitrogen) is a locally confined event. By contrast, in ^{1}H NERRD, relaxation rates around the HORROR (half-rotary resonance) condition are considered, which denote a partial homonuclear recoupling. This recoupling always concerns multiple protons that are part of the regional proton spin bath in the wider surrounding of the proton in focus. At the same time, not only auto- but also the associated cross relaxation is partly induced. The effects lead to strongly enhanced signal decay in ^{1}H $R_{1\rho}$ experiments

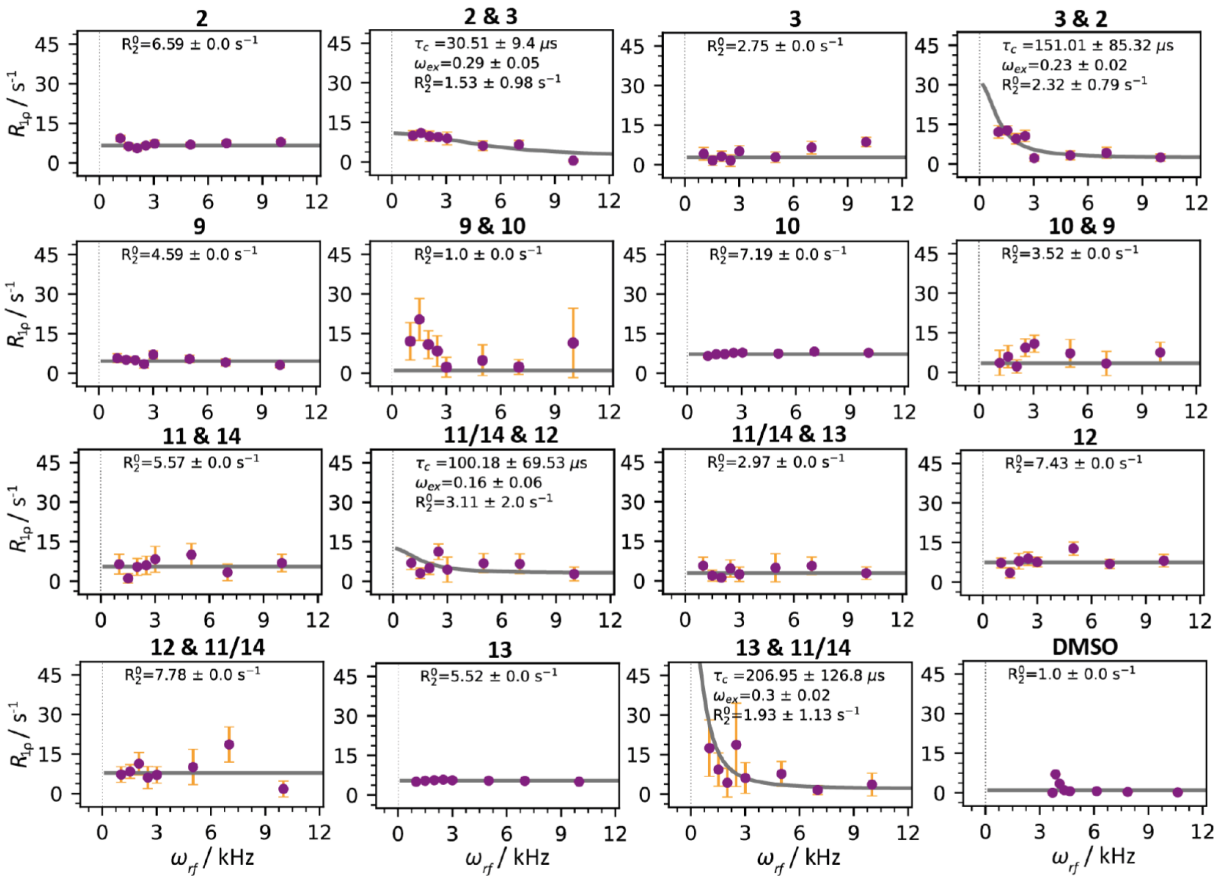

Figure 3.13: ^{1}H BMRD relaxation dispersion curves for the ligand in complex with hCAII, based on a ^{1}H–^{1}H 2D correlation using a refocussed INEPT from a 200 µM ligand sample in D_2O and a 480 µM hCAII sample with ligand in a 1:1 molar ratio in 80 mM phosphate and 20 mM NaCl at pH 8 in D_2O. The relaxation that is encoded in the peak heights of cross peaks refers to the first number mentioned (direct dimension). Cross peaks can be considered to be the more faithful source of information, as with this pulse scheme they bear higher signal to noise than the diagonal peaks and are not overlapped. Profiles for nuclei within the same ring (in particular cross peak between 3 and 2 and cross peak between 2 and 3) would naively be expected to yield identical rates. This is mostly not the case due to the error involved but can also derive from differential contributions to the effective correlation times of the individual chemical shifts from structural elements that are further away from one type of proton than to another type of proton in the same ring. Within error, protons of the indole ring generally tend to have flat profiles at higher "baseline" rates. Errors with the value 0.0 s^{-1} denote numbers smaller than 0.05 s^{-1}.

with spin lock field strengths significantly outside the HORROR conditions (compare Fig. 3.18B/C).[187,228] A quantitative description of the NERRD effect for proton $R_{1\rho}$, covering homonuclear and heteronuclear interactions, has been derived elsewhere.[241] Here, on-resonance proton NERRD experiments ($\beta_e = \pi/2$) were performed using the pulse scheme in Fig. 3.15. Accordingly, the expected rates are:

$$R_{1\rho}^{AB} = \frac{d_{II}^2}{4}\Big(\frac{1}{6}J(\omega_r) + \frac{1}{12}J(2\omega_r) + \frac{9}{2}J(\omega_I) + 3J(2\omega_I) + \frac{3}{8}J(2\omega_e - 2\omega_r) + \frac{3}{4}J(2\omega_e - \omega_r) + \frac{3}{4}J(2\omega_e + \omega_r) + \frac{3}{8}J(2\omega_e - 2\omega_r)\Big) \tag{3.4}$$

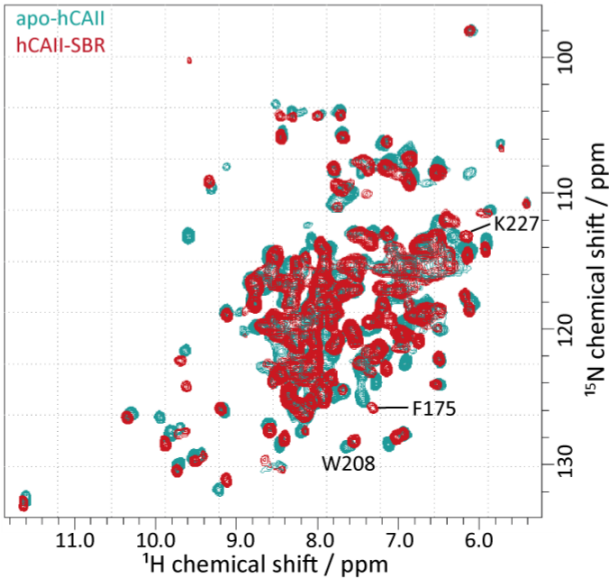

Figure 3.14: hNH correlations of crystalline hCAII. Overlay of ^{2}H, ^{13}C,^{15}N-labelled apo-hCAII (cyan) and ^{2}H,^{13}C,^{15}N-labelled hCAII crystals soaked with SBR in 1:2 molar ratio with ≈ 0.5 mg crystallized ^{2}H,^{13}C,^{15}N-labelled soaked with SBR with < 10 % deuterated DMSO (red). Both crystalline samples were kept in 2.9 M ammonium sulphate, 50 mM deuterated TRIS buffer at pH 8.0. The apo-hCAII sample was filled in a 1.3 mm rotor and spun at 55 kHz at 296 K while the hCAII–SBR sample was kept in a 0.7 mm rotor, spun at 100 kHz MAS and 298 K.

with the homonuclear dipolar coupling constant $d_{\mathrm{II}} = (\mu_0\gamma_I^2\hbar)/(4\pi r_{\mathrm{eff}}^3)$, magic-angle spinning frequency ω_r, and the effective field strength ω_e. r_{eff} is the effective distance characterizing the overall proton-proton dipolar interactions of the regional dipolar network. ^{1}H NERRD profiles in main Text Fig. 3.2B were simulated in Python using the above analytical derivation for on-resonance homonuclear proton recoupling as a function of the spin lock field strength.[187,241]

A consequence for the quantitative interpretation of proton NERRD data is the unknown interaction strength d_{II} of the regional proton-proton dipolar-coupling network. By much contrast, d_{II} for ^{15}N NERRD of amide groups or in ^{13}C NERRD applications for amino acid sidechains is accurately defined from the one-bond ^{1}H-^{15}N or the ^{1}H-^{13}C distance, respectively, both being sufficiently consistent throughout the entire protein. Whereas for heteronuclear NERRD only populations and amplitude are entangled, in proton NERRD even the combination of the two is uncertain, as the effective proton-proton distance remains an approximation. Homonuclear, dipolar couplings which are partly restored when the effect of MAS is counteracted by (almost) rotor-synchronised spin inversion (i.e., partial recoupling) both, at the rotor-resonance condition ($\omega_e = \omega_r$) and at the HORROR condition ($\omega_e = \frac{1}{2}\omega_r$). The latter condition, the *first* NERRD bump, which is also in focus here (around the half-rotor-resonance/HORROR condition) is associated with only homonuclear terms being recoupled. For this condition, higher relaxation rates are obtained experimentally than predicted theoretically from the effect of a homonuclear dipolar cou-

pling within an IS spin pair according to eq. 3.4.[187] In addition to the homonuclear cross relaxation contribution, this deviation between predicted rates and experimental rates has been interpreted to reflect the extension of the individual proton-proton dipolar-coupling interactions into a multi-site dipolar network, which amplifies the local effect. Whereas this makes ^{1}H $R_{1\rho}$ a very sensitive reporter on µs timescale motion, the accuracy of amplitude information has remained a conceptual challenge.

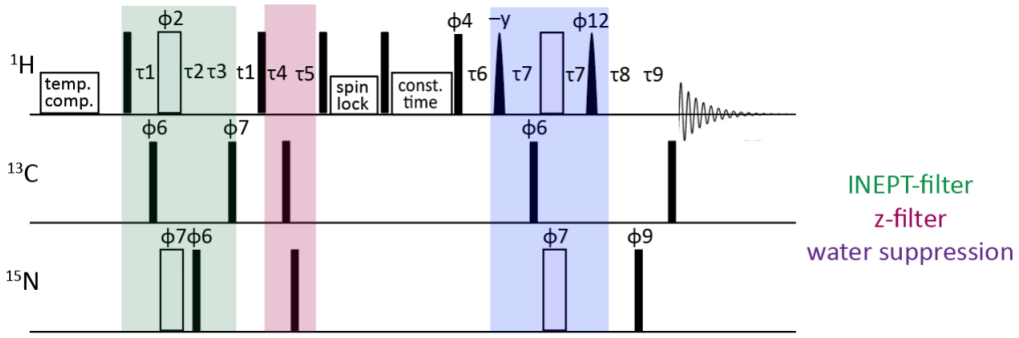

Figure 3.15: Ligand $R_{1\rho}$ relaxation dispersion (^{1}H NERRD) solid-state NMR pulse sequence, with the indirect chemical-shift evolution period t_1. The pulse sequence was based on a NOESY sequence and included temperature compensation, a ^{13}C/^{15}N-INEPT-filter (green), a z-filter or NOESY mixing time (magenta), an on-resonance spin lock, and water suppression modules (purple). The spin lock times amounted each to 2, 5, 10, 20 and 40 ms. Constant-time compensation was implemented as well. The delays τ were defined as $\tau_1 = 3.57$ ms, $\tau_2 = 1.97$ ms, $\tau_3 = 1.59$ ms, $\tau_4 = 2.5$ ms, $\tau_5 = 2.49$ ms, $\tau_6 = 2.54$ ms, $\tau_7 = 20$ ms, $\tau_8 = 942$ ms, and $\tau_8 = 1.59$ ms. The phase cycling was performed with $\phi_2 = 0\ 1\ 0\ 1\ 2\ 3\ 2\ 3$, $\phi_4 = 0\ 0\ 1\ 1\ 2\ 2\ 3\ 3$, $\phi_6 = 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$, $\phi_7 = 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2$, $\phi_9 = 0\ 0\ 0\ 0\ 2\ 2\ 2\ 2$, and $\phi_{12} = 0\ 2\ 3\ 3\ 0\ 0\ 1\ 1$; $\phi_{\mathrm{rec}} = 0\ 2\ 3\ 1\ 2\ 0\ 1\ 3$.

Fig. 3.15 shows the heteronuclear-filtered, ^{1}H $R_{1\rho}$-edited proton-proton correlation experiment used to probe ligand proton NERRD behaviour. Fig. 3.16 shows a representative 2D (overlaid to best-match H2), in which those (partly overlapped) diagonal peaks are annotated with assignments that, based on solution NMR, are certain. Fig. 3.18 shows all NERRD profiles for SBR in comparison to the profile for DMSO as a reference case as well as replotted profiles (a rather rigid residue and one with pronounced µs timescale motion) from Rovó et al.[187] for comparison.

Fig. 3.17 demonstrates the X-ray-based determination of the effective proton-proton distances in the ligand, taking into account all dipolar interactions with neighbouring protons, each weighted by r^{-6}, in a cumulative way. By contrast, fitting of each experimental ligand ^{1}H NERRD profile was done for a wide distribution of effective distances r_{eff}, and each time, the goodness of fit was determined (Fig. 3.19). A reasonable goodness of fit plateaus at ≈ 2 Å, reflecting the entanglement of the motional amplitude (S^2) with the underlying interaction strength.

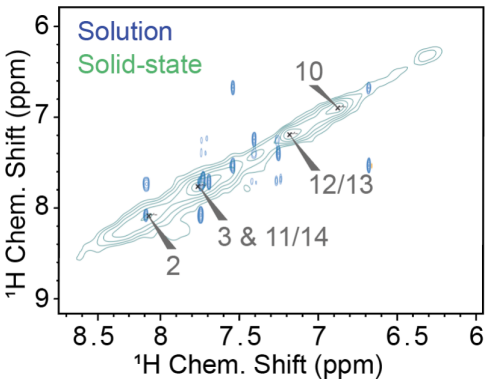

Figure 3.16: Filtered ligand spectrum recorded at 55.555 kHz MAS and 298 K on a co-crystallized hCAII – SBR sample containing ≈ 1 mg ^{2}H,^{13}C,^{15}N-labelled protein using the sequence shown in Fig. 3.15 (green). In contrast to the filtered correlations in solution (overlaid in blue), ligand properties were assessed here using the diagonal peaks (using 2D correlations with insignificant mixing time). In principle, the sequence grants 2D resolution, which would ameliorate peak overlap, but cross peaks with high signal to noise were neither obtained by *J*-based, NOE, nor RFDR mixing. Similarly, cross peaks between hCAII and SBR were not detected within the sensitivity limits as, assumingly, ligand dynamics are too fast for RFDR transfer but insufficiently slow for NOE build-ups.

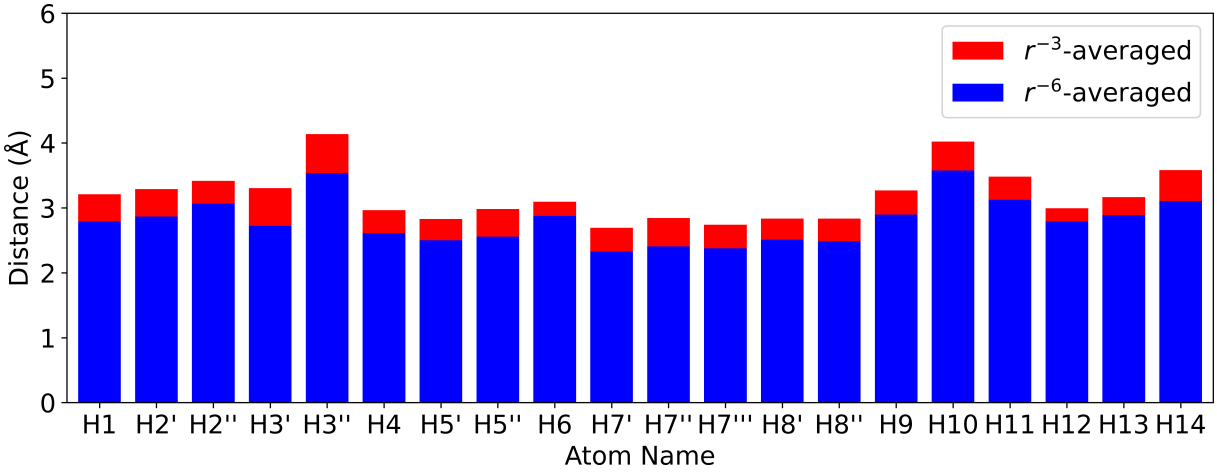

Figure 3.17: Approximation of the effective proton-proton distance, taking all ligand proton-proton distances up to 6 Å (assuming an elongated conformation) into account. In reality, the ligand is known to bend around dynamically, which renders the proton density site- and time dependent (compare Fig. 3.26). The effective distances are dominantly determined by the closest neighbour due to the r^{-6}-weighting. Nevertheless, due to the chains of strong couplings in the network, local changes are expected to modulate the entire network and thus have a long-range influence even for more distant protons. The effective distance commonly amounts to around 3 Å.

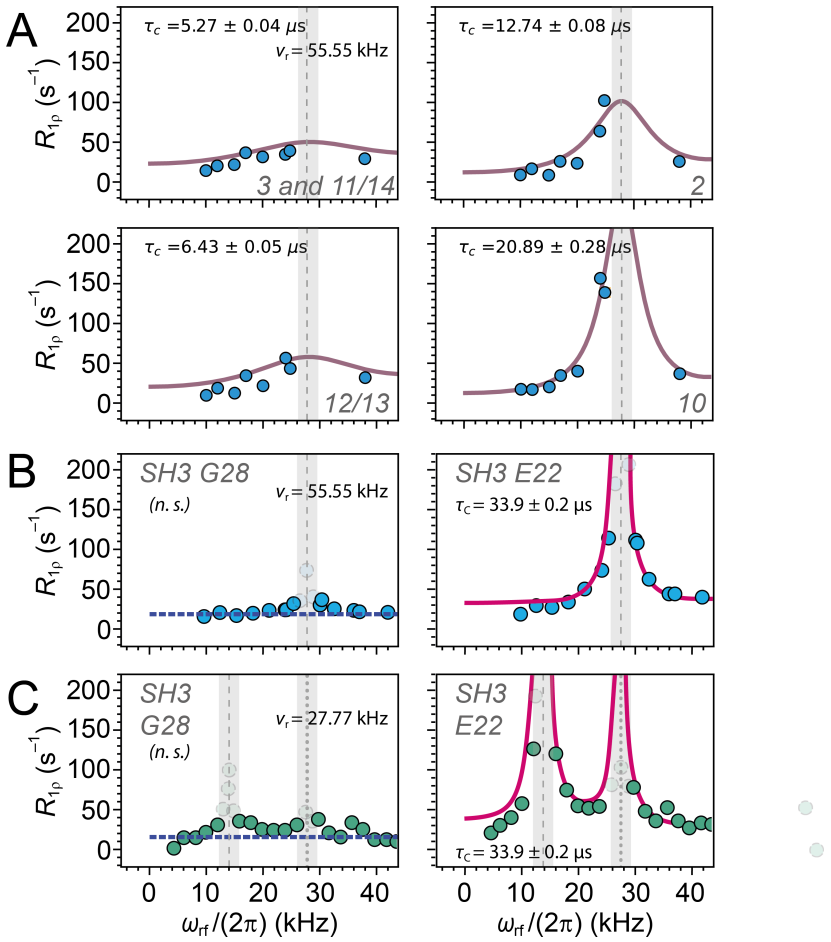

Figure 3.18: (A) NERRD behaviour in the co-crystallized hCAII – SBR sample shown for all discernable ligand proton signals for the co-crystallised hCAII/SBR sample at 55.555 kHz MAS. For part of the peaks, multiple ligand proton contributions overlap. Fits were obtained using 1.5 Å r_{eff}. The uncertainty associated with r_{eff}, does not allow for reliable S^2 determination. For comparison, a relatively rigid residue (G28) and one undergoing pronounced µs motion (E22) in the SH3 domain of chicken α-spectrin are shown for 55.555 kHz (B) and 27 kHz MAS (C). Data in the gray zone comprises coherent contributions, a sharp increase in apparent relaxation rates at the HORROR (dashed lines) and rotor resonance conditions (dotted lines), as seen nicely in the SH3 G28 data (semi-transparent dots in the gray areas) for reference. (Data in panels B and C replotted from Rovó et al.[187])

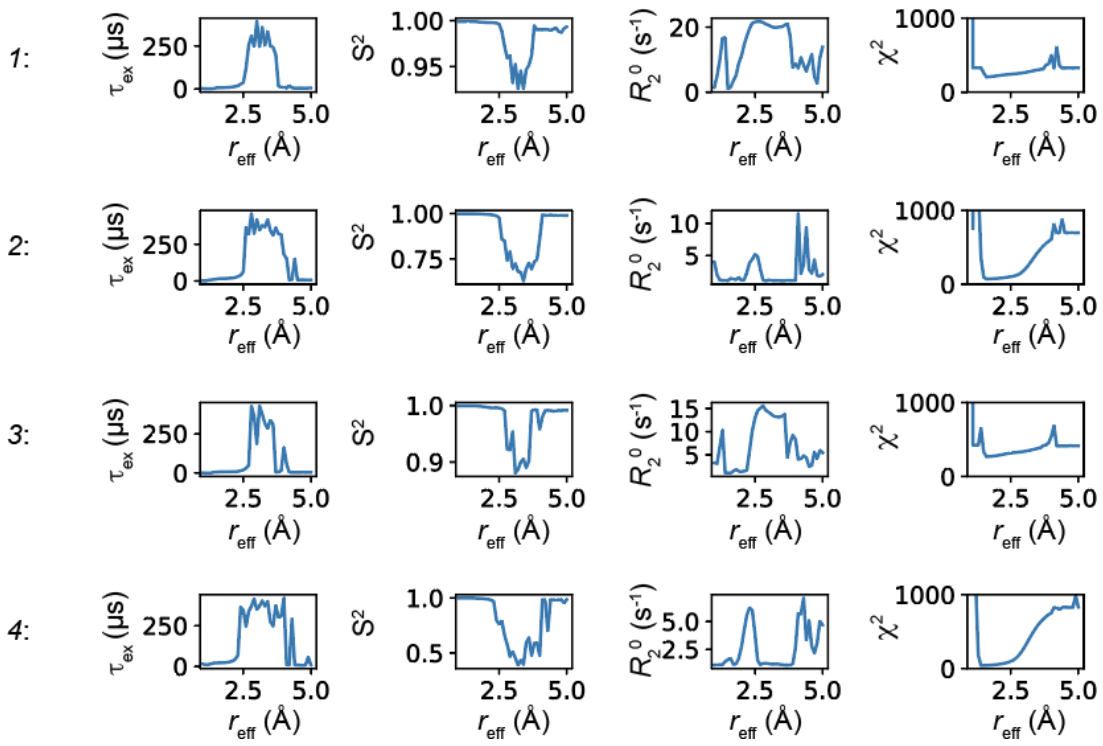

Figure 3.19: Dependence of ^{1}H-NERRD fitting parameters τ_{ex}, S^2, R_2^0, and χ^2 on the assumed r_{eff} (the distance reflecting the cumulative, r^{-6}-weighted proton-proton interactions), shown for the different ligand signals, with a flat plateau of minimal χ^2 values obtained for effective-distance values around 1.5 Å.

3.5.2 MD Simulations

System Setup

The MD simulations of hCAII in SBR-bound and apo form were carried out with GROMACS 2021.4[200] using the Amber ff19SB force field.[99] Initial atom coordinates were taken from the X-ray structure of the prepared co-crystallized hCAII sample. The two missing N-terminal residues 1Met and 2Ser were added with the software MODELLER.[244] For parametrising the coordination of the zinc ion, the program MCPB.py[229] (metal centre parameter builder) from the AmberTools22 package was used. After adding hydrogen atoms to the protein with protonation states according to pH 7, both systems were solvated with about 10,000 OPC water molecules in a periodic cubic box.

After energy minimization, both systems were equilibrated for 100 ps in the NVT ensemble and 500 ps in the NpT ensemble with harmonic position restraints on all heavy atoms of protein, ligand, and zinc (force constants of 1000 kJ mol^{-1} nm^{-2}). The temperature was kept constant at 300 K using the V-rescale thermostat developed by Bussi and coworkers.[111] Pressure was kept constant at 1 bar using the C-rescale barostat[113] and Parrinello-Rahman barostat during equilibration in the NpT ensemble and the production MD run, respectively. All bond lengths in protein and ligand were constrained with LINCS, and water molecules were kept rigid with SETTLE, allowing to integrate the equations of motion with 2 fs time steps. Lennard-Jones 12-6 interactions were smoothly shifted to zero at a 1.2 nm cut-off. Long-range electrostatic interactions were treated with the particle mesh Ewald algorithm[201] using a distance of 1.2 nm for switching between short- and long-range electrostatic interactions. The final production MD simulations were performed for 2 µs and (three times) 4 µs in case of apo and SBR-bound form of hCAII, respectively.

Backbone and Ligand Dynamics

RMSFs of hCAII in ligand-free and ligand-bound form show no substantial changes in protein dynamics apart from the loop around 234Gly and the helices around 21Pro and 133Ala (3.20). For analysing the dynamics of the ligand SBR with respect to the protein, RMSDs of different ligand atoms were determined from the trajectory aligned with respect to the protein backbone (N, CA, CB, C, O; main text 3.3). The RMSD of the nitrogen at the sulfonamide warhead stays constantly on a low level of around 0.5 Å due to the stable bond to the zinc ion. The RMSDs of the benzylic carbon atom depicted in red in main text Fig. 3.3B shows two distinct jumps at ca. 1.67 µs and 2.85 µs, during which the benzene ring flips about 180°. These rotameric jumps are on the µs timescale and hence in accordance with the timescale detectable with $R_{1\rho}$) relaxation dispersion experiments.

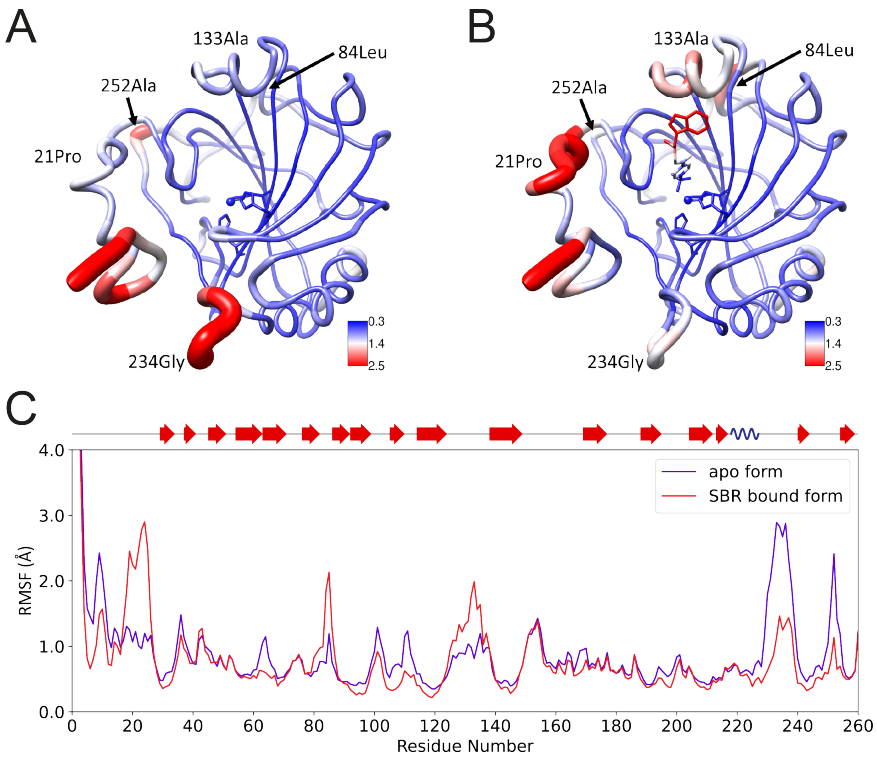

Figure 3.20: Root-mean-square fluctuations (RMSFs) obtained from MD simulations of hCAII in the apo and SBR-bound form for 2 and 4 µs, respectively. (A) and (B) RMSFs visualized by ribbon thickness and color in the X-ray structure of hCAII in apo and SBR-bound form, respectively. (C) Plot of C^{α} RMSFs vs. sequence number.

For analysing the rotation of the benzene ring quantitatively, a vector perpendicular to the rotation axis was constructed, which is in-plane with the average orientation of the benzene ring during the time interval from 3 µs to 4 µs in the first bound-form simulation (green vector in Fig. 3.21A). Practically, this time-averaged in-plane vector was determined by taking the mean of two time-averaged C-H bond vectors on one side of the benzyl ring (blue and yellow vectors in Fig. 3.21A). The time period from 3 µs to 4 µs was chosen because here the fluctuations of the RMSD were lowest. Measuring the angle between the in-plane vector at some time point and the time-averaged in-plane vector allows to determine the flipping angle (Fig. S18). Apart from flipping of the benzene ring about 180°, small dips during the time periods 35 ns – 65 ns and 2.57 µs – 2.75 µs are observable in the plot of the flipping angle as well as in the RMSD of the benzene carbon atom (Fig. 3.21C and red curve in main text Fig. 3.3B). During these time periods, the benzene ring is in a tilted conformation and the pathway of the successive alkyl chain is shifted (Fig. 3.22A). In the third simulation of the ligand-bound form, a 180° flip in the major conformation is observed at 3.63 µs, and two fast 180° flips are seen at ≈ 2.4 µs in the minor conformation of the benzylic ring (Fig. 3.21E).

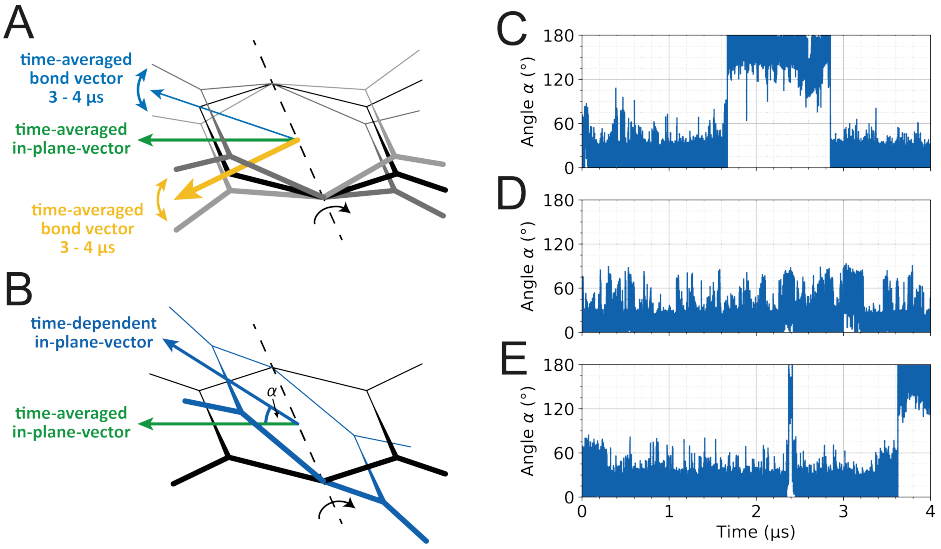

Figure 3.21: Aromatic ring flip of the SBR ligand. (A) Construction of the vector perpendicular to the rotation axis and in-plane with the average orientation of the benzene ring (green vector) by averaging the time-averaged C-H-bond vectors 2 (blue) and 3 (yellow, numeration according to main text Fig. 3.1). (B) Measurement of the flipping angle αbetween the in-plane vector at a given time point (blue) and the time-averaged in-plane vector (green). (C) Flipping angle αas function of time. D) and E) The flipping angle αin two additional trajectories of 4 µs each, produced under identical conditions as the trajectory analysed in panel C and in the main text.

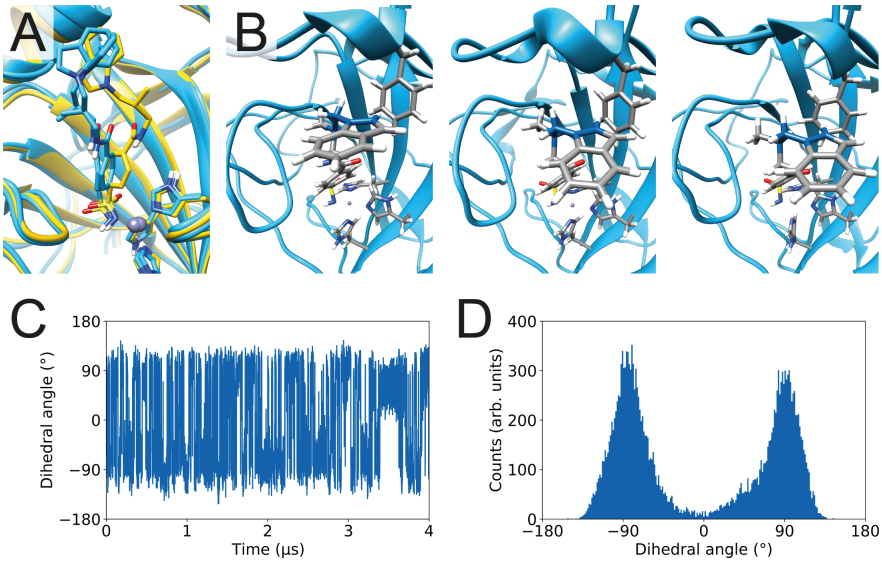

Figure 3.22: Conformations of minor importance. (A) Tilted conformation of the benzene ring with displaced alkyl chain, e.g., around 50 ns (compared to conformations at 0 ns and 100 ns) and 2.6 µs. (B) T-shaped π–πinteraction of the ligand SBR with 130Phe around 3.5 µs. (C) and (D) Dihedral angle between methylene group and indole ring (marked blue in B) as a function of time and as a histogram, respectively.

By contrast, the RMSDs and fluctuations of the RMSDs of indole ring atoms (N17 and C12, marked blue and yellow in main text Fig. 3.3) are much larger/faster. The reasons for the high flexibility of the indole ring are lower steric hindrance in the opening of the ligand binding pocket and the lack of strong and attractive interactions like π–π interactions or cation–π interactions to nearby aromatic or positively charged amino acids. The strongest interaction of the indole ring observable in the simulation are π–π interactions with 130Phe, especially T-shaped π–π interactions between 3.39 µs and 3.55 µs, indicated by strongly decreased torsional fluctuations of the bond towards the indole ring (Fig. 3.22B-D). With only 160 ns, such interactions are however expected to be too fast to be detectable with $R_{1\rho}$) experiments.

Time-Correlation Functions for Different H-H Vectors

In the following, TCFs were determined from the MD trajectory to estimate the timescale of dynamics for different parts of the ligand more quantitatively.

In NMR, relaxation resulting from orientational motions of 2$^{\text{nd}}$-rank tensors describing homo- or heteronuclear dipolar couplings (DD) or chemical-shift anisotropy (CSA) can be derived from TCFs. Considering the time-dependent reorientation of such a 2$^{\text{nd}}$-rank tensor aligned along some interaction vector $\vec{v}$ within a protein, the corresponding internal TCF $C_{\text{int}}(\tau)$ can be determined as a function of the lag time τ by averaging over all initial times t_0 the second Legendre polynomial $P_2 = 0.5 \cdot (3x^2 - 1)$ of the scalar product of a normalized interaction vector $\vec{v}$ at time t_0 and $t_0 + \tau$:

$$C_{\text{int}}(\tau) = \langle P_2 \left(\cos\theta(t_0, t_0 + \tau)\right)\rangle_{t_0} = \langle P_2 \left(\vec{v}(t_0) \circ \vec{v}(t_0 + \tau)\right)\rangle_{t_0} \tag{3.5}$$

The scalar product of two normalized interaction vectors is the cosine of the angle between both vectors. For ^{1}H–^{15}N R_1, $R_{1\rho}$, R_2, or hetNOE experiments, the 2$^{\text{nd}}$-rank tensor and hence also the interaction vector can be assumed to align with the ^{1}H–^{15}N inter-spin vector. In case of ^{1}H-NERRD, the interaction vector is determined by the homonuclear dipolar couplings to manifold other protons in the surrounding. Numerically, the TCF between a pair of two spins was calculated by

$$C_{\text{int}}(n) = \frac{1}{N-n} \cdot \sum_{i=1}^{N-n} \frac{1}{2} \cdot \left[3 \left(\vec{v}(t_i) \circ \vec{v}(t_{i+n})\right)^2 - 1\right] \tag{3.6}$$

N corresponds to the number of simulated time points, n to the time point of the correlation function and i to the initial time point. Please note that with increasing time difference τ, less contributions are averaged and hence, for long time differences τ, the determined correlation function becomes statistically noisy. In order to determine only the internal correlation time $C_{\text{int}}(\tau)$, the overall motion (tumbling) of the protein was removed for the whole trajectory by structural superposition of the protein backbone (N, CA, CB, CO, O) atoms with the initial structure of the simulation.

A time-dependent effective interaction strength per MD frame, determined by the homonuclear dipolar couplings to other protons in the surrounding and correctly reflecting the interactions determining the ^{1}H-NERRD experiments, is difficult to construct – among others – due to the dependence of the homonuclear dipolar coupling on the varying angle between H–H vectors and the magnetic field. Whereas in theory, a representative distribution of arbitrary z-axes of a crystal powder could be constructed and averaged by a Monte-Carlo approach, in practice this becomes a challenge given the large number of frames/time steps required for a sufficiently defined correlation function. As an alternative, TCFs of individual H–H vectors were determined from the MD simulation, which represent the time behaviour of bilateral proton-proton interactions. Whereas this does not allow a direct simulation of the NERRD data, it demonstrates that ring flipping of the benzyl ring of SBR corresponds to µs timescale dynamics detectable via ^{1}H $R_{1\rho}$) experiments for both rings (compare Fig. 3.23 – 3.25).

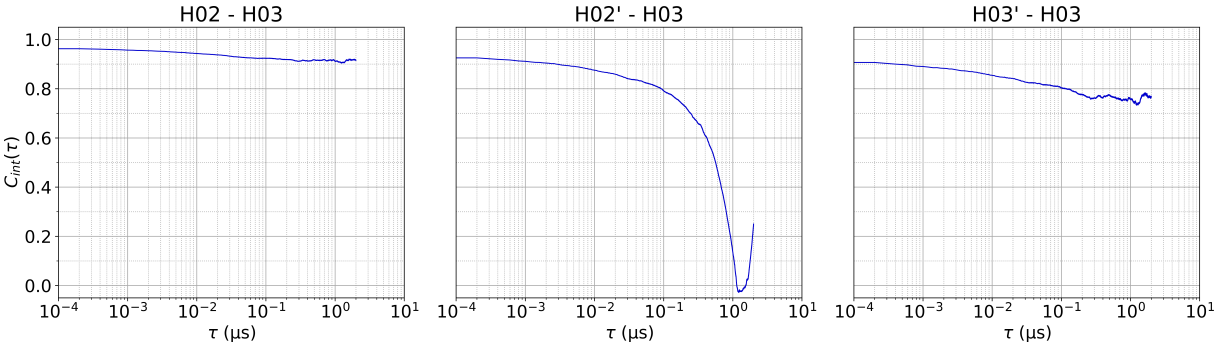

Figure 3.23: Internal time correlation functions of different intra-benzyl proton-proton vectors of the ligand SBR.

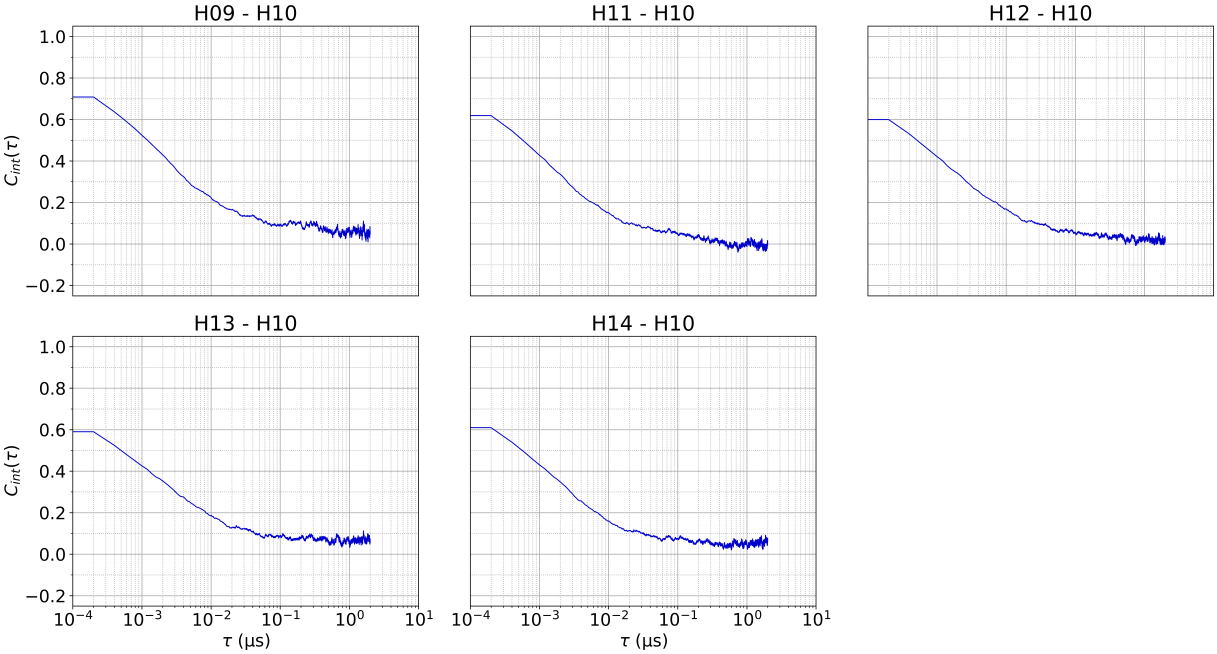

Figure 3.24: Internal time correlation functions of different intra-indole proton-proton vectors of the ligand SBR.

The distinct minima between 1 µs and 2 µs observed in Fig. 3.23 and 3.25 are in line with the experimental finding that the ring flipping events of the benzyl ring induce measurable $R_{1\rho}$ relaxation dispersion (^{1}H-NERRD) for both, indole and benzyl protons. By contrast, for intra-indole H–H vectors, no µs timescale contributions are observed, precluding measurable $R_{1\rho}$ relaxation dispersion (^{1}H-NERRD) in the absence of benzyl flipping. TCF in 3.23 – 3.25 and main Text Fig. 3.3 are clipped at half the simulation time (i.e., 2 µs).

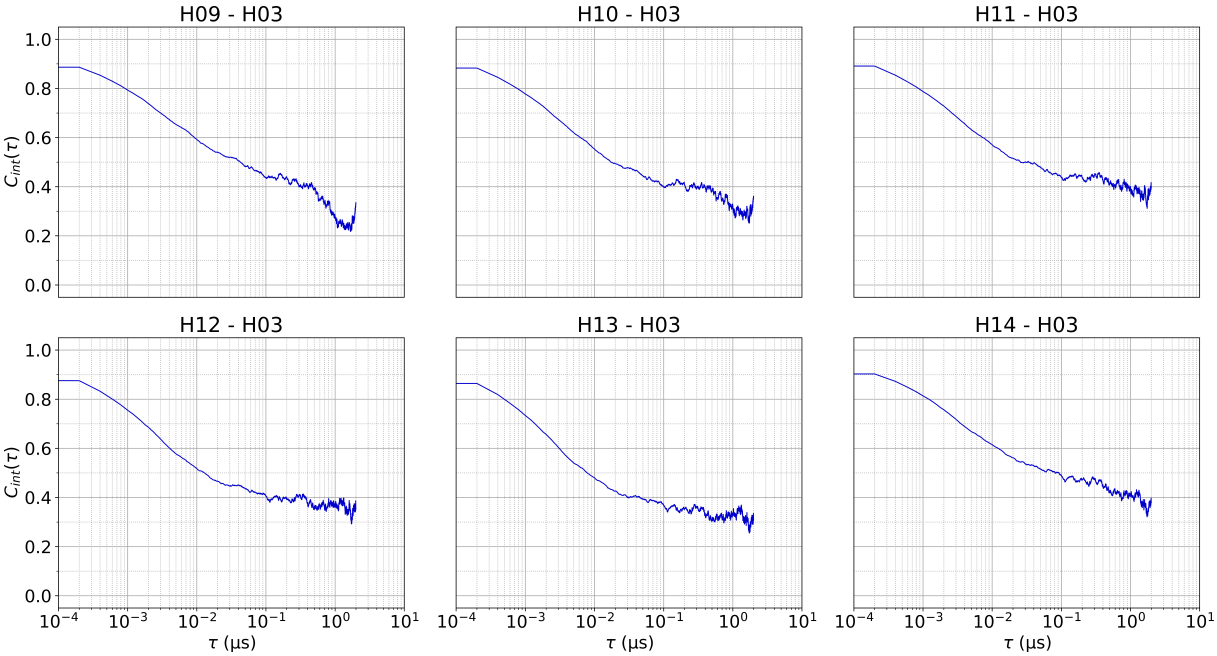

Figure 3.25: Internal time correlation functions of different inter-ring proton-proton vectors connecting the indole and benzyl moieties of the ligand SBR.

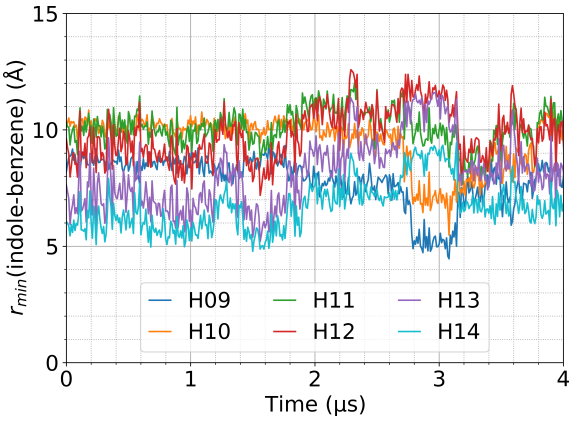

Figure 3.26: Distance between different indole protons (color code below) and benzene ring in the MD simulation as a function of simulation time. r^{-6}-averaged distances (representing the effective interaction strength) over the simulation accord to for 5.6 Å, 7.8 Å, 9.7 Å, 9.7 Å, 7.5 Å, and 5.8 Å for H09, H10, H11, H12, H13, and H14, respectively.

4

Nonlinear Impact of Electrolyte Solutions on Protein Dynamics

The work presented in this chapter was published as H. G. Daronkola[†], B. Söldner[†], H. Singh, R. Linser, A. V. Verde, *ChemBioChem* **2024**, *25*, e202400057.
[†]: These authors contributed equally to this work.

My contributions to this work:

- I prepared and measured all NMR samples with the help of Himanshu Singh.
- I assigned all spectra based on reference chemical shifts from the Biological Magnetic Resonance Bank (BMRB), transferred the assignments to spectra measured at high salt concentration and determined chemical-shift perturbations (CSPs).
- I determined and evaluated all ^{1}H-^{15}N hetNOE experiments and all constant-time ^{15}N-CPMG relaxation dispersion experiments.

4.1 Abstract

Halophilic organisms have adapted to multi-molar salt concentrations, their cytoplasmic proteins functioning despite stronger attraction between hydrophobic groups. These proteins, of interest in biotechnology because of decreasing fresh-water resources, have excess acidic amino acids. It has been suggested that conformational fluctuations – critical for protein function – decrease in the presence of a stronger hydrophobic effect, and that an acidic proteome would counteract this decrease. However, our understanding of the salt- and acidic amino acid dependency of enzymatic activity is limited. Here, using solution NMR relaxation and molecular dynamics simulations for in total 14 proteins, we show that salt concentration has a limited and moreover non-monotonic impact on protein dynamics. The results speak against the conformational-fluctuations model, instead indicating that maintaining protein dynamics to ensure protein function is not an evolutionary driving force behind the acidic proteome of halophilic proteins.

4.2 Introduction

Aqueous environments with NaCl concentrations up to multi-molar are home to halophilic ("salt-loving") organisms. To avoid membrane disruption by high osmotic pressure, many archae halophiles accumulate molar concentrations of KCl in their cytoplasm.[246–248] Cytoplasmic proteins of these organisms, here termed halophilic proteins, hence remain functional and soluble at KCl concentrations far higher than the typical KCl cytoplasmic concentration ($\approx 0.15\,\mathrm{M}$) observed in most other organisms on Earth. Proteins (here called mesophilic) of non-halophilic organisms have typically low solubility at molar KCl concentrations.[249–251] Conversely, halophilic proteins often do not remain folded, stable or functional at low KCl concentration.[246,252–255]

Halophilic cytoplasmic proteins contain fewer large + hydrophobic and large + cationic amino acids, but contain more small + polar, small + apolar and acidic amino acids (predominantly at the surface,[253] than mesophilic proteins.[252,254,256] As a result, halophilic proteins have higher net negative charges than mesophilic ones (an average of $-8\,e$ to $-10\,e$ vs. $-1\,e$ to $-3\,e$, respectively).[254] Multiple hypotheses have been advanced to explain this difference, via the impact it will have on protein solubility (i.e., propensity for protein aggregation), thermodynamic stability of the folded protein conformation, and biological function of the molecule (e.g., enzymatic activity). Prior work has focused on clarifying the impact of acidic amino acids and salt on protein solubility and on the thermodynamic stability of folded proteins.[257] In contrast, our understanding of the mechanism(s) leading to strong salt dependency of enzymatic activity – a dependency that is often different from the salt-dependency of protein aggregation and protein conformational stability – and the eventual contribution of acidic amino acids to this dependency, is limited.

A key player in enzymatic activity is enzyme flexibility, defined as the amplitude and frequency of fluctuations in conformation experienced by the folded enzyme.[258–261] Higher protein flexibility is more often associated with motions of higher amplitude but may also be associated with faster motions (of the same amplitude). Enzyme activity involves physical steps necessary to bring substrates and cofactors to their binding sites and to release the product, as well as the chemical step itself.[262] Whereas the direct relation between enzyme flexibility and the chemical step is not fully elucidated, experimental and simulation studies have clearly demonstrated a connection between enzyme flexibility and the physical steps.[262,263]

According to the conformational-fluctuations model, the increased magnitude of the hydrophobic effect at high salt concentrations would decrease protein flexibility relative to lower salt concentrations.[264] In this scenario the increased intramolecular repulsion brought by the acidic proteome of halophilic proteins helps maintaining conformational fluctuations key for function at high salt concentration.[264] Studies of a limited number of non-halophilic proteins have suggested that a decrease in enzyme function correlates with increased thermodynamic stability in some cases, and that this correlation arises because of a reduction of protein flexibility.[259] To date, however, a comprehensive assessment of how protein structural flexibility responds to changes in salt concentration, whether that response differs between halophilic and mesophilic proteins, and the role of charged amino acids in this response does not exist.

Here we use NMR, molecular dynamics (MD) simulations, and analytical models to understand how protein structural flexibility is affected by the concentration of KCl. Simulations are performed on 5 pairs of halophilic and mesophilic proteins, and experiments are performed on 4 mesophilic proteins. Covering a wide range of sizes, net charges, and degrees of flexibility at their native electrolyte concentration, generalized conclusions on the dependency of flexibility on changes in salt concentration and the role of acidic amino acids in halophilic adaptation are obtained.

4.3 Results

4.3.1 Experimental Results

We first probed protein motion as a function of KCl concentration within four different – ^{15}N- or ^{15}N/^{13}C-labelled – mesophilic proteins with a molecular-weight range between 7 kDa and 42 kDa using NMR spectroscopy in solution. In particular, we targeted the SH3 domain of chicken α-spectrin (7.4 kDa), ubiquitin (8.5 kDa), human carbonic anhydrase II (29 kDa), and maltose binding protein (42 kDa). The first two proteins are small but contain all different forms of secondary-structural motifs; a very high signal-to-noise can be achieved. The last two proteins are reasonably complex structures with sizeable hydrophobic cores.

To comprehensively capture changes in protein flexibility as a function of salt concentration, we turned to experiments addressing two different time regimes. Fast (ps-ns timescale) motion was targeted using the [^{15}N, ^{1}H] heteronuclear nuclear Overhauser effect (hetNOE). hetNOE values approach 1 for rigid amino acids whereas values down to zero, sometimes lower, are found for very flexible elements of the protein primary sequence. The latter case is a representative of substantial fast (ps-ns) timescale fluctuations of the ^{1}H-^{15}N bond vector, with decreasing values both for faster and for higher-amplitude motion (lower order parameters). Fig. 4.1A-B show the hetNOE values of the SH3 domain and of hCAII, respectively, as a function of residue number, for no-salt and high-salt conditions. For SH3, several local differences are observed with and without KCl, but no global changes are apparent. With addition of KCl, hetNOE values of the N-terminus, 28Gly and 41Trp are slightly reduced, indicating slightly increased mobility on the ps-ns timescale, while hetNOE values are strongly increased for 38Asn (n-Src loop, from 0.079 to 0.813) and 47Asn (distal loop, from 0.097 to 0.725), indicating a considerable rigidification in the ps-ns timescale regime upon addition of KCl. In case of hCAII, hetNOE values increase for several residues upon addition of salt (e.g., residues 55–57, 59, 62, 92–94, 96 and 106–107), however, no global shift is visible. As the highest hetNOE values are expected to be uniform among the rigid residues and are defined by the overall tumbling correlation time, the outliers are likely due to a higher degree of uncertainty. For ubiquitin and MBP, similar trends were observed for hetNOE values (compare Fig. 4.10).

To assess global differences of hetNOE between low and high concentration of KCl, the sequential plots of all four proteins were statistically analyzed using box plots (Fig. 4.1C). In these plots the region between the first and third quartile is depicted as a box. Fig. 4.11 depicts the same data in the form of a histogram. For SH3 and ubiquitin, the average of hetNOE values for the 3 secondary structure categories investigated are very close to 1, indicating that these proteins are not very flexible in the ps-ns time scale. The distributions of hetNOE values are very narrow, especially for α-helical and β-sheet-like secondary structural regions. Box sizes and median values are virtually identical at low and high salt concentration, revealing no impact of salt concentration on dynamics of SH3 and ubiquitin. For hCAII and MBP, more hetNOE values deviate from 1, indicating that these proteins bear more flexible residues. Moreover, hetNOE values are much more dispersed at 2.0 M KCl concentration, which may reflect some impact of salt on the flexibility of these proteins but at least in part derives from the decreased sensitivity of the measurements under high-salt conditions. In agreement with the latter, despite the larger distributions, the median hetNOE values for the three secondary structure categories analysed here remain untouched by the high salt concentration. Overall, the hetNOE data show that changes in salt concentration have either no impact or only minor impact on protein dynamics, these effects being more pronounced for protein regions with higher mobility compared to those that are motionally restricted. The dependence of protein flexibility

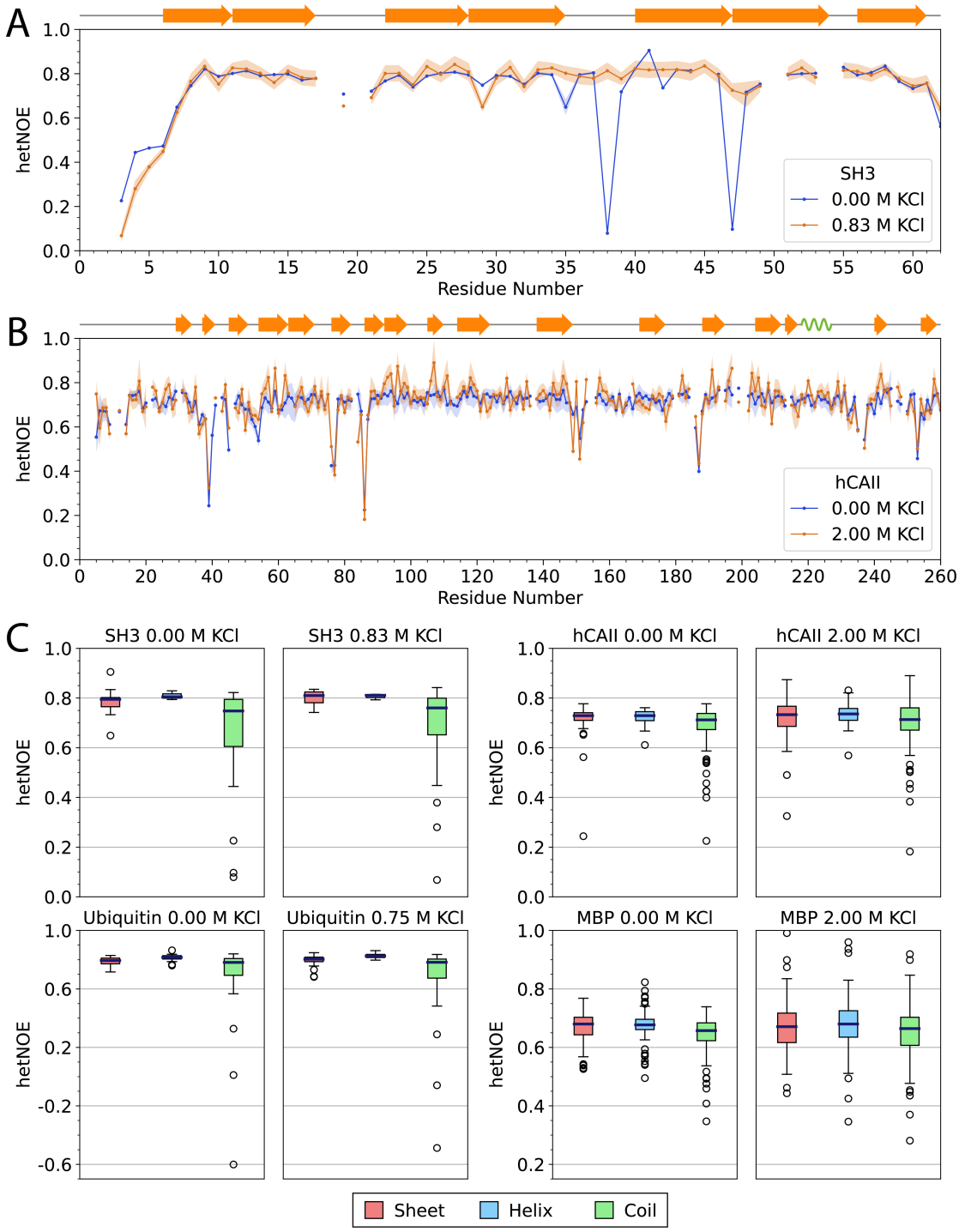

Figure 4.1: Fast (ps-ns) timescale motion with and without KCl, measured by [^{15}N, ^{1}H] hetNOE, defined as ratio $I_{\mathrm{sat}}/I_{\mathrm{eq}}$, where I_{sat} and I_{eq} are the peak intensities with and without proton saturation, respectively. (A) hetNOE of the SH3 domain of chicken α-spectrin. (B) hetNOE of human carbonic anhydrase II. The shaded areas around the curve represent the propagated error estimated from the spectral noise as described in section 4.5.1. (C) Distribution of hetNOE values (box plots) in the proteins SH3, ubiquitin, hCAII, and MBP for different secondary structure motifs. Boxes extend from the first to the third quartile with a line at the median. The whiskers extend over 1.5× the inter-quartile range from the box. Outliers are marked as white circles. All means and distributions are highly similar between no-salt and high-KCl conditions.

on salt concentration, whenever present, is not systematic: We observe both increases and decreases in residue flexibility with increasing salt concentration. Similar trends are observed in our chemical-shift perturbation study, described in section 4.5.1.

We investigated protein dynamics at a slower (µs-ms) timescale using Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion measurements. These quantify the excess contribution (R_{ex}) to the transverse relaxation rate R_2 that arises for a ^{15}N nucleus through sampling of differential isotropic chemical shifts upon conformational exchange. R_{ex} increases as the population of the minor state increases and as the difference in chemical shifts between the conformations increases. Whereas such measurements can also inform on conformational-exchange timescales and even populations, in this work we restrict our interpretation of differences in R_{ex} between two salt concentrations as indicating a different extent of conformational exchange. Note that caution has to be taken when interpreting relaxation dispersion data for different buffers due to their impact on the chemical shifts of ground and excited-state conformations[265–267]

Fig. 4.2A-B show the R_{ex} rates for the proteins ubiquitin and MBP. For ubiquitin, which shows an overall extremely low extent of motion at the µs-ms timescale, no significant alterations for exchange rate values are observable between the two salt concentrations. Likewise, even though MBP displays a larger range of slow motions than ubiquitin, systematic changes in dynamics with salt concentration over the entire protein are also not observed. The higher degree of fluctuations observed for MBP at the high-salt condition likely derives from measurement uncertainties; the standard deviation of the R_{ex} values over the entire protein is 10.5 Hz. Several residues show a higher exchange contribution at low salt (e.g., residues 62, 63 and some between 150 and 175; blue). In contrast, other residues show a higher exchange contribution at high salt (e.g., residues 12, 13, 223, 251). However, no clear connection between the structural motif in which these residues are embedded or their extent of surface exposure is discernible. Fig. 4.12 shows the R_{ex} rates as a function of sequence for the remaining proteins SH3 and hCAII under no-salt and high-salt conditions. Similarly to MPB and ubiquitin, SH3 and hCAII do not show global changes in dynamics with salt concentration, with only a few residues in hCAII showing statistically different values of R_{ex} between the two salt concentrations.

We statistically assessed the distributions of all four proteins again in the form of box plots (Fig. 4.2C). For SH3 and ubiquitin, the distributions of R_{ex} rates are again very small. Apart from a minor change in the average R_{ex} rates for β-sheet-like secondary structure elements in hCAII with salt concentrations, the box plots show no significant changes in R_{ex} rates with salt concentration. The histograms of exchange contributions (Fig. 4.13) of hCAII and MBP differ quantitatively between the two salt conditions whereas those for ubiquitin and for SH3 are practically identical. Nevertheless, a systematic impact of salt on the dynamics of particular structural motifs or on the dynamics of the whole protein is not found for any of the 4 proteins.

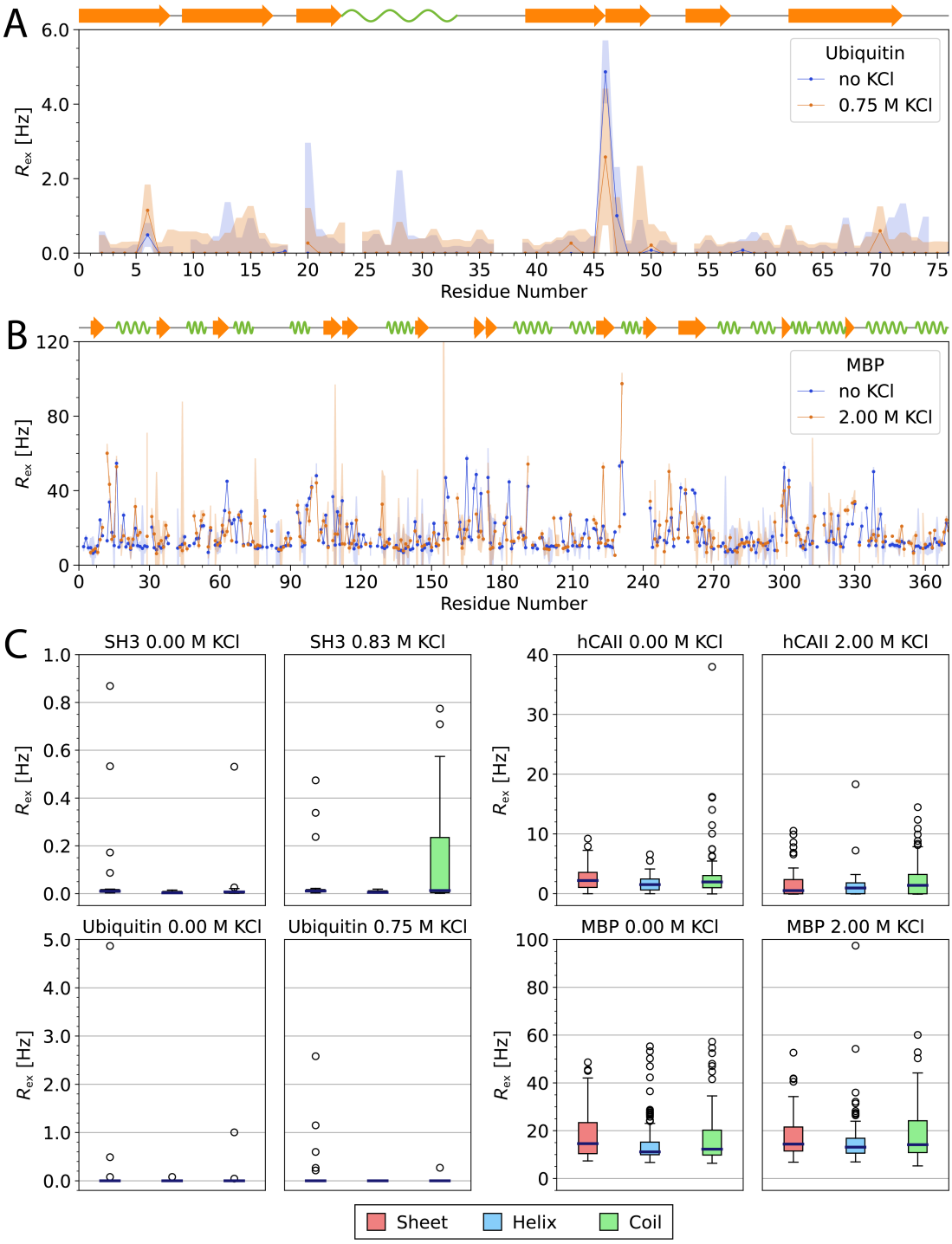

Figure 4.2: Fast (ps-ns) timescale motion with and without KCl, measured by [^{15}N, ^{1}H] hetNOE, defined as ratio I_{sat}/I_{eq}, where I_{sat} and I_{eq} are the peak intensities with and without proton saturation, respectively. (A) hetNOE of the SH3 domain of chicken α-spectrin. (B) hetNOE of human carbonic anhydrase II. The shaded areas around the curve represent the propagated error estimated from the spectral noise as described in section 4.5.1. (C) Distribution of hetNOE values (box plots) in the proteins SH3, ubiquitin, hCAII, and MBP for different secondary structure motifs. Boxes extend from the first to the third quartile with a line at the median. The whiskers extend over 1.5× the inter-quartile range from the box. Outliers are marked as white circles. All means and distributions are highly similar between no-salt and high-KCl conditions.

4.3.2 MD Simulations

To investigate how atomic fluctuations respond to salt concentration and the potential role of acidic amino acids in the maintenance of protein flexibility at high salt concentration, we performed atomistic MD simulations comparing 5 halophilic proteins with their mesophilic counterparts. An illustrative simulation box is shown in Fig. 4.3; names and PDB IDs given in Fig. 4.4). Each halophilic-mesophilic pair was selected to ensure high similarity in amino acid sequence and structure between the two proteins while having a wide range of sizes and net charges in each category, as discussed in Section 4.6.1. This diverse and representative group of proteins enables us to generalize conclusions based on our observations.

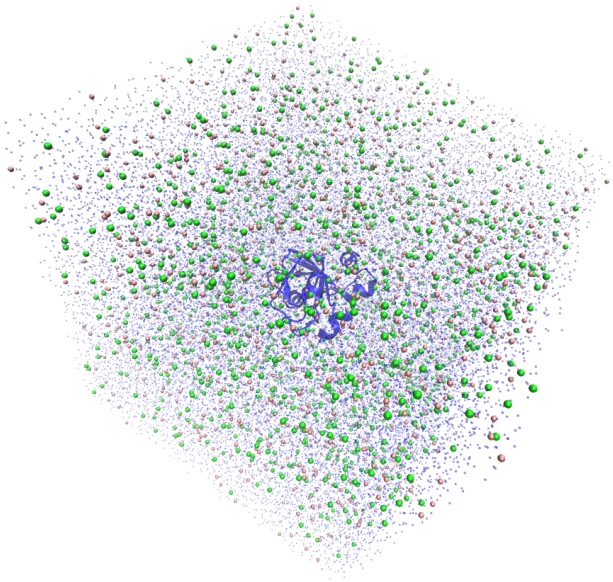

Figure 4.3: Example simulation box, of the halophilic ferredoxin protein (PDB ID: 1DOI). The dark blue shape corresponds to the cartoon representation of the protein, K^+ ions are shown as pink spheres, Cl^- ions are shown as green spheres and water molecules depicted as transparent small blue dots.

Each protein was simulated in aqueous KCl solutions with molality $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ and $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$. The high salt concentration is thus non-optimal for the mesophilic proteins; the low salt concentration is likewise non-optimal for the halophilic ones. In section 4.6.3 we show that protein structures remain stable during the entire simulation.

Conformational Flexibility

The Root Mean Square Fluctuation (RMSF) quantifies the amplitude of atomic displacements relative to the average position, as detailed in section 4.6.3. Fig. 4.4 shows the RMSF per residue (heavy atoms only) of each protein, at high and low salt concentration; Fig. 4.16 shows the same data in histogram form. Similarly to what was observed in the hetNOE values and R_{ex} rates, different protein sections and even adjacent residues in the protein sequence can have dramatically different dynamics. Increasing salt concentration

impacts protein flexibility only locally, for protein residues that show high flexibility (as measured by RMSF larger than 2 Å) at low salt concentration, which are predominantly found in coil structure (Fig. 4.15). The majority of flexible protein sections (found in proteins 2L28, 2ITH, 1FRD, 1HZ6 and 4CNX) becomes less flexible at high salt concentration. The opposite salt dependency is also seen, but more rarely: some sections of the mesophilic proteins 1V9E and 2L28 and of the halophilic protein 3WRT show larger RMSF at higher salt concentration. In contrast, the dynamics of residues without large-amplitude motions is only marginally sensitive to salt concentration; this weak dependence is easily seen in the the RMSF histograms (Fig. 4.16) of proteins 1ZKJ, 3WRT, 2KAC and 1V9E. These results are in line with a prior computational study of a ubiquitin-like protein,[268] a neutron scattering study of bovine serum albumine,[269] and our NMR experiments.

Mechanism of Flexibility Modulation by Salt

To understand the origin of the non-monotonic effect of salt on protein dynamics, we consider the impact of salt concentration on the hydrophobic effect and on electrostatics. We consider only salts without strong specific interactions with protein charged groups, as is the case for KCl.[270]

The Hydrophobic Effect and Protein Flexibility

Fig. 4.5A shows normalised hydrophobic solvent-accessible surface area (SASA) histograms, $P(A)$, for two of the proteins, obtained from simulation; the histograms for the other proteins are shown in Fig. 4.17. The SASA distributions are mono-peaked and show little or no skewness, so the distributions can be characterized by their mean and width. Fig. 4.5B shows the mean hydrophobic SASA of the proteins studied in simulation; numerical values are shown in Table 4.5. The mean hydrophobic SASA varies between 1000 Å^2 and 4000 Å^2. For 8 of the 10 proteins studied here, the hydrophobic SASA is lower at high salt concentration, by between 10 Å^2 and 60 Å^2. Only the proteins 2L28 and 2ITH have higher mean hydrophobic SASA, by ≈ 60 Å, at the higher salt concentration. For either type of response, the mean hydrophobic SASA of each protein at the two salt concentrations differs less than the surface area of a single methane molecule, indicating that the mean hydrophobic SASA depends only weakly on salt concentration. But, what is the impact of salt on hydrophobic SASA fluctuations?

We quantify SASA fluctuations through the width of the histogram at 1/10 of the histogram maximum. Fig. 4.5C indicates that the width of the hydrophobic SASA histograms is very weakly reduced at high salt concentration for seven of the 10 proteins studied. Only for three of the proteins – the halophilic 2ITH and the mesophilic 2L28 and 1HZ6 – the hydrophobic SASA histogram becomes markedly narrower at high salt concentration. These results correlate well with the salt-dependent RMSF values presented above, in which proteins 1HZ6, 2ITH and 2L28 show the strongest salt-dependent flexibility.

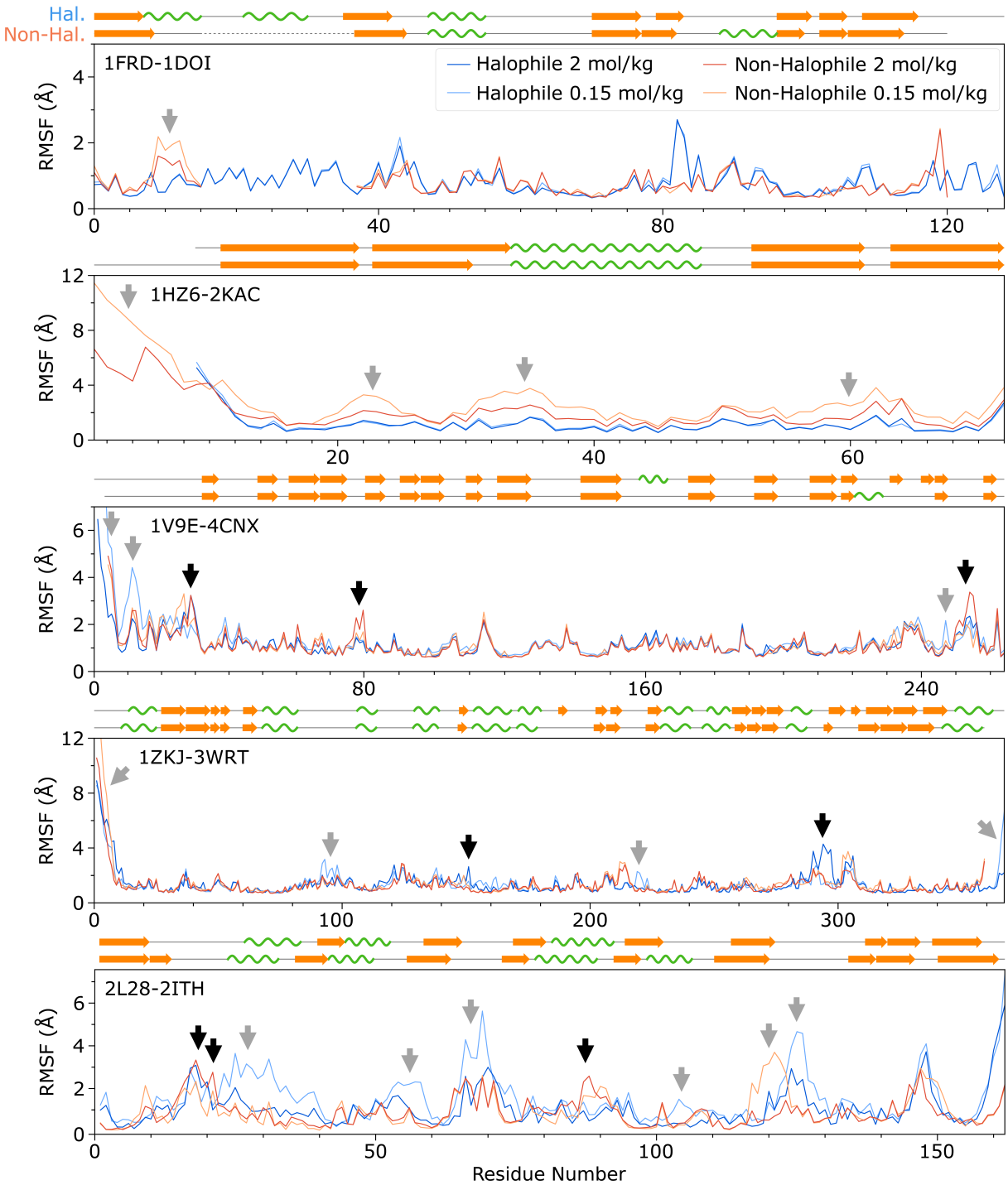

Figure 4.4: RMSF of each residue of the indicated mesophilic-halophilic protein pairs, at high and low KCl concentration, calculated from MD simulation. (A) ferredoxin (PDB ID 1DOI, 1FRD); (B) protein L (PDB ID 2KAC, 1HZ6); (C) carbonic anhydrase (PDB ID 4CNX, 1V9E); (D) β-lactamase (PDB ID 3WRT, 1ZKJ); (E) dihydrofolate reductase (PDB ID 2ITH, 2L28). Black arrows point to some of the residues for which the RMSF of a protein is higher at the high salt concentration; grey arrows point to some of the residues for which the RMSF is higher at the low salt concentration.

To understand the extent to which changes in SASA fluctuations with salt concentration can be explained by the hydrophobic effect, we take the commonly used approximation[257,271] that the free-energy cost ΔG_{cav} of introducing an apolar group in aqueous solution scales linearly with the surface tension γ of the solution and with the area A, of the group according to:

$$\Delta G_{\mathrm{cav}} = \gamma \cdot A \tag{4.1}$$

The impact that a change in salt concentration has on fluctuations in the hydrophobic SASA via a change in the hydrophobic effect can be estimated from equation 4.1 and from the probability density, $P_{\mathrm{ref}}(\Delta A)$, of those fluctuations at a reference salt concentration, as

$$P_c(A) = D_2 \cdot P_{\mathrm{ref}}(A) \cdot \exp\left(-\frac{\Delta\gamma A}{k_{\mathrm{B}}T}\right) \tag{4.2}$$

where $P_c(A)$ is the probability density of the SASA fluctuations at the new salt concentration, $\Delta\gamma = \gamma_c - \gamma_{\mathrm{ref}}$ is the difference of the surface tension at the two concentrations, T is the temperature, k_{B} is the Boltzmann constant and D_2 is a normalization constant (see section 4.6.3 for further details).

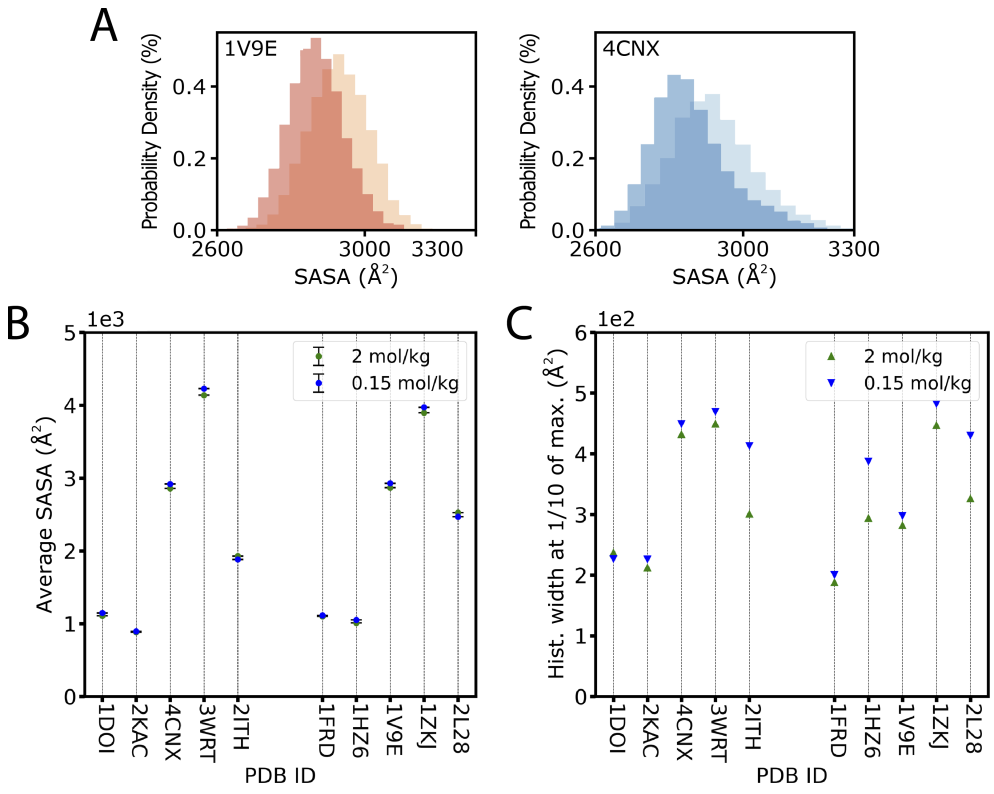

Figure 4.5: Hydrophobic SASA associated with hydrophobic residues at high and low salt concentration, calculated from MD simulations: (A) illustrative histograms for proteins with the indicated PDB IDs at $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$ (dark blue and orange) and $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ (light blue and orange); protein 4CNX is halophilic and protein 1V9E is mesophilic; (B) Average SASA; the error bars show the standard error of the mean; (C) histogram width at 1/10 of the maximum height. Differences below $50\,\mathrm{Å}^2$ are not statistically significant.

Figure 4.6 shows distributions of hydrophobic SASA obtained through this model. Modifying that reference distribution according to equation 4.2 for estimates of $\Delta\gamma$ between $0.15\,\mathrm{mol\,kg^{-1}}$ and $2.00\,\mathrm{mol\,kg^{-1}}$ leads to the $P_c(A)$ curves shown in blue when considering macroscopic values of the surface tension of electrolyte solutions, and in orange when considering that the microscopic surface tension is roughly half of the macroscopic one.[257,272] The increased hydrophobic effect systematically lowers the mean hydrophobic SASA by $40\,\mathrm{Å^2}$ when using microscopic γ values, and by $100\,\mathrm{Å^2}$ when using macroscopic γ values. The decrease in mean hydrophobic SASA at high salt concentration predicted by this model is consistent with simulation results for 8 of the 10 proteins investigated here (see Table 4.5). The model also predicts that salt concentration has no impact on the amplitude of the hydrophobic SASA fluctuations relative to the mean. This response occurs because the energetic penalty associated with changes in salt concentration depends linearly on A. This simple model suggests that the enhanced hydrophobic effect at high salt concentration will not impact protein flexibility as measured by the width of hydrophobic SASA fluctuations, although it will slightly reduce the mean hydrophobic SASA, consistent with the results in Figure 4.5B. It follows that the narrowing of the width of the SASA histograms of proteins 2ITH, 1HZ6 and 2L28 must be the result of other mechanisms.

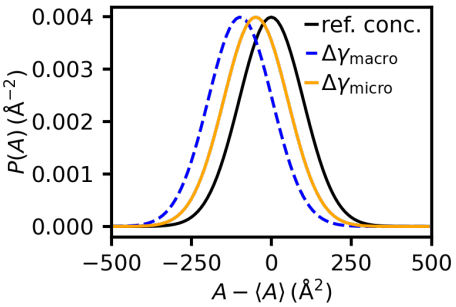

Figure 4.6: Probability distribution of the hydrophobic SASA at $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$ considering macroscopic ($\Delta\gamma_{\mathrm{macro}}$) and microscopic ($\Delta\gamma_{\mathrm{micro}}$) values of surface tension predicted through equations 4.1 and 4.2. For the prediction of surface tensions, a probability density $P_{\mathrm{ref}}(A)$ at the reference salt concentration was modelled as a normal distribution with mean $\langle A\rangle$ and standard deviation $100\,\mathrm{Å^2}$.
$\Delta\gamma_{\mathrm{micro}} = 0.5 \cdot \Delta\gamma_{\mathrm{macro}} = 0.5 \cdot (76 - 72) \cdot 10^{-3}\,\mathrm{N\,m^{-1}}$.[257,272,273]

Electrostatic Interactions and Protein Flexibility

We consider two elementary charges as a model of charged amino acids in a folded protein. In the simulation environment these charges will be under the influence of multiple other potentials (bond, angle, dihedral, Lennard-Jones potentials, electrostatic interactions with other charges). The impact of these interactions on the inter-charge distance is implicitly captured by the distribution $P_{\mathrm{ref}}(r)$ of inter-charge distance at a reference salt concentration, and is assumed to be independent of salt concentration. Applying a formalism analogous to that used to model the hydrophobic SASA fluctuations, the distribution of inter-charge distance $P_{\mathrm{c}}(r)$ at a given salt concentration c can be estimated

from $P_{\mathrm{ref}}(r)$ by considering the impact that the change in salt concentration has on the electrostatic interactions between the two charges in question:

$$P_c(r) = E \cdot \exp\left[-\frac{U_{\mathrm{electr,c}}(r) - U_{\mathrm{electr,ref}}(r)}{k_{\mathrm{B}}T}\right] \cdot P_{\mathrm{ref}}(r) \tag{4.3}$$

E is a normalization constant and $U_{\mathrm{electr}}(r)$ is modelled using the Debye-Hückel theory. The salt dependence of electrostatic interactions is present in the salt-dependent values of the relative permittivity of the electrolyte solution and of the Debye length. Further details are given in section 4.6.3.

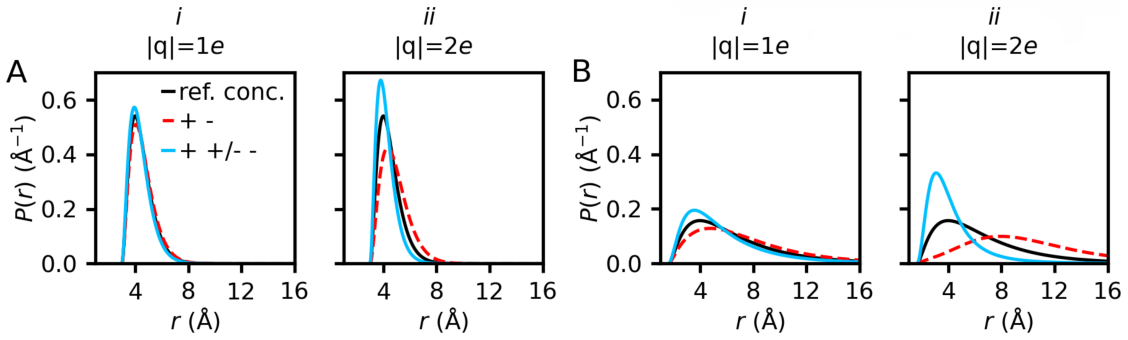

Figure 4.7: Probability distribution of inter-charge distance fluctuations at $2.00\,\mathrm{mol\,kg^{-1}}$ (blue and dashed red curves) predicted through equation 4.2 from $P_{\mathrm{ref}}(r)$ at the reference concentration $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ for (A) narrow and (B) wide model $P_{\mathrm{ref}}(r)$ distributions, for (*i*) mono and (*ii*) divalent charges.

Figure 4.7 shows predictions from equation 4.3. The reference distributions are modelled as asymmetric potentials because the inter-charge distance cannot be smaller than the van der Waals radius of atoms. Panel A.*i* shows predictions when the distance distribution at the reference concentration is narrow and thus does not allow substantial fluctuations. In this case, increasing salt concentration barely impacts the inter-charge distance, reflecting the fact that the other interactions affecting the charges are much more intense than their electrostatic interaction. In panel A.*ii*, we show the response of the fluctuations if the two charges have magnitude $2e$, as an approximation to the scenario of a pair of charged amino acids interacting with a nearby pair of charged amino acids. In this case, the distribution of inter-charge distances responds, albeit still weakly, to the change in salt concentration. Because of stronger charge screening at high salt concentration, like-charges on average come closer together *and* they experience narrower fluctuations around the new mean distance. In contrast, unlike-charges on average move further apart *and* fluctuate more amply around the new mean distance. The scenario of changes under the influence of a soft potential, with FWHM $\approx 5\,\text{Å}$, is examined. Panel B.*ii* shows predictions for single point charges. Distance fluctuations now clearly reflect an impact of salt concentration, with unlike-charges substantially increasing the range of their distance fluctuations as salt increases and like-charges showing the opposite behaviour. Panel B.*ii* clarifies that changes in fluctuations in response to salt concentration become dramatically more intense, for both like and unlike charges, when the two charges have magnitude $2e$.

The results from these analytical models qualitatively explain both the salt-dependent and the salt-independent protein dynamics observed in simulations and experiment, as illustrated in Figure 4.8. Only minimal changes in *global* protein flexibility with salt concentration are seen for many folded proteins because they are rigid. In this case, changes in the magnitude of the hydrophobic effect or of electrostatic interactions remain small relative to the other interactions governing protein flexibility, so flexibility changes little with salt concentration; this is seen in the small changes ($< 100\,\text{Å}^2$, less than one methane molecule) in the width of the hydrophobic SASA histograms as a function of salt concentration experienced by all proteins investigated in simulation (Figure 4.5). The hydrophobic effect can strongly reduce the flexibility of flexible protein sections that contain hydrophobic amino acids and are solvent-exposed (orange arrow in Figure 4.8; hydrophobic amino acids are not shown). More often, however, a strong reduction in local protein flexibility is associated with protein regions rich in like-charges, which at high salt concentration no longer strongly repel (black arrows in Figure 4.8). Likewise, a strong local increase in protein flexibility at high salt concentration may also occur when stabilizing salt bridges at low salt concentration become screened at high salt concentration (red arrows in Figure 4.8).

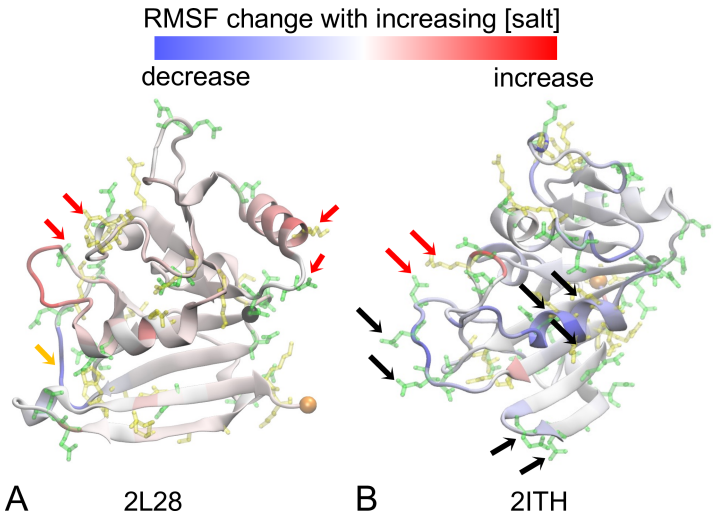

Figure 4.8: Local change in protein flexibility (ΔRMSF), as measured by the difference in the RMSF values of each amino acid (i) at high and low salt concentration ($\Delta\text{RMSF}(i) = \text{RMSF}_{2\text{m}}(i) - \text{RMSF}_{0.15\text{m}}(i)$; values from Fig. 4.4), for the (A) mesophilic and (B) halophilic dihydrofolate reductase proteins investigated in simulations. The blue-white-red color scale indicates a reduction-maintenance-increase in protein flexibility with increasing salt concentration. Acidic amino acids are depicted in green, basic basic amino acids in yellow and amino acids at the termini are shown grey, orange spheres. Red and black arrows indicate unlike-charge/like-charge interactions that are screened at high salt concentrations and contribute to positive/negative values of ΔRMSF, respectively. The orange arrow indicates a negative ΔRMSF likely resulting from an increase in the hydrophobic effect.

4.4 Conclusion

Our results demonstrate that changes in salt concentration barely impact the dynamics of rigid protein sections, but may noticeably impact the dynamics of flexible protein regions. The stronger hydrophobic effect at high salt concentrations has limited bearing on the flexibility of folded proteins, contrary to expectations.[259,264] Instead, changes in salt concentration influence protein flexibility primarily through electrostatics. As salt concentration increases, flexible protein regions rich in like-charges will experience an increase in low-amplitude motions because charge screening will enable the protein to explore conformations where the charges come closer together, at the expense of expanded conformations. We note that lower-amplitude motions may manifest as an enhancement in *short* timescale dynamics, because low amplitude motions are typically faster. An enhancement of fast dynamics at high salt concentration has been reported from neutron scattering experiments on tRNA.[274,275] By contrast, flexible regions rich in unlike-charges may increase the amplitude of their motions when salt concentration increases because of screening of salt bridges at high salt concentration.

Our results do not support the scenario that an enrichment in acidic amino acids and the corresponding high negative net charge is necessary to ensure protein flexibility in a halophilic environment.[276] More likely, excess acidic amino acids are involved in other mechanisms that enable halophilic proteins to function at high salt concentration, such as reduction of solvent-accessible surface area,[277] preferential solvation of the folded state,[276,278–280] cooperative carboxylate-cation-water networks[276,281] or prevention of aggregation.[277,280,282–284]

Our results identify which physical information must be considered to efficiently predict protein conformational ensembles and protein dynamics as a function of salt concentration. They also highlight the complex and non-linear interplay between all factors contributing to protein dynamics, which makes it difficult to predict it using simplified physical models. Strong predictive capabilities may be more easily obtained with machine learning,[285] which can efficiently handle non-linearities, in particular physics-informed machine learning approaches[286] which can incorporate the physical insights revealed here. The beneficial impact of these insights on future predictions will be critical to guide the modification of biotechnologically useful enzymes[287–290] so they can remain functional in saline solutions thus reducing dependency on our fresh water resources, to obtain improved understanding of the connection between salt-dependent protein dynamics and protein-protein interactions in the course of evolution,[291–294] and to foster improved strategies for formulations of protein pharmaceutics and in structural biology.

4.5 Materials and Methods

4.5.1 Experimental Methods

Proteins Used for NMR Spectroscopy

Nuclear magnetic resonance (NMR) experiments were performed on 4 proteins: α-spectrin SH3 domain (SH3), ubiquitin, human carbonic anhydrase II (hCAII), and maltose binding protein (MBP). All proteins were purchased from Giotto Biotech, Italy. Table 4.2 lists the PDB IDs and amino acid content of each protein, and Table 4.3 details the experimental conditions used for each protein.

NMR Sample Preparation and Measurement

For NMR studies, uniformly ^{15}N or ^{13}C/^{15}N-labelled protein samples were prepared in a mixed solvent of 90 % $^{1}H_2O$ and 10 % $^{2}H_2O$ (sample conditions summarized in Table 4.3). All NMR experiments were carried out with protein concentrations of 0.5 mM on a Bruker Avance 800 MHz NMR spectrometer equipped with a 5 mm cryogenically cooled triple-resonance probe and a pulse-field gradient. The ^{1}H, ^{15}N resonance assignments of SH3, ubiquitin, hCAII and MBP proteins were taken from the Biological Magnetic Resonance Bank (BMRB, accession numbers 34420, 1541, 34308 and 4986, respectively). In the case of MBP, assignments were further completed using a 3D HNCA triple-resonance experiment. Chemical shifts for high-salt conditions were obtained using a titration approach, upon which original H/N HSQC or HNCA peaks could be tracked upon increase of KCl concentrations and associated with their new positions.

Chemical-Shift Perturbations

Chemical-shift perturbations (CSPs) were calculated from ^{1}H and ^{15}N chemical-shift differences between protein solutions with and without KCl according to the equation $\sqrt{\Delta\delta_{\mathrm{H}}^2 + (0.14 \cdot \Delta\delta_{\mathrm{N}})^2}$.

[^{15}N, ^{1}H] heteronuclear nuclear Overhauser effect

Steady-state [^{15}N, ^{1}H] heteronuclear nuclear Overhauser effect (hetNOE) measurements were carried out for 4 protein samples (see Tables 4.2 and 4.3) with and without proton saturation during the relaxation delay. Either 5 s of relaxation delay and 3 s of proton saturation, or 8 s of relaxation delay without successive proton saturation were used. hetNOE values were determined as ratios of peak intensities measured from spectra acquired with and without proton saturation. NMR spectra were processed using Bruker TopSpin 4.0.8 and analysed using CCPN.[295] For estimation of intensity errors, 10 randomly distributed noise peaks were picked in both spectra and the absolute intensity values were averaged. Errors of hetNOE values were calculated using error propagation assuming that the distributions underlying all data are normally distributed.

^{15}N Relaxation Dispersion

Constant-time ^{15}N-CPMG (CT-CPMG) relaxation dispersion experiments[296] were measured for all four protein samples (experimental details summarized in Table 4.1). Experiments were performed with a constant-time delay as mentioned in Table 4.1 for different protein samples. Nine variable CPMG frequencies (ν_{CPMG}) ranging from 50 Hz to 2000 Hz were collected. Besides, for each dataset the frequencies 50 Hz and 750 Hz were repeated for estimation of the error in $R_{2,\mathrm{eff}}$, and a reference spectrum without constant time delay ($T_{\mathrm{CPMG}} = 0$) was recorded. For each 2D dataset corresponding to one CPMG frequency, 128 and 2048 complex points in the indirect and direct dimensions, respectively, were collected with 32 scans. 2D datasets for all frequencies were measured in a scan-interleaved fashion. A recycle delay of 1.5 s was used, giving rise to a net acquisition time between 15 h and 52 h per sample. NMR relaxation data were processed using Topspin 4.0.8. Peak intensities were quantified and visualized with CCPN.[295] The effective amide ^{15}N transverse relaxation rate at each CPMG frequency was calculated according to the relation of effective transverse relaxation rate ($R_{2,\mathrm{eff}}$),

$$R_{2,\mathrm{eff}} = -\frac{1}{T_{\mathrm{CPMG}}} \cdot \ln\left(\frac{I}{I_0}\right) \tag{4.4}$$

where I is the peak intensity, I_0 is the reference peak intensity of a spectrum recorded without CT-CPMG relaxation period, and T_{CPMG} is a constant time delay.
Evaluation of CPMG data was performed using ChemEx.[80] Relaxation dispersion curves were fitted analytically according to the Bloch-McConnell equations on the basis of a two-state exchange model. Errors were estimated from duplicate points measured for 50 Hz and 750 Hz field strength. The exchange contribution to the exchange rate R_{ex} was determined as the difference between the maximum calculated effective transverse relaxation rate $R_{2,\mathrm{eff}}$ (maximum of the fit curve) and the fitted intrinsic transverse relaxation rate R_2^0.

Evaluation of hetNOE and R_{ex} values

Sequential hetNOE and R_{ex} plots, histograms and box plots were generated using the Matplotlib module of Python 3; the sequential secondary structure representations on top of the plots were generated using the Biotite module of Python 3. To better identify differences in behaviour between different secondary structure elements, hetNOE and CPMG values were grouped into three secondary-structure classes (sheet, helix and coil) and normalised for each secondary-structure class. Secondary-structure elements were determined based on the X-ray structures with PDB ID 2NUZ, 1UBQ, 2CBA and 1ANF using DSSP[297] and grouped into three instead of the default eight DSSP secondary structure types (3-, 4- and 5-turn helices (G, H and I) as helix; extended strands in β-sheet conformation (B) and residues in isolated β-bridges (E) as sheet; bends (S), hydrogen bonded turns (T) and unstructured regions (U) as coil).

Table 4.1: Conditions of hetNOE and CPMG experiments.

	SH3	**Ubiquitin**	**hCAII**	**MBP**
T_{CPMG} (ms)	80	80	40	30
δ_{O1P} (ppm)	116.0	116.0	116.0	119.0
PW_{90} without KCl (µs)	7.30	7.19	7.35	7.10
PW_{90} with KCl (µs)	9.47	10.12	13.36	15.23

4.5.2 Simulation Methods

We used the TIP3P water model,[298] the AMBER ff14SB[98,299] force field for proteins and the potassium and chloride parameters of Joung and Cheatham[300] for TIP3P water. The Lennard-Jones (LJ) parameters for interactions involving carboxylate groups were modified from the default values obtained using combination rules. The self-interaction LJ parameters for all carboxylate oxygen atoms use the values derived by Kashefolgheta et al.,[301] which approximate the hydration free energy of acetate in TIP3P water better than the original AMBER parameters. The LJ parameters for the interaction between carboxylate and the side chain of the positively charged amino acid lysine were also modified,[301] thus avoiding unphysically strong salt bridges. The LJ parameters for the interaction between carboxylate groups and potassium were modified to improve the solution activity derivative and the carboxylate-potassium distance in a protein crystal.[302] The modified parameter values are given in Section 4.6.3.

The $[2\mathrm{Fe\text{-}2S}]^{2+}$ ligand present in the ferredoxin proteins was simulated using the parameters developed by Carvalho et. al.[303] for the [2Fe-2S] ferredoxin from *Mastigocladus laminosus*, with some modifications: i) The equilibrium angles in our simulations are those in the crystal structure of our proteins. ii) the force constant for the Fe-S-C angle, missing in the parameter set of Carvalho et al. for the $[2\mathrm{Fe\text{-}2S}]^{2+}$ from *Mastigocladus laminosus*, is given the value for the desulforedoxin protein from *Desulfovibrio gigas*.[303] This change is acceptable because the ferredoxins studied here contains Fe(III) coordinated by four cysteines in the same configuration as those in the desulforedoxin protein. The zinc metal center of the carbonic anhydrase proteins is connected to three histidine residues and one water molecule. That metal center was simulated using ZAFF (Zinc AMBER force field)[304,305] and the van der Waals (vdW) parameters for the zinc ion from Li et al..[305] Simulations were performed using the GPU version of the pmemd engine in AMBER 2018.[284] Only the l-bfgs minimization step was performed on the Sander engine of AMBER 2018 because it was not available in the pmemd engine. Each protein was simulated at $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ and $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$. The simulation boxes are cubic and have edge lengths of $\approx 100\,\text{Å}$; the minimum distance between the protein surface and the box face was $\approx 20\,\text{Å}$. Periodic boundary conditions were applied in all directions. The SHAKE[306] algorithm was used to constrain all bonds with hydrogen atoms. A 1 fs

time step was used to simulate the carbonic anhydrase proteins because of the water molecule that coordinates the zinc metal center, whereas a 2 fs time step was used for all other simulations. A Langevin thermostat with a collision frequency of 0.01 ps^{-1} was used to keep the average system temperature at 298 K. This low collision frequency was necessary to reduce the impact of the thermostat on the dynamics of the system, and was sufficient to stabilize the average temperature. This low collision frequency ensures that the self-diffusivity, rotational correlation time, and shear viscosity stay within 2 % of those obtained from NVE simulations at the same temperature.[110] The production simulations were done in the NVT ensemble and lasted 0.5 µs for carbonic anhydrase and 1 µs for all other proteins, with configurations saved every 100 ps for analysis. Further simulation details can be found in Section 4.6.3 and in our prior publication.[302] These simulations were previously analysed to understand the structure and dynamics of the protein solvation shells; those results are reported in that prior publication.[302] Here we focus exclusively on the dynamics of the protein.

Protein conformational dynamics was characterized in simulations by calculating the root mean square fluctuations (RMSF) and the hydrophobic surface area (SASA) as function of time. This analysis was done with VMD[307] and the Cpptraj package[308] from AMBER. The median protein structure was calculated using the software package Theseus.[309,310] Details of these calculations can be found in Section 4.6.3.

4.6 Supporting Information

4.6.1 Studied Proteins

Relevant characteristics of the proteins investigated in experiment and in simulation are displayed in Table 4.2. For the simulations, each halophilic-mesophilic pair was selected to ensure high similarity in amino acid sequence and structure between the two proteins while having a wide range of sizes and net charges in each category. The pair sequence identity varies between 22 % and 90 %. The proteins in each pair share a common structure, visible in the narrow range of the Root Mean Square Deviation (RMSD) of the common C^{α} atoms between protein pairs, after structural alignment: $0.95 < \mathrm{RMSD/\AA} < 2.41$. The halophilic protein group contains the highly charged ferredoxin ($-29e$), with 27 % acidic and only 5 % basic amino acids, as well as less charged species like dihydrofolate reductase ($-15e$), which has only 19 % and 9 % acidic and basic amino acids, respectively. In contrast, the mesophilic proteins are weakly negative and have between 8 % (beta-lactamase) and 18 % (ferredoxin) acidic amino acids. Halophilic proteins have typically 17 % to 20 % acidic amino acids and only 8 % to 10 % basic amino acids, whereas mesophilic proteins are weakly negatively charged, having 12 % to 14 % acidic and 10 % to 12 % basic amino acids, so our selected proteins are good representatives of these categories.

Table 4.2: Total number of amino acids (n_{aa}), number of acidic (n_{acidic}) and basic (n_{basic}) amino acids and charge of the proteins investigated in simulation and in experiment. The information for proteins investigated in simulation was originally summarized in [302] and is repeated here for the convenience of readers.

Halophiles					**Mesophiles**				
PDB	n_{aa}	n_{acidic}	n_{basic}	charge	PDB	n_{aa}	n_{acidic}	n_{basic}	charge
1DOI*	128	34	7	$-29e$	1FRD*	98	18	8	$-12e$
2KAC*	64	16	1	$-15e$	1HZ6*	72	10	8	$-2e$
3RWT*	367	57	30	$-27e$	1ZKJ*	359	29	32	$3e$
4CNX*	259	48	29	$-17e$	1V9E*	256	30	28	$0e$
2ITH*	162	30	15	$-15e$	2L28*	162	22	18	$-4e$
					2NUZ**	62	11	11	$0e$
					1UBQ**	76	11	12	$1e$
					2CBA**	260	32	43	$11e$
					1ANF**	370	51	46	$-5e$

* Investigated in simulations.

** Investigated experimentally.

4.6.2 Experimental Methods

Resonance assignments under low-salt conditions were pursued via transfer of chemical shifts tabulated in the BMRB, in conjunction with additional 3D backbone experiments in the case of MBP. Highest possible salt conditions compatible with both protein integrity and sensitivity of the cryo probe were assessed via a stepwise addition of KCl. Final conditions were chosen to be slightly below the KCl concentrations at which each protein started to precipitate. Likewise, chemical-shift assignments under these conditions were obtained via tracking H/N resonances upon KCl titration using a series of 2D HSQC spectra. For MBP, to disambiguate the strongly overlapping HSQC peaks, changes in chemical shifts in the course of increasing KCl concentrations were tracked using 3D HNCA correlations.

Table 4.3: Conditions of protein samples investigated experimentally.

	SH3	**Ubiquitin**	**hCAII**	**MBP**
Full name	α-spectrin SH3 domain	Ubiquitin	human Carbonic Anhydrase II	Maltose Binding Protein
Labelling	^{15}N	^{15}N	^{15}N	^{13}C/^{15}N
Amount (mg)	5	5	10	10
Buffer	Citrate buffer	Phosphate buffer	Tris buffer	Phosphate buffer
Buffer conc. (mmol/l)	10	10	20	10
pH	3.5	6.5	7.5	6.5
NaN_3 (v/v)	0.02	0.02	0.00	0.02
KCl conc. (mmol/l)	0.83	0.75	2.00	2.00

4.6.3 Computational Methods

Force Field Parameters

The van-der-Waals (vdW) interatomic interaction is mimicked in the simulations through the Lennard-Jones potential:

$$V_{ij}(r) = \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] \tag{4.5}$$

In expression 4.5, $R_{\min,ij}$ and ε_{ij} are Lennard-Jones (LJ) parameters governing the interaction between different atom types, and are typically obtained from their self-interaction parameters using Lorentz-Berthelot (LB) combination rules:

$$R_{\min,ij} = \frac{R_{\min,ii} + R_{\min,jj}}{2}, \quad \varepsilon_{ij} = \sqrt{\varepsilon_{ii}\,\varepsilon_{jj}} \tag{4.6}$$

where the indices ii and jj denote the self-interaction parameters.

Table 4.4: Recommended values of LJ ε and $R_{\min}$ for the indicated interactions involving carboxylate groups (e.g., in the side chain of acidic amino acids), primary amines (e.g., in the side chain of amino acid lysine) and potassium ions, when using TIP3P water, the Joung and Cheatham[300] parameters for K^+ and the GAFF[311] or AMBER[98] parameters for amino acids.

Applied to	**Parameter**	**Source**	**Value**
carboxylate O$^{\#}$	$\varepsilon^{\dagger}_{\mathrm{O,O}}$ (kcal mol^{-1})	[301]	0.1617
	$f^{\ddagger}_{\varepsilon_{\mathrm{O,O}}}$	[301]	0.77
	$\varepsilon^{*}_{\mathrm{O,O,AMBER}}$ (kcal mol^{-1})	AMBER/GAFF	0.21
carboxylate O-primary amine N	$R^{\dagger}_{\mathrm{min,N,O}}$ (Å)	[301]	3.5549
	$f^{\ddagger}_{R_{\mathrm{min,N,O}}}$	[301]	1.02
	$R^{*}_{\mathrm{min,N,O,LB}}$ (Å)	AMBER/GAFF	3.4852
carboxylate O-primary amine H	$R^{\dagger}_{\mathrm{min,H\,O}}$ (Å)	[301]	2.3064
	$f^{\ddagger}_{R_{\mathrm{min,H,O}}}$	[301]	1.02
	$R^{*}_{\mathrm{min,H\,O,LB}}$ (Å)	AMBER/GAFF	2.2612
carboxylate O-potassium ion	$R^{\dagger}_{\mathrm{min,K^+O}}$ (Å)	[302]	3.6355
	$f^{\ddagger}_{R_{\mathrm{min,K^+O}}}$	[302]	1.08
	$R^{*}_{\mathrm{min,K^+O,LB}}$ (Å)	AMBER/GAFF	3.3662

† Optimized parameter value.

‡ Optimized scaling factor, defined in eq. 4.7 and eq. 4.8.

* From original AMBER/GAFF values, or combination rules (see eq. 4.6);

$^{\#}$ Applied to all interactions through LB rules.

In the simulations reported here, the carboxylate oxygen self-interaction LJ parameter, $\varepsilon_{O,O}$, was modified from its original value in the AMBER/GAFF force field,[98,311] as described in [301]. The modification was done through a scaling factor, f, applied to the value of $\varepsilon_{O,O}$ in the AMBER force field:

$$\varepsilon_{O,O} = f_{\varepsilon_{O,O}} \cdot \varepsilon_{O,O,AMBER} \tag{4.7}$$

The optimized $\varepsilon_{O,O}$ was used in the simulations, through combination rules (eq. 4.6), for all LJ interactions involving this atom type.
The LJ parameters obtained from combination rules, however, proved insufficient to describe the (ij) interaction between the optimized carboxylate oxygen atom type and some ions. For those cases, the force field was substantially improved by modifying $R_{\min,i,O}$ values from their LB value, while keeping the corresponding ε_{iO} at their LB values (already including the optimized $\varepsilon_{O,O}$). As described in [302], the modification was done through a scaling factor, f, applied to the LB value of $R_{\min,iO}$:

$$R_{\min,i,O} = f_{R_{\min,i,O}} \cdot R_{\min,i,O,LB} \tag{4.8}$$

Modifications to the LB values were done to the Lennard-Jones potentials governing the interactions between carboxylate (R–COO^-) and potassium ions, and between the carboxylate groups and the alkylamine group (R–NH_3^+) of the amino acid lysine.
Table 4.4 shows the original values, scaling factors, and modified values of all modified parameters used in the simulations reported here.

Simulation Details

The initial configurations were set up in the *tLeap* module of the AMBER software,[284] by inserting potassium and chloride ions up to the target concentration and then inserting extra ions (using the *addIons* tool of AMBER) until neutralization was achieved. Subsequently, *tLeap* was used to insert TIP3P water molecules;[298] a water layer of minimum thickness 20 Åbetween the surface of the protein and the edge of the simulation box was created. LJ and electrostatic interactions were cut-off at 12 Å. Beyond this distance, the particle mesh Ewald (PME) scheme with a grid spacing of 1.0 Å, and 4^{th} order interpolation[312] was used to calculate electrostatic interactions. We used long range dispersion corrections for both energy and pressure. Bonds with H-atoms were constrained using the SHAKE algorithm[306] during the NpT equilibration and the production simulations, with the exception of one water molecule that coordinates the zinc center of the carbonic anhydrase (for technical reasons, SHAKE cannot be applied to this molecule). We use a 2 fs time step for all simulations except for those of the carbonic anhydrase proteins, which use a 1 fs time step.

Each system went through 4 minimization cycles, each consisting of 2500 minimization steps using the steepest-descent algorithm and 7500 steps using the Conjugate Gradient algorithm. The protein atoms were restrained to their initial positions, with decreasing values of the restraint bond constants (500, 300, 100, and 50 kcal mol^{-1} Å^{-2}) for each cycle. Subsequently the restraints were lifted and each system was minimized for 10000 steps using the l-bfgs algorithm. The target system temperature was monotonically increased from 0 to 298 K during a 500 ps simulation in the canonical ensemble (NVT), during which the protein atoms were restrained to their positions (restraint bond constant of 10 kcal mol^{-1} Å^{-2}) and the temperature was controlled using the Langevin thermostat with a collision frequency of 1.0 ps^{-1}. Each system was equilibrated for 10 ns in 10 steps of 1 ns in the NpT ensemble, with the temperature controlled using the Berendsen barostat[112] with a relaxation time of 2.0 ps. Because the GPU version of the pmemd engine is very sensitive to changes in the box size, the *skinnb* parameter was increased to 5 Å. This change prevents unnecessary rebuilds of the pair list when there is no possibility of missing atoms in the pairwise calculation within the cut-off. The saved configuration from the last 1 ns of the NpT equilibration simulations that has the box density closest to the average was selected as the starting structure for each production simulation.

Obtaining the Median Protein Structure using Theseus

The Theseus[309,310] software package, which uses the maximum likelihood (ML) method instead of the common least-squares criteria, can be used to superimpose multiple macromolecular structures. When using the ML method, various regions of the superimposed structures are corrected for correlations among atoms. This method produces much more accurate results than conventional least-squares criteria.

Conventional methods discard residues that are aligned with gaps when macromolecules that are being superimposed have different lengths. In the ML method which is used in Theseus, all the residues are considered. Using a wrapper script "theseus_align", amino acid sequences are extracted from the PDB files, then they are aligned, and the superposition is performed using this alignment.

We have used Theseus to produce Figure 4.14. At first, for the production trajectories for each of the 10 proteins at $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ and $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$, we superimposed all the saved configurations (frames) to the fist frame using Theseus. The output of Theseus for this superposition of each protein at each concentration indicates which frame may be considered the *median structure*, i.e., that is the saved configuration that is overall most similar to the average structure; it can also can be considered the most "typical" structure in the ensemble of saved configurations. In the second step, we used the "theseus_align" script to align the median structure for each protein at $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ and $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$ to produce each panel in Figure 4.14.

Root Mean Square Fluctuations

We characterized the dependence of protein structural flexibility on salt concentration by calculating the root mean square fluctuation (RMSF) of non-hydrogen atoms in each protein at $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ and $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$. The RMSF of an atom i is a measure of the average displacement of that atom over time relative to its average position:

$$\mathrm{RMSF}_i = \sqrt{\frac{1}{t_{\max}} \sum_{t=0}^{t_{\max}} \left(x_i\left(t\right) - \bar{x}_i\right)^2} \tag{4.9}$$

$x_i\left(t\right)$ are the 3-dimensional coordinates of atom i in configuration t of $t_{\max}$ saved configurations; $\bar{x}_i$ are the coordinates of atom i averaged over all input frames. The RMSF over a group of N atoms, each with mass m_i, is calculated as a mass-weighted average:

$$\langle \mathrm{RMSF} \rangle_N = \frac{\sum_{i=1}^{N} \mathrm{RMSF}_i \cdot m_i}{\sum_{i=1}^{N} m_i} \tag{4.10}$$

The RMSF per residue shown in Figure 4.4 in the main text, the histogram of RMSF in Figure 4.16 and the RMSF of secondary structures in Figure 4.15 were calculated using the Cpptraj package[308] from AMBER software. For each protein, after aligning each frame of a trajectory against the first frame to remove the contribution of rotation and translation, the RMSF per residue was calculated by averaging over all non-hydrogen atoms of each residue using eq. 4.10 for all saved configurations of the production run. The result of this calculation is shown in Figure 4.4 in the main text. We have also shown the results of this calculation as a histogram in Figure 4.16, where the histogram is normalised, and distributed over 20 bins.

We also used the Cpptraj package[308] to classify the residues of each protein as a function of their secondary structure using the DSSP method[297] of Kabsch and Sander. This method assigns secondary structure types for residues based on backbone amide (N−H) and carbonyl (C=O) atom positions. We grouped in category “Alpha” all residues with 3_{10} helix, α helix, and π (3_{14}) helix configuration; in category “Beta” those with extended β (parallel β) and isolated β (anti-parallel β) configuration; in category “Coil” those in turn or bend configuration, as well as those with a configuration that could not be assigned to any of the other groups. For the residues in each category, we calculated the RMSF for non-hydrogen atoms as one group using equation 4.10. The result is shown in Figure 4.15.

Protein Surface Area

The hydrophobic SASA (Solvent Accessible Surface Area) shown in Figure 4.5 and Figure 4.17 is calculated using VMD through an own-program written in TCL/TK.[307] We have calculated the SASA for the hydrophobic residues with the three-letter code Phe, Met, Trp, Ile, Val, Leu, Pro, Ala, considering only the non-hydrogen atoms, for all saved configurations of the production runs of each protein at $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ and $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$. The SASA measurement starts from the distance of 2 Åfrom the surface of a protein, and to avoid the internal protein voids and pockets affecting our results, we used the "-restrict" option to only consider solvent-accessible points near our selected hydrophobic residues. These SASA results for each protein at each concentration are shown as normalised histograms in Figure 4.17 distributed over 20 bins. In Figure 4.5B the average value of SASA for each protein is shown with its associated standard error of the mean, and the corresponding values are also presented in Table 4.5. In Figure 4.5C, we show the width of the SASA histograms shown in Figure 4.17 at the height of 1/10 of its maximum value.

4.6.4 Experimental Results

Chemical-Shift Perturbations

Fig. 4.9 shows differences in amide ^{1}H and ^{15}N chemical shifts (chemical-shift perturbations, CSPs) between high-salt and low-salt conditions. In addition to changes in direct solute-solvent interactions (e.g., for surface-exposed amides), CSP analysis is used to monitor residue-resolved changes in both structure and dynamics. The highest CSPs ($> 0.10\,\mathrm{ppm}$) can be observed for different kinds of residues. Remarkably, most of the CSP above the threshold arise from lysine (20.8 %), indicating a structural effect of salt concentration especially for charged residues. On the other hand, high CSPs are also often observed for the uncharged residues glycine (16.7 %) and leucine (12.5 %). Regarding secondary structure, 54.2 % of these high CSPs are observed in coil regions, even though on average only 40.6 % of the secondary structure of the four studied proteins belongs to coil regions, indicating a stronger effect of KCl concentration on flexible regions of the protein.

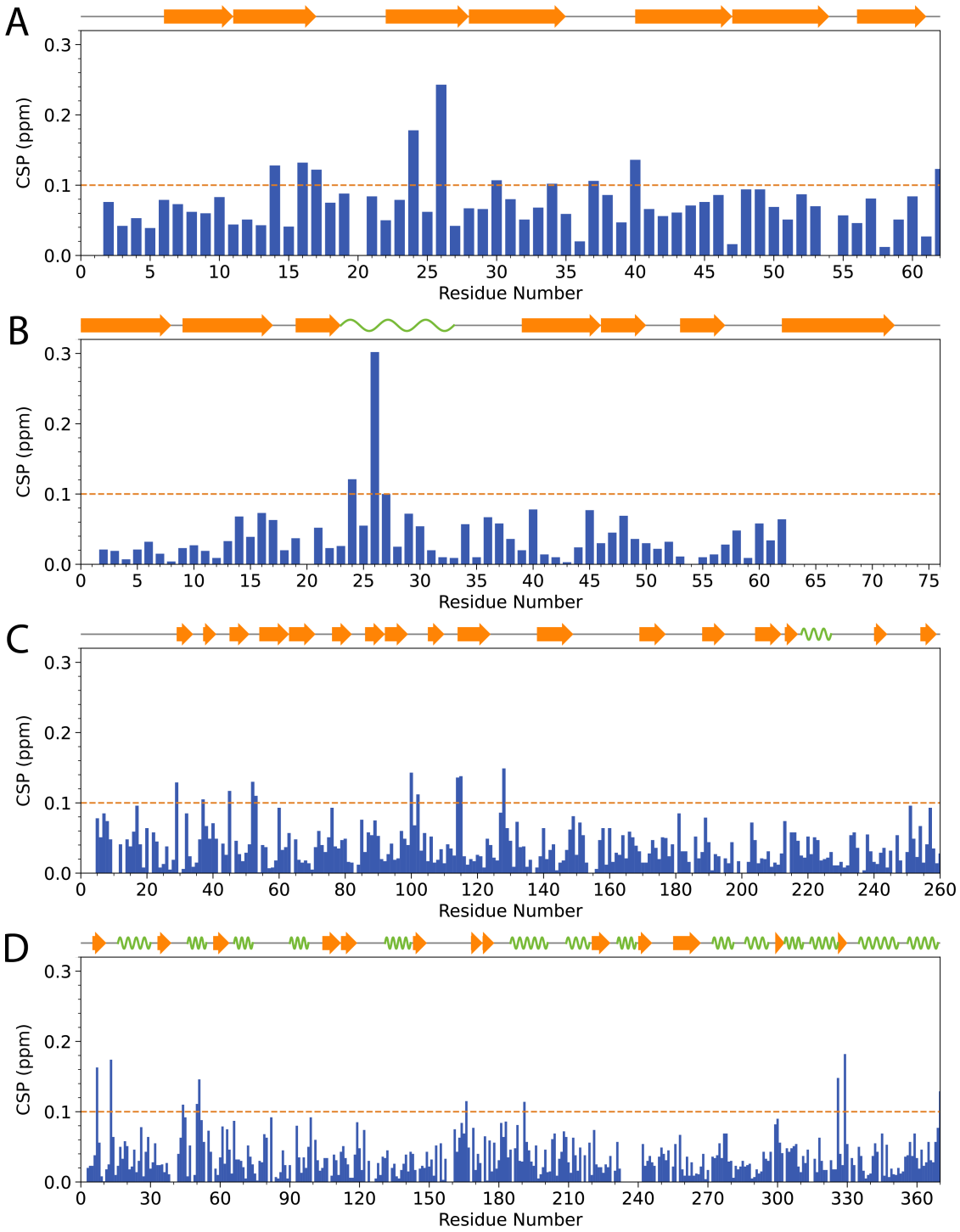

Figure 4.9: Chemical-shift perturbations (CSPs) between no-salt and high-salt conditions for the four proteins in focus, i.e., the SH3 domain of chicken α-spectrin (A), ubiquitin (B), hCAII (C), and MBP (D). The CSPs were determined using the equation $\sqrt{\Delta\delta_{\mathrm{H}}^2 + (0.14 \cdot \Delta\delta_{\mathrm{N}})^2}$.

Protein Dynamics as Measured by hetNOE and Exchange Contributions

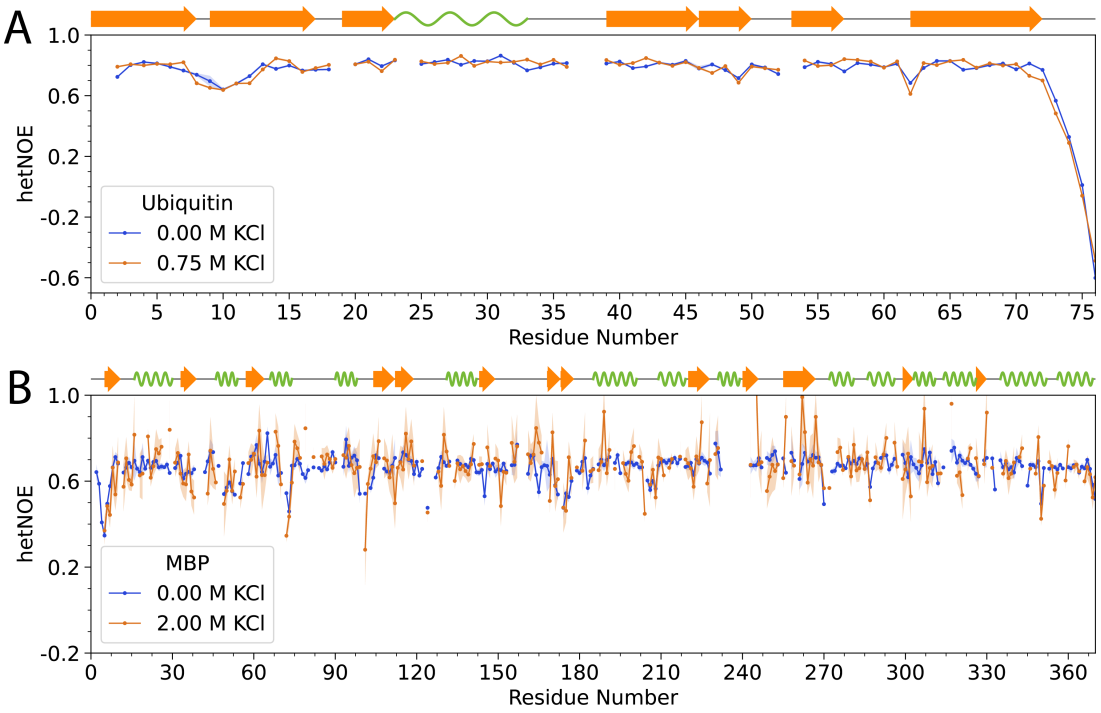

Figure 4.10: Fast (ps-ns) timescale motion with and without KCl, measured by [^{15}N, ^{1}H] hetNOE, defined as ratio I_{sat}/I_{eq}, where I_{sat} and I_{eq} are the peak intensities with and without proton saturation, respectively. The shaded areas around the curve represent the propagated error estimated from the spectral noise. Low hetNOE values indicate flexibility in the ps-ns timescale. For ubiquitin, no significant alterations for hetNOE and exchange rate values are observable under high-salt concentrations. For MBP, hetNOE values are much more dispersed in case of the 2.0 mol/l KCl concentration, which probably reflects the lower sensitivity of the experiment in the presence of salt.

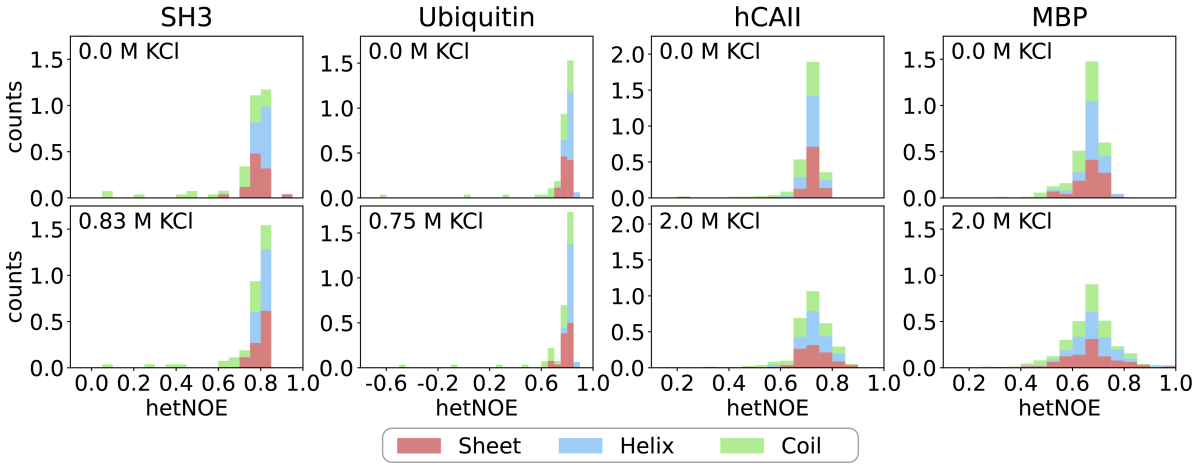

Figure 4.11: Histograms of [^{15}N, ^{1}H] hetNOE values of proteins SH3, ubiquitin, hCAII and MBP with and without KCl. [^{15}N, ^{1}H] hetNOE values are defined as ratio I_{sat}/I_{eq}, where I_{sat} and I_{eq} are the peak intensities with and without proton saturation, respectively. Low hetNOE values indicate flexibility in the pico- to nanosecond timescale.

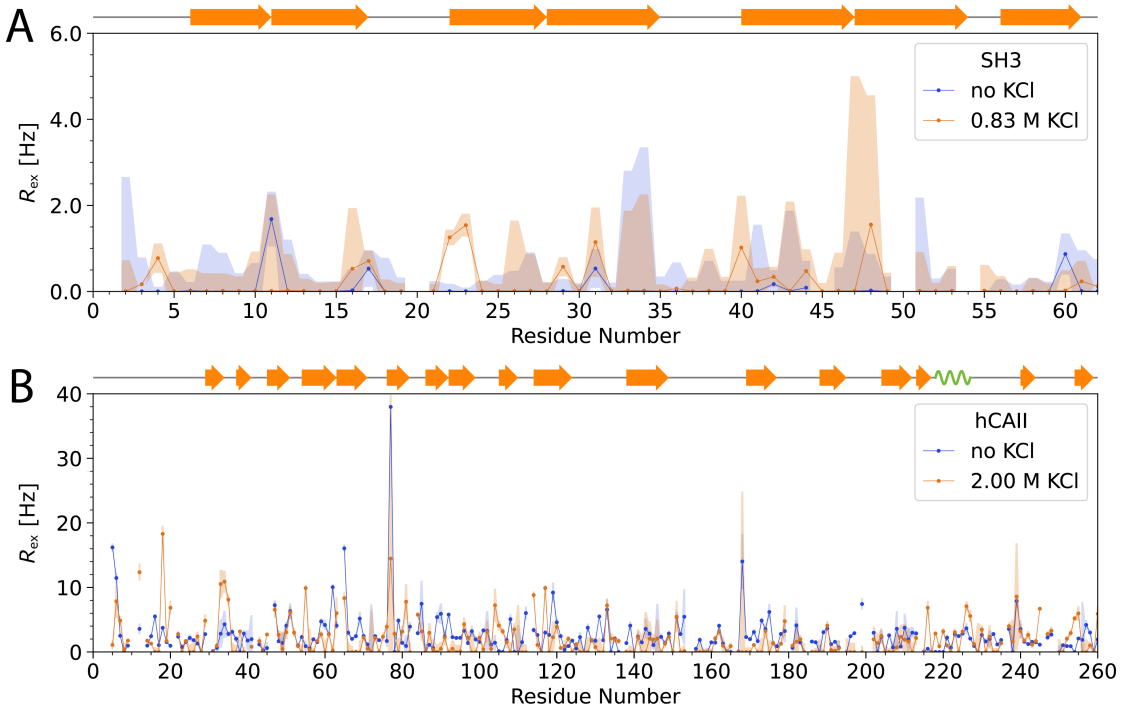

Figure 4.12: Exchange contributions R_{ex} extracted from ^{15}N-CPMG relaxation dispersion profiles of proteins SH3 and hCAII with and without KCl. High R_{ex} values indicate flexibility in the micro- to millisecond timescale. Shades denote the error of the data. For ubiquitin, all exchange rates R_{ex} are quite small, indicating no remarkable dynamics on the micro- to millisecond timescale. When comparing the behaviour with and without KCl, differences in R_{ex} values are all within error except for 22Glu and 23Val in the RT loop. Here contributions from intermediate-timescale dynamics are slightly increased under addition of KCl (from 0.008 Hz to 1.255 Hz for 22Glu and from 0.004 Hz to 1.540 Hz for 23Val). For MBP, trends in the exchange rates R_{ex} are less obvious due to many fluctuations and compromised tuning of the probe under high-salt conditions.

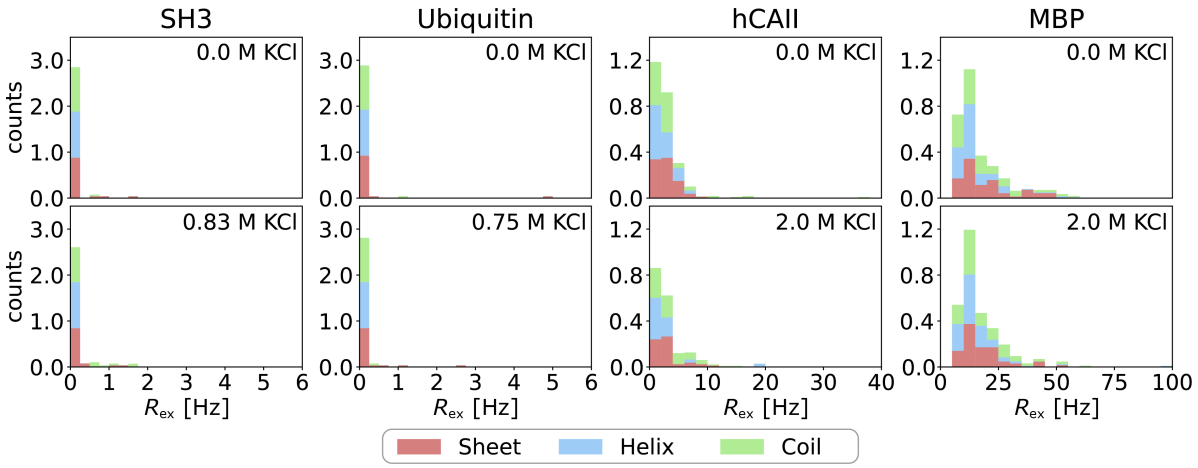

Figure 4.13: Histograms of exchange contributions R_{ex} of proteins SH3, ubiquitin, hCAII and MBP with and without KCl. Exchange contributions R_{ex} were obtained from CPMG measurements. High R_{ex} values indicate flexibility in the micro- to millisecond timescale.

4.6.5 Computational Results

Structural Stability

To assess how the structure of the proteins is affected by salt concentration, in Figure 4.14 we compare the median structure of each protein at high and low salt concentration during the production runs. The median structure is defined as the saved configuration that most resembles the average structure within a given time interval, and was identified using the software package Theseus.[309,310]

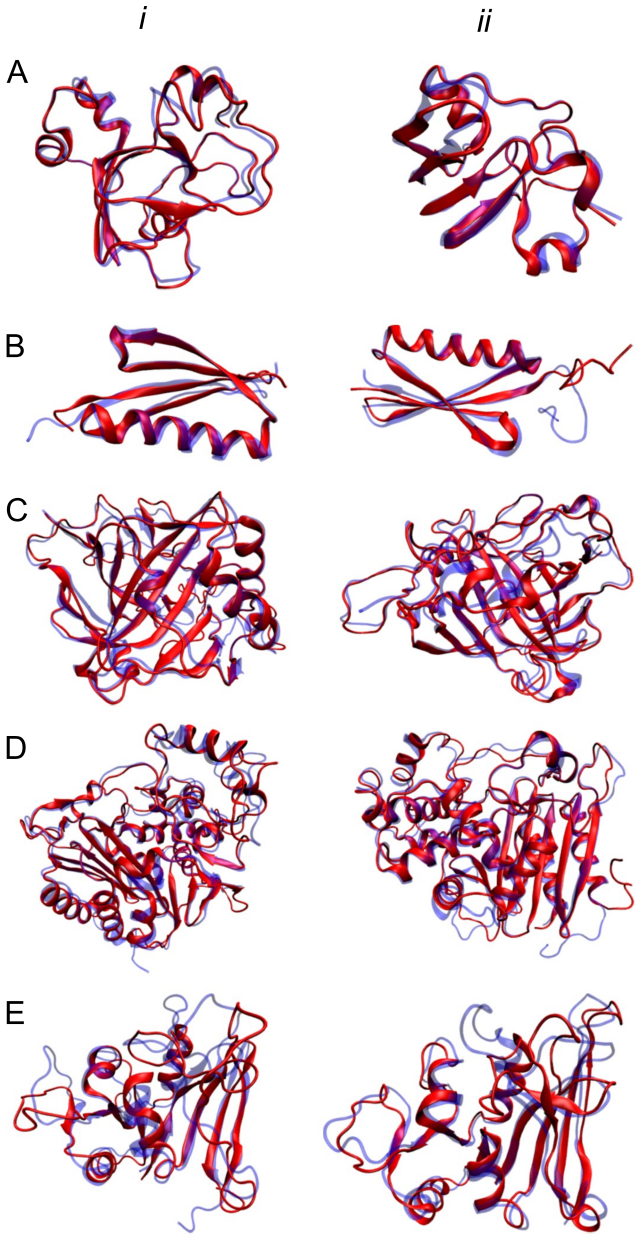

Figure 4.14: Superposition of the median structures calculated from the production run simulations of each protein at $b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$ (red) and $b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$ (blue). Halophilic proteins are shown in column *i*; their mesophilic counterparts are shown in column *ii*. (A) ferredoxin (PDB ID 1DOI, 1FRD); (B) protein L (PDB ID 2KAC, 1HZ6); (C) carbonic anhydrase (PDB ID 4CNX, 1V9E); (D) β-lactamase (PDB ID 3WRT, 1ZKJ); (E) dihydrofolate reductase (PDB ID 2ITH, 2L28).

The structures at the two salt concentrations substantially overlap, with differences occurring primarily in sections in coil conformation. For both forms of the carbonic anhydrase (PDB 4CNX and 1V9E)[313] and for the halophilic β-lactamase (PDB 3WRT)[314] studied here, there is experimental evidence that the proteins are active – and thus probably remain folded – outside of their optimal KCl concentration range. In contrast, the halophilic protein L (PDB 2KAC) is known to unfold at low salt concentration.[315] Because unfolding times of even small proteins are much beyond our simulated time scale, the proteins remain folded for the entire duration of the simulations. Our simulations thus enable the study of protein conformational flexibility either within a stable or a metastable state.

Dynamics

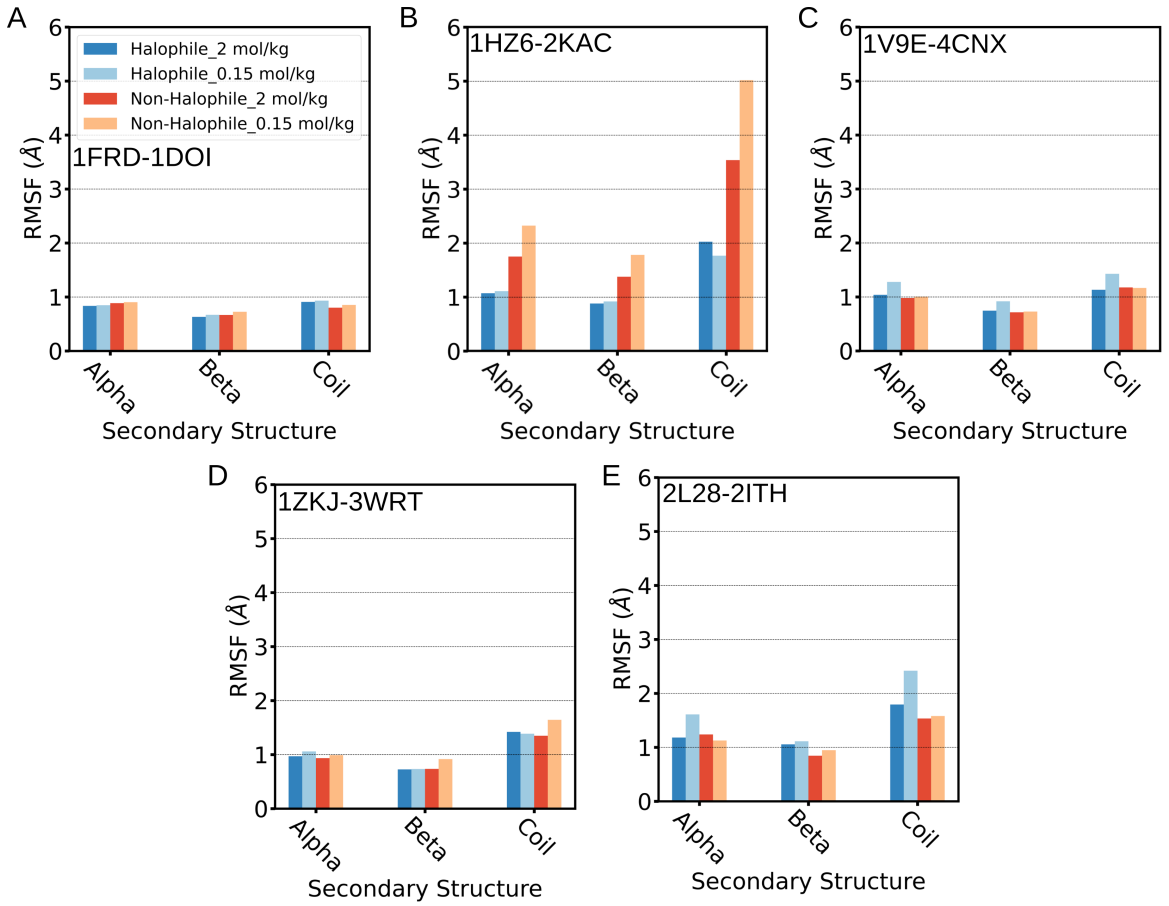

Figure 4.15: Average residue RMSF for residues in α-helix, β-sheet and coil configurations, for the indicated halophilic-mesophilic protein pairs at high and low KCl concentration, calculated from MD simulation. (A) ferredoxin; (B) protein L; (C) carbonic anhydrase; (D) β-lactamase; (E) dihydrofolate reductase. Compared to Figure 4.4 in the main text, the results show that salt concentration has a small but measurable impact on the dynamics of every protein, as measured by the residue RMSF. In agreement with the experiments, residues belonging to coil conformation experience the highest RMSF. Regions with higher RMSF have more pronounced salt-dependent changes.

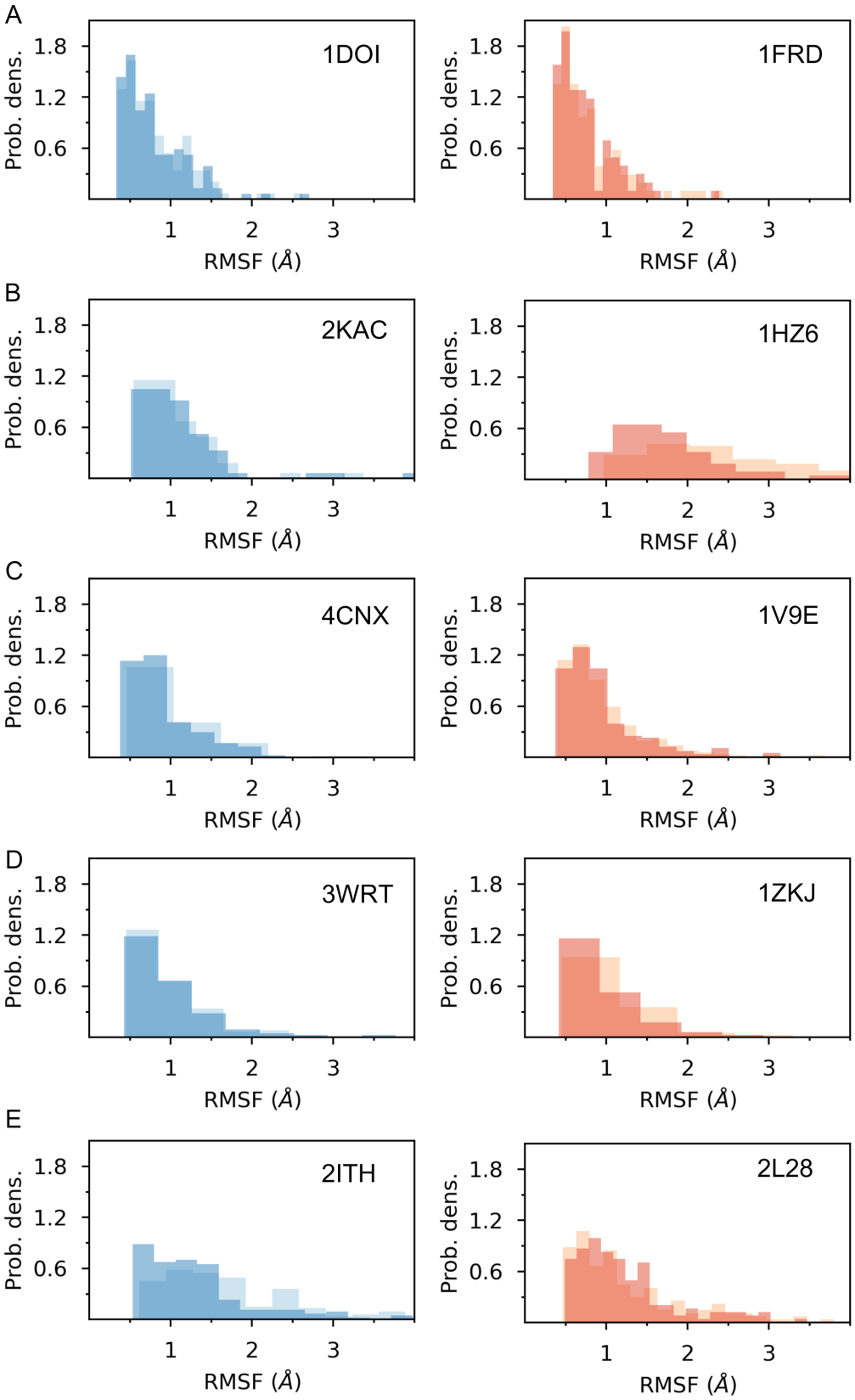

Figure 4.16: Histogram of the RMSF of each residue of the indicated halophilic-mesophilic protein pairs, at high and low KCl concentration calculated from MD simulations. Related to Figure 4.4. (A) ferredoxin; (B) protein L; (C) carbonic anhydrase; (D) β-lactamase; (E) dihydrofolate reductase.

Modelling the Impact of Salt Concentration on Fluctuations of the Hydrophobic Solvent-Accessible Surface Area

Consider a generic effective potential, $F_\Phi(\xi)$, that depends on collective variable $\Phi = \Phi(q)$ where q denotes the $3N_{\mathrm{at}}$-dimensional vector containing all the atomic positions at a given time. The probability density, $P_\Phi(\xi)$, that the collective variable Φ assumes value ξ is related to the effective potential as

$$F_\Phi(\xi) = -k_\mathrm{B}T \ln P_\Phi(\xi) + k_\mathrm{B}T \ln D \tag{4.11}$$

or equivalently, as

$$P_\Phi(\xi) = D \exp\left(-\frac{F_\Phi(\xi)}{k_\mathrm{B}T}\right) \tag{4.12}$$

where T is temperature, k_B is the Boltzmann constant and D is a constant normalising the probability density distribution.[316] D only shifts the potential $F_\Phi(\xi)$ by a constant, i.e., the absolute value of $F_\Phi(\xi)$ has no physical meaning; only differences between U values for different ξ are meaningful. In the remaining discussion, the collective variable will be identifiable by context, such that the subscript Φ identifying it will be dropped.

Table 4.5: Average value of hydrophobic SASA (with standard error of the mean) associated with the heavy atoms of hydrophobic residues (Phe, Met, Trp, Ile, Val, Leu, Pro, Ala) for the indicated proteins.

Protein	$\mathbf{SASA_{hydrophobic}}$ **(Å)**	
	$b_{\mathrm{KCl}} = 2.00\,\mathrm{mol\,kg^{-1}}$	$b_{\mathrm{KCl}} = 0.15\,\mathrm{mol\,kg^{-1}}$
1DOI	1108.1 ± 0.6	1147.0 ± 0.6
2KAC	883.2 ± 0.6	894.4 ± 0.6
4CNX	2858.6 ± 1.7	2918.3 ± 1.1
3WRT	4137.9 ± 1.1	4228.4 ± 1.2
2ITH	1928.4 ± 0.8	1882.7 ± 1.1
1FRD	1100.4 ± 0.5	1113.7 ± 0.5
1HZ6	1010.7 ± 0.8	1051.2 ± 1.0
1V9E	2867.7 ± 0.8	2927.9 ± 0.8
1ZKJ	3897.0 ± 1.1	3971.4 ± 1.2
2L28	2524.1 ± 0.9	2468.8 ± 1.1

The probability density that a protein has a given hydrophobic SASA, A, at a reference salt concentration is governed by an effective potential, $F_{\mathrm{eff,ref}}(A)$, that includes the contribution $\gamma_{\mathrm{ref}}\,A$ (equation 4.1 of the main text). Replacing the $\gamma_{\mathrm{ref}}A$ contribution by that at a different salt concentration, c, and assuming that all other terms of the effective

potential are independent of salt concentration leads to an estimate of the effective interaction potential, $F_{\mathrm{eff},c}(A)$, governing hydrophobic SASA at salt concentration c:

$$F_{\mathrm{eff},c}(A) = F_{\mathrm{eff,ref}}(A) - \gamma_{\mathrm{ref}} A + \gamma_c A = F_{\mathrm{eff,ref}}(A) + \Delta\gamma\, A \tag{4.13}$$

where $\Delta\gamma = \gamma_c - \gamma_{\mathrm{ref}}$ is the difference in surface tension between the two salt concentrations. From equations 4.12 and 4.13 it follows that the probability density, $P_c(A)$, of observing a given SASA, A, at salt concentration c can be determined as

$$P_c(A) = D_1 \exp\left(-\frac{F_{\mathrm{eff,ref}}(A) + \Delta\gamma\, A}{k_{\mathrm{B}}T}\right) = D_2\, P_{\mathrm{ref}}(A)\, \exp\left(-\frac{\Delta\gamma\, A}{k_{\mathrm{B}}T}\right) \tag{4.14}$$

(i.e., equation 4.2 in the main text) from the probability density, $P_{\mathrm{ref}}(A)$, at the reference salt concentration. The constants D_1 and D_2 are introduced to normalise the probability density distribution; D_2 is different from D_1 because $P_{\mathrm{ref}}(A)$ is assumed to include its own normalization constant.

Modelling the Impact of Salt Concentration on the Distribution of Inter-Charge Distances

The electrostatic interaction, $U_{\mathrm{electr}}(r)$, between two point charges in an electrolyte solution can be estimated using Debye-Hückel theory:[317]

$$U_{\mathrm{electr}}(r) = \frac{1}{4\pi\epsilon_0\epsilon_{\mathrm{r}}} \frac{q_1\, q_2}{r} \exp(-\kappa r) \tag{4.15}$$

with the Debye length κ^{-1}, the vacuum permittivity ϵ_0, the relative permittivity ϵ_{r} and the electric charges $q_{1/2}$ at distance r. Both κ^{-1} and ϵ_{r} depend on the salt concentration. The effective potential, $F_{\mathrm{eff,ref}}$, between two charges of a protein at a reference salt concentration includes the electrostatic component, $U_{\mathrm{electr,ref}}(r)$. Replacing that component by that at a different salt concentration c and assuming that all other contributions to the effective potential are independent of salt concentration leads to an estimate of the effective interaction potential between the two charges at concentration c:

$$F_{\mathrm{eff},c}(r) = F_{\mathrm{eff,ref}}(r) + U_{\mathrm{electr},c(r)} - U_{\mathrm{electr,ref}}(r) \tag{4.16}$$

From equations 4.12 and 4.16, the probability density, $P_c(r)$, of observing charges at distance r at concentration c can be obtained from the probability density, $P_{\mathrm{ref}}(r)$, at the reference concentration as

$$\begin{aligned} P_c(r) &= D_3 \exp\left[-\frac{F_{\mathrm{eff,ref}}(r) + U_{\mathrm{electr},c}(r) - U_{\mathrm{electr,ref}}(r)}{k_{\mathrm{B}}T}\right] \\ &= D_4 P_{\mathrm{ref}}(r) \exp\left[-\frac{U_{\mathrm{electr},c}(r) - U_{\mathrm{electr,ref}}(r)}{k_{\mathrm{B}}T}\right] \end{aligned} \tag{4.17}$$

The constants D_3 and D_4 are introduced to normalise the probability density distribution; D_4 is different from D_3 because $P_{\mathrm{ref}}(r)$ is also normalised.

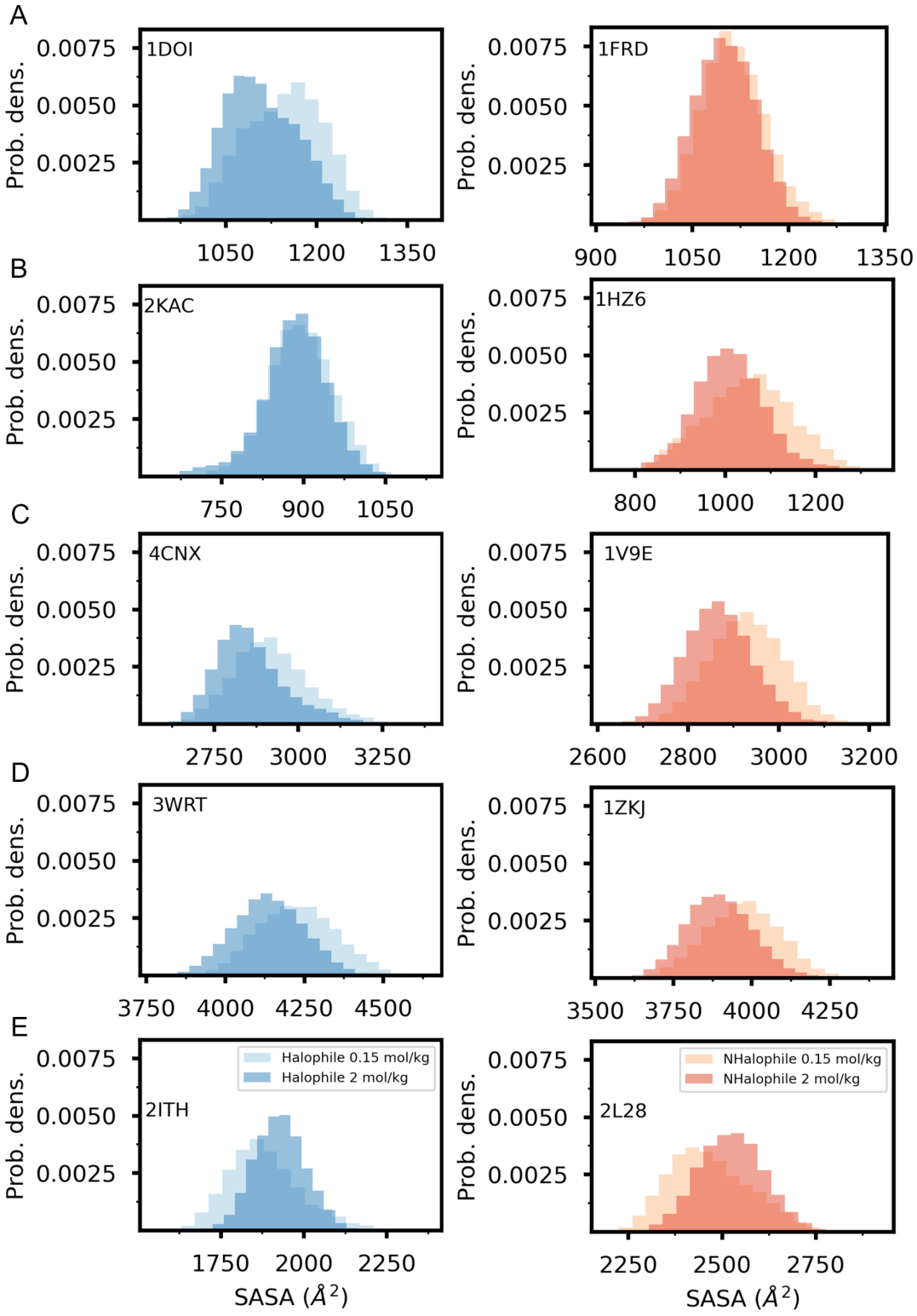

Figure 4.17: Hydrophobic SASA of the indicated proteins, at high and low KCl concentration, calculated from simulations. Related to Figure 4.5 in the main text. (A) ferredoxin; (B) protein L; (C) carbonic anhydrase; (D) β-lactamase; (E) dihydrofolate reductase.

5

Activation of a Secondary-Messenger Receptor via Allosteric Modulation of a Dynamic Conformational Ensemble

The work presented in this chapter was published as B. Söldner[†], H. Singh[†], E. Akoury[†], G. Witte, R. Linser, *Angew. Chem. Int. Ed.* **2025**, *64*, e202509394 and B. Söldner[†], H. Singh[†], E. Akoury[†], G. Witte, R. Linser, *Angew. Chem.* **2025**, *64*, e202509394.
[†]: These authors contributed equally to this work.

Contributions to this work:

- Together with E. Akoury and H. Singh, I assigned the protein backbone chemical shifts and determined chemical-shift perturbations (CSPs). The published, final assignments and CSPs presented in this chapter were determined by me using all available backbone experiments together.
- Together with E. Akoury and H. Singh, I analysed the NMR relaxation experiments. The published, final analysis of the NMR relaxation experiments presented in this chapter was performed by me.
- Molecular dynamics simulations were solely performed and analysed by me.

5.1 Abstract

Bacterial signalling cascades have recently become of great relevance in the context of bacterial antibiotics resistance. Cyclic diadenylate monophosphate (c-di-AMP) is a key bacterial secondary messenger involved in growth, biofilm formation, virulence gene expression and others. The activation mechanisms of c-di-AMP receptors like the trimeric P_{II}-like proteins upon messenger binding have, however, remained elusive due the pivotal role of highly flexible protein regions. Here, using solution NMR spectroscopy to elucidate the interplay between the ordered and disordered structural elements of the apo and messenger-bound forms of the 44 kDa homotrimeric P_{II}-like signal transduction protein A (PstA), we reveal a sensitive modulation of the conformational ensemble of those extended loops thought to bind the downstream interaction partners by messenger association at the receptor core. The orchestration of the spatial properties of the loops, despite their retained internal dynamics, reveals the importance of allosteric effects even for disordered structural elements, whose steerable ensemble properties have long escaped the classical structural-biology understanding.

5.2 Introduction

The ubiquitous role of intrinsically disordered proteins/regions (IDPs/IDRs) for protein function in almost any context of cellular organization and activity has widely been recognized in the last decade, owing to a soaring number of experimental studies and simulations.[151,157,320,321] Their diverse functional aspects, long elusive due to the lack of any defined conformations, can be rationalized today by an ensemble description,[322–324] which quantitatively rationalizes their selectivities and overall functionalities derived from partially preformed transient structures, folding upon binding, or mixtures of both.[325–327] Despite the rapid interchange of conformations, the properties of these ensembles are well-defined, context-specifically tailored, and evolutionarily fine-tuned. While retaining a high degree of plasticity, these still define important biophysical features, e.g., condensation tendencies or pore permeabilities via distinct steric and electrostatic effects.[328–330]

Loop structures within folded domains, beyond their obvious role for interconnecting the classical secondary structure elements, would be expected to constitute equally important elements for tailored affinity, cooperativity, and propagation of information across signalling cascades. However, how events within the flexible loops can possibly impact functional features of the protein core or, vice versa, the possible dependency of loop properties on events in the core, has largely escaped the current structural-biology understanding and requires careful analysis of interdependencies in the dynamic ensemble.

Here we interrogate this cross-coupling in a bacterial secondary messenger receptor, where messenger association happens to the core, but binding to downstream interaction partners hinges on long, disordered loops, whose structural properties have escaped structural

biology and, on the first glance, seem fully unrelated to messenger binding in the core of the trimeric complex.[331] Secondary messengers are intracellular messengers that act in response to an extracellular signal, which are generated at cell surface receptors to initiate downstream processes within the cell while prohibiting the primary messenger from overcoming the cell membrane.[332] Bacterial secondary-messenger systems, many of which are non-existent in humans, have seen tremendous scientific interest in recent years in particular due to the rise of bacterial drug resistance, a soaring problem with a large social impact, as they provide a means to selectively combat procaryotes. Cyclic diadenylate monophosphate (c-di-AMP, see Fig. 5.1A), catalysed by diadenylyl cyclase (DAC) enzymes through the condensation reaction of two ATP molecules,[333] is an important secondary messenger widely found in prokaryotes, regulating sensing of DNA integrity, cell wall metabolism, and potassium transport.[333–335] Despite intense efforts in identifying the mechanisms of cyclic dinucleotides with their protein binding partners for therapeutic intervention, however, the activation of the respective receptors by c-di-AMP secondary messenger binding are poorly understood.

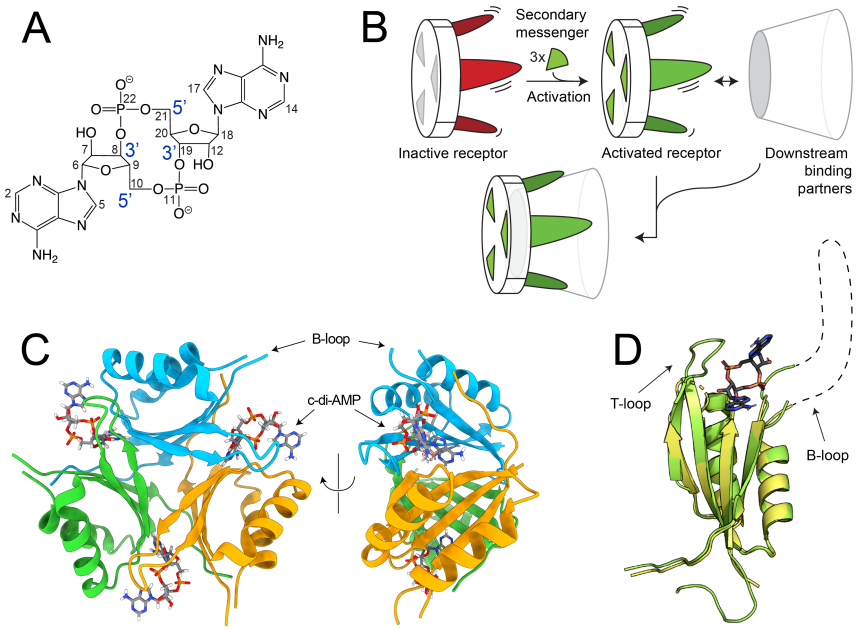

Figure 5.1: PstA structure and secondary-messenger binding. A) Cyclic di-AMP (c-di-AMP), a bacterial secondary messenger. B) Current model of activation of P_{II}-like secondary messenger receptors by c-di-AMP binding. A binding-incompetent trimeric receptor becomes activated by threefold association of c-di-AMP, allowing its long, disordered loops to coordinate C_3-symmetric downstream interaction partners.[336] C) Trimeric arrangement and structural elements of PstA represented by the X-ray coordinates of the holo complex (PDB 4WK1). D) Overlay of crystal structures of apo-PstA (PDB 4WK3) and $(\text{c-di-AMP})_3$:$(\text{PstA})_3$ complex (PDB 4WK1), depicting a single monomer.

Commonly found among Gram-positive firmicutes and actinobacteria,[334] c-di-AMP binds several receptors including the transcription factor BusR, the cytoplasmic cation trans-

porters KtrA and CpaA, and the histidine kinase KdpD.[337,338] The family of signal transduction proteins is among the most widespread in bacteria, archaea, and plants. Upon contact with the vital effector molecules ATP/AMP and 2-oxoglutarate, it signals on nitrogen availability and interacts with other enzymes, transcription factors, and membrane transport proteins.[339–341] Trimeric proteins are known to bind trimeric target proteins like amphotericin B (AmtB), N-acetylglucosamine kinase (NAGK), PipX protein, and dinitrogenase reductase-activating glycohydrolase (DRAG).[342] They are highly conserved homotrimers (of around 40 kDa overall molecular weight) exhibiting a double βαβ ferredoxin-like fold (compare Fig. 5.6).[343,344] Each monomer comprises a compact core of two α-helices and four β strands prominently connected by two loops that accommodate intermolecular ligand and ATP-binding sites.[345] With an arrangement of three long protruding loops (called either B-loop or T-loop), one from each of the monomers, the trimer is able to capture downstream interaction partners (compare Fig. 5.7, PDB 2NUU[336]), thereby activating the associated signalling cascades (scheme in Fig. 5.1B). The activation mechanism of these receptors has, however, remained elusive as the disordered loops have largely been invisible in crystal structures. In all those cases in which a reasonably defined electron density for part of the loops was obtained (compare Fig. 5.6), their observed structural features likely originate from the intermolecular crystal contacts and must be treated with care.

P_{II}-like signal transduction protein A (PstA, Fig. 5.1C) from *Staphylococcus aureus* (named DarA in *B. subtilis*) is one of the various c-di-AMP receptors of this fold.[331,337] Ligand incorporation in P_{II} proteins generally is thought to be cooperative, meaning that association of the second and third ligand happens with altered affinity to be able to sample a wider range of ligand concentrations.[346,347] Messenger binding in PstA happens with a dissociation constant K_D of around 20 nM (see ITC data in Fig. 5.9 and compare Müller et al.[331]). The mode of activation by c-di-AMP has remained elusive hitherto.[331,348,349] Upon ligand binding, the short T-loop now produces a well-defined electron density, but the core retains a virtually identical structure with or without the ligand (Fig. 5.1D). The longer B-loops on the other hand, invisible in both, apo and bound forms, remain fully disordered, posing the puzzle how messenger binding would induce their binding competence. Whereas the role of disordered regions often escapes X-ray crystallography or cryo electron microscopy, NMR spectroscopy in solution[215,350–355] or in the solid state[224,356–358] can assess both, well-structured regions as well as those that bear no or little (transient) structural properties, facilitating insights into the well-tempered interplay of structure and dynamics for biological functionality. Here we use NMR spectroscopy in solution, supported by simulations, of the dynamic ensemble to unveil an intimate coupling between the structural features of the disordered loops and those of the core as a function of secondary-messenger binding.

5.3 Results and Discussion

5.3.1 Residual Secondary-Structural Propensity of PstA Loops

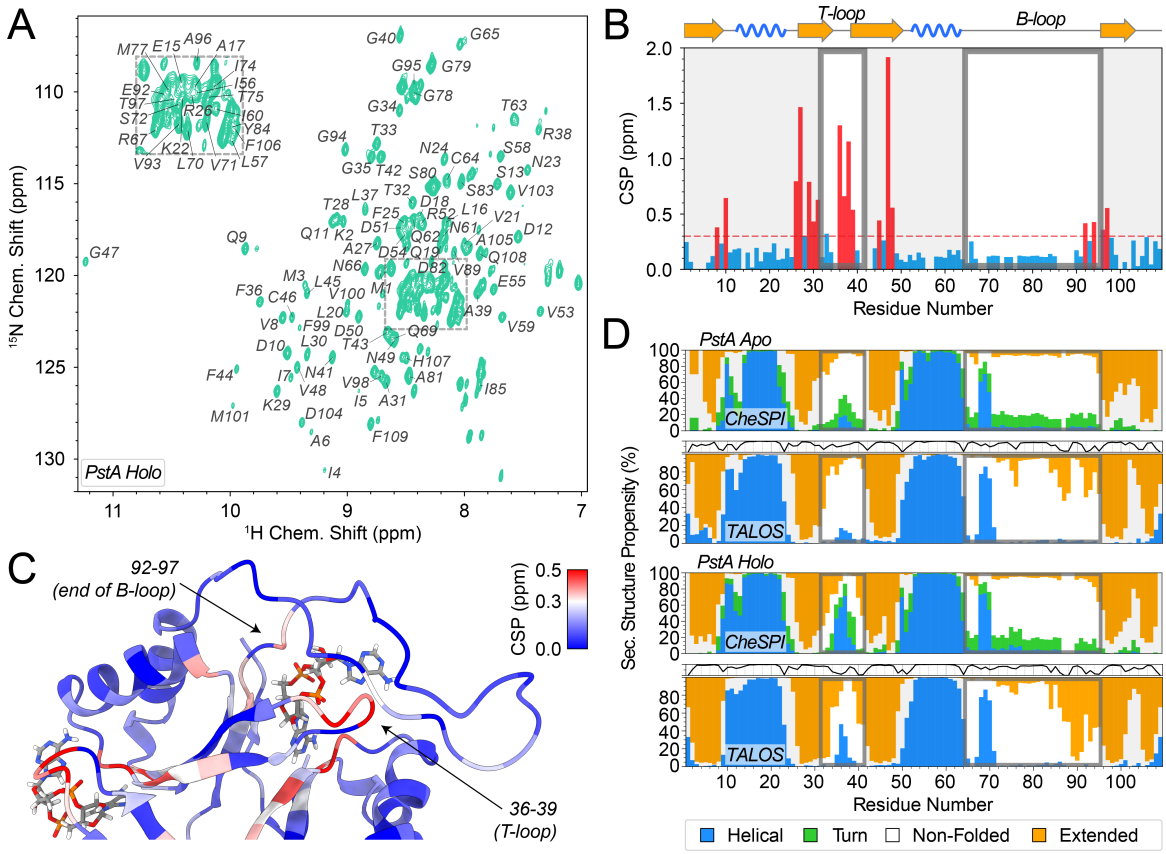

Figure 5.2: Chemical-shift perturbations (CSPs) between apo and c-di-AMP-bound PstA and NMR-based secondary structural features. A) 2D [^{1}H, ^{15}N]-HSQC NMR spectrum of perdeuterated, ^{13}C/^{15}N-labelled c-di-AMP-PstA complex, recorded at 800 MHz Larmor frequency at 25 °C, together with residue-specific H/N assignments. B) CSPs as a function of residue and C) plotted on the structure. The red line in B) was arbitrarily set to 0.3 ppm. The color code for C) is depicted in the top right. CSPs based on amide proton and amide nitrogen chemical shifts as $\sqrt{\Delta\delta_{\mathrm{H}}^2 + (0.14 \cdot \Delta\delta_{\mathrm{N}})^2}$. D) Secondary-structural propensities of apo protein (top) and complex (bottom), assessed either via CheSPI or via TALOS-N, the loops in focus being highlighted by gray boxes. The TALOS-N profiles also denote the certainty with which the residue-specific prediction is associated on top the respective plot, with a confidence scale ranging from 0 (bottom) to 100 (top).

To assess structural and dynamical features of the PstA receptor and their modulation by secondary-messenger binding experimentally, we subjected the trimeric complex in the presence and absence of the ligand c-di-AMP to solution NMR spectroscopy. Sample preparation was done as described for crystallography previously, however, using either doubly (^{15}N/^{13}C) or triple-labelled (^{2}H/^{15}N/^{13}C) minimal media. (See details on sample preparation in the Materials and Methods). Resonance assignment was achieved using

a set of standard triple-resonance backbone experiments. (See details in the Materials and Methods.) In total 504/437 shifts (CO, C^{α}, C^{β}, N^{H}, and H^{N}) from 106/98 amino acids (95.8 %/83.1 % of all CO, C^{α}, C^{β}, N^{H}, and H^{N} shifts; 96.2 %/85.6 % of all backbone amide shifts) could be unambiguously assigned for ligand-bound and ligand-free PstA, respectively. (Fig. 5.2A shows a ^{15}N-HSQC spectrum of the receptor in c-di-AMP-bound form. All resonance assignments were deposited in the BMRB under accession codes 52882 and 52883.) Upon binding of the secondary messenger, large chemical-shift differences (CSPs) are observed in the T-loop and in the end of β3, i.e., sites that are in direct contact with the ligand and particularly through interactions involving G47, F36, L37 and R26. (See an overlay of the ^{15}N-HSQC spectra, residues with CSPs > 1.0 ppm being highlighted in Fig. 5.11, as well as chemical-shift perturbations plotted on the monomeric protein structure and as a function of residue in Figs. 5.2B and C, respectively. Fig. 5.12 shows a 3D HSQC-NOESY strip highlighting G47 to ligand cross peaks.) Interestingly, in addition, residues 92 to 97, the C-terminal residues of the B-loop, show sizeable perturbations after addition of the ligand.

Secondary-structural features of the trimer core, assessed using either CheSPI[6] or TALOS-N[5,359] (see Fig. 5.2D), exactly match the observations from previous crystallographic studies. The T-loop, compared to the apo form, attains a substantial shift from non-folded towards structured behaviour. Interestingly, however, apart from a slight increase of extended at the very end of the B-loop, which is in relative proximity to the ligand, and a slightly more helical tendency for residues 75 – 77, the very beginning of the B-loop, the highly dynamic properties of the B-loops seem to be largely maintained upon binding of the messenger. Only the residual, transient structural propensities in these regions seem to be influenced by messenger binding, while their high degree of disorder is very clear throughout residues 73 – 96. A rather steady helical stretch, irrespective of messenger binding, is seen for residues 68 – 70, even though these residues are disordered in crystallography.

5.3.2 PstA Dynamics from Relaxation Experiments

To assess, in more quantitative detail, the dynamics of individual amino acids in the complexes, we characterized motion on different timescales, the intermediate (µs-ms) and fast-timescale (ps-ns) dynamics, both represented in Fig. 5.3. To characterize the site-specific µs-ms conformational-exchange dynamics of the PstA trimer, we engaged ^{15}N constant-time CPMG relaxation dispersion experiments (see Experimental Details in the SI). Interestingly, for the apo trimer, strong relaxation dispersion (with an exchange contribution of up to 38 s^{-1}), denoting slow µs-timescale motion on the 100 µs timescale (see Table 5.2), is found for the majority of residues in β2, one residue in the beginning of the T-loop, and in one β3-residue (see Figs. 5.3A-C). Most of these residues span the trimeric interface, which is formed predominantly by the residues shortly before the

T-loop (residues 27 – 34). Exchange contributions R_{ex} from individual fits are depicted as a function of sequence in Fig. 5.3B and color-encoded on the structure in Fig. 5.3C. This presence of relaxation dispersion demonstrates a conformational exchange in the core of the apo form. By contrast, this exchange is fully abolished upon ligand binding, where no dispersion is witnessed over the entire protein (Figs. 5.3A, B, and C). Given the large size of the trimer, the data do not allow an accurate determination of the timescale, but the presence/absence of strong conformational exchange is unambiguous. Given the colocalization of residues with an exchange contribution and the intermolecular interactions, the observations likely point to a temporary loosening of the trimer interface, which might serve a facilitated uptake of the bulky ligand. In the bound form, by contrast, the intermolecular interactions are significantly strengthened through multiple additional indirect intermolecular contacts mediated by the ligand (Fig. 5.13, hence rigidifying not only the T-loop, expected from crystallography and described below, but also abrogating conformational exchange of the trimer interface. Interestingly, the CPMG experiments do not show a significant extent of dispersion for the residues of the B-loop. A contribution to the R_{ex} rates from conformational exchange here (which are different between the apo and holo forms, see below) must hence derive from intermediate-timescale processes faster than accessible via relaxation dispersion methods, i.e., beyond the $\approx 50\,\mu s$ timescale.
In addition, fast-timescale (ps-ns) dynamics for apo and ligand-bound PstA were obtained using ^{15}N R_1, R_2, and heteronuclear NOE experiments (Fig. 5.3D, top to bottom, respectively). In the apo form, the T-loop is characterized by hetNOE values much lower (down to ≈ 0.5) and R_1 values much higher (up to $4\,s^{-1}$) than the prototypical core residues locally, suggesting high flexibility on the ps-ns timescale. By contrast, in the presence of c-di-AMP, the low hetNOE and high R_1 values of the apo form are completely flattened out, all parameters now matching those of the other core residues. This suggests a complete rigidification of the T-loop in the bound form due to the interactions with the ligand. Ligand binding, involving a total of 8 H-bonds to the receptor (compare Fig. 5.13) hence locks the T-loop into a single, well-defined position. In the apo form, just before and after the T-loop close to the ligand binding site, very high R_2 values up to $\approx 50\,s^{-1}$ are found, which is in congruency with the relaxation dispersion data above. These rates exceed the values of around $35\,s^{-1}$ found throughout the remainder of core residues (which are in congruency with what would be expected for a perdeuterated $\approx 40\,kDa$ globular protein at 30 °C), confirming the presence of the intermediate-timescale conformational exchange process described above.
In coherence with their disordered character in crystallographic studies, most of the B-loop residues (including amino acids 69 – 93) show the signatures of fast-time scale mobility. In contrast to the well-defined density for the end and beginning of the loop in part of the X-ray structures, the NMR data suggest steadily increasing disorder towards the center of

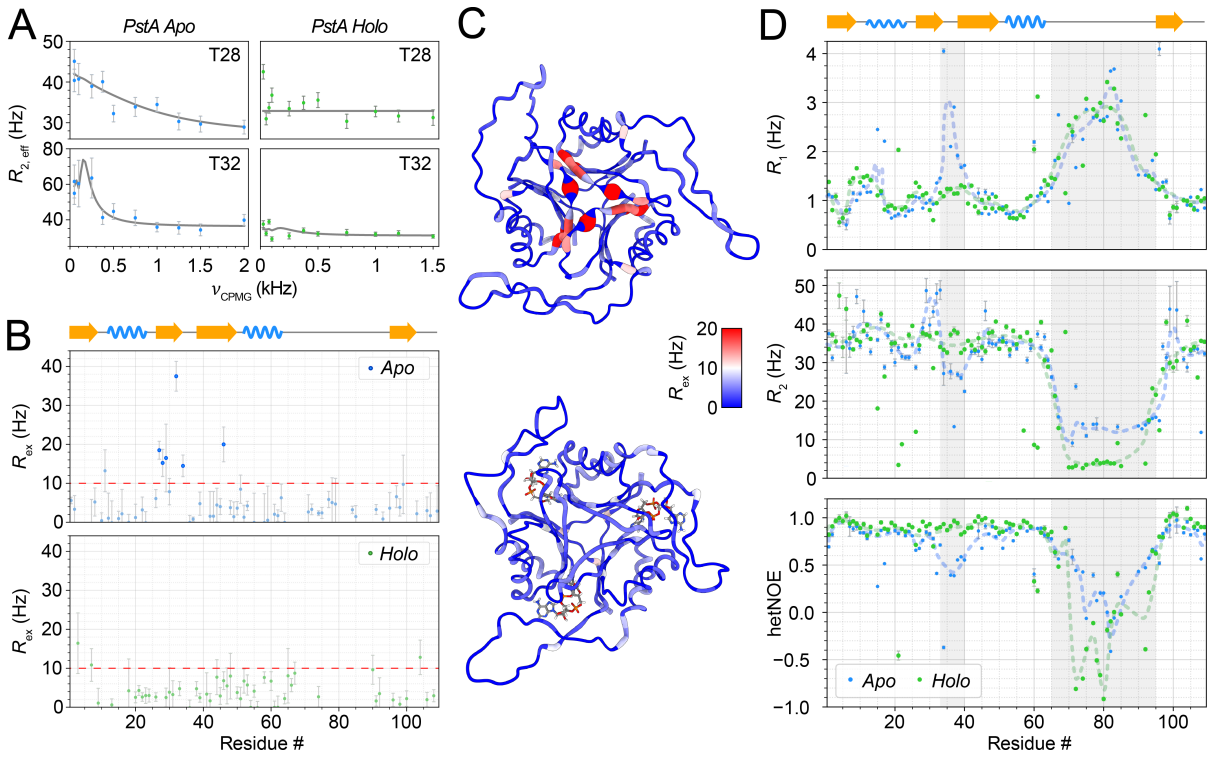

Figure 5.3: Conformational dynamics in apo and messenger-bound PstA. A) Representative relaxation dispersion profiles for apo PstA (left), compared to those of the same residues of the complex (right). B) Exchange contributions R_{ex} as a function of residue, comparing complex (bottom) and apo protein (top). The points clearly above the red line reflect strong dispersion as opposed to noise-like fluctuations. C) R_{ex} contributions plotted on the structure, comparing apo protein (top) and complex (bottom). D) Longitudinal relaxation rates R_1, transverse relaxation rates R_2, and heteronuclear steady-state NOE for apo protein (blue) and complex (green). Gray shades denote residues in the T-loop (residues 33 to 40) and B-loop (residues 65 to 95). Dotted trend lines disregard local outliers found for individual residues such as in short turns.

the loop in both, the apo and holo form. Residues at the very center of the B-loop have similarly high R_1 values for both forms, and also hetNOE values confirm the high degree of local motion for both forms, with slightly less extreme tip dynamics for apo than for holo PstA. Maybe most interestingly, differences are observable with respect to the R_2 rates of the loop center: Whereas the holo form yields values of around $2 - 4\,s^{-1}$, such rates in the B-loop tip (but not for the rest of the protein) are $\approx 4\times$ higher (by around $10\,s^{-1}$) in the apo form. Given the hetNOE and R_1 data, fast-timescale dynamics with very low effective correlation times still clearly apply for both forms. This can be rationalized intuitively by dissecting the two individual underlying contributions (timescales) to R_2, where the higher R_2 rates in the apo loops likely point to differential chemical-exchange (R_{ex}) contributions. Without any surface interactions, apo and holo loops would bear identical contributions from fast, internal ps-ns timescale reorientation of the individual H-N bond vectors. However, an extra contribution from (µs timescale) chemical exchange (varying fast-timescale averages) due to short-lived but repeated interactions with the

structurally variable ligand binding site, modulating the chemical shifts on a slower timescale and hence adding to dephasing, will depend on the presence/absence of the c-di-AMP ligand just below it, while retaining the fast-motional profiles regarding R_1 and hetNOE. By contrast, such conformational-exchange contributions seem different (lower) for the holo loop tip, suggesting that the ligand somehow drives the center of the loop into a more homogeneous situation with fewer surface contacts, dominated by the fast internal motion. Even though these are slight differences between apo and holo trimers, the data are compatible neither with defined loop conformations nor with loops that, irrespective of messenger binding, move fully independently. Instead, the NMR experiments hint to alterations of transient interactions between dynamic loops and the receptor core upon messenger binding, hence modulating the spatial properties of the loop and redefining interaction sites available in the ensemble for downstream signalling.

5.3.3 Simulation of PstA Dynamics

In order to complement the differences in loop behaviour witnessed experimentally in the conformational ensemble in the holo form compared to the apo receptor with more mechanistic details, we pursued simulations both, for the apo as well as for the ligand-bound protein. (See the SI for building of ligand topology and parametrization as well as further details on the setup of these simulations.) Gaussian-accelerated molecular dynamics (GaMD) simulations with a high boost potential (See details in the SI.) were used to obtain six complex structures of three monomers each that sufficiently cover the available conformational space. These starting structures (in total 18 loop conformations under otherwise identical conditions) hence obtained were used to start conventional MD simulations (yielding in total 18 µs of simulation time with respect to the conformational space sampled by the loops). This way, the bias associated with a single starting structure, the danger of being trapped in local energy wells, as well as the potential pitfalls associated with reconstructing motional properties from accelerated simulations were circumvented. Fig. 5.14 shows the root mean square deviation (RMSD) with respect to the respective starting structure as a function of time. As during the first 100 ns most RMSDs reach a plateau, all further analysis was performed based on the remaining 900 ns, also for reducing even further a potential bias by the starting structure.

As expected, the fast overall motion of the B-loop observed experimentally is well-reflected in both forms. However, extended structure is observed as temporary events that form sporadically in the beginning and the end of the B-loop in holo form. (See residue-specific secondary-structural propensities as a function of time in Fig. 5.15.) Interestingly, the Cartesian space covered by overall loop motion is significantly higher in the apo form compared to the holo trimer. Fig. 5.4A shows this deviation (root-mean-square fluctuations, RMSFs) in Cartesian space as a function of residue, where higher values for both,

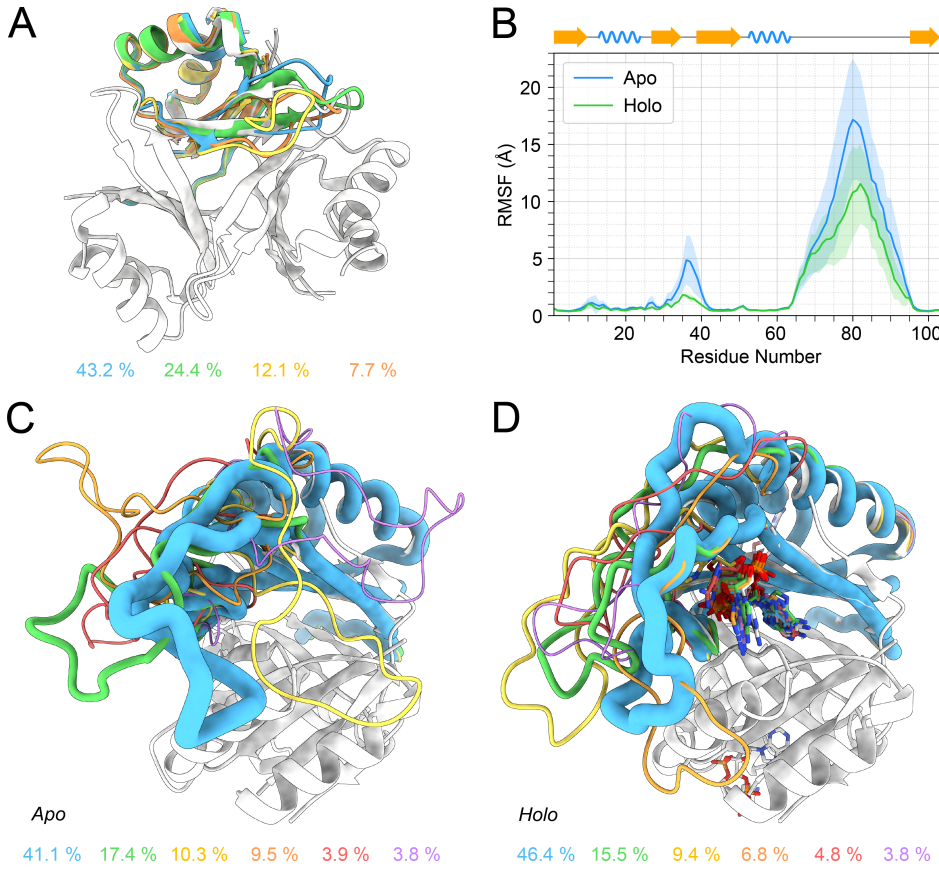

Figure 5.4: Conformational space sampled by the PstA core and B-loop as obtained by simulations. A) Root mean square fluctuations with respect to the average structure of both trajectories as a function of residue, comparing the apo complex (blue) with the messenger-bound form (green). B) and C) Cluster analysis of the conformational ensemble of the apo (B) and holo receptor (C), using RMSD cut-offs of 5.0 Å and 3.5 Å, respectively. For cluster analysis, a concatenated trajectory of all three single monomers was used. D) Conformational distribution of the apo form core and T-loop, assessed by clustering with an RMSD cut-off of 1.2 Å. Whereas the behaviour of the tip of the T-loop matches the fast-timescale motion seen experimentally, structural changes of the interface, as expected on the 100 µs timescale from dispersion measurements, are not captured within the simulations. A peptide flip is observed for residue 27 – 28, compare Figs. S12 and S13. (The conformations of the holo form are captured within a single such cluster, which is hence not shown.) For B), C), and D), the relative populations of each cluster are denoted in descending order below.

the T- and the B-loop, are obtained in the apo form. To elucidate the nature of these conformational dynamics in more detail, the trajectories were clustered based on backbone atoms (N, C^{α}, C^{β}, C^{O}, and O) with the GROMOS algorithm described by Daura et al.[360] The cut-off values for clustering the conformational space of the B-loop using concatenated monomer trajectories was set such that 85 % of the conformational space could be captured within six clusters, i.e., a 5.0 Å cut-off for the more divergent apo form and a 3.5 Å cut-off for the less divergent holo form. Fig. 5.16 demonstrates that transitions between the clusters frequently occur. Figs. 5.4B and C visualize the clusters obtained for apo and holo

form. Whereas the holo form is dominated by a common directionality of the loop, the apo form samples a wide range of directions and surface interactions. The high degree of *local* disorder (i.e., bond vector reorientations) towards the middle of the loop is, however, retained for both forms, as expected from the experimental R_1, R_2 and hetNOE values. For analysing the conformational space sampled by the core and the small T-loop, the coordinates of apo and holo trimers *without* the B-loop were clustered using a cut-off of 1.3 Å. For the apo form, for this cut-off, 85 % of the frames were captured within four clusters (depicted in Fig. 5.4D), whereas for the holo form, a single cluster already covers the entire conformational space sampled. The motions within the monomer-monomer interface of the core, a conformational exchange experimentally observed on a $> 100\,\mu s$ timescale, is not seen in the simulations in spite of the high boost potential applied in the preceding GaMD simulation. By contrast, the behaviour of the tip of the T-loop in the apo form matches the fast-timescale motion seen experimentally.

To elucidate in more detail in which way loop dynamics are impacted by the presence of the ligand, we interrogated the *in silico* ensembles in various ways. Most effectively, a plane was constructed from which the distance of any loop residues could be measured. This plane is perpendicular to the C_3 symmetry axis of the trimer and located in the center of mass of the core, such that *large positive* distances measured for the tips of all three loops, according to the current model (Fig. 5.1B), represent a *binding-competent conformation*, while low or negative distances for one or more loops denote binding-incompetent conformations (Fig. 5.5A). Fig. 5.5B shows the distributions obtained for the apo and holo receptor with respect to the loop position (using residue A81 as a reporter, averaged over the three monomers). The width of the distribution in holo form is significantly narrower and a strong, 7 Å shift of the average distance of this distribution towards the putative interaction side is witnessed. The presence of the ligand hence achieves an effective distributional shift of the space in which the loops prevail. Even though only subtle, a small decrease of R_g is associated with the loops' directional bias induced by the messenger, which makes the complex slightly more compact (Fig. 5.19). This observation is in qualitative consistence with SAXS data, which confirm the subtle decrease of the diameter in the presence of c-di-AMP also from an experimental standpoint (Fig. 5.20, deposited into the SASBDB under accession codes SASDXD5 and SASDXE5 for apo and holo PstA, respectively).

On a first glance, it is hard to imagine that a subtle repositioning could have an impact on binding competency. However, all three loops are thought to be involved upon binding, leading to a C_3-symmetric complex. Conversely, according to the rules of conformational preselection, an encounter complex will translate into a binding event with a higher probability in case of a preformed binding-competent conformation, which kinetic advantage may be particularly important for the low concentrations of the receptor in the cell. To assess

the relative probabilities for a proper interface to form (i.e., *all* loops positioned towards the interaction partner) in the apo vs. holo trimer, we interrogated the ensemble by applying a series of different axial threshold values assumed for sufficient forward protrusion. Indeed, the subtle distributional effect for the average B-loop position as a single event (Fig. 5.5B) translates into a severalfold probability difference when taking the trimeric nature of the interface into account, which hence exerts a strong amplification of the modulating effect of the ligand (Fig. 5.5C, Fig. 5.21). Assuming $\approx 16\,\text{Å}$ as a representative cut-off for minimum forward protrusion, the chance of forming a binding-competent interface (three loops pointing forward) is ≈ 5 times higher than in the apo form. A conformational preselection of the ensemble by the trimeric architecture and its avidity would hence have a substantial kinetic effect, while pre-orientation of the large loops would further facilitate complex formation via reduced entropic costs of binding, enhancing also the thermodynamic feasibility of binding in addition to the kinetic advantage.

To better rationalize the grounds of the redirectional effects upon messenger association in particular for the disordered B-loop, we performed contact analyses within the dynamic ensemble (compare Figs. 5.5D/E, 5.22, and Table 5.3). Indeed, various recurring contacts are observed that, as a whole, embody the shift of the conformational distribution observed in experiments and simulation. Among the most prominent transient contacts are the intermolecular hydrogen bonding contacts between the end of the B-loop (residue E92 and the conserved motif G94-G95) and hydroxyls and phosphate of the ligand, which are naturally absent in the apo form. These hydrogen bonds are relatively persistent, with a percental occurrence between 26.3 % and 49.4 % of the time (compare Table 5.3). Interestingly, upon ligand binding, G94 and G95 at the end of the B-loop now also form more persistent *intramolecular* hydrogen bonds with Q9 at the beginning of the α1-helix (18.1 % vs. 64.0 % of the time for apo and holo form, respectively, for the contact 95GlyO-9GlnN, but also hydrogen bonds to the sidechain of Q9 are now observed). In addition to this "latch", a rather persistent salt bridge between the conserved residues R67 in the beginning of the B-loop and D12 in the core is found in apo and holo form (56.7 % vs. 67.1 % of the time, respectively). This contact tends to act as a *hinge* about which the stem of the loop can rotate in apo, whereas in the holo form, with the "latch" closed, this reorientation is largely restricted. Importantly, all of the residues mentioned above are highly conserved (compare Fig. 5.23 and see below), confirming their dedicated role in restraining the direction of the B-loop. Furthermore, inter-loop contacts (mostly hydrophobic contacts, but also patterns of transient hydrogen bonds) are found in the holo form, e.g., interactions between G79 and F36 and between M77 and L37). Lastly, upon stabilisation of the T-loop by ligand binding, manifold transient intermolecular contacts between the T-loop, the ligand, and the end of the helix α1 are traceable. These include both, hydrophobic contacts as well as the formation of various hydrogen bonds.

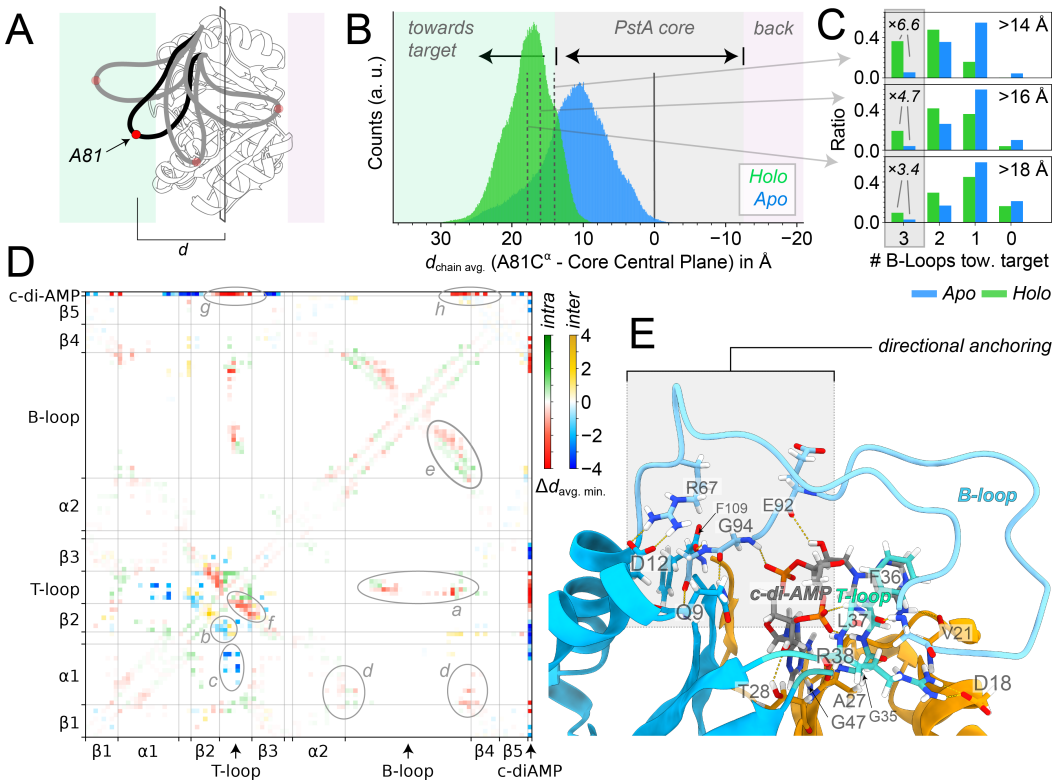

Figure 5.5: Mechanistic assessment of loop pre-direction. A) Quantification of loop extrusion via distance measurement between residue 81 at the outer tip of the B-loop and the central plane through the center of mass of the core and perpendicular to the C_3 axis. B) Distribution of loop protrusion, depicted as average over all three monomers at any given time, measured as depicted in A). C) Distribution of *collective* forward positioning (3, 2, 1, or none of the loops being beyond a given threshold) as a function of the threshold chosen (on the upper right, also compare Fig. 5.21). The numbers on the bottom left denote the ratio of occurrence of three forward-positioned loops between holo and apo receptor. D) Differences in intra- and intermolecular interactions between apo and holo PstA, measured as the difference in minimum distance between residues. (Red and blue show decreased distances for intra- and intermonomer contacts and hence closer proximity, compare Fig. 5.22 for the individual minimal distances.) Ligand association entails an increase for contacts between T-loop and B-loop (*a*), between T-loop and β2 (intermolecular, *b*), between T-loop and α1 (intermolecular contact, *c*), between the ends of the B-loop and the beginning of α1 (*d*), a shift of the intra-B-loop contact distribution (*e*) as well as stronger intra-T-loop contacts (*f*), in addition to direct interactions between loops and ligand (*g*, *h*). E) Important H-bonds (>10 % relative occurrence in the MD simulations), depicted on a representative conformation of holo PstA, including a first monomer (blue, with the B- and T-loop represented in light blue and cyan, respectively), the c-di-AMP ligand (gray), and a second monomer (orange). The network of recurring contacts at the end and beginning of the B-loop is highlighted with a gray box.

The overall network of these transient inter- and intramolecular contacts around the ligand binding site and near to the B-loop, created or at least strongly reshaped by ligand binding, are consistent with the ensemble being shifted into a spatially more occluded situation whenever the messenger is present. (See Fig. 5.24 for a simplified sketch of

these interactions and the effect they entail.) Interestingly, part of the temporary H-bond network observed as transient features in the holo ensemble also match the ones formed permanently in the crystal structure of holo *L. monocytogenes* PstA (Fig. 5.25, also compare again Fig. 5.6). Generally, the simulation results are in full agreement with the experimental data for the beginning and end of the B-loop, its moderate chemical-shift perturbations, and a slight increase in helical/extended propensities despite its retained high degree of internal dynamics.

While the identical apo and messenger-bound structures of various P_{II}-like proteins from crystallography have consistently made it difficult to associate the event of messenger binding with any sort of biophysical impact, our results – representing a slightly qualitative approach but fraternizing a multitude of complementary viewpoints – reveal a series of interesting differences for both, the core and the loops. On one hand, the above data reveal a so-far elusive plasticity of the *core* of the apo trimer. It was unforeseen in the original crystallographic studies that the bulky c-di-AMP ligand with its extended structure was found to be buried inside a complex with a structure identical to the apo protein. The spontaneous conformational exchange witnessed of for the inter-monomer ties now likely facilitates a partial deformation that will allow quick substrate uptake, while ligand uptake would face strong energetic hurdles in case of a fully rigid geometry. These dynamics, which in the apo form lead to the described dispersion effects on the 100 µs timescale, are quenched upon ligand association as multiple added H-bonds form, hence further stabilizing the trimeric structure of the holo form.

Even more interestingly, the above analyses also show the characteristics of the disordered and previously "invisible" *loops* to be sensitively modulated by ligand association, which would imply kinetic and thermodynamic differences in associating to downstream interaction partners and hence rationalizes the differences of the complex with respect to an initiation of downstream signalling cascades. Given their seemingly random motional characteristics, the biophysical features of loops in general, other than connecting other structural elements, have largely escaped understanding from a structural-biology viewpoint and hence not received much interest. The above data show that even with this high degree of local flexibility, overall loop behaviour – and hence any functional properties that the loops contribute to – can still be orchestrated from the distance by binding events in the protein core. The B-loop conformational ensemble is inherently influenced by ligand association, which induces a switch from a rather random loop behaviour, stochastically sampling a heterogeneous distribution of transient intramolecular interactions around the ligand binding site, to an effectively directed behaviour, where the "stem" of the loop (the C- and N-terminal ends) is rather fixated by fast recurring interactions to the newly shaped and chemically very different ligand binding site, and the tip of the loop instead becomes less involved.

The allosteric modulation is noteworthy as the individual amino acids in the loop invariably stay highly flexible with or without the messenger being bound and the regulatory effect solely occurs via a network of concatenated, transient interactions to a redesigned pocket interaction surface. Interestingly, the loop redirection model is also in full congruency with the cooperative character of messenger association assumed for P_{II} proteins earlier, which seems to be an enabling feature for activation of such trimeric receptors. As the properties of the conformational ensemble are a direct consequence of the presence or absence of the messenger, the facilitation of fully activated (triple-liganded) receptors, favoured even at low ligand concentrations, directly translates into predominant formation of a fully binding-competent interface with respect to the three loops.

To verify our interpretations about the directional role of the loop stem residues, we interrogated the extent of residue type conservation as a function of primary structure over different PstA homologues, as observed in multiple-sequence-alignment analysis (Fig. 5.23). Indeed, we obtain perfect agreement with the above model. A highly conserved amino acid motif is identified between PstA residues T28 and T43, with the consensus sequence ^{28}TKLXXXGGFLXXGNTT^{43}. This represents the conserved binding interface to the secondary messenger. More interestingly, amino acids K2, D12, R67, and the amino acid stretch ^{94}GGA^{96} toward the C-terminal end of the protein (compare the contacts described above) are highly conserved and occur in more than 95 % of the P_{II}-like protein sequences. While, importantly, these do not bind the ligand, their conservation is congruent with the above architecture of transient loop:core contacts that modulate the B-loop directional properties as a function of ligand binding. The conserved interactions between the conserved amino acids Q92 to G94 and the ligand, aided by a network of local hydrogen bonding contacts, are able to constrain into which direction the B-loop will most likely protrude from the core and at the same time drive the tip of the loop far away from the core. Whereas in the *presence* of the ligand, the (temporary) formation of these H-bonds acts like a latch, directing the beginning and end towards one predominant side, the simulations nicely show that in the *absence* of the ligand, the stem of can flip over and samples a heterogeneous pattern of random orientations. Residues 69 to 89, by contrast, show high sequence variation, hence giving further support to the kinetically driven model based on conformational preselection and avidity of multiple weak interactions. The intricacies of PstA have been prohibiting a more in-depth biological assessment.[331] Systems that are more tractable both from a structural-biology as well as a cell-biological perspective, like *E. coli* GlnK and its target, the ammonia channel AmtB (see Fig. 5.7), may hence be useful to overcome the limitations of this study with respect to assays and mutation studies to verify the here assumed connectivity between the biophysical observations and the actual biological impact.[336,361]

The allosteric modulation of spatial residence for overall flexible structural elements via concatenated transient interactions, amplified by a multitude of weak individual effects, adds to the emerging picture of a fine-tuned interplay of disorder and order in conformational ensembles throughout the cell. Entirely (intrinsically) disordered proteins (IDPs) or long disordered regions (IDRs) terminal to well-folded globular domains with a breadth of differentially stable and stabilisable structural properties, in particular secondary structural features, have seen soaring biophysical attention. The case of flexible loops within folded domains, with their ability to orchestrate interactions with interaction partners via allosteric modulation by events in the core via sequential short-lived interactions hence appears as an important aspect that has not been specifically addressed in detail. It comes to no surprise, and has been well-known, that short loops, as the T-loop of this study, which practically folds directly *onto* the ligand here, can drastically change their motional properties (i.e., rigidify) upon ligand binding. By contrast, we have demonstrated here how the spatial properties of the dynamic ensembles formed by much longer loops and hence their interaction profiles and overall properties can be sensitively tweaked by alteration/creation of new sets of transient interactions and translation of the effect through coupled structural interdependencies and transient intramolecular interaction networks within a dynamic ensemble. This ensemble modulation represents another possibility – adding to the growing range of allosteric mechanisms being explored hitherto – to tailor functionality via spatially distant binding sites. Even though the effect of quaternary interactions, providing interaction surfaces that influence structural propensities of other proteins, is well known, the spatial redistribution of a conformation ensemble with retained dynamic properties via subtly redirecting, transient interactions will be a relevant and likely general addition to our understanding of how ensembles can be modulated.

5.4 Conclusion

Here we have shown how association of a secondary messenger to its receptor, for which discernable differences upon activation have been elusive due to the conformational plasticity of the relevant interaction sites, can change its biophysical properties through a network of transient interactions. Apart from quenching of µs timescale conformation exchange dynamics in the core, the association of the ligand is found to impact the spatial properties of the dynamic conformational ensemble formed by the 30 amino acids long loop thought to orchestrate the binding to downstream interaction partners. The subtle, indirect impact by which the properties of the dynamic ensemble are altered upon ligand binding to the protein core via a network of transient intra- and intermolecular interactions acts as an allosteric modulation of basic biophysical features and may be of interest for understanding the functional roles of disordered structural elements and their behaviour as a function of upstream regulatory events generally.

5.5 Materials and Methods

5.5.1 Recombinant Protein Expression and Purification

A pET28a expression vector was used to clone *Staphylococcus aureus* PstA (SACOL0525) via NdeI and NotI restriction sites. Unlabeled, $^{1}H/^{15}N$-, $^{1}H/^{15}N/^{13}C$-, and $^{2}H/^{15}N/^{13}C$-uniformly labelled PstA constructs in conjunction with an N-terminal His_6-tag first without, later including a thrombin cleavage site (to avoid overlapping peaks in the central region of the H/N plane) were all expressed in *Escherichia coli* Rosetta BL21(DE3) cells and purified by Ni-NTA affinity and size exclusion chromatography (GE Healthcare) as the following: Cells were grown in protonated or deuterated M9 minimal medium containing ^{15}N-NH_4Cl and ^{13}C-glucose as sole sources of nitrogen and carbon, respectively. After reaching OD_{600} of 0.6, protein expression was induced by 1 mM IPTG and the cell culture was grown for 18 hours at 18 °C. Cells were harvested and re-suspended in lysis buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 mM Imidazole, 10 % v/v glycerol) and flash frozen. Cells were then thawed, homogenized and cell lysates were centrifuged at 17000 g for 30 min at 4 °C. The supernatant was incubated with His-Select Nickel affinity gel resin Ni-NTA (Qiagen) for 30 min at 4 °C. After washing with buffer B (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 40 mM Imidazole, 10 % v/v glycerol), buffer C (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 300 mM Imidazole, 5 % v/v glycerol) was used to elute the protein from the resin. After an overnight dialysis with buffer D (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 % v/v glycerol), anion exchange chromatography followed. PstA was loaded onto a HiTrap Q HP column (GE Healthcare) and eluted with a linear gradient to 100 % of buffer E (50 mM Tris-HCl pH 7.5, 1 M NaCl, 5 % v/v glycerol). Samples without thrombin cleavage site were finally concentrated and loaded onto a HiLoad Superdex 75 column (GE Healthcare) equilibrated with buffer F (50 mM phosphate buffer, pH 6.8, 50 mM NaCl, 1 mM DTT) for further purification by size exclusion chromatography. For the construct with a thrombin cleavage site, the protein was incubated with thrombin at room temperature for 10 h for His-tag removal. Protein solutions with the cleaved His-tag were then concentrated and loaded onto a HiLoad Superdex 75 column (GE Healthcare) equilibrated with buffer F (50 mM phosphate buffer, pH 6.4, 50 mM NaCl, 1 mM DTT) for further purification by size exclusion chromatography. The resulting protein fractions were concentrated and analysed by 15 % SDS-PAGE.

5.5.2 NMR Backbone Assignment

All NMR experiments were performed at 30 °C on a Bruker Avance III 800 MHz spectrometer equipped with a cryogenic probe. A set of $^{1}H/^{15}N/^{13}C$ and $^{2}H/^{15}N/^{13}C$ PstA protein samples (0.5 mM in 50 mM phosphate buffer pH 6.8 or 6.4, respectively, 150 mM NaCl, 1 mM DTT and 10 % v/v D_2O) were used to measure three-dimensional (3D) HNCA, HNCACB, HNCO, HN(CA)CO, CBCA(CO)NH TROSY-based triple-resonance experiments for the NMR backbone assignments of free and c-di-AMP-bound forms (see details in Table 5.1). A 3D ^{15}N-NOESY-HSQC experiment was recorded with 2048 (F1) × 48 (F2) × 256 (F3) complex points, 16 scans per increment with spectral widths of 16025 Hz, 4055 Hz, and 16025 Hz in the ^{1}H, ^{15}N, ^{1}H dimensions, respectively, and a total experiment time of 2.5 days.

Table 5.1: Experiments recorded for assignments. The index His denotes samples of a first iteration with a His_6-tag at pH 6.8. The abbreviation SW stands for spectral width.

Sample	Spectrum	Dimension			Points			Scans	SW (Hz)		
		F1	F2	F3	F1	F2	F3		F1	F2	F3
Apo	^{15}N-HSQC	H	N		4096	256		16	12500	2839	
Apo^{His}	^{15}N-TROSY	H	N		2048	512		16	14423	2839	
$Holo^{His}$	^{15}N-TROSY	H	N		2048	512		40	14423	2839	
Apo	^{15}N-TROSY	H	N		2048	128		8	14706	2839	
Holo	^{15}N-TROSY	H	N		2048	256		16	14706	2839	
$Holo^{His}$	CBCA(CO)NH	H	N	C	2048	64	220	32	11161	2839	16077
Apo	CBCA(CO)NH	H	N	C	2048	48	128	32	11364	2839	16100
Apo^{His}	HN(CA)CO	H	N	C	2048	120	128	16	11161	2839	2818
$Holo^{His}$	HN(CA)CO	H	N	C	2048	128	128	16	11161	2839	2818
Holo	HN(CA)CO	H	N	C	2048	48	128	8	11364	2839	2818
Apo^{His}	HNCA	H	N	C	2048	48	230	16	11161	2839	6039
$Holo^{His}$	HNCA	H	C	N	2048	120	40	8	14423	6039	2839
Apo	HNCA	H	N	C	2048	64	128	16	11364	2839	6037
Holo	HNCA	H	N	C	2048	48	110	8	11364	2839	6037
Apo^{His}	HNCACB	H	N	C	2048	120	144	16	14423	2839	16077
$Holo^{His}$	HNCACB	H	N	C	2048	120	144	16	14423	2839	16077
Apo	HNCACB	H	N	C	2048	124	128	16	14706	2839	16100
Holo	HNCACB	H	N	C	2048	110	128	16	14706	2839	16100
Apo^{His}	HNCO	H	N	C	2048	40	108	16	14423	2839	2818
$Holo^{His}$	HNCO	H	N	C	2048	114	56	8	14423	2839	2818
Holo	HNCO	H	N	C	2048	40	128	8	12500	2839	3220

NMR titrations contained 150 µM ^{15}N-labelled protein in the same buffer with increasing amount of c-di-AMP (Biolog) to achieve PstA:c-di-AMP ratios of 1:0.5, 1:1, 1:2 and 1:5. 2D ^{1}H–^{15}N TROSY-HSQC experiments were acquired using 600 complex points and 32 scans

per increment with spectral widths of 14423 Hz and 2839 Hz in the ^{1}H and ^{15}N dimensions, respectively, and a total experiment time of 6 hours. All NMR experiments were processed with TopSpin (Bruker BioSpin) and analyzed with CCPN Analysis.[295] Averaged chemical-shift perturbations were calculated according to the equation $\sqrt{\Delta\delta_{\mathrm{H}}^2 + (0.14 \cdot \Delta\delta_{\mathrm{N}})^2}$.

5.5.3 Backbone ^{15}N Relaxation Experiments

For NMR relaxation studies, uniformly ^{15}N-labelled (u-^{15}N) and ^{2}H/^{15}N/^{13}C-triply-labelled (u-^{15}N/^{13}C) PstA samples were prepared in a mixed solvent of 90 % $^{1}H_2O$ and 10 % $^{2}H_2O$ (50 mM sodium phosphate, 50 mM NaCl, pH 6.4). All NMR experiments were carried out at 30 °C with protein concentrations between 0.5 and 0.6 mM on a Bruker Avance 800 MHz NMR spectrometer.

^{15}N Longitudinal (R_1) and transverse (R_2) relaxation rates and Steady-state [^{15}N, ^{1}H] heteronuclear nuclear Overhauser effect (hetNOE) experiments on ^{15}N spins of the backbone were performed on apo and c-di-AMP-bound PstA at a B_0 field of 18.8 T and 30 °C. ^{15}N backbone relaxation rates R_1 and R_2 were measured using relaxation delays of 1, 10, 50, 100, 200, 300, 600, 800, and 1000 ms for R_1 and 17.0, 33.9, 50.9, 67.8, 84.8, 101.8, 135.7, 152.6, 169.6, 203.5, and 254.4 ms for R_2. Peak intensities were fitted to a single exponential decaying function. Steady-state ^{1}H–^{15}N hetNOE intensities were measured in an interleaved manner using ^{1}H saturation time of 3 s and a recycle delay of 5 s, which was adequate to completely saturate amide proton resonances. Steady-state [^{15}N, ^{1}H] heteronuclear-NOE measurements were carried out with and without proton saturation during the relaxation delay. In these NOE-experiments, 5 s of relaxation delay and 3 s of proton saturation (or 8 s of relaxation delay only) were used. The heteronuclear-NOE values were determined as the ratio of the peak intensities measured from the spectra acquired with and without proton saturation. All relaxation measurements were performed at 303 K. NMR spectra were processed using TopSpin (Bruker BioSpin) and analysed using CARA[362] and CCPN.[295] The added contribution to the dephasing seen for the R_2 rates of the apo form compared to the holo receptor cannot stem from processes slower than the µs timescale, as those would be reflected in the BMRD experiments and as the rates were recorded using a CPMG pulse train of 2 kHz.

Constant-time ^{15}N-CPMG (CT-CPMG) relaxation dispersion experiments[363] were measured at 303 K. Experiments were performed with a constant-time delay of 40 ms, and 9 variable CPMG frequencies (ν_{CPMG}) ranging from 50 to 2000 Hz were collected. In addition, a reference spectrum without constant time delay ($T_{\mathrm{CPMG}} = 0$) was recorded. Each 2D dataset corresponding to one CPMG frequency, 128 and 2048 complex points in the indirect and direct dimensions, respectively, was collected with 64 scans. 2D datasets for all frequencies were measured in a scan-interleaved fashion. The recycle delay of 1.5 s was used, giving rise to a net acquisition time of approximately 5.2 h per data set. The

^{15}N carrier during the CPMG period was set to 118 ppm and 117 ppm for PstA in the apo and ligand-bound form, respectively. The ^{15}N 90° pulse width of CPMG pulses was set to 32.0 µs. NMR spectra were processed using TopSpin (Bruker BioSpin). Peak intensities were quantified and visualised with CCPN.[295] The effective amide ^{15}N transverse relaxation rate at each CPMG frequency was calculated according to the relation of effective transverse relaxation rate

$$R_{2,\text{eff}} = -\frac{1}{T_{\text{CPMG}}} \cdot \ln\left(\frac{I}{I_0}\right) \tag{5.1}$$

where I is the peak intensity, I_0 is the intensity in a reference spectrum recorded without the CT-CPMG relaxation period and T_{CPMG} is a constant time delay. Evaluation of CPMG data was performed using ChemEx.[80] Relaxation dispersion curves were fitted numerically according to the Block-McConnell equations on the basis of a two-state exchange model in a residue-specific and, in addition, for several residues showing relaxation dispersion also in a global manner. The noise level of each spectrum was used as error for the intensities and propagated in the fitting procedure. The exchange rate R_{ex} was determined as difference between the maximum calculated effective transverse relaxation rate $R_{2,\text{eff}}$ (maximum of the fit curve) and the fitted intrinsic transverse relaxation rate R_2^0.

Table 5.2: Exchange contribution and timescales of the individual relaxation dispersion fits (for residues with significant exchange only).

Residue	$\boldsymbol{\tau_{\text{ex}}}$	$\boldsymbol{R_{\text{ex}}}$
27	(5.7 ms ± 38.1 ms)	18.5 Hz ± 2.3 Hz
28	166.2 µs ± 83.4 µs	15.2 Hz ± 3.5 Hz
29	(222.0 µs ± 819.7 µs)	16.5 Hz ± 8.7 Hz
32	38.6 ms ± 8.9 ms	37.5 Hz ± 3.8 Hz
34	(10.7 ms ± 60.3 ms)	14.5 Hz ± 2.8 Hz
46	166.4 µs ± 67.5 µs	20.0 Hz ± 4.5 Hz

A global fit (of k_{ex} and population p_{B} with interchanging states A and B) of residues 28, 29 and 46 yields $k_{\text{ex}} = 190.8\,\mu\text{s} \pm 94.1\,\mu\text{s}$, i.e., a 100 µs timescale conformational exchange. However, while the presence/absence of exchange is unambiguous, all of these timescales of the individual dispersion profiles are difficult to fit in an accurate fashion due to the low signal to noise obtained for the trimeric complexes.

5.5.4 Molecular Dynamics Simulations

All moldecular dynamics (MD) simulations were performed using the CHARMM36m force field,[100] which is well established for both ordered and disordered proteins, together with the TIP3P water model.

For the Gaussian-accelerated molecular dynamics (GaMD) simulations performed with NAMD 2.14,[364] starting structures obtained from clustering of preliminary, not further described, classical MD simulations of PstA apo and ligand bound form were used. These starting structures differed in each chain and the B-loops were in extended conformation close to the protein surface and did not contain any well-defined secondary structure elements. Both systems were prepared using CHARMM-GUI.[365,366] Both systems were placed into an octahedral box with unit cell lengths of 10.5 nm, solvated with TIP3P water[298] at pH 7 and neutralized with NaCl at a physiological concentration of 50 mM. After energy minimization, the respective system was equilibrated for 500 ps in the NVT ensemble and for 2.5 ns in 5 steps in the NpT ensemble, during which the position restraints were gradually released. Temperature was adjusted to 303.15 K using the Langevin thermostat and the pressure was kept constant at 1 bar using the Nosé-Hoover Langevin piston barostat.[114,367] All bond lengths in protein and ligand were kept constant with SHAKE, allowing to integrate the equations of motion with 2 fs time steps. Lennard-Jones 12-6 and electrostatic interactions were smoothly shifted to zero at a 1.2 nm cut-off. Long-range electrostatic interactions were treated with the particle mesh Ewald algorithm.[201] The dual-boost biasing potential (on both, dihedral and total potential energy) was equilibrated in two steps:[368] At first, 10 ns of conventional MD were performed, in which statistics about the potential energy of the system were collected. In the second 50 ns long step, collection of potential statistics was continued, but an additional, concurrently updated biasing potential was applied. In the end of the equilibration, the potential statistics were fixed for calculating the boost potential applied at the production run. The upper limit $\sigma_{0\mathrm{D/P}}$ for the standard deviation of the potential energy was set to 18 kcal/mol, which corresponds to acceleration factors k_0 of 0.9 and 0.1 for the dihedral and total boost potential. The final GaMD production runs were conducted for 1 µs. Although this high biasing potential does not allow for accurate reweighting of the potential energy surface anymore, it still ensures that the large conformational space of the B-loop is extensively sampled and provides more reliable and representative starting structures for replica MDs than simple loop modellers, which are only suited for short and rather rigid loops.[369]

The classical MD simulations of PstA in apo and c-di-AMP-bound form were carried out with Gromacs 2024.2.[200] As starting structures, snapshots were taken every 200 ns from the preceding GaMD simulation. After energy minimization, the respective system was equilibrated for 500 ps in the NVT ensemble and for 2.5 ns in 5 steps in the NpT ensemble, during which the position restraints were gradually released. Temperature was kept

constant at 303.15 K using the V-rescale thermostat developed by Bussi and coworkers[111] and pressure was kept constant at 1 bar using the Parrinello-Rahman barostat.[370,371] All bond lengths in protein and ligand were constrained with LINCS, and water molecules were kept rigid with SETTLE, allowing to integrate the equations of motion with 2 fs time steps. Lennard-Jones 12-6 interactions were smoothly shifted to zero at a 1.2 nm cut-off. Long-range electrostatic interactions were treated with the particle mesh Ewald algorithm[201] using a distance of 1.2 nm for switching between short- and long-range electrostatic interactions. The final production runs were performed for 1 µs each.

5.5.5 Protein Sequence Alignment

The protein sequence alignment was generated by the constraint-based alignment tool COBALT[372] for 50 PstA homologues, including *Staphylococcus aureus* PstA (PDB codes 4WK1 and 4D3H), *Listeria monocytogenes* PstA (PDB code 4RWW) and *Bacillus subtilis* DarA (PDB code 4RLE) (compare Fig. 5.6).

5.5.6 Small Angle X-ray Scattering

Small angle X-ray scattering (SAXS) data were collected at EMBL DESY (Hamburg, Germany). PstA with His_6-tags was additionally purified by size exclusion chromatography in Tris buffer to ensure a sample free of aggerates. The c-di-AMP-bound PstA complex was formed after addition of filtered c-di-AMP in identical buffer conditions to the apo protein. To ensure absence of free c-di-AMP that might interfere with the buffer reference, we used a minimal under-saturation (99 %) for complex preparation. Samples were centrifuged prior to measurement and were measured in different concentrations, alternating with two buffer measurements for each sample in batch mode. After buffer correction, scattering data of the different concentrations were analysed and merged using ATSAS[373] as previously described.[374,375] The radius of gyration R_g was determined by Guinier plot analysis ($\ln I(s)$ vs. s^2) including points with $s \cdot R_g < 1.3$ in the linear regression (Guinier approximation for globular proteins) and pair-distribution analysis was performed using the GNOM option in ATSAS.[376] The Guinier plots for apo and c-di-AMP complex indicate that the samples are free of aggregates with $R_g^{\text{PstA apo}} = 2.67\,\text{nm} \pm 0.01\,\text{nm}$ and $R_g^{\text{PstA holo}} = 2.60\,\text{nm} \pm 0.01\,\text{nm}$. The error distribution shows that the sample is free of aggregates.

5.5.7 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) data were collected at 20 °C using a Malvern PEAQ-ITC system with 15 µM PstA in Tris buffer in the cell. c-di-AMP was dissolved in the same buffer to 177 µM and titrated into the cell by 19 injections of 2 µl, spaced 150 s apart. Data evaluation was done using the Malvern PEAQ-ITC software. Experiments were repeated at least once to confirm the robustness of the assay.

5.6 Supplementary Figures and Tables

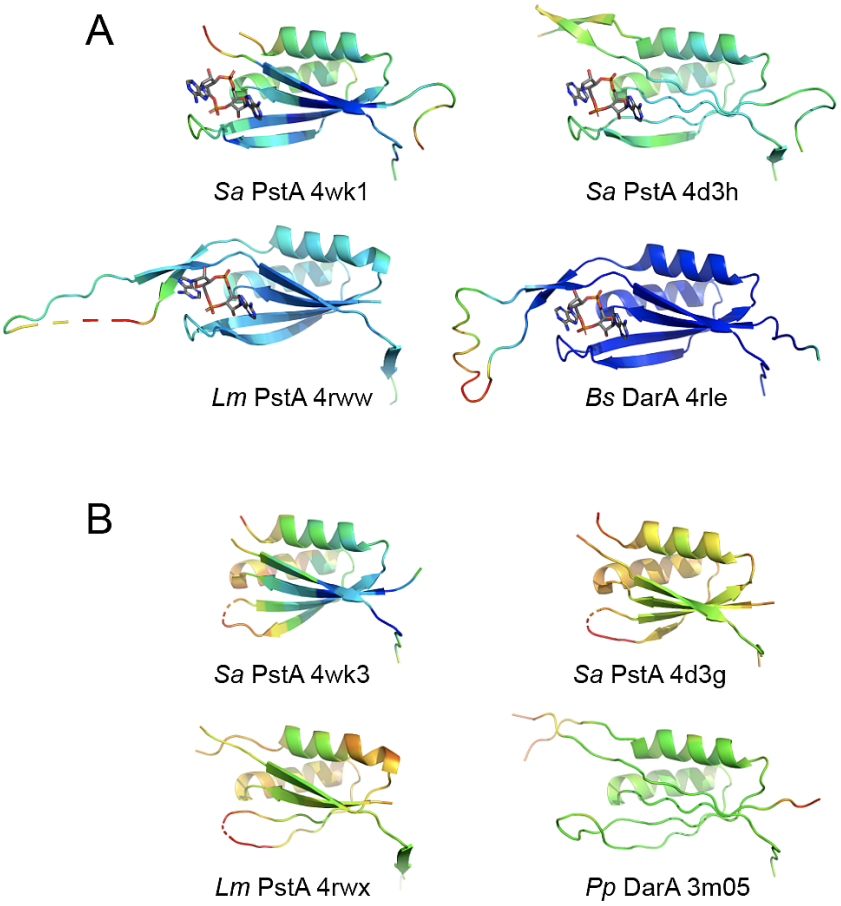

Figure 5.6: Monomeric units of the crystal structures of PstA homologs. A) c-di-AMP-bound species, in particular PstA from *Staphylococcus aureus* (*Sa*) and *Listeria monocytogenes* (*Lm*), as well as DarA from *Bacillus subtilis* (*Bs*). B) Apo species, in particular again PstA from *S. aureus* and *L. monocytogenes*, as well as DarA from *P. pentosaceus* (*Pp*). Colours are according to relative B-factor colouring in Pymol (blue: highly defined density, low density). The B-loop protrudes from the upper left.

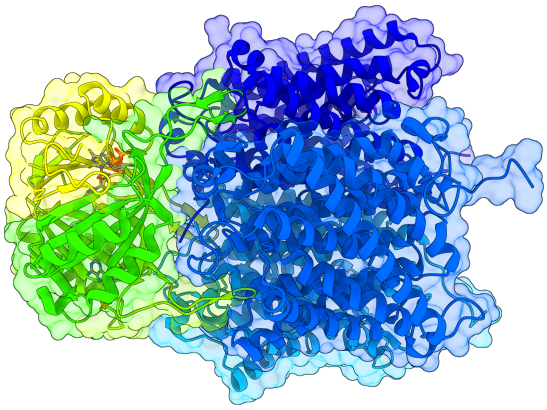

Figure 5.7: Association of the homologue P_{II} signal transduction protein GlnK of *E. coli* to its target, the ammonia channel AmtB (PDB 2NUU) mediated by the three long T-loops.[336]

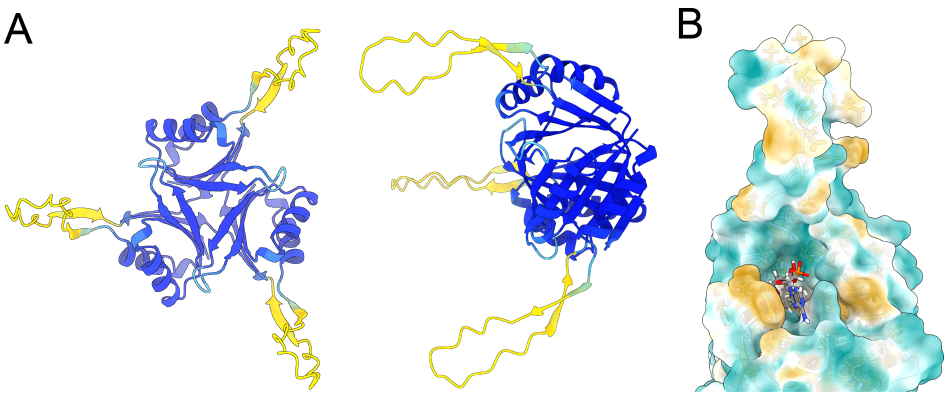

Figure 5.8: Overall topology of PstA and B-loop hydrophobicity. A) AlphaFold structure of the *S. aureus* PstA monomer, showing the confidence of structural predictions as a color gradient (blue: confident, yellow: not confident). B) Hydrophobic surface of PstA in ligand-bound form. The B-loop is modelled here as extended structure ranging into the solvent. A high number of hydrophobic residues in the B-loop suggest potential interactions to the underlying surface (the ligand binding site), which is the only other region on the protein surface with high hydrophobicity.

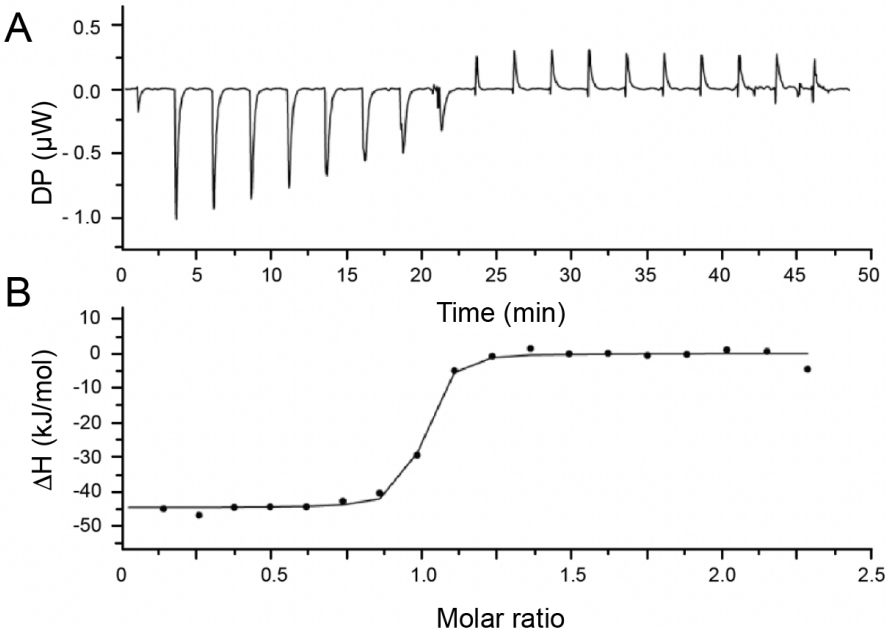

Figure 5.9: Energetics of binding of c-di-AMP PstA. A) Isothermal titration calorimetry (ITC) analysis of PstA (15 µM) with c-di-AMP (177 µM) monitored at 20 °C. B) Binding isotherm of the calorimetric titration of c-di-AMP into PstA, showing the integrated heats (dots) and the respective fit (solid line). The fitting marks a K_D of 19 ± 8 nM, $n = 0.95$, and $\Delta H = -44.7\,\mathrm{kJ\,mol^{-1}}$.

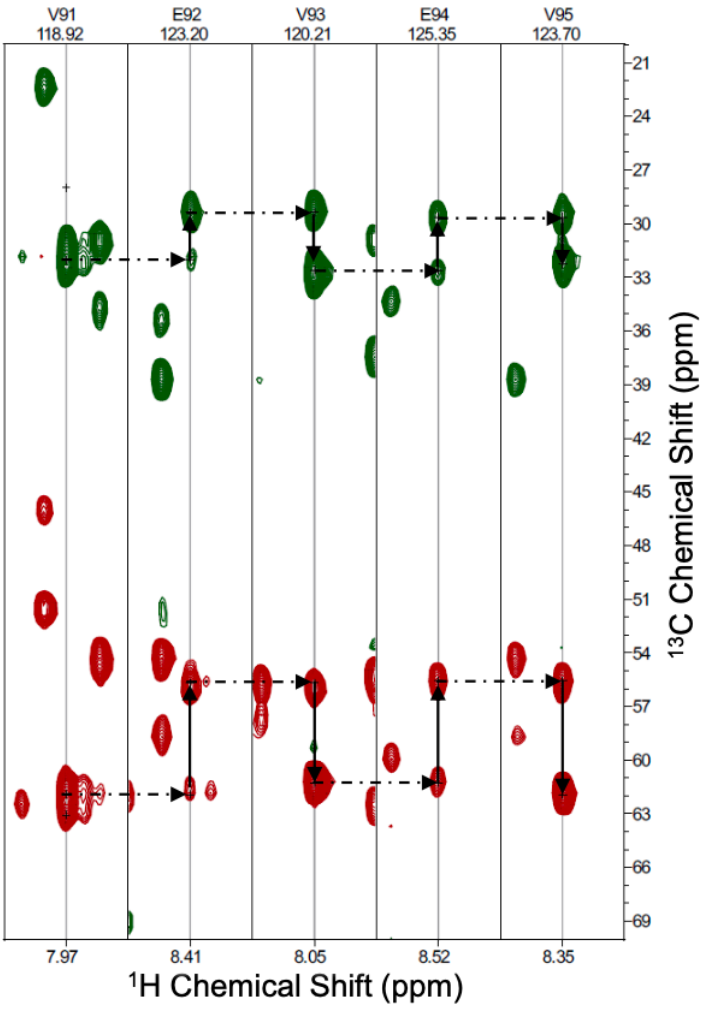

Figure 5.10: Exemplary walk through the primary sequence, shown here for an HNCACB experiment for residues V91 to V95.

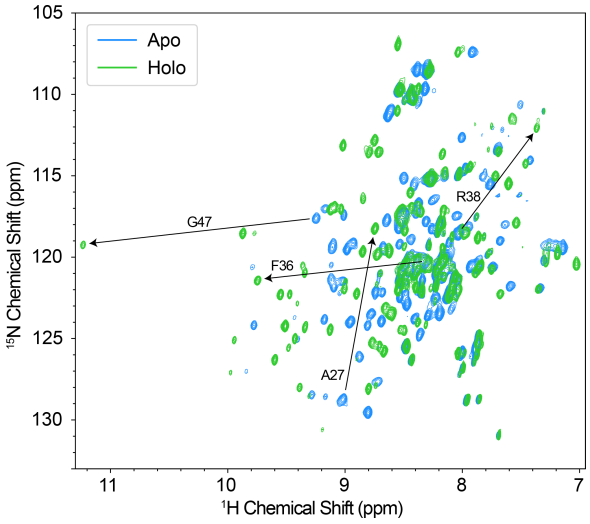

Figure 5.11: Overlay of 2D [^{1}H, ^{15}N]-HSQC spectra for apo PstA (blue) and c-di-AMP-PstA complex (green) under similar experimental conditions, together with residue assignments containing CSPs > 1.0 ppm.

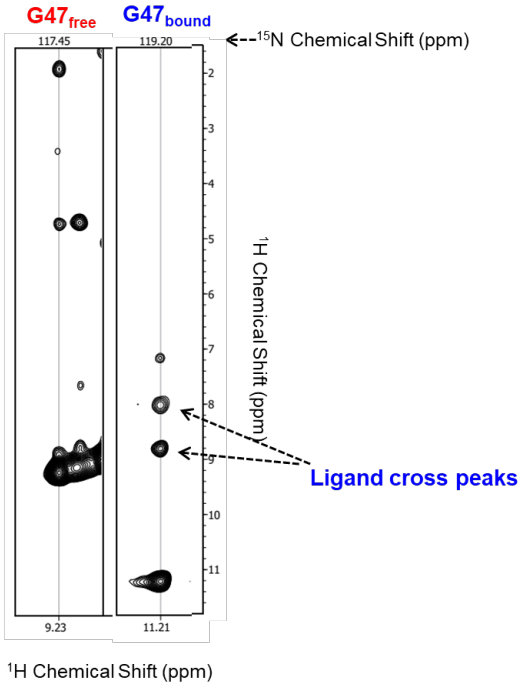

Figure 5.12: ^{15}N-edited NOESY strips for ligand:protein contacts in close proximity to G47.

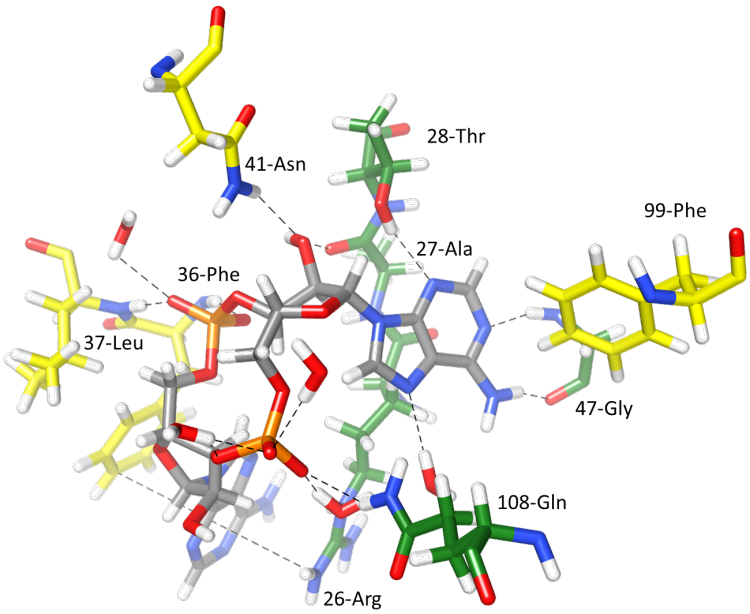

Figure 5.13: Embedding of the c-di-AMP ligand as seen in crystallography. Hydrogen bonds and π interactions between c-di-AMP and PstA (PDB code 4WK1) in the binding cavity are shown as dashed and solid lines, respectively. Residues from monomer A are shown with green carbon atoms and residues from monomer B are shown with yellow carbon atoms. T28, L37, F38, N41, G47 and Q108 form hydrogen bonds with different parts of c-di-AMP, R26 forms a cation-π interaction with adenine, F36 a face-to-face π-stacking with adenine and F99 might form a weak edge-to-face π-stacking with adenine (in the crystal structure of SaPstA according to I. Campeotto et al., PDB code 4D3H[348]).

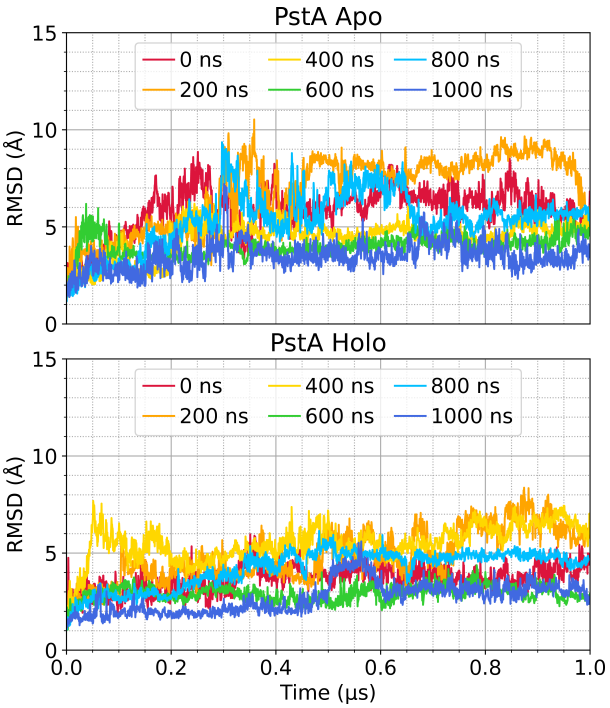

Figure 5.14: RMSD of the cMD simulations to the respective starting structure as a function of time. As these starting structures were obtained from initial GaMD simulations, the individual cMD runs (different colors) are labelled according to the timepoint at which they were initially found in the GaMD simulations.

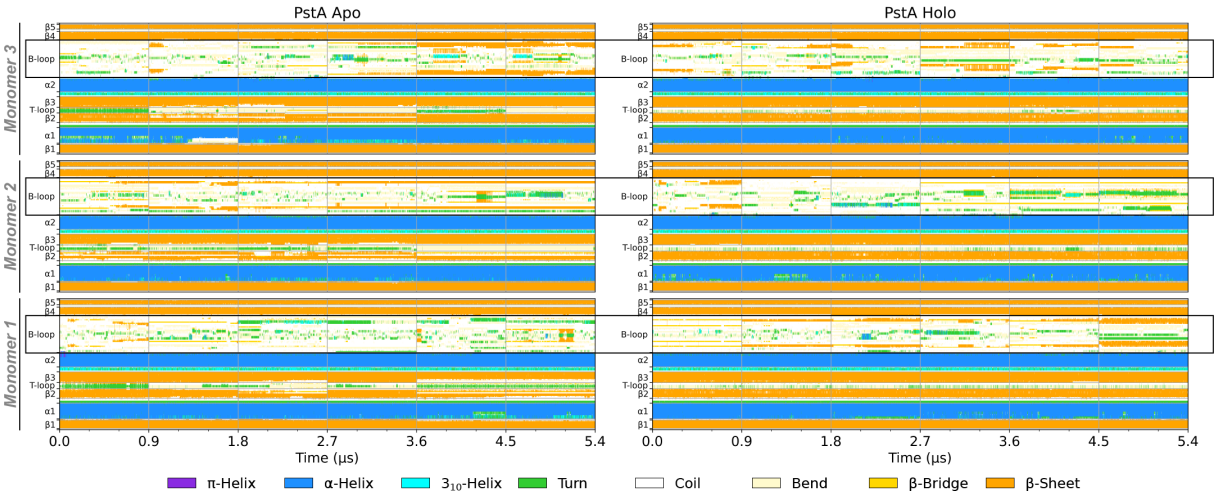

Figure 5.15: Secondary structural propensities as seen in the MD simulation for apo (left) and holo form (right). The three monomers are plotted on top of each other, the individual trajectories (omitting the first 100 ns) are shown one after the other. The secondary-structure properties are denoted by colors as shown at the bottom.

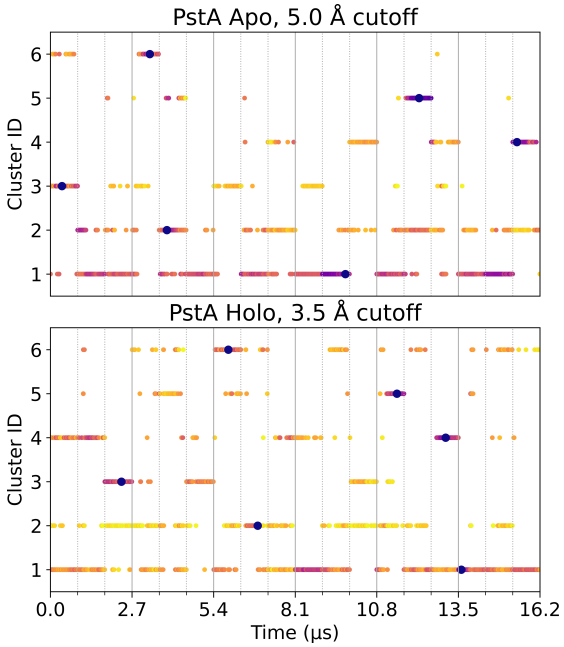

Figure 5.16: Projection of the cluster IDs resulting from clustering of the concatenated trajectories of the extracted monomers ($6 \cdot 3 \cdot 900$ ns per trajectory analysed) with respect to the simulation time using RMSD cut-offs of 5.0 Å and 3.5 Å for the apo and holo receptor. The dark blue dots indicate the cluster center and the colouring denotes the RMSD with respect to the respective cluster center. The high number of transitions between the different clusters substantiates a sufficient sampling of the performed conventional MD simulations and a low to negligible bias of the starting structures.

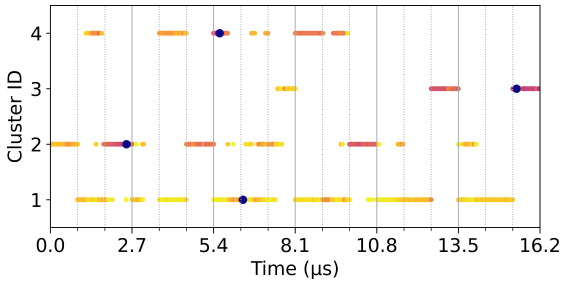

Figure 5.17: Projection of the cluster IDs resulting from clustering the T-loop and core (disregarding B-loop conformations, with a cut-off of 1.2 Å), as in Main Text Fig. 5.4D. Represented in similar way as Fig. 5.16.

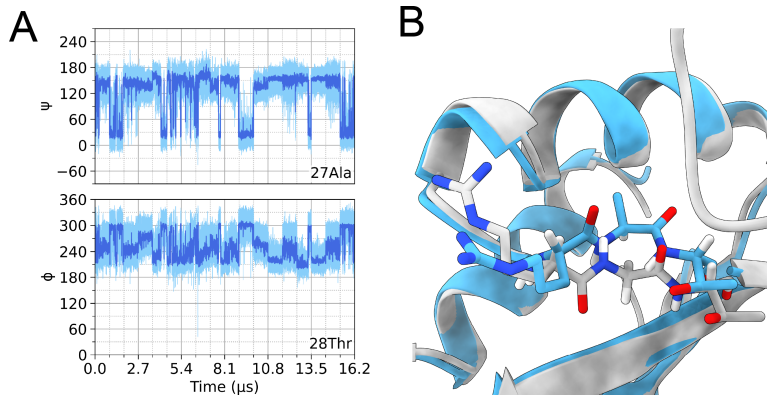

Figure 5.18: Peptide flip in residues 26 – 28 on the fast µs timescale. This local motion is in addition to the chemical exchange witnessed for the entire region (27 – 34) in the relaxation dispersion experiments, which can only capture exchange phenomenons $> 100\,\mu s$.

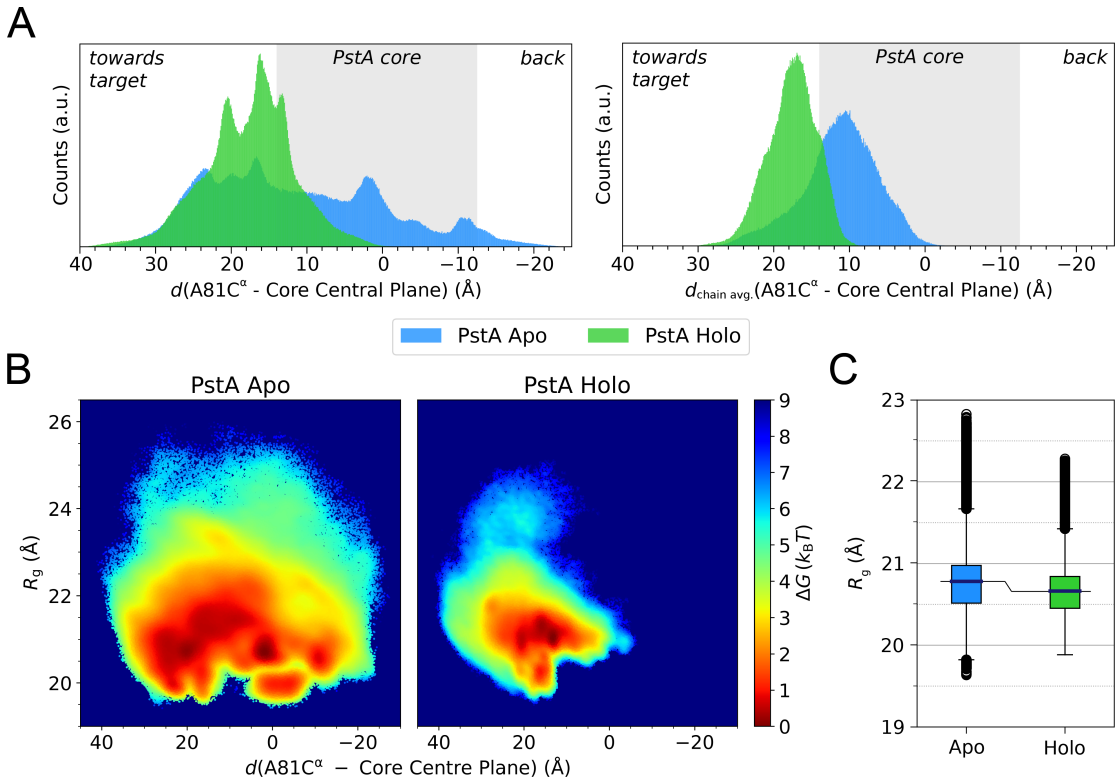

Figure 5.19: Distribution of loop orientations. A) Distribution of individual (top) and chain-averaged loop orientations (below). B) Free energy surface of the individual orientations together with the corresponding radius of gyration (determined with respect to the center of mass of the trimer). C) Average radius of gyration (R_g) as obtained from the simulation of apo and holo PstA,

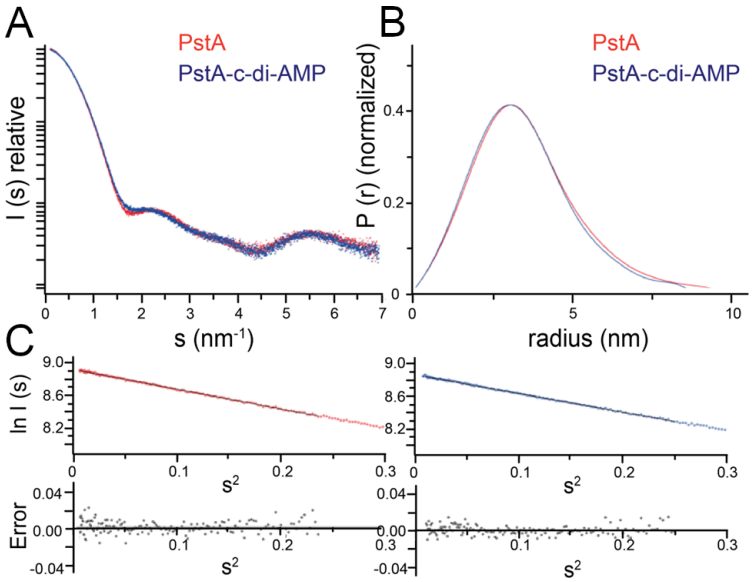

Figure 5.20: Small-angle X-ray scattering curves of c-di-AMP-bound (blue) and free PstA (red). Despite similar overall shape, the complex shows differences in the local minimum at $s = 1.7\,\text{nm}$. Pair distribution functions obtained using GNOM of PstA (red) and the complex (blue) indicate that the latter has a reduced maximum particle diameter ($D_{\text{max}}^{\text{PstA apo}} = 9.24\,\text{nm}$; $D_{\text{max}}^{\text{PstA apo}} = 8.59\,\text{nm}$). The error is estimated from the linear regression vs. the measured data. Accordingly, the Guinier plots are fitted to an experimental radius of gyration R_g of $2.67 \pm 0.01\,\text{nm}$ and $2.60 \pm 0.01\,\text{nm}$ for apo and holo form, which qualitatively matches with the simulation (Fig. 5.19C) and again confirms the modulation of overall loop redirection towards a more restricted space.

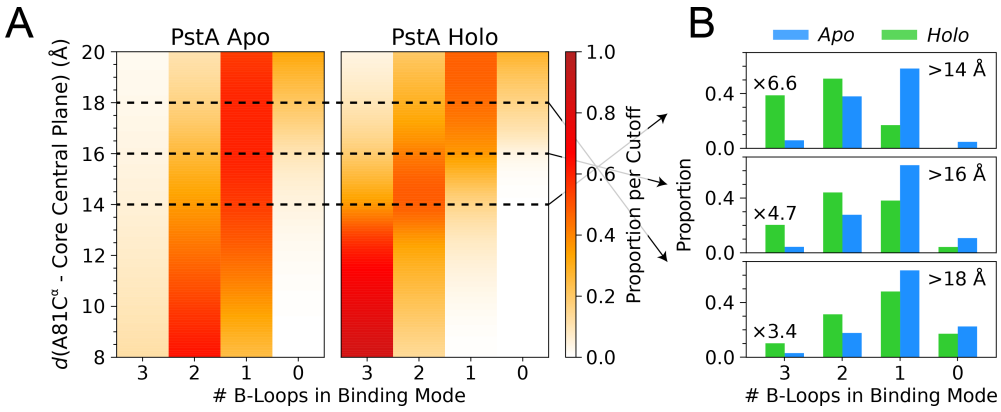

Figure 5.21: Distribution of collective forward positioning (3, 2, 1, or none of the loops are beyond a given threshold) as a function of the threshold (minimum distance between A81C$^{\alpha}$ and the center of mass) chosen. In A), a continuous variation of the distance threshold is chosen, whereas in B), three discrete thresholds (14, 16, and 18 Å, same as in Main Text Fig. 5.5C) are specifically read out.

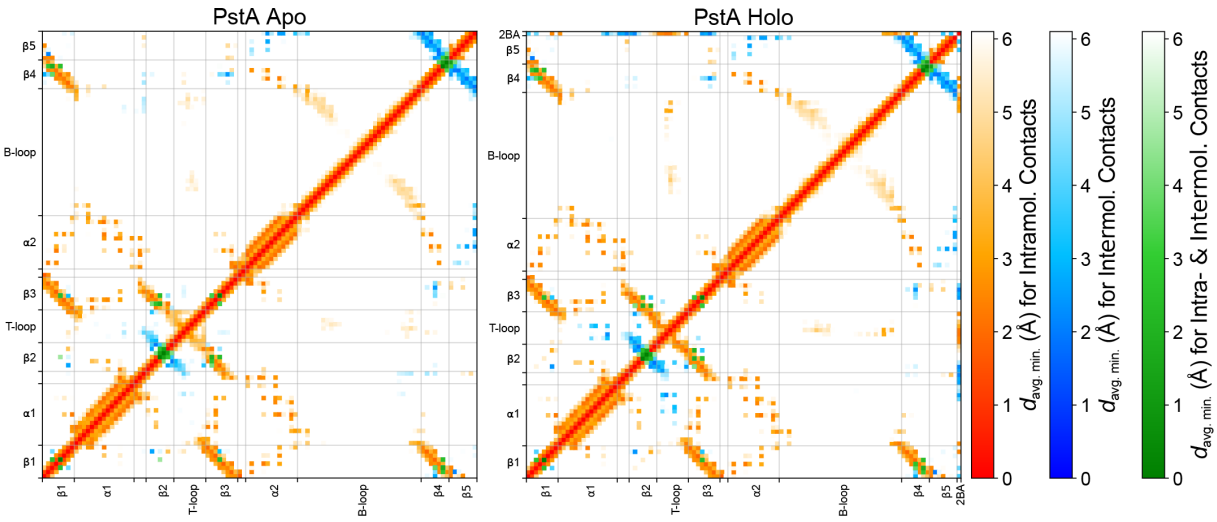

Figure 5.22: Contact analyses for apo PstA and the PstA:c-di-AMP complex. Intramonomer contacts are shown in orange/red colors, intermonomer contacts are shown in blue colors, green denotes the presence of both, intra- and intermonomer contacts. PstA:c-di-AMP contacts (shown as the rightmost column/top row in the plot for the holo form) distinguish contacts to two different PstA monomers.

Table 5.3: Most important hydrogen bonds formed in MD and their relative occurrence.

H-bond			Percentage H-bond exists Apo	Holo	Type
67ArgHH*	–	12AspOD*	57 %	67 %	B-loop – Core
95GlyO	–	9GlnH	18 %	64 %	B-loop – Core
95GlyH	–	109PheOT*	23 %	52 %	B-loop – Core intermol.
65GlyH	–	12AspOD*	64 %	50 %	B-loop – Core
38ArgHH*	–	18AspOD*	12 %	50 %	T-loop – Core intermol.
94/95GlyO	–	9GlnHE*	27 %	34 %	B-loop – Core
35GlyH	–	27GlyO	2 %	28 %	T-loop – Core intermol.
79GlyH	–	36PheO	0 %	14 %	B-loop – T-loop
34GlyO	–	29LysH	19 %	0 %	T-loop – Core intermol.
34GlyO	–	28ThrOD1	19 %	0 %	T-loop – Core intermol.
40GlyO	–	10AspH	36 %	1 %	T-loop – Core
2BA.N1	–	47GlyH	-	91 %	Ligand – Core
2BA.P*P1	–	36PheH	-	89 %	Ligand – T-Loop
2BA.O2P1	–	37LeuH	-	75 %	Ligand – T-Loop
2BA.H6*	–	47GlyO	-	68 %	Ligand – Core
2BA.H61*	–	25PheO	-	60 %	Ligand – Core
2BA.O2P	–	94GlyH	-	49 %	Ligand – B-Loop
2BA.N3	–	28ThrHG1	-	40 %	Ligand – Core
2BA.HO2’	–	28ThrOG1	-	38 %	Ligand – Core
2BA.H61*	–	21ValO	-	38 %	Ligand – Core
2BA.HO21	–	92GluO	-	37 %	Ligand – B-Loop
2BA.O1P1	–	35GlyH	-	29 %	Ligand – T-Loop
2BA.O1P	–	26ArgHH11	-	27 %	Ligand – Core
2BA.O2P	–	95GlyH	-	26 %	Ligand – B-Loop
2BA.O2’	–	28ThrHG1	-	25 %	Ligand – Core

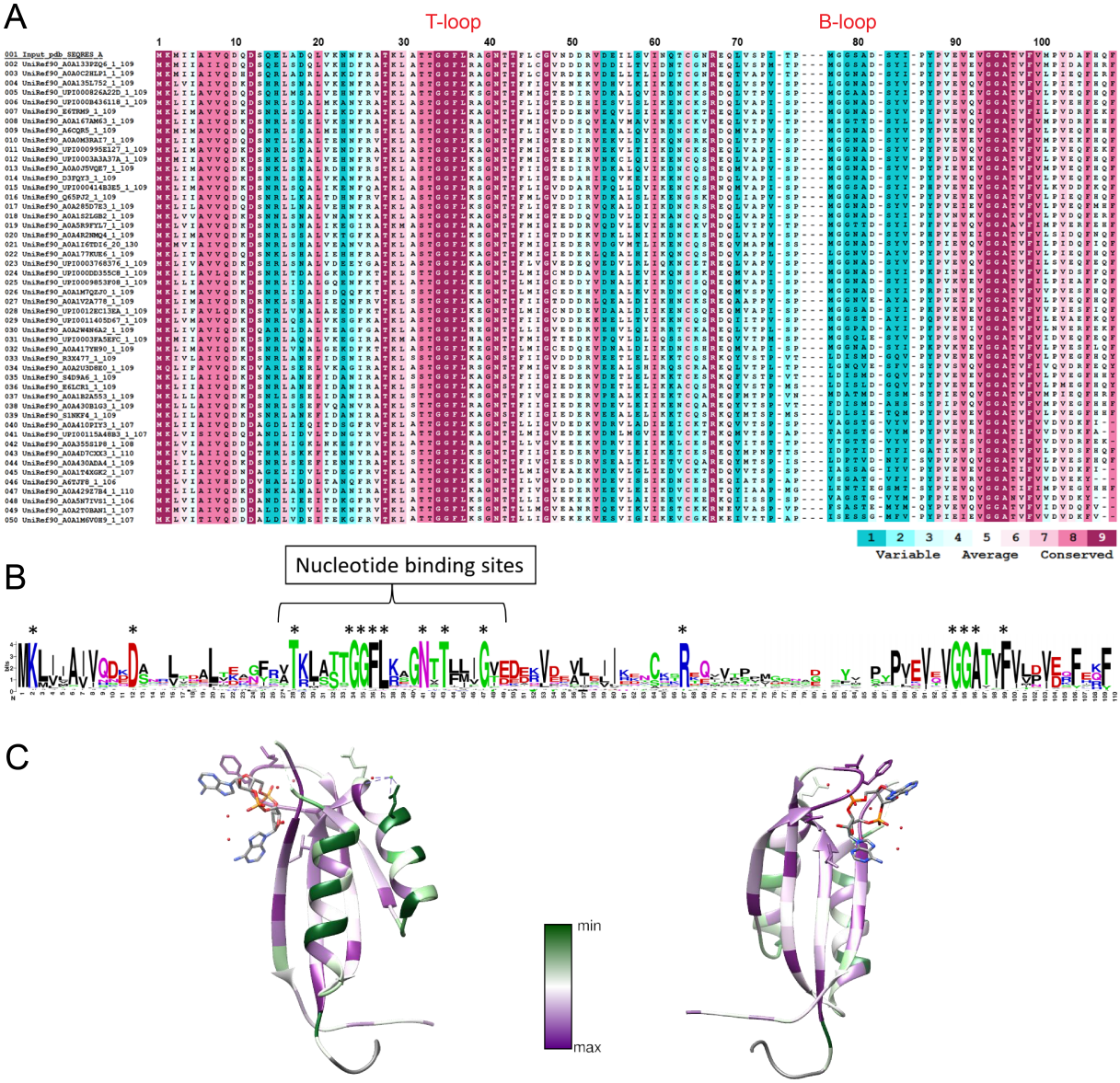

Figure 5.23: Conservation of the PstA primary sequence, the size of the letter representing the extent of sequence conservation. A) Alignment of all sequences compared. B) Conservation represented by the size of the respective letters in the primary sequence. C) Conservation color coded on the monomeric, holo PstA structure.

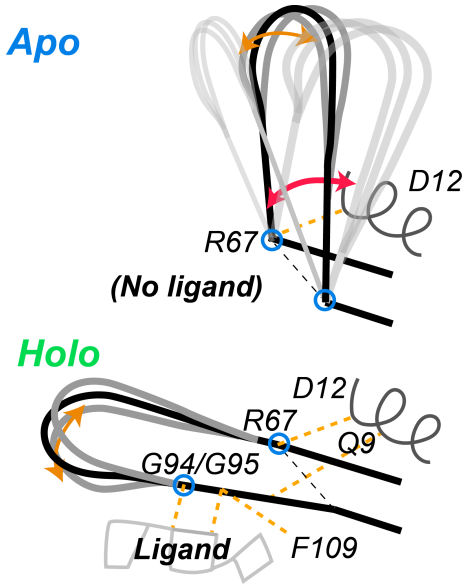

Figure 5.24: Simplified representation of the transient interactions found to effectively redirect the B-loop conformational ensemble in the presence of the ligand. Whereas the salt bridge between R67 and D12 acts like a hinge (blue circles) both, in the apo and holo form, the newly arising network of transient interactions with and around the ligand act like a latch that effectively reduces angular motion of the loop at its root (green).

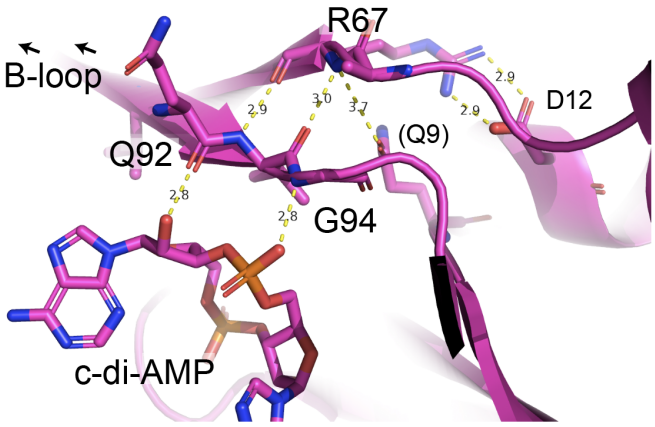

Figure 5.25: H-bonds seen in the crystal structure of holo PstA of *L. monocytogenes* (PDB 4RWW). The stabilization of the loop's beginning and end upon crystallization, in particular between c-di-AMP and conserved residues Q92 and G94 at the end of the B-loop, as well as the salt bridge between the sidechain of R67 with D12 at the end of α-helix 1 seen here as a static architecture, coincides at least with part of the transient H-bonds seen in the dynamic conformational ensemble of the MD simulations.

6

Appendix

6.1 Publications

An up-to-date list of publications is available via my ORCiD 0000-0003-2738-8263.

Publications with major/equal contributions:

- B. Söldner, K. Grohe, P. Neidig, J. Auch, S. Blach, A. Klein, S. K. Vasa, L. V. Schäfer, R. Linser, *J. Phys. Chem. Lett.* **2023**, *14*, 1725–1731
- J. Kotschy†, B. Söldner†, H. Singh, S. K. Vasa, R. Linser, *Angew. Chem. Int. Ed.* **2024**, *63*, e202313947
- J. Kotschy†, B. Söldner†, H. Singh, S. K. Vasa, R. Linser, *Angew. Chem.* **2024**, *136*, e202313947
- H. G. Daronkola†, B. Söldner†, H. Singh, R. Linser, A. V. Verde, *ChemBioChem* **2024**, *25*, e202400057
- B. Söldner†, H. Singh†, E. Akoury†, G. Witte, R. Linser, *Angew. Chem. Int. Ed.* **2025**, *64*, e202509394
- B. Söldner†, H. Singh†, E. Akoury†, G. Witte, R. Linser, *Angew. Chem.* **2025**, *64*, e202509394

†: These authors contributed equally to this work.

Publications with minor contributions:

- K. Grohe, E. Nimerovsky, H. Singh, S. K. Vasa, B. Söldner, B. Vögeli, C. M. Rienstra, R. Linser, *Chem. Commun.* **2019**, *55*, 7899–7902
- A. Klein, S. K. Vasa, B. Söldner, K. Grohe, R. Linser, *J. Phys. Chem. Lett.* **2022**, *13*, 1644–1651
- T.-C. Lin, L. Engelhard, B. Söldner, R. Linser, D. Summerer, *Adv. Sci.* **2024**, *11*, 2307930
- H. Aucharova†, A. Klein†, S. Medina Gomez, B. Söldner, S. K. Vasa, R. Linser, *Chem. Commun.* **2024**, *60*, 3083–3086

†: These authors contributed equally to this work.

6.2 Conferences and Workshops

Conferences:

- Alpine Conference on Magnetic Resonance in Solids,
  4 – 8 September 2022, Chamonix Mont-Blanc, France.
  Round-table talk: "Assessment of Protein Structure and Dynamics Enabled by Accurate Solid-State NMR Distance Information", B. Söldner, K. Grohe, P. Neidig, J. Auch, S. Blach, A. Klein, S. V. Kumar, L. Schäfer, R. Linser.

- 43rd FGMR Annual Discussion Meeting,
  12 – 15 September 2022, Karlsruhe, Germany.
  Poster presentation: "Structure and Dynamics of Solid Proteins via Accurate Solid-State NMR Distance Information", B. Söldner, K. Grohe, P. Neidig, J. Auch, S. Blach, A. Klein, S. V. Kumar, L. Schäfer, R. Linser.

- Annual *bio*-N^3*MR* Network Meeting,
  19 July 2022, Dortmund, Germany.
  Talk: "Accurate Solid-State NMR Distance Information from Magnetization Build-Up Analysis", B. Söldner, K. Grohe, P. Neidig, J. Auch, S. Blach, A. Klein, S. V. Kumar, L. Schäfer, R. Linser.

- 19th European Magnetic Resonance Congress (EUROMAR),
  9 – 13 July 2023, Glasgow, United Kingdom.
  Promoted talk: "Structure and Dynamics of Proteins from Accurate Solid-State NMR Restraints", B. Söldner, K. Grohe, P. Neidig, J. Auch, S. Blach, A. Klein, S. V. Kumar, L. Schäfer, R. Linser.

- 30th International Conference on Magnetic Resonance in Biological Systems (ICMRBS), 18 – 23 August 2024, Seoul, South Korea.
  Poster presentation: "Allosteric Orchestration of Loop Dynamics of P_{II}-like Signal Transduction Protein A by the Secondary Messenger c-di-AMP", B. Söldner, H. Singh, E. Akoury, G. Witte, R. Linser.

- Gordon Research Conference on Computational Aspects of Biomolecular NMR,
  1 – 6 June 2025, Les Diablerets, Switzerland.
  Poster presentation: "Activation of the Bacterial Secondary-Messenger Receptor PstA via Allosteric Modulation of a Dynamic Conformational Ensemble", B. Söldner, H. Singh, E. Akoury, G. Witte, R. Linser.

Workshops:

- 4^{th} G-NMR School,
  17 – 21 February 2020, Göttingen, Deutschland.
- ISMAR NMR Summer School 1 on “General NMR”,
  23 – 29 February 2020, Windischleuba, Germany.
- Online Course “Biomolecular NMR: Advanced Tools”,
  27 September – 7 October 2021, University of Gothenburg.
- EMBO/Pasteur Practical Course on Integral Structural Biology,
  4 – 9 July 2022, Paris, France.
- ISMAR NMR Summer School 2 on “Theory of NMR”,
  26 March – 1 April 2023, Windischleuba, Germany.

6.3 References

[1] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, et al., *Protein NMR Spectroscopy*, 2nd ed., Academic Press, Burlington, **2007**.

[2] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, 2nd ed., Wiley-VCH, Chichester, **2011**.

[3] R. K. Harris, E. D. Becker, S. M. C. de Menezes, et al., *Pure Appl. Chem.* **2001**, *73*, 1795–1818.

[4] N. E. Hafsa, D. S. Wishart, *J. Biomol. NMR* **2014**, *60*, 131–146.

[5] Y. Shen, A. Bax, *J. Biomol. NMR* **2013**, *56*, 227–241.

[6] J. T. Nielsen, F. A. A. Mulder, *J. Biomol. NMR* **2021**, *75*, 273–291.

[7] H. Friebolin, *Ein- und zweidimensionale NMR-Spektroskopie*, 5th ed., Wiley-VCH, Weinheim, **2013**.

[8] C. P. Slichter, *Principles of Magnetic Resonance*, 3rd ed., Springer, Berlin, **1996**.

[9] M. Karplus, *J. Chem. Phys.* **1959**, *30*, 11–15.

[10] Martin. Karplus, *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.

[11] B. Vögeli, J. Ying, A. Grishaev, et al., *J. Am. Chem. Soc.* **2007**, *129*, 9377–9385.

[12] F. Li, J. H. Lee, A. Grishaev, et al., *ChemPhysChem* **2015**, *16*, 572–578.

[13] G. A. Morris, R. Freeman, *J. Am. Chem. Soc.* **1979**, *101*, 760–762.

[14] M. J. Duer, *Introduction to Solid-State NMR Spectroscopy*, Blackwell Publishing, Oxford, **2004**.

[15] A. E. McDermott, T. Polenova, *Solid State NMR Studies of Biopolymers*, Wiley-VCH, Hoboken, **2012**.

[16] J. C. Paniagua, *Concept Magn. Reson. A* **2006**, *28A*, 384–409.

[17] M. Mehring, *Principles of High Resolution NMR in Solids*, 2nd ed., Springer, Berlin, **1983**.

[18] H. W. Spiess in *NMR Basic Principles and Progress*, *15*, (red.: P. Diehl, E. Fluck, R. Kosfeld), Springer, Berlin, **1978**, 85–214.

[19] T. Polenova, R. Gupta, A. Goldbourt, *Anal. Chem.* **2015**, *87*, 5458–5469.

[20] J. Klinowski, *New Techniques in Solid-State NMR*, Springer, Berlin, **2005**.

[21] Y. Nishiyama, G. Hou, V. Agarwal, et al., *Chem. Rev.* **2023**, *123*, 918–988.

[22] J. R. Lewandowski, J.-N. Dumez, Ü. Akbey, et al., *J. Phys. Chem. Lett.* **2011**, *2*, 2205–2211.

[23] S. K. Vasa, P. Rovó, R. Linser, *Acc. Chem. Res.* **2018**, *51*, 1386–1395.

[24] V. Tugarinov, V. Kanelis, L. E. Kay, *Nat. Protoc.* **2006**, *1*, 749–754.

[25] M. Huber, S. Hiller, P. Schanda, et al., *ChemPhysChem* **2011**, *12*, 915–918.

[26] R. Kerfah, M. J. Plevin, R. Sounier, et al., *Curr. Opin. Struct. Biol.* **2015**, *32*, 113–122.

[27] K. Xue, R. Sarkar, C. Motz, et al., *J. Phys. Chem. C* **2018**, *122*, 16437–16442.

[28] M. Callon, D. Luder, A. A. Malär, et al., *Chem. Sci.* **2023**, *14*, 10824–10834.

[29] M. Schledorn, A. A. Malär, A. Torosyan, et al., *ChemBioChem* **2020**, *21*, 2540–2548.
[30] A. Samoson, *J. Magn. Reson.* **2019**, *306*, 167–172.
[31] I. Bertini, K. S. McGreevy, G. Parigi, *NMR of Biomolecules*, Wiley-VCH, Weinheim, **2012**.
[32] B. Elena, A. Lesage, S. Steuernagel, et al., *J. Am. Chem. Soc.* **2005**, *127*, 17296–17302.
[33] B. Alonso, D. Massiot, *J. Magn. Reson.* **2003**, *163*, 347–352.
[34] G. De Paëpe, B. Eléna, L. Emsley, *J. Chem. Phys.* **2004**, *121*, 3165–3180.
[35] N. Agnieszka, Lund University, Lund, **2012**, 59 pp.
[36] E. Barbet-Massin, A. J. Pell, J. S. Retel, et al., *J. Am. Chem. Soc.* **2014**, *136*, 12489–12497.
[37] S. R. Hartmann, E. L. Hahn, *Phys. Rev.* **1962**, *128*, 2042–2053.
[38] B. H. Meier, *Chem. Phys. Lett.* **1992**, *188*, 201–207.
[39] S. Hediger, B. H. Meier, N. D. Kurur, et al., *Chem. Phys. Lett.* **1994**, *223*, 283–288.
[40] S. Hediger, B. H. Meier, R. R. Ernst, *Chem. Phys. Lett.* **1995**, *240*, 449–456.
[41] A. Pines, M. G. Gibby, J. S. Waugh, *J. Chem. Phys.* **1973**, *59*, 569–590.
[42] Y. Ji, L. Liang, X. Bao, et al., *Solid State Nucl. Magn. Reson.* **2021**, *112*, 101711.
[43] N. M. Szeverenyi, M. J. Sullivan, G. E. Maciel, *J. Magn. Reson.* **1982**, *47*, 462–475.
[44] T. Gullion, J. Schaefer, *J. Magn. Reson.* **1989**, *81*, 196–200.
[45] Y. Pan, T. Gullion, J. Schaefer, *J. Magn. Reson.* **1990**, *90*, 330–340.
[46] A. E. Bennett, J. H. Ok, S. Vega, et al., *J. Chem. Phys.* **1992**, *96*, 8624–8627.
[47] A. E. Bennett, C. M. Rienstra, J. M. Griffiths, et al., *J. Chem. Phys.* **1998**, *108*, 9463–9479.
[48] N. C. Nielsen, H. Bildso/e, H. J. Jakobsen, et al., *J. Chem. Phys.* **1994**, *101*, 1805–1812.
[49] R. Verel, M. Ernst, B. H. Meier, *J. Magn. Reson.* **2001**, *150*, 81–99.
[50] K. Takegoshi, S. Nakamura, T. Terao, *Chem. Phys. Lett.* **2001**, *344*, 631–637.
[51] K. Takegoshi, S. Nakamura, T. Terao, *J. Chem. Phys.* **2003**, *118*, 2325–2341.
[52] M. Baldus, A. T. Petkova, J. Herzfeld, et al., *Mol. Phys.* **1998**, *95*, 1197–1207.
[53] V. Chevelkov, K. Giller, S. Becker, et al., *J. Magn. Reson.* **2013**, *230*, 205–211.
[54] C. Shi, H. K. Fasshuber, V. Chevelkov, et al., *J. Biomol. NMR* **2014**, *59*, 15–22.
[55] J. Kowalewski, L. Mäler, *Nuclear Spin Relaxation in Liquids*, Taylor & Francis, New York, **2006**.
[56] J. J. Keeler, *Understanding NMR Spectroscopy*, 2nd ed., Wiley-VCH, Chichester, **2011**.
[57] Y. Jin, Y. Cui, T. Yuwen, *Magn. Reson. Lett.* **2025**, 200195.
[58] F. Bloch, *Phys. Rev.* **1946**, *70*, 460–474.

[59] I. Solomon, *Phys. Rev.* **1955**, *99*, 559–565.

[60] B. Vögeli, T. F. Segawa, D. Leitz, et al., *J. Am. Chem. Soc.* **2009**, *131*, 17215–17225.

[61] B. Vögeli, M. Friedmann, D. Leitz, et al., *J. Magn. Reson.* **2010**, *204*, 290–302.

[62] J. Orts, B. Vögeli, R. Riek, *J. Chem. Theory Comput.* **2012**, *8*, 3483–3492.

[63] D. Strotz, J. Orts, C. N. Chi, et al., *J. Chem. Theory Comput.* **2017**, *13*, 4336–4346.

[64] V. Kharchenko, M. Nowakowski, M. Jaremko, et al., *J. Biomol. NMR* **2020**, *74*, 707–716.

[65] M. Goldman, *J. Magn. Reson.* **1984**, *60*, 437–452.

[66] P. Schanda, M. Ernst, *Prog. Nuc. Magn. Reson. Spec.* **2016**, *96*, 1–46.

[67] M. W. F. Fischer, A. Majumdar, E. R. P. Zuiderweg, *Prog. Nuc. Magn. Reson. Spec.* **1998**, *33*, 207–272.

[68] M. Goldman, *J. Magn. Reson.* **2001**, *149*, 160–187.

[69] M. Goldman in *123*, Proc. Intl. School of Physics Enrico Fermi, Amsterdam, **1992**.

[70] R. Brüschweiler, D. A. Case, *Prog. Nuc. Magn. Reson. Spec.* **1994**, *26*, 27–58.

[71] K. Zumpfe, A. A. Smith, *Front. Mol. Biosci.* **2021**, *8*, 727553.

[72] P.-c. Chen, M. Hologne, O. Walker, et al., *J. Chem. Theory Comput.* **2018**, *14*, 1009–1019.

[73] M. L. Gill, R. A. Byrd, I. I. I. Arthur G. Palmer, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5839–5849.

[74] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.

[75] G. M. Clore, A. Szabo, A. Bax, et al., *J. Am. Chem. Soc.* **1990**, *112*, 4989–4991.

[76] D. M. LeMaster, *J Biomol NMR* **1995**, *6*, 366–374.

[77] M. Kovermann, P. Rogne, M. Wolf-Watz, *Q. Rev. Biophys.* **2016**, *49*, 1–43.

[78] J. Jeener, B. H. Meier, P. Bachmann, et al., *J. Chem. Phys.* **1979**, *71*, 4546–4553.

[79] N. L. Fawzi, J. Ying, R. Ghirlando, et al., *Nature* **2011**, *480*, 268–272.

[80] P. Vallurupalli, G. Bouvignies, L. E. Kay, *J. Am. Chem. Soc.* **2012**, *134*, 8148–8161.

[81] D. Marion, D. F. Gauto, I. Ayala, et al., *ChemPhysChem* **2019**, *20*, 276–284.

[82] D. Ban, C. A. Smith, B. L. de Groot, et al., *Arch. Biochem. Biophys.* **2017**, *628*, 81–91.

[83] F. Massi, E. Johnson, C. Wang, et al., *J. Am. Chem. Soc.* **2004**, *126*, 2247–2256.

[84] P. Rovó, *Solid State Nucl. Magn. Reson.* **2020**, *108*, 101665.

[85] H. Y. Carr, E. M. Purcell, *Phys. Rev.* **1954**, *94*, 630–638.

[86] S. Meiboom, D. Gill, *Rev. Sci. Instrum.* **1958**, *29*, 688–691.

[87] M. Tollinger, N. R. Skrynnikov, F. A. A. Mulder, et al., *J. Am. Chem. Soc.* **2001**, *123*, 11341–11352.

[88] D. F. Hansen, P. Vallurupalli, L. E. Kay, *J. Phys. Chem. B* **2008**, *112*, 5898–5904.

[89] R. Ishima, S. Bagby in *Modern Magnetic Resonance*, (Ed.: G. A. Webb), Springer, Cham, **2018**, 435–452.
[90] A. G. Palmer, C. D. Kroenke, J. Patrick Loria, *Meth. Enzymol.* **2001**, *339*, (Eds.: T. L. James, V. Dötsch, U. Schmitz), 204–238.
[91] P. J. Farber, A. Mittermaier, *Biophys. Rev.* **2015**, *7*, 191–200.
[92] J. P. Carver, R. E. Richards, *J. Magn. Reson.* **1972**, *6*, 89–105.
[93] D. G. Davis, M. E. Perlman, R. E. London, *J. Magn. Reson. Series B* **1994**, *104*, 266–275.
[94] H. M. McConnell, *J. Chem. Phys.* **1958**, *28*, 430–431.
[95] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed., Oxford University Press, Oxford, **2017**.
[96] D. C. Young, *Computational Chemistry*, Wiley-VCH, New York, **2001**.
[97] K. Lindorff-Larsen, S. Piana, K. Palmo, et al., *Proteins* **2010**, *78*, 1950–1958.
[98] J. A. Maier, C. Martinez, K. Kasavajhala, et al., *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
[99] C. Tian, K. Kasavajhala, K. A. A. Belfon, et al., *J. Chem. Theory Comput.* **2020**, *16*, 528–552.
[100] J. Huang, S. Rauscher, G. Nawrocki, et al., *Nat. Methods.* **2017**, *14*, 71–73.
[101] L. D. Schuler, X. Daura, W. F. Van Gunsteren, *J. Comput. Chem.* **2001**, *22*, 1205–1218.
[102] C. Oostenbrink, A. Villa, A. E. Mark, et al., *J. Comput. Chem.* **2004**, *25*, 1656–1676.
[103] N. Schmid, A. P. Eichenberger, A. Choutko, et al., *Eur. Biochem. J.* **2011**, *40*, 843–856.
[104] M. M. Reif, P. H. Hünenberger, C. Oostenbrink, *J. Chem. Theory Comput.* **2012**, *8*, 3705–3723.
[105] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
[106] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
[107] L. Verlet, *Phys. Rev.* **1967**, *159*, 98–103.
[108] W. C. Swope, H. C. Andersen, P. H. Berens, et al., *J. Chem. Phys.* **1982**, *76*, 637–649.
[109] R. W. Hockney, S. P. Goel, J. W. Eastwood, *J. Comput. Phys.* **1974**, *14*, 148–158.
[110] J. E. Basconi, M. R. Shirts, *J. Chem. Theory Comput.* **2013**, *9*, 2887–2899.
[111] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
[112] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, et al., *J. Chem. Phys.* **1984**, *81*, 3684–3690.
[113] M. Bernetti, G. Bussi, *J. Chem. Phys.* **2020**, *153*, 114107.
[114] S. E. Feller, Y. Zhang, R. W. Pastor, et al., *J. Chem. Phys.* **1995**, *103*, 4613–4621.

[115] R. C. Bernardi, M. C. R. Melo, K. Schulten, *BBA - Gen. Subjects* **2015**, *1850*, 872–877.

[116] D. Hamelberg, J. Mongan, J. A. McCammon, *J. Chem. Phys.* **2004**, *120*, 11919–11929.

[117] J. Song, Y. Li, C. Ji, et al., *Sci. Rep.* **2015**, *5*, 7906.

[118] Y. Miao, V. A. Feher, J. A. McCammon, *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595.

[119] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141–151.

[120] L. Wang, R. A. Friesner, B. J. Berne, *J. Phys. Chem. B* **2011**, *115*, 9431–9438.

[121] G. Bussi, *Mol. Phys.* **2014**, *112*, 379–384.

[122] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.

[123] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 020603.

[124] P. Robustelli, K. Kohlhoff, A. Cavalli, et al., *Structure* **2010**, *18*, 923–933.

[125] C. Camilloni, A. Cavalli, M. Vendruscolo, *J. Phys. Chem. B* **2013**, *117*, 1838–1843.

[126] J. J. Chou, D. A. Case, A. Bax, *J. Am. Chem. Soc.* **2003**, *125*, 8959–8966.

[127] L. E. Kay, *Nat. Struct. Biol.* **1998**, *5*, 513–517.

[128] K. Lindorff-Larsen, R. B. Best, M. A. DePristo, et al., *Nature* **2005**, *433*, 128–132.

[129] G. M. Clore, A. M. Gronenborn, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5891–5898.

[130] A. Bax, *Protein Sci.* **2003**, *12*, 1–16.

[131] O. F. Lange, N.-A. Lakomek, C. Farès, et al., *Science* **2008**, *320*, 1471–1475.

[132] G. M. Clore, *Mol. BioSyst.* **2008**, *4*, 1058–1069.

[133] G. M. Clore, J. Iwahara, *Chem. Rev.* **2009**, *109*, 4108–4139.

[134] A. J. Baldwin, L. E. Kay, *Nat. Chem. Biol.* **2009**, *5*, 808–814.

[135] M. Nilges, M. J. Macias, S. I. O'Donoghue, et al., *J. Mol. Biol.* **1997**, *269*, 408–422.

[136] W. Rieping, M. Habeck, B. Bardiaux, et al., *Bioinformatics* **2007**, *23*, 381–382.

[137] P. Güntert, *Eur. Biophys. J.* **2009**, *38*, 129–143.

[138] P. Güntert, L. Buchner, *J. Biomol. NMR* **2015**, *62*, 453–471.

[139] C. Schwieters, J. Kuszewski, G. Mariusclore, *Prog. Nuc. Magn. Reson. Spec.* **2006**, *48*, 47–62.

[140] J. Kuszewski, C. D. Schwieters, D. S. Garrett, et al., *J. Am. Chem. Soc.* **2004**, *126*, 6258–6273.

[141] A. Cavalli, C. Camilloni, M. Vendruscolo, *J. Chem. Phys.* **2013**, *138*, 094112.

[142] E. Papaleo, C. Camilloni, K. Teilum, et al., *PeerJ* **2018**, *6*, e5125.

[143] F. Palazzesi, M. K. Prakash, M. Bonomi, et al., *J. Chem. Theory Comput.* **2015**, *11*, 2–7.

[144] S. Rauscher, V. Gapsys, M. J. Gajda, et al., *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524.

[145] J. Henriques, C. Cragnell, M. Skepö, *J. Chem. Theory Comput.* **2015**, *11*, 3420–3431.

[146] M. H. Koch, P. Vachette, D. I. Svergun, *Q. Rev. Biophys.* **2003**, *36*, 147–227.
[147] S. Skou, R. E. Gillilan, N. Ando, *Nat. Protoc.* **2014**, *9*, 1727–1739.
[148] M. Bernetti, G. Bussi, *Eur. Phys. J. B* **2021**, *94*, 180.
[149] G. Jeschke, *Annu. Rev. Phys. Chem.* **2012**, *63*, 419–446.
[150] B. Schuler, *J. Nanobiotechnol.* **2013**, *11*, S2.
[151] B. Schuler, H. Hofmann, *Curr. Opin. Struct. Biol.* **2013**, *23*, 36–47.
[152] W. Boomsma, J. Ferkinghoff-Borg, K. Lindorff-Larsen, *PLoS Comput. Biol.* **2014**, *10*, e1003406.
[153] E. Ravera, L. Sgheri, G. Parigi, et al., *Phys. Chem. Chem. Phys.* **2016**, *18*, 5686–5701.
[154] M. Bonomi, G. T. Heller, C. Camilloni, et al., *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116.
[155] B. Söldner, K. Grohe, P. Neidig, et al., *J. Phys. Chem. Lett.* **2023**, *14*, 1725–1731.
[156] M. Bonomi, C. Camilloni, A. Cavalli, et al., *Science Adv.* **2016**, *2*, e1501177.
[157] P. Sormanni, D. Piovesan, G. T. Heller, et al., *Nat. Chem. Biol.* **2017**, *13*, 339–342.
[158] F. Kümmerer, S. Orioli, D. Harding-Larsen, et al., *J. Chem. Theory Comput.* **2021**, *17*, 5262–5275.
[159] G. M. Clore, C. D. Schwieters, *J. Am. Chem. Soc.* **2004**, *126*, 2923–2938.
[160] R. B. Fenwick, C. D. Schwieters, B. Vögeli, *J. Am. Chem. Soc.* **2016**, *138*, 8412–8421.
[161] K. Grohe, S. Patel, C. Hebrank, et al., *Structure* **2020**, *28*, 1024–1034.e3.
[162] D. F. Gauto, L. F. Estrozi, C. D. Schwieters, et al., *Nat. Commun.* **2019**, *10*, 2697.
[163] M. Lu, R. W. Russell, A. J. Bryer, et al., *Nat. Struct. Mol. Biol.* **2020**, *27*, 863–869.
[164] M. Zinke, K. A. A. Sachowsky, C. Öster, et al., *Nat. Commun.* **2020**, *11*, 5759.
[165] B. Reif, *Chem. Rev.* **2022**, *122*, 10019–10035.
[166] T. Le Marchand, T. Schubeis, M. Bonaccorsi, et al., *Chem. Rev.* **2022**, *122*, 9943–10018.
[167] R. Linser, B. Bardiaux, V. Higman, et al., *J. Am. Chem. Soc.* **2011**, *133*, 5905–5912.
[168] M. J. Knight, A. J. Pell, I. Bertini, et al., *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11095–11100.
[169] M. Huber, A. Böckmann, S. Hiller, et al., *Phys. Chem. Chem. Phys.* **2012**, *14*, 5239.
[170] R. Linser, B. Bardiaux, L. B. Andreas, et al., *J. Am. Chem. Soc.* **2014**, *136*, 11002–11010.
[171] L. B. Andreas, K. Jaudzems, J. Stanek, et al., *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9187–9192.
[172] S. K. Vasa, H. Singh, K. Grohe, et al., *Angew. Chem. Int. Ed.* **2019**, *58*, 5758–5762.

[173] M. G. Jain, D. Lalli, J. Stanek, et al., *J. Phys. Chem. Lett.* **2017**, *8*, 2399–2405.
[174] E. Nimerovsky, E. E. Najbauer, K. T. Movellan, et al., *J. Phys. Chem. Lett.* **2022**, *13*, 1540–1546.
[175] V. Agarwal, S. Penzel, K. Szekely, et al., *Angew. Chem. Int. Ed.* **2014**, *53*, 12253–12256.
[176] G. Hou, S. Yan, J. Trébosc, et al., *J. Magn. Reson.* **2013**, *232*, 18–30.
[177] X. Cheng, S. Jo, Y. Qi, et al., *Biophys. J.* **2015**, *108*, 1954–1962.
[178] A. De Simone, K. R. Mote, G. Veglia, *Biophys. J.* **2014**, *106*, 2566–2576.
[179] K. Mueller, *J. Magn. Reson. Series A* **1995**, *113*, 81–93.
[180] K. Grohe, E. Nimerovsky, H. Singh, et al., *Chem. Commun.* **2019**, *55*, 7899–7902.
[181] V. Ladizhansky, *Solid State Nucl. Magn. Reson.* **2009**, *36*, 119–128.
[182] M. Bak, J. T. Rasmussen, N. C. Nielsen, *J. Magn. Reson.* **2000**, *147*, 296–330.
[183] J. Jumper, R. Evans, A. Pritzel, et al., *Nature* **2021**, *596*, 583–589.
[184] R. B. Best, G. Hummer, *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
[185] K. Takemura, A. Kitao, *J. Phys. Chem. B* **2012**, *116*, 6279–6287.
[186] V. Chevelkov, Y. Xue, R. Linser, et al., *J. Am. Chem. Soc.* **2010**, *132*, 5015–5017.
[187] P. Rovó, C. A. Smith, D. Gauto, et al., *J. Am. Chem. Soc.* **2019**, *141*, 858–869.
[188] D. S. Cerutti, D. A. Case, *WIREs Comput. Mol. Sci.* **2019**, *9*, e1402.
[189] P. Ma, Y. Xue, N. Coquelle, et al., *Nat. Commun.* **2015**, *6*, 8361.
[190] D. F. Gauto, O. O. Lebedenko, L. M. Becker, et al., *J. Struc. Biol. X* **2023**, *7*, 100079.
[191] R. Linser, U. Fink, B. Reif, *J. Am. Chem. Soc.* **2010**, *132*, 8891–8893.
[192] N. Kulminskaya, S. K. Vasa, K. Giller, et al., *Chem. Commun.* **2015**, *52*, 268–271.
[193] N. Kulminskaya, S. K. Vasa, K. Giller, et al., *J. Biomol. NMR* **2015**, *63*, 245–253.
[194] A. Klein, S. K. Vasa, B. Söldner, et al., *J. Phys. Chem. Lett.* **2022**, *13*, 1644–1651.
[195] R. Linser, V. Chevelkov, A. Diehl, et al., *J. Magn. Reson.* **2007**, *189*, 209–216.
[196] J. P. Linge, M. Habeck, W. Rieping, et al., *J. Magn. Reson.* **2004**, *167*, 334–342.
[197] R. Boelens, T. M. G. Koning, R. Kaptein, *J. Mol. Struc.* **1988**, *173*, 299–311.
[198] M. M. Jolly, J. A. Jarvis, M. Carravetta, et al., *J. Biomol. NMR* **2017**, *69*, 197–205.
[199] S. C. Lovell, I. W. Davis, W. B. Arendall III, et al., *Proteins* **2003**, *50*, 437–450.
[200] M. J. Abraham, T. Murtola, R. Schulz, et al., *SoftwareX* **2015**, *1–2*, 19–25.
[201] U. Essmann, L. Perera, M. L. Berkowitz, et al., *J. Chem. Phys.* **1995**, *103*, 8577–8593.
[202] P. Rovó, R. Linser, *ChemPhysChem* **2018**, *19*, 34–39.
[203] J. Kotschy, B. Söldner, H. Singh, et al., *Angew. Chem. Int. Ed.* **2024**, *63*, e202313947.
[204] J. Kotschy, B. Söldner, H. Singh, et al., *Angew. Chem.* **2024**, *136*, e202313947.
[205] D. Ringe, G. A. Petsko, *Meth. Enzymol.* **1986**, *131*, 389–433.

[206] C. Gerlach, M. Smolinski, H. Steuber, et al., *Angew. Chem. Int. Ed.* **2007**, *46*, 8511–8514.
[207] G. E. Bacon, N. A. Curry, S. A. Wilson, *Proc. R. Soc. Lond. A* **1964**, *279*, 98–110.
[208] N. Fischer, A. L. Konevega, W. Wintermeyer, et al., *Nature* **2010**, *466*, 329–333.
[209] C. W. Murray, M. L. Verdonk, *J. Comput. Aided Mol. Des.* **2002**, *16*, 741–753.
[210] M. Nazaré, H. Matter, D. W. Will, et al., *Angew. Chem. Int. Ed.* **2012**, *51*, 905–911.
[211] C.-e. A. Chang, W. Chen, M. K. Gilson, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1534–1539.
[212] G. Klebe, *Nat. Rev. Drug Discov.* **2015**, *14*, 95–110.
[213] L. Chan, G. M. Morris, G. R. Hutchison, *J. Chem. Theory Comput.* **2021**, *17*, 2099–2106.
[214] V. M. Krishnamurthy, B. R. Bohall, V. Semetey, et al., *J. Am. Chem. Soc.* **2006**, *128*, 5802–5812.
[215] L. Mariño Pérez, F. S. Ielasi, L. M. Bessa, et al., *Nature* **2022**, *602*, 695–700.
[216] K. Wüthrich, G. Wagner, *FEBS Lett.* **1975**, *50*, 265–268.
[217] M. Akke, U. Weininger, *J. Phys. Chem. B* **2023**, *127*, 591–599.
[218] S. K. Vasa, H. Singh, P. Rovó, et al., *J. Phys. Chem. Lett.* **2018**, *9*, 1307–1311.
[219] B. A. Grzybowski, A. V. Ishchenko, C.-Y. Kim, et al., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1270–1273.
[220] B. S. Avvaru, S. A. Busby, M. J. Chalmers, et al., *Biochemistry* **2009**, *48*, 7365–7372.
[221] D. F. Gauto, P. Macek, A. Barducci, et al., *J. Am. Chem. Soc.* **2019**, *141*, 11183–11195.
[222] A. D. Gossert, W. Jahnke, *Prog. Nuc. Magn. Reson. Spec.* **2016**, *97*, 82–125.
[223] H. Singh, C. K. Das, S. K. Vasa, et al., *Angew. Chem. Int. Ed.* **2020**, *59*, 22916–22921.
[224] H. Singh, S. K. Vasa, H. Jangra, et al., *J. Am. Chem. Soc.* **2019**, *141*, 19276–19288.
[225] A. L. Breeze, *Prog. Nuc. Magn. Reson. Spec.* **2000**, *36*, 323–372.
[226] T.-L. Hwang, S. Mori, A. J. Shaka, et al., *J. Am. Chem. Soc.* **1997**, *119*, 6203–6204.
[227] A. G. I. Palmer, *Chem. Rev.* **2004**, *104*, 3623–3640.
[228] D. F. Gauto, A. Hessel, P. Rovó, et al., *Solid State Nucl. Magn. Reson.* **2017**, *87*, 86–95.
[229] P. Li, K. M. Merz, *J. Chem. Inf. Model.* **2016**, *56*, 599–604.
[230] S. Glöckner, K. Ngo, C. P. Sager, et al., *ACS Chem. Biol.* **2020**, *15*, 675–685.
[231] F. Hoffmann, F. A. A. Mulder, L. V. Schäfer, *J. Chem. Phys.* **2020**, *152*, 084102.
[232] W. Kabsch, *Acta Crystallogr. D* **2010**, *66*, 125–132.
[233] P. Evans, *Acta Crystallogr. D* **2006**, *62*, 72–82.

[234] P. D. Adams, P. V. Afonine, G. Bunkóczi, et al., *Acta Crystallogr. D* **2010**, *66*, 213–221.
[235] P. Emsley, B. Lohkamp, W. G. Scott, et al., *Acta Crystallogr. D* **2010**, *66*, 486–501.
[236] X. Guo, D. He, L. Huang, et al., *Comput. Theor. Chem.* **2012**, *995*, 17–23.
[237] M. Bieri, P. R. Gooley, *BMC Bioinform.* **2011**, *12*, 421.
[238] J. Iwahara, J. M. Wojciak, R. T. Clubb, *J. Biomol. NMR* **2001**, *19*, 231–241.
[239] K. Ogura, H. Terasawa, F. Inagaki, *J. Biomol. NMR* **1996**, *8*, 492–498.
[240] C. Zwahlen, P. Legault, S. J. F. Vincent, et al., *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.
[241] P. Rovó, R. Linser, *J. Phys. Chem. B* **2017**, *121*, 6117–6130.
[242] V. Kurauskas, S. A. Izmailov, O. N. Rogacheva, et al., *Nat. Commun.* **2017**, *8*, 145.
[243] T. Zinkevich, V. Chevelkov, B. Reif, et al., *J. Biomol. NMR* **2013**, *57*, 219–235.
[244] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779–815.
[245] H. G. Daronkola, B. Söldner, H. Singh, et al., *ChemBioChem* **2024**, *25*, e202400057.
[246] J. K. Lanyi, *Bacteriol. Rev.* **1974**, *38*, 272–290.
[247] J. Peters, R. Oliva, A. Caliò, et al., *Chem. Rev.* **2023**, *123*, 13441–13488.
[248] C. Brininger, S. Spradlin, L. Cobani, et al., *Semin. Cell Dev. Biol.* **2018**, *84*, 158–169.
[249] P. M. Tessier, A. M. Lenhoff, *Curr. Opin. Biotechnol.* **2003**, *14*, 512–516.
[250] A. C. Dumetz, A. M. Snellinger-O'brien, E. W. Kaler, et al., *Protein Sci.* **2007**, *16*, 1867–1877.
[251] B. Guo, S. Kao, H. McDonald, et al., *J. Cryst. Growth* **1999**, *196*, 424–433.
[252] G. Graziano, A. Merlino, *BBA - Proteins Proteomics* **2014**, *1844*, 850–858.
[253] S. P. Kennedy, W. V. Ng, S. L. Salzberg, et al., *Genome Res.* **2001**, *11*, 1641–1650.
[254] S. Paul, S. K. Bag, S. Das, et al., *Genome Biol.* **2008**, *9*, R70.
[255] A. Siglioccolo, A. Paiardini, M. Piscitelli, et al., *BMC Struct. Biol.* **2011**, *11*, 50.
[256] D. Madern, G. Zaccai, *Biochimie* **2004**, *86*, 295–303.
[257] M. S. Date, B. N. Dominy, *Commun. Comput. Phys.* **2013**, *13*, 90–106.
[258] P. A. Fields, *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **2001**, *129*, 417–431.
[259] K. Teilum, J. G. Olsen, B. B. Kragelund, *BBA - Proteins Proteomics* **2011**, *1814*, 969–976.
[260] P. Arroyo-Mañez, D. E. Bikiel, L. Boechi, et al., *BBA - Proteins Proteomics* **2011**, *1814*, 1054–1064.
[261] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, et al., *Nature* **2005**, *438*, 117–121.
[262] L. Y. P. Luk, E. J. Loveridge, R. K. Allemann, *Phys. Chem. Chem. Phys.* **2015**, *17*, 30817–30827.
[263] P.-h. Wang, R. B. Best, J. Blumberger, *J. Am. Chem. Soc.* **2011**, *133*, 3548–3556.

[264] M. Mevarech, F. Frolow, L. M. Gloss, *Biophys. Chem.* **2000**, *86*, 155–164.
[265] M. Wong, G. Khirich, J. P. Loria, *Biochemistry* **2013**, *52*, 6548–6558.
[266] M. Wong, G. Khirich, J. P. Loria, *Biochemistry* **2013**, *52*, 7160–7160.
[267] B. Mateos, O. Millet, *J. Magn. Reson. Open* **2022**, *12–13*, 100072.
[268] Q. Li, M. Li, C. Li, et al., *FEBS Lett.* **2021**, *595*, 521–531.
[269] M. Tehei, D. Madern, C. Pfister, et al., *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14356–14361.
[270] B. Hess, N. F. A. van der Vegt, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13296–13300.
[271] D. Sitkoff, K. A. Sharp, B. Honig, *Biophys. Chem.* **1994**, *51*, 397–403.
[272] K. A. Sharp, A. Nicholls, R. F. Fine, et al., *Science* **1991**, *252*, 106–109.
[273] K. Ali, A.-H. A. Shah, S. Bilal, et al., *Colloids Surf. A* **2009**, *337*, 194–199.
[274] J. H. Roh, M. Tyagi, P. Aich, et al., *Soft Matter* **2015**, *11*, 8741–8745.
[275] J. H. Roh, *Macromol. Chem. Phys.* **2016**, *217*, 256–265.
[276] G. Zaccai, F. Cendrin, Y. Haik, et al., *J. Mol. Biol.* **1989**, *208*, 491–500.
[277] X. Tadeo, B. López-Méndez, T. Trigueros, et al., *PLoS Biol.* **2009**, *7*, e1000257.
[278] D. Madern, C. Ebel, G. Zaccai, *Extremophiles* **2000**, *4*, 91–98.
[279] G. Ortega, T. Diercks, O. Millet, *Chem. Biol.* **2015**, *22*, 1597–1607.
[280] P. Herrero-Alfonso, A. Pejenaute, O. Millet, et al., *Protein Sci.* **2024**, *33*, e5003.
[281] H. Geraili Daronkola, A. Vila Verde, *Biophys. J.* **2023**, *122*, 2577–2589.
[282] A. H. Elcock, J. A. McCammon, *J. Mol. Biol.* **1998**, *280*, 731–748.
[283] F. Frolow, M. Harel, J. L. Sussman, et al., *Nat. Struct. Biol.* **1996**, *3*, 452–458.
[284] D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, et al., *Amber 2018*, University of California, San Francisco, **2018**.
[285] G. Janson, G. Valdes-Garcia, L. Heo, et al., *Nat. Commun.* **2023**, *14*, 774.
[286] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, et al., *Nat. Rev. Phys.* **2021**, *3*, 422–440.
[287] G. K. Meghwanshi, N. Kaur, S. Verma, et al., *Biotechnol. Appl. Biochem.* **2020**, *67*, 586–601.
[288] G. Rossino, M. S. Robescu, E. Licastro, et al., *Chirality* **2022**, *34*, 1403–1418.
[289] S. J. Horn, G. Vaaje-Kolstad, B. Westereng, et al., *Biotechnol. Biofuels* **2012**, *5*, 45.
[290] A. Bangaru, K. A. Sree, C. Kruthiventi, et al. in *Bio-Clean Energy Technologies*, *1*, (Eds.: P. Chowdhary, N. Khanna, S. Pandit, et al.), Springer, Singapore, **2022**, 81–112.
[291] I. Dundas, *Extremophiles* **1998**, *2*, 375–377.
[292] B. M. Rode, *Peptides* **1999**, *20*, 773–786.
[293] L. M. Longo, J. Lee, M. Blaber, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2135–2139.
[294] L. M. Longo, M. Blaber, *Front. Microbiol.* **2014**, *4*, 418.
[295] W. F. Vranken, W. Boucher, T. J. Stevens, et al., *Proteins* **2005**, *59*, 687–696.

[296] J. P. Loria, M. Rance, A. G. Palmer, *J. Am. Chem. Soc.* **1999**, *121*, 2331–2332.
[297] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577–2637.
[298] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, et al., *J. Chem. Phys.* **1983**, *79*, 926–935.
[299] W. D. Cornell, P. Cieplak, C. I. Bayly, et al., *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
[300] I. S. Joung, T. E. Cheatham, *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
[301] S. Kashefolgheta, A. Vila Verde, *Phys. Chem. Chem. Phys.* **2017**, *19*, 20593–20607.
[302] H. Geraili Daronkola, A. Vila Verde, *Biophys. J.* **2021**, *120*, 2746–2762.
[303] A. T. P. Carvalho, A. F. S. Teixeira, M. J. Ramos, *J. Comput. Chem.* **2013**, *34*, 1540–1548.
[304] M. B. Peters, Y. Yang, B. Wang, et al., *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947.
[305] P. Li, B. P. Roberts, D. K. Chakravorty, et al., *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748.
[306] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
[307] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38, 27–28.
[308] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
[309] D. L. Theobald, P. A. Steindel, *Bioinformatics* **2012**, *28*, 1972–1979.
[310] D. L. Theobald, D. S. Wuttke, *PLoS Comput. Biol.* **2008**, *4*, e43.
[311] J. Wang, R. M. Wolf, J. W. Caldwell, et al., *J. Comput. Chem.* **2004**, *25*, 1157–1174.
[312] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[313] A. C. Warden, M. Williams, T. S. Peat, et al., *Nat. Commun.* **2015**, *6*, 1–10.
[314] S. Arai, Y. Yonezawa, N. Okazaki, et al., *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71*, 541–554.
[315] J. Qvist, G. Ortega, X. Tadeo, et al., *J. Phys. Chem. B* **2012**, *116*, 3436–3444.
[316] C. Hartmann, J. C. Latorre, G. Ciccotti, *Eur. Phys. J. Spec. Top.* **2011**, *200*, 73–89.
[317] P. Debye, E. Hückel, *Phys. Z.* **1923**, *24*, 185–206.
[318] B. Söldner, H. Singh, E. Akoury, et al., *Angew. Chem. Int. Ed.* **2025**, *64*, e202509394.
[319] B. Söldner, H. Singh, E. Akoury, et al., *Angew. Chem.* **2025**, *64*, e202509394.
[320] D. Neri, M. Billeter, G. Wider, et al., *Science* **1992**, *257*, 1559–1563.
[321] E. Akoury, M. D. Mukrasch, J. Biernat, et al., *Protein Sci.* **2016**, *25*, 1010–1020.
[322] A. Chowdhury, D. Nettels, B. Schuler, *Annu. Rev. Biophys.* **2023**, *52*, 433–462.
[323] B. Fu, M. Vendruscolo in *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, (Eds.: I. C. Felli, R. Pierattelli), Springer, Cham, **2015**, 35–48.

[324] A. S. Holehouse, B. B. Kragelund, *Nat. Rev. Mol. Cell. Biol.* **2024**, *25*, 187–211.
[325] J. Kragelj, T. Orand, E. Delaforge, et al., *Biomolecules* **2021**, *11*, 1204.
[326] S. Medina Gomez, I. Visco, F. Merino, et al., *Angew. Chem. Int. Ed.* **2024**, *63*, e202403941.
[327] T. R. Sisk, P. Robustelli, *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2313360121.
[328] L. Mollica, L. M. Bessa, X. Hanoulle, et al., *Front. Mol. Biosci.* **2016**, *3*, 52.
[329] B. R. Sabari, A. Dall'Agnese, A. Boija, et al., *Science* **2018**, *361*, eaar3958.
[330] F. Jonas, Y. Navon, N. Barkai, *Nat. Rev. Genet.* **2025**, *26*, 424–435.
[331] M. Müller, K.-P. Hopfner, G. Witte, *FEBS Lett.* **2015**, *589*, 45–51.
[332] D. Kalia, G. Merey, S. Nakayama, et al., *Chem. Soc. Rev.* **2012**, *42*, 305–341.
[333] G. Witte, S. Hartung, K. Büttner, et al., *Mol. Cell* **2008**, *30*, 167–178.
[334] U. Römling, *Sci. Signal.* **2008**, *1*, pe39.
[335] J. He, W. Yin, M. Y. Galperin, et al., *Nucleic Acids Res.* **2020**, *48*, 2807–2829.
[336] M. J. Conroy, A. Durand, D. Lupo, et al., *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1213–1218.
[337] R. M. Corrigan, I. Campeotto, T. Jeganathan, et al., *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 9084–9089.
[338] A. M. Bandera, J. Bartho, K. Lammens, et al., *Nucleic Acids Res.* **2021**, *49*, 10166–10177.
[339] K. Forchhammer, *FEMS Microbiol. Rev.* **2004**, *28*, 319–333.
[340] L. F. Huergo, G. Chandra, M. Merrick, *FEMS Microbiol. Rev.* **2013**, *37*, 251–283.
[341] K. Forchhammer, J. Lüddecke, *FEBS J.* **2016**, *283*, 425–437.
[342] C. Rajendran, E. C. M. Gerhardt, S. Bjelic, et al., *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18972–18976.
[343] C. Palanca, V. Rubio, *Environ. Microbiol. Rep.* **2017**, *9*, 290–299.
[344] Y. Xu, P. D. Carr, T. Huber, et al., *Eur. J. Biochem.* **2001**, *268*, 2028–2037.
[345] Y. Xu, E. Cheah, P. D. Carr, et al., *J. Mol. Biol.* **1998**, *282*, 149–165.
[346] S. Helfmann, W. Lü, C. Litz, et al., *J. Mol. Biol.* **2010**, *402*, 165–177.
[347] S. Maier, P. Schleberger, W. Lü, et al., *PLoS One* **2011**, *6*, e26327.
[348] I. Campeotto, Y. Zhang, M. G. Mladenov, et al., *J. Biol. Chem.* **2015**, *290*, 2888–2901.
[349] P. H. Choi, K. Sureka, J. J. Woodward, et al., *MicrobiologyOpen* **2015**, *4*, 361–374.
[350] D. D. Boehr, H. J. Dyson, P. E. Wright, *Chem. Rev.* **2006**, *106*, 3055–3079.
[351] R. Sprangers, L. E. Kay, *Nature* **2007**, *445*, 618–622.
[352] K. W. East, J. C. Newton, U. N. Morzan, et al., *J. Am. Chem. Soc.* **2020**, *142*, 1348–1358.
[353] Y. Toyama, R. W. Harkness, L. E. Kay, *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2203172119.
[354] H. Singh, C. K. Das, B. C. Buchmuller, et al., *Nucleic Acids Res.* **2023**, *51*, 6495–6506.

[355] J. F. Schmuck, J. Borggräfe, M. Etzkorn, *Nat. Commun.* **2024**, *15*, 5145.
[356] T. Schubeis, T. Le Marchand, C. Daday, et al., *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 21014–21021.
[357] D. F. Gauto, P. Macek, D. Malinverni, et al., *Nat. Commun.* **2022**, *13*, 1927.
[358] S. Medina Gomez, T. I. Gonzalez, S. K. Vasa, et al., *Angew. Chem. Int. Ed.* **2024**, *63*, e202411472.
[359] Y. Shen, F. Delaglio, G. Cornilescu, et al., *J Biomol NMR* **2009**, *44*, 213–223.
[360] X. Daura, K. Gademann, B. Jaun, et al., *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.
[361] K. H. R. MacPherson, Y. Xu, E. Cheah, et al., *Acta Cryst. D* **1998**, *54*, 996–998.
[362] R. L. J. Keller, ETH Zurich, **2005**, 147 S.
[363] F. A. Mulder, N. R. Skrynnikov, B. Hon, et al., *J. Am. Chem. Soc.* **2001**, *123*, 967–975.
[364] J. C. Phillips, D. J. Hardy, J. D. C. Maia, et al., *J. Chem. Phys.* **2020**, *153*, 044130.
[365] S. Jo, T. Kim, V. G. Iyer, et al., *J. Comput. Chem.* **2008**, *29*, 1859–1865.
[366] J. Lee, X. Cheng, J. M. Swails, et al., *J. Chem. Theory Comput.* **2016**, *12*, 405–413.
[367] G. J. Martyna, D. J. Tobias, M. L. Klein, *J. Chem. Phys.* **1994**, *101*, 4177–4189.
[368] Y. T. Pang, Y. Miao, Y. Wang, et al., *J. Chem. Theory Comput.* **2017**, *13*, 9–19.
[369] J.-J. Feng, J.-N. Chen, W. Kang, et al., *J. Chem. Theory Comput.* **2021**, *17*, 4614–4628.
[370] M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, *52*, 7182–7190.
[371] S. Nosé, *J. Chem. Phys.* **1984**, *81*, 511–519.
[372] J. S. Papadopoulos, R. Agarwala, *Bioinformatics* **2007**, *23*, 1073–1079.
[373] D. Franke, M. V. Petoukhov, P. V. Konarev, et al., *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225.
[374] C. D. Putnam, M. Hammel, G. L. Hura, et al., *Q. Rev. Biophys.* **2007**, *40*, 191–285.
[375] H. D. T. Mertens, D. I. Svergun, *J. Struct. Biol.* **2010**, *172*, 128–141.
[376] K. Manalastas-Cantos, P. V. Konarev, N. R. Hajizadeh, et al., *J. Appl. Crystallogr.* **2021**, *54*, 343–355.
[377] T.-C. Lin, L. Engelhard, B. Söldner, et al., *Adv. Sci.* **2024**, *11*, 2307930.
[378] H. Aucharova, A. Klein, S. Medina Gomez, et al., *Chem. Commun.* **2024**, *60*, 3083–3086.

6.4 Eidesstattliche Erklärung (Affidavit)

Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Ort, Datum
(Place, date)

Unterschrift
(Signature)