Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Development and Application of Novel NMR Methodology for Elucidation of Protein Structure and Dynamics

Kristof Grohe

aus

Göttingen

2020

Erklärung
Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Rasmus Jan Linser betreut.

Eidesstattliche Versicherung
Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 18.07.2020

Kristof Grohe

Dissertation eingereicht am: 03.08.2020

1. Gutachter: Prof. Dr. Rasmus Jan Linser

2. Gutachter: Prof. Dr. Konstantin Karaghiosoff

Mündliche Prüfung am: 17.09.2020

Danksagung

Mein besonderer Dank gilt Prof. Dr. Rasmus Linser, Anfang des Jahres 2014 trat ich in seine Forschungsgruppe im Rahmen einer Masterarbeit kombiniert mit einem Forschungs-Praktikum ein. Nicht nur sein Wissenschaftliches Format und die NMR-Spektroskopie an sich, sondern insbesondere seine hingebungsvolle Betreuung ließen mir keinen Zweifel, dass ich auch meine Doktorandenzeit mit ihm verbringen will. Ich sollte es nicht bereuen, über die Jahre erwuchs sowohl wissenschaftlich als auch privat ein enges Verhältnis. Sein Führungsstiel war nie durch Zwang oder ähnliches geprägt, sondern ein Führen durch Vorbild. Er hat den Rahmen dafür geschaffen, dass wissenschaftliche Hürden immer dadurch gemeistert werden konnten, dass jeder im Team seine Stärken optimal einbringen konnten. Hierbei hat er immer neben der organisatorischen, eine aktive Rolle gespielt. Vielen Dank.

Generell herrschte in unserer Arbeitsgruppe immer ein sehr gutes Klima. Hierbei möchte ich meinem Kollegen Suresh Kumar Vasa danken, der durch seine Hilfsbereitschaft, verbunden mit seinem enormen und unersetzlichen Fachwissen, für alle Gruppenmitglieder eine Bereicherung war und ist. In diesem Zusammenhang danke ich ebenso Petra Rovó, die mich während meiner Masterarbeit sehr gut betreut hat und auch danach immer ein offenes Ohr für mich hatte. Ein sehr inspirierender und angenehmer Kollege und Kooperationspartner war Himanshu Singh. Er hat durch seine aufgeschlossene und freundliche Art die fruchtvolle wissenschaftliche Zusammenarbeit beflügelt. Ebenfalls möchte ich meinen Doktoranden-Kolleginnen und Kollegen Alex, Katja und Sara danken. Nachdem sie dazugestoßen sind, ist das Arbeitsklima noch heiterer und fröhlicher geworden.

Ich möchte auch den Kollaborationspartnern aus aller Welt danken. Durch ihr Engagement und ihre Brillanz, konnten wir gemeinsam unsere Projekte zum erfolgreichen Abschluss bringen.

Mein Dank gilt auch meinen Eltern. Sie haben mir in meiner Kindheit und Jugend, durch interessante Gespräche und durch gemeinsame Unternehmungen, meine naturwissenschaftlich-technische Begeisterung geschenkt und mich Zeit meines Lebens bestmöglich unterstützt.

Ich danke meiner Freundin Bianca, ohne sie wäre ich nicht da wo ich bin. Sie musste viel zurückstecken und hat mich immer in allen Ebenen unterstützt.

Parts of this thesis are published in the following academic journals:

Kristof Grohe, Snehal Patel, Cornelia Hebrank, Sara Medina, Alexander Klein, Petra Rovo´, Suresh K. Vasa, Himanshu Singh, Beat Vögeli, Lars V. Schäfer, Rasmus Linser, Protein Motional Details Revealed by Complementary Structural Biology Techniques, *Structure*, 28, 1–11, **2020**, Elsevier.

Kristof Grohe, Evgeny Nimerovsky, Himanshu Singh, Suresh K. Vasa, Benedikt Söldner, Beat Vögeli, Chad M. Rienstra, Rasmus Linser, Exact distance measurements for structure and dynamics in solid proteins by fast-magic-anglespinning NMR, *Chem. Commun.*,55, 7899-7902, **2019**, The Royal Society of Chemistry.

Kristof Grohe, Kumar Tekwani Movellan, Suresh Kumar Vasa,·Karin Giller, Stefan Becker, Rasmus Linser, Non-equilibrium hydrogen exchange for determination of H-bond strength and water accessibility in solid proteins*, J. Biomol. NMR*, 68, 7–17, **2017**, Springer.

Suresh K. Vasa, Himanshu Singh, **Kristof Grohe**, and Rasmus Linser, Assessment of a Large Enzyme–Drug Complex by Proton-Detected Solid-State NMR Spectroscopy without Deuteration, *Angew. Chem. Int. Ed*., 58, 5758–5762, **2019**, Wiley.

Himanshu Singh, Chandan K. Das, Suresh K. Vasa, **Kristof Grohe**, Lars V. Schäfer, Rasmus Linser, The active site of a prototypical "rigid" drug target is marked by extensive conformational dynamics*, accepted in Angew. Chem. Int. Ed.,* **2020**, Wiley, DOI: 10.1002/anie.202009348.

Jonathan Cramer, Johannes Schiebel, Tobias Wulsdorf, **Kristof Grohe**, Eszter Eva Najbauer, Frederik R. Ehrmann, Nedyalka Radeva, Nina Zitzer, Uwe Linne, Rasmus Linser, Andreas Heine, Gerhard Klebe, A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring, *Angew. Chem. Int. Ed.*, 56, 1908–1913, **2017**, Wiley.

Summary

The design of a novel drug is a creative act. The difference between a researcher and an artist, besides the pursuit of a particular goal, is that his/her creativity is based on scientific knowledge and technology. Until today, drug discovery was mainly dominated by trial and error, based on empirically derived rules. Today, the "trial and error" approach is being replaced by a conscious design based on improved predictions. These improvements in the prediction of the structure of a drug molecule are mainly based on the increasing knowledge about molecular mechanisms. [1] This increase of mechanistic knowledge is also caused by the progress of structure elucidation techniques like NMR-spectroscopy or X-ray crystallography. So far, our understanding of molecular mechanisms is based on the justified principle: Structure determines function. The number of deposited protein structures grows exponentially. Nevertheless, the ratio between solved and unsolved human protein structures is in the single-digit percentage range. [2] This relative number is even lower for proteins that are insoluble and at the same time amorphous in the solid-state, such as fibrils or membrane proteins. The structure of such proteins of great importance is not accessible by X-ray crystallography or liquid-state NMR-spectroscopy. In order to close this gap, a new NMR methodology has been developed over the last two decades, the proton-detected fast-magic-angle-spinning solid-state NMR-spectroscopy. This new method enables the structure-elucidation of such amorphous and insoluble proteins. In this context, Bernd Reif et al., Rasmus Linser et al., Guido Pintacuda et al., and others did the pioneering work. In this line, an important objective of this thesis was to contribute to the continuous development of this methodology. The key to well-resolved protein structure from NMR-spectroscopy is precise distance restraints. However, till date, solid state NMR-spectroscopy has only been able to provide qualitative restraints, grouping the internuclear distance as close, medium, or far. Opposed to the general picture, with the help of Suresh K. Vasa, Evgeny Nimerovsky, Himanshu Singh, Beat Vögeli, and others, I developed a user-friendly approach to determine accurate distance restraints in solid-state NMR. Hereby, all site-specific errors that occur during magnetization transfer are addressed by an integrated approach. The approximations to be made are carefully validated by numerical simulations. Further, in order to address challenging protein targets where the assignment might be ambiguous and incomplete, I have performed the first kinetic hydrogen-deuterium exchange measurements in solid-state NMR-spectroscopy in collaboration with Suresh K. Vasa, Himanshu Singh, and others. The method reports on the over-all positioning of exchangeable protons within the protein by information on the presence in the hydrophobic core or on the hydrophilic surface and on the presence and strength of structural hydrogen bonds. Therefore, this information can be used for assignment purposes and for structure determination. Hereby, it was crucial to be able to separate exchange hindrance due to hydrophobic shielding from the hindrance due to hydrogen bonding. Experimental site-specific information about hydrogen bond strength is valuable for determination of protein structure and dynamics simultaneously. In addition to a well-defined distance restraint, a hydrogen bond also reports on the rigidity of structural elements. In particular, the stabilization of functionally important loop tips by sidechain-to-backbone hydrogen bonds is investigated in this work.

While the interplay of protein structure and function is well accepted and understood, the current state of understanding how local motion contribute to functional mechanisms leaves much to be discovered. In this context, NMR spectroscopy is one of the techniques of choice. Nowadays detailed information on directional protein motions is provided primarily by the theoretically based MD simulations. Another part of this work is dedicated to advance and apply NMR-spectroscopic methods, which have been developed to determine these directional dynamics and lead towards interpretations of the biological function. Along these lines, in liquid-state protein NMR, Vögeli et al. recently developed an approach that provides distance restraints with such accuracy that differences between distance restraints and average atomic positions are no longer associated with experimental error but with protein dynamics. With the help of Cornelia Hebrank, Snehal Patel, and Lars V. Schäfer, this method was applied to the protein binding domain SH3, yielding interesting mechanistic insights regarding ligand binding. For the first time, the novel method was evaluated using state-of-the-art MD simulations and further compared with the dynamics data using other NMR approaches. In addition to providing mechanistical insides for the protein SH3, the value of this new approach, which is the first experimental method for the determination of spatial dynamics, is emphasized.

In general, the main objective of this work was to develop applicable methods for the benefit of other researchers and thus for the benefit of science and humanity.

Table of Contents

1 Introduction

1.1 Example of an Important Protein Structure Determined by Solid-State NMR

A protein structure can reveal the biological function, including binding affinity, [3] and thus, the molecular structure of a potential drug candidate can be approximated. By contrast, if the information about protein structure is leaking, any drug discovery, if possible at all, requires extensive, time-consuming, and costly target screening. With a known structure of the drug target, however, time and costs can often be reduced to a minimum. In addition, knowledge of protein structure and function can enable successful drug design where no medical cure has been in sight so far. Over the last two decades, a new NMR methodology, the proton-detected fast-magic-angle-spinning solid-state NMR, has been developed that enables the investigation of proteins such as fibrils, membrane proteins, or other insoluble amorphous samples [4] that cannot be investigated by X-ray, cryo-EM, conventional solid-state NMR, or liquid-state NMR.

An example of a recent breakthrough is the structural elucidation of the outer-membrane protein G (ompG) in the environment of lipid bilayers. [5] OmpG was found to be able to import mono-, di- and tri-sacharides with a pH-dependent change from open and close conformation.

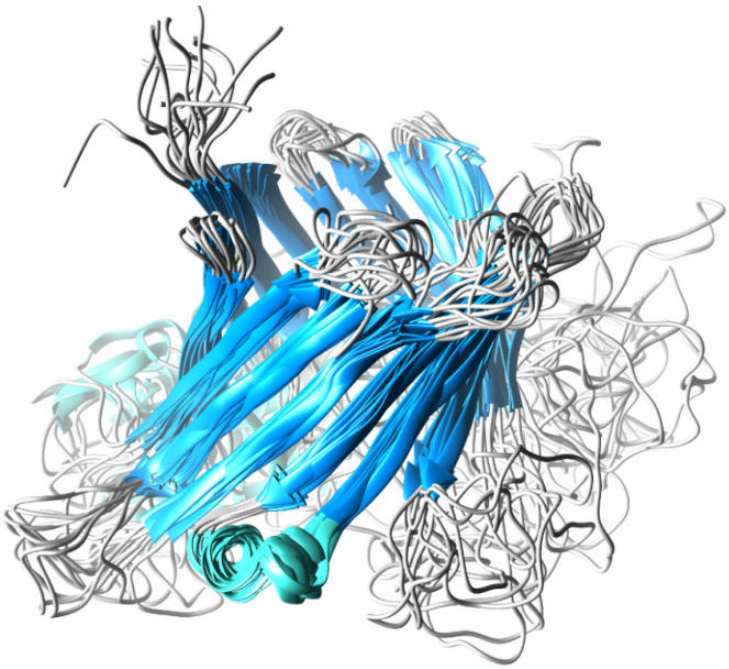


Figure 1: Structure of the outer membrane protein G (OmpG), elucidated by proton-detected fast-magic angle-spinning solid-state NMR spectroscopy (PDB identifier: 5MWV). [5] β-sheets are depicted in blue, α-helices in green, and loops are shown in light gray.

The understanding of the functional mechanisms is constantly revised and reshaped. The influence of local fluctuations on different timescales on larger allosteric motions or breathing motions is more and more confirmed. Understanding these influences of site-specific protein dynamics leads to significant improvements in mechanistic understanding and thus to faster progress in drug discovery for De-Novo. In order to determine these site-specific protein dynamics, NMR spectroscopy is able to provide unique tools. [6]

1.2 Fundamental Biochemical Background

1.2.1 Protein structural motives[7]

A living cell is the most diverse and sophisticated chemical production plant. An unimaginable variety of different chemicals, each dead by itself, together form life in its interactions. These self-organized, prosperous, and productive molecular interactions are beginning to be understood in the field of supramolecular chemistry; however, the state of knowledge is more similar to scratching the surface of Mother Nature's functional principles. The most diverse class of molecules, the proteins, determine the processes of life. Proteins serve as molecular machines that fulfill innumerable functions. Examples are enzymatic reactions in the metabolic pathways through the synthesis of natural products, xenobiotics, or even other proteins; providing mechanisms for information transfer across complex signaling pathways; regulation and assurance of active and passive molecular transport and the structural integrity of the cell. In short, proteins are involved in almost every process of the living cell and usually play a decisive role. The functional variety of proteins is made possible by the structural arrangement in various specific three-dimensional structures (folds), wherein, the primary structure is represented by the sequence of amino acids connected throughout by amide bonds (peptide bonds), forming a polymer chain. The three-dimensional arrangement of this chain is called the secondary structure. The secondary structure shows two main elements: The alpha-helix and the beta-sheet (**Figure 2**). The β-sheet consists of two adjacent peptide strands in either parallel or antiparallel orientation (**Figure 2 B**). The α-helix has a right-handed helical chirality, at every turn, the sequence continues with 3.6 amino acids per 5.1 Å length. Each amino acid of the helix forms a backbone hydrogen bond with the third nearest amino acid in the sequence (**Figure 2 A**).

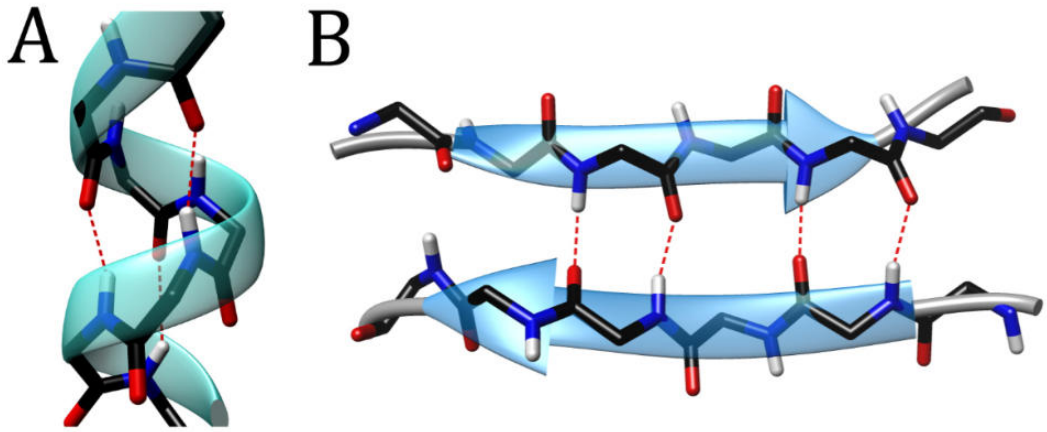


Figure 2: Protein of secondary structural elements. **A:** The α-helix is a right-handed spiral with 3.6 amino acids per turn. The amino acid *n* forms a hydrogen bond with the amino acid *n+4*. **B:** β-sheets are adjacent peptide strands that are connected via a hydrogen bond network. They can be parallel or antiparallel (as shown here).

From secondary structural elements, as described above, a function-specific overall fold of the protein is assembled. An example of a tertiary protein fold is shown in **Figure 3**, where the protein endothiapepsin (PDB code 3T6I)[8] is depicted, which is a target of the research presented in Chapter 2.2.3.

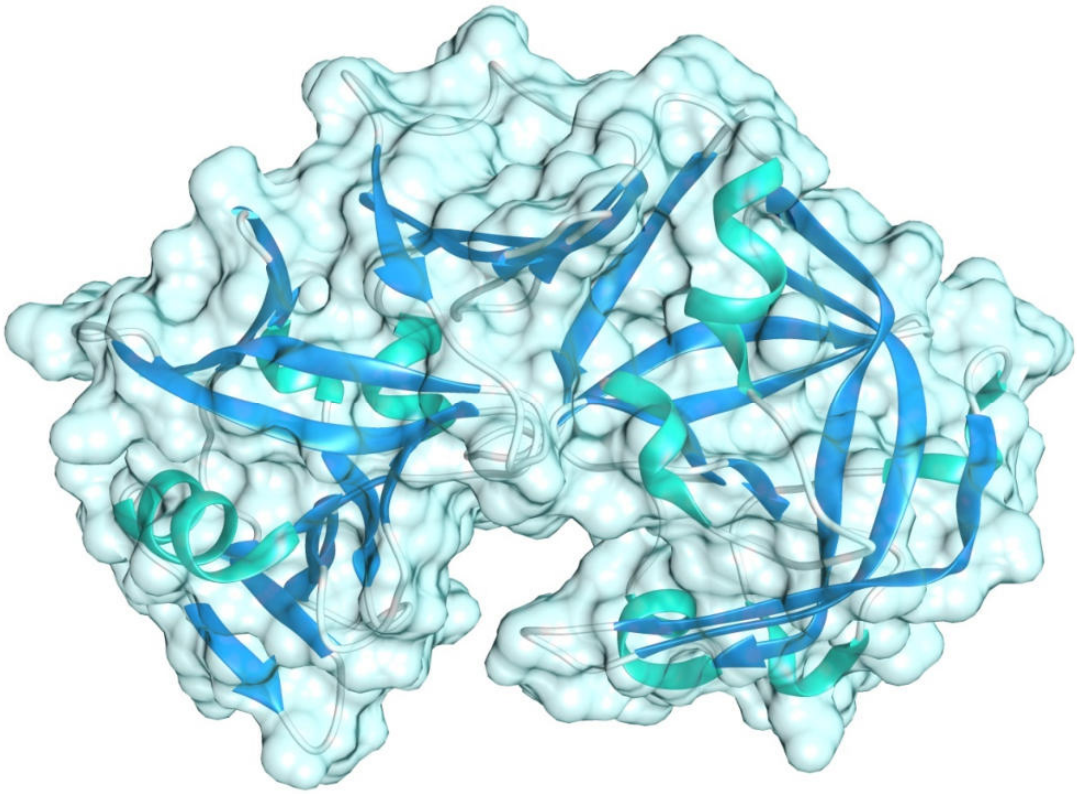


Figure 3: Structure of endothiapepsin (PDB code 3T6I). [8] β-sheets are shown in blue, α-helices in green, loops in white, and the surface in translucent light blue.

The 3D structure, consisting of antiparallel β-sheets and α-helices, forms a channel with the active center in the middle. The protein endothiapepsin is an aspartyl protease, the members of this family catalyze the hydrolytic fragmentation of peptides and proteins. The peptide strand of the reactant penetrates the active channel. [9] As such, the tertiary structure of endothiapepsin is decisive for substrate specificity following the key lock principle and is thus very important for its catalytic function.

Beyond the tertiary structure some proteins even consists of a supramolecular assembly of a multitude of proteins. This is called quarternary structure, and its components are connected by H-bonds and van-der-Waals forces. One of the most prominent examples of such a supramolecular assembly is hemoglobin. [10] Another example is the protease of the HIV virus (**Figure 4**, PDB entry 2b60[11]), which catalyzes the hydrolysis of peptide bonds analogous to endothiapepsin (compare the structural similarity: **Figures 3** and **4**). In contrast to endothiapepsin, which is a monomeric protein, HIV protease is a dimer of two tertiary structured proteins and therefore has a quarternary structure. The two domains are bound by intercalated antiparallel β-strands. Thus, as described above (**Figure 2**), intermolecular H-bonds are formed as shown in **Figure 4 B** (red dotted lines).

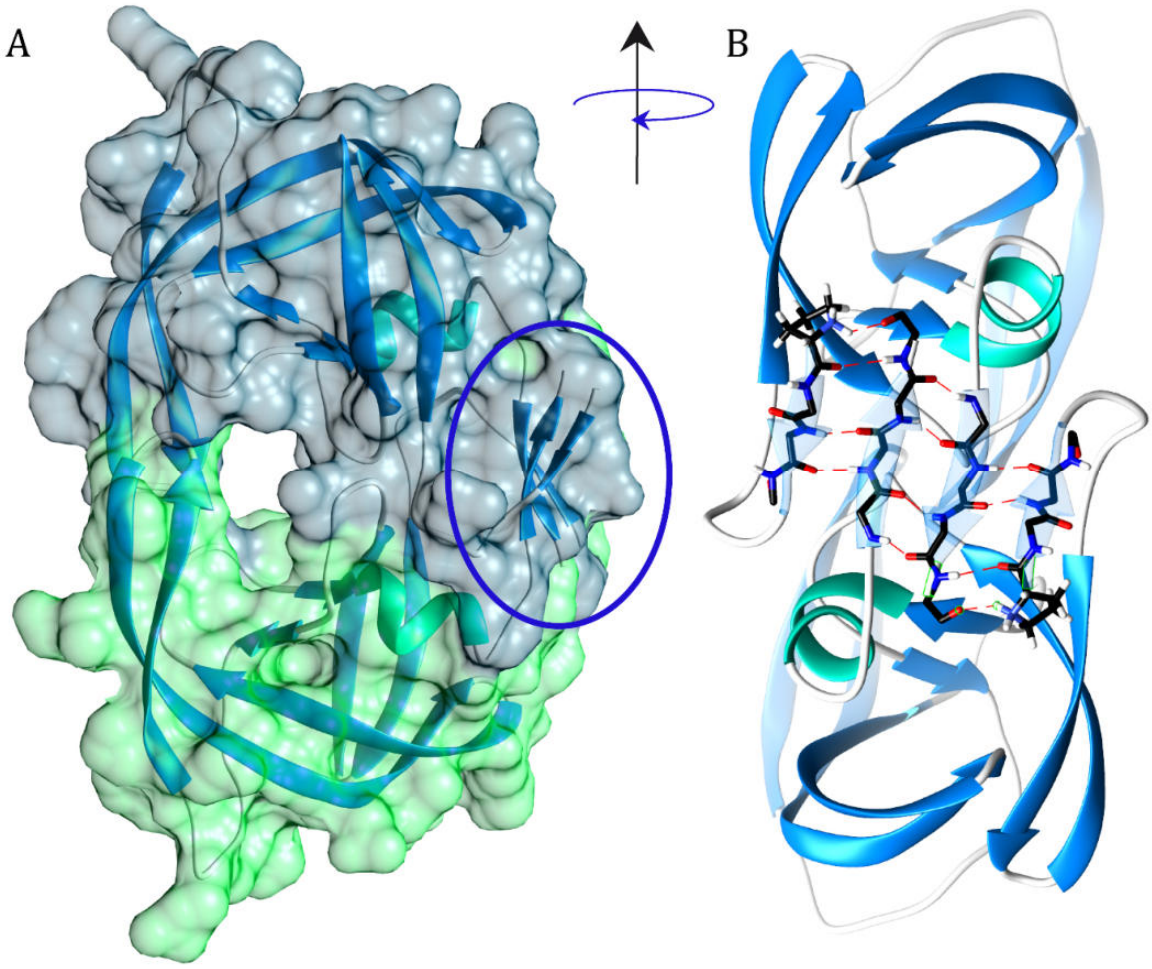


Figure 4: X-ray structure of the peptidase of HIV-1, PDB entry 2b60. **A:** the translucent surface of the structure shows the two protein domains. One is represented in blue, the other in green. **B:** Representation of the binding region of the domains, the structure is rotated by 90 degrees. The β-sheets are shown in blue, the α-helices in green, and the loops are shown in white.

1.2.2 Enzymes

Enzymes are proteins which enhance the speed of a chemical reaction, similar to catalysts in chemistry, but in a highly selective manner. The catalytic unit providing the selectivity is formed by the active center, which is buried in a tertiary or even quarternary protein structure as described above. An example is human carbonic anhydrase II (hCAII), also known as carbonate dehydratase, which is a research objective of this thesis (Chapters 2.1.1, 2.1.2, and 2.2.1). **Figure 5** shows the average of the hCAII structure, calculated from the 20 minimum-energy structures determined by ultra-fast magic-angle-spinning solid-state NMR (Chapter 2.1.1, PDB code: 6qeb, residues 1-25 hidden) superimposed on the crystal structure [12] (PDB code: 2cab). This enzyme has a so-called beta-barrel shape, which consists of antiparallel beta-strands, as shown in **Figure 2 B**. The protein forms a cone-like cavity. Structure and dynamics of this cavity including the active center is the key to the specificity and efficiency of catalysis of the chemical reaction.

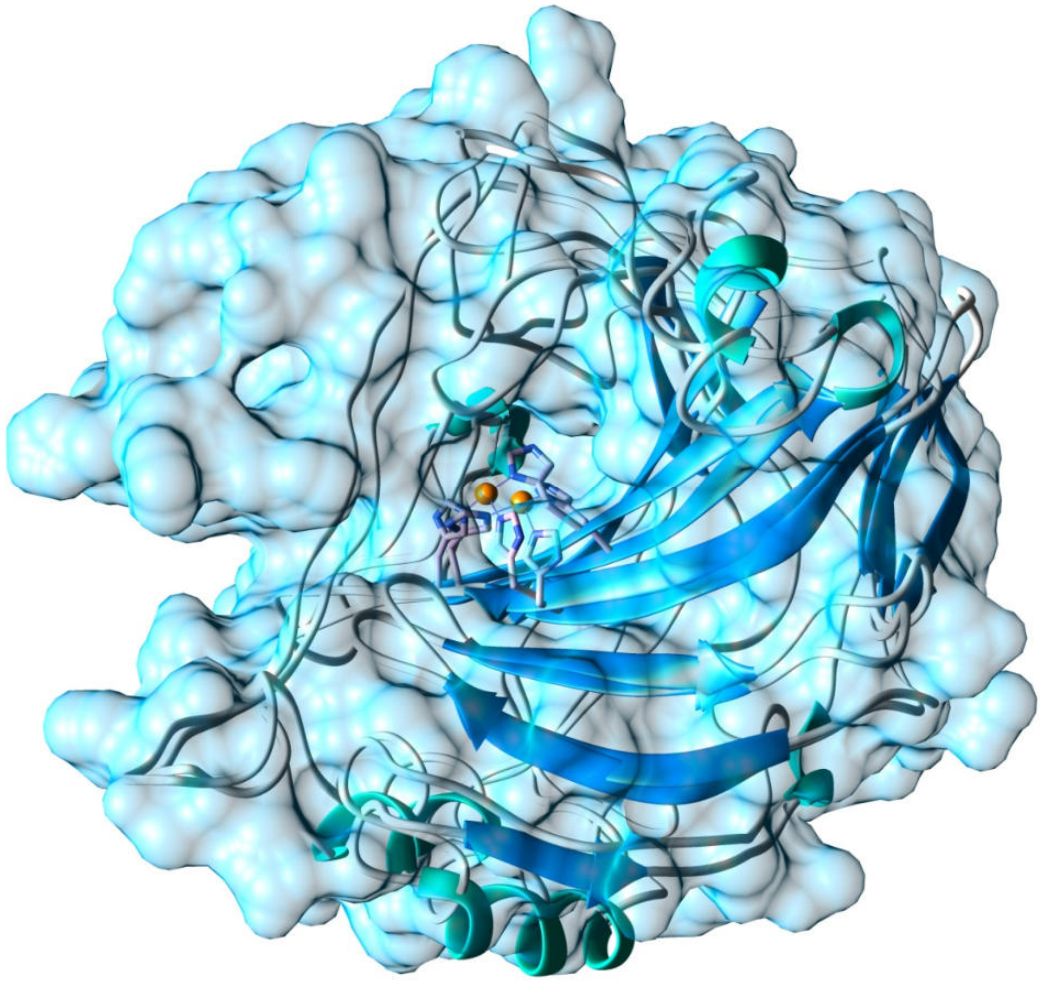


Figure 5: Structure of the human carbonic anhydrase II. The average structure of the 20 lowest energy structures elucidated by ultra-fast magic-angle-spinning solid-state NMR (Chapter 2.1.1 PDB code 6qeb, residues) superimposed on the crystal structure 2cba[12]. The β-sheets are depicted in blue, the α-helices in green, and the loops in white. The zinc ion of the active center is shown in orange, residues 1-25 are not shown.

The enzyme carbonic anhydrase II catalyzes the interconversion of carbon dioxide and water to bicarbonate and protons. [13] The active center consists of a zinc anion (**Figure 5**, orange sphere), complexed by three histidine sidechains and a hydroxide ion. If CO_2 enters the cavity via a hydrophobic wall (compare Chapter 2.2.1), it is attacked nucleophilically by the hydroxide ion. The resulting bicarbonate is replaced by a water molecule, which is deprotonated and, as such, regenerates the hydroxide (compare Chapter 2.2.1). hCAII is the enzyme with the highest known turnover rate with about 10^6 reactions per second. [13a] To enable this reaction speed, a special composition of structure and dynamics is required via a chain of rigid water molecules acting as a proton shuttle. This chain of firmly bound water molecules transports protons from the active center to the solvent water (for further details, see Chapter 2.2.1). As such, it plays an important role in metabolism and pH-regulation. Its dysfunction can cause a variety of diseases, so it is a target for many drugs. [14]

1.2.3 The binding subunit sarcoma homology 3 (SH3) of chicken α-spectrin

The protein domain SH3 (sarcoma homolog 3) was initially found in the oncogenic tyrosine kinase c-Src, where Src stands for the Pous-Sarcom-Virus. [15] The domain is part of more than 300 human proteins as well as other eukaryotic and viral proteins. SH3 plays an important role in substrate recognition, regulation of kinase activity, and membrane anchoring. It binds to proline-rich sequences, in particular carrying the PxxP motif [16] which forms the polyproline II helix also being the main structural motif of collagen [17] (**Figure 6**).

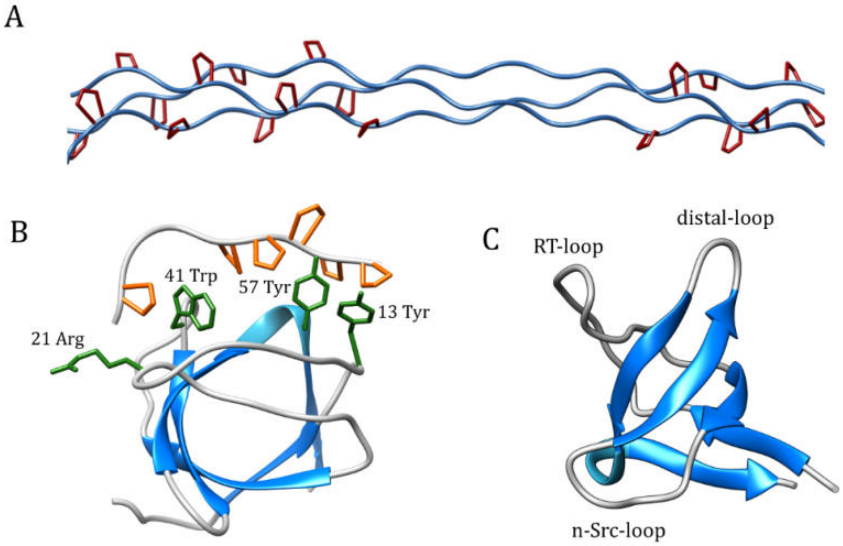


Figure 6: SH3 domain of chicken α-spectrin. **A:** Molecular structure of collagen with the polyproline II helix as a structural motif. Prolines are displayed in dark red (PDB code: 1BKV) [18]. **B:** Representation of the SH3 binding. The SH3 structure from Chapter 2.2.2 (PDB code: 6SCW, the average structure of the 10 lowest target functional structures derived by eNOE restraints) is aligned with the ligand from the SH3-ligand complex solved by NMR[19] (PDB code: 2JMA). The sidechains shown in green play important roles for the binding. The proline sidechains of the

ligand are shown in orange. **C:** Cartoon representation of the eNOE single-state average structure from Chapter 2.2.2 (PDB code:6SCW).

Although, the more than 300 members of the SH3 family differ significantly in protein sequence, their tertiary structure is highly conserved and forms a tight β-barrel shape. The β-strands are connected by three loops (**Figure 6 C**), the RT-, the distal, and the n-Src-loop, as well as a small α-helix between 55Ala and 57Tyr. The hydrophobic binding pocket is formed by the groove between the RT- and the n-Src-loop. [16] Hereby, the aromatic sidechains 15Tyr, 41Trp, and 57Tyr form hydrophobic interactions with the prolines of the ligand (**Figure 6 B**). 41Trp, placed right in front of the n-Src-loop, plays a special role in that it additionally forms a hydrogen bond between the sidechain amide proton and a carbonyl oxygen of the ligand backbone (see also Chapter 2.2.2). [20] The sidechain of the RT-loop Arg21 is believed to play an important role, acting as a "gatekeeper", by blocking the binding groove in the closed conformation. [6a]

1.2.4 H/D exchange for structure determination

Hydrogen bonds are representing the most important forces constituting the secondary, tertiary, and quaternary structure of a protein. A hydrogen bond is a special case in chemistry, also called two-electron three-center bond. [21] The strength of an H-bond lies between a classical covalent, two-electrons two-center bond, and a salt bridge. [22] This type of bond is formed by a proton donor in combination with a proton acceptor. The proton acceptor requires a free electron pair. In the case of proteins, this is usually the oxygen of a carbonyl group. The proton donor must have a strongly polarized covalent bond to the hydrogen, usually the HN-bond. Amide protons are acidic and can, therefore, be exchanged. In general, the presence of an H-bond prevents this exchange.

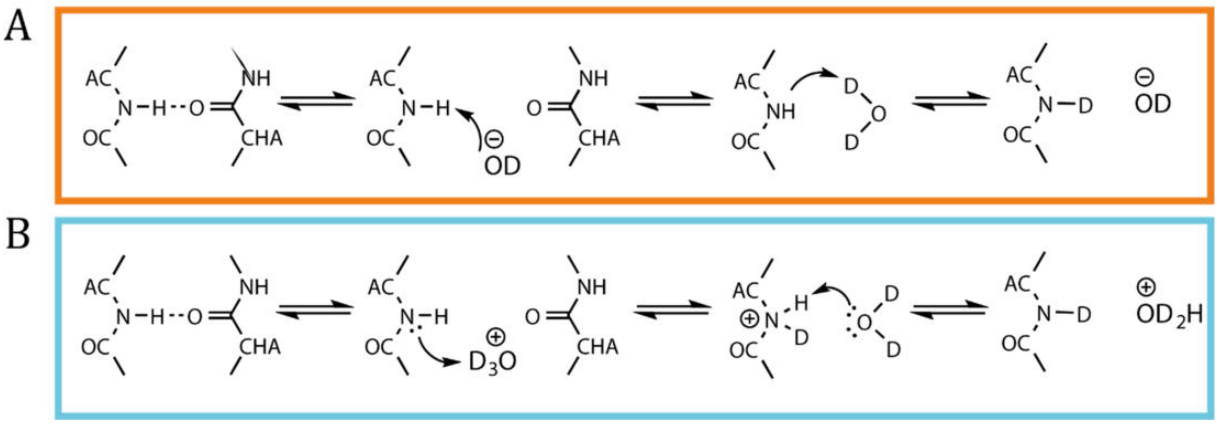


Figure 7: Chemical exchange mechanisms for fully protonated proteins in a deuterated buffer. **A:** Dissociative mechanism. After the H-bond is disrupted, the amide proton is transferred to a base, followed by reprotonation/redeuteration. This reaction mechanism takes places in a base solution. **B:** Associative mechanism. The amide nitrogen is protonated/deuterated before or after

disrupting the H-bond, followed by a deprotonation reaction. Exchange is only possible if the H-bond is disrupted. The mechanism takes place in an acidic environment.

The stronger the bond, the less likely it is that the proton will be exchanged with a proton of the solvent water. However, experiments show that even protons involved in H-bonds show exchange. [23] This exchange can follow two possible mechanisms, one that is associative proton/deuteron catalyzed (**Figure 7 B**) and one that is dissociative base-catalyzed (**Figure 7 A**). [24] The associative mechanism only contributes to the exchange under acidic conditions. Here, a proton/deuteron binds to the free electron of the amide nitrogen. Either before or after protonation, the H-bond dissociates, and in every second case, the previously bound proton is released. Under neutral or basic conditions, the dissociative mechanism is dominant. Here, the proton that forms the H-bond, must dissociate first, followed by a reprotonation/redeuteration. In both cases, the H-bond must break. This breaking can be caused by extreme events of local fluctuations. [25] It follows from this that the exchange rate of protons is a measure for the opening rate of the H-bond and thus also for the stability and rigidity of the structural element (see Chapter 2.1.3).

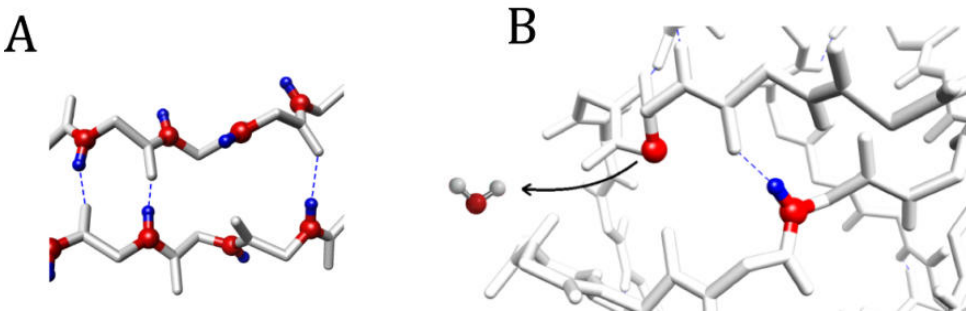


Figure 8: Exchange of amide protons with water. Nitrogen atoms are shown in red, hydrogen atoms as blue spheres. **A:** Due to vibrational modes, the amide bond breaks and exposes the proton to the solvent. **B:** After deprotonation, the nitrogen, which serves as a base, is protonated again.

In addition, both exchange mechanisms require the presence of a water/hydroxyl molecule. This means the exchange kinetics are a measure for water accessibility in addition to H-bond strength. The obstruction of water accessibility is mainly dominated by hydrophobic shielding and is not necessarily proportional to the distance to the protein surface (see Chapter 2.1.3). This means that since not only the near-surface amide-protons but also, on a slower timescale, acidic protons in the hydrophobic core region of the protein are exchanged with the solvent protons, larger structural rearrangements are necessary. The mechanism by which water approaches the protein core has not yet been fully clarified. Two possible mechanisms are discussed in the literature [26]: (1) the so-called penetration model, in which water reaches the amide-protons of the hydrophobic core by "slipping" through gaps opened by local fluctuations, and (2) the local unfolding model, in which constructively multiplexed local motions lead to a slower globular motion with large amplitude, resulting in a reversible local unfolding. Nowadays, it is assumed that the second mechanism is more likely. [27]

1.3 Fundamental Theoretical Background of NMR Spectroscopy

1.3.1 General physical background[28]

Already at the end of the 19th century (1896), an optical splitting caused by a magnetic field was discovered by Pieter Zeeman (Zeeman effect). This effect was interpreted by Hendrik Anton Lorentz as a shift in light frequency by the precession frequency (Larmor-frequency) of atoms in a magnetic field acting as magnetic roundabouts. In 1916, Arnold Sommerfeld discovered the directional quantization of the atomic angular momentum. According to the light-quantum hypothesis (Einstein 1906), the frequency shift $\pm \nu_L$ corresponds to the energy difference as follows:

$$\Delta E = \pm h\nu \tag{1}$$

In 1922, the Stern-Gerlach experiment confirmed that the smallest atomic angular momentum (I=±0.5) can only have two possible orientations to the external magnetic field. Accordingly, the angular momentum P can be described as follows:

$$P = \sqrt{I(I+1)} \tag{2}$$

Hereby, I is the spin-quantum-number, which can take values from 0.5 to 6 in steps of 0.5. The relation between angular and magnetic momentum can be understood by following the simplified notion that the nuclear spin causes a circular motion of an electronic charge.

$$\mu = \gamma P \tag{3}$$

The constant γ is called the gyromagnetic ratio and is specific for every active NMR nucleus. The combination of Equations 2 and 3 shows the quantum-mechanical description of the magnetic moment.

$$\mu = \gamma\sqrt{I(I+1)} \tag{4}$$

The projection of the magnetic moment on the z-axis, parallel to the external field B_0, can be described as a function of the directional quantum number m. This is a function of the spin quantum number I and can take values from m=I, I-1 to -I.

$$\mu_z = \gamma m\hbar = \gamma I_z \tag{5}$$

The energy of the Zeeman splitting can be described accordingly.

$$\Delta E = 2\mu_z B_0 = 2\gamma m\hbar B_0 \tag{6a}$$

In case of a spin ½ system, as shown in **Figure 9**, the z-projection of the magnetic moment can be either parallel or antiparallel to the external magnetic field B_0, which can be defined as α- and β-state, respectively. The Zeeman splitting simplifies to:

$$\Delta E = \gamma \hbar B_0 \tag{6b}$$

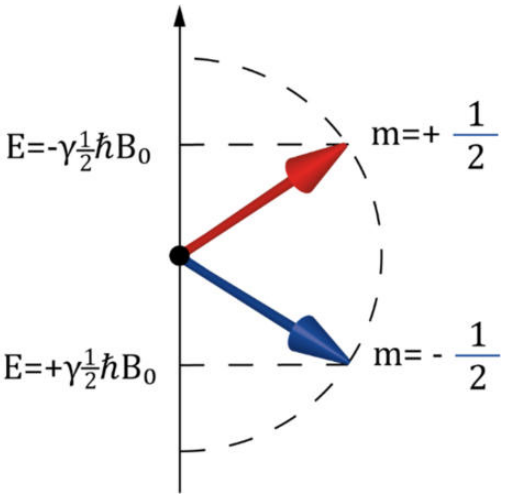


Figure 9: Directional quantization of the nuclear spin in the external field B_0. In this example, the spin quantum number I has the value ½. Accordingly, the directional quantum number m can occupy the values m=±½. This holds true for the nuclei ^{1}H, ^{15}N, and ^{13}C, most widely used in protein NMR. The spin state with m = ½ is also called α-state or spin up, the spin state with m = -½ is called β-state or spin down.

Figure 9 shows a vector picture of the α- and β-states. The relation between the magnetic moment and its projection on the z-axis is given by the angle of the spin-vector to the external field B_0:

$$\mu_z = \mu \cos\Theta \tag{7}$$

This angle Θ is given by a constant value of 54.7°. In a deeply simplified picture, one could imagine the nuclear spin as a roundabout precessing around the B_0-field (**Figure 10**). The angular momentum of the precession compensates the magnetic force caused by the B_0-field, resulting in an equilibrium. Since only the angle Θ is quantum-mechanically permitted (compare Equations 3.5 and 7), a stronger central force will be compensated by faster precession or a higher precession frequency (Larmor frequency). Accordingly, the Larmor frequency ν_L is proportional to both the external field and the gyromagnetic ratio.

$$\nu_L = \left|\frac{\gamma}{2\pi}\right| B_0 \tag{8}$$

From Equation 6, the relation between Larmor-frequency and magnetic energy can be derived.

$$\Delta E = h\nu_L \tag{9}$$

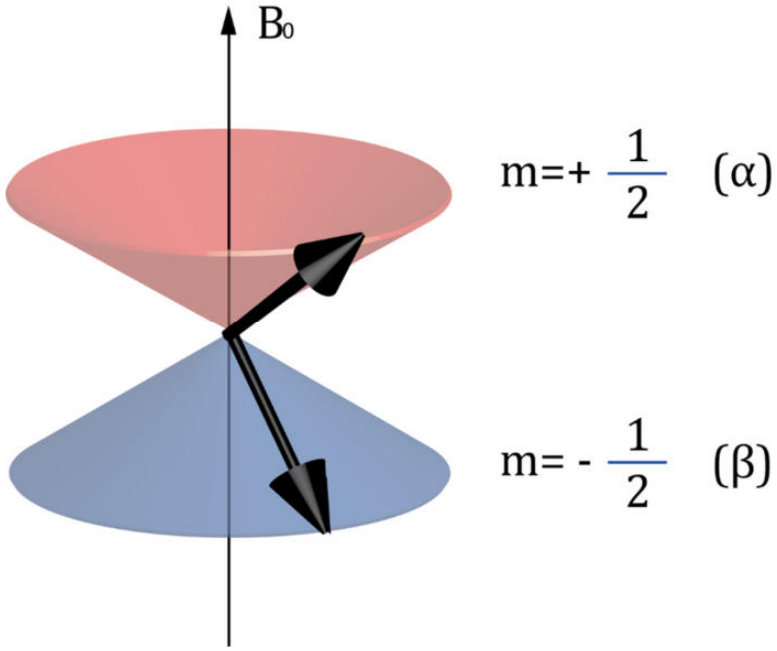


Figure 10: Depiction of the precession fulfilled by the α- and β-state relative to the external field B_0.

1.3.2 Resonance-condition pulsed Fourier Transform NMR[29]

Nuclear magnetic resonance means that a nucleus changes its spin state in response to an external promotion. This external promotion can be achieved by applying a radio frequency. If this frequency matches the Larmor frequency ν_L (see above), the spin absorbs part of the applied energy by changing its spin state between α and β according to Equation 9. If the spin states are unevenly occupied, as caused by an external magnetic field, the RF-frequency leads to a detectable magnetization (**Figure 11**). Hereby, the difference in the occupation of α- and β-state is proportional to the magnetic flux density reaching the nucleus and can be described by the Boltzmann-distribution (Equation 10).

$$\frac{N_\beta}{N_\alpha} = \exp\left(\frac{-\Delta E}{k_B T}\right) \approx 1 - \frac{\Delta E}{k_B T} = 1 - \frac{\gamma h B_0}{2\pi k_B T} \qquad (10)$$

With N_α and N_β as the occupation numbers of the corresponding spin states, T as the temperature, h as the Planck constant, k_B as the Boltzmann constant, ΔE as the energy difference between the α- and β-state, B_0 as the magnetic flux density of the external field, and γ as the gyromagnetic ratio of the nucleus.

Historically, continuous radiofrequency irradiation was used for resonance detection, while the magnetic field was varied. Since the Larmor frequency is proportional to the external field, a resonance can be detected at a certain field strength. Today, superconducting electric magnets are used that reach a field-strength of up to 28.2 Tesla, which is about 100,000 times stronger than the Earth's magnetic field. Such magnets provide a static field that cannot be changed for the experiment. Therefore, a new and

even more efficient approach had to be developed, called pulsed Fourier-transform NMR. Here, all expected resonance frequencies are applied simultaneously in the form of a short (in the order of µs) radio-frequency pulse (RF pulse). The pulse rotates the macroscopic magnetization around the axis aligned with the emitter coil. Hereby, the angle of rotation is proportional to the amplitude and duration of the pulse. At an angle of 90°, the excitation is optimal. After the excitation pulse, the magnetization precesses in the xy-plane and induces a current which is modulated by its Larmor frequency and detected by the receptor coil. The receptor coil is arranged at 90° to the transmitter coil. **Figure 11 C** shows the so-called free induction decay, which is, for a single spin, the combination of a sine and an exponential function. The exponential function describes the relaxation of the magnetization back to its equilibrium.

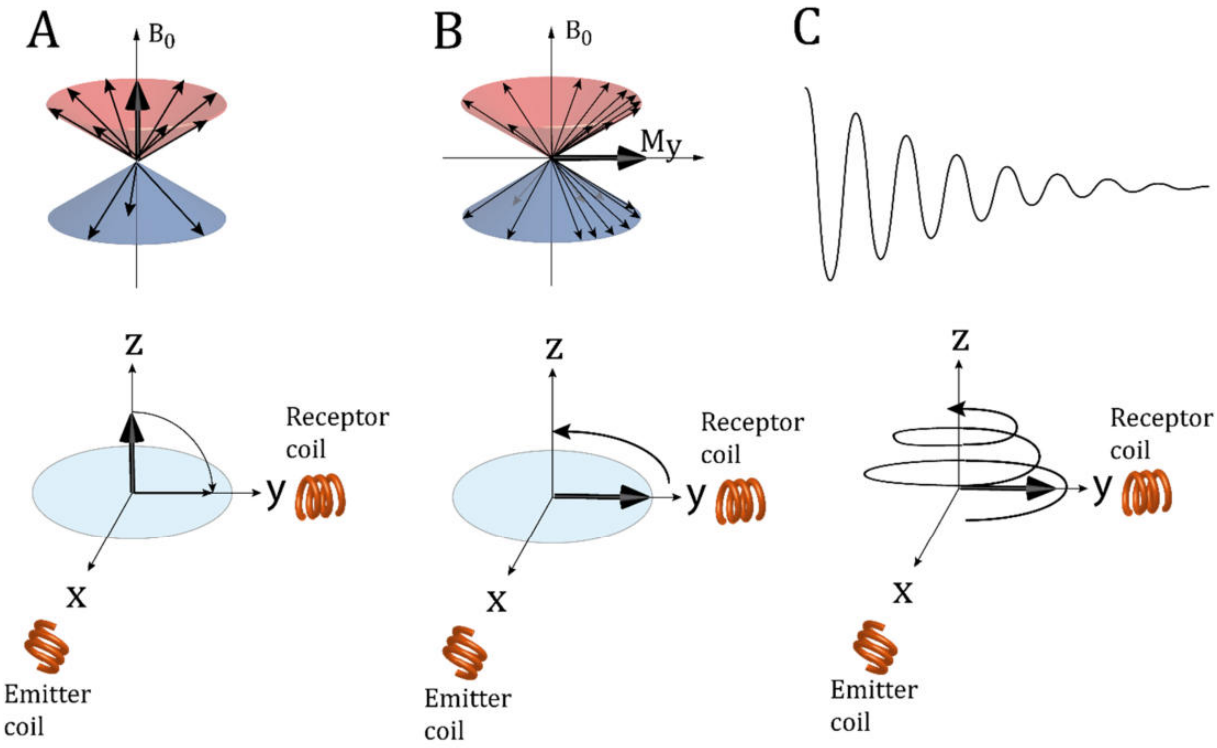


Figure 11: Depiction of excitation and detection of the simplest NMR experiment. **A:** Prior to the excitation pulse, the macroscopic magnetization of the sample is aligned along the z-axis, according to the Boltzmann distribution. **B:** The radiofrequency pulse excites the magnetization such that a vector bundle is precessing in the *xy*-plane, the macroscopic magnetization is turned by 90°. **C:** The precession frequency is detected by the receptor coil in the form of an RF frequency. The amplitude of the signal decreases exponentially over time, and the magnetization undergoes relaxation.

1.3.3 Chemical shift, scalar and dipolar coupling[29a, 30]

Above, the strength of the external field B_0 was mentioned as the magnetic flux density at the position of the nucleus. Indeed, that is what NMR spectroscopy measures. In addition to the strength of the NMR magnet, other parameters influence this observable. Predominantly, the electron shell around the core is shielding the magnetic field. The higher the electron density around the nucleus, the smaller the fraction of the external

field that reaches it. The electron density of the shell is influenced by the chemical surrounding. Depending on many factors such as electronegativity or electronic configuration, the electron density of the shell is reduced or increased by bound residues. This influence of the chemical surrounding is called chemical shift.

$$\delta = 1 - \frac{B_{eff}}{B_0} \tag{11}$$

With δ as the chemical shift, B_0 as the external magnetic field, and B_{eff} as the effective magnetic field at the position of the nucleus.

Besides electronegativity, a further influence on the Larmor frequency is caused by delocalized π-electrons. In an aromatic ring, these electrons are allowed to fulfill circular movements caused by the B_0-field, which induces a retarding magnetic field (ring current effect). This weakens the external field at the top and under the aromatic ring, while next to the ring, where the aromatic protons are placed, the field is strengthened. For this reason, aromatic protons have a relatively low Larmor frequency (resonance frequency), which is known as downfield-shift (according to the continuous-wave-NMR) in old terms. Another important influence on the Larmor frequency is exerted by other spins in the surroundings. Imagined as small magnets themselves, they change the magnetic field at the position of the observed nucleus relative to their orientation. This happens through space, in terms of dipolar coupling, and through the chemical bond that is relayed by the binding electrons, known as scalar coupling. Electrons have a gyromagnetic ratio which is about 658 times higher compared to ^{1}H. Depending on the spin state of the neighboring nuclei, the magnetic field at the observed atom is therefore increased or decreased. This reduces or increases the Larmor frequency. In order to illustrate these effects more clearly, the rotating frame formalism was developed. This formalism allows the conversion of the laboratory coordinate system or polar coordinate system into the rotating frame system, which rotates at the Larmor frequency of the uncoupled spin. Using this formalism, the scalar coupling can be represented more clearly as shown in **Figure 12**: In a two-spin *IS* system, the Larmor frequency of the observed spin *I* is decreased when the coupled spin *S* is in the α-state. This leads to an anticlockwise precession in a coordinate system that rotates at the initial Larmor frequency. By contrast, when the spin state of spin *S* is β, the Larmor frequency of *I* is increased, the spin vector of *I* moves clockwise (**Figure 12 C**).

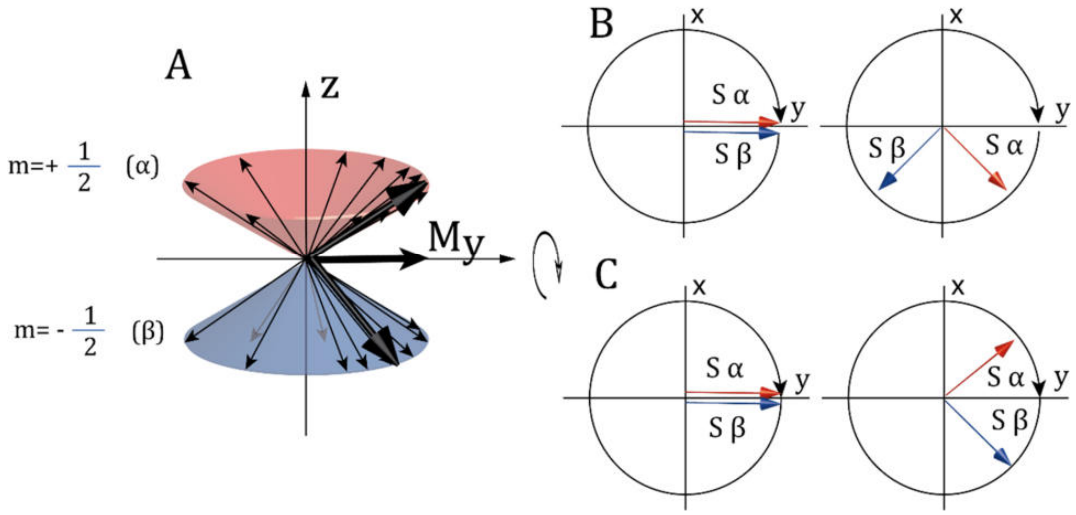


Figure 12: Visualization of the scalar coupling. **A:** Magnetization of spin *I* precessing in the xy-plane, as a vector bundle. **B:** Behavior of spin *I* coupled with spin *S* in *α*- and *β*-state, respectively. On the left, the timepoint zero is shown, on the right, a quarter of the time, representing the inverse Larmor-frequency, has passed. While the coupling to spin *S* in the *α*-state decreases the precession frequency, the frequency is increased by coupling to spin *S* in the *β*-state. **C:** Representation of coupling in the rotating frame coordinate system. Since the system rotates with the Larmor frequency according to an uncoupled situation, Spin *I* precesses anti-clockwise when coupled with spin *S* in *α*-state and clockwise when coupled to spin *S* in *β*-state.

For the situation described above, the term “scalar coupling” is used because this type of coupling does not depend on the orientation to the external field and, therefore, has no vector component. By contrast, the Hamiltonian of the homonuclear dipolar coupling through space has the following dependence on the angle of the internuclear vector relative to the external magnetic field.

$$H_D = \frac{\mu_0 \gamma_I \gamma_S h^2}{16\pi^3 r^3} \cdot \frac{1}{2}(3\,cos^2\theta - 1)\left[3\hat{I}_z\hat{S}_z - \hat{I}\cdot\hat{S}\right] \qquad (12)$$

This angular dependence can be understood figuratively. **Figure 13** shows the relative orientation of two dipolar coupled spins represented as bar magnets. Whether the magnets repel or attract each other depends on their relative orientation. Since the rod magnets represent spins located in an external field B_0, their orientation is either parallel or antiparallel to this field. Accordingly, their relative orientation depends on the internuclear vector (**Figure 13**, dotted line). If the internuclear vector is parallel to B_0, the spin in the center couples most strongly to the spin on the circle (**Figure 13 A**). If the angle of the internuclear vector is orthogonal to the external field, the coupling is inverse and weaker than in **A** (**Figure 13 C**). Logically, there is a situation in-between where the coupling is zero. This is the case when the internuclear vector is arranged to the B_0-field at 54.74°, known as the magic angle (**Figure 13 B**).

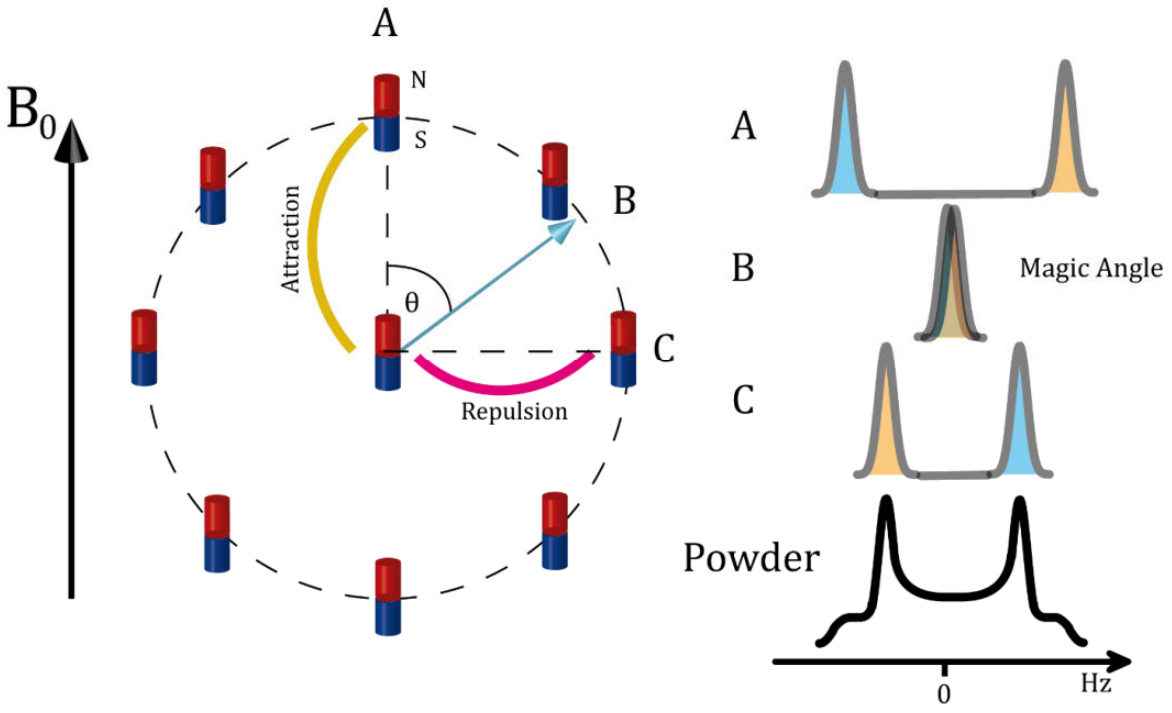


Figure 13: Representation of the directional dependence of the dipolar coupling. On the left side, a two-spin system is represented by rod magnets aligned parallel to the external magnetic field B_0. On the right side, the contributions, of the orientations shown on the left side, to an NMR spectrum are depicted. **A:** The alignment of the internuclear vector is parallel to the external field, and the rod magnets attract each other. **B:** The angle of the internuclear vector to the external field has 54.74 degrees, which is the magic angle. The coupling is zero. **C:** The angle of the internuclear vector relative to B_0 is 90°, the coupling is inverse, but weaker, as compared to **A**. The rod magnets repel each other. In a powder, all orientations are present simultaneously, resulting in a pattern like the one on the bottom right (in the case of a two-spin system).

1.3.4 Magic-angle spinning[30]

Dipolar couplings, as shown above, cover many kHz. Theoretically, each spin is coupled with each other spin in the sample through space. This leads to multiple splitting of peaks, increasing ambiguity to an extent where the spectrum becomes useless. A solution for this problem is the averaging of dipolar couplings by rotation as is provided by Brownian motion in solution-state NMR. However, with regard to the nature of the dipolar coupling (see above), in contrast to Brownian motion, a rotation about a particular axis is sufficient. When the entire sample is rotated about an axis 54.74° relative to the external field, all internuclear vectors are averaged to meet the magic angle condition, and the dipolar coupling becomes zero. The faster the sample is rotated, the better this averaging works. For this purpose, technically sophisticated solid-state NMR probe heads are used.

1.3.5 Magnetization transfer in solid- and liquid-state NMR

In the case of protein NMR, a multitude of resonances causes spectral ambiguity. Therefore, an improvement of the spectral resolution is necessary, which can be achieved by multidimensional NMR spectroscopy. Here, in addition to the frequency detected directly by the receptor coil, resonance frequencies of additional nuclei are also detected indirectly. In order to combine such Larmor frequencies of different nuclei to result in a peak of a multidimensional spectrum, the magnetic polarization must be transferred from one nucleus to the other. The most common way to transfer polarization in NMR is the INEPT transfer. INEPT means: Insensitive Nuclei Enhanced by Polarization Transfer and was originally developed by Morris and Freeman [31] to increase the sensitivity of one-dimensional ^{13}C-spectroscopy. The gyromagnetic ratio and thus the polarization of ^{1}H is by four times higher than the polarization of ^{13}C. By transferring the polarization from ^{1}H to ^{13}C, carbon NMR thus becomes four times more sensitive. Nowadays, the INEPT transfer is mainly used for multidimensional NMR spectroscopy. The polarization is transferred through the scalar coupling (compare Chapter 1.3.3) via the electrons of the chemical bond acting as transmitters. **Figure 14** explains the INEPT sequence step by step. First, an RF pulse on the proton channel excites the spin I (^{1}H) such that the magnetization vector is rotated 90° around the x-axis, resulting in an overall magnetization parallel to the y-axis. The pulse is followed by a delay of ¼ J^{-1}, where J stands for the scalar-coupling constant between I (^{1}H) and S (^{13}C, ^{15}N, or any other hetero nucleus). In view of the rotating frame formalism as shown in **Figure 12 C**, spins coupled to S in the α-state (red) have a reduced precession frequency and thus move anticlockwise, while spins coupled to S in the β-state have accelerated precession and thus move clockwise. Subsequently, a simultaneous pulse on both channels rotates the magnetization vector of the spins I and S by 180° so that the magnetization vectors are on the opposite side of the xy-plane, and the spin state of the hetero-nucleus is flipped. The result of these pulses could be regarded as a point reflection of the spin vectors. After a further delay of ¼ J^{-1}, the proton spins I, coupled to a hetero-nucleus S in the α and β-states, are antiparallel to each other, aligned to the x-axis. Now, a 90° pulse is applied simultaneously on both channels, which rotates the magnetization around the y-axis in terms I-spins and around the x-axis in terms of the S-spins. Thus, the magnetization vector of the hetero-nuclei S coupled to a proton I in an α-state moves to the y-axis, while the magnetization of nuclei S coupled to a proton I in β-state moves to the y-axis. As such, the polarization transfer is performed using the scalar coupling between I and S as a sorting element. For every spin pair, in which I was excited after the first pulse, the magnetization of the hetero-nucleus S is now either parallel or antiparallel to the y-axis. The next period is called refocusing. Each of the steps described above, with the exception of excitation, is now inverted. By contrast, the I-magnetization is now on z and the S-magnetization on the xy-plane. As a final result, the detectable magnetization of the spin S is refocused on the x-axis.

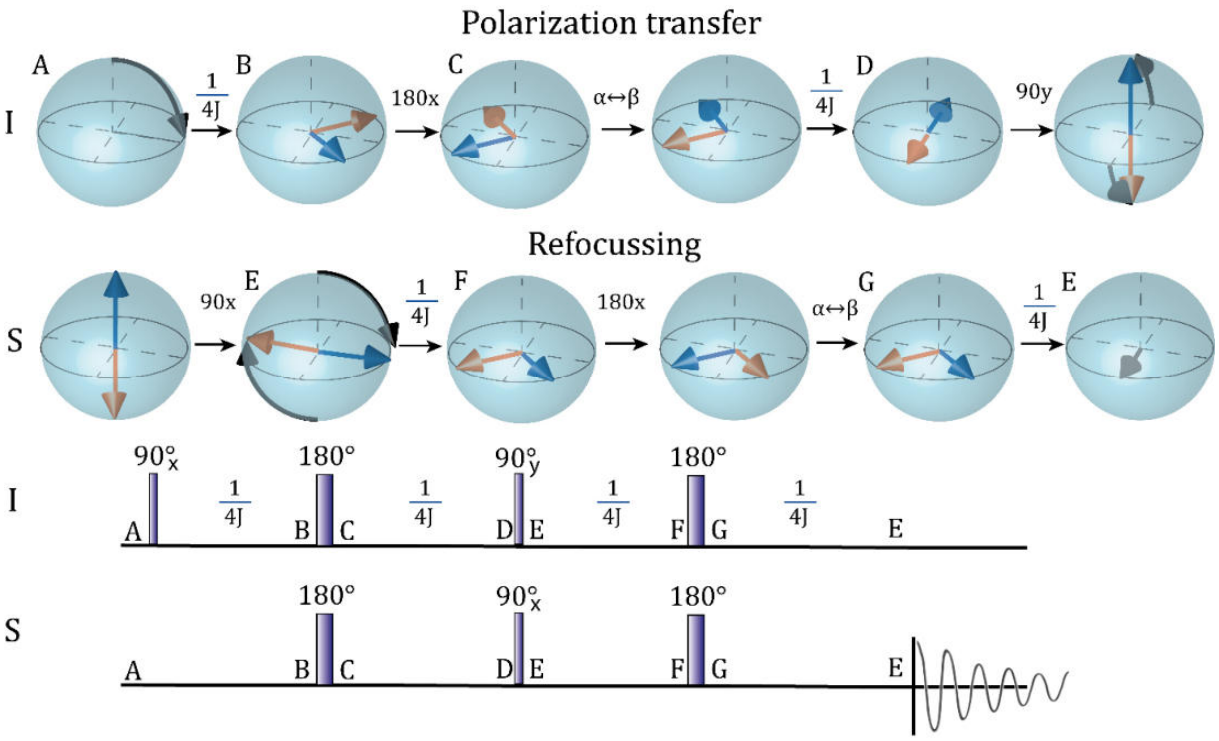


Figure 14: Representation of the INEPT magnetization transfer. Spins that are coupled to a spin in α-state are colored red, those that are coupled to a spin in β-state are colored blue. The coordinate system rotates with the Larmor frequency according to the rotating frame formalism. The sequence starts with a 90°-excitation pulse around the x-axis, on the *I*-channel (usually the ^{1}H channel). The Larmor frequency of spins *I*, coupled to a spin *S* in α-state is reduced, while the Larmor frequency of a spin *I*, coupled to a spin *S* in β-state is increased. After a delay optimized to a fourth of the inverse *J* coupling, the spins are split on the *xy*-plane. **B**: A 180° pulse, applied simultaneously on *I* and *S*, reverses the orientation of the spin-vectors. **C**: Accordingly, their running direction is also reversed. Thus, after another delay of ¼ J^{-1}, the magnetization vectors are aligned antiparallel to the x-axis. At this point, simultaneously, a $90°_y$ and $90°_x$-pulse are applied on the *I*- and *S*-channel respectively. At point **D**, by "sorting the spins" using the *J*-coupling, an antiphase term on *I* is created ($2I_xS_z$). The two 90° pulses can now transfer the antiphase term from *I* to *S*, resulting in $2I_zS_y$, the polarization is transferred. The antiphase magnetization can be refocused by applying steps **A-D** in reverse order (**E-H**), resulting in the in-phase operator S_x.

1.3.6 Cross-polarization transfer[30]

In solution-state NMR, the INEPT, as explained above, represents the main approach to transfer magnetic polarization. By contrast, for solid-state NMR, the INEPT is mostly not applicable due to short T_2-relaxation times. As the magnetization precesses in the xy-plane, most of the polarization is lost. Therefore, Pines at al. [32] developed an approach for an efficient polarization transfer in solid-state NMR, known as cross-polarization. In contrast to the INEPT, cross-polarization (CP) transfers magnetization directly through space. Coming back to the *IS* nomenclature, where the spins of polarization origin, usually the protons, are called *I*, and the heteronuclear destination spins are called *S*, the transfer sequence begins with a 90° pulse on the *I*-channel (^{1}H-channel). This is followed by an adiabatic spinlock on both channels in the form of an RF field. Hereby, the RF field can be

regarded as an additional field B_1, which is applied perpendicular to the z-axis. Similar to the Zeeman-splitting along the z-axis, which is proportional to the B_0-field, the spins undergo an additional splitting along the new B_1-field. The new energy-splitting is proportional to the B_1-field and to the gyromagnetic ratio of them simultaneously. Therefore, the energy difference between the spin states is equalized by the application of B_1-fields whose intensities between channel I and S show the inverse ratio in comparison to the ratio of gyromagnetic ratios of both nuclei. Therefore, the polarization can be transferred in an adiabatic energy-conserving manner. This is called Hartman-Hahn condition. [33] Even in an ideal case, the efficiency of polarization transfer has a maximum of 50%. This upper limit exists because, according to the adiabatic condition, only zero-quantum transfers are permitted. All spins I are in α-state after excitation, and since the orientation of S is random, the spin operator $I_\alpha S_\beta$ has only 50% occupancy. The other spin-operator $I_\alpha S_\alpha$ cannot contribute to the polarization transfer. Therefore, at least half of the polarization is lost.

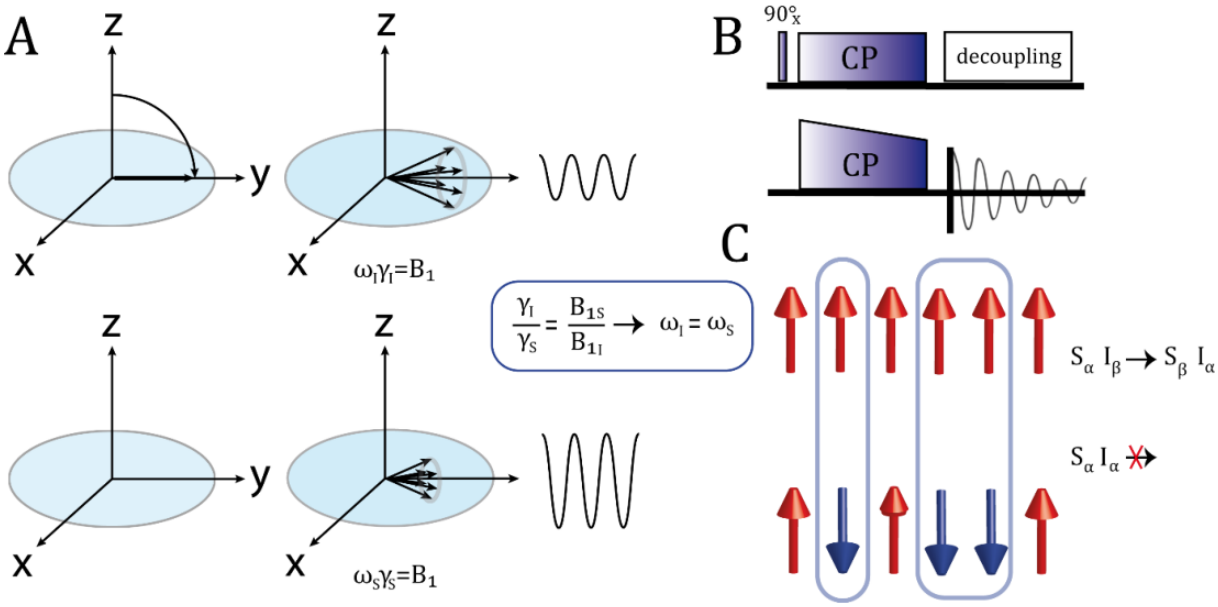


Figure 15: Illustration of the cross-polarization transfer. **A:** An adiabatic spinlock is applied after a 90° pulse on I around the x-axis. This creates an additional magnetic field B_1. If the ratio of field strength between the I- and S-channel is the inverse of the ratio between the gyromagnetic ratios of the nuclei, the Hartmann-Hahn condition is fulfilled, and the magnetization is transferred. **B:** Pulse sequence of the CP-building block. **C:** Since an adiabatic spinlock allows only zero-quantum transitions, and the spin state of S is random, the maximal magnetization transfer efficiency is 50%.

1.3.7 Distance determination in solid- and liquid-state NMR spectroscopy

1.3.7.1 Nuclear Overhouser effect[28]

The Nuclear Overhouser Effect Spectroscopy (NOESY) is the most common and therefore most important approach for measuring internuclear distances. The principles behind the nuclear Overhouser-effect, which describes the magnetization transfer through space, can

be understood using the Solomon equations. The Solomon equations transcribe differential equations of populations in magnetization. For a simplified explanation, a two-spin system consisting of spin *I* and *S* is used. The energy levels of the z-magnetization are shown in **Figure 16 D-E**. The magnetization of spins *I* and *S* can be derived from the population *N* of the states shown in the figure as follows:

$$I_z = N_1 - N_3 + N_2 - N_4 \tag{13}$$

$$S_z = N_1 - N_2 + N_3 - N_4 \tag{14}$$

In addition to the single-quantum operators above, an additional multi-quantum operator must be introduced that describes the difference in the population differences between the two permitted single-quantum transitions of spin *I* and spin *S* (related to the zero and double-quantum transitions W_0 and W_2, **Figure 16 D-E**).

$$2I_zS_z = N_1 - N_3 - N_2 + N_4 \tag{15}$$

In general, the transitions between the energy levels can be departed into allowed transitions, i.e. the single-quantum transitions of *I* (W_I^1 and W_I^2) and *S* (W_S^1 and W_S^2), and forbidden transitions, i.e. the zero- and double-quantum transitions W_0 and W_2. The magnetization flow is represented by the time derivatives of the spin-operators. Equation 16 exemplarily shows the time derivative of Equation 13:

$$\frac{dI_z}{dt} = \frac{N_1}{dt} - \frac{N_3}{dt} + \frac{N_2}{dt} - \frac{N_4}{dt} \tag{16}$$

$\frac{N}{dt}$ can be substituted by using the transition rates as shown in **Figure 16**, exemplarily shown for N_1.

$$\frac{N_1}{dt} = -N_1W_S^1 - N_1W_I^1 + N_2W_S^1 + N_3W_I^1 \tag{17}$$

For normalization purposes, the unity operator is used, which is the sum of all populations:

$$E = N_1 + N_2 + N_3 + N_4 \tag{18}$$

With the population, expressed using spin operators,

$$N_1 = \frac{1}{4}(E + I_z + S_z + 2I_zS_z) \tag{19}$$

$$N_2 = \frac{1}{4}(E + I_z - S_z - 2I_zS_z) \tag{20}$$

$$N_3 = \frac{1}{4}(E - I_z + S_z - 2I_zS_z) \tag{21}$$

$$N_4 = \frac{1}{4}(E - I_z - S_z + 2I_zS_z) \tag{22}$$

The Solomon equations, which are the time derivates of these spin operators, can now be expressed using rate constants.

$$\frac{dI_z}{dt} = -(W_I^1 + W_I^2 + W_2 + W_0)I_z - (W_2 - W_0)S_z - (W_I^1 - W_I^2)2I_zS_z \tag{23}$$

$$\frac{dS_z}{dt} = -(W_2 + W_0)I_z - (W_S^1 + W_S^2 + W_2 - W_0)S_z - (W_S^1 - W_S^2)2I_zS_z \tag{24}$$

$$\frac{d2I_zS_z}{dt} = -(W_I^1 + W_I^2)I_z - (W_S^1 + W_S^2)S_z - (W_I^1 + W_I^2 + W_S^1 + W_S^2)2I_zS_z \tag{25}$$

The Solomon equations 21 to 23 show the (dipolar) connectivity between spins *I* and *S*. The magnetization-change rate of I_z, for instance, depends not only on I_z-I_z^0 but also on S_z-S_z^0 and $2I_zS_z$. This phenomenon is called cross-relaxation. The cross-relaxation rate σ_{IS}, which describes the magnetization transfer from *S* to *I*, is given by the rate constant of the "forbidden" zero- and double-quantum transitions.

$$\sigma_{IS} = (W_2 - W_0)(S_z - S_z^0) \tag{26}$$

Looking at Equation 25, it becomes clear that the rate of magnetization transfer depends on the deviation of spin S_z from its equilibrium S_z^0. This fact is particularly helpful for understanding the pulse sequence in **Figure 16** (top). In detail, the standard 2D NOE-pulse sequence consists of only 3 pulses, all $90°_x$. The first pulse rotates the magnetization of *I* and *S* on the y-axis (**B**), followed by a period of chemical shift evolution. After the evolution, a second $90°_x$ pulse rotates the magnetization back to the z-axis. Hereby, only the magnetization remains, which is given by the projection of the magnetization vector onto the y-axis. This means that if the incremented t_1-time coincides with the time required for a half rotation of the *I*-spin, which gives the diagonal peak of *I*, the spin vector of *S* is partially on the x-axis and thus its z-magnetization is out of equilibrium after the second pulse. Vice versa, this is also true when t_1 satisfies the condition for the diagonal peak of *S*.

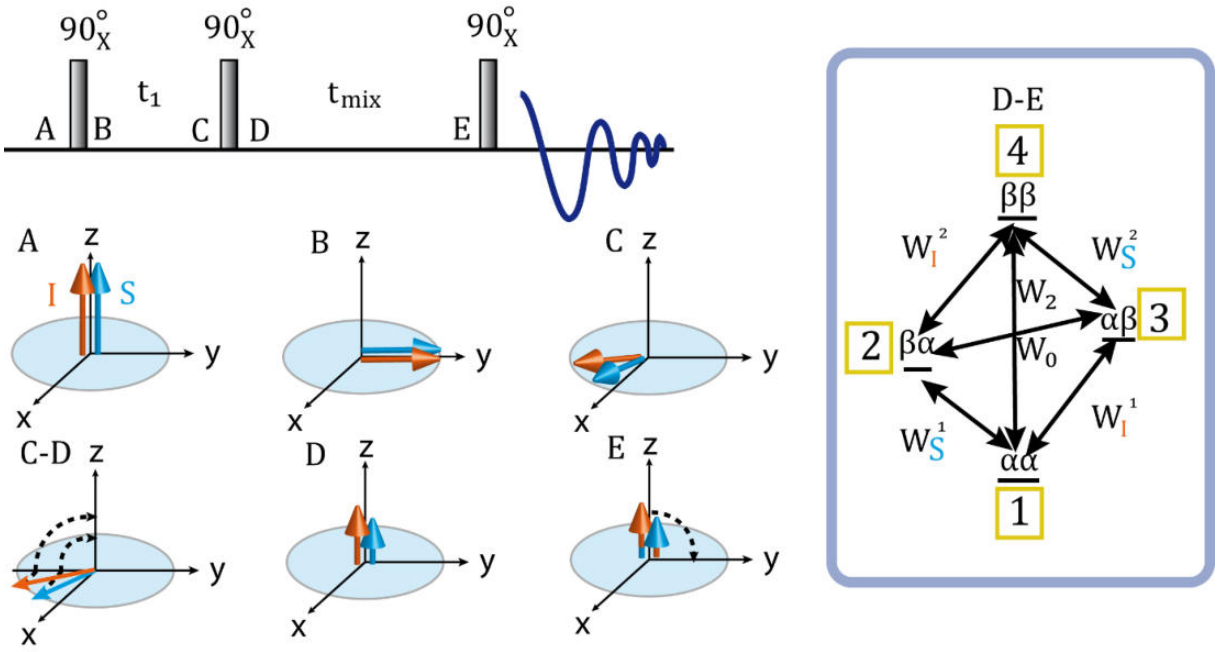


Figure 16: Representation of the magnetization transfer by the nuclear Overhouser effect. At the top, the simplest version of a 2D NOESY pulse sequence is displayed. The first 90° pulse excites both nuclei (**A-B**). During the following evolution time, the different chemical shift of *I* and *S* evolves (**B-C**). The second 90° pulse returns the magnetization along the z-axis (**C-D**). Hereby, whenever the evolution time for *I* results in a maximum y-magnetization due to the different offsets, *S* cannot have maximum y-component and must have a nonzero x-component. Therefore, if the *I*-magnetization leads to a diagonal peak, the magnetization of *S* is out of equilibrium during **D-E** and vice versa. Thus, the requirements for magnetization transfer during **D-E** are fulfilled. The last 90° pulse generates an observable magnetization along the y-axis. The box on the right shows the energy levels of a two-spin system. W_0 and W_2 are the zero and double-quantum transitions, respectively, leading to magnetization transfer during mixing.

In addition to longitudinal cross-relaxation, the magnetization change is also caused by longitudinal auto-relaxation R_1, which generally occurs in NMR-spectroscopy.

$$\frac{dI_z}{dt} = -R_1(I_z - I_z^0) \tag{27}$$

By combining Equations 24 and 25, the change in the difference to equilibrium magnetization over time, can now be expressed. For simplicity's sake, it is assumed that cross-correlated relaxation with respect to the term $2I_zS_z$ is zero.

$$\frac{d(I_z - I_z^0)}{dt} = -R_1(I_z - I_z^0) - \sigma_{IS}(S_z - S_z^0) \tag{28}$$

Equation 26 shows that the cross-relaxation rate σ is proportional to the difference between double and zero quantum relaxation rates (W_2-W_0). These, in turn, are caused by the dipolar relaxation mechanism. In general, longitudinal relaxation depends on field fluctuations around the Larmor frequency. Therefore, molecular vibrations are usually much too fast to cause such a kind of relaxation. Consequently, the dipolar interactions that cause longitudinal cross-relaxation are due to a molecular reorientation in the

magnetic field, the molecular tumbling. In order to generate double quantum transitions (W_2), the tumbling frequency must be in the range of the sum of the Larmor frequencies of the two nuclei involved. In order to cause W_0-relaxation, the tumbling frequency must be in the range of the difference between the two Larmor frequencies; in the homonuclear case, this is about zero. For a fast-tumbling molecule, W_2-transitions are predominant, and the cross-peak intensity becomes positive, for a slow tumbling molecule W_0 is predominant, and the peak intensity becomes negative (compare **Figure 16 D-E**). In addition to the tumbling time of the molecule (called correlation time), the cross-relaxation depends on the internuclear distance. Since the cross-relaxation is driven by the dipolar interaction of the two nuclei *I* and *S*, the dipolar coupling Hamiltonian has to be squared. The dipolar coupling is anti-proportional to the cubic distance (Equation 12) so that the dipolar cross-relaxation is proportional to the inverse sixth power of the distance. Finally, the magnetization-transfer rate, due to NOE, can be expressed within the narrowing limit of $\omega_0\tau_c \ll 1$ (ω_0: Larmor frequency, τ_c: correlation time) as follows:

$$\sigma_{IS}^{NOE} = \frac{\hbar^2\mu_0^2\gamma^4\tau_c}{8\pi^2 r_{IS}^6} \tag{29}$$

where $\hbar$ is the Planck constant over 2π, μ_0 is the magnetic permeability, γ is the gyromagnetic ratio, r_{IS} is the internuclear distance, and τ_c the molecular correlation time.

1.3.7.2 Rotor-synchronized dipolar recoupling[34]

Similar to liquid-state NOESY, also in solid-state NMR, approaches for measuring internuclear distances for protein structure-determination exist. As with NOESY, these approaches make use of the dipolar coupling between the spins. In solid-state, as discussed in Chapter 1.3.4, these dipolar interactions are spun out by the magic-angle spinning. For the selective reintroduction of the dipolar interactions, there are a variety of pulse-sequence building blocks (DARR[35], DREAM[36], PDSD[37]). In this work, the Radio Frequency Driven Dipolar Recoupling (RFDR[38]) approach was used. The RFDR-building block consists of a series of rotor-synchronized 180° pulses. The 180° pulses are placed in the middle of the rotor period, which corresponds to the inverse of the spinning frequency. This means the first pulse is applied after the rotor has turned 180°. In a homonuclear two-spin system, the RFDR sequence reintroduces the zero-quantum part of the dipolar coupling operator. Approximating ideal 180° pulses, the Hamiltonian of chemical shift and dipolar coupling is divided into the following terms.

$$H_{int}(t) = H_0(t) + H_1(t) \tag{30}$$

Hereby $H_{int}(t)$ is the internal Hamiltonian, containing all changes of the spin system except of those referring to the RF-irradiation (external Hamiltonian). While $H_0(t)$ representing the zero-order parts of the internal Hamiltonian, which are commuting with themselves over the RFDR sequence and thus average to 0, the 180° prevent the first order Hamiltonian H_1 from commuting at every timepoint of the RFDR building-block.

$$H_1(t) = H_{CS,1}(t) + H_{D,1}(t)$$

$$H_1(t) = \frac{1}{2}\big(\delta_I(t) + \delta_S(t)\big)(I_z - S_z) - d_{IS}(t)\big(I_x S_x + I_y S_y\big) \tag{31}$$

In general, the theoretical details of how magnetization is transferred via RFDR are not completely understood. Further details are omitted in this description. An easy way how to imagine dipolar recoupling via RFDR is that generally, pulses inverting all spins of the same kind in the middle of a rotor-period, thus in a rotor synchronized manner, prevent their homonuclear dipolar coupling from complete averaging due to Magic-Angle-Spinning (MAS). As with NOESY, the active dipolar coupling between the nuclei can be employed in order to measure the transfer efficiency of the magnetization of a spin pair. Hereby, the distance dependence of the dipolar coupling constant can be used to determine internuclear distances, which is useful for structure calculation.

$$d_{IS} = \left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_I \gamma_S}{r_{IS}^3}\hbar \tag{32}$$

where $\hbar$ is the Planck constant over 2 π, μ_0 is the magnetic permeability, r_{IS} is the internuclear distance, and γ the gyromagnetic ratio.

RFDR is the most common approach for distance determination in proton-detected MAS solid-state NMR spectroscopy.

1.3.8 Approach for exact distance determination[39]

In Chapter 1.3.7.1, the distance determination using NOE, which is used for structure determination, was described. Commonly, in protein NMR, magnetization-transfer rates and thus distances are read out from a single 3D NOESY-HSQC experiment. In a multidimensional NMR spectrum, the peak intensity and thus the accuracy of distance determination is impeded by various sources of error: Differential T_1 relaxation during NOE-mixing; site-specific INEPT transfer efficiencies, and additional magnetization transfer via third spins placed in close proximity, in terms of spin diffusion. These sources of error can be addressed as explained below. Instead of a single spectrum, a series of NOE spectra are recorded with increasing mixing time. Hereby, in order to avoid increasing spin diffusion at long mixing times, the maximum mixing time of the series for deuterated and completely back-exchanged proteins can be determined using the correlation time τ_c[40]:

$$T_{mix}^{deut} = 5 \cdot 10^{-10}\,\frac{s^2}{\tau_c} \tag{33}$$

And correspondingly for fully protonated proteins:

$$T_{mix}^{prot} = 2.5 \cdot 10^{-10} \frac{s^2}{\tau_c} \quad (34)$$

By using Equations 31 or 32, the NOESY transfer stays in the approximately linear initial rate regime [41]. The recording of such buildup series allows a more accurate determination of the polarization transfer rate. However, the fitted buildup-rate is still hampered by the above-mentioned sources of error, including residual spin diffusion. Therefore, Vögeli et al. [39a] developed an integrative approach to dealing with these errors, called eNOE (exact-NOE). The interrelated analysis steps are implemented in the freely available MATLAB-based program eNORA2[42] by Strotz et al. Hereby, the magnetization loss due to auto-relaxation during mixing can be corrected using a mono-exponential fit of the diagonal decay. The cross-peak buildup is corrected by using the diagonal decay rate ρ as a fixed parameter for double exponential fitting (**Figure 17 A).**

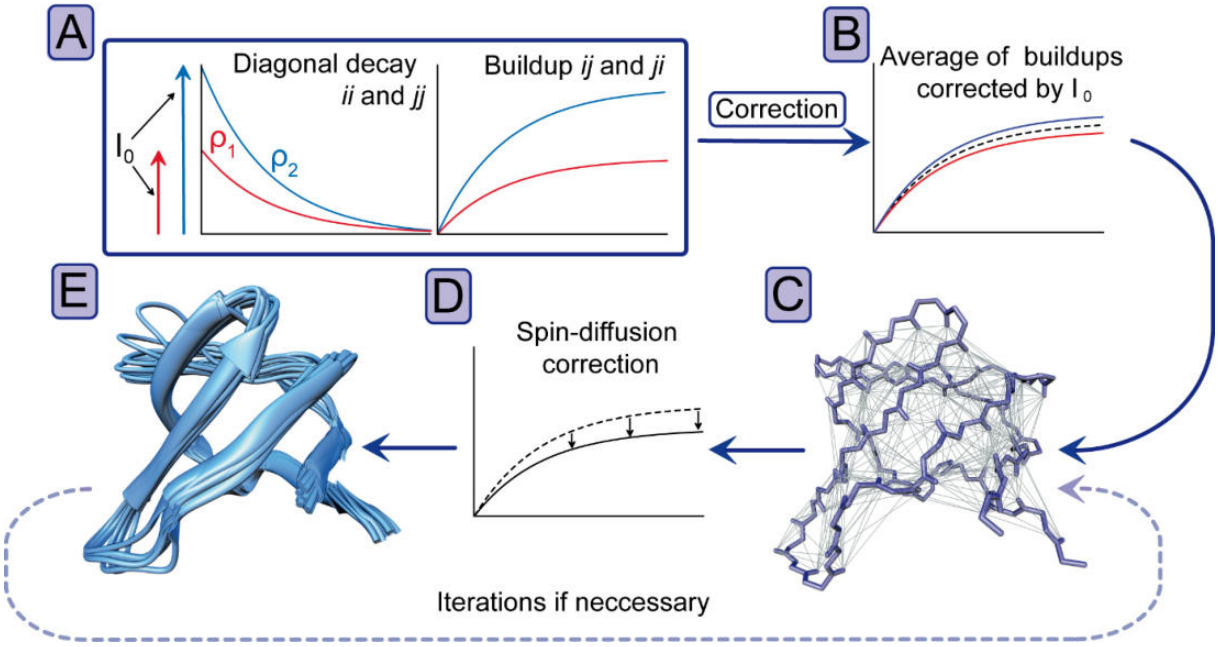


Figure 17: Error-correction flowchart for exact distance determination. **A:** Representation of diagonal peak decays (left) and cross-peak buildups (right). I_o is the diagonal peak intensity extrapolated at zero mixing time, and ρ is the diagonal decay rate. **B:** Correction by normalizing the buildup intensities by I_0. **C:** All internuclear contacts closer than 8 Å are shown on the backbone of the SH3 solid-state structure. **D:** Correction of spin diffusion in terms of magnetization, which is transferred via third spins in the proximity (**C**) by a relaxation matrix approach. **E:** The structural ensemble is calculated using the corrected restraints. If further refinement is required, the resulting structure can be used as input prior to step **D**. The figure is brought forward from Chapter 2.1.2.

The differential efficiencies of heteronuclear magnetization transfer can be addressed by normalizing the cross-peak intensities by the extrapolated diagonal intensity at zero mixing time (I^0, **Figure 17 A**). This correction is valid because, in terms of NOESY-HSQC, the diagonal peak of the nucleus, which is the destination of magnetization, undergoes the

same heteronuclear magnetization transfers as the cross-peak. Vice versa, in terms of HSQC-NOESY, this is the case for the diagonal peak of the nucleus where the magnetization has its origin. A minor source of error is the differential magnetization recovery during the d_1-delay (relaxation delay). By recording HSQC-NOESY spectra, this error is corrected by normalization with the diagonal peak of magnetization origin. However, in contrast to NOESY-HSQC, HSQC-NOESY is hindered by diminished water suppression. If a pair of cross- and diagonal peaks is unambiguously available , the average value of the corrected buildup rate is taken. On the one hand, this is advantageous because an average value is always more accurate than a single value; on the other hand, in NOESY mixing, the magnetization is on both nuclei at a fraction of the time, so that a correction with both diagonal peaks is necessary for the ideal addressing of differential auto-relaxation. In addition, if both diagonal peaks are available, the error of differential recovery during d_1-delay can also be corrected by using NOESY-HSQC spectra, resulting in superior water suppression. The most important correction performed by the eNOE approach is the correction for spin diffusion in terms of relayed magnetization transfer using a relaxation matrix approach. For this correction, an initial structural model is needed, which can be determined using conventional NOE structure determination protocols. During mixing, in addition to the direct transfer from *I* to *S*, magnetization is also transferred through third spins acting as transmitters. The initial structural model is used (**Figure 17 D**) to determine the ratio between the polarization transferred directly form *I* to *S* and the magnetization transferred through relay transfer. This ratio is then used to correct the cross-peak buildup. The described eNOE approach leads to very accurate distance restraints, also for 3D-NOESY, up to an error below 0.07 Å.

The application of the eNOE approach to the SH3-domain of chicken-alpha spectrin is shown in Chapter 2.2.2. Above, it is described that the NOE has a counterpart in proton-detected solid-state NMR spectroscopy, the RFDR. Up to now, the accuracy of the RFDR-distance restraints was only qualitatively grouped as close, medium-range, or far, which resulted in an upper limit of 4 Å, 6 Å, or 8 Å, respectively. Contrary to common belief, in Chapter 2.1.2, an approach for exact-RFDR restraints (eRFDR), based on the eNOE-approach, was successfully developed, yielding accurate upper and lower distance limits with an accuracy of 0.38 Å.

1.3.9 Experimental elucidation of directional motion using eNOE-based multistate structure determination

Accurate distance restraints, determined as described above, are obviously very beneficial for structure determination. But maybe even more important is the application of the eNOE approach for experimental determination of spatial dynamics. [43] If the error of the determined distance restraints is smaller than the amplitude of molecular motion, the spatial dynamics can be determined by applying exact distance restraints to a specialized structure determination protocol. How the directional dynamics can be obtained by using exact distance restraints can be explained exemplarily: Due to the exponential relation between the NOE transfer rate and distance, a relative motion parallel to the vector connecting two nuclei results in a distance determined by eNOE being shorter than the actual average distance (**Figure 18 A**). Accordingly, the majority of distances are determined to be too short due to protein motion. Exceptions are distances between successive backbone protons that are located on the same site of the backbone (**Figure 18 C**), like consecutive amide protons. Here a motion around the dihedral angle can lead to an overestimation of the average distance (compare Chapters 2.1.2 and 2.2.2). Upon structure determination, the mostly underestimated distance restraints will be violated. The reason for such violation is the three-dimensional nature of the structure, in which the nuclei are restrained simultaneously in several directions in addition to chemical restraints.

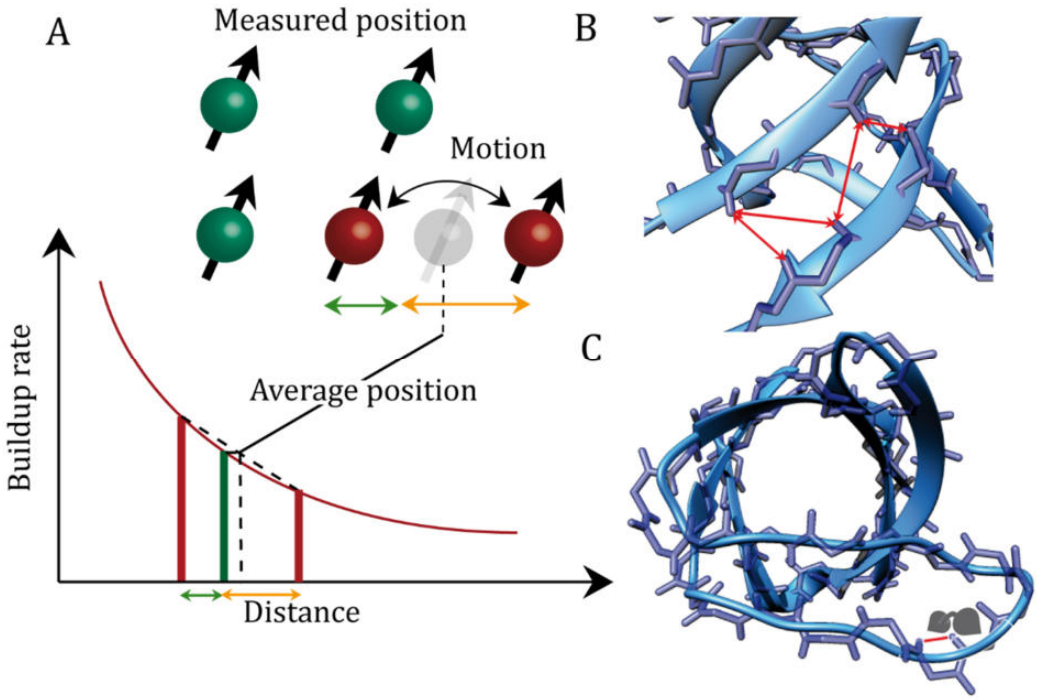


Figure 18: Representation of under- or overestimations with regard to the motional average of the internuclear distance. **A:** The NOE build-up rate decreases exponentially with the internuclear distance. If the spins show a motion parallel to their internuclear vector, the average distance between the nuclei is underestimated due to the r^{-6} relation. Such kind of motions are predominantly present between strands as shown in **B**. **C:** Overestimated internuclear distance. The average distance represents the minimum distance allowed by the chemical restraints. Angular motion in both directions lead to increasing distance and thus lower build-up rates.

In the following, it is described how these violations can be made use of. Vögeli et al. and Güntert et al. developed a structure determination protocol that reports on the spatial protein dynamics form eNOE-restraints using the software program CYANA [44]. Instead of a single structure, a multitude of structures are calculated simultaneously while the eNOE distance restraints are averaged with an r^{-6}-weighting. In other words, the distance restraint is calculated back to a rate, proportional to the magnetization-transfer efficiency by using Equation 32. This value is then averaged over the structures, calculated simultaneously, and finally, by using the individual transfer rates, distances are determined for each of the calculated structures individually. Where distance restraints are violated in case of conventional structure calculation, the individual distance restraints for the different substructures of the so-called "multistate structure calculation" will differ and thus the substructures. These simultaneously calculated sub-structures represent landmarks of the protein dynamics in a directional manner. The multistate structure determination was applied in an integrative elucidation of directional motion of the protein-binding domain SH3 in Chapter 2.2.2, representing the first experimental study on biologically relevant spatial sidechain motion.

2 Results

2.1 Methods Development for Proton-Detected Solution-like Fast- and Ultrafast Magic-Angle-Spinning Solid-State NMR Spectroscopy

2.1.1 Protein expression of the SH3 domain of chicken α-spectrin, dedicated as a common sample for NMR methods development, including the work: Assessment of a Large Enzyme-Drug Complex by Proton-Detected Solid-State NMR Spectroscopy Without Deuteration

The NMR protein sample of the SH3 domain was decisive for all projects presented here (except Chapter 2.2.3), as well as for other projects of my colleagues. As a well-behaving test protein, the samples were made available to all group members to perform NMR-methodology development. The SH3 domain was recombinantly expressed in E. coli BL21 as described before. [45] The plasmid pET3d, which encodes for the wildtype SH3 domain, was used as a vector. In addition to the SH3-Genome, this vector is equipped with an ampicillin-resistant gene. Prior to the DNA-sequence encoding for SH3, a lactose promotor is placed so that the addition of the lactose mimicry isopropyl-β-D-1-thiogalactopyranoside (IPTG) induces the expression of SH3. Cell growth took place in the minimal medium M9 at 37 °C up to an optical density (OD_{600}) of 0.6. Hereby, SH3 was expressed in double-labeled form using 1 g/L ^{15}N-labeled ammonium-chloride and uniformly ^{13}C labeled glucose. Likewise, in triple-labeled form, ^{15}N-labeled ammonium-chloride and uniformly ^{13}C- and ^{2}H-labeled Glucose were used in 100% deuterated M9 medium. When the OD_{600} of 0.6 was reached, the sample was induced with IPTG. After induction, the temperature was lowered to 22 °C for at least 12 hours. The final OD^{600} after expression was, on average, approximately 2.2. The cells were harvested by centrifugation at 3300 G and lysed by French-press. Purification by anion-exchange chromatography was performed using a Q-sepharose fast flow (QFF) column in Tris buffer at pH 8.5 (**Table 1**) prior to size-exclusion chromatography with the Superdex 75 column in citric acid buffer with a pH of 3.5 (**Table 1**).

For the liquid-state NMR sample, the protein was concentrated in a citric acid buffer to up to 2 mmol/l. Hereby, 0.03 % azide (NaN3), as well as a protease inhibitor, were added to yield the final protein sample. For solid-state NMR, after concentrating the protein to 12 µg/ml, the citric acid buffer was exchanged with water (pH 3.5) by centrifugation. A 350-mM copper EDTA (copper ethylene diamine tetra acetate) solution at pH 7.5 was added to yield a final copper EDTA concentration of 50 mM. The pH was titrated to 7.5 using 3% ammonia-solution. Crystallization started immediately (and takes at least 12 hours). Copper EDTA enhances T_1-relaxation, which reduces the interscan delay and therefore results in enormous savings in measurement time. [45]

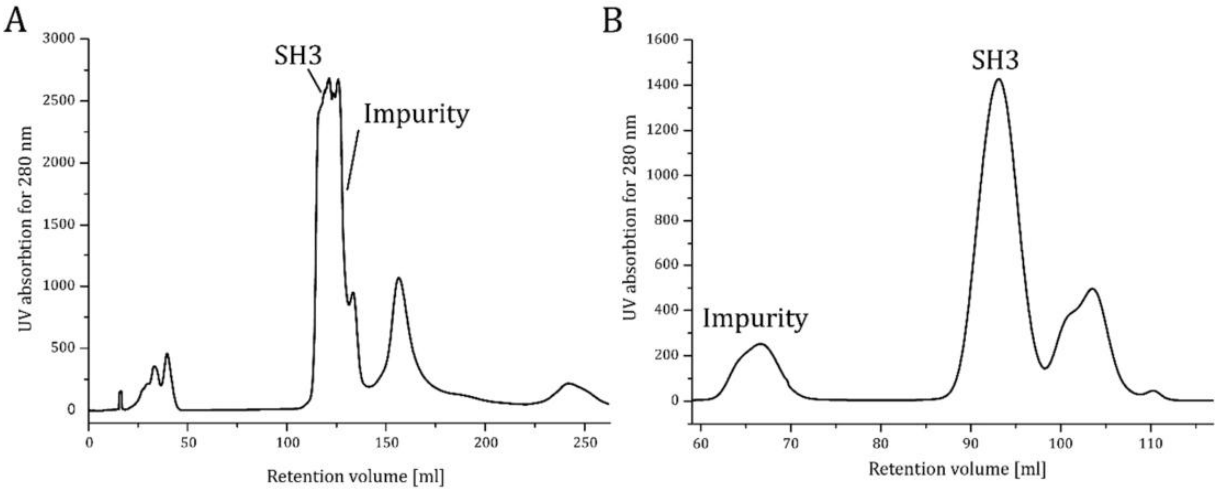


Figure 19: Two-step Fast-Flow Liquid Chromatography (FPLC) purification of SH3 using anion-exchange and size-exclusion chromatography. **A:** Anion-exchange chromatography using initially a 20-mM Tris-based buffer at pH 8.5. The initial buffer A was gradually exchanged with buffer B, containing 20 mM Tris-base and 1M NaCl at pH 8.5, over 10 column volumes (one of which is 20 ml). The peak, starting from 120 ml, was identified as SH3 using SDS-PAGE (**Figure 20**). The gradient from 0% to 100% buffer B had to be used due to special column conditions. Usually, a 0% to 10% gradient would have been advisable. **B:** Size-exclusion chromatography using a 20 mM citric acid buffer containing 150 mM NaCl, at pH 3.5. The peak at 95 ml retention volume was identified using SDS-PAGE to correspond to the final product SH3 (**Figure 20**).

Table 1: Buffer recipes for FPLC chromatography.

Buffer		pH
Anion exchange A	20 mM Tris-base	8.5
Anion exchange B	20 mM Tris-base, 1M NaCl	8.5
Size exclusion	20 mM citric acid, 150 mM NaCl	3.5

All steps of sample production were monitored by SDS-PAGE gel electrophoresis. Hereby, acrylamide gels of 12-16 % were used. Instead of the conventional running buffers, it was crucial to use a cathode buffer specifically dedicated to proteins of lower molecular weight, the Tris-Tricine buffer of **Table 2**.

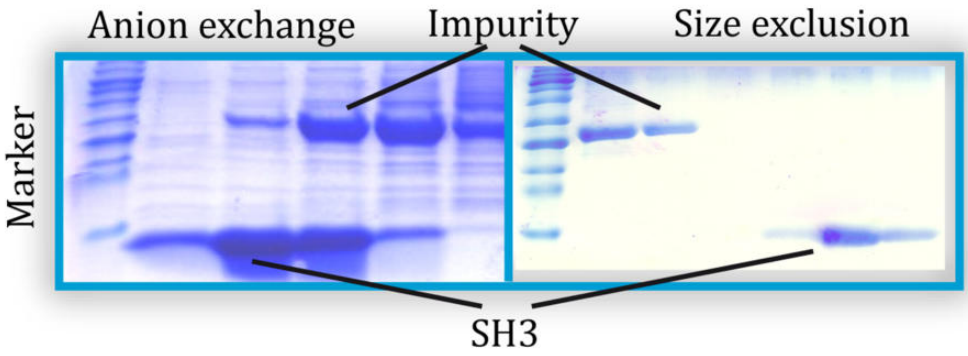


Figure 20: SDS-PAGE using Tris-tricine running buffer. **Left:** Gels after anion exchange, the lowest marker band corresponds to 11 kDa protein weight. The tracks in the center correspond to the

peak at 120 ml retention volume of the anion exchange chromatogram (**Figure 19**). **Right:** Gel after size-exclusion purification, the left tracks correspond to the peak at 90 ml retention volume in the chromatogram (**Figure 19**).

Table 2: Small molecular weight Tris-tricine gel buffers.

Buffer			pH
10X Anode buffer	1M Tris-base		8.9
10X Cathode buffer	1M Tris-base	1M Tricine	8.25

Among other projects in this thesis and the work of colleagues, the sample was used by Vasa et al. to advance biochemistry and structural biology by development and application of novel methodology in the field of ultrafast magic-angle-spinning (111 kHz) solid-state NMR spectroscopy. The methodology presented enables the structure and dynamics of large, fully protonated proteins to be elucidated. Therefore, this technique is able to fill gaps left by the analysis using other methods, like liquid-state NMR, which is limited in protein size and crystallography, and cryo-EM, both of which provide only static information. In the following work, Vasa et al. analyzed the active-site dynamics of the 29-kDa enzyme, human carbonic anhydrase II (see also Chapter 1.2.2). Besides the unique information on carbonic anhydrase II, the methods developed can be used for many other targets that fall outside the scope of other techniques.

All methods used for the large 29-kDa carbonic anhydrase II were developed and tested at the SH3 domain of chicken α-spectrin. This is exemplary shown in **Figure 21**, where the dynamics measurements of fully protonated samples under ultrafast magic-angle-spinning conditions are verified using the established approaches at fast magic-angle-spinning and deuteration. Therefore, data on deuterated and 100% back-exchanged SH3 under fast magic-angle-spinning (55.6 kHz) conditions were compared with data from protonated SH3 under ultrafast magic-angle-spinning (111 kHz) conditions. The results verify the novel approach.

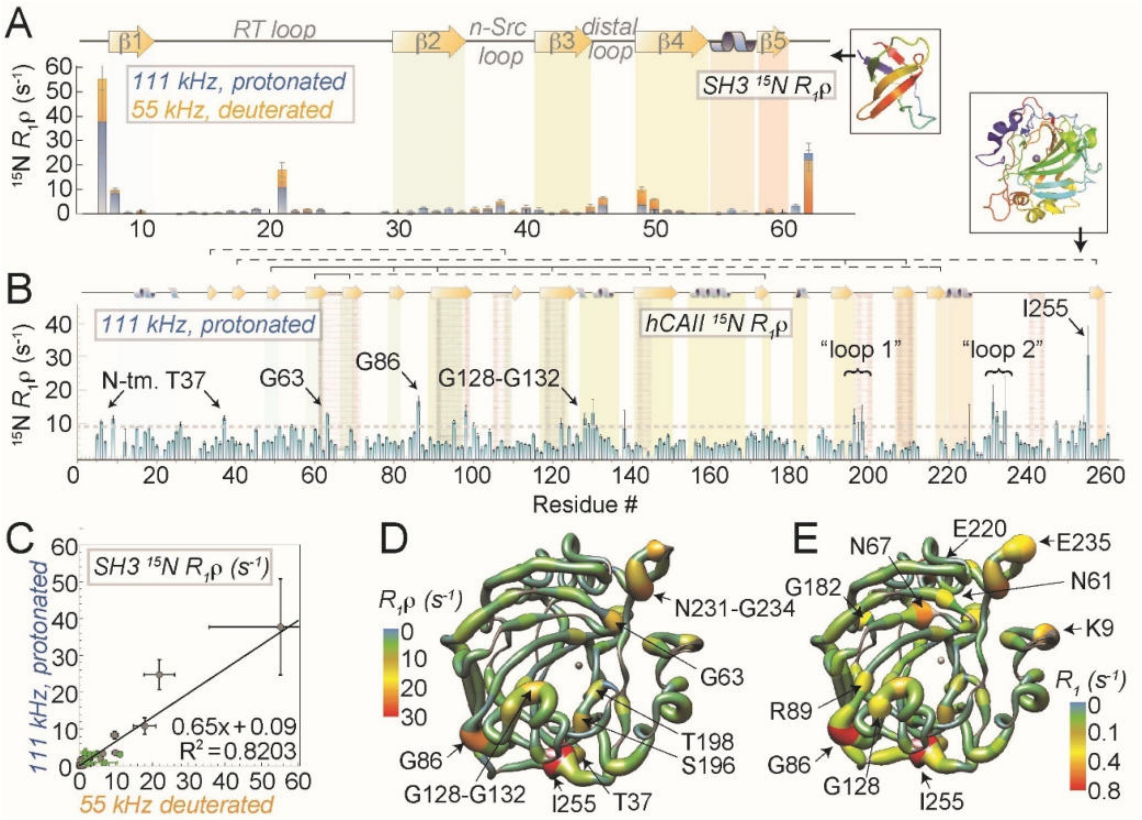


Figure 21: Backbone dynamics of fully protonated samples of SH3 and human carbonic anhydrase II (hCAII). **A:** ^{15}N $R_1\rho$-relaxation rates measured to compare fast magic-angle-spinning conditions (55.6 kHz, deuterated, 100% back-exchanged), shown in orange, and ultra-fast magic-angle-spinning conditions (111 kHz, fully protonated), shown in blue, on the model protein α-spectrin SH3 (both using 10-kHz spin-lock field strength). Secondary structural elements are shown at the top. The gray shades in **B** mark the active-site cone of hCAII. **B:** $R_1\rho$ relaxation of the hCAII-acetazolamide complex, depicted in blue. The dashed line set to 10 s^{-1} is a qualitative threshold. **C:** Correlation of the two ^{15}N $R_1\rho$-relaxation date sets from **A** (sidechain nitrogens shown in green). **D:** $R_1\rho$ rates represented in the structure of hCAII. **E:** R_1 rates represented in the structure of hCAII. This figure was created by Vasa et al. and is used with their kind permission.

Proton Detection

International Edition: DOI: 10.1002/anie.201811714
German Edition: DOI: 10.1002/ange.201811714

Assessment of a Large Enzyme–Drug Complex by Proton-Detected Solid-State NMR Spectroscopy without Deuteration

*Suresh K. Vasa, Himanshu Singh, Kristof Grohe, and Rasmus Linser**

Abstract: *Solid-state NMR spectroscopy has recently enabled structural biology with small amounts of non-deuterated proteins, largely alleviating the classical sample production demands. Still, despite the benefits for sample preparation, successful and comprehensive characterization of complex spin systems in the few cases of higher-molecular-weight proteins has thus far relied on traditional ^{13}C-detected methodology or sample deuteration. Herein we show for a 29 kDa carbonic anhydrase:acetazolamide complex that different aspects of solid-state NMR assessment of a complex spin system can be successfully accessed using a non-deuterated, 500 µg sample in combination with adequate spectroscopic tools. The shown access to protein structure, protein dynamics, as well as biochemical parameters in amino acid sidechains, such as histidine protonation states, will be transferable to proteins that are not expressible in E. coli.*

Solid-state NMR spectroscopy has been used for protein samples not amenable to solution NMR spectroscopy, such as supramolecular structures,[1] amyloid fibrils,[2] and membrane proteins.[3] A disadvantage of solid-state NMR spectroscopy is the necessity for either large sample amounts (on the order of 40 mg) or, recently, deuteration and labile-proton back-exchange of the target protein.[4] Both approaches have imposed extreme demands for protein preparation, such as being recombinantly expressible in large amounts or in deuterated fashion (usually in *E. coli*) combined with being refoldable into a native state. The dream of using minimal amounts of non-deuterated samples has been pushed forward by the development of very fast magic-angle spinning (MAS), which has been demonstrated to be useful in particular for assessments of secondary structure and shift perturbations. However, in particular for aliphatic protons, the homogeneous linewidths technically obtainable to date (relative to chemical shift dispersion) are still far from those achieved upon deuteration.[6] As such, even though full structure determination of an intermediate-size protein has been accomplished at fast MAS in the presence of protonation,[7] deuteration has seemed to be the more realistic route for assignments, structure, and dynamics.[8] To date, in-depth structural studies generally explore relatively simple scaffolds, with the few demonstrated exceptions for large proteins[8d,9] still implementing either ^{13}C detection or deuteration.

Carbonic anhydrases (CAs), a family of enzymes of around 30 kDa molecular weight, catalyze the conversion between dissolved CO_2 and bicarbonate in living cells. Small-molecule inhibitor:protein complexes of carbonic anhydrases are a representative for the large number of such samples that 1) comprise highly complex spin systems (more than 4000 nuclei), 2) are difficult to be prepared in sufficient amounts, and 3) do not completely back-exchange their amide protons in a native state.

With methodology such as higher-dimensionality experiments[4d,9c,10] and time-shared approaches,[11] developed by us and other groups in recent years, solid-state NMR spectroscopy has been expanding its capacities steadily. As such, we have recently obtained a large set of sidechain carbon and backbone assignments in perdeuterated hCAII.[10b] Here, we use a suite of tailor-made tools for exploitation of the non-exchangeable protons present in a non-deuterated complex of hCAII:acetazolamide at 111 kHz MAS. This yields comprehensive data as the basis for a wide range of structural-biology questions, including protein structure, dynamics, and characterization of biologically interesting sidechain moieties of large and complex proteins, from minimal amounts of uniformly ^{15}N/^{13}C-labeled sample. Apart from making an unprecedented range of target proteins amenable for NMR analysis, the methodology outperforms established approaches spectroscopically by adding new sources of information content.

We prepared a sample of 500 µg ^{13}C/^{15}N-labeled microcrystalline enzyme from sitting-drop crystallization trials, soaked in a multi-step procedure with the small-molecule active-site blocker acetazolamide (see details on the sample preparation procedure, assignments, general NMR methods, and pulse sequences in the Supporting Information text and Figures S1–S6). For proton-based structure calculation in the absence of deuteration, the presence of aliphatic protons leads to dipolar truncation effects,[12] but the absolute number of short- and intermediate-range restraints is expected to be high compared with deuterated samples, such that structural convergence has previously been achieved for proteins of 8 and 14 kDa.[7] For hCAII with more than twice the size, we could assign a large extent of the sidechain protons (around 700 unambiguous sidechain proton shifts) by 3D scalar-transfer-based sidechain correlations HccCH[7,13] and hCccCH using Waltz mixing (see the pulse sequences in

[*] Dr. S. K. Vasa, Dr. H. Singh, K. Grohe, Prof. Dr. R. Linser
Faculty for Chemistry and Pharmacy
Ludwig-Maximilians-University Munich
Butenandtstr. 5–13, 81377 Munich (Germany)
E-mail: rasmus.linser@lmu.de
Dr. H. Singh, Prof. Dr. R. Linser
Faculty of Chemistry and Chemical Biology
Technical University Dortmund
Otto-Hahn-Straße 4a, 44227 Dortmund (Germany)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201811714.

Figure S1 and representative strips in Figures S4–S6). In addition to the proton shifts listed in Table S3, all assignments have been deposited into the BMRB under accession code 34347. Among these, approximately 1300 homonuclear correlation peaks could be exploited for proton–proton distance-based structure calculation. These distance restraints were obtained using a single time-shared[11,14] radio-frequency-driven recoupling[15] experiment H-RFDR-hN/CH (see Figures S1 and S9 for the pulse sequence and exemplary strips, respectively). This experiment yields proton chemical shift correlations between any proximal protons irrespective of chemical nature and disperses their correlations by the shift of either the attached ^{13}C or ^{15}N atom. Using ARIA,[16] that is, without manual crosspeak assignment, 341 unambiguous and 252 ambiguous crosspeaks for the ^{15}N-edited part, in addition to 423 unambiguous and 220 ambiguous crosspeaks for the ^{13}C-edited part of the spectrum, were obtained, including 354 intra-residue, 288 medium-range, and 376 long-range contacts in total, and their intensities were converted into restraints for structure calculation. In addition, 18 chemical-shift-perturbation-based restraints were obtained from 3D hCANH shift comparison of protein in the absence and presence of ligand, enabling correct positioning of the ligand (see details on the structure calculation, including shift-perturbation-based restraints, in the Supporting Information text).

Aliphatic proton linewidths obtained at 111 kHz MAS are between 50 and 150 Hz (see Figures S7 and S8). Even though the distance restraints obtained correspond to only short contacts between 2.5 and 5 Å, the presence of aliphatic protons turned out to be a precondition for the correct convergence of the complex structure. Structure calculation of a deuterated protein (without ligand) solely based on amide-to-amide distances (data not shown) did not yield an acceptable backbone fold, with more than 8 Å deviation from the X-ray structure.

Figure 1 displays the bundle of ten lowest-energy backbone structures with a backbone RMSD of 2.2 Å, in conjunction with a (simplified) depiction of the long-range restraints obtained from the time-shared RFDR experiment, as well as the minimal-energy structure. The NMR structure bundle (PDB accession code 6QEB) is very similar to the cryogenic crystallographic structure (Figure 1B), with a well-defined core structure; however, most of the external loops are poorly restrained. Considering the large spin system, the overall structure of the protein, including the loops and turns, is remarkably well characterized by the NMR restraints (see Table S4 for structural statistics). Only the N-terminus, which yields few and poorly defined internuclear distances owing to lower-intensity cross peaks, is not sufficiently restrained to reach a converging fold. Also the position of the ligand (compare Figure 1C–E) is more ambiguous (apart from its sulfonamide warhead) on the basis of CSP-based restraints (see Figure S10), which likely reflects shortcomings of the synthetic restraints. (Acetazolamide unfortunately does not bear any non-exchanging hydrogen atoms, which would allow for further distance restraints.)

Figure 2A shows ^{15}N $R_1\rho$ relaxation data of the u-^{15}N/^{13}C-labeled hCAII:acetazolamide complex obtained at 111 kHz MAS as a function of residue (plotted on the protein structure

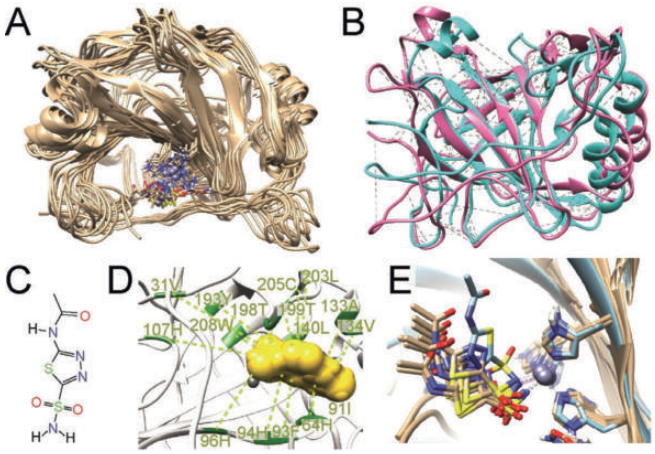


Figure 1. Structural assessment of the ^{15}N/^{13}C-hCAII:acetazolamide complex at 111 kHz MAS. A) Bundle of the ten lowest-energy structures with a backbone RMSD of 2.2 Å. B) Representation of distance restraints obtained by solid-state NMR spectroscopy (with all individual restraints from/to one amino acid depicted at the respective C^α positions), shown on an overlay between the lowest-energy NMR structure (magenta) and the crystallographic structure (PDB 3hs4, cyan). The N-terminal amino acids 1–25 are not shown. C) Lewis formula of acetazolamide, used in unlabeled form. D) Protein:ligand distance restraints derived from chemical shift perturbation. E) Close-up of the active site in the NMR structure.

in Figure 2B and shown in numerical form in Figure S16). Relaxation data in the 29 kDa complex with sufficient dispersion could successfully be recorded using a tailor-made pseudo-4D relaxation-modulated hCANH (see details in the Supporting Information). Despite the small sample amount of around 500 µg, this approach yielded high signal-to-noise relaxation decays (see Figure S16 for decay curves)

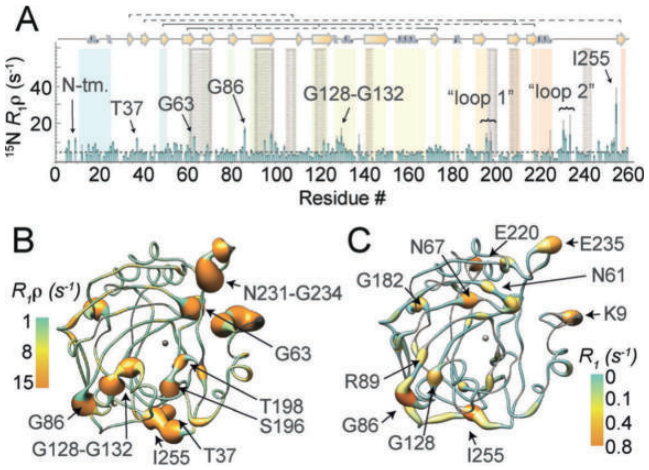


Figure 2. Backbone dynamics of hCAII assessed without deuteration. A) ^{15}N $R_1\rho$ relaxation rates of the hCAII:acetazolamide complex at 111 kHz in the absence of deuteration, using a 10 kHz ^{15}N spin lock. B) $R_1\rho$ rates plotted on the structure of hCAII. C) R_1 rates plotted on the structure of hCAII. In (A), the top row depicts secondary structural elements and sheet connectivities. Gray shades denote the active-site cone in hCAII. A dashed line, tentatively set to 5 s^{-1}, approximates the rigid-residue $R_1\rho$ baseline value.

within only one week of measurement time. Both the residues in turn regions as well as in the sheets comprising the protein core show high overall rigidity with $R_1\rho$ values of around 5 s^{-1}. Significantly higher rates of up to 30 s^{-1} are found in the entrance to the active site (residues 129–131), the tip of a C-terminal loop (residues 231–234, "loop 2"), as well as for the individual residues 37, 86, and 255. All of these residues also show elevated B-factors in both the apo and complex forms (see Figure S12). Loop 2, residues 255 and 86, as well as 9 and 220 (fast $R_1\rho$ decay, non-convergent fitting) are also prominent in the R_1 experiments, accessible in a similar fashion as $R_1\rho$ (Figures 2C, S11, and S17; see the pseudo-4D pulse sequence in Figure S1). This indicates motion on faster timescales in addition to the slow-timescale motion apparent from $R_1\rho$ rates. Interestingly, in addition to these various external sites, also G63 as well as T198 and S196 (previously coined "loop 1"[17]) show elevated $R_1\rho$ rates, indicating the existence of µs–ms timescale dynamics right within the active site. T198 is the very residue to which both the attacking nucleophile OH^- as well as the "deep water" molecule, from which the conserved water network in the pocket emanates, are hydrogen-bonded, and it is one of the crucial residues for catalysis.[18] Interestingly, the residues 231 to 234 ("loop 2") as well as residues around T198 in the active site ("loop 1") had previously been hypothesized to undergo conformational exchange in MD studies.[17] Calling for more involved studies, the observed motion under room temperature conditions suggests an influence of active-site protein dynamics on reaction specificity and kinetics, which has thus far not been considered in drug discovery studies using hCAII as a model system.[19]

Sidechain protons represent a rich source of information particularly for enzyme active sites. For example, histidine sidechain protons play a role in many catalysis processes but cannot be seen easily in X-ray crystallography. It has been shown in solution NMR spectroscopy that the tautomerization/protonation state of histidines can be assessed based on combinations of ^{15}N shifts and $^1J_{H\varepsilon/C\varepsilon}$ couplings alone.[20] Figure 3 demonstrates that a combination of short- and long-range CP-based experiments (similar to ^1J and ^2J H/N HSQC used in solution, see experimental details in the Supporting Information) as well as HNC triple-resonance experiments gives access to individual shifts as well as couplings (see the standard expectation values for these parameters in Figure 3A) of (reasonably rigid) protonated His sidechains under fast MAS. This is demonstrated for hCAII in Figure 3B and D for two exemplary histidines. Hε1/Cε1 assignments had been obtained previously by specific labeling.[5] Based on two (time-shared) experiments (Figure 3E,F; see also the Supporting Information text and Figure S13), providing ^{13}C-filtered one- or two-bond H/N correlations as well as ^{15}N-filtered one- or two-bond H/C

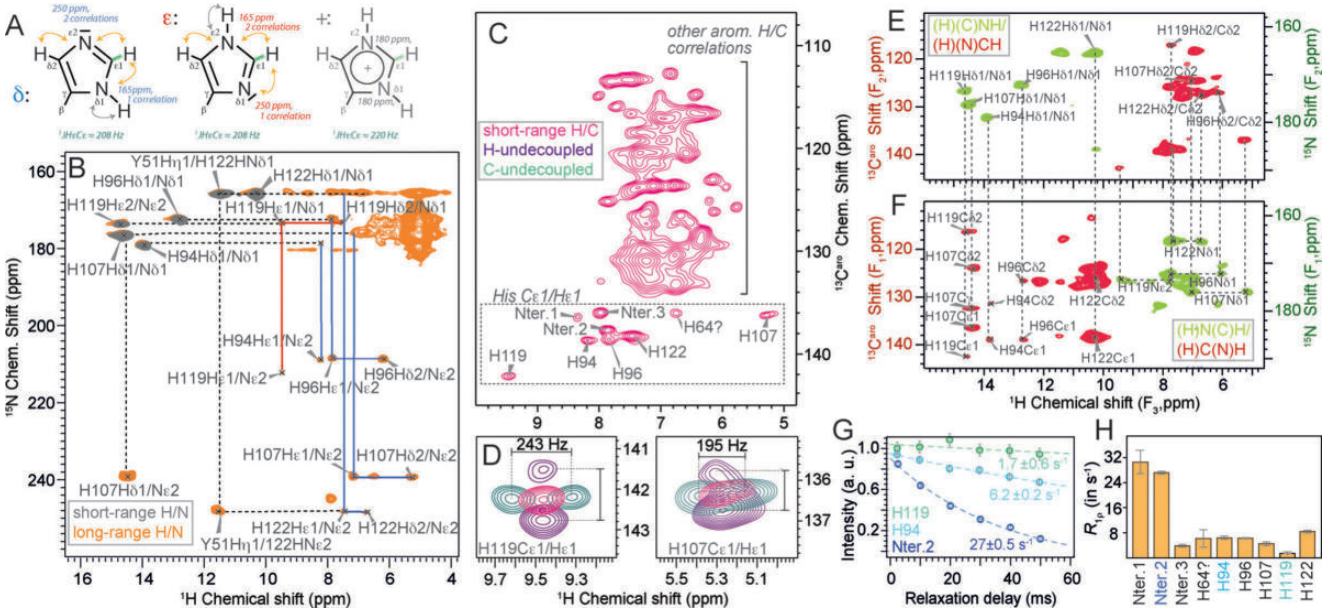


Figure 3. Access to His sidechain protonation states via non-exchangeable protons at 111 kHz MAS. A) Neutral histidine tautomers δ and ε, as well as the protonated form + with expected $^1J_{H\varepsilon C\varepsilon}$ couplings, ^{15}N shifts, and likely H/N correlations. Gray and orange arrows denote one- and two-bond correlations, respectively. B) Tautomerization states assessed via ^{15}N shifts in H/N correlations using long-range dipolar transfers (orange), overlaid with a short-CP correlation (gray). Patterns denoting δ and ε states are marked in blue and red, respectively. C) Aromatic H/C correlation with assigned histidine Cε1/Hε1 peaks. Residue-specific assignments according to Shimahara et al.[5] D) Non-decoupled H/C spectra, obtained either without ^1H (purple) or without ^{13}C decoupling (green) for assessment of Hε1/Cε1 ^1J couplings. E) Filtered (one-bond) H/N and H/C correlations, obtained by a time-shared (H)(C)NH/(H)(N)CH experiment, revealing His Cδ2/Hδ2 correlations. F) Time-shared two-bond correlation (H)N(C)H/(H)C(N)H, correlating ring carbon shifts with ring H^N shifts and ring nitrogen shifts with ring H^C shifts. In (E) and (F), the N-edited correlations are depicted in green, the C-edited ones in red (see Figure S13 for transfer pathways of the experiments). The two experiments were conducted using different 2D versions based on the same 3D pulse sequence (Figure S1). G,H) Representative $^{13}C\varepsilon1$ $R_1\rho$ decay curves and bar chart of His $R_1\rho$ rates, respectively. Whereas the five core histidine residues H94, 96, 107, 119, and 122 were unambiguously assigned on the basis of the above experiments, the N-terminal His residues are ambiguous, and H64 is a tentative assignment. All experiments were conducted at an effective temperature of 15 °C.

correlations, we can here unambiguously assign all δ1, δ2, ε1, and ε2 ^{1}H, ^{15}N, and ^{13}C shifts for the core histidine residues 94, 96, 107, 119, and 122. As one example, H119, with no uncoordinated-nitrogen shifts (expected at ca. 240 ppm) and an increased $^{1}J_{\mathrm{He/C\varepsilon}}$ coupling, is one of the Zn^{2+}-bonded bases besides H94 and H96, rendering it similar to the protonated form. By contrast, with a small *J* coupling and two carbon-bonded protons in direct vicinity to an unprotonated ^{15}N, H107 is a neutral histidine and in the δ state, with insignificant fraction of protonated form at physiological pH. This neutral protonation state, which allows it to form a hydrogen bond to E117 and to be an acceptor towards Y193-OH, explains the rigidity of the involved amino acids and thus helps inducing a well-defined protein core. Indeed, mutation of this residue has been linked to a CA misfolding disease and abrogates function, even though H107 is not directly involved in catalysis.[21] As shown in Figure S14, the protonation state of neither of the residues changes upon ligand binding.

In addition, particularly the non-exchangeable protons become accessible as valuable reporters on histidine sidechain dynamics. ^{13}Cε $R_1\rho$ rates are assessed in Figure 3 G, H (see Figure S15 for all decay curves). Whereas the (incompletely assigned) N-terminal His residues expectedly show larger mobility, remarkably, slight differences in dynamics can be detected also within the three Zn-coordinating residues 94, 96, and 119. Even though mutations may still facilitate unambiguous aromatic sidechain assignments in larger proteins, the possibility of the above assessments owing to non-exchangeable protons, which are neither feasible for X-ray crystallography nor for solid-state NMR spectroscopy with either of the established approaches, may be of great value for addressing mechanistic questions in future structural-biology studies.

These results show that with suitable spectroscopic tools the pool of non-exchangeable protons can become exploitable for a comprehensive range of structural-biology aspects in proteins of noticeable size. Part of the demonstrated aspects will be transferable to chemical compounds and materials for which proton shift accessibility will be valuable because of limitations in carbon/nitrogen labeling. Whereas proton linewidths for fully protonated samples at 111 kHz MAS are not reaching the values possible upon deuterium dilution (RAP[22]) or spinning even faster, such (future) alternatives also decrease the available signal-to-noise ratios. Additionally, apart from the preparative feasibility of fully protonated samples, the possibility for partial deuteration or even faster MAS to further reduce proton linewidths (counteracting a signal-to-noise reduction) is strongly dependent on the sample preparation. These arguments in mind, with aliphatic protons becoming available for solid-state NMR-based structural biology, the development of spectroscopic methods to overcome the current hurdles, such as dipolar truncation or signal overlap, seems to be the most fruitful endeavor.

We have reported on the fast-MAS solid-state NMR-based characterization of structure, dynamics, and sidechain protonation states of a u-^{13}C/^{15}N-labeled sample of human carbonic anhydrase II in complex with a small-molecule inhibitor, using 500 μg of sample at 111 kHz MAS. Despite the complexity of the spin system comprising more than 4000 atoms, a well-resolved complex structure was obtained based on structural restraints from non-exchangeable protons and ligand shift perturbations. Residue-specific relaxation properties demonstrate not only dynamics in the outer loops but also the presence of conformational exchange in the active site. Owing to the availability of sidechain protons and a set of tailored aromatic experiments, access was also obtained to sidechain chemistry information. The methodology employed here will be transferable to proteins that cannot be obtained in high amounts or deuterated/refolded fashion, such as those expressed in insect cells or tissue culture.

Acknowledgements

We are grateful to Petra Rovó for constructive interactions as well as the Gerhard Klebe group (Uni Marburg) for the plasmid, protocols, and helpful discussions and Karin Giller, Stefan Becker, and Eszter Najbauer (MPIbpc Göttingen) for initial efforts in this project. Financial support from the Deutsche Forschungsgemeinschaft (SFB 749, TP A13, SFB 1309, TP 03, Emmy Noether program), the Verband der Chemischen Industrie (VCI, Liebig program), the Excellence Clusters CiPS-M and RESOLV, and the Center for Nano-Science (CeNS) is acknowledged.

Conflict of interest

The authors declare no conflict of interest.

Keywords: protein function · protein structure · proton detection · solid-state NMR spectroscopy · ultrafast MAS

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 5758–5762
Angew. Chem. **2019**, *131*, 5814–5819

[1] a) A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, A. Lange, *Nature* **2012**, *486*, 276–279; b) H. Zhang, G. Hou, M. Lu, J. Ahn, I.-J. L. Byeon, C. J. Langmead, J. R. Perilla, I. Hung, P. L. Gor'kov, Z. Gan, W. W. Brey, D. A. Case, K. Schulten, A. M. Gronenborn, T. Polenova, *J. Am. Chem. Soc.* **2016**, *138*, 14066–14075; c) L. He, B. Bardiaux, M. Ahmed, J. Spehr, R. König, H. Lünsdorf, U. Rand, T. Lührs, C. Ritter, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E272–E281.

[2] a) M. T. Colvin, R. Silvers, Q. Z. Ni, T. V. Can, I. Sergeyev, M. Rosay, K. J. Donovan, B. Michael, J. Wall, S. Linse, R. G. Griffin, *J. Am. Chem. Soc.* **2016**, *138*, 9663–9674; b) M. D. Tuttle, G. Comellas, A. J. Nieuwkoop, D. J. Covell, D. A. Berthold, K. D. Kloepper, J. M. Courtney, J. K. Kim, A. M. Barclay, A. Kendall, W. Wan, G. Stubbs, C. D. Schwieters, V. M. Y. Lee, J. M. George, C. M. Rienstra, *Nat. Struct. Mol. Biol.* **2016**, *23*, 409–415; c) M. Aulikki Wälti, F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall, A. Böckmann, P. Güntert, B. H. Meier, R. Riek, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E4976–E4984; d) T. Theint, P. S. Nadaud, D. Aucoin, J. J. Helmus, S. P. Pondaven, K. Surewicz, W. K. Surewicz, C. P. Jaroniec, *Nat. Commun.* **2017**, *8*, 753.

[3] a) O. Saurel, I. Iordanov, G. Nars, P. Demange, T. Le Marchand, L. B. Andreas, G. Pintacuda, A. Milon, *J. Am. Chem. Soc.* **2017**, *139*, 1590–1597; b) D. Good, C. Pham, J. Jagas, J. R. Lewan-

dowski, V. Ladizhansky, *J. Am. Chem. Soc.* **2017**, *139*, 9246 – 9258; c) L. Joedicke, J. Mao, G. Kuenze, C. Reinhart, T. Kalavacherla, H. R. A. Jonker, C. Richter, H. Schwalbe, J. Meiler, J. Preu, H. Michel, C. Glaubitz, *Nat. Chem. Biol.* **2018**, *14*, 284.

[4] a) V. Chevelkov, K. Rehbein, A. Diel, B. Reif, *Angew. Chem. Int. Ed.* **2006**, *45*, 3878 – 3881; *Angew. Chem.* **2006**, *118*, 3963 – 3966; b) M. Huber, S. Hiller, P. Schanda, M. Ernst, A. Böckmann, R. Verel, B. H. Meier, *ChemPhysChem* **2011**, *12*, 915 – 918; c) M. J. Knight, A. J. Pell, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, T. Herrmann, L. Emsley, G. Pintacuda, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11095 – 11100; d) M. Zinke, P. Fricke, C. Samson, S. Hwang, J. Wall, S. Lange, S. Zinn-Justin, A. Lange, *Angew. Chem. Int. Ed.* **2017**, *56*, 9497 – 9501; *Angew. Chem.* **2017**, *129*, 9625 – 9629; e) S. K. Vasa, P. Rovó, R. Linser, *Acc. Chem. Res.* **2018**, *51*, 1386 – 1395; f) P. Rovó, C. A. Smith, D. Gauto, B. L. de Groot, P. Schanda, R. Linser, *J. Am. Chem. Soc.* **2019**, *141*, 858 – 869.

[5] H. Shimahara, T. Yoshida, Y. Shibata, M. Shimizu, Y. Kyogoku, F. Sakiyama, T. Nakazawa, S.-i. Tate, S.-y. Ohki, T. Kato, H. Moriyama, K.-i. Kishida, Y. Tano, T. Ohkubo, Y. Kobayashi, *J. Biol. Chem.* **2007**, *282*, 9646 – 9656.

[6] a) A. Böckmann, M. Ernst, B. H. Meier, *J. Magn. Reson.* **2015**, *253*, 71 – 79; b) K. Xue, R. Sarkar, C. Motz, S. Asami, D. C. R. Camargo, V. Decker, S. Wegner, Z. Tosner, B. Reif, *Sci. Rep.* **2017**, *7*, 7444.

[7] L. B. Andreas, K. Jaudzems, J. Stanek, D. Lalli, A. Bertarello, T. Le Marchand, D. Cala-De Paepe, S. Kotelovica, I. Akopjana, B. Knott, S. Wegner, F. Engelke, A. Lesage, L. Emsley, K. Tars, T. Herrmann, G. Pintacuda, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9187 – 9192.

[8] a) N.-A. Lakomek, S. Penzel, A. Lends, R. Cadalbert, M. Ernst, B. H. Meier, *Chem. Eur. J.* **2017**, *23*, 9425 – 9433; b) V. Agarwal, S. Penzel, K. Szekely, R. Cadalbert, E. Testori, A. Oss, J. Past, A. Samoson, M. Ernst, A. Böckmann, B. H. Meier, *Angew. Chem. Int. Ed.* **2014**, *53*, 12253 – 12256; *Angew. Chem.* **2014**, *126*, 12450 – 12453; c) H. R. W. Dannatt, M. Felletti, S. Jehle, Y. Wang, L. Emsley, N. E. Dixon, A. Lesage, G. Pintacuda, *Angew. Chem. Int. Ed.* **2016**, *55*, 6638 – 6641; *Angew. Chem.* **2016**, *128*, 6750 – 6753; d) B. Bersch, J. M. Dörr, A. Hessel, J. A. Killian, P. Schanda, *Angew. Chem. Int. Ed.* **2017**, *56*, 2508 – 2512; *Angew. Chem.* **2017**, *129*, 2549 – 2553.

[9] a) S. H. Park, B. B. Das, F. Casagrande, Y. Tian, H. J. Nothnagel, M. Chu, H. Kiefer, K. Maier, A. A. De Angelis, F. M. Marassi, S. J. Opella, *Nature* **2012**, *491*, 779 – 783; b) J. S. Retel, A. J. Nieuwkoop, M. Hiller, V. A. Higman, E. Barbet-Massin, J. Stanek, L. B. Andreas, W. T. Franks, B.-J. van Rossum, K. R. Vinothkumar, L. Handel, G. G. de Palma, B. Bardiaux, G. Pintacuda, L. Emsley, W. Kühlbrandt, H. Oschkinat, *Nat. Commun.* **2017**, *8*, 2073; c) H. Fraga, C.-A. Arnaud, D. F. Gauto, M. Audin, V. Kurauskas, P. Macek, C. Krichel, J.-Y. Guan, J. Boisbouvier, R. Sprangers, C. Breyton, P. Schanda, *ChemPhysChem* **2017**, *18*, 2697 – 2703; d) T. Wiegand, R. Cadalbert, C. Gardiennet, J. Timmins, L. Terradot, A. Böckmann, B. H. Meier, *Angew. Chem. Int. Ed.* **2016**, *55*, 14164 – 14168; *Angew. Chem.* **2016**, *128*, 14370 – 14375.

[10] a) S. Xiang, J. Biernat, E. Mandelkow, S. Becker, R. Linser, *Chem. Commun.* **2016**, *52*, 4002 – 4005; b) S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J. Phys. Chem. Lett.* **2018**, *9*, 1307 – 1311.

[11] R. Linser, B. Bardiaux, V. Higman, U. Fink, B. Reif, *J. Am. Chem. Soc.* **2011**, *133*, 5905 – 5912.

[12] M. G. Jain, D. Lalli, J. Stanek, C. Gowda, S. Prakash, T. S. Schwarzer, T. Schubeis, K. Castiglione, L. B. Andreas, P. K. Madhu, G. Pintacuda, V. Agarwal, *J. Phys. Chem. Lett.* **2017**, *8*, 2399 – 2405.

[13] V. Agarwal, B. Reif, *J. Magn. Reson.* **2008**, *194*, 16 – 24.

[14] C. Shi, P. Fricke, L. Lin, V. Chevelkov, M. Wegstroth, K. Giller, S. Becker, M. Thanbichler, A. Lange, *Sci. Adv.* **2015**, *1*, e1501087.

[15] A. E. Bennett, J. H. Ok, S. Vega, R. G. Griffin, *J. Chem. Phys.* **1992**, *96*, 8624 – 8627.

[16] W. Rieping, M. Habeck, B. Bardiaux, M. Bernard, T. E. Malliavin, M. Nilges, *Bioinformatics* **2007**, *23*, 381 – 382.

[17] H. Ma, A. Li, K. Gao, *ACS Omega* **2017**, *2*, 8414 – 8420.

[18] S. Z. Fisher, M. Aggarwal, A. Y. Kovalevsky, D. N. Silverman, R. McKenna, *J. Am. Chem. Soc.* **2012**, *134*, 14726 – 14729.

[19] R. Gaspari, C. Rechlin, A. Heine, G. Bottegoni, W. Rocchia, D. Schwarz, J. r. Bomke, H.-D. Gerber, G. Klebe, A. Cavalli, *J. Med. Chem.* **2016**, *59*, 4245 – 4256.

[20] a) A. L. Hansen, L. E. Kay, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1705 – E1712; b) N. Shimba, Z. Serber, R. Ledwidge, S. M. Miller, C. S. Craik, V. Dötsch, *Biochemistry* **2003**, *42*, 9227 – 9234.

[21] a) K. Almstedt, M. Lundqvist, J. Carlsson, M. Karlsson, B. Persson, B.-H. Jonsson, U. Carlsson, P. Hammarström, *J. Biol. Chem.* **2004**, *342*, 619 – 633; b) P. J. Venta, R. J. Welty, T. M. Johnson, W. S. Sly, R. E. Tashian, *J. Hum. Genet.* **1991**, *49*, 1082 – 1090.

[22] S. Asami, P. Schmieder, B. Reif, *J. Am. Chem. Soc.* **2010**, *132*, 15133 – 15135.

Manuscript received: October 11, 2018
Revised manuscript received: January 7, 2019
Accepted manuscript online: January 28, 2019
Version of record online: February 28, 2019

2.1.2 Exact distance measurements for structure and dynamics in solid proteins by fast magic-angle-spinning NMR

Progress regarding determination of protein structure and dynamics address an urgent need in structural biology and biochemistry. In this context, magic-angle-spinning solid-state NMR spectroscopy has emerged as an important tool complementing other methods like X-ray crystallography, solution-state NMR spectroscopy, and cryo-electron microscopy. Thus, many proteins, such as membrane proteins, proteins that exceed certain size limits, or protein fibrils, can be uniquely addressed by solid-state NMR. The determination of protein structure by NMR is mainly based on the determination of internuclear distances, which are converted into restraints for structure calculation. So far, solid-state NMR spectroscopy has only been able to provide qualitative upper-limit restraints by sorting the determined distances into categories such as near (1.8-4 Å), medium (1.8-6 Å), or far (1.8-8 Å). The development of more precise approaches has been circumvented by the complex nature of through-space magnetization transfer, which is not yet fully understood yet in theory. Contrary to common belief in the field, the following work shows that an interrelated approach for error minimization during all steps of polarization transfer in practice can determine highly accurate distance restraints with tight upper and lower boundaries of 0.38 Å accuracy. In the future, this approach could be combined with the ultrafast magic-angle-spinning approaches presented in chapter 2.1.1 and the elucidation of spatial dynamics of Chapter 2.2.2, yielding novel information about structure and dynamics of large-scale protein targets.

Contributions by co-authors:

Dr. Evgeny Nimerovsky provided the software for numerical simulations; Dr. Himanshu Singh produced the human carbonic anhydrase II sample (supporting information); Dr. Suresh Kumar Vasa performed the assignments of carbonic anhydrase II (supporting information); Benedikt Söldner performed the assignments of RFDR by spatial correlations of carbonic anhydrase II (supporting information); Prof. Beat Vögeli assisted in the software application of eNORAII; Prof. M. Rienstra designed the fundamentals of the research pursued by Dr. Evgeny Nimerovsky.

ChemComm

COMMUNICATION

View Article Online
View Journal

Cite this: DOI: 10.1039/c9cc02317h

Received 25th March 2019,
Accepted 28th May 2019

DOI: 10.1039/c9cc02317h

rsc.li/chemcomm

Exact distance measurements for structure and dynamics in solid proteins by fast-magic-angle-spinning NMR†

Kristof Grohe,[ab] Evgeny Nimerovsky,[c] Himanshu Singh,[ab] Suresh K. Vasa,[ab] Benedikt Söldner,[a] Beat Vögeli,[d] Chad M. Rienstra[c] and Rasmus Linser*[ab]

Fast-magic-angle-spinning solid-state NMR is a developing technique for determination of protein structure and dynamics. Proton–proton correlations usually lead to rough distance restraints, a serious hurdle towards high-resolution structures. Analogous to the "eNOE" concept in solution, an integrative approach for more accurate restraints enables improved structural accuracy with minimal analytical effort.

In the past decades, magic-angle-spinning (MAS) solid-state NMR spectroscopy has made fast progress regarding determination of structure and dynamics of insoluble proteins or large protein complexes.[1,2] Recently, proton-detected solid-state NMR on perdeuterated and proton back-exchanged or even fully protonated samples has enabled solid-state NMR structures based on miniscule sample amounts.[3,4] Till-date, homonuclear magnetization transfer has represented the most important tool for structure determination generally.[5–7] In case of proton detection, sensitive ^{1}H–^{1}H through-space correlations are obtained *via* mixing like RFDR[8] or DREAM,[9] yielding proton–proton inter-nuclear distance restraints like in solution NMR spectroscopy.[10–13] Similarly as in both solution and conventional solid-state NMR, polarization is transferred among the interacting spins through space, with the efficacy of the transfer being modulated by the inter-nuclear distance. Commonly, the amount of transferred magnetization is read out from cross peak intensities or volumes at one particular mixing time and translated into qualitative restraints (distance ranges). This approach leads to rough estimation of distances and is hampered by various errors. The accuracy is compromised by site-specific relaxation, differential transfer efficiency during CP steps, and offset-dependent pulse imperfections for each of the involved nuclei. Additionally, magnetization is transferred in substantial amounts *via* third spins in terms of spin diffusion (or relay transfers). In conventional solid-state NMR, time-resolved analysis of dephasing curves upon recoupling of isolated spin pairs has been a more accurate alternative.[14,15] In solution NMR, determination of the time-resolved magnetization buildup upon compensation of artifacts has been used for more accurate distance restraints.[16] We wondered if proton-detected solid-state NMR structure elucidation could benefit from similar concepts.

Fig. 1 depicts the buildup of such a homonuclear proton–proton correlation, as obtained *via* a series of 3D ^{15}N-edited RFDR spectra (H-RFDR-hNH)[10,11] from a deuterated ^{15}N, ^{13}C-labeled sample of chicken α-spectrin SH3 domain, micro-crystallized in

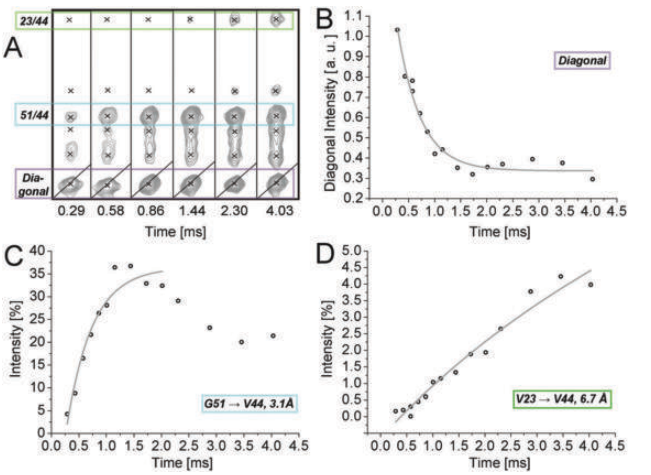


Fig. 1 Experimental RFDR buildup in a micro-crystalline sample of chicken α-spectrin SH3 domain. (A) Strips from ^{15}N-edited RFDR spectra (H-RFDR-hNH) recorded at different mixing times. The peaks highlighted in green and blue correspond to G51 (3.1 Å) and V23 (6.7 Å) amide magnetization, respectively, transferred to the amide of V44. (B) Diagonal decay of the amide proton magnetization of V44. (C and D) Cross-peak intensity (relative to the diagonal peak intensity at zero mixing time) of cross peaks 51 → 44 and 23 → 44, respectively, as a function of mixing time. (Data recorded on perdeuterated protein at 55 kHz MAS.)

[a] *Faculty for Chemistry and Pharmacy, Ludwig-Maximilians-University Munich, Butenandtstr. 5-13, 81377 Munich, Germany. E-mail: rasmus.linser@lmu.de*
[b] *Faculty of Chemistry and Chemical Biology, Technical University Dortmund, Otto-Hahn-Straße 4a, 44227 Dortmund, Germany*
[c] *Department of Chemistry, University of Illinois, 600 South Mathews Avenue, Urbana, IL 61801, USA*
[d] *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, 12801 East 17th Avenue, Aurora, CO 80045, USA*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9cc02317h

100% H_2O and spun at 55.5 kHz at 700 MHz proton Larmor frequency (see below for more details).

The time-resolved experimental intensities reflect the expected dependence of the buildup rates on internuclear distances. Even for rate-based restraints, still, the above-mentioned sources of errors have to be addressed in order to turn qualitative through-space correlations into accurate distance restraints. Taking the eNOE framework developed by Vögeli and Riek for solution NMR NOEs[16] as a template, this can be achieved as described in the following (see flowchart of data processing in Fig. 2): (i) the differential polarization transfer efficiency during the heteronuclear correlation part of the experiment is compensated for by normalizing the cross-peak intensities by the corresponding extrapolated diagonal-peak intensities at zero mixing time, using a mono-exponential fit. (ii) Site-specific magnetization loss during mixing is compensated by taking the diagonal-decay rate as a fixed parameter correcting the corresponding cross-peak buildup. (iii) The normalized intensities are then corrected for indirect polarization transfer mediated by nearby spins (relay transfers). The correction factors for each cross-peak intensity can be estimated by simulations using a transfer matrix (ESI†). The correction factor represents the ratio of the simulated two-spin buildup and the sum of all simulations, taking third spins into account (eqn (S4.1) and (S4.2), ESI†). The matrix contains decay rates as diagonal and buildup rates as off-diagonal elements. Buildup rates are simulated using the structural model and decay rates are taken from the diagonal peaks if available. This has been described in detail for eNOEs.[17] (iv) In case cross peaks from both transfers are available, the average of the build-up rate of cross-peak $H_j \rightarrow H_i$ and $H_i \rightarrow H_j$ is taken and converted into a ''bidirectional'' distance restraint. (This yields improved reliability over the uni-directional restraint, also compare Fig. 4A.) In terms of correction for relayed magnetization transfer an initial (approximate) structural model has to be provided, which can be generated from the data by using uncorrected restraints. The resulting restraints should now bear high accuracy, and a structure with improved resolution can be calculated.

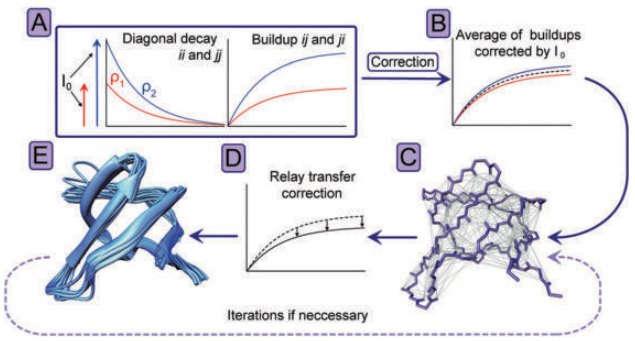


Fig. 2 Processing flowchart. (A) Schematic representation of diagonal decays (left) and cross-peak buildups (right), where I_0 stands for the extrapolated intensity at zero mixing-time and ρ for the diagonal-peak decay-rate. (B) Representation of buildups corrected by I_0. (C) All internuclear contacts within 8 Å depicted on the backbone of SH3. (D) Correction for relayed magnetization transfer *via* third spins using a transfer matrix. (E) Solid-state NMR structure ensemble generated with the distance restraints corrected for relayed magnetization transfer. For further iterative refinement, the improved average structure can be again used for correction of relayed magnetization transfer, resulting in improved restraints for structure calculation.

Such analysis has similarly been implemented in an automated manner for solution state NMR in the freely available (and editable) MATLAB-based program eNORA2.[18] We wondered whether the routines for artefact compensation of the eNOE-approach can be applied as such for solid-state NMR, despite the obvious differences in physical and technical details as to how the magnetization transfer is achieved (see transfer-theoretical details in the ESI†). *E.g.*, the NOE is two to three orders of magnitude slower and void of pulses (and related losses) during transfer. Secondly, whereas for NOEs, zero- and double-quantum relaxation terms imply a distance-proportionality of r^{-6}, the first-order dipolar recoupling Hamiltonian during RFDR is dependent on r^{-3} terms. (This also applies to any corrections for relayed magnetization transfer.) Starting from the eNORA framework, the relation between rates and distances, as well as the start parameters for fitting, need thus to be modified. Maybe most interestingly, even though the experimental diagonal and cross peaks show a seemingly exponential decay and buildup behavior on first glance, from a theoretical point of view, the complex RFDR polarization transfer rather shows a Bessel function-like behavior.[19,20] At this point, such complex functions and their fitting are constitutively impractical for the automated framework in focus. In order to validate that a simple function like an mono-exponential buildup represents a good approximation to the more complicated behavior during RFDR, we performed numerical powder-averaged simulations of two-spin RFDR magnetization buildup and diagonal-decay curves as done before,[21] using an in-house MATLAB-based program[22] especially dedicated for finite-pulse RFDR[23] (see Fig. 3 and the ESI†).

Successively, the simulated curves were ''test-fitted'' in the initial regime (buildup until the first maximum) with linear, exponential, and trigonometric functions (see Fig. S7A–C, ESI†). Focusing on this regime omits the oscillations at long mixing times, which are more corrupted by dipolar truncation and spin diffusion. In fact, in all cases the fit (of the slope, the exponential rate, or the inverse frequency, respectively) over distance showed a correlation with an R^2 value of above 0.99 (see more details to this analysis in the ESI†). Evidently, the reason for the nearly perfect correlation of simple functions with the simulated data in the initial regime is that any errors, *i.e.*, any dissimilarities between buildup behavior and fitting function, are similar for all buildups and therefore eliminated by analyzing relative trends. This holds true as long as the fitting is performed uniformly up to a comparable point like the first maximum.

The above simulations prove that in order to convert buildups from RFDR into distances, simple measures like exponential fitting in the initial regime are a good practical approximation. Consequently, exact distance restraints can be determined in a straightforward way using a suitably modified version of the program eNORA2 (for details of modification see the ESI†). As a test case, the RFDR buildup data of the SH3 sample was processed within this framework, using a modified eNORA2 routine with an

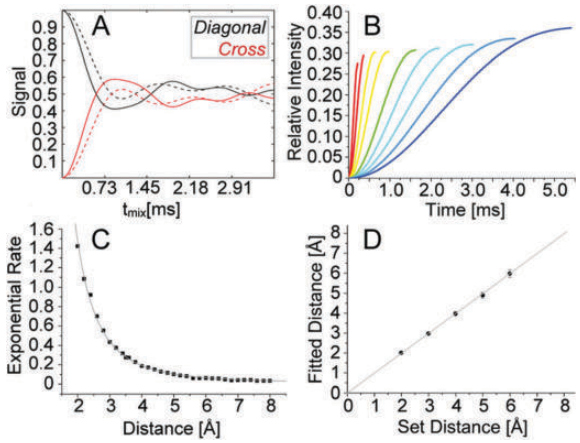


Fig. 3 Validation of simplified fitting using two-spin simulations. (A) Simulated diagonal-peak intensity (black) and cross peak buildup (red) for a proton–proton spin pair of 3 Å distance, assuming only the first-order term (dashed line) or taking into account higher-order terms (solid line). (B) Simulated initial-regime buildups (including higher-order terms) for the proton–proton distances of 2.0, 2.4, 2.8, 3.2, 3.8, 4.2, 4.6, 5.0 and 5.4 Å. (C) Depiction of extracted buildup rates as obtained for various distances by simple mono-exponential fitting of simulations (black symbols). The distance dependence of the fit parameter matches a calibrated r^{-3} function (gray) with a correlation coefficient R^2 of 0.99. (D) Verification of the fitting procedure of eNORA2 modified for RFDR using simulated buildup and decay curves. Simulations for 2 to 6 Å were used as the "experimental" input to verify the fitting procedure. Determined distances and distances set for the simulation correlate linearly.

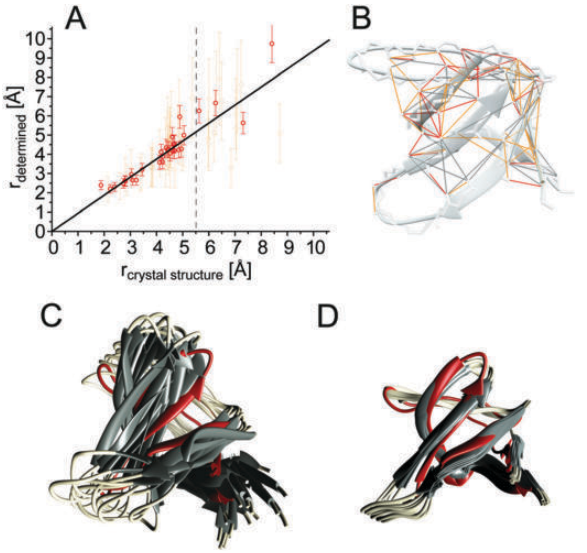


Fig. 4 Improvement of structure determination using the eRFDR-approach. (A) Correlation between bidirectional (red) as well as unidirectional (orange) exact-RFDR restraints (depicted on the structure in (B)) and the corresponding distances read out from the crystal structure (2NUZ). Distance errors are employed as described for eNORA2 previously.[25] The R^2 value of the fit is 0.96 for distances smaller than 5.5 Å. Values for larger distances are less accurate. The deviation of the short distance HN18 ↔ HN19 (leftmost point) is likely due to molecular dynamics in this loop[26] (also compare Fig. S10, ESI†). (B) Depiction of the obtained eRFDR-restraints, color-coded as red (bidirectional) and orange (uni-directional). (C) Ensemble of the 10 lowest-energy structures from 300 calculated structures using 112 unambiguous conventional upper-distance restraints in addition to angular restraints. For determining the upper distance limit of conventional restraints, the restraints were grouped with respect to the ratio of cross and diagonal peak. Proton distances corresponding to large cross-peak intensity ratios were treated as close (5 Å), medium ratios as 7.5 Å, and small ratios as far (9 Å), in addition to angular restraints. (D) Structure calculation based on 47 conventional RFDR upper distance restraints (for peaks with insufficient resolution in their diagonal), 35 unidirectional eRFDR restraints and 30 bidirectional eRFDR restraints.

r^{-3} distance dependency (based on exponential fitting functions). For correcting for relayed magnetization transfers based on the transfer matrix approach, we used the average structure of the 10 lowest-energy structures (see Fig. 4C) obtained from the standard RFDR structure elucidation protocol. The correction involved all nearby spins within spheres of 12 Å radius centered at each of the two spins of interest.

We indeed obtained distances that correlate very well with the distances read out from the corresponding crystal structures 2NUZ. The RMSD for all restraints shorter than 5.5 Å that both transfers ($H_j \leftrightarrow H_i$) are available for is 0.38 Å (also compare Fig. 4A). Very similar results were obtained for a second, larger test case, the human carbonic anhydrase II (hCAII)[24] with a molecular weight of 29 kDa. Here the obtained exact ("eRFDR") distances show a good correlation (RMSD of 0.29 Å for bidirectional restraints shorter than 5.5 Å) with the distances read out from the crystal structure (pdb 2CBA, see Fig. S9, ESI†).

Generally, distances above 5.5 Å show a larger deviation, which is on one hand likely due to artifacts like dipolar truncation and non-trivial spin diffusion contributions. On the other hand, even without such effects, the uncertainty would increase due to the exponential decay of transfer efficiency as a function of distance. Interestingly, whereas the theory for a two-spin system suggests an r^{-3} dependence of the cross peak intensities with higher-order terms being comparably small, in praxis additional effects may come to action that are complex to grasp. As such, even though we stuck to the theoretically sound r^{-3} dependence for structure calculation in the following, the optimal correlation would rather be achieved using an r^{-5} weighting (see Fig. S2, ESI†). In practice, however, the accuracies for structure calculation are rather indistinguishable for the different exponents (see Fig. S3, ESI†).

To demonstrate the success of improved structure determination using eRFDR restraints (see Fig. 4A), the determined distances were used as restraints for structure calculation. The resulting (backbone) structural ensemble (Fig. 4D) was compared with an ensemble calculated using conventional RFDR restraints (see Fig. 4C), determined from the spectrum with 2 ms of mixing time (for details see ESI†). The 10 minimal-energy structures were aligned (with regard to the backbone of structured regions) with the crystal structure 2NUZ. Whereas the RMSD with respect to the average backbone structure of the ensemble (precision of the structure) is 2.40 Å for the conventional distance restraints, this RMSD is decreased to 0.56 Å in the presence of eRFDRs. Similarly, the deviation from the crystal structure (accuracy of the NMR structure) is decreased from 3.48 Å to 1.69 Å. Just as for conventional structure calculation, areas with few restraints due to dynamics (the RT-loop and n-Src loop) deviate most strongly.

Structure calculation within the eNORA framework can be performed as an iterative refinement process. Any obtained

structure can be used again as an improved template for constructing relay transfer corrections. In our hands, convergence is reached, however, upon using the conventional RFDR structure for relay transfer correction already, further refinement cycles did not significantly improve the structure. This is shown in Fig. S11 (ESI†). Whether the relay transfer correction leads to improvements will depend on a decent quality of the initial structural model. As an initial structure could be insufficient in different ways, a threshold for its quality is difficult to determine. For perdeuterated proteins, we feel, however, that convergence will be reached as long as the initial structure has at least a qualitatively correct fold.

Although solid-state NMR-specific sources of error, including dipolar truncation and coherent effects cannot be completely circumvented, the eRFDR-approach leads to greatly enhanced restraint precision and, accordingly, more accurate solid-state NMR protein structures. Even though in the solid state, linewidths are unaffected by higher molecular weight, the main limiting factor is the increasing diagonal peak overlap. Diagonal peaks in 3D data bear the resolution of an H/N 2D plane. These problems can potentially be addressed by higher dimensionality (or, more laboriously, selective labeling). Also, eNORA is at this point not capable of dealing with ambiguous cross peak assignments. Finally, the eRFDR approach is more time consuming than single-point RFDR due to the need for multiple data points, of which the ones with short mixing times bear relatively low signal-to-noise ratios. In principle, the approach is equally amenable for non-crystalline samples. However, these are potentially more challenging due to larger linewidths and hence peak overlap (*e.g.*, in fibrils) or due to lower sensitivity (*e.g.*, in membrane protein preparations). An additional solid-state specificity are intermolecular contacts within the crystal lattice. Whereas for partly back-exchanged, crystalline samples usually an insignificant fraction of these contacts exist, both for fully protonated and for proteins with high content of interleaved structural elements (*e.g.*, fibrils) their consideration seems advisable for any relay transfer corrections. Even though eNORA is not equipped for protein lattices yet, single-chain constructs could be created manually after each round of oligomeric structure calculation, which then suit the same purpose. In prospect of ongoing development towards faster MAS, the presented approach might be particularly valuable for fully protonated samples. Due to higher spin concentration, polarization transfer is compromised here by the effects targeted in eRFDR even more than in deuterated and amide-back-exchanged samples. Samples will yield a higher number of RFDR restraints with similar resolution of the diagonal peak (coverage of a larger spectral space). Ambiguities and truncation problems might partially be alleviated by band-selective polarization transfer.[7]

Here we have demonstrated acquisition of solid-state NMR proton–proton distance restraints of high accuracy. Facilitated determination of structures with atomic resolution without (fundamental) limitations regarding protein size will be valuable for structural biology and pharmaceutical sciences. In the long run, multi-state models elucidated by eRFDR instead of single average structures, representing slow motions (as demonstrated for eNOEs in solution), might complement existing dynamics methods in the solid state.

Financial support is acknowledged from the Deutsche Forschungsgemeinschaft (funds from SFB 749, TP A13, SFB 1309, TP 03, the Emmy Noether program, and the Excellence Clusters CiPS-M and RESOLV to R. L.), as well as the National Institutes of Health grants R01-GM123455 and R01-GM112845 to C. M. R.

Conflicts of interest

There are no conflicts to declare.

References

1 R. Linser, *Solid State Nucl. Magn. Reson.*, 2017, **87**, 45–53.
2 C. M. Quinn and T. Polenova, *Q. Rev. Biophys.*, 2017, **50**, e1.
3 S. K. Vasa, H. Singh, K. Grohe and R. Linser, *Angew. Chem., Int. Ed.*, 2019, **58**, 5758–5762.
4 L. B. Andreas, K. Jaudzems, J. Stanek, D. Lalli, A. Bertarello, T. Le Marchand, D. Cala-De Paepe, S. Kotelovica, I. Akopjana, B. Knott, S. Wegner, F. Engelke, A. Lesage, L. Emsley, K. Tars, T. Herrmann and G. Pintacuda, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 9187–9192.
5 M. J. Knight, A. J. Pell, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, T. Herrmann, L. Emsley and G. Pintacuda, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 11095–11100.
6 V. Agarwal, S. Penzel, K. Szekely, R. Cadalbert, E. Testori, A. Oss, J. Past, A. Samoson, M. Ernst, A. Böckmann and B. H. Meier, *Angew. Chem., Int. Ed.*, 2014, **53**, 12253–12256.
7 M. G. Jain, D. Lalli, J. Stanek, C. Gowda, S. Prakash, T. S. Schwarzer, T. Schubeis, K. Castiglione, L. B. Andreas, P. K. Madhu, G. Pintacuda and V. Agarwal, *J. Phys. Chem. Lett.*, 2017, **8**, 2399–2405.
8 A. E. Bennett, J. H. Ok, S. Vega and R. G. Griffin, *J. Chem. Phys.*, 1992, **96**, 8624–8627.
9 R. Verel, M. Ernst and B. H. Meier, *J. Magn. Reson.*, 2001, **150**, 81–90.
10 D. H. Zhou, J. J. Shea, A. J. Nieuwkoop, W. T. Franks, B. J. Wylie, C. Mullen, D. Sandoz and C. M. Rienstra, *Angew. Chem., Int. Ed.*, 2007, **46**, 8380–8383.
11 R. Linser, B. Bardiaux, V. Higman, U. Fink and B. Reif, *J. Am. Chem. Soc.*, 2011, **133**, 5905–5912.
12 M. Huber, A. Böckmann, S. Hiller and B. H. Meier, *Phys. Chem. Chem. Phys.*, 2012, **14**, 5239–5246.
13 R. Linser, B. Bardiaux, S. G. Hyberts, A. H. Kwan, V. K. Morris, M. Sunde and G. Wagner, *J. Am. Chem. Soc.*, 2014, **136**, 11002–11010.
14 T. Gullion and J. Schaefer, *J. Magn. Reson.*, 1989, **81**, 196–200.
15 Y. Ishii, *J. Chem. Phys.*, 2001, **114**, 8473–8483.
16 B. Vögeli, T. F. Segawa, D. Leitz, A. Sobol, A. Choutko, D. Trzesniak, W. van Gunsteren and R. Riek, *J. Am. Chem. Soc.*, 2009, **131**, 17215–17225.
17 J. Orts, B. Vögeli and R. Riek, *J. Chem. Theory Comput.*, 2012, **8**, 3483–3492.
18 D. Strotz, J. Orts, C. N. Chi, R. Riek and B. Vögeli, *J. Chem. Theory Comput.*, 2017, **13**, 4336–4346.
19 K. T. Mueller, *J. Magn. Reson.*, 1995, **113**, 81–93.
20 C. P. Jaroniec, C. Filip and R. G. Griffin, *J. Am. Chem. Soc.*, 2002, **124**, 10728–10742.
21 M. Leskes, Ü. Akbey, H. Oschkinat, B.-J. van Rossum and S. Vega, *J. Magn. Reson.*, 2011, **209**, 207–219.
22 E. Nimerovsky and A. Goldbourt, *J. Magn. Reson.*, 2012, **225**, 130–141.
23 R. Zhang, Y. Nishiyama, P. Sun and A. Ramamoorthy, *J. Magn. Reson.*, 2015, **252**, 55–66.
24 S. K. Vasa, H. Singh, P. Rovó and R. Linser, *J. Phys. Chem. Lett.*, 2018, **9**, 1307–1311.
25 D. Strotz, J. Orts, M. Minges and B. Vögeli, *J. Magn. Reson.*, 2015, **259**, 32–46.
26 P. Rovó, C. A. Smith, D. Gauto, B. L. de Groot, P. Schanda and R. Linser, *J. Am. Chem. Soc.*, 2019, **141**, 858–869.

Supporting information

For the manuscript

Exact distance measurements for structure and dynamics in solid proteins by fast-magic-angle-spinning NMR.

*Kristof Grohe[a,b], Evgeny Nimerovsky[c], Himanshu Singh[a,b], Suresh K. Vasa[a,b], Benedikt Söldner[a], Beat Vögeli[d], Chad M. Rienstra[c] and Rasmus Linser[a,b]**

[a] K. Grohe, Dr. H. Singh, Dr. S. K. Vasa, B. Söldner, Prof. Dr. R. Linser

Faculty for Chemistry and Pharmacy
Ludwig-Maximilians-University Munich
Butenandtstr. 5-13, 81377 Munich, Germany
*E-mail: rasmus.linser@lmu.de

[b] K. Grohe, Dr. H. Singh, Dr. S. K. Vasa, Prof. Dr. R. Linser

Faculty for Chemistry and Chemical Biology,
Technical University Dortmund,
Otto-Hahn-Straße 4a, 44227 Dortmund, Germany

[c] Dr. E. Nimerovsky, Prof. Dr. C. M. Rienstra

Department of Chemistry
University of Illinois
600 South Mathews Avenue
Urbana, IL 61801, United States of America

[d] Prof. Dr. B. R. Vögeli

Department of Biochemistry and Molecular Genetics,
University of Colorado Denver,
12801 East 17th Avenue,
Aurora, CO 80045, United States of America.

Experimental section

The perdeuterated, ^{13}C- and ^{15}N-labeled samples of chicken α-spectrin SH3 domain and human carbonic anhydrase II (hCAII) were expressed and purified as described earlier[1,2], with exchangeable sites back-exchanged along purification in 100 % H_2O. For NMR experiments approximately 1 mg of Cu-EDTA-doped protein[3] was needed each. NMR experiments were carried out at 700 MHz proton Larmor frequency and 55.55 kHz MAS frequency, using a standard-bore 1.3 mm triple-resonance probe on a Bruker AVANCE NEO spectrometer at approximately 20 °C. RFDR spectra were recorded for 0.29, 0.43, 0.58, 0.72, 0.86, 1.01, 1,15, 1.44, 1.73, 2.02, 2.30, 2.88, 3.46, and 4.03 ms RFDR mixing time for SH3 and 0.18, 0.31, 0.61, 0.86, 1.15, 1.44, 2.02, 2.88, 4.03, 4.90, and 5.76 ms mixing time for hCAII. The 180° pulse during RFDR had a pulse length of 3 µs, resulting in a duty factor of 0.16 and an *rf* frequency of 166.6 kHz. 25 ms t_{1max} in the ^{15}N, 20 ms in the indirect ^{1}H, and 50 ms acquisition time in the direct ^{1}H-dimension were used. The experimental time of each 3D spectrum was approximately 20.5 hours. (The first 4 mixing times (up to 1 ms) were sufficient for the analysis (see Fig S4), which leads to a total experimental time of 82 hours or 3.5 days.) Data processing and assignments were carried out using Topspin 4 and CcpNMR[4], respectively with determination of peak intensities as peak heights.

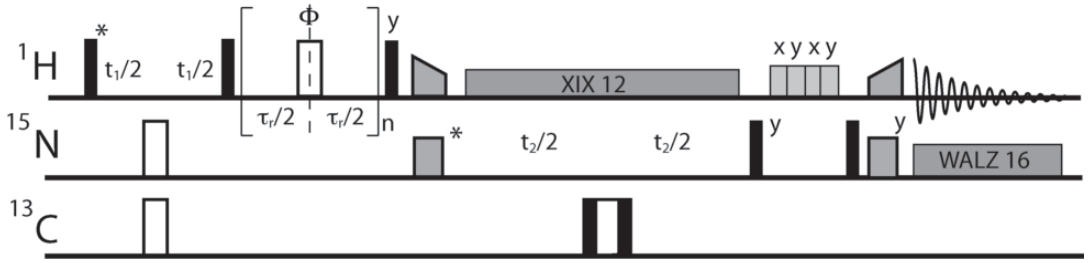


Fig. S1: Pulse scheme for the 3D ^{15}N-edited ^{1}H-^{1}H RFDR experiment.[5,6] Filled and open bars represent 90 and 180 ° pulses, respectively. Water suppression (light gray shapes) was achieved using similar means as proposed by Zhou et al.[7] Asterisks denote phase-sensitive incrementation according to States-TPPI, the phase ϕ represents an XY-8[8] phase cycle during the RFDR mixing. CP (non-annotated dark gray shapes) duration amounted to 300 µs and used an 80-100% ramp on the proton channel. Decoupling during indirect and direct acquisition was performed using XiX-12[9] at 14 kHz and WALTZ-16[10] at 10 kHz, respectively. t_{1max} and t_{2max} amounted to 20 and 25 ms, respectively. The counter *n* was chosen such that overall RFDR mixing times were achieved as denoted in the main text.

Numerical powder-averaged two-spin-model simulations of ^{1}H-RFDR cross-peak buildups were obtained using an in-house MATLAB[11] script, setting inter-proton distances from 2.0 to 8 Å, and normalizing the intensity to the diagonal-peak, which was simulated equally.

Structure calculations were performed using CNS1.1[12] calculating 300 trial structures with 15000 steps in the first and 220000 steps in the second slow-cooling annealing

stage. The eRFDR structure calculation depicted in Figure 4C (main text) was carried out using the available 45 unidirectional and 30 bidirectional eRFDR restraints (as shown on the average structure in Figure 4B, main text), in addition to 47 conventional RFDR restraints as well as dihedral-angle restraints. In terms of comparison, a structure calculation using 122 conventional RFDR restraints in addition to dihedral-angle restraints, as shown in Figure 4A, was performed. In order to create an objective measure for the quality of structural improvement using eRFDR restraints, we grouped all unambiguous cross-peaks from the 2 ms mixing time dataset by their intensity relative to the diagonal peak, creating three tiers: Peaks showing an intense relative cross peak were considered close and converted into a restraint with an upper distance limit of 5 Å. Peaks showing a medium or a small relative cross peak intensity were considered intermediate or far, with an upper distance limit of 7.5 Å and 9 Å, respectively. The computational time for structure calculation amounted to approximately 1.5 days on one core (an Intel i7 processor).

Data processing to generate eRFDR-restraints by help of eNORA2

In order to generate exact distance restraints from RFDR buildup the following error sources have to be addressed: *i*) Site specificity of the polarization transfer efficiency during the heteronuclear correlation part. *ii*) Site-specific relaxation and pulse imperfections. *iii*) Relayed polarization transfer via third spins.

In order to correct for differential polarization transfer efficiency, the cross-peak buildup $H_j \rightarrow H_i$ is divided by the diagonal-peak intensity at zero mixing time $I_{ii}(0)$,

$$I_{ji,norm}(t) = \frac{I_{ji}(t)}{I_{ii}(0)} \quad , \tag{S1}$$

which is extrapolated from a mono-exponential fit.

$$I_{ii}(t) = I_{ii}(0) \cdot e^{-k_{ii}t} \, . \tag{S2}$$

$I_{ii}(t)$ is the diagonal-peak intensity of spin i measured at mixing time t, $I_{ii}(0)$ is the diagonal-peak intensity of spin i at zero mixing time, and k_{ii} represents the decay rate of the diagonal peak. While in liquid-state NOESY k_{ii} is the auto-relaxation rate of spin i, this behavior is dominated by losses induced by the RFDR pulse train in the solid-state counterpart. In the normalized intensity the polarization transferred via neighboring spins in terms of spin diffusion is corrected using an initial structural model, considering all relay pathways in addition to the direct transfer:

$$I_{ji,corr}(t) = p_{ji}(t) \cdot I_{ji,norm}(t) \, , \tag{S3}$$

$I_{norm}(t)$ being the cross-peak intensity normalized with the diagonal-peak intensity at zero mixing time, $I_{corr}(t)$ the corrected cross-peak intensity, and $p(t)$ the correction coefficient for intensities at mixing time t. (For more details see Orts et al.[13])

The correction coefficient *p(t)* can be derived on the basis of the spin network geometries using either the simulated cross-peak intensity originating from polarization relay-transferred through individual third spins (option "three spin" in the masterscript of eNORA2):[13]

$$p(t) = \frac{I_{twospin}(t)}{I_{twospin}(t) + \sum_n (I_{threespin}{}^n(t) - I_{twospin}(t))} \quad \text{(S4.1)}$$

Alternatively, *p(t)* can be derived from the simulated ratio of the transfers in a two-spin system and in a multi-spin system, if the option "full matrix" is chosen:[14]

$$p(t) = \frac{I_{twospin}(t)}{I_{multispin}(t)} \quad \text{(S4.2)}$$

$I_{twospin}(t)$ is the simulated intensity using an ideal two-spin model and $I_{threespin}$ or $I_{multispin}{}^{(n)}(t)$ (using the option "three spin" or "full matrix", respectively) the simulated intensity originating from polarization transfer through the neighboring spins within spheres of a user-chosen radius (12 Å in this study). Whereas the data shown in the manuscript is based on the "three-spin approach", no significant differences were observed upon usage of the "full-matrix approach"

In order to further correct for polarization loss during application of the RFDR pulse train, the diagonal-peak decay k_{ii} values obtained from fits in equation S2 are used as fixed input values to extract the cross-relaxation rates from fits of the buildup intensities $I_{ij}(t)$ and $I_{ji}(t)$.

For description of any magnetization transfers, eNORA uses the solution for the Solomon equation, which for a two-spin system is:

$$\frac{I_{ij}(t)}{I_{ii}(0)} = \frac{I_{ji}(t)}{I_{jj}(0)} = \frac{-\sigma_{ij}}{\lambda_+ - \lambda_-}\left[e^{-\lambda_- t} - e^{-\lambda_+ t}\right] \quad \text{(S5)}$$

with

$$\lambda_\pm = \frac{\left(k_{ii} + k_{jj}\right)}{2} \pm \sqrt{\left(\frac{k_{ii} - k_{jj}}{2}\right)^2 + \sigma_{ij}^2} \quad , \quad \text{(S6)}$$

where σ_{ij} is the magnetization transfer rate between protons H_i and H_j, and k_{ii} was the diagonal decay rate of the proton H_i.

For simulation of relay transfers, eNORA2 makes use of the transfer matrix describing the time evolution of all possible cross peak intensities.

$$I(t) = I(0) \cdot e^{-Rt} \quad \text{(S7)}$$

With

$$R = \begin{pmatrix} k_{11} & \sigma_{12} & \dots \sigma_{1N} \\ \sigma_{21} & k_{22} & \dots \sigma_{2N} \\ \sigma_{N1} & \sigma_{N2} & \dots k_{NN} \end{pmatrix} \quad \text{(S8)}$$

"Bidirectional" distance restraints were obtained using the average of the build-up rate σ of cross-peaks $H_j \rightarrow H_i$ and $H_i \rightarrow H_j$:

$$\sigma_{bidir} = sqrt(\sigma_{ij}\sigma_{ji}) \quad \text{(S9)}$$

The first-order dipolar recoupling Hamiltonian during RFDR being dependent on r^{-3} terms, the buildups were associated with distances using the equation

$$y = a \cdot r^{-3}. \quad \text{(S10)}$$

The computational time for the relay transfer correction is approximately 2 minutes using one core of an Intel i7 processor.

Comparison of distance correlation functions

In solution state NMR, the mechanism for through-space polarization transfer is precisely known as cross relaxation according to zero- and double-quantum transitions and described by the Solomon equation.[15]

$$\sigma_{ij} = \left(\frac{\mu_0}{4\pi}\right)^2 \cdot \frac{\gamma^4 h^2}{10} \tau_c \cdot r^{-6} \quad \text{(S11)}$$

By contrast, in the case of RFDR, the practical consequences of polarization transfer with respect to the relation between effective transfer rate and internuclear distance is not fully understood. The effective Hamiltonian of (finite-pulse) RFDR scales inversely with the MAS rotor frequency and is dependent on various other parameters. Artifacts resulting from varying chemical shift differences between recoupled nuclei are removed by super-cycles (in this case XY-8), which eliminate higher-order (cross) terms in the average Hamiltonian. From a theoretical analysis using a quantized-field approach/average Hamiltonian theory, or Floquet theory, transfer probability of RFDR in a two-spin system is expected to scale with r^{-3}, as a coherent-driven type of transfer.[16] However, for multi-spin interactions in praxis a less clear dependency, effectively somewhere between r^{-3} and r^{-6}, has been described.[17] As the correlation known from proton-driven spin diffusion[18] has been experimentally found to be proportional to r^{-6}, we were interested to experimentally determine the practically most relevant expression, assuming any potency between r^{-3} and r^{-6} (Figure S2 and S3)

$$\sigma_{ij} = a \cdot r^{-x} + b \quad \text{(S12)}$$

Figure S3 shows uni- and bidirectional eRFDR restraints analyzed using an adapted version of the freely available program eNORA2[19]. In order to determine the most useful correlation between buildup rate and distance we modified the program by changing the functions for distance determination and relay transfer correction according to equation S12 with x = 3, 4, 5, and 6. Interestingly, the R^2 values of the four different correlations of measured versus X-ray distances of SH3 show very similar values of around 0.9. Regarding this consistency, for further analysis we still used the correlation function containing r^{-3}, since this represents the first-order average Hamiltonian for dipolar recoupling.

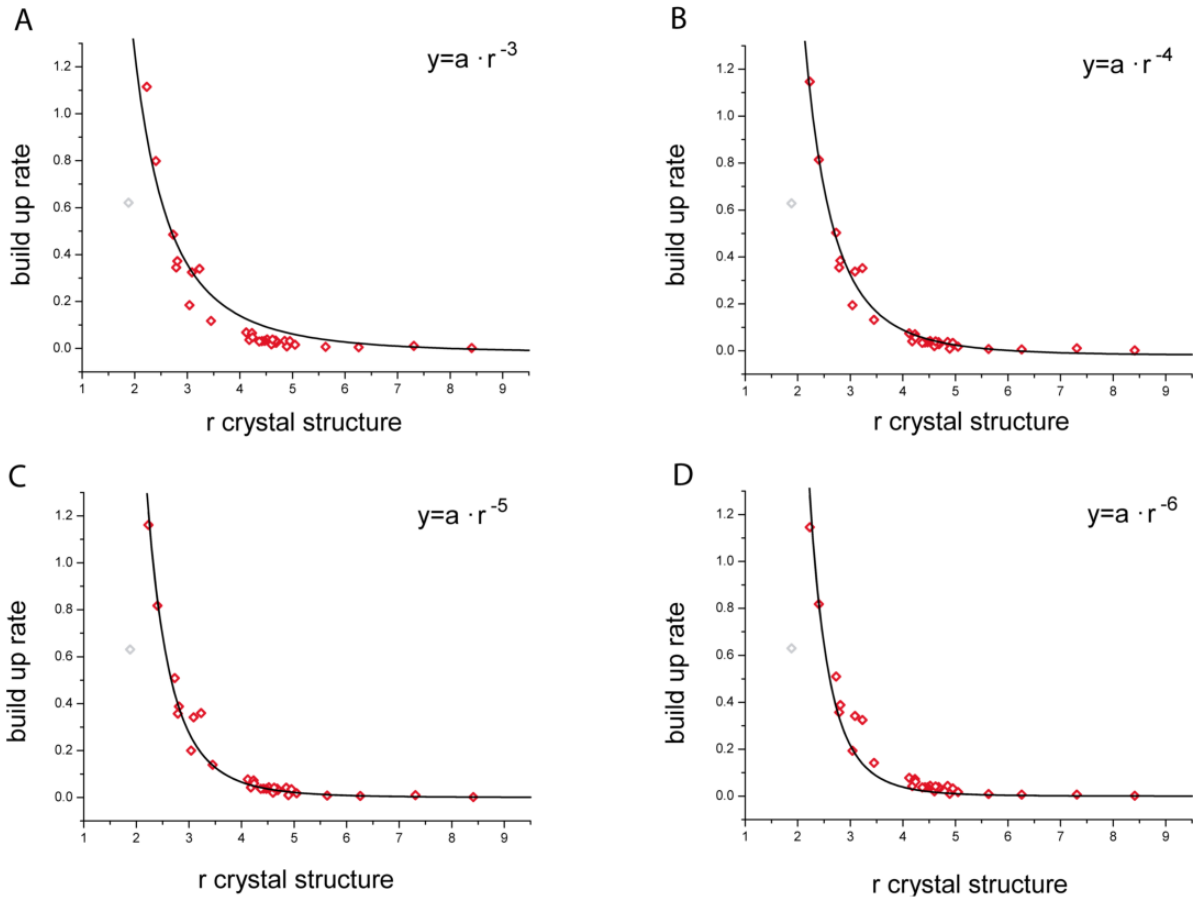


Figure S2. Buildup rates obtained by eNORA2 as a function of the distance read out from the crystal structure 2NUZ. **A-D:** The plotted buildup curves were corrected for spin diffusion using the full-transfer-matrix approach considering an r^{-3}, r^{-4}, r^{-5}, or r^{-6} buildup-distance correlation and fitted with the equation shown in the respective figure. The data point with shortest amide-amide distance (shown in gray) stands out from the expected behavior. For this pair with unusually short distance (residues 18 and 19), the protons are at the tip of a dynamic loop.[20] As such, at room temperature, both their effective distance is expected to be increased and their dipolar interaction is reduced. Compare Figures S3 and S10A. (This pair is omitted from the fit, also see below.)

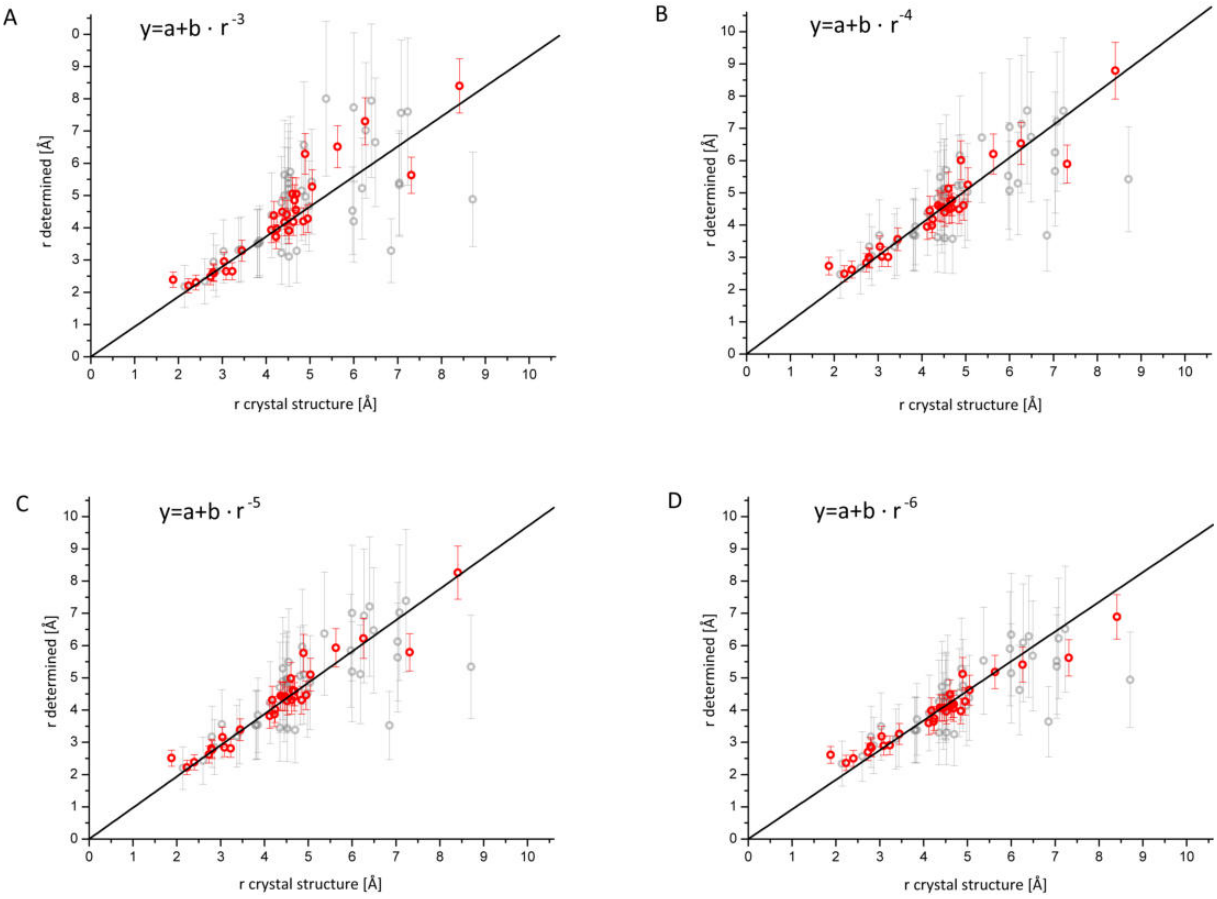


Figure S3. Correlations between distances measured by eRFDR and distances read out from the crystal structure (2NUZ). In **A**, **B**, **C**, and **D** the eRFDR distances were determined with the equation shown in the respective figure. In all cases the measured distances correlate well with the ones read out from the crystal structure. Distances which are greater than 5.5 Å in the crystal structure show a larger deviation probably caused by dipolar truncation and increasing spin diffusion contributions. The comparison shows that the exponent of the distance dependency of the RFDR buildup is less relevant than expected. Other errors have a higher impact on the accuracy. For the data point with shortest amide-amide distance, which stands out from the expected behavior, this is most probably due to loop dynamics (see above, also compare Figures S2 and S10).

Validation of uniform fitting up to 1 ms of mixing time

In order to practically apply the initial-regime approach to the experimental data, the transfers were grouped into “fast buildup”, in which intensities were fitted till 1 ms mixing time, and “slow-buildup”, which were fitted till 2 ms mixing time, based on the time point when the maximum intensity is reached. For purely automated data processing we also verified if buildups could be fitted absolutely uniformly (without manual interference), therefore we fitted all cross peaks only till 1 ms mixing time. In fact, this only shows minor differences regarding the determined distances (see Figures S4 and S8) and at the same time demonstrates that RFDR buildups consisting of 4 to 5

spectra (here with a total experimental time of around 80 hours) are fully sufficient to determine eRFDR restraints.

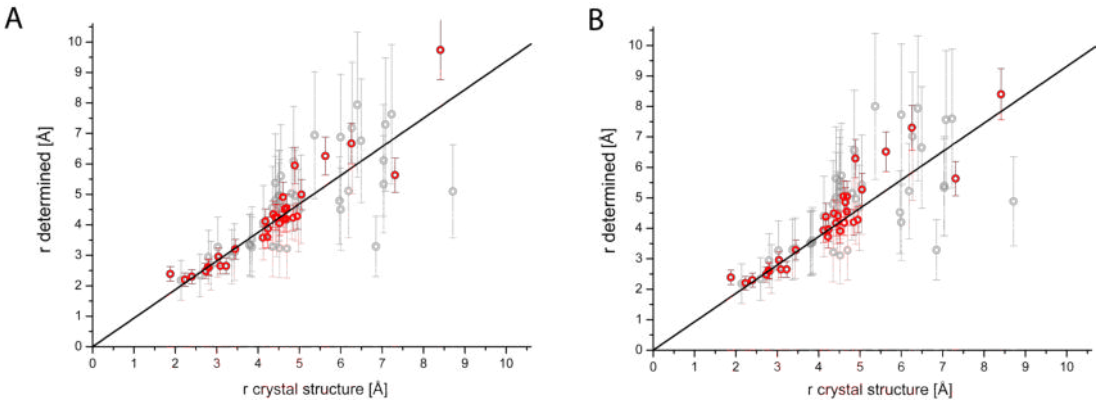


Figure S4. Validation of uniformly fitting up to 1 ms of mixing time. **A)** Bidirectional (red) and unidirectional (gray) eRFDR restraints, determined by fitting the cross peaks of fast buildup with a fit up to 1ms and those with slower buildup with a fit up to 2 ms of mixing time. **B)** eRFDR restraints determined by fitting all peaks uniformly up to 1 ms of mixing time.

Verification of the applied analytical routines using powder-averaged numerical simulations

In the course of verifying the used data processing approach, we run powder-averaged 2- and 3-spin simulations and used the simulated buildup and diagonal-decay curves as an input for the modified eNORA2 program. We used an in-house written MATLAB script including the following propagators:

$$U_1 = exp\left\{-i \int_0^{0.5T_R-0.5t_p} dt\,[1.5\omega_D(t)I_{z1}I_{z2}]\right\} \quad \text{(S13a)}$$

$$U_2 = \hat{T}exp\left\{-i \int_{0.5T_R-0.5t_p}^{0.5T_R+0.5t_p} dt\,\left[1.5\omega_D(t)I_{z1}I_{z2} + \omega_{rf}(I_{x1}+I_{x2})\right]\right\} \quad \text{(S13b)}$$

$$U_3 = exp\left\{-i \int_{0.5T_R+0.5t_p}^{T_R} dt\,[1.5\omega_D(t)I_{z1}I_{z2}]\right\} \quad \text{(S13c)}$$

$$U_4 = \hat{T}exp\left\{-i \int_{0.5T_R-0.5t_p}^{0.5T_R+0.5t_p} dt\,\left[1.5\omega_D(t)I_{z1}I_{z2} + \omega_{rf}\left(I_{y1}+I_{y2}\right)\right]\right\} \quad \text{(S13d)}$$

$$U_{RFDR} = U_3 \cdot U_4 \cdot U_1 \cdot U_3 \cdot U_2 \cdot U_1 \quad \text{(S13e)}$$

where $\omega_D(t)$ is the dipolar function[21,22], dependent on time, spinning speed, and Euler angles[7], T_R is the rotor period, t_p is the width of the π – pulses, and $\hat{T}$ is the time-ordered Dyson operator.

U_2 and U_4 are corresponding to the x- and y-π pulses of an XY-8 phase cycle, respectively. A Dyson-dependent propagator can be departed in a series of N Dyson-independent propagators[23] as follows:

$$U = \hat{T}exp\{-i\int_0^\tau dtH(t)\} \approx e^{-i\int_{(N-1)\tau/N}^{\tau} dtH(t)} \dots e^{-i\int_{\tau/N}^{2\tau/N} dtH(t)} e^{-i\int_0^{\tau/N} dtH(t)} \quad \text{(S14)}$$

Figure S5 shows exemplary comparisons of the simulated buildup curve with the experimental data from the SH3 domain. The shape of the simulated curves is in good agreement with the experimental values, however, with this direct fitting (without any correction) the fitted distance is significantly off in each case.

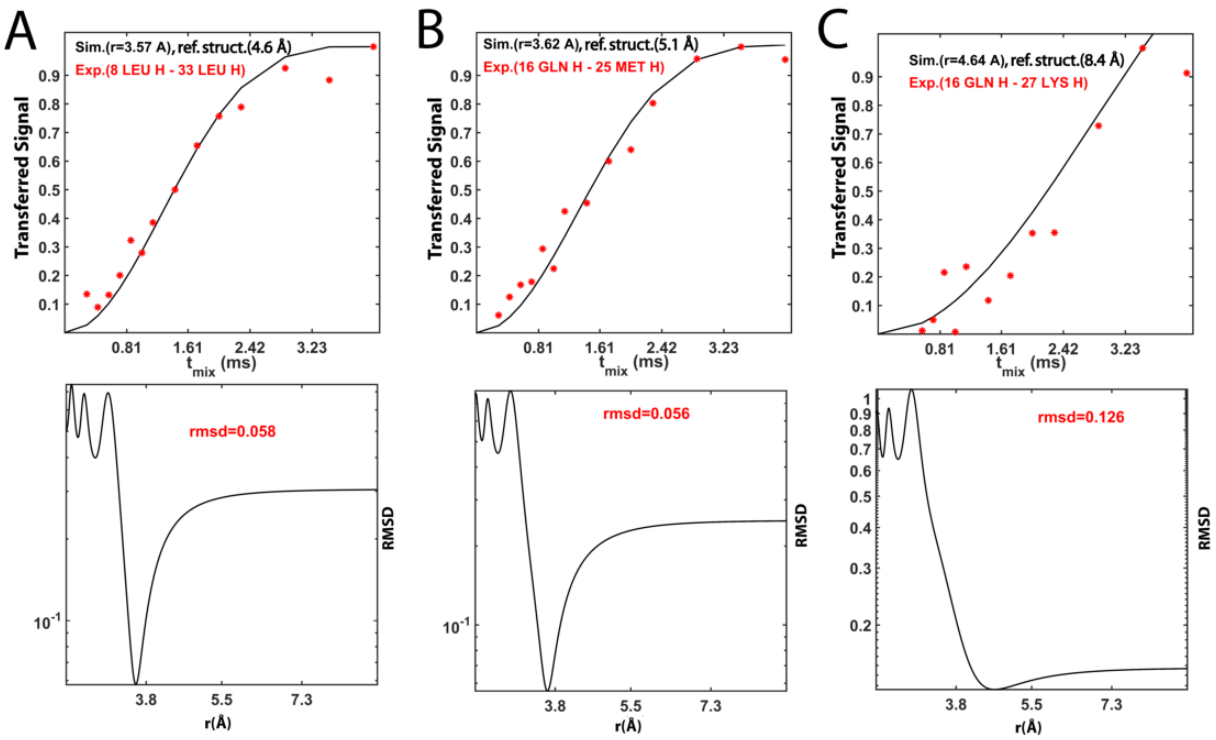


Figure S5. Comparison of experimental data (red stars) with the exact numerical 2-spin model solution (black line), directly fitting the data without correction. RFDR buildups for the exemplary distances: **A)** Leucines 8 and 33, with a distance for the best-correlating simulation of 3.6 Å and a distance read out from the crystal structure of 4.6 Å. **B)** Gln 16 and Met 25, with a distance for the best-correlating simulation of 3.6 Å and a distance read out from the crystal structure of 5.1 Å. **C)** Gln 16 and Lys 27, with a distance for the best-correlating simulation of 4.6 Å and a distance read out from the crystal structure of 8.4 Å. Bottom panels depict RMSDs between experimental data and simulated curve as a function of simulation distance.

The best-correlating simulation distances are significantly shorter than the distances read out from the crystal structure 2NUZ. This deviation, which cannot be accounted for

by simple scaling using a constant factor, underlines that an uncorrected 2-spin simulation is not suitable for exact distance determination. In order to confirm the influence of spin diffusion via third spins, we performed 3-spin simulations. Figure S6 shows the simulation of Leu 8 – Leu 33, the same pair of nuclei also used in Figure S5A. In fact, the deviation of the simulated distance from the crystal structure distance is reduced, demonstrating the significance of spin diffusion via third spins. Simulations involving all spins playing a role in the polarization transfers, also involving refinement cycles that are potentially necessary, are not suitable due to computational effort. This makes the simplified analytical solution presented in this work more favorable.

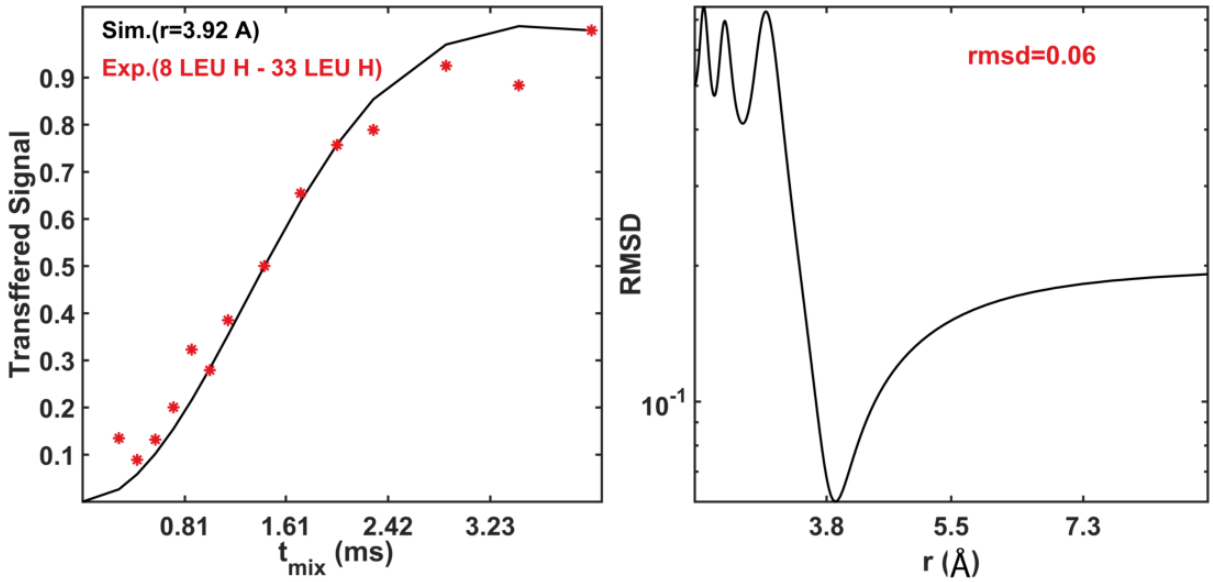


Figure S6. Powder averaged 3-spin simulation (black line) and experimental intensities (red stars) of the buildup Leu 8 – Leu 33. The distance of the third spin (to either of the two others) was set to 4.6 Å. The minimum RMSD regarding simulated curve and experimental data changed from 3.6 Å in case of the 2-spin simulation to 3.9 Å, which is closer to the crystal structure distance of 4.6 Å.

For verification of the presented analytical approach with a simplified buildup function, we conducted 2-spin model simulations as described above with inter-nuclear distances from 2 to 8 Å, spinning speed of 55.555 kHz, and a Larmor frequency of 700 MHz. The resulting buildup curves (partly shown in Figure 2B main text) were fitted up to the first maximum using a linear (Figure S7A), a trigonometric (Figure S7B), and an exponential function (Figure S7C). The resulting rates were fitted using equation S15. All fits correlate nearly perfectly with an R^2 value of above 0.99, meaning that all 3 fitting procedures would be practically suitable to determine polarization buildup rates in the initial regime. Such fitting of the initial regime only has been used in other contexts previously.[24]

$$\sigma_{ij} = a \cdot r^{-3} \qquad \text{(S15)}$$

To examine further the applicability of the combined exponential fitting routines used in eNORA2, simulated cross-peak buildup and diagonal-decay intensity values were used as an input for the eNORA2 program, modified as described above. The distances determined using eNORA2 and the distances set for creating the simulated input show a good correlation (Figure S7D). According to the results of the simulations, fitting the experimental intensities exponentially represents a sufficient approximation.

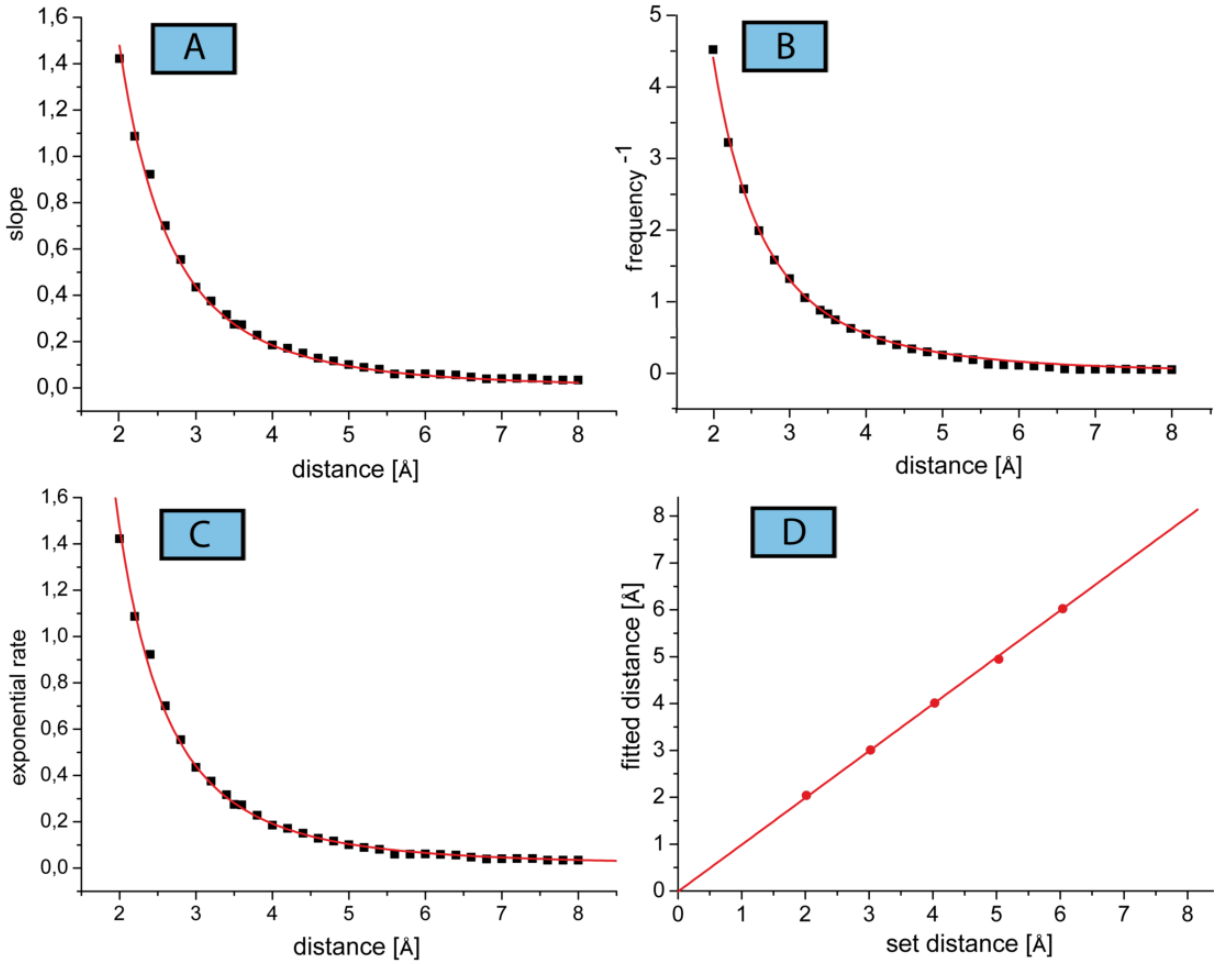


Figure S7. Comparison of different fitting routines to determine the polarization buildup rates. **A)** Linearly fitted buildup rates. **B)** Buildup rates as the inverse frequencies of a trigonometric function. **C)** Exponentially fitted buildup rates. In **A-C** simulations (compare main text Figure 2B) according to inter-nuclear distances from 2 Å up to 8 Å were fitted till the first maximum and correlated using equation S15 with a resulting R^2 value above 0.99 in all cases. **D)** Verification of the modified (exponential) fitting procedure of eNORA2 using simulated buildup and decay curves. Distances from 2 to 6 Å were used as the "experimental" input for the program eNORA2 to verify the fitting procedure. Determined distances and distances set for the simulation correlate linearly.

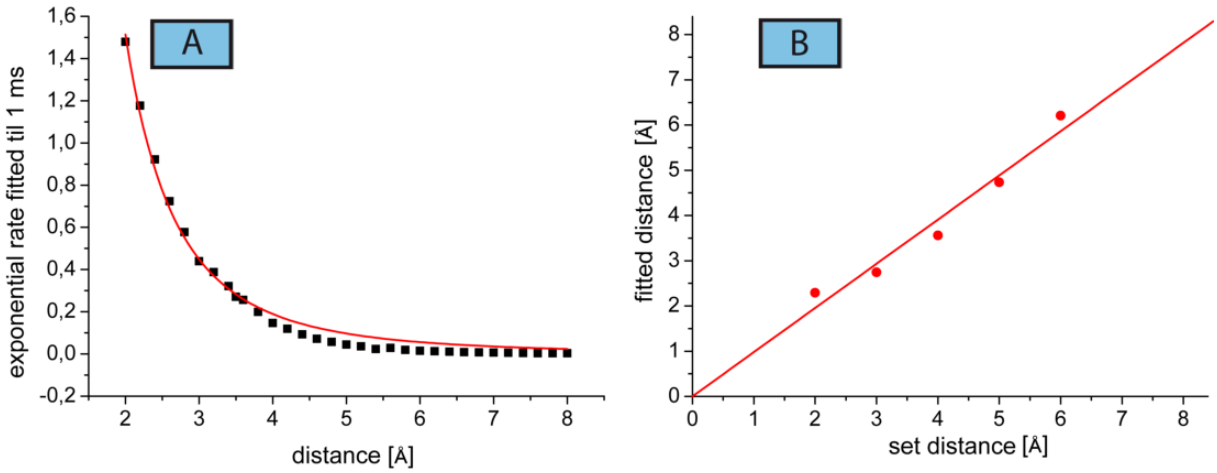


Figure S8. Fitting of the simulated buildups uniform up to 1 ms of mixing time only. **A)** Exponential buildup rates according to Figure S7C, fitted uniformly up to 1 ms of mixing time. **B**) Verification of the fitting procedure in eNORA2. Simulated buildup and decay curves up to 1 ms of mixing time, for distances from 2 to 6 Å, were used as "experimental" input for the program eNORA2, according to Figure S7D.

Modifications of eNORA for solid-state NMR RFDR

With the data processing strategies for solid-state NMR demonstrated to be applicable in the described way, it is possible to make use of the eNORA2[19] program, which was originally written for liquid-state eNOE, if modified for solid-state NMR as described in the following. The program allows to generate eRFDR restraints in a fast and convenient manner. Differences between the exact-NOE approach and the exact-RFDR approach concern the buildup behavior and the correlation of polarization transfer with internuclear distance. For the buildup behavior no simple analytical function (approximated rather via more complicated Bessel-functions sometimes) is known. However, we showed that this obstacle can be circumvented by assessing the relative magnetization transfer rate using a mono-exponential buildup function followed by a calibration. This holds perfectly true only as long as all buildups are fitted up to a comparable (relative) time point, like the first maximum of the buildup. If all buildups are fitted equally up to a fixed time, for example 1 ms, discrepancies between the fitting function and the nature of the buildup do invoke errors. However, as shown in Fig. S8, for buildups corresponding to internuclear distances in the range of 2-6 Å, in practice these errors are reasonably low. Regarding the scaling between buildup and distance, the dipolar recoupling in RFDR follows a cubic correlation when exclusively taking the first-order Hamiltonian into account. By including higher-order terms, the correlation gets less clean. Additionally, using finite-pulse RFDR, also mechanisms contribute to the polarization transfer that are different from dipolar recoupling. Consequently, our experimental data seems to obey a correlation between r^{-4} and r^{-5}. However, these differences have only a small impact on the accuracy of the determined distances, such

that a cubic correlation as predicted by the first-order Hamiltonian was chosen as a good approximation.

The code of eNORA2 was modified regarding the equations for the spin diffusion correction and conversion of buildup rates into distances, which can be found in the following eNORA2 MATLAB scripts:

eNORA2/eNOEprogram/shared/R.m,

eNORA2/eNOEprogram/shared/S.m,

eNORA2/eNOEprogram/shared/sigma2distance.m.

Since the initial parameters for converging the fit are set for NOE-buildup, which covers hundreds of ms, the values of the applied RFDR mixing times should be multiplied by 1000. This factor is equated by the calibration using the input parameter correlation time τ_c in the master script.

Note:

As the distance dependency scaling exponent does not largely influence the accuracy of the distances obtained (as shown in Figures S2 and S3), it is also possible to successfully use a completely unmodified version of eNORA2 (assuming an r^{-6} dependency), which is kindly provided at the homepage of Prof. Dr. Beat Vögeli.

eRFDR distance restraints recorded for the 29kDa human carbonic anhydrase II

The triply labeled, 100% back-exchanged sample of human carbonic anhydrase II (hCAII) was prepared as described earlier[25]. NMR experiments were carried out as described in the Experimental Section. From the RFDR-spectra 176 through-space connections were assigned unambiguously. 43 of these through-space connections contain one unambiguous pair of cross- and diagonal peak, leading to unidirectional eRFDR restraints (in Figure S9A**:** orange connections, in B**:** orange cycles), and 33 contain 2 pairs of unambiguous cross- and diagonal peaks, yielding bidirectional eRFDR restraints (in Figure S9A: red connections, in B: red cycles). The determined distances and the distance read out from the crystal structure (2CBA) are in very good agreement, with an RMSD of 0.29 Å regarding bidirectional restraints shorter than 5.5 Å.

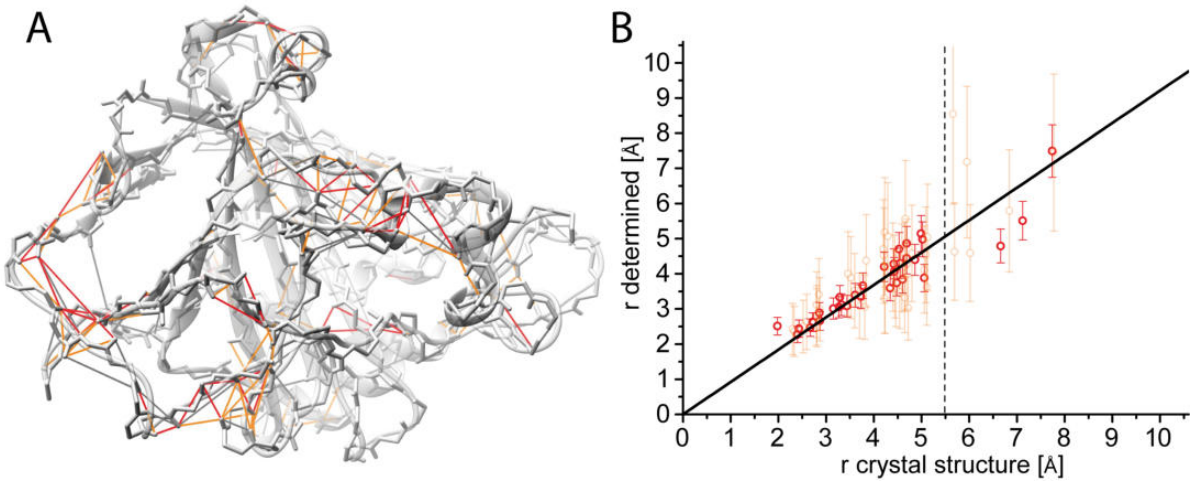


Figure S9. **A)** Bidirectional (red) and unidirectional (orange) eRFDR as well as conventional (gray) RFDR restraints depicted on the X-ray structure of carbonic anhydrase II. **B)** Correlation between bidirectional eRFDR restraints (shown in red) or unidirectional eRFDR restraints (shown in light orange) and the corresponding distances read out from the X-ray structure. The buildup curves were fitted uniformly up to 1 ms of mixing time. Distances larger than 5.5 Å show larger deviation due to error caused by dipolar truncation and strong spin diffusion. The shortest distance corresponding to HN37->HN38 shows a larger deviation presumably caused by dynamics (compare Figure S10B).

Comment to amide proton distances of residue 18 to 19 (SH3) and 37 to 38 (hCAII) according to the X-ray structures 2NUZ and 2CBA

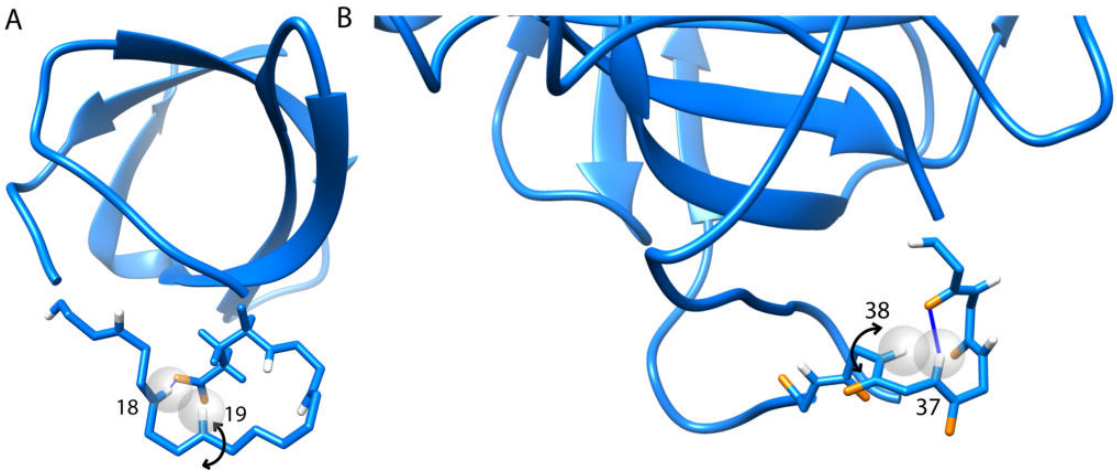


Figure S10. Depiction of the amide protons of **A)** residue 18 and 19 in the SH3 X-ray structure 2NUZ and **B)** 37 and 38 in the hCAII X-ray structure 2CBA (The distances are as small as the sum of the two van-der-Waals radii, represented by spheres). Apart from uncertainties in the reconstruction of proton locations from (heteronuclear-only) electron densities generally, both the distal loop in **A** as well as the respective loop in **B** are also known to undergo motions on different timescales[2,20,26,27]. Both effects will be

sensed within the RFDR experiment, rendering the effective distance larger than what is suggested by the reference distance determined from 2NUZ and 2CBA crystal structures.

Iterative structural improvement using the eRFDR approach

Figure S11 shows that in case of deuterated and 100% backexchanged SH3, a standard solid-state NMR structure is already sufficient to successfully perform the corrections for relayed magnetization transfer via third spins. An improved structure using the eRFDR approach can be obtained in the second step. The structural precision and accuracy are equal to those of the eRFDR structure obtained when using the crystal structure in terms of correction for relayed transfer via third spins.

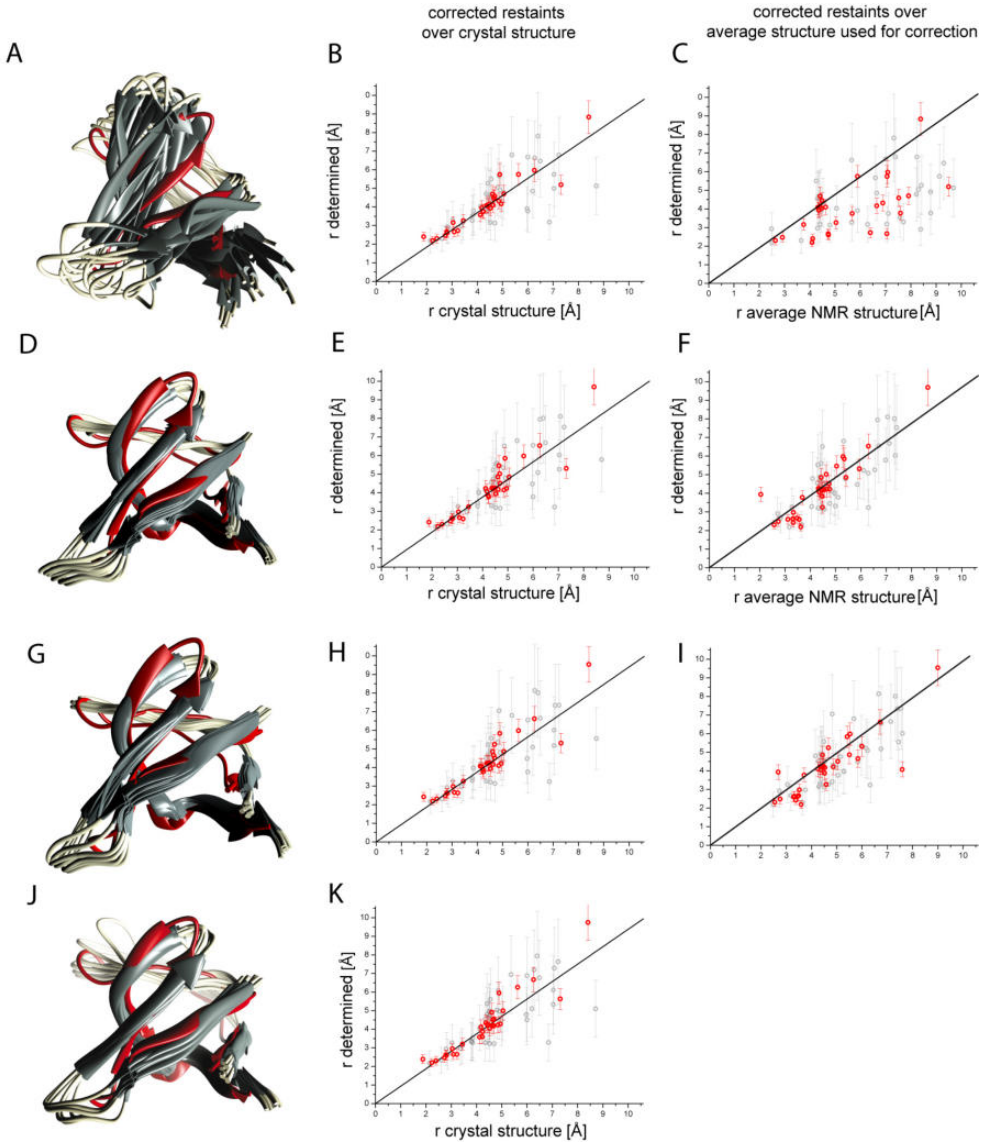


Figure S11: Structure determination using the eRFDR approach starting from standard RFDR restraints. **A)** Structural ensemble of the 10 lowest energy structures determined by conventional RFDR upper limit restraints depicted in gray, aligned with the crystal structure 2NUZ shown in red. **B)** Distance restraints obtained by using the eRFDR approach with the average structure of A employed for spin diffusion correction (in this terms spin diffusion means relayed magnetization transfer via third spins) plotted over

the corresponding internuclear distances of the crystal structure 2NUZ. **C)** Distance restraints obtained by using the eRFDR approach plotted over the internuclear distances of the average structure of the ensemble in A. **D)** Structural ensemble of the 10 lowest energy structures determined by the eRFDR-restraints obtained by using the average structure of A for spin diffusion correction. **E)** Distance restraints obtained by using the average structure of D for relay transfer correction plotted over the corresponding internuclear distances of the crystal structure 2NUZ. **F)** Distance restraints obtained by using the eRFDR approach plotted over the internuclear distances of the average structure of the ensemble in D. **G-I)** Additional refinement iteration just as described for E-F. **J)** Structural ensemble of the 10 lowest-energy structures obtained by using the eRFDR approach employing the crystal structure 2NUZ for relay transfer correction. **K)** Distance restraints obtained by using the eRFDR approach with the crystal structure employed for relay transfer correction.

Supplementary literature:

1. R. Linser, V. Chevelkov, A. Diehl and B. Reif, *J. Magn. Reson.*, 2007, **189**, 209-216.
2. S. K. Vasa, H. Singh, P. Rovó and R. Linser, *J. Phys. Chem. Lett.*, 2018, **9**, 1307−1311.
3. R. Linser, V. Chevelkov, A. Diehl and B. Reif, *J. Magn. Reson.*, 2007, **189**, 209–216.
4. W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins: Structure, Function, and Bioinformatics*, 2005, **59**, 687-696.
5. D. H. Zhou, J. J. Shea, A. J. Nieuwkoop, W. T. Franks, B. J. Wylie, C. Mullen, D. Sandoz and C. M. Rienstra, *Angew. Chem., Int. Ed.*, 2007, **46**, 8380-8383.
6. R. Linser, B. Bardiaux, V. Higman, U. Fink and B. Reif, *J. Am. Chem. Soc.*, 2011, **133**, 5905-5912.
7. Z. Zhou, R. Kümmerle, X. Qiu, D. Redwine, R. Cong, A. Taha, D. Baugh and B. Winniford, *J. Magn. Reson.*, 2007, **187**, 225-233.
8. T. Gullion, D. B. Baker and M. S. Conradi, *J. Magn. Reson.* , 1990, **89**, 479.
9. M. Ernst, A. Samoson and B. H. Meier, *J. Magn. Reson.*, 2003, **163**, 332-339.
10. A. J. Shaka, J. Keeler, T. Frenkiel and R. Freeman, *J. Magn. Reson.*, 1983, **52**, 335-338.
11. Matlab, *The MathWorks, Inc.*, R2011b.
12. P. D. A. A.T. Brünger, G.M. Clore, P.Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, N. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, *Acta Cryst.D54*, 1998, 905-921.
13. J. Orts, B. Vögeli and R. Riek, *J. Chem. Theory Comput.*, 2012, **8**, 3483-3492.
14. B. Vögeli, *J. Magn. Reson.*, 2013, **226**, 52-63.
15. I. Solomon, *Phys. Rev.*, 1955, **99**, 559-565.
16. G. J. Boender, S. Vega and H. J. M. de Groot, *J. Chem. Phys.*, 2000, **112**, 1096-1106.
17. K. H. Mroue, Y. Nishiyama, M. K. Pandey, B. Gong, E. McNerny, D. H. Kohn, M. D. Morris and A. Ramamoorthy, *Sci. Rep.*, 2015, **5**, 1-10.
18. T. Manolikas, T. Herrmann and B. H. Meier, *J. Am. Chem. Soc.*, 2008, **130**, 3959-3966.
19. D. Strotz, J. Orts, C. N. Chi, R. Riek and B. Vögeli, *J. Chem. Theory Comput.*, 2017, **13**, 4336-4346.
20. P. Rovó, C. A. Smith, D. Gauto, B. L. de Groot, P. Schanda and R. Linser, *J. Am. Chem. Soc.*, 2019, **141**, 858–869.
21. T. Gullion, *Chem. Phys. Lett.*, 1995, **246**, 325-330.
22. E. T. Olejniczak, S. Vega and R. G. Griffin, *The J. Chem. Phys.*, 1984, **81**, 4804-4817.
23. E. Nimerovsky and A. Goldbourt, *J. Magn. Reson.*, 2012, **225**, 130-141.
24. R. Zhang, Y. Nishiyama, P. Sun and A. Ramamoorthy, *J. Magn. Reson.*, 2015, **252**, 55-66.
25. S. K. Vasa, H. Singh, P. Rovó and R. Linser, *J. Phys. Chem. Lett.*, 2018, **9**, 1307-1311.
26. P. Rovó and R. Linser, *ChemPhysChem*, 2018, **19**, 34-39.
27. S. K. Vasa, H. Singh, K. Grohe and R. Linser, *Angew. Chem., Int. Ed.*, 2019, **58**, 5758-5762.

2.1.3 Non-equilibrium hydrogen exchange for determination of H-bond strength and water accessibility in solid proteins

The time-resolved analysis of peak intensity after changing the deuteration-degree of the solvent water provides information on water accessibility and H-bond strength (see Chapter 1.2.4). However, non-equilibrium hydrogen exchange was measured by solution state NMR-before, the information mentioned above are provided in an ambiguous manner. Additionally, es explained in chapter 2.1.1 and 2.1.2, advancing solid state NMR-spectroscopy is vital for structural biology.

Both, information about solvent accessibility and H-bond strength provide important information about protein structure. Water accessibility reports about the hydrophobicity of the surroundings, e.g., if the exchangeable proton is located in the hydrophobic core or on the hydrophilic surface of the protein. The method is widely applicable, especially for challenging targets. Hydrogen bonds are the most important interactions that determine the protein structure. Usually, the presence and strength of hydrogen bonds are only read out from atomic distances in a structural model. However, the relationship between bond length and strength is not always proportional, making dedicated experimental data indispensable. Detailed measurements of H-bond strength were performed on small molecules [46] and studies on the hydrogen exchange of proteins were previously performed by mass spectrometry. However, mass spectrometry is not able to provide site-specific data about hydrogen bonds.

In the following work, I was able to perform an analysis that separates information about H-bond strength and water accessibility. Only the separate information about hydrophobic shielding and H-bonds provide optimal data that can be used for structure calculation, dynamics determination, and assignment each. Strikingly, the data is in good agreement with dedicated predictions from molecular dynamics measurements (see Conclusions). The experimental data in combination with MD simulations could shed light on the detailed chemical mechanism of hydrogen exchange in proteins, which is still enigmatic. Furthermore, to the best of my knowledge, this work represents the first kinetic study performed in solid-state NMR spectroscopy.

Contributions by co-authors:

Kumar Tekwani Movellan, Suresh Kumar Vasa, and Karin Giller were involved in the sample production. Dr. Stefan Becker provided the organizational background for the sample production.

J Biomol NMR (2017) 68:7–17
DOI 10.1007/s10858-017-0110-0

ARTICLE

Non-equilibrium hydrogen exchange for determination of H-bond strength and water accessibility in solid proteins

Kristof Grohe[1,2] · Kumar Tekwani Movellan[1,2] · Suresh Kumar Vasa[1,2] · Karin Giller[2] · Stefan Becker[2] · Rasmus Linser[1,2]

Received: 6 December 2016 / Accepted: 2 April 2017 / Published online: 9 April 2017

Abstract We demonstrate measurement of non-equilibrium backbone amide hydrogen–deuterium exchange rates (HDX) for solid proteins. The target of this study are the slowly exchanging residues in solid samples, which are associated with stable secondary-structural elements of proteins. These hydrogen exchange processes escape methods measuring equilibrium exchange rates of faster processes. The method was applied to a micro-crystalline preparation of the SH3 domain of chicken α-spectrin. Therefore, from a 100% back-exchanged micro-crystalline protein preparation, the supernatant buffer was exchanged by a partially deuterated buffer to reach a final protonation level of approximately 20% before packing the sample in a 1.3 mm rotor. Tracking of the HN peak intensities for 2 weeks reports on site-specific hydrogen bond strength and also likely reflects water accessibility in a qualitative manner. H/D exchange can be directly determined for hydrogen-bonded amides using ^{1}H detection under fast magic angle spinning. This approach complements existing methods and provides the means to elucidate interesting site-specific characteristics for protein functionality in the solid state.

Keywords Solid-state NMR · H-bond determination · Water accessibility · Protein structure · Proton detection · MAS

Introduction

Quantification of amide hydrogen deuterium exchange (HDX) rates has been proven to be an important method to obtain direct information on protein structure, stability, dynamics and protein–protein interactions (Kato et al. 2009; Raschke and Marqusee 1998; Wagner and Wüthrich 1979). Important examples are the analysis of the secondary structure of β-amyloid fibers related to Alzheimer's disease (Whittemore et al. 2005) or the elucidation of the folding mechanism of the protein ubiquitin (Yi and Baker 1996). Dependent on their participation in H-bonds, hydrogen exchange rates of backbone amides can be modulated by protein structure and dynamics. By contrast, side chain protons directly involved in active-site function, like hydroxyls involved in enzymatic catalysis, or residues important for intermolecular interactions on protein surfaces, are often exposed to enable H-bond interactions and proton transfer. (Gaspari et al. 2016; Wydrzynski et al. 1996; Zhang et al. 2006). As such, their exchange rates are modulated by the accessibility of the site and the local pK_A (Cook and Cleland 2007). In this manuscript we show that H/D exchange in the solid state leads to site-specific determination of H-bond strength and likely reflects solvent accessibility. These parameters can help to improve structural understanding of proteins which are difficult to elucidate by standard techniques, like membrane proteins (Raschke and Marqusee 1998; Weingarth et al. 2014b), intrinsically disordered proteins, and supramolecular structures (Andronesi et al. 2008; He et al. 2016).

Electronic supplementary material The online version of this article (doi:10.1007/s10858-017-0110-0) contains supplementary material, which is available to authorized users.

✉ Rasmus Linser
rasmus.linser@lmu.de

[1] Department Chemie und Pharmazie, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

[2] Department for NMR-Based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

Water accessibility in solid samples has been characterized employing water-protein magnetization transfer (Ader et al. 2009; Andronesi et al. 2008; Chevelkov et al. 2005; Lesage and Böckmann 2003; Lesage et al. 2008; Linser et al. 2009; Luo and Hong 2010; Weingarth et al. 2014a; Williams and Hong 2014) or by redissolution and solution NMR readout (Cho et al. 2011; Daebel et al. 2012; Dasari et al. 2011; Gallagher et al. 1992; Lührs et al. 2005; Olofsson et al. 2006). Water magnetization is mostly transferred via chemical exchange and subsequent spin diffusion processes (slow magic angle spinning, MAS) or NOE/ROE mechanisms (fast MAS) (Williams and Hong 2014). Therefore, protein-intrinsic magnetization is eliminated concomitantly by recoupling methods. Alternatively, the source of magnetization (for example from water) upon transfer of polarization to protein sites can be identified via an indirect evolution. In addition, environmental paramagnetic effects on T_1 or T_2 relaxation at protein residues can be determined in a site-specific way. Examples for such experiments include Cu-edta (T_1 measurements) and caged Gd (T_2 measurements) in the solvent or in lipids. In these approaches, water accessibility is also deduced from spatial proximities. Again, hydroxyl groups can act as a mediator of the effect (Linser et al. 2009; Pintacuda and Otting 2002).

H/D exchange is a more complex phenomenon than the mere water accessibility of protein sites. Given a general exposure of the site to bulk water, the exchange and its rate depend on various site-specific parameters, including the amide bond strength and how well the proton can be attacked by a Brönstedt base. This is linked to the bond angle with respect to the attacking water, distance, and also the dynamics of the protein site (Qin et al. 2012). The detailed chemical mechanisms for hydrogen exchange are specific to the case, but in general there are two possibilities (Eigen 1964; Englander et al. 1997; Miller and Dill 1995): In an acid-catalyzed mechanism of HDX, the exchange partner (deuteron) first binds to the nitrogen and only then the amide hydrogen is abstracted. Following a base-catalyzed mechanism, first the amide group is deprotonated, followed by binding a deuteron. The amide hydrogen exchange rate is the sum of the acid- and the base-catalyzed exchange rates, whereby both rates strongly depend on the pH. In most in vivo situations or under conditions representing those (with a pH of around 7), the acid-catalyzed share of the hydrogen exchange rate is negligibly small and only the base-catalyzed mechanism remains. Considering the predominant dissociative mechanism, the H-bond has to be disrupted temporarily in order to exchange the amide proton. Therefore, HDX experiments have been used to understand kinetics of H-bonding and obtain information about local stability and dynamics of secondary structure elements (Grantcharova and Baker 1997; Maity et al. 2003; Milne et al. 1998).

Non-equilibrium H/D exchange rates in *solution* NMR are usually determined via rapid dissolution of a protonated protein in deuterated buffer or similar (Sadqi et al. 2002). The measurements are usually performed under exact pH and temperature control. In *solid-state* NMR, time-resolved H/D exchange measurements (more than two steps, i.e., before and after D_2O treatment) are more difficult to achieve. This is due to the fact that the preparation and measurement of a sample are more time consuming (Brown et al. 2011) and need to be done under particular conditions (generating micro-crystals, fibrillization, or membrane protein lipid reconstitution etc.).

Further interaction with the sample (in the sense of solvent exchange) can be associated with drawbacks, for instance re-dissolution of microcrystalline preparations. Such handling is also generally restricted by the low rotor volume and sample viscosity, which can result in a loss of the sample. In addition, the presence of non-exchanging protons for doubly labeled samples of traditional solid-state NMR has made quantitative determination of the amide protonation state difficult, if not quenching of HDX and redissolution for solution-NMR readout was pursued. For fast exchange in perdeuterated proteins, a steady-state exchange experiment has been proposed that monitors sites undergoing H/D exchange on the time scale of seconds (Lopez del Amo et al. 2010). This approach is based on samples that contain a mixture of ^1H and ^2H in the buffer and the fact that protonated and deuterated amides can be treated differently in the experiment. Using this approach, five residues could be unambiguously characterized in the SH3 domain of α-spectrin (the protein also of this study), yielding fast exchange rates between 0.2 and 3.0 s^{-1}.

Here we present a method to directly access and characterize the behavior of H-bonded amide protons that slowly exchange with the solvent over a course of days to weeks. The experiments are based on proton detection in the solid state. As such, they are in contrast to exchange trapping or simple 2-step exchange quantification proposed earlier. This method is able to characterize exact time courses for exchange on different time scales in only 1 mg of protein, even though difficulties for absolute quantification remain.

Materials and methods

Sample preparation

The NMR spectra of the SH3 domain of chicken α-spectrin were recorded using approximately 1 mg of Cu-edta-"doped" protein, which was expressed and purified largely as described earlier (Linser et al. 2007). Two labeling

variants were employed, bacterial growth in deuterated ^{13}C-glucose and protonated water (sample 1), or uniformly deuterated ^{13}C-glucose and deuterated water (sample 2). Micro-crystallization was performed by pH shift from 3.5 to 7.5 using a buffer containing 100% D_2O or H_2O, respectively. In both variants, ^{15}N labeling was used. In addition, for non-equilibrium H/D exchange experiments, approximately 1 ml of the supernatant buffer of yet unpacked crystals from sample 2 was removed and lyophilized. Then the lyophilisate was redissolved in 85% D_2O and 15% H_2O. The treated buffer was added to the residual wet microcrystalline protein slurry such that the resulting proton concentration was approximately 20% based on total volume ratios. This slurry of microcrystals was vortexed briefly to expose the microcrystals to the new solvent conditions. It was then incubated for 5 min in order to allow diffusion of the new buffer into the crystallite water channels. The supernatant was taken off and the sample center-packed into a 1.3 mm rotor (sample 3). Using fluorinated rubber plugs in the bottom and top of the rotor, the filled sample was subjected to NMR experiments as follows.

NMR experiments

NMR-experiments were carried out at 800 MHz proton Larmor frequency at 55.55 kHz MAS and 32 °C effective temperature, using a standard-bore 1.3-mm triple-resonance probe on a Bruker Avance III spectrometer.

Hydrogen exchange was determined by following intensity changes of HN peaks from consecutively measured dipolar-transfer 2D HN spectra (Zhou et al. 2007) as a function of time. The maximum evolution in the indirect dimension was set to 55 ms, and WALTZ-16 (Shaka et al. 1983) and XiX decoupling (Ernst et al. 2003) were used in the direct and indirect dimension, using 2.9 and 2.4 kHz, respectively.

The total duration of the 2D experimental series was 2 weeks. Due to the lag time between exposure of the protein to the new buffer and completion of rotor filling and measurement setup, the first spectrum was started 115 min after adding the deuterated buffer. The next 10 HN-spectra were recorded with 8 scans, amounting to an experimental time of 14 min, each. The following 66 time points were measured with a growing number of scans from 16 up to 128 scans with a final experimental time of 443 min per experiment. After 1 week of consecutively measured HN-spectra, the sample was taken out and stored at 32 °C for 7 days before additional points of the series were recorded. Consequently, despite taking trials to accurately establish all original measurement conditions, the final points have to be taken into account with a higher error. The spectra were processed with Bruker Topspin. Apodization was performed using QSINE both in the direct and indirect dimension with a sine-bell-shift of $\pi/4$ and $\pi/2$, respectively. Peaks were analyzed using the program CCPNmr (Vranken et al. 2005) and intensities fitted as a function of time using the programs Origin and Mathematica (OriginLab; Wolfram Research 2010).

Even though fitting of the decay rates in an exponential decay is irrespective of absolute intensities, experimental peak heights were translated into approximate relative values for better overview. This was based on the assumption that after 2 weeks at 37 °C all residues have reached their equilibration values, as derived from the exchange curves.

Structure calculations

Structure calculation was carried out using the software CNS-Solve 1.1. For comparison two NMR structure ensembles were calculated. The first structure ensemble was calculated using 294 ^{1}H-^{1}H- RFDR restraints and 69 dihedral angle restraints as published before (Linser et al. 2011). The second structure ensemble was calculated using the same restraints in addition to 15 H-bond restraints (set to 1.9 ± 0.2 Å). These stemmed from H-bonds denoted as "strong" as judged in this study. Simulated annealing was performed in torsion angle space with 50,000 cooling steps in the high-temperature annealing stage and 150,000 cooling steps in the slow cooling annealing stage. From the 150 generated structures the 10 lowest-energy structures were used for the structure ensemble.

Results

Qualitative assessment

In a first context of H/D exchange in a microcrystalline SH3 sample we crystallized several mg of H_2O-purified and lyophilized protein in D_2O. This is a procedure similar to those commonly used for preparation of partially back-exchanged microcrystals. Despite the small size and the absence of a larger hydrophobic core, the resulting preparation showed clear signs of differential incomplete exchange of amide protons against deuterons. However, upon this procedure, only a few residues maintain protonated after dissolution in D_2O. Most residues, including for example T32 and L33, which belong to the slowest exchanging in the crystalline state (see below), seem to exchange relatively fast. This qualitative result is shown in Fig. 1a.

We wondered if the differential protonation levels obtained were purely due to differences in the exchange kinetics in solution, i.e., before attaching to the solid phase, or if and how exchange continues in the crystal. We thus attempted to quantify such exchange behavior in the crystalline state in its site-dependent time course. Therefore, we

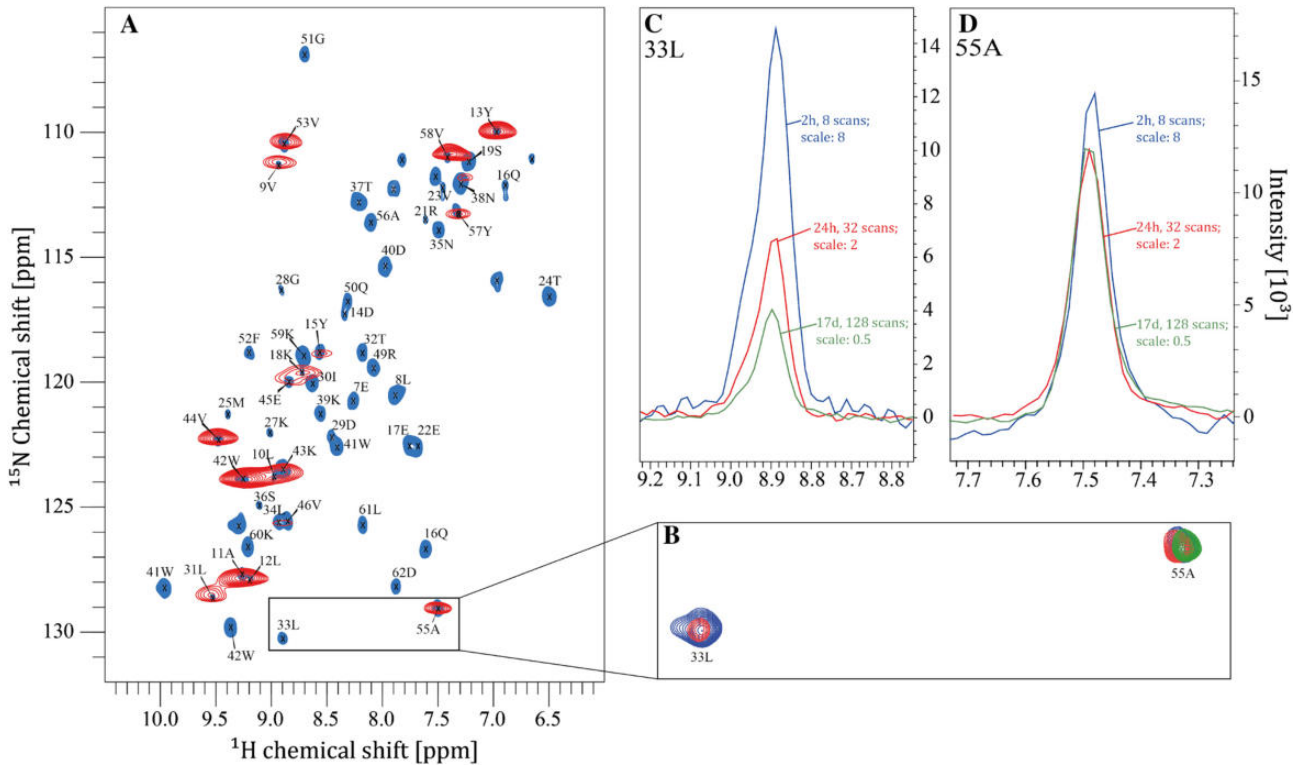


Fig. 1 H-D exchange as observed in different contexts of solid-state NMR. **a** The *red* spectrum was obtained from a partly protonated SH3 preparation that was purified and lyophilized in H_2O before dissolution and crystallization overnight in 100% D_2O (sample 1). The *blue* reference spectrum was recorded on a deuterated and 100% amide back-exchanged SH3 sample in H_2O (sample 2). Assuming immediate H/D-exchange upon dissolution in D_2O, in the red spectrum no peaks should be visible. Peaks which remain intense must undergo slower H/D-exchange over the process of dissolution and crystallization, **b** shows the overlay of three spectra chosen out of an hNH-series recorded for 17 days on sample 3. Here, fully amide-protonated crystals were subjected to partially deuterated buffer, reflecting HDX in a context *post* crystal formation. The spectrum in blue represents the time point after 2 h, that one in red after 24 h and the green spectrum after 17 days. **c** and **d** show the 1D cross sections of peaks 33L and 55A. The peak intensity of 33L slowly decays while the intensity of 55A has nearly reached its equilibrium value already at the first time point. Both of these residues, however, turn out to belong to the slowest exchanging ones (see below)

prepared the same amount of a perdeuterated but 100% proton-backexchanged sample and lyophilized part of its crystallization mother liquor. The D_2O-redissolved lyophilisate was mixed with the slurry of the protonated crystallization mixture as described in the Methods section.

Time-resolved H/D exchange was observed by a series of consecutive H/N correlations. This series started after a lag time of 144 min after exposure of the protein to the partly H-depleted buffer. Altogether, we obtained qualitatively interpretable decay curves for 29 out of 41 non-ambiguous (spectrally well-resolved) SH3 residues. Generally, most residues are comparably fast and have decayed almost to a steady state after 40 h. This is in most cases a plateau, which represents the final protonation level of the sample around 20%. A set of 16 different residues have already reached their final equilibrium protonation before the first NMR measurement, as judged by a flat, non-decaying profile. A subset of residues reaches their steady state much more slowly. Here the exchange has not come to equilibrium at the end of the experiment after 2 weeks. Figure 2 shows a selection of representative curves for some residues. Figures S1 and S3 (Supporting Information) display the obtained curves that were used for further analysis.

It is expected from HDX approaches in solution NMR or mass spectrometry monitoring protein water interactions that many of the amide protons instantaneously exchange to their final equilibrium protonation level. These are the water-exposed residues on the protein surface that are not or weakly H-bonded. For stronger H-bonds, from a naive point of view, one would expect a simple mono-exponential intensity decay over the incubation time with different rates. This is, however, not the case for all of the peaks. For several residues, the exponential decay functions are supplemented by an additional, very slow exponential contribution to the intensity decay (Fig. 2d, e). Opposed to global effects from instabilities like wobb, shim, temperature, etc. (see below), we observe here a residue-specific, systematic additional decrease of intensity for these residues.

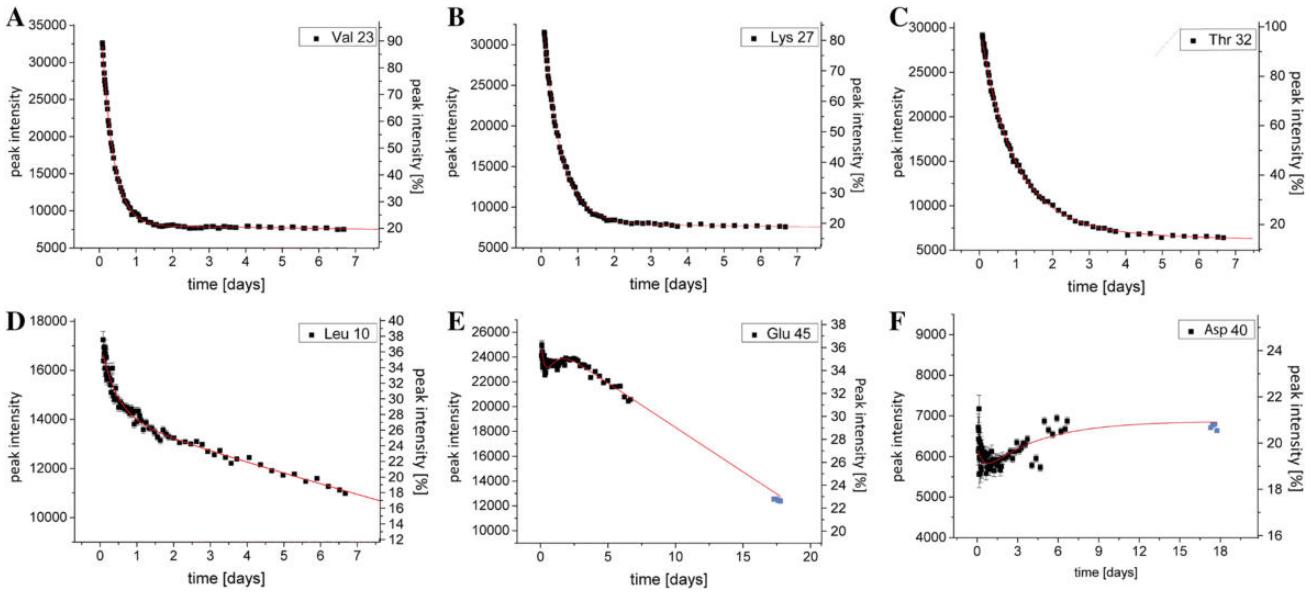


Fig. 2 Representative examples for H/D exchange, observed by consecutive 2D HN-spectra as described above. **a, b**, and **c** show a regular exponential decay, **d** shows an exponential decay regarding the first time points, while after 1 day of measurement time, the signal decay follows a second, much slower component and apparently linear function, **e** shows a fast initial decay of signal intensity within the lag period of the experiment, followed by a intensity decay in a linear manner. **f** The initial decay nearly reached its equilibrium whereas the signal slowly rises again in the following days (rise within the error range). The last points shown only in **e** and **f** (*light blue*) were measured after the sample had been taken out and stored at the same temperature for 7 days. No quantitative data was extracted from the trend lines in **e** and **f**. Protonation percentage scales on the right are only approximate due to unknowns upon conversion between individual peak heights (*left* scale, as extracted from peak heights in HN 2D spectra) and actual protonation degrees. Peak intensity values (*left* y-axis) are direct readouts from H/N correlation peak heights

In previous studies, a double exponential decay function was described for the H/D exchange of sidechain hydroxyl groups (Agarwal et al. 2010). Measuring at a timescale of milliseconds, Agarwal et al. contribute the two exponential exchange rates to the opening rate of the corresponding H-bond and to the intrinsic exchange rate. Here, using non-equilibrium H/D exchange, the intrinsic rates of the exchange (on the order of 1 min^{-1}) would be too fast to be observed (Bai et al. 1993; Englander et al. 1972; Koide et al. 1995). For non-solvent-accessible residues, however, the exchange will be additionally slowed down by the barrier between H-bond and water proton. As such, we think that the biexponential decay apparent for some residues here might represent the two processes *opening of the H-bond* and *hindered water accessibility*.

The additional exponential, seemingly linear contribution tends to occur for residues which are shielded by hydrophobic side chains of neighboring residues (see below). For solvent accessibility in the hydrophobic core of proteins different mechanisms have generally been proposed. One of the two major models is the so-called penetration model: Here water molecules are able to penetrate the hydrophobic core due to fluctuations in the protein structure. The other proposed mechanism is the local-unfolding model: Concerted motions of secondary structural elements, resembling a breathing motion, enable the otherwise shielded amide protons of hydrophobic regions to exchange (Englander et al. 1980; Jardetzky and Finucane 1998; Tomita et al. 2009). Both models are extreme cases of proton trajectories associated with higher-energy (and thus low-probability) transition states.

For individual sites, the exchange can be reminiscent of multiple-steps reaction kinetics with individually different, coupled pathways between amide and bulk water. This hurdle is associated with numerous unknowns. As such, a fully quantitative description of the processes on the basis of the here existing information is not possible unfortunately.

Semi-quantitative analysis of exchange rates:

The eventual equilibration with bulk water described above is slightly different from the solution state case: In the liquid state, there is a high excess of bulk water. In solid-state NMR rotors, the increase of the proton concentration in the bulk solvent stemming from the amide H/D exchange in the sample is more significant and can potentially appear in the exchange curves. Roughly 500 mg ml^{-1} would be obtained for optimal filling (pure crystal). Assuming on the order of 100 exchangeable protons of the 7 kDa protein, we reach maximally 7 moles of those compared with 55.5 moles of

water protons (in 100% H_2O) per liter crystal. (These contain 500 mg ml^{-1} water and two protons per water molecule.) Realistically, we assume 200 mg ml^{-1} protein in the rotor in addition to bulk water due to imperfectly dense filling. Consequently, these numbers transform into 2.8 and 88.8 M of exchangeable protein protons and water protons, respectively. We start with fully protonated (in exchangeable sites) protein and approximately 20% water protonation. Thus, the additional 2.8 M protons would mean a slow rise of the water protonation level from 20% to a final of 23%, or a deviation by 3%. We neglect this minor (and difficult to quantify) contribution to the baseline of the exponential decays, considering that the actual amide decay spans a range (from 100% to ca. 20%) that is approximately 25 times larger.

As detailed above, in non-equilibrium H/D exchange experiments hindered water accessibility of the amide-proton can interfere with the actual exchange. This is the case when it becomes comparably slow as the chemical exchange. For the residues next to the water channels of the micro-crystalline sample, we expect a non-perturbed exchange decay. For residues where the opening rate of the H-bond and the accessibility hindrance are on the same time scale, the two will lead to inseparable bi-exponential decays. For residues with two very different contributions (i.e., for example a strongly hindered water accessibility) both could be disentangled qualitatively by fitting the curves with a doubly exponential equation (Agarwal et al. 2010):

$$\left[A_a\right](t) = n_1 \times e^{-k_1 \times t} + n_2 \times e^{-k_2 \times t} + y_0 \quad (1)$$

In addition to the uncertainties about the exact mathematical description of the complex exchange process, a doubly exponential fit (like using Eq. 1, having five fit parameters) can easily over-parameterize experimental data. For completeness, fitting of the data using a biexponential decay function is shown in Fig. S3. Here we involved an upper limit for the additional (slow) exponential rate, which was set to 8.3×10^{-6} min^{-1}. This was done in order to prevent perturbation of the first (fast) exponential rate (which are the data of interest).

An alternative way to overcome the problem of an over-parameterized fit is to treat additional slow exponential components as infinitively slow and dispense quantification of k_2. This leads back to a mono-exponential equation, now with an additional linear function:

$$[A_a](t) = n \times e^{(-k \times t)} + D \times t + y_0 \quad (2)$$

(Results of the fitting using Eq. 2 are displayed in Fig. S1 and Table 1 of the Supporting Information). This simplification procedure is justified since any continuous function inducing slow changes (like the slow exponential process expected for this effect) can be approximated to have a linear behavior over small time intervals. As such, similar to a linear baseline correction, the more relevant parameters for faster processes do not experience any differences. This is despite the fact that a linear correction obviously represents non-physical behavior for the expected effects and cannot deliver absolute data for the slower decay component.

Both simplified approaches will yield qualitative data only. A complete treatment of the exchange scenario in the sense of differential equations on the other hand leads to equations dependent on a multitude of unknowns and did not turn out to be any helpful for us.

Smaller experimental artifacts are apparent in the profiles. This includes the uncertainty of the last measurement point as well as a slight non-sensible increase of intensity (on the order of up to 2%) for some residues. Decreasing homogeneous line widths due to decreasing levels of protonation are a potential source of a minor increase in overall sensitivity over the experiment. We observe, however, maintained signal intensity for the spontaneously exchanging residues. Instrument instability, for instance field drift or changes in tuning or rf powers, would cause mostly detrimental effects (peak decay). Such factors are difficult to monitor for experiments with chemical changes in the sample over time. In general, measurement artifacts seem to be of little effect for the qualitative assessment here. This is in line with what is expected from the performance on the instrument. Even if slight magnet instability occurred, differences between residues would still be untouched. As such, conclusions about *relative* exchange behavior are expected to be correct.

Hydrogen bond strength from HDX rates

Figure 3 shows the measured HDX-rates for each residue obtained using the simplistic mathematical interpretation (in Eq. 1) outlined above. These rates are very similar (mostly identical) to those obtained when using Eq. 2. Table 1 (Supporting Information) again shows both numbers for all exchange rates obtained. Despite the fact that we can only observe such events that are slower than the 2-h dead time of the experimental approach (which excludes amides not H-bonded or in very weak H-bonds), we see significant differences between different sites for this parameter. Figure 3 differentiates slow and fast exchange processes by red and blue colors. This color coding refers to a deliberate threshold of 2×10^{-3} min^{-1}, below which we denote the H-bond as "strong" (slow exchange, depicted also as blue lines in B). Such qualitative procedure has been done in the literature previously (Stefanowicz et al. 2009). The H-bonds denoted as "strong" all show a distance in the crystal structure closely around 1.9 Å between carbonyl oxygen and amide proton. Those residues

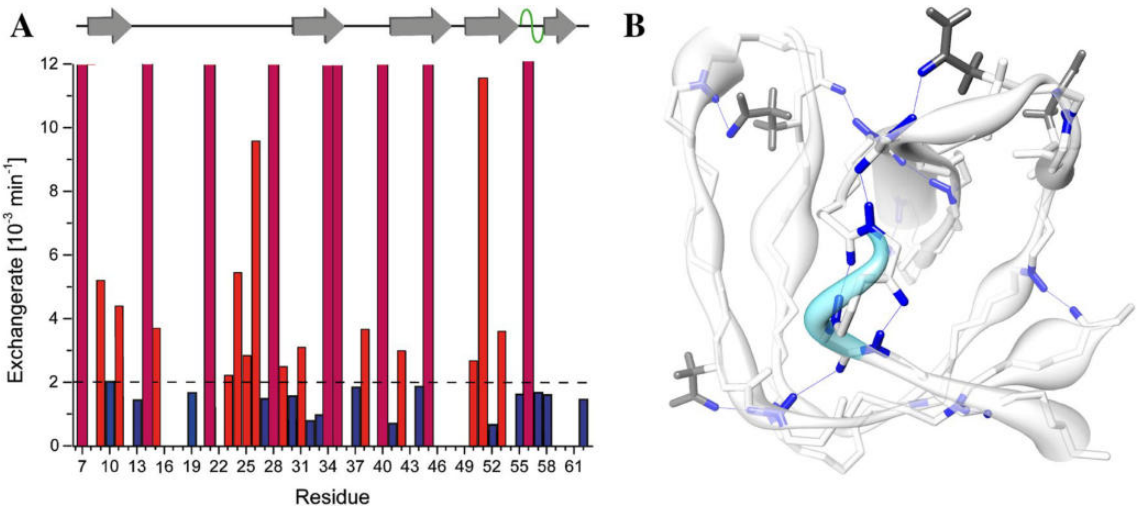


Fig. 3 Hydrogen deuterium exchange rates in min^{-1}. **a** The ruby-colored columns (reaching the *top* of the figure) depict residues with exchange rates too fast for recording using the experimental setup described here. Their protonation levels have already reached equilibrium by the first time point (after 115 min). The shorter red and blue columns represent fast (but measurable) and slow hydrogen deuterium exchange caused by intermediate and strong H-bonding, respectively. We deliberately distinguished between fast and slow by drawing a line at 2×10^{-3} min^{-1}. **b** Strong H-bonds, related with the slowly exchanging residues, depicted as blue lines on the crystal structure of SH3. The structure is drawn such that ribbon thickness represents S^2-values from dipolar recoupling (Chevelkov et al. 2009) in addition. Thinner ribbon reflects a larger S^2 value and therefore denotes a more rigid part of the protein. Even though some shared tendencies might be interpreted, a consistent match between exchange rates and dipolar order parameters is not apparent. Side chains (glutamic acid, asparagine and aspartic acid) involved in strong hydrogen bonds with backbone amides are depicted in *gray*

depicted as ruby bars reaching the top of the plot ("fast" exchange, i.e. "weak" H-bonds) are amides exchanging faster than the detection limit and have a flat exchange profile. Red bars (not reaching the upper end of the plot) are rates that we classify as intermediate.

Several residues show invariable intensities close to approximately 20%, which hints to complete exchange already within the lag time (i.e., a very fast exchange). Given the missing zero-point of the measurement for intensity calibration, the relative intensities can only be approximate (see Materials and Methods). As such, a completely flat intensity profile (within error) over the course of the measurement could potentially be mistaken with a very slow residue, too. This potential pitfall, however, is very unlikely here: All of these residues are clearly water-exposed and they involve those residues previously identified in the steady-state exchange experiment by Lopez del Amo et al. (Lopez del Amo et al. 2010). In addition, qualitative comparison with intensities of a standard sample separately reveals again low peak intensities and thus hints to complete equilibration having taken place.

As mentioned in the introduction, the amide proton (at a pH around 7) can only be exchanged whenever it is not forming an H-bond, at least temporarily. The amides observed are almost all part of H-bonds in the SH3 crystal structure 2NUZ (Chevelkov et al. 2007). As such, the intrinsic HD-exchange rate of H-bonded amides observed to undergo exchange is dependent of the opening rate of the H-bond. Therefore, the exchange behavior should be affected by protein dynamics. We thus asked if we would see any correlation of the rates with motional order parameters. The plot in Fig. 3b reflects dipolar order parameters (Chevelkov et al. 2009) as the thickness of the protein backbone. The comparison shows the tendency that regions with fast exchanging residues (high rates fitted by Eqs. 1, 2) tend to show lower order parameters and therefore higher motional amplitude. It is obvious, however, that the kind of dynamics involved in H/D exchange is not represented adequately by this measure. This is not surprising given that the time scale of exchange is off by orders of magnitude.

Based on the results presented here, the solid-state NMR non-equilibrium exchange experiment is not suitable for exchange rates faster than $2–5\times10^{-3}$ min^{-1}. It is thus complementary to the equilibrium method for HDX quantification by Lopez del Amo et al., where only the peaks K60, S36, R21, $W41_{\varepsilon}$, $N35_{sc}$ and $N38_{sc}$, comprising fast exchange, could be measured accurately (Lopez del Amo et al. 2010).

Water accessibility from H/D-exchange behavior

As mentioned above, the details about how solvent water molecules get access to the hydrophobic protein core are not clear. Also, they are complex and individually different for the various amides. We assume, nevertheless, that the presence of an additional slow exponential part of the

decay function (Eq. 1) might reflect the hindrance of the amide site to be in contact with water protons. Equation 2 introduces the approximation of an infinitively slow second exponential by a linear function, reflecting the hindrance of water accessibility by its slope. We only evaluated the existence of a linear component of this fitting to analyze the water accessibility hindrance in a tentative "yes or no" manner. If the fitted slope (Fig. S2 supporting information) exceeds 1.9×10^{-5}% min^{-1} in magnitude, the residue was deliberately counted as hindered solvent accessible. If, on the other hand, the fitted value for the slope is lower than 5.2×10^{-6}% min^{-1} in magnitude (the additional "linear" contribution being absent), the residue was deliberately counted as highly water accessible. All values between these borders were excluded due to the mechanistic unknowns and the extremely simplified fitting equation. Figure 4a shows a comparison between hydrophobicity trends over sequence and the presence of accessibility hindrance as obtained by this binary analysis. Hydrophobicity is plotted as a moving average of three. The color code refers to the presence or absence of a linear decay part with criteria as detailed above. (Blue colored columns: accessibility hindrance, red coloring: high water accessibility.) Even though there are some exceptions, a qualitative trend is indeed observed for regions of many hydrophobic residues in which access to bulk water can be assumed to be hindered. Going through the amides one by one reveals that most of the "hindered" amides are within the core of the protein (see Fig. 4b) or have hydrophobic groups shielding them from the solvent (see Fig. 4 c). This suggests that the additional contribution to the exponential decay curve might reflect a residue-specific water accessibility feature. Despite a qualitative match, some exceptions are obvious, as for instance for the residues 15, 25, and 41, which show a slope exceeding the border of 1.9×10^{-5}% min^{-1}, even though they are surrounded by hydrophilic residues. Such deviations would be expected, however, as hydrophobicity of the primary sequence is only a rather weak representation of the hydrophobic shielding from residues close in space. Individual slopes for each residue are shown in Fig. S2 (Supplementary Information).

In a previous study of water accessibility of SH3 amide groups in the solid state (Chevelkov et al. 2005), the accessibility was measured by a magnetization transfer from water to the amide nitrogen as described in the Introduction. Chevelkov et al. compared their measured magnetization buildup with distances from amides to refined water molecules in the crystal structure. They assumed that magnetization transfer originates mainly from rigid water molecules. This comparison revealed SH3 surfaces that are accessible to such (mobile) water molecules which do not provide magnetization in transfer experiments. Such residues close to water molecules that must be able to exchange with the bulk solvent are K60, S36, Q50, D14, and K39. Also in our study, residues K60, S36, D14, and K39 show clearly no linear-decay behavior and are denoted as solvent accessible. For Q50 no clear trend can be read out.

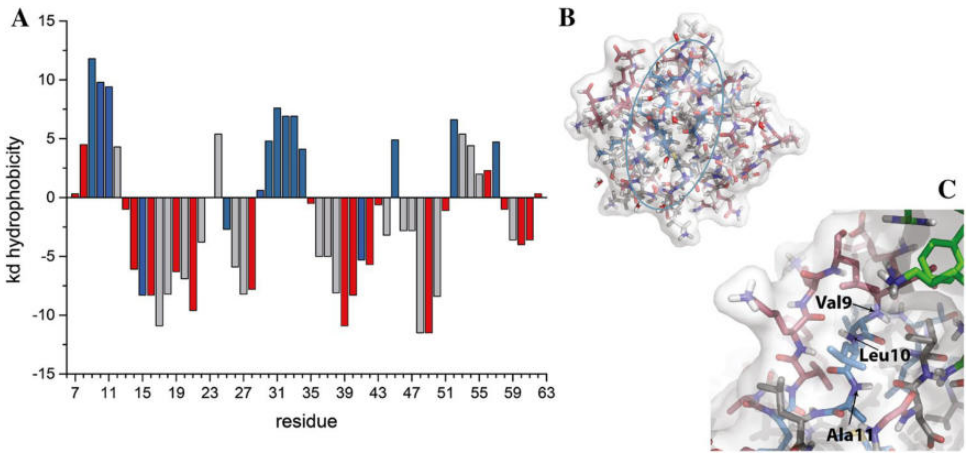


Fig. 4 Qualitative agreement between the presence of an accessibility hindrance component of the decay curve and hydrophobicity of the surrounding residues. a Kyte-Doolittle hydrophobicity by residue (moving average of three). Blue color marks the residues with distinct linear decay-behavior. The criterion is a slope D, (as fitted using Eq. 2) with a magnitude larger than 1.9×10^{-5}% min^{-1}. Red color marks the residues with clearly no linear decay-behavior, the criterion set to be a slope lower than 5.2×10^{-6}% min^{-1} in magnitude. Gray colors denote ambiguous or missing residues and residues with no distinct properties, showing a fitted slope between 5.2×10^{-6} and 1.9×10^{-5}% min^{-1} in magnitude (Fig: S2, Supporting Information). b Depiction of the residues with distinctly slanted decay curves (colored in *blue*, mostly in the center of the protein, *blue* ellipse) and those with clearly no slant (colored in *red*) in the SH3 monomer structure. c Close-up into the amides of V9-A11 in the crystal lattice (*green*), which are effectively protected against the surface water by hydrophobic side chains

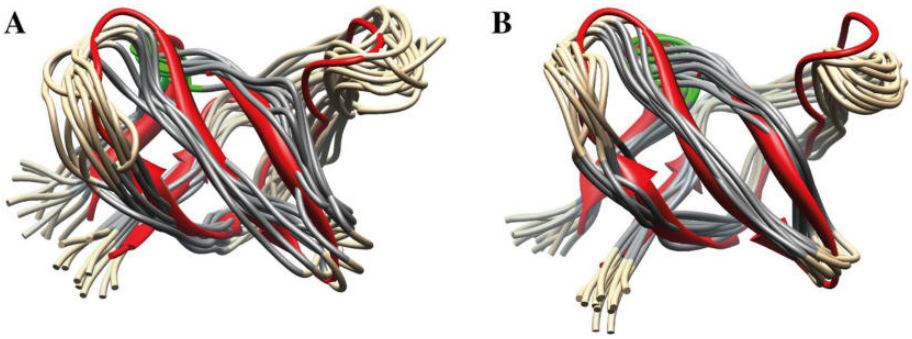


Fig. 5 H-bonds for structure calculation. **a** NMR structure model (*brown, gray* (strands), and *green* (helices)) calculated using 294 RFDR- and 69 TALOS-restraints aligned to the crystal structure (*red*). **b** NMR structure model calculated using 294 RFDR-, 69 TALOS- and 15 H-bond-restraints aligned to the crystal structure (same colors). The Backbone RMSD (all residues) excluding H-bond restraints is 1.72 (**a**) and including H-bond restraints 1.03 Å (**b**). The average derivation from the crystal structure is **a** 2.31 and **b** 2.33 Å. All alignments were pursued using backbone atoms CA, CO, and N of the residues 7–61

Use of hydrogen bond information for NMR-based structure calculation

H-bonds are generally used as a restraint for various aspects of structure calculation (Hagn et al. 2013; Lührs et al. 2005; Meier et al. 2006). Dependent on how structurally favorable the H-bond is, the bond will be characterized as a strong or weak one in the HDX experiments above. Accordingly, if a strong H-bond is found experimentally, this can be reflected as a dedicated restraint. Of course this necessitates that RFDR distances etc. unambiguously ascertain which are the two residues that may actually be involved in the bond. For strong H-bonds, we used a distance between amide and carbonyl oxygen of 1.9 ± 0.2 Å as restraints for structure calculation. In addition, we used 294 published RFDR restraints and TALOS backbone dihedral angles (Linser et al. 2011). As expected, it turns out that the precision of the structure calculation becomes significantly better (decrease of the rmsd from 1.7 to 1.0 Å), whereas the accuracy is naturally little affected (see Fig. 5). As such, meaningful H-bond restraints in solid-state NMR spectroscopy from quantification of HDX-exchange can be helpful for protein structure determination. This is particularly true for otherwise underdetermined structures. Even though the exchange rate data only denote the existence of an H-bond, they provoke the search for the partner moiety. In future algorithms it might be useful to implement qualitative restraints "H-bonded amide" or "non-H-bonded amide". This could be done instead or in addition to a precise distance specified between two nominated H-bonded atoms.

Conclusions

The time-resolved measurement of proton deuterium exchange in solid-state NMR is able to deliver qualitative data on the strength of hydrogen bonds. This is true even though it still harbors more difficulties here than in liquid-state NMR regarding the practical measurement and data interpretation. As such, we can only extract qualitative information on the strength of protein H-bonds, which also turns out to be useful for structure calculation. The effects that cause the additional difficulties in quantitative data analysis, however, potentially contain interpretable information that can be useful for structural biology. One such feature is an additional slow component in the exchange profiles. Its existence might be an indirect reporter on the hydrophobicity of the surrounding of a residue or on the residue being buried in the core. We believe that future solid-state NMR structural-biology studies will increasingly benefit from data of this kind in various ways. This particularly includes membrane proteins and protein fibrils.

Acknowledgements RL acknowledges support from the Deutsche Forschungsgemeinschaft (SFBs 803, project A04, and 749, as well as the Emmy Noether program), the Verband der Chemischen Industrie (VCI) in terms of a Liebig junior group fellowship, and the Excellence Cluster CIPSM. We thank Anne Diehl and Kristina Rehbein (FMP Berlin) for their kind support regarding protein purification.

References

Ader C, Schneider R, Baldus M (2009) Structural rearrangements of membrane proteins probed by water-edited solid-state NMR spectroscopy. J Am Chem Soc 131:170–176

Agarwal V, Linser R, Fink U, Faelber K, Reif B (2010) Identification of hydroxyl protons, Determination of their exchange dynamics, and characterization of hydrogen bonding in a microcrystallin protein. J Am Chem Soc 132:3187–3195

Andronesi OC, von Bergen M, Biernat J, Seidel K, Griesinger C, Mandelkow E, Baldus M (2008) Characterization of Alzheimer's-like paired helical filaments from the core domain of tau protein using solid-state NMR spectroscopy. J Am Chem Soc 130:5922–5928

Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. Proteins 17:75–86

Brown LS, Shi LC, Ladizhansky V (2011) Conformation of a seven-helical transmembrane photosensor in the lipid environment. Angew Chem Int Ed 50:1302–1305

Chevelkov V, Faelber K, Diehl A, Heinemann U, Oschkinat H, Reif B (2005) Detection of dynamic water molecules in a microcrystalline sample of the SH3 domain of a-spectrin by MAS solid-state NMR. J Biomol NMR 31:295–310

Chevelkov V, Faelber K, Schrey A, Rehbein K, Diehl A, Reif B (2007) Differential line broadening in MAS solid-state NMR due to dynamic interference. J Am Chem Soc 129:10195–10200

Chevelkov V, Fink U, Reif B (2009) Accurate determination of order parameters from 1 H,15 N dipolar couplings in MAS solid-state NMR experiments. J Am Chem Soc 131:14018–14022

Cho MK, Kim HY, Fernandez CO, Becker S, Zweckstetter M (2011) Conserved core of amyloid fibrils of wild type and A30P mutant α-synuclein. Protein Sci 20:387–395

Cook PF, Cleland WW (2007) Enzyme kinetics and mechanism. Garland Science, New York

Daebel V et al (2012) β-Sheet core of tau paired helical filaments revealed by solid-state NMR. J Am Chem Soc 134:13982–13989

Dasari M et al (2011) Bacterial inclusion bodies of Alzheimer's disease β-amyloid peptides can be employed to study native-like aggregation intermediate states. ChemBioChem 12:407–423

Eigen M (1964) Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. Angew Chem Int Ed Engl 3:1–19

Englander SW, Downer NW, Teitelbaum H (1972) Hydrogen exchange. Annu Rev Biochem 41:903–924

Englander SW et al (1980) Individual breathing reactions measured in hemoglobin by hydrogen exchange methods. Biophys J 32:577–589

Englander SW, Mayne L, Bai Y, Sosnick TR (1997) Hydrogen exchange: the modern legacy of Linderstrøm-Lang. Protein Sci 6:1101–1109

Ernst M, Samoson A, Meier BH (2003) Low-power XiX decoupling in MAS NMR experiments. J Magn Reson 163:332–339

Gallagher W, Tao F, Woodward C (1992) Comparison of hydrogen-exchange rates for bovine pancreatic trypsin-inhibitor in crystals and in solution. BioChemistry 31:4673–4680

Gaspari R et al (2016) Kinetic and structural insights into the mechanism of binding of sulfonamides to human carbonic anhydrase by computational and experimental studies. J Med Chem 59:4245–4256

Grantcharova VP, Baker D (1997) Folding dynamics of the src SH3 domain. BioChemistry 36:15685–15692

Hagn F, Etzkorn M, Raschle T, Wagner G (2013) Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. J Am Chem Soc 135:1919–1925

He L et al (2016) Structure determination of helical filaments by solid-state NMR spectroscopy. Proc Natl Acad Sci 113:E272–E281

Jardetzky O, Finucane MD (1998) NMR relaxation measurements of backbone proton exchange in proteins permit a distinction between different mechanisms of exchange. Mol Phys 95:1127–1136

Kato H, Gruschus J, Ghirlando R, Tjandra N, Bai Y (2009) Characterization of the N-terminal tail domain of histone H3 in condensed nucleosome arrays by hydrogen exchange and NMR. J Am Chem Soc 131:15104–15105

Koide S, Jahnke W, Wright PE (1995) Measurement of intrinsic exchange rates of amide protons in a 15N-labeled peptide. J Biomol NMR 6:306–312

Lesage A, Böckmann A (2003) Water-protein interactions in microcrystalline Crh measured by 1 H-13C solid-state NMR spectroscopy. J Am Chem Soc 125:13336–13337

Lesage A et al (2008) Polarization transfer over the Water–Protein interface in solids. Angew Chem Int Ed 47:5851–5854

Linser R, Chevelkov V, Diehl A, Reif B (2007) Sensitivity enhancement using paramagnetic relaxation in MAS solid-state NMR of perdeuterated proteins. J Magn Reson 189:209–216

Linser R, Fink U, Reif B (2009) Probing surface accessibility of proteins using paramagnetic relaxation in solid-state NMR spectroscopy. J Am Chem Soc 131:13703–13708

Linser R, Bardiaux B, Higman V, Fink U, Reif B (2011) Structure calculation from unambiguous long-range amide and methyl (1)H-(1)H distance restraints for a microcrystalline protein with MAS solid-state NMR spectroscopy. J Am Chem Soc 133:5905–5912

Lopez del Amo J-M, Fink U, Reif B (2010) Quantification of protein backbone hydrogen-deuterium exchange rates by MAS solid-state NMR spectroscopy. J Biomol NMR 48:203–212

Lührs T et al (2005) 3D structure of Alzheimer's amyloid-β(1–42) fibrils. Proc Natl Acad Sci USA 102:17342–17347

Luo WB, Hong M (2010) Conformational changes of an ion channel detected through water-protein interactions using solid-state NMR spectroscopy. J Am Chem Soc 132:2378–2384

Maity H, Lim WK, Rumbley JN, Englander SW (2003) Protein hydrogen exchange mechanism: local fluctuations. Protein Sci 12:153–160

Meier S, Strohmeier M, Blackledge M, Grzesiek S (2006) Direct observation of dipolar couplings and hydrogen bonds across a β-hairpin in 8 M urea. J Am Chem Soc 129:754–755

Miller DW, Dill KA (1995) A statistical mechanical model for hydrogen exchange in globular proteins. Protein Sci 4:1860–1873

Milne JS, Mayne L, Roder H, Wand AJ, Englander SW (1998) Determinants of protein hydrogen exchange studied in equine cytochrome c. Protein Sci 7:739–745

Olofsson A, Sauer-Eriksson AE, Öhman A (2006) The solvent protection of Alzheimer amyloid-β-(1–42) fibrils as determined by solution NMR spectroscopy. J Biol Chem 281:477–483

Pintacuda G, Otting G (2002) Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. J Am Chem Soc 124:457–471

Qin H, Lim L, Song J (2012) Protein dynamics at Eph receptor-ligand interfaces as revealed by crystallography, NMR and MD simulations. BMC Biophys 5:2

Raschke TM, Marqusee S (1998) Hydrogen exchange studies of protein structure. Curr Opin Biotechnol 9:80–86

Sadqi M, Casares S, López-Mayorga O, Conejero-Lara F (2002) The temperature dependence of the hydrogen exchange in the SH3 domain of α-spectrin. FEBS Lett 527:86–90

Shaka A, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broadband decoupling: WALTZ-16. J Magn Reson (1969) 52:335–338

Stefanowicz P, Petry-Podgorska I, Kowalewska K, Jaremko L, Jaremko M, Szewczuk Z (2009) Electrospray ionization mass spectrometry as a method for studying the high-pressure denaturation of proteins. Biosci Rep 30:91–99

Tomita A et al (2009) Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. Proc Natl Acad Sci 106:2612–2616

Vranken WF et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

Wagner G, Wüthrich K (1979) Correlation between the amide proton exchange rates and the denaturation temperatures in globular proteins related to the basic pancreatic trypsin inhibitor. J Mol Biol 130:31–37

Weingarth M, van der Cruijsen EAW, Ostmeyer J, Lievestro S, Roux B, Baldus M (2014a) Quantitative analysis of the water occupancy around the selectivity filter of a K+ channel in different gating modes. J Am Chem Soc 136:2000–2007

Weingarth M, van der Cruijsen EAW, Ostmeyer J, Lievestro S, Roux B, Baldus M (2014b) Quantitative analysis of the water occupancy around the selectivity filter of a K+ channel in different gating modes. J Am Chem Soc 136:2000–2007

Whittemore NA, Mishra R, Kheterpal I, Williams AD, Wetzel R, Serpersu EH (2005) Hydrogen–deuterium (H/D) exchange mapping of Aβ1-40 amyloid fibril secondary structure using nuclear magnetic resonance spectroscopy. BioChemistry 44:4434–4441

Williams JK, Hong M (2014) Probing membrane protein structure using water polarization transfer solid-state NMR. J Magn Reson 247:118–127

Wolfram Research (2010) Mathematica, Version 8.0, Champaign, IL

Wydrzynski T, Hillier W, Messinger J (1996) On the functional significance of substrate accessibility in the photosynthetic water oxidation mechanism. Physiol Plant 96:342–350

Yi Q, Baker D (1996) Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. Protein Sci 5:1060–1066

Zhang Y-M, Hurlbert J, White SW, Rock CO (2006) Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of streptococcus pneumoniae. J Biol Chem 281:17390–17399

Zhou Z et al (2007) A new decoupling method for accurate quantification of polyethylene copolymer composition and triad sequence distribution with 13 C NMR. J Magn Reson 187:225–233

Non-equilibrium hydrogen exchange for determination of H-bond strength and water accessibility in solid proteins.

Supporting Information

Kristof Grohe[a,b], **Kumar Tekwani Movellan**[b], **Suresh Kumar Vasa**[a,b], **Karin Giller**[b], **Stefan Becker**[b] **and Rasmus Linser**[a,b,1]

[a]Department Chemie und Pharmazie, Ludwig-Maximilians-Universität München, 81377 München, Germany;

[b]Department for NMR-Based Structural Biology Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany;

[1]**To whom correspondence should be addressed. E-mail:** rasmus.linser@lmu.de

Site-resolved hydrogen non-equilibrium exchange rates

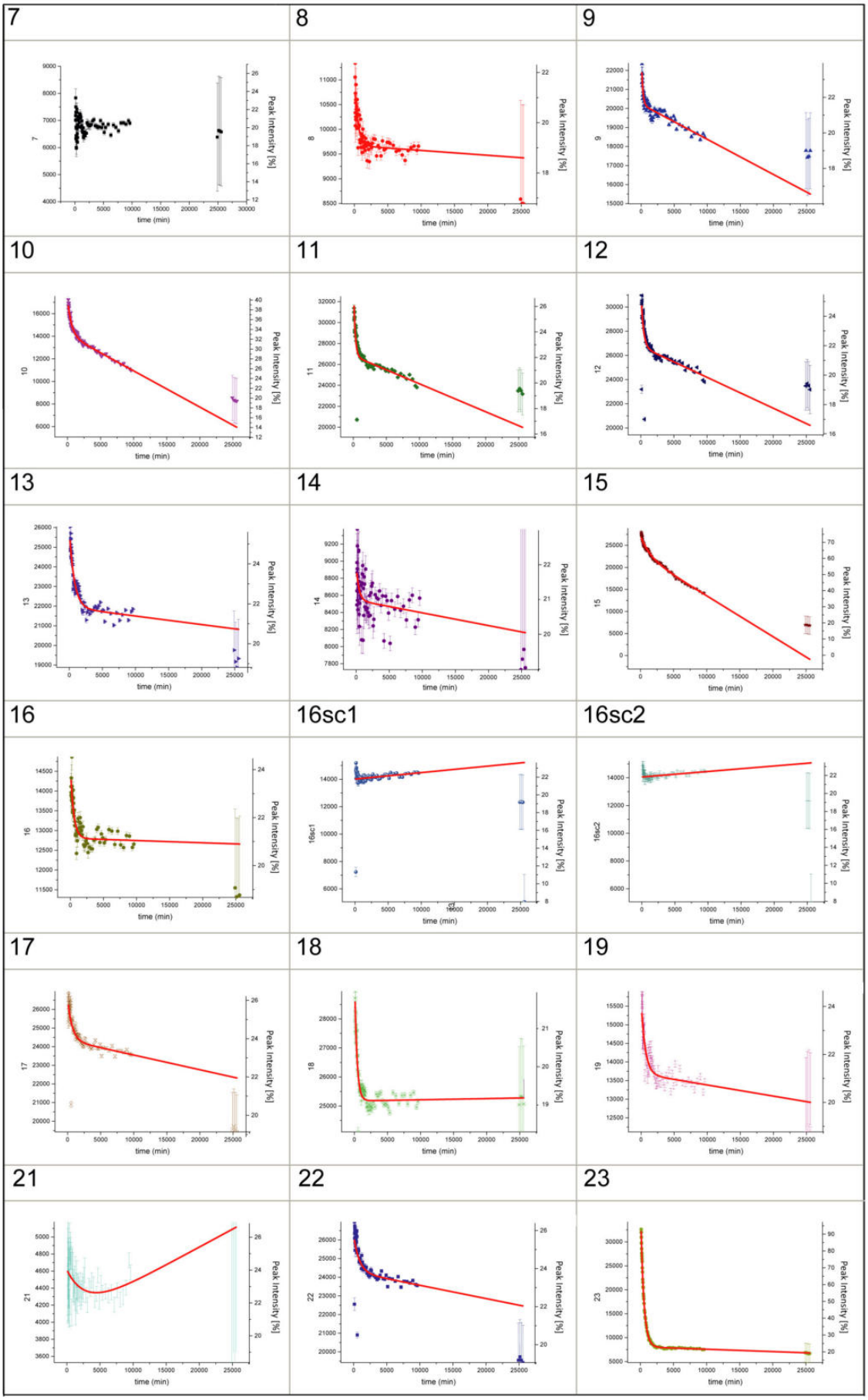
residue	k (Eq. 2) [10^{-3} ·min^{-1}]	Error [10^{-3} · min^{-1}]	Adj. R^2	k_1 (Eq. 1) [10^{-3} ·min^{-1}]	Error [10^{-3} · min^{-1}]	Adj. R^2
7	eq.*			eq.*		
8	1.92#	0.424	0.66	1.92#	0.427	0.65
9	5.05	1.39	0.87	5.20	1.57	0.89
10	1.63	0.162	0.98	2.03	0.299	0.98
11	4.26	1.23	0.71	4.39	1.45	0.70
12	overlap			overlap		
13	1.35	0.185	0.87	1.40	0.336	0.87
14	eq.*			eq.*		
15	0.671	0.116	0.99	3.69	0.791	0.99
16	2.07#	0.447	0.63	2.07#	0.451	0.63
17	overlap			overlap		
18	overlap			overlap		

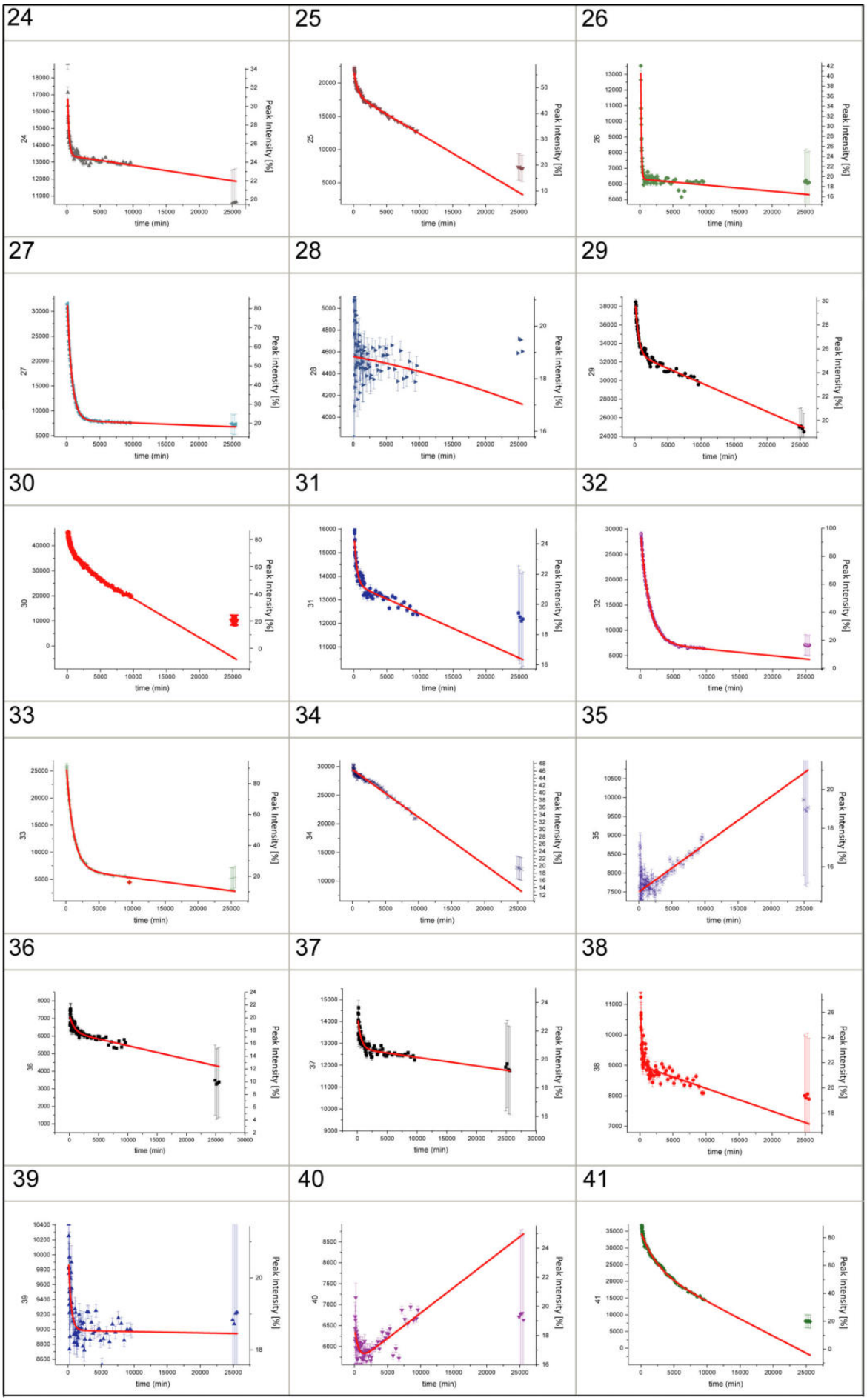
19	1.63	0.349	0.72	1.63	0.351	0.72
21	eq.*			eq.*		
22	overlap			overlap		
23	2.19	0.0231	1.0	2.20	0.0338	1.0
24	4.73	0.600	0.84	5.48	0.805	0.85
25	1.77	0.198	0.99	2.86	0.464	0.99
26	9.61	0.992	0.85	9.87	1.10	0.85
27	1.45	0.0153	1.0	1.46	0.0219	1.0
28	eq.*			eq.*		
29	2.05	0.191	0.93	2.49	0.343	0.96
30	0.780	0.0860	0.99	1.57	0.244	1.0
31	2.51	0.416	0.90	3.12	0.688	0.90
32	0.748	0.0150	1.0	0.790	0.0422	1.0
33	0.935	0.0193	1.0	0.963	0.0381	1.0
34	eq.*			eq.*		
35	eq.*			eq.*		
36	overlap			overlap		
37	1.79	0.340	0.84	1.79	0.342	0.80
38	3.63	0.892	0.73	3.63	0.898	0.73
39	2.54#	0.838	0.39	2.47#	0.692	0.38
40	eq.*			eq.*		
41	0.421	0.0625	0.99	0.719	0.240	0.99
42	2.95	0.397	0.83	3.00	0.488	0.83
43	4.37#	1.20	0.65	3.70#	1.14	0.65
44	1.69	0.276	0.90	1.85	0.438	0.90
45	eq.*			eq.*		
46	7.77#	1.68	0.61	7.81#	1.80	0.61

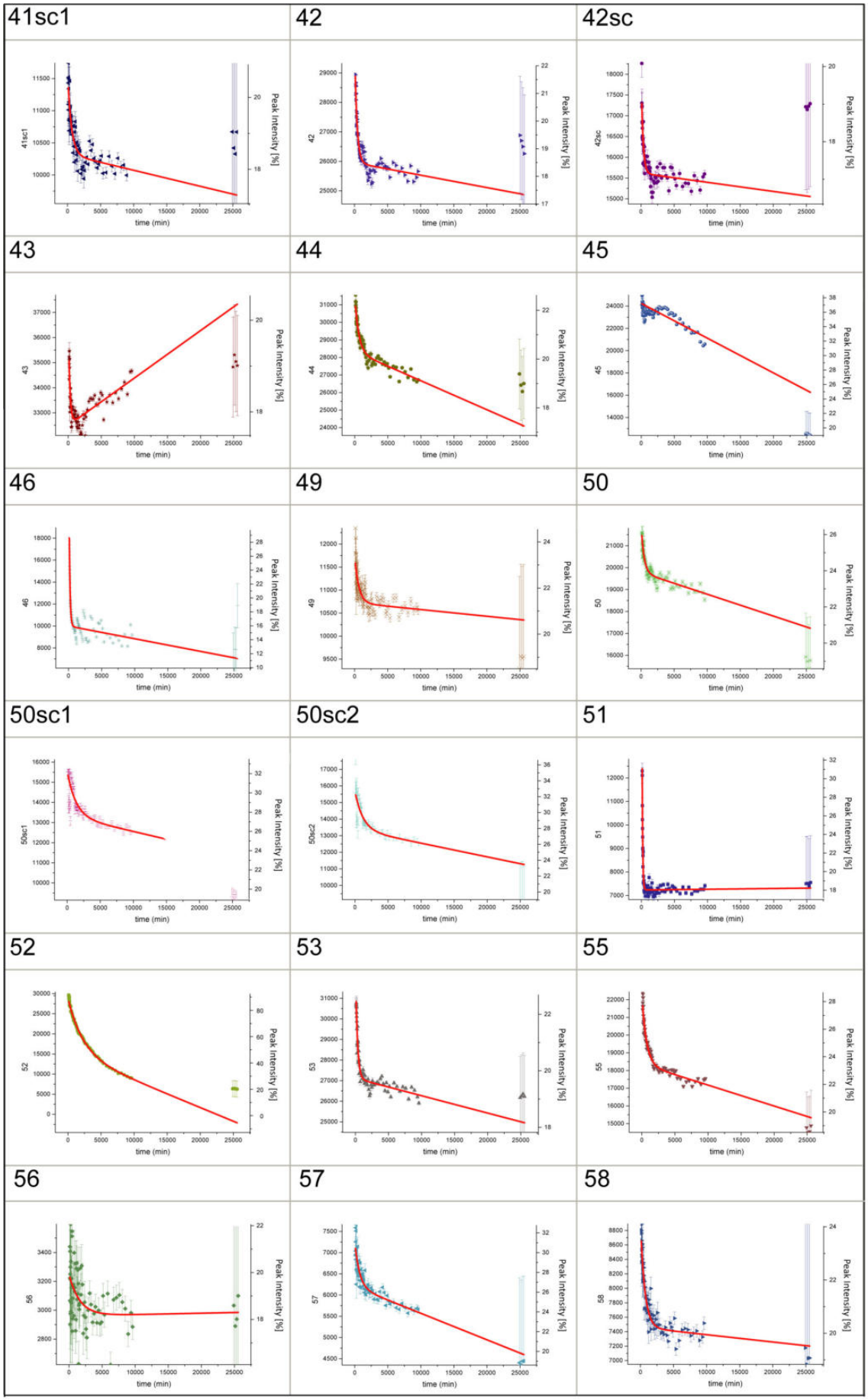
49	1.62#	0.470	0.58	1.62#	0.473	0.57
50	2.34	0.493	0.83	2.34	0.496	0.83
51	11.5	0.955	0.89	11.5	0.937	0.88
52	0.444	0.0321	1.0	0.700	0.0997	1.0
53	3.57	0.461	0.86	3.57	0.465	0.86
55	1.46	0.149	0.94	1.63	0.283	0.94
56	eq.*			eq.*		
57	1.40	0.322	0.85	1.72	6.94	0.85
58	1.52	0.227	0.83	1.59	0.395	0.83
59	overlap			overlap		
60	2.03#	0.524	0.44	3.45#	1.01	0.35
61	1.70#	0.329	0.70	1.76#	0.544	0.70
62	1.42#	0.194	0.58	1.47	0.343	0.84

Table S1: Site-resolved non-equilibrium hydrogen exchange rates expressed in 10^{-3} min^{-1}. The table contains all the exchange rates from unambiguous peaks with an exchange slow enough for non-equilibrium measurements but faster than the rate originating from water accessibility hindrance. *k* (Eq. 1) and *k* (Eq. 2) are fitted using equation 1 and 2 (main text), respectively. Data colored in red and marked with a hash are fitted with an adjusted R^2 value lower than 0.70 and therefore need to be excluded from consideration. These can be considered to represent fast exchange, where equilibrium protonation levels have almost been reached at the first time point. Residues marked with eq.* were already at their equilibrium intensity before the first measured time point and were counted as exchanging too fast for recording. These are depicted as ruby-colored columns in Fig. 3 (main text). Residues marked as “overlap” are ambiguous.

Peak intensities from 76 consecutive 2D HN-spectra




24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
time (min)
Peak Intensity [%]


41sc1
42
42sc
43
44
45
46
49
50
50sc1
50sc2
51
52
53
55
56
57
58
time (min)
Peak Intensity [%]

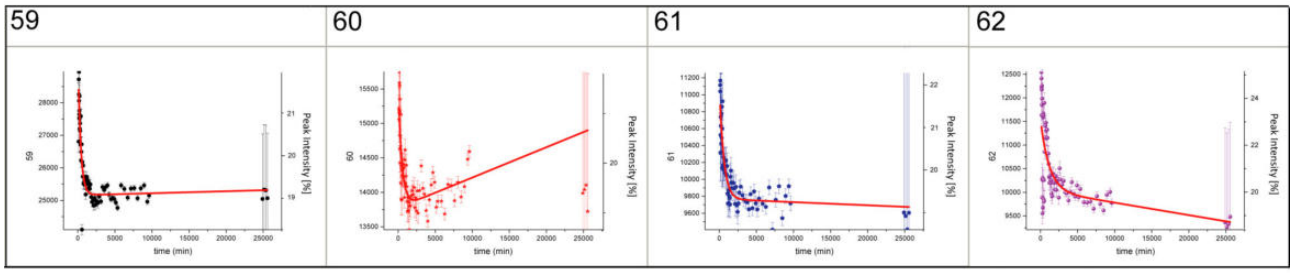


Fig S1: Peak intensities from 76 consecutive 2D H/N spectra measured over two weeks with increasing numbers of scans as described in the main text. Numbers 7 till 62 correspond to residue E7 till D62 of chicken α-spectrin. 16Sc1, 16sc2, 50sc1, 50sc2, 41sc and 42sc are corresponding to side chain amides. All graphs are fitted using equation 2 (main text). Protonation percentage scales on the right are approximate due to unknowns in the conversion from the individual intensities.

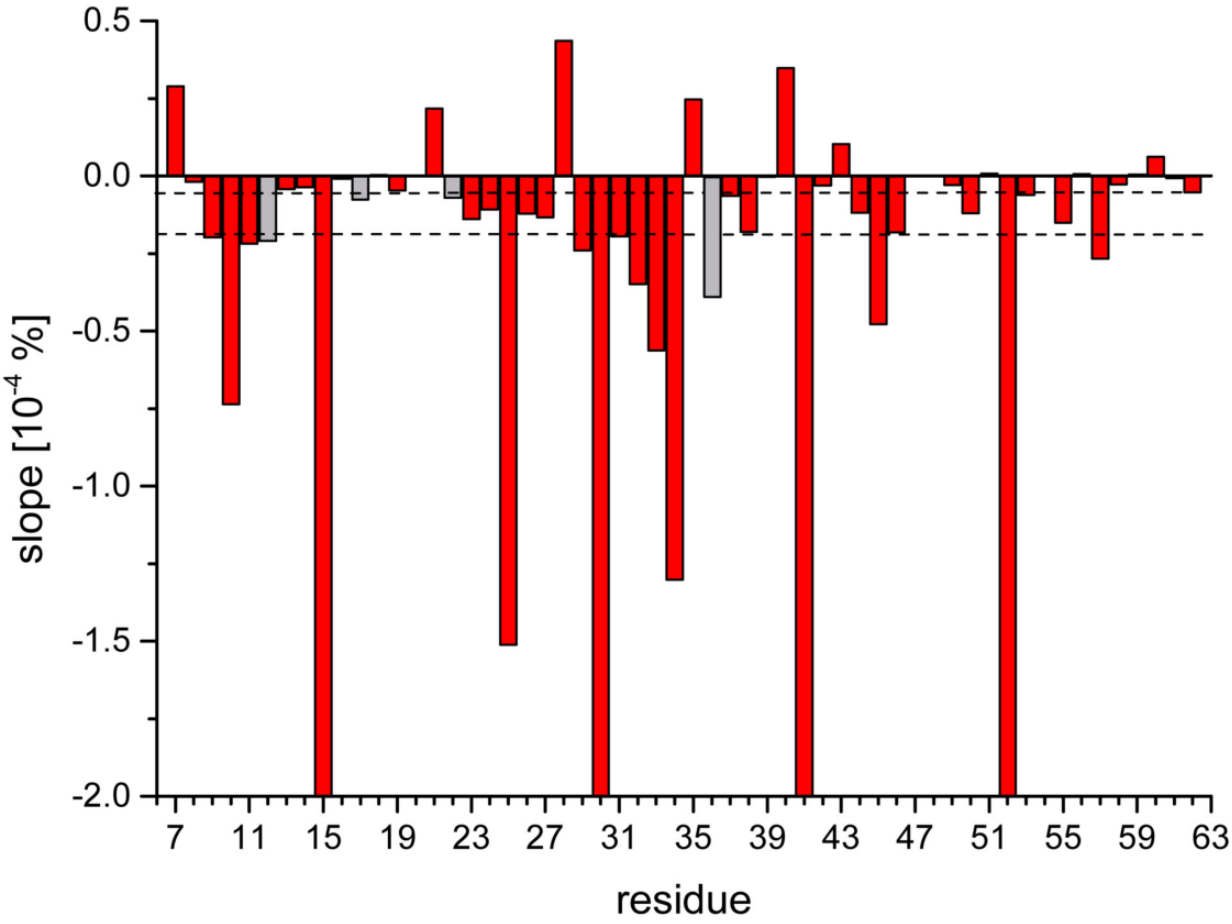
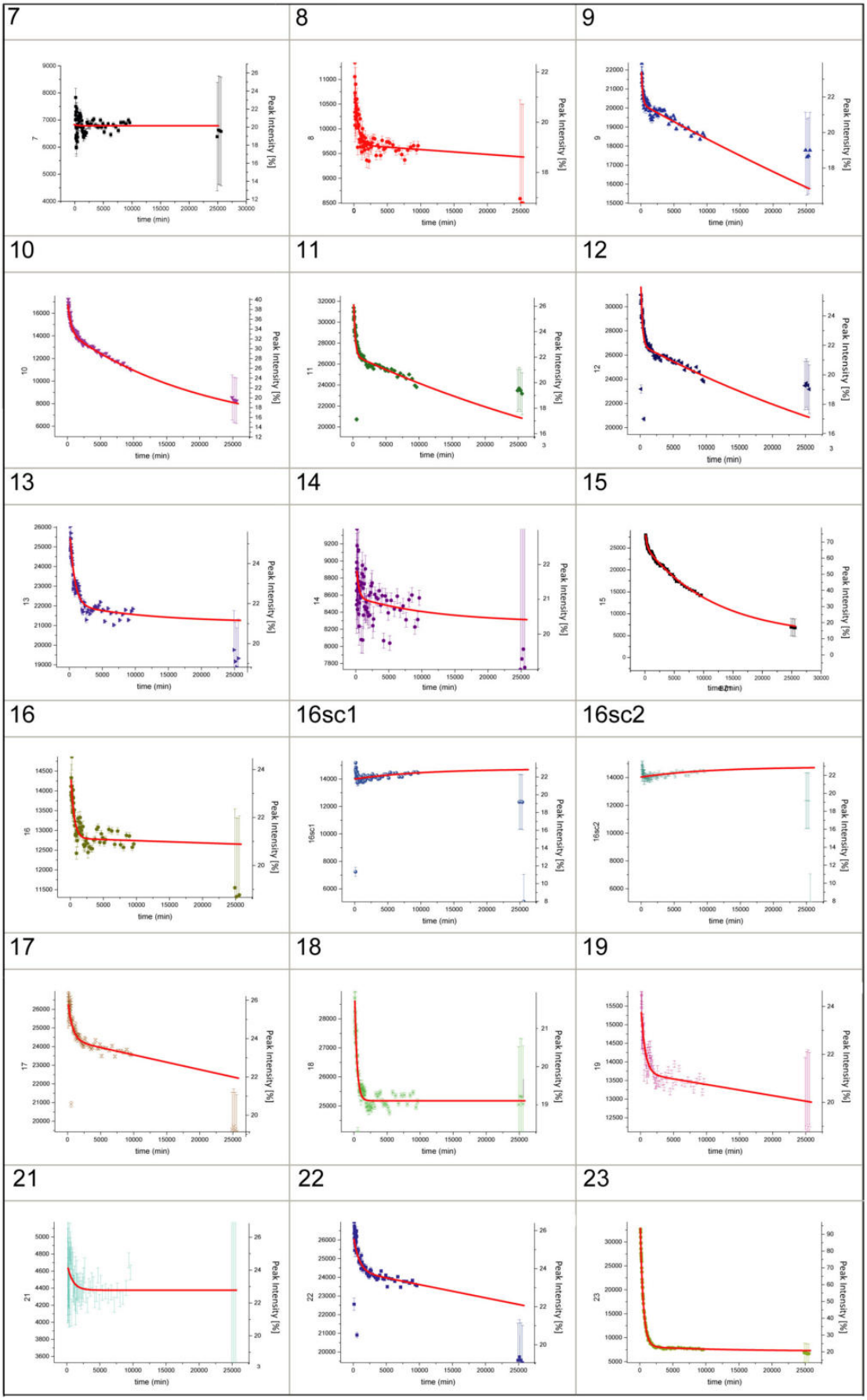
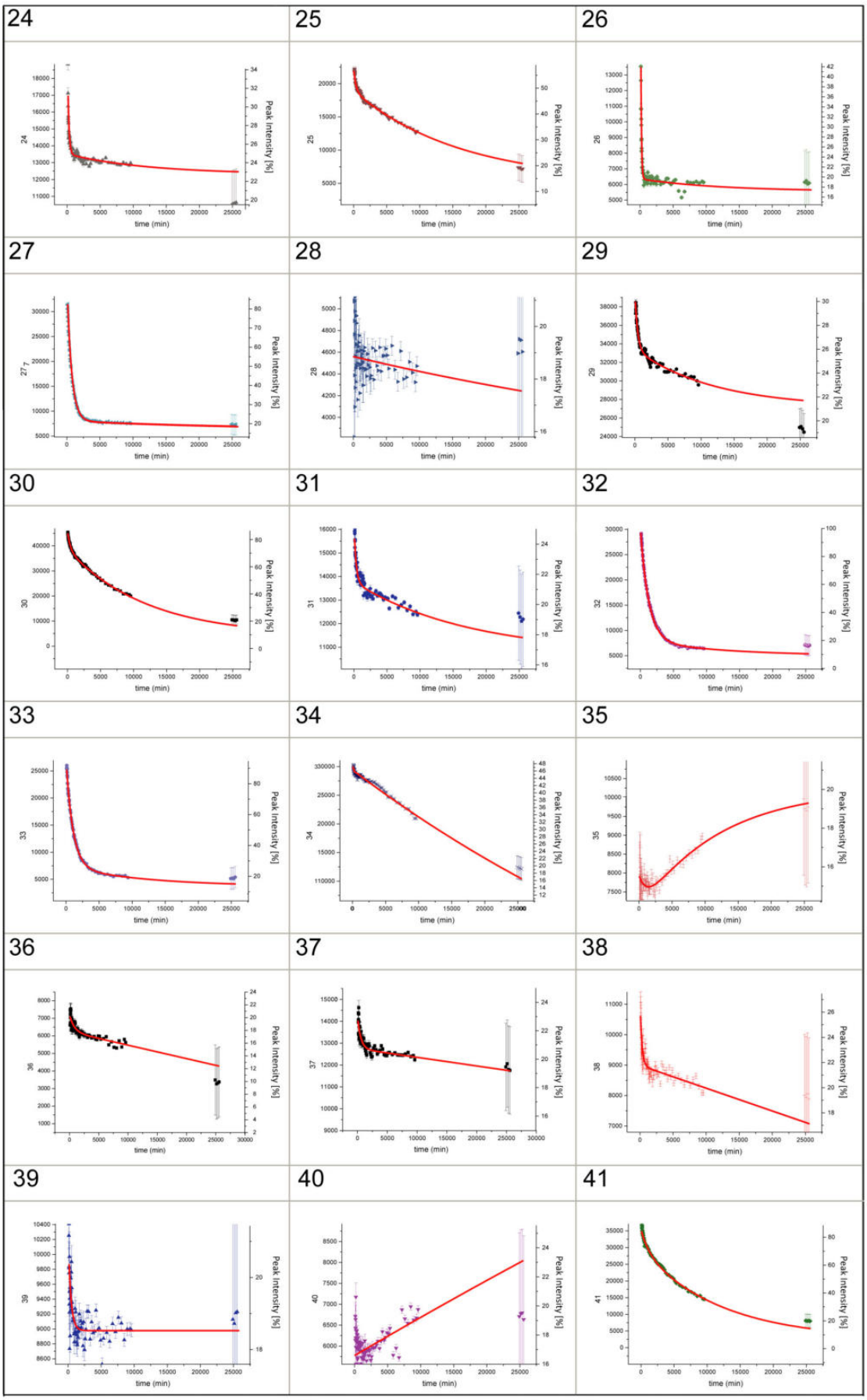
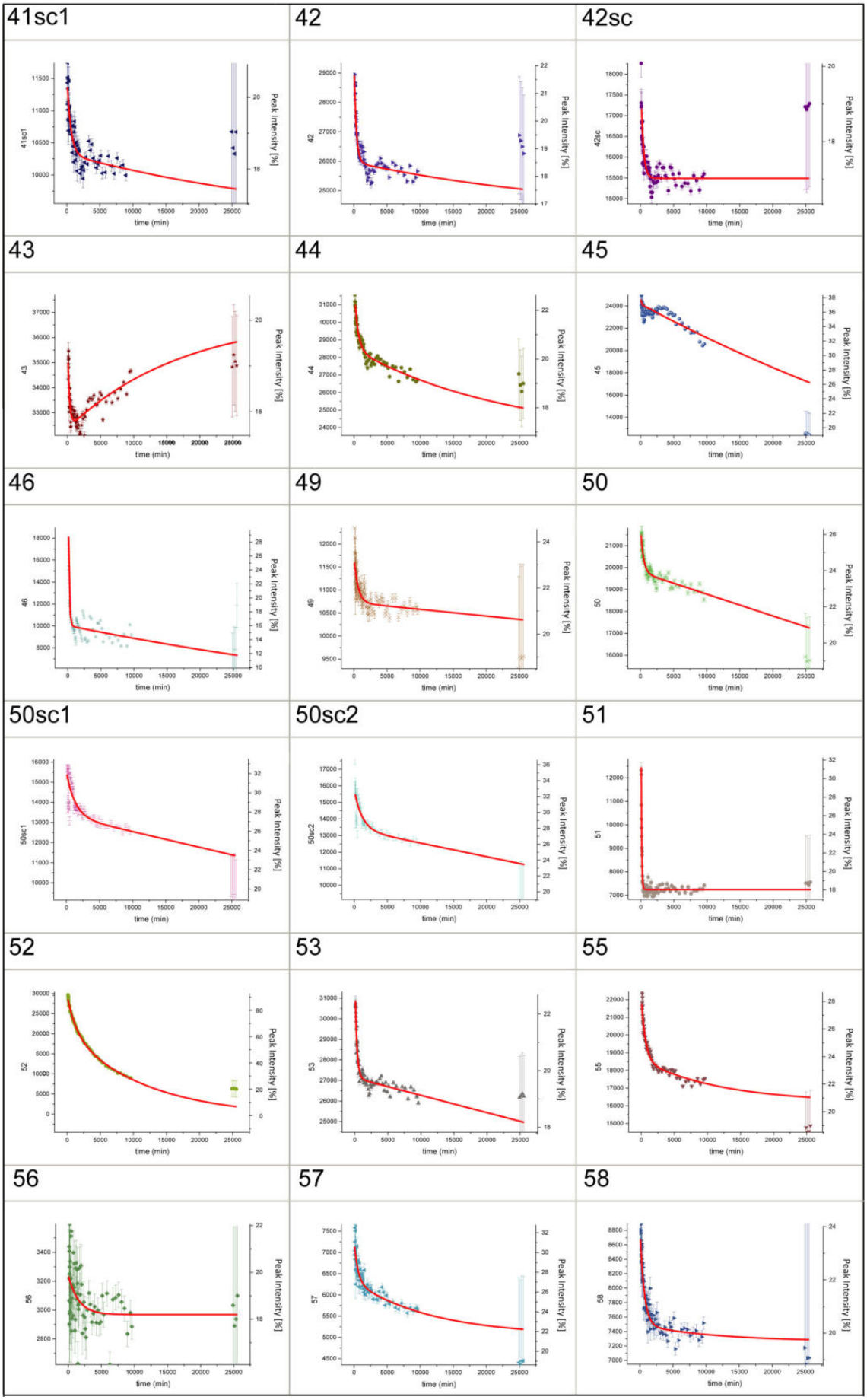


Fig. S2: Slope of the decay curves shown in figure S1, fitted using equation 2 (main text). In order to classify the residues qualitatively as accessibility-hindered or non-accessibility-hindered, two borders of slope D were defined. An upper border of $1.9 \cdot 10^{-5}$ % · min^{-1} in magnitude identifies all decay curves with a slow exchange component, as belonging to diffusion-hindered residues. A lower border of $5.2 \cdot 10^{-6}$ % · min^{-1} in magnitude categorizes decay curves bearing no such slow-decay component. These curves belong to residues that are water-accessible. Some decay curves show a positive slope, this originates mostly from low signal to noise and therefore inaccurate fitting. Another explanation for the existence of positive slopes is the slight increase of the proton concentration in the buffer due to the release of amide protons (as described in the main text).


7
8
9
10
11
12
13
14
15
16
16sc1
16sc2
17
18
19
21
22
23
time (min)
Peak Intensity [%]


24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
time (min)
Peak Intensity [%]


41sc1
42
42sc
43
44
45
46
49
50
50sc1
50sc2
51
52
53
55
56
57
58
time (min)
Peak Intensity [%]

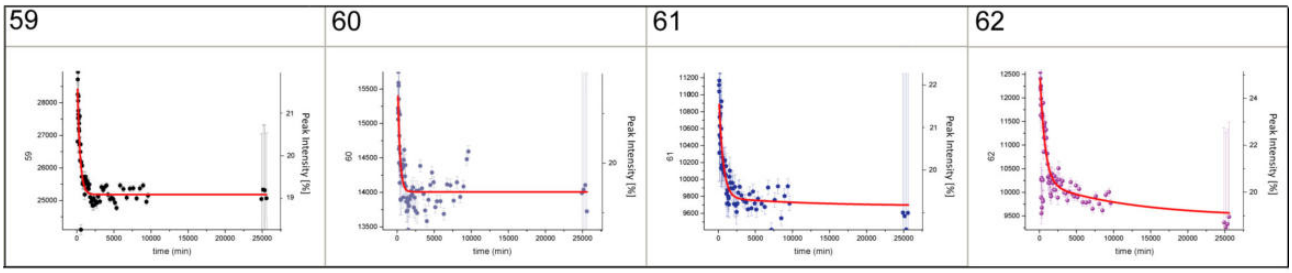


Fig S3: Peak intensities from 76 consecutive 2D H/N spectra measured over two weeks with increasing numbers of scans as described in the main text. Numbers 7 till 62 correspond to residue E7 till D62 of chicken α-spectrin. 16Sc1, 16sc2, 50sc1, 50sc2, 41sc and 42sc correspond to side chain amides. All graphs are fitted using equation 1 (main text). Protonation percentage scales on the right are approximate due to unknowns in the conversion from the individual intensities.

2.2 Application of Solid-State- and Solution-State NMR Spectroscopy for the Benefit of Biochemistry and Structural Biology

2.2.1 Fit as a fiddle: The active site of human carbonic anhydrase II is marked by extensive conformational dynamics

The work of Singh et al. (doi.org/10.1002/anie.202009348) was dedicated to provide data for mechanistic insights on the pharmaceutically important enzyme human carbonic anhydrase II, in physiological conditions. My contribution was the application of the non-equilibrium H/D-exchange method developed in Chapter 2.1.3. Both the resulting data on the presence and strength of H-bonds and water accessibility provided steps toward a better mechanistic understanding of the enzyme's reaction mechanism. Water accessibility is particularly interesting for this protein because the active center cone buries a chain of rigid water molecules that are functionally important. These water molecules prevent the exchange of acidic protein sites with water. As such, even the amide protons around the active cone are close to the protein surface, they are protected from exchange by the tightly bound water. Therefore, methods measuring water accessibility as a function of protein-water distance, e.g. by CP, would provide misleading information. By contrast, the non-equilibrium H/D-exchange was able to proof the presence of the vital water chain under physiological conditions (**Figure 3**, following publication).

As developed in Chapter 2.1.3, the individual analysis of H-bond strength contributed to the determination of a well-defined solution structure of human carbonic anhydrase II, which is deposited in the PDB database (PDB access-code: 6HD2). The importance of some H-bonds stands out. An example is shown in **Figure 22**, where the H-bond between the oxygen of the sidechain 62Asn and the backbone amide proton of the residue 64His is shown. This H-bond stabilized the residue 64His, whose sidechain represents the endpoint of the proton shuttle performed by the rigid water chain (see submitted publication below). **Figure 22 B** shows the two orientations of the histidine sidechain in the crystal structure 2CBA[12], which is believed to be responsible for the proton release to the bulk water. **Figure 22 A** shows the double exponential decay of the amide-peak intensity of 64His. From the decay, a hydrophobic shielding, as well as a strong H-bond is detectable. The crystalline structure confirms this H-bond by the distance of the amide proton to the oxygen of the 62Asn sidechain.

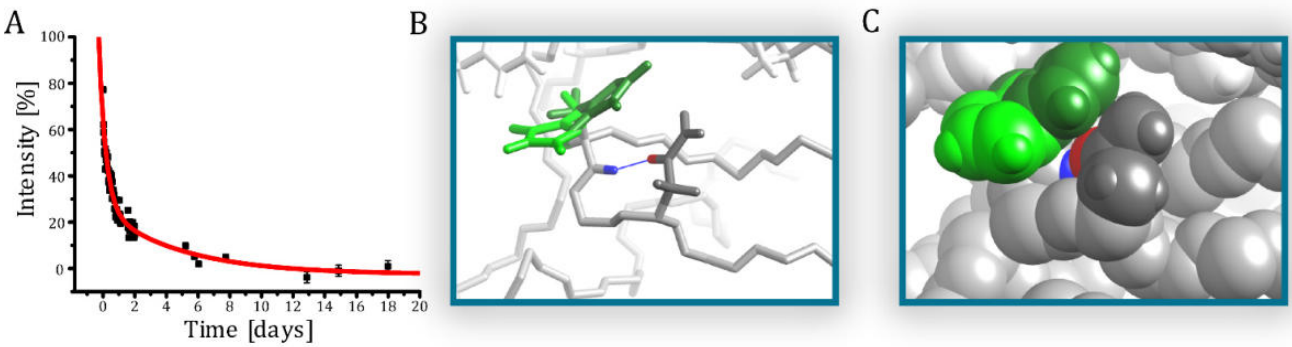


Figure 22: Mechanistically important hydrogen bond between the amide proton of 64His and the sidechain oxygen of 62Asn. **A**: Intensity decay of the amide proton peak after changing the protonation level of the buffer. The double exponential behavior of the curve denotes the presence of a strong H-bond and slight hindrance of water accessibility due to hydrophobic shielding. **B**: The two crystallographic states of the sidechain 64His are shown in dark and light green. It is assumed that these structures represent the motion responsible for proton release. The H-bond, depicted as a blue line, seems to be decisive for stabilizing the directional sidechain motion. **C**: In order to visualize the solvent protection due to shielding with respect to the 64His amide-proton, the nuclei of the loop are shown as shares (amide proton shown in blue)

.

Fit as a fiddle: The active site of human carbonic anhydrase II is marked by extensive conformational dynamics

Himanshu Singh[a, b], Chandan K. Das[c], Suresh K. Vasa[a, b], Kristof Grohe[a, b], Lars V. Schäfer[c], Rasmus Linser[a, b]*

[a] Dr. Himanshu Singh, Dr. S. K. Vasa, Prof. Dr. R. Linser
Faculty for Chemistry and Pharmacy
Ludwig-Maximilians-University Munich
Butenandtstr. 5-13, 81377 Munich, Germany
E-mail: rasmus.linser@tu-dortmund.de
[b] Dr. Himanshu Singh, Dr. S. K. Vasa, Prof. Dr. R. Linser
Faculty of Chemistry and Chemical Biology
Technical University Dortmund
Otto-Hahn-Str. 4a, 44227 Dortmund, Germany
[c] Dr. Chandan K. Das, Prof. Dr. Lars V. Schäfer
Theoretical Chemistry,
Ruhr University Bochum,
Universitätsstr. 150, 44801 Bochum, Germany

Abstract: As drug targets for numerous diseases, carbonic anhydrases are a class of proteins with ubiquitous pharmacological importance. On the basis of hundreds of crystal structures, the accepted mechanistic view has been based on exceptional stiffness of the active-site geometry. Using solution NMR and molecular dynamics (MD) simulations, here we show that in its physiological form in solution the residues crucial for the catalytic reaction comprise conformational-exchange dynamics between open and closed states on the timescale of the enzymatic turnover. The presence of conformational plasticity, which is abolished upon binding of an active-site small-molecule inhibitor, is accompanied by a deviation of the pocket ground-state structure compared to crystallography. The findings, which will be of particular significance for the many medicinal-chemistry studies involving hCAII inhibitors, also show that assessment by NMR spectroscopy and MD is crucial as a complement to crystallography.

Human carbonic anhydrases (hCAs, EC 4.2.1.1) are able to catalyze rapid interconversion between CO_2 and HCO_3^- and thus play an important role in almost all living organisms and tissues.[3] The ubiquitous isoform hCAII belongs to the fastest enzymes known, reaching 10^4 to 10^6 turnovers per second, and has been serving as a model system for examining proton transfers in more complex systems such as the photosynthetic reaction center[3b], bacteriorhodopsin[4], and cytochrome c oxidase[5]. Understanding the molecular mechanism of hCAII catalysis has constituted an intense scientific investigation for decades.[6] Innumerable structural studies, mainly via X-ray[4-5], neutron diffraction,[1, 7] and MD simulations[8] have been pursued, at the same time providing generally applicable tools for elucidation of drug-protein and water-protein interactions. Despite information with an outstanding level of detail, the high catalytic turnover rate as well as the mechanistic details of its catalytic reaction have remained a matter of debate till-date. In particular, motional aspects associated with the reaction and CAII compatibility with substrates other than CO_2, are subjects of current investigation.[9] With a diverse range of crystallography studies (> 750 structures) providing very detailed structural data for native CA and a plethora of inhibitors, apart from the H64 sidechain as a mobile proton

shuttle towards the solvent, the protein has been representing a textbook example for a rigid and perfectly placed active-site geometry and derived water network (Fig. 1A) as the basis for both, catalytic activity und druggability.[6c, 10] Fig. 1B shows B-factors in unliganded hCAII in crystalline form, and Fig. 1C shows a crystallography-derived (i. e., structural-diversity-based) assessment of protein plasticity; from both views the protein active site (arrows) has been appearing absolutely rigid.

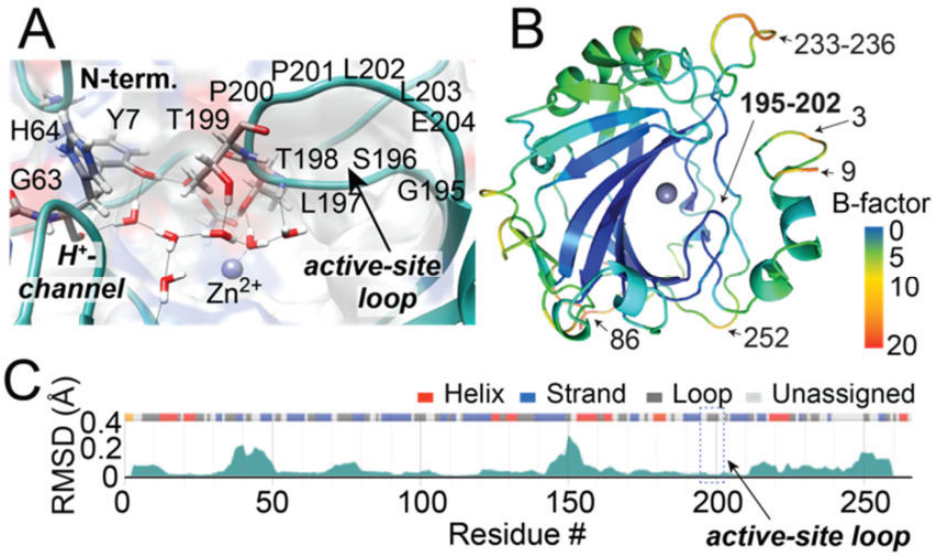


Figure 1: Rigidity of hCAII in crystallographic studies. **A)** Depiction of the conserved active site and water network spanning the active-site loop, N-terminal Y7, and proton shuttle H64, prepared from pdb 4Y0J.[1] **B)** B-factors in pdb 2CBA, color-coded on the protein structure. **C)** Structural conservation expressed as residue-specific RMSD between the available X-ray structures with > 95% sequence identity. Graphic obtained from the PDBflex server.[2]

Catalysis and selective interactions between biomolecules can be leveraged and regulated by conformational-exchange dynamics of the involved residues. Whereas such dynamics on the µs-ms timescale are difficult to spot in crystallography and under cryogenic and relaxation dispersion techniques under near-physiological conditions as well as residual dipolar couplings (RDCs) are able to report on site-specific µs-ms conformational exchange processes.[11] For hCAII, dynamics studies by NMR were long elusive after early work by Venters and coworkers yielded assignments that were incomplete for the active site.[12] However, near-complete backbone and sidechain carbon assignments of the 260-amino-acid protein and its active site could be obtained recently using solid-state NMR.[13] Using additional backbone assignment experiments on deuterated ^{13}C, ^{15}N-labeled hCAII, as well as 3D HcCH- and hCCH-TOCSY experiments on protonated hCAII, in solution, we were able here to obtain complete resonance assignments of hCAII in its physiological form. (See spectroscopic and preparative details in the SI; chemical shift assignments are listed in the BMRB under accession code 34308.) Even though backbone and sidechain shifts are similar between the crystalline form and the monomeric solution form for most residues, we find significant differences for a subset of residues. Fig. 2A shows a solution NMR H/N-HSQC, overlaid with a proton-detected solid-state H/N correlation. Fig. S3

shows shift differences as a function of residue, as well as a correlation between monomer and crystal shifts, and deviating residues ($\Delta\delta$> 0.25 ppm) highlighted on the crystal structure (pdb 2CBA). T198 is visible in solids but is exchange-broadened in solution at the same temperature. Fig. S2 shows that temperature-induced differences could be excluded for this assessment.

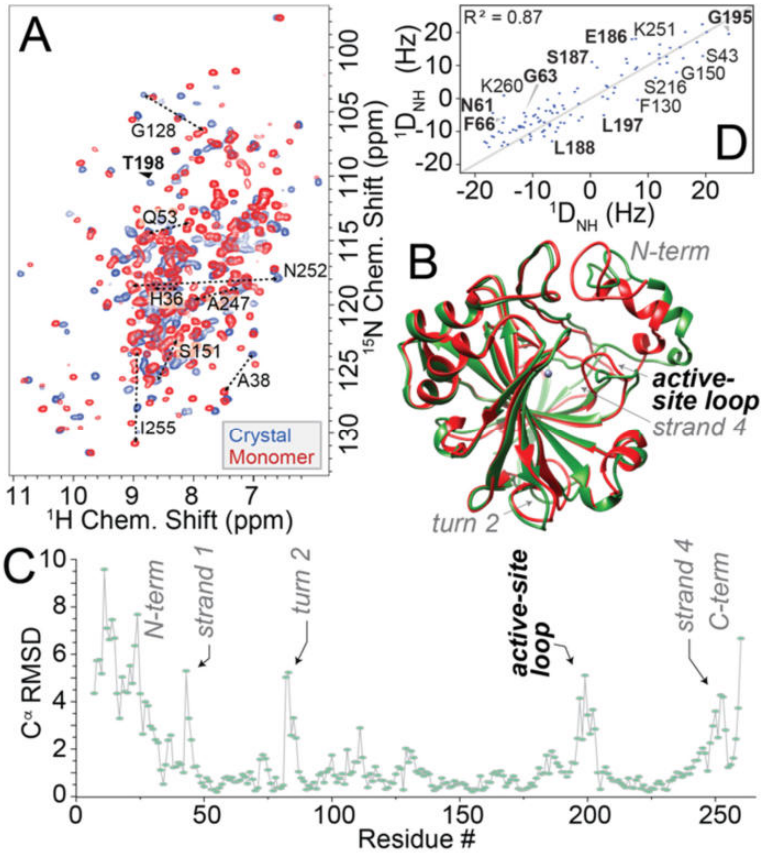


Figure 2: Comparison of hCAII chemical shifts under crystalline and monomeric conditions and deviation between monomeric structure in solution and crystallographic structures. **A)** ^{1}H/^{15}N correlation spectra of monomer (red) and crystal (blue) at 25 °C, also compare Figs. S3-5. Several peaks shift, T198 is exchange-broadened in the solution NMR spectrum. Solid-state spectra recorded on the same spectrometer on a protonated, uniformly ^{13}C, ^{15}N-labeled microcrystalline sample of hCAII at 111 kHz magic-angle spinning and similar temperature. **B)** Superposition of the structure in the crystalline state (PDB 2CBA, depicted in red) with the minimum-energy solution structure (green). Strongly deviating regions are denoted. **C)** C$^{\alpha}$ RMSD between solution structure and crystalline form as a function of residue. Active-site residues are annotated in bold. **D)** Correlation of experimentally measured RDCs in solution with back-calculated RDCs based on the static X-ray structure (pdb 2CBA), showing a comparably poor correlation. (The most strongly deviating residues are annotated.)

Based on the differences between the aggregation states, we wondered whether the crystallographic findings deviated from the enzyme structure in solution. For this reason, we first determined the solution NMR structure, using a combination of ^{13}C- (aliphatic and aromatic) and ^{15}N-edited 3D NOESY experiments for ^{1}H-^{1}H distance restraints in addition to $^{1}H/^{15}N$ residual dipolar couplings (RDCs) induced by partial alignment in Pf1 filamentous phages, as well as H-bond restraints from H/D-exchange and chemical-shift-based dihedral-angle restraints, using the program ARIA 2.3[14]. With the total number of constraints per residue amounting to ~17 (compare Figs. S6-8 and Tables S1-3), a total of 500 structures were calculated, from which an ensemble of 20 structures (shown in Fig. S8A) were selected for structure analysis. A correlation of 0.97 between back-calculated and experimental RDCs (see Fig. S8B) indicates good agreement between observed data and calculated structure. The atomic coordinates of hCAII were deposited in the PDB (pdb id: *6HD2*). Even though expectedly, the structural organization observed in the three-dimensional fold of hCAII in solution is overall very similar to the structure determined by crystallography (see Fig. 2B), with an RMSD (NMR vs. X-ray) of each secondary structural element of around 0.2 to 0.7 Å, differences with respect to the crystallographic structure (see Figs. 2B and C) are found for residues in regions around residue 40 (coined "strand 1", a disordered region with high B-factors above 25), around residue 86 ("turn 2", B-factors around 20), and around residue 255 ("strand 4"). All these residues exhibit chemical-shift perturbations compared to the crystalline state and show obtrusive behavior also in the dynamics experiments (see below). It is thus likely that for these residues variations in the structure can be adopted, allowing for a spatially restricted positioning in the crystal lattice. The N-terminus unfortunately is poorly defined due to a low number of distance restraints. (See residue-resolved precision in Fig. S7.) Most importantly, for the active-site loop around the door keeper residue T198, with reasonable structural precision, the RMSD with respect to the X-ray structure is high (see Figs. 2B and C). The differential placement of the active-site loop, from which the conserved H-bond network emanates, in the solution structure has a substantial impact on the average active-site geometry (compare Fig. S9). To verify the above structural differences between crystalline and monomeric solution state, we also subjected the static X-ray structure (2CBA) to the program PALES[15], generating back-calculated RDC values as they would be expected if the H-N bond vector orientations of the crystalline structure were representative for those under solution conditions (Fig. 2D). Indeed, the correlation coefficient with regard to the experimentally measured RDCs in solution is only 0.87, with residues in the active site (e. g., G63, G195, L197) being among the most deviating ones, suggesting a differential, slightly more open average structure under monomeric conditions in the absence of a crystal lattice.

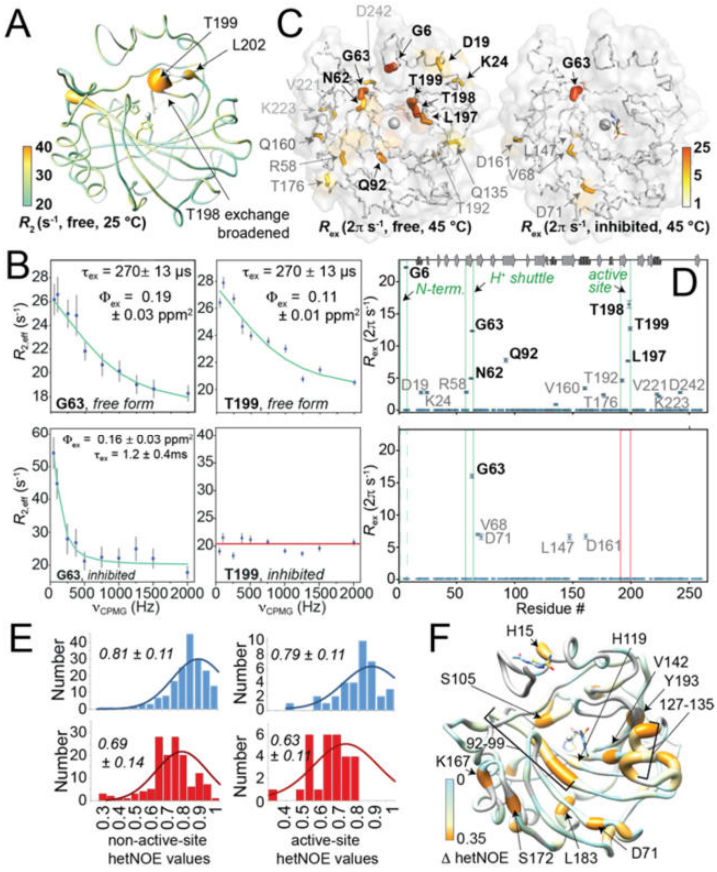


Figure 3: hCAII dynamics. **A)** R_2 rates exceeding 20 Hz (cyan, as the value expected for a globular 29 kDa protein in the absence of conformational exchange). **B)** Relaxation dispersion profiles from a collective fit of the outer pocket close to H64 and active-site loop in the absence (upper row) and presence of inhibitor (lower row). **C)** Exchange contribution from RD in the absence (left) and presence of inhibitor (right) depicted on the structure. **D)** Residues with significant exchange contributions in the absence (upper row) and presence (lower row) of inhibitor as a function of sequence. G6 also shows dispersion with inhibitor, which however has too low signal-to-noise ratio for fitting and therefore is not shown in the plot. Data in C) and D) represent individual fits. **E)** Shift of hetNOE distributions (histograms and Gaussians fits as well as raw-data mean and standard deviation) for non-active-site (left) and active-site residues (right) of ligand-bound (red) and non-liganded protein (blue). **F)** Site-specific decrease in hetNOE of liganded with respect to free form.

Elucidating the nature of the active-site water network by solid-state NMR, we have previously reported the existence of subtle relaxation dispersion effects around the water network for crystalline hCAII.[16] Whereas all residues of the active-site pocket were visible in solid-state NMR experiments at room temperature, under physiological conditions in solution, pronounced dynamics appear in the active site, and in particular the catalytically important residue T198 is completely exchange-broadened at 25 °C. Hence, dynamics appear under native conditions that are far less restricted than in the presence of a crystal lattice. Slightly increasing the temperature to 37 °C and 45 °C, we

were now able here to undertake a detailed assessment of the physiological active-site dynamics, including ^{15}N longitudinal and transverse relaxation rates R_1 and R_2 and the [^{15}N, ^{1}H]-heteronuclear Overhauser effect (hetNOE), as well as constant-time Carr-Purcell-Meiboom-Gill[17] relaxation dispersion experiments capturing site-specific µs-ms timescale motion. (See all relaxation data in Figs. S10-14 and S16-18.) Most interestingly, slow motion on the 100 µs timescale for the unliganded protein with elevated R_2 rates up to 37 s^{-1} (see Figs. 3A and S10) and strong relaxation dispersion (Figs. 3B-D and S11-12) were found locally for residues in the active-site loop. Whereas the backbone conformational exchange we observe on the same timescale for the residues around H64, thought to exert proton shuttling via its sidechain rotation, could be reconciled with the established mechanistic picture, we find the strongest dispersion and highest R_2 rates at the very bottom of the active site (Fig. 3B-D). The conformational exchange can be fitted individually (Fig. 3C/D) or collectively over the active site (Fig. S11) and does not only involve T198, but the whole active-site loop from S196 to E204 (see Figs. S11 and S12 for the dispersion curves and peak shapes at 45 °C and 37 °C, respectively). Fitting the dispersion profiles within the loop as a global fit yields an exchange lifetime of 270 µs, for which exchange contributions R_{ex} are depicted by differential coloring in Fig. 3C. Interestingly, residues G63 (H64 is unfortunately overlapped) and G6 at the very N-terminus, which are coupled to the active-site loop through the H-bond network, show strong dispersion on the same timescale and could be included in the global fit (Figs. 3B/D and S14).

In order to mechanistically assess the nature of the active-site conformational exchange in more detail, we carried out molecular dynamics (MD) simulations in Gromacs[18]. (See details in the SI.) Dynamics on the 100 µs timescale are very challenging to capture in MD, and a full assessment of the motion on that timescale was thus not possible. The active site loop is packed against and tightly interacts with the N-terminal part of hCAII via adjacent hydrophobic surfaces. Whereas these contacts would not abolish loop motion, they are expected to slow down the dynamics of the active site loop via coupling to the N-terminus. Previously, N-terminal truncation of hCAII up to residue 24 was shown to largely retain catalytic activity, with a penalty of around 1 kcal/mol on the activation energy for catalysis and a remaining 10^5 turnovers per second.[19] In addition to the intended acceleration of the active-site loop dynamics, coupling between the N-terminus and the active site, which is also evident from our relaxation dispersion data above, will naturally remain elusive in MD simulations using such an N-terminally truncated protein. Nevertheless, these simulations can still grasp the different intrinsic interactions and plasticity of the active site of apo vs. inhibitor-bound hCAII and unravel the mechanics underlying the experimental observations. Indeed, without the increased barrier due to the N-terminal interactions, the dynamics of the active-site loop is witnessed in MD within a 500 ns timescale. For hCAII in the absence of an inhibitor, the loop was found to easily and reversibly unlatch its H-bonds to the Zn-bound hydroxide and E106 sidechain to detach from the catalytic center in a collective hinge motion, the tip of the loop around T198 showing the largest displacement. Fig. 4A-C shows MD simulations of the

unliganded protein as well as with the inhibitor dorzolamide, in which this pronounced plasticity is not observed (see below).

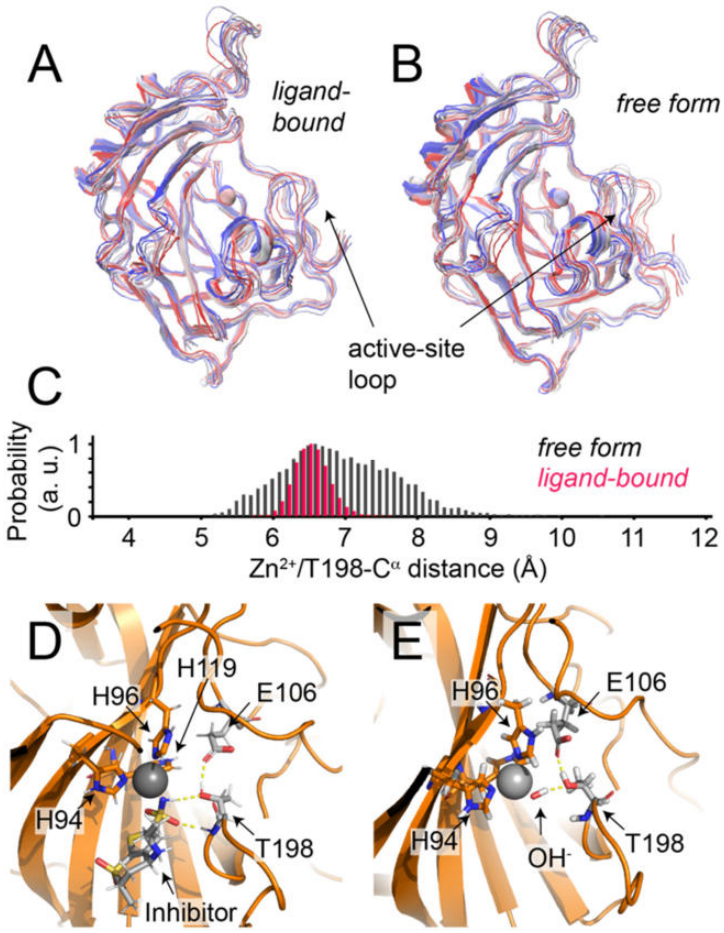


Figure 4: Assessment of loop interactions via MD simulation of N-terminally truncated hCAII. **A)** In the ligand-bound form, the active-site loop is locked in its closed position. **B)** Without inhibitor, the loop shows pronounced open/close dynamics. **C)** Histogram of Zn-T198 distance in the presence (red) and absence (black) of inhibitor. **D)** Stabilization of the closed loop conformation via multiple H-bonds to the inhibitor. **E)** Zn-OH^--based H-bonds of the active-site loop in the apo state.

Sulfonamide inhibitors of carbonic anhydrases are known to replace the Zn-bound catalytic water molecule. Hence, they also impair the conserved water network, one of the peculiar features of the CA core.[1, 20] Changes in active-site B-factors induced by anti-glaucoma drugs like acetazolamide or dorzolamide are, however, insignificant in the crystal, and structures are virtually identical to the apo form (see Fig. S15).[21] Given the above results, we wondered what impact binding of such substrate analogues would have on the monomeric protein in solution. In fact, as shown in Fig. 3B-D, in the inhibitor-bound state the dynamics of the protein are substantially altered: Now, in fact, motion of the active-site loop ceases. (Also compare Fig. S16 and S17.) This is observed consistently in the experimental data as well as in the MD simulations of the N-terminally truncated protein (Fig. 4A/C). By contrast, the N-terminal residue G6 still shows unambiguous

dispersion with the ligand bound, which hints to spontaneous conformational exchange of the N-terminus irrespective of the active-site situation. Unfortunately, due to exchange broadening, *quantitative* fitting for G6 is compromised. Also G63 conformational exchange is not quenched, however, it does become slowed down to the ms regime in the presence of the inhibitor. In addition, changes regarding ps-ns timescale motion upon ligand binding can be deduced from R_1 and hetNOE data. In the uninhibited case, fast-timescale motion is present at the tips of some external loops, like residues 40 and 86, which are associated with structural deviations to the crystalline state (see above). hetNOE values in the absence of inhibitor scatter around 0.8 for the whole sequence, again showing slightly elevated fast-timescale motional (low hetNOE and high R_1 values) for the loop around T198. (hetNOE values are also part of Fig. S10 and S16.) By contrast, upon inhibition, hetNOE values are decreased for residues in parts of the primary sequence. (Statistics and differences, liganded vs. unliganded form, are shown in Fig. 3E and plotted on the protein structure in Fig. 3F, respectively. All statistics are shown in Fig. S17.) S18 also shows the effect of CO_2 binding to the active site as the natural substrate. Its affinity is approximately 100x lower than the inhibitor, such that these data only show subtle effects. The trends, however, seem to be in line with the sulfonamide as a high-affinity (covalent) substrate analogue.

Our various observations unambiguously demonstrate the existence of conformational-exchange dynamics in the active site in hCAII under physiological conditions, which contradict previous, X-ray crystallography-based studies[21-22]. The presence of strong, spontaneous conformational exchange in the active site of CAs challenges the mechanistic model of a highly rigid active-site reaction chamber. Whereas the loop remains in the closed position known from the X-ray structures, when a high-affinity inhibitor providing multiple H-bonds is bound in the active site, the weak H-bonding between T198 and the Zn-bound catalytic water/OH^- in the apo form (visualized in Fig. 4D/E) is easily opened at room temperature in the absence of a crystal lattice, such that the active-site loop undergoes pronounced open/close dynamics. The timescale of the active-site loop motion is further modulated by the hydrophobic interactions with the N-terminus, which couples the dynamics of these two structural regions. As such, the protein has the possibility to dynamically adjust the active site towards a suitable geometry for different steps of the catalysis by varying the active site loop around T198. The conformational-exchange timescales are comparable to the catalytic turnover rate of the enzyme, underlining its possible relevance for biological activity. The sampling of open and closed conformations in the absence of a substrate may be advantageous for substrate intake and also adds to understanding hCAII substrate flexibility[9b]. Conversely, the results for the substrate analogue inhibitor and trends for bicarbonate as a substrate suggest that such conformational changes of this region may be switched off in the event of substrate binding, where the increased fast-timescale motion observed might be beneficial for conversion and product release instead. The conversion of local slow-timescale motion into fast-timescale fluctuations in large parts of the protein upon binding of an active-site inhibitor also suggests coupling of active-site plasticity with the overall protein architecture. Such coupled motions have previously been observed in other proteins,[23]

adding a new nuance to the assumed simple scaffolding function of the global architecture for active-site catalysis.

Here, we have demonstrated important differences for properties in the active site of hCAII, a drug target in various pathological contexts and unmatched model system in drug discovery research, under biologically representative conditions in solution compared to previous crystallography-based studies. In particular, using NMR relaxation, relaxation dispersion, RDCs, and MD simulations, we have demonstrated that μs timescale conformational exchange between open and closed forms is clearly observed for the important active-site loop under physiological conditions. Active-site plasticity and its modulation by pocket occupancy can be expected to be of major importance for future drug design for this and related targets.

Acknowledgements

Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the context of SFB 749, TP A13, SFB 1309, TP 03 (project number 325871075), and the Emmy Noether program. This work was funded under Germany's Excellence Strategy – EXC 2033 – 390677874 and EXC 114 – 24286268.

Keywords: carbonic anhydrase II • protein structure • conformational exchange dynamics • drug discovery • solution-state NMR

[1] Z. Fisher, A. Y. Kovalevsky, M. Mustyakimov, D. N. Silverman, R. McKenna, P. Langan, *Biochemistry* **2011**, *50*, 9421-9423.

[2] T. Hrabe, Z. Li, M. Sedova, P. Rotkiewicz, L. Jaroszewski, A. Godzik, *Nucleic Acids Res.* **2016**, *44*, D423-D428.

[3] a) S. Lindskog, *Pharmacology & therapeutics* **1997**, *74*, 1-20; b) S. C. Frost, R. McKenna, *Carbonic anhydrase: mechanism, regulation, links to disease, and industrial applications, Vol. 75*, Springer Science & Business Media, **2013**.

[4] C. U. Kim, H. Song, B. S. Avvaru, S. M. Gruner, S. Park, R. McKenna, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 5257-5262.

[5] a) A. Liljas, K. Kannan, P.-C. Bergsten, I. Waara, K. Fridborg, B. Strandberg, U. Carlbom, L. Järup, S. Lövgren, M. Petef, *Nature New Biology* **1972**, *235*, 131; b) M. I. Hassan, B. Shajee, A. Waheed, F. Ahmad, W. S. Sly, *Bioorganic & medicinal chemistry* **2013**, *21*, 1570-1582.

[6] a) Y. J. Zheng, K. M. Merz Jr., *J. Am. Chem. Soc.* **1992**, *114*, 10498-10507; b) R. Michalczyk, C. J. Unkefer, J.-P. Bacik, T. E. Schrader, A. Ostermann, A. Y. Kovalevsky, R. McKenna, S. Z. Fisher, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 5673-5678; c) A. Liljas, *IUCrJ* **2018**, *5*, 4-5.

[7] a) S. Z. Fisher, M. Aggarwal, A. Y. Kovalevsky, D. N. Silverman, R. McKenna, *J. Am. Chem. Soc.* **2012**, *134*, 14726−14729; b) N. Niimura, M. Takimoto-Kamimura, I. Tanaka, *Encyclopedia of Analytical Chemistry* **2016**.

[8] a) S. Toba, G. Colombo, K. M. Merz, *J. Am. Chem. Soc.* **1999**, *121*, 2290-2302; b) S. Taraphder, C. M. Maupin, J. M. Swanson, G. A. Voth, *J. Phys. Chem. B* **2016**, *120*, 8389-8404.

[9] a) S. Paul, S. Taraphder, *J. Phys. Chem. B* **2015**, *119*, 11403–11415; b) J. T. Andring, C. U. Kim, R. McKenna, *IUCrJ* **2020**, *7*, 287-293.

[10] a) G. Klebe, in *Small Molecule — Protein Interactions* (Ed.: M. K. Herbert Waldmann), Springer-Verlag Berlin Heidelberg, **2003**; b) P. W. Snyder, J. Mecinović, D. T. Moustakas, S. W. Thomas, M. Harder, E. T. Mack, M. R. Lockett, A. Héroux, W. Sherman, G. M. Whitesides, *Proc. Natl. Acad. Sci. U.S.A.* **2011**.

[11] a) J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, *Nat. Struct. Mol. Biol.* **1997**, *4*, 292–297; b) D. D. Boehr, H. J. Dyson, P. E. Wright, *Chem. Rev.* **2006**, *106*, 3055-3079; c) H. van den Bedem, J. S. Fraser, *Nat. Methods* **2015**, *12*, 307-318; d) P. Rovó, R. Linser, *ChemPhysChem* **2018**, *19*, 34 – 39.

[12] R. A. Venters, B. T. Farmer, C. A. Fierke, D. S. Leonard *J. Mol. Biol.* **1996**, *264*, 1101–1116.

[13] S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J. Phys. Chem. Lett.* **2018**, *9*, 1307-1311.

[14] B. Bardiaux, T. Malliavin, M. Nilges, in *Protein NMR techniques*, Springer, **2012**, pp. 453-483.

[15] M. Zweckstetter, *Nat. Protoc.* **2008**, *3*, 679-690.

[16] H. Singh, S. K. Vasa, H. Jangra, P. Rovó, C. Päslack, C. K. Das, H. Zipse, L. V. Schäfer, R. Linser, *J. Am. Chem. Soc.* **2019**, *141*, 19276-19288.

[17] S. Meiboom, D. Gill, *Rev. Sci. Inst.* **1958**, *29*, 688-691.

[18] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19-25.

[19] G. Aronsson, L.-G. Mrtensson, U. Carlsson, B.-H. Jonsson, *Biochemistry* **1995**, *34*, 2153-2162.

[20] S. Z. Fisher, A. Y. Kovalevsky, J. F. Domsic, M. Mustyakimov, R. McKenna, D. N. Silverman, P. A. Langan, *Biochemistry* **2009**, *49*, 415-421.

[21] K. H. Sippel, A. H. Robbins, J. Domsic, C. Genis, M. Agbandje-McKenna, R. McKenna, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2009**, *65*, 992–995.

[22] K. Hakansson, M. Carlsson, L. A. Svensson, A. Liljas, *J. Mol. Biol.* **1992**, *227*, 1192.

[23] a) G. Bouvignies, P. Bernadó, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler, M. Blackledge, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885–13890; b) C. A. Smith, D. Ban, S. Pratihar, K. Giller, M. Paulat, S. Becker, C. Griesinger, D. Lee, B. L. de Groot, *Proc. Natl Acad. Sci. U.S.A.* **2016**, *doi/10.1073/pnas.1519609113*.

Supporting information

for the manuscript

Fit as a fiddle: The active site of human carbonic anhydrase II is marked by extensive conformational dynamics

Himanshu Singh[a, b], Chandan K. Das[c], Suresh K. Vasa[a], Kristof Grohe[a, b], Lars V. Schäfer[c], Rasmus Linser[a, b]*

[a]Department Chemistry Ludwig-Maximilians-University Munich Butenandtstr. 5-13, 81377 Munich, Germany

[b]Faculty of Chemistry and Chemical Biology, Technical University Dortmund, Otto-Hahn-Straße 4a, 44227 Dortmund, Germany

[c]Theoretical Chemistry, Ruhr University Bochum, Universitätsstr. 150, 44801 Bochum, Germany

*To whom correspondence should be addressed: Prof. Dr. Rasmus Linser, Faculty of Chemistry and Chemical Biology, Technical University Dortmund, Otto-Hahn-Straße 4a, 44227 Dortmund, Germany

Tel.: +49 (0) 231 755-3910; Email: rasmus.linser@tu-dortmund.de

Table of Contents

Materials and Methods

Overexpression and purification of hCAII: The cDNA encoding *hCAII* was cloned into a pGEX expression vector and transformed into *E. coli* strain BL21(DE3) codon plus cells for over-expression, isolation and purification of the protein as described earlier.[1] A representative size-exclusion chromatography profile of purified (protonated $^{13}C/^{15}N$) hCAII is shown in Fig. S1.

NMR Spectroscopy: For NMR studies, uniformly ^{15}N-labelled (u-^{15}N), $^{13}C/^{15}N$-doubly-labelled (u-$^{13}C/^{15}N$), and uniformly $^{2}H/^{13}C/^{15}N$-triple-labeled (u-$^{2}H/^{13}C/^{15}N$) hCAII were prepared in a mixed solvent of 90% H_2O and 10% $^{2}H_2O$ (50 mM sodium phosphate, 50 mM NaCl (pH = 7.4)). All triple resonance NMR experiments were carried out at 25 ºC with protein concentrations between 0.5 to 0.6 mM on a Bruker Avance 800 MHz NMR spectrometer equipped with a 5 mm cryogenically cooled triple-resonance probe and a pulsed-field gradient. A suite of 3D double- and triple-resonance NMR experiments were performed for sequence-specific ^{1}H, ^{13}C and ^{15}N backbone resonance assignments as discussed earlier.[2] In addition, we recorded 3D experiments such as HCCH-TOCSY (mixing time = 24 ms), [^{15}N, ^{1}H]-NOESY-HSQC (mixing time = 100 ms) and [^{13}C, ^{1}H]-NOESY-HSQC (mixing time = 80 ms) for almost complete assignment of ^{1}H, ^{13}C and ^{15}N side-chain resonances and for the determination of NOE distance constrains used in the 3D structure calculation of the enzyme. $^{1}H/^{15}N$-RDCs were collected using Pf1 filamentous phage (purchased from Asla Biotech) with a concentration of 15 mg/mL as an alignment medium, resulting in $^{1}H^{N}$-$^{15}N^{H}$ residual dipolar couplings between -15 and +18 Hz. The RDCs were determined via IPAP HSQC experiments.[3] Alignments of the hCAII sample was confirmed by a D_2O quadrupole splitting of 14 Hz. PALES[4] was used to calculate the alignment tensor, resulting in a tensor magnitude of 13.91 Hz and a rhombicity of 0.252. The complete backbone and sidechain ^{1}H, ^{13}C and ^{15}N resonance assignments of hCAII were deposited into the BMRB (http://www.bmrb.wisc.edu) under the accession number 34308. The assignments of backbone amide ^{1}H and ^{15}N chemical shifts of hCAII in complex with the sulfonamide inhibitor acetazolamide were obtained by comparison to the assignments of free hCAII and confirmation via 3D ^{15}N-edited NOESY and TOCSY experiments. The ^{1}H chemical shifts were referenced with respect to the external standard 2,2-dimethyl-2-silapentene-5-sulfonates (DSS), while the ^{13}C and ^{15}N chemical shifts were referenced indirectly.

NMR Relaxation: The backbone ^{15}N T_1 relaxation measurements at 800 MHz were acquired using recovery delays of 10, 50, 100, 200, 300, 500 and 700 ms. The ^{15}N T_2 measurements were

carried out with the same acquisition parameters using a CPMG pulse sequence[5] with relaxation delays of 0.5, 1, 5, 10, 20, 30, and 40 ms. 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. NMR spectra were processed using TopSpin3.5 (Bruker BioSpin) and analyzed using CARA[6] and CCPN[7].

Constant-time ^{15}N-CPMG (CT-CPMG) relaxation dispersion experiments[8] were measured at 298, 310, and 318 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. Besides, for each data set the frequencies 700 and 150 Hz were repeated for estimation of error in $R_{2,eff}$ and a reference spectrum without constant time delay ($T_{CPMG} = 0$) 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. The recycle delay of 1.5 s was used, giving rise to a net acquisition time of approximately 2.6 h per data set.

NMR relaxation data were processed using Topspin 4.0.8. Peak intensities were quantified and visualized by using SPARKY[9]. The effective amide ^{15}N transverse relaxation rate at each CPMG frequency was calculated according to the relation of effective transverse relaxation rate ($R_{2,\,eff}$),

$$R_{2,eff} = \frac{-\ln\left(\frac{I}{I_0}\right)}{T}, \qquad (1)$$

where I is the peak intensity, Io is the corresponding intensity in a reference spectrum recorded without the CT-CPMG relaxation period and T is a constant time delay. Data were analyzed individually using the NESSY software package[10] to obtain the kinetic parameters of interest, corresponding to a two-site exchange process. NESSY fits the profiles to no-exchange and fast-exchange mathematical models. It then chooses the best model for each residue based on the corrected Akaike information criterion.

Model 1: no exchange

$$R_{2,\,eff} = R_2^0 \qquad (2)$$

Model 2: two states, fast exchange

$$R_{2,\,eff} = R_2^0 + \phi_{ex}/k_{ex}[1.0 - (4\nu_{CPMG}/k_{ex})*\tanh(k_{ex}/4\nu_{CPMG})] \qquad (3)$$

For global analysis, a two-state model (*i.e*. with the same kinetic parameters for all dispersion profiles considered) was fitted to the data using the program SHEREKHAN[11], which minimizes the target function by numerically propagating mathematical equations such as the Bloch-McConnell, Carver-Richards and Luz-Meiboom models.

Solid-state ^{15}N $R_1\rho$ measurements were recorded as pseudo-3D, proton-detected HN correlation spectra and data fitting was done as described previously.[12]

NMR Structure Calculation: The 3D solution structure of hCAII was determined using the following NMR constraints: (i) Dihedral angle constraints derived using TALOS-N[13] with the knowledge of individual $^{1}H^{N}$, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO chemical shift values as inputs. A total of 387 ϕ and ψ dihedral angle constraints were used. (ii) Generic hydrogen bond (H-bond) constraints were imposed for residues located at well-defined α-helical and β-strand regions. In addition, H-bond restraints originating from H/D-exchange studies (indicating presence of a strong H-bond in solution) were used as shown in Table S3. (H-bond acceptors were defined as found in crystals, using pdb 2cba and HBOND, cib.cf.ocha.ac.jp/bitool/HBOND.) In all cases, an upper limit of 2.4 Å was used for the H-O distance. A total of 156 H-bond constraints were used. (iii) Cross peaks in NOESY spectra were identified and automatically assigned using ARIA 2.3[14]. The upper-bound distance constraints were set to 6.0 Å, respectively. An initial extended model consisting of a Zn^{2+} ion, tetrahedrally coordinated with H94, H96, H119 residues and a water molecule, was created using the CNS program[15]. A total of 3892 distance constraints, which included 764 intra-residue, 1098 inter-residue (sequential), 676 medium-range, and 1354 long-range distance constraints, were used in the 3D structure calculation. With all these restraints as input, the 3D structure of hCAII was calculated using the simulated annealing protocol in ARIA 2.3. A total of 500 structures were calculated, from which 20 structures with lowest target function and no distance or angle violations were selected. These 20 conformers with lowest target function were further refined in explicit water with 115 $^{1}H/^{15}N$-RDC restraints, NMR-derived distance restraints and angle restraints using the inbuilt ARIA 2.3 CNS program. The program PSVS-1.4 (http://www.psvs-1_4.nesg.org) was used to validate the quality of the selected ensemble of lowest-energy structures of hCAII. The 3D coordinates of individual atoms of hCAII thus obtained were deposited in the PDB (pdb id: 6HD2). The structure figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) and UCSF Chimera.

H/D exchange studies: Non-equilibrium H/D-exchange was measured on u-^{15}N hCAII (lyophilized from aqueous phosphate buffer at pH 7.5 and redissolved in D_2O) by recording

consecutive ^{15}N-^{1}H HSQC spectra over a period of 18 days (7 days continuous measurement acquisition time, using 15 to 45 min per spectrum). The intensity decay was fitted with a double exponential function as shown in equation 1.

$$I(t) = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} \quad (1)$$

The exchange rate at physiological pH generally depends on two different properties, the H-bond opening rate and the water accessibility of the site. The opening-rate of the H-bond is represented by an exponential decay rate of the HSQC-peak intensity, while hindered water accessibility, if present, results in a disturbance of this exponential function, either making it a multi-exponential decay (fitted as double exponentially decay) or effectively resulting in no decay at all.[16]

MD simulations:

All MD simulations were carried out with Gromacs 2019.2[17]. For apo hCAII, the X-ray crystal structure 3KS3[18] was used as a starting structure, whereas the dorzolamide-bound simulations were initiated from the X-ray structure 4M2U[19]. To create the N-terminally truncated proteins, the coordinates of the first 24 residues were deleted from 3KS3 and 4M2U, respectively, i.e., the first residue in the chain is Gly25. After adding hydrogen atoms to the protein, inhibitor, and the crystal water (the protonation states of all ionizable residues were set corresponding to pH=7), the systems were solvated with about 10.000 SPC/Eb water molecules in a periodic rhombic dodecahedron box. The Amber ff15ipq[20] protein force field was used. For the coordination of the Zn ion, we followed the metal center parametrization protocol of Merz and coworkers[21], using the MCPB.py program as available in the Amber18 package. Lennard-Jones 12-6 interactions were smoothly shifted to zero at a 1.0 nm cut-off; this distance was also used for switching between short- and long-range electrostatic interactions, which were treated with the particle mesh Ewald algorithm[22]. After energy minimization, the systems were heated up to 300 K during 200 ps in the NVT ensemble; in these equilibration simulations harmonic position restraints with force constants of 1000 kJ/mol/nm^2 were applied to all protein heavy atoms. Afterwards, two subsequent 200 ps NpT simulations with position restraints were carried out. In the first run, the position restraints were applied to all protein heavy atoms, and in the second run the position restraints on the protein side chain atoms were released. Finally, all position restraints were switched off. Temperature and pressure were kept constant at 300 K and 1 bar using the thermostat of Bussi and coworkers[23] and the Berendsen barostat, respectively. The SETTLE and LINCS constraint algorithms were applied to constrain internal degrees of freedom of the water molecules and all protein bonds with H-atoms, respectively,

allowing to integrate the equations of motion with 2 fs time steps. Finally, for both apo and dorzolamide-bound hCAII, 8 independent 500 ns MD simulations were initiated using different random seeds for the atomic velocities drawn from a Maxwell-Boltzmann distribution at 300 K. These sets of simulations were carried out for both the full-length proteins as well as the N-terminally truncated ones.

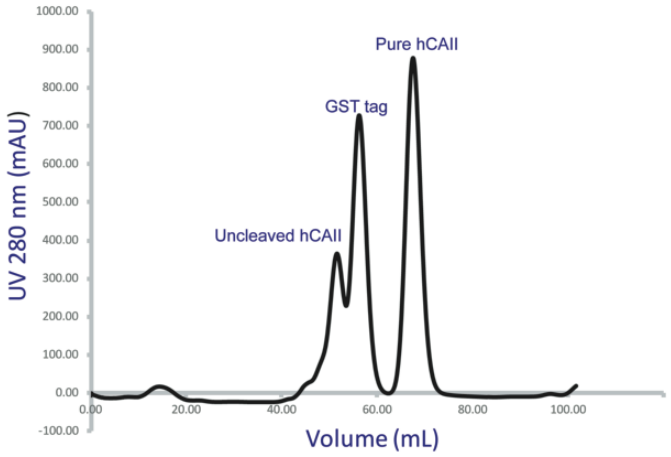


Fig. S1: Size-exclusion chromatography profile of purified hCAII.

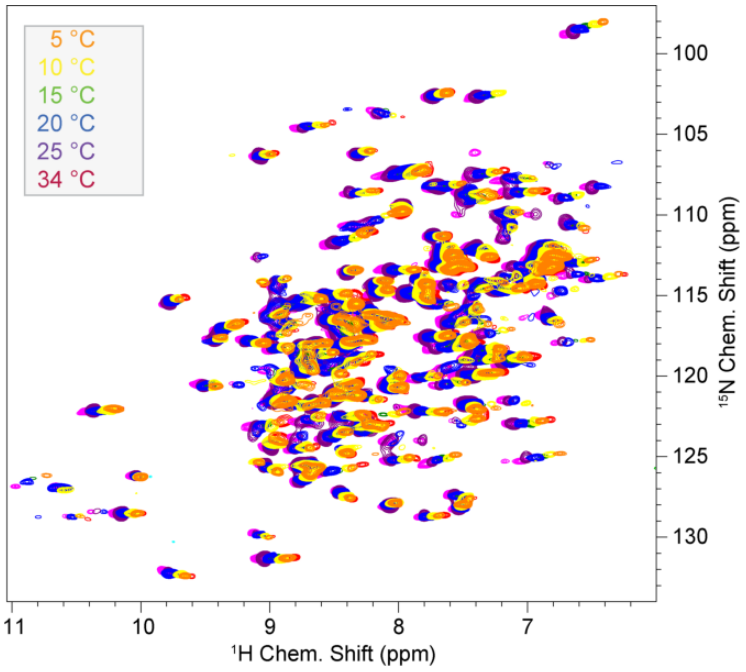


Fig. S2: Comparison of hCAII solution [^{1}H,^{15}N]-HSQC spectra under various temperatures.

100

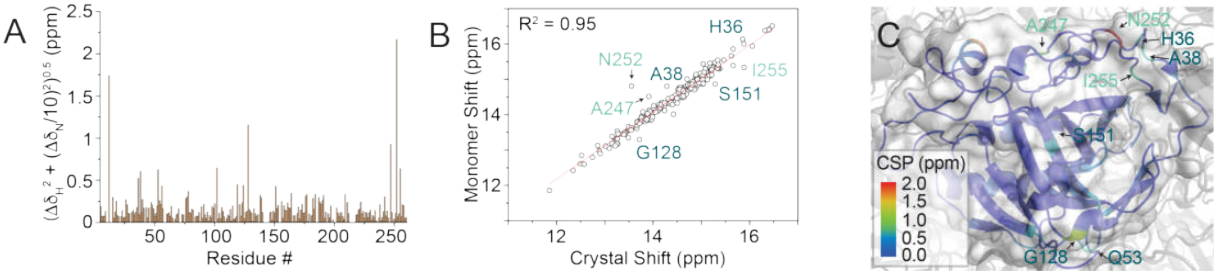


Fig. S3: Chemical shift differences between solution hCAII and hCAII in the crystalline state[1]. **A)** CSPs as a function of residue. **B)** Correlation plot of chemical shifts in monomeric and crystalline hCAII, with effective amide shifts $\sqrt{(\delta_H^2 + \delta_N^2/100)}$. **C)** Shift differences and their extent, color-coded by origin on pdb 2CBA, dark green representing direct changes obviously from crystal-crystal contacts and cyan such that must be indirectly affected by conformational changes. Solid-state spectra recorded on the same spectrometer on a protonated, uniformly ^{13}C, ^{15}N-labeled microcrystalline sample of hCAII at 111 kHz magic-angle spinning and similar temperature.

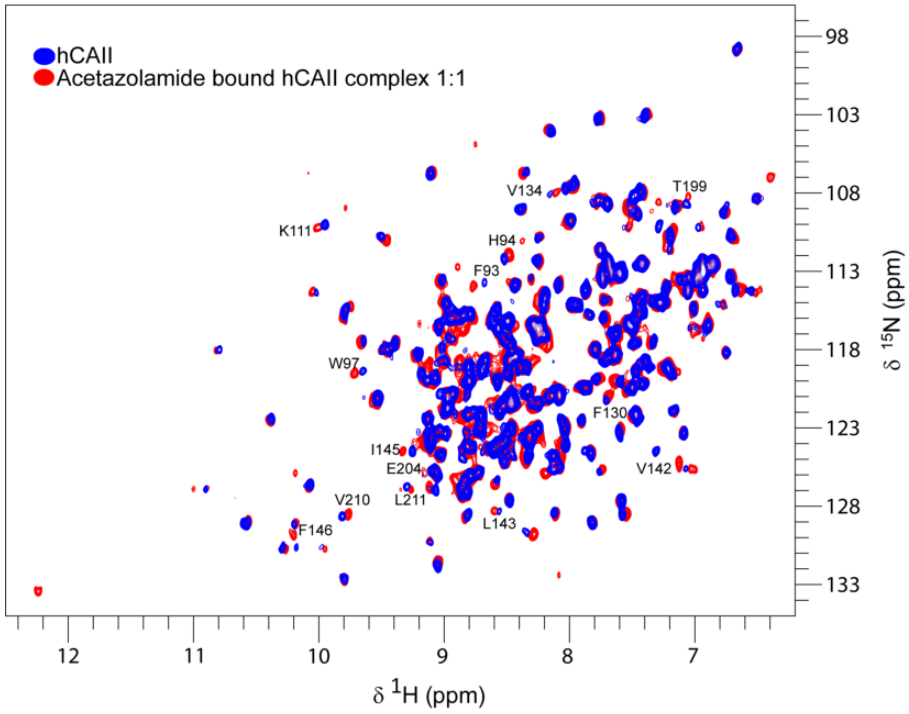


Fig. S4: Overlay of 2D [^1H, ^{15}N]-HSQC spectra for hCAII (blue) and hCAII:acetazolamide complex (red) under similar experimental conditions at 25 ºC.

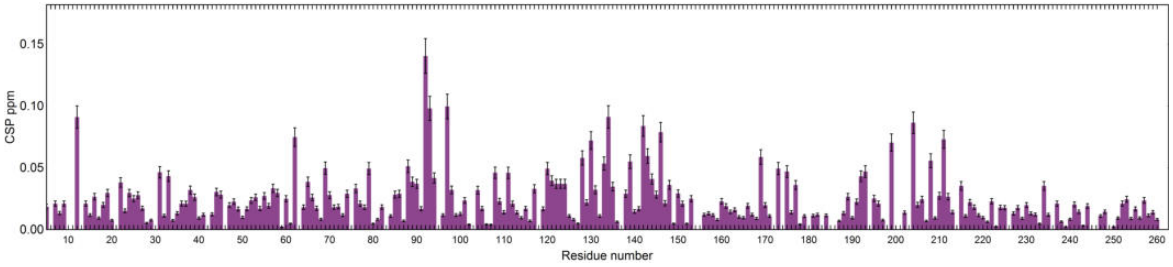


Fig. S5: Residue-wise plot of chemical shift perturbations (considering only amide proton and amide nitrogen chemical shifts) upon hCAII:acetazolamide complex formation.

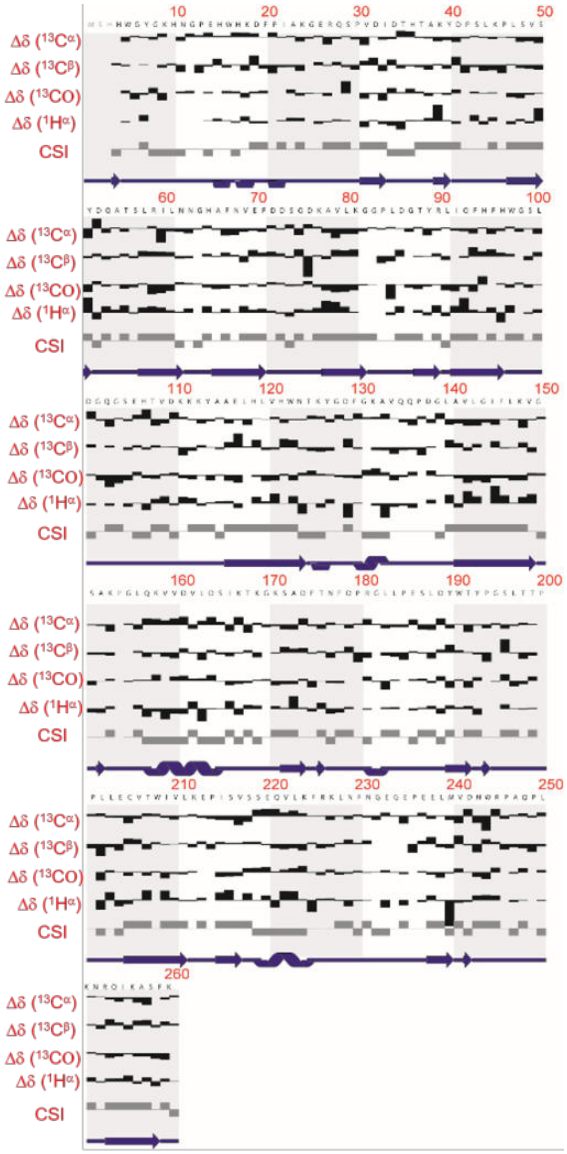


Fig. S6: Chemical shift index derived from the complete resonance assignments of hCAII.

Table S1: Experimental RDCs measured using Pf1 phages (Tensor: 1, magnitude 13.91, rhombicity 0.252).

Res.	Type	1	2	RDC		Res.	Type	1	2	RDC
7	Tyr	N	H	-8.3400		127	Tyr	N	H	2.730
8	Gly	N	H	8.070		128	Gly	N	H	-14.320
9	Lys	N	H	-10.520		129	Asp	N	H	2.140
14	Glu	N	H	-5.730		130	Phe	N	H	8.190
16	Trp	N	H	-6.970		131	Gly	N	H	13.330
17	His	N	H	13.200		135	Gln	N	H	16.450
20	Phe	N	H	18.480		136	Gln	N	H	-8.000
25	Gly	N	H	-11.200		150	Gly	N	H	14.950
27	Arg	N	H	-8.030		158	Lys	N	H	-15.340
32	Asp	N	H	-4.800		159	Val	N	H	-11.760
36	His	N	H	-8.990		163	Leu	N	H	-12.900
38	Ala	N	H	3.000		166	Ile	N	H	-2.230
39	Lys	N	H	-8.670		167	Lys	N	H	-6.220
40	Tyr	N	H	12.010		172	Ser	N	H	-7.050
41	Asp	N	H	-15.860		177	Asn	N	H	-2.720
43	Ser	N	H	19.490		178	Phe	N	H	-3.440
47	Leu	N	H	16.090		179	Asp	N	H	-4.580
48	Ser	N	H	-11.970		183	Leu	N	H	10.220
49	Val	N	H	-5.630		184	Leu	N	H	12.050
50	Ser	N	H	-10.320		186	Glu	N	H	7.870
55	Thr	N	H	-9.570		187	Ser	N	H	0.260
56	Ser	N	H	-11.350		188	Leu	N	H	-6.640
59	Ile	N	H	-13.410		189	Asp	N	H	-7.700
60	Leu	N	H	-16.760		190	Tyr	N	H	-11.990
61	Asn	N	H	-16.920		195	Gly	N	H	24.000
62	Asn	N	H	-15.370		197	Leu	N	H	2.340
63	Gly	N	H	-9.100		199	Thr	N	H	11.930
66	Phe	N	H	-15.190		202	Leu	N	H	-5.190
67	Asn	N	H	-17.060		204	Glu	N	H	-4.530
68	Val	N	H	-14.980		212	Lys	N	H	2.970
70	Phe	N	H	-6.300		213	Glu	N	H	13.840
73	Ser	N	H	-12.340		216	Ser	N	H	11.300
74	Gln	N	H	6.210		219	Ser	N	H	-9.870
81	Gly	N	H	4.240		221	Gln	N	H	-8.000
82	Gly	N	H	-8.040		222	Val	N	H	-1.440
84	Leu	N	H	-2.000		224	Lys	N	H	-10.420
86	Gly	N	H	-17.990		228	Leu	N	H	-15.000
87	Thr	N	H	5.400		231	Asn	N	H	1.150
90	Leu	N	H	-5.380		232	Gly	N	H	-2.050
92	Gln	N	H	-12.130		235	Glu	N	H	-11.560
96	His	N	H	-18.340		238	Glu	N	H	19.410
99	Ser	N	H	4.030		241	Val	N	H	-7.060
100	Leu	N	H	-8.940		243	Asn	N	H	-6.620
104	Gly	N	H	-11.500		244	Trp	N	H	11.910
112	Lys	N	H	-6.000		245	Arg	N	H	7.810
113	Lys	N	H	15.550		251	Lys	N	H	7.200
115	Ala	N	H	-13.960		254	Gln	N	H	20.170
117	Glu	N	H	-17.570		257	Ala	N	H	-10.320
122	His	N	H	6.380		259	Phe	N	H	-7.420
125	Thr	N	H	-12.330		260	Lys	N	H	-15.000

Table S2: H-bond restraints for strong H-bonds identified through HDX, according to Table S4 and Fig. S20 and S21. The first column shows the amide proton while the second one shows the oxygen. (Acceptor residues were derived from the hCAII crystal structure 2CBA coordinates using HBOND, cib.cf.ocha.ac.jp/bitool/HBOND.) Amide protons 199 and 213 seem to have no binding partner in 2cba. Column 3 to 5 show the restraint value for structure calculation.

Residue	nucleus 2	Distance	error	error
57	69 CO	1.9 Å	+0.5 Å	-0.2 Å
58	69 CO	1.9 Å	+0.5 Å	-0.2 Å
64	62 OD1 Asn	1.9 Å	+0.5 Å	-0.2 Å
67	60 CO	1.9 Å	+0.5 Å	-0.2 Å
98	115 CO	1.9 Å	+0.5 Å	-0.2 Å
112	109 CO	1.9 Å	+0.5 Å	-0.2 Å
114	105 OG Serin	1.9 Å	+0.5 Å	-0.2 Å
125	87 CO	1.9 Å	+0.5 Å	-0.2 Å
127	124 CO	1.9 Å	+0.5 Å	-0.2 Å
133	129 CO	1.9 Å	+0.5 Å	-0.2 Å
134	132 CO	1.9 Å	+0.5 Å	-0.2 Å
136	133 CO	1.9 Å	+0.5 Å	-0.2 Å
139	138 CO	1.9 Å	+0.5 Å	-0.2 Å
160	156 CO	1.9 Å	+0.5 Å	-0.2 Å
184	181 CO	1.9 Å	+0.5 Å	-0.2 Å
223	219 CO	1.9 Å	+0.5 Å	-0.2 Å
248	30 CO	1.9 Å	+0.5 Å	-0.2 Å

Table S3: NMR structural statistics for the ensemble of 20 refined conformers of hCAII (*pdb id:6HD2*)

Conformationally restricting restraints[a]	
Distance Restraint List	
Total	4550
Intraresidue (i = j)	764
Sequential (| i-j | = 1)	1098
medium-range (2 ≤ | i-j | ≥ 4)	676
long-range (| i-j | ≥ 5)	1354
Hydrogen bonds	156
Dihedral angle restrains (ϕ and φ)	387
$^{1}H/^{15}N$-RDCs	115
No. of restraints per residue[b]	17.5
Model quality[c]	
Rmsd backbone atoms (Å)	0.89
Rmsd heavy atoms (Å)	2.2
Rmsd bond lengths (Å)	0.005
Rmsd bond angles (°)	0.8
MolProbity Ramachandran statistics[d]	
Most favored region (%)	83.6
Allowed region (%)	15.6
Additionally allowed region (%)	0.5
Disallowed region (%)	0.6
Global quality scores (raw/Z score)	
Verify3D	0.42
PROCHECK (ϕ-ψ)[e]	-0.61
PROCHECK (all)[e]	-0.46
MolProbity clash score	26.03
Model contents	
Total no. of residues	260
BMRB accession number	34308
PDB ID code	6HD2

[a] Analyzed for residues 1 to 260
[b] There are 254 residues with conformationally restricting constraints
[c] *Ordered residue ranges: 13-20,33-41,45-62,64-83,87-96,107-109,112-128,130-138,140-149,152-183,188-195,204-245,247-251,254-259*
[d] Residues selected based on: Dihedral angle order parameter, with S(*ϕ*)+S(*ψ*)>=1.8
[e] Residues with sum of phi and psi order parameters > 1.8
Generated using PSVS 1.5

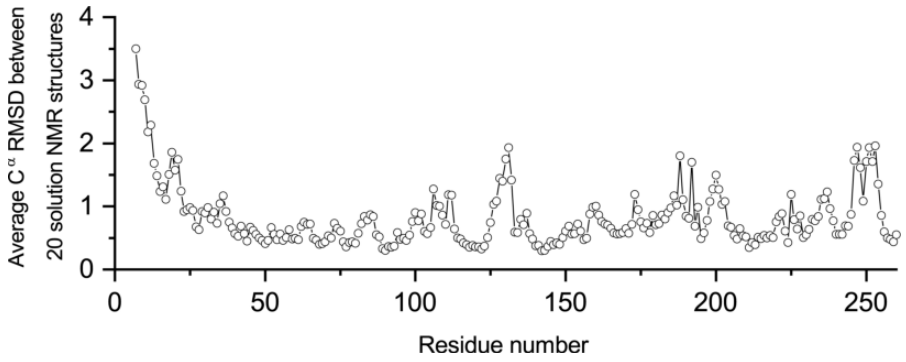


Fig. S7: Residue-resolved RMSD of C^{α} positions over a bundle of 20 lowest-energy structures.

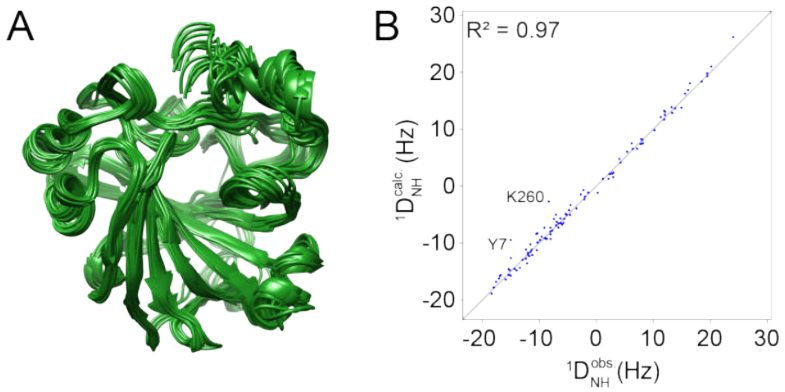


Fig. S8: hCAII monomeric structure in solution. **A)** An ensemble of 20 superimposed minimum-energy structures of monomeric hCAII. **B)** Good correlation between calculated N-H RDCs based on the average NMR structure and experimentally measured N-H RDCs in solution. The N-terminal amino acids 1-7 are not shown in **A)**.

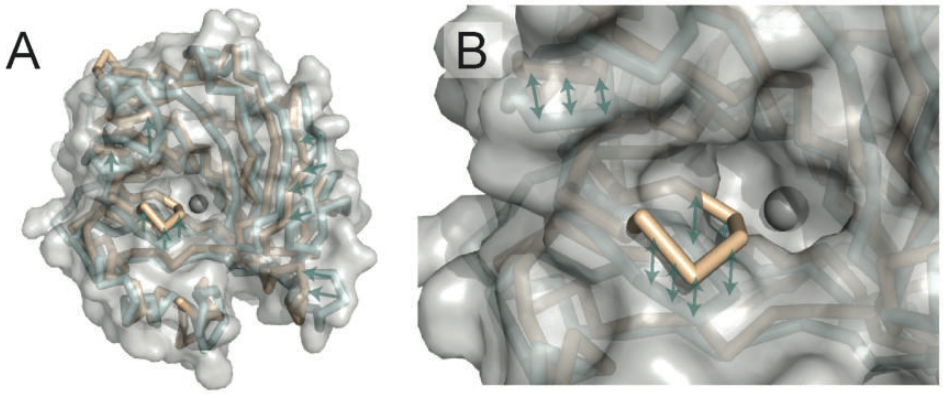


Fig. S9: Comparison of average-NMR-structure active-site-loop conformation with the conformation in the crystal (2cba). Green arrows denote changes going from X-ray to NMR structure. **A)** Overview, **B)** zoom into the active site.

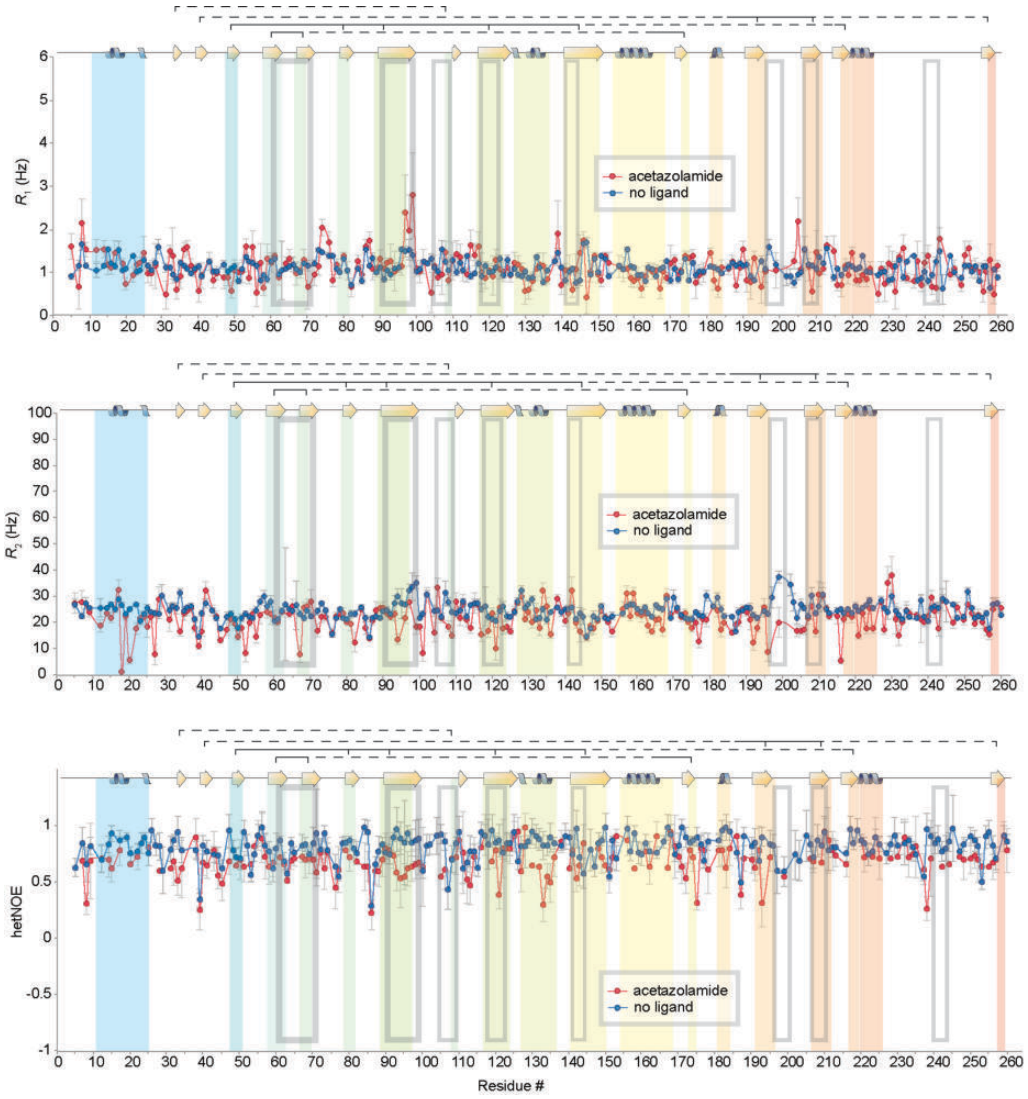


Fig. S10: Site-specific ^{15}N relaxation properties (R_1, R_2, and hetNOE) of non-ligand-bound (blue) and acetazolamide-bound hCAII (red).

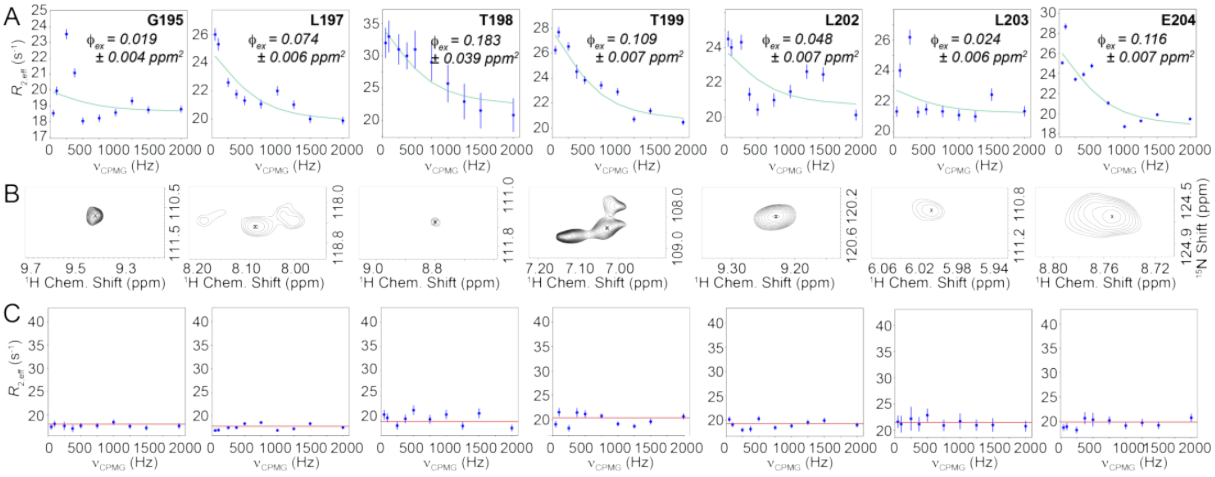


Fig. S11: Relaxation dispersion profiles **(A)** and peak shapes **(B)** for the active-site residues 195 to 204 at a temperature of 45 °C. (Residues 200 and 201 are prolines.) The exchange life time was globally fitted to τ_{ex} = 270 ± 13 µs. **C)** RD profiles for the inhibited case.

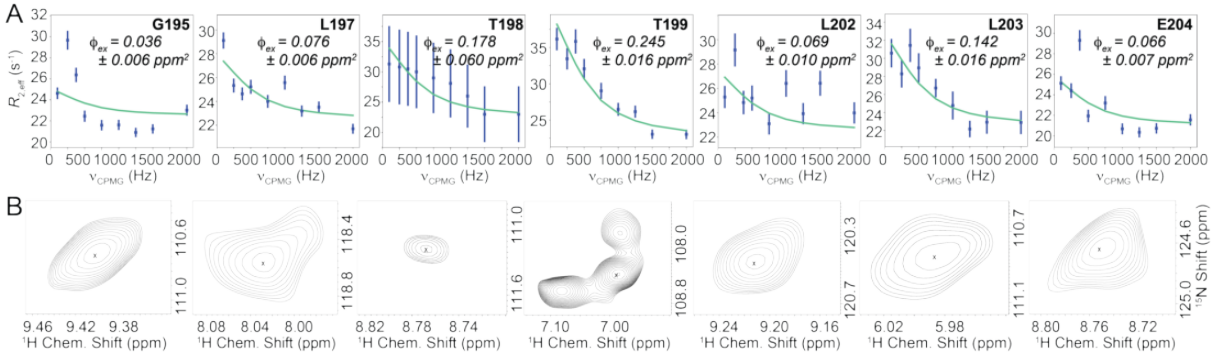


Fig. S12: Relaxation dispersion profiles **(A)** and peak shapes **(B)** for the active-site residues as in Fig. S10, but at 37 °C. The exchange life time was globally fitted to τ_{ex} = 290 ± 23 µs.

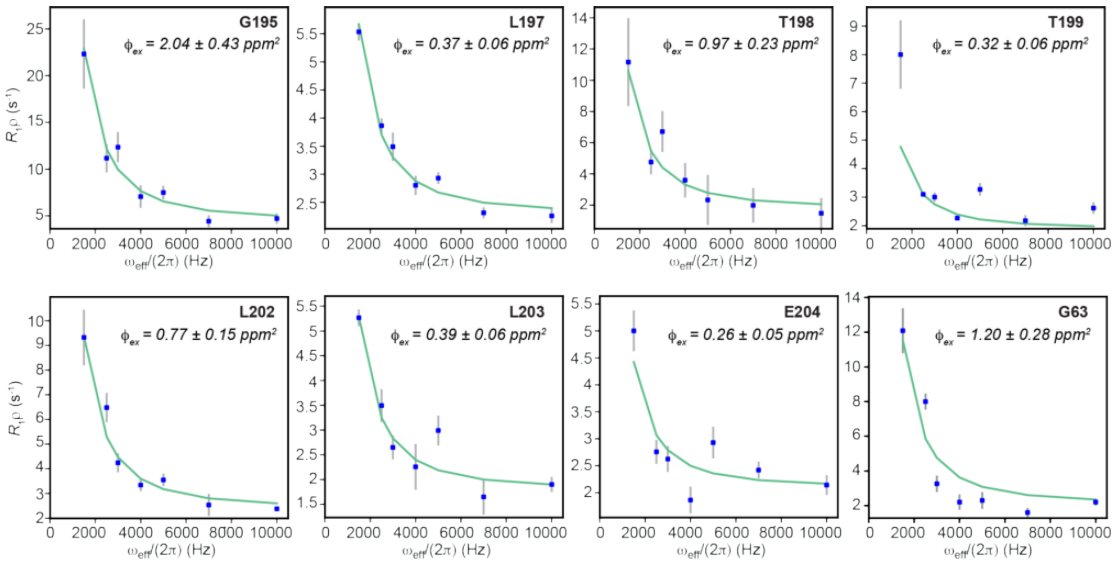


Fig. S13: Solid-state NMR relaxation dispersion of the active-site loop, with an exchange life time globally fit to 204 ± 46 µs. G63 was fit individually, yielding an exchange life time of 250 ± 67 µs.

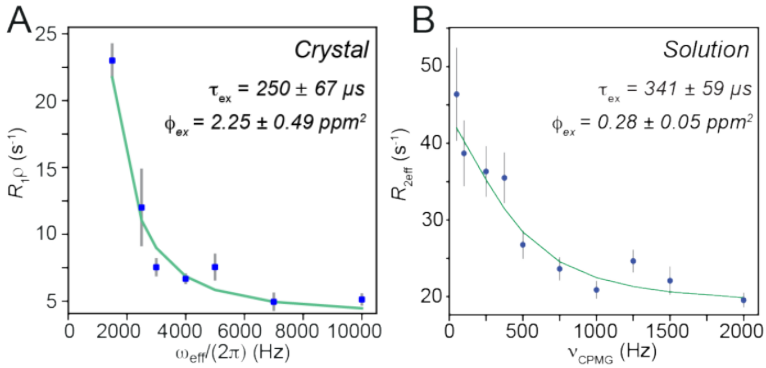


Fig. S14: Comparison of G6 dispersion profiles in crystalline form **(A)** and in solution at 45 °C **(B)**.

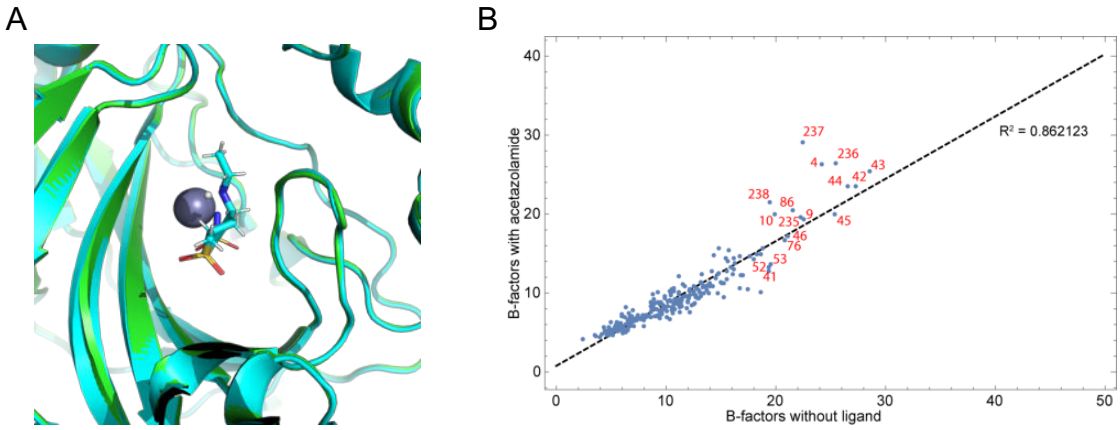


Fig. S15: A) Overlay of non-liganded (green, pdb 4Y0J) and dorzolamide-bound X-ray structures (blue, pdb 6BC9) of hCAII. **B)** Correlation between C^{α} B-factors of non-liganded (pdb 2cba) hCAII and hCAII soaked with acetazolamide (pdb 3hs4). Flexible residues with B-factors larger than 20 in any of the two structures are labeled. (None of them associated with the active-site loop.)

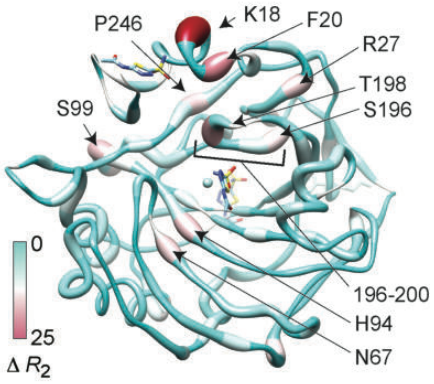


Fig. S16: Sausage representation of R_2 difference values between holo and AZ-bound hCAII.

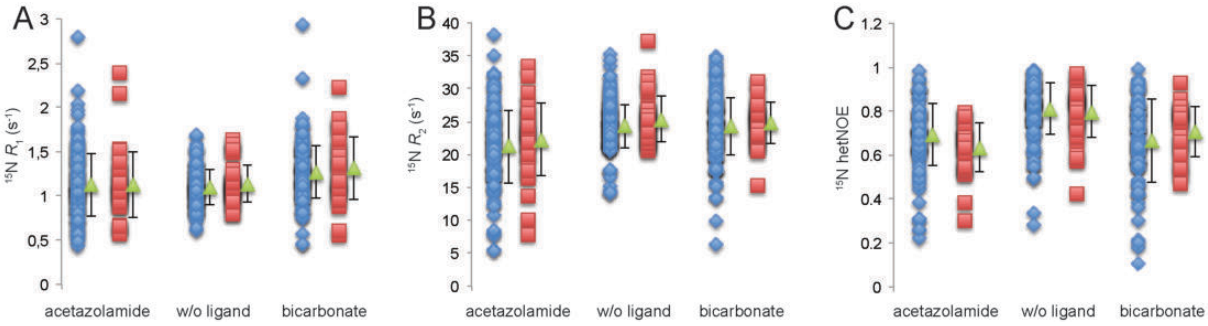


Fig. S17: Distributions of relaxation values compared between different ligand binding situations. **A)** ^{15}N R_1 relaxation, **B)** ^{15}N R_2 relaxation, **C)** ^{1}H-^{15}N het-nOe. Red and blue distributions denote active-site and non-active-site residues, respectively. Green triangles denote the average and standard deviation of each distribution.

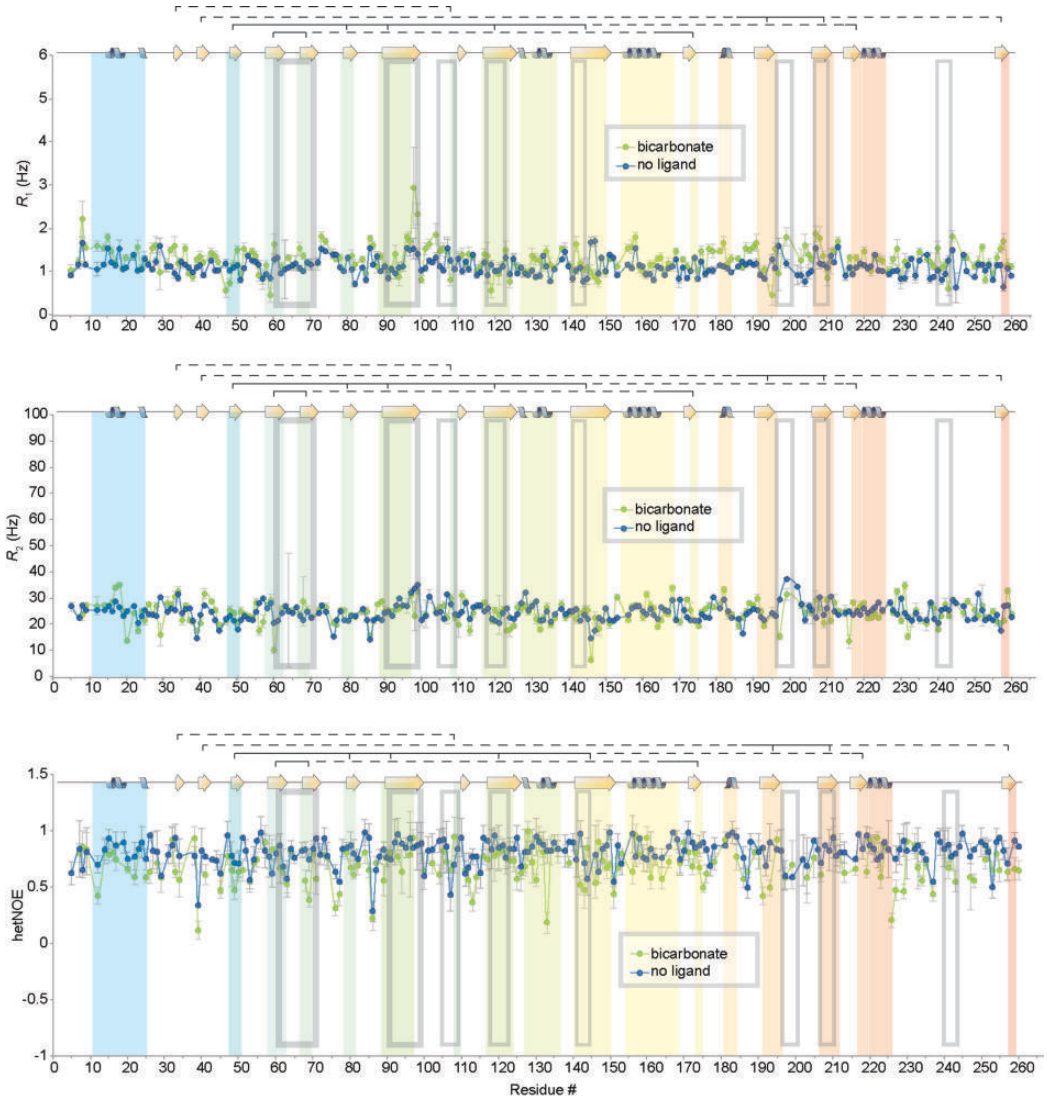


Fig. S18: Site-specific ^{15}N relaxation properties (R_1, R_2, and hetNOE) of non-ligand-bound (blue, as in Fig. S9) vs. bicarbonate-bound hCAII (green).

Assessment of protection of active-site regions of hCAII against H/D exchange

Site-resolved structural stability in (native) hCAII was assessed via a combination of H/D exchange experiments, in which lyophilized, protonated protein was dissolved in D_2O and – while protons steadily exchanged against deuterons – monitored as a function of time using HSQC experiments. The amide protons in hCAII can be divided into three main classes according to their exchange properties: *i)* The rapidly exchanging amide protons (red in Fig. S19) are, as expected, mainly located in loop regions and on the surface of the protein. More importantly, the exchange rates of the buried amides depend on both H-bond opening and accessibility of the site by breathing[16d, 24]. Here, we find *ii)* regions which show slow exchange (half-life times between a few hours to several days, turquoise in Fig. S19B), as well as *iii)*

residues that are fully protected and do not exchange over several months (W191 H^N in Fig S19A and blue color in Fig. S19B). Most group-ii residues show a double-exponential H/D exchange profile, representing the two components, H-bond opening and diffusion to the site. (Table S4 shows exchange rates obtained from mono- or biexponential fitting. Fig. S20 and S21 show the faster exponential rate as a function of sequence.) Interestingly, group-iii residues are localized close to both the active site as well as to the hydrogen-bonded water networks of hCAII [25]. (Fig. S22 depicts the presence of double exponentiality as a function of residue.) The intermediate lifetime of these amide protons – despite their direct proximity to the cone surface – confirms exchangeability within the pocket. T199 is shown in Fig. S19A in comparison to an exchange-protected residue further inside the core.

Also H64, which has been implicated in conformational dynamics, displays a similar amide exchange behavior. The sidechain of H64 at the entrance to the active site shows two orientations in various X-ray structures under cryogenic conditions. On the basis of ^{15}N relaxation dispersion (see main text), the residues around H64 backbone amide display dynamics on the 100 µs timescale both with and without inhibitor. Despite the exposed nature at the tip of a loop pointing to the bulk water, H/D exchange of the H64 amide is on the order of 1d (Fig. S19A). Accordingly, as expected, its H-bond is not broken by the exchange dynamics of the pocket architecture.

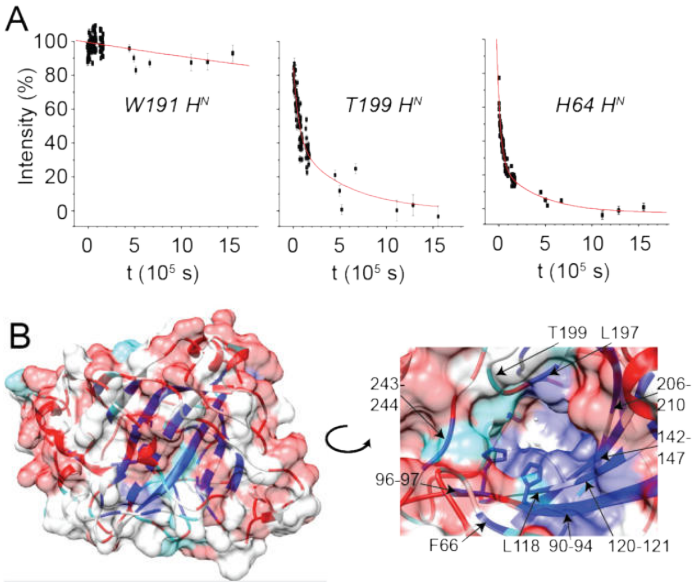


Fig. S19: hCAII active-site HD exchange in solution NMR. **A)** Slow H/D exchange profile of W191 H^N in comparison to intermediate rates in T199 H^N and H64 H^N. **B)** Water accessibility depicted on crystal structure 2CBA. Deep blue colored residues: no signal decay during the measured time period; turquoise residues: clearly double exponential behavior; red colored residues: either purely fast monoexponential behavior or equilibration before the first spectrum was measured. Right: close-up of the active site.

Table S4: Exponential rates fitted using eq. 1. The faster exponential rate is interpreted as the opening rate of the individual H-bond. In order to analyze the faster rate qualitatively we set a border of $2 \cdot 10^{-3}$ min^{-1}. Residues corresponding to rates slower than this are colored in blue, residues corresponding to rates faster than this are colored in yellow. Residues corresponding to peaks that have reached equilibrium before the first measurement (peaks gone "missing") are colored in red. Residues with very slow decay, such that a fit using eq. 1 is not possible, are shaded in gray.

Residue	H-bond opening rate k_1 [min^{-1}]	Second exponential rate k_2 [min^{-1}]
5	Missing	Missing
6	Missing	Missing
7	Missing	Missing
8	Missing	Missing
16*	5,26E-06	5,25E-06
17	Missing	Missing
20	Missing	Missing
25	Missing	Missing
28	0,00232	4,83E-05
29	2,19E-50	2,14E-50
31	1,40E-05	1,39E-05
33	7,52E-06	7,50E-06
34	Missing	Missing
37	Missing	Missing
38	Missing	Missing
39	Missing	Missing
41	Missing	Missing
43	Missing	Missing
44	Missing	Missing
45	Missing	Missing
49	Missing	Missing
50	Missing	Missing
53	Missing	Missing
54	Missing	Missing
56	(2,91E-05) missing peak, noise level	(2,91E-05) missing peak, noise level
57	0,00179	3,40E-04
58	1,70E-04	1,70E-04
59	1,70E-06	1,70E-06
61	0,01116	1,15E-137
62	Missing	Missing
63	Missing	Missing
64	0,00162	1,71E-04
66	6,82E-06	6,82E-06
67	0,0017	1,57E-04
69	8,32E-63	1,10E-64
72	Missing	Missing
74	Missing	Missing
76	Missing	Missing
81	Missing	Missing
82	Missing	Missing
84	Missing	Missing
86	Missing	Missing
87	Missing	Missing
88	0,00337	1,33E-22

Residue	H-bond opening rate k_1 [min^{-1}]	Second exponential rate k_2 [min^{-1}]
93	-7,76E-07	-7,95E-07
94	9,17E-50	1,37E-50
95	Missing	Missing
96	3,68E-39	3,13E-39
97	5,51E-44	5,51E-44
98	0,00104	1,27E-84
99	Missing	Missing
100	7,50E-05	7,50E-05
101	0,01549	1,02E-04
102	Missing	Missing
103	Missing	Missing
104	Missing	Missing
105	-8,23E-147	-8,23E-147
106 *	3,94E-06	3,94E-06
107	3,36E-50	1,06E-57
108	1,07E-06	1,06E-06
109	5,53E-05	5,53E-05
110	0,00667	0,00667
111	Missing	Missing
112	0,00138	1,67E-04
113	Missing	Missing
114	0,0015	2,10E-04
115	Missing	Missing
116	4,14E-06	4,13E-06
117 *	1,81E-6	1,81E-6
118	1,93E-06	1,93E-06
(119) *	(3,70E-33)	(9,50E-38)
120	-	2,55E-05
121	2,43E-06	2,41E-06
123	1,04E-05	1,04E-05
124	6,22E-06	6,22E-06
125	0,00206	2,11E-04
126	Missing	Missing
127	0,00133	1,78E-04
128	Missing	Missing
130	Missing	Missing
131	Missing	Missing
132	Missing	Missing
133	0,00152	4,50E-05
134	1,44E-04	-
135	0,03984	0,00343
136	0,00167	1,00E-04
138	Missing	Missing
139	0,00168	0,00168

113

90	4,43E-05	4,43E-05
91	7,05E-06	7,04E-06
92	5,93E-39	2,83E-39
144	2,01E-49	1,35E-49
145	5,80E-36	1,97E-36
146	3,44E-06	3,44E-06
147	8,91E-06	8,89E-06
150	Missing	Missing
153	1,83E-66	1,83E-66
156	Missing	Missing
157	Missing	Missing
159	0,03158	0,03158
160	1,02E-04	-
161	0,02679	-5,92E-117
163	Missing	Missing
164	1,99E-04 missing peak, noise level	1,99E-04 missing peak, noise level
165	Missing	Missing
166	0,00281	1,28E-86
167	0,00371	1,98E-04
168	0,00819 missing peak, noise level	0,00818 missing peak, noise level
169	Missing	Missing
171	Missing	Missing
172	Missing	Missing
175	Missing	Missing
176	Missing	Missing
179	Missing	Missing
181	(7,83E-05) missing peak, noise level	(3,30E-155) missing peak, noise level
182	Missing	Missing
183	0,00372	6,39E-05
184	1,90E-04	1,90E-04
187	Missing	Missing
188	Missing	Missing
191	5,86E-06	5,85E-06
192	1,79E-06	1,79E-06
193	7,81E-07	7,68E-07
195	-3,08E-154	-8,48E-142
196 *	5,11E-49	5,11E-49
197	6,93E-33	2,18E-33
198	Missing	Missing
199	9,42E-04	9,97E-05
203	1,15E-90	2,81E-99
204	Missing	Missing
205	0,00366	1,32E-04
206	2,95E-05	2,95E-05
207	4,07E-06	4,06E-06
208	3,36E-40	2,17E-40
209 *	1,12E-05	1,12E-05
210	2,92E-06	2,92E-06
144	2,01E-49	1,35E-49
145	5,80E-36	1,97E-36
146	3,44E-06	3,44E-06

140	6,19E-30	1,94E-30
142	-1,42E-06	-1,42E-06
143	1,35E-52	3,68E-55
156	Missing	Missing
157	Missing	Missing
159	0,03158	0,03158
160	1,02E-04	-
161	0,02679	-5,92E-117
163	Missing	Missing
164	1,99E-04 missing peak, noise level	1,99E-04 missing peak, noise level
165	Missing	Missing
166	0,00281	1,28E-86
167	0,00371	1,98E-04
168	0,00819 missing peak, noise level	0,00818 missing peak, noise level
169	Missing	Missing
171	Missing	Missing
172	Missing	Missing
175	Missing	Missing
176	Missing	Missing
179	Missing	Missing
181	(7,83E-05) missing peak, noise level	(3,30E-155) missing peak, noise level
182	Missing	Missing
183	0,00372	6,39E-05
184	1,90E-04	1,90E-04
187	Missing	Missing
188	Missing	Missing
191	5,86E-06	5,85E-06
192	1,79E-06	1,79E-06
193	7,81E-07	7,68E-07
195	-3,08E-154	-8,48E-142
196 *	5,11E-49	5,11E-49
197	6,93E-33	2,18E-33
198	Missing	Missing
199	9,42E-04	9,97E-05
203	1,15E-90	2,81E-99
204	Missing	Missing
205	0,00366	1,32E-04
206	2,95E-05	2,95E-05
207	4,07E-06	4,06E-06
208	3,36E-40	2,17E-40
209 *	1,12E-05	1,12E-05
210	2,92E-06	2,92E-06
211	3,76E-07	3,73E-07
213	0,00149	1,78E-04
215	1,95E-06	1,94E-06
217	3,57E-06	3,57E-06
222*	5,42E-05	5,42E-05
223	1,39E-04	1,39E-04
224	0,0136	9,65E-05
227	0,0044	2,20E-97
229	0,00892	4,03E-84

147	8,91E-06	8,89E-06
150	Missing	Missing
153	1,83E-66	1,83E-66
233	(9,89E-05) already decayed	(9,89E-05) already decayed
235	Missing	Missing
238	Missing	Missing
239	Missing	Missing
240	2,71E-06	2,71E-06
241	0,02177	-5,71E-108
243	2,20E-05	2,20E-05
244	8,59E-06	8,58E-06
245	Missing	Missing
247	Missing	Missing
248	8,80E-04	3,70E-05
250	Missing	Missing
255	Missing	Missing
258	0,00238	5,802E-96
259	Missing	Missing
260	Missing	Missing

230	(6,76E-06) missing peak, noise level	(5,85E-06) missing peak, noise level
231	Missing	Missing
232	Missing	Missing

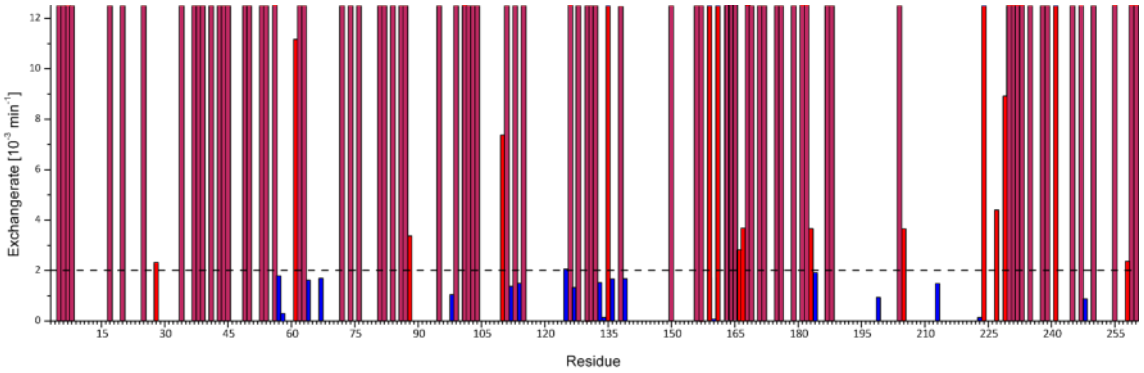


Fig. S20: Fast and intermediate H/D-exchanging residues. The faster exchange rate of the double-exponential function 1 is interpreted as the H-bond opening rate, corresponding to the H-bond strength. The dashed line marks an arbitrary border below which the residues are counted as H-bonded and are colored in blue. Residues with a rate faster than this are colored in red. Residues which are already fully decayed before the first measurement point are colored in ruby, they are not at all or weakly H-bonded. Very long-lived residues (gray in Table S4) cannot be fitted accurately and hence are not shown.

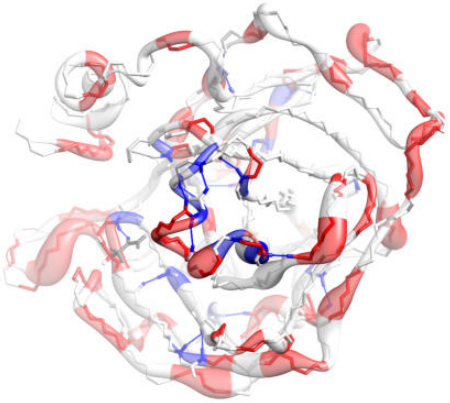


Fig. S21: H-bonds depicted on the crystal structure of hCAII. Strong H-bonds according to Table S4 and Fig. S20 are shown as blue lines. In this plot ^{15}N $R_1\rho$ relaxation data from crystalline hCAII[1] is shown as ribbon thickness. The ribbon of strongly H-bonded residues is colored in blue, while the ribbon of not or weakly H-bonded residues, according to the "missing" peaks in Table S4 and Fig. S20, is colored in red. Sidechains are colored in gray.

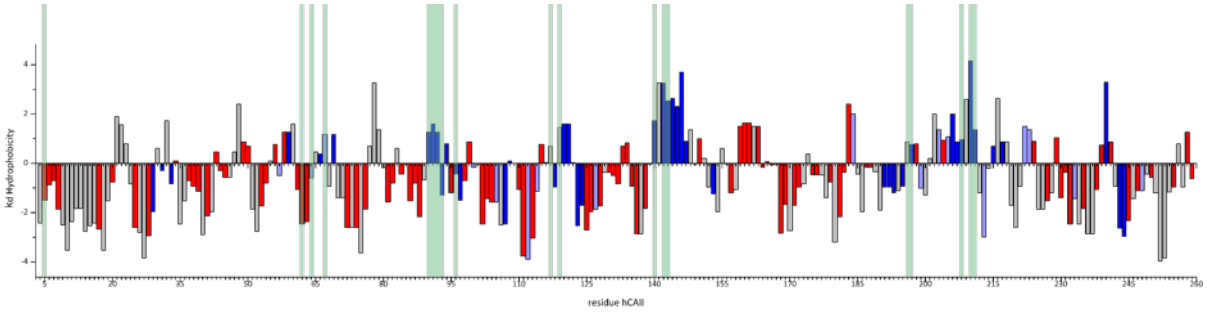


Fig. S22: Exchange-hindered residues (dark blue) and exchanging residues (red) as a function of sequence, presented with their Kite-Doolittle hydrophobicity as a moving average of 3. Exchange-hindered residues show nearly no signal decay during the measured time period. Light-blue-labeled residues show double exponential behavior according to Table S4. Red-labeled residues show either purely fast mono-exponential behavior or are already equilibrated before the first spectrum was obtained (dead time of around 15 min.). The green-shaded residues denote the active site.

References:

[1] S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J. Phys. Chem. Lett.* **2018**, *9*, 1307−1311.
[2] R. A. Venters, B. T. Farmer, C. A. Fierke, D. S. Leonard *J. Mol. Biol.* **1996**, *264*, 1101–1116.
[3] M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* **1998**, *131*, 373-378.
[4] M. Zweckstetter, *Nat. Protoc.* **2008**, *3*, 679-690.
[5] S. Meiboom, D. Gill, *Rev. Sci. Inst.* **1958**, *29*, 688-691.
[6] R. Keller, PhD thesis (ETH Zurich), **2004**.
[7] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, P. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687–696.
[8] F. A. A. Mulder, N. R. Skrynnikov, B. Hon, F. W. Dahlquist, L. E. Kay, *J. Am. Chem. Soc.* **2001**, *123*, 967-975.
[9] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325-1327.
[10] M. Bieri, P. R. Gooley, *BMC Bioinformatics* **2011**, *12*, 421.
[11] A. Mazur, B. Hammesfahr, C. Griesinger, D. Lee, M. Kollmar, *Bioinformatics* **2013**, *29*, 1819-1820.
[12] H. Singh, S. K. Vasa, H. Jangra, P. Rovó, C. Päslack, C. K. Das, H. Zipse, L. V. Schäfer, R. Linser, *J. Am. Chem. Soc.* **2019**, *141*, 19276-19288.
[13] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213-223.
[14] W. Rieping, M. Habeck, B. Bardiaux, M. Bernard, T. E. Malliavin, M. Nilges, *Bioinformatics* **2007**, *23*, 381-382.
[15] A. T. Brünger, *Nat. Protoc.* **2007**, *2*, 2728-2733.
[16] a) S. W. Englander, D. B. Calhoun, J. J. Englander, N. R. Kallenbach, R. K. Liem, E. L. Malin, C. Mandal, J. R. Rogero, *Biophys. J.* **1980**, *32*, 577-589; b) O. Jardetzky, M. D. Finucane, *Molecular Physics* **1998**, *95*, 1127-1136; c) A. Tomita, T. Sato, K. Ichiyanagi, S. Nozawa, H. Ichikawa, M. Chollet, F. Kawai, S.-Y. Park, T. Tsuduki, T. Yamato, S.-y. Koshihara, S.-i. Adachi, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2612-2616; d) K. Grohe, K. T. Movellan, S. K. Vasa, K. Giller, S. Becker, R. Linser, *J. Biomol. NMR* **2017**, *68*, 7-17.
[17] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19-25.
[18] B. S. Avvaru, C. U. Kim, K. H. Sippel, S. M. Gruner, M. Agbandje-McKenna, D. N. Silverman, R. McKenna, *Biochemistry* **2009**, *49*, 249-251.
[19] M. A. Pinard, C. D. Boone, B. D. Rife, C. T. Supuran, R. McKenna, *Bioorg. Med. Chem.* **2013**, *21*, 7210-7215.
[20] K. T. Debiec, D. S. Cerutti, L. R. Baker, A. M. Gronenborn, D. A. Case, L. T. Chong, *J. Chem. Theory Comput.* **2016**, *12*, 3926-3947.
[21] a) P. Li, K. M. Merz, *J. Chem. Inf. Model.* **2016**, *56*, 599-604; b) M. B. Peters, Y. Yang, B. Wang, L. Füsti-Molnár, M. N. Weaver, K. M. Merz, *J. Chem. Theory Comput.* **2010**, *6*, 2935-2947.
[22] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8592.
[23] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys. Lett.* **2007**, *126*, 014101.
[24] C. J. Craven, N. M. Derix, J. Hendriks, R. Boelens, K. J. Hellingwerf, R. Kaptein, *Biochemistry* **2000**, *39*, 14392-14399.
[25] K. M. Merz Jr., M. A. Murcko, P. A. Kollman, *J. Am. Chem. Soc.* **1991**, *113*, 406-411.

2.2.2 Protein motional details revealed by complementary structural-biology techniques

The SH3 domain of chicken α-spectrin (see Chapter 2.2.1) is mainly used as a test protein for methods development by the Linser group, due to its favorable behavior in crystalline form. However, in addition to its suitability as a target for methods development, the SH3-family also has great biological relevance as a protein-binding domain. When using the SH3 samples for the development of state-of-the-art assignment-, structure determination-, and dynamics-methodology in the solid-state, more and more data on this protein was collected as a side-effect. Therefore, I chose SH3 as a biological target by applying the recently developed spatial dynamics determination by eNOE-based multistate structure calculation in combination with NMR-restrained MD-simulation. The eNOE-based multistate structure determination, developed by Vögeli et al. and Güntert et al., is the first approach for the experimental determination of spatial protein motion that complements the theoretical MD simulation. Therefore, the combination of NMR-multistate structure determination, MD simulation, and existing dynamics data from method development projects led to a deeper understanding of spatial motions of the SH3 domain and is a step towards a better understanding of fundamental protein dynamics.

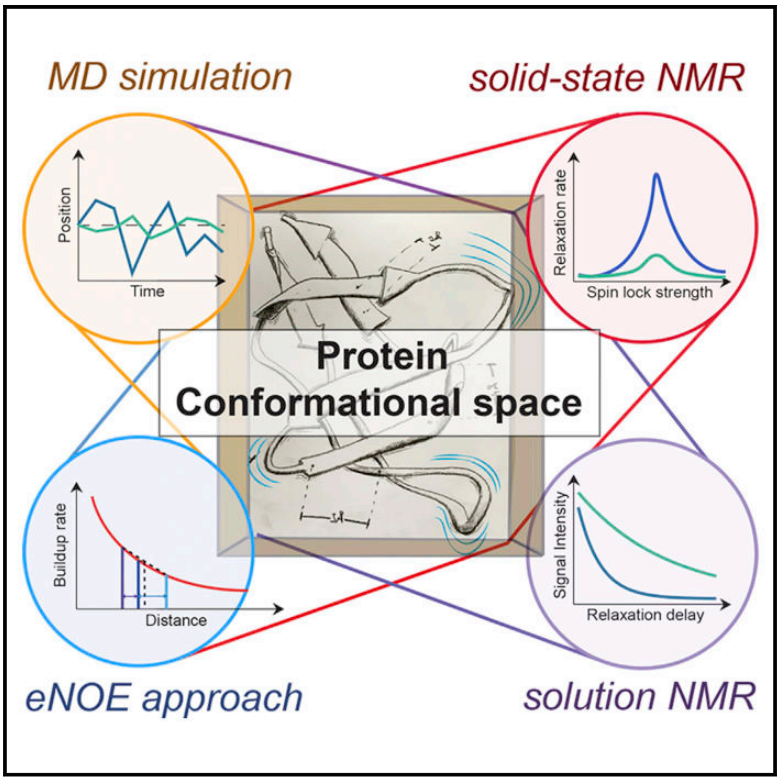
Contributions by co-authors:

Cornelia Hebrank performed the assignment of sidechain shifts of SH3-domain in solution; Snehal Patel performed the MD simulations; Sara M. Gomez, Alexander Klein, Dr. Suresh Kumar Vasa, and Dr. Himanshu Singh are responsible for recording the solution-state dynamics of SH3; Petra Rovó recorded the solid-state dynamics of SH3; Prof. Dr. Beat Vögeli contributed additional information for the application of the software package for multistate structure determination (developed by Vögeli and Güntert); Prof. Dr. Lars V. Schäfer is responsible for the supervision of the MD setup and the organizational background of Snehal Patel.

Structure

Protein Motional Details Revealed by Complementary Structural Biology Techniques

Graphical Abstract



Highlights

- A wide range of complementary structural biology techniques used in conjunction
- Experimental and computational assessment as mutual validation for protein dynamics
- Integrated characterization using ensemble MD simulations restrained by exact NOE restraints

Authors

Kristof Grohe, Snehal Patel, Cornelia Hebrank, ..., Beat Vögeli, Lars V. Schäfer, Rasmus Linser

Correspondence

lars.schaefer@ruhr-uni-bochum.de (L.V.S.), rasmus.linser@tu-dortmund.de (R.L.)

In Brief

Structural biology requires robust data on both structure and dynamics. Here, Grohe et al. combine different experimental and computational approaches to showcase which complementary information is obtained for a small protein. The study demonstrates how integrative approaches, where different methodologies mutually validate each other, can provide the most complete picture.

Grohe et al., 2020, Structure *28*, 1–11
September 1, 2020
https://doi.org/10.1016/j.str.2020.06.001

Structure

CellPress

Article

Protein Motional Details Revealed by Complementary Structural Biology Techniques

Kristof Grohe,[1,2] Snehal Patel,[3] Cornelia Hebrank,[1] Sara Medina,[1,2] Alexander Klein,[1,2] Petra Rovó,[1] Suresh K. Vasa,[1,2] Himanshu Singh,[1,2] Beat Vögeli,[4] Lars V. Schäfer,[3,*] and Rasmus Linser[1,2,5,*]

[1]Faculty for Chemistry and Pharmacy, Ludwig-Maximilians-University Munich, 81377 Munich, Germany

[2]Faculty of Chemistry and Chemical Biology, Technical University Dortmund, 44227 Dortmund, Germany

[3]Theoretical Chemistry, Ruhr University Bochum, 44801 Bochum, Germany

[4]Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Aurora, CO 80045, USA

[5]Lead Contact

*Correspondence: lars.schaefer@ruhr-uni-bochum.de (L.V.S.), rasmus.linser@tu-dortmund.de (R.L.)

https://doi.org/10.1016/j.str.2020.06.001

SUMMARY

Proteins depend on defined molecular plasticity for their functionality. How to comprehensively capture dynamics correctly is of ubiquitous biological importance. Approaches commonly used to probe protein dynamics include model-free elucidation of site-specific motion by NMR relaxation, molecular dynamics (MD)-based approaches, and capturing the substates within a dynamic ensemble by recent eNOE-based multiple-structure approaches. Even though MD is sometimes combined with ensemble-averaged NMR restraints, these approaches have largely been developed and used individually. Owing to the different underlying concepts and practical requirements, it has remained unclear how they compare, and how they cross-validate and complement each other. Here, we extract and compare the differential information contents of MD simulations, NMR relaxation measurements, and eNOE-based multi-state structures for the SH3 domain of chicken α-spectrin. The data show that a validated, consistent, and detailed picture is feasible both for timescales and actual conformational states sampled in the dynamic ensemble. This includes the biologically important side-chain plasticity, for which experimentally cross-validated assessment is a significant challenge.

INTRODUCTION

Protein function is enabled by a combination of structural features with a well-defined and responsive extent of conformational plasticity. Atomic resolution information on motion is traditionally obtained via liquid-state nuclear magnetic resonance (NMR) relaxation (see Figure 1B). Here, the lifetime of a non-equilibrium distribution of nuclear spin states, prepared and monitored in a multitude of different ways, can be translated into information on site-specific amplitudes (model-free order parameters) and timescales of motion (Abragam, 1961; Palmer and Massi, 2006). In solid-state NMR, recoupling of anisotropic interactions can be exploited for the same purpose. For example, the (motion-modulated) size of a dipolar interaction for a nuclear pair of known distance can be probed, and recoupling leveraged by internal motion under otherwise incomplete recoupling conditions can be assessed in detail via techniques recently developed (Chevelkov et al., 2009; Schanda et al., 2010; Schanda and Ernst, 2016a; Rovó and Linser, 2017). However, whereas timescales and Lipari-Szabo order parameters (Lipari and Szabo, 1982) can be determined from NMR relaxation, there is no structural information on the states sampled in the course of such motion, information which is indispensable for understanding the biological implications of protein plasticity.

In static (frozen or otherwise solidified) preparations, where different distinct conformations can be captured, physiological protein dynamics are usually difficult to distinguish from other kinds of sample inhomogeneity. In addition, B factors in crystallography, representative of either residual motion or different solidified states, are only accessible for the most rigid protein elements and often do not reflect the room temperature distributions. Variability in single-particle cryoelectron microscopic data has been reassembled to reflect molecular motion (Fischer et al., 2010); however, significantly different shape classes (i.e., states of a biological mechanism) need to be present. At the same time, molecular dynamics (MD) simulations have seen increasing faith and popularity owing to continuous improvements of algorithms and computational capacities, together with ongoing refinement of the force fields used to describe the interatomic interactions (Lindorff-Larsen et al., 2005; Dror et al., 2012). Even though the timescales accessed in atomistic MD simulations are often too short to fully sample all relevant conformational substates, the mechanistic pictures obtained are increasingly robust and have been invaluable to explain functional features. Whereas the full set of time-dependent atomic coordinates is obtained, reduced representations are usually given, such as, e.g., root-mean-square deviations (RMSDs) of the atoms from their average position (i.e., root-mean-square

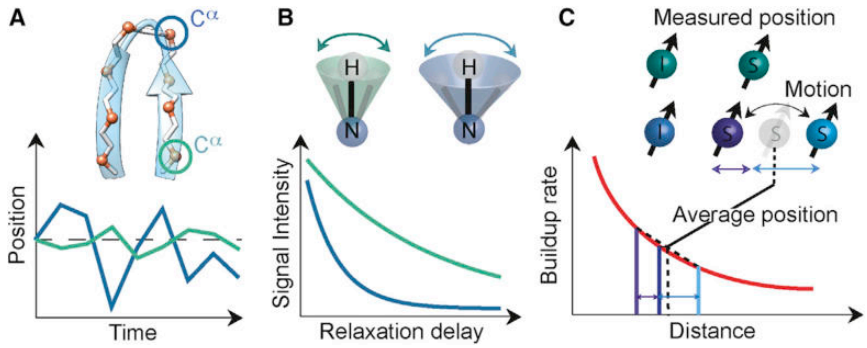


Figure 1. Complementary Methods Assessing Dynamics

(A) Simulated dynamics from molecular dynamics. Spontaneous movements are represented by displacement curves.

(B) Experimental determination of the (model-free) degree of order and timescales based on NMR relaxation. Dependent on the experiment and motional timescales, dynamics can increase or decrease relaxation effects in NMR spectroscopy.

(C) Multi-state eNOE structure calculation. The r^{-6} distance dependence of the NOE intensities leads to a single-state distance restraint usually shorter than the motional average distance. If the experimental error of distance determination is lower than the distance variation due to dynamics, this shortening can be used to decipher plasticity.

fluctuations [RMSFs]), as shown in Figure 1A. However, with its great level of detail, it is often challenging to validate MD simulations by experimental data in a direct and meaningful way.

Experimental data are time averaged and/or ensemble averaged, often sparse, and comprise statistical and maybe also systematic errors (Bonomi et al., 2017). In recent years, multiple NMR-MD hybrid approaches toward atomic resolution protein motion have been suggested, which allow the experimental determination of conformational substates in a dynamic ensemble with reasonable confidence. These approaches are, rather than on NMR relaxation as the classical method to assess dynamics, typically based on distance restraints from the nuclear Overhauser effect (NOEs), scalar couplings, and residual dipolar couplings (RDCs) (Clore and Schwieters, 2004; Lange et al., 2008; Lindorff-Larsen et al., 2005). There are two opposing philosophies with regard to the number of members of such ensembles (Bonomi et al., 2017; Ravera et al., 2016): In the maximum parsimony approach, the minimum number of structures that is sufficient to explain the experimental data is used, while in the maximum entropy approach MD trajectories or large ensembles of structures sampling the entire theoretically allowed conformational space are biased in the minimal possible way to match the experimental data. In more recent work, chemical shifts are used as directly accessible experimental information on ensemble conformational distributions, which can be incorporated into MD simulations (Camilloni et al., 2013; Robustelli et al., 2010). Other innovations also include the use of paramagnetic relaxation enhancement (Clore and Iwahara, 2009; Russo et al., 2013) or even cross-correlated relaxation rates between dipolar interactions to RDCs or combined RDC and relaxation dispersion data (Fenwick et al., 2016; Pratihar et al., 2016).

A different recent concept assessing conformational space is based on the accurate measurement of (conflicting) proton-proton internuclear distances (exact NOEs or eNOEs) (Vögeli, 2014; Vögeli et al., 2009, 2014) and relies mainly on the exponential nature of the relationship between magnetization transfer efficiency and intermolecular distance. This over-proportional relation leads to the shortening of distance restraints, and thus inconsistencies with other restraints, in the presence of conformational motions (Figure 1C). The difference between the internuclear distances of the single average structure and the precise restraints can now be exploited to yield information about the conformational space of the protein (Vögeli et al., 2016). Therefore, a number of static structures (structural states) are calculated simultaneously using the eNOE restraints averaged over these states with an r^{-6} weighting (according to the Solomon equation [Solomon, 1955]). For instance, if a non-averaged restraint would lead to an underestimated distance as depicted in green in Figure 1C, the exponentially weighted restraints resolving the conflict would lead to the two different structural states shown in purple and blue. Importantly, even though "directional dynamics" information is obtained, the structural "states" determined by eNOE-based multi-state structure calculation are governed by a reduction of distance restraint violation via averaging. Therefore, the states are not whole-protein conformers representing a local energy minimum, but local landmarks of proton residences describing the direction and amplitude of motion at given sites. Also, the timescale of the dynamics is elusive from the kind of data acquired.

SH3 domains (sarcoma homolog 3) were initially found in the oncogenic tyrosine kinase c-Src, with a cDNA closely related to the Pous sarcoma virus (Bishop, 1985). The domain is part of more than 300 human proteins in addition to eukaryotic and virus proteins, maintaining a conserved structure (Blanco et al., 1997). SH3 domains play an important role in substrate recognition, regulation of kinase activity and membrane anchoring. Hereby, SH3 binds to proline-rich sequences, in particular those carrying the PxxP motive (Saksela and Permi, 2012), forming the poly proline II helix. Whereas the dynamics of this important class of protein domains has been the target of a multitude of methodological NMR relaxation studies (reporting on amplitudes and timescales) (Asami et al., 2015; Chevelkov et al., 2009, 2010; Korzhnev et al., 2004; Rovó and Linser, 2018), the conformational space constituting their biological fitness has remained largely elusive. This is of interest particularly for amino acid side chains lining the specificity pocket, responsible for binding to interaction partners within signaling cascades. For this class of proteins, as well as for many other current targets in structural biology research, grasping the individual dynamics-function relationship will be possible only if a detailed picture of the structural dynamics is available. This includes ground-state protein structure, the timescale of site-specific dynamics, as well as the actual conformational space. Only on the basis of all three aspects can protein folding, stability, and function be understood in detail.

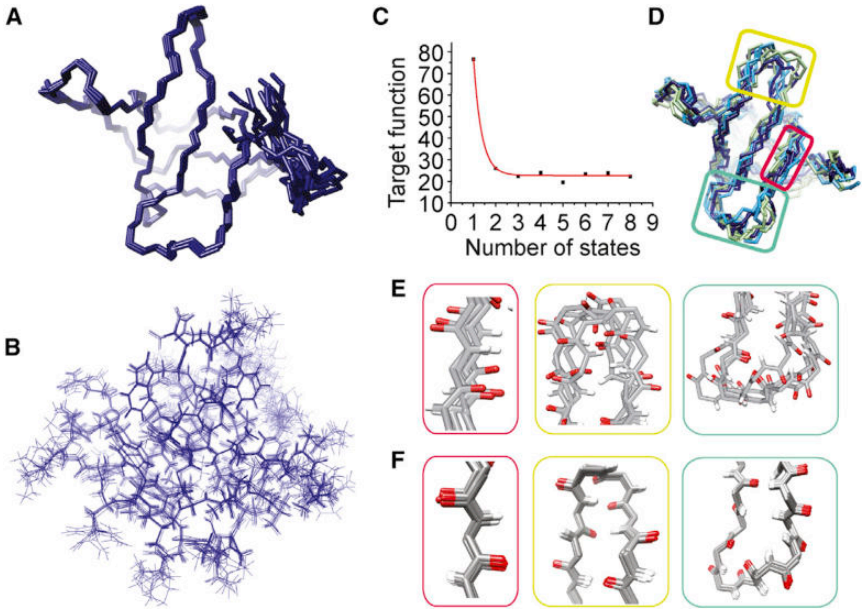


Figure 2. Single-State and Multi-State Structure Calculation of the SH3 Domain Determined via Solution NMR eNOEs

(A) eNOE single-state structure shown as a ten minimum-energy structural bundle in ribbon representation for comparison with multi-state structure calculations. Deviations due to experimental errors are small for all backbone atoms (compare Figure S2A for site-specific RMSDs).

(B) All-atom single-state structural bundle shown as wire.

(C–E) Multi-state structure calculation, represented by a sub-bundle of three conformers for each state. (C) Target function for the multi-structure approach as a function of the number of states considered. (D) Representation of the three computed states, colored in purple, green, and cyan, by sub-bundles of three structures each. The three substates from each of the minimum energy structures in a three-state structure calculation are grouped by structural similarity. (E) Enlargements of a β strand (left), the distal loop (center), and the n-SRC loop (right) of the three-state bundle showing four representative substates total for visual impression.

(F) Enlargements of the same β strand (left), distal loop (center), and n-SRC-loop (right) of the single-state eNOE structural bundle for comparison.

Here, we are deliberately selecting the three non-integrative techniques, NMR relaxation, MD simulation, and multi-state eNOE structure calculation, to compare the outcome of the individual approaches without mutual bias. Subsequently, we strengthen our comparative assessment by comparison of the individual techniques with NMR-restrained ensemble MD simulations that use the eNOE-derived distance restraints as bias potentials. We show that, together with the timescale information from regular NMR relaxation, the multi-state approach experimentally and independently validates the details of MD simulations (and vice versa). Consequently, the parallel assessment of the SH3 domain of chicken α-spectrin via these four methodological frameworks yields a consistent picture and detailed insights into protein motion, which underlines the high value of existing and future integrative approaches.

RESULTS

Here, we verify the added value of combining methods assessing and mutually validating protein dynamics from purposely independent perspectives to arrive at a multifaceted and cross-validated picture of the motional behavior of the SH3 domain of chicken α-spectrin, including both timescales and conformational space. In particular, all-atom MD simulations of the atomic motions within the global structure were combined with an experimental, eNOE-based picture of the time-independent conformational aspect of the dynamics, complemented by a set of NMR relaxation data from both solution and solid-state NMR for timescale information as well as by a hybrid eNOE/MD method. The previously described, orthogonal approaches are specifically selected here for the validation sought, which will aid devising and improving hybrid approaches for future integrative structural biology studies.

Single-State Protein Structure Calculation from Solution NMR

To obtain a baseline for multi-state eNOE characterization of the SH3 domain, we began by determining a single-state eNOE solution NMR reference structure. All restraints used can be found under BMRB: 34420, and chemical shifts and experimental details are given in Table S1 and in the STAR Methods, respectively. In contrast to conventional NOEs, where the cross-relaxation rates are estimated from the intensities of a single NOE spectrum, using the eNOE approach the magnetization transfer efficiency for structure calculation was read out from the buildup rate from a series of NOESY spectra with increasing mixing time. To account for errors due to relayed magnetization transfer through third spins and auto relaxation, the NOE mixing times were selected such that the NOE buildups were in the initial, near-linear regime. Correction of cross-peak intensities with respect to differential relaxation losses during the through-bond magnetization transfer steps was pursued using the program eNORA2 (Strotz et al., 2017). Generally, the procedures are almost identical to the multi-state calculation later on. In particular, final refinement emphasizing non-experimental information on solvent interactions, interatomic distances, and dihedral angles, was deliberately not pursued. As such, the single-state structure acts as a pure representation of site-specific tightness of the eNOE distance restraints (see Figure S2A) that define spatial distributions of protons in the multi-state ensemble. The obtained single-state structure (PDB: 6SCW) is shown in Figures 2A and 2B. Using all of these structural data, we expectedly obtain very high backbone precision of 0.06 Å and an all-heavy-atom RMSD of 0.42 Å for the structured regions. The details of multi-state structure calculation (Figures 2C–2E) are described below.

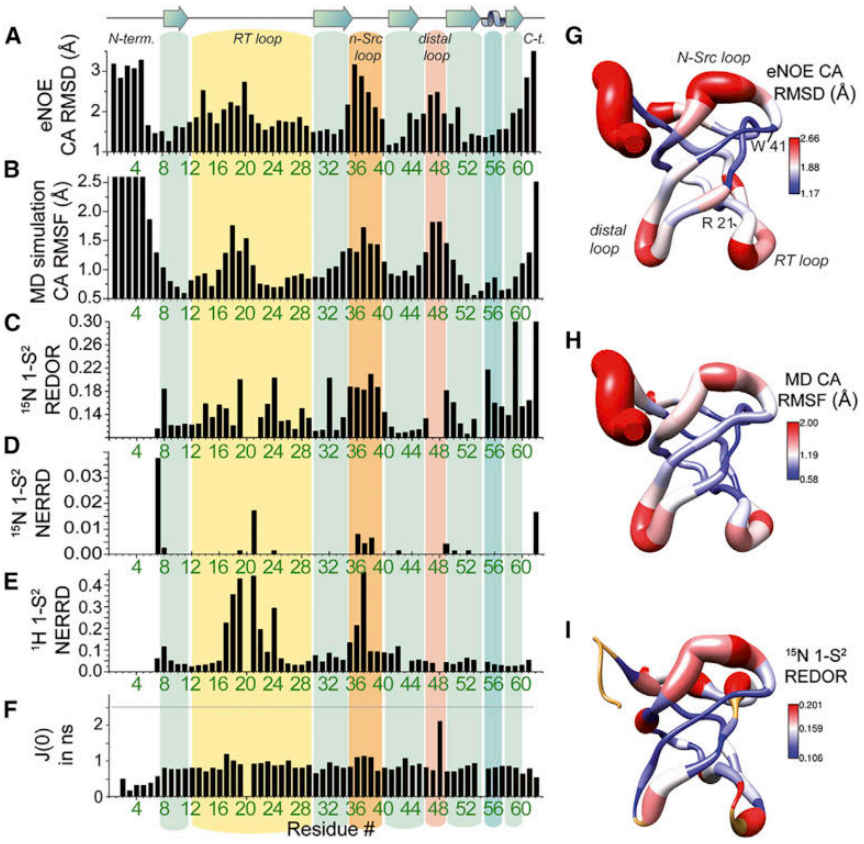


Figure 3. Complementary and Cross-Validating Assessment of Backbone Dynamics from MD Simulations, eNOE-Based Multi-State Structure Determination, Model-Free Order Parameters from Fast Magic Angle Spinning Solid-State NMR, and Solution NMR Relaxation

(A) C^{α}-RMSD relative to the average structure of all 30 substructures from a three-state eNOE structure calculation at pH 3.5. These numbers represent the conformational space of the dynamic ensemble, in contrast to the precision of the NMR structure calculation per se, which is shown in Figure S2A.

(B) C^{α} RMSF obtained from averaging over four 500 ns MD simulations with the Amber ff15ipq force field at pH 3.5.

(C) ^{15}N dipolar order parameters determined by REDOR MAS solid-state NMR at 40 kHz spinning speed, measured at 25°C and pH 7.0. This recoupling method covers a wide range of timescales.

(D) ^{15}N dipolar order parameters determined from NERRD measurements in MAS solid-state NMR. Missing bars other than for residues 20, 48, and 54 (Pro or missing signals), represent data with insignificant dispersion (Rovó et al., 2019).

(E) ^{1}H dipolar order parameters determined from proton NERRD in MAS solid-state NMR. The NERRD data cover micro- to millisecond timescale motion only (Rovó et al., 2019).

(F) Motion from solution state relaxation (covering motion faster than 4 ns only) for comparison, represented as the spectral density values at zero frequency. β sheet regions of the sequence are shaded in light green, the 3_{10} helix is shown in darker green, loops are shaded in yellow, orange, and red. RMSD and RMSF values are given in Å.

(G–I) Dynamics features depicted on the average of the eNOE-based single-state structure, shown simultaneously as ribbon thickness and via color coding. (G) Three-state eNOE-based C^{α} RMSFs as shown in (A). (H) MD-derived RMSFs as shown in (B). (I) ^{15}N dipolar order parameters as shown in (C).

Detailed Mechanistic Views from MD Simulations

Proceeding toward protein dynamics, we first performed unbiased all-atom MD simulations of the SH3 domain. These computations add *in silico* details on mechanistic aspects of protein motion and a global view on the conformational transitions, albeit within a relatively short time window. MD simulations were carried out with three different Amber force fields (and associated water models, see the STAR Methods). For each force field, two sets of MD simulations were performed, one with all Asp and Glu side chains modeled as protonated (SH3 does not contain His residues) to mimic the pH 3.5 conditions of the solution NMR experiments, and the other set with all titratable side chains assigned to their standard protonation states to mimic pH 7 conditions. For each force field/pH combination, four repeats of length 500 ns (i.e., 12 μs in total) were carried out (see the STAR Methods for details).

To break down the complex picture of the full atomistic dynamics, we computed the RMSD of every C^{α} atom from its average position over time (i.e., the RMSF). Site-specific C^{α} RMSF from the MD simulation with the ff15ipq force field at pH 3.5 are shown in Figure 3B. The results obtained with the alternative force fields and at the two different pH conditions simulated (pH 3.5 and 7) are all in close agreement with each other, see Figure S1. In line with previous NMR studies (Chevelkov et al., 2010; Rovó and Linser, 2018), the tips of the RT (17–22), n-Src (36–40), and distal loop (47–48) show higher flexibility of the backbone, while β sheet regions forming the β barrel core are more rigid. The most flexible residues, in terms of C^{α} atoms in the backbone, are Lys18, Thr37, Asn47, and Asp48, excluding the N and C termini.

Timescale-Specific Information from NMR Relaxation

Next, we set out to compare the *in silico* results from MD simulations with experimental assessments. We first involved a set of NMR relaxation and recoupling data (Figures 3C–3E). We compile here a broader selection of assessments, part of which very recent approaches (involving solid-state proton relaxation dispersion) have been described in more detail previously (Rovó et al., 2019). (Still, given the many other elements involved in this study, we limit ourselves to an exemplary subset of the available techniques.) The upper limit of the timescales assessed via regular solution state NMR relaxation is the tumbling correlation time (in this case 4 ns), overshadowing any (internal) interactions on slower timescales. In solution, we determined spectral densities at 0, $\sim\nu_H$, and ν_N frequencies, $J(0)$ being exemplarily represented in Figure 3F, from R_1, R_2, and hetNOE measurements and spectral density mapping (Farrow et al., 1995; Peng and Wagner, 1992) (see the complete data in Figure S3), in which the extent of fast motion can be assessed. Inconspicuous profiles for loop regions confirm the insensitivity to slower motion.

In the $J(0)$ data, fast motion (faster than the correlation time, i.e., picosecond to nanosecond motion) is reflected in short effective correlation times. Low $J(0)$ values show up at the N and C termini. On the other hand, intermediate timescale chemical exchange contributions to the R_2 relaxation rates (that are faster than what is eliminated by the CPMG pulse train) to some extent induce higher $J(0)$ values in the loops, most obviously in the distal loop.

$R_1\rho$ relaxation dispersion studies, which specifically characterize the slower motions, were pursued using solid-state NMR on micro-crystalline protein under magic angle spinning. As shown previously, the dynamics of the SH3 domain should be very similar in liquid and solid state, despite the experimentally inevitable pH difference between the samples (pH 3.5 in solution versus pH 7.0 in the solid state) (Chevelkov et al., 2010). However, the motion of the N terminus, involved in crystal-crystal contacts, is thought to be slowed down in the crystalline state. We involve ^{15}N as well as ^{1}H near-rotary-resonance relaxation dispersion (NERRD) order parameters (described in detail in Rovó et al., 2019) in addition to dipolar recoupling (via REDOR [Gullion and Schaefer, 1989]) order parameters (see below). In brief, in the absence of motion, the recoupling of anisotropic interactions in ^{1}H and ^{15}N NERRD takes place only at the rotor resonance condition (and at the half-rotary resonance or HORROR condition for homonuclear interactions). In the presence of motion on the timescale of the rotor period, this condition is broadened and can be assessed by $R_1\rho$ relaxation dispersion compared with analytically or numerically simulated data. ^{15}N relaxation dispersion reflects motion of the N-H bond vector on the fast-μs timescale locally and can be measured without significant influence of third spins. ^{1}H relaxation dispersion at the HORROR condition, by contrast, reflects motion of the amide proton with respect to the bulk of other protons nearby and serves a "regional" assessment of motion, where the event of local conformational exchange is also reported by residues nearby. (^{1}H NERRD cannot properly be quantified due to the strong presence of coherent interactions. As such, order parameters obtained via comparison with analytical data [in which coherent contributions are absent] are thought of as qualitatively sound but quantitatively off.) Dipolar recoupling via REDOR (Gullion and Schaefer, 1989), the third solid-state dynamics approach used, depends on the internuclear distance as well as the orientation of the internuclear vector. As such, it reflects motion in a spin pair of known distance via the obtained dipolar coupling constant, which is decreased upon motional averaging in a quantitative manner. Simulations of the dephasing curves reveal the dipolar order parameter, irrespective of the timescale of motion (as long as the dynamics are faster than the dephasing curves are recorded, usually over hundreds of microseconds).

Site-specific REDOR dephasing curves for an amide proton-back-exchanged ^{2}H, ^{13}C, ^{15}N sample of micro-crystalline SH3 are shown in Figure 3C. The very flexible residues, e.g., the N terminus, escape the solid-state NMR assessment based on cross-polarization magnetization transfer. The qualitative trends of the dipolar order parameters are consistent with the MD simulation. Exceptions are the residues Leu8, Thr32, Ala55, and Lys59, where a relatively high motional amplitude is measured, while those residues seem comparably rigid in the MD simulation, possibly due to the limited simulation timescale. Figures 3D and 3E show results from ^{15}N and ^{1}H NERRD measurements, respectively (see details in the STAR Methods). In both NERRD measurements the data match the MD data; however, in contrast to faster timescale motion, microsecond timescale motion is present only at residue Arg21, the specificity pocket site structurally adjacent to Arg21 that is affected by RT loop motion, and the N and C termini interacting with the latter via crystal-crystal contacts. Whereas strong ^{15}N dispersion is largely limited to these sites, proton NERRD as a longer-range reporter of microsecond timescale motion shows strong dispersion also for residues more distant from Arg21. On one hand, this reflects the stronger (longer-range) changes in the proton dipolar coupling network sensed by ^{1}H NERRD compared with amide bond fluctuations, the measure of ^{15}N NERRD, in highly mobile regions. On the other hand, as assessed in detail previously (Rovó et al., 2019), the RT loop is dominated by slow μs motion, the distal loop moves on a faster timescale, and the n-Src-loop covers motions on intermediate timescales. Residues 47 and 48 in the distal loop (compare the solution relaxation data) are too mobile to be detected in cross-polarization-based experiments, and residues 20 and 54 are prolines. The faster fluctuations of the distal loop clearly exceed the timescale window assessed in both NERRD experiments.

Experimental Conformational Assessment via eNOE-Based Multi-State Structure Elucidation

To complement the NMR dynamics measurements with experimental data on the spatial distribution of conformational flexibility, we pursued multi-state structure determination from ensemble-averaged eNOE restraints, reapplying the protocols established in the original literature (Vögeli, 2014; Vögeli et al., 2013, 2014). The eNOE multi-state approach as an experimental, almost exclusively distance restraint-based technique is sensitive to timescales up to the experimental time of the NOE assessment (hundreds of milliseconds if chemical shifts of different states are similar, and otherwise up to the chemical shift timescale). As such, whereas usual solution NMR relaxation measurements aiming at slow motional processes are overshadowed by molecular tumbling and only isotropic chemical shift changes can inform on micro- to millisecond motion via relaxation dispersion methods, the eNOE-based approach captures internal dynamics over a broad range of timescales in solution. First, to determine the number of structural states following the maximum parsimony principle to represent the local protein conformational space, a series of structure calculations were performed with a number of states ranging from 1 to 8 (Vögeli et al., 2012). The NOE-based multi-state procedure assesses a basis set of states needed to fulfill any conflicting distance restraints from eNOE measurements. Similar to a principal-component analysis, the number of states sufficiently representing the ensemble characteristics is an integer number identified by a ceasing improvement of the target function (see Figure 2C). The target function (Güntert and Buchner, 2015; Vögeli et al., 2016, also see the STAR Methods) represents the sum of all restraint violations. It tangentially approaches a minimum for two to three states, whereas incorporation of additional states leads to no improvement. For the calculations performed here, the target function is reduced by approximately 65% for the two-state with respect to the single-state calculation and reaches a minimum

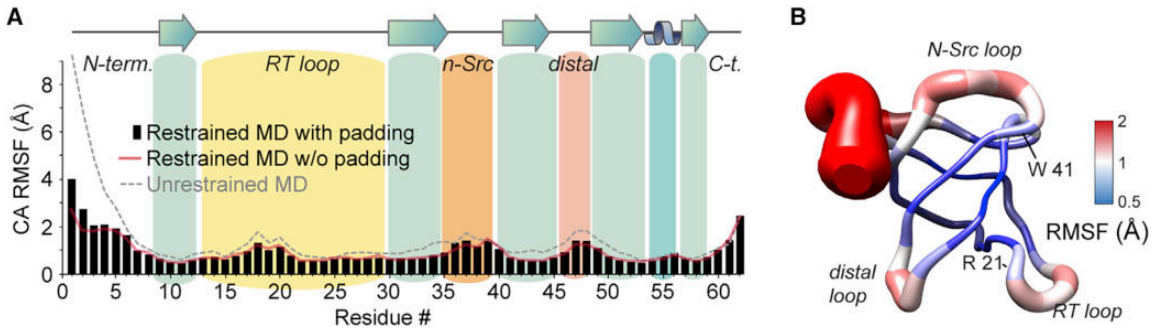


Figure 4. Comparison of Cα RMSF in the SH3 Domain in Restrained Ensemble MD Simulations with Unbiased MD Simulations

(A) RMSFs as a function of sequence, comparing restrained simulations with (black bars) or without a 1 Å padding on the eNOE distance restraints (red line) with the unbiased MD simulation (gray dashed line).

(B) RMSFs of the run with paddings shown on the SH3 domain structure. Figure S2L also includes a correlation between the restrained MD run (with paddings) and the multi-state eNOE calculation.

for three states (shown in Figure 2C). As shown earlier, this does not describe how many structural states in terms of energy minima are present and is needed only for sufficiently describing the spatial distribution of proton sites used. Since the approach does not comprise the density of data desirable for internal validation from a subset of the acquired data, left out of the structure calculation and treated separately as proposed previously (Brunger et al., 1993), the approach has been designed as a stand-alone method that is validated here externally via the MD and relaxation data mentioned above. The resulting 3 × 10 lowest-target function structures of the final three-state structure calculation, containing in total 30 structural states, were grouped together forming three sub-bundles of motional states. The deviations between the three structures, as obtained from differences in H^N, C^α, or all-heavy-atom coordinates, are shown in Figures 3A and S2B/S2C. Representing the site-specific spatial distribution as assessed by eNOE inconsistency, this parameter can be directly compared with the RMSFs from the MD trajectories. Indeed, the trends observed in these data (backbone and side-chain atoms) have very high similarity with the fluctuations observed in the MD simulations. (Figure S2L depicts a correlation between regular MD RMSFs as well as restrained MD RMSFs [see below] and the multi-state RMSD, opposed to a correlation with the *single*-state structure as the measure of mere structural precision of the calculation.) The good agreement both represents an experimental validation of the simulations (of up to hundreds of nanoseconds motion) as well as a validation of the multi-state eNOE concept and, furthermore, allows for the visualization of the conformational space sampled by the protein (see below). The loop regions show significantly higher differences of structural states than the β strand regions (see Figures 2D and 2E). The residues showing the highest sub-structural differences regarding C^α are Asp14, Pro20, Ser36-Asn38, Asn47, and Asp48. The residues Lys18 (which shows highest RMSFs in the MD simulations) and Ser19 also show strong spatial divergence in the multi-state ensemble. See Figure S4 for a three-state morph of Trp41 and Arg21, with a blue line between Lys18 and Ser19 (Grohe et al., 2019). The smallest conformational space is sampled by Trp41, Trp42, and Lys43, in line with the solid-state NMR (in particular REDOR) measurements and similar to the MD simulation. Importantly, site-specific RMSDs read out from the very well-defined single-state eNOE structure determination (Figures 2A and 2B) do *not* show the systematic increase in loop regions (Figures S2A and S2L), ruling out a data insufficiency for the described trends. However, it is important to point out that the substructures of the three-state calculation are sorted into the three-state bundles, shown in Figures 2D and 2E, by spatial similarity only. Likewise, angular information is not included the eNOE data, and TALOS dihedral angle restraints are purposely switched off in the final calculation step to reflect assessment of *spatial* plasticity in a way as unbiased as possible.

Ensemble MD Simulations Based on eNOE Restraints

MD simulations have previously been combined with experimental data in an integrative manner, where MD trajectories are biased by experimental restraints, such as, e.g., NOE data, chemical shifts, or order parameters S^2 (Lindorff-Larsen et al., 2005; Papaleo et al., 2018; Richter et al., 2007). For validation of data consistency within the set of methodology shown here, we carried out additional, restrained ensemble MD simulations that used eNOE-derived distance restraints as bias potentials (see the STAR Methods for details.) Our restrained ensemble simulation scheme consisted of carrying out parallel MD simulations of 20 replicas for 100 ns each under the influence of flat-bottom harmonic biasing potentials that penalize deviations of the ensemble average of the distances from the experimentally derived reference values (see the STAR Methods). As all force fields tested behaved highly similar in the unbiased simulations (see above and Figure S1), we chose the ff15ipq force field at pH 3.5 conditions for these restrained ensemble simulations.

Figure 4 shows the C^α-RMSF of the restrained MD ensemble in comparison with the unbiased (standard) MD described above. As expected, the eNOE-derived, average distance restraints slightly dampen the fluctuations, but overall, the unbiased MD and restrained MD ensembles show the same qualitative trends over the primary sequence. This is the case irrespective of whether the eNOE-derived distance bounds were increased by a 1 Å "padding" for the onset of the energy penalty due to the flat-bottom harmonic restraint, or whether the eNOE-derived distance bounds were used directly. Out of the 671 eNOE restraints used, 162/53 were violated on average, with a low

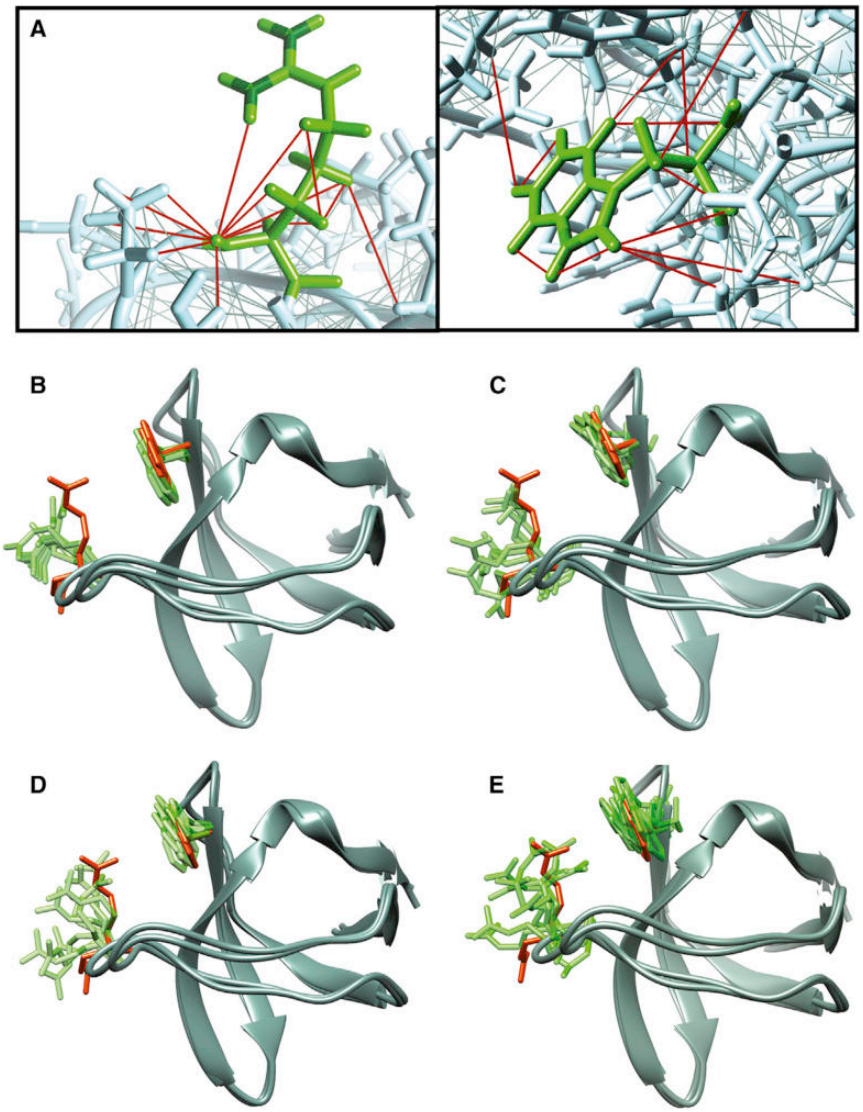


Figure 5. Experimental and *In Silico* Characterization of Side-Chain Conformational Spaces Elucidated by Ensemble-Based eNOE Assessment and MD Simulations

(A) Depiction of eNOE restraints on the residues Arg21 (left) and Trp41 (right) drawn as red lines. The distance restraints are forming several tripods that define the three-dimensional orientation of the side chains.

(B) Superposition of the X-ray structures PDB: 2NUZ and 1SEM showing binding-relevant residues Arg21 (left) and Trp41 (right: ribbon depicted in gray, side chains in orange) with the 10 lowest-target-function structures from the eNOE single-state structure calculation (ribbon, hidden; side chains, green). The Arg21 side-chain orientation of the eNOE solution structure resembles an intermediate of the X-ray structures.

(C) Superposition of X-ray structures PDB: 2NUZ and 1SEM as in (B) with representative substructures of the ensemble elucidated by eNOE three-state structure calculation, depiction as in (B). The side chain of residue Trp41, which forms an important hydrogen bond with a backbone carbonyl of the poly proline ligand, stays comparably rigid (compare Figure 3A), while the side chain of Arg21 has substantial motional freedom.

(D) Superposition of X-ray structures PDB: 2NUZ and 1SEM with frames from MD simulation (12 frames, 40 ns apart, taken from one of the four 500 ns simulations with the ff15ipq force field).

(E) Superposition of PDB: 2NUZ and 1SEM with frames from the eNOE-restrained MD simulation (10 frames, taken after 100 ns from each of the 20 replicas). Also compare Figure S4.

average violation per restraint of only 0.178/0.062 Å in the ensemble simulations without/with the additional distance padding, respectively. The RMSD from the starting structure of the simulations (i.e., the energy-minimized X-ray structure) was on average approximately 2 Å for C^{α} atoms of residues 6–61 (leaving out the flexible N-terminal residues that are not resolved in the X-ray crystal structure), and thus comparable with the unbiased MD simulations (approximately 1 Å RMSD). Taken together, the restrained ensemble, which has been deposited to the PDB-Dev: PDBDEV_00000046, agrees with our above findings and completes a consistent picture of the structural dynamics of the SH3 domain.

The Conformational Space of Side-Chain Motion

Besides information about mechanistic details of motion for the backbone, MD simulations as well as ^{13}C- and ^{15}N-edited eNOEs provide information about the conformational space sampled by side chains. This is particularly important since the side chains are responsible for biological functionality of various kinds. As described above, the SH3 domain plays an important role as a binding domain in a plethora of proteins (Teyra et al., 2017). To make use of the conformational characterization of motions in side chains, we focused on two exemplary residues in the specificity pocket, Arg21 and Trp41, which play a biologically important role due to their participation in ligand binding (Massenet et al., 2005). For all SH3:ligand interactions, the formation of an H bond between the imide-proton of a particular tryptophan side chain (in this case Trp41) and the backbone carbonyl functionality of the ligand is highly conserved. Matching experimental with simulated solid-state dynamics data has previously suggested large motion in the RT loop (around Arg21) that might play a role for binding selectivity through side-chain conformational exchange (Rovó et al., 2019): two structures, PDB: 2NUZ and 1SEM, are representative for the most open and closed states found in the PDB (Rovó et al., 2019). The side-chain conformations of Arg21 pointing toward the ligands in PDB: 2NUZ generate steric clashes while the other rotamers do not. In Figure 5B, the structural bundle of the single-state structure calculation is superimposed on the X-ray structures PDB: 2NUZ and 1SEM. The substructures of the bundle show side-chain rotamers that represent the average position between the side chain Arg21 of structures PDB: 2NUZ and 1SEM.

Figure 5C shows an excerpt of the substructures from the three-state calculation (cf. Figure 2D). Here, the side chain of Arg21 shows a larger conformational space, resembling a motion in between the open and closed state. (Note that the multi-state eNOE data only represent proton-based spatial distributions, whereas specific bond angles are poorly characterized.) By contrast, Trp41—known to form a functionally important H bond with the backbone of the ligand that is well conserved within the SH3:ligand complexes) (Teyra et al., 2017)—shows a decreased conformational space (also see Figure S2C, including the all-heavy-atoms RMSD). Figure 5D shows MD frames sampled at uniformly incremented simulation time points. Indeed, the MD simulation yields a very similar picture as the multi-structure eNOE-based ensemble (Figure 5C). Again, Arg21 widely samples conformations ranging in between the boundaries spun by PDB: 2NUZ and 1SEM, whereas for Trp41 these boundaries are considerably narrower. (The MD representation between all states is shown in Figure S4.) Figure 5E shows the respective picture from the eNOE-restrained MD run, which is again in congruency.

DISCUSSION

Experimental data are usually error prone, which requires careful validation by independent data. In past works that combine different types of data into a single model-building algorithm, Bayesian approaches have been used to assess the best-fitting models as well as their reliabilities (Rieping et al., 2005). By contrast, in this work the assessment of protein motion is pursued in principle by a deliberately non-integrative combination of orthogonal, independent techniques that yield different kinds of information. A large degree of consistency in the nature and extent of motion confirms that the data are perfectly compatible and support the same overall picture. For example, the REDOR-derived dipolar order parameters are in line with the MD-derived or eNOE multi-state pictures of plasticity, also allowing to assess whether the experimental conditions among the analyses are sufficiently consistent (which is the case). Since the conventional (unbiased) MD simulations did not draw on any experimental NMR information, the overall consistency with the various time-scale-independent NMR data is mutually reassuring. This concerns both the backbone dynamics as well as the extent of spatial features of side-chain motion from eNOE multi-state structures. This consistency is further supported by the restrained ensemble MD simulations with NOE-derived distance restraints. The detailed spatial features of protein motion (e.g., the space sampled by backbone motion and the extent of side-chain movements) obtained from these two techniques have been lacking in many previous studies, but they provide valuable information for understanding the dynamic characteristics of a protein.

The explicit comparison shows that the individual techniques each add details that are not assessable from other methods. As mentioned in more detail in the respective sections, these differences concern (1) responsiveness to all or only specific timescales (large range of timescales in REDOR and eNOE multi-state structures; fast motion in regular solution NMR data and MD; microsecond timescale motion in NERRD), (2) local versus longer-range response to motion (^{15}N and ^{1}H NERRD, respectively), (3) backbone motion only versus all-atom assessment (relaxation data and multi-state structure/MD, respectively), as well as (4) the difference between simulated detail and experimental verification.

In addition to complementarity, each of the individual techniques has flaws that can be largely alleviated in the course of the complemented assessment. All of the solid-state NMR characterization shown here lacks the residues that display the highest degree of dynamics, due to dipolar-based magnetization transfers, which fail for residues with high mobility. (A partial remedy is the use of scalar transfers for dynamics studies [Linser et al., 2010], which has been implemented in some of the recent ssNMR studies [Schanda and Ernst, 2016b].) Residues missing in dipolar experiments are the N-terminal ones as well as those in the distal loop. These residues with motion faster than the tumbling time (less than nanosecond timescales) are well characterized in the solution NMR experiments. MD simulation timescales are often too short to fully sample the relevant conformational space, and MD force fields can still be improved. Especially time constants associated with barrier crossings between conformational substates might often not be very accurate because barriers are often not included in the force field parametrization in a similar manner as the differences between free energy minima (Vitalini et al., 2015). Nevertheless, especially in combination with ensemble NMR data, there is little doubt about the value of MD simulations as a fast and robust approach to protein dynamics (Hoffmann et al., 2018a, 2018b). The eNOE-based multi-state structure calculation only offers (local) directional motion aspects based on proton-proton distance information, whereas a global view on transitions between conformers is elusive and bond angle properties are underdefined. The energy-based information on (global) motional aspects as well as local details from MD data, however, bridges this weakness of the multi-state eNOE structures. The global assessment of motional trajectories is obviously needed for correct interpretation of large-scale motion (e.g., in larger proteins), which are often important in the context of biological function.

The combined assessment of complementary features of motion, including the aspect of conformational space sampled, brings about a comprehensive, more relevant and valid picture of protein dynamics. The conclusions demonstrated in this work are derived from a rather small domain, but with increased spectrometer and computation time, similar results will be obtained for larger or otherwise more intricate molecules. These possibilities, largely facilitated by collaborative scientific efforts, are likely to be of high value for understanding aspects of protein:protein interactions, enzymatic catalysis, protein stability, and folding, which are otherwise difficult to grasp by any individual technique alone.

Here, we have demonstrated the constructive interplay of complementary techniques for assessment of protein dynamics in a small protein domain, focused on addressing the biologically important aspects of protein motion. The work, involving MD simulations, NMR relaxation/recoupling measurements, and spatial dynamics by eNOE-based multi-state structure determination principally as stand-alone techniques, shows a robust overall agreement between all methods. Particular details of the dynamics, such as global mechanistic features (MD), side-chain conformational space (MD and multi-state structures), as

well as information on timescales for fast (regular solution NMR relaxation) and slow motion (relaxation dispersion measurements), however, are brought upon only by a combination of techniques. Most importantly, the mutual validation of simulated (MD) and experimental aspects of spatial motion (multi-state eNOE structures) with model-free order parameters from NMR relaxation as well as the combination with timescale-resolved data are necessary to arrive at a complete and validated picture. On this basis, faithful mechanistic conclusions relevant for answering biological questions can be drawn. Generally, obtaining relevant dynamics information in a consistent way should make the characterization of protein motion, including timescales, energies, occupancies of exited structural states, as well as dynamic intermolecular interactions, increasingly attractive. This development toward multifaceted and collaborative approaches may gratifyingly turn structural biology from static views to (experimentally validated) moving scenarios within the next years.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Protein Expression and NMR-Spectroscopy
  - Structure Calculation
  - MD Simulations
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2020.06.001.

ACKNOWLEDGMENTS

Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the context of SFB 749, TP A13, SFB 1309, TP 03, and the Emmy Noether program, from the Verband der Chemischen Industrie (VCI, Liebig program), the Excellence Clusters CiPS-M and RESOLV, and the Center for NanoScience (CeNS) is acknowledged. This work was funded under Germany's Excellence Strategy— EXC 2033 – 390677874 and EXC 114 – 24286268. B.V. was supported by a start-up package from the University of Colorado Denver. We thank David van der Spoel (Uppsala University) for help with developing a Python script that converts experimental NOE data from NMR-STAR format into Gromacs distance restraints. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – SFB 1309 – 325871075; SFB 749 – 27112786. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder – EXC 2033 – Projektnummer 390677874; EXC 114–24286268.

AUTHOR CONTRIBUTIONS

K.G., S.P., C.H., S.M., A.K., P.R., S.K.V., and H.S. conducted the experiments. K.G., B.V., L.V.S., and R.L. designed the experiments and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 5, 2019
Revised: May 5, 2020
Accepted: June 3, 2020
Published: June 23, 2020

REFERENCES

Abragam, A. (1961). The Principles of Nuclear Magnetism (Oxford University Press).

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1-2*, 19–25.

Asami, S., Porter, J.R., Lange, O.F., and Reif, B. (2015). Access to Cα backbone dynamics of biological solids by 13C T1 relaxation and molecular dynamics simulation. J. Am. Chem. Soc. *137*, 1094–1100.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide protein data bank. Nat. Struct. Mol. Biol. *10*, 980.

Bishop, J.M. (1985). Viral oncogenes. Cell *42*, 23–38.

Blanco, F.J., Ortiz, Á.R., and Serrano, L. (1997). 1H and 15N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure. J. Biomol. NMR *9*, 347–357.

Bonomi, M., Heller, G.T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. Curr. Opin. Struct. Biol. *42*, 106–116.

Brunger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R., and Nilges, M. (1993). Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. Science *261*, 328.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.

Camilloni, C., Cavalli, A., and Vendruscolo, M. (2013). Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins. J. Phys. Chem. B *117*, 1838–1843.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

Chevelkov, V., Faelber, K., Schrey, A., Rehbein, K., Diehl, A., and Reif, B. (2007). Differential line broadening in MAS solid-state NMR due to dynamic interference. J. Am. Chem. Soc. *129*, 10195–10200.

Chevelkov, V., Fink, U., and Reif, B. (2009). Accurate determination of order parameters from 1H,15N dipolar couplings in MAS solid-state NMR experiments. J. Am. Chem. Soc. *131*, 14018–14022.

Chevelkov, V., Xue, Y., Linser, R., Skrynnikov, N., and Reif, B. (2010). Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. J. Am. Chem. Soc. *132*, 5015–5017.

Chi, C.N., Strotz, D., Riek, R., and Vögeli, B. (2015). Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals. J. Biomol. NMR *62*, 63–69.

Clore, G.M., and Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem. Rev. *109*, 4108–4139.

Clore, G.M., and Schwieters, C.D. (2004). How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? J. Am. Chem. Soc. *126*, 2923–2938.

Debiec, K.T., Cerutti, D.S., Baker, L.R., Gronenborn, A.M., Case, D.A., and Chong, L.T. (2016). Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. J. Chem. Theory Comput. *12*, 3926–3947.

Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H., and Shaw, D.E. (2012). Biomolecular simulation: a computational microscope for molecular biology. Annu. Rev. Biophys. *41*, 429–452.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. J. Chem. Phys. *103*, 8577–8592.

Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A., and Kay, L.E. (1995). Spectral density function mapping using 15N relaxtaion data exclusively. J. Biomol. NMR *6*, 153–162.

Fenwick, R.B., Schwieters, C.D., and Vögeli, B. (2016). Direct investigation of slow correlated dynamics in proteins via dipolar interactions. J. Am. Chem. Soc. *138*, 8412–8421.

Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature *466*, 329.

Grohe, K., Nimerovsky, E., Singh, H., Vasa, S.K., Söldner, B., Vögeli, B.H., Rienstra, C.M., and Linser, R. (2019). Exact distance measurements for structure and dynamics in solid proteins by fast magic angle spinning NMR. Chem. Commun. *55*, 7899–7902.

Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization. J. Magn. Reson. Ser. B *101*, 114–119.

Gullion, T., and Schaefer, J. (1989). Rotational-echo double-resonance NMR. J. Magn. Reson. *81*, 196–200.

Güntert, P., and Buchner, L. (2015). Combined automated NOE assignment and structure calculation with CYANA. J. Biomol. NMR *62*, 453–471.

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

Hoffmann, F., Mulder, F.A.A., and Schäfer, L.V. (2018a). Accurate methyl group dynamics in protein simulations with AMBER force fields. J. Phys. Chem. B *122*, 5038–5048.

Hoffmann, F., Xue, M., Schäfer, L.V., and Mulder, F.A.A. (2018b). Narrowing the gap between experimental and computational determination of methyl group dynamics in proteins. Phys. Chem. Chem. Phys. *20*, 24577–24590.

Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M., and Kay, L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. Nature *430*, 586–590.

Lange, O.F., Lakomek, N.-A., Fares, C., Schröder, G.F., Walter, K.F.A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science *320*, 1471–1475.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lindorff-Larsen, K., Best, R.B., DePristo, M.A., Dobson, C.M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. Nature *433*, 128–132.

Linser, R., Chevelkov, V., Diehl, A., and Reif, B. (2007). Sensitivity enhancement using paramagnetic relaxation in MAS solid-state NMR of perdeuterated proteins. J. Magn. Reson. *189*, 209–216.

Linser, R., Fink, U., and Reif, B. (2010). Assignment of dynamic regions in biological solids enabled by spin-state selective NMR experiments. J. Am. Chem. Soc. *132*, 8891–8893.

Lipari, G., and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J. Am. Chem. Soc. *104*, 4546–4559.

Logan, T.M., Olejniczak, E.T., Xu, R.X., and Fesik, S.W. (1992). Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. FEBS Lett. *314*, 413–418.

Massenet, C., Chenavas, S., Cohen-Addad, C., Dagher, M.-C., Brandolin, G., Pebay-Peyroula, E., and Fieschi, F. (2005). Effects of p47phox C terminus phosphorylations on binding interactions with p40phox and p67phox: structural and functional comparison of p40phox and p67phox SH3 domains. J. Biol. Chem. *280*, 13752–13761.

Palmer, A.G., and Massi, F. (2006). Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem. Rev. *106*, 1700–1719.

Papaleo, E., Camilloni, C., Teilum, K., Vendruscolo, M., and Lindorff-Larsen, K. (2018). Molecular dynamics ensemble refinement of the heterogeneous native state of NCBD using chemical shifts and NOEs. PeerJ *6*, e5125.

Peng, J.W., and Wagner, G. (1992). Mapping of spectral density functions using heteronuclear NMR relaxation measurements. J. Magn. Reson. *98*, 308–332.

Pratihar, S., Sabo, T.M., Ban, D., Fenwick, R.B., Becker, S., Salvatella, X., Griesinger, C., and Lee, D. (2016). Kinetics of the antibody recognition site in the third IgG-binding domain of protein G. Angew. Chem. Int. Ed. *55*, 9567–9570.

Ravera, E., Sgheri, L., Parigi, G., and Luchinat, C. (2016). A critical assessment of methods to recover information from averaged data. Phys. Chem. Chem. Phys. *18*, 5686–5701.

Reichel, K., Fisette, O., Braun, T., Lange, O.F., Hummer, G., and Schäfer, L.V. (2017). Systematic evaluation of CS-Rosetta for membrane protein structure prediction with sparse NOE restraints. Proteins *85*, 812–826.

Richter, B., Gsponer, J., Várnai, P., Salvatella, X., and Vendruscolo, M. (2007). The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. J. Biomol. NMR *37*, 117–135.

Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. Science *309*, 303–306.

Robustelli, P., Kohlhoff, K., Cavalli, A., and Vendruscolo, M. (2010). Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. Structure *18*, 923–933.

Robustelli, P., Piana, S., and Shaw, D.E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. Proc. Natl. Acad. Sci. U S A *115*, E4758.

Rovó, P., and Linser, R. (2017). Microsecond timescale proton rotating-frame relaxation under magic angle spinning. J. Phys. Chem. B *121*, 6117–6130.

Rovó, P., and Linser, R. (2018). Microsecond timescale protein dynamics: a combined solid-state NMR approach. ChemPhysChem *19*, 34–39.

Rovó, P., Smith, C.A., Gauto, D., de Groot, B.L., Schanda, P., and Linser, R. (2019). Mechanistic insights into microsecond time-scale motion of solid proteins using complementary ^{15}N and ^{1}H relaxation dispersion techniques. J. Am. Chem. Soc. *141*, 858–869.

Russo, L., Maestre-Martinez, M., Wolff, S., Becker, S., and Griesinger, C. (2013). Interdomain dynamics explored by paramagnetic NMR. J. Am. Chem. Soc. *135*, 17111–17120.

Saksela, K., and Permi, P. (2012). SH3 domain ligand binding: what's the consensus and where's the specificity? FEBS Lett. *586*, 2609–2614.

Schanda, P., and Ernst, M. (2016a). Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: principles and applications to biomolecules. Prog. Nucl. Magn. Reson. Spectrosc. *96*, 1–46.

Schanda, P., and Ernst, M. (2016b). Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: principles and applications to biomolecules. Prog. Nucl. Mag. Res. Sp *96*, 1–46.

Schanda, P., Meier, B.H., and Ernst, M. (2010). Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. J. Am. Chem. Soc. *132*, 15957–15967.

Schrödinger, L. (2015). The PyMOL Molecular Graphics System, Version-1.8.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Solomon, I. (1955). Relaxation processes in a system of two spins. Phys. Rev. *99*, 559–565.

Strotz, D., Orts, J., Chi, C.N., Riek, R., and Vögeli, B. (2017). eNORA2 exact NOE analysis program. J. Chem. Theory Comput. *13*, 4336–4346.

Takemura, K., and Kitao, A. (2012). Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. J. Phys. Chem. B *116*, 6279–6287.

Teyra, J., Huang, H., Jain, S., Guan, X., Dong, A., Liu, Y., Tempel, W., Min, J., Tong, Y., Kim, P.M., et al. (2017). Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities. Structure *25*, 1598–1610.e3.

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., et al. (2008). BioMagResBank. Nucleic Acids Res. *36*, D402–D408.

Vallat, B., Webb, B., Westbrook, J.D., Sali, A., and Berman, H.M. (2018). Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. Structure *26*, 894–904.e2.

van Rossum, B.-J., Castellani, F., Rehbein, K., Pauli, J., and Oschkinat, H. (2001). Assignment of the nonexchanging protons of the a-Spectrin SH3 domain by two- and three-dimensional ^{1}H-^{13}C solid-state magic-angle spinning NMR and comparison of solution and solid-state proton chemical shifts. ChemBioChem *2*, 906–914.

Vitalini, F., Mey, A.S.J.S., Noé, F., and Keller, B.G. (2015). Dynamic properties of force fields. J. Chem. Phys. *142*, 084101.

Vögeli, B. (2014). The nuclear Overhauser effect from a quantitative perspective. Prog. Nucl. Magn. Reson. Spectrosc. *78*, 1–46.

Vögeli, B., Güntert, P., and Riek, R. (2013). Multiple-state ensemble structure determination from eNOE spectroscopy. Mol. Phys. *111*, 437–454.

Vögeli, B., Kazemi, S., Güntert, P., and Riek, R. (2012). Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. Nat. Struct. Mol. Biol. *19*, 1053.

Vögeli, B., Olsson, S., Güntert, P., and Riek, R. (2016). The exact NOE as an alternative in ensemble structure determination. Biophys. J. *110*, 113–126.

Vögeli, B., Olsson, S., Riek, R., and Güntert, P. (2015a). Compiled data set of exact NOE distance limits, residual dipolar couplings and scalar couplings for the protein GB3. Data Brief *5*, 99–106.

Vögeli, B., Olsson, S., Riek, R., and Güntert, P. (2015b). Complementarity and congruence between exact NOEs and traditional NMR probes for spatial decoding of protein dynamics. J. Struct. Biol. *191*, 306–317.

Vögeli, B., Orts, J., Strotz, D., Chi, C., Minges, M., Wälti, M.A., Güntert, P., and Riek, R. (2014). Towards a true protein movie: a perspective on the potential impact of the ensemble-based structure determination using exact NOEs. J. Magn. Reson. *241*, 53–59.

Vögeli, B., Segawa, T.F., Leitz, D., Sobol, A., Choutko, A., Trzesniak, D., van Gunsteren, W., and Riek, R. (2009). Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. J. Am. Chem. Soc. *131*, 17215–17225.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Vuister, G.W., and Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in ^{15}N-enriched proteins. J. Am. Chem. Soc. *115*, 7772–7777.

Wang, L.-P., McKiernan, K.A., Gomes, J., Beauchamp, K.A., Head-Gordon, T., Rice, J.E., Swope, W.C., Martínez, T.J., and Pande, V.S. (2017). Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. J. Phys. Chem. B *121*, 4023–4039.

Zhou, D.H., and Rienstra, C.M. (2008). High-performance solvent suppression for proton-detected solid-state NMR. J. Magn. Reson. *192*, 167–172.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli BL21(DE3) Competent Cells	Novagen/Millipore	Cat# 70235-4
Chemicals, Peptides, and Recombinant Proteins		
^{15}N Ammonium Chloride	Cambridge Isotope Laboratories	Cat# NLM-467-50
^{13}C D-Glucose	Cambridge Isotope Laboratories	Cat# CLM-1396
Deposited Data		
Chicken α-spectrin SH3 crystal structure	Chevelkov et al., 2007	PDB: 2NUZ
Chicken α-spectrin SH3 single-state eNOE reference structure	This study	PDB: 6SCW
NMR chemical shifts	This study	BMRB: 34420
NMR-restrained MD ensemble	This study	PDBDEV_00000046
Software and Algorithms		
CCPNmr	Vranken et al. 2005	https://www.ccpn.ac.uk/v2-software/downloads
Cyana	Güntert et al., 1997	https://www.las.jp/english/products/cyana.html
Pymol	Schrödinger, 2015	https://pymol.org/2/
eNORA	Strotz et al., 2017	http://www.ucdenver.edu/academics/colleges/medicalschool/departments/biochemistry/Faculty/PrimaryFaculty/Pages/vogeli.aspx
Topspin 3.5	Bruker Corporation	https://www.bruker.com/products/mr/nmr/nmr-software/software/topspin
Gromacs	Abraham et al., (2015)	http://www.gromacs.org/Downloads
PROCHECK	Laskowski et al. (1993)	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/download.html
Molprobity	Chen et al. (2010)	http://molprobity.biochem.duke.edu/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rasmus Linser (rasmus.linser@tu-dortmund.de).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The datasets generated during this study are available at the BMRB (Ulrich et al., 2008), the PDB (Berman et al., 2003), and the PDB-Dev (Vallat et al., 2018) . The accession numbers for the NMR chemical shifts, the NMR single-state structure, and the NMR-restrained MD structural ensemble reported in this paper are BMRB: 34420, PDB: 6SCW, and PDB-Dev: PDBDEV_00000046, respectively.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The protein used in this study was obtained via recombinant expression in *E. coli* BL21 (DE3).

METHOD DETAILS

Protein Expression and NMR-Spectroscopy

The recombinant expression and purification of the ^{13}C and ^{15}N labeled SH3 domain of chicken α-spectrin for solution and solid-state samples was carried out as described before (Linser et al., 2007; van Rossum et al., 2001). In brief, a pET-3a vector containing the

cDNA of the protein was transformed into competent *E. coli* BL21 (DE3) cells and grown in M9 medium containing $^{13}C_6$-glucose (2 g/l) and ^{15}N NH4Cl (1 g/l) at a temperature of 37 °C to an OD of 0.6, when expression was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside. Cells were harvested after over-night expression at 22 °C and lysed using an Emulsiflex homogenizer. After purification via anion exchange on a Q-SEPH FF column, the protein-containing fractions were pH-adjusted to 3.5, concentrated, and subjected to a Superdex 75 size exclusion column in a citric acid buffer of pH 3.5. After concentration to around 500 μM, the protein was filled into a 5 mm Shigemi tube and used for NMR studies. NMR spectra were recorded at a temperature of around 25 °C on a Bruker 800 MHz Avance III spectrometer equipped with a triple-channel cryo-probe or a 1.3 mm MAS probe, respectively. For solution assignment of sidechain chemical shifts, TOCSY-based C(CC)CONH (Grzesiek et al., 1993) and H(CC)CONH (Logan et al., 1992) experiments were recorded on a ^{15}N, ^{13}C-labeled sample at pH 3.5, connecting sidechain and amide chemical shifts. In case of the H(CC)CONH experiment, MLEV17-mixing was employed, with a mixing time of 80 ms. In case of the C(CC)CONH experiment, DIPSI2 mixing was used, with a mixing time of 12 ms (assignments are shown in Table S1 and also deposited in the BMRB under accession code 34420). In total, 620 chemical shifts could be assigned of which 370 are sidechain shifts. A series of ^{15}N- and ^{13}C-resolved 3D [^1H-^1H]-NOESY-HSQC experiments were recorded using 20, 30, 40, 50, 60, and 70 ms and 20, 40, 50, 60, 70, and 80 ms of NOESY mixing time, respectively. A formula for the maximal mixing time optimal for eNOE is available on the homepage of the Vögeli group. NOE peak assignment was performed using the spectrum with longest mixing time via the semi-automated NOESY assignment function of CCPNmr (Vranken et al., 2005). In total, 689 distance restraints were obtained, from which 530 are eNOEs and 159 generic eNOEs (generic eNOEs lack a resolved diagonal signal, and a generic diagonal decay (average of the resolved diagonal signals) was employed) (Chi et al., 2015). In addition, 119 TALOS ϕ and ψ angles as well as 44 ^3J couplings were obtained. ^3J-coupling constants between H^N and H^α were determined via a quantitative J-coupling experiment (HNHA) (Vuister and Bax, 1993) with the coupling constants read out from the ratio of cross- and diagonal-peak intensities. Solution NMR T_1 and T_2 times (in the presence of CPMG pulse trains) were determined using relaxation delays of 10, 50, 100, 200 (2x), 300, 400, 500, 600 (2x), 800, 1020 ms and 10, 30, 50 (2x), 70, 90, 110 (2x), 130, 150, 170, 190 ms, respectively. hetNOE spectra were recorded with 8 s of ^{15}N d1 and no or 3 s of saturation alternatingly. ^1H-^{15}N REDOR data were recorded on a perdeuterated and 100% back-exchanged sample of SH3 of pH 7.0 at 40 kHz MAS, 800 MHz proton Larmor frequency, and roughly 30 °C in a 1.3 mm rotor via recoupling-edited MISSISSIPPI correlations (Zhou and Rienstra, 2008), using rotor-synchronized proton 180 ° pulses of 71 kHz B_1 for up to 850 μs recoupling duration, incremented from 6 to 34 rotor periods in steps of 2, and fitted individually assuming a fixed H-N distance. ^{15}N and ^1H NERRD data were recorded at a temperature of approximately 25 °C. Full experimental details of $^1H/^{15}N$ NERRD are described in Rovó et al. (Rovó et al., 2019). All spectra were processed using the program Topspin 3.6 and analyzed using CCPNmr (Vranken et al., 2005). The eNOE analysis was performed using eNORA2 (Strotz et al., 2017).

Structure Calculation

eNOE structure calculations were carried out with the program Cyana (Güntert et al., 1997) using 300 initial structures and 300000 torsion angle annealing steps for final structure calculations. Restraints were used as shown in Table S1 and as deposited in the BMRB under accession code 34420. Hereby, the dihedral angle restraints from TALOS+ (Shen et al., 2009) weighted with 1.0 during the first and the second annealing steps and with 0 during the two last annealing steps. The ^3J-HAHN coupling constants were turned off in the first annealing stage, weighted with 0.5 during the second and with 0.2 during the third and last annealing stage. Distance restraints were always weighted fully. In order to determine the number of structural states representing the molecular dynamics, 1 to 8 structural states of the entire protein were calculated simultaneously, each using 100 starting structures and 100000 torsion angle annealing steps for determination of the number of states as shown in Figure 2C. The determined eNOE distance restraints were the r^{-6}-averages of the corresponding distances in the individual states. Similarly, the measured ^3J-coupling constants are representing the arithmetic means of the individual states according to the Karplus function. By contrast, angular restraints from TALOS+ were used in the initial stage of structure calculation only and omitted for the final annealing. In order to keep the structural states as close as permitted by the experimental restraints, bundling restraints were applied between all state-structures. These distance restraints with an upper limit of 1.2 Å are weighted with 0.1 in terms of backbone heavy atoms and with 0.01 in terms of sidechain heavy atoms. The overall approach, as well as the CYANA target function, has previously been described in the literature (Güntert and Buchner, 2015; Vögeli et al., 2016).

In brief, averaged distances D are defined as:

$$D(\text{measured}) = \text{distance fitted from NOESY},$$

$$D(calculated) = [sum(k)\{R(calculated,k)^{-6}\}]^{-1/6},$$

where k runs over all members of the ensemble.

$$TF = 0 \text{ if } abs[D(\text{measured}) - D(\text{calculated})] < \text{error}$$

$$TF = w\ \{D(\text{measured}) - D(\text{calc}) - \text{error}\}^2 \text{ if } D(\text{measured}) > D(\text{calc})$$

$$TF = w\ \{D(\text{measured}) - D(\text{calc}) + \text{error}\}^2 \text{ if } D(\text{measured}) < D(\text{calc})$$

where *w* is the weight of the restraint type and *R* stands for inter-proton distances.

Weights were applied in a run-dependent manner as follows:

Iteration #: 1 2 3 4)

`anneal_weight_aco := 1.0, 1.0, 0.0, 0.0` (TALOS)

`anneal_weight_cco := 0.0, 0.5, 0.2, 0.2` (^3J)

(NOESY data are weighted with 100% in any iteration.)

The weight of each piece of the experimental data used was applied as originally established on GB3 (Vögeli et al., 2015a) and tested on a set of other systems (Vögeli et al., 2013; Vögeli et al., 2016; Vögeli et al., 2015b). eNOE multi-state structure calculations were run using 100 initial random-structures and 100,000 torsion angle annealing steps. The final 3-state structure calculation used 300 starting structures and 300 000 torsion angle annealing steps. PROCHECK and Molprobity evaluation results are shown in Table S1C.

MD Simulations

All MD simulations were carried out with Gromacs version 2019.2 (Abraham et al., 2015). The 1.85 Å resolution X-ray crystal structure (PDB 2NUZ) was used as a starting structure of the simulations, after adding the six missing N-terminal residues (Met1-Asp2-Glu3-Thr4-Gly5-Lys6) and the C-terminal residue Asp62, which were not resolved in the X-ray structure; the coordinates of these residues were taken from the solution structure (PDB 6SCW). After adding hydrogens to the protein and crystal water heavy atoms, the system was solvated with ca. 4500 water molecules in a periodic dodecahedron simulation box. The overall charge of the box was neutralized by Cl^- ions. The following three Amber protein force fields were used: ff15ipq (Debiec et al., 2016), ff99SB-disp (Robustelli et al., 2018), and FB15 (Wang et al., 2017). In the ff15ipq simulations, the SPC/E_b water model was used (Takemura and Kitao, 2012), whereas the TIP4PD-derived ff99SB-disp 4-site water model (Robustelli et al., 2018) was used in the ff99SB-disp simulations. In the FB15 simulations, the TIP3P-FB water model (Wang et al., 2017) was used. 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 (Essmann et al., 1995) with a 0.12 nm grid spacing. After steepest-descent energy minimization, the system was equilibrated for 100 ps in the NpT ensemble with harmonic position restraints on all protein heavy atoms, with force constants of 1000 kJ/mol/nm^2. Temperature and pressure were kept constant at 300 K and 1 bar, respectively, using the thermostat of Bussi and coworkers (Bussi et al., 2007) and the Berendsen barostat, respectively. The use of the SETTLE and LINCS constraint algorithms to constrain all internal degrees of freedom of the water molecules and all protein bonds, respectively, allowed to integrate the equations of motion with 2 fs time steps. Finally, the position restraints were switched off, and four individual 500 ns MD simulations were initiated from different random seeds to generate the initial atomic velocity distribution at 300 K.

For the restrained ensemble MD simulations, back-calculation of the proton-proton distances included r^{-6} averaging over the ensemble, and restraining was only applied when the calculated ensemble averages deviated from the experimentally derived ensemble averages. Thus, instead of modeling the eNOE distances as static, distance restraints were included in the force field such that the distance restraint needed only to be satisfied as an ensemble average. An MD ensemble comprised of 20 replicas was run for 100 ns in the NpT ensemble at 300 K and 1 bar (with the same thermostat and barostat as described above for the unbiased MD simulations). The eNOE-derived distance bounds were converted into distance restraints and included in the Gromacs protein topology. Flat-bottom harmonic distance restraints with force constants of 1000 kJ/mol/nm^2 were used as biasing potentials. The onset of the harmonic energy penalty was shifted by +1 Å and -1 Å for the upper and lower distance bounds, respectively, relative to the eNOE-derived distances; this ''padding'' resolves some possible ambiguities with the precise distance values and was applied previously, e.g., to determine membrane protein structures from sparse NOE restraints (Reichel et al., 2017). We repeated the restrained ensemble MD simulations also without applying this 1 Å padding. PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010) evaluation results are shown in Table S1C.

QUANTIFICATION AND STATISTICAL ANALYSIS

Backbone and all-atom conformational distributions were determined using RMSD between either bundle members (see main text for details), snapshots in MD calculations (see Method Details), or comparing outcomes of replica runs. Bundle angular properties shown in Table S1 were evaluated using PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010). Validity of individual methods is assessed by external validation using complementary technical approaches (see main text). This study does not employ significance criteria for proofing or rejecting hypotheses made.

Structure, Volume *28*

Supplemental Information

Protein Motional Details Revealed

by Complementary Structural Biology Techniques

Kristof Grohe, Snehal Patel, Cornelia Hebrank, Sara Medina, Alexander Klein, Petra Rovó, Suresh K. Vasa, Himanshu Singh, Beat Vögeli, Lars V. Schäfer, and Rasmus Linser

Supporting information for the manuscript:

Protein motional details revealed by complementary structural-biology techniques

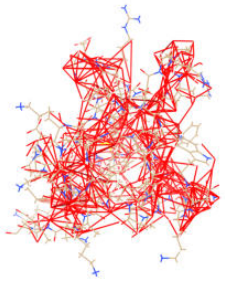
Kristof Grohe[a,b], Snehal Patel[c], Cornelia Hebrank[a], Sara Medina[a,b], Alexander Klein[a,b], Petra Rovó[a], Suresh K. Vasa[a,b], Himanshu Singh[a,b], Beat Vögeli[d], Lars V. Schäfer[c] and Rasmus Linser[a,b].

[a.] *Faculty for Chemistry and Pharmacy, Ludwig-Maximilians-University Munich, Butenandtstr. 5-13, 81377 Munich, Germany, E-mail: rasmus.linser@lmu.de*
[b.] *Faculty of Chemistry and Chemical Biology, Technical University Dortmund, Otto-Hahn-Straße 4a, 44227 Dortmund, Germany.*
[c.] *Theoretical Chemistry, Ruhr University Bochum, Universitätsstr. 150, D-44801 Bochum, Germany.*
[d.] *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, 12801 East 17th Avenue, Aurora, CO 80045, United States of America.*

Content

eNOE data, solution relaxation, statistics and structure calculation

Table S1. Restraints used for multi-state structure determination and underlying chemical shift assignments (**A**), as well as angular evaluation results (**B**), Related to STAR Methods. All restraints except of the TALOS+ restraints were averaged over all states. Angular restraints from TALOS prediction were exclusively considered for initial folding and omitted for final minimization. The top right side shows eNOE restraints for structure determination shown on a conformer of the single-state model. Assignment of backbone ^{15}N and $^{1}H^{N}$ chemical shifts (pH 3.5 and 298 K) was performed based on existing literature (Blanco et al., 1997), $^{1}H^{\alpha}$ and sidechain ^{1}H as well as C^{α} and sidechain ^{13}C were assigned using C(CC)CONH (Grzesiek et al., 1993) and H(CC)CONH (Logan et al., 1992) experiments.



A) Restraints	3-state model	Single-state model
eNOE restraints	530	530
Generic eNOE restraints	159	159
Restraints of strong H-bonds	19	19
Total number of eNOE	689	689
^{3}J-HNHA restraints	44	44
TALOS restraints	119	119

B) Shifts		H	N	HA	HB	HG	HD	CA	CB	CG	CD	
1	Met			4.06	2.04	2.47 2.35		54	31.64	29.82		
2	Asp	8.8	123.2	4.69	2.69 2.74			53.05	38.93			
3	Glu	8.58	122	4.36	2.09 1.88	2.37 2.36		55.02	27.86	32.37		
4	Thr	8.13	114.3	4.2	4.18	1.14		61.52	68.65	20.71		
5	Gly	8.36	111.5	3.89				44.25				
6	Lys	7.94	120	4.3	1.73 1.67	1.30 1.29	1.56 1.60	54.94	32.73	23.68	27.84	HE: 2.91
7	Glu	8.6	122.9	4.44	1.94 1.96	2.39 2.29		54.47	29.12	32.69		
8	Leu	8.29	123	5.33	1.27 1.59	1.26	0.72	52.33	44.05	24.31	22.17	
9	Val	9.1	111.9	5.14	1.93	0.96		56.92	35.05	22.99 19.43		
10	Leu	8.93	123.3	5.05	1.64 1.30		0.78	51.43	45.7	25.57	23.01 23.96	
11	Ala	9.02	126.9	4.53	1.58			51.48	18.46			
12	Leu	9.17	127.6	3.82	0.66 1.11		0.65	54.6	41.78	23.97	20.32	
13	Tyr	7.02	111.6	4.55	2.99 2.00		7.27	53.49	41.89			HE: 6.59 CE: 117.18
14	Asp	8.33	117.7	4.55	2.61 2.81			53.18	39.81			
15	Tyr	8.61	120.3	4.61	2.79 2.97			58.55	42.63			HE: 6.78 CE 118.18
16	Gln	7.53	127.3	4.47	1.73 1.71	2.19 2.23		52.67	28.56	32.71		HE: 6.78;7.44 NE: 112.36
17	Glu	7.99	123.2	4.07	2.17	2.21		55.08	27.46	31.34		
18	Lys	8.6	120.7	4.34	1.91 1.95	1.55 1.54	1.72.1.69	54.09	32.41	23.16	27.34	
19	Ser	7.64	115.7	4.8	3.62 4.04			56.52	64.29			
20	Pro			4.5	1.86 2.38	1.91 2.07	3.72	63.85	31.09	26.41	50.91	
21	Arg	7.66	113.7	4.51	1.86 1.58	1.33 1.31	2.92 2.93	55.13	29.73	26.46	42.33	HH:7.09
22	Glu	7.77	121	5.42	2.17 2.08	2.64 2.63		54.09	32.13	34.7		
23	Val	7.38	114.4	4.5	1.68	0.61 0.59		58.84	33.99	20.04		
24	Thr	7.46	119.3	4.98	3.94	1.27 1.27		60.24	69.32	21.5		
25	Met	9.46	121.8	4.84	2.51	2.11 1.82		53.1	35.25	30.28		HE: 2.1

26	Lys	8.5	123.8	4.81	1.61 1.65	1.38 1.35	1.63	52.69	33.61	23.57	27.77	
27	Lys	8.88	122.7	3.2	1.53 1.43	1.03		57.46	31.3	23.37	28.62	
28	Gly	8.79	115.6	4.39 3.45				43.67				
29	Asp	8.42	121.3	4.47	2.52 2.81			53.89	39.69			
30	Ile	8.09	120.7	4.92	1.73	0.81	0.81	59.06	37.36	27.14 18.03	17.58	
31	Leu	9.25	127.2	4.92	1.53 1.37	1.17		52.63	42.69	26.14		
32	Thr	8.39	117.2	4.53	3.99	1.08 1.06		62.18	68.61	21.55		
33	Leu	8.98	129	4.31	1.75 1.93	1.07	0.63	53.79	42.13	24.26	22.58	
34	Leu	8.98	126.3	4.45	1.37 1.08	1.54	0.26	53.93	41.4	24.66	20.81	
35	Asn	7.55	114	4.68	2.74 2.60		7.14 7.90	53.22	39.89			ND: 115.04
36	Ser	9.06	123.5	3.9	1.97 2.77			55.96	61.11			
37	Thr	8.13	115.4	3.9	4.2	1.24		64.48	68.33	21.18		
38	Asn	8.66	122.6	4.78	2.84 3.67		7.90 7.26	52.29	39.47			ND: 114.53
39	Lys	8.44	120.9	4.18	1.74 1.76	1.39 1.37	1.6	57.5	32.02	23.12	27.91	
40	Asp	8.21	114.4	4.5	2.26 2.64			53.89	41.52			
41	Trp	8.1	123	5.02	2.89 2.74		7.02	55.04	31.22		127.3	HE: 10.00;7.02 HZ: 7.31;6.55 HH:7.16 CE: 119.34 CZ3: 120.48 CH2: 124.28 NE: 128.84
42	Trp	9.29	124.8	5.47	2.90 2.74			53.1	30.46			HE: 6.63;9.23 HZ: 7.14 HH:7.47 CE: 119.89 CZ: 120.48 CH2: 125.27 NE: 129.25
43	Lys	8.84	124.2	4.32	1.06 1.46	0.93 0.95	1.29	54.29	33.00	24.44	27.83	
44	Val	9.33	122.1	5.32	2.03	0.76 0.71		57.98	35.86	18.13 20.73		
45	Glu	8.59	118.6	5.33	1.89 1.84	2.18 2.20		53.1	31.52			
46	Val	8.83	125	4.44	2.05	0.99 0.91		59.63	32.61	20.12 18.03		
47	Asn	9.37	126.6	4.25	2.81 2.97		7.58 6.97	54.23	36.41			ND: 113.73
48	Asp	8.66	112.5	4.42	2.93 2.98			53.62	37.54			
49	Arg	8.12	120.5	4.66	1.83 1.85	1.65 1.68	3.26 3.23	54.04	31.31	26.16	42.14	
50	Gln	8.46	118.7	5.48	1.79 1.77	2.17 2.17		52.38	30.57	32.57		HE: 7.17;6.65 NE: 110.33
51	Gly	8.57	107	3.89				44.38				
52	Phe	9.18	119.1	5.57	3.08 2.55		7.08	57.41	41.2		130.92	HE: 7.38 CE: 131.00
53	Val	9.01	111.1	4.73								
54	Pro			3.49	1.22 0.71		3.72	60.79	29.3		50.87	
55	Ala	7.41	129.2	2.6	0.15			53.51	15			
56	Ala	7.78	113.4	3.94	1.12			52.07	17.3			
57	Tyr	7.66	116.1	4.66	2.94 3.30		6.64	55.12	36.85		133.8	HE: 6.83 CE: 118.31
58	Val	7.35	111.1	5.46	1.82	0.68 0.68		57.47	34.92	18.67		
59	Lys	8.54	118.6	4.77	1.68 1.65	1.36 1.38	1.62	53.12	35.8	23.00	28.2	
60	Lys	9.17	125.8	4.4	1.79 1.97	1.3		57.47	32.01	24.93		
61	Leu	8.43	125	4.39	1.52 1.49		0.78	53.8	41.8	24.84	22.03	
62	Asp	7.97	122.8	4.45	2.88 2.87							

C)		eNOE multi-state	Standard MD	Restrained MD
PROCHECK*	Most favored regions	49.1 %	81.8 %	96.4 %
	Additionally allowed regions	43.6 %	14.5 %	3.6 %
	Generously allowed regions	7.3 %	1.8 %	0 %
	Disallowed regions	0.0 %	1.8 %	0 %
MolProbity*	Favored regions	63.3 %	90.1 %	91.1 %
	Allowed regions	92.2 %	98.4 %	99.2 %
	Outliers	7.8 %	1.6 %	0.8 %

***Angle evaluation of both, MD snapshots as well as the eNOE multi-state structure calculation will not result in the high scores usually obtained from ground-state structures as in X-ray coordinates since the motion itself (including higher-free energy transitions) is explicitly represented. For the eNOE multi-state structure calculation in particular, only the NOE-derived space of probable residence is obtained, which does not equate real structural states, and specifically sidechain angular properties cannot be deduced at all from the experimental data employed.

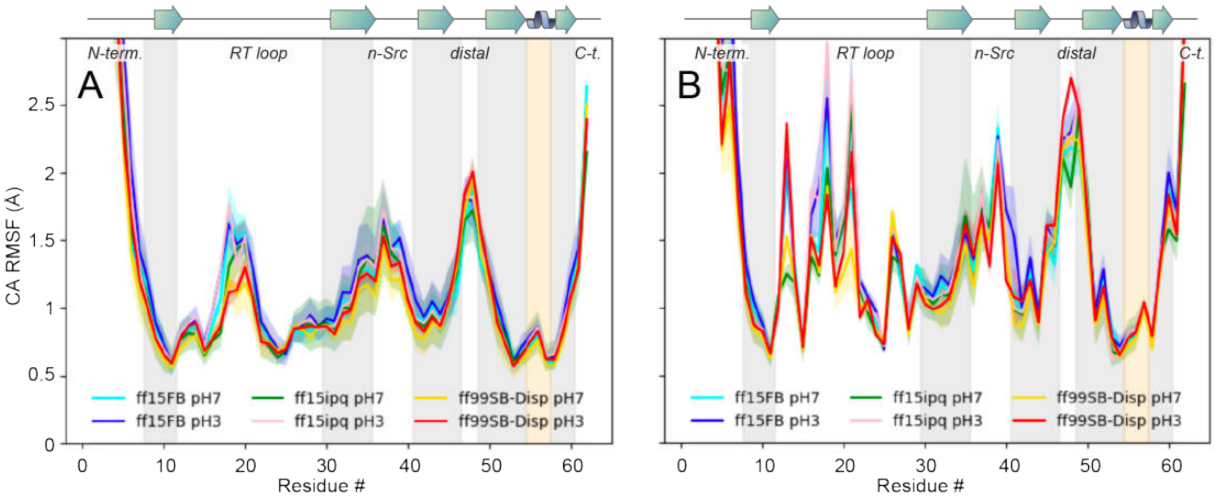


Figure S1. Results of MD simulations with different force fields, Related to STAR Methods. **A)** Cα RMSF plot of the SH3 domain for the ff15ipq, ff15FB, and ff99SB-disp force fields at pH 3.5 and pH 7, respectively. **B)** RMSF calculated for all heavy atoms including side chains of the SH3 domain for the ff15ipq, ff15FB and ff99SB-disp force fields at pH 3.5 and pH 7, respectively. The shaded areas indicate the standard deviations of the four 500 ns simulations carried out for each force field/pH.

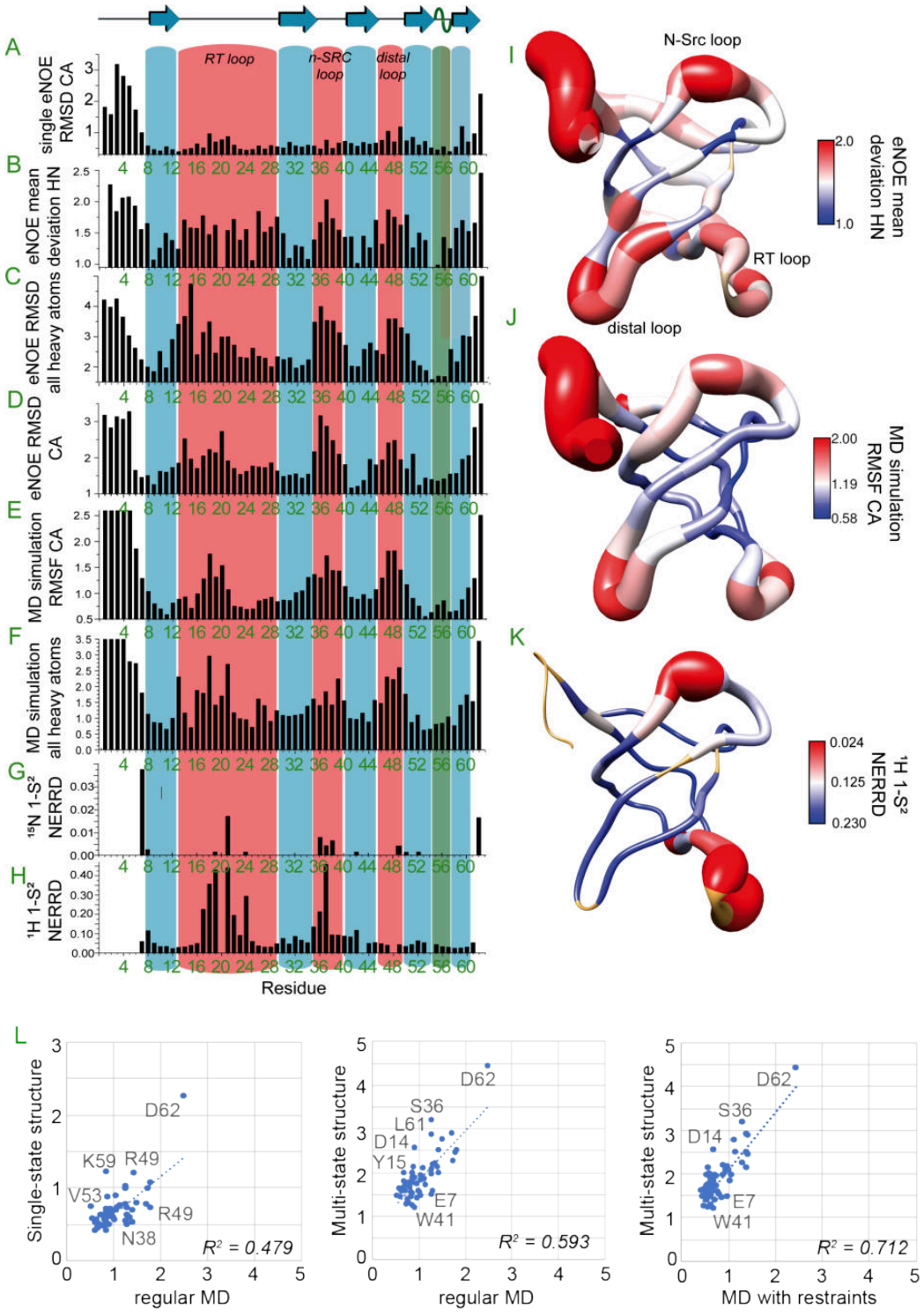


Figure S2. Extended comparison of the data from eNOE-based structure determination, order parameters from fast-magic-angle-spinning solid-state NMR and data from MD-simulations, Related to Figure 3. **A)** Site-resolved RMSD of C^{α} within the 10 minimum target function structures of the *single-state* eNOE structure calculation at pH 3.5. The bar plots **B-D** depict the site-specific mobility as determined by *three-state* structure determination at pH 3.5. The 10 lowest-target-function structures including three states each were combined, and the average structure from all 30 substructures was determined. **B)** Site-specific average deviation of backbone proton positions of the *three-state* bundle to the average structure. **C)** RMSD of all heavy atoms including side chains of the *three-state* bundle relative to the average structure. **D)** C^{α}-RMSD of the *three-state* bundle relative to the average

structure. **E)** C^{α}-RMSF determined by molecular dynamics simulations. **F)** RMSF of all heavy atoms including side chains determined by MD simulations as in E. **G)** ^{15}N dipolar order parameters determined by NERRD MAS solid-state NMR. This relaxation method covers fast microsecond timescale motion. **H)** ^{1}H dipolar order parameters determined by NERRD MAS solid-state NMR. This relaxation method covers slow microsecond timescale motion. **I-K** depict the dynamics measured by the different techniques, visualized on the average (single-state) SH3 structure, both color-coded and as ribbon thickness. **I)** Backbone proton displacement as shown in B. **J)** RMSF of the C^{α} atoms determined by MD simulations as shown in E. **K)** ^{1}H effective order parameters determined by NERRD as shown in H. Structured regions of the sequence are shaded in blue, mobile loops are shaded in red. **L)** Correlation of (regular) MD C^{α} RMSFs with the eNOE *multi-state* RMSDs, reflecting the conformational ensemble in a distinctive but comparable way (middle), compared with the correlation between (regular) MD C^{α} RMSFs and eNOE *single-state* RMSDs (left), which is a measure of structural precision, as well as with a correlation in which the restrained ensemble MD simulation is used (right). All values in Å, fully flexible residues 1-6 are not shown.

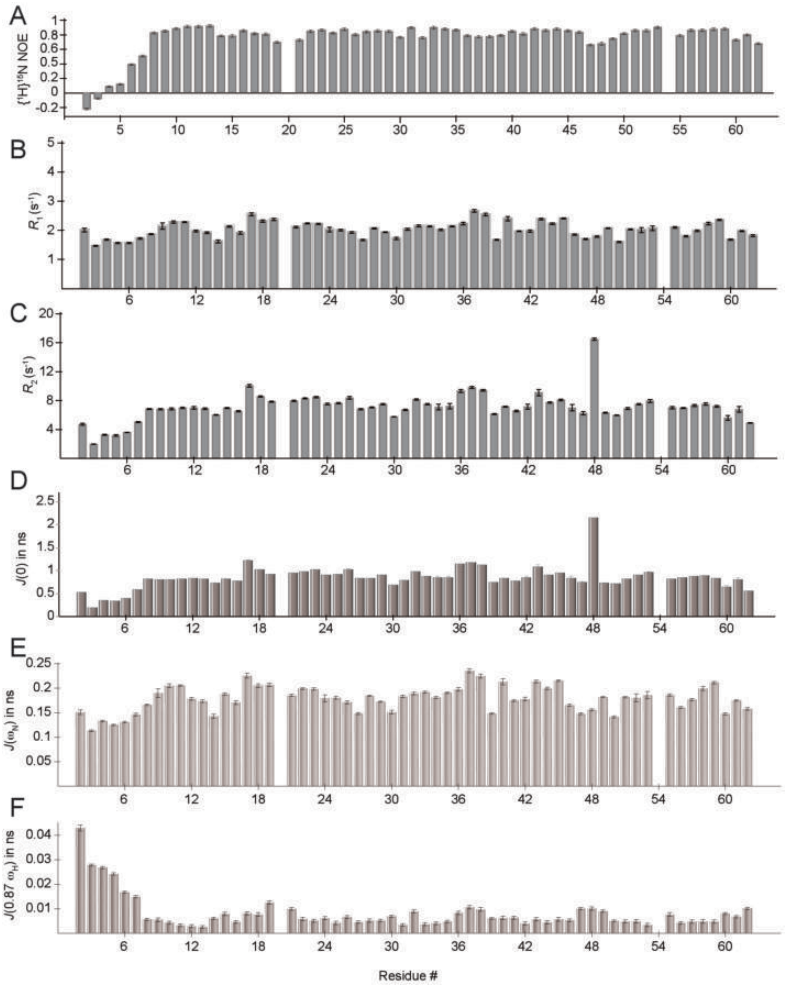


Figure S3. Regular solution state ^{15}N relaxation experiments and spectral density mapping as shown in main text Fig. 3F for comparison with solids data, Related to STAR Methods. **A)** Heteronuclear NOE, **B)** R_1, **C)** R_2, **D)** $J(0)$, **E)** $J(\omega_N)$, **F)** $J(0.87\omega_H)$. Relaxation data were recorded on a 1 mM SH3 domain sample at 800 MHz Larmor frequency at roughly 25 °C and pH 3.5. See experimental details in the Methods section of the main text.

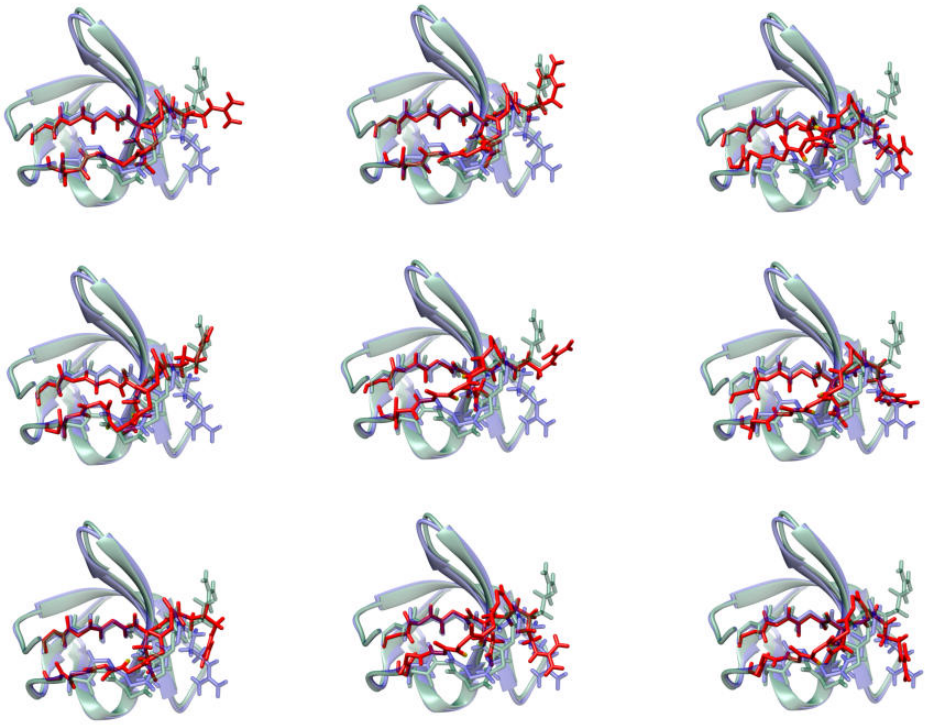


Figure S4. Snapshots from an NMR-derived molecular movie created by merging a three-state eNOE ensemble containing 10 conformers for each of the 3 sub-bundles (see details in the main text Methods section), Related to Figure 5. The structures in green and blue are showing the X-rays structures 2NUZ and 1SEM, representing different states within the dynamics according to conformations found in the pdb databank (Rovó et al., 2019). The red loop structure represents the morph from the eNOE three-state bundle. The three-state eNOE structure shows large-scale motion of Arg21.

Manual for multi-state structure calculation using Cyana

This manual is written for using the eNOE restraints as determined by the MATLAB-based version of eNORA2[42] as an input for Cyana[83] equipped with multi-state structure calculation scripts[103].

The upper and lower limit of the distance restraints have to be provided as separate files with the file type .upl and .lol fulfilling the format:

```
2  ASP    H       1  MET   HA            3.54
2  ASP    H       2  ASP  HB2            4.43
2  ASP    H       3  GLU  HB3            5.46
3  GLU    H       2  ASP  HB3            5.13
```

The input file for the ^3J HAHN coupling restraints needs to have the following format. In the third column the coupling constant is listed, the last three columns are the Karplus coefficients in the order C, B, A. Note that the order of Karplus coefficients is often reverse.

```
2 ASP  H      2 ASP  HA     7.10309      0.23875      1      0.111  -1.222 8.754
3 GLU  H      3 GLU  HA     7.35598      0.11675      1      0.111  -1.222 8.754
4 THR  H      4 THR  HA     15.53475     0.52979      1      0.111  -1.222 8.754
6 LYS  H      6 LYS  HA     7.54164      0.11755      1      0.111  -1.222 8.754
```

Dihedral angle restraints from TALOS can be used as follows:

```
3      GLU    PHI    -154.875      -115.715
4      THR    PHI    -166.946      -83.946
7      GLU    PSI    100.51 139.04
8      LEU    PSI    117.331       174.037
```

After preparing the input files and copying them into the same folder as the scripts “CALC_multistate.cya” and “PREP.cya”, from where Cyana is executed, one has to modify the PREP as explained in the following.

In this case the PREP-script prepares the restraints for being averaged over the *n* states of the structure calculation. An important information is that TALOS restraints are not averaged, therefore they should be omitted in the final stage of structure calculation.

The very first part of the script determines the number of states which are calculated:

```
syntax nbundle=@i=3 togetherweight=@r=0.01 multitensor

multitensor=.true.
together=.true.
moloffset=100
```

where `nbundle=@i=3` defines the number of states, in the displayed case 3, and `moloffset=100` defines the numbering offset for the different sub-structures. For instance, in case of a 3-state calculation of SH3, the first state would have the residue numbers of 1-62, the second 101-162 and the third 201-262.

The next part defines the input files and how they are averaged. The sequence file has a common format and is given as:

```
# ------ Sequence file ------

read seq name_of_sequence_file.seq
```

The next input file refers to the TALOS restraints as shown above

```
# ------ Make bundle angle restraints ------

read aco name_of_TALOS_input.aco
```

The ^{3}J HAHN coupling restraints can be introduced next:

```
# ------ Make bundle coupling constant restraints ------

read bundle.seq
read cco name_of_coupling_constant_input_file.cco
```

If RDC restraints are available (here commented out, because no RDC restraints are available):

```
# ------ Make bundle RDC restraints ------

#read bundle.seq
#read rdc name_of_rdc_input.rdc
```

In the next part of the script the eNOE input will be defined:

```
# ------ Make ambiguous bundle distance restraints ------

#subroutine PURGE
     #distance delete "HA 9, HB2 9"
#end


init

read upl name_of_ upper_limit_input_file.upl

#PURGE
#distance modify info=full
```

…

```
init

read lol name_of_lower_limit_input_file.lol
```

As described above, this manual is dedicated for using the MATLAB-based version of eNOE2 prior to Cyana. Instead, the newest version of the program also provides an inbuilt eNORA. When using ready-to-use eNOE-restraints, the `subroutine PURGE` and distance modifications should be commented out.

The last part of the script is about making the bundling restraints, which hold the structural states together during structure calculation. Generally, nothing has to be modified. The upper limit of the bundling restraints is given as 1.2 Å, which can be changed by the user:

```
${$rnum(iar(i))+moloffset}" upl=1.2 weight=$togetherweight info=none
    end if
```

The sidechain atoms are down-weighted by a factor of 10, which can be also modified:

```
  distances set "* - N CA C CB, * - N CA C CB" weight=weight*0.1
```

After all user-defined values of the script PREP are set, the script “CALC_multistate.cya” has to be adjusted. The very first part defines the number of cores (`nproc = number_of_cores`) which are used for computation.

```
syntax inputseed=@i=3771
nproc = 8

if (master) then
PREP
end if
```

In the next part the input files, as written by the PREP-script, are defined. The names should not be modified. If the restraints are not available, comment them out.

```
# ------ Structure calculation ------
read upl bundle.upl
read lol bundle.lol
read aco bundle.aco
read cco bundle.cco
#read rdc bundle.rdc
if (existfile('together.upl')) read upl together.upl append
```

In the following part, the weighting of the restraints above during the four stages of simulated annealing is defined.

```
anneal_weight_rdc := 0.0, 0.0
anneal_weight_aco := 1.0, 1.0, 0.0, 0.0
```

```
anneal_weight_cco := 0.0, 0.5, 0.2, 0.2
```

In this example, RDC restraints are weighted always 0, because they are absent. The TALOS-restraints are weighted fully during stage one and two and are omitted during the last two stages. The ^{3}J coupling restraints are weighted with 0 during the first, with 50 % during the second, and with 20 % during the third and fourth stage. The eNOEs are always weighted fully. As mentioned above, TALOS restraints are not averaged over the states and should only be used for the initial fold in order to save computational time.

The next lines define the number of starting structures and the number of annealing steps, respectively.

```
seed=inputseed
calc_all 300 steps=300000
```

The last part of the script defines the evaluation of the calculation as printed in the output file

```
if (master) then
  cut_cco=1.0
  cut_rdc=3.0
  weight_aco = 0.0
```

This provides the cutoff value below which the restraint violation is not printed in the output file

```
rmsdrange:=residue-residuey
  overview bundle structures=10 pdb

  read pdb bundle.pdb
  rmsdrange:= residue-residuey
  overview bundleSec structures=10 pdb # reference=eNOE190115_3.pdb
```

2.2.3 A false-positive screening hit in fragment-based lead discovery: Watch out for the Red Herring

An example of a fruitful collaboration between crystallographic and NMR research groups is shown here. The two methods complement each other, and by their combination, biologically important results can be obtained, pushing forward pharmaceutical research.

In this work, Cramer and Schiebel et al. carried out a comparative study on fragment screening on the aspartic protease endothiapepsin (see Chapter 1.2.1). Fragment-based drug discovery is one of the essential pillars of pharmaceutical research. [104] After the detection of a screening hit obtained by using different methods, the X-ray structure determination revealed a molecule binding to the protein, which was different from the initial fragment used for screening. After reproducing the possible reaction mechanism that converts the fragment into the molecule found to be bound to the protein, Cramer and Schiebel turned to us for further NMR-based investigations. By reaction monitoring of the compound in the absence and presence of endothiapepsin by using 1D-NMR and NOESY, followed by NMR-structure determination, we were able to prove parts of the postulated reaction mechanism. Furthermore, by detecting the predominant NOE polarization-transfer mechanism in the presence of endothiapepsin to be zero-quantum, we have confirmed protein binding in solution for the initial screening compound as well as for the last intermediate of the reaction mechanism.

False-positive screening hits can lead to an enormous loss of time and resources. This work examines such false-positive hits in detail and demonstrates the need for structure determination to be involved in the process in order to prevent misleading results.

Drug Discovery

International Edition: DOI: 10.1002/anie.201609824
German Edition: DOI: 10.1002/ange.201609824

A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring

*Jonathan Cramer[+], Johannes Schiebel[+], Tobias Wulsdorf, Kristof Grohe, Eszter Eva Najbauer, Frederik R. Ehrmann, Nedyalka Radeva, Nina Zitzer, Uwe Linne, Rasmus Linser, Andreas Heine, and Gerhard Klebe**

Abstract: *With the rising popularity of fragment-based approaches in drug development, more and more attention has to be devoted to the detection of false-positive screening results. In particular, the small size and low affinity of fragments drives screening techniques to their limit. The pursuit of a false-positive hit can cause significant loss of time and resources. Here, we present an instructive and intriguing investigation into the origin of misleading assay results for a fragment that emerged as the most potent binder for the aspartic protease endothiapepsin (EP) across multiple screening assays. This molecule shows its biological effect mainly after conversion into another entity through a reaction cascade that involves major rearrangements of its heterocyclic scaffold. The formed ligand binds EP through an induced-fit mechanism involving remarkable electrostatic interactions. Structural information in the initial screening proved to be crucial for the identification of this false-positive hit.*

Fragment-based approaches are very efficient in the generation of new lead molecules required for the design of drug candidates. However, the screening methods that are commonly applied to identify promising hits for subsequent X-ray crystallographic experiments can give misleading results. The identification of false-positives, often caused by so called Pan Assay Interference Compounds (PAINS),[1] is a major challenge in any library screening program. It has been shown that the molecular mechanisms for the generation of false-positive signals include unspecific binding, colloidal aggregation, inherent reactivity, and interference with the assay detection method.[2,3] Additionally, the presence of impurities from synthesis or degradation of the compound can give misleading results.[4] The pursuit of a false-positive binder can lead to a significant loss of time and resources. Therefore, the identification of such "red herrings" has become a major concern in library design.

Recently, we performed an extensive screening of a fragment library containing 361 compounds against the aspartic protease endothiapepsin (EP).[5,6] In this context, we compared the outcome of various state-of-the-art screening methods. Additionally, the entire library was subjected to a crystallographic screen with subsequent characterization of the hits by isothermal titration calorimetry (ITC). Compound **1** (fragment number 177) emerged as one of the most potent binders across multiple screening methods. It was found among the top ten fragments in four out of six and as a hit in five out of six screening techniques, showing virtually complete inhibition of the enzyme in two different biochemical assays, the highest thermal shift of 3.4 ± 0.3 °C, and a K_d of 115 ± 8 μM by ITC.[5,6] Importantly, during the selection of molecules for the STD-NMR experiments, the fragment passed a filter process based on purity, aggregation behavior, and solubility criteria that is typically applied in pharmaceutical companies.[6] Moreover, **1** is not predicted to be a PAINS compound when using the approach of Baell and Holloway.[2] Therefore, **1** seemed to be an attractive starting point for fragment-based lead optimization.

With our established crystallographic method, we were able to collect a dataset for **1** in complex with EP to a resolution of 1.25 Å. To our surprise the ligand, which was bound near the active site of the enzyme, showed little resemblance to the original compound (**1**) in four independent crystallization experiments. The electron density clearly indicates the presence of a much larger molecule accommodated in the binding cleft (Figure 1 A). This molecule features a tricyclic core substituted with a bicyclic and a monocyclic moiety. The characteristic methyl substitution pattern and the unambiguous presence of a chlorine atom, as indicated by its anomalous scattering, suggest a chemical relationship to **1**. Fortunately, the high resolution of the crystallographic model, hereafter referred to as EP–**2**, allowed us to determine the atomic connectivity of the unknown binder **2**.

Supplier specifications, as well as our own HPLC and NMR experiments, excluded the possibility of an impurity in the sample that could account for the crystallographic results. Therefore, we proposed that **2** is generated from **1** under the applied assay conditions. To verify this hypothesis, we

[*] J. Cramer,[+] Dr. J. Schiebel,[+] T. Wulsdorf, F. R. Ehrmann, N. Radeva, N. Zitzer, Prof. Dr. A. Heine, Prof. Dr. G. Klebe
Institut für Pharmazeutische Chemie, Philipps-Universität Marburg
Marbacher Weg 6, 35032 Marburg (Germany)
E-mail: klebe@staff.uni-marburg.de

K. Grohe, E. E. Najbauer, Prof. Dr. R. Linser
Abteilung für NMR-basierte Strukturbiologie
Max Planck Institut für Biophysikalische Chemie
Am Faßberg 11, 37077 Göttingen (Germany)

K. Grohe, Prof. Dr. R. Linser
Fakultät für Chemie und Pharmazie
Ludwig-Maximilians-Universität München
Butenandtstr. 5–13, Haus F, 81377 München (Germany)

Dr. U. Linne
Fachbereich Chemie, Philipps-Universität Marburg
Hans-Meerwein-Straße 4, 35043 Marburg (Germany)

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/anie.201609824.

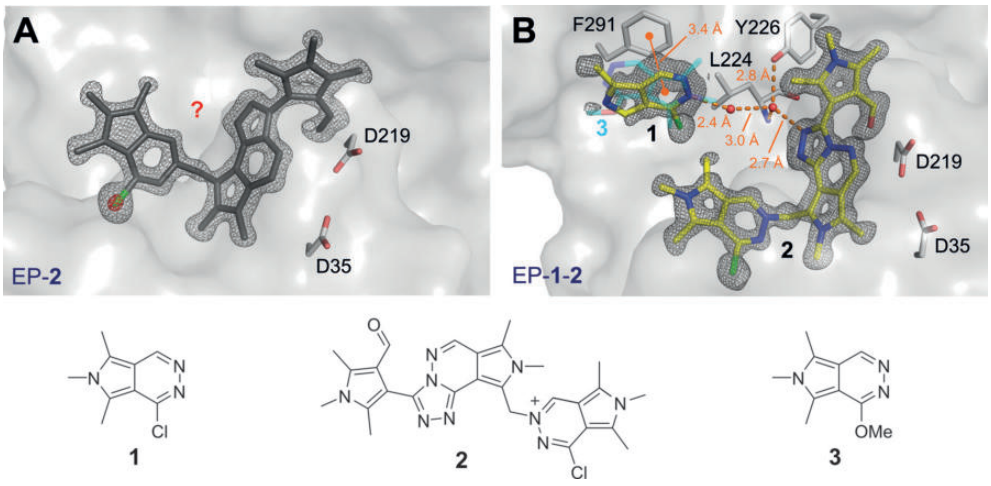


Figure 1. Crystal structures of EP in complex with different reaction products of **1** (also see Figure S2). The mF_o-DF_c electron densities for bound ligands are depicted as gray mesh at the 3σ level [2.8σ for **1** in (B)] prior to the inclusion of each respective ligand into the model. A) Structure of an unknown contaminant in complex with EP from the crystallographic screening campaign (EP–**2**). The observed electron density clearly cannot be explained by the chemical formula of **1**, but instead reveals the presence of a larger molecule that is characterized by the connectivity defined by gray sticks and contains a chlorine atom as indicated by the anomalous signal shown in red (5σ level). B) Structure generated by soaking an EP crystal with the HPLC isolate $m/z = 196.1$ (EP–**1**-**2**). The specific interaction of **1** with its environment is highlighted in orange while the methoxy analogue of **1** (molecule **3**) is shown with its primary binding mode from a superimposition of the EP–**3** structure onto the EP–**1**-**2** structure in transparent cyan (also see Figure S2A).

monitored a solution of **1** in the assay buffer by HPLC. This experiment clearly showed the time-dependent formation of multiple species in the reaction mixture. Since some material precipitated from aqueous solution, we repeated the experiment in methanol at 50 °C, which gave equivalent results (Figure S1A in the Supporting Information). Mass spectrometric analysis revealed molecules with *m*/*z* values of 192.3, 196.1, 337.3, 355.3, and 533.1 as the main components of the mixture. Subsequently, these compounds were isolated by preparative HPLC and subjected to HR-MS, MS/MS, NMR, and protein X-ray crystallography, leading to the assignment of putative structures (Scheme 1).

m/z 192.3: MS analysis indicated the absence of a chlorine substituent and thus allowed the identification of **3** as the alcoholysis product of **1**. This finding suggests that the chloropyridazine heterocycle in **1** is prone to react in a nucleophilic substitution reaction. Remarkably, the methoxy derivative **3** was found to bind to EP. Four copies of the molecule, two of which had sufficient occupancy to assign them to the electron density in the crystal structure, bind on top of Phe291, which is known to be a binding hotspot in EP,[7] in a 5-fold stacking arrangement (Figure S2A, structure referred to as EP–**3**).

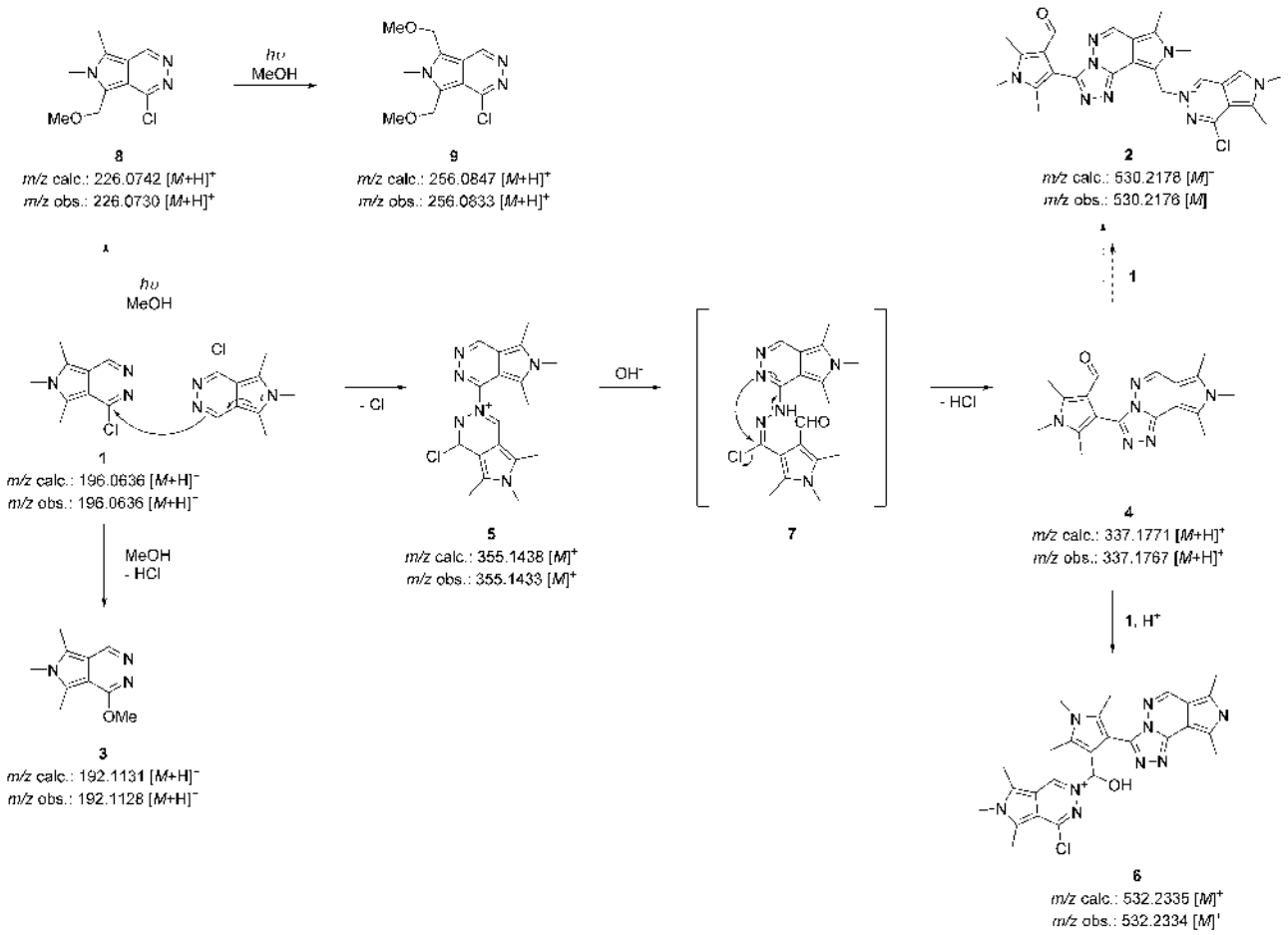
m/z 196.1: HR-MS and NMR results confirmed that this molecule corresponds to the parent compound **1**. Furthermore, we were able to reproduce the above-mentioned HPLC reaction pattern through repeated exposure of the isolate to methanol at 50 °C. This observation verifies the reactive nature of **1** in solution. Soaking experiments with this isolate resulted in a structure containing both **2** and some density for additionally bound **1**, presumably due to the higher concentration or resolution (Figure 1 B, structure referred to as EP–**1**-**2**). It could therefore be excluded that trace impurities from the synthesis of **1** accumulate in the binding pocket of EP or that such impurities are required for the formation of **2**.

m/z 337.3: From the mass spectrum, it was obvious that no chlorine atoms are present in this molecule. In agreement with our HR-MS and NMR experiments, we assigned structure **4** to this compound. Importantly, this structure closely resembles the central scaffold of **2**, the putative molecule initially found in the binding pocket of EP.

m/z 355.5: The isotopic distribution of this molecule suggests the presence of a single chlorine atom, leading to the assignment of structure **5**. NMR, HR-MS, and MS/MS analyses supported this proposal (Scheme S1 in the Supporting Information). Crystallographic experiments did not reveal electron density in the EP active site that might correspond to **5**. Instead, density for parts of **4** could be detected (Figure S2B, structure referred to as EP–**4**), which might be explained by the finding that the more water-soluble **5** is readily converted into **4**.

m/z 533.1: The associated NMR spectrum indicates the presence of a trimeric form of **1**. However, the absence of a CH_2 group excludes a composition that could explain the electron density observed in the crystal structure. Instead, in agreement with HR-MS and MS/MS experiments (Scheme S1), structure **6** was assigned to this molecule. Compound **6**, which we could provide for interaction studies after HPLC isolation, consistently did not bind to EP in crystallographic experiments.

From the chemical composition of the isolated compounds, we deduced a mechanism for the reaction of **1** in methanolic solution, which is depicted in Scheme 1. The reaction pathway is initiated by nucleophilic attack by an imino nitrogen of **1** on the 3-chloropyridazine ring of a second molecule. This dimerization results in the formation of the charged intermediate **5**. In a competitive reaction, the halogen atom is exchanged against a methoxy group upon nucleophilic attack of a solvent molecule, which forms **3**. After its generation, **5** is transformed into **4** by an intramolecular substitution reaction. This reaction is proposed to proceed via the hypothetical intermediate **7**, which is formed through hydrolytic cleavage of the iminium moiety in **5**. A similar reaction cascade has been described for unsubstituted 6*H*-pyrrolo[3,4-*d*] pyridazines.[8] The suggested reaction pathway is supported by the fact that treatment of **5** with an aqueous ammonia solution leads to the quantitative transformation of **5** into **4**, as shown by HR-MS analysis. Under



Scheme 1. Proposed mechanism for the reaction of **1** in methanolic solution. Experimentally observed HR-MS data are compared to calculated *m/z* values of the postulated reaction intermediates.

low-pH conditions, however, the transformation of **5** into **4** is substantially slower. Compound **4** finally reacts with an additional molecule of **1** to give the trimeric form **6**, a charged hemiaminal species.

To explain the inherent reactivity of **1**, we calculated Nucleus-Independent Chemical Shifts (NICS), natural atomic charges, Wiberg bond indices, and electrophilicity indices for this molecule and several derivatives thereof using QM methods at a high level of theory (Table S2). The aromatic nature of the pyridazine ring was investigated through comparison with related isoindole, [2′,3′]-pyrrolo annelated indole-like and monocyclic analogues (Table S2). The NICS-(1) calculations (Table S11) revealed a loss of aromaticity on the six-membered ring of **1** and other isoindoles compared to indole analogues. At the same time, an increased bond order for the C=N bond in **1** compared to pyridazine and its [2′,3′]-pyrrolo-annelated indole-like analogues is indicative of an imine-like C=N bond, which facilitates the initial reaction. The high global electrophilicity index of **1** (Table S16) and the positive atomic charge of the carbon atom attached to the chlorine (Table S3) further indicate that this position is prone to nucleophilic attack. A putative positive charge, as proposed for compound **2**, should lead to an additional gain in reactivity, as indicated by increased electrophilicity and a reduced HOMO–LUMO gap (P-**4** in Tables S15, S16). This might explain why **2** is only formed in minute quantities, resulting in a partial binding-site occupancy (67 %).

To further investigate which of the molecules from the reaction mixture bind to EP, we performed different NMR experiments in the presence of EP. In a time-resolved series of standard 1D ^{1}H-NMR experiments, we were able to corroborate the reactivity of **1** in an EP-containing sample (Figure S1B). A 2D ^{1}H-^{1}H-NOESY NMR experiment confirmed the correct structure assignment of **4** and demonstrated that this molecule binds to EP not only in the crystal but also in solution, as indicated by positive NOE cross-peaks (Figures 2 and S2B), which were negative in the absence of EP. An additional ^{1}H-^{13}C-HSQC experiment underlined that this molecule in fact contains an aldehyde function (Figure S3). Furthermore, the NOESY spectra confirmed that **1** can also interact with EP (Figures 1B and 2). Overall, our NMR experiments identified **1** and **4**, but not **2**, as EP binders. Presumably, the strong interactions of **2** in the EP binding site, its low level of formation, or its high reactivity in solution prevented spectroscopic detection. The binding affinity of **2** can be estimated to be better than 10 μM, as judged by the crystallographic observation of this ligand despite its presence at only μM or even lower concentrations in the soaking solution. The determination of an exact binding constant, however, seems to be hardly possible because **2** is formed in

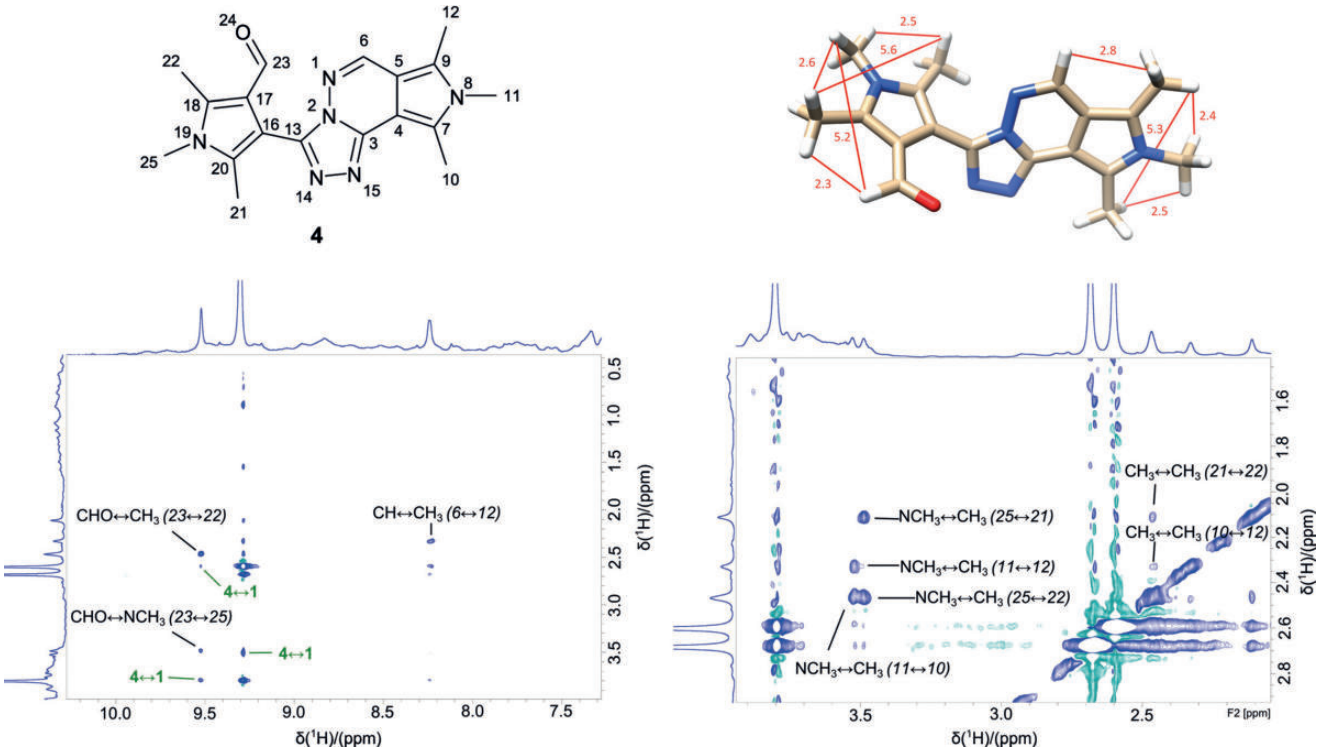


Figure 2. NMR analysis of the reaction mixture in the presence of EP. An ^{1}H-^{1}H-NOESY NMR spectrum was collected from an EP sample incubated with **1** for 12 hours. In both excerpts, positive peaks are shown in blue, negative peaks in cyan. The cross-peaks annotated in black correspond to intramolecular NOE contacts of **4**. Additional intermolecular contacts (green) between **4** and **1** indicate spatial proximity of the two weak binders in the binding pocket of EP. Observed NOE peak intensities match the interatomic distances of the three-dimensional model of **4** (shown in red in Å). High-intensity peaks in both spectrum excerpts result from the excess of **1**.

only minute quantities and is very reactive as shown by our QM-calculations, both of which indicate that the ligand might bind even with sub-micromolar affinity to EP.

While the finding that **1** and **4** can both weakly interact with EP supported our hypothesis that the unknown EP binder **2** is composed of these two compounds, as suggested by its electron density (Figure 1 A), it was crucial to show that these two molecular parts are actually covalently linked in the crystal structure. Indeed, the electron density observed in the initial crystal structure could not be explained by the presence of either **1** or **4** in partial occupancy but only by the presence of both parts at the same time at a distance too close to represent separated molecules. Fortunately, the EP–**1**-**2** structure, subsequently obtained from a crystal treated with the HPLC isolate **1**, was of sufficient resolution (1.03 Å, Figure 1 B) to enable a least-squares refinement using SHELXL and thereby the determination of geometric parameters of the bound molecule **2**, along with estimated standard deviations.[9] In particular, we analyzed the geometry of the methylene bridge that must be formed upon the connection of **1** and **4**. Comparison to values derived from small-molecule crystal structures clearly indicated that a covalent linkage is chemically consistent and thus very likely (Figures 3 A and Figure S4). Based on our EP–**1**-**2** structure, the presence of an aldehyde in **2** could also be confirmed by its derived geometric characteristics, such as a typical CO bond length of 1.24 ± 0.03 Å (Figures 3 A and S5). These observations allowed us to confidently assign structure **2** to the initially unknown EP inhibitor. Importantly, an LC/HR-MS experiment indicated the presence of a compound with the proposed molecular composition of **2** in the crude reaction mixture. An MS/MS fragmentation of this isolate, moreover, supported this assignment (Scheme S1).

For a hypothetical reaction of **1** and **4** to give **2**, the activation of an aliphatic CH_3 group and the subsequent substitution of a hydrogen atom would be a mechanistic requirement. Since 2-methylpyrroles and related 2-methylindoles are known to react at their respective CH_3 groups after irradiation with visible light, or anodic or catalytic oxidation,[10] a radical mechanism was assumed to be involved in the formation of **2**. To investigate the photochemical reactivity of the methyl groups, a sample of the parent compound **1** was irradiated in a photochemical reactor. Compounds **8** and **9** were identified as major products of the photochemical activation of **1** in methanol (Scheme 1). While this indicated a certain photochemical reactivity, the formation of **2** was also observed in the absence of light. Since **2** was traceable in solution in minute quantities, EP might selectively trap and thereby stabilize this highly reactive molecule. Alternatively, it seems possible that the juxtaposition of **1** and **4** in the EP binding pocket may directly facilitate the final reaction between these two species upon the activation of **4**.

EP specifically recognizes the final reaction product **2** through three hydrogen bonds, one of which is mediated by water (Figure 4 A). In particular, the aldehyde oxygen of **2** accepts an H-bond from Thr222, although this residue usually acts as an acceptor.[11] This phenomenon is part of several

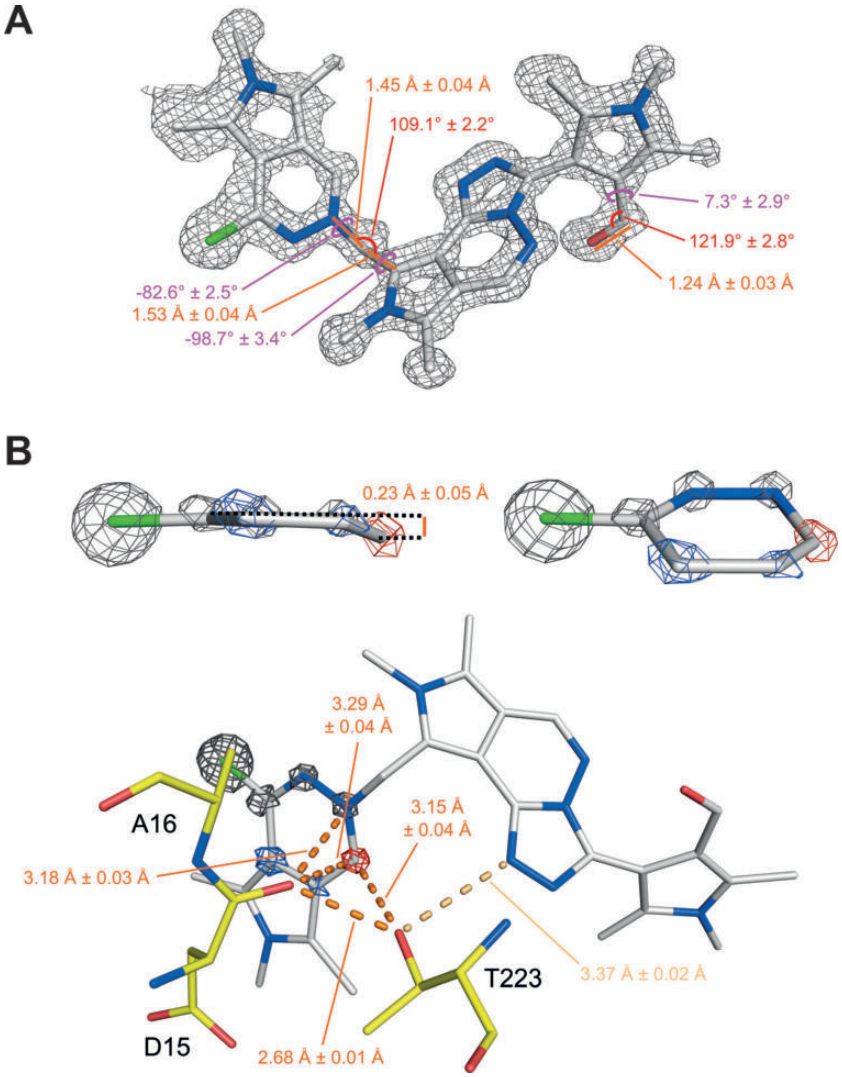


Figure 3. Crystallographic elucidation of the structure of **2**. A) Selected geometric parameters and estimated standard deviations underlining the presence of an aldehyde function and methylene bridge in **2**. The depicted values and $2mF_o-F_c$ map (1σ level) were derived using the SHELXL-based refinement strategy described in the Supporting Information and compared to CSD queries (Figures S4 and S5). B) Non-planarity of the chloropyridazinium ring. Our SHELXL refinement revealed that one of the pyridazinium carbon atoms is significantly shifted out of the ring plane (4.3σ). To further highlight the positional accuracy achieved by the refinement of this high-resolution structure, the $2mF_o-F_c$ electron density is shown at 1.7σ for this out-of-plane atom in red, as well as for two other pyridazinium carbon atoms in blue. A slightly higher σ-cutoff was chosen for the residual atoms of this positively charged heterocycle (2.3σ, gray mesh) in order to better visualize the location of each atom, as indicated by the density maxima. How the partial positive charge on the slightly pyramidalized carbon atom and attached nitrogen might be compensated by the protein environment is indicated in orange. Based on distance criteria, Thr223 likely points with its hydroxy hydrogen atom toward one of the Asp15 carbonyl oxygen lone pairs so that one of the Thr223 lone pairs is directed toward the pyramidalized carbon atom of **2**. The second Asp15 CO lone pair, in contrast, can compensate the partial positive charge on the nearby pyridazinium nitrogen.

structural adaptations that take place upon binding of the ligand. Structure **2** would sterically interfere with the *apo* protein, which therefore experiences an induced fit as shown in Figure 4B and simultaneously takes up a glycerol molecule from the soaking buffer at the position of the catalytic water. Finally, the question arises as to how the positive charge of **2** is stabilized within the EP binding pocket. This charge will be partially located within the pyrrole heterocycle, which can interact electrostatically with the negatively charged Asp15 (Figure 4A). In addition, least-squares refinement of the EP–**1**-**2** structure revealed a very interesting chemical phenomenon that further clarifies the charge stabilization. The carbon atom attached to the formally positively charged pyridazinium nitrogen is significantly shifted out of the ring plane (4.3 σ, Figure 3B). A partial transfer of the negative charge from Thr223 to this imine carbon and the accompanying hybridization change might explain the local breakdown of ring planarity. Since the presence of a positive charge in the ring system of the pyridazine moiety in **2** leads to its destabilization according to our quantum chemical calculations, these electrostatic interactions seem to be key for its stabilization within the binding pocket. In small-molecule crystal structures, analogous pyridazinium and pyridinium rings are generally found to be fully planar (Figure S6A,B). Two host–guest complexes, however, are exceptions to this trend and bear astonishing similarities to the EP–**1**-**2** protein-ligand complex (Figure S6C–E). All of these complexes exhibit two partially negatively charged oxygen atoms in the host next to the positively charged guest molecule. Partial electron transfer from the host to the guest might explain the observed local deviation from planarity.

In summary, we conclude that the presumably weak affinity of **1** for EP cannot fully explain the astonishing biological activity observed in multiple assays. Instead, a number of other molecules, which are generated from **1** in a reaction cascade that leads to a profound rearrangement of its heterocyclic core structure, contribute to EP inhibition. The unequivocally non-fragment-like molecule **2** seems to exhibit the strongest interaction of all the reaction products, with an estimated affinity of at least 10 μM, if not decidedly better. It contacts EP through multiple H-bonds and electrostatic interactions that are enabled by the unique chemistry of the inherently cationic compound. This study shows that in addition to a thorough investigation of the chemical stability of fragment library members, structural information is essential for the identification of false-positive binders, which is of utmost importance in the early stages of a fragment-based drug discovery project to avoid investments made to follow up futile leads. Based on crystallographic data, we were able to identify the 1-chloropyrrolopyridazine core of **1** as a moiety that represents an intriguing case of a self-reacting fragment that may be generally prone to react under conditions commonly employed in biological assays.

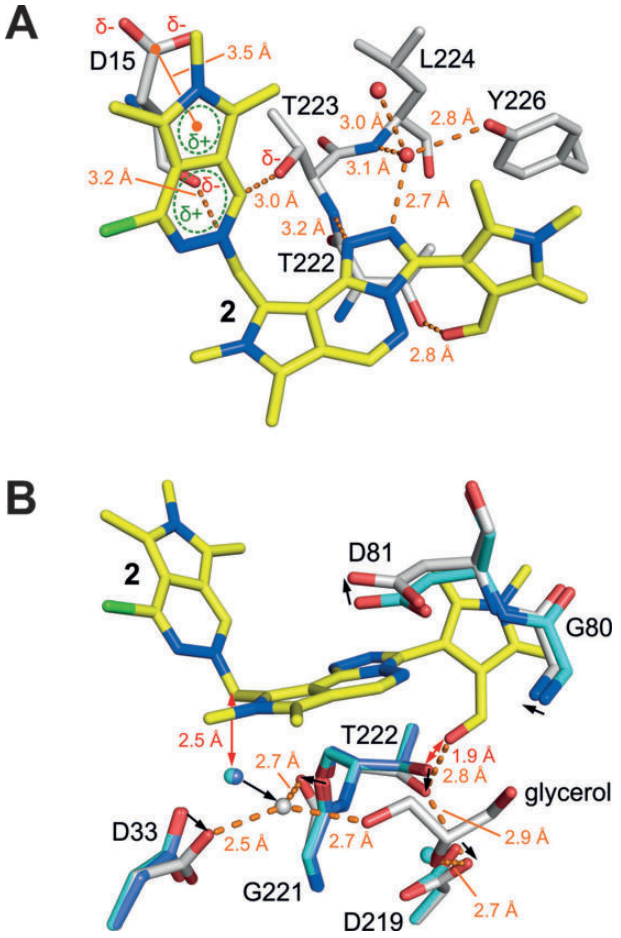


Figure 4. Recognition of **2** by EP. A) Interactions between **2** (yellow) and EP (gray), assessed by means of intermolecular distances (orange). For more details concerning the proposed electrostatic interactions between Asp15, Thr223, and **2**, see Figure 3 and Figure S6. B) Induced-fit binding of **2**. Compared to the EP apo form (cyan, PDB ID: 4Y5L), the ligand-bound protein underwent significant conformational changes, as indicated by black arrows (RMSD of depicted protein atoms = 0.44 Å). Since the binding site of **2** is not fully populated in the EP–**1**-**2** structure (occupancy of 63 %), some of the depicted residues are found in the apo (violet) as well as in the ligand-bound (gray) conformation (e.g., Asp33). Red arrows indicate clashes that would occur between the apo state and the ligand without conformational adaptation of the protein.

Experimental Section

The reaction cascade emanating from **1** was simply initiated by dissolving the compound. Purchased **1**, as well as all HPLC isolates, were soaked into EP crystals prepared as described previously, before the collection of diffraction data and structure determination.[5,7] Additional experimental details are provided in the Supporting Information.

Acknowledgements

We thank the Helmholtz-Zentrum Berlin for the allocation of synchrotron radiation beamtime and travel support. Furthermore, we are grateful for ideas and feedback on the proposed mechanism and refinement results provided by Ali Al-Mourabit, Andrea Vasella, and co-workers, as well as George M. Sheldrick. The work described herein was financially supported by the BMBF-Project Frag2Xtal (No. 05K13RM1) and by the European Research Council (ERC) of the European Union (grant 268145-DrugProfilBind). RL acknowledges the Liebig- (VCI) and Emmy-Noether-Programs (DFG).

Conflict of interest

The authors declare no conflict of interest.

Keywords: drug discovery · fragment-based lead discovery · medicinal chemistry · PAINS · reactivity

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 1908–1913
Angew. Chem. **2017**, *129*, 1934–1940

[1] J. B. Baell, *J. Chem. Inf. Model.* **2013**, *53*, 39–55.
[2] J. B. Baell, G. A. Holloway, *J. Med. Chem.* **2010**, *53*, 2719–2740.
[3] a) B. J. Davis, D. A. Erlanson, *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2844–2852; b) S. L. McGovern, E. Caselli, N. Grigorieff, B. K. Shoichet, *J. Med. Chem.* **2002**, *45*, 1712–1722; c) G. M. Rishton, *Drug Discovery Today* **2003**, *8*, 86–96.
[4] a) J. C. Hermann, Y. Chen, C. Wartchow, J. Menke, L. Gao, S. K. Gleason, N.-E. Haynes, N. Scott, A. Petersen, S. Gabriel, et al., *ACS Med. Chem. Lett.* **2013**, *4*, 197–200; b) C. J. McElhinny, A. H. Lewin, S. W. Mascarella, S. Runyon, L. Brieaddy, F. I. Carroll, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 6661–6664; c) J. Schiebel, S. G. Krimmer, K. Röwer, A. Knörlein, X. Wang, A. Y. Park, M. Stieler, F. R. Ehrmann, K. Fu, N. Radeva, et al., *Structure* **2016**, *24*, 1398–1409.
[5] J. Schiebel, N. Radeva, S. G. Krimmer, X. Wang, M. Stieler, F. R. Ehrmann, K. Fu, A. Metz, F. U. Huschmann, M. S. Weiss, et al., *ACS Chem. Biol.* **2016**, *11*, 1693–1701.
[6] J. Schiebel, N. Radeva, H. Köster, A. Metz, T. Krotzky, M. Kuhnert, W. E. Diederich, A. Heine, L. Neumann, C. Atmanene, et al., *ChemMedChem* **2015**, *10*, 1511–1521.
[7] N. Radeva, S. G. Krimmer, M. Stieler, K. Fu, X. Wang, F. R. Ehrmann, A. Metz, F. U. Huschmann, M. S. Weiss, U. Mueller, et al., *J. Med. Chem.* **2016**, *59*, 7561–7575.
[8] T. Hernandez de la Figuera Gomez, J. S. Arques, R. A. Jones, H. M. Dawes, M. B. Hursthouse, *J. Chem. Soc. Perkin Trans. 1* **1985**, 899–902.
[9] G. M. Sheldrick, *Acta Crystallogr. Sect. C* **2015**, *71*, 3–8.
[10] a) L. Eberson, T. Laerum, K. Undheim, S.-K. Kan, M. Lounasmaa, *Acta Chem. Scand. Ser. B* **1980**, *34*, 747–751; b) V. Karunaratne, D. Dolphin, *J. Chem. Soc. Chem. Commun.* **1995**, 2105; c) S. Ohmiya, M. Noguchi, S. Ina, H. Kubo, H. Otomasu, *Chem. Pharm. Bull.* **1992**, *40*, 854–857; d) N. Sakurai, S. Ohmiya, *J. Chem. Soc. Chem. Commun.* **1993**, 297–298.
[11] N. Radeva, J. Schiebel, X. Wang, S. Krimmer, K. Fu, M. Stieler, F. Ehrmann, A. Metz, T. Rickmeyer, M. Betz, J. Winquist, A. Y. Park, F. U. Huschmann, M. S. Weiss, U. Mueller, A. Heine, G. Klebe, *J. Med. Chem.* **2016**, *59*, 9743–9759 .

Manuscript received: October 11, 2016
Revised: November 24, 2016
Final Article published: January 18, 2017

3 Comparative assessment of data from different projects and Conclusions

This thesis consists of a multitude of different projects, all of which share the same purpose of enabling a deeper understanding of protein structure and dynamics, leading to characterization of protein functionality. In the projects presented, this purpose is fulfilled either by creating or improving analytical tools that are freely accessible to all researchers or by applying such tools developed by our or other groups. In the following it is shown what the different projects have in common and how they complement each other.

The projects "Exact distance measurements for structure and dynamics in solid proteins by fast magic-angle-spinning NMR" (2.1.2) and "Non-equilibrium hydrogen exchange for determination of H-bond strength and water accessibility in solid proteins" (2.1.3) are both primarily intended to facilitate structure elucidation by proton-detected solid-state NMR spectroscopy. On the one hand, by enhancing the precision of distance restraints from through-space correlations, on the other hand, by enabling the detection of H-bonds and the global position of the residue (by determining the water accessibility). For both projects, the goal was to develop an easy-to-use, low-effort method providing high-quality output. Therefore, no special sample preparation such as selective labeling, mutagenesis, or posttranslational chemical modification is required for either approach. This allows the user to apply these methods in addition to any other planned measurement. In addition to providing restraints for structure determination, both methods can also be used to determine protein dynamics. Since the relation between through-space correlation transfer and distance is over-proportional, differences between measured and equilibrium distance can be used in order to decipher spatial dynamics (compare Chapter 2.2.2). The site-resoled proton deuterium exchange simultaneously reports on local vibrations with respect to the H-bond opening rates and on larger, concerted motions in terms of water accessibility of hydrophobic regions. Both novel NMR tools are advances in structural biology in general and proton-detected magic-angle-spinning NMR spectroscopy in particular. While the exact-RFDR approach (2.1.2) is equally beneficial for nearly all NMR samples, the non-equilibrium H/D exchange (2.1.3) is especially suited for challenging samples. The high sensitivity with 2D-heteronuclear correlation, which is the only type of measurement required, makes it particularly valuable for otherwise insensitive protein samples. Furthermore, the information about the hydrophobic shielding, more precisely, the information about being on the surface, exposed to the solvent, or being in the hydrophobic core, shielded by other residues of the protein, provides first ideas for resonance assignment and protein fold, respectively.

The projects "Carbonic anhydrase structure and dynamics under physiological conditions." (2.2.1) and "Protein motional details revealed by complementary structural-biology techniques" (2.2.2) can be considered as the applications of the two projects mentioned above. While this holds perfectly true for the H/D-exchange measurements, Project 2.2.2 is objectively rather an application of the methods developed by Vögeli et al.[39a]

Even though the non-equilibrium hydrogen-deuterium exchange was performed in liquid-state NMR previously, Project 2.2.1 is based on the H/D-exchange Project 2.1.3, because it relies on the same detailed analysis that separates the two information on hydrophobic shielding, or water accessibility, and H-bond strength.

Since the two Projects 2.2.2, for the determination of spatial protein dynamics, and Project 2.1.3, H/D-exchange in solid-state NMR, are performed on the same protein, a comparison of their outcome regarding protein dynamics stands to reason. This comparison is shown in **Figure 23**. **Figure 23 A** shows the H-bond opening rate according to the analysis in the H/D-exchange Project 2.1.3, strong H-bonding is indicated by blue color, **Figure 23 B** shows the Root-Mean-Square Deviation (RMSD) of the substructures from the eNOE-based multistate structure determination from Project 2.2.2, and **Figure 23 C** shows the Root-Mean-Square Fluctuation (RMSF) of the MD simulation performed for cross-validation within eNOE-based dynamics. The comparison is especially meaningful for residues with strong H-bonds in **A** since weak, absent, or undetectable (ambiguous) H-bonding does not automatically reveal the presence of large-amplitude motion. By contrast, a strong H-bond prevents movements with large amplitudes.

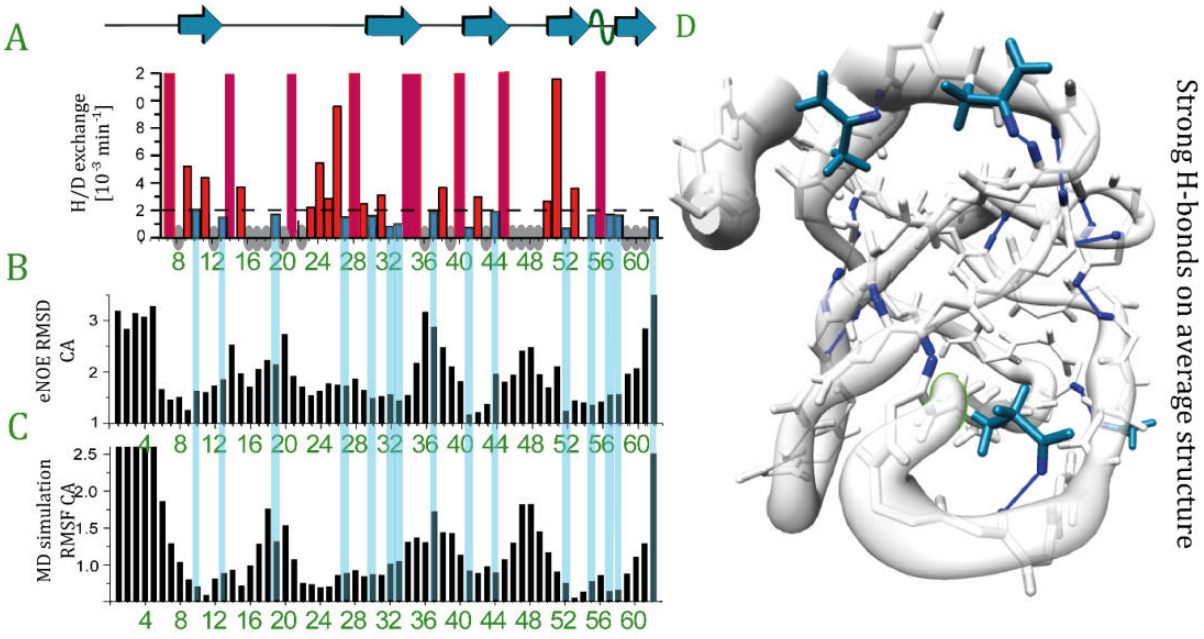


Figure 23: Comparison between dynamics information from Projects 2.1.3 and 2.2.2. **A:** H-bond opening determined by solid-state H/D exchange at pH 7 (Project 2.1.3). Blue columns indicate H-bonds that are recognized as strong. **B:** RMSD of the eNOE-based multistate structure (Project 2.2.2). The amplitude of the deviation between the substructures is comparable to the amplitude of motion fluctuation. The high columns, therefore, stand for dynamic residues and vice versa. **C:** RMSF from the molecular dynamics-simulation from Chapter 2.2.2, main text **Figure 3**. As with **B**, the high columns stand for mobile, the low for rigid residues. **D:** Strong H-bonds as shown in A (blue columns) depicted on the single-state structure from Project 2.2.2 (compare **Figure 3 B**, Chapter 2.1.3). H-bonds are drawn as blue lines; sidechains involved in H-bonds with backbone amides are shown in cyan. Blue shadings connect columns **B** and **C** with the corresponding ones in **A** and represent strong H-bonding.

Affirmatively, all sides where strong H-bonds could be detected show small motion amplitudes in the standard MD simulation and experimental determination of spatial motion from Chapter 2.2.2, except for residues 19Ser, 37Thr, and 62Asp. The data divergence of these sides can be explained. The amide protons of 19Ser and 37Thr are located in loop regions, they are not bound to backbone carbonyl oxygens, but to oxygens from glutamine and asparagine sidechains, respectively (**Figure 23 D**, shown in cyan). In contrast to the backbone, these sidechains have many bonds with rotational freedom, with the only rotatable bond being between N and C^{α}. The residue 62Asp is the final one at the N-terminus, which is much more rigid in solid-state (H-bond detection Chapter 2.1.3) than in solution (eNOE based multistate struct determination, Chapter 2.2.2). The higher rigidity of the N-terminus in a solid state is caused by crystal contacts with neighboring proteins. The good correlation between the experimental spatial dynamics-determination, its cross-validating MD simulations, and the H/D-exchange analysis is an indication that the H-bond detection is trustworthy. This suggested a comparison of the experimental H/D-exchange data with dedicated molecular dynamics simulations for H/D-exchange prediction. Therefore, Snehal Patel and Prof. Dr. Lars V. Schäfer kindly conducted MD simulations in which the H/D-exchange rate is predicted by the formation of hydrogen bonds between the backbone amide protons and solvent water (unpublished data). [105] This is the most direct theoretical approach for H/D-exchange determination since the H/D-exchange, besides a broken backbone H-bond, requires water to be present at the amide site (compare Chapter 1.2.4). The simulations were carried out using standard approaches with sidechain charges that resemble the situation of pH7 and pH3 (solid- and solution-state condition, respectively) and REST (Replica Exchange with Solute Tempering) simulations with sidechain charges resembling pH 3[106] . Simulations at pH3 and 7 showed only negligible differences. The exchange prediction using MD simulations as shown in **Figure 24**, with an H-bond between amide and solvent water as a requirement, all parameters responsible for H/D-exchange hinderance are taken into account simultaneously. Therefore, the two separated parameters of the experimental H/D-exchange analysis of Chapter 2.1.3 must be recombined. The plot showing the intramolecular H-bond strength (**Figure 24 A**) is stacked with the plot showing the presence of hydrophobic shielding (**Figure 24 B**). Blue-colored columns in **A** show strong H-bonding, and blue-colored columns in **B** the presence of hydrophobic protection against solvent water. **Figures C** and **D** show the H/D-exchange probability predicted by MD-simulations. High columns denote fast exchange while low columns denote slow exchange. Sites where either hydrophobic shielding, strong H-bonding, or both is detected from the experimental analysis, are shaded blue. Sites with weaker hydrogen bonding and no detectable hydrophobic shielding are shaded red. No shading means no data from the solid-state analysis is available. With the exception of the residues 30Ile, 32Thr, and 62Asp (shown as yellow columns in **Figure 24 C** and **D**), all sites showing exchange hindrance due to strong H-bonding or hydrophobic shielding also show a slow exchange as predicted by MD simulation.

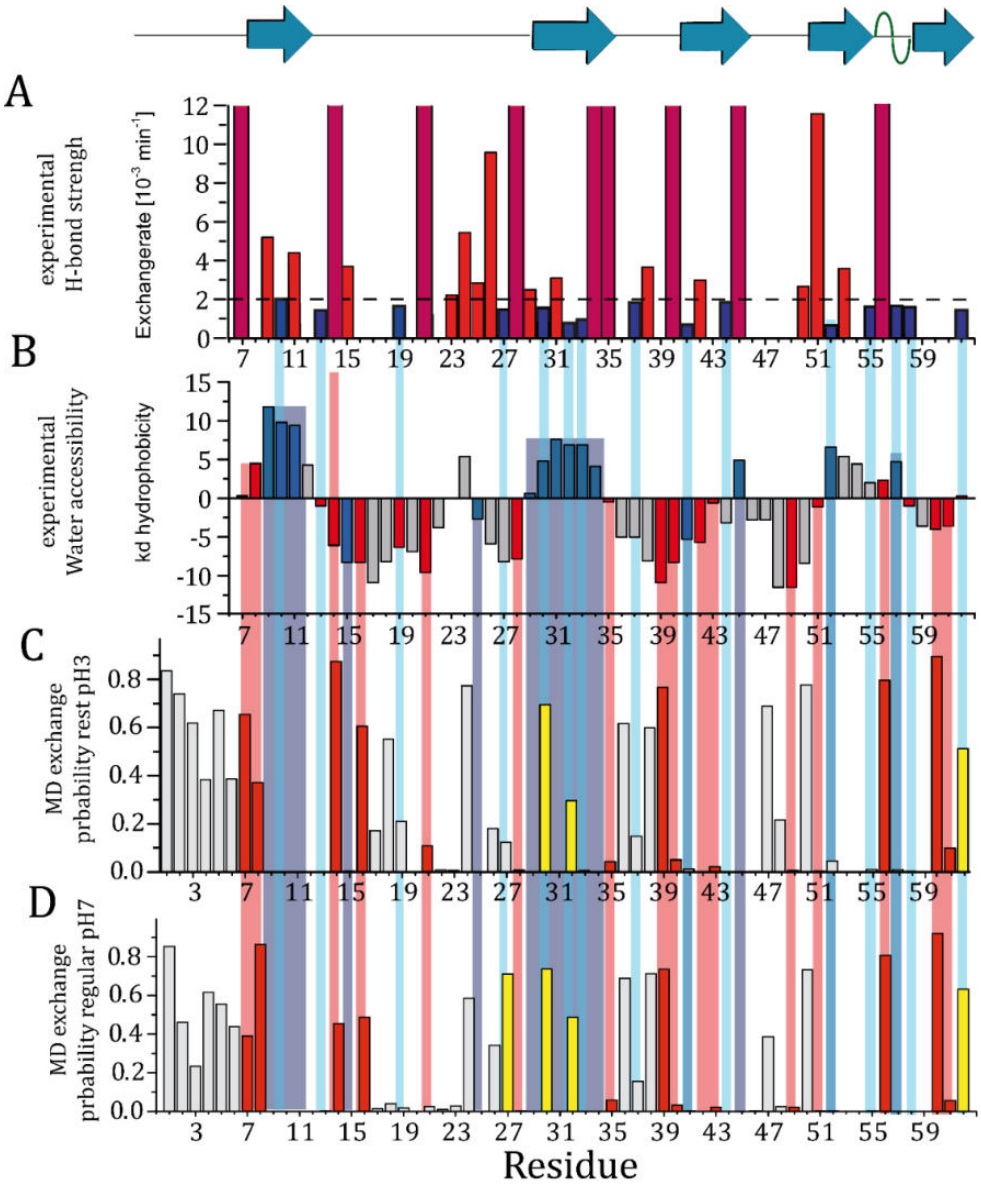


Figure 24: Comparison between experimental H/D exchange recorded with solid-state NMR and MD simulations dedicated to predicting H/D exchange. **A:** Proton-deuterium exchange rate as determined from an exponential fit of the amide peak intensity decay. Columns corresponding to slowly decaying peaks are colored blue and are denoted to report on strong H-bonding, columns corresponding to faster-decaying residues are colored red and denote to report on weaker H-bonding. Purple-red bars correspond to peaks with a decay, too fast to be recorded (already decayed after 20 min). **B:** Kyte-Doolittle hydrophobicity as a running average of 3. Columns corresponding to residues showing hindered water accessibility, as shown by the presence of a multiexponential behavior of peak-intensity decay (compare supporting information in Chapter 2.1.3) are colored blue. Columns corresponding to clearly monoexponentially decaying peaks, denoting the absence of hydrophobic shielding, are colored red. A high value means that the residue and its two sequential neighbors are hydrophobic on average, negative values denote hydrophilicity. **C:** H/D exchange predicted by Replica Exchange with Solute Tempering (RSET) simulations. An H-bond between the backbone-amide proton and water is counted as an exchange. **D:** H/D-exchange predicted by a MD-simulation at pH 7. As in **C**, H-bonds between backbone and solvent water are counted as exchanges.

In summary, the correlation between H/D exchange, experimentally measured by solid-state NMR and predicted by MD simulation, correlates quite well. The differences for the residues 30Ile and 62Asp can be explained. The amide of 30Ile is detected to form an intermolecular hydrogen bond within the protein crystal (**Figure 25**); this crystal contact is responsible for exchange hindrance caused by the hydrogen bond itself and the solvent protection caused by the adjacent protein.

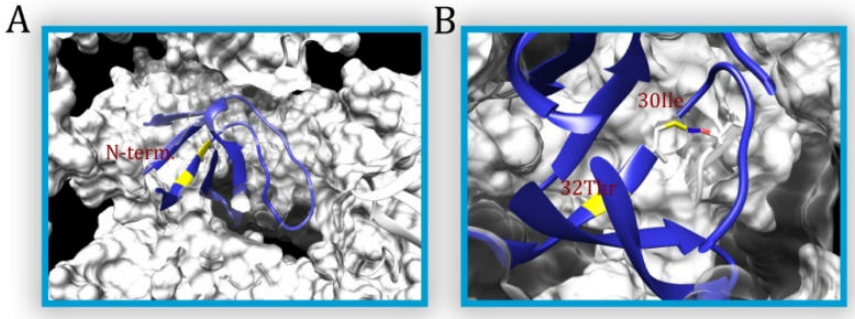


Figure 25: Depiction of exchange hindrance due to crystal contacts. **A:** The β-sheet including the residues 30Ile and 32Thr (colored yellow) and the N-terminus, are close to the surface of the neighboring protein. **B:** Representation of the intermolecular H-bond between the backbone amide-proton of 30Ile and the sidechain oxygen of 37Thr, from the neighboring protein.

This is most likely also the explanation for the data divergence of 32Thr. MD simulations are performed for a single protein embedded in water that mimics the situation of a single molecule in solution. The reason for the divergence at 62Asp is similar to that explained above (**Figure 23**), most likely due to motional restrictions of the N-terminus due to crystal-contacts. Residues, shaded in red, for which no exchange hindrance is detected by NMR usually also show a high exchange, as predicted by MD simulations. However, there are some divergences which may be due to differences in dynamic behavior in the solid and liquid phase. In general, it is more difficult to detect the absence of exchange hinderance using NMR, as it also means the absence of data.

Besides being the first kinetics measurement in protein-solid-state NMR, the study of Chapter 2.1.3, which aims to separate the different forms of exchange hindrances, provides restraints for assignment, determination of structure, and dynamics simultaneously. In order to validate the verisimilitude of the separation of H-bond data from water accessibility hindrance, an additional MD analysis (pH7 for the data shown) of the backbone H-bond opening rate was carried out by Snehal Patel and Prof. Dr. Lars V. Schäfer (unpublished data). For the prediction of H-bond opening rates, 250 MD-simulations of 1 ns length were performed each. H-bonds between Backbone amide and carbonyl groups were calculated and averaged over the 250 simulations. An autocorrelation function was calculated by GROMACS for each residue and fitted exponentially. **Figure 26** compares the H-bond strength determined by solid-state NMR via exponential fitting of the peak decay (**Figure 26 A**) with the H-bond opening rates predicted by MD simulations (**Figure 26 B**). The H-bond opening rates determined experimentally (**Figure 26 A**) and theoretically (**Figure 25 B**) show a good correlation.

Whenever a strong H-bond is detected by solid-state NMR, the H-bond opening rate determined by MD is also low, with the sole exception of the H-bond formed by residue 57Tyr. This divergence is again most likely due to differences regarding the N-terminus in the solid- and liquid-state.

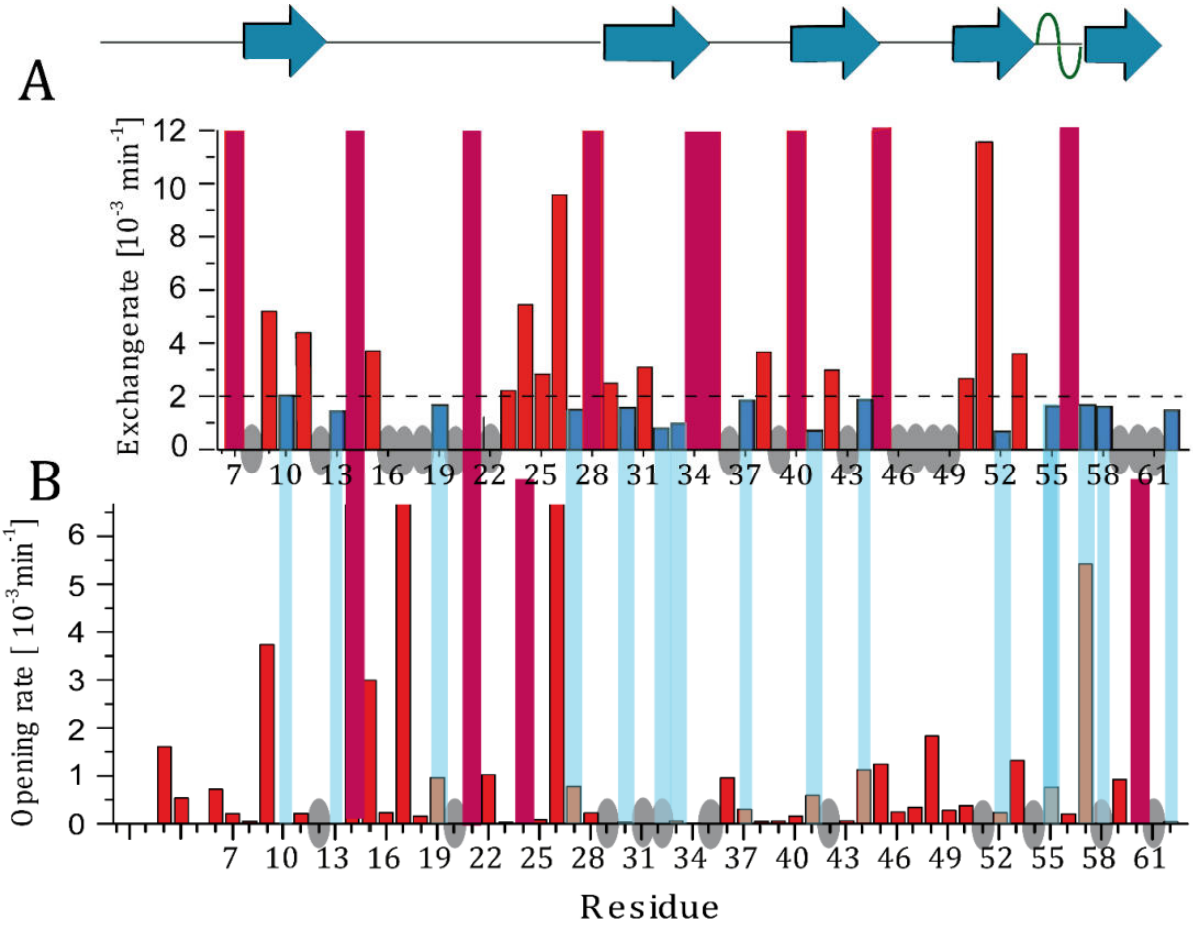


Figure 26: Comparison of H-bond opening rates determined by solid-state NMR and MD simulation. **A:** H-bond strength read out from the exponential decay rates of backbone amide peaks. Blue columns correspond to backbone amides forming strong H-bonds. **B:** Opening rates determined by MD simulation. The gray ellipsoids indicate the absence of unambiguous data.

The good agreement of the comparison between experimentally and theoretically determined H-bond strength indicate that both rate-determining parts of the exchange mechanism can indeed be analyzed separately as shown in Project 2.1.3. Thus, the "purified" data can be used for assignment by using the information on the hydrophobicity of residues close in space, for structural refinement, with the H-bonds as structural restraints and also for the analysis of protein dynamics.

The cross-validation between the two projects about H/D-exchange measurement (2.1.3 and 2.2.1) and the two projects about exact distance determination and experimental determination of spatial dynamics (2.1.2 and 2.2.2) shows encouraging agreement. The application of both methods revealed new mechanistic insights of biological importance. The H/D exchange of the enzyme human carbonic anhydrase II, an important drug target for many diseases, provided further evidence for the mechanistically important chain of rigid water molecules in the active center and detected H-bonds, stabilizing loops that

carry functionally crucial sidechains. The eNOE-based multistate structure determination, which is the first NMR approach for determination of protein motion in a spatial manner (developed by Vögeli et al.), shed light on the ligand-binding mechanism of the SH3 domain. Nevertheless, the main achievements of this work are represented by the advancement of solid-state NMR spectroscopy by facilitating the structure and dynamics methods. Hereby, the improvement of through-space restraint precision stands out. Against the common opinion in the field, affected by the theoretical complexity of the RFDR transfer, it was possible to turn qualitative into precise-distance restraints. The resulting improvement of protein structures in terms of precision and accuracy is demonstrated in Chapter 2.1.2.

Liquid state NOESY is the most widely used method providing distance restraints for protein structure determination. However, proteins can only be analyzed by liquid-state NOESY up to a molecular weight limit of ca. 300 kDa. In contrast, by using solid-state RFDR there is no absolute physical limit to the protein size. Nevertheless, the necessity of deuteration limits the number of distance restraints and prohibit detailed information about sidechains. Further advances in solid-state NMR technology, like Magic-Angel-Spinning of 111111 rotations per second or even faster, will enable measuring exact ^{1}H-^{1}H distance restraints of protonated protein samples. In addition to the backbone amide protons of deuterated and 100% back-exchanged samples, fully protonated samples contain also H^{α}- and sidechain-protons. Besides simply providing a higher number of restraints these contacts contain valuable information about structure and dynamics of sidechains. By using these distance restraints, determined by the eRFDR concept, in addition to multistate structure determination, or NMR restrained MD-simulation, spatial sidechain dynamics of proteins can be determined in the solid state. This is particularly valuable, because the physics of sidechains govern the biochemical function of the protein. In general, considering that a large fraction of drug targets are membrane proteins or protein fibrils, the facilitated determination of structure and dynamics assessment of insoluble proteins serves an urgent need in structural biology and pharmaceutical science.

4 References to the Introduction

[1] in *Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen*, Spektrum Akademischer Verlag, Heidelberg, **2009**, pp. 9-21.

[2] J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nature Reviews Drug Discovery* **2006**, *5*, 993-996.

[3] H. Hegyi, M. Gerstein, *Journal of Molecular Biology* **1999**, *288*, 147-164.

[4] a) M. D. Tuttle, G. Comellas, A. J. Nieuwkoop, D. J. Covell, D. A. Berthold, K. D. Kloepper, J. M. Courtney, J. K. Kim, A. M. Barclay, A. Kendall, W. Wan, G. Stubbs, C. D. Schwieters, V. M. Y. Lee, J. M. George, C. M. Rienstra, *Nature Structural &Amp; Molecular Biology* **2016**, *23*, 409; b) M. G. Zagorski, C. J. Barrow, *Biochemistry* **1992**, *31*, 5621-5631.

[5] J. S. Retel, A. J. Nieuwkoop, M. Hiller, V. A. Higman, E. Barbet-Massin, J. Stanek, L. B. Andreas, W. T. Franks, B.-J. van Rossum, K. R. Vinothkumar, L. Handel, G. G. de Palma, B. Bardiaux, G. Pintacuda, L. Emsley, W. Kühlbrandt, H. Oschkinat, *Nature Communications* **2017**, *8*, 2073.

[6] a) P. Rovó, C. A. Smith, D. Gauto, B. L. de Groot, P. Schanda, R. Linser, *Journal of the American Chemical Society* **2019**, *141*, 858-869; b) S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J. Phys. Chem. Lett.* **2018**, *9*, 1307-1311; c) R. Petra, L. Rasmus, *ChemPhysChem* **2018**, *19*, 34-39.

[7] J. D. Watson, *Molekularbiologie*, Pearson Studium, **2011**.

[8] P. T. Erskine, L. Coates, S. Mall, R. S. Gill, S. P. Wood, D. A. A. Myles, J. B. Cooper, **2003**, *12*, 1741-1749.

[9] in *Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen*, Spektrum Akademischer Verlag, Heidelberg, **2009**, pp. 381-402.

[10] M. F. Perutz, *Scientific American* **1978**, *239*, 92-125.

[11] J. C. Clemente, Stow, L.R., Janka, L.K., Jeung, J.A., Desai, K.A., Govindasamy, L., Agbandje-McKenna, M., McKenna, R., Goodenow, M.M., Dunn, B.M. CRDT, *pdb-one database* **2005**.

[12] K. Håkansson, M. Carlsson, L. A. Svensson, A. Liljas, *Journal of Molecular Biology* **1992**, *227*, 1192-1204.

[13] a) S. C. Frost, R. McKenna, *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications*, Springer Netherlands, **2013**; b) *Pharmacology & Therapeutics* **1997**, *74*, V.

[14] C. T. Supuran, *Journal of Enzyme Inhibition and Medicinal Chemistry* **2012**, *27*, 759-772.

[15] J. M. Bishop, *Cell* **1985**, *42*, 23-38.

[16] K. Saksela, P. Permi, *FEBS Letters* **2012**, *586*, 2609-2614.

[17] L. Pauling, R. B. Corey, *Proceedings of the National Academy of Sciences of the United States of America* **1951**, *37*, 272-281.

[18] R. Z. Kramer, J. Bella, P. Mayville, B. Brodsky, H. M. Berman, *Nature Structural Biology* **1999**, *6*, 454-457.

[19] S. Casares, E. Ab, H. Eshuis, O. Lopez-Mayorga, N. A. J. van Nuland, F. Conejero-Lara, *BMC Struct Biol* **2007**, *7*, 22-22.

[20] J. Teyra, H. Huang, S. Jain, X. Guan, A. Dong, Y. Liu, W. Tempel, J. Min, Y. Tong, P. M. Kim, G. D. Bader, S. S. Sidhu, *Structure* **2017**, *25*, 1598-1610.e1593.

[21] A. Ø. Madsen, in *Electron Density and Chemical Bonding I: Experimental Charge Density Studies* (Ed.: D. Stalke), Springer Berlin Heidelberg, Berlin, Heidelberg, **2012**, pp. 21-52.

[22] J. Overgaard, B. B. Iversen, in *Electron Density and Chemical Bonding I: Experimental Charge Density Studies* (Ed.: D. Stalke), Springer Berlin Heidelberg, Berlin, Heidelberg, **2012**, pp. 53-74.

[23] Ü. Akbey, S. Lange, W. Trent Franks, R. Linser, K. Rehbein, A. Diehl, B.-J. van Rossum, B. Reif, H. J. J. o. B. N. Oschkinat, **2009**, *46*, 67.

[24] a) S. W. Englander, T. R. Sosnick, J. J. Englander, L. Mayne, *Curr Opin Struct Biol* **1996**, *6*, 18-23; b) A. G. Doukas, A. Pande, T. Suzuki, R. H. Callender, B. Honig, M. Ottolenghi, *Biophysical journal* **1981**, *33*, 275-279.

[25] a) S. W. Englander, L. Mayne, Y. Bai, T. R. Sosnick, *Protein science : a publication of the Protein Society* **1997**, *6*, 1101-1109; b) H. Maity, W. K. Lim, J. N. Rumbley, S. W. Englander, *Protein Science* **2003**, *12*, 153-160.
[26] S. W. Englander, D. B. Calhoun, J. J. Englander, N. R. Kallenbach, R. K. Liem, E. L. Malin, C. Mandal, J. R. Rogero, *Biophysical Journal* **1980**, *32*, 577-589.
[27] a) A. Tomita, T. Sato, K. Ichiyanagi, S. Nozawa, H. Ichikawa, M. Chollet, F. Kawai, S.-Y. Park, T. Tsuduki, T. Yamato, S.-y. Koshihara, S.-i. Adachi, *Proceedings of the National Academy of Sciences* **2009**, *106*, 2612-2616; b) I.-H. Park, J. D. Venable, C. Steckler, S. E. Cellitti, S. A. Lesley, G. Spraggon, A. Brock, *Journal of Chemical Information and Modeling* **2015**, *55*, 1914-1925.
[28] J. Keeler, *Understanding NMR Spectroscopy*, Wiley, **2011**.
[29] a) H. Friebolin, *Ein- und zweidimensionale NMR-Spektroskopie: Eine Einführung*, Wiley, **2013**; b) J. Cavanagh, W. J. Fairbrother, A. G. Palmer, N. J. Skelton, M. Rance, *Protein NMR Spectroscopy: Principles and Practice*, Elsevier Science, **2010**.
[30] in *Solid State NMR Spectroscopy for Biopolymers: Principles and Applications*, Springer Netherlands, Dordrecht, **2006**, pp. 7-29.
[31] G. A. Morris, R. Freeman, *Journal of the American Chemical Society* **1979**, *101*, 760-762.
[32] A. Pines, M. G. Gibby, J. S. Waugh, **1973**, *59*, 569-590.
[33] S. Hediger, B. H. Meier, N. D. Kurur, G. Bodenhausen, R. R. Ernst, *Chemical Physics Letters* **1994**, *223*, 283-288.
[34] N. C. Nielsen, L. A. Strassø, A. B. Nielsen, in *Solid State NMR* (Ed.: J. C. C. Chan), Springer Berlin Heidelberg, Berlin, Heidelberg, **2012**, pp. 1-45.
[35] K. Takegoshi, S. Nakamura, T. Terao, *Chemical Physics Letters* **2001**, *344*, 631-637.
[36] R. Verel, M. Ernst, B. H. Meier, *Journal of Magnetic Resonance* **2001**, *150*, 81-99.
[37] J. Pauli, M. Baldus, B. van Rossum, H. de Groot, H. Oschkinat, *ChemBioChem* **2001**, *2*, 272-281.
[38] A. E. Bennett, R. G. Griffin, J. H. Ok, S. Vega, **1992**, *96*, 8624-8627.
[39] a) B. Vögeli, T. F. Segawa, D. Leitz, A. Sobol, A. Choutko, D. Trzesniak, W. van Gunsteren, R. Riek, *J. Am. Chem. Soc.* **2009**, *131*, 17215-17225; b) B. Vögeli, S. Olsson, P. Güntert, R. Riek, *Biophysical Journal* **2016**, *110*, 113-126.
[40] B. Vögeli, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2014**, *78*, 1-46.
[41] R. Zhang, Y. Nishiyama, P. Sun, A. Ramamoorthy, *J. Magn. Reson.* **2015**, *252*, 55-66.
[42] D. Strotz, J. Orts, C. N. Chi, R. Riek, B. Vögeli, *J. Chem. Theory Comput.* **2017**, *13*, 4336-4346.
[43] B. Vögeli, S. Kazemi, P. Güntert, R. Riek, *Nature Structural &Amp; Molecular Biology* **2012**, *19*, 1053.
[44] P. Güntert, C. Mumenthaler, K. Wüthrich, *Journal of Molecular Biology* **1997**, *273*, 283-298.
[45] R. Linser, V. Chevelkov, A. Diehl, B. Reif, *J. Magn. Reson.* **2007**, *189*, 209-216.
[46] E. Arbely, I. T. Arkin, *Journal of the American Chemical Society* **2004**, *126*, 5362-5363.
[47] a) S. Lindskog, *Pharmacology & therapeutics* **1997**, *74*, 1-20; b) S. C. Frost, R. McKenna, *Carbonic anhydrase: mechanism, regulation, links to disease, and industrial applications, Vol. 75*, Springer Science & Business Media, **2013**.
[48] C. U. Kim, H. Song, B. S. Avvaru, S. M. Gruner, S. Park, R. McKenna, *Proceedings of the National Academy of Sciences* **2016**, *113*, 5257-5262.
[49] a) M. I. Hassan, B. Shajee, A. Waheed, F. Ahmad, W. S. Sly, *Bioorganic & medicinal chemistry* **2013**, *21*, 1570-1582; b) A. Liljas, K. Kannan, P.-C. Bergsten, I. Waara, K. Fridborg, B. Strandberg, U. Carlbom, L. Järup, S. Lövgren, M. Petef, *Nature New Biology* **1972**, *235*, 131.
[50] Z. Fisher, A. Y. Kovalevsky, M. Mustyakimov, D. N. Silverman, R. McKenna, P. Langan, *Biochemistry* **2011**, *50*, 9421-9423.
[51] a) S. Z. Fisher, M. Aggarwal, A. Y. Kovalevsky, D. N. Silverman, R. McKenna, *Journal of the American Chemical Society* **2012**, *134*, 14726-14729; b) N. Niimura, M. Takimoto-Kamimura, I. Tanaka, *Encyclopedia of Analytical Chemistry* **2016**; c) S. Toba, G. Colombo, K. M. Merz, *Journal of the American Chemical Society* **1999**, *121*, 2290-2302; d) S. Taraphder, C. M.

Maupin, J. M. Swanson, G. A. Voth, *The Journal of Physical Chemistry B* **2016**, *120*, 8389-8404.
[52] a) R. A. Venters, B. T. Farmer II, C. A. Fierke, L. D. Spicer, *Journal of molecular biology* **1996**, *264*, 1101-1116; b) R. A. Venters, B. E. Coggins, D. Kojetin, J. Cavanagh, P. Zhou, *Journal of the American Chemical Society* **2005**, *127*, 8785-8795.
[53] S. K. Vasa, H. Singh, P. Rovó, R. Linser, *The journal of physical chemistry letters* **2018**, *9*, 1307-1311.
[54] S.-R. Tzeng, C. G. Kalodimos, *Current opinion in structural biology* **2011**, *21*, 62-67.
[55] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, D. Kern, *Nature* **2007**, *450*, 913.
[56] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, D. Kern, *Nature* **2005**, *438*, 117.
[57] B. Bardiaux, T. Malliavin, M. Nilges, in *Protein NMR techniques*, Springer, **2012**, pp. 453-483.
[58] H. Ma, A. Li, K. Gao, *ACS Omega* **2017**, *2*, 8414−8420.
[59] C. J. Craven, N. M. Derix, J. Hendriks, R. Boelens, K. J. Hellingwerf, R. Kaptein, *Biochemistry* **2000**, *39*, 14392-14399.
[60] K. Grohe, K. T. Movellan, S. K. Vasa, K. Giller, S. Becker, R. Linser, *Journal of biomolecular NMR* **2017**, *68*, 7-17.
[61] K. M. Merz Jr, *Journal of the American Chemical Society* **1991**, *113*, 406-411.
[62] a) A. G. Palmer III, *Chemical reviews* **2004**, *104*, 3623-3640; b) L. E. Kay, D. A. Torchia, A. Bax, *Biochemistry* **1989**, *28*, 8972-8979.
[63] a) K. H. Sippel, A. H. Robbins, J. Domsic, C. Genis, M. Agbandje-McKenna, R. McKenna, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2009**, *65*, 992–995; b) K. Hakansson, M. Carlsson, L. A. Svensson, A. Liljas, *J. Mol. Biol.* **1992**, *227*, 1192.
[64] S. K. Vasa, H. Singh, P. Rovó, R. Linser, *J. Phys. Chem. Lett.* **2018**, *9*, 1307−1311.
[65] S. Z. Fisher, A. Y. Kovalevsky, J. F. Domsic, M. Mustyakimov, R. McKenna, D. N. Silverman, P. A. Langan, *Biochemistry* **2009**, *49*, 415-421.
[66] a) L. Jaremko, M. Jaremko, K. Giller, S. Becker, M. Zweckstetter, *Science* **2014**, *343*, 1363–1366; b) N. Bertheleme, P. Seok Chae, S. Singh, D. Mossakowska, M. M. Hann, K. J. Smith, J. A. Hubbard, S. J. Dowell, B. Byrne, *Biochim. Biophys. Acta* **2013**, *1828*, 2583-2591.
[67] a) D. M. West, University of Florida **2012**; b) T. O. Wambo, The University of Texas at San Antonio **2017**.
[68] H. Shimahara, T. Yoshida, Y. Shibata, M. Shimizu, Y. Kyogoku, F. Sakiyama, T. Nakazawa, S.-i. Tate, S.-y. Ohki, T. Kato, H. Moriyama, K.-i. Kishida, Y. Tano, T. Ohkubo, Y. Kobayashi, *J. Biol. Chem.* **2007**, *282*, 9646–9656.
[69] R. A. Venters, B. T. Farmer, C. A. Fierke, D. S. Leonard *J. Mol. Biol.* **1996**, *264*, 1101–1116.
[70] a) O. Jardetzky, M. D. Finucane, *Molecular Physics* **1998**, *95*, 1127-1136; b) K. Grohe, K. T. Movellan, S. K. Vasa, K. Giller, S. Becker, R. Linser, *J. Biomol. NMR* **2017**, *68*, 7-17.
[71] N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, H. Stark, *Nature* **2010**, *466*, 329.
[72] a) A. Abragam, *The Principles of Nuclear Magnetism*, Oxford University Press, London, **1961**; b) A. G. Palmer, F. Massi, *Chem. Rev.* **2006**, *106*, 1700–1719.
[73] a) P. Schanda, M. Ernst, *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *96*, 1-46; b) V. Chevelkov, U. Fink, B. Reif, *J. Am. Chem. Soc.* **2009**, *131*, 14018-14022; c) P. Schanda, B. H. Meier, M. Ernst, *J. Am. Chem. Soc.* **2010**, *132*, 15957–15967; d) P. Rovó, R. Linser, *J. Phys. Chem. B* **2017**, *121*, 6117–6130.
[74] O. D. Ron, M. D. Robert, J. P. Grossman, X. Huafeng, E. S. David, *Annu. Rev. Biophys.* **2012**, *41*, 429-452.
[75] a) B. Vögeli, J. Orts, D. Strotz, C. Chi, M. Minges, M. A. Wälti, P. Güntert, R. Riek, *Journal of Magnetic Resonance* **2014**, *241*, 53-59; b) B. Vögeli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *78*, 1-46.
[76] I. Solomon, *Phys. Rev.* **1955**, *99*, 559-565.
[77] K. Saksela, P. Permi, *FEBS Lett.* **2012**, *586*, 2609-2614.
[78] O. F. Lange, N.-A. Lakomek, C. Fares, G. F. Schroeder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmueller, C. Griesinger, B. L. d. Groot, *Science* **2008**, *320*, 1471-1475.

[79] G. W. Vuister, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 7772-7777.
[80] S. Grzesiek, J. Anglister, A. Bax, *J. Magn. Reson. Ser. B* **1993**, *101*, 114-119.
[81] T. M. Logan, E. T. Olejniczak, R. X. Xu, S. W. Fesik, *FEBS Lett.* **1992**, *314*, 413-418.
[82] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins: Structure, Function, and Bioinformatics* **2005**, *59*, 687-696.
[83] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283-298.
[84] K. Takemura, A. Kitao, *J. Phys. Chem. B* **2012**, *116*, 6279-6287.
[85] K. T. Debiec, D. S. Cerutti, L. R. Baker, A. M. Gronenborn, D. A. Case, L. T. Chong, *J. Chem. Theory Comput.* **2016**, *12*, 3926-3947.
[86] P. Rovó, C. A. Smith, D. Gauto, B. L. de Groot, P. Schanda, R. Linser, *J. Am. Chem. Soc.* **2019**, *141*, 858-869.
[87] P. Rovó, R. Linser, *ChemPhysChem* **2018**, *19*, 34-39.
[88] N. A. Farrow, O. Zhang, A. Szabo, D. A. Torchia, L. E. Kay, *J. Biomol. NMR* **1995**, *6*, 153-162.
[89] V. Chevelkov, Y. Xue, R. Linser, N. Skrynnikov, B. Reif, *J. Am. Chem. Soc.* **2010**, *132*, 5015-5017.
[90] T. Gullion, J. Schaefer, *J. Magn. Reson.* **1989**, *81*, 196-200.
[91] K. Grohe, E. Nimerovsky, H. Singh, S. K. Vasa, B. Söldner, B. H. Vögeli, C. M. Rienstra, R. Linser, *Chem. Commun.* **2019**, *55*, 7899-7902.
[92] C. Massenet, S. Chenavas, C. Cohen-Addad, M.-C. Dagher, G. Brandolin, E. Pebay-Peyroula, F. Fieschi, *J. Biol. Chem.* **2005**, *280*, 13752-13761.
[93] R. Linser, U. Fink, B. Reif, *J. Am. Chem. Soc.* **2010**, *132*, 8891-8893.
[94] P. Schanda, M. Ernst, *Prog. Nucl. Mag. Res. Sp.* **2016**, *96*, 1-46.
[95] A. A. Smith, M. Ernst, B. H. Meier, *J. Chem. Phys.* **2018**, *148*, 045104.
[96] F. Vitalini, A. S. J. S. Mey, F. Noé, B. G. Keller, *J. Chem. Phys.* **2015**, *142*, 084101.
[97] a) F. Hoffmann, F. A. A. Mulder, L. V. Schäfer, *J. Phys. Chem. B* **2018**, *122*, 5038-5048; b) F. Hoffmann, M. Xue, L. V. Schäfer, F. A. A. Mulder, *Phys Chem Chem Phys* **2018**, *20*, 24577-24590.
[98] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213-223.
[99] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19-25.
[100] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8592.
[101] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
[102] F. J. Blanco, Á. R. Ortiz, L. Serrano, *J. Biomol. NMR* **1997**, *9*, 347-357.
[103] B. Vögeli, P. Güntert, R. Riek, *Mol. Phys.* **2013**, *111*, 437-454.
[104] M. Congreve, C. W. Murray, T. L. Blundell, *Drug Discovery Today* **2005**, *10*, 895-907.
[105] S. Hwang, C. Öster, V. Chevelkov, K. Giller, S. Lange, S. Becker, A. J. J. o. B. N. Lange, **2019**, *73*, 281-291.
[106] L. Wang, R. A. Friesner, B. J. Berne, *The Journal of Physical Chemistry B* **2011**, *115*, 9431-9438.