DISS. ETH NO. 23867

eNOE Method Development And Applications To Protein Allostery

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

DEAN STROTZ

MSc ETH in Biology, ETH Zurich

born on 17.09.1973

citizen of

Zurich, Switzerland

accepted on the recommendation of

Prof. Dr. Roland Riek, examiner

Prof. Dr. Gunnar Jeschke, co-examiner

PD. Dr. Beat Vögeli, co-examiner

2016

This work is dedicated to my partner Pauline and our wonderful and courageous children Arthur and Leo. My achievement is hopefully also some reward for your sacrifices, may your independent spirits help you to find and follow your personal calling!

Acknowledgments

I would like to thank our supervisor and group leader Prof. Dr. Roland Riek for giving me the freedom to follow my own instinct and develop my own approach to solving scientific problems, while at the same time keeping things in perspective. Thank you Roland for your reinvigorating ideas, with helpful twists and turns whenever the process got sticky.

I want to thank the person I worked with the closest over the past few years and who was a great supervisor and friend, offering steadiness and continuity: Beat Vögeli. Together we managed to achieve a great many things, complementing each other in some skills, guided however by your steady and never relenting scientific stride.

Then, there is of course Julien Orts. With your friendly presence and your ingenious science and thorough advice, you are an inspiring person to be around. Thank you Julien for sharing these past few years with me.

Simon, your contributions are invaluable. I feel very privileged to have had the opportunity to collaborate with you and hope we can continue to do so in the future.

There are also Chi Celestine and Stefan Bibow, with whom I had the pleasure to collaborate on scientific projects and who also kept spirits flying high.

Of course, I must not forget to thank my fellow PhD students and all other group members and our patient and helpful secretary Kristina Comiotto. Thank you!

Last but not least, I would like to thank my parents, Robert and Beatrice, and my in-laws, Stanislaus and Irène: without your kind support, encouragement and baby-sitting it would not have been possible to be a father while first gaining a bachelor degree in Biophysics in Los Angeles, then a masters in Biology at ETH and afterwards work on a doctorate!

Contents

Abstract

Three-dimensional structure elucidation at atomic resolution provides key information to understanding biomolecular function. However, detailed understanding of enzyme activity, protein folding and signal transduction throughout biological molecules requires information of internal motion and dynamics in addition to the three-dimensional structure.

There are essentially two techniques to determine molecular structure at atomic resolution; X-ray and NMR. Mainly due to the relatively fast execution, X-ray is the preferred structure determination tool for many applications, if the molecule forms diffracting crystals. NMR however, is the only technique that can determine molecular structures and dynamics at atomic resolution in solution and at near physiological conditions. In the early 1980's the Wüthrich group at ETHZ was developing the conceptional framework to calculate protein structures in solution using NMR spectroscopy. Their approach to calculate structures required sequence-specific resonance assignment and was based on scalar couplings, hydrogen bond information obtained from amide exchange rates and short-range nuclear Overhauser enhancement (NOE) measurements. The structures were calculated using distance geometry computation.

20 years after the Wütherich's Group seminal achievement and the work on NOE based structure elucidation by many others that followed, the Riek group was picking up the loose ends on the interpretation of NOE data in order to obtain more exact NOE (eNOE) data. The eNOE data is reflective not only of a single static structure but an occupied conformational space and dynamics. To reveal this information a novel ensemble based structure calculation protocol in combination with the analysis of NOE data was required and developed.

In the first part of this thesis (I. Methods) we present the current state of development in methods pertaining to the analysis of distance restraint information from NOE measurements and structure calculation. First we analyze the experimental accuracy of our eNOE data, dissecting in detail the experimental setup to measure the data and the errors contained therein. We then show an extension of our data set that allows analysis of NOE data in more crowded spectra resulting from large proteins with overlapping resonance signals. In order to extract the eNOE data from the large NOE data sets, a comprehensive software package was realized with eNORA2. The software package offers two complementary ways to analyze the NOE data and correct for spin

diffusion, the unwanted and indirect transfer of magnetization during NOE measurements. According to project and sample specific conditions a full-matrix (FM) or three-spin approach for spin diffusion correction may be used. The approach allows for various labeling shemes in deuterated samples enabling the eNOE analysis of large biomolecular systems. Structure calculation is presented in a novel and complementary protocol to the ensemble based structure calculation protocol previously established in the Riek group (minimal conformational space required to fullfil the data). In this novel approach we use replica exchange simulation and maximum entropy reweighing (maximum conformation space required to fullfil the data).

In the second part of this thesis (II. Applications) we illustrate the potential impact on the comprehensive elucidation of the action of biomolecules at atomic resolution using our high-resolution solution state NMR data. One of the major challenges presents itself in the accurate representation, interpretation and analysis of the data. We present the current state of our approach on the examples of Pin1-WW and Cyclophilin A. The Pin1-WW is a peptidyl-prolyl cis/trans isomerase of phospho-Ser/Thr-Pro motifs and key cell-cycle regulator playing a role in human diseases such as Alzheimer's disease, hepatitis C infection and cancer. In order to study information transduction in human Pin1-WW we use replica exchange simulations as basis for a detailed analysis. An extension of methods for the analysis in a statistical manner was required in order to determine atomistic thermodynamic models. Our analysis provides a basis for the understanding of the multi-functionality of the Pin1 protein. We then proceed to employ the ensemble based structure calculation protocol to analyze the allosteric coupling of the ligand-binding event at loop1 to the interface between the Pin1-WW domain and the Pin1-PPIase. Cyclophilin A is another peptidyl-prolyl cis/trans isomerase involved in signal transduction, protein folding, responding to inflammatory stimuli and playing a key role in several human diseases such as viral infections, neurodegeneration, cancer, HIV-1, hepatitis C and asthma. With our eNOE based ensemble structure calculation protocol and analysis of Cyclophilin A, we observe the presence of an open and closed state in the apo form of the protein. This finding infers conformational sampling of Cyclophilin A and a preorganization for catalysis. Furthermore we illustrate the allosteric coupling network connecting the ligand-binding loop to the active site.

Zusammenfassung

Strukturaufklärung steht nur am Anfang eines Verständnisses von molekularen Interaktionen in atomarer Auflösung. Um enzymatische Aktivität, Proteinfaltungsprozesse und Signaltransport durch Biomoleküle zu verstehen, benötigt es zusätzlich detaillierte Kenntnisse über interne Bewegung und Dynamik.

Prinzipiell gibt es zwei Techniken zur Aufklärung von molekularen Strukturen in atomarer Auflösung; Kristallographie und Kernspinresonanz (NMR). Hauptsächlich wegen der raschen Ausführung ist Kristallographie die bevorzugte Technik für viele Anwendungen, unter der Voraussetzung, dass das Molekül sich gut kristallisieren lässt. Kernspinresonanz ist allerdings die einzige Technik, mit welcher molekulare Struktur und Dynamik in molekularer Auflösung nahe an physiologischen Bedingungen ermittelt werden können. In den frühen 80er Jahren hat die Wüthrich Gruppe an der ETHZ die konzeptionellen Rahmenbedingungen geschaffen für die Berechnung von Proteinstrukturen in Lösung mittels Kernspinresonanz. Ihre Methode der Berechnung von Strukturen verlangte die kettenspezifische Resonanzbestimmung und stützte sich auf skalare Kopplungen, Informationen von Wasserstoffbrücken aus Amidaustauschraten und Kurzdistanzen aus Nuklearen Overhauser Effekt (NOE) Messungen. Die Strukturen wurden mittels Distanz-Matrix-Geometrie Berechnungen ermittelt.

20 Jahre nach den zukunftsweisenden Errungenschaften der Wüthrich Gruppe und den Beiträgen von vielen anderen Wissenschaftlern zur Weiterentwicklung von NOE basierten Strukuturberechnungen hat die Riek Gruppe die losen Enden der Interpretation von NOE Daten neu aufgegriffen mit dem Ziel einer exakteren Auswertung der Daten. Es hat sich gezeigt, dass diese Daten nicht nur representativ für ein einziges und statisches Ensemble von Strukturen sind, sondern Einsicht in den belegten konformellen Raum und die Dynmik vermitteln. Damit ist ein neues Protokol für die Analyse von exakten NOE-Daten (eNOE) und die Multi-Zustands-Ensembles basierte Strukturberechnung geschaffen worden.

Im ersten Teil dieser Doktorarbeit (I. Methods) präsentieren wir den gegenwärtigen Stand der Entwicklungen in Methoden zur Analyse der Distanzinformation, die in NOE Daten enthalten sind, aber auch Entwicklungen in Strukturberechnungsmethoden. Zuerst analysieren wir die experimentelle Genauigkeit userer eNOE Daten, indem wir den experimentellen Aufbau unserer Experimente und die darin enthaltenen Fehler im Detail analysieren. Danach zeigen wir eine Möglichkeit zur Vergrösserung des Datensatzes, welche die Analyse der Daten von grossen Proteinen mit vielen überlappenden

Resonanzsignalen erlaubt. Wir realisierten die umfassende Software eNORA2, um die Extraktion von eNOE Daten aus so grossen Datensätzen, wie sie in NOE Experimenten anfallen, zu ermöglichen. Die Software offeriert zwei sich ergänzende Wege für die Analyse der Daten und die Korrektur von Spindiffusion, den unerwünschten, indirekten Übertrag von Magnetisierung während der Messung von NOE Experimenten. Aufgrund von projekt- und probenspezifischen Kriterien kann entweder die Vollmatrix- oder die Dreispinkorrektur für Spindiffusion gewählt werden. Mit der Dreispinkorrektur für Spindiffusion, ermöglichen wir die Berücksichtigung von verschiedenen Markierungsmöglichkeiten in deuterierten Proben und bereiten damit einen Weg zur Analyse von grossen biomolekularen Systemen mittels eNOE Daten. Komplementär zur bereits etablierten, auf Multi-Zustands-Ensembles basierten Strukturberechnung (Ermittlung des minimalen konformellen Raumes der die Daten reproduziert), präsentieren wir auch ein neues Strukturberechnungsprotokoll wozu wir Replika-Austausch Simulationen und Maximum-Entropie-Gewichtung (Ermittlung des maximalen konformellen Raumes der die Daten zu reproduziert) verwenden.

Im zweiten Teil der Doktorarbeit (II. Applications) illustrieren wir den potentiellen Einfluss unserer hochaufgelösten Strukturdaten auf das Verständnis der Funktion von Biomolekülen in atomarer Auflösung. Eine grosse Aufgabe stellt sich dabei in der korrekten Darstellung, Interpretation und Analyse der Daten. Wir präsentieren unser gegenwärtiges Verständnis der Aussagekraft der eNOE Daten anhand der Pin1-WW Domäne und Cyclophilin A. Die Pin1-WW ist eine Peptidyl-Prolyl-cis/trans-Isomerase für phosphorylierte Serin/Threonin-Prolin-Motive. Diese Domäne ist ein Schlüsselprotein in der Regulierung des Zell-Zyklus und spielt eine signifikante Rolle in menschlichen Krankheiten wie Alzheimer, Hepatitis-C-Infektionen und Krebs. Um die Übertragung von Information in der menschlichen Pin1-WW Domäne zu studieren, verwenden wir Replika-Austausch Simulationen als Basis für unsere Analyse. Um atomare, thermodynamische Modelle entwicklen zu können, verwenden wir eine statistische Analyse. Unsere Analyse bereitet damit eine Basis zum Verständnis der Multifuntionalität des Pin1-Enzyms. Wir benutzen das auf Multi-Zustands-Ensembles basierte Strukturberechnungsprotokol, um die Bindung des Liganden an die Schleife (Loop 1) und die allosterische Kopplung dieses Events zur Interaktionsoberfläche zwischen der Pin1-WW Domäne und der Pin1-PPIase zu analysieren. Cyclophilin A ist eine weitere Peptidyl-Prolyl-cis/trans-Isomerase, welche ein Bestandteil von Signalübertragungswegen und Proteinfaltungsmechanismen ist, auf Entzündungen regiert und eine wichtige Rolle spielt in verschiedenen menschlichen Krankheiten wie

viralen Infektionen, Neurodegeneration, Krebs, HIV-1, Hepatitis-C und Asthma. Mit unseren eNOE Daten und dem auf Multi-Zustands-Ensembles basierten Strukturberechnugsprotokol analysieren wir Cyclophilin A und können eine offene und geschlossene Konformation in der Apoform feststellen. Diese Festellung suggeriert eine konformelle Abtastung von Cyclophilin A und eine Vororganisation für die Katalyse. Im weiteren illustrieren wir mit unserer Analyse die allosterische Kopplung zwischen der Bindungsstelle für den Liganden in der Schleife (Loop 1) und der enzymatisch aktiven Stelle des Enzyms.

1. General Introduction

This chapter is in part based on the following publications:

Beat Vögeli, Julien Orts, **Dean Strotz**, Peter Güntert, Roland Riek, Discrete Three-Dimensional Representation of Macromolecular Motion from eNOE-based Ensemble Calculation, CHIMIA 2012, doi: 10.2533/chimia.2012.787. All Authors contributed to writing of the intitial manuscript.

Beat Vögeli, Julien Orts, **Dean Strotz**, Celestine Chi, Martina Minges, Marielle Aulikki Wälti, Peter Güntert, Roland Riek, Towards a true protein movie: A perspective on the potential impact of the ensemble-based structure determination using exact NOEs, J Magn Reson 2014, doi: 10.1016/j.jmr.2013.11.016. All Authors contributed to writing of the intitial manuscript.

1.1. Structural biology and exchange dynamics by NMR

Proteins are inherently dynamic systems with inter-change between structural states on time scales between 10^{-12} and 10^{5} seconds and beyond. The dynamic exchanges between these states have been omnipresent during evolution and play an important role in the action of biomolecules. Furthermore, it has been suggested that concerted motions may be a pivotal factor in the enzymatic function of proteins and in protein-ligand interaction. This includes synchronization of the internal motions of a protein, which have been shown to influence the kinetics of catalysis [1, 2] or to minimize entropic losses due to complex formation upon ligand binding by reducing the conformational entropy of the protein [3, 4]. Internal dynamics may also be involved in allosteric mechanisms [5, 6]. The protein folding process is another example since folding comprises dynamics of complex nature on large scales in both time and space. Molecular dynamics simulations, thermodynamic and kinetic studies have often been used to describe protein folding on a molecular basis [7, 8].

One of the major challenges in structural biology is thus a comprehensive description of the 3D structures and the exchange dynamics between structural states at atomic resolution with the ultimate goal of an experimental data-based movie of a

biomolecule. While structure determination of biomolecules at atomic resolution by NMR or X-ray crystallography is well established, the description of structural landscapes of proteins as well as the dynamic interchange between the various conformations are still largely incomplete. The average conformation, as obtained from NMR spectroscopy by the established structure determination protocols, has been described as representing "… the shape of the molecule as it would be seen on a photograph taken at low shutter speeds" [9], and furthermore, "...the average derived from spectroscopic data represents a virtual structure devoid of physical meaning" [9]. This view certainly reflects the fact that a NMR observable is an averaged property rather than a function of an averaged structure. For example, if a molecule fluctuates between two states, the observed Nuclear Overhauser Enhancement (NOE) is the average of the two individual NOEs. The extracted distance then would yield a value that is between those of the two states. In reality, the extracted distance is never sampled at all. Furthermore, the well established standard structure determination protocols used in NMR spectroscopy make use of the overabundant number of experimentally readily accessible Nuclear Overhauser Enhancement (NOE) rate constants – typically up to 20 per residue in small proteins [10]. These rate constants are employed in a semi-quantitative manner at most because the measurement of NOEs is flawed by spin diffusion, which is the indirect magnetization transfer between the two spins of interest via other nearby spins, and is also affected by peak overlap, low signal-to-noise ratio and technical limitations. The practice of translating NOE cross peaks from NOESY spectra semi-quantitatively into upper limit distances, following the proportionality of the NOE cross-relaxation rate with the inverse 6^{th} power of the distance between two (isolated) interacting spins, originates from the 1980s when it proved difficult to determine NOE rates and to convert them into exact distances [10, 11].

Routine analysis of fast and slow local dynamics is done by means of relaxation measurements and is mostly restricted to backbone ^{15}N-^{1}H moieties and methyl groups [12, 13]. However, even though NMR relaxation phenomena provide a great deal of insight into local motion the dynamic picture is still largely incomplete because it is difficult to detect translational or concerted motion. But, there is exciting progress in NMR-based methods towards a more holistic description of structural landscapes of proteins and the transitions between the various states. This includes measurements of residual dipolar couplings (RDCs) [14], relaxation dispersion (CPMG) [15], cross-correlated relaxation (CCR) [16, 17], paramagnetic relaxation enhancement (PRE) [18], and exact nuclear Overhauser enhancement (eNOE) data [19, 20], in part combined with

molecular dynamics simulation, structure prediction software or ensemble-based structure calculations [21-26] .

The Riek Group developed an ensemble-based structure determination protocol using a modified version of CYANA and ensemble-averaged distance restraints obtained from eNOE rates [20, 25-29], whereby the ensemble covers the minimal conformational space required to fullfil the experimental data. We recently developed another ensemble-based structure determination protocol (presented in chapter 5) using a combination of replica-exchange simulations with chemical shift data through maximum entropy reweighing [30]. In this approach the ensemble covers the maximal conformational space allowed by the experimental data and the ensemble-averaged distance restraints obtained from exact NOE (eNOE) rates [29] are used to validate the states of the calculated ensemble [30].

These developments were possible in light of several advances to measure and interpret NOEs but also due to progress in NMR. This progress includes (i) the availability of high magnetic fields and heteronuclear spectroscopy (i.e. ^{15}N and ^{13}C-resolved [^{1}H, ^{1}H]-NOESY) which reduce peak overlap significantly, (ii) considerable increase in sensitivity due to cryoprobe technology, which enables the use of shorter NOE mixing times in order to reduce the spin diffusion substantially (note that the desired contribution to the cross peak grows linearly with the mixing time, while the contribution from spin diffusion increases initially with the square of the mixing time), and (iii) increased computer power enabling analytical advances such as the full relaxation matrix formalism applied to the Solomon equation in order to correct NOE restraints for spin diffusion [28, 31, 32] and (iv) recent progress in simulation techniques [33-41].

1.2. Extraction of distance restraints

The Riek Group demonstrated that it is possible to obtain very precise and accurate H^N–H^N NOEs in both deuterated and protonated protein samples with the eNOE protocol [20, 27]. NOE buildups are measured with optimized 3D-^{15}N/^{13}C-resolved [^{1}H, ^{1}H]-NOESY experiments on cryogenic probes and converted into precise distances. For example, distances up to 5 Å obtained from a perdeuterated ubiquitin sample have a random error of only ≈ 0.07 Å [20]. This is considerably smaller than the 0.24 Å pairwise rms deviations from distances extracted from corresponding high-resolution NMR or X-ray structures.

The extraction of exact eNOEs between spins *i* and *j* is severely hampered by spin diffusion, which is relayed magnetization transfer via neighboring spins [42, 43]. Two related approaches to extract exact distances that are distinguished mainly by the methods to calculate the effects of spin diffusion were developed. The experimental procedure encoded in the MATLAB program eNORA2 is described in chapter 4. In brief summary: The full matrix approach, which is used for fully protonated samples, takes into account spin diffusion in that the magnetization transfer pathways between all spins are simultaneously active. In the three-spin approach [20, 27], suited for partially deuterated proteins, the individual correction contributions from each neighboring spin are obtained from the exact solution of three spin systems and summed up. The implementation of eNORA2 [to be published] allows the analysis of various methyl-labeling schemes for partially deuterated, large biomolecular systems.

1.2.1. The mathematical basis of the eNOE

The NOE cross-relaxation rate between two spins K=1/2 and L=1/2 is given by [44, 11]

$$\sigma_{KL} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma^4 h^2}{40\pi^2} \frac{1}{(r_{KL}^{\text{rigid}})^6} \left[J(0) - 6J(2\omega)\right] \tag{1}$$

where γ is the gyromagnetic ratio of nucleus K, ω is the spectral frequency of the nuclei, μ_0 is the permeability in vacuum, and h denotes Planck's constant. r_{KL}^{rigid} is the internuclear distance in a hypothetically rigid structure. A simple expression for the spectral density J obtained under the assumption of isotropic molecular tumbling [44] is

$$J(\omega) = S_{KL}^{\text{fast}\,2} \frac{\tau_c}{1 + (\tau_c \omega)^2} + \left((r_{KL}^{\text{rigid}})^6 \left\langle \frac{1}{r_{KL}^6} \right\rangle - S_{KL}^{\text{fast}\,2} \right) \frac{\tau_{\text{tot}}}{1 + (\tau_{\text{tot}} \omega)^2} \tag{2}$$

with

$$\frac{1}{\tau_{\text{tot}}} = \frac{1}{\tau_c} + \frac{1}{\tau_{\text{int}}} \tag{3}$$

where τ_c is the rotational correlation time of the molecule and τ_{int} is the correlation time for internal motion. The angular brackets denote a Boltzmann ensemble average and $S_{KL}^{\text{fast}\,2}$ is an order parameter for fast internal motion [45],

$$S_{KL}^{\text{fast}\,2} \equiv (r_{KL}^{\text{rigid}})^6 \frac{4\pi}{5} \sum_{q=-2}^{2} \left\langle \frac{Y_{2q}(\theta_{KL}^{\text{mol}}, \phi_{KL}^{\text{mol}})}{(r_{KL})^3} \right\rangle^2 \tag{4}$$

(Note that the influence of anisotropic tumbling can be neglected in most cases as discussed in ref. [27]).

It is evident from equations 1-4 that the NOE is a time- and ensemble-averaged observable containing both structural and dynamical information. However, the most common way to extract distances from the measured NOE cross-relaxation rate is to assume the presence of a rigid molecule (i.e. $S_{KL}^{\text{fast}\,2} = 1$ and $\left\langle \frac{1}{r_{KL}^6} \right\rangle = \frac{1}{\left(r_{KL}^{\text{rigid}}\right)^6}$). This simplification of the theory is usually accompanied by the assumption that the NOE rate is proportional to the cross peak intensity in NOESY spectra. The NOE rates are converted into distances by making use of the relative insensitivity of the NOE to fast motion (i.e. $S_{KL}^{\text{fast}\,2} = 1$ in equations 1-4), which was shown to be valid for H-H spin pairs if the local H-X order parameters for fast motion are larger than 0.5 [46]. This is most often the case in folded proteins.

1.3. From eNOE-derived distance restraints to 3D structural ensemble

1.3.1 CYANA based ensemble calculation (minimal conformational space required)

The exact nature of the eNOE results in an extremely tight structure bundle, here shown on the example of GB3 (with a root-mean-square deviation for the backbone atoms of 0.11 Å), in particular when compared to a conventional structure calculation using semi-quantitative NOEs (compare Figures 1 left with middle). The structure bundle coincides closely with the RDC-optimized X-ray structure [47, 48]. However, there are many distance restraint violations indicating that the structure does not agree with the experimental data. The large number of violations of experimental restraints is attributed to the motion-averaged nature of the measured NOE (equations 1-4) since the structure calculation protocol is based on a single static structure.

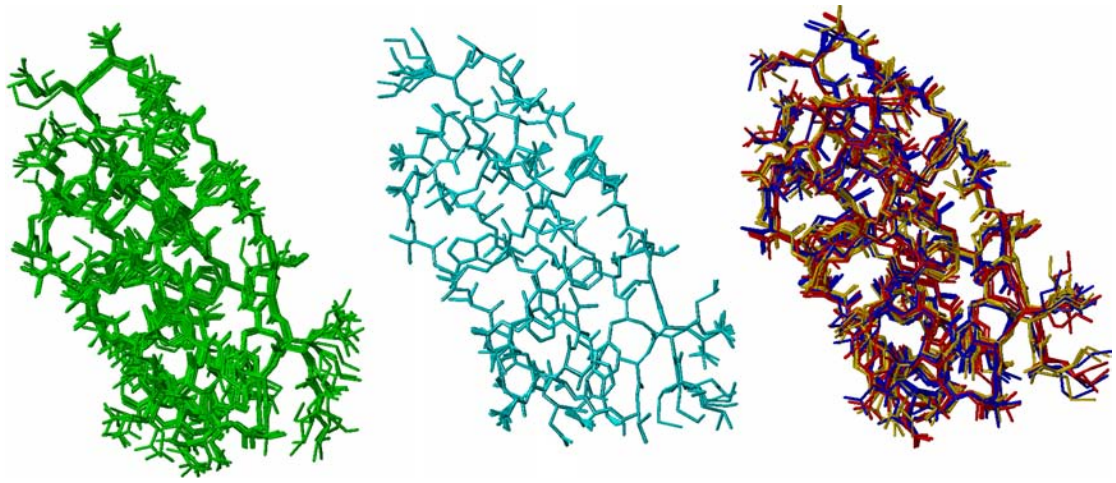


Figure 1: Heavy-atom structural representations of GB3 following either the conventional protocol with NOEs as experimental input, the conventional protocol with eNOEs or the ensemble-based protocol with eNOEs. Left: Bundle calculated with a conventional protocol based on standard NOE measurements. Nine conformers are shown. Middle: Single-state bundle calculated with eNOEs. Nine conformers are shown. Right: 3 three-state ensembles obtained from eNOEs. The three most similar structures from each three-state conformer are grouped into sub-bundles shown in gold, red and blue. Reprinted by permission from Macmillan Publishers Ltd: Vögeli, Kazemi, Güntert and Riek, Nat. Struct. Mol. Biol. 2012, 19, 1053-1057, copyright 2012. [25]

It is obvious that exact NOEs reveal a wealth of information about internal motion that is sacrificed in routine structure calculation protocols. The resulting effect on the structure calculation when using traditional methods is depicted in Figure 2. It is clear that an ensemble of structural states of which the averaged NOEs have to match the experimental eNOEs would be a better representation of the structure [49].

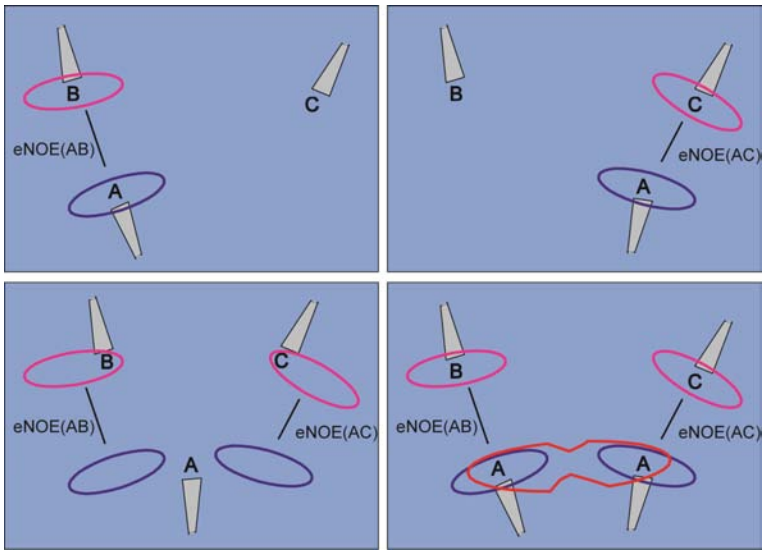


Figure 2. Schematic representation of the effect of multiple eNOEs on structure calculation. eNOEs between spins A and B, and A and C are measured. Use of only the A-B eNOE leads to the situation shown in the left top panel and similarly for the A-C eNOE in the right top panel. The blue ellipsoid shows the expected position of spin A and the pink ellipsoid of spins B and C. Simultaneous use of both eNOEs in a single-state calculation would place spin A between the two positions and pull spins B and C towards each other (left bottom panel). In a two-state ensemble, spin A is restricted to the red area, where it is once located in the left and once in the right ellipsoid (right bottom panel).

According to the ergodic hypothesis the impact of protein motion on the NOE measurement can be described by an ensemble of structural states. To include the dynamic dependence of eNOEs in structure calculation (equations 1-4) an ensemble-based protocol was established (within the software package CYANA) [25, 26]. The protocol requests that the experimental restraints are fulfilled by a set of structural states rather than by a single structure. In contrast to the standard structure determination protocol, it therefore takes into account that the NOE is a time- and ensemble-averaged parameter.

To clarify the discussion we provide the following definitions:

- A structure is defined by a bundle (or an ensemble) of conformers fulfilling the experimental data.
- A conformer is the result of one individual structure calculation that fulfills the experimental data and may be composed of one or more states.
- A state is one set of coordinates for all atoms of a molecule. If there are multiple states they fulfill the experimental data on average and not individually.
- Sub-bundles are formed by sorting the states according to structural similarity in the region of interest. There are as many sub-bundles as there are states in a conformer, and each sub-bundle comprises as many conformers as the original structure bundle. This requires for each state to belong to exactly one sub-bundle. The sub-bundle for each structural state is a measure of the precision of the individual structural states similar to the conventional bundle representation.

For instance, a multi-state structure calculation that aims at fulfilling the experimental data with 3 states per conformer could be started from 100 initial random conformers. The 20 conformers with lowest CYANA target function are selected to represent the solution structure. Each of these 20 conformers comprises 3 states. These states are grouped into 3 sub-bundles according to structural similarity in a given loop region of interest, i.e. each of the 3 states of a conformer is assigned to a different sub-bundle. Hence, each of the 3 sub-bundles comprises 20 conformers. These "sub-conformers" consist of a single state.

To avoid divergence among the structural states that is not implied by the experimental restraints, we impose “bundling restraints”, i.e. weak harmonic restraints that minimize the distances between corresponding atoms in different states [25, 50]. Using this protocol, it is found that the ensemble consisting of three three-state structures

shown in Figure 1 describes the experimental data well. This structural ensemble is a compact experiment-based representation of GB3 covering its conformational space in solution. Because of the bundling restraints used for the calculation, the ensemble covers the minimal conformational space required to fulfill the experimental data.

A detailed inspection of the structural ensemble shows that the three structural states are distinct from each other (Figure 1, right) with pronounced differences for the β-sheet and attached loops (Figure 3). The timescale of exchange is most likely on the sub-ms time scale, because slower motions would result in line broadening or resonance doubling that is not observed in the spectra, and because the three-state ensemble is in fair agreement with the RDC-derived order parameters that are sensitive to motion faster than milliseconds [48]. The ensemble reveals further insights into structure and dynamics such as side chain rotamer states, conformational exchange dynamics of the side chains, and correlation between backbone and side chain configurations (for more details see ref. [26]).

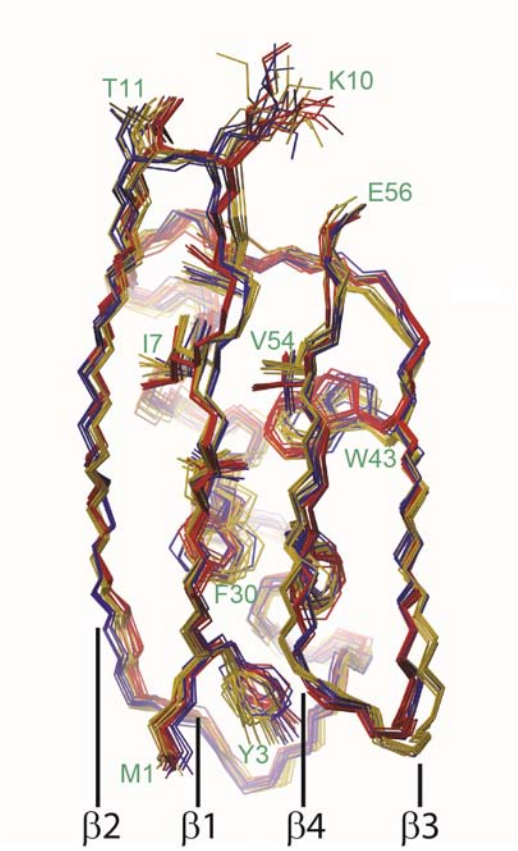


Figure 3: The 3D ensemble structure of GB3. Representation of three states of GB3 obtained from the ensemble-based protocol using eNOEs. The three most similar structures from each three-state conformer are grouped into sub-bundles shown in gold, red and blue. For each sub-bundle (ensemble) 9 conformers are shown. The backbone, the side chains of the hydrophobic core and the two solvent-exposed residues K10 and T11 are also shown. The termini and the side chains are labeled with the residue number. Figure is taken from B. Vögeli, J. Orts, D. Strotz, P. Güntert, R. Riek, Chimia, 2012, 66, 787-790. [51]

1.3.2 Replica-exchange simulations combined with chemical shift data through maximum entropy reweighing (maximum allowed conformational space)

Our ensemble-based structure determination using a combination of replica-exchange molecular dynamics simulations with chemical shift data applying maximum entropy reweighing, yields an ensemble covering the maximal conformational space allowed by the experimental data [30]. In this case the ensemble-averaged distance restraints obtained from eNOE rates do not guide the simulation but are used to validate the states of the calculated ensemble (chapter 5 [30]).

The conformational phase space is segmented into discrete states using K-means clustering and reweighed via the backbone chemical shift data applying the maximum entropy principle (chapter 5 [30]). Additional experimental data, such as the averaged translational and radial information obtained from eNOEs, cross-correlated relaxation rates [52], and backbone and side-chain scalar couplings, is used to validate the reweighed ensembles [30].

Having obtained representative structure ensembles at multiple temperatures, we are in a position to analyze equilibrium thermal unfolding, determine free energy landscapes and temperature dependent probabilities of conformational clusters (K-means) to be in a certain conformational space [30]. The analysis does not currently allow the dissection of chemical exchange kinetics even though this may be possible in the future using an appropriate statistical framework [53-55]. The chosen approach further illustrates the sensitivity to and the potential of eNOEs to resolve sparsely populated states (chapter 5 [30]). Analysis using the statistical approach of Bayesian Markov state models [56, 57] reveals dynamic information of the obtained structure ensembles (chapter 6 [30]).

1.4. References

[1] Bruice TC, Benkovic SJ, Chemical basis for enzyme catalysis, Biochemistry 2000, 39, 6267-6274.

[2] Osborne MJ, Schnell J, Benkovic SJ, Dyson HJ, Wright PE, Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism, Biochemistry 2001, 40, 9846-9859.

[3] Frederick KK, Marlow MS, Valentine KG, Wand AJ, Conformational entropy in molecular recognition by proteins, Nature 2007, 448, 325-329.

[4] Tzeng S-R, Kalodimos CG, Protein activity regulation by conformational entropy, Nature 2012, 488, 236-240.

[5] Jain RK, Ranganathan R, Local complexity of amino acid interactions in a protein core, Proc Natl Acad Sci USA 2003, 101,111-116.

[6] Weinkam P, Pons J, Sali A, Structure-based model of allostery predicts coupling between distant sites, Proc Natl Acad Sci USA 2012, 109, 4875-4880.

[7] Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W, Atomic-level characterization of the structural dynamics of proteins, Science 2011, 133, 341-346.

[8] Eichmann C, Preissler S, Riek R, Deuerling E, Cotranslational structure acquisition os nascent polypeptides monitored by NMR spectroscopy, Proc Natl Acad Sci USA 2010, 107, 9111-9116.

[9] Jardetzky O, On the nature of molecular conformations inferred from high-resolution, NMR Biochim Biophys Acta 1980, 621, 227-32.

[10] Wüthrich K, NMR of Proteins and Nucleic Acid, Wiley, New York, 1986.

[11] Neuhaus D, Williamson MP, The Nuclear Overhauser Effect in Structural and Conformational Analysis, Wiley, New York, 2000.

[12] Kay LE, Torchia DA, Bax A, Backbone Dynamics of Proteins as Studied by Nitrogen-15 Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease, Biochemistry, 1989, 28, 8972-8979.

[13] Kay LE, NMR studies of protein structure and dynamics, J Magn Reson 2005, 173, 193-207.

[14] Salmon L, Bouvignies G, Markwick P, Blackledge M, Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales, Biochemistry 2011, 50, 2735-2747.

[15] Mittermaier A, Kay LE, New tools provide new insights in NMR studies of protein dynamics, Science 2006, 312, 224-228.

[16] Frueh D, Internal motions in proteins and interference effects in nuclear magnetic resonance, Prog Nucl Magn Reson Spectrosc 2002, 41, 305-324.

[17] Vögeli B, Yao L, Correlated Dynamics between Protein HN and HC Bonds Observed by NMR Cross Relaxation, J Am Chem Soc 2009, 131, 3668-3678.

[18] Iwahara J, Clore GM, Structure-independent analysis of the breadth of the positional distribution of disordered groups in macromolecules from order parameters for long variable-length vectors using NMR paramagnetic relaxation enhancement, J Am Chem Soc 2010, 132, 13346-13356.

[19] Koning TMG, Boelens R, Kaptein R, Calculation of the nuclear overhauser effect and the determination of proton-proton distances in the presence of internal motion, J Magn Reson 1990, 90, 111-123.

[20] Vögeli B, Segawa TF, Leitz D, Sobol A, Choutko A, Trzesniak D, van Gunsteren W, Riek R, Exact Distances and Internal Dynamics of Perdeuterated Ubiquitin from NOE Buildups, J Am Chem Soc 2009, 131, 17215-17225.

[21] Clore GM, Schwieters CD, Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: Correspondence of dipolar coupling and heteronuclear relaxation measurements, Biochemistry 2004, 43, 10678-10691.
[22] Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M, Simultaneous determination of protein structure and dynamics, Nature 2005, 433, 128-132.
[23] Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL, Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics, Science 2008, 320, 1471-1475.
[24] Fenwick RB, Esteban-Martin S, Richter B, Lee D, Walter KFA, Milanovic D, Becker S, Lakomek NA, Griesinger C, Salvatella X, Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition, J Am Chem Soc 2011, 133, 10336-10339.
[25] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19, 1053-1057.
[26] Vögeli B, Güntert P, Riek R, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol Phys 2012, 111, 1-18.
[27] Vögeli B, Friedmann M, Leitz D, Sobol A, Riek R, Quantitative determination of NOE rates in perdeuterated and protonated proteins: Practical and theoretical aspects, J Magn Reson 2010, 204, 290-302.
[28] Orts J, Vögeli B. Riek R, Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs, J Chem Theory Comput 2012, 8, 3483-3492.
[29] Orts J, Vögeli B, Riek R, Güntert P, Stereospecific assignments in proteins using exact NOEs, J Biomol NMR 2013, 57, 211 -218.
[30] Olsson S, Strotz D, Vögeli B, Riek R, Cavalli A, The dynamic basis for signal propagation in human Pin1-WW, Structure 2016, pii: S0969-2126(16)30159-9.
[31] Solomon I, Relaxation Processes in a System of Two Spins, Phys Rev 1955, 99, 559-565.
[32] Boelens R, Koning TMG, Vandermarel GA, Vanboom JH, Kaptein R, Iterative Procedure for Structure Determination from Proton-Proton NOEs Using a Full Relaxtion Matrix Approach. Application to a DNA Octamer, J Magn Reson 1989, 82, 290-308.
[33] Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M, Simultaneous determination of protein structure and dynamics, Nature 2005, 433, 128–132;
[34] Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Combining experiments and simulations using the maximum entropy principle, PLoS Comput Biol 2014, 10, e1003406.
[35] Pitera JW, Chodera JD, On the use of experimental observations to bias simulated ensembles, J Chem Theory Comput 2012, 8, 3445.
[36] Roux B, Weare J, On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method, J Chem Phys 2013, 138, 084107.
[37] Cavalli A, Camilloni C, Vendruscolo M, Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle, J Chem Phys 2013, 138, 094112.
[38] Olsson S, Frellsen J, Boomsma W, Mardia KV, Hamelryck T, Inference of structure ensembles from sparse, averaged data, PLoS One 2013, 8, e79439.
[39] Olsson S, Vögeli B, Cavalli A, Boomsma W, Ferkinghoff-Borg J, Lindorff-

Larsen K, Hamelryck T, Probabilistic determination of native state ensembles of proteins, J Chem Theory Comput 2014, 10, 3484–3491.
[40] Olsson S, Ekonomiuk D, Sgrignani J, Cavalli A, Molecular dynamics of biomolecules through direct analysis of dipolar couplings, J Am Chem Soc 2015, 137, 6270–6278.
[41] Hummer G, Köfinger J, Bayesian ensemble refinement by replica simulations and reweighting, J Chem Phys 2015, 143, 243150.
[42] Keepers JW, James TL, A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra, J Magn Reson 1984, 57, 404-426.
[43] Kalk A, Berendsen HJC, Proton magnetic relaxation and spin diffusion in proteins, J Magn Reson 1976, 24, 343-366.
[44] Brüschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR, Influence of Rapid Intramolecular Motion on NMR Cross-Relaxation Rates. A Molecular Dynamics Study of Antamanide in Solution, J Am Chem Soc 1992, 114, 2289-2302.
[45] Lipari G, Szabo A, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity, J Am Chem Soc 1982, 104, 4546-4559.
[46] Leitz D, Vögeli B, Greenwald J, Riek R, Temperature Dependence of 1HN-1HN Distances in Ubiquitin As Studied by Exact Measurements of NOEs, J Phy. Chem B 2011, 115, 7648-7660.
[47] Ulmer TS, Ramirez BE, Delaglio F, Bax A, Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy, J Am Chem Soc 2003, 125, 9197-9191.
[48] Yao L, Vögeli B, Torchia DA, Bax A, Simultaneous NMR Study of Protein Structure and Dynamics Using Conservative Mutagenesis, J Phys Chem B 2008, 112, 6045-6056.
[49] Brüschweiler R, Blackledge M, Ernst RR, Multi-conformational peptide dynamics derived from NMR data: a new search algorithm and its application to antamanide, J Biomol NMR 1991, 1, 3-11.
[50] Clore GM, Schwieters CD, How much backbone motion in ubiquitin is required to acount for dipolar coupling data measured in mutliple alignment media as assessed by independent cross-validation?, J Am Chem Soc 2004, 126, 2923-2938.
[51] Vögeli B, Orts J, Strotz D, Güntert P, Riek R, Discrete Three-dimesional Representation of Macromolecular Motion from eNOE-based Ensemble Calculation, Chimia 2012, 66, 787-790.
[52] Reif B, Hennig M, Griesinger C, Direct measurement of angles between bond vectors in high-resolution NMR, Science 1997, 276, 1230– 1233.
[53] Mey ASJS, Wu H, Noé F, xTRAM: estimating equilibrium expectations from time-correlated simulation data at multiple thermodynamic states, Phys Rev 2014, X 4, 041018.
[54] Wu H, Mey ASJS, Rosta E, Noé F, Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states, J Chem Phys 2014, 141, 214106.
[55] Wu H, Paul F, Wehmeyer C, Noé F, Multiensemble Markov models of molecular thermodynamics and kinetics, Proc Natl Acad Sci USA 2016, 113, E3221–E3230.
[56] Trendelkamp-Schroer B, Wu H, Paul F, Noé F, Estimation and uncertainty of reversible Markov models, J Chem Phys 2015, 143, 174101.

[57] Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F, Markov models of molecular kinetics: generation and validation, J Chem Phys 2011, 134, 174105.

Part I.

Methods

2. The experimental accuracy of the uni-directional exact NOE

This chapter is based on the following publication:

Dean Strotz, Julien Orts, Martina Minges and Beat Vögeli, The experimental accuracy of the uni-directional exact NOE, J Magn Reson 2015, doi: 10.1016/j.jmr.2015.07.007. Beat Vögeli advised on the underlying work and on the writing of the manuscript. Julien Orts advised on the creation of figures and analysis. Martina Minges did the analysis of the pulses. Dean Strotz recorded all experimental data (except as mentioned above), did the analysis, prepared the figures and wrote the initial manuscript.

2.1. Introduction

Conversion of the nuclear Overhauser enhancement (or effect, NOE) into approximate upper distance limits is the most valuable tool in biomolecular structure determination in the liquid state [1, 2]. Building on previous work [3-10] protocols to calculate exact NOEs (eNOE) from NOE data were established [11-13]. eNOEs lend unprecedented precision to the calculation of distance restraints used for structure calculation [14, 15]. Moreover, when the eNOE's nature of an averaged quantity [16-22] is accounted for in structure determination, it also provides information of the sampled conformational space [11-13, 23, 24].

The exact determination of an eNOE requires the analysis of the buildups of both cross peaks [14]. In practice, many NOEs can only be determined from one cross peak due to asymmetry in the resolution of the ^{1}H-resolved dimensions or in the peak intensities [14, 23]. As a consequence, the distance derived from a single cross-peak buildup has an additional error, which turns into an elevated upper limit and a reduced lower limit restraint in the structure determination [11, 23]. It is not a trivial task to estimate the order of the error. In previous studies, the limits of the distances have been set to 15 or 20 % higher and lower values. These limits were approximate estimates from comparisons with high-resolution structures [23]. Here, we rationalize the choice of these limits by an in-depth analysis of the asymmetry of the magnetization pathways leading to the two cross peaks.

NOE buildups that could not be normalized to the intensity of the diagonal peak of the spin of the origin magnetization have in the past been discarded in the eNOE

protocol. Alternatively, we proposed to derive a more conservative upper distance restraint limit ('generic normalized eNOE', presented in chapter 3) [25]. To define tighter distance limits for a subset of these distance restraints, we compare uni-directional eNOEs normalized to the origin of magnetization to such normalized to the destination of magnetization.

2.1.1. Dissecting multi-dimensional NOESY pulse sequences

In this work, the fates of the magnetization pathways leading to the cross and diagonal peaks, that ultimately determine the accuracy of the extracted cross-relaxation rate, are analyzed. Typical 2D and 3D NOESY pulse sequences and the spectral peaks are depicted in Figure 1.

Two-dimensional (2D) [^{1}H, ^{1}H]-NOESY pulse sequences [2, 9, 10, 27, 28] are the fundamental building blocks of the 3D (or 4D) NOESY experiments typically recorded for macromolecules [29-33]. Such 2D NOESY experiments allow the analysis of NOESY effects isolated from other influences, not possible in higher multi-dimensional experiments. More specifically, one of the appeals of 2D NOESY experiments lies in their theoretical spectral symmetry [34]. In practice, however, even the symmetry of 2D NOESY is broken. This is illustrated with correlation plots of cross-relaxation rates obtained from cross peaks above the diagonal versus those below the diagonal (as demonstrated for the WW domain in Figure 2).

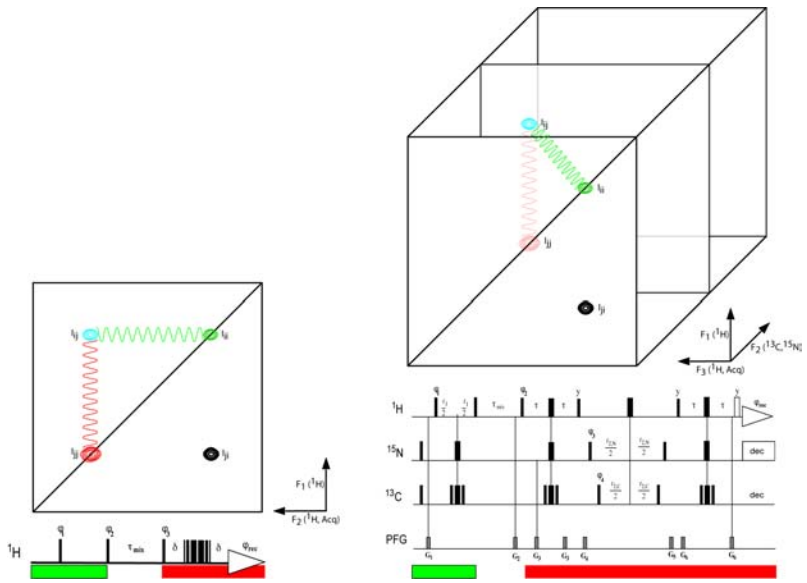


Figure 1: Illustration of the NOESY pulse sequence segments and their relation to the recorded spectra. A 2D [^{1}H, ^{1}H]-NOESY sequence with water suppression [26] is shown in the left panel and a 3D simultaneous [^{15}N, ^{13}C]-resolved [^{1}H, ^{1}H]-NOESY-HMQC sequence in the right panel, respectively. Segments of the pulse sequences, during which the pathways giving rise to a specific cross and diagonal-peak pair are shared, are related by color code, illustrating the two possible ways to normalize cross-peak intensities thereby accounting for different experimental artifacts (see 2.2. Theory).

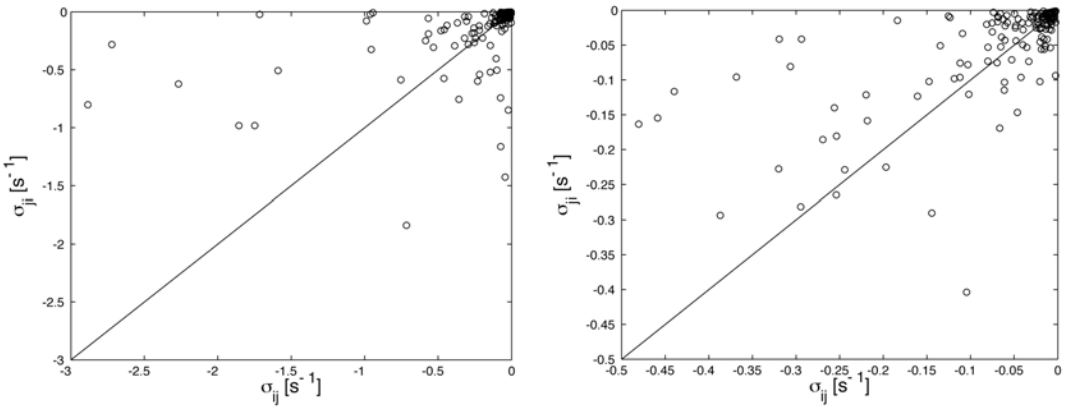


Figure 2: Broken symmetry of apparent cross-relaxation rates in 2D [^{1}H, ^{1}H]-NOESY. A full comparison of cross-relaxation rates σ obtained from above and below the diagonal of a NOESY of a WW domain (4.0 kDa) is shown in the left panel, and an extension of cross-relaxation rates smaller than 1.5 s^{-1} in absolute value in the right panel. Calculated σ values are corrected for spin-diffusion and ^{2}H exchange.

The effect can be understood as different attenuation of the individual magnetization pathways throughout the entire pulse sequence. The cumulative attenuation without the relaxation processes during NOESY mixing can be visualized by the analysis of the dispersion of back-predicted diagonal-peak intensities at zero mixing time (Figure 3a). The intensity dispersion over the entire domain is very broad (s = 0.61, see definition below). Including only H^{α} and H^{N} values in the β-sheet backbone, the back-predicted diagonal-peak intensities have a dispersion of 0.55 (Figure 3b). For H^{N} spins, the dispersion is reduced to 0.40, since there is only one relevant J coupling, $^3J_{HN,H\alpha}$, that impacts the peak intensity during the evolution periods and the H^{N} frequencies are also far removed from the suppressed water frequency.

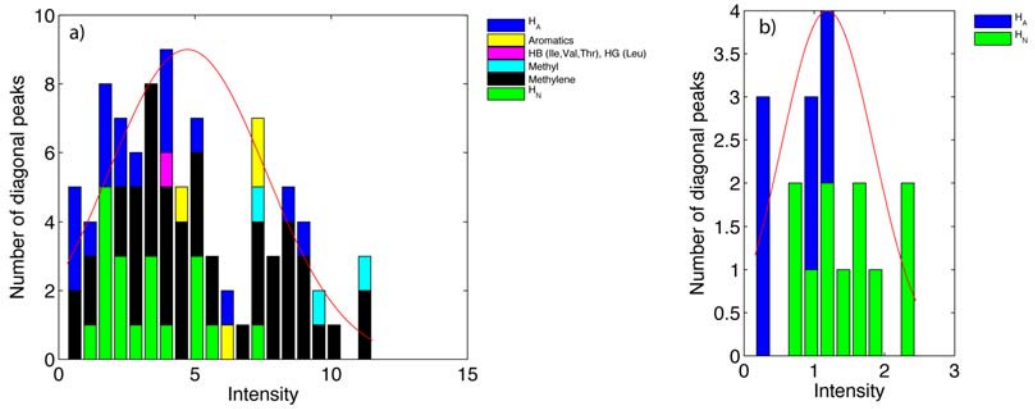


Figure 3: Stacked bar plot of back-predicted diagonal-peak intensities at zero mixing time $I^{diag}(0)$ for the WW domain. The intensity values from pseudo atoms are scaled by proton multiplicities. The fitted normal distribution is shown in red. In the left panel, all available $I^{diag}(0)$ values are shown. In the right panel, β -sheet $I^{diag}(0)$ of H^{N} and H^{α}, for which the overall dynamics and the auto-relaxation are comparable, are selected. The intensity values have an arbitrary scale as produced by Bruker Biospin hardware, for simplicity they are downscaled by a factor of 10^8.

In the 3D ^{15}N- or ^{13}C-resolved [^{1}H, ^{1}H]-NOESY-HXQC (where X stands for M or S), the different excitation profiles of ^{15}N and ^{13}C pulses and the different relaxation of ^{15}N and ^{13}C nuclei will enhance the non-symmetry of the experimental pathways (Figure 1).

In order to analyze the attenuation of the detected signal intensities, we introduce scaling factors α and magnetization transfer factors T of the individual pulse sequence segments (yielding equation 1). $M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1)$ represents the magnetization of spin i prior to the first pulse of a scan and is a function of the interscan delay d_{int} and the longitudinal auto-relaxation rate constant R_1. The scaling factor associated with the xth evolution period (indirect or direct) ending up in the NOESY plane and the subsequent Fourier transforms, $\alpha_{x,i}^{\mathrm{EP}}\left(proc, J, R_2\right)$, is determined by the user-set processing parameters *proc*, J-coupling induced FID modulation and the transverse relaxation rate constant R_2. $T_{ij}^{\mathrm{NOESY}}(\tau_{\mathrm{mix}})$ represents the transfer factor during NOESY mixing t_{mix}. If a diagonal-peak is considered, i is equal to j, while i is different from j for cross-peaks. T_i^{HXQC} is the transfer factor (attenuation) of the magnetization of spin i during HXQC including processing effects. α_i^{WS} quantifies signal attenuation during water suppression. The attenuated intensity of a signal originating from spin i and being detected on spin j can be expressed as a product of attenuation factors (pulse sequence elements are shown in Figure 1) and the starting magnetization $M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1)$:

$$I_{ij} = M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1) \times \alpha_{1,i}^{\mathrm{EP}}\left(proc, J, R_2\right) \times T_{ij}^{\mathrm{NOESY}}(\tau) \times T_j^{\mathrm{HXQC}} \times \alpha_j^{\mathrm{WS}} \times \alpha_{2,j}^{\mathrm{EP}}\left(proc, J, R_2\right) \quad (1)$$

For 2D [^{1}H, ^{1}H]-NOESY, the transfer factor $T^{\mathrm{HXQC}} = 1$, which simplifies the analysis. Depending on the design, the element order of other NOESY pulse sequences may be rearranged, or some elements may be merged.

The evaluation protocol of eNOEs involves normalization of cross-peak intensities to diagonal-peak intensities (see 2.2. Theory). The scope of this work is to estimate the uncertainties related to each factor in equation 1 and propagate them through the procedure for extraction of cross-relaxation rates from I_{ij} and I_{ii} or I_{jj}. The WW domain with a tumbling time of ca. 4 ns serves as our model protein. The resulting overall errors for the cross-relaxation rates define the corresponding upper and lower distance limits.

2.2. Theory

The Solomon equations describe the development of the peak intensities over the NOESY mixing time to very good approximation for a pair of spins-½ [35]. Generalization to a system consisting of N spins leads to the following differential equation for the N x N intensity matrix $I(t)$ [11, 36]:

$$\frac{d}{dt} I(t) = -R\, I(t) \tag{2}$$

The solution is:

$$I(t) = \mathrm{e}^{-Rt}\, I(0) \tag{3}$$

with the relaxation matrix R [10, 36]:

$$R = \begin{pmatrix} \rho_1 & \sigma_{12} & \sigma_{13} & \cdot & \cdot \\ \sigma_{21} & \rho_2 & \sigma_{23} & \cdot & \cdot \\ \sigma_{31} & \sigma_{32} & \rho_3 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \tag{4}$$

The diagonal matrix elements R_{ii} are the auto-relaxation rates ρ_i of the spins. The off-diagonal elements, the cross-relaxation rate constants σ_{ij} between spins i and j, depend on the internuclear distance r_{ij} and they are given by:

$$R_{ij} = \sigma_{ij} = \frac{b^2}{r_{ij}^{\,6}} \left(6J(2\omega) - J(0) \right) \tag{5}$$

where the spectral density function is

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1 + \left(\omega \tau_c \right)^2} \right) \tag{6}$$

and

$$b = \frac{1}{2}\frac{\mu_0}{4\pi}\hbar\gamma_{\mathrm{H}}^2 \tag{7}$$

with μ_0 the permeability of vacuum, $\hbar$ the reduced Planck constant and γ_{H} the gyromagnetic ratio of the nucleus.

In practice, mono-exponential curves are used to fit the diagonal-peak intensities of NOESY series in order to obtain the magnetization at the onset of NOESY mixing $I_{\mathrm{i}}^{\mathrm{diag}}(0)$ and the auto-relaxation rate ρ_i of spin i (and analogously for spin j) [15]:

$$I_i^{\mathrm{diag}}(\tau_{\mathrm{mix}}) = e^{-\rho_i\tau_{\mathrm{mix}}} \cdot I_i^{\mathrm{diag}}(0) \tag{8}$$

$I_{\mathrm{i}}^{\mathrm{diag}}(0)$ is back-predicted from:

$$I_i^{\mathrm{diag}}(0) = I_i^{\mathrm{diag}}(\tau_{\mathrm{mix}}) \cdot e^{\rho_i\tau_{\mathrm{mix}}} \tag{9}$$

The normalized cross-peak intensity for magnetization transferred from spin i to spin j, and normalizing to the origin of the magnetization, spin i, is

$$T_{ij}^{\mathrm{NOE}}(\tau_{\mathrm{mix}}) = \frac{I_{ij}^{\mathrm{NOE}}(\tau_{\mathrm{mix}})}{I_i^{\mathrm{diag}}(0)} \tag{10.1}$$

or when normalizing to the destination of magnetization, spin j,

$$T_{ij}^{\mathrm{NOE}}(\tau_{\mathrm{mix}}) = \frac{I_{ij}^{\mathrm{NOE}}(\tau_{\mathrm{mix}})}{I_j^{\mathrm{diag}}(0)} \tag{10.2}$$

.

See Figure 1 for the magnetization pathways in the respective dimensions. The cross-relaxation rate constant σ_{ij} between spins i and j is then calculated by the two-spin approximation, fitting:

$$T_{ij}^{\mathrm{NOE}}(\tau_{\mathrm{mix}}) = -\frac{\sigma_{ij}}{\left(\lambda_+ - \lambda_-\right)}\left[e^{-\lambda_-\tau_{\mathrm{mix}}} - e^{-\lambda_+\tau_{\mathrm{mix}}}\right] \tag{11}$$

with

$$\lambda_{\pm} = \frac{\left(\rho_i + \rho_j\right)}{2} \pm \sqrt{\left(\frac{\rho_i - \rho_j}{2}\right)^2 + \sigma_{ij}^2} \tag{12}$$

If the spectral symmetry is broken, then in order to obtain the eNOE, both cross peaks need to be observed since equation 10 is not strictly valid. With the same logic as applied before, the cross-relaxation rate must be calculated using a normalization procedure involving both spins *i* and *j*, as follows easily from equation 1:

$$T_{ij}^{\text{NOE-bi}}(\tau_{\text{mix}}) = T_{ji}^{\text{NOE-bi}}(\tau_{\text{mix}}) = \sqrt{\frac{I_{ij}^{NOE}(\tau_{\text{mix}}) I_{ji}^{NOE}(\tau_{\text{mix}})}{I_i^{\text{diag}}(0) I_j^{\text{diag}}(0)}} \tag{13}$$

For simplicity, if both diagonal peaks are resolved the true cross-relaxation rate is approximated as [14]

$$\sigma_{bi}^{fit} = \sqrt{\sigma_{ij}^{\text{fit}} \sigma_{ji}^{\text{fit}}} \tag{14}$$

from the fitted cross-relaxation rates σ_{ij}^{fit} and σ_{ji}^{fit}.

For an *N*-spin system, the two-spin approach is a coarse approximation. We subsequently correct either the intensities or the obtained cross-relaxation rate for spin diffusion. For each pair, we estimate a correction factor p_{ij} using a previously known structure or iteratively [37]. The corrected cross-relaxation rate therefore is

$$\sigma_{ij}^{\text{corrected}} = p_{ij}\, \sigma_{ij} \tag{15}$$

In the following, we analyze the case when equation 10 is used for the extraction of distance limits rather than equation 13.

2.3. The model system

The WW domain is the N-terminal domain of the human peptidyl-prolyl cis-trans isomerase (PPIase) Pin 1 [38]. The WW domain, which is separated by a flexible linker from a C-terminal catalytic domain, is responsible for the specificity of binding [39]. The 34-residue mutant S18N-W34F of the WW domain is less prone to aggregation than the native WW domain. It folds into the three-stranded anti-parallel β -sheet fold characteristic of WW domains. The 34-residue mutant has a MW of 4011.4 g/mol. Using T_1 and $T_{1\rho}$ ^{13}C-relaxation measurements of C^{α} atoms in the β -sheet at natural abundance with a sample concentration of 1.2 mM, the rotational tumbling time τ_c was determined from T_1/T_2 ratios [40] to be 3.6 ns at 5˚C. At 5˚C all NOEs are in the negative regime ($\omega\tau_c > 1.12$) and peak dispersion is good (see Figure S1 in appendix 10.1.). No baseline rolls were present in any spectrum. From the 2D NOESY, 89 well-resolved diagonal-peaks (out of 269 protons) are available. 141 bi-directional eNOEs (see equation 14) can be evaluated. Normalizing to the origin (I_{ij}/I_{ii}), an additional 515 uni-directional eNOEs can be evaluated. Normalizing to the destination (I_{ij}/I_{jj}), an additional 602 uni-directional eNOEs can be evaluated. From the 3D NOESY, 152 well-resolved diagonal-peaks (out of 269 protons) are available. 282 bi-directional eNOEs can be evaluated. Normalizing to the origin (I_{ij}/I_{ii}), an additional 413 uni-directional eNOEs can be evaluated. Normalizing to the destination (I_{ij}/I_{jj}), an additional 381 uni-directional eNOEs can be evaluated.

2.4. Results

In the following, we analyze and quantify the attenuation of the peak signal during each element of the pulse sequence as defined in equation 1. Rather than a detailed analysis of individual magnetization transfer pathways, we aim at obtaining overall statistics that provide guidelines for reasonable upper and lower distance limits in structure calculation. Because the cross-relaxation rates are obtained from normalized intensities, the dispersions of the average attenuation (of I_{ij}, equation 1) cause errors rather than the attenuation *per se*. We quantify the dispersion of the attenuation of n pathways during a segment of a pulse sequence by the relative standard deviation s from the average value $\bar{x}$:

$$s = \frac{std}{\overline{x}} \quad , \tag{16}$$

where the standard deviation is

$$std = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^{\frac{1}{2}} \tag{17}$$

2.4.1. Equilibration during interscan delay: $M_i^{SS}(d_{\text{int}}, R_1)$

At the short interscan delays usually used in NOESY experiments (we typically use 0.8 s) equilibrium magnetization is not reestablished. As such, by reducing the interscan delay, the contribution to the signal intensity from a single scan of the pulse sequence is lower, but cycling the experiment many times compensates the signal reduction. Therefore, some dispersion of the magnetizations of all spins is present prior to the first pulse of each scan. In order to estimate the effect of a short interscan delay, we compare back-predicted diagonal magnetization at time zero, $I^{\text{diag}}(0)$ from 2D [^{1}H, ^{1}H]-NOESY experiments to those obtained from experiments with very long interscan delays (Figure 4).

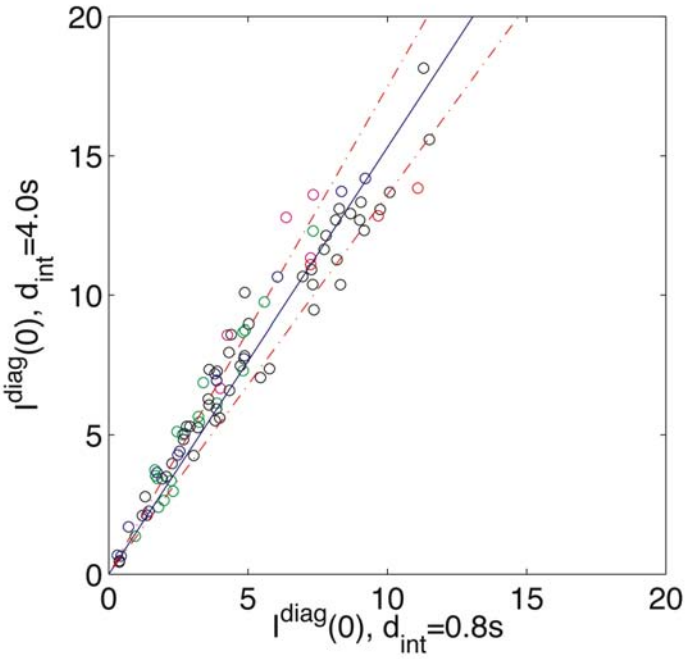


Figure 4: Correlation plot of back-predicted $I^{\text{diag}}(0)$ from 2D [^{1}H, ^{1}H]-NOESY series with interscan delays d_{int} of 0.8 s and 4.0 s. The error boundaries are ±12 % (red lines), which are obtained from the root-mean-square deviation of the normalized values. Amide, alpha, methyl, methylene and side-chain methine intensities are shown in green, blue, red, black and magenta, respectively. The intensity values have an arbitrary scale as produced by Bruker Biospin hardware, for simplicity they are downscaled by a factor of 10^8.

To reach a near-complete reestablishment of the equilibrium polarization, we chose an interscan delay of 4.0 s, where ca. 99% of the equilibrium polarization is reached (see 2.8.3. Methods) (Figure 4). The dispersion of 0.61 for the short interscan delay is reduced to 0.54 upon extension of the interscan delay. The remaining dispersion is caused by the other factors in equation 1. Taking the mean of the ratios of the back-predicted $I^{diag}(0)$ values from experiments recorded with interscan delays (d_{int}) of 0.8 s and 4.0 s shows that 38% of the steady state magnetization is lost due to the 5 times shorter d_{int}. An upper limit for the standard deviation of $M_i^{SS}(d_{int},R_1)$ introduced by short interscan delays is the standard deviation from the average rescaling due to the difference in d_{int}. The error boundary for $M_i^{SS}(d_{int},R_1)$ is ±12 %.

2.4.2. Evolution periods: $\alpha_{x,i}^{EP}(proc,J,R_2)$

A. Line width

Measured line widths are a combination of physical (R_2 rates) and non-physical parameters such as the window function or acquisition time (the resolution). In a first attempt, we focus on a simple analysis of the line widths.

We evaluate the peak height rather than the volume to obtain the signal intensity. This procedure is not strictly correct, but is more convenient in practice. In order to identify the error introduced by that method we approximate volumes by multiplication of the heights by the line width (note that this relationship is exact for Lorentzian line shapes). We estimated line widths by manually measuring at half maximum (full width at half maximum, FWHM) in both dimensions of the 2D [^{1}H,^{1}H]-NOESY spectra. Diagonal intensities scaled with the line width in comparison to the raw intensities provide clues about the effects of the non-uniformity of R_2 rates on the measured intensities (Figure 5).

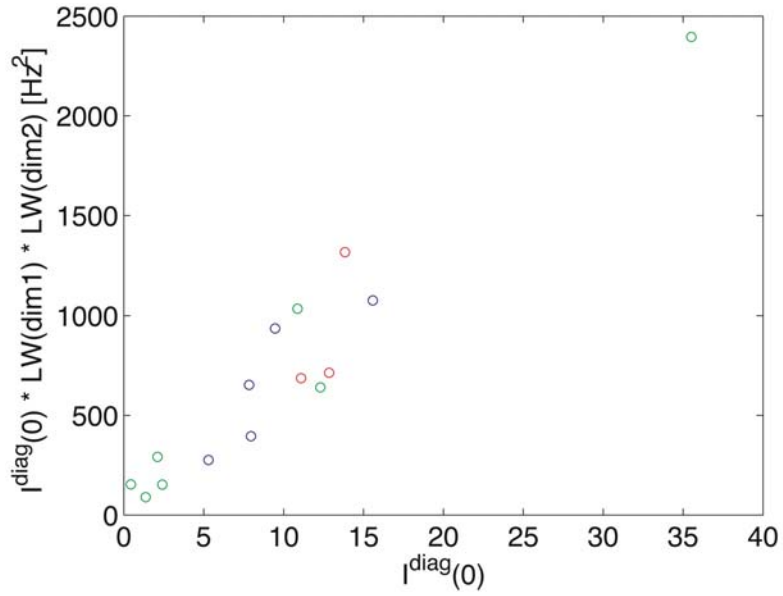


Figure 5: Relationship of peak volumes and intensities. Plotted are intensities multiplied by the line widths versus intensities from well-separated diagonal-peaks. Amide, methylene and methyl intensities are shown in green, blue and red, respectively. The intensity values have an arbitrary scale as produced by Bruker Biospin hardware, for simplicity they are downscaled by a factor of 10^8.

The analysis of an arbitrary collection of well-separated peaks shows no narrowing of the dispersion when scaled with line width (*s* is 0.70 in both cases). Line width alone does therefore not suffice to narrow the dispersion. Apparently, assuming that the product of the line width and the intensity reproduces the volume is too simple. Plausible explanations are that the shape is not Lorentzian, or that *J* couplings, sampling truncation and the window function modify the line width at least to a similar extent as transverse relaxation does. Similar results were obtained when using the volume instead of the peak height (data not shown). Because the values of the R_2 and *J* couplings are not known, and due to the apparently complex interplay between transverse relaxation, *J* coupling evolution and FID modulation, we chose to simulate the effects in order to analyze each contribution in more detail.

B. Simulations of impact of relaxation, *J* coupling evolution, resolution and window function

Using simulations of free induction decay (FID) modulations and Fourier transformations, we illustrate the dependency of the spectral peak heights on the digital resolution (set by the number of points and time increments), the transverse relaxation rate constant R_2 and the number and amplitudes of the *J* couplings in a single spectral dimension (Figure 6). In order to simulate the experimental scenario as realistically as possible, we simulated the free induction decay of a single spin in the time domain

modulated by scalar coupling to one or two other spins, assuming a zero Larmor frequency in a in-house written MatLab routine. The scalar couplings were chosen in the applicable range of $^3J_{HN,H\alpha}$ and $^3J_{H\alpha,H\beta}$ couplings. Furthermore, we use the same time increment of 80 μs as in the experimental setup to produce an array of complex data points. The number of real points was set to 512, 1024 and 2048 (acquisition times of 41, 82 and 164 ms, corresponding to resolutions of 48.83, 24.41 and 12.21 Hz, respectively). The resulting raw FID was then treated the same as the experimental data, that is, the window function applied and zero filled. The window function used is an adjustable sine bell window function, with an offset of 0.5 (cosine), endpoint of 0.98 and exponent 2 (squared bell function). Then, the maximum value of the Fourier transformed data is determined.

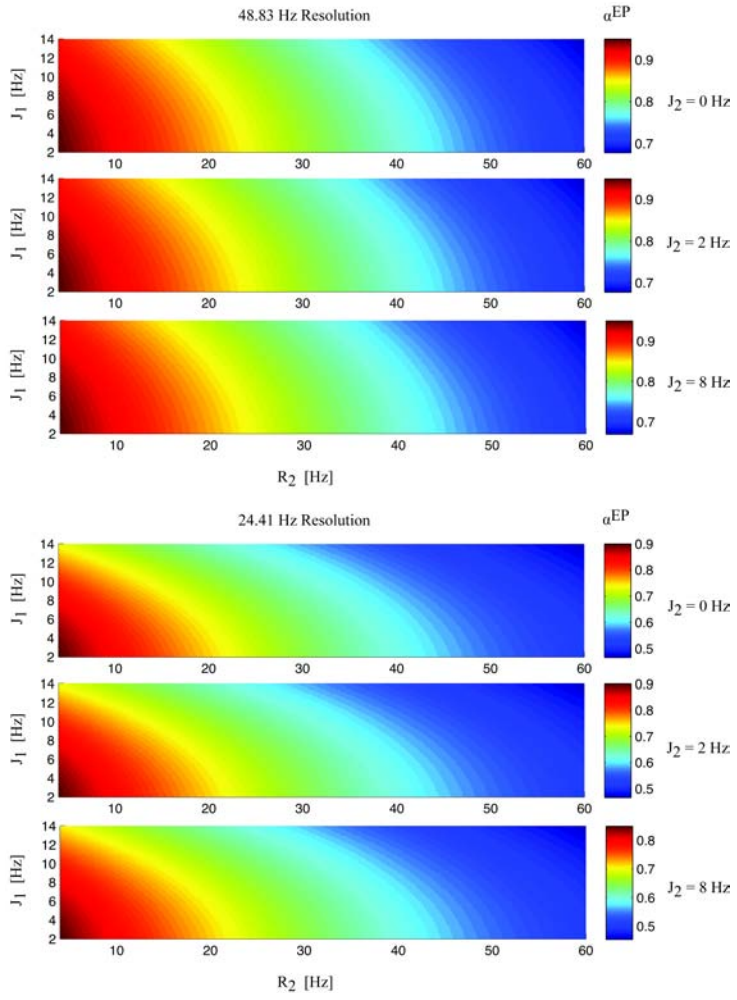
Overall, the simulated attenuation factor ranges between 0.90 and 0.26. These two extremes are obtained for the smallest and largest values of the three parameters (R_2, J coupling and number of points). This is a general trend as the results show that increasing R_2, J coupling and resolution (number of points) enhance the attenuation of the peak intensities. Due to symmetry, the introduction of a second J coupling has a similar effect as the one caused by the first J coupling. When one of any of the parameters is large, the dependency on the others is more pronounced such that a mutual enhancement is obtained. Within the given ranges, the dependence on the J coupling is the weakest.

The typical dispersion of the attenuation factor in a real experiment is much narrower than the separation of the simulated extremes. In the following, we attempt to estimate the dispersion range of the scaling factor expected for our experimental 2D NOESY setup for the WW domain. The number of points prior to zero filling (TD) was 1024 in both dimensions. Because it is difficult to read out R_2 values from line width analysis, we simulated them assuming a rigid model (see Supporting Material in appendix 10.1.). The simulated rates are rather approximate and are 34 s^{-1} on average with a standard deviation of 20 s^{-1} (range from ca. 15 to 50 s^{-1}). The calculated values are likely an overestimate of the true R_2 values due to motion. To correct for this effect, R_2 of protons located in backbone segments with secondary structure may be scaled by an order parameter of 0.8 and in loops and side chains by a smaller factor. Thus, we assume a range of 10 - 40 s^{-1} to estimate the scaling factor of peak intensities. Using the relatively narrow range of the $^3J_{HN,H\alpha}$ couplings measured for the WW domain (see Table S1 in appendix 10.1.) results in virtually identical statistics as when using the full possible range of proton-proton J couplings. At the experimental resolution of 24.41 Hz,

the scaling factors α^{EP} is ca. 0.70 on average with a standard deviation of 0.13. This results in a dispersion of 18%.

Since the simple correction using the line width as proposed in A) does not improve the data, and the necessary precision of R_2 and J values required for an exact simulation is not readily available, correction is not applicable to individual peak intensities.

As mentioned previously, the dispersion of the scaling factor is smaller for lower resolution. On the one hand, this is an unfavorable trend if the resolution is optimized in NOESY spectra. On the other hand, it benefits three- and higher-dimensional experiments as the number of points in the indirect dimensions must be reduced in order to maintain reasonable experimental times. In our 3D NOESY experiment, we typically use 200 complex points. In that case, the scaling factors α^{EP} is ca. 0.82 on average with a standard deviation of 0.10 (see Figure S3 in appendix 10.1.). The dispersion is then reduced to 12%.



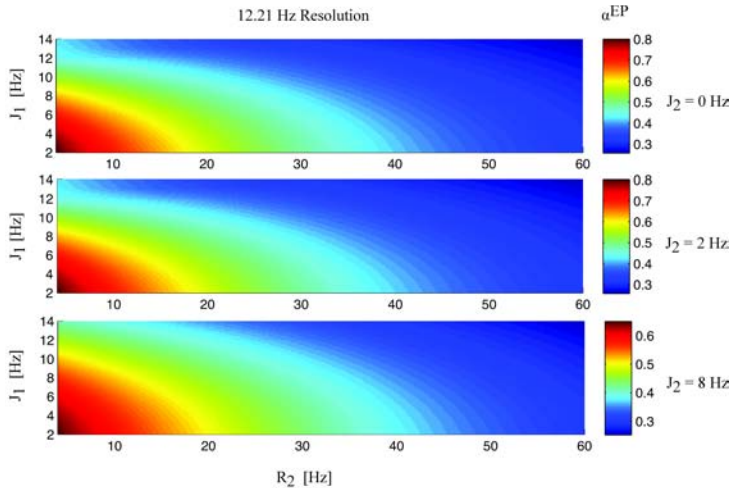


Figure 6: Simulation of the dependency of the scaling factor α^{EP} on the spectral resolution, the transverse relaxation rates constant R_2 and the amplitudes of the J couplings. Each contour plot shows α^{EP} depending on R_2 (x axis) and one J coupling (y axis). In all cases, the range of R_2 is 5 – 60 s^{-1}. In the three top, middle and bottom panels the number of real points are set at 512, 1024 and 2048, respectively, which were doubled by zero-filling (acquisition times of 41, 82 and 164 ms, corresponding to resolutions of 48.83, 24.41 and 12.21 Hz). In each top panel, the range of the only involved J coupling (J_1) is varied between 2 and 14 s^{-1}. In each middle panel, two J couplings are simulated, one of which (J_1) has a range of 2 - 14 s^{-1}, while the second one (J_2) is held constant at 2 s^{-1}. In each bottom panel, two J couplings are simulated, one of which (J_1) has a range of 2 - 14 s^{-1}, while the second one (J_2) is held constant at 8 s^{-1}.

2.4.3. NOESY mixing period: $T_{ij}^{NOESY}(\tau_{mix})$

Error sources affecting the scaling/transfer factor for the NOESY mixing period, $T_{ij}^{NOESY}(\tau_{mix})$, are the assumption of a mono-exponential decay of the diagonal peaks (equation 8), the possibly unknown value of the auto-relaxation rate constant of the second spin and shortcomings in the correction of the buildup intensities for spin diffusion (equation 15). Note that erroneous $I(0)$ and ρ_i are not related to the intrinsic symmetry break of the NOESY.

In the normalization procedure, auto-relaxation processes during the NOESY mixing are accounted for by calculation of the auto-relaxation rate constant ρ_i and back-prediction of the signal intensity to time zero. Auto-relaxation rates within particular types (such as H^N, H^α, methylene, methyl, etc.) of atoms are quite uniform for the WW domain (see Table S2 in appendix 10.1.).

Back-prediction of the intensity to time zero is done according to equation 9. The error introduced by this first order approximation is judged by:

$$\frac{I_i^{diag}(\tau_{mix})}{I_i^{diag}(0)} = 1 - \rho_i \tau_{mix} + \frac{1}{2}(\rho_i^2 + \sum_{k \neq i} \sigma_{ik}^2)\tau^2_{mix} + O(\tau^3_{mix}) \cong e^{-\rho_i \tau_{mix}} + \frac{1}{2}(\sum_{k \neq i} \sigma_{ik}^2)\tau_{mix}^2 \qquad (18)$$

Using the minimum ρ_i of 2 s^{-1} and one large σ_i^{fit} of 1 s^{-1}, which absorbs the contributions from all σ_{ik} (see Figure S4 and Table S3 in appendix 10.1.), the first and the second

terms on the right hand side after a mixing time of 80 ms are 0.852 and 0.003, respectively, which translates into a relative error of ca. 0.4 %. This results in an error of the fitted ρ_i of 3 %, and it is reduced for larger auto-relaxation rates. Obviously, the error in

$I^{diag}(0)$ introduced by the mono-exponential is vanishingly small. This error, in turn is propagated into the error of the fitted cross-relaxation rate. This effect is also very small and therefore not relevant for the generation of distance limits.

However, if only one diagonal peak can be evaluated one has to use an estimate for the auto-relaxation rate constant of the second spin as evident from equations 11 and 12. In extreme cases, the extracted cross-relaxation rate can be over- and underestimated by ca. 10 % and 20 %, respectively (see Table 1). Generally, the error is expected to be lower than 10 %. In our approach, we usually calculate average auto-relaxation rate constants from separated diagonal peaks for each atom type listed in Table 1 and use them for an estimation of the unknown values. The standard deviation within one specific group is typically 1 s^{-1} with a maximum of 2 s^{-1} for H^N. Therefore our true errors are significantly reduced and 10 % constitutes an upper limit. Note that this error is not the one of the extracted transfer function $T_{ij}^{NOESY}(\tau_{mix})$ defined in equation 11, but the error of the extracted σ_{ij} from $T_{ij}^{NOESY}(\tau_{mix})$.

Table 1: Error of σ_{ij} fitting due to error in ρ_i.

σ_{ij} (s^{-1})	σ_{ij} fitted (s^{-1})	ρ_i (s^{-1})	ρ_j (s^{-1})	error (%)	error (s^{-1})
1	1.000	4.21	4.21	0.0	0.000
0.01	0.010	4.21	4.21	0.0	0.000
1	1.077	4.21	1.93	7.2	0.077
0.01	0.011	4.21	1.93	7.0	0.001
1	0.825	4.21	10.5	-21.3	-0.176
0.01	0.008	4.21	10.5	-21.2	-0.002

Theoretical peak intensities were predicted using equations 11 and 12, with ρ_i and ρ_j as given in Table 1. σ_{ij} was fitted with $\rho_i = \rho_j = 4.21$ s^{-1}. The mixing times were 20, 40, 60 and 80 ms.

A potential error in the spin diffusion correction factor has been discussed in reference [37]. This error is not the scope of this work as it is equally applied to cross-relaxation rates obtained from both peaks and does not cause deviations between uni- and bi-directional upper distance limits. Therefore we use an error of 0 % for $T_{ij}^{NOESY}(\tau_{mix})$ in our final analysis, but 10 % for the extraction of σ_{ij} from $T_{ij}^{NOESY}(\tau_{mix})$ (see Table 2).

2.4.4. Effects of water suppression: α^{WS}

When dealing with proton-detected experiments water suppression and its effects on measured peak intensities have to be considered. Resonances close to the water frequency such as some H^{α} and H^{β} are heavily suppressed in intensity by water suppression schemes. In this study, we choose the w5 WATERGATE block as a model scheme because it is very convenient and efficient in practice [26]. Note that resonances at the edges of the spectrum such as methyl resonances may also be suppressed in poorly adjusted w5 water suppression schemes.

In order to investigate the w5 suppression profile, we performed an nmrsim (Bruker Biospin, NMR-SIM Experiment Simulator) simulation with a single spin system (Figure 7). Strong suppression (> 40 %) is achieved up to 1 ppm up- and downfield from the water frequency (see Supporting Material in appendix 10.1., for additional information).

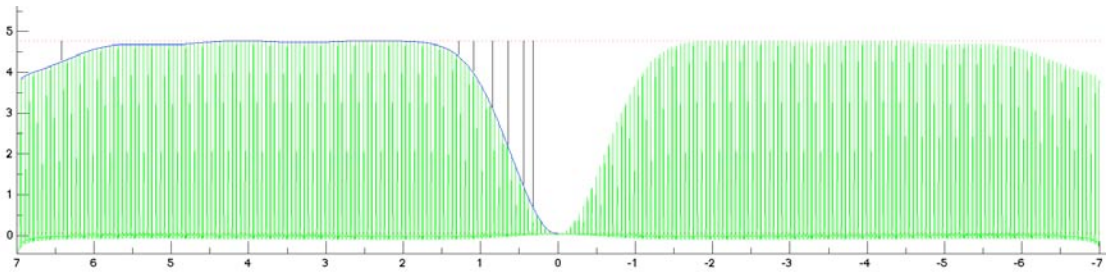


Figure 7: Signal attenuation profile of the w5 WATERGATE scheme. The simulated intensities are plotted versus the ^{1}H spectral frequency in ppm for a 900 MHz field. Black lines indicate (left to right) 10%, 10%, 20%, 40%, 60%, 80% and 90% signal suppression. Intensities were simulated in nmrsim (Bruker Biospin) and the free induction decays processed and fitted in Matlab. Signal intensities are scaled by a factor of 10^{8}.

In principle, it is possible to correct peak intensities close to the water frequency based on the simulated suppression profile. We used the profile to correct our experimental diagonal intensities. As we did not achieve a better dispersion of the intensities at zero mixing time, we conclude that this procedure is not feasible in practice. Not only are the resonances close to the water frequency heavily suppressed; the heavy suppression amplifies the relative influence of the noise on the accuracy of a correction.

Due to the continuous character of the error imposed on α^{WS}, it is not possible to quantify an upper limit. Rather, the user must decide on the upper limit (and therefore on the frequency range from which no peaks are analyzed) based on the attenuation profile of the water suppression element. As is the case for the errors of all scaling factors

caused by spectral symmetry breakdown, partial suppression does in principle not constitute a problem for the accuracy of bi-directional eNOEs since cross-peak intensities are normalized by their respective diagonal-peak intensities (provided there is sufficient S/N). The resonances at particular frequencies are subject to the same power of water suppression as the respective cross-peaks they give rise to and the effects therefore cancel out.

2.4.5. Extension to a third dimension: T^{HXQC}

With $^{13}C/^{15}N$ isotope enriched samples it is preferable from the perspective of resolution to acquire multidimensional experiments [29, 41]. Optimization of higher-dimensional NOESY experiments is still an active field [31, 42-45]. Furthermore, in multidimensional experiments more subtle water suppression schemes other than WATERGATE allow the detection of otherwise suppressed resonances. However, in the ^{15}N- and/or ^{13}C-resolved [^1H, ^1H]-NOESY experiments the excitation profiles of ^{15}N- and ^{13}C-pulses and the different relaxation rates of ^{15}N and ^{13}C nuclei will inevitably lead to additional non-symmetry of the experimental pathways.

To gain an understanding of T_i^{HXQC}, the transfer factor (attenuation) of the magnetization of spin i during HXQC, we isolated the HSQC element of our 3D experiment and performed it as a 2D experiment. The interscan delay d_{int} was set to 10 s in order to ensure complete recovery of all spins to equilibrium. Figure 8 shows the dispersion of intensities normalized to the strongest peak (Q^e of Met15). The methyl intensities (magenta) are on average about double the strength of the others. Aromatic intensities (yellow) are in general below the average of the rest of the intensities. To calculate the mean and standard deviation of the peak intensities for which we also calculated diagonal intensities in the NOESY, we have excluded Q^e of Met15 as an outlier, giving 0.37 ± 0.20. The intensities larger than the average plus standard deviation (0.57) are labeled by atom type. The majority are methyl groups, supplemented by other side-chain protons and some H^α. This is not unexpected, as it is well known that methyl groups typically undergo slower relaxation. In addition, the relaxation of side-chain spins is also reduced by the fact that they are also highly flexible. This effect impacts on the observed intensities in a twofold manner. On the one hand, slow transverse relaxation rates reduce intensity loss during the polarization transfer elements. On the other hand, the α^{EP} factors are attenuated for both the direct ^1H detection dimension as well as the indirect heteronuclear dimension. The amide intensities are reduced by 20 % due to the forward and reversed INEPT elements delays optimized for proton-carbon transfer (3.6

ms). In principle, optimal conversion of inphase to anti-phase coherence and vice versa can be obtained by INEPT elements with the inversion pulses on ^{13}C and ^{15}N shifted relative to each other. However, this choice comes at the cost of intensity loss due to additional relaxation, which may become severe for large systems.

To get an estimate of the T_i^{HSQC}, we have to subtract the impact of the direct ^1H dimension, α_2^{EP}, from the intensities extracted from the 2D experiment (Note that the influence of the heteronuclear dimension is included in T_i^{HSQC}). Thus, an overall standard deviation of 54 % of the observed intensities and a standard deviation of 18 % for α_2^{EP} taken from above result in a standard deviation of 51 % for T_i^{HSQC}.

Given this very large contribution to the overall intensity distribution through a 3D NOESY experiment, a significant correlation should be observed between the diagonal-peak intensities $I^{diag}(0)$, back-calculated from 3D NOESY experiments (with an interscan delay d_{int} of 0.8 s), and the intensities obtained from the HSQC spectrum. Indeed, Figure 9 shows that the correlation is high with a correlation coefficient r of 0.87 and a standard deviation of 36 % (excluding Q^{ε} of Met15, aromatic protons and H^a close to water).

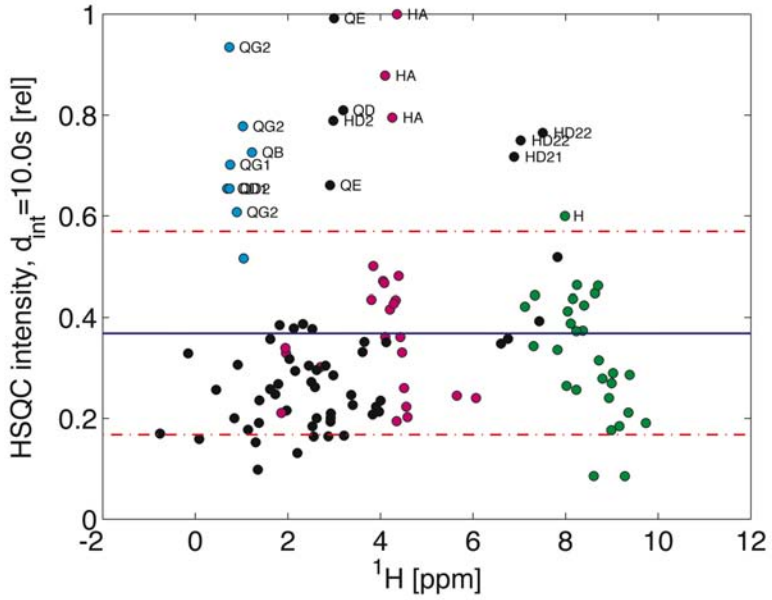


Figure 8. Intensity distribution of the HSQC element recorded in 2D mode. The dispersion is a combined effect of those of T_i^{HSQC} and α_2^{EP}. The intensities (normalized to the strongest value and by their spin multiplicity) are plotted on the y-axis versus the chemical shifts on the x-axis. The spectrum was recorded with an interscan delay d_{int} of 10 s. The mean and standard deviation of the intensities is 0.37 and 0.20, respectively (excluding Q^{ε} of Met15, aromatic protons and H^a close to the water frequency). The color code is H^N, green; methylene, black; methane, magenta; methyl, cyan.

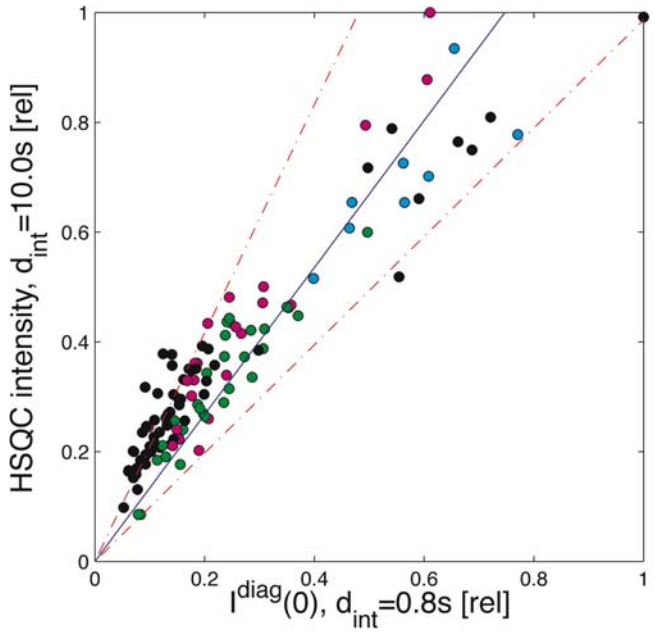


Figure 9. Comparison of the intensities of NOESY-HSQC and HSQC diagonal peaks. The NOESY diagonal-peaks $I^{diag}(0)$ (y axis) were back-predicted to zero mixing time for an interscan delay of 0.8 s, and normalized to the largest value. The HSQC intensities (x axis) were obtained with an interscan delay d_{int} of 10 s. The standard deviation is 36 % and Pearson's correlation coefficient r is 0.87 (excluding Q^e of Met15, aromatic protons and H^a close to the water frequency). The color code is H^N, green; methylene, black; methane, magenta; methyl, cyan.

In the following, we quantify the contribution of heteronuclear pulse imperfection to the distribution of T_i^{HXQC}. HXQC elements contain at least two 90° pulses on the heteronuclei, namely those flanking the heteronuclear evolution period. In the case of HSQC, the two INEPT elements contain also (at least) two 180° inversion pulses. In our previous studies, we employed rectangular pulses for the 90° rotations and composite 90°-180°-90° pulses for inversion. It is clear that part of the low T_i^{HXQC} of aromatic intensities is caused by the large chemical shift offset of their resonances. On the other hand, the excitation and inversion profile of the aliphatic spins is rather complete and shows that only a small fraction of the distribution of T_i^{HXQC} is caused by imperfection (at least at 700 MHz fields).

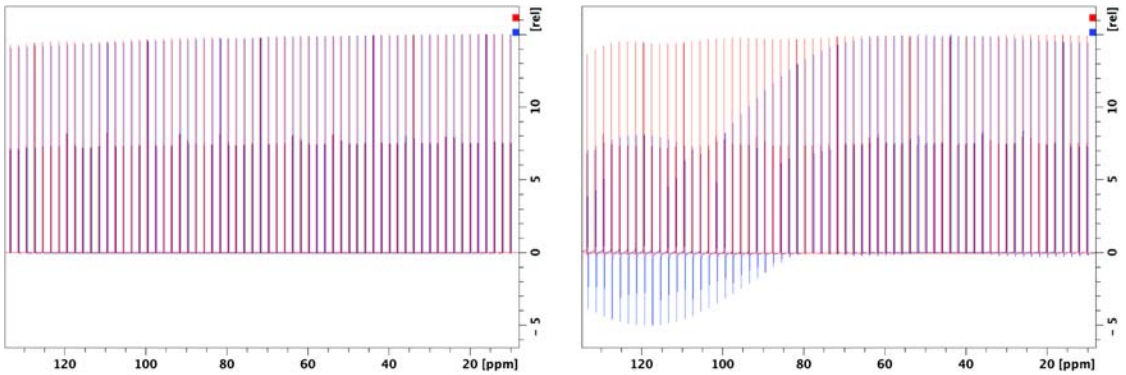


Figure 10. Comparison of simulated profiles of rectangular and shaped pulses. The profiles of the 90° excitation pulses are shown on the left. The rectangular pulse profile is shown in blue, and the one of BEBOP(+zy) in red, respectively. The BEBOP(+zy) pulse was executed with 10'000 Hz power, 550 ms length, a theoretical excitation profile of 37.5 kHz corresponding to 214.3 ppm @ 700MHz, and an offset of 0 Hz. It was shifted by 50 Hz relative to the rectangular pulse for visual comparison of signal intensities. The profiles of the 180° inversion pulses are shown on the right. The rectangular pulse profile is shown in blue, and the one of the BIBOP 180º inversion pulse in red, respectively. The BIBOP pulse was executed with 10'000 Hz power, 252.5 ms length, a theoretical excitation profile of 20 kHz corresponding to 114.3 ppm @ 700 MHz, and an offset of 4'375 Hz. The profile was shifted by 50 Hz relative to the rectangular pulse. Simulations were made in nmrsim, Bruker Biospin.

The question arises as to whether shaped pulses may provide more optimal profiles and thus minimize the dispersion of T^{HXQC}. We analyze the performance of conventional rectangular 90° and 180° pulses and compare them to broadband shaped pulses using both simulated profiles and experimental data. We selected the computationally optimized broadband 90° shaped excitation pulses BEBOP(+zy), BEBOP(-zy) and the BIBOP 180° inversion pulse [46, 47] because they perform the desired operation at a 99 % level within the shortest possible time such that unwanted relaxation and evolution processes are minimized. Note that very clean inversion/excitation profile can be obtained from adiabatic pulses, which are, however typically in order of magnitude longer than those used here.

The expectation inspired by the simulated pulse profiles (Figure 10) correlates well with the analysis of experimental data (Figure 11). Here, the rectangular pulses are applied at 45 ppm, which was originally chosen to obtain good excitation and inversion over the entire aliphatic region. Replacing the rectangular 90° pulses of our HSQC element with shaped pulses (blue data points) results in an overall loss in signal intensity of 7 %, with up to 10 % in the aliphatic region (< 46 ppm, mean value 6 %) and up to 19 % in the aromatic region (>115 ppm, mean value 14 %). Replacing the 90°-180°-90° inversion pulse trains of our HSQC element with a shaped pulse (red data points) results in a signal gain in the aromatic region of up to 107 % but comes with a signal loss in the aliphatic region of up to 39 %. The average gain in the aromatic region is 47 %, while the average loss in aliphatic region is 15 %. With both, the rectangular 90° and 90°-180°-

90° inversion pulses of our HSQC element replaced by shaped pulses (black data points) the net effects are less pronounced. There is less signal gain in the aromatic region and slightly more 'tailing'/loss in the aliphatic region. The average gain in the aromatic region is 32 % (maximum 92 %), and the average loss in the aliphatic region is 18 % (maximum 42 %).

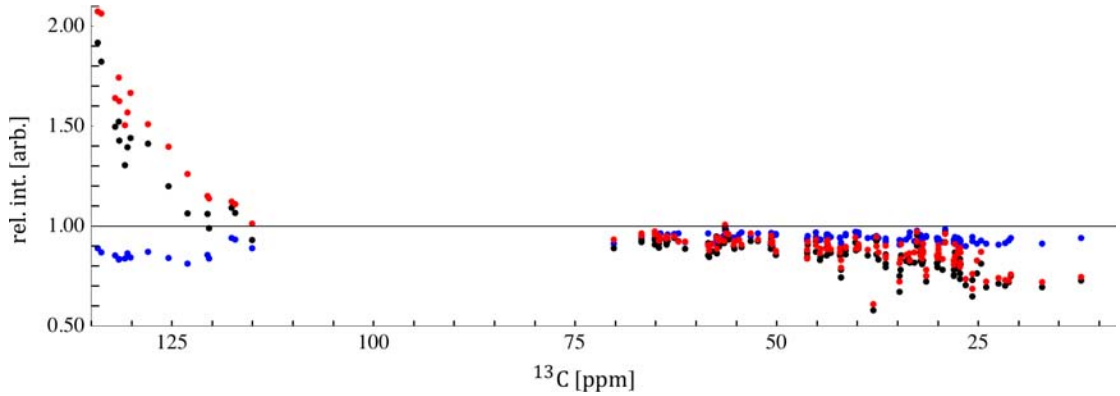


Figure 11. Comparison of experimental performance of a NOESY-HSQC using rectangular and shaped pulses. A 3D [^{15}N, ^{13}C]-resolved [^{1}H,^{1}H]-NOESY-HSQC was run in a 2D mode without sampling the indirect proton dimension t_1. Relative peak intensities obtained when the two 90° rectangular are replaced by 90° shaped pulses are shown in blue, when the two 90°-180°-90° composite pulses are replaced with 180° shaped pulse in red, and when all four pulses are replaced in black, respectively. Intensities are scaled to the reference spectrum with standard rectangular pulses applied at 45 ppm (reference line at 1). The BEBOP power level was 6.4 dB with length 550 ms, the theoretical excitation profile 37.5 kHz corresponding to 214.3 ppm @ 700 MHz. The BIBOP power level was 6.4 dB with length 252.5 ms, a theoretical inversion profile of 20 kHz corresponding to 114.3 ppm @ 700 MHz.

2.5. Discussion

2.5.1. Estimation of errors

In the following, we aim at establishing upper and lower limits of experimentally determined cross-relaxation rates and the distance restraints derived from those. To that purpose, we calculate the cumulative error obtained from the individual elements of pulse sequences shown in equation 1. To calculate the overall error, we first use the ratio of the intensities of a cross peak and the diagonal peak either from which the magnetization originates or to which the magnetization is transferred. An additional error arises from the fit of the cross-relaxation rate σ_{ij} from equations 10 and 11 because ρ_j is not known if the diagonal-peak of spin j cannot be evaluated. All numerical estimations derived in the previous sections and the resulting errors of the fitted cross-relaxation rate constants and distance limits are summarized in Table 2.

2.5.2. Cumulative error in 2D NOESY

We calculate an overall error expected for a cross-relaxation rate constant extracted from a 2D [^{1}H, ^{1}H]–NOESY experiment. In this case, T_i^{HXQC} is trivially 1 with an error of 0 %. Furthermore, we assume that the intensities are not affected by water suppression, that is, the error of α_i^{WS} is also 0 %. As the errors of T_{ij}^{NOESY} are vanishing, the only contributions to the error of I_{ij}/I_{ii} are α_{2i}^{EP} and α_{2j}^{EP}. For 1024 points in both dimensions, as used in our experimental setup, the resulting error is ca. 25 %.

Instead, one may extract the cross-relaxation rate constant from I_{ij}/I_{jj}. An additional error comes from the dispersion of the initial magnetization M_i^{SS}, which we determined to be ca. 12 %. The overall error then is slightly larger than above (31 %). However, this difference is virtually eliminated when the errors are further propagated into the error of the extracted distance (5 % in both cases). We conclude that the cross-peak intensities may be normalized by the intensities of the peak of magnetization origin or destination with equal success. An advantage of this approach is that errors introduced by the water suppression are cancelled. For example, the error of I_{ij}/I_{ii} would increase to 38 % if a peak were reduced by 20 % via water suppression. In that case, the use of I_{ij}/I_{jj} would be preferable (although the error of the distance is only increased to 7 %).

Table 2. Relative errors of various elements of NOESY pulse schemes, the extracted cross-relaxation rate constants and distance limits.

factor/quantity	error 2D NOESY		error 3D NOESY-HXQC	error 3D HXQC-NOESY[a]	
$M_i^{SS}(d_{int}, R_1)$	12 %		12 %	12 %	
$\alpha_{1,i}^{EP}(proc, J, R_2)$	18 % ($TD_{1,H}$ 1024)		12 % ($TD_{1,H}$ 200)	12 % ($TD_{1,H}$ 200)	
$T_{ij}^{NOESY}(\tau_{mix})$	0 %		0 %	0 %	
T_j^{HXQC}	0 %		51 % ($TD_{C/N}$ 80)	51 % ($TD_{C/N}$ 80)	
α_j^{WS}	0 %	20 % (water-gate)	0 % (purge pulse)	0 %	20 % (water-gate)
$\alpha_{2,j}^{EP}(proc, J, R_2)$	18 % ($TD_{2,H}$ 1024)		18 % ($TD_{2,H}$ 1024)	18 % ($TD_{2,H}$ 1024)	
I_{ij}	28 %	35 %	57 %	57 %	60 %
I_{ij} / I_{ii} (scaled to diagonal of spin of origin)	25 %	38 %	76 %	25 %	38 %
I_{ij} / I_{jj} (scaled to diagonal of spin of destination)	31 %		24 %	76 %	
extraction of σ_{ij} from $T_{ij}^{NOESY}(\tau_{mix})$	10 %		10 %	10 %	
σ_{ij} (scaled to diagonal of spin of origin)	27 %	39 %	77 %	27 %	39 %
σ_{ij} (scaled to diagonal of spin of destination)	32 %		26 %	77 %	
r_{ij} (scaled to diagonal of spin of origin) [b]	5 %	7 %	13 %	5 %	7 %
r_{ij} (scaled to diagonal of spin of destination) [b]	5 %		4 %	13 %	
$r_{ij}^{upl/lol}$ (scaled to diagonal of spin of origin) [c]	+5 % / -4 %	+9 % / -5 %	+28 % / -9 %	+5 % / -4 %	+9 % / -5 %
$r_{ij}^{upl/lol}$ (scaled to diagonal of spin of destination) [c]	+7 % / -5 %		+5 % / -4 %	+27 % / -9 %	

[a] Order of the factors in equation 1 is changed.

[b] Error propagated as $\Delta r/r = \Delta s/6s$.

[c] Error determined as the difference between r obtained from s and from $\sigma+\Delta\sigma$ or $\sigma-\Delta\sigma$.

2.5.3. Cumulative error in 3D NOESY

First, we analyze the experimental errors obtained from the 3D [^{15}N, ^{13}C]-resolved [^{1}H, ^{1}H]-NOESY-HSQC experiment. In contrast to 2D NOESY, T_i^{HXQC} contributes the major part to the overall error (51 %). This pulse sequence allows water suppression by a purge pulse, which relies on the evolution of heteronuclear scalar coupling in the protein giving independent control over the orientation of the operators of the protein and the uncoupled water operator. This procedure proved to be the best choice in practice because it offers the most effective water suppression and is the only one that does not contribute to further attenuation of the signals. Thus the impact on the error of α_i^{WS} can be set to 0 %. The number of points in the indirect ^{1}H dimension is much smaller than the one typically used for 2D experiments. Since a smaller number of points results in a lower resolution the dispersion during evolution periods is reduced. Therefore, the disadvantage of more pulse sequence elements in higher dimensional experiments is compensated to some extent. For 200 points as used in our experimental setup, the overall error of I_{ij}/I_{ii} is 76 %.

Interestingly, the error is significantly lower for I_{ij}/I_{jj} (24 %) because the impact of the main contributor to the error, T_i^{HXQC}, is cancelled out. The difference is sufficiently large to maintain a notable difference in the errors of the extracted distance (13 % versus 4 %). This indicates first that our previously employed normalization to the diagonal-peak of the magnetization origin indeed produces errors within the limits we chose (15 % or 20 %). However, it also shows that the pool of uni-directional eNOE distance limits can be extended by inclusion of those that are normalized to the magnetization destination peaks, possibly with tighter upper and lower limits.

When the order of the NOESY and HXQC elements is reversed as in a 3D [^{15}N, ^{13}C]-resolved HXQC-[^{1}H, ^{1}H]-NOESY experiment, the contribution of T_i^{HXQC} to the overall error cancels in I_{ij}/I_{ii} rather than in I_{ij}/I_{jj}. As a consequence, the errors of the distance are 5 % and 13 %, respectively. However, with the HXQC preceding the NOESY element, a separate water suppression element is required after the NOESY element. If we assume an additional error of 20 % from this element, the distance error obtained from I_{ij}/I_{ii} analysis increases to 7 %.

One of the main contributions to the intensity dispersion during the HXQC element, the suboptimal performance of ^{13}C rectangular pulses caused by offset effects, was analyzed in greater detail. We tested some of the most sophisticated shaped pulses. On the one hand, if a more quantitative analysis of the aromatic region is desired, these

pulses are those of choice. On the other hand, this comes at the cost of loss in the aliphatic region. Therefore, we recommend to employ the rectangular 90° and 90°-180°-90° inversion pulses for the HSQC element with the carrier frequencies set to 45 ppm in order not to attenuate the methyl region signals. In principle, the loss in the aromatic region could be reduced for rectangular pulses by shifting the carrier frequency. However, the eNOEs of the methyl region are of great importance for structure calculation and the methyl residues typically outnumber the aromatic residues. We choose to sacrifice the possible signal gain in the aromatic region for an improvement in the methyl region.

Another popular 3D NOESY pulse sequence is the ^{15}N-resolved NOESY. As the detected pathways of magnetization are almost identical to the counterparts in the [^{15}N, ^{13}C]-resolved 3D NOESY, we repeated the calculations using the distributions of the attenuation factors pertaining to that specific case (see Table S4 in appendix 10.1.). The main reduction of the distributions is the one of T_i^{HXQC}, which leads to a substantial reduction of the distance errors of r_{ij} (9 instead of 13 %) and r_{ji} (8 instead of 13 %) when the NOESY-HXQC and the HXQC-NOESY scheme, respectively, are used. Their two symmetry-related pathways (amide-amide NOEs) lead to identical errors. A further reduction of the errors could be achieved by deuteration of the sample, in which case, however NOEs between an amide and a non-amide proton would be sacrificed.

2.5.4. Experimental verification

In the following, we wish to verify our estimates of distance errors for 2D and 3D NOESY using both normalization strategies. A reference-free way to validate the estimated errors is direct comparison of the cross-relaxation rates and distances, which can be obtained for both cross-peaks from a single experiment. For the WW domain distances obtained from 2D NOESY, the overall pairwise difference is 11 % and 15 % of the actual distances when normalizing to the origin and destination magnetization intensity, respectively. This translates into errors of the individual data sets of 8 % and 11 %, which is in fair agreement with the predicted 5 %. For 3D NOESY-HSQC with normalization to the origin intensity, the pairwise difference is 16 %, whereas it is reduced to 11 % when normalized to the destination intensity. These figures translate into 12 % and 8 % for individual data sets. Again, they correlate well with the predicted 13 % and 4 %. We cannot exclude that additional small errors that we did not account for in our analysis contribute to the generally slightly larger differences in experiment. For example, water saturation effects may be transferred to the protein protons.

Next, we compare effective distances r^{eff}_{ij} (derived with equations 5 and 6 assuming a rigid molecule) obtained from 2D [^{1}H, ^{1}H]–NOESY and a 3D [^{15}N, ^{13}C]-resolved [^{1}H, ^{1}H]-NOESY-HSQC experiments to independently determined distances from an X-ray structure of Pin1 WW (pdb code: 2ZQT). The results are shown in Figure 12.

Distances from 2D NOESY data normalized to the origin of magnetization have a root-mean-square deviation (r.m.s.d.) from the X-ray distances of 0.96 Å with a Pearson's correlation coefficient of 0.67, comparable to the respective values of 0.98 Å and 0.67 when normalizing to the destination of magnetization. This similarity is in agreement with our predictions.

For the 3D NOESY derived distances however, there is a significant difference between normalizing to the origin, which results in an r.m.s.d. of 0.91 Å and a correlation coefficient of 0.55, or destination of magnetization with 0.82 Å and 0.65, respectively. Again, this trend agrees well with our estimations suggesting that normalizing to the destination magnetization reduces the error for this type of 3D NOESY.

The differences between the eNOE-derived distances are much smaller than those between any of the eNOE-derived ones and those from the X-ray structure (compare Figure 12 and Figure S5 in appendix 10.1.). This suggests that a substantial part of the r.m.s.d. between the eNOE and X-ray distances is due to altered distances caused by the crystallization conditions and the idealized geometry assumed when the proton were added to the heavy atoms. When comparing effective distances obtained from 2D NOESY to those of 3D NOESY, the smallest r.m.s.d. is observed between 2D data normalized to the origin and 3D data normalized to the destination of magnetization (see Figure S5 in appendix 10.1.), further supporting the predicted relatively large error introduced during the HSQC element. The r.m.s.d. of 0.64 Å (correlation coefficient 0.83) for distances that are on average ca. 4 Å amounts to ca. 16 %, which gives individual errors of ca. 11 % when attributed equally to both data sets (Figure S5 in appendix 10.1.). This would be in good agreement with our predictions of 8 % for both. Note that the difference between 2D- and 3D-derived distances must be seen as an upper limit for the accuracy as they may be at least partially caused by different samples and uncertainties in their respective τ_c of 3.6 ns ($^{13}C^{\alpha}$-relaxation measurements, at natural abundance) and 4.25 ns (^{15}N-relaxation measurements). Furthermore, because the 2D NOESY experiments were measured at higher field strength and longer mixing times than the 3D experiments (900 MHz and 0.08 s versus 700 MHz and 0.06 s), errors in

correction for spin diffusion may be propagated differently [37].

To test the generality of our findings, we calculated the same statistics for two other proteins for which we recorded 3D NOESY-HSQC. Our previously recorded eNOE set for GB3 is expected to show similar behavior because of a similar tumbling time as the one of the WW domain (ca. 4 ns) [23]. The comparison shows that the error for the cross-relaxation rates extracted with normalization to the origin is ca. 56 %, which is substantially smaller than the expected 77 %. As a consequence, the error in the distance is also smaller (8 % instead of the predicted 13 %). Presumably, some of the individual factors presented in Table 2 are somewhat different for GB3. To check the validity of our findings on a larger system, we calculated the distance errors for another protein currently studied in our laboratory, the second PDZ (post-synaptic density-95/discs large/zonula occludens-1) domain of the human tyrosine phosphatase 1E with a tumbling time of 6.9 ns. Normalizing to the origin, the pairwise difference of the distances is 17 %, whereas it is reduced to 14 % when normalized to the destination intensity (see Figure S7 in appendix 10.1.). These numbers translate into 12 % and 10 % for individual data sets. Again, they correlate reasonably well with the predicted 13 % and 4 %.

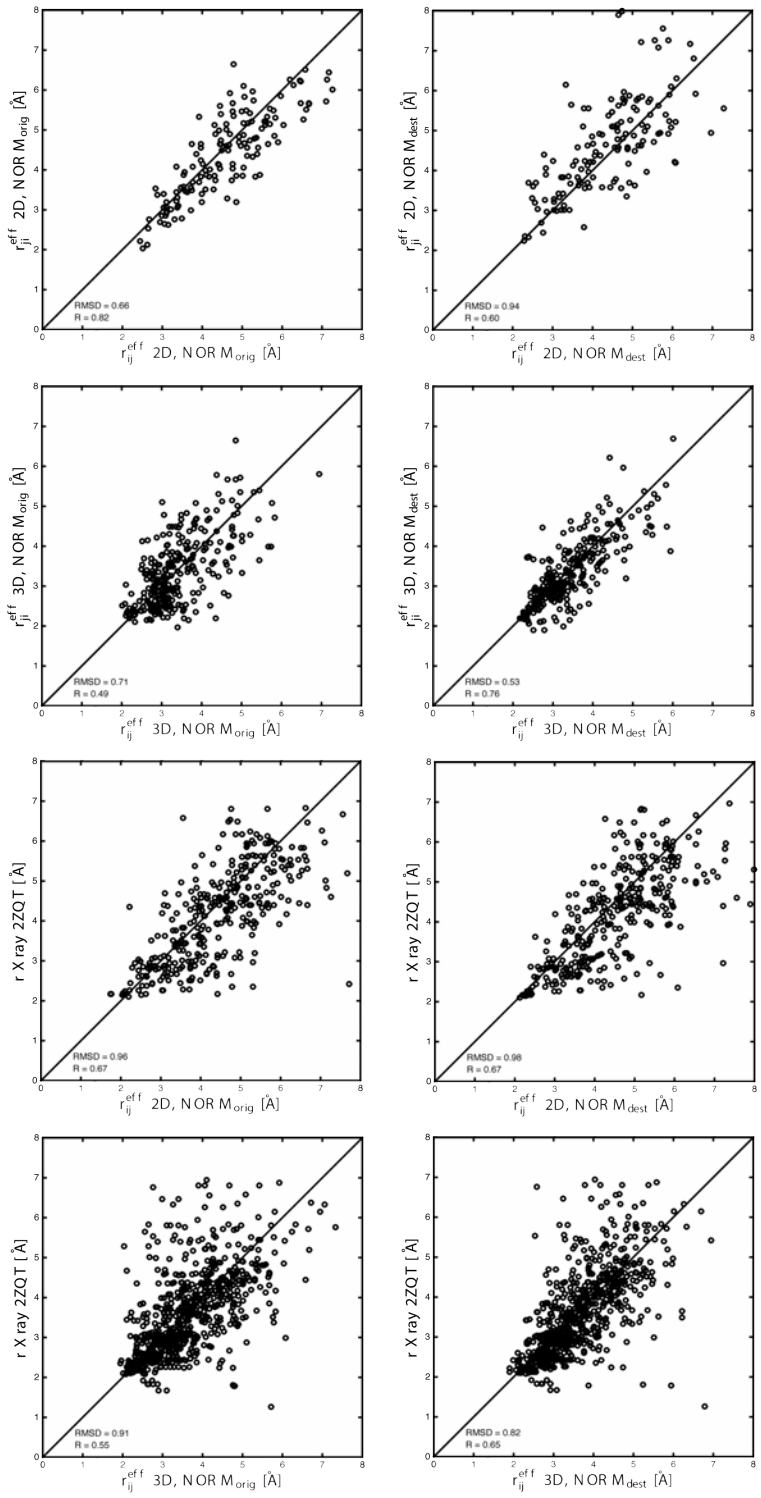


Figure 12: Validation of eNOE-derived distances r^{eff}_{ij}. In the first row, effective distances r^{eff}_{ji} versus r^{eff}_{ij} derived from 2D NOESY data normalized to the origin of magnetization (using $I_i^{diag}(0)$, $NORM_{orig}$) are shown on the left, and to the destination of magnetization (using $I_j^{diag}(0)$, $NORM_{dest}$) on the right. In the second row, r^{eff}_{ji} versus r^{eff}_{ij} derived from 3D NOESY data normalized to origin of magnetization ($NORM_{orig}$) are shown on the left, and to the destination of magnetization ($NORM_{dest}$) on the right. In the third row, X-ray-derived distances effective distances versus r^{eff}_{ij} derived from 2D NOESY data normalized to the origin of magnetization (using $I_i^{diag}(0)$, $NORM_{orig}$) are shown on the left, and those derived from data normalized to the destination of magnetization (using $I_j^{diag}(0)$, $NORM_{dest}$) on the right. In the fourth row, X-ray-derived distances effective distances versus r^{eff}_{ij} derived from 3D NOESY data normalized to the origin of magnetization (using $I_i^{diag}(0)$, $NORM_{orig}$) are shown on the left, and those derived from data normalized to the destination of magnetization (using $I_j^{diag}(0)$, $NORM_{dest}$) on the right. For plots showing only the subset of the distances common to the 2D and 3D sets, see Figure S6 in appendix 10.1.

2.5.5. Upper and lower distance limits

The relationship between the internuclear distance and the cross-relaxation rate constant is given by equation 5. Propagation of the experimental error of the cross-relaxation rate constant into the error of the distance yields:

$$\frac{\Delta r_{ij}}{r_{ij}} = \frac{1}{6}\frac{\Delta\sigma_{ij}}{\sigma_{ij}} \tag{19}$$

If the error of the cross-relaxation rate constant has a symmetric distribution, the error of the derived distance is highly asymmetric due to the r^{-6} dependence. As a consequence, the lower limit should be set much closer to the derived distance than the upper limit. We calculate those limits as:

$$\frac{\Delta^{\mathrm{upl}} r_{ij}}{r_{ij}} = \frac{r_{ij}(\sigma_{ij} - \Delta\sigma_{ij}) - r_{ij}(\sigma_{ij})}{r_{ij}(\sigma_{ij})} = \left(1 - \Delta\sigma_{ij}/\sigma_{ij}\right)^{-1/6} - 1 \tag{20}$$

$$\frac{\Delta^{\mathrm{lol}} r_{ij}}{r_{ij}} = \frac{r_{ij}(\sigma_{ij} + \Delta\sigma_{ij}) - r_{ij}(\sigma_{ij})}{r_{ij}(\sigma_{ij})} = \left(1 + \Delta\sigma_{ij}/\sigma_{ij}\right)^{-1/6} - 1 \tag{21}$$

where $r(.)$ is the distance as a function of the cross-relaxation rate constant, and 'upl' and 'lol' indicate the tolerances to be applied to the distance to obtain the upper and lower distance limits, respectively, for a structure calculation. The numerical values corresponding to our estimated errors of the cross-relaxation rates for the various NOESY experiments are listed in Table 2. While they are almost symmetric with ranges between -5 and +7 % for the 2D NOESY, 3D NOESY-HXQC using I_{ij}/I_{jj} and 3D HXQC-NOESY using I_{ij}/I_{ii}, they are strongly skewed for 3D NOESY-HXQC using I_{ij}/I_{ii} and 3D HXQC-NOESY using I_{ij}/I_{jj}, with +28 and +27% for the upper distance limits, and -9 % for both lower distance limits.

2.6. Conclusion

In this work, we have justified the previously published choice of upper and lower distance limits of ±15 to ±20 % when obtained from one cross-peak buildup. As a matter of fact, the errors introduced by the break of symmetry of the NOESY pathways are typically much smaller (ca. 4 % - 15 % depending on the extraction procedure). However, we recommend using the previously established limits because there are other error sources, such as errors in the correction for spin diffusion or in the overall tumbling time.

A second important finding is that, as opposed to common practice, normalization of the cross-peak intensities to the diagonal-peak intensity of the spin of the destination magnetization rather than the origin of magnetization is equally good for 2D NOESY

and even better for 3D NOESY-HXQC. This finding opens up the possibility to collect many more uni-directional eNOEs. As will be discussed in chapter 3, we have proposed to extract 'normalized generic' eNOEs from NOE buildups for which no diagonal-peak decay of the spin of magnetization origin could be fitted [25]. Such lower NOE restraints can be converted into upper limit restraints that are more realistic than those from conventional NOEs. Many of these 'normalized generic' eNOEs are expected to have a diagonal of the destination spin whose decay could be evaluated. Thus, application of the normalization to the destination intensity would result in tighter upper limit distance restraints, and also in lower distance restraints.

2.7. Experimental section

2.7.1. Preparation of samples

The unlabeled S18N mutants were prepared by solid-phase peptide synthesis [48] and purified either in the Jeffery W. Kelly lab or ordered from GL Biochem Ltd. The purified samples were suspended in NMR buffer (10mM K_2PO_4, 100mM NaCl, 0.02% NaN_3, pH 6.0) with 3% D_2O. The final NMR sample concentration was 1.2 mM.

In order to produce ^{15}N- and $^{15}N/^{13}C$-labeled S18N samples the S18N genetic sequence was ordered from GL Biochem Ltd. already preassembled in peT32 with a HIS_6-tag cleavable by TEV cleavage site. Transformed BL21/DE3 cells were grown in pre-cultures started from glycerol stock. In 2L M9 cultures (with either ^{15}N NH_4Cl or ^{15}N NH_4Cl/ ^{13}C glucose) the cells were grown form OD_{600} of 0.1 to OD_{600} of 0.7 at 37˚C and shaken at 120 RPM, then induced with IPTG. Following induction the temperature was reduced to 25˚C and left for expression for another 4 hrs before harvesting. Following two-step Ni-column purification the sample was desalted and TEV protease added in 1:50 (m/m) ratio and the sample left over night at room temperature. A further Ni-column purification step then provided the clean NMR sample. The sample's buffer was exchanged to the NMR buffer (10mM K_2PO_4, 100mM NaCl, 0.02% NaN_3, pH 6.0) using dialysis, then concentrated using 2kDa cutoff concentrator tubes (Sartorius Vivaspin 15R). The final NMR sample concentration was 1.2mM.

2.8. NMR spectroscopy

2.8.1. NOESY experiments

Series of 2D–[^{1}H, ^{1}H] NOESY experiments [27] were recorded on a Bruker 900 MHz spectrometer.

Series of 3D [^{15}N, ^{13}C]-resolved [^{1}H, ^{1}H]-NOESY-HSQC experiments were recorded to measure NOE buildups [23] on a Bruker 700 MHz spectrometer. The inter-scan delay was 0.8 s. Simultaneous [^{15}N, ^{1}H]- HSQC and [^{13}C, ^{1}H]-HSQC elements were employed, following indirect proton chemical shift evolution and [^{1}H, ^{1}H]-NOE mixing (τ_m). Diagonal-peak decays and cross-peak buildups were measured with τ_m of 20, 30, 40, 50, and 60 ms. The spectra were recorded with 400(t_1) x 80(t_2) x 2048(t_3) real points on a Bruker 700 MHz spectrometer. The maximal evolution times were $t_{1max,1H}$ = 22.0 ms, $t_{2max,15N}$ = 14.4 ms, $t_{2max,13C}$ = 7.6 ms, and $t_{3max,1H}$ = 102.4 ms, spectral widths $SW_{1,1H}$ = 13.0 ppm, $SW_{2,15N}$ = 39.2 ppm, $SW_{2,13C}$ = 29.8 ppm, $SW_{3,1H}$ = 14.3 ppm. The time-domain data were multiplied with a squared cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to 1024 x 128 x 512 real points.

2.8.2. τ_c measurements

^{15}N/^{13}C-labeled samples at 1.2 mM were used to perform T_1 and $T_{1\rho}$ ^{15}N-relaxation measurements of backbone amides in the β-sheet backbone ([^{15}N, ^{1}H]-TROSY experiments [30]). Relaxation delays for T_1 measurements were τ_m = 32, 112, 192, 352, 512, 673 and 833ms and for $T_{1\rho}$ measurements τ_m = 30, 60, 90, 120, 150 and 180 ms. Using T_1/T_2 ratios [40] τ_c at 1.2 mM and 278K was determined to be 4.25 ns. The relaxation experiments were recorded on a Bruker 700 MHz spectrometer.

For τ_c measurements with unlabeled samples, we used T_1 and $T_{1\rho}$ ^{13}C-relaxation measurements of C_α atoms [49] in the β-sheet at natural abundance. The experiments were modified from standard Bruker HSQC experiments to contain the $T_{1\rho}$ spin lock and spin lock-heating compensation. Relaxation delays for T_1 measurements were τ_{rel} = 10, 300, 600 and 800 ms and for $T_{1\rho}$ measurements τ_{rel} = 10, 40 and 70 ms. Using T_1/T_2

ratios [40], τ_c at 1.2 mM and 5˚C was determined to be 3.6 ns. The relaxation experiments were recorded on a Bruker 700 MHz spectrometer.

2.8.3. Auto-relaxation rate

To estimate an effective auto-relaxation rate during the interscan delay, the auto-relaxation times R_1 of well-resolved methyl peaks were calculated from an inversion recovery series and found to be ca. 1.4 s. Because methyl relaxation rates may be limited in representing all atom types, we also estimated the effective R_1 from NOESY experiments with two different interscan delays d_{int} since for a diagonal peak it is true that:

$$I_i^{\text{eq}}(d_{\text{int}}) = I_i^{\text{eq}}(0) \cdot (1 - e^{-R_1 \cdot d_{\text{int}}}) \tag{26}$$

Then setting $I_i^{\text{eq}}(d_{\text{int}}) = I_i^{\text{diag}}(0)$ and fitting

$$\overline{\left(I_{d_{\text{int},2}}^{\text{eq}}(0) / I_{d_{\text{int},1}}^{\text{eq}}(0) \right)} = \frac{(1 - e^{-R_1 \cdot d_{\text{int},2}})}{(1 - e^{-R_1 d_{\text{int},1}})} \tag{27}$$

yields an average R_1 of 1.2 s^{-1}, which is in good agreement with the result from the inversion recovery experiment. With $d_{\text{int},1}$ of 0.8 s and $d_{\text{int},2}$ of 4.0 s, respectively, 62 % and 99 % of the equilibrium polarization is reached.

2.8.4. $^3J_{HN,H\alpha}$ scalar couplings

The $^3J_{HN,H\alpha}$ couplings of the WW domain were measured by water-flip-back 2D constant time HMQC-J experiments [50] with ~110 μM of ^{15}N labeled sample at 5˚C. Experiments were recorded on a Bruker 700 MHz spectrometer.

2.8.5. Data analysis

Data was processed and analyzed with NMRpipe [51] by parabolic interpolation and assigned in CCPnmr [52]. We choose to use the peak intensities in preference to peak volumes. Reading the peak height directly out of the spectrum, as done in many programs, is in general a very robust, reliable and simple method. Previous experience has shown that peak volumes are in practice not more reliable than peak intensities (oral communication with Frank Delaglio, nmrPipe developer). In general, to obtain peak volumes a shape has to be fitted (in trivial cases a Lorentzian). In nmrPipe, peak picking

and fitting is performed using parabolic models. These are the intensities we chose as the basis for our analysis.

Upper and lower distance restraints from eNOEs were determined with the eNORA software package [37]. From the 2D NOESY, 89 well-resolved diagonal peaks (out of 269 protons) were fitted. 141 bi-directional eNOEs (see equation 14) were evaluated. Normalizing to the origin (I_{ij}/I_{ii}), an additional 515 uni-directional eNOEs were evaluated. Normalizing to the destination (I_{ij}/I_{jj}), an additional 602 uni-directional eNOEs were evaluated. From the 3D NOESY, 152 well-resolved diagonal-peaks (out of 269 protons) were fitted. 282 bi-directional eNOEs were evaluated. Normalizing to the origin (I_{ij}/I_{ii}), an additional 413 uni-directional eNOEs were evaluated. Normalizing to the destination (I_{ij}/I_{jj}), an additional 381 uni-directional eNOEs were evaluated. We were following the previously established protocol: i) The diagonal intensity of each residue was fitted mono-exponentially to back-predict the intensity at zero mixing time and to obtain the auto-relaxation rate. ii) The initial intensity was then used to normalize the cross-peak intensity measured at the various mixing times. iii) These normalized values were fitted to determine the cross-relaxation rate. iv) A correction for spin diffusion effects was estimated from simulations of apparent cross-relaxation rates from an input structure (lowest energy conformer from pdb code 1Pin [53]) using the full relaxation matrix method [37]. The upper and lower distance restraints were then set following the established protocol [23]. Magnetically equivalent protons were treated by r^{-6}-summation rather than a pseudo-atom approach [23]. Stereo specific assignments were determined as described previously [54].

2.9. References

[1] Neuhaus D, Williamson MP, The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2 end. New York: Wiley; 2000.

[2] Wüthrich K, NMR of Proteins and Nucleic Acids, New York: Wiley; 1986.

[3] James TL, Relaxation matrix analysis of two-dimensional nuclear Overhauser effect spectra, Curr Opin Struct Biol 1991, 1(6):1042-1053.

[4] Borgias BA, Gochin M, Kerwood DJ, James TL, Relaxation Matrix Analysis of 2D NMR Data, Prog Nucl Magn Reson Spectrosc 1990, 22:83-100.

[5] Boelens R, Koning T, Van der Marel GA, Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer, J Magn Reson 1989, 82(2):290-308.

[6] Boelens R, Koning TMG, Kaptein R, Determination of biomolecular structures from proton-proton NOE's using a relaxation matrix approach, J Mol Struct 1988, 173:299-311.

[7] Keepers JW, James TL, A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra, J Magn Reson 1984, 57(3):404-426.

[8] Dobson CM, Olejniczak ET, Poulsen FM, Ratcliffe RG, Time development of proton nuclear overhauser effects in proteins, J Mag Res 1982, 48(1):97-110.

[9] Kumar A, Wagner G, Ernst RR, Wüthrich K, Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: Implication for studies of protein conformation, J Am Chem Soc 1981, 103:3654-3658.

[10] Macura S, Ernst RR, Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy, Mol Phys 1980, 41(1):95-117.

[11] Vögeli B, The nuclear Overhauser effect from a quantitative perspective, Prog Nucl Magn Reson Spectrosc 2014, 78(C):1-46.

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

[13] Vögeli B, Orts J, Strotz D, Güntert P, Riek R, Discrete Three-dimensional Representation of Macromolecular Motion from eNOE-based Ensemble Calculation, CHIMIA International J Chem 2012, 66(10):787-790.

[14] Vögeli B, Friedmann M, Leitz D, Sobol A, Riek R, Quantitative determination of NOE rates in perdeuterated and protonated proteins: Practical and theoretical aspects, J Magn Reson 2010, 204(2):290-302.

[15] Vögeli B, Segawa TF, Leitz D, Sobol A, Choutko A, Trzesniak D, van Gunsteren W, Riek R, Exact Distances and Internal Dynamics of Perdeuterated Ubiquitin from NOE Buildups, J Am Chem Soc 2009, 131(47):17215-17225.

[16] Post CB, Internal motional averaging and three-dimensional structure determination by nuclear magnetic resonance, J Mol Biol 1992, 224(4):1087-1101.

[17] Bruschweiler R, Roux B, Blackledge M, Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution, J A Chem Soc 1992, 114(7):2289-2302.

[18] Koning TMG, Boelens R, Kaptein R, Calculation of the nuclear overhauser effect and the determination of proton-proton distances in the presence of internal motions, J Magn Reson 1990, 90(1):111-123.

[19] Torda AE, Scheek RM, van Gunsteren WF, Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat, J Mol Biol 1990, 214(1):223-235.
[20] Kessler H, Griesinger C, Lautz J, Mueller A, Van Gunsteren WF, Berendsen HJC, Conformational dynamics detected by nuclear magnetic resonance NOE values and J coupling constants, J Am Chem Soc 1988, 110(11):3393-3396.
[21] Olejniczak ET, Poulsen FM, Dobson CM, Proton nuclear Overhauser effects and protein dynamics, J Am Chem Soc 1981, 103(22):6574–6580.
[22] Tropp J, Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances, J Chem Phys 1980, 72:6035-6043.
[23] Vögeli B, Güntert P, Riek R, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol Phys 2013, 111(3):437-454.
[24] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19(10):1053-1057.
[25] Chi CN, Strotz D, Riek R, Vögeli B, Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals, J Biomol NMR 2015:1-7.
[26] Liu ML, Mao XA, Ye CH, Huang H, Nicholson JK, Lindon JC, Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy, J Magn Reson vol. 132; 1998: 125-129.
[27] Kumar A, Ernst RR, Wüthrich K, A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules, Biochem Biophys Res Commun 1980, 95(1):1-6.
[28] Meier BH, Ernst RR, Elucidation of chemical exchange networks by two-dimensional NMR spectroscopy: The heptamethylbenzenonium ion, J Am Chem Soc1979, 101(21):6441-6442.
[29] Fesik SW, Zuiderweg ERP, Heteronuclear three-dimensional nmr spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra, J Magn Res 1988, 78(3):588-593.
[30] Cavanagh J, Fairbrother WJ, Palmer III AG, Skelton NJ, Protein NMR spectroscopy: principles and practice, San Diego: Academic Press; 2007.
[31] Mishra SH, Harden BJ, Frueh DP, A 3D time-shared NOESY experiment designed to provide optimal resolution for accurate assignment of NMR distance restraints in large proteins, J Biomol NMR 2014, 60(4):265-274.
[32] Kay L, Clore G, Bax A, Gronenborn A, Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution, Science 1990, 249(4967):411-414.
[33] Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY, High-Resolution Four-Dimensional 1H− 13C NOE Spectroscopy using Methyl-TROSY, Sparse Data Acquisition, and Multidimensional Decomposition, J Am Chem Soc 2005, 127(8):2767-2775.
[34] Griesinger C, Gemperle C, Sørensen OW, Symmetry in coherence transfer, Mol Phys 1987, 62(2):295-332.
[35] Solomon I, Relaxation processes in a system of two spins, Phys Rev 1955, 993:559-565.
[36] Ernst RR, Bodenhausen G, Wokaun A, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Oxford: Claredon Press; 1987.
[37] Orts J, Vögeli B, Riek R, Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs, J Chem Theory Comput 2012, 8:3483-3492.

[38] Kowalski JA, Liu K, Kelly JW, NMR solution structure of the isolated Apo Pin1 WW domain: Comparison to the x-ray crystal structures of Pin1, Biopolymers 2002, 63(2):111-121.
[39] Lu KP, Hanes SD, Hunter T, A human peptidyl-prolyl isomerase essential for regulation of mitosis, Nature 1996, 380: 544-547.
[40] Kay LE, Torchia DA, Bax A, Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, Biochemistry 1989, 28: 8972-8979.
[41] Pascal SM, Muhandiram DR, Yamazaki T, Formankay JD, Kay LE, Simultaneous Acquisition of 15N- and 13C-Edited NOE Spectra of Proteins Dissolved in H2O, J Magn Reson, Series B 1994, 103(2):197-201.
[42] Guo C, Tugarinov V, Identification of HN-methyl NOEs in large proteins using simultaneous amide-methyl TROSY-based detection, J Biomol NMR 2009, 43(1):21-30.
[43] Coggins BE, Werner-Allen JW, Yan A, Zhou P, Rapid Protein Global Fold Determination Using Ultrasparse Sampling, High-Dynamic Range Artifact Suppression, and Time-Shared NOESY, J Am Chem Soc 2012, 134(45):18619-18630.
[44] Bostock MJ, Holland DJ, Nietlispach D, Compressed sensing reconstruction of undersampled 3D NOESY spectra: application to large membrane proteins, J Biomol NMR 2012, 54(1):15-32.
[45] Wang S, Zhou P, Sparsely-sampled, high-resolution 4-D omit spectra for detection and assignment of intermolecular NOEs of protein complexes, J Biomol NMR 2014, 59(2):51-56.
[46] Ehni S, Luy B, BEBEtr and BUBI: J-compensated concurrent shaped pulses for 1H-13C experiments, J Magn Reson 2013, 232(C):7-17.
[47] Kobzar K, Ehni S, Skinner TE, Glaser SJ, Luy B, Exploring the limits of broadband 90Â° and 180Â° universal rotation pulses, J Magn Reson 2012, 225(C):142-160.
[48] Kaul R, Angeles AR, Jäger M, Powers ET, Kelly JW, Incorporating β-Turns and a Turn Mimetic out of Context in Loop 1 of the WW Domain Affords Cooperatively Folded β-Sheets, J Am Chem Soc 2001, 123: 5206-5212.
[49] Korzhnev DM, Billeter M, Arseniev AS, NMR studies of Brownian tumbling and internal motions in proteins, Prog Nucl Magn Reson Spectrosc 2001, 38.
[50] Kuboniwa H, Grzesiek S, Delaglio F, Bax A, Measurement of HN-Hα J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods, J BiomolNMR 1994, 4: 871-878.
[51] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol NMR 1995, 6(3):277-293.
[52] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 2005, 59(1):687-696.
[53] Ranganathan R, Lu KP, Hunter T, Noel JP, Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent, Cell 1997, 89(6):875-886.
[54] Orts J, Vögeli B, Riek R, Güntert P, Stereospecific assignments in proteins using exact NOEs, J Biom NMR 2013, 57(3):211-218.
[55] Schorn C, Acquisition and Processing, NMR Spectroscopy: Data Acquisition, Wiley Online Library: 63-110.

3. Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals

This chapter is based on the following publication:

Celestine N. Chi, **Dean Strotz**, Roland Riek, Beat Vögeli, Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals, J Biomol NMR 2015, doi: 10.1007/s10858-015-9917-8. The idea of the paper was conceived and as an extension of the work described in chapter 2: The experimental accuracy of the uni-directional exact NOE, J Magn Reson 2015, doi:10.1016/j.jmr.2015.07.007. Dean Strotz wrote the software elements for the analysis, advised on the analysis and participated in writing of the manuscript.

3.1. Introduction

The fundament of NMR-based structure determination of biomolecules is the Nuclear Overhauser Effect (NOE) between hydrogens [1, 2]. Its traditional use however is only semi-quantitative. We have recently demonstrated that for small proteins distance restraints with accuracy smaller than 0.1 Å can be obtained by careful analysis of a series of NOESY experiments (eNOE) [3-9]. Subsequently, we used these exact restraints to calculate structural ensembles of the small model protein GB3 consisting of multiple rather than a single state [9-11]. Such ensembles were in significantly better agreement with the eNOEs and thus represent a more realistic sampling of the conformational space than single-state calculations [12-17]. These representations open an avenue for a comprehensive description of a protein's structural landscape and dynamics at atomic resolution [9, 18-20]. In order to extend the eNOE-based ensemble structure determination to larger biomolecules, we propose here a simplified approach to translate a NOE into a correct upper distance restraint and demonstrate it for the 16 kDa cis/trans isomerase human cyclophilin A [21-24], for which 3471 such restraints were collected.

In the eNOE analysis, the NOESY cross-peak intensities have to be normalized to back-predicted intensities of the diagonal peaks for zero mixing time [3-9]. As a consequence, cross peaks arising from spin pairs with overlapped diagonal peaks cannot be converted into exact distance restraints, which reduces the number of eNOE-derived distance restraints significantly. Supplementing eNOE with conventional NOE restraints would adversely impact the multiple-state ensembles as the restraints are calibrated in an

overall statistical manner. A comparison of upper distance limits either obtained from conventional NOEs following the standard CYANA [25] procedure or from our eNOE protocol is shown in Figure 1. Counting bi-directional NOEs once, 10% of the individual restraints potentially violate the true distances and would enforce false separation of states in multi-state ensembles. On the other hand, 55% of the restraints are at least 1 Å less restrictive than those obtained from eNOEs, which results in a large loss of information.

Omission of the conventional NOEs typically reduces the overall NOE data set by hundreds of restraints and the relative loss increases with increasing protein size. This effect is most pronounced for large protein systems because the chemical shift overlap increases approximately exponentially with increasing protein size [26, 27]. For example, while for the 56-residue protein GB3 823 eNOE- and 1041 NOE-derived distance restraints were collected [10, 11], the respective numbers are 1254 and 3471 for the 165-residue protein cyclophilin. In the following, we introduce a protocol to rescue the conventional NOEs and translate them into quantitative upper distance restraints. We show that restraints from cross peaks with overlapped diagonal peaks can provide a relatively tight upper limit, while never being lower than the true value. In such a way, multiple-state ensembles can be significantly better defined without causing artificial state separation.

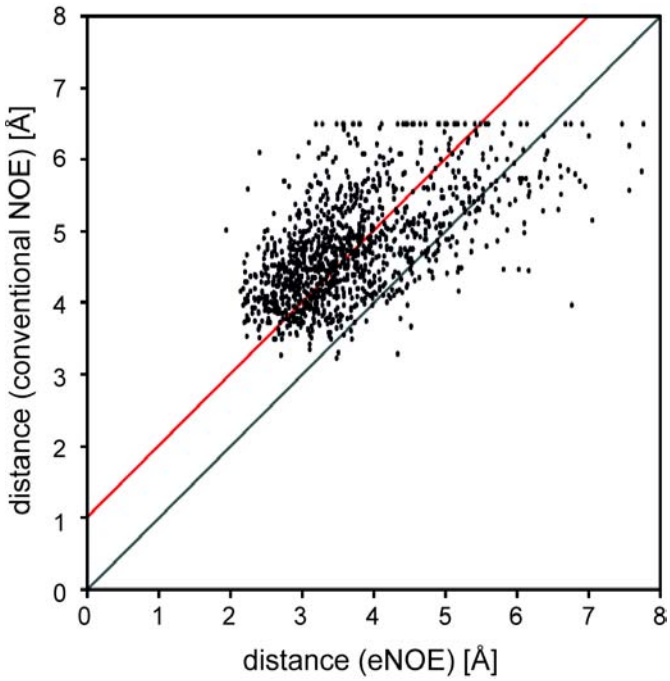


Figure 1. Comparison of upper distance limits obained from conventional NOEs with CYANA [25] and limits derived from eNOEs for cyclophilin A. Data points below the black line denote limits that potentially violate the true distances. Data points above the red line denote limits that are at least 1 Å less restrictive than those obtained from eNOEs.

3.2. Results and discussion

We find that the intensities of non-overlapped diagonal peaks in a series of 3D [^{15}N, ^{13}C]-resolved [^{1}H, ^{1}H]-NOESY spectra of human cyclophilin back-calculated to zero mixing time are dispersed over more than one order of magnitude (Figure 2). Non-uniform relaxation properties, scalar couplings, incomplete equilibration of the intrinsic pre-scan polarization (although an interscan delay of 1 s was chosen here), and suboptimal performance of the pulses on the hetero-nuclei due to offset effects cause differentiation of the intensities. Interestingly, all outliers with large intensities are located to the highly mobile termini and to the outmost positions of long side chains. Most of them are atoms in the N-terminal residues 1 and 2, the C-terminal residue 165 and the Q$^{\varepsilon}$ methyl groups of methionine (colored red in Figure 2).

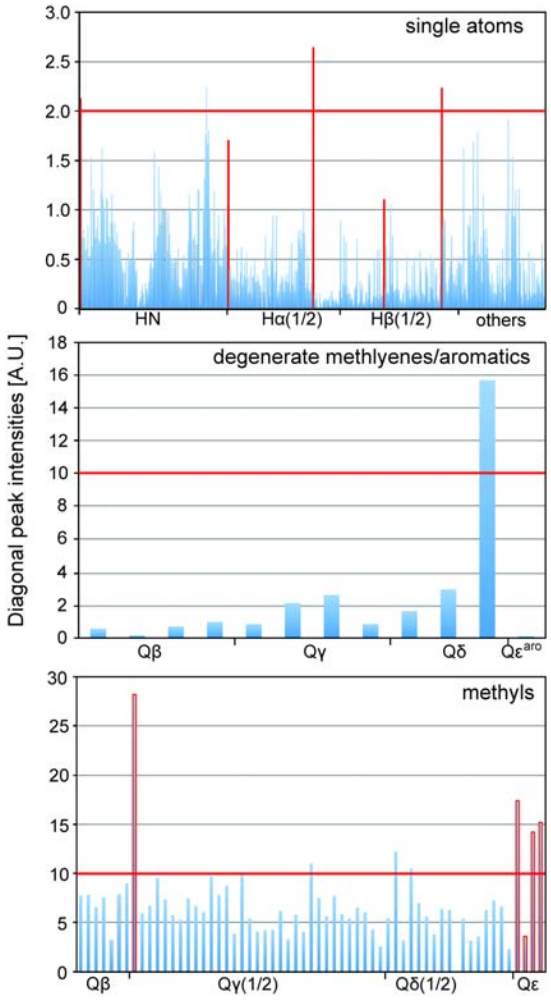


Figure 2. Back-calculated intensities of diagonal peaks in a 3D [^{15}N,^{13}C]-resolved [^{1}H,^{1}H]-NOESY spectrum of human cyclophilin A at zero mixing time. Intensities of single ^{1}H, degenerate methylene or aromatic, and methyl spins are shown in the top, middle, and bottom panel, respectively. The red horizontal lines indicate the chosen intensities used for the calculation of the upper distance limits from generic normalized eNOEs. Intensities of spins located in the termini and Q$^{\varepsilon}$ methyls of methionine are framed in red.

Given the simple identification rule for maximum outliers, it is straightforward to set an upper practical limit to the expected intensities even if the diagonal peaks cannot be analyzed due to peak overlap. The horizontal red lines in Figure 2 indicate such limits that are chosen larger than almost any intensity not corresponding to the identified risk group (i.e. terminal residues 1, 2, and 165, and the four methionine methyl groups Q^{ε} of residues 61, 100, 136, and 142). It is our rational that if a cross-peak buildup can be fitted, the curve may be normalized to the corresponding generic diagonal intensity resulting in a lower limit for the cross-relaxation rate. The cross-relaxation rate, in turn, can then be converted into an upper distance limit using the rotational correlation time measured from ^{15}N relaxation data. This upper distance limit is still of high quality since an offset in the diagonal intensity is greatly reduced in the distance due to the $1/r^6$ dependency of the NOE. For example, if the generic diagonal intensity is twice the true value, the upper distance limit is only 12% larger than the true value. In the following, we refer to such upper limits as distances derived from generic normalized eNOEs.

In principle, it is also possible to define a generic lower limit for the diagonal intensities. The use of such a limit would result in lower distance limits from cross peaks without available diagonal peak intensities. As opposed to upper intensity limits, lower limits may easily be violated in specific cases due to intensity reduction via exchange broadening or suboptimal pulses. Therefore, it is not recommended to use lower distance limits from generic normalized eNOEs.

To test the validity of the approach, we compare in the following upper distance limits extracted from generic normalized eNOEs to those from regular eNOEs of human cyclophilin A, respectively. 1254 upper limit restraints were obtained from ^{15}N,^{13}C-resolved [^{1}H, ^{1}H]-NOESY experiments with mixing times 15, 20, 25, 30, 40 and 50 ms. Both types of experimental eNOEs were corrected for spin diffusion using the eNORA program [28]. The correlations between the resulting upper distance limits are plotted in Figure 3. The 44 comparisons of the risk group are colored red. Excluding those, among the remaining 1210 distances there are no violations larger than 0.5 Å, two with violations larger than 0.3 Å, and three more restraints are violated by more than 0.2 Å. The largest one is caused by the $Q^{\gamma 2}$ (Val20) - H^{N} (Ser21) spin pair with 0.43 Å. Inspection reveals that it is caused by the fact that only one cross peak could be used for the eNOE, while both were used for the generic normalized eNOE. Indeed, comparison of the cross-relaxation rate corresponding to the same cross peak shows a nearly identical value. To be cautious, these upper limit restraints are multiplied by a factor of 1.2 for the structure calculation, as done for the unidirectional eNOE-derived distances

[10, 11]. Thus, no restraints are violated and the presented approach is highly reliable. It is worth to mention that it is possible that the risk group for another system may also include other highly flexible segments such as a protein loop. Such segments should be easy to identify by resolved diagonal peaks originating from protons in the segment. Although not the case in our study (methione being the exception), other candidates are ends of long side chains such as those of lysines. These candidates are expected to cause violations in the structure determination process and thus should also be identified.

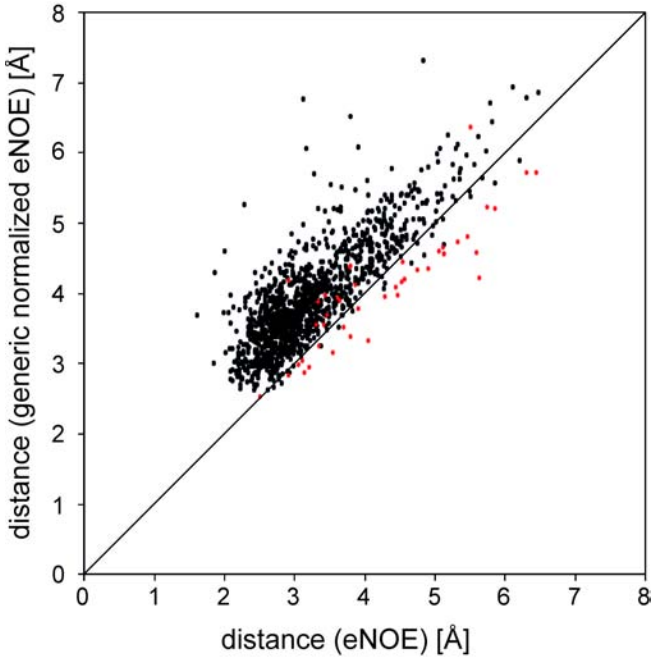


Figure 3. Upper distance limits obtained from generic normalized eNOEs versus those from eNOEs. If the cross-relaxation rate is normalized to the diagonal peak intensity of a terminal residue or Q^{ε} of methionine, it is marked red. The black line shows slope 1.

Importantly, 2217 additional generic normalized eNOEs were determined without an eNOE counterpart enlarging the number of restraints for cyclophilin A by a factor of almost 2. To show their impact, we calculated structures with the software package CYANA [25] using either only the 1254 upper and lower distance limits from eNOEs or supplemented with the limits from the generic normalized eNOEs (omitting the ones from the risk group, Tables 1 and 2). Although the 1254 restraints from the eNOEs alone define the structure of cyclophilin A relatively well, the r.m.s.d from the mean structure in the ordered segments is reduced by 0.25 Å for the backbone, and by 0.30 Å for all heavy atoms, when the complete data set is used (Figure 4). A structure calculation with 4537 upper distance limits from conventional NOEs produces r.m.s.d. values that are 0.06 Å and 0.03 Å larger than those obtained from eNOEs only, and 0.31 Å and 0.33 Å larger than those derived from the complete eNOE data set. These effects are largely

masked by the high flexibility of the termini and the loop comprising residues 66-76 if the statistics are taken over the entire molecule (Table 2).

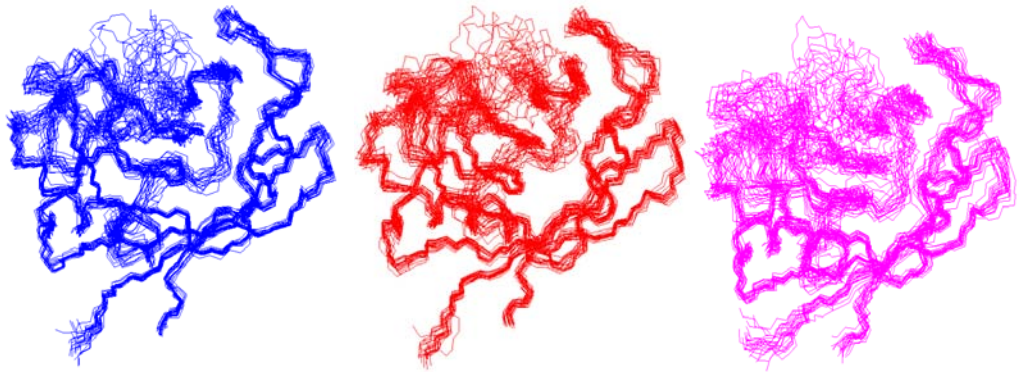


Figure 4. Impact of supplement of the generic normalized eNOEs on the structure calculation of cyclophilin A. Structures calculated using 1254 eNOEs supplemented with 2217 generic normalized eNOEs (left), 1254 eNOE alone (middle) and 4537 conventional NOEs (right).

An increased fraction of the generic normalized eNOEs is expected for larger systems due to more resonance overlap and as verified by ongoing work in our laboratory. We simulated such a scenario for cyclophilin by deletion of 400 or 800 upper and lower distance limits from the original eNOE data set. The impact of the generic normalized eNOEs becomes very pronounced as demonstrated in Figure 4 and Tables 1 and 2. When using a 5:2 (2144:854) instead of the original 5:3 (2144:1254) ratio of the counts of generic normalized eNOEs to regular eNOEs, the structural r.m.s.d. in the ordered segments of cyclophilin comprising residues 3-65 and 77-163 is reduced by ca. 0.6 Å for both the backbone as well as all the heavy atoms upon exclusion of the generic normalized eNOEs (Table 1). The r.m.s.d. reduction is more than 1.2 Å (backbone) and 1.4 Å (all heavy atoms) for a 5:1 (2144:454) ratio. Although the r.m.s.d. reductions are weaker for the entire molecule (Table 2), similar trends as those for the ordered segments are observed.

Table 1. Impact of generic normalized eNOEs on CYANA structures of cyclophilin A: Statistics of the ordered residues

# upper/lower limits from eNOEs	# upper limits from generic normalized eNOEs	r.m.s.d. to the mean[a], backbone	r.m.s.d. to the mean[a], all heavy atoms
20 single-state structures			
1254/1254	0	0.84 ± 0.11 Å	1.27 ± 0.13 Å
1254/1254	2144	0.59 ± 0.07 Å	0.97 ± 0.07 Å
854/854	0	1.23 ± 0.17 Å	1.69 ± 0.16 Å
854/854	2144	0.65 ± 0.12 Å	1.00 ± 0.11 Å
454/454	0	1.97 ± 0.22 Å	2.63 ± 0.19 Å
454/454	2144	0.72 ± 0.11 Å	1.11 ± 0.09 Å
4537 conventional NOEs		0.90 ± 0.11 Å	1.30 ± 0.12 Å
20 two-states ensembles			
1254/1254	0	1.19 ± 0.14 Å	1.71 ± 0.17 Å
1254/1254	2144	0.97 ± 0.12 Å	1.56 ± 0.20 Å
4537 conventional NOEs		1.23 ± 0.14 Å	1.79 ± 0.16 Å

[a] Residues 3-65,77-163

Table 2. Impact of generic normalized eNOE data set on CYANA structures of cyclophilin A: Statistics of all residues

# upper/lower limits from eNOEs	# upper limits from generic normalized eNOEs	r.m.s.d. to the mean[a], backbone	r.m.s.d. to the mean[a], all heavy atoms	eNOE + generic normalized eNOE violations
20 single-state structures				
1254/1254	0	1.21 ± 0.21 Å	1.66 ± 0.21 Å	57
1254/1254	2144	1.23 ± 0.18 Å	1.52 ± 0.15 Å	153
854/854	0	1.52 ± 0.22 Å	2.00 ± 0.21 Å	24
854/854	2144	1.32 ± 0.31 Å	1.62 ± 0.27 Å	76
454/454	0	2.40 ± 0.33 Å	3.02 ± 0.28 Å	7
454/454	2144	1.27 ± 0.22 Å	1.58 ± 0.16 Å	29
4537 conventional NOEs		1.39 ± 0.34 Å	1.79 ± 0.30 Å	15
20 two-states ensembles				
1254/1254	0	1.57 ± 0.26 Å	2.11 ± 0.25 Å	10
1254/1254	2144	1.32 ± 0.16 Å	1.89 ± 0.21 Å	12
4537 conventional NOEs		1.60 ± 0.27 Å	2.16 ± 0.25 Å	0

[a] Residues 2-164

While the lower r.m.s.d for the eNOE data provides a measure of the increased completeness over the conventional NOE networks, it does not quantify the quality of the data. We observe similar trends for the r.m.s.d in two-states ensembles (Table 1). However, we observe that bundles derived from conventional NOEs do not provide clear state separation, while only the ensemble calculated from eNOEs and generic normalized eNOEs produce consistently two states even in the loops (to be published elsewhere).

In conclusion, we introduced a protocol to enlarge the eNOE data set substantially without adding semi-quantitative restraints such as those derived from conventional NOEs that may contradict true distances. Because this data does not enforce false

separation of states, eNOE-based multiple-state ensemble structure calculation becomes applicable to larger, biologically relevant proteins.

3.3. Materials and methods

3.3.1. Sample preparation

The gene enconding the full-length human cyclophilin was sub-cloned in a pRSET/A vector (Invitrogen) containing an N-terminal His-tag with a thrombin cleavage site (MHHHHHHLVPRGS). Expression and purification will be described elsewhere (Chi et al., manuscript submitted). Briefly, the cDNA containing the cyclophilin gene was transformed into *E. coli* BL21 plyS cells and plated on an ampicillin/chloramphenicol-containing plate. A 1-liter culture in M9 medium for $^{15}N/^{13}C$- or ^{15}N-labeling was initiated and cells were grown to an OD_{600} of 0.9. Protein expression was initiated by adding 1 mM IPTG (isopropyl β-D-thiogalactopyranoside). Cells were then allowed to express over night at 18 ^{o}C. Cells were harvested, re-suspended in purification buffer (10 mM Tris/HCl) and lysed by French press and spun at 40,000 *g* for 20 minutes. The supernatant was filtered and loaded onto a nickel (II)-charged chelating sepharose FF column (Amersham Biosciences), equilibrated with purification buffer as above and washed with 400 ml of the same buffer. The bound protein was eluted with 250 mM imidazole at pH 7.9, in aliquots of 12 ml. Fractions containing partially pure proteins were pooled, dialyzed for two hours and passed through a DEAE S-column equilibrated with purification buffer. The His-tag was cleaved off by incubating in thrombin for three hours at room temperature. Pure cyclophilin was collected as flow-through by passing the digested fractions through a nickel (II)-charged chelating sepharose FF column equilibrated with purification buffer. The purity was checked on SDS PAGE stained with coomassie brilliant blue and its identity confirmed by Matrix Assisted Laser Desorption Ionization time-of-flight mass spectrometry (MALDI-TOF). The final NMR samples contained protein at a concentrations of 1 to 2 mM (as determined by absorption measurements) in 10 mM sodium phosphate, pH 6.5, 5 mM DTT, 0.01 % $NaNO_3$ and 3% D_2O.

3.3.2. NMR spectroscopy and data analysis

All NMR experiments were run on Bruker 600 and 700 MHz spectrometers equipped with triple resonance cryogenic probes at 299 K. Data were processed and analyzed with NMRpipe [29] and CCPnmr [30].

For assignment purposes 3D HNCACB, ^{15}N-resolved [^{1}H,^{1}H]-NOESY-HSQC, ^{15}N-resolved HMQC-[^{1}H,^{1}H]-NOESY, ^{15}N-resolved [^{1}H,^{1}H]-TOCSY-HSQC and ^{13}C-resolved HCCH-TOCSY spectra were recorded [31]. Cross peak buildups and diagonal peak decays were probed with 3D ^{15}N/^{13}C-resolved [^{1}H,^{1}H]-NOESY-HMQC experiments using mixing times of 15, 20, 25, 30, 40 and 50 ms [11].

1254 upper and lower distance restraints from eNOEs were determined with the eNORA software package [28] following the previously established protocol: i) The diagonal intensity of each residue was fitted mono-exponentially to back-predict the intensity at zero mixing time and to obtain the autorelaxation rate. ii) The initial intensity was then used to normalize the cross peak intensity measured at the various mixing times. iii) These normalized values were fitted to determine the cross-relaxation rate. iv) A correction for spin diffusion effects was estimated from simulations of apparent cross-relaxation rates from an input structure (lowest energy conformer from pdb code 1oca [23]) using the full relaxation matrix method [28]. v) The upper and lower distance restraints were then set following the established protocol [11]. Magnetically equivalent protons were treated by r^{-6}-summation rather than a pseudo-atom approach as used previously [11].

3471 generic normalized eNOEs were generated with the same eNORA protocol as used for the eNOEs with the following exceptions: i) The back-predicted diagonal peak intensities at zero mixing time were plotted as shown in Figure 2. Separate generic diagonal peak intensities were chosen for three different classes of peaks, the first including all single atoms, the second the degenerate and aromatic atoms, and the third the methyl groups, respectively. The values were set to 2, 10 and 10 (note, the scale for peak intensities comes in arbitrary units). The values were selected such that they are larger than all diagonal intensities except for a few outliers that are easily categorized as the risk group (see main text). The resulting upper distance limits are rather insensitive to the particular choice due to the r^{-6} dependency of the cross-relaxation rate.

4537 conventional NOEs were extracted from the 3D ^{15}N/^{13}C-resolved [^{1}H,^{1}H]-NOESY-HMQC experiment with a mixing time of 50 ms. The NOE intensities I were converted into upper distance limits r^{CYANA} using the standard CYANA procedure [25]:

$$r^{\mathrm{CYANA}} = r^{\mathrm{reference}} * [\mathrm{median}(I \text{ of all cross peaks})/I]^{1/6} \qquad (1)$$

For $r^{\mathrm{reference}}$, we chose the default setting of 4.0 Å. Distances larger than 6.5 Å were set to 6.5 Å.

^{15}N R_1, $R_{1\rho}$, and R_2 relaxation rates were recorded with standard [^{15}N, ^{1}H]-TROSY type experiments [31] and resulted in a overall tumbling time τ_c of 9.2 ns.

3.3.3. Structure calculation

Upper and lower distance restraints from eNOEs and generic normalized eNOEs or from conventional NOEs were used together with conservative Φ and Ψ dihedral angle restraints from C^{α} chemical shifts for structure determination. The weight of the dihedral angle restraints were reduced to zero in the final part of the calculation. Structural coordinates were determined with version 3.97 of the program CYANA [25]. Calculations were done with 200'000 torsion angle dynamics steps for 200 conformers with random torsion angles by simulated annealing. The 20 conformers with the lowest final target function values were selected and analyzed. The coordinates of the single state bundle calculated from the complete data set consisting of the eNOEs and the generic normalized eNOEs together with the upper/lower distance limit tables are deposited in the PDB/BMRB (pdb code 2MZU; bmrb accession code 25502). The two structural states were calculated simultaneously and averaged for the two-state ensemble. A weak harmonic well potential with bottom width of 1.2 Å was used to keep identical heavy atoms from the different states together following the previous proposal [10, 11].

3.4. References

[1] Wüthrich K, NMR of proteins and nucleic acids, Wiley, New York, 1986.

[2] Neuhaus D, Williamson MP, The Nuclear Overhauser Effect in Structural and Conformational Analysis, Wiley, New York, 2000.

[3] Kumar A, Wagner G, Ernst RR, Wüthrich K, Buildup rates of the nuclear Overhauser effect measured by two-dimensional proton magnetic resonance spectroscopy: Implication for studies of protein conformation, J Am Chem Soc 1981, 103:3654-3658.

[4] Keepers JW, James TL, A theoretical study of distance determinations from NMR, Two-dimensional nuclear Overhauser effect spectra, J Magn Reson 1984, 57:404-426.

[5] Boelens R, Koning TMG, Kaptein R, Determination of biomolecular structures from proton-proton NOEs using a relaxation matrix approach, J Mol Struct 1988, 173:299-311.

[6] Boelens R, Koning TMG, Van der Marel GA, Van Boom JH, Kaptein R, Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach, Application to a DNA octamer, J Magn Reson 1989, 82:290-308.

[7] Vögeli B, Segawa T, Leitz D, Sobol A, Choutko A, Trzesniak D, Van Gunsteren W, Riek R, Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups, J Am Chem Soc 2009, 131:17215-17225.

[8] Vögeli B, Friedmann M, Leitz D, Sobol A, Riek R, Quantitative determination of NOE rates in perdeuterated and protonated proteins: practical and theoretical aspects, J Magn Reson 2010, 204:290-302.

[9] a) Vögeli B, The nuclear Overhauser effect from a quantitative perspective, Prog Nucl Magn Reson Spectrosc 2014, 78:1-46;
b) Vögeli B, Orts J, Strotz D, Chi C, Minges M, Wälti MA, Güntert P, Riek R, Towards a true protein movie: A perspective on the potential impact of the ensemble-based structure determination using exact NOEs, J Magn Reson 2014, 241:53-59.

[10] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19:1053-1057.

[11] Vögeli B, Güntert P, Riek R, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol Phys 2013, 111:437-454.

[12] Brüschweiler R, Blackledge M, Ernst RR, Multi-conformational peptide dynamics derived from NMR data: A new search algorithm and its application to antamanide, J Biomol NMR 1991,1:3-11.

[13] Brüschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR, Influence of rapid intramolecular motions on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution, J Am Chem Soc 1992,114:2289-2302.

[14] Bonvin AMJJ, Boelens R, Kaptein R, Time- and ensemble-averaged direct NOE restraints, J Biomol NMR 1994,4:143-149.

[15] Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M, Simultaneous determination of protein structure and dynamics, Nature 2005, 433:128-132.

[16] Olsson S, Vögeli B, Cavalli A, Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Hamelryck T, Probabilistic determination of native state ensembles of proteins, J Chem Theory Comput 2014, 10:3484-3491.

[17] Mantsyzov AB, Maltsev AS, Ying J, Shen Y, Hummer G, Bax A, A maximum entropy approach to the study of residue-specific backbone angle distributions in a-synuclein, an intrinsically disordered protein, Prot. Science 2014, 23:1275-1290.
[18] Fenwick RB, Esteban-Martin S, Salvatella X, Understanding biomolecular motion, recognition and allostery by use of conformational ensembles, Eur Biophys J 2011, 40:1339-1355.
[19] Ravera E, Salmon L, Fragai M, Parigi G, Al-Hashimi H, Luchinat C, Insights into domain-domain motions in proteins and RNA from solution NMR, Acc Chem Res 2014, 47:3118-3126.
[20] Torchia DA, NMR studies of dynamic biomolecular conformational ensembles, Prog Nucl Magn Reson Spectrosc 2015, 84-85:14-32.
[21] Handschumacher E, Harding MW, Rice J, Drugge RJ, Speicher DW, Cyclophilin: A specific cytosolic binding protein for cyclosporin A, Science 1984, 226:544-547.
[22] Clubb RT, Ferguson SB, Walsh CT, Wagner G, Three-dimensional solution structure of Escherichia coli periplasmic cyclophilin, Biochemistry 1994, 33:2761-2772.
[23] Ottiger M, Zerbe O, Güntert P, Wüthrich K, The NMR solution conformation of unligated human cyclophilin A, J Mol Biol 1997, 272:64-81.
[24] Wang P, Heitman J, The cyclophilins. Genome Biol. 2005, 6:226.
[25] Güntert P, Automated structure determination from NMR Spectra, Eur Biophys J 2009, 38:129-143.
[26] Mumenthaler C, Güntert P, Braun W, Wüthrich K, Automated combined assignment of NOESY spectra and three-dimensional protein structure determination, J Biomol NMR 1997, 10:351-362.
[27] Güntert P, Methods and Principles in Medicinal Chemistry, Wiley-VCH, Weinheim, Germany, 2003, 16, 39-66.
[28] Orts J, Vögeli B, Riek R, Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs, J Chem Theory Comput 2012, 8:3483-3492.
[29] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, NMRPipe: A multidimensional spectral processing system based on UNIX pipe, J Biomol NMR 1995, 6:277-293.
[30] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED, The CCPN data model for NMR spectroscopy: Development of a software pipeline, Proteins 2005, 59:687-696.
[31] Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skleton NJ, Protein NMR spectroscopy. Principles and practice, Academic Press, San Diego, 2007.

4. The eNORA2 exact NOE analysis program

This chapter is based on the following publication (to be submitted):

Dean Strotz, Julien Orts, Celestine Chi, Roland Riek, Beat Vögeli, The eNORA2 exact NOE analysis program. Dean Strotz wrote the software except for the FM approach (implemented by Julien Orts), prepared figures and wrote the manual and participated in writing of the manuscript.

4.1. Introduction

Almost all 3D protein structures solved by NMR are modeled with upper proton-proton distance limits derived from nuclear Overhauser enhancement (NOEs) [1, 2]. With the aim of correct average conformations in mind, simple interpretations of the NOESY cross peak volumes or intensities have become the method of choice. However, more quantitative interpretations of NOESY spectra can result in very accurate distance restraints [3-7]. The potential of such restraints for studies of the spatial sampling in addition to the average conformation of a protein has been recognized [8-14]. We have developed protocols for exact NOE determination from NOESY series [3, 15, 16] and for generation of multiple-state structural ensembles from the derived upper and lower distance limits [17-21].

Although our approaches are relatively simple, initial evaluation of the data was time consuming due to the large amount of data (typically thousands of peaks per protein), and due to the book-keeping requirements for uni- and bi-directional NOEs, magnetically equivalent atoms, etc. Here, we present a Matlab package that automatizes all data processing steps required to convert intensities of assigned peaks in NOESY series into upper and lower distance limits for structure calculation. The program includes previously published features of eNORA (*e*xact *NO*E by *r*elaxation matrix *a*nalysis) [22], but contains important new functions and integrates the NMR data analysis with CYANA structure calculation [23, 24]. Most prominently, partially deuterated molecules can be treated now, and options are implemented for correction for spin diffusion without stereospecific assignment, saving plots of fits to cross-peak and diagonal-peak intensities as separate files, edition of mixing times, buildup normalization to diagonal peak intensities of spins of magnetization origin or destination, generation of

upper and lower limit distance files for structure calculation with the CYANA package and generation of upper limits from buildups for which the diagonal peaks cannot be evaluated. The program is intended for experts and non-experts alike and a deeper understanding of the spin dynamics is not necessary.

eNORA2 builds in part on ideas implemented in previously published programs. The most notable programs that use a full matrix approach to extract distances from NOESY employ hybrid approaches, where intensity or relaxation matrices combine entries from experiment and estimated quantities from temporary structural models. Examples are IRMA (*I*terative *r*elaxation *m*atrix *a*pproach; extraction of distances from a hybrid intensity matrix composed of measured and theoretical intensities using restrained molecular dynamics simulations) [25-27], DSPACE (similar to IRMA) [28], MARDIGRAS (*M*atrix *a*nalysis of *r*elaxation for *di*scerning *g*eomet*r*y of an *a*queous *s*tructure; extraction of distances from a hybrid intensity matrix composed of measured and theoretical intensities building on CORMA, see below) [29], MORASS (*M*ultispin/*m*ultiple *O*verhauser *r*elaxation *a*nalysi*s* and *s*imulation) [30,31] and FIRM (*F*ull *i*terative *r*elaxation *m*atrix) [32]. Some programs, such as MIDGE (*M*odel-*i*ndependent *d*istance *ge*neration) [33] and DISCON (*Di*stribution of *s*olution *con*formations) [34, 35], do not employ temporary estimates of the distances. Others again use algorithms that couple energy terms to NOESY intensities [36] such as DINOSAUR (*Di*rect *NO*E *s*imulation *a*pproach for *u*nbelievable structure *r*efinement) [37] and COMATOSE (*Co*mplete *m*atrix *a*nalysis *t*orsion *o*ptimized *s*tructur*e*) [38]. Implementations using the full matrix approach were also developed for the popular structure calculation programs CYANA [39], X-PLOR [40,41], DISCOVER [42] and ARIA [43]. These sophisticated approaches were predated by simple predictions of peak intensities based on structural models, which then could be compared to experimental intensities, for example with NOEMOT [44], CORMA (*Co*mplete *r*elaxation *m*atrix *a*nalysis) [45] and CROSREL (*Cro*ss *rel*axation) [46].

4.2. Methods

eNORA2 is an object-oriented program written in Matlab [47] format.

4.2.1. Theory

The mixing time-dependence of the NOESY intensities may be described by multi-spin Solomon equations [3, 48]. For a system with *N* spins, a two-dimensional NOESY element produces *N* x *N* peaks of which *N* are diagonal peaks and *N* x (*N-1)* are cross peaks. Their intensities may be written in matrix form:

$$I(\tau_{mix}) = \begin{pmatrix} I_{11}(\tau_{mix}) & \dots & I_{1i}(\tau_{mix}) & \dots & I_{1N}(\tau_{mix}) \\ \dots & & \dots & & \dots \\ I_{i1}(\tau_{mix}) & \dots & I_{ii}(\tau_{mix}) & \dots & I_{iN}(\tau_{mix}) \\ \dots & & \dots & & \dots \\ I_{N1}(\tau_{mix}) & \dots & I_{Ni}(\tau_{mix}) & \dots & I_{NN}(\tau_{mix}) \end{pmatrix} \quad (1)$$

In practice, most of these peaks are buried in spectral noise since only a fraction of all proton pairs have a sufficiently short internuclear distance to be detected.

The time evolution of $I(t_{mix})$ is given by

$$I(\tau_{mix}) = I(0)e^{-\boldsymbol{R}\tau_{mix}} \quad (2)$$

where ***R*** is the relaxation matrix containing the auto- and cross-relaxation rate constants ρ_i and σ_{ij}:

$$R = \begin{pmatrix} \rho_1 & \sigma_{12} & \dots & \sigma_{1i} & \dots & \sigma_{1N} \\ \sigma_{21} & \rho_2 & & \sigma_{2i} & & \sigma_{2N} \\ \dots & \dots & & \dots & & \dots \\ \sigma_{i1} & \sigma_{i2} & \dots & \rho_i & \dots & \sigma_{iN} \\ \dots & \dots & & \dots & & \dots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{Ni} & \dots & \rho_N \end{pmatrix} \quad (3)$$

For extraction of the cross-relaxation rate constants from a NOESY series, eNORA2 makes use of a nonlinear ISPA (isolated spin-pair approach) with correction for spin diffusion [22]. Under the assumption that all spin pairs giving rise to a cross peak are ideal two-spin systems *i-j*, the exact analytical solution for the cross-peak buildup of equation 2 is

$$\frac{I_{ij}(t)}{I_{ii}(0)} = \frac{I_{ji}(t)}{I_{jj}(0)} = \frac{-\sigma_{ij}}{\lambda_+ - \lambda_-}\left[e^{-\lambda_- t} - e^{-\lambda_+ t}\right] \tag{4}$$

with

$$\lambda_\pm = \frac{\left(\rho_i + \rho_j\right)}{2} \pm \sqrt{\left(\frac{\rho_i - \rho_j}{2}\right)^2 + \sigma_{ij}{}^2} \tag{5}$$

For practical purposes, we approximate the analytical solutions for the diagonal peak decays by single-exponential functions:

$$\frac{I_{ii}(t)}{I_{ii}(0)} = e^{-\rho_i t} \tag{6}$$

$$\frac{I_{jj}(t)}{I_{jj}(0)} = e^{-\rho_j t} \tag{7}$$

4.2.2. Extraction of the cross-relaxation rates

The general workflow for the extraction of the cross-relaxation rate is similar to the previously reported method [22], namely first fitting the autorelaxation rate decays of both diagonal peaks, second correcting the NOE buildup for spin diffusion and third fitting the normalized and spin diffusion corrected NOE buildup.

In both the three-spin system (TSS) and the full relaxation matrix (FRM) approaches magnetically or chemically equivalent spins are treated as a single spin with spin multiplicity factors in the according equations. The equivalent spins are designated $\vec{i}$, embracing spins i_1, … , i_N, and the factor $N_{\vec{i}}$ is the N-fold multiplicity. Implemented examples are methylene groups ($N_{\vec{i}}$ = 2), methyl groups ($N_{\vec{i}}$ = 3), degenerate aromatic groups ($N_{\vec{i}}$ = 4), and degenerate methyl groups in valine and leucine ($N_{\vec{i}}$ = 6). Thus, the solutions given in equations 4-7 are generalized to the following approximate forms:

$$\frac{I_{\vec{i}\vec{j}}(t)}{I_{\vec{i}\vec{i}}(0)} = \frac{-N_{\vec{j}}\sigma_{\vec{i}\vec{j}}}{\lambda_+^{\vec{i}\vec{j}} - \lambda_-^{\vec{i}\vec{j}}}\left[e^{-\lambda_-^{\vec{i}\vec{j}} t} - e^{-\lambda_+^{\vec{i}\vec{j}} t}\right] \tag{8.1}$$

$$\frac{I_{\vec{j}\vec{i}}(t)}{I_{\vec{j}\vec{j}}(0)} = \frac{-N_{\vec{i}}\sigma_{\vec{j}\vec{i}}}{\lambda_+^{\vec{j}\vec{i}} - \lambda_-^{\vec{j}\vec{i}}}\left[e^{-\lambda_-^{\vec{j}\vec{i}} t} - e^{-\lambda_+^{\vec{j}\vec{i}} t}\right] \tag{8.2}$$

with

$$\lambda_{\pm}^{\mu\nu} = \frac{\left(\rho_{\mu} + \rho_{\nu}\right)}{2} \pm \sqrt{\left(\frac{\rho_{\mu} - \rho_{\nu}}{2}\right)^2 + N_{\mu}^{fast} N_{\nu}^{fast} \sigma_{\mu\nu}{}^2} \qquad (9)$$

Here, N^{fast}_{m} and N^{fast}_{n} (m, n = i, j) are the fast-exchanging subgroups of the chemically equivalent groups. The values are 3 for methyl groups and degenerate pairs of methyl groups in valine and leucine, and 1 for all other cases, respectively. Note that if the magnetization starts on a group of equivalent spins, the normalized intensities are those expected for a single spin (although we still modify $\lambda_{\pm}^{\mu\nu}$).

We have shown that, for practical reasons, normalization to the diagonal peak intensity of the spin of the magnetization destination yields similar cross-relaxation rates [49]. Therefore, eNORA2 offers an optional alternative to equations 8.1 and 8.2:

$$\frac{I_{\bar{i}\bar{j}}(t)}{I_{\bar{j}\bar{j}}(0)} \approx \frac{-N_{\bar{i}}\sigma_{\bar{i}\bar{j}}}{\lambda_{+}^{\bar{i}\bar{j}} - \lambda_{-}^{\bar{i}\bar{j}}}\left[e^{-\lambda_{-}^{\bar{i}\bar{j}}t} - e^{-\lambda_{+}^{\bar{i}\bar{j}}t}\right] \qquad (8.3)$$

$$\frac{I_{\bar{j}\bar{i}}(t)}{I_{\bar{i}\bar{i}}(0)} \approx \frac{-N_{\bar{j}}\sigma_{\bar{j}\bar{i}}}{\lambda_{+}^{\bar{j}\bar{i}} - \lambda_{-}^{\bar{j}\bar{i}}}\left[e^{-\lambda_{-}^{\bar{j}\bar{i}}t} - e^{-\lambda_{+}^{\bar{j}\bar{i}}t}\right] \qquad (8.4)$$

Although even the decay of diagonal peaks of isolated groups of magnetically equivalent spins is multi-exponential, the analytical solutions for the diagonal peak decays are still approximated by mono-exponential functions:

$$\frac{I_{\bar{i}\bar{i}}(t)}{I_{\bar{i}\bar{i}}(0)} = e^{-\rho_{\bar{i}}t} \qquad (10.1)$$

$$\frac{I_{\bar{j}\bar{j}}(t)}{I_{\bar{j}\bar{j}}(0)} = e^{-\rho_{\bar{j}}t} \qquad (10.2)$$

The cross-relaxation rate constants are now extracted in three distinct steps. First, the intensities at zero mixing time, $I_{\bar{i}\bar{i}}(0)$ and $I_{\bar{j}\bar{j}}(0)$, and the autorelaxation rate constants, $\rho_{\bar{i}}$ and $\rho_{\bar{j}}$, are fit to equations 10.1 and 10.2.

The impact of relayed magnetization transfer via additional neighboring spins in a real spin system (spin diffusion) cannot be neglected. Thus, in a second step, we estimate the relative contributions from spin diffusion to the spectral intensities from a previously known approximate structure. The contributions can be expressed as correction factors for the apparent NOESY peaks intensities at each mixing time [22]:

$$I_{\bar{\mu}\bar{\nu}}(t) = p_{\bar{i}\bar{j}}(t) I^{\mathrm{app}}_{\bar{\mu}\bar{\nu}}(t) \qquad (11)$$

In a third step, the cross-relaxation rate constants, $\sigma_{\bar{i}\bar{j}}$ and $\sigma_{\bar{j}\bar{i}}$, are obtained by fitting equations 8.1 and 8.2 (or alternatively equations 8.3 and 8.4) with the corrected intensities while keeping $I_{\bar{i}\bar{i}}(0)$ and $\rho_{\bar{i}}$, and $I_{\bar{j}\bar{j}}(0)$ and $\rho_{\bar{j}}$, respectively, fixed.

The resulting distances may be used to calculate a new structure with which the quantification of spin diffusion would be improved (step 2) and more accurate cross-relaxation rate constants can be obtained from a new fit (step 3).

4.2.3. Estimation of the spin-diffusion correction factor

We have developed two different approaches to determine *p*, both of which require PDB coordinates of a previously known structure (such as a conventionally determined NMR structure or an X-ray structure). In the FRM approach, the buildup intensities containing spin diffusion, I_{ij}^{app}, are directly calculated with equation 2 using the full relaxation matrix and compared to the intensities of the two-spin buildup following equations 8.1 and 8.2 to derive corrections, $p_{ij}(t)$, for each mixing time [22]. In the TSS approach, we follow a strategy in which individual correction contributions from each neighboring spin *k* are obtained from the exact solutions of three-spin systems *ijk* and are then summed up [15,16]. Importantly, this method can also be applied to partially deuterated proteins.

4.2.3.1. Correction for spin diffusion with the TSS approach

In the TSS approach, instead of calculating the simulated intensities via the full relaxation matrix approach, individual corrections for each neighboring spin *k* obtained from the exact (numerical) solution of three-spin systems *mnk* are summed up (mn = *ij* or *ji*). Three-spin build-up curves are simulated using equation 2 for all cases where proton *k* is located within spheres of a user-defined radius centered at *m* and *n* (default value 6.5 Å), using cross-relaxation rates predicted from the distances in a structure, and autorelaxation rates taken from experiment (when available) or also predicted.

Analog to the FRM approach, these numerically simulated intensities are then compared with those simulated for the two-spin systems. The simulated three-spin system intensities mimic the values that one would observe in an experiment (if *m*, *n* and *k* were the only spins present), and are referred to as the apparent intensities $I_{\mathrm{mn}}^{\mathrm{app}(k)}(t)$ at time *t*. The contribution from each spin *k* to the spin diffusion is the difference between

the apparent and the exact theoretical two-spin intensities, $\left(I_{\overline{\mu\nu}}^{\mathrm{app(k)}}(t) - I_{\overline{\mu\nu}}(t)\right)$. The *overall* apparent intensity is the sum of the exact two-spin intensity and all the three-spin contributions involving *k* spins. The correction factor is thus:

$$p_{\overline{\mu\nu}}(t) = \frac{I_{\overline{\mu\nu}}(t)}{I_{\overline{\mu\nu}}(t) + \sum_k (I_{\overline{\mu\nu}}^{\mathrm{app}(k)}(t) - I_{\overline{\mu\nu}}(t))} \tag{12}$$

4.2.3.2. Spin diffusion in deuterated molecules

In the TSS approach, the impact of the deuteration levels is accounted for by scaling the contribution to spin diffusion from spin *k* by its protonation level ζ_k.

$$p_{\overline{\mu\nu}}(t) = \frac{I_{\overline{\mu\nu}}(t)}{I_{\overline{\mu\nu}}(t) + \sum_k \zeta_k (I_{\overline{\mu\nu}}^{\mathrm{app}(k)}(t) - I_{\overline{\mu\nu}}(t))} \tag{13}$$

4.2.3.3. Corrections for stereo spin pairs

Stereocenters such as methylene or methyl groups are treated differently depending on their assignment status. If the stereo assignment of such a group is known, the user can supply this information and the program will compute the spin diffusion correction factor *p* individually. On the other hand, if the assignment is unknown, eNORA2 will compute all possibilities of stereo assignments and average the correction factors subsequently before converting the cross-relaxation rates into distances.

4.2.4. CYANA input file

The extracted cross-relaxation rates can be converted into upper and lower distance limit restraints for CYANA structure calculations [23, 24]. The user defines a treshold χ for the quality of fits of the intensities to the experimental ones at the *N* mixing times [19, 22].

$$\chi = \frac{1}{\max(\mathrm{abs}(I_{\mathrm{exp}}))} \sqrt{\frac{\sum_{i=1}^{N} \left[I_{\mathrm{exp}}(\tau_{\mathrm{mix}}(i)) - I_{\mathrm{fit}}(\tau_{\mathrm{mix}}(i)) \right]^2}{N-1}} \tag{14}$$

Those cross-relaxation rates s_{ij} or s_{ji} that are obtained with χ lower than the treshold are converted into upper and lower limits following [3]:

$$r_{ij}^{\text{eff}} = \left(\left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma^4 h^2}{10} \frac{\tau_c}{\sigma_{ij}} \right)^{1/6} = \left(56.94 \frac{\tau_c / \text{ns}}{\sigma_{ij} / \text{s}^{-1}} \right)^{1/6} \text{ Å} \qquad (15)$$

The eNORA2 output lists cross-relaxation rates as if it were between two single protons even for chemically equivalent groups, such that the corresponding distance has a physical meaning. However, if all protons in a chemically equivalent group are used rather than a pseudo atom [50], CYANA interprets distances derived from NOEs involving magnetically or chemically equivalent spins as effective distances corresponding to the sum of all cross-relaxation rates between the individual spin pairs. Therefore, the distances are shortened as [20, 51]:

$$r_{\bar{i}\bar{j}}^{\text{CYANA}} = \frac{r_{\bar{i}\bar{j}}}{\left(N_{\bar{i}} N_{\bar{j}} \right)^{1/6}} \qquad (16)$$

For example, for a methyl group the distance is reduced by a factor $3^{-1/6} = 0.833$. These factors are used in Table 1.

If the distance restraint is obtained from one rather than from both cross peaks, there is an additional error due to the asymmetry of the NOESY pathways [49]. eNORA2 adds an additional user-defined tolerance to such restraints in the CYANA input files (see Table 1). We recommend 20% [49].

4.2.4.1. Corrections for chemically equivalent spins

Although equations 8 and 9 take into account chemically equivalent spins, appearing as superpositions of signals from multiple spins, these equations are approximations. eNORA2 adds corrections to the extracted distances. Per methyl group, upper and lower limits obtain an additional 8.5% tolerance to account for fast internal motions of methyl groups [20, 51].

Independent buildups of chemically equivalent groups of non-exchanging spins have an additional tolerance of 5%. This is the case, for example, for a buildup involving a methylene group $\vec{j}$, where the fit function for $\sigma_{i\bar{j}}$ is in first order a sum of the individuals, $s_{ij(1)} + s_{ij(2)}$, but $\sigma_{\bar{j}i}$ is in first order a weighted sum, $(I_{j(1)}s_{ij(1)} + I_{j(2)}s_{ij(2)})/(I_{j(1)} + I_{j(2)})$.

The complete set of spin types to which these corrections apply is listed in Table 1.

Table 1. Factors applied to the distances used for CYANA input files

condition	correction lol	correction upl
per methyl $\vec{i}$ or $\vec{j}$	0.833 x 0.915	0.833 x 1.085
per pseudo di-methyl $\vec{i}$ or $\vec{j}$	0.742 x 0.915 x 0.95	0.742 x 1.085 x 1.05
per methylene $\vec{i}$ or $\vec{j}$	0.89 x 0.95	0.89 x 1.05
per aromatic (2-fold degernerate) $\vec{i}$ or $\vec{j}$	0.891	0.891
per aromatic (4-fold degernerate) $\vec{i}$ or $\vec{j}$	0.794	0.794
bi-directional NOE	1	1
uni-directional NOE	1 – X[a] (default 0.8)	1 + X[a] (default 1.2)

[a] user-defined, recommended 20 % [49]

4.2.4.2. Generic normalized eNOEs

eNORA2 has an option to convert buildups into upper distance limits that do not violate the true distances without using the corresponding diagonal peaks [52]. We have recently shown that this is a valuable feature when large molecules are analyzed that typically suffer from severe overlap of the NOESY diagonal peaks [52]. Analysis of the $I_{\vec{i}\vec{i}}(0)$ intensities of the resolved diagonal peaks allows the user to set save upper limits for single protons, methyl groups and chemically equivalent spins such as those of methylene groups and aromatics. eNORA2 then converts them into CYANA upper distance limits.

4.3. Results and discussion

A flow chart presenting the architecture of eNORA2 is shown in Figure 1.

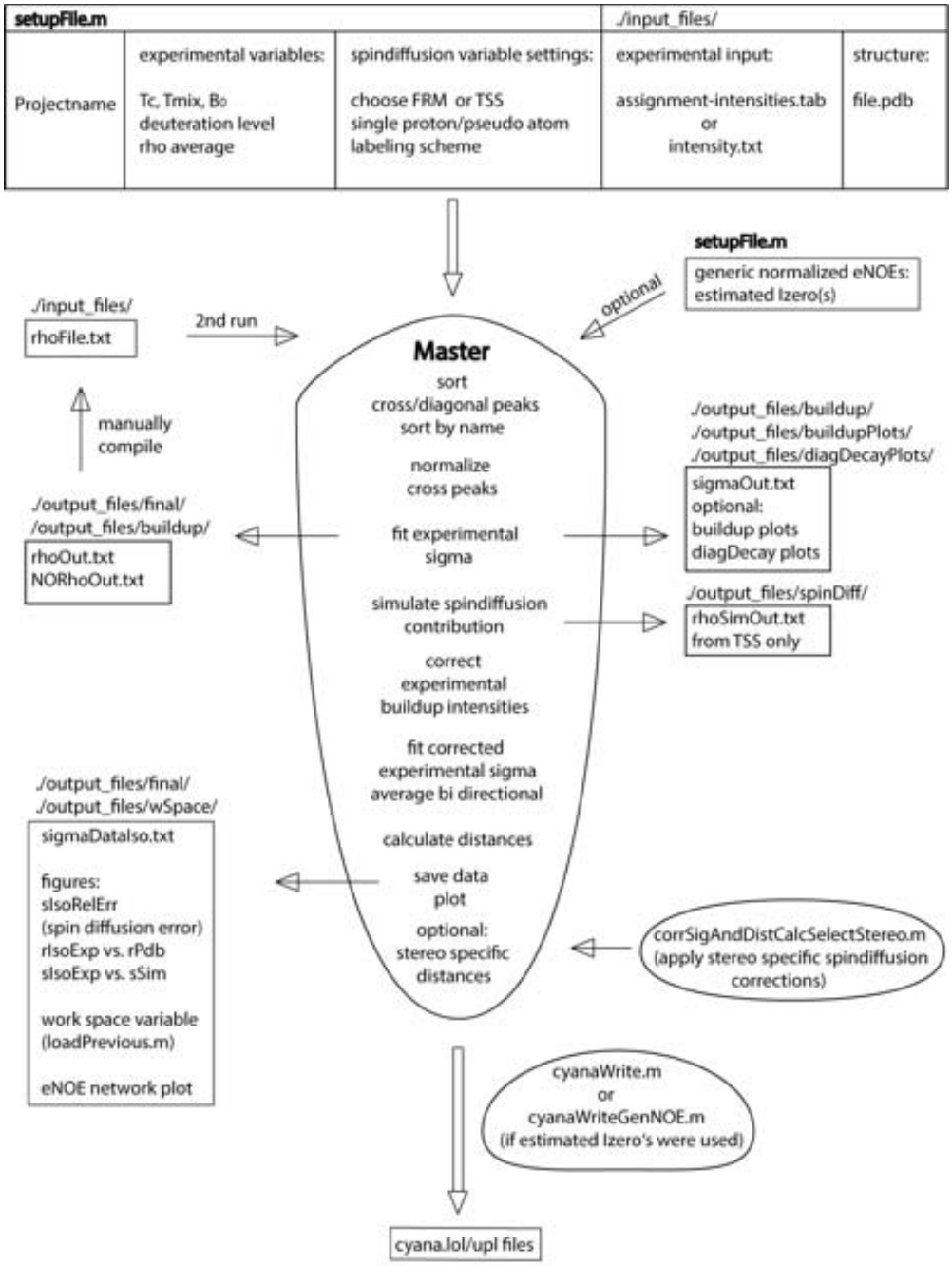


Figure 1. Flow chart of eNORA2. Extensions 'm' indicate Matlab scripts. Subfolders are shown with '/' starting from the 'home' folder.

4.3.1. Setup

There are a number of user variables to be set before eNORA2 is initiated. They are labeled with user.variableName. Strings are given with single quotation marks, such as user.projectname = 'PROJECTNAME' and multiple numbers are given in vector form such as user.mixing_times = [0.02 0.03 0.04 0.05]. All input data such as pdb structures, eNOE and autorelaxation rate ('rho') data is to be placed into the input_files directory. All output is found in the output_files/PROJECTNAME directory. There is the possibility to save the current status of the project and to continue the analysis later. A standard setup script to run eNORA2 is shown in Figure 2.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%    THIS IS THE SETUP FILE    %%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear
clc

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%      USER SPECIFIC INPUT     %%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%      MODIFY THESE VARIABLES TO FIT PROJECT      %%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Projectname
user.projectname                = 'standardProject';

% Specify pdb file name naming convention to use (markley, cyana, oldcyana):

user.pdbFileName                = 'startingStructure.pdb';
user.pdbNameConvention          = 'cyana.lib';


%%%%%%%%%%%%%%          nmrDraw.tab file or intensity file          %%%%%%%%%%%%%%

% The number of dim of your NOESY if you read a nmrDrawTabFile (2/3) dimensions:
user.specDimens                 = 3;

user.nmrDrawTabFileName     = 'assignedAndFittedPeaks.tab';
%% user.intensityFileName          = 'fileName';

user.assNameConvention          = 'cyana.lib';

user.magOriginSpin              = 1;
user.magNormSpin                = 2;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%                EXPERIMENTAL INPUT                %%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Specify the mixing times used (seconds):
user.mixing_times               = [0.02 0.03 0.04 0.05 0.06];

% Percent D2O in sample for deuterium locking:
user.perCentD2O                 = 3;

% The overall correlation time (in units of ns):
user.tauC                       = 4.25;
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Set to 'n' if not provided
user.buildupRhoFileName      = 'n';

% Best guess of the autorelaxation rate by user
user.rhoAverage              =  5.3;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% user.spinDiffApp           ='threeSpin';
user.spinDiffApp             ='fullMatrix';

% Enter the polarizing field strength (in units of MHz):
user.B0field                 = 700;

% If stereo specific assignments are not all known, set to 1, otherwise 0 if the stereo specific assignments are
known for all spins.
user.stereoAvg               = 1;

% If some stereo specific assignments are known, compile file and set
user.corrSigStereo           = 'n';

user.upperlimit              = 6.5;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

addpath('./eNOEprogram/setupAndMaster/');
master %Located in the master folder, executes the routine's elements

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Figure 2. Standard setup file. Double commented (%%) commands are optional commands.

4.3.2. Essential user variables

4.3.2.1. Project name

The project name is typically set to a string identifying the data used. It is the name the project will be given trailing the time stamp, i.e. user.projectname = 'PROJECTNAME' will be applied to create a directory such as /output_files/160322_15:33_PROJECTNAME.

4.3.2.2. Structure for spin diffusion calculation

The spin diffusion calculation requires a structure coordinate file. Currently, a pdb file with just one model and with protons and pseudo atoms is required. Usually the lowest energy model is used. In order to read the structure file the naming convention of the protons has to be given. The three implemented options are the conventions proposed by reference [54] (markley.lib), and the current and an old CYANA convention (cyana.lib and oldcyana.lib). If protons are missing, i.e. in X-ray crystallography derived structures, computer programs like CYANA can build the protons and the pseudo atoms into the structures.

4.3.2.3. Reading the buildup data

For buildup input files, there are two options. Either a NMRDRAW file or a generic intensity file described below may be used. Peak assignment of an NMRDRAW.tab file needs to conform to the following assignment standard:

res.proton;res.proton;res.heavyatom,

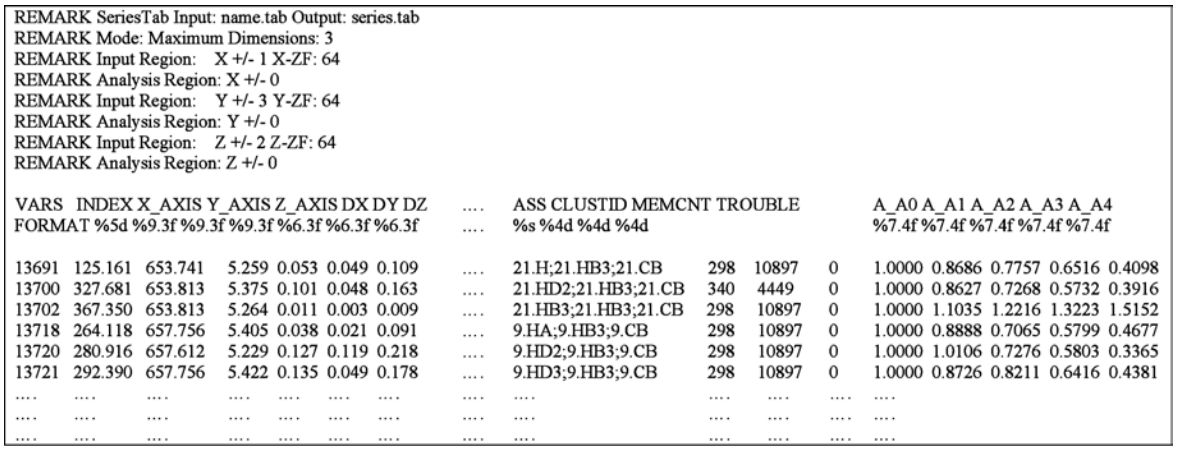
i.e.

16.H;22.HA;22.CA

(stored in sigmaData.input.ASS) for a three dimensional NOESY experiment and

16.H;22.HA

for a two dimensional NOESY experiment. Once an NMRDRAW.tab file containing the complete assignment has been generated, use of the fitting tool of nmrPipe (series3D.com for 3D NOESY and fit.com for 2D NOESY) adds the relative intensity variables A0, … , AN for N+1 mixing times) as shown in Figure 3 [55]. The setting is user.nmrDrawTabFileName = 'NMRDRAWFILE.tab' and user.assNameConvention = ' namingConvention.lib' according to the assignment standard used (see above; the naming convention of the structure file, 'pdbName.pdb', and the buildup input file, 'NMRDRAWFILE.tab', are not required to be the same).



```
REMARK SeriesTab Input: name.tab Output: series.tab
REMARK Mode: Maximum Dimensions: 3
REMARK Input Region:   X +/- 1 X-ZF: 64
REMARK Analysis Region: X +/- 0
REMARK Input Region:   Y +/- 3 Y-ZF: 64
REMARK Analysis Region: Y +/- 0
REMARK Input Region:   Z +/- 2 Z-ZF: 64
REMARK Analysis Region: Z +/- 0

VARS   INDEX X_AXIS Y_AXIS Z_AXIS DX DY DZ     ....  ASS CLUSTID MEMCNT TROUBLE          A_A0 A_A1 A_A2 A_A3 A_A4
FORMAT %5d %9.3f %9.3f %9.3f %6.3f %6.3f %6.3f ....  %s %4d %4d %4d                    %7.4f %7.4f %7.4f %7.4f %7.4f

13691  125.161  653.741  5.259 0.053 0.049 0.109  ....  21.H;21.HB3;21.CB      298  10897  0  1.0000 0.8686 0.7757 0.6516 0.4098
13700  327.681  653.813  5.375 0.101 0.048 0.163  ....  21.HD2;21.HB3;21.CB    340  4449   0  1.0000 0.8627 0.7268 0.5732 0.3916
13702  367.350  653.813  5.264 0.011 0.003 0.009  ....  21.HB3;21.HB3;21.CB    298  10897  0  1.0000 1.1035 1.2216 1.3223 1.5152
13718  264.118  657.756  5.405 0.038 0.021 0.091  ....  9.HA;9.HB3;9.CB        298  10897  0  1.0000 0.8888 0.7065 0.5799 0.4677
13720  280.916  657.612  5.229 0.127 0.119 0.218  ....  9.HD2;9.HB3;9.CB       298  10897  0  1.0000 1.0106 0.7276 0.5803 0.3365
13721  292.390  657.756  5.422 0.135 0.049 0.178  ....  9.HD3;9.HB3;9.CB       298  10897  0  1.0000 0.8726 0.8211 0.6416 0.4381
....   ....     ....     ....  ....  ....  ....   ....  ....                   ....  ....   ....  ....
....   ....     ....     ....  ....  ....  ....   ....  ....                   ....  ....   ....  ....
....   ....     ....     ....  ....  ....  ....   ....  ....                   ....  ....   ....  ....
```

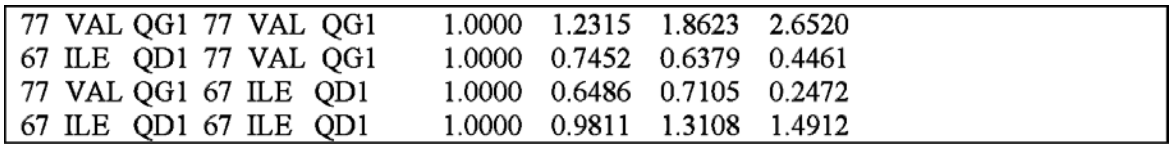
Figure 3. Example of the nmrDraw tab file entry format (user.nmrDrawTabFileName).

The second NOESY data input format allowed in eNORA by setting user.intensityFileName = 'intensity_file' and user.assNameConvention = 'namingConvention.lib ' is in the form of an intensity file and needs to conform to the following assignment standard (Figure 4):

res aminoacid proton res aminoacid proton relativeIntensityList

i.e.

13 LYS H 16 SER HA 0.2 0.4 0.6 0.8 1



```
77 VAL QG1 77 VAL QG1   1.0000  1.2315  1.8623  2.6520
67 ILE QD1 77 VAL QG1   1.0000  0.7452  0.6379  0.4461
77 VAL QG1 67 ILE QD1   1.0000  0.6486  0.7105  0.2472
67 ILE QD1 67 ILE QD1   1.0000  0.9811  1.3108  1.4912
```

Figure 4. Example of intensity file entry format (user.intensityFileName). Columns 1 to 6 are the cross and diagonal peak assignments and 7 to10 the measured intensities at four different mixing times (in this specific case scaled to 1.0, which is not required).

If the measurement for a given mixing time is erroneous the user can remove that measurement point for all buildups and decays of diagonals simultaneously without editing the input file by setting user.runThrough = 0. This user setting temporarily interrupts the routine following an initial fitting of the buildup intensities and allows the user to edit the mixing time. Accepting the changes will then refit the buildups with the chosen time points and run the routine until completion.

4.3.2.4. Flow of the magnetization and normalization

The assignment of the cross peaks in a NOESY experiment may be done such that the first entry indicates from which spin the magnetization originates and the destination of magnetization is the second entry (res.originSpin;res.destinationSpin) or vice versa (res.destinationSpin;res.originSpin). In a three dimensional experiment the last entry is

then to be assigned to heavy atom such as res.originSpin;res.destinationSpin;res.heavyAtom or res.destinationSpin;res.originSpin;res.heavyAtom. If the first entry is the origin of magnetization the setting is user.magOriginSpin = 1, otherwise user.magOriginSpin = 2. The diagonal peak to be used for normalization may be the one of the origin or destination of the spin magnetization flow and is chosen by setting user.magNormSpin = 1 or 2.

4.3.2.5. User supplied information for fitting of experimental data

In order to fit the buildups the mixing times are given in units of seconds (i.e. user.mixing_times = [0.02 0.03 0.04 0.05]).

Sample specific information requires the overall correlation time (i.e. user.tauC = 4.25) given in units of nanoseconds and the percent D_2O added to the sample for deuterium locking (i.e. user.perCentD2O = 3).

Ideally, all autorelaxation rates are obtained from fitting the diagonal peak decays. Upon setting user.buildupRhoFileName = 'n', these autorelaxation rates will be written into ‘RhoOut.txt’. However, in practice many diagonal peak decays cannot be fitted (due to spectral overlap, artifacts, etc.). Therefore, a general average autorelaxation rate for missing values may also be set by the user in units of s^{-1} (user.rhoAverage) in order to increase the number of eNOEs calculated. In practical cases, the influence of a value off by 1-2 s^{-1} is maximally 10% on the cross-relaxation rate [49].

Following the initial run, the estimation of the missing autorelaxation rates may be improved by using averages of the fitted experimental values (in the initial run) obtained for the same spin types written into the ‘NORhoOutNumCode.txt’ file and provided by user.buildupRhoFileName (Figure 5) in later runs. These average autorelaxation rates may be calculated using the script calcAverageRhoConst.m found in the eNOEprogram/handTools directory or may be chosen manually by analyzing the values in the ‘RhoOut.txt’ file following the intial run of eNORA found in the buildup directory of the output_files/PROJECTNAME/buildup directory.

6	5	1.238	1.82e+08	0.0549
6	30	2.087	1.12e+09	0.0173
6	2011	2.673	8.96e+08	0.0203
6	2015	1.665	1.82e+09	0.0163
6	6	3.3556	999	999
9	30	3.3556	999	999
10	4	3.3556	999	999
12	18	3.3556	999	999
13	2	2.3873	999	999
13	29	3.3556	999	999

Figure 5. Example entries in an initial intensity/autorelaxation rate ('rho') input file (user.buildupRhoFileName). The first column is the residue number, the second column the proton atom type given in numeric code, the third column is the autorelaxation rate in s^{-1}, the fourth the $I_{ii}(0)$ value calculated and the fifth column is the χ calculated for the fit value. The value 999 in the $I_{ii}(0)$ column and χ column specify rho values calculated from experimental averages, but not fitted from diagonal peak decay.

4.3.2.6. Spin diffusion method

The spin diffusion method used by default is set to full matrix approach (user.spinDiffApp = 'fullMatrix'). To apply the three-spin approach, the setting is changed to 'threeSpin'. For deuterated samples, the three-spin approach is necessary [22]. In the basic setting, the deuteration levels of HN (given by the deuteration of D_2O in the solution), HA, methylene and all other protons are chosen separately with the variables user.perCentD2O, user.deut_HA, user.deut_Met and user.deut_Prot.

In addition, the user can also provide a methyl-group specific labeling scheme for methyl-bearing amino acids specified with the variable user.labelingScheme = 'RES_MethGroup%0'. The available user input vocabulary covers all currently known selective labeling schemes [56] and includes ALA_B, ILE_G2, ILE_D1, LEU_D1, LEU_D2, LEU_D1_D2, MET_E, THR_G2, VAL_G1, VAL_G2, VAL_G1_G2. As an example, a protein selectively labeled as ILV would therefore be set as user.labelingScheme = 'VAL_G1_G2%0 LEU_D1_D2%0 ILE_D1%0', if the deuteration level for the selectively labeled residues is 0%.

For the spin diffusion calculation, the upper limit for the distance of spins pairs to be included in the pathways is set to 6.5 Å. This limit can be changed with user.upperlimit.

If stereo specific information is not known for all spins (user.stereoAvg = 1 and user.corrSigStereo = 'n') the individual values of the calculated spin diffusion correction factor are an average of the stereo partners. Partially known stereospecific assignment information may be applied at the end of the run by running the calcStereoSel.m script which generates an output file 'sigmaDataOutIsoStereoSpec.txt', before generating CYANA restraint files (see below). Note that such an assignment may be obtained from the eNOEs after an initial CYANA calculation [57].

Spin diffusion via methyl groups may be calculated by representing the group by a single pseudo atom (user.pseudoMethyl = 1) or by considering the three pathways given by the three individual protons (user.pseudoMethyl = 0) [50].

4.3.2.7. Experimental cross-relaxation rate constants and distance restraints

Running the eNORA2 master file generates the output file sigmaDataOutIso.txt in the output_files/TIMESTAMP_PROJECTNAME/final folder (Figure 6). This file contains the cross-relaxation rate constants $\sigma_{\bar{ij}}$ and $\sigma_{\bar{ji}}$, and the internuclear distances $r_{\bar{ij}}$ and $r_{\bar{ji}}$, their averages if both pathways are used, along with other parameters such as the factor by which that rate is scaled as compared to the value obtained from a simple fit with the two-spin model, $p_{ij}(t)$, and the quality factors of the buildup fits, χ. Some figures are generated by default, the most important ones of which show the eNOE network superimposed on the input PDB structure, spin-diffusion correction factors and comparisons of the experimental cross-relaxation rates and distances to those predicted from the input structure (see Figure 7).

RESULTS ARE DERIVED FROM ISOTROPIC CALCULATIONS

ASSi	ATOMi	ASSj	ATOMj	AvgSigma	AvgrEff	...	SigmaExpCorrij	SigmaExpCorrji	rEffij	rEffji	Pcorrij	Pcorrji	SigmaPredij	SigmaPredji	rPdbij	rPdbji	...
6	HA	6	H	NA	NA	...	-2.0956	NA	2.31	NA	0.96	NA	-0.6641	NA	2.80	NA	...
8	HG2	37	HG2	-0.0638	4.14	...	-0.0519	-0.0889	4.29	3.92	0.41	0.42	-2.1634	-2.1634	2.86	2.86	...
11	HA	12	H	-2.8284	2.20	...	-1.2804	-6.2480	2.51	1.93	0.98	0.98	-2.6055	-2.6055	2.23	2.23	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 6. Example entries in a sigmaDataOutIso.txt file.

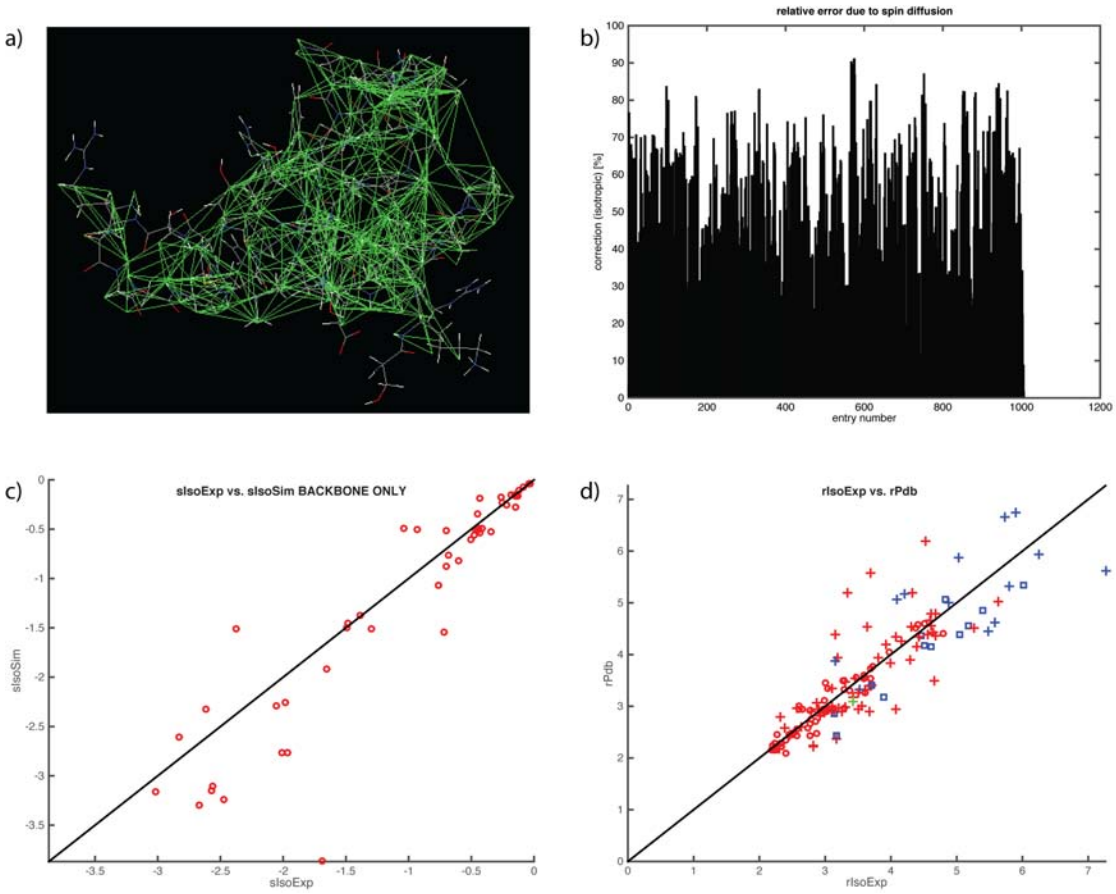


Figure 7. Output figures generated by the Master file. Top left, spin pairs giving rise to analyzed eNOEs are indicated by lines drawn on the input PDB structure; top right, correction factors *p* in a bar plot; bottom left, cross-relaxation rates simulated from the input structure versus experimental ones; bottom right, distances extracted from the input structure versus experimental effective distances.

Once the master file is run, the user has the option to create upper and lower limit distance restraint files in CYANA format (upl and lol files, see Figure 8). The cyanaWrite.m command generates these files by applying all the corrections listed in Table 1.

6 LYS HA	6 LYS H	2.77	# rEff= 2.31	uni-biDir = 1	Ass-Norm = 6H
7 LEU H	7 LEU HA	2.65	# rEff= 2.65	uni-biDir = 2	Ass-Norm = 7HA
13 LYS HB2	14 ARG H	5.27	# rEff= 4.39	uni-biDir = 1	Ass-Norm = 14H
14 ARG HA	15 MET H	2.45	# rEff= 2.45	uni-biDir = 2	Ass-Norm = 15H
…	…	…	…	…	…
…	…	…	…	…	…

Figure 8. Example of entries in a typical CYANA lol/upl file.

Generic limits for normalization may be provided for non-normalized cross peaks. The generic *I*(0) values for each group are set through user.IzeroMethyl, user.IzeroSingle, and user.IzeroMethylenePseudo in the setup file. These values are determined as described in reference [52]. The generic upper limit distance restraints are then written into the upl file in CYANA format.

4.3.2.8. Non-standard amino acids and other molecules

Non-standard amino residues or molecules other than proteins can be added to the numcode.lib file in the eNOEprogram/read_in folder in order to allow the program to recognize them. Simultaneously, an additional nomenclature convention file has to be created and placed in the eNOEprogram/lib folder (see eNORA2 manual for more details).

4.3.2.9. Plotting buildup curves and diagonal decays

By default auto-plotting of all buildups and diagonal decays is disabled. One can enable this option by setting user.autoPlot = 0. Plots in png format will be saved into the buildupPlots and diagDecayPlots folders.

4.4. Conclusions

We have introduced a Matlab-based program for facile evaluation of NOESY buildups. eNORA2 can be downloaded from our webpage (www.bionmr.ethz.ch). We provide an extensive user manual and example files. Finally, we note that eNORA2 is an open source program such that further methodological developments can be accounted for. Indeed, optimal extraction of cross-relaxation rates from NOESY spectra and the subsequent interpretation of those rates is an active field of research [58-64].

4.5. References

[1] Kaptein R, Boelens R, Scheek RM, van Gunsteren WF, Biochemistry 1988, 27, 5389-5395.
[2] Wüthrich K, Science 1989, 243, 45-50.
[3] Vögeli B, Prog Nucl Mag Reson Spectrosc 2014, 78, 1-46.
[4] Bell RA, Saunders JK, Can J Chem 1970, 48, 1114-1122.
[5] Schirmer RE, Noggle JH, Davis JP, Hart PA, J Am Chem Soc 1970, 92, 3266-3273.
[6] Boelens R, Koning TMG, Kaptein R, J Mol Struct 1988, 173, 299-311.
[7] James TL, Curr Opin Struct Biol 1991, 1, 1042-1053.
[8] Kessler H, Griesinger C, Lautz J, Müller A, van Gunsteren WF, Berendsen HJC, J Am Chem Soc 1988, 110, 3393-3396.
[9] Bonvin AMJJ, Boelens R, Kaptein R, J Biomol NMR 1994, 4, 143-149.
[10] Bonvin AMJJ, Brünger AT, J Mol Biol 1995, 250, 80-93.
[11] Bonvin AMJJ, Brünger AT, J. Biomol. NMR 1996, 7, 72-76.
[12] Bürgi R, Pitera J, van Gunsteren WF, J. Biomol. NMR 2001, 19, 305-320.
[13] Zagrovic B, van Gunsteren WF, Proteins 2006, 63, 210-218.
[14] Richter B, Gsponer J, Varnai P, Salvatella X, Vendruscolo M, J Biomol NMR 2007, 37, 117-135.
[15] Vögeli B, Segawa T, Leitz D, Sobol A, Choutko A, Trzesniak D, van Gunsteren WF, Riek R, J Am Chem Soc 2009, 131, 17215-17225.
[16] Vögeli B, Friedmann M, Leitz D, Sobol A, Riek R, J Magn Reson 2010, 204, 290-302.
[17] Vögeli B, Orts J, Strotz D, Chi C, Minges M, Wälti MA, Güntert P, Riek R, J Magn Reson 2014, 241, 53-59.
[18] Vögeli B, Kazemi S, Güntert P, Riek R, Nat Struct Mol Biol 2012, 19, 1053-1057.
[19] Vögeli B, Güntert P, Riek R, Mol Phys 2013, 111, 437-454.
[20] Vögeli B, Olsson S, Riek R, Güntert P, J Struct Biol 2015, 191, 306-317.
[21] Vögeli B, Olsson S, Güntert P, Riek R, 2016, Biophys J 2016, 110, 113-126.
[22] Orts J, Vögeli B, Riek R, J Chem. Theory Comput 2012, 8, 3483-3492.
[23] Güntert P, Prog Nucl Magn Reson Spectrosc 2003, 43, 105-125.
[24] Güntert P, Eur Biophys J 2009, 38, 129-143.
[25] Boelens R, Koning TMG, Kaptein R, J Mol Biol 1988, 173, 299-311.
[26] Boelens R, Koning TMG, Van der Marel GA, Van Boom JH, Kaptein R, J Magn Reson 1989, 82, 290-308.
[27] Bonvin AMJJ, Rullmann JAC, Lamerichs RMJN, Boelens R, Kaptain R, Proteins Struct Func Genet 1993, 15, 385-400.
[28] Summers MF, South TL, Kim B, Hare DR, Biochemistry 1990, 29, 329-340.
[29] Borgias BA, James TL, J Magn Reson 1990, 87, 475-487.
[30] Post CB, Meadows RP, Gorenstein DG, J Am Chem Soc 1990, 112, 6796-6803.
[31] Nikonowicz EP, Meadows RP, Gorenstein DG, Biochemistry 1990, 29, 4193-4204.
[32] Edmondson SP, J Magn Reson 1992, 98, 283-298.
[33] Madrid M, Llinas E, Llinas M, J Magn Reson 1991, 93, 329-346.
[34] Lai XN, Chen CP, Andersen NH, J Magn Reson B 1993, 101, 271-288.
[35] Constantine KL, Mueller L, Andersen NH, Tong H, Wandler CF, Friederichs MS, Bruccoleri RE, J Am Chem Soc 1995, 117, 10841-10854.
[36] Yip P, Case DA, J Magn Reson 1989, 83, 643-648.
[37] Bonvin AMJJ, Boelens R, Kaptain R, J Biomol NMR 1994, 4, 143-149.

[38] Borgias BA, Gochin M, Kerwood DJ, James TL, Prog Nucl Magn Reson Spectrosc 1990, 22, 83-100.
[39] Mertz JE, Güntert P, Wüthrich K, Braun W, J Biomol NMR 1991, 1, 257-269.
[40] Nilges M, Habbazettl J, Brunger A, Holak TA, J Mol Biol 1991, 219, 499-510.
[41] Schwieters CD, Kuszewski JJ, Clore GM, Prog Nucl Magn Reson Spectrosc 2006, 48, 47-62.
[42] Yip PF, J Biomol NMR 1993, 3, 361-365.
[43] Linge JP, Habeck M, Rieping W, Nilges M, J Magn Reson 2004, 167, 334-342.
[44] Lane A, J Magn Reson 1988, 78, 425-439.
[45] Borgias BA, James TL, J Magn Reson 1988, 79, 493-512.
[46] Leeflang BR, Kroon-Batenburg LMJ, J Biomol NMR 1992, 2, 495-518.
[47] MATLAB and Statistics Toolbox Release 2014b, The MathWorks Inc, Natick, Massachusetts, United States.
[48] Solomon I, Phys Rev 1955, 993, 559-565
[49] Strotz D, Orts J, Minges M, Vögeli B, J Magn Reson 2015, 259, 32-46.
[50] Fletcher CM, Jones DNM, Diamond R, Neuhaus D, J Biomol NMR 1996, 8, 292-310.
[51] Vögeli B, Olsson S, Riek R, Güntert P, Data in Brief 2015, 5, 99-106.
[52] Chi C, Strotz D, Riek R, Vögeli B, J Biomol NMR 2015, 62, 63-69.
[53] Chi C, Vögeli B, Bibow S, Strotz D, Güntert P, Riek R, Angew Chem Int Ed Engl 2015, 54, 11657-11661.
[54] Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K, J Mol Biol 1998, 280, 933-952.
[55] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, J Biomol NMR 1995, 6, 277-293.
[56] Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J, Curr Opin Struct Biol 2015, 32, 113-122.
[57] Orts J, Vögeli B, Riek R, Güntert P, J Biomol NMR 2013, 57, 211-218.
[58] Efimov SV, Khodov IA, Ratkova EL, Kiselev MG, Berger S, Klochkov VV, J Mol Struct 2016, 1104, 63-69.
[59] Kolmer A, Edwards LJ, Kuprov I, Thiele CM, J Magn Reson 2015, 261, 101-109.
[60] Chalmers G, Glushka JN, Foley BL, Woods RJ, Prestegard JH, J Magn Reson 2016, 265, 1-9.
[61] Dass R, Kasprzak P, Koźmiński W, Kazimierczuk K, J Magn Reson 2016, 265, 108-116.
[62] Baishyaa B, Vermaa A, J Magn Reson 2016, 252, 41-48.
[63] Kuprov I, Hodgson DM, Kloesges J, Pearson CI, Odell B, Claridge TDW, Angew Chem Int Ed Engl 2015, 54, 3697-3701.
[64] Mantsyzov AB, Maltsev AS, Ying J, Shen Y, Hummer G, Bax A, Protein Sci 2014, 23, 12765-1290.

5. Replica exchange simulations

This chapter is based on the following publication:

Simon Olsson, **Dean Strotz**, Beat Vögeli, Roland Riek, Andrea Cavalli, The dynamic basis for signal propagation in human Pin1-WW, Structure 2016, doi: 10.1016/j.str.2016.06.013. Simon Olsson did the programming for the simulations and its analysis. Dean Strotz recorded all experimental data, did the analysis of the experimental data, prepared figures concerning the experimental data and participated in the writing of the manuscript. The publication 'The dynamic basis for signal propagation in human Pin1-WW' was split into two parts, this is part one, pertaining to method development.

5.1 Introduction

With the advent of special purpose supercomputers [1] simulations combined [2-10] with high-resolution NMR data [11] have become increasingly important. In combination with sophisticated statistical models of molecular kinetics [12-14] a complete view of dynamics at the atomistic scale is within reach. Here, we present a new approach to generate representative structure ensembles of proteins. In this approach, we combine replica-exchange simulations with chemical shift data through maximum entropy reweighing. Our resulting ensembles correlate well with high-accuracy NMR measurements. In particular, the recently introduced exact nuclear Overhauser enhancement (eNOE) enables us to validate our ensembles.

5.2. Results

5.2.1. Temperature-dependent structure ensembles of Pin1-WW

To understand the conformational properties of Pin1-WW, we determined the free-energy landscapes of our model system, the double mutant (S18N/W34F, introduced in chapter 2.3) and the wild-type (WT) at 15 temperatures ranging from 278 K to 348 K, well above and below the melting temperatures of the constructs. We first generated canonical ensembles (CEs) at each of the 15 temperatures. This was achieved by performing replica-exchange molecular dynamics (MD) simulations of the two constructs with a cumulative simulation time of 15 μs per construct. Next we validated our simulations with special emphasis on the S18N/W34F construct. First we compared

back-predicted and experimental chemical shift data from backbone and side-chain resonances at all the corresponding temperatures. We found remarkably good agreement; nearly all of the CEs agreed with the experimental chemical shift data within the random-prediction error (Figures S1A–S1E in appendix 10.2.). To ensure we had ensemble models that were in complete agreement with experimental data, we segmented the phase space into 20 discrete states using K-means clustering (5.4. Experimental Procedures). These clusters were used as a basis for reweighing the ensembles at each temperature using the backbone chemical shift data through the maximum entropy principle (appendix 10.2.). This yielded the reweighed ensembles (REs). Apart from improving the agreement with the backbone chemical shift data used to perform the reweighing, we also observe an improvement in the agreement with side-chain chemical shift data not used during the reweighing (Figure S1 in appendix 10.2.). This result is consistent regardless of whether a Gaussian or a flat-bottom error model was used during the reweighing (5.4. Experimental Procedures).

For validation of the REs, we acquired additional experimental data at two select temperatures (278 K and 303 K), corresponding to the lowest and highest temperatures at which we could measure these at high resolution (Figure S2 in appendix 10.2.). Specifically, we recorded eNOEs [15], cross-correlated relaxation rates (CCRs) [16], and backbone and side-chain scalar couplings (^{3}J), providing a wide range of averaged translational and radial information (5.4. Experimental Procedures). Again, we found remarkably good agreement between our ensembles and the experimental data (Figure S3 and S4 in appendix 10.2.). Importantly, we found that the agreement with these data was either better or on par in the REs compared with the CEs. In general, the REs using the flat-bottom error model displayed the best overall quality, and we thus used this in all subsequent analyses.

5.2.2. Free-energy landscape analysis

Having thoroughly validated CEs of the doubly mutated Pin1-WW domain at hand, we performed free-energy landscape analysis. We generated the free-energy landscapes by projecting the ensembles onto the first two principal components (5.4. Experimental Procedures). For reference, we show two of the free-energy landscapes at 278 K and 303 K along with the cluster centers (Figures 1A and 1B). Subsequently, we quantified temperature-dependent population changes in the cluster probabilities defined above (5.4. Experimental Procedures, Figure 1E). The cluster probability reflects the chance of being in a certain segment of configurational space. We identified two clusters with a

significant fraction of the probability mass at 278 K, a native cluster (> 90%) and a near-native cluster (>5%). Both of these clusters have the well-known three-strand anti-parallel β sheet native fold of the WW domain. However, there are three important differences. A ring flip in Trp11 allows for a topological rearrangement of the N- and C-termini in the near-native cluster (Figures 1D and S5 in appendix 10.2.). This difference is consistently correlated with a transition in the preferential backbone dihedral angles in the binding loop (Figures 1C and 2). Finally, the two clusters show distinct patterns of linearly correlated motions (Figure 3). We performed an analysis of previously deposited structures of human Pin1-WW in the PDB and identified that near-native topology appeared in a significant number of the analyzed structures (see appendix 10.2. and Figure S6 in appendix 10.2.). The 18 other clusters (U) correspond to different partially unfolded or completely unfolded states.

Following the populations of native and near-native clusters and the sum of unfolded clusters as a function of temperature, we see that the native cluster is monotonously declining and the unfolded clusters are monotonously increasing in population. Interestingly, the near-native cluster shows a distinct maximum between 303 K and 310 K, close to physiological temperature (Figure 1E). The temperature at which the unfolded conformations constitute 50% of the population coincides closely with the experimentally determined melting temperature (326 K).

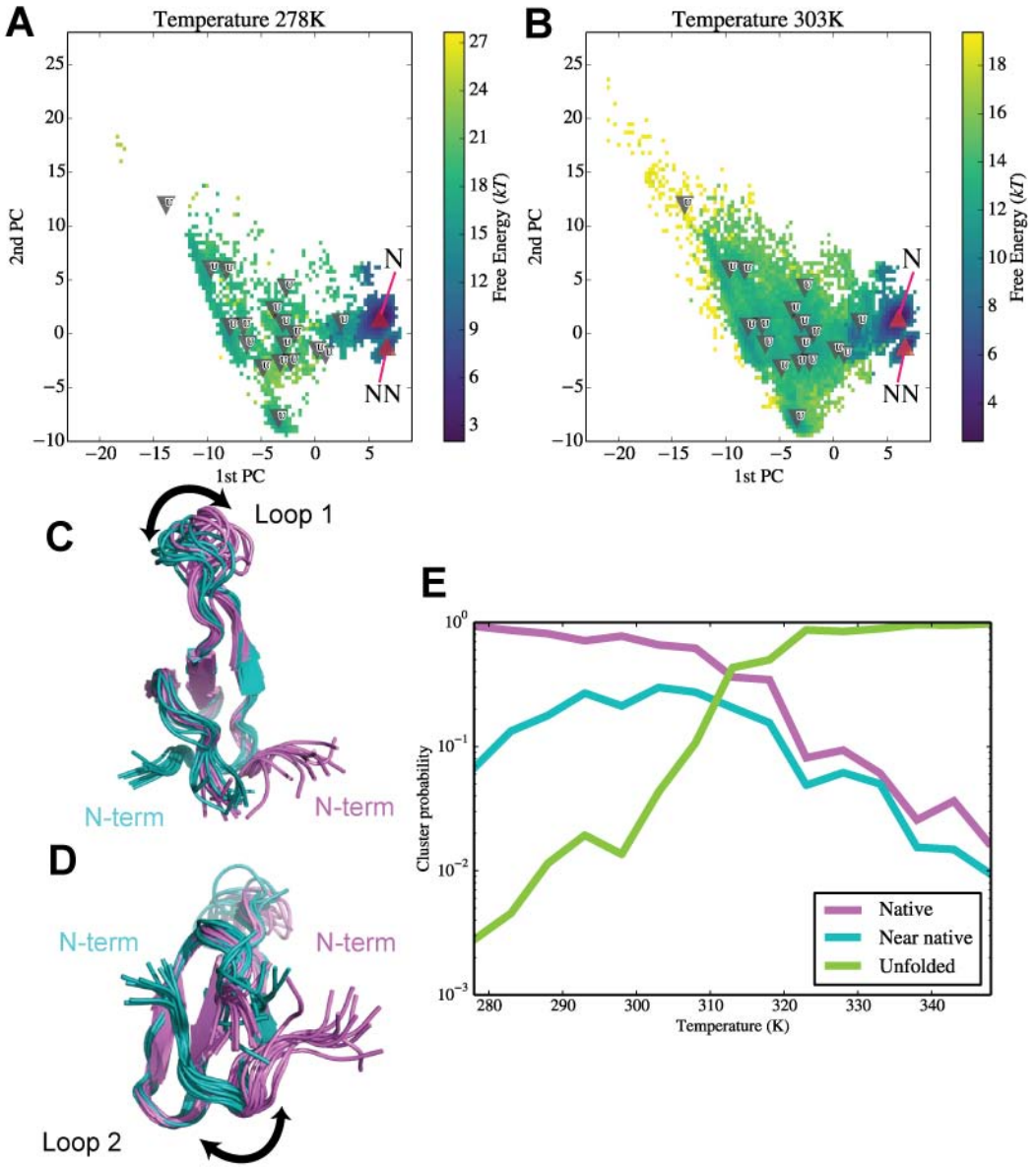


Figure 1. Summary of generated ensembles at 278 K and 303 K. Free-energy landscapes at (A) 278 K and (B) 303 K and cluster populations as a function of temperature in the REs (E). Cluster centroids are shown with upward pointing red triangles for clusters that have a higher population in the RE compared with the CE, and a downward pointing triangle for the clusters having a lower population. The native and near-native clusters are red triangles annotated with N and NN, respectively. The unfolded clusters are black triangles annotated with U. Renderings of conformational changes in loop 1 (residues 17–20) (C) and the topological reorientation of the N- and C-termini (D) are shown for ten random conformations from each cluster. The native cluster is purple and the near-native cluster is teal.

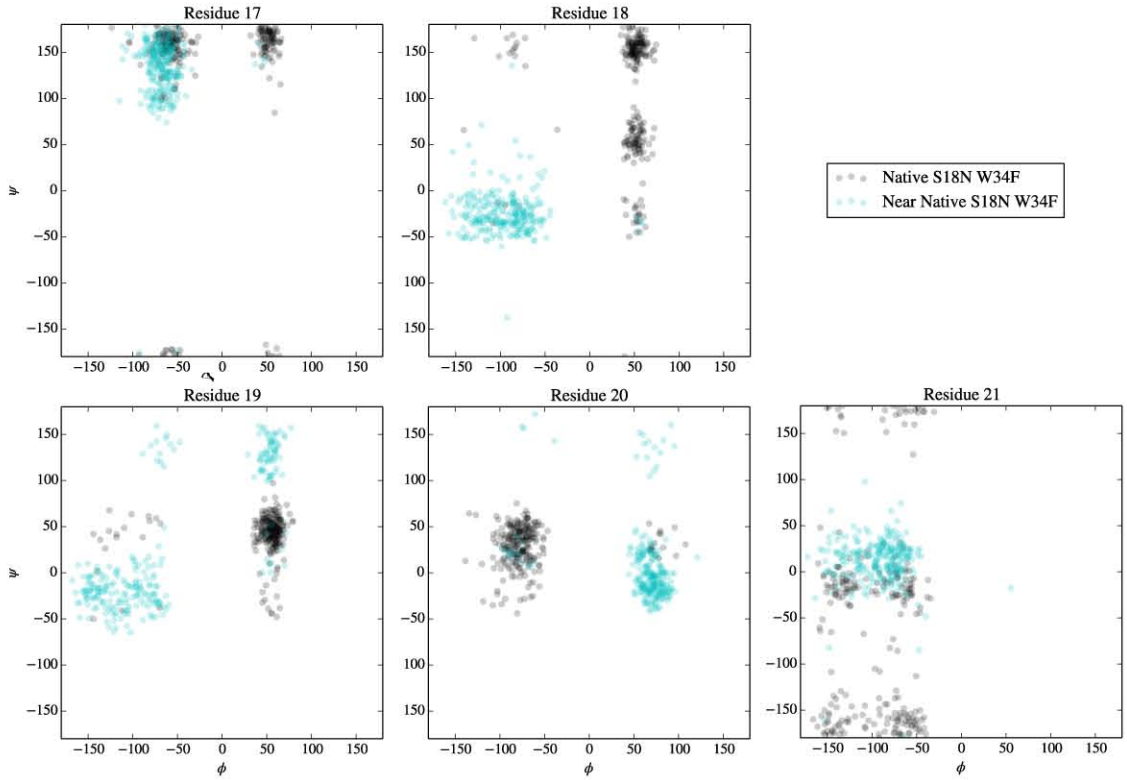


Figure 2. Φ, Ψ torsion angle distributions of loop 1 in the near-native and native states scatter points of Φ, Ψ torsion pairs in native (black) and near-native (cyan) states identified by cluster analysis.

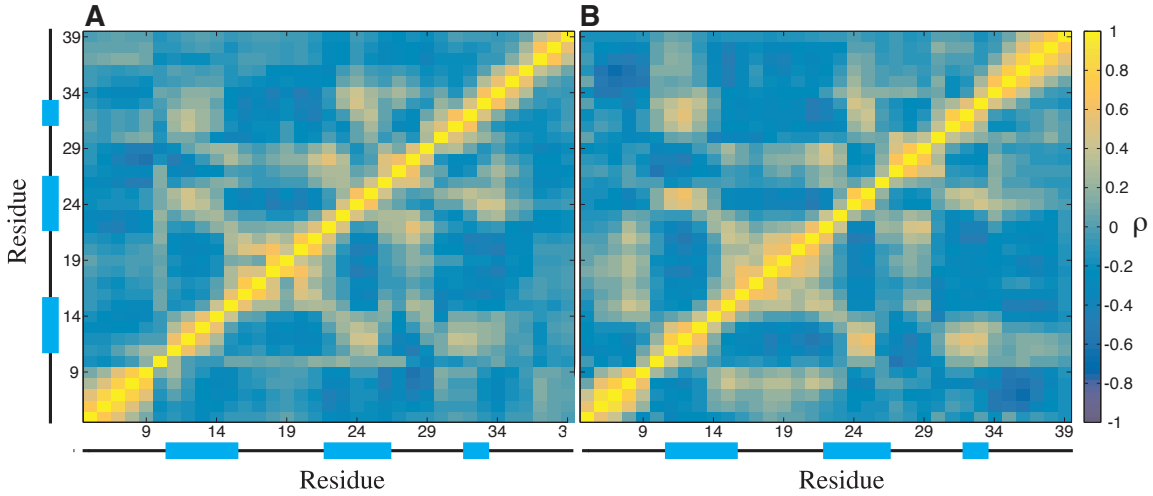


Figure 3. Linearly correlated motions in Pin1-WW, S18N/W34F native (A) and near-native (B) conformational states at 278 K identified using cluster analysis. Correlation maps were computed using THESEUS [17]. Secondary structure is shown on residue axes: black lines are loop, turn, and coil regions, and blue blocks are β strands.

5.2.3. eNOE data is sensitive to sparsely populated near-native state

At 278 K, the near-native cluster is only sparsely populated (~5%–10%). Due to this, it is not obvious that any of the ensemble-averaged experimental observables would be sensitive to the presence of this state, due to experimental uncertainties and because the two states are quite similar. To test this, we computed the residue-wise absolute restraint violations of eNOEs back computed from CEs, the REs, and native and near-native clusters independently. The eNOE data were separated into bidirectional (Figures 4A and 4B) and unidirectional classes (Figures 4C and 4D). The former is the more precise probe as it is based upon two NOE cross peaks, whereas the latter is only derived from either one [15]. We computed overall average violations of bidirectional eNOEs at 278 K for the different ensembles and states. We found that the REs ($\bar{\varepsilon}_{\text{bidir;278K}} = 0.207$ s^{-1}) provided a comparable agreement compared with the CEs ($\bar{\varepsilon}_{\text{bidir;278K}} = 0.205$ s^{-1}) but better than either of the individual states alone (native, 0.219 s^{-1}; near native, 0.243 s^{-1}) (Figure 4). Interestingly, the native cluster alone seems to describe the entire ensemble well apart from residues around loops 1 and 2 as well as the very C-terminal residues as judged from bidirectional eNOE data (Figure 4A). On the other hand, the near-native state overall correlates less well with the data but provides a superior agreement in these particular regions compared with the native cluster alone. In addition, the near-native state has a much better agreement at Phe34 and the two adjacent residues. However, as the probability of the near-native state is small compared with the native state, it only influences the CEs and REs only modestly, but systematically. A similar pattern was observed at 303 K, strongly suggesting that the bidirectional eNOE is sufficiently sensitive to detect the presence of sparsely populated states. Finally, we note how the RE and CE models both represent loop 2 better at 278 K than either of the near-native or native states do independently. In the unidirectional eNOEs, we see a similar pattern but not as strong: at 278 K residues in loop 1 are better described by the near-native state while the native state agrees well in the rest of the structure. Such sensitivity to the near-native state was not observed for either of the other high-resolution experimental data types used in the validation.

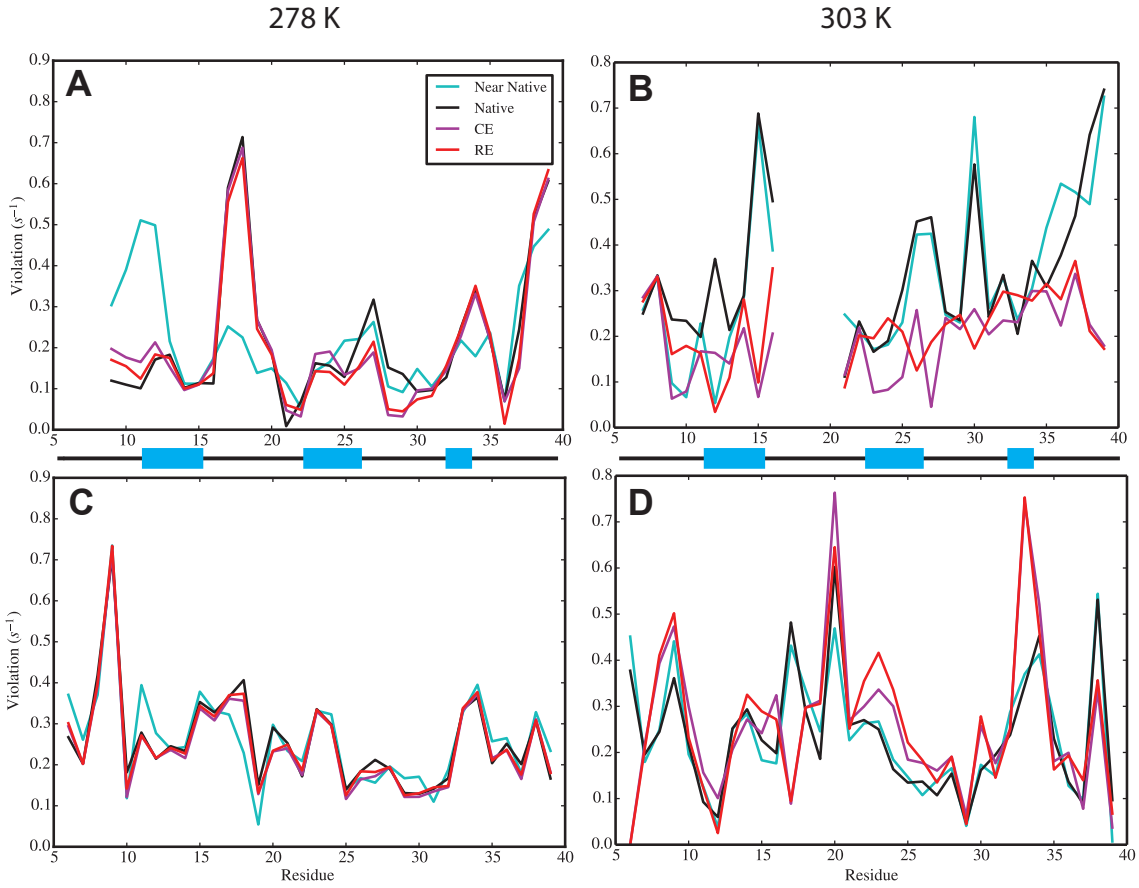


Figure 4. Residue-wise eNOE violations of different models of Pin1-WW, S18N/W34F violations (root-mean-square deviation [RMSD] from the experimental value) of bidirectional (A and B) and unidirectional (C and D) eNOEs at 278 K and 303 K, respectively. Comparison of violations in canonical and reweighed ensembles (CE and RE, respectively) and the two major conformational clusters identified (native and near native). CE and RE include all clusters (native, near native, and unfolded). Secondary structure is shown on the residue axes: black lines are loop, turn, and coil regions, and blue blocks are β strands. RMSD violations of 0.1–0.2 s^{-1} translate into distance errors of approximately 0.3–0.4 Å.

5.2.4. Native and near-native state exchange correlates with a microsecond-millisecond timescale process

Peng et al. (2007) [18] detected the presence of a microsecond-millisecond exchange process in Arg17 of loop 1 at 278 K by backbone ^{15}N NMR relaxation dispersion in Pin1-WW WT. Assuming a slowly exchanging two-state model, they predicted an amide ^{15}N chemical shift difference between the two states of 2.5 ± 0.2 ppm. Strikingly, in our S18N/W34F simulation, we find the back-computed average chemical shift difference between the native and near-native conformational clusters correlate well with this (2.2 ppm; Figure 5). We also observe large ^{15}N chemical shift differences in the residues Trp11 and Arg21 for which no previous data have been reported (Figure 5). Performing a chemical shift variance analysis on our WT simulation [19], we found a very similar pattern; Trp11 and Arg17 again took the largest values along with Ser38 (Figure 5).

Performing this analysis using the S18N/W34F simulation shows similar results (5.4. Experimental Procedures and Figure S7 in appendix 10.2.).

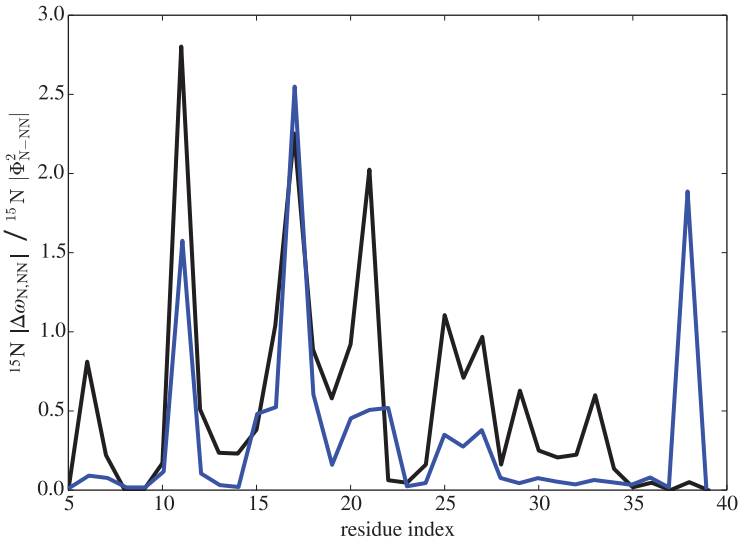


Figure 5. Chemical shift variability in WT and S18N/W34F simulations of Pin1-WW. Comparison ^{15}N chemical shift differences ($|\Delta_{\omega N,NN}|$ in ppm) back computed using the S18N/W34F simulation (black) and chemical shift variances ($|\Psi_{N\text{-}NN}|$ in ppm2) back computed using WT simulation (blue) at 278 K. Experimental value for Arg17, 2.5 ± 0.2 ppm [18]. Chemical shifts back computed using CamShift [20, 21].

5.3. Discussion

5.3.1. Analysis of equilibrium thermal unfolding data using replica-exchange molecular dynamics simulations

Following biophysical quantities as a function of temperature is a classical approach in protein folding studies. Here we present a way in which one may integrate such data with replica-exchange MD simulations and thereby obtain atomistic ensembles of proteins as a function of temperature. Using this method, we detect multiple conformational states in a double mutant of Pin1-WW (S18N/W34F), which are corroborated by complementary experimental data. While the approach readily allows for comparison with many equilibrium experiments, it does not, at this stage, allow for dissection of chemical exchange kinetics. However, a future extension of this may be possible by complementing this type of simulations with regular MD simulations in an appropriate statistical framework [22-24].

5.3.2. The potential of eNOE data to resolve excited-state populations

We have previously shown that the improved precision of eNOE data allows for the determination of multi-conformer bundles of proteins [11, 15, 26]. The current study suggests that the application of eNOEs in structural analysis can likely go beyond the determination of multi-conformer bundles. Specifically, we generated ensembles using a combination of replica-exchange MD simulations and maximum entropy-based chemical shift reweighing, without considering the eNOE data, and detected two states with considerable probability. However, we observed that the eNOE data are highly sensitive to the presence of both of these states, which may be surprising considering one of them appears to have a low population (p ~5%–10%). Thus, we show how the eNOE experiments may directly complement other experiments classically associated with detection of transient states, such as paramagnetic relaxation enhancement (PRE) [27] and relaxation dispersion. Particularly, the eNOEs are sensitive to shorter distances than what is typically amenable for PRE analysis, and this distance information may be complementary to information derived from relaxation dispersion.

5.4. Experimental procedures

5.4.1. Constant bias replica-exchange molecular dynamics simulations

WT Pin1-WW coordinates were extracted from a previously reported crystal structure (PDB: 1PIN) [28]. The initial coordinates of the double mutant (S18N/W34F) were constructed by manual mutagenesis in PyMOL (DeLano Scientific, LCC) of the WT coordinates. The initial coordinates of the WT and S18N/W34F constructs were solvated in 3236 and 2881 TIP3P [29] water molecules, respectively. The systems were neutralized by the addition of four chloride atoms and energy minimized in the force field, ff99sb-ildn [30] followed by 500 ps of equilibration in the NPT ensemble at 300 K and 1 bar. An unfolding simulation was carried out for 2 ns and at 600 K for the S18N/W34F construct; no cis/trans isomerizations were observed. Subsequently, 15 structures were sampled linearly from the unfolding trajectory and used as starting conformations in each of the 15 temperature replicas. Each replica was equilibrated at the target temperature for 500 ps. To ensure reasonable acceptance rates in the replica-exchange scheme [31], free-energy histograms were estimated for each of the replicas using 1 ns of well-tempered metadynamics (NVT ensemble) with the potential energy as the reaction coordinate [32]. Bias factors were chosen to follow the geometric distribution. In the production simulation, each histogram was used as a constant bias in

its corresponding temperature replica, yielding an average acceptance rate of 53.6%. All NVT simulations were coupled to a Bussi thermostat [33] and NPT simulations to a Berendsen barostat [34]. All simulations employed p-LINCS constraints on bonds involving protons and 2 fs time steps [35]. Coulomb interactions were cut off at 9.5 Å and long-range electrostatics were treated by particle mesh Ewald [36]. For the WT simulations, an identical protocol was used but the unfolding trajectory was 10 ns and the average acceptance rate was 32.5%. Each replica was run for 1 μs to a cumulative simulation time of 2 x 15 μs. All simulations were carried out using Gromacs [37] linked to PLUMED2 [21, 38]. All subsequent analyses were facilitated by the MDTraj python library [39].

5.4.2. Back-prediction of NMR parameters from structural ensembles

Below the averaged quantity $\langle Q \rangle$ is given by the expectation

$$\langle Q \rangle = \int dx Q(x)\, p(x) \approx \overline{w(\ell(x)) Q(x)}\,, \tag{1}$$

where $Q(x)$ is a function that computes a quantity (such as a distance or angle) from a structure x with probability p(x), and w(•) is a function that returns the normalized weight of a cluster ℓ , for example in the CE or the RE. $\bar{r}$ denotes the sample mean of a random variable *r*.

eNOE cross-relaxation rates were back predicted using the expression assuming slow conformational exchange and isotropic tumbling,

$$\sigma_{ij} = \frac{\mu_0^2}{4\pi} \gamma_H^4 \hbar^2 \langle r_{ij}^{-6} \rangle \left[6J(2\omega_H) - J(0) \right] , \tag{2}$$

where σ_{ij} is one extracted cross-relaxation rate for unidirectional eNOEs and the signed geometrical average in case of a bidirectional eNOE ($\mathrm{sgn}\sigma_{ij}\sqrt{\sigma_{ij}\sigma_{ji}}$). μ_0 is the permeability of free space, γ_H and ω_H are the gyromagnetic ratio and Larmor frequencies of the proton, respectively, and $\hbar$ is the reduced Planck's constant. The spectral density is $J(w) = \frac{2}{5} \frac{\tau_C}{1 + \omega^2 \tau_C^2}$, where τ_c is the experimentally determined autocorrelation time.

CCRs were back predicted assuming isotropic tumbling as [40]

$$R_{H-N,H\alpha-C\alpha}^{meas} = R_{H-N,H\alpha-C\alpha} + R_{H-C\alpha,H\alpha-N} \tag{3}$$

with

$$R_{X-Y,U-V} = \gamma_X \gamma_Y \gamma_U \gamma_V \left\langle r_{U-V}^{-3} \right\rangle \left\langle r_{X-Y}^{-3} \right\rangle \left(\frac{\hbar \mu_0}{4\pi} \right)^2 \frac{1}{5} \left\langle 3\cos^2\theta - 1 \right\rangle \tau_c \quad (4)$$

where θ is the angle between the two spin-pair vectors X-Y and U-V of length $r_{U\text{-}V}$ and $r_{X\text{-}Y}$, respectively.

Side-chain and backbone ^{3}J-coupling data were back computed using the ensemble-averaged Karplus equation [41],

$$^3J_\theta = A\left\langle \cos^2(\theta) + \right\rangle + B\left\langle \cos^2(\theta) + \right\rangle + C \quad (5)$$

where θ is the dihedral angle defined by the four nuclei giving rise to the scalar coupling, $^3J_\theta$. Previously reported Karplus parameters (A, B, and C) were used for side-chain [42] and backbone angles [43], respectively.

Methyl and aromatic chemical shifts were back-predicted using CH3Shift [44] and ArShift [45] through Almost [21].

Comparison with ^{15}N relaxation dispersion experiments was carried out in two different ways for the S18N/W34F simulation. First, absolute chemical shift differences between the native and near-native clusters was computed, ^{15}N $|\Delta_{\omega N,NN}|$. Second, we performed a calculation where we assumed the chemical exchange between the two states was Markovian with equal exchange rates, by computing the chemical shift variance for each resonance $^{15}N|_{\Phi N,NN}|$ [19]. Since no clear separation of the native and near-native clusters was present in the WT simulation we quantified the chemical exchange contribution to the relaxation by using only the latter of these approaches.

5.4.3. NMR spectroscopy

Details on sample preparation and nuclear Overhauser effect spectroscopy (NOESY), τ_c and $^3J_{H,H\text{-}\alpha}$ scalar coupling experiments at 278 K have been described previously [46]. The data at 303 K were measured in a similar manner with the following differences: For NOESY experiments, diagonal-peak decays and cross-peak buildups were measured at mixing times τ_m = 40, 60, 80, 100, 120 ms. τ_c was determined to be 2.5 ns using ratios of cross-correlation rates determined from constant-time evolution transverse relaxation optimized spectroscopy (TROSY) and anti-TROSY relaxation [47]. In addition to these experiments, we acquired $^3J_{C',C_\gamma}$ and $^3J_{N,C_\gamma}$ [48] and CCR data [40, 49] at both temperatures. Stereo-specific assignments were determined as described previously [50]. Temperature-dependent chemical shift changes were followed by ^{13}C resolved constant-time heteronuclear single quantum coherence experiments with

gradient selection and ^{15}N-resolved 2D heteronuclear multiple quantum coherence experiments with selective ^{1}H pulses [51]. Measurements were performed from 278 K to 348 K at 5 K increments. The raw data were processed with NMRpipe [52] and assigned in CCPnmr Analysis [53]. We used a previously established protocol [46] for the determination of eNOE cross-relaxation using an updated version of the eNORA software package [54], to be published elsewhere. Spin diffusion effects were compensated for using the full relaxation matrix method [54] with the lowest energy conformer of a previously determined crystal structure of 1PIN [28]. Magnetically equivalent protons were treated by r^{-6} summation [55].

5.5. References

[1] Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE, Biomolecular simulation: a computational microscope for molecular biology, Annu Rev Biophys 2012, 41, 429–452.

[2] Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M, Simultaneous determination of protein structure and dynamics, Nature 2005, 433, 128–132;

[3] Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Combining experiments and simulations using the maximum entropy principle, PLoS Comput Biol 2014, 10, e1003406.

[4] Pitera JW, Chodera JD, On the use of experimental observations to bias simulated ensembles, J Chem Theory Comput 2012, 8, 3445.

[5] Roux B, Weare J, On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method, J Chem Phys 2013, 138, 084107.

[6] Cavalli A, Camilloni C, Vendruscolo M, Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle, J Chem Phys 2013, 138, 094112.

[7] Olsson S, Frellsen J, Boomsma W, Mardia KV, Hamelryck T, Inference of structure ensembles from sparse, averaged data, PLoS One 2013, 8, e79439.

[8] Olsson S, Vögeli B, Cavalli A, Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Hamelryck T, Probabilistic determination of native state ensembles of proteins, J Chem Theory Comput 2014, 10, 3484–3491.

[9] Olsson S, Ekonomiuk D, Sgrignani J, Cavalli A, Molecular dynamics of biomolecules through direct analysis of dipolar couplings, J Am Chem Soc 2015, 137, 6270–6278.

[10] Hummer G, Köfinger J, Bayesian ensemble refinement by replica simulations and reweighting, J Chem Phys 2015, 143, 243150.

[11] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19, 1053–1057.

[12] Swope WC, Pitera JW, Suits F, Describing protein folding kinetics by molecular dynamics simulations. 1. Theory, J Phys Chem B, 2004, 108, 6571–6581.

[13] Noé F, Krachtus D, Smith JC, Fischer S, Transition networks for the comprehensive characterization of complex conformational change in proteins, J Chem Theory Comput 2006, 2, 840–857.

[14] Buchete N-V, Hummer G, Coarse master equations for peptide folding dynamics, J Phys Chem B 2008, 112, 6057–6069.

[15] Vögeli B, The nuclear Overhauser effect from a quantitative perspective, Prog Nucl Magn Reson Spectrosc, 2014, 78, 1–46.

[16] Reif B, Hennig M, Griesinger C, Direct measurement of angles between bond vectors in high-resolution NMR, Science 1997, 276, 1230– 1233.

[17] Theobald DL, Wuttke DS, THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures, Bioinformatics 2006, 22, 2171–2172.

[18] Peng T, Zintsmaster JS, Namanja AT, Peng JW, Sequence- specific dynamics modulate recognition specificity in WW domains, Nat Struct Mol Biol 2007, 14, 325–331.

[19] Ban D, Funk M, Gulich R, Egger D, Sabo TM, Walter KFA, Fenwick RB, Giller K, Pichierri F, de Groot BL, et al., Kinetics of conformational sampling in ubiquitin, Angew Chem Int Ed Engl 2011, 50, 11437–11440.

[20] Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M, Fast and accurate predictions of protein NMR chemical shifts from interatomic distances, J Am Chem Soc 2009, 131, 13894–13895.
[21] Fu B, Sahakyan AB, Camilloni C, Tartaglia GG, Paci E, Caflisch A, Vendruscolo M, Cavalli A, Almost: an all atom molecular simulation toolkit for protein structure determination, J Comput Chem 2014, 35, 1101– 1105.
[22] Mey ASJS, Wu H, Noé F, xTRAM: estimating equilibrium expectations from time-correlated simulation data at multiple thermodynamic states, Phys Rev 2014, X 4, 041018.
[23] Wu H, Mey ASJS, Rosta E, Noé F, Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states, J Chem Phys 2014, 141, 214106.
[24] Wu H, Paul F, Wehmeyer C, Noé F, Multiensemble Markov models of molecular thermodynamics and kinetics, Proc Natl Acad Sci USA 2016, 113, E3221–E3230.
[25] Vögeli B, Olsson S, Riek R, Güntert P, Complementarity and congruence between exact NOEs and traditional NMR probes for spatial decoding of protein dynamics, J Struct Biol 2015, 191, 306–317.
[26] Vögeli B, Olsson S, Güntert P, Riek R, The exact NOE as an alternative in ensemble structure determination, Biophys J 2016, 110, 113–126.
[27] Iwahara J, Clore GM, Detecting transient intermediates in macromolecular binding by paramagnetic NMR, Nature 2006, 440, 1227–1230.
[28] Ranganathan R, Lu K, Hunter T, Noel J, Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent, Cell 1997, 89, 875–886.
[29] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, Comparison of simple potential functions for simulating liquid water, J Chem Phys 1983, 79, 926.
[30] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE, Improved side-chain torsion potentials for the Amber ff99sb protein force field, Proteins 2010, 78, 1950–1958.
[31] Sugita Y, Okamoto Y, Replica-exchange molecular dynamics method for protein folding, Chem Phys Lett 1999, 314, 141–151.
[32] Barducci A, Bussi G, Parrinello M, Well-tempered metadynamics: a smoothly converging and tunable free-energy method, Phys Rev Lett 2008, 100, 020603.
[33] Bussi G, Donadio D, Parrinello M, Canonical sampling through velocity rescaling, J Chem Phys 2007, 126, 014101.
[34] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR, Molecular dynamics with coupling to an external bath, J Chem Phys 1984, 81, 3684–3690.
[35] Hess B, P-LINCS: a parallel linear constraint solver for molecular simulation, J Chem Theory Comput 2008, 4, 116–122.
[36] Darden T, York D, Pedersen L, Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems, J Chem Phys 1993, 98, 10089.
[37] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, et al., Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 2013, 29, 845–854.
[38] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G, Plumed 2: new feathers for an old bird, Comput Phys Commun 2014, 185, 604–613.
[39] McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang L-P, Lane TJ, Pande VS, Mdtraj: a modern open

library for the analysis of molecular dynamics trajectories, Biophys J 2015, 109, 1528–1532.
[40] Vögeli B, Yao L, Correlated dynamics between protein HN and HC bonds observed by NMR cross relaxation, J Am Chem Soc 2009, 131, 3668–3678.
[41] Lindorff-Larsen K, Best RB, Vendruscolo M, Interpreting dynamically-averaged scalar couplings in proteins, J Biomol NMR 2005, 32, 273–280.
[42] Pérez C, Löhr F, Rüterjans H, Schmidt JM, Self-consistent Karplus parametrization of ^{3}J couplings depending on the polypeptide side-chain torsion chi1, J Am Chem Soc 2001, 123, 7081–7093.
[43] Vögeli B, Ying J, Grishaev A, Bax A, Limits on variations in protein backbone dynamics from precise measurements of scalar couplings, J Am Chem Soc 2007, 129, 9377–9385.
[44] Sahakyan AB, Vranken WF, Cavalli A,Vendruscolo M, Structure-based prediction of methyl chemical shifts in proteins, J Biomol NMR 2011, 50, 331–346.
[45] Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M, Using side-chain aromatic proton chemical shifts for a quantitative analysis of protein structures, Angew Chem Int Ed Engl 2011, 50, 9620–9623.
[46] Strotz D, Orts J, Minges M, Vögeli B, The experimental accuracy of the uni-directional exact NOE, J Magn Reson 2015, 259, 32–46.
[47] Pervushin K, Riek R, Wider G, Wüthrich K, Attenuated t2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, Proc Natl Acad Sci USA 1997, 94, 12366–12371.
[48] Hu J-S, Bax A, Determination of f and c-1 angles in proteins from ^{13}C-^{13}C three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond?, J Am Chem Soc 1997, 119, 6360– 6368.
[49] Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G, Simultaneous determination of j and f angles in proteins from measurements of cross-correlated relaxation effects, J Biomol NMR 1999, 14, 277–280.
[50] Orts J, Vögeli B, Riek R, Güntert P, Stereospecific assignments in proteins using exact NOEs, J Biomol NMR 2013, 57, 211–218.
[51] Cavanagh J, Fairbrother WJ, Palmer AG III, Rance M, Skelton NJ, Protein NMR Spectroscopy: Principles and Practice, Academic Press 2007.
[52] Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol NMR 1995, 6, 277–293.
[53] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley J.L, Ionides J, Laue ED, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 2005, 59, 687–696.
[54] Orts J, Vögeli B, Riek R, Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs, J Chem Theory Comput 2012, 8, 3483–3492.
[55] Nilges M, O'Donoghue SI, Ambiguous NOEs and automated NOE assignment, Prog Nucl Magn Reson Spectrosc 1998, 32, 107–139.

Part II.

Applications

6. The dynamic basis for signal propagation in human Pin1-WW

This chapter is based on the following publication:

Simon Olsson, **Dean Strotz**, Beat Vögeli, Roland Riek, Andrea Cavalli, The dynamic basis for signal propagation in human Pin1-WW, Structure 2016, doi: 10.1016/j.str.2016.06.013. Simon Olsson did the programming for the simulations and its analysis. Beat Vögeli advised and participated in writing of the manuscript. Dean Strotz recorded all experimental data, did the analysis of the experimental data, prepared figures concerning the experimental data and participated in the writing of the manuscript.
The publication 'The dynamic basis for signal propagation in human Pin1-WW' was split into two parts, this is part two, pertaining to biologically relevant research.

6.1 Introduction

The idea of allostery as the transmission of information through a molecule is a continuously developing topic within structural and molecular biology [1]. Atomistic descriptions of this fundamental biophysical mechanism are critical to our understanding of processes such as signaling and regulation within and between cells. Over the past decades, our view of allostery has undergone a transition from switching between discrete and static [2, 3] conformational states to a more refined thermodynamic picture that is characterized by skewing of populations in the conformational ensemble [4, 5]. Examples of the latter include highly flexible [6] and even disordered states [7]. Nuclear magnetic resonance (NMR) relaxation dispersion experiments have proven to be particularly powerful to study this aspect of allostery [8, 9]. Such experiments yield structural, thermodynamic, and kinetic information under favorable circumstances [10,11]. Furthermore, with the advent of sophisticated statistical models of molecular kinetics [12-14] and special purpose supercomputers [15], a complete view of allostery at the atomistic scale is within reach; in particular when combined [16-24] with high-resolution NMR data [25].

The mitotic regulator Pin1 is composed of an N-terminal binding domain (WW) flexibly tethered to a larger C-terminal peptidyl-prolyl cis-trans isomerase (PPIase)

domain (Figure 1) [26, 27]. Both domains contain binding sites that specifically recognize phosphorylated Ser/Thr-Pro motifs. Such motifs act as regulatory switches often found in intrinsically disordered regions [28]. However, a recent report indicates that association of unphosphorylated substrates to Pin1 may be possible through the formation of a fuzzy complex [29]. Understanding the molecular underpinnings of Pin1 functional action has been the subject of intense research. Pin1 has been observed to interact with a wide range of ligands involved in biological processes such as mitosis [26] and signaling, as well as medical conditions such as cancer and neurodegeneration [30]. It is now clear that substantial cross-talk between the two domains takes place by the means of loop 2 of the WW domain (residues 27–30) and residues 138 and 140–142 in the PPIase domain [31-35]. In particular, it has been shown that the structural and dynamic properties of the PPIase domain are sensitive to the presence of the N-terminal WW domain and/or ligand [31, 36, 37]. However, a lot of important questions about the functional mechanism of Pin1 remain unanswered. Do bound ligands migrate from the WW domain to the PPIase domain? If so, how? Many substrates of Pin1 contain multiple pSer/pThr-Pro motifs. This may suggest a different mechanism where the high-affinity binding site on the WW domain acts as an anchor, and catalysis is carried out by the PPIase elsewhere [34]. Some data suggest that ligand binding either enhances [36] or reduces [34, 35] the interdomain cross-talk in Pin1. Further, hypervariability of the binding site in the WW domain family (loop 1, residues 17–20 in Pin1-WW) has been shown to modulate binding specificity [38]. Still, Pin1 binds promiscuously [28], which may suggest that loop 1 in Pin1-WW is particularly malleable [39] and may adopt conformations that suit a wide variety of motifs. Such a view could also explain the disparate binding poses of different ligands in complex structures [40], different structural and dynamic responses to binding of different ligands, and in turn also provide the means for switching between different functional actions.

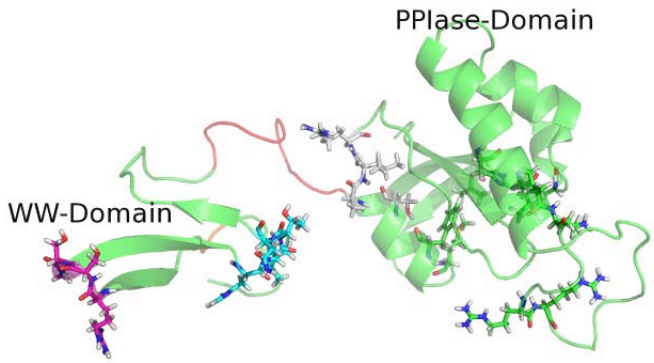


Figure 1. Schematic illustration of human Pin1. The binding loop (loop 1) of Pin1-WW is highlighted in magenta sticks, loop 2 in cyan sticks, and the conserved prolines are colored orange. The residues important for catalysis and binding in the PPIase domain are highlighted with green sticks and the WW domain interaction surface is highlighted with gray sticks. The flexible interdomain linker is red.

In chapter 5 we presented a new approach to generate representative structure ensembles of proteins. In this approach, we combined replica-exchange simulations with chemical shift data through maximum entropy reweighing. Our resulting ensembles correlated well with high-accuracy NMR measurements. In particular, the recently introduced exact nuclear Overhauser enhancement (eNOE) enabled us to validate a significantly populated excited state detected in our ensembles (Figure 1, chapter 5). The excited-state population has the same core β-sheet as the well-known ground state but is characterized by a topological reorganization of the N- and C-termini and a transition in the local structure propensities of loop 1. The latter correlates with a chemical exchange process that is quenched upon ligand binding [39]. In addition, the two states exert different patterns of backbone-backbone structural correlations involving sites for ligand binding (loop 1), domain-domain association (loop 2), and highly conserved proline residues in the N-terminus. These differences provide the basis for a spectrum of functional responses. Our ensembles provide the thermodynamic means to understand how different ligands may act as different allosteric effectors and thereby trigger different functional responses in Pin1.

6.2. Results

6.2.1. Ligand binding shifts native/near-native state equilibrium

As the binding loop of Pin1-WW (loop 1) is a dynamic hotspot, it has been the fulcrum in many previous studies trying to understand binding and catalysis in Pin1. The flexibility of loop 1 is essential for efficient ligand binding [39] and couples to two other

key sites in Pin1-WW: loop 2 involved in the association to the catalytic PPIase domain [41] and in the near-native state, the highly conserved proline residues at position 7 and 8 (Figure 3, chapter 5). We use the S18N/W34F construct as it was more amenable to NMR measurements compared to the WT [42, 43]. The construct involves conservative mutagenesis of residues shown to be involved in ligand binding that could hamper the binding function of the domain. Consequently, we verified that this construct is still able to bind a phosphorylated Cdc25C motif (Figure S1 in appendix 10.3.).

In our WT Pin1-WW simulation, we observe a continuous spectrum of conformational states consistent with the native and near-native configurations we observe in S18N/W34F. This suggests that loop 1 may function as a dial that enables different ligands to skew the conformational populations in different ways. Thus, when a particular ligand binds, it may control the equilibrium between associated and dissociated states of the WW and PPIase domains. This mechanism is likely as loop 1 where ligands bind and the topology of the N- and C-termini are tightly coupled. The topology of the near-native state could potentially interfere with the WW:PPIase association interface and thereby modulate this equilibrium. Consequently, residues flanking the phosphorylated binding motifs may encode specific information in which Pin1 is recruited to a specific purpose.

To test this hypothesis, we used previously published NOE data [40] of free WT Pin1-WW and of the complexes with phosphorylated fragments of the tau-protein (Pin1-WW:pτ) or of Cdc25C (Pin1-WW:pCdc25C). Specifically, we selected subsets of clusters of our CE at 278 K in the WT simulation that gave the best agreement with each of the experimental datasets (see appendix 10.3.). We found a modest difference between the free form and the bound forms. The two ligand-bound forms were represented by the same three clusters. The free form of Pin1-WW was best represented by four clusters, three of which were the same as the bound form.

The structural differences between the set of clusters selected using the dataset of the free and complex constructs are subtle but clear. The free form has a higher population of native state topology of the N- and C-termini, consistent with what one would predict on the basis of the data presented herein and previously [31, 39, 44]. In addition, we find differences in the distributions of dihedral angles in the binding loop (Figure 2). Collectively, these results suggest that ligand binding indeed perturbs the conformational equilibrium of Pin1-WW. To test whether there was a potential for differential modulation by different ligands, we checked for any particular preferences for residues

flanking pSer/pThr-Pro motifs. We split Pin1 ligands into two classes: ligands that only have a single binding site and ligands that have at least one binding site (Figure 3 and appendix 10.3.). In the pool of ligands with only a single binding site, we see clear biases in the residue positions upstream of the binding motif. These trends are clearly stronger than any trends seen at the same positions from the multiple binding site motifs. Downstream (positions 10–16), no strong sequential biases were found for either of the two ligand classes (Figure 3). This finding is consistent with the potential for substrate dependent functional response in Pin1-WW.

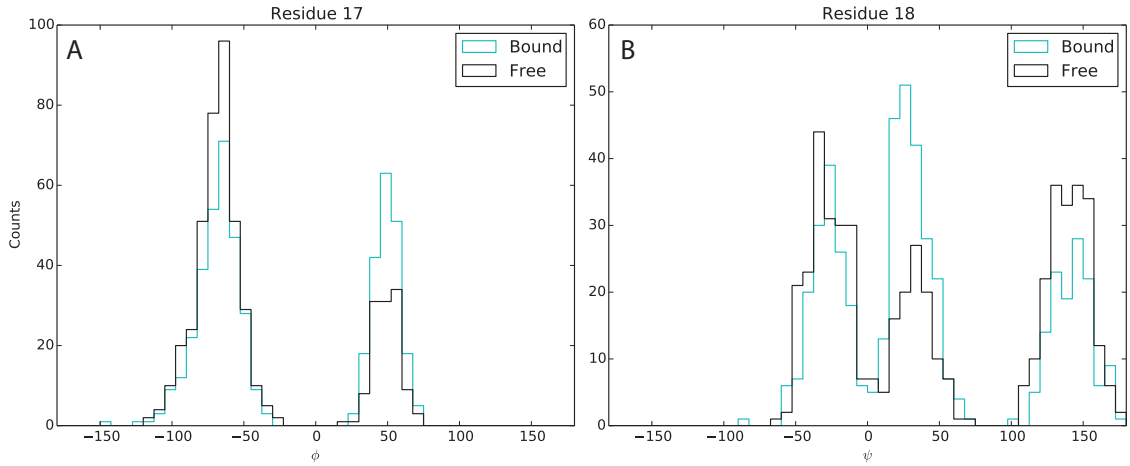


Figure 2. Changes in backbone dihedral angles upon ligand binding in Pin1-WW. Histograms of representative backbone dihedral angles in the bound and unbound ensembles generated using WT simulation and previously published data [40]. (A) Φ angle of residue 17, (B) Ψ angle of residue 18, both located in the binding loop.

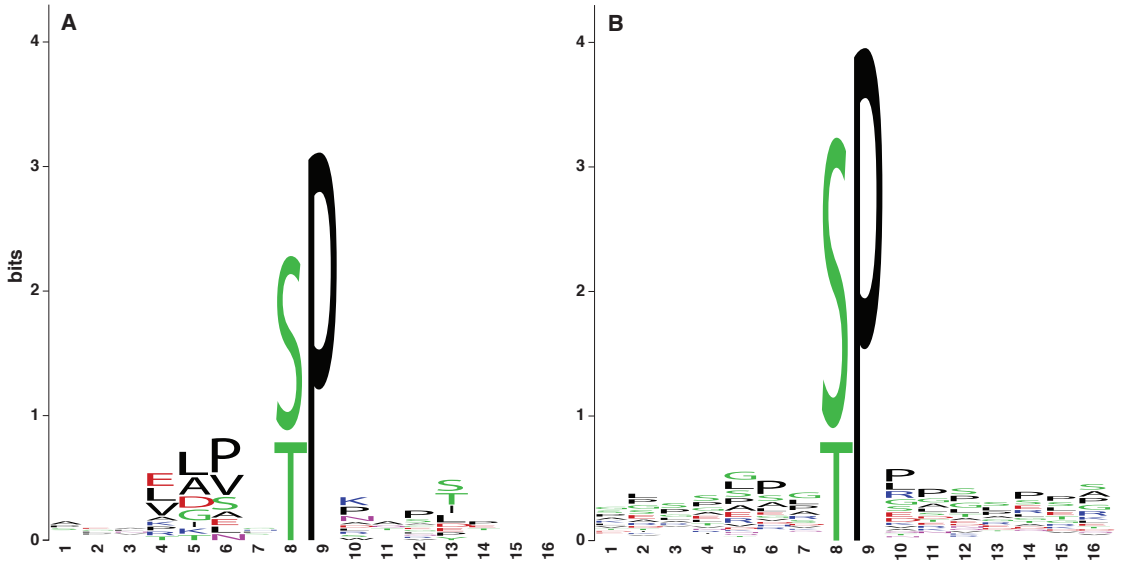


Figure 3. Comparison of two classes of Pin1 binding motifs. (A and B) Logo plots of (A) binding motifs in Pin1 ligands with only one binding motif and (B) binding motifs in ligands with at least one binding motif.

6.2.2. Native and near-native topologies exert different propensities to associate with the catalytic PPIase domain of human Pin1

Finally, we measured the relative propensity to form the WW:PPIase interdomain contact in Pin1 observed in crystallography [27, 45] between the near-native and native topologies of the WW domain. This was done by carrying out 48 MD simulations of full-length Pin1 WT – 24 of each of the two WW domain topologies. In these simulations, the Pin1 coordinates were initialized such that the PPIase and WW domains were dissociated. The simulations were used to estimate two Markov state models (6.4. Experimental Procedures). We found both topologies had some domain association propensity, consistent with previously reported NMR relaxation data [44]. However, these propensities were more pronounced in the native topology compared with the near-native topology. Contributing to our assessment of these propensities were two factors: the average shortest distance between the Cα of residue 28 and those of residues 138, 140–142, $\bar{r}$, and the mean association and dissociation timescales (t_A and t_D) (6.4. Experimental Procedures). The latter measures how long it takes to arrive at an associated state from a dissociated state, or the reverse, on average. $\bar{r}$ was substantially shorter in the native topology simulations (11.9 ± 2.44 Å) compared with the near-native simulations (21.6 ± 1.3 Å) (6.4. Experimental Procedures). Similarly, we found the t_D = 8.6 ± 6.1 ms and t_A = 3.9 ± 4.6 ms for the native topology simulation, whereas t_D = 362.1 ± 99.29 ns and t_A = 1,383 ± 294.1 ns was found for the near-native topology. These values suggest that the exchange between associated and dissociated states overall is faster in the near-native topology compared with the native topology. The values furthermore suggest the dissociation to be fast compared with the association rate in the near-native topology, whereas comparable rates are found for the native topology.

6.3. Discussion

6.3.1. Different ligands may induce alternate allosteric responses in Pin1-WW through an allosteric dial in loop 1

The experimental data presented and discussed (chapter 5) strongly suggest the presence of an energetically excited near-native conformation. The main characteristic of this near-native state is a topological rearrangement of the N- and C-termini and an altered preference of backbone dihedral angles in the binding loop. Multiple previously reported structures of human Pin1-WW had a similar topology (Figure S3 in appendix

10.3.), but what is the biological significance of this configuration? Early NMR studies suggest that Pin1 samples configurations where the PPIase and WW domains are in equilibrium between associated and dissociated states [31, 44, 46]. Furthermore, this equilibrium may be perturbed in different ways by the presence of ligands [36, 44, 47] or mutagenesis [32]. The topological arrangement associated with the transition from the native to the near-native state would obstruct the WW:PPIase domain-domain interaction interface as observed by crystallography [27, 45]. In turn, this may suggest that this minor configuration is primarily populated when the two domains are dissociated, while the major configuration may in principle be compatible with both the associated and dissociated states. Our simulations of Pin1 suggest that this is indeed the case; compared with the native topology, the near-native topology has a lower propensity to form the interaction previously shown to facilitate information transfer between the two domains (cyan and gray sticks in Figure 1). This may be seen as another example of the population-shuffling mechanism [48], but here involving a translational order parameter.

The topology dependent difference in propensity to form specific interactions between the WW and PPIase domains in Pin1 provides the foundation for a potential equilibrium between at least two disparate functional mechanisms of Pin1 (Figure 4). In the first, Pin1 interacts with two pSer/pThr-Pro motifs simultaneously and the second involves a single motif that may be transferred in between the domains [34]. Our results and those presented by others suggest that ligands alter the conformational properties of the WW domain upon binding. This also includes the skewing of populations between the two topological states. Consequently, there is a potential for ligands to recruit Pin1 for a particular purpose by different patterns in residues flanking the recognition motifs. Indeed, in an analysis of known Pin1 interaction partners, we observe some trends in single-motif ligands that are not present in multi-motif ligands. While this analysis does not account for a number of higher order effects (e.g., temporal correlations of phosphorylation of different motifs), it does suggests that flanking residues may play a crucial role for deciding downstream functional response in Pin1.

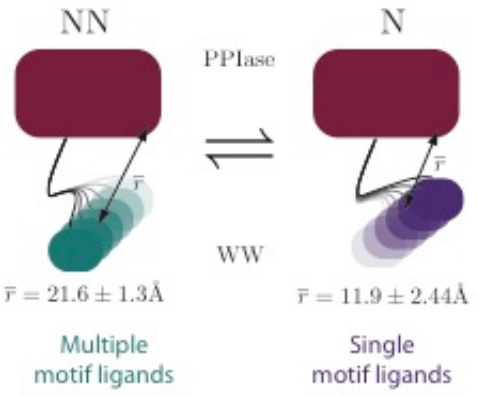


Figure 4. A Model of differential interdomain contact propensity in Pin1 with the WW domain in the native and near-native states. The PPIase domain is red while the WW domain is purple in the native state and teal in the near-native state. The double harpoon represents a slow exchange between the two topologies of the WW domain. Fluctuations in the interdomain distance are represented by multiple states of WW domain, with their transparency as qualitative representation of state probability. The exchange rate between these states is fast compared with the interconversion of the near-native and native topologies. The average shortest distance between the interaction sites of the WW and PPIase domains $\bar{r}$ (±SD) are shown for both states.

6.4. Experimental procedures

6.4.1. Pin1 simulations and analysis

Initial coordinates were generated using the primary sequence of human Pin1 (UniProt: Q13526) and the Phyre 2 web server [49]. A molecular simulation box was prepared as described for the WW domain simulations by solvation in 12,850 TIP3P [50] water molecules, and the charge was neutralized using three chloride ions. The WW and PPIase domains were dissociated using steered MD [51, 52] in the FF99SB-ILDN force field [53]. The resulting dissociated state corresponds to the starting coordinates for the native-like topology simulations. Starting structures of the near-native topology were generated by mechanically reorganizing the WW domain termini using PyMOL; following this intervention, the Pin1 coordinates were re-solvated in 11,312 TIP3P water molecules, and the potential energy was minimized and re-equilibrated in the NPT ensemble. For each of these starting configurations, 24 independent 40 ns simulations were performed using different starting momenta sampled from the Maxwell-Boltzmann distribution at 300 K. The simulation sets were analyzed independently using PyEMMA2 [54]. The Cα-Cα distances between I28 in the WW domain and either of the residues 138, 140–142 in the PPIase domain were used to construct two 64 state Bayesian Markov state models (MSMs) [55]. We found lag times of 3.25 ns yielded models that accurately predict contact dynamics on time scales on the order of 100s of ns, as supported by Chapman-Kolmogorov tests [56] (Figure S2 in appendix 10.3.). We did not observe any transitions between the WW domain topologies during the

simulations performed and conclude that the exchange process likely is much slower than the dynamics described by the MSMs we obtain here. We can therefore confidently assume that the stationary distribution of two MSMs represents conformational distributions within two meta-stable configurations of Pin1. These distributions were used to compute the expected shortest Cα-Cα distances between I28 in the WW domain and either of the residues 138, 140–142 in the PPIase domain, $\bar{r}$. The association and dissociation timescales (t_A and t_B) were computed as the mean first passage time (MFPT) between associated clusters (r < 10 Å) and dissociated clusters (r > 15 Å), where r is defined in the same way as the distance above. The MFPT was computed as previously described [57].

6.4.2. NMR spectroscopy

Titration of Cdc25C peptide with sequence EQPLpTPVTDL (BACHEM) and human Pin1-WW S18N/W34F was followed by ^{15}N-resolved HSQC experiments [58] with relative concentrations of 0.0:1, 0.4:1, 1.2:1, 2.0:1, 2.8:1, 3.6:1, 4.4:1. All spectra were acquired on a Bruker 900 MHz spectrometer. Assuming fast exchange and a two-state binding process, we estimated the dissociation constant to be K_d = 392 ± 201.7 mM [59]. The raw data were processed with NMRpipe [60] and assigned in CCPnmr Analysis [61].

6.5. References

[1] Papaleo E, Saladino G, Lambrughi M, Lindorff-Larsen K, Gervasio FL, Nussinov R, The role of protein loops and linkers in conformational dynamics and allostery. Chem Rev 2016, 116, 6391–6423.

[2] Monod J, Wyman J, Changeux JP, On the nature of allosteric transitions: a plausible model, J Mol Biol 1965, 12, 88–118.

[3] Koshland DE Jr, Némethy G, Filmer D, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 1966, 5, 365–385.

[4] Cooper A, Dryden DT, Allostery without conformational change. A plausible model, Eur Biophys J 1984, 11, 103–109.

[5] Motlagh HN, Wrabl JO, Li J, Hilser VJ, The ensemble nature of allostery, Nature 2014, 508, 331–339.

[6] Freiburger LA, Baettig OM, Sprules T, Berghuis AM, Auclair K, Mittermaier AK, Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme, Nat Struct Mol Biol 2011, 18, 288–294.

[7] Hilser VJ, Thompson EB, Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins, Proc Natl Acad Sci USA 2007, 104, 8311– 8315.

[8] Popovych N, Sun S, Ebright RH, Kalodimos CG, Dynamically driven protein allostery, Nat Struct Mol Biol 2006, 13, 831–838.

[9] Brüschweiler S, Schanda P, Kloiber K, Brutscher B, Kontaxis G, Konrat R, Tollinger M, Direct observation of the dynamic process underlying allosteric signal transmission, J Am Chem Soc 2009, 131, 3063–3068.

[10] Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE, Studying excited states of proteins by NMR spectroscopy, Nat Struct Biol 2001, 8, 932–935.

[11] Palmer AG 3rd, Chemical exchange in biomacromolecules: past, present, and future, J Magn Reson 2014, 241, 3–17.

[12] Swope WC, Pitera JW, Suits F, Describing protein folding kinetics by molecular dynamics simulations. 1. Theory, J Phys Chem B, 2004, 108, 6571–6581.

[13] Noé F, Krachtus D, Smith JC, Fischer S, Transition networks for the comprehensive characterization of complex conformational change in proteins, J Chem Theory Comput 2006, 2, 840–857.

[14] Buchete N-V, Hummer G, Coarse master equations for peptide folding dynamics, J Phys Chem B 2008, 112, 6057–6069.

[15] Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE, Biomolecular simulation: a computational microscope for molecular biology, Annu Rev Biophys 2012, 41, 429–452.

[16] Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M, Simultaneous determination of protein structure and dynamics, Nature 2005, 433, 128–132;

[17] Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Combining experiments and simulations using the maximum entropy principle, PLoS Comput Biol 2014, 10, e1003406.

[18] Pitera JW, Chodera JD, On the use of experimental observations to bias simulated ensembles, J Chem Theory Comput 2012, 8, 3445.

[19] Roux B, Weare J, On the statistical equivalence of restrained- ensemble simulations with the maximum entropy method, J Chem Phys 2013, 138, 084107.

[20] Cavalli A, Camilloni C, Vendruscolo M, Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle, J Chem Phys 2013, 138, 094112.

[21] Olsson S, Frellsen J, Boomsma W, Mardia KV, Hamelryck T, Inference of structure ensembles from sparse, averaged data, PLoS One 2013, 8, e79439.
[22] Olsson S, Vögeli B, Cavalli A, Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Hamelryck T, Probabilistic determination of native state ensembles of proteins, J Chem Theory Comput 2014, 10, 3484–3491.
[23] Olsson S, Ekonomiuk D, Sgrignani J, Cavalli A, Molecular dynamics of biomolecules through direct analysis of dipolar couplings, J Am Chem Soc 2015, 137, 6270–6278.
[24] Hummer G, Köfinger J, Bayesian ensemble refinement by replica simulations and reweighting, J Chem Phys 2015, 143, 243150.
[25] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19, 1053–1057.
[26] Lu K, Hanes S, Hunter T, A human peptidyl-prolyl isomerase essential for regulation of mitosis, Nature 1996, 380, 544–547.
[27] Ranganathan R, Lu K, Hunter T, Noel J, Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent, Cell 1997, 89, 875–886.
[28] Zhou XZ, Kops O, Werner A, Lu PJ, Shen M, Stoller G, Küllertz G, Stark M, Fischer G, Lu KP, Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins, Mol Cell 2000, 6, 873–883.
[29] Helander S, Montecchio M, Pilstkal R, Su Y, Kuruvilla J, Elvén M, Ziauddin JM, Anandapadamanaban M, Cristobal S, Lundström P, et al., Preanchoring of Pin1 to unphosphorylated c-Myc in a fuzzy complex regulates c-Myc activity, Structure 2015, 23, 2267–2279.
[30] Lu KP, Finn G, Lee TH, Nicholson LK, Prolyl cis-trans isomerization as a molecular timer, Nat. Chem. Biol. 2007, 3, 619–629.
[31] Bayer E, Goettsch S, Mueller JW, Griewel B, Guiberman E, Mayr LM, Bayer P, Structural analysis of the mitotic regulator hpin1 in solution: insights into domain architecture and substrate binding, J Biol Chem 2003, 278, 26183–26193.
[32] Wilson KA, Bouchard JJ, Peng JW, Interdomain interactions support interdomain communication in human Pin1, Biochemistry 2013, 52, 6968– 6981.
[33] Guo J, Pang X, Zhou H-X, Two pathways mediate interdomain allosteric regulation in Pin1, Structure 2015, 23, 237–247.
[34] Peng JW, Investigating dynamic interdomain allostery in Pin1, Biophys Rev 2015, 7, 239–249.
[35] Wang X, Mahoney BJ, Zhang M, Zintsmaster JS, Peng JW, Negative regulation of peptidyl-prolyl isomerase activity by interdomain contact in human Pin1, Structure 2015, 23, 2224–2233.
[36] Namanja AT, Wang XJ, Xu B, Mercedes-Camacho AY, Wilson KA, Etzkorn FA, Peng JW, Stereospecific gating of functional motions in Pin1, Proc Natl Acad Sci USA 2011, 108, 12289–12294.
[37] Guo J, Zhou H-X, Dynamically driven protein allostery exhibits disparate responses for fast and slow motions, Biophys J 2015, 108, 2771–2774.
[38] Sudol, M., and Hunter, T., New wrinkles for an old domain, Cell 2000, 103, 1001–1004.
[39] Peng T, Zintsmaster JS, Namanja AT, Peng JW, Sequence- specific dynamics modulate recognition specificity in WW domains, Nat Struct Mol Biol 2007, 14, 325–331.
[40] Wintjens R, Wieruszeski JM, Drobecq H, Rousselot-Pailley P, Buée L, Lippens G, Landrieu I, 1H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides, J Biol Chem 2001, 276, 25150–25156.

[41] Morcos F, Chatterjee S, McClendon CL, Brenner PR, López-Rendón R, Zintsmaster J, Ercsey-Ravasz M, Sweet CR, Jacobson MP, Peng JW, Izaguirre JA, Modeling conformational ensembles of slow functional motions in Pin1-WW, PLoS Comput Biol 2010, 6, e1001015.
[42] Crane JC, Koepf EK, Kelly JW, Gruebele M, Mapping the transition state of the WW domain beta-sheet, J Mol Biol 2000, 298, 283–292.
[43] Price JL, Shental-Bechor D, Dhar A, Turner MJ, Powers ET, Gruebele M, Levy Y, Kelly JW, Context-dependent effects of asparagine glycosylation on Pin1 WW folding kinetics and thermodynamics, J Am Chem Soc 2010, 132, 15359–15367.
[44] Jacobs DM, Saxena K, Vogtherr M, Bernado P, Pons M, Fiebig KM, Peptide binding induces large scale changes in inter-domain mobility in human Pin1, J Biol Chem 2003, 278, 26174–26182.
[45] Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel JP, Structural basis for phosphoserine-proline recognition by group IV WW domains, Nat Struct Biol 2000, 7, 639–643.
[46] Bernadó P, Fernandes MX, Jacobs DM, Fiebig K, García de la Torre J, Pons M, Interpretation of NMR relaxation properties of Pin1, a two- domain protein, based on Brownian dynamic simulations, J Biomol NMR 2004, 29, 21–35.
[47] Namanja AT, Peng T, Zintsmaster JS, Elson AC, Shakour MG, Peng JW, Substrate recognition reduces side-chain flexibility for conserved hydrophobic residues in human Pin1, Structure 2007, 15, 313–327.
[48] Smith CA, Ban D, Pratihar S, Giller K, Schwiegk C, de Groot BL, Becker S, Griesinger C, Lee D, Population shuffling of protein conformations, Angew Chem Int Ed Engl 2014, 127, 209–212.
[49] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE, The Phyre2 web portal for protein modeling, prediction and analysis, Nat Protoc 2015, 10, 845–858.
[50] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML, Comparison of simple potential functions for simulating liquid water, J Chem Phys 1983, 79, 926.
[51] Grubmüller H, Heymann B, Tavan P, Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force, Science 1996, 271, 997–999.
[52] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G, Plumed 2: new feathers for an old bird, Comput Phys Commun 2014, 185, 604–613.
[53] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE, Improved side-chain torsion potentials for the Amber ff99sb protein force field, Proteins 2010, 78, 1950–1958.
[54] Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H, Noé F, Pyemma 2: a software package for estimation, validation, and analysis of Markov models, J Chem Theory Comput 2015, 11, 5525–5542.
[55] Trendelkamp-Schroer B, Wu H, Paul F, Noé F, Estimation and uncertainty of reversible Markov models, J Chem Phys 2015, 143, 174101.
[56] Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F, Markov models of molecular kinetics: generation and validation, J Chem Phys 2011, 134, 174105.
[57] Noé, F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, Proc Natl Acad Sci USA 2009, 106, 19011–19016.
[58] Cavanagh J, Fairbrother WJ, Palmer AG III, Rance M, Skelton NJ, Protein NMR Spectroscopy: Principles and Practice, Academic Press 2007.

[59] Williamson MP, Using chemical shift perturbation to characterize ligand binding, Prog Nucl Magn Reson Spectrosc 2013, 73, 1–16.
[60] Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol NMR 1995, 6, 277–293.
[61] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley J.L, Ionides J, Laue ED, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 2005, 59, 687–696.

7. The Allosteric Coupling Of Ligand-Binding Via The Interaction Interface in Pin1

This chapter is based on the following work in progress:

Dean Strotz, Celestine N. Chi, Julien Orts, Simon Olsson, Peter Güntert, Beat Vögeli, Roland Riek, The Allosteric Coupling Of Ligand-Binding Via The Interaction Interface in Pin1, work in progress. Celestine N. Chi produced the full length Pin1 and recorded the corresponding titration with pCdc25C. Julien Orts measured the STD experiment. Dean Strotz did the analysis, recorded all experimental data (except as mentioned above), prepared the figures and wrote the initial manuscript.

7.1. Introduction

Allostery in proteins describes the process by which an effect (such as ligand binding) on one site of a biomolecule or bimolecular complex is transmitted to another distal functional site. This 'action at a distance' phenomenon can lead to regulation of biological activity (see the review by Motlagh et al. [1]). However, the elucidation of the nature of allostery appears to be rather difficult and "remains a biophysical enigma eluding a general, quantifiable and predictive atomic description" (citation from [1]). Nonetheless, on the basis of biophysical measures including NMR and the determination of distinct structures of proteins stabilized/kept in given states, several models on the mechanism of allostery have been postulated including the sequential mechanism (induced fit) by Koshland, Nemethy and Filmer [2, 3], the conformational selection mechanism (originally termed the symmetric model, now called the MWC model) by Monod, Wymann, and Changeux [4], and the dynamic allostery model introduced by Cooper and Dryden [5]. While the sequential mechanism assumes adaptability of the structure upon ligand binding, the MWC model assumes the existence of two pre-existing states whose equilibrium shifts toward the ligand-binding competent form upon ligand binding. In other words, a protein may be required to sample a broad conformational space in the absence of ligand out of which conformations are selected upon their relative affinities for the ligand [6]. The dynamic allostery model assumes that ligand binding changes the frequency and amplitude of thermal fluctuations within a protein without perturbing the average structure. Trying to combine these proposed

mechanisms of allostery on the basis of a thermodynamic argumentation, it was suggested that allosteric proteins should be viewed as ensembles of interchanging states and that ligand binding may skew the populations in the conformational ensemble [1, 7].

The coordinated structural and dynamic changes that allostery is proposed to build on are however elusive to most structural biology techniques and rarely accessible on a molecular level with detailed structural information (beyond kinetic and thermodynamic data) because the measured NMR parameters, such as NOEs and RDCs, are ensemble averaged. Native-state HD exchange and relaxation dispersion [8, 9] opened an avenue to study "invisible" intermediate states and states with low populations [10]. However, as relaxation dispersion describes timescales of motion, no amplitudes of motion can be deduced. The methods to detect correlated motion experimentally over broad time and length scales are very challenging [11-16] and hence MD simulation are often used to model and explore correlated motions [17].

The biological significance and function of Pin1 includes involvement in the regulation of mitosis [18] and a protective function against Alzheimer's disease [19]. Pin1 modulates hepatitis C infection [20] and is overexpressed in many human cancer cells [21]. The human peptidyl-prolyl cis-trans isomerase (PPIase) Pin1 contains the N-terminal Pin1-WW subunit separated by a flexible linker from the C-terminal catalytic domain. The 34-residue-long Pin1-WW folds into a signature, three-stranded, anti-parallel and twisted β-sheet. The domain is unusually resistant to aggregation for a small β-sheet fold. The high stability likely arises due to a large number of hydrogen bonds, a hydrophobic mini-core and proline rich N- and C-termini that lock the fold [22]. The Pin1-WW subunit contains with loop1 (residues 16-20) an unusually long, highly flexible [23, 24], hyper variable and high affinity phosphate-binding site. The binding site is responsible for localization and recognition of phospho-Ser/Thr-Pro motives (KEGG Database, [EC 5.2.1.8]). Containing a second binding site, the C-terminal peptidyl-prolyl cis-transe Pin1-PPIase competes (although with lower affinity) with the Pin1-WW subunit for the same substrate, but only the Pin1-PPIase domain catalyzes the cis/trans isomerization [25].

The dynamics of WW-domains have been extensively studied in order to understand their substrate recognition preferences [17, 24, 26, 27], the dynamic allosteric coupling to the C-terminal catalytic domain [17, 26, 28], folding [29, 30] and stability [22]. The dynamic coupling of the two domains signifies a differential response of Pin1 to a wide variety of ligands. This differential cross-talk between the two domains, having been shown to take place between loop2 (residues 27-30) of the Pin1-WW and residues 138

and 140-142 of the Pin1-PPIase [31, 32], is therefore of particular interest (Figure 1a). Previously, we suggested promiscuous binding of the Pin1-WW to its ligands and used replica exchange simulations combined with chemical shift data through maximum entropy reweighing to show that Pin1-WW:loop1 may act as an allosteric dial, conducting the interaction with the Pin1-PPIase through loop2 [26] and furthermore demonstrated the sensitivity of the eNOE data to sparsely populated states [26]. In the present study we use our complementary two-states structure calculation protocol [33, 34] applied to the Pin1-WW in isolation. With our data we aim to further analyze the allosteric coupling of the ligand-binding event at loop1 to the interaction interface between the Pin1-WW and the Pin1-PPIase. Indeed, our atomic resolution data, derived from eNOE data, backbone and side-chain scalar couplings, allows us to illustrate the internal correlation of the apo form of the Pin1-WW domain. We have clear evidence of the apo form sampling the holo state in line with conformational sampling. By comparing our apo and holo structural ensembles we show how the ligand-binding (pCdc25C) event at loop1 leads to a breaking of the internal correlation present in the apo form and a decoupling of the allosteric network from parts of the inter-domain interface, concurrent with a redistribution of states at the beginning of β-strand3. These changes lead to a weaker association between the Pin1-WW domain and the Pin1-PPIase domain.

7.2. Results and discussion

7.2.1. Pin1-WW binding to ligand pCdc25C: Chemical shift analysis and Saturation Transfer Difference (STD) NMR

The Pin1-WW construct S18N/W34F behaves experimentally more favorable and even more resistant to aggregation than the wild-type [35, 36]. Due to the conservative mutagenesis of residues shown to be directly involved in ligand binding [23, 24, 37] it was verified that the construct is able to bind the pCdc25C fragment (EQPLpTPVTDL, Figure S1 in appendix 10.4.). The differential chemical shift mapping of the isolated Pin1-WW domain is shown in Figure 1a, confirming ligand binding as expected [23, 24, 31, 32, 37]. Using the isotherms of the chemical shift perturbation and two-state exchange models (chemical shift perturbation versus concentration) the pCdc25C affinity (K_d) was determined to be 392 +/- 202 μM in line with isothermal titration calorimetry (ITC) measurements (supplementary Figure S2 in appendix 10.4.) that yielded a similar affinity (K_d) of 158 +/- 70 μM.

Next, the modulation of the Pin1-WW:PPIase interaction interface, as suggested by the crystal structure of Pin1 (1Pin.pdb, [38]), was investigated in the full length Pin1 mutant protein (S18N/W34F) upon ligand-binding of pCdc25C (Figure S3 in appendix 10.4.). In the Pin1-WW domain residues His27, Ile28, Thr29, Asn30, Ala31, Ser32 and Gln33 display chemical shift perturbations, while in the Pin1-PPIase domain residues Lys97, Ala137, Ser138, Phe139, Ala140, Leu141, Arg142 Ser147 and Gly148 are affected by chemical shift changes. This analysis shows that the ligand binding event at loop1 of the Pin1-WW affects the Pin1-WW:PPIase interaction interface in close agreement with previous results [31, 32].

Saturation transfer difference (STD) experiments were used to analyze changes in the strength of interaction between the Pin1-WW and Pin1-PPIase domains upon addition of ligand pCdc25C (Figure S4 and S5 in appendix 10.4.). The difference spectra (control – irradiated experiment) show a less efficient transfer of magnetization upon addition of ligand at residues Arg142, Ser147 and Val150 in the vicinity of loop2 and beginning of β -strand2. This loss of transfer efficiency is direct evidence that the interaction via the Pin1-WW:PPIase interface is less tight/rigid upon addition of pCdc25C, indicating a ligand-induced decoupling of the proposed allostery between the ligand-binding site on loop1 and the domain interface. Our findings are supported by the previous research of Peng and coworkers, who projected a decrease in inter-domain contact upon addition of pCdc25C [31, 32] from chemical shift perturbations and using Pin1 mutant constructs [32]. To rationalize the above findings atomic resolution structures are necessary.

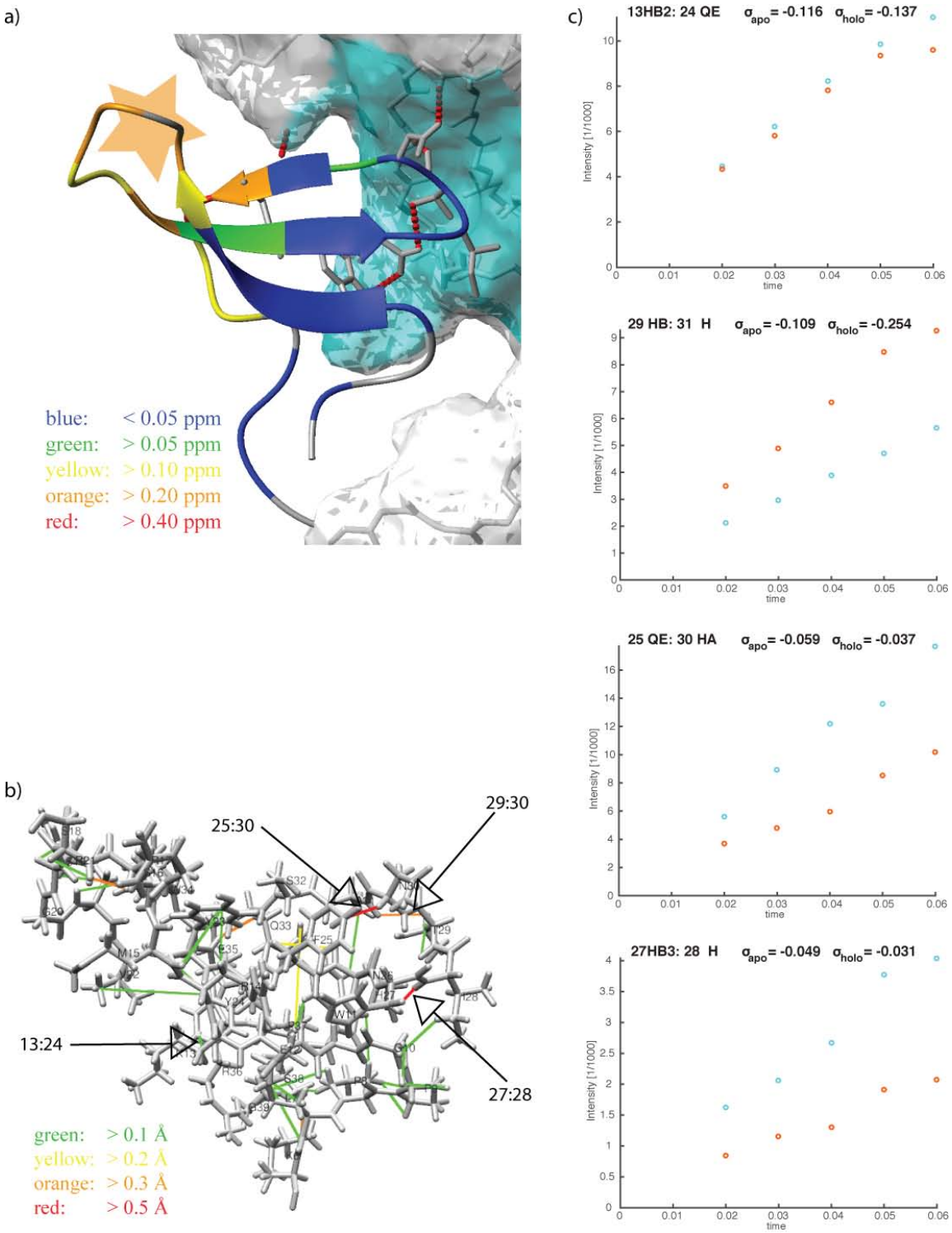


Fig. 1: a) Differential chemical shift mapping of Pin1-WW upon pCdc25C ligand binding: The orange star denotes the ligand binding region of loop1. Residue 17 (grey) is broadened beyond detection. The grey side chains are residues in the interaction interface part of the Pin1-WW, the cyan surface area approximates the interaction interface part of the Pin1-PPIase (grey). The Figure was created with molmol [39]. b) Mapping of changes in bidirectional eNOE distances upon ligand binding, illustrating the network of structural changes relaying the ligand-binding event at loop1 to the interaction interface. The Figure was created with Chimera [40]. c) For the restraints highlighted in the structure (b) an overlay comparing apo (cyan) and holo (red) experimental NOE buildup intensities are shown to illustrate their experimental difference. The experimental buildups are scaled with the ratio of the correlation times. The buildup derived NOE rates are given in 1/s.

7.2.2. The two-states structure of apo Pin1-WW

The two-states ensemble for the apo form was calculated from 279 well resolved bidirectional eNOEs and 450 unidirectional eNOEs, 62 scalar (backbone and side chains, Figure 2, 3, 4, S6 and Table S1 in appendix 10.4.). The eNOE data is well represented by the two-states ensemble as evidenced by the CYANA target function (TF) analysis (Figure 3c, S6, S7a in appendix 10.4.), jack-knife back-calculation (S7b in appendix 10.4.) and the correlation between back-predicted and experimental cross-correlation rates, which were not used in the structure calculation (Figure S8 in appendix 10.4.). For purposes of illustration, we further show that the back-predicted buildups of the two-states ensemble fit the experimental NOE buildups better than the ones of the single-state ensemble (Figure S9, calculated using eNORA2 [41, 42] in appendix 10.4.). The bundle representation of the backbone of apo Pin1-WW shown in Figure 3b illustrates the presence of two distinct structural states reaching from the ligand binding site at loop1 up to the Pin1-WW:PPIase interaction interface at loop2. It is interesting to note that the cyan state of the apo two-states ensemble superimposes well with the holo X-ray structure (1F8A [37]) (Figure 2). In particular, the side chains of residues Trp11, Phe25, Asn26, Ile28, Thr29, Asn30, Asn33 and Phe34, which are building the hydrophobic core as well as the interaction-interface to the catalytic domain. This finding indicates that the apo Pin1-WW domain exchanges in the micro-millisecond time range between a catalytic domain-interaction competent and incompetent state.

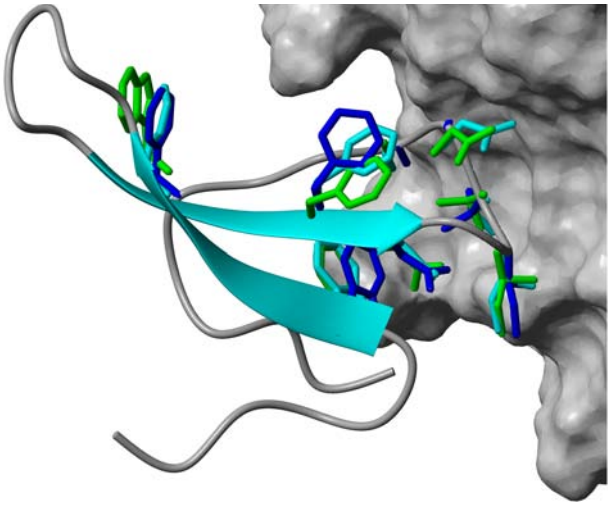


Fig 2: Overlay of representative conformers of the apo two-states ensemble (cyan/blue) with the holo X-ray (green) (18FA.pdb [37]), illustrating conformational sampling.

7.7.3. The two-states structure of holo Pin1-WW in complex with the peptide pCdc25C

The preliminary two-states ensemble for the holo form of the Pin1-WW domain in complex with the peptide pCdC25C was calculated from 269 well resolved bidirectional eNOEs and 458 unidirectional eNOEs, 59 scalar (backbone and side chains, Figure 3, 4, S6, Table S2 in appendix 10.4.). As in the apo state, the eNOE data is well represented by the two-states ensemble as evidenced by TF analysis (Figure 3c, S6, S7a in appendix 10.4.), jack-knife back-calculation (S7b in appendix 10.4.), the cross-correlation rates for the complex were not measured. The significantly higher drops in TF value with increasing states for the holo ensemble indicate more dynamic movement than for the apo ensemble. We are currently analyzing the three-states bundles to support this argument.

The mapping of changes in bidirectional eNOE distance restraints (Figure 1b) upon ligand binding illustrates a network of eNOEs that relay the binding event at loop1 to the Pin1-WW:PPIase interaction interface. An overlay comparing the apo and holo experimental buildup intensities illustrates the validity of the calculated differences in structural restraints (Figure 1c). The apo and holo Pin1:WW domain have different correlation times (τ_c) of 4.25 ns versus 5.67 ns (at 5° C), therefore the experimental buildups are scaled with the ratio of the correlation times (a correct representation to first order) for visual analysis.

It is apparent (Figure 3b/4a) that one of the conformers of the apo two-states ensemble (cyan/blue) also samples the conformational space of the holo form (orange) in β-strand2 (residues 21-26) and loop2 (residues 27-30), in line with conformational sampling. We interpret deviations of this result in loop1 (residues 16-20) as induced fit upon ligand binding (Figure 4a). The tightening of the binding cleft in the holo structure is illustrated with a representative conformer of the apo (cyan) and holo (orange) two-states ensemble in Figure 4b. It is reasonable to speculate that the ligand specific tightening of the structure around the ligand-binding site directly modulates the Pin1-WW:PPIase interaction, but due to ligand proximity to the Pin1-WW:PPIase interaction interface, bound ligands may modulate the interaction further with direct contacts to the interaction interface [28].

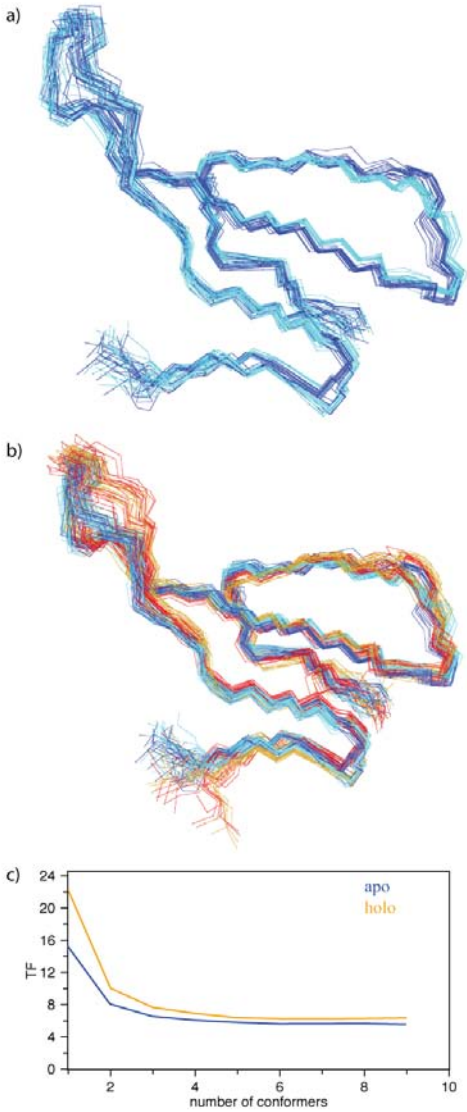


Fig. 3: Pin1-WW. a) apo two-states ensemble b) overlay of apo two-states ensemble (cyan /blue) and the holo two-states ensemble (orange/red). c) The overall target function (TF) as function of number of states. The two-states ensemble reflects the data better than the single-state representation.

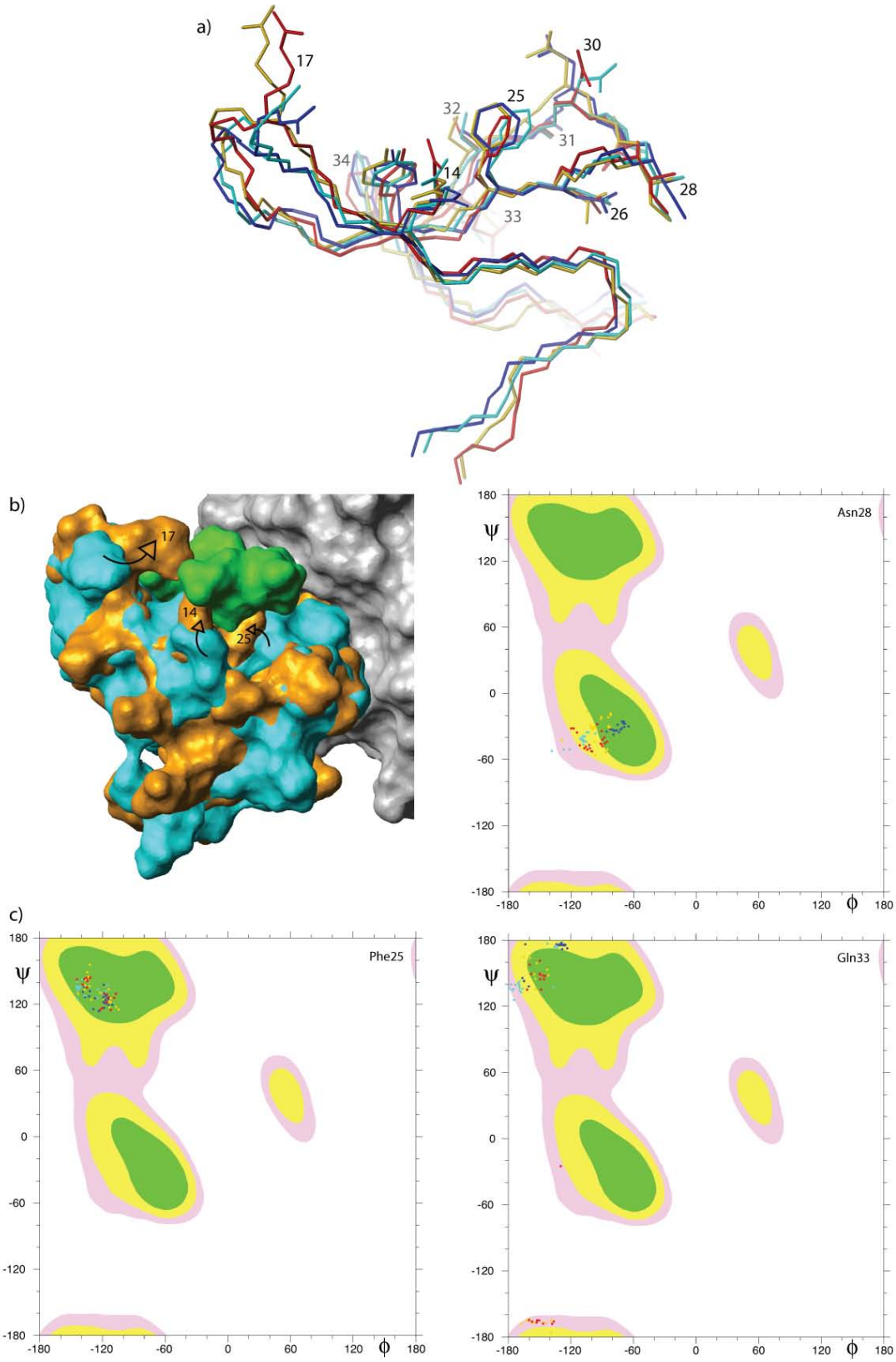


Fig 4: a) Representative conformers of the apo (cyan/blue) and holo (orange/red) two-states ensembles, with stick representation of side-chains important to ligand-binding and at the Pin1-WW:PPIase interaction interface. b) Surface plot with a representative conformer of the apo (cyan) and holo (orange) two-states ensembles. The Pin1-PPIase interaction interface (grey) and the ligand (green) are shown (18FA.pdb, [ref]). Tightening of the holo Pin1-WW domain around the ligand-binding site and induced fit of residue 17. c) Ramachandran plots of backbone Ψ and Φ angles for select residues, illustrating the loss of correlation upon ligand binding in residues following loop2 (residues 27-30) at the beginning of β-strand 3 (residues 31-33). The dots are following the apo (cyan/blue) and holo (orange/red) color scheme. All Figures created with molmol [39].

Ramachandran plots of backbone Ψ and Φ angles (Figure 4c, S10 in appendix 10.4.) show a loss in internal correlation for specific residues, in a manner independent of structural super-positioning. For residue Phe25 the states are not separable, for residue Asn26 the holo two-states ensemble gravitates towards the cyan apo-state (Figure S10 in appendix 10.4.), while in Asn28 and Ile29 the states are clearly separable in the apo as well as the holo two-states ensemble. For residues Ser32 the separation of the states is less clear and for residue Gln33 the separation of the states is lost. An overlay of the RMSD to mean structure for the apo and holo ensembles (states were sorted by analysis of Asn26, Figure S11 in appendix 10.4.) conveys the same information. The figure shows a general decrease of the RMSD to the mean structure. This is indicative for a loss in correlation. There are subtle differences between the residues. In the key residues implicated in ligand binding (Arg14, Arg17, Tyr23, Phe25) the RMSD to the mean structure remains approximately constant except for residue Phe25. For the residues (Trp11, Asn26) implicated in allosteric signaling of the binding event from loop1 to the interface, the RMSD to the mean structure decreases clearly. For the residues in the Pin1-WW:PPIase interface, Ile28, Thr29 and Asn30, the RMSD to the mean remains approximately constant, whereas it decreases for residues Ala31, Ser32 and Gln33. In conclusion, both the apo and holo forms of the Pin1-WW domain can be sorted into two states (Figure 3) but only for the apo two-states ensemble is the separation evident over the whole Pin1-WW domain. The ligand-binding event of pCdc25C at loop1 therefore breaks the internal correlation present in the apo Pin1-WW domain, resulting in several, only locally correlated states and an allosteric redistribution of states at the beginning of β-strand3.

7.2.4. Proposed Pin1-WW:PPIase interface interaction in response to pCdc25C binding

The measured chemical shift perturbations of full-length Pin1 upon addition of pCdc25C illustrate the effect of ligand-binding on the Pin1-WW:PPIase interaction and based on STD measurements we have direct experimental evidence of a weakening of this interaction. On a structural basis we illustrate the breaking of the internal correlation present in apo Pin1-WW, resulting in several, only locally correlated states and an allosteric redistribution of states at the beginning of β-strand3, which leads to a loss of the coherent inter-domain binding motif and concomitantly a decrease of the inter-domain interaction affinity. Based on our observations and on previous research by Peng

and coworkers [31, 32], it is reasonable to project that other ligands influence the coupling network differently.

7.3. Experimental procedures

7.3.1. Preparation of samples

Preparation $^{15}N/^{13}C$ labeled samples was described previously in detail [43], the phosphorylated ligand was ordered from Bachem AG, Switzerland. The NMR buffer was 10mM K_2PO_4, 100mM NaCl, 0.02% NaN_3, pH 6.0 with sample concentrations of 1.2mM Pin1-WW. For the holo sample a 4-fold excess of pCdc25C was used.

7.3.2. NMR experiments

All experiments were recorded on a Bruker 700 MHz spectrometer and at 5° C, except where described otherwise. All spectra were processed and analyzed using the software package NMRPipe [44], assignment was done in CcpNMR [45]. The measurement and analysis of eNOEs using eNORA2 [41, 42] was described previously in detail [43], in short: series of 3D [^{15}N, ^{13}C]-resolved [^{1}H, ^{1}H]-NOESY-HSQC experiments were recorded to measure NOE buildups [54]. The inter-scan delay was 0.8 s. Simultaneous [^{15}N, ^{1}H]- HSQC and [^{13}C, ^{1}H]-HSQC elements were employed, following indirect proton chemical shift evolution and [^{1}H, ^{1}H]-NOE mixing (τ_m). Diagonal-peak decays and cross-peak buildups were measured with τ_m of 20, 30, 40, 50, and 60 ms for both apo and holo samples.

7.3.3. τ_c measurements

τ_c measurements were described previously in detail [43], in short: τ_c for the apo Pin1-WW domain was previously determined to be 4.25 ns at 5° C and 1.2mM concentration [43]. τ_c for the holo Pin1-WW domain was determined to be 5.67 ns at 5° C and 1.2mM concentration.

7.3.4. $^{3}J_{HN,H\alpha}$ scalar coupling

$^{3}J_{HN,H\alpha}$ scalar coupling measurements were described previously in detail [43].

7.3.5. $^{3}J_{H\alpha,H\beta}$ scalar couplings

$^{3}J_{H\alpha,H\beta}$ scalar couplings were obtained from 3D $^{13}C\alpha$-separated H^{α}-H^{β} in-phase COSY (HACAHB-COSY) experiments [46] in D_2O. The experiment was recorded with 50(MQ[Cα], t1) × 54(Hβ, t2) × 2048(Hα, t3) complex points, giving t1max = 22.5 ms, t2max = 10.8 ms, t3max = 204.8 ms. The time domain data were multiplied with a square cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to 256 × 512 × 2048 complex points. The Karplus parameters used in structure calculations were from [47].

7.3.6. $^{3}J_{C',C}$ and $^{3}J_{N,C}$ scalar couplings (aromatic residues)

$^{3}J_{C',C^{\gamma}}$ and $^{3}J_{N,C^{\gamma}}$ scalar couplings for aromatic side chains were obtained from $^{13}C'$-$\{^{13}C^{\gamma}\}$ and ^{15}N-$\{^{13}C^{\gamma}\}$ spin-echo difference ^{1}H-^{15}N HSQC experiments [48] performed on a Bruker 600 MHz spectrometer. The experiment were recorded with 100(N, t1) or 200(N, t1) × 512(HN, t2) complex points, giving t1max = 50 ms or t1max = 100ms and t2max = 51.2 ms, respectively. The time domain data were multiplied with a square cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to 512 × 2048 complex points. The Karplus parameters used in structure calculations were from [47].

7.3.7. Cross-correlated relaxation (CCR)

$\Gamma_{HNiNi/H\alpha iC\alpha i} + \Gamma_{H\alpha iNi/HNiC\alpha i}$ were obtained from two experiments performed on a Bruker 600 MHz spectrometer equipped with a z-axis gradient cryogenic probe. A DIAI (double in-phase/anti-phase inter-conversion) method was realized with a pair of 3D HNCA pulse sequences (“reference” and “trans”) [49] for the first experiment. A 3D ct-HNCA MMQ (mixed multi-quantum, with zero- and double-quantum coherence evolution averaged) experiment was used for the second experiment. The ZQ and DQ coherences were superimposed, resulting in four components to be evaluated [50]. The experiments were recorded with τMQ = 31.0 ms or τMQ = 33.5 ms, 50(MQ[N,Cα], t1) or 55(MQ[N,Cα], t1) × 36(N, t2) × 512(HN, t3) complex points, t1max = 25.0 or 27.5 ms, t2max = 18.0 ms, t3max = 51.2 ms. The time domain data were multiplied with a square cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to 256 × 128 × 2048 complex points.

7.3.8. Prediction of CCR rates

The back-calculation has been described previously in detail [16].

7.3.9. Structure calculation

A comprehensive description of the various sub-states based on experimental data can be obtained by following an established, ensemble based protocol [33, 34] using the software packages eNORA2 [41, 42] and CYANA [51, 52]. As input for structure calculation we used upper and lower distance restraints from eNOEs together with backbone, Hβ and aromatic side chain scalar couplings and conservative Φ and Ψ dihedral angle restraints derived from C_α chemical shifts [53]. The weight of the dihedral angle restraints was reduced to zero in the final steps of the structure calculation. Calculations were done with 50'000 torsion angle dynamics steps for 100 conformers with random torsion angles by simulated annealing. The multi-states structural ensembles were each calculated simultaneously and averaged. A weak harmonic well potential with bottom width of 1.2 Å was used to keep identical heavy atoms from the different states spatially together and avoid divergence among the structural states that is not implied by the experimental restraints [33, 54]. The 20 conformers with the lowest final target function values were selected and analyzed. The calculated coordinates and the complete data set consisting of the eNOEs together with the upper/lower distance limit tables will be deposited in the PDB/BMRB.

7.4. References

[1] Motlagh HN, Wrabl JO, Li J, Hilser VJ, The ensemble nature of allostery, Nature 2014, 508(7496), 331-9.

[2] Koshland DE, Némethy G, Filmer D, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 1966, 365–385.

[3] Koshland DE, The Key-Lock Theory and the Induced Fit Theory, Angew Chem 1994, 33, 2375 -2378.

[4] Monod J, Wyman J, Changeux JP, On the nature of allosteric transitions: a plausible model, J Mol Biol 1965, 88–118.

[5] Cooper A, Dryden DT, Allostery without conformational change. A plausible model, Eur Biophys J 1984,11(2), 103-9.

[6] Gianni S, Dogan J, Jemth P, Distinguishing induced fit from conformational selection, J Biophys Chem 2014, 189, 33-39.

[7] Liu J, Nussinov R, Allostery: An Overview of Its History, Concepts, Methods, and Applications, PLoS Comput Biol 2016, 12(6):e1004966.

[8] Vallurupalli P, Hansen DF, Kay LE, Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy, Proc Natl Acad Sci USA 2008, 105, 11766-11771.

[9] Neudecker P, Zarrine-Afsar A, Davidson AR, Kay LE, Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy, Proc Natl Acad Sci USA 2007, 104, 15717-15722.

[10] Baldwin AJ, Kay LE, NMR spectroscopy brings invisible protein states into focus, Nat Chem Biol 2009, 5, 808-814.

[11] Cusack S, Doster W, Temperature dependence of the low frequency dynamics of myoglobin, Biophys J 1990, 58, 243-251.

[12] Faure P, Micu A, Perahia D, Doucet J, Smith JC, Benoit JP, Correlated intramolecular motions and diffuse x-ray scattering in lysozyme, Nat Struct Biol 1994, 1, 124-128.

[13] Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH, NMR evidence for slow collective motions in cyanometmyoglobin, Nat Struct Biol 1997, 4, 292-297.

[14] Mayer KL, Earley MR, Gupta S, Pichumani K, Regan L, Stone MJ, Covariation of backbone motion throughout a small protein domain, Nat Struct Biol 2003, 10, 962-965.

[15] Vogeli B, Yao L, Correlated Dynamics between Protein HN and HC Bonds Observed by
NMR Cross Relaxation, J Am Chem Soc 2009, 131, 3668.

[16] Vogeli B, Comprehensive description of NMR cross-correlated relaxation under anisotropic molecular tumbling and correlated local dynamics on all time scales, J Chem Phys 2010, 133(1), 014501.

[17] Morcos F, Chatterjee S, McClendon CL, Brenner PR, López-Rendón R, Zintsmaster J, Ercsey-Ravasz M, Sweet CR, Jacobson MP, Peng JW et al, Modeling Conformational Ensembles of Slow Functional Motions in Pin1-WW, PLoS Comput Biol 2010, 6(12), e1001015.

[18] Lu KP, Hanes SD, Hunter T, A human peptidyl-prolyl isomerase essential for regulation of mitosis, Nature 1996, 380, 544-547.

[19] Ma SL, Pastorino L, Zhou XZ, Lu KP, Prolyl isomerase Pin1 promotes amyloid precursor protein (APP) turnover by inhibiting glycogen synthase kinase-3 (GSK3) activity: Novel mechanism for Pin1 to protect against Alzheimer disease, J Biol Chem 2012, 287, 6969-6973.

[20] Lim YS, Tran HTL, Park SJ, Yim SA, Hwang SB, Peptidyl-prolyl isomerase Pin1 is a cellular factor required for Hepatitis C virus propagation, J Virol 2011, 85, 8777-8788.

[21] Lu KP, Prolyl isomerase Pin1 as a molecular target for cancer diagnostics and therapeutics, Cancer Cell 2003, 4, 175-180.
[22] Jäger M, Dendle M, Kelly JW, Sequence determinants of thermodynamic stability in a WW domain-An all-β-sheet protein, Protein Sci 2009, 18(8), 1806-1813.
[23] Jäger M, Zhang Y, Bieschke J, Nguyen H, Dendle M, Bowman M, Noel JP, Grueble M, Kelly JW, Structure-function-folding relationship in a WW domain, PNAS 2006, 10648-10653.
[24] Peng T, Zintsmaster JS, Namanja AT, Peng JW, Sequence-specific dynamics modulate recognition specificity in WW domains, Nat Struct Mol Biol 2007, 14(4), 325-331.
[25] Jacobs DM, Saxena K, Vogtherr M, Bernado P, Pons M, Fiebig KM, Peptide binding induces large scale changes in inter-domain mobility in human Pin1, J Biol Chem 2003, 278(28), 26174-82.
[26] Olsson S, Strotz D, Vögeli B, Riek R, Cavalli A, The Dynamic Basis for Signal Propagation in Human Pin1-WW, Structure 2016, ii: S0969-2126(16)30159-9.
[27] Peng JW, Wilson BD, Namanja AT, Mapping the dynamics of ligand reorganization via 13CH3 and 13CH2 relaxation dispersion at natural abundance, J Biomol NMR 2009, 45(1-2), 171-183.
[28] Guo J, Pang X, Zhou HX, Two pathways mediate interdomain allosteric regulation in pin1, Structure 2015, 23(1), 237-47.
[29] Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M, The folding mechanism of a beta-sheet: the WW domain, J Mol Biol 2001, 311(2), 373-393.
[30] Jäger M, Zhang Y, Bieschke J, Nguyen H, Dendle M, Bowman ME, Noel JP, Gruebele M, Kelly JW, Structure-function-folding relationship in a WW domain, Proc Natl Acad Sci USA 2006, 103(28), 10648-10653.
[31] Wilson KA, Bouchard J, Peng J, Interdomain interactions support interdomain communication in human Pin1, Biochemistry 2013, 52, 6968-81.
[32] Peng J, Investigating Dynamic Interdomain Allostery in Pin1, J Biophys Rev 2015, 7, 239-249.
[33] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19,1053-7.
[34] Vögeli B, The nuclear Overhauser effect from a quantitative perspective, Prog Nucl Magn Reson Spectrosc 2014, 78(C), 1-46.
[35] Crane JC, Koepf EK, Kelly JW, Gruebele M, Mapping the transition state of the WW domain beta-sheet, J Mol Biol 2000, 298, 283–292.
[36] Price JL, Shental-Bechor D, Dhar A, Turner MJ, Powers ET, Gruebele M, Levy Y, Kelly JW, Context-dependent effects of asparagine glycosylation on Pin WW folding kinetics and thermodynamics, J Am Chem Soc 2010, 132, 15359–15367.
[37] Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel J, Structural basis for phosphoserine-proline recognition by group IV WW domains, Nature 2000, 7, 639-643.
[38] Ranganathan R, Lu KP, Hunter T, Noel JP, Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent, Cell 1997, 89(6), 875-86.
[39] Koradi R, Billeter M, Wüthrich K, MOLMOL: A program for display and analysis of macromolecular structures, J Mol Graph 1996, 14(1), 51-55.
[40] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE J, UCSF Chimera--a visualization system for exploratory research and analysis, Comput Chem 2004, 25(13), 1605-12.

[41] Orts J, Vögeli B, Riek R, Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs, J Chem Theory Comput 2012, 8(10), 3483-3492.
[42] Strotz D, Orts J, Chi C, Riek R, Vögeli B, The eNORA2 exact NOE analysis program, to be submitted.
[43] Strotz D, Orts J, Minges M, Vögeli B, The experimental accuracy of the uni-directional exact NOE, J Magn Reson 2015, doi: 10.1016/j.jmr.2015.07.007
[44] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol NMR 1995, 6(3):277-293.
[45] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 2005, 59(1):687-696.
[46] Grzesiek S, Kuboniwa H, Hinck AP, Bax A, Multiple-Quantum Line Narrowing for Measurement of Hα-Hβ J Couplings in Isotopically Enriched Proteins, J Am Chem Soc 1995, 117, 5312-5315.
[47] Pérez C, Löhr F, Rüterjans H, Schmidt JM, Self-Consistent Karplus Parametrization of 3JCouplings Depending on the Polypeptide Side-Chain Torsion χ 1, J Am Chem Soc 2001, 123(29), 7081-7093.
[48] Hu JS, Grzesiek SS, Bax A, Two-dimensional NMR methods for determining Xi angles of aromatic residues in proteins from thee-bond JCCy and JNCy couplings, J Am Chem Soc 1997, 119, 1803-1804.
[49] Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G, Simultaneous determination of ψ and Φ angles in proteins from measurements of cross-correlated relaxation effects, J Biomol NMR, 1999, 14, 277-280.
[50] Fenwick RB, Schwieters CD, Vögeli B, Direct Investigation of Slow Correlated Dynamics in Proteins via Dipolar Interactions, JACS 2016, 138(27), 8412-21.
[51] Guntert P, Mumenthaler C, Wuthrich K, Torsion angle dynamics for NMR structure calculation with the new program DYANA, J Mol Biol 1997, 273(1), 283-298.
[52] Güntert P, Automated structure determination from NMR spectra, Eur Biophys J 2009, 38(2), 129-143.
[53] Luginbuhl P, Szyperski T, Wüthrich K, Statistical basis for the use of C-13-alpha chemical shifts in protein structure determination, J Magn Reson Ser B 1995, 109, 229–233.
[54] Vögeli B, Güntert P, Riek R, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol Phys 2013, 111(3), 437-454.

8. A structural ensemble of the enzyme cyclophilin reveals an orchestrated mode of action at atomic resolution

This chapter is based on the following publication:

Celestine N. Chi, Beat Vögeli, Stefan Bibow, **Dean Strotz**, Julien Orts, Peter Güntert, Roland Riek, A Structural Ensemble of the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution, Angew Chem Int Ed 2015, doi: 10.1002/anie.201503698. Beat Vögeli and Roland Riek advised on the underlying work of the paper and on the writing of the manuscript. Celestine N. Chi recorded all experimental data and did the analysis. Stefan Bibow advised on the recording of RDC experimental data and its analysis. Dean Strotz wrote the software (eNORA2) for and advised on the analysis of the eNOE data and participated in the writing of the manuscript.

8.1. Introduction

The catalytic mechanisms of enzymes are believed to rely on a dynamic interplay between well-arranged structural states [1]. The magnitude of the conformational change may cover a large range in both space and time. The most relevant time scale for protein action is believed to be in the μs–ms range. Evidence of such dynamics has been found for the well-studied human cyclophilin A [1a-c, 2], a peptidylprolyl cis–trans isomerase. For cyclophilin A, NMR relaxation experiments revealed ms motions both during catalysis and in the apo state that can be interpreted as a two-state interconversion process [2c-e, 4]. In combination with room temperature X-ray crystallography [2d] and mutagenesis studies [5], it has been suggested that the presence of a dynamic network encompassing the active site and its close neighborhood is key for activity. This finding has been complemented by a proposed mode of action of cyclophilin A, derived from molecular dynamics simulation restricted by NMR restraints in combination with density functional theory calculations [6]. The calculations indicate that cyclophilin A acts through an electrostatic handle mechanism at the carbonyl of the residue preceding the proline in the substrate.

Traditional approaches dedicated towards elucidating such conformational dynamics are Carr-Purcell-Meiboom-Gill (CPMG) type NMR relaxation rate measurements, which capture timescales of protein motions in the μs–ms time range, and from fluorescence-

based techniques, which can measure rates in the μs–s time window [7]. However, it is difficult to represent the spatial sampling of these slow motions [1d]. New methodologies combining NMR probes with molecular dynamics simulations are being advanced to unravel this problem [1d, 6]. Recently, we introduced another concept that makes use of exact Nuclear Overhauser Effect (NOE)-derived distance restraints [8]. In order to obtain a plausible description of the various sub-states of cyclophilin A at atomic resolution, we employed an ensemble structure calculation with the use of eNOEs and residual dipolar couplings (RDCs).

8.2. Results and discussion

8.2.1. The experimental data is well explained with two distinct states

The structural ensembles were calculated following an established protocol [8] using an updated version of the software eNORA [9] and CYANA [10]. Ensemble structure calculations were performed with a total of 3629 eNOE-based distance restraints, 396 H-N RDCs derived from four alignment media, 279 scalar couplings, and 128 angle restraints from $^{13}C\alpha$ chemical shifts (Table S1 in appendix 10.5.). The CYANA target function, the weighted sum of all squared violations of the experimental restraints, acts as a measure of the quality of the calculated structures. It drops significantly from one state to two states and levels off after three states (Figure 1c). This observation indicates that, in contrast to the single-state structure, multistate ensembles describe the experimental data well (Figure 1c and Table S1 in appendix 10.5.). In order to test for self-consistency of the experimental data, a cross validation test was performed with a jackknife procedure that repeats structure calculation ten times with 10 % of the experimental input data deleted at random such that each distance restraint is omitted exactly once. The back-calculated target function of the omitted data then represents the entire data set. The decrease in this target function for higher-state ensembles (Figure 1c) further confirms that the experimental data describe two or more states well. Comparable cross-validations were also done with the RDCs and the $^{3}J_{HNHA}$ couplings (Figure 1d and 1e). Again, a significant and a moderate drop in the target function values for the $^{3}J_{HNHA}$ couplings and the RDCs is observed when increasing the number of states from one to two.

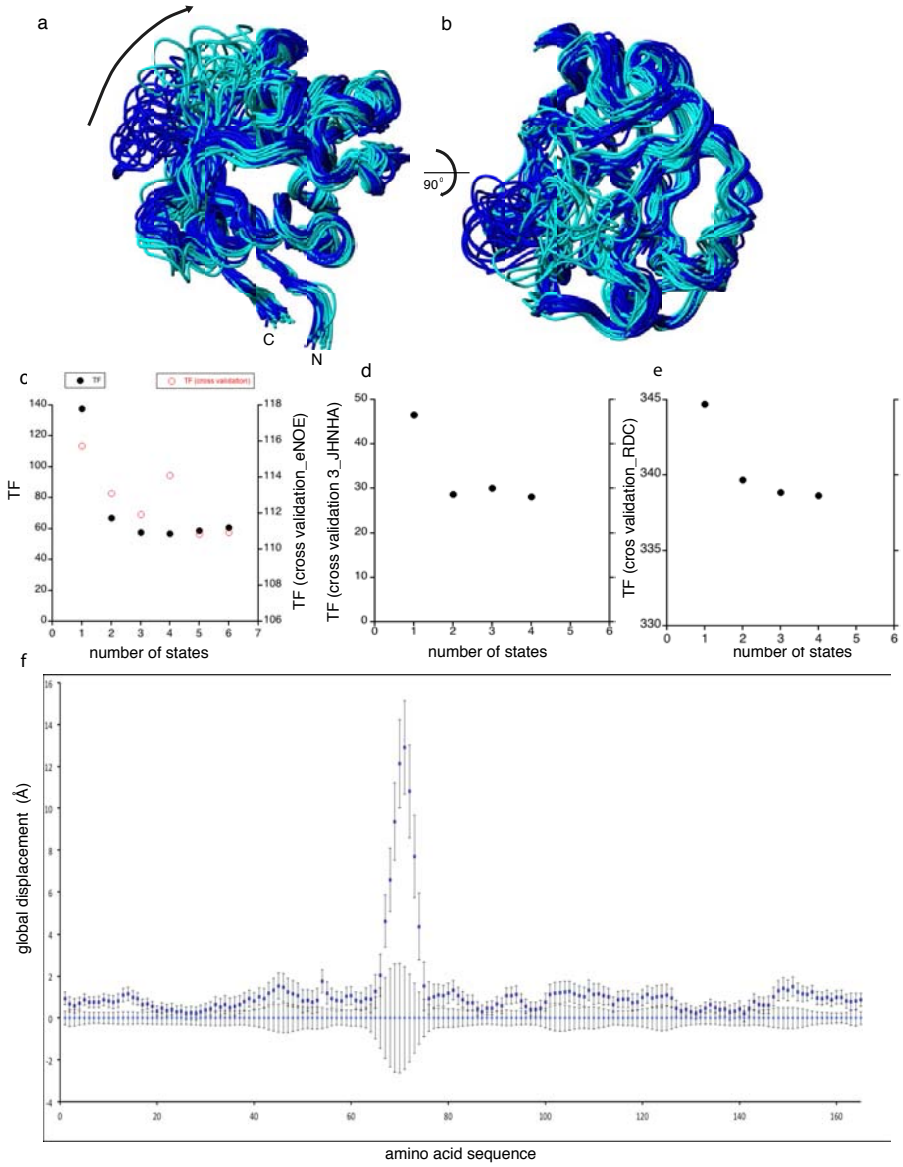


Figure 1. Structural ensemble of cyclophilin A in its apo form highlighting the presence of two distinct states. (a) and (b) Backbone trace of 20 structural ensembles each representing two different states. States are color-coded as open (blue) and closed (cyan). Two distinct states are observed throughout most of the structure. The orientation shown in (b) is a 90˚ rotation from the top as indicated. (c) Dependence on the number of states of the CYANA target function (TF) and overall TF from the jackknife-type cross validation are shown. The CYANA TF value drops from one to two states and levels off thereafter, indicating the presence of two states as shown in (a) and (b). There appears to be an outlier for the four states ensemble of unknown origin. d) and e) Dependence on number of states of the CYANA TF on jackknife-type cross-validation of RDCs data and $^{3}J_{HNHA}$ couplings are shown. The TF values drop from one to two states and levels off thereafter, again suggesting that the two-states representation of the ensemble reflects the experimental data well. (f) The backbone global displacement (r.m.s.d.) between the mean structures of the two sub-states is plotted against the amino acid sequence. The error bars (grey) are the root mean square deviations (r.m.s.d.) of the individual sub-states of backbone atoms for the residue of interest and its neighboring residues. States are color-coded as open (blue) and closed (cyan). The r.m.s.d. was calculated with MolMol [11].

As a representative for the following discussion, the two-state ensemble described by a structural bundle of 2 x 20 conformers (PDB ID: 2n0t, Figure 1a, b) is used in order to prevent over-fitting of the data (i.e., minimalizing the number of degrees of freedom)

and because the same two states are also observed in the higher-state structure calculations (compare Figure 1 and 2 with Figures S3 and S9 in appendix 10.5.). Note that the latter finding shows that the additional degrees of freedom given for higher-state structure calclulations are not used, excluding the potential issue that the presence of two states might be an artifact of constraining the number of states to two. Moreover, the deletion of 120 distance restraints violated in the single state structure calculation (Table S3 in appendix 10.5.) resulted in a loss of two distinct structural states in the two-state structure calculation (Figure S5b in appendix 10.5.).

8.2.2. Correlated patterns are observed throughout the two-states ensemble

Inspection of the Cyclophilin A two-state ensemble reveals several remarkable features. Most prominently, the ligand-binding loop comprising residues 64–74 samples two spatially well-separated states (Figure 1a, b, f, Figure 2 and Table S2 in appendix 10.5.). The two states are referred to as "open" and "closed" (Figure 1 and Figure 2) because the closed state (cyan) is slightly more compact (average total surface area: 8937 $Å^2$ calculated with the program MolMol [11]) compared to the open state (blue; average total surface area: 9211 $Å^2$; Figure 1). The two states are also distinct at the active site and in surrounding regions, thus indicating long-range correlations (Figure 1–3 and Table S2 in appendix 10.5.). In the backbone, the two states are considerably different for residues 9–16, 34–42, 54– 57, 64–78, 89–94, 101–107, and 118–127. The presence of two distinct interchanging states at these locations concurs with documented slow conformational exchange measured by NMR relaxation data both in the apo state and during catalysis (Figure S1 in appendix 10.5.) [2c, e].

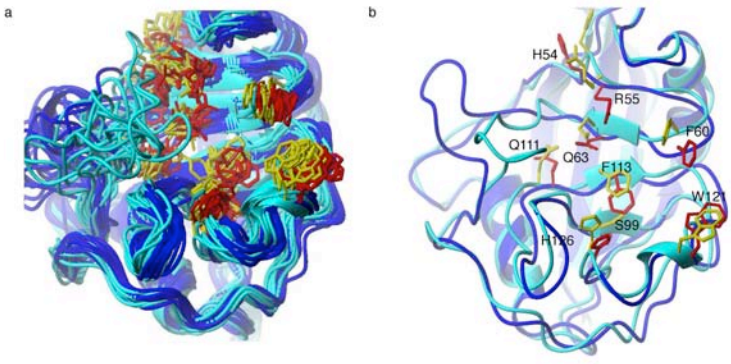


Figure 2. The two-states ensemble of the active site residues of cyclophilin. a) A ribbon representation of the 2 x 20 structural ensembles is shown color coded individually for the two conformational sub-states: the closed state is color coded with cyan for the backbone and yellow for the side chains of the active site, while the open state is color coded with blue for the backbone and red for the side chains in the active site, respectively. b) A single representative of (a) having the side chains labeled according to the single letter code. The lowest energy two-states conformers were selected.

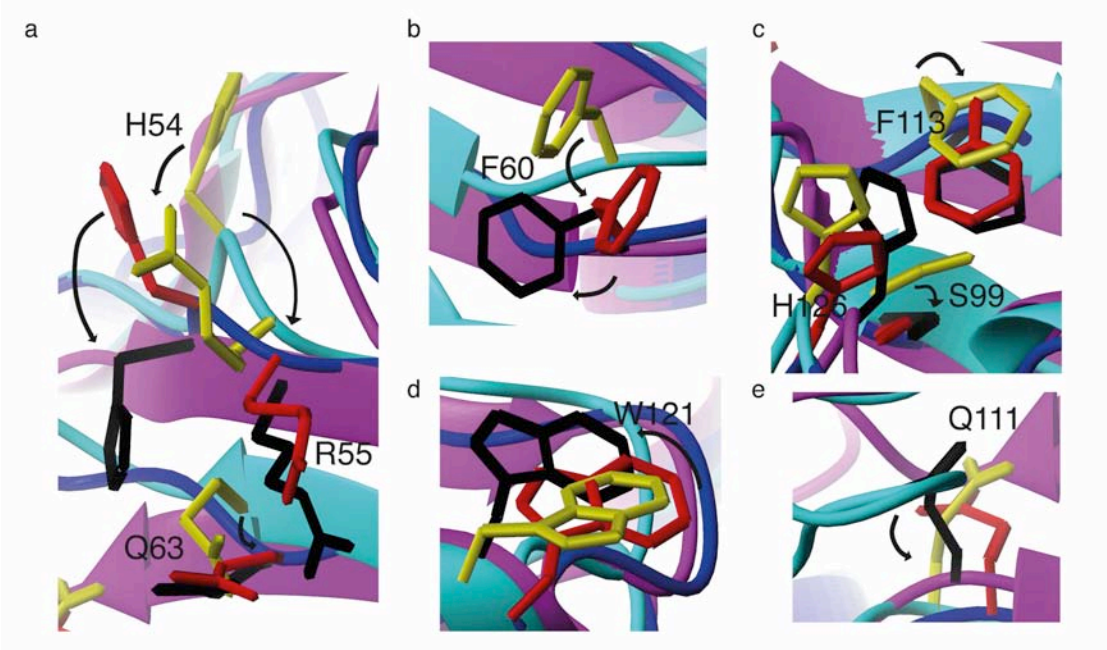


Figure 3. Proposed mechanism of action of cyclophilin at atomic resolution. The X-ray structure from Cyclophilin A in complex with the HIV-1 capsid protein (pdb: 1ak4) was superimposed with the presented two-states ensemble highlighting that the open state fits well the ligand-bound state. The closed state is color coded with cyan for the backbone ribbon and yellow for the side chains, the open state is color coded with blue for the backbone ribbon and red for the side chains, and the X-ray structure is color coded with magenta for the backbone ribbon and black for the side chains, respectively. (a)-(e) show individual close ups of the super-positions. The side chains of interest are labeled with a single amino acid residue code. The potential modes of action for catalysis of the individual residues are highlighted by arrows.

8.2.3. Ligand binding occurs in the open state via conformational sampling

In concert with the backbone, many side chains show two distinct states. This is primarily caused by a propagation of different peptide plane orientations into the side chains since the side-chain rotamer angles between the two states are similar (Figure S7 in appendix 10.5.). Of particular interest is that also the side chains at the active site show two distinct states, that is, the side chains H54, R55, F60, Q63, S99, Q111, F113, W121 and H126 (Figure 2, Figure 3, and Table S2 in appendix 10.5.). These include side chains important for ligand binding (i.e., H54, F60, Q111, F113, W121 and H126 [12]), as well as side chains essential for activity (i.e., Q63, S99, F113, and R55 [5]). The presented two-state correlation of the latter side chains resolves the proposed activity-related dynamic network at atomic resolution [2d], which guides the charged side chain of R55 into position to create an electrostatic potential that acts on the carbonyl group of the proline-preceding residue of the ligand [6]. Interestingly, the orientations of the discussed side chains of the open state closely match those of the crystal structure of cyclophilin A in complex with the HIV-1 capsid protein (Figure 3; PDB ID: 1ak4 [12].), thus highlighting that cyclophilin A closely samples the ligand-bound state at the active

site in its apo state. This finding implies that the mechanism of conformational sampling [13] is key for its ligand recognition as suggested earlier [2c, e]. By contrast, the ligand-binding loop comprising residues 64–74 is not in its ligand-bound state. Thus, an induced fit mechanism [7, 14] must prevail for this part of the protein upon complex formation (Figure S2 in appendix 10.5.).

8.2.4. The closed state is 15-40% populated

In contrast to the open state, the closed state appears to be far from the active ligand-bound conformation. The loop residues 64–74 penetrate the space the bound ligand would cover (Figure S2 in appendix 10.5.) and the side chains of R55, Q63, F113, and S99 are not well positioned for activity (Figure 3 and Figure S2 in appendix 10.5.). Overall, large conformational differences between the two states are observed for residues 64–74, the side chains of H54, F60, S99, Q111, W121, and H126, and the side-chain conformation and configurations of R55 (Figure 2 and Figure 3). Furthermore, subtle dissimilarities in the side-chain conformations and configurations are apparent in 60–70% of all the side chains that are not solvent exposed, thus indicating that the differences between the two states are spread over almost the entire protein structure (Figure S4 in appendix 10.5.). Interestingly, each of the individual states of the six-states ensemble can be grouped to one of the two sub-states determined in the two-states ensemble (Figures S3b and S9 in appendix 10.5.). This finding allows a rough estimate of the population of the two states by counting the number of each of the two states in the six-states structure calculation (Figure S3b in appendix 10.5.). With this analysis, the closed state appears to be less populated (between 15–40%) than the open state, which is in line with previous findings based on relaxation measurements (i.e., ca. 15 % population at 10˚C [2c]). In summary, from the observation that cyclophilin A interchanges on the ms timescale, as deduced from a single set of cross-peaks in the spectra [2e], relaxation measurements [2e], and the presented two-states ensemble, the presence of large-scale concerted motion between two states is suggested for the apo form of cyclophilin A.

8.2.5. Concerted movements in the side-chains are associated with ligand binding

Based on the two-states ensemble presented herein, as well as other data [5, 6, 12] it is suggested that cyclophilin A interchanges between a closed and an open sub-state on the millisecond timescale both in its apo state and during catalysis. Owing to the intrinsic

exchange between the two states, there is a time window during which the ligand can bind to the open state. In detail, the loop opens up accompanied by an associated outward movement of H54 and an inward movement of R55 to its place for catalysis. F113 turns about 60˚ degrees downwards and S99 rotates 180˚ downwards (Figure3a, c). F60, W121, and H126 flip outwards, while Q63 moves upward to keep R55 in place (Figure 3a, b, d, e). These rearrangements, accompanied by many subtle changes throughout most of the protein structure, create the ligand-binding cavity, position the active site residue R55 for activity, and prepare the active site for ligand binding. Upon binding of the ligand, only small, subtle side-chain rearrangements at the active site occur (Figure 3). The switch back to the closed state requires release of the peptide by a mode of action reminiscent of a catapult, with the ligand-binding loop being the handle.

8.2.6. Mutagenesis studies

In order to find further support for the role of the loop motion in protein activity and the presence of two states, we attempted to lock the loop into the open state by replacing the Gly residues of the N-terminal hinge in the peptide-binding loop (i.e. Gly64-Gly65) with the more conformationally restricted residue Ala. It is thus expected that the double mutant is close to a ligand-bound state (open state) and concomitantly should show increased activity. Indeed, the well-folded double mutant (Figure S8 in appendix 10.5.) shows an increased affinity for both peptide ligands studied when compared to wild-type cyclophilin A (Figure 4a, b). Furthermore, the mutation introduces chemical shifts of ^{15}N-^{1}H moieties in the same direction as the addition of ligand to a sample of wild-type cyclophilin A (Figure 4c–h), thus indicating that the mutant is indeed in a more open state than free wild-type enzyme.

8.3. Conclusions

The presented ensemble structure calculation of the apo state of the enzyme cyclophilin, complemented by mutagenesis and affinity measurements, reveals a long-range well- orchestrated conformational interchange between substrates important for its catalytic activity and highlights a synergistic induced fit and conformational sampling mechanism of action. The complexity unraveled reflects the adaptation and optimization power of evolution, as well as the beauty of these types of biological machineries, which are composed of several hundreds of atoms moving in concert.

8.4. References

[1] a) Henzler-Wildman KA, Kern D, Dynamic personalities of proteins, Nature 2007, 450, 964–972;
b) Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis, Nature, 2007, 450, 913 – 916;
c) Wolf-Watz M, Thai V, Henzler-Wildman KA, Hadjipavlou G, Eisenmesser ZE, Kern D, Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair, Nat Struct Mol Biol 2004, 11, 945–949;
d) Lange OF, Lakomek N-A, Fares C, Schroder GF, Walter KFA, Becker S, Meiler J, Grubmuller JH, Griesinger C, B. de Groot BL, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, Science 2008, 320, 1471 – 1475.

[2] a) Kern D, Eisenmesser EZ, Wolf-Watz M, Enzyme Dynamics During Catalysis Measured by NMR Spectroscopy, Methods in Enzymology, Vol. 394 (Ed.: L. J. Thomas), Academic Press, San Diego, 2005, pp. 507 – 524;
b) Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D, Intrinsic motions along an enzymatic reaction trajectory, Nature 2007, 450, 838 – 844;
c) Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D, Intrinsic dynamics of an enzyme underlies catalysis, Nature 2005, 438, 117-121;
d) Fraser JS, Kern D, Alber T, Hidden alternative structures of proline isomerase essential for catalysis, Nature 2009, 462, 669–673;
e) Eisenmesser EZ, Bosco DA, Akke M, Kern D, Enzyme dynamics during catalysis, Science 2002, 295, 1520 – 1523.

[3] a) Handschumacher RE, Harding MW, Rice J, Drugge RJ, Speicher DW, Cyclophilin: a specific cytosolic binding protein for cyclosporin A, Science 1984, 226, 544 – 547;
b) de Villiers EM, Wagner D, Schneider A, Wesch H, Miklaw H, Wahrendorf J, Papendick U, zur Hausen H, Human papillomavirus infections in women with and without abnormal cervical cytology, Lancet 1987, 330, 703 – 706;
c) Fischer G, Wittmann-Liebold B, Lang K, Kiefhaber T, Schmid FX, Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins, Nature 1989, 337, 476 – 478;
d) Clubb RT, Ferguson SB, Walsh CT, Wagner G, Three-dimensional solution structure of Escherichia coli periplasmic cyclophilin, Biochemistry 1994, 33, 2761 – 2772.

[4] Finkelstein AV, Ptitsyn OB, ProteinPhysics, AcademicPress, Amsterdam, 2002.

[5] Zydowsky LD, Etzkorn FA, Chang HY, Ferguson SB, Stolz LA, Ho SI, Walsh CT, Active site mutants of human cyclophilin A separate peptidyl-prolyl isomerase activity from cyclosporin A binding and calcineurin inhibition, Protein Sci 1992, 1, 1092 – 1099.

[6] Camilloni C, Sahakyan AB, Holliday MJ, Isern NG, Zhang F, Eisenmesser EZ, Vendruscolo M, Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism, Proc Natl Acad Sci USA 2014, 111, 10203 – 10208.

[7] a) Chi CN, Bach A, Engstrom A, Wang H, Stromgaard K, Gianni S, Jemth P, A sequential binding mechanism in a PDZ domain, Biochemistry 2009, 48, 7089 – 7097;

b) Chi CN, Elfstrom L, Shi Y, Snäll T, Engström A, Jemth P, Reassessing a sparse energetic network within a single protein domain, Proc Natl Acad Sci USA 2008, 105, 4679 – 4684.

[8] a) Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19, 1053–1057;
b) Vögeli B, The nuclear Overhauser effect from a quantitative perspective, Prog. Nucl Magn Reson Spectrosc 2014, 78, 1 – 46.

[9] Orts J, Vögeli B, Riek R, J. Chem., Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs, Theory Comput. 2012, 8, 3483 – 3492.

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

[11] Koradi R, Billeter M, Wüthrich K, MOLMOL: a program for display and analysis of macromolecular structures, J Mol Graphics 1996, 14, 29–32.

[12] Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, Hill CP, Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid, Cell 1996, 87, 1285 – 1294.

[13] Monod J, Wyman J, Changeux JP, On the nature of allosteric transitions: A Plausible model, J Mol Biol 1965, 12, 88 – 118.

[14] Koshland DE Jr., Nemethy G, Filmer D, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 1966, 5, 365 – 385.

[15] Chi CN, Engström A, Gianni S, Larsson M, Jemth P, Two conserved residues govern the salt and pH dependencies of the binding reaction of a PDZ domain, J Biol Chem 2006, 281, 36811 – 36818.

9. Conclusions and Outlook

This chapter is **in part** based on the following publications:

Beat Vögeli, Julien Orts, **Dean Strotz**, Peter Güntert, Roland Riek, Discrete Three-Dimensional Representation of Macromolecular Motion from eNOE-based Ensemble Calculation, CHIMIA 2012, doi: 10.2533/chimia.2012.787. All Authors contributed to writing of the manuscript.

Beat Vögeli, Julien Orts, **Dean Strotz**, Celestine Chi, Martina Minges, Marielle Aulikki Wälti, Peter Güntert, Roland Riek, Towards a true protein movie: A perspective on the potential impact of the ensemble-based structure determination using exact NOEs, J Magn Reson 2014, doi: 10.1016/j.jmr.2013.11.016. All Authors contributed to writing of the manuscript.

Simon Olsson, **Dean Strotz**, Beat Vögeli, Roland Riek, Andrea Cavalli, The dynamic basis for signal propagation in human Pin1-WW, Structure 2016, doi: 10.1016/j.str.2016.06.013. Simon Olsson did the programming for the simulations and its analysis. Dean Strotz recorded all experimental data, did the analysis of the experimental data, prepared figures concerning the experimental data and participated in the writing of the manuscript.

9.1.1. Analysis of experimental buildup data

In this thesis we presented the current state of development in methods pertaining to the analysis of distance restraint information from NOE measurements. We analyzed the experimental accuracy of our eNOE data, dissecting in detail the experimental setup to measure the data and the errors contained therein [1]. In our analysis of the experimental accuracy of eNOE data we did however not discuss the complex time dependence of the NOE. In the present translation from eNOE rates to distances it is assumed that fast motion (i.e. faster than the rotational correlation time of the protein studied) does not perturb the NOE (i.e. $S_{KL}^{\text{fast}\,2} = 1$ in equations 1-4, chapter 1). The rationale of this assumption is based on the finding that for ^{1}H-^{1}H NOEs between two H-X moieties (X being a heavy atom, ^{15}N or ^{13}C) with order parameters larger than 0.5 the NOE is indeed

not significantly affected by fast motions. This low impact is attributed to the significant cancellelations between the fast motion-induced angular and distance effects on the NOE rate [2]. Similar conclusions have been drawn from molecular dynamics studies, but they also reveal that a few percent of the NOEs may violate the assumption considerably yielding distances with an error of more than 10 % [3, 4]. A detailed analysis showed that these critical NOEs are from side chain atoms of a phenylalanine [3]. In a second study [4], half of the critical NOEs involve side-chain arginine and lysine protons and not surprisingly, the most extreme averaging involves dihedral transitions. Conclusively, NOEs involving atoms located at the far end of very long and highly flexible side chains are at the moment used with caution. If a combination of exact NOESY and ROESY measurements [3] or the measurements of local order parameters of each ^{1}H-^{15}N and ^{1}H-^{13}C moiety by ^{15}N- and ^{13}C-relaxation measurements, respectively, and ^{13}C-^{13}C NOESY [2] may resolve these potential sources of error remains to be demonstrated.

Furter improvements for the eNOE determination may be done by (i) fitting the experimental data differently, (ii) incorporating anisotropic tumbling and (iii) using lower limits for missing resonances in structure calculation. For a new implementation of eNORA2 [5] adding up the intensities of bidirectional buildups before fitting has the potential to yield even more precise distance restraints than currently available when averaging the calculated cross relaxation rates post-fitting. However, the mathematics involved in such a fitting procedure are of non trivial nature and would have to be established. Second, currently in eNORA2, the analysis and spin diffusion correction is implemented for isotropically tumbling molecules. If access to the anisotropic diffusion tensors is available, a possible avenue to higher precision in eNOE analysis for some molecules is to enable anisotropic analysis. In further combination with spin specific order-parameters to include internal motion this might be beneficial for protein folding studies for instance, as unfolded states of proteins are highly likely to tumble anisotropically. Third, for improved structural resolution and analysis of coupling networks, one could expand eNORA2 by generating anti-NOEs [6] for missing peaks realized in the form of lower limit restraints. Such restraints have the potential to improve resolution of the conformational space sampling especially in regions with motion (where NOE data is more sparse) if used cautiously. Anti-NOEs would essentially act like a repulsive potential and thereby counteract the artifically introduced, but necessary symmetry restraints (keeping the states together) used in multi-states ensemble structure calculation. One has to use anti-NOEs with caution due to missing order parameters. However, since the diagonal peak can be used as an estimate of the

order parameter, defining its upper limit, if one estimates the order parameter cautiously, the calculated lower limit restraint would also be small enough as not to introduce artifacts into the multi-states ensemble structure calculation.

9.1.2. Limitations of eNOE analysis due to system size

One of the major limitations of eNOE analysis is increased peak overlap in large systems. The problem is particularly pressing with respect to the diagonal peaks, which must be analyzed in order to obtain the magnetization at the onset of NOESY mixing and the auto-relaxation rate. With the generic normalized eNOE we presented an extension of the eNOE data set by enabling the analysis of NOE data in crowded spectra with overlaping resonance signals resulting from large proteins. Another approach reducing spectral overlap as well as reducing spin diffusion was implemented by enabling the analysis of deuterated samples with various labeling schemes using the three-spin approach. Labeling schemes reduce spectral overlap very efficiently and selectively, putting the analysis of such large proteins as the proteasome subunit alpha7alpha7 (360kDa [7]) within reach. Recording NOESY spectra with additional dimensions, such as 4D HMQC-NOESY-HMQC [8] would be another method with great potential to resolve diagonal peak overlap by separating the proton shifts with two heavy atom resonances [8]. The main issue of 4D NOESY when applied to large protein systems is with the transverse relaxation that increases proportionally with molecular size causing losses of magnetization in HMQC elements, while the NOE transfer relies on longitudinal magnetization and is proportionally more efficient with increasing overal tumbling times.

9.1.3 Limitations due to available measurement time

The measurement time for NOESY buildups increases on average three to four times over a single NOESY experiment due to the measurement of a minimum of three to four different mixing times. Whether sparse sampling techniques [9, 10] in combination with special spectral transformation schemes (such as compressed sensing reconstruction [10] or the SCRUB algorithm [11]) may be applicable to address the time requirement effectively remains to be demonstrated because the relative cross peak and diagonal peak intensities must be preserved for the extraction of eNOE rates.

9.1.4. Structure calculation

We have illustrated the application of the protocol for ensemble based structure calculation established in the Riek group. As mentioned in the introduction, in order to avoid divergence among the structural states that is not implied by the experimental restraints, "bundling restraints", i.e. weak harmonic restraints that minimize the distances between corresponding atoms in different states [12, 13] are imposed. These bundling restraints are limiting the conformational space sampled in a somewhat artificial manner. However due to the $1/r^6$ dependence it is difficult to argue for or against unfolded states since their contribution to the eNOE would be minimal. Nevertheless, this constitutes one of the limitations of the structure calculation protocol and of the NMR measurable, the NOE. Using the ensemble based structure calculation protocol, as established, one cannot obtain unfolded or partially unfolded states despite other experimental data such as CD indicating a significant population of unfolded states. If measurements are performed with systems showing considerable populations of unfoldes states, it is proposed to subtract the contribution to the eNOE of the well-folded conformers from the data obtained at unfolding conditions, giving more weight to the contribution of the unfolded conformational sampling space. Another future improvement of the method would come with the ability to calculate populations of states. This would be possible by allowing the weight of the bundling restraints to be a free variable instead of using fixed equal weights, allowing population wheighted multi-state ensembles to be calculated. Last but not least, inclusion of cross-correlation rates (CCR) data could act to better define angular sampling in analogy to scalar couplings or RDC data.

We applied a novel approach using replica exchange (MD) simulation with chemical shift data through maximum entropy reweighing (maximum conformation space required to fullfil the data)[14]. The obtained ensembles correlated well with exact nuclear Overhauser enhancement (eNOE) measurements used for validation. The chosen approach further illustrates the sensitivity of eNOE data to states with low populations. We have thereby shown how the eNOE experiments may directly complement other experimental data traditionally associated with the detection of transient states, such as paramagnetic relaxation enhancement (PRE) and relaxation dispersion. The eNOEs are particularly sensitive to shorter distances than what is typically amenable for PRE analysis. The eNOE distance information may be complementary to temporal information (exchange rates) derived from relaxation dispersion (and less reliably, the population information). Relaxation dispersion data may therefore be used to further validate our models. Using the full set of heteronuclear shifts for reweighing would

further enhance the chosen approach of maximum entropy by chemical shifts. However, ultimately we would like to use eNOE data to directly drive the simulations through the maximum entropy approach. The inclusion into structure calculations of other NMR measurables having distinct time dependence and sensitivity to the sampled conformational space is of great interest. Therefore, we would like to include also RDC and CCR in our structure calculations using the maximum entropy approach.

9.1.5. Cross-validation of eNOEs and incorrect assignments

NMR structure determination is a multi-probe method that requires hundreds to thousands of experimental restraints for a reliable structure determination. Among the large number of restraints, there will in general, be some incorrect ones. Wrong restraints may originate, for example, from wrong assignments or/and wrong integration caused by peak overlap. In a standard structure determination such restraints are usually exposed by distance restraint violations in the resulting 3D structure. In an ensemble-based structure calculation, on the other hand, they may be hidden and even lead to the appearance of an additional structural state. A cross-validation procedure for the ensemble-based structure calculation using eNOEs is therefore mandatory. In our recent work, this test consists of the arbitrary deletion of 10% of all the eNOE-derived distance restraints and a subsequent consistency check by evaluating their violations with a structural ensemble obtained from the remaining 90% of the restraints (i.e. contribution to the target function) [15]. Although this approach is sound, more profound studies should be designed for the identification of wrong eNOEs and examine the cause for the remaining structure violations further by, for example, Bayesian inference to derive probability distributions [16, 17].

9.1.6. The information content of calculated structures and their analysis

We used the protocol for ensemble based structure calculation to illustrate the potential impact on the comprehensive elucidation of the action of biomolecules at atomic resolution using our high-resolution solution state NMR data. We analyzed the interface between the Pin1-WW domain and the allosteric coupling to the Pin1-PPIase. We also demonstrated the relevance of the protocol in determining the presence of an open and closed state in Cylophilin A infering conformational sampling of Cyclophilin A and a preorganization for catalysis. Furthermore we illustrated an allosteric coupling

network connecting the ligand-binding loop to the active site. We used the replica exchange simulations as basis for a detailed analysis yielding an understanding of the multi-functionality of the Pin1 protein and to study information transduction in human Pin1-WW.

One of the major challenges we found is the accurate representation, interpretation and analysis of the data. In the replica exchange simulations, an extension of methods for the analysis in a statistical manner was applied to determine atomistic thermodynamic models using Markovian modeling [14]. While the approach readily allows for comparison with many equilibrium experiments, it does not, at this stage, allow for dissection of chemical exchange kinetics. However, a future extension of this may be possible by complementing this type of simulations with regular MD simulations in an appropriate statistical framework [18-20] as suggested in Simon et al. 2016 [14].

9.1.7. Biological relevance of the research aim

To increase the biological relevance of our model system Pin1:WW wild type should be used. In an extension to the data already analyzed the structure of the complex in isolation and in the full length Pin1 should be solved. The interdomain interaction interface can be analyzed on a structural basis in more detailed than so far. Furthermore and excitingly, there are ligands previously determined to increase interaction between the two domains and others that weaken this interaction [23, 24]. We have already preliminary results using other ligands than xenopus pCdc25C that are in support of this avenue of research (unpublished).

9.5. References

[1] Strotz D, Orts J, Minges M, Vögeli B, The experimental accuracy of the uni-directional exact NOE, J Magn Reson 2015, doi: 10.1016/j.jmr.2015.07.007.

[2] Leitz D, Vögeli B, Greenwald J, Riek R, Temperature Dependence of 1HN-1HN Distances in Ubiquitin As Studied by Exact Measurements of NOEs, J Phys Chem B 2011, 115, 7648-7660.

[3] Brüschweiler R, Roux B, Blackledge M, Griesinger C, Karplus M, Ernst RR, Influence of Rapid Intramolecular Motion on NMR Cross-Relaxation Rates. A Molecular Dynamics Study of Antamanide in Solution, J Am Chem Soc 1992, 114, 2289-2302.

[4] Post CB, Internal Motional Averaging and Three-dimensional Structure Determination by Nuclear Magnetic Resonance, J Mol Biol 1992, 224, 1087-1101.

[5] Strotz D, Orts J, Chi CN, Riek R, Vögeli B, The eNORA2 exact NOE analysis program (to be submitted).

[6] Brüschweiler R, Blackledge M, Ernst RR, Multi-conformational peptide dynamics derived from NMR data: A new search algorithm and its application to anamanide, J Biomol NMR 1991, 1(1), 3-11.

[7] Sprangers R, Kayk L, Quantitative dynamics and binding studies of the 20S proteasome by NMR, Nature 2007, 445, 618-622.

[8] Kay LE, Clore GM, Bax A, Gronenborn AM, Four-Dimensional Heteronuclear Triple-Resonance NMR Spectroscopy of Interleukin-1b in Solution, Science 1990, 249, 411-414.

[9] Bostock MJ, Holland DJ, Nietlispach D, Compressed sensing reconstruction of undersampled 3D NOESY spectra: application to large membrane proteins, J Biomol NMR 2012, 54 15-32.

[10] Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY, High-Resolution Four-Dimensional 1H-13C NOE Spectraoscopy using Methyl-TROSY, Sparse Data Acquisition, and Multtdimensional Decomposotion, J Am Chem Soc 2005, 127, 2767-2775.

[11] Coggins BE, Werner-Allen JW, Yan A, Zhou P, Rapid Protein Global Fold Determination Using Ultrasparse Sampling, High-Dynamic Range Artifact Suppression, and Time-Shared NOESY, J Am Chem Soc 2012, 134, 18619-18630.

[12] Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19, 1053-1057.

[13] Clore GM, Schwieters CD, How much backbone motion in ubiquitin is required to acount for dipolar coupling data measured in mutliple alignment media as assessed by independent cross-validation?, J Am Chem Soc 2004, 126, 2923-2938.

[14] Olsson S, Strotz D, Vögeli B, Riek R, Cavalli A, The dynamic basis for signal propagation in human Pin1-WW, Structure 2016, pii: S0969-2126(16)30159-9.

[15] Brünger AT, Clore GM, Gronenborn AM, Saffrich R, Nilges M, Assessing the Quality of Solution Nuclear Magnetic Resonance Structures by Complete Cross-Validation, Science 1993, 261, 328-331.

[16] Rieping W, Habeck M, Nilges M, Inferential Structure Determination, Science 2005, 309, 303-306.

[17] Olsson S, Frellsen J, Boomsma W, Mardia KV, Hamelryck T, Inference of structure ensembles of flexible Biomolecules from sparse, averaged data, PLoS ONE 2013, online journal.

[18] Mey ASJS, Wu H, Noé F, xTRAM: estimating equilibrium expectations from time-correlated simulation data at multiple thermodynamic states, Phys Rev 2014, X 4, 041018.
[19] Wu H, Mey ASJS, Rosta E, Noé F, Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states, J Chem Phys 2014, 141, 214106.
[20] Wu H, Paul F, Wehmeyer C, Noé F, Multiensemble Markov models of molecular thermodynamics and kinetics, Proc Natl Acad Sci USA 2016, 113, E3221–E3230.
[21] Pande VS, Beauchamp K, Bowman GR, Everything you wanted to know about Markov State Models but were afraid to ask, Methods 2010, 52(1), 99-105.
[22] Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, Proc Natl Acad Sci USA 2009, 106(45), 19011-6.
[23] Wilson KA, Bouchard J, Peng J, Interdomain interactions support interdomain communication in human Pin1, Biochemistry 2013, 52, 6968-81.
[24] Peng J, Investigating Dynamic Interdomain Allostery in Pin1, J Biophys Rev 2015, 7, 239-249.

10. Appendix

10.1. Appendix to chapter 2: The experimental accuracy of the uni-directional exact NOE

The model system

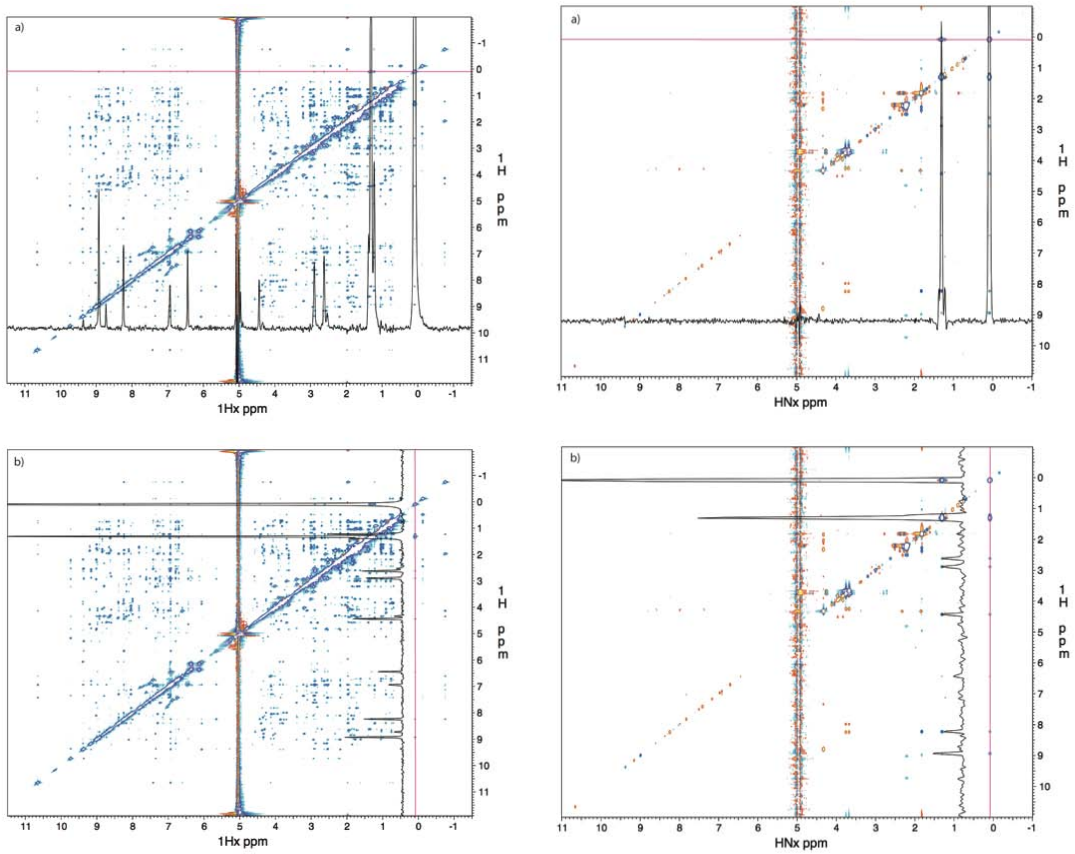


Figure S1: 2D (left panel) and 3D NOESY spectra (right panel) of the WW domain of human Pin1. The mixing times were 80 (2D, 900 MHz) and 60 ms (3D, 700 MHz). The plane in the 3D NOESY was taken at 34.452 ppm (when labeled with the ^{13}C frequency). Representative slices were taken through a well-separated diagonal peak along both ^{1}H dimensions. Note that no baseline rolls are present.

Simulation of $\alpha_{x,i}^{LW}\left(proc,J,R_2\right)$

For protons, R_2 was calculated by summing up the individual contributions from dipolar interactions with atoms within a defined distance r_{ik} i.e. 6.5 Å (atom clouds) as such [30]:

$$R_{2i} = \sum_{i \neq k} \frac{b}{8r_{ik}^6}\left(5J(0) + 9J(\omega) + 6J(2\omega)\right) \tag{S1}$$

$$b = \left(\frac{\mu_0}{4\pi}\right)^2 \hbar^2 \gamma_H^4 \tag{S2}$$

$$J(\omega) = S^2\left(\frac{t_c}{1+\left(\omega t_c\right)^2}\right) \tag{S3}$$

with μ_0 the permeability of free space, $\hbar$ the reduced Planck constant and γ the gyromagnetic ratio of the nucleus. R_2 is calculated for all atoms then individually averaged over all models in the structure obtained from the eNOE distance limits and scalar couplings. Contributions from intra-methylene and methyl are excluded, and a correction for mobility was applied using $S^2 = 0.8$ (here, we assume a correlation time for internal motion of 0 ps). Chemical shift anisotropy (CSA) was not accounted for as it is very small for protons.

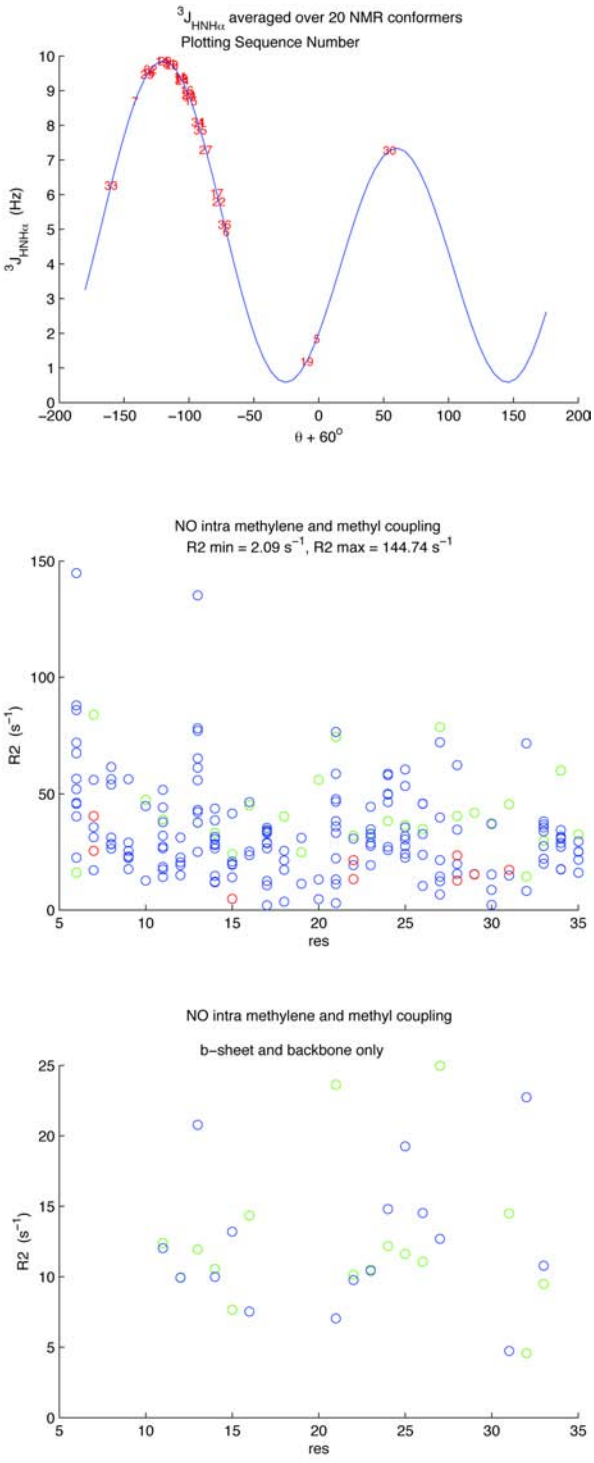


Figure S2: a) $^3J_{HN,H\alpha}$ coupling simulation. b) Simulation of R_2 for all protons. The color code is H^N, green; methylene, blue; methyl, red. The values are not corrected for motion ($S^2 = 0.8$). c) Simulation of R_2 for backbone protons in the β-sheet. The simulations define the approximate range of R_2 that needs to be considered. The values are not corrected for motion ($S^2 = 0.8$).

Table S1: Measured $^3J_{HN,H\alpha}$ at 5ºC

residue #	$^3J_{HN,H\alpha}$ [Hz]	error [Hz]
11	7.963	0.132
13	8.351	0.282
14	9.936	0.226
15	9.571	0.110
16	6.403	0.488
17	7.533	2.300
18	7.204	0.022
19	13.676	7.422
21	6.702	0.282
22	6.309	0.246
23	7.902	0.348
24	10.047	1.050
26	9.538	0.212
27	6.072	0.236
28	8.708	0.212
29	8.837	0.216
30	6.882	0.130
31	6.732	0.188
32	9.621	0.082
33	6.584	0.402
34	7.409	0.286
35	6.745	0.180
36	3.716	0.764
38	7.990	0.060

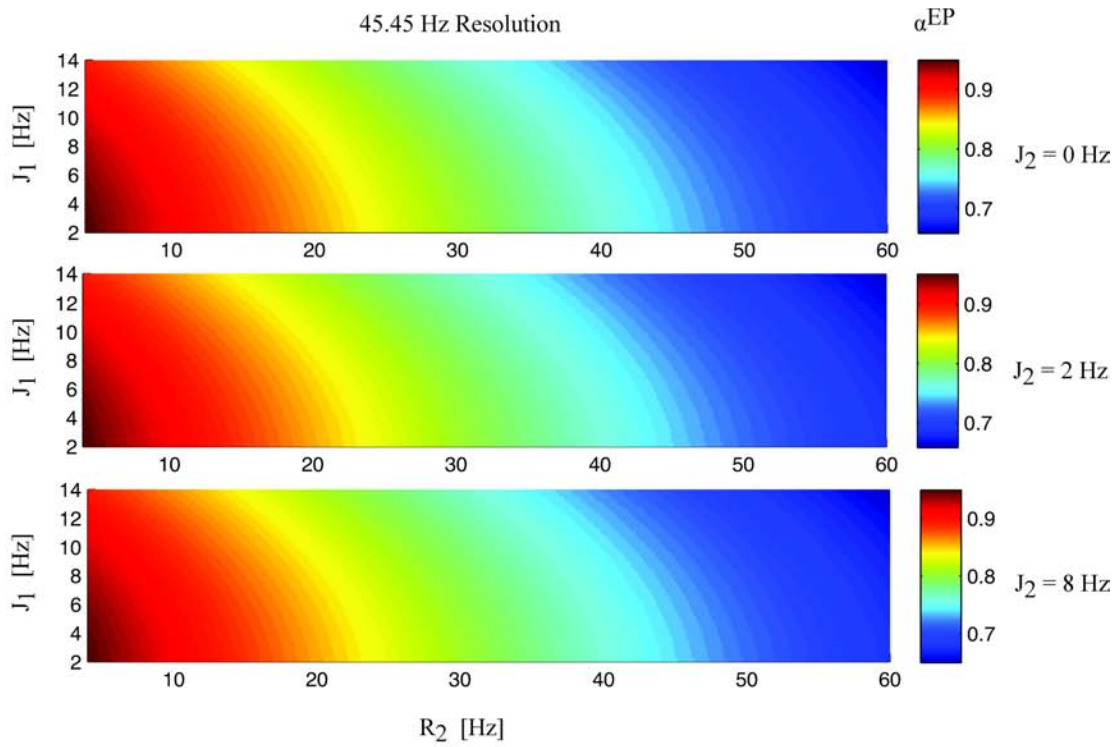


Figure S3: Simulation of the dependency of the scaling factor α^{EP} on the spectral resolution, the transverse relaxation rates constant R_2 and the amplitudes of the J couplings. Each contour plot shows α^{EP} depending on R_2 (x axis) and one J coupling (y axis). The range of R_2 is 5 – 60 s^{-1}. The number of real points is set at 400, which was doubled by zero-filling (acquisition time of 22 ms, corresponding to a resolution of 45.45 Hz). The resolution of 45.45 Hz corresponds to the experimental resolution of the indirect dimension of the ^{15}N- or ^{13}C-resolved [^{1}H, ^{1}H]-NOESY-HXQC. In the top panel, the range of the only involved J coupling (J_1) is varied between 2 and 14 s^{-1}. In the middle panel, two J couplings are simulated, one of which (J_1) has a range of 2 - 14 s^{-1}, while the second one (J_2) is held constant at 2 s^{-1}. In the bottom panel, two J couplings are simulated, one of which (J_1) has a range of 2 - 14 s^{-1}, while the second one (J_2) is held constant at 8 s^{-1}.

Error estimation for NOESY mixing period: $T_{ij}^{\mathrm{NOESY}}(\tau_{\mathrm{mix}})$

For the WW domain, 89 diagonal-peaks have been fitted and the spin type specific values obtained (Table S2).

Table S2: Spin type specific analysis of ρ_i.

	number	average ρ_i [s^{-1}]	std [s^{-1}]	min [s^{-1}]	max [s^{-1}]
all	89	4.2	1.7	1.9	10.5
H^N	19	5.1	2.1	2.0	10.5
methine	18 (17 H^a)	3.7	1.4	2.1	7.7
CH_2	34	4.7	1.4	2.5	7.7
CH_2 pseudo	11	3.2	0.9	2.3	5.1
CH_3	3	3.0	0.8	2.1	3.4
aromatics	4	2.3	0.4	1.9	2.7

Values of the WW domain at 5°C.

$$\frac{I_{ii}^{\mathrm{diag}}(\tau_{\mathrm{mix}})}{I_{ii}^{\mathrm{diag}}(0)} = 1-\rho_i\tau_{\mathrm{mix}}+\frac{1}{2}(\rho_i{}^2+\sum_{k\neq i}\sigma_{ik}^2)\tau_{\mathrm{mix}} \cong e^{-\rho_i\tau_{\mathrm{mix}}}+\frac{1}{2}(\sigma_{\mathrm{eff}}^2)\tau_{\mathrm{mix}} \qquad \text{(S4)}$$

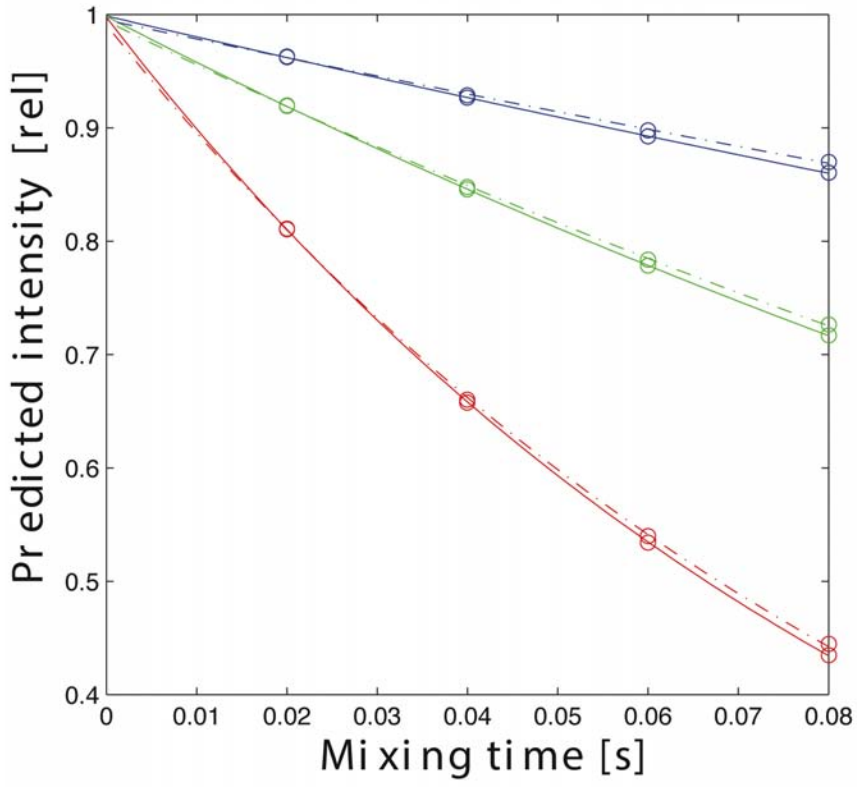


Figure S4: Error of ρ fitting due to 2nd order tailor expansion approximation according to equation S4. The color code is: ρ = 1.93, blue; for ρ = 4.21, green; for ρ =10.50, red. The line style code is: σ = 1, solid line; σ = 2, broken line.

Table S3: Error of ρ_i fitting due to the approximation shown in equation S4. The mixing times were 20, 40, 60 and 80 ms.

ρ_i (s^{-1})	σ_{eff} (s^{-1})	ρ_i fitted (s^{-1})	error (%)	abs error (s^{-1})	$I(0)$
1.93	1	1.87	3.0	0.058	0.999
4.21	1	4.15	1.6	0.068	0.999
10.50	1	10.40	1.0	0.105	0.998
1.93	2	1.70	11.9	0.230	0.995
4.2	2	3.94	6.5	0.27	0.994
10.50	2	10.08	4.0	0.419	0.991

Simulated water suppression profile

The simulations were run with nmrsim simulations (Bruker Biospin). The spin system is a uncoupled proton with variable chemical shift with no scalar coupling and T_2 set to 0.1 s [55]. The sweep width is 14 ppm with the center frequency set to 0 ppm (water frequency) for simulation purposes, with a variable chemical shift increment of ~50 Hz and 257 increments. Relaxation was set to active during acquisition only. The WATERGATE (w5) solvent suppression has a duration of 3.35 ms in total, containing two gradients of 1000 us with 100 us homospoil delay each, proton pulses adding up to 65 us and the nine d19 delays adding up to 1080 us. The Bruker script of the pulse sequence is shown below.

spin system:

```
; proton with variable chemical shift
proton a var1 t2=0.1
```

pulse sequence w5:

```
1 ze
2 d1
3 (p1 ph3):f1
  p16:gp2
  d16 pl18:f1
  p27*0.087 ph4
  d19*2
  p27*0.206 ph4
  d19*2
  p27*0.413 ph4
  d19*2
  p27*0.778 ph4
  d19*2
  p27*1.491 ph4
  d19*2
  p27*1.491 ph5
  d19*2
  p27*0.778 ph5
  d19*2
  p27*0.413 ph5
  d19*2
  p27*0.206 ph5
  d19*2
  p27*0.087 ph5
```

```
  p16:gp2
  d16
  go=2 ph0
  ihc
  lo to 2 times l3
  wr #0
exit

ph3= 0 1 2 3
ph4= 0 1 2 3
ph5= 2 3 0 1
ph0= 0 1 2 3

; pl1  : f1 channel - power level for pulse
; p1   : f1 channel -  90 degree high power pulse
; pl18: f1 channel - power level for water suppression
; p16 : homospoil/gradient pulse
; p27 : f1 channel -  90 degree pulse at pl18
; d1   : relaxation delay; 1-5 * T1
; d16 : delay for homospoil/gradient recovery
; d19 : delay for binomial water suppression
; d19 = (1/(2*d)), d = distance of next null (in Hz)
; NS : 4 * n, total number of scans: NS *TD0

; for z-only gradients:
; gpz2: 34%
```

Estimation of errors

We estimate the cumulative error obtained for the individual elements of pulse sequences using equation 1:

$$I_{ij} = M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1) \times \alpha_{1,i}^{\mathrm{EP}}\left(proc, J, R_2\right) \times T_{ij}^{\mathrm{NOESY}}(\tau) \times T_j^{\mathrm{HXQC}} \times \alpha_j^{\mathrm{WS}} \times \alpha_{2,j}^{\mathrm{EP}}\left(proc, J, R_2\right) \quad (1)$$

To calculate to overall error, we use the ratio of the intensities of a cross and a diagonal peak. If the cross peak is scaled by the intensity of the diagonal of the spin from which the magnetization originates, only the last four factors of the right-hand side of equation 1 contribute to the error.

$$I_{ij} / I_{ii} = \frac{T_{ij}^{\mathrm{NOESY}}(\tau) \times T_j^{\mathrm{HXQC}} \times \alpha_j^{\mathrm{WS}} \times \alpha_{2,j}^{\mathrm{EP}}\left(proc, J, R_2\right)}{T_{ii}^{\mathrm{NOESY}}(\tau) \times T_i^{\mathrm{HXQC}} \times \alpha_i^{\mathrm{WS}} \times \alpha_{2,i}^{\mathrm{EP}}\left(proc, J, R_2\right)} \quad \text{(S5)}$$

Another possibility is to normalize to the intensity of the diagonal of the spin to which the magnetization is transferred.

$$I_{ij} / I_{jj} = \frac{M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1) \times \alpha_{1,i}^{\mathrm{EP}}\left(proc, J, R_2\right) \times T_{ij}^{\mathrm{NOESY}}(\tau)}{M_j^{\mathrm{SS}}(d_{\mathrm{int}}, R_1) \times \alpha_{1,j}^{\mathrm{EP}}\left(proc, J, R_2\right) \times T_{jj}^{\mathrm{NOESY}}(\tau)} \quad \text{(S6)}$$

Propagation of the error of each factor then leads to the relative errors:

$$\frac{\Delta\left(I_{ij} / I_{ii}\right)}{I_{ij} / I_{ii}} = \left\{ \begin{aligned} &\left[\frac{\Delta T_{ij}^{\mathrm{NOESY}}(\tau)}{T_{ij}^{\mathrm{NOESY}}(\tau)}\right]^2 + \left[\frac{\Delta T_j^{\mathrm{HXQC}}}{T_j^{\mathrm{HXQC}}}\right]^2 + \left[\frac{\Delta \alpha_j^{\mathrm{WS}}}{\alpha_j^{\mathrm{WS}}}\right]^2 + \left[\frac{\Delta \alpha_{2,j}^{\mathrm{EP}}\left(proc, J, R_2\right)}{\alpha_{2,j}^{\mathrm{EP}}\left(proc, J, R_2\right)}\right]^2 \\ &+ \left[\frac{\Delta T_{ii}^{\mathrm{NOESY}}(\tau)}{T_{ii}^{\mathrm{NOESY}}(\tau)}\right]^2 + \left[\frac{\Delta T_i^{\mathrm{HXQC}}}{T_i^{\mathrm{HXQC}}}\right]^2 + \left[\frac{\Delta \alpha_i^{\mathrm{WS}}}{\alpha_i^{\mathrm{WS}}}\right]^2 + \left[\frac{\Delta \alpha_{2,i}^{\mathrm{EP}}\left(proc, J, R_2\right)}{\alpha_{2,i}^{\mathrm{EP}}\left(proc, J, R_2\right)}\right]^2 \end{aligned} \right\}^{1/2} \quad \text{(S7)}$$

$$\frac{\Delta\left(I_{ij} / I_{jj}\right)}{I_{ij} / I_{jj}} = \left\{ \begin{aligned} &\left[\frac{\Delta M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1)}{M_i^{\mathrm{SS}}(d_{\mathrm{int}}, R_1)}\right]^2 + \left[\frac{\Delta \alpha_{1,i}^{\mathrm{EP}}\left(proc, J, R_2\right)}{\alpha_{1,i}^{\mathrm{EP}}\left(proc, J, R_2\right)}\right]^2 + \left[\frac{\Delta T_{ij}^{\mathrm{NOESY}}(\tau)}{T_{ij}^{\mathrm{NOESY}}(\tau)}\right]^2 \\ &+ \left[\frac{\Delta M_j^{\mathrm{SS}}(d_{\mathrm{int}}, R_1)}{M_j^{\mathrm{SS}}(d_{\mathrm{int}}, R_1)}\right]^2 + \left[\frac{\Delta \alpha_{1,j}^{\mathrm{EP}}\left(proc, J, R_2\right)}{\alpha_{1,j}^{\mathrm{EP}}\left(proc, J, R_2\right)}\right]^2 + \left[\frac{\Delta T_{jj}^{\mathrm{NOESY}}(\tau)}{T_{jj}^{\mathrm{NOESY}}(\tau)}\right]^2 \end{aligned} \right\}^{1/2} \quad \text{(S8)}$$

Similar expressions are obtained for the 3D HXQC-NOESY experiments with rearranged order of the factors in equation 1 and accordingly modified equations S7 and S8.

An additional error arises from the fit of the cross-relaxation rate σ_{ij} from equations 10 and 11 because ρ_j is not known if the diagonal-peak of spin j cannot be evaluated.

All numerical estimations derived in the previous sections and the resulting errors of the fitted cross-relaxation rate constants and distance limits are summarized in Table 2 (chapter 2).

Table S4. Relative errors of ^{15}N-resolved 3D NOESY pulse schemes, the extracted cross-relaxation rate constants and distance limits.

factor/quantity	error 3D NOESY-[^{15}N,^{1}H]-HXQC	error 3D -[^{15}N,^{1}H]-HXQC-NOESY[a]	
I_{ij}	40 %	40 %	44 % (20% watergate)
I_{ij} / I_{ii} (scaled to diagonal of spin of origin)	52 %	25 %	38 %
I_{ij} / I_{jj} (scaled to diagonal of spin of destination)	24 %	50 %	
σ_{ij} (scaled to diagonal of spin of origin)	53 %	27 %	39 %
σ_{ij} (scaled to diagonal of spin of destination)	26 %	51 %	
r_{ij} (scaled to diagonal of spin of origin) [b]	9 %	5 %	7 %
r_{ij} (scaled to diagonal of spin of destination) [b]	4 %	8 %	
$r_{ij}^{upl/lol}$ (scaled to diagonal of spin of origin) [c]	+13 % / -7 %	+5 % / -4 %	+9 % / -5 %
$r_{ij}^{upl/lol}$ (scaled to diagonal of spin of destination) [c]	+5 % / -4 %	+13 % / -7 %	

[a] Order of the factors in equation 1 is changed.

[b] Error propagated as $\Delta r/r = \Delta s/6s$.

[c] Error determined as the difference between r obtained from s and from $\sigma+\Delta\sigma$ or $\sigma-\Delta\sigma$.

Experimental verification of predicted errors in extracted distances

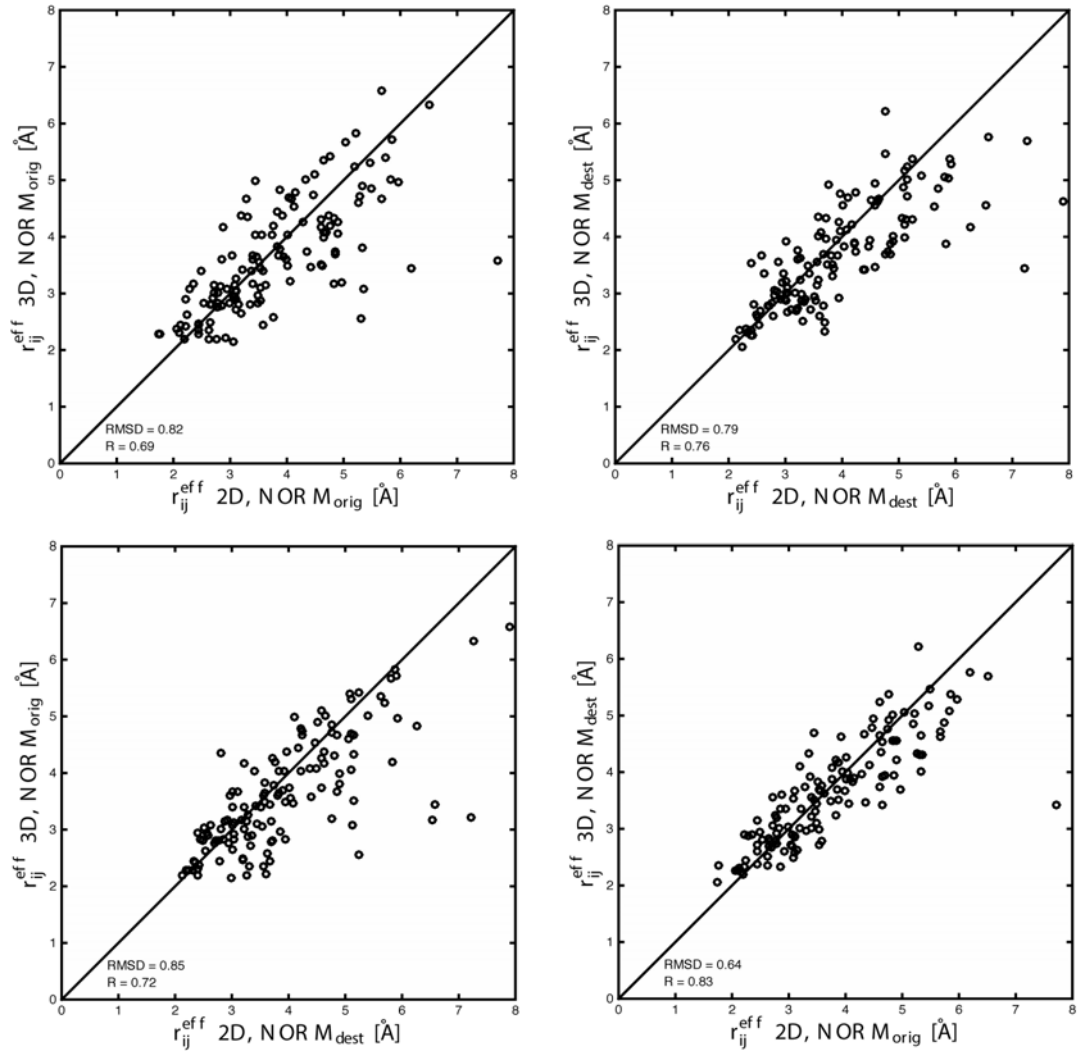
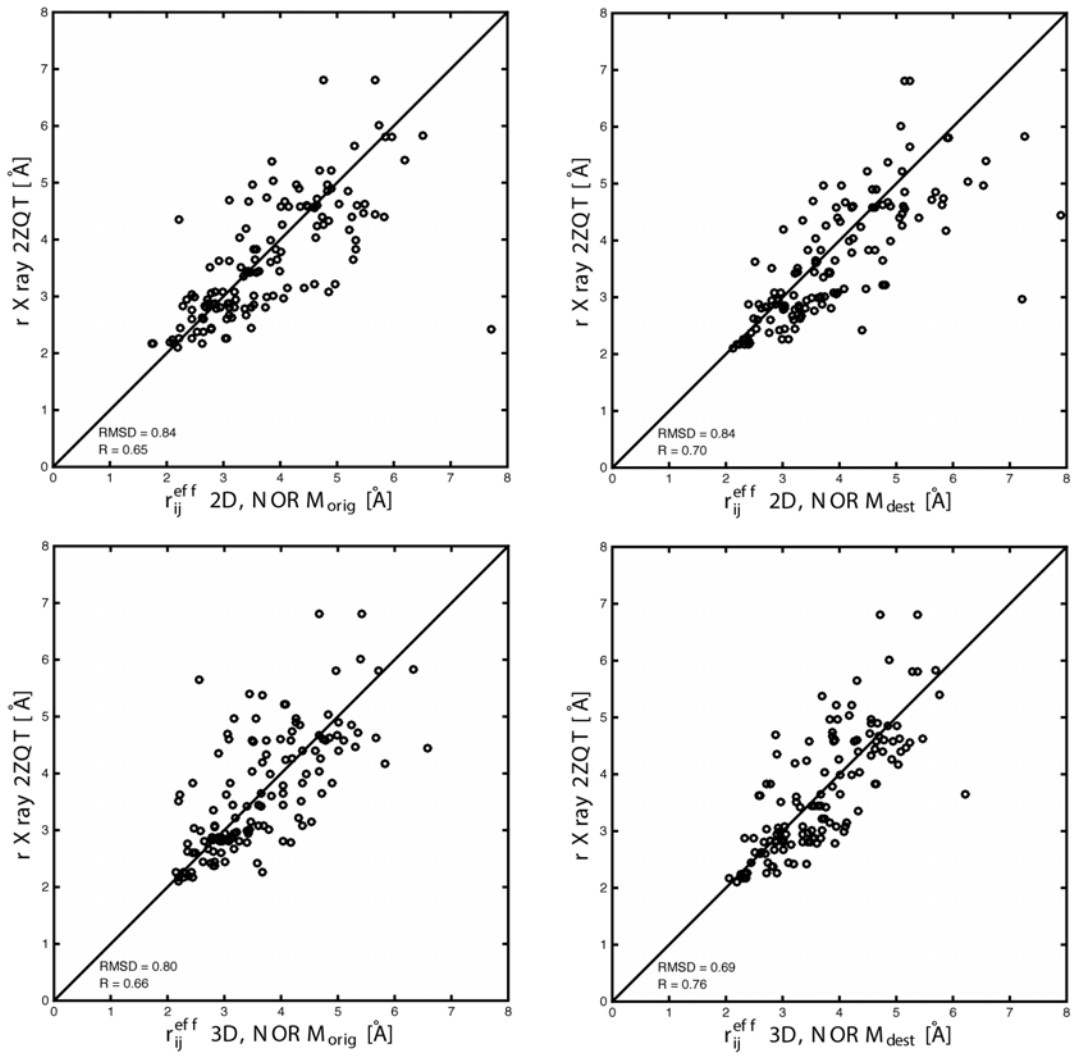


Figure S5: Comparison of the effective distances derived from 2D and 3D NOESY, when normalizing to origin or destination of magnetization. First row: r^{eff}_{ij} distances derived from 3D NOESY data normalized to origin vs. 2D NOESY data normalized to origin of magnetization (left) and 2D NOESY data normalized to destination of magnetization (right). Second row: r^{eff}_{ij} distances derived from 3D NOESY data normalized to origin vs 2D NOESY data normalized to destination of magnetization (left) and 2D NOESY data normalized to origin of magnetization (right).



S

Figure S6: Comparison of the eNOE- and X-ray-derived distances. eNOEs found in the 3D experiment as well as the 2D experiment are selected. In the first row, effective distances r^{eff}_{ij} derived from 2D NOESY data normalized to the origin of magnetization (using $I_i^{\mathrm{diag}}(0)$, $\mathrm{NORM_{orig}}$) are shown on the left, and to the destination of magnetization (using $I_j^{\mathrm{diag}}(0)$, $\mathrm{NORM_{dest}}$) on the right. In the second row, r^{eff}_{ij} derived from 3D NOESY data normalized to origin of magnetization ($\mathrm{NORM_{orig}}$) are shown on the left, and to the destination of magnetization ($\mathrm{NORM_{dest}}$) on the right.

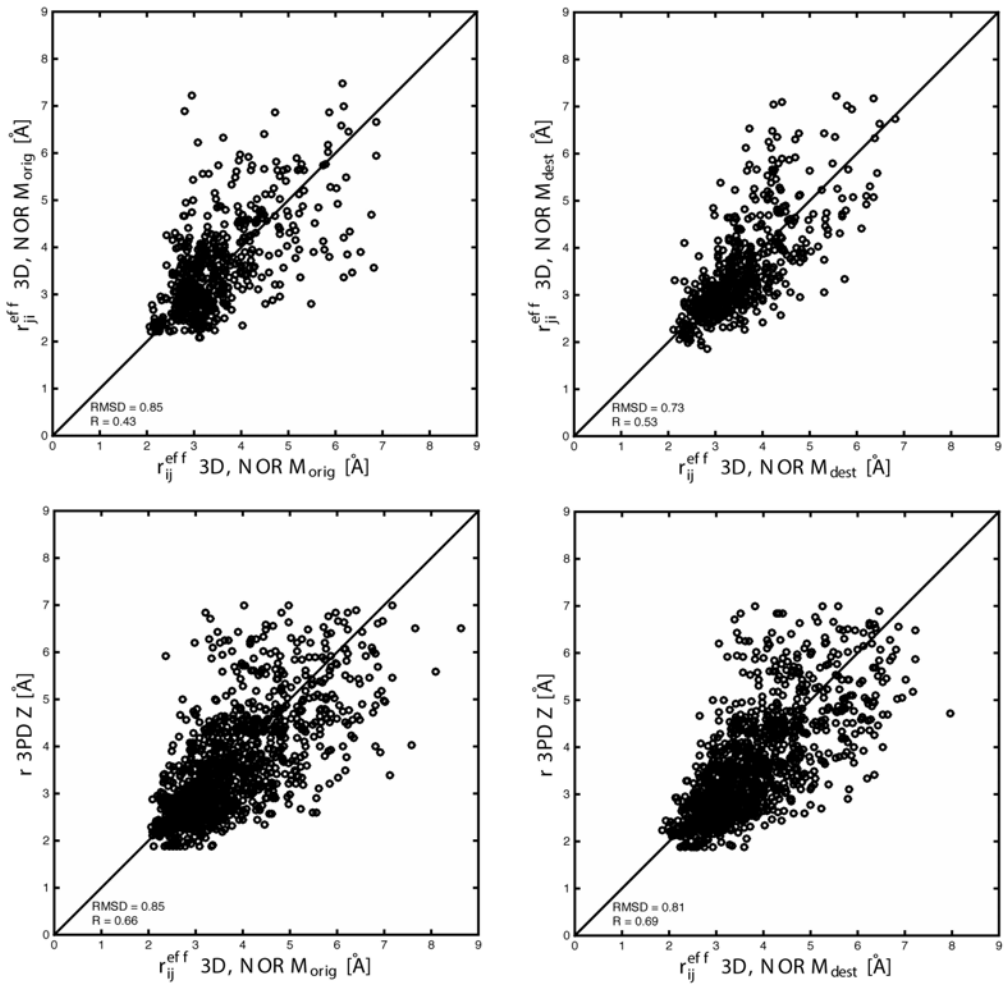


Figure S7: eNOE-derived distances from the second PDZ (post-synaptic density-95/discs large/zonula occludens-1) domain of human tyrosine phosphatase 1E, in support of better normalization by normalizing to the destination spin in 3D HSQC-NOESY. In the first row, effective distances r^{eff}_{ij} derived from data normalized to the origin of magnetization (using $I_i^{diag}(0)$, $NORM_{orig}$) are shown on the left, and to the destination of magnetization (using $I_j^{diag}(0)$, $NORM_{dest}$) on the right. In the second row, eNOE- and X-ray-derived distances are compared. r^{eff}_{ij} derived from data normalized to origin of magnetization ($NORM_{orig}$) are shown on the left, and to the destination of magnetization ($NORM_{dest}$) on the right.

10.2. Appendix to chapter 5: Replica-exchange simulations combined with chemical shift data

Conformational clustering and Maximum Entropy reweighing of structural ensembles using backbone chemical shifts

The samples generated using the constant biased simulations (see 5.4. Experimental procedures) do not follow the Boltzmann distribution of the target force field. We consequently reweighed the samples into the generalized canonical distribution (CEs),

$$q(x,T) \propto \exp\left(-\frac{E_{phys}(x)}{\kappa T}\right), \quad (1)$$

where T denotes temperature and k Boltzmann's constant. The temperatures here take on values where experimental data was acquired. $E_{phys}(x)$, denotes the force field ff99SB-ILDN [1].

All the conformations sampled were represented by all Cα-Cα distances and projected into the first 10 principal components of their covariance matrix. Subsequently, the conformational space was segmented into 20 clusters using the K-means algorithm. Adding additional clusters to the analysis yielded highly redundant clusters in the high-density regions of the phase-space. Without loss of generality we may cast equation 1 as,

$$q(x,C,T) \propto \theta(C - C(x))\exp\left(-\frac{E_{phys}(x)}{\kappa T}\right), \quad (2)$$

where $\theta(\cdot)$ is the indicator function, C denotes a cluster index and the function $C(\cdot)$ returns the cluster annotation of a structure x. We aimed at using the clusters C as the basis for obtaining a new generalized distribution, $p(x, T \mid \lambda_{i,T})$, by biasing through the *maximum entropy principle* (MEP) using backbone chemical shifts (REs, from above). By applying the MEP we obtain

$$p(x,C,T \mid \lambda_{i,T}) \propto q(x,C,T)\prod_{i=1}^{N}\exp\left(-\lambda_{i,T}\delta_i(x)\right), \quad (3)$$

where $\delta_i(x)$ is the chemical shift i back-predicted using CamShift [2, 3] and $\lambda_{i,T}$ is a Lagrange multiplier corresponding to experimental observable I at temperature T. We now assume that $\delta_i(x) \approx \delta_i(y)$ for C(x) = C(y) for all our observations. This allows us to simplify equation 3 to,

$$\hat{p}(x,C,T \mid \lambda^{C}_{i,T}) \propto q(x,C,T)\prod_{i=1}^{N}\exp\left(-\lambda^{C}_{i,T}\overline{\delta_{i,T}(C)}\right), \quad (4)$$

where $\overline{\delta_{i,T}(C)}$ is the chemical shift in cluster C, with the corresponding augmented Lagrange multiplier $\lambda^{C}_{i,T}$. Ensemble averages were calculated using the expectations,

$$\overline{\delta^{calc}_{i,T,\lambda_{i,T}}} = \varepsilon_T + \sum_{C=1}^{20} \overline{\delta_{i,T}(C)} \hat{p}(x,T \mid C, \lambda_{i,T}) Z^{-1}_{\hat{p}} \quad (5)$$

and

$$\overline{\delta^{calc}_{i,T}} = \varepsilon_T + \sum_{C=1}^{20} \overline{\delta_{i,T}(C)} \hat{q}(x,T \mid C) Z^{-1}_{\hat{q}} \quad (6)$$

for the REs and CEs respectively. Z^. is an appropriate normalization factor. Systematic prediction errors were compensated for with a 0th order correction factor for each temperature, ε_T. The correction factors were calculated as mean difference between all experimental chemical shifts and the corresponding expectations using the unrestrained simulation at each temperature and isotope independently.

Estimation of the Lagrange multipliers $\lambda_{i,T}$, necessary to evaluate p(x, T | $\lambda_{i,T}$), was carried out using a variation of the stochastic optimization scheme previously proposed [4], using two different error models. For the Gaussian, we minimize the temperature dependent χ^2_T,

$$\chi^2_T = \sum_i \left(\frac{\delta^{\exp}_{i,T} - \overline{\delta^{calc}_{i,T,\lambda_{i,T}}}}{\sigma_{pred}} \right)^2 \quad (7)$$

where σ_{pred} is the random prediction error of CamShift in ppm. Each atom type has their own random prediction error as reported in [2]. For the flat-bottom, we minimize the following expression,

$$F^2_T = \sum_i \hat{\Delta}_{i,T} \left(\Delta_{i,T} - \operatorname{sgn} \Delta_{i,T} \sigma_{pred} \right)^2 \quad (8)$$

with $\hat{\Delta}_{i,T} = \delta^{\exp}_{i,T} - \overline{\delta^{calc}_{i,T,\lambda_{i,T}}}$ and $\hat{\Delta}_{i,T}$ is 1 if $\Delta_{i,T} > \sigma_{pred}$ and 0 otherwise. The Lagrange multipliers were updated by adding samples drawn from a Gaussian distribution with zero mean and a small variance, evaluating equations 7 or 8 and accepting changes according to the Metropolis-Hastings criterion.

Analysis of Pin1-WW domain structures in the Protein Data Bank

We recovered 49 complete (no missing residues or chemical modifications) Pin1-WW domain structures from the 61 indexed on Uniprot (UniProt: Q13526) [5]. These

were manually classified into native-like (44 structures) or near native-like (6 structures) topologies, based upon the relative orientation of the N- and C-terminal configuration (see Figure S6).

Additional figures

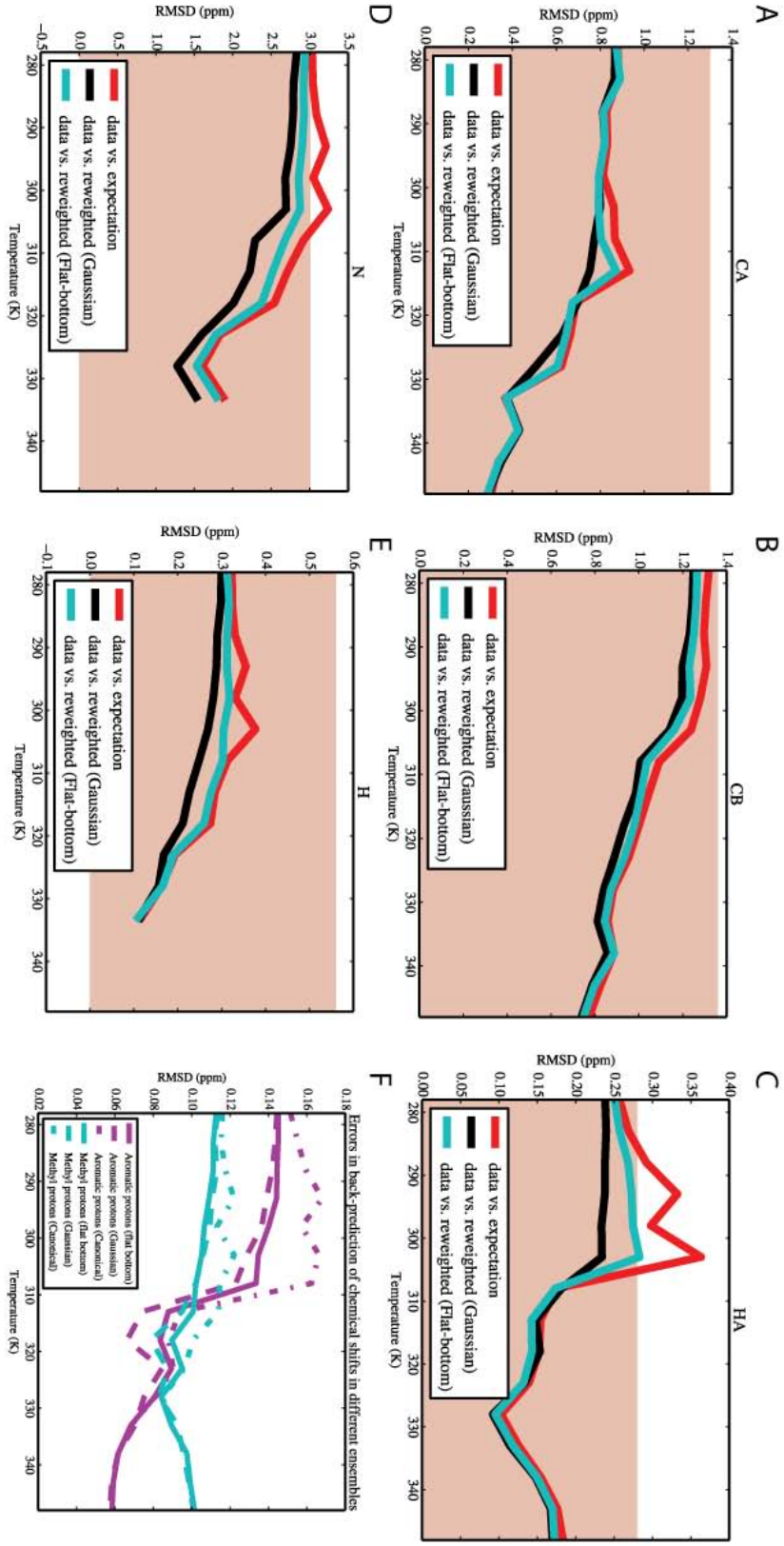


Figure S1, related to Figure 1, chapter 5: Root mean square error of average backbone and side chain chemical shifts back-computed using the differently weighed ensembles at all temperatures. The red shaded area illustrates prediction error of CamShift. Root mean square error of average side-chain methyl proton and aromatic proton chemical shifts back-computed using the differently weighed ensembles at all temperatures.

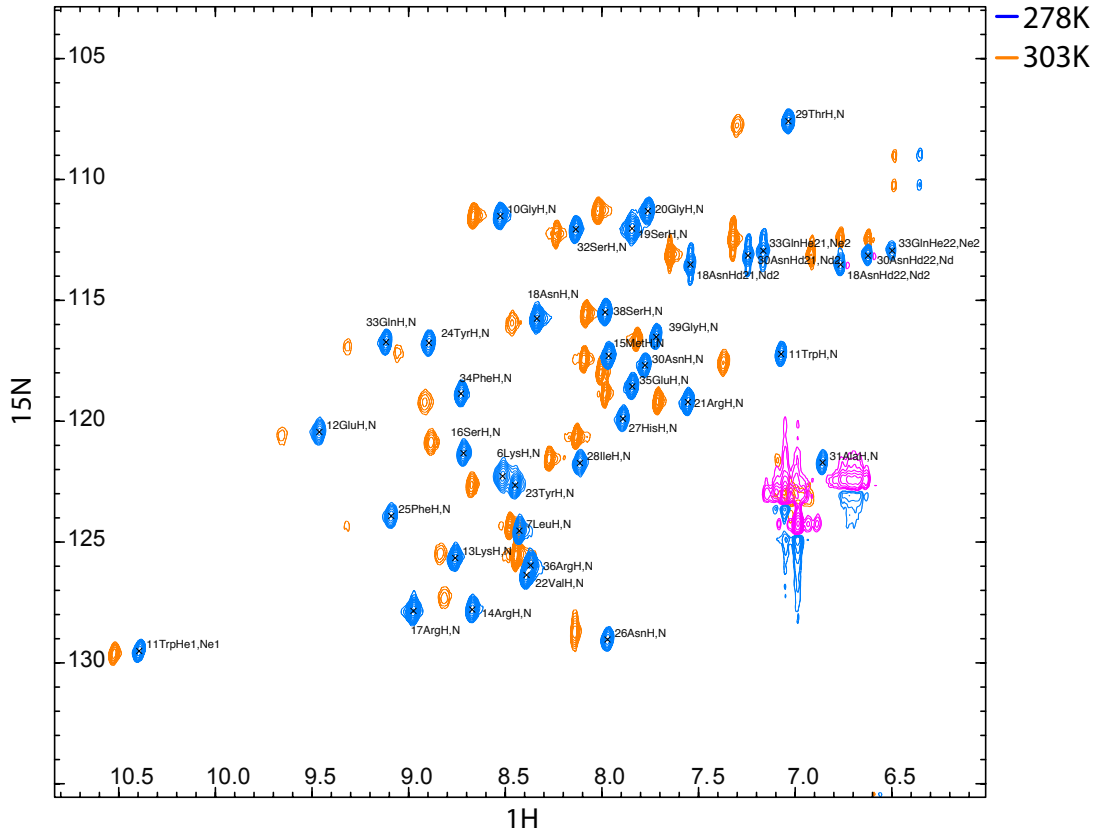


Figure S2, related to Figure 1, chapter 5: ^{1}H-^{15}N HMQC spectra of Pin1-WW S18N/W34F at 278K (blue) and 303 K (orange). Resonance assignments are shown as labels on signals at 278 K.

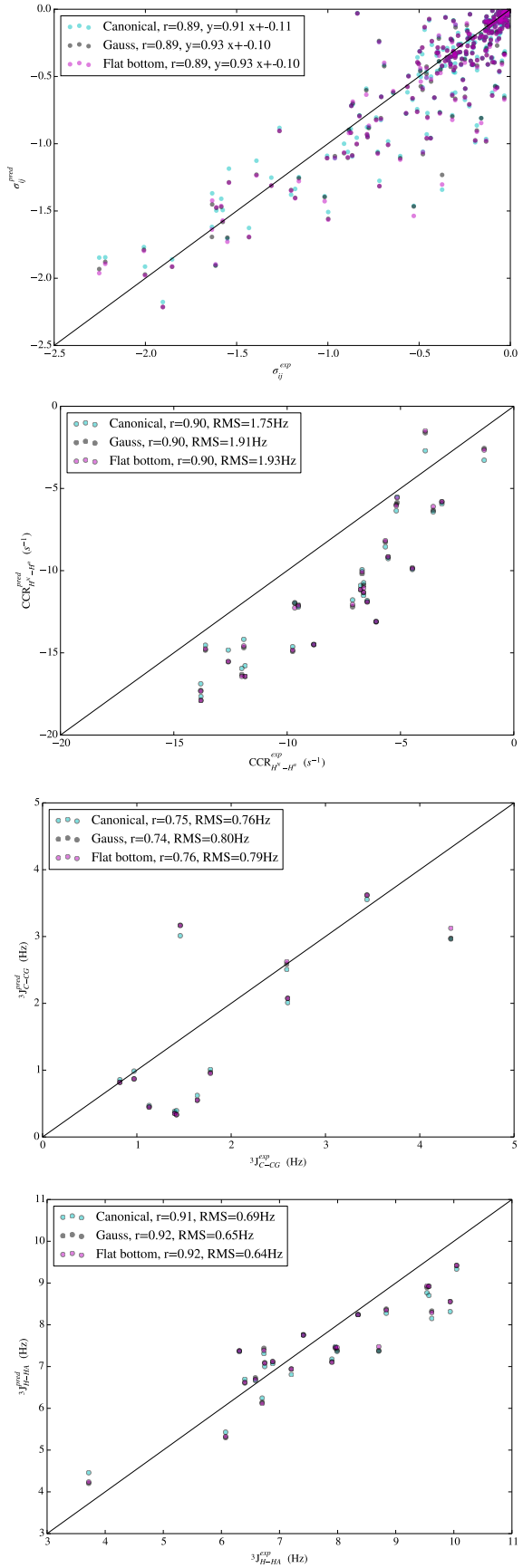


Figure S3, related to Figure 3A,C, chapter 5: Validation of the three differently weighed ensembles at 278 K. Top- left, comparison to bi-directional eNOE rates, top-right comparison to CCR data, bottom-left side-chain ^{3}J-couplings, bottom-right back-bone ^{3}J-couplings.

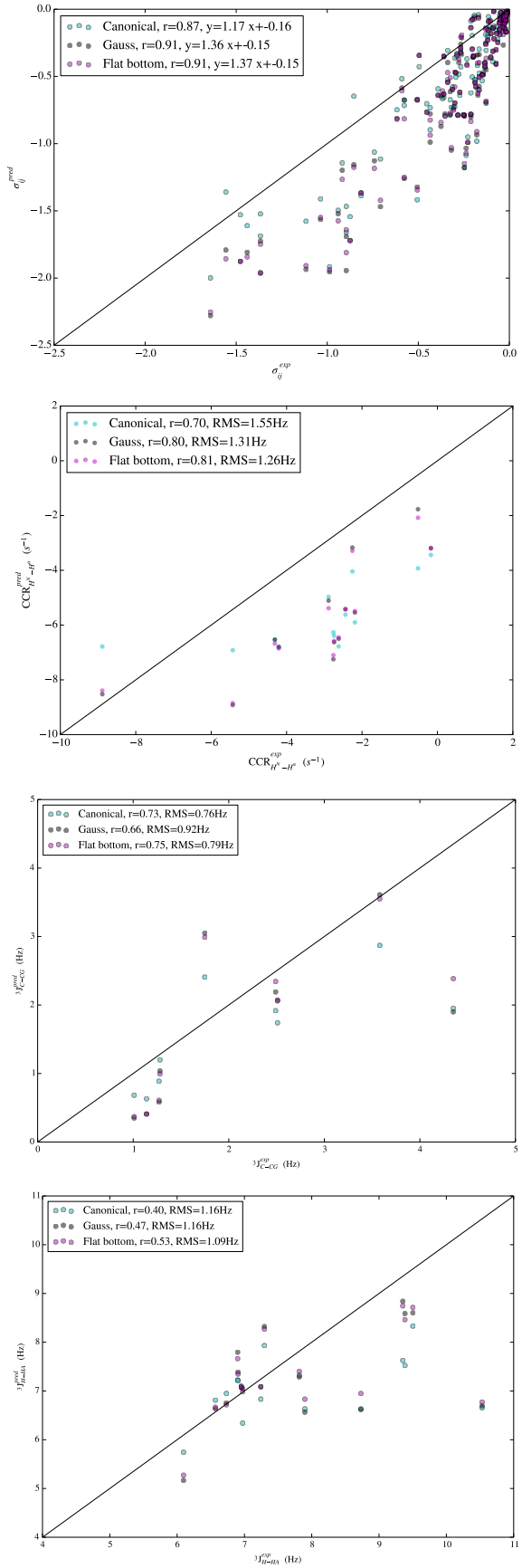


Figure S4, related to Figure 3B,D, chapter 5: Validation of the three different ensembles at 303 K. Top-left, comparison to bi-directional eNOE rates, top-right comparison to CCR data, bottom-left side-chain ^{3}J-couplings, bottom-right backbone ^{3}J-couplings.

Figure S5, Related to Figure 4: Illustration of ring-flip of Trp-11 in WW-Pin1 between the native (purple) and near-native (teal) configurations identified by cluster analysis.

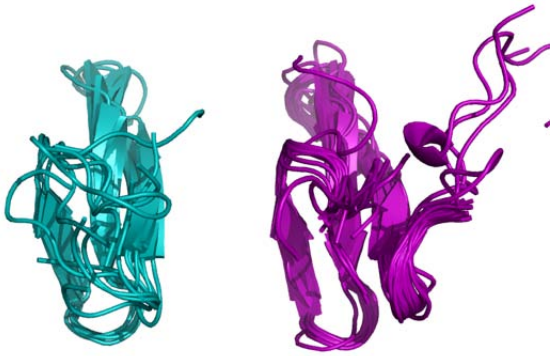


Figure S6, related to Figure 1C,D, chapter 5: Render of Pin1-WW structures listed in Uniprot (UniProt: Q13526) with native-like topology (purple) and near native-like topology (teal).

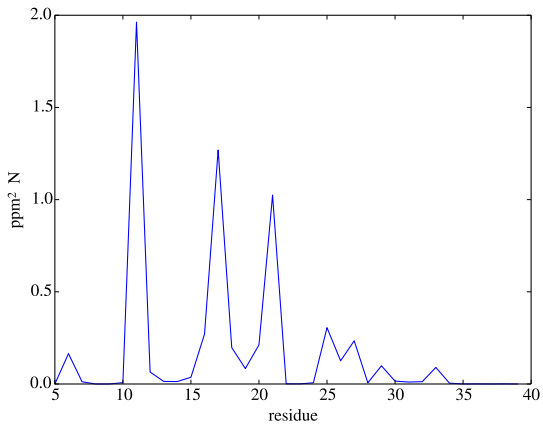


Figure S7, related to Figure 5, chapter 5: Chemical shift variances computed for the S18N/W34F simulation at 278 K, using native and near-native clusters.

References

[1] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE, Improved side-chain torsion potentials for the amber ff99sb protein force field, Proteins 2010, 78(8), 1950–8.

[2] Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M, Fast and accurate predictions of protein nmr chemical shifts from interatomic distances, J Am Chem Soc 2009, 131(39), 13894–5.

[3] Fu B, Sahakyan AB, Camilloni C, Tartaglia GG, Paci E, Caflisch A, Vendruscolo M, Cavalli A, Almost: An all atom molecular simulation toolkit for protein structure determination, J Comput Chem 2014, 35(14), 1101–1105.

[4] Olsson S, Vögeli BR, Cavalli A, Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K, Hamelryck T, Probabilistic determination of native state ensembles of proteins, J Chem Theory Comput 2014, 10(8), 3484–3491.

[5] UniProt Consortium, Uniprot: a hub for protein information, Nucleic Acids 2015, Res. 43 (Database issue), D204–12.

10.3. Appendix to chapter 6: The dynamic basis for signal propagation in human Pin1-WW

Sequence analysis of ligand binding motifs

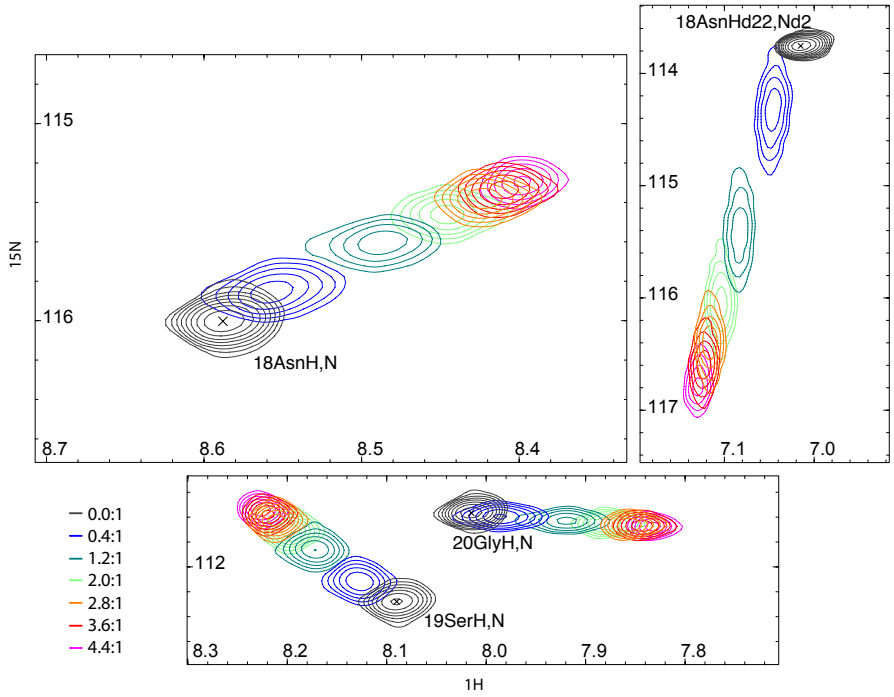
Possible binding partners of Pin1 were identified using the IntAct database [1], results were filtered using only reviewed UniProtKB entries of human proteins [2]. These results were further filtered to include only ligands with at least one pSer/pThr-Pro motif. A total of 78 candidate ligands was identified, with between 1 and 27 pSer/pThr-Pro motifs each. These motifs and the seven residues immediately flanking residues up and down stream of these motifs were pooled for further analyses. We used these 16-residue fragments to create two sequence logo plots: one only using single binding site ligands and one using all ligands. Sequence logos represent sequence conservation by letter stacks where the heights of the letters at each position represent the occurrence frequency [3].

Selection of WT ensembles of Pin1-WW, Pin1-WW:τ and Pin1-WW:pCdc25C using NOE datasets

Like for the S18N/W34F simulation we clustered the WT simulation into 20 clusters using K-means. From these clusters we identified six with near-native or native topology which were pooled for the current analysis excluding all unfolded and partially unfolded clusters. For each of these clusters we computed the average NOEs for previously reported data [4] within each of the clusters independently as $\overline{r^{-6}}$ where r is an inter-proton distance. We then used a greedy Monte Carlo approach to optimize the subset of clusters whose average would give us the best agreement with the experimental data using a square well potential as defined in XPLOR-NIH. [5] The Monte Carlo approach randomly selects between 1 and 6 of the clusters and computes an average and compares this to the experimental data - if the agreement is better than a previous subset this is saved as the best. The procedure converged within a few hundred iterations.

Additional figures

A



B

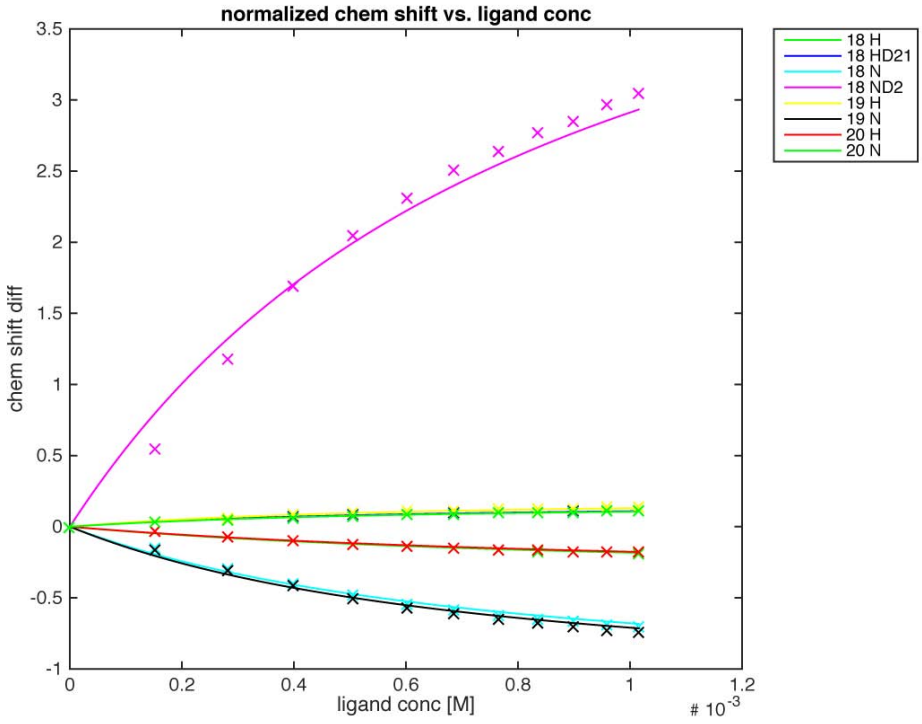


Figure S1. Related to experimental procedures and Figure 2, chapter 6: Titration of Pin1-WW S18N/W34F with pCdc25C. (A) Excerpts of ^{1}H-^{15}N HMQC spectra of Pin1-WW S18N/W34F at 278 K with increasing relative concentrations of pCdc25C ligand. (B) Corresponding titration isotherms as a function of absolute concentration.

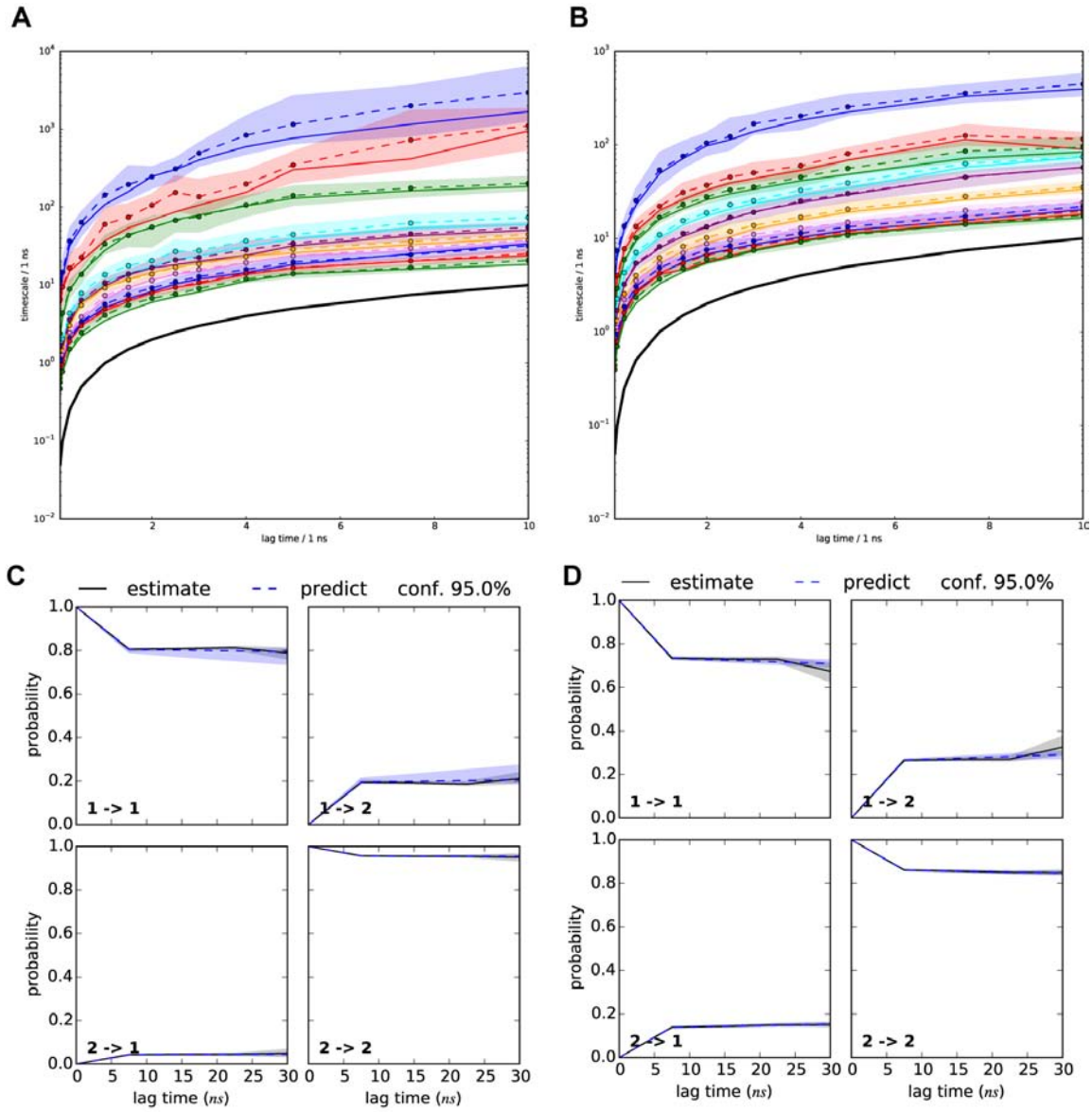


Figure S2, related to experimental procedures and Figure 4, chapter 6: Validation of Markov state models. Implied time-scale plots of Markov state models generated using (A) simulations with a native like topology and (B) Near-native topology of the WW-domain in Pin1. Colored lines and their shaded area indicate the 10 slowest relaxing processes and their uncertainty as a function of MSM lag-time. The thick black line indicates the temporal resolution limit. Chapman-Kolomogorov tests of Markov state models (MSMs) generated using (C) simulations with a native like topology and (D) Near-native topology of the WW-domain in Pin1. Black line and grey shaded area is estimated transition probability and its uncertainty, whereas the blue dashed lined and blue shaded area is the predicted values using a MSM with a lag-time 3.25 ns.

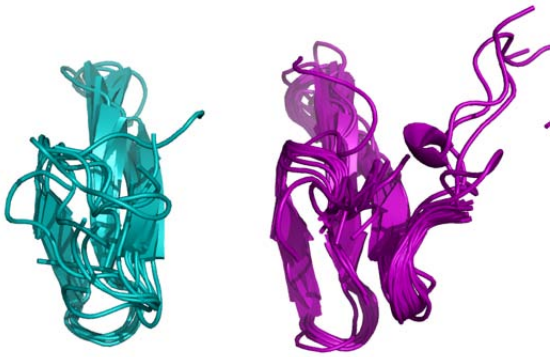


Figure S3, related to Figure 1C,D, chapter 5: Render of Pin1-WW structures listed in Uniprot (UniProt: Q13526) with native-like topology (purple) and near native-like topology (teal).

References

[1] Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H, The mintact project–intact as a common curation platform for 11 molecular interaction databases, Nucleic Acids 2014, Res. 42(Database issue), D358–63.

[2] UniProt Consortium, Uniprot: a hub for protein information, Nucleic Acids 2015, Res. 43 (Database issue), D204–12.

[3] Crooks GE, Hon G, Chandonia J-M, Brenner SE, Weblogo: a sequence logo generator, Genome Res 2004, 14(6), 1188–90.

[4] Wintjens R, Wieruszeski JM, Drobecq H, Rousselot-Pailley P, Buée L, Lippens G, Landrieu I, 1h nmr study on the binding of pin1 trp-trp domain with phosphothreonine peptides, J Biol Chem 2001, 276(27), 25150–6.

[5] Schwieters CD, Kuszewski J, Clore GM, Using xplor–NIH for NMR molecular structure determination, Prog

10.4. Appendix to chapter 7: The allosteric coupling of ligand-binding via the interaction interface in Pin1

Pin1-WW binding to ligand pCdc25C

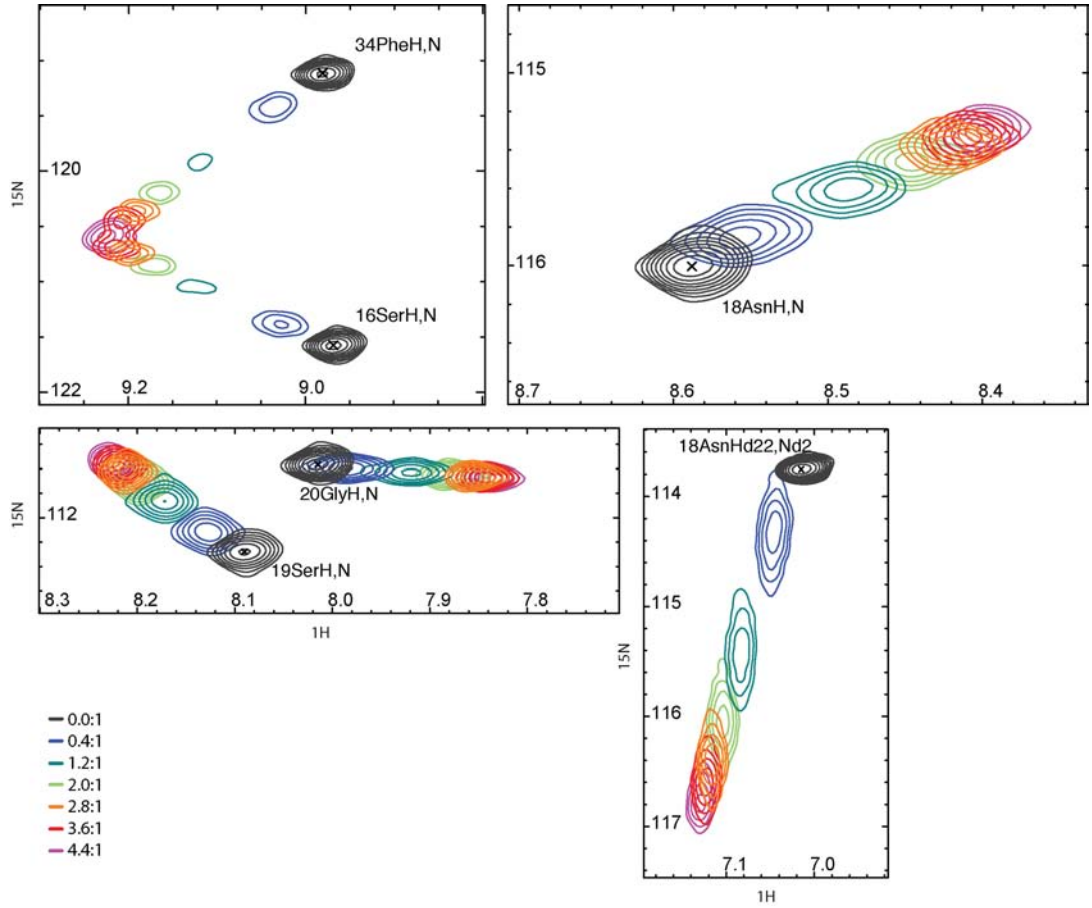


Fig. S1: Excerpts of [^{15}N, ^{1}H]- HSQC showing chemical shift perturbations upon ligand-binding (pCdc25C) of the Pin1-WW subunit in loop1. The spectra were recorded at 278K with increasing relative concentrations of pCdc25C as indicated in the legend. The initial concentration of Pin1-WW was 0.4 mM, with step-wise addition of 30 µL increments of 2.4 mM pCdc25C.

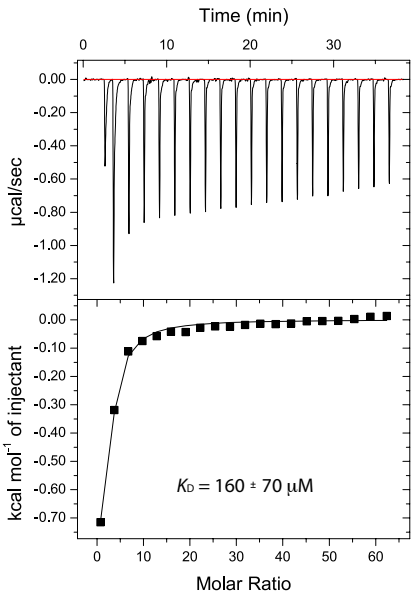


Fig. S2: ITC data of Pin1-WW domain and pCdc25C. The upper plot shows the thermal power required to maintain a zero temperature difference between reference and sample cell. Each peak corresponds to a ligand injection. The lower plot shows the binding isotherm, with normalized heats per peak as a function of molar ratio.

Full length Pin1 binding to ligand pCdc25C

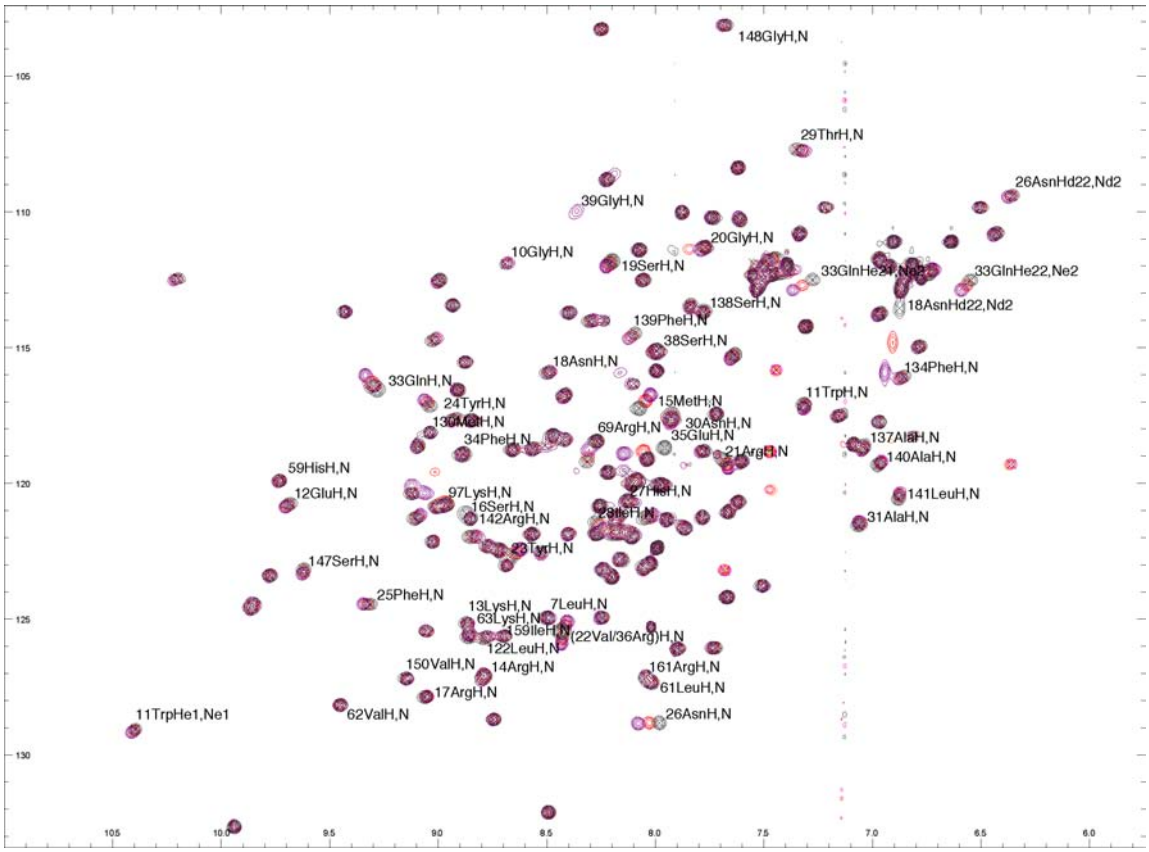


Fig. S3: Preliminary [^{15}N, ^{1}H]- HSQC experiment, showing the chemical shift perturbations upon ligand binding for full-length Pin1. The residues at the domain interface of Pin1-PPIase as revealed by X-ray (1Pin.pdb [1] structure analysis (Lys97, Ala137, Ser138, Phe139, Ala140, Leu141, Arg142, Ser147, Gly148 and Val150 [2, 3] are labeled. In addition, the Pin1-WW loop2 (His27, Ile28, Thr29, Asn30) as well as the substrate recognition site [4] residues on Pin1-WW:loop1 (Ser16, Arg17, Ser18, Ser19, Gly20, Thr23, Phe25, Phe34) are labeled. Furthermore, the residues responsible for substrate recognition on Pin1-PPIase (His59, Leu61, Val62, Lys63, Arg69, Leu122, Met130, Phe134) [2, 3] are labeled.

Saturation transfer difference (STD) experiments with full length Pin1

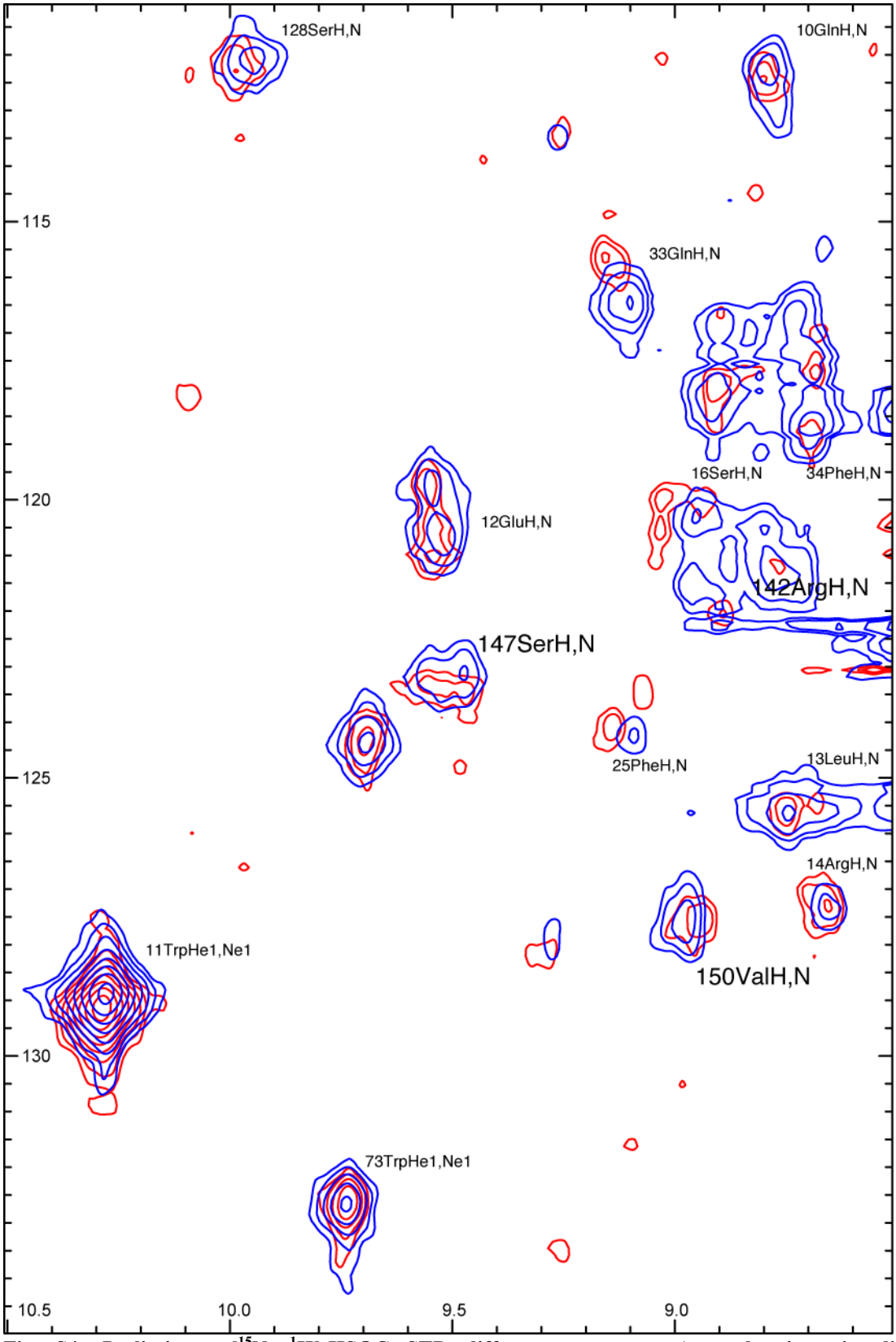


Fig. S4: Preliminary [^{15}N, ^{1}H]-HSQC STD difference spectrum (control minus irradiated experiment) of full length Pin1, irradiated at 11TrpHe1, Ne1. The apo (blue) and holo (red) spectra are shown. Upon addition of ligand, the transfer efficiency of the magnetization decreases for residues Arg142, Ser147 and Val150, indicative of a less tight, less rigid interaction between the two domains.

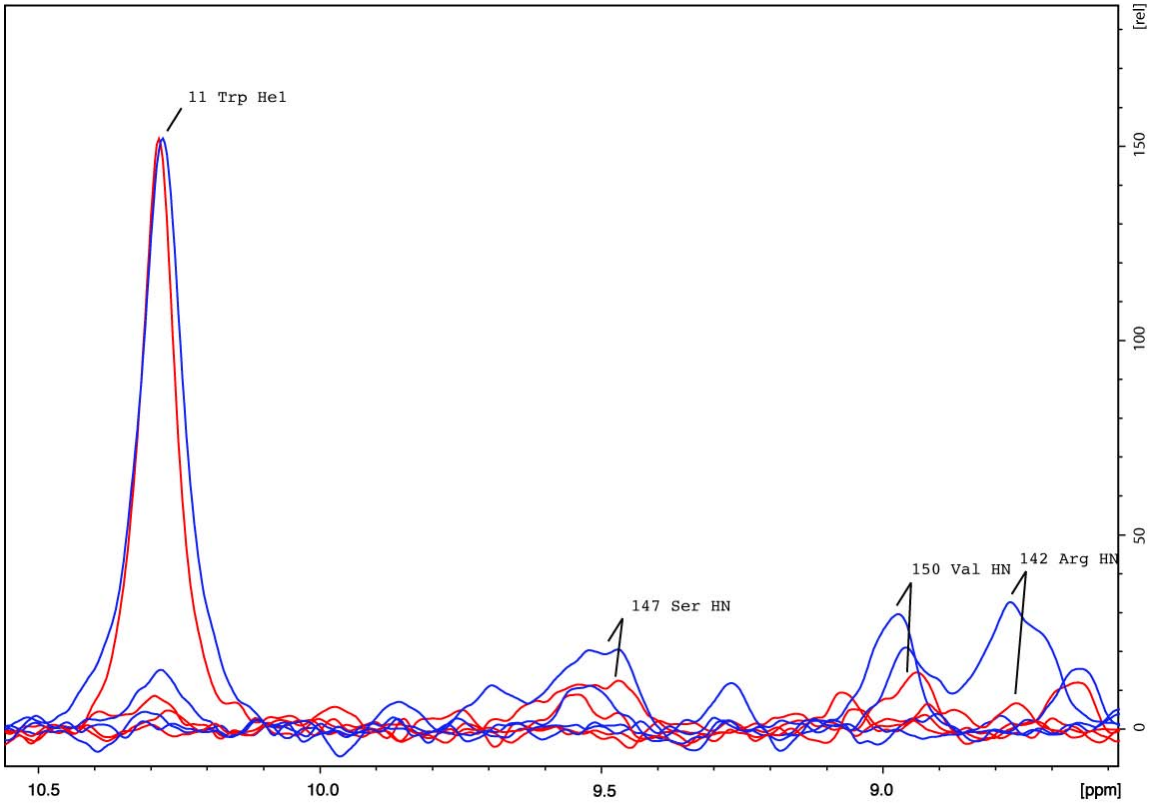


Fig. S5: Preliminary data of 1D slices of [^{15}N, ^{1}H]-HSQC STD difference spectrum (control minus irradiated experiment) of full length Pin1, irradiated at 11TrpHe1, Ne1. The black lines indicate the labeled peaks on the extracted slices. The apo (blue) and holo (red) spectra are shown. Upon addition of ligand pCdc25C, the transfer efficiency of the magnetization is decreasing for residues Arg142, Ser147 and Val150, indicative of a less tight, less rigid interaction between the two domains.

Structure validation

The CYANA target function is a measure for the quality of the computed structural ensembles given in terms of the squared violation of the experimental restraints. The multi-state ensembles reflect the data significantly better than the single-state ensemble and there is an improvement in the target function up to four states until the value levels off (Figure S7a). Specifically, we observe a significantly better quality of the two-states versus the single-state ensemble (Figure S6). We are developing various methods to verify the accuracy and validity of our structure calculations. A jack-knife procedure applied to the eNOEs provides a general check against over-fitting and self-consistency of the experimental data [5]. The jack-knife protocol repeats a structure calculation ten times, each time 10% of the experimental input data are deleted at random such that every restraint is omitted once. The back-calculate the target function from the omitted data then represents the entire data set. If the back-calculated target function for the higher state ensembles (Fig. S7b) decreases this provides further support for the conclusion that the experimental data is better described by the multi-states ensemble than the single-state ensemble.

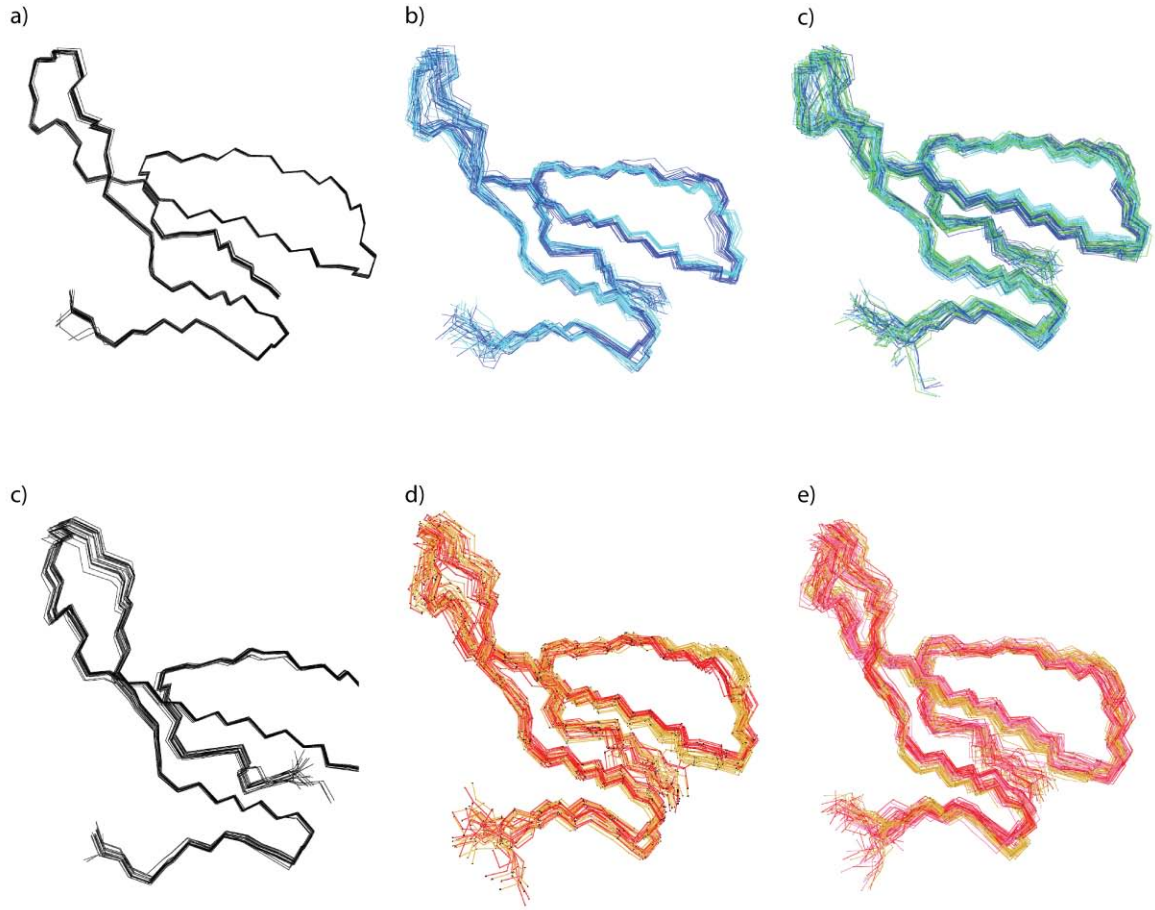


Fig. S6: Bundles of apo Pin1-WW a) single-state, b) two-states and c) three-states ensembles. Bundles of holo Pin1-WW d) single-state, e) two-states and f) three-states ensembles.

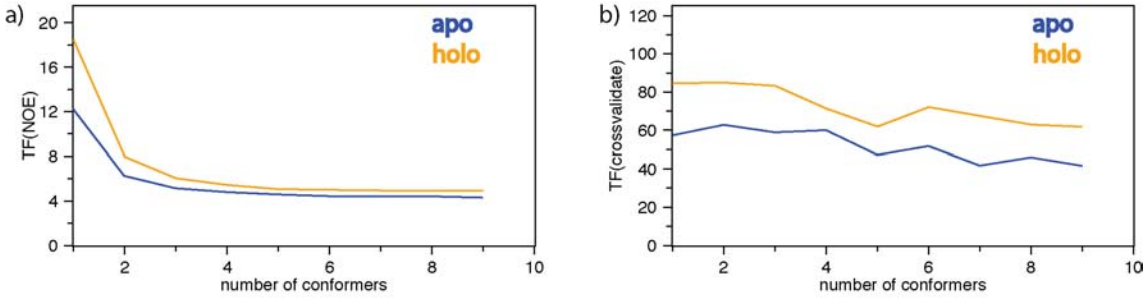


Fig. S7: a) drop of target function (TF, NOEs only) for apo and holo Pin1-WW from one to multi-states ensemble b) the back-calculated TF for the jack-knife procedure for apo and holo Pin1-WW.

To further illustrate the quality of the eNOE multi-states ensemble one can compare experimental cross-correlation rates to back-calculated ones from the different ensembles. Cross-correlation measurements established by Bodenhausen [6] have been further refined [7]. Two complementary experiments are measured to derive the experimental cross-correlation and the experimental error. We use cross-correlations to illustrate self-consistency and improvement of the cross-correlation rates when going from single-state ensemble to multi-states ensembles, showing that the sampled conformational space is realistic (Figure S8). However, we thereby also illustrate how the NMR structural ensembles are more representative of the true structure than static X-ray structures (Figure S8). The simulated cross-correlation rates for the apo Pin1-WW of the single-state structural ensemble have a better Pearson's correlation coefficient (R 0.74, rmsq 0.29) than the ones derived from the holo X-ray structures (R 0.69, rmsq 1.39 Å and R 0.70, rmsq 0.88 Å) 1Pin.pdb (1.35 Å) [1] and 2ZQT.pdb (1.46 Å, M130A) [8], respectively (Figure S6). The cross-correlation rates of the two-states and three-states ensembles have an even better Pearson's correlation coefficient (R 0.80, rmsq 0.26 Å) and (R 0.85, rmsq 0.24 Å) (Figure S8). For higher than three-states structural ensembles the Pearson's correlation coefficient remains approximately constant at an average of R 0.82 +/- 0.01, with an average rmsq of 0.26 +/- 0.00 Å.

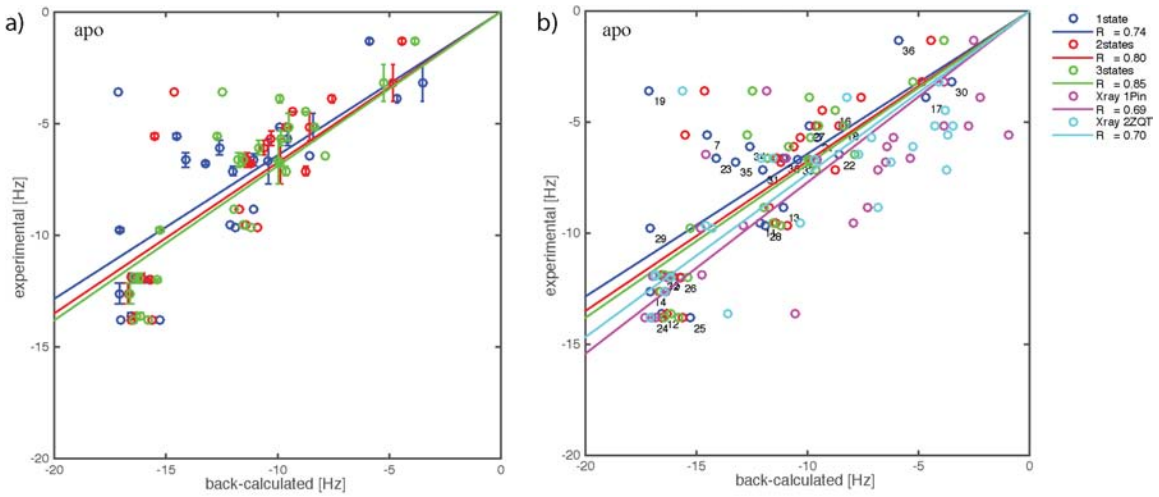


Fig. S8: Comparison of experimental to back-predicts cross-correlation rates for apo and holo Pin1-WW. a) Cross-correlation rates of the single-state, two-states ensembles and three-states ensembles with the respective experimental error bars given, when available. b) Both the two-states and three-states cross-correlation rates have a higher coefficient than those from the X-ray structures. Numbers are sequence specific residue labels.

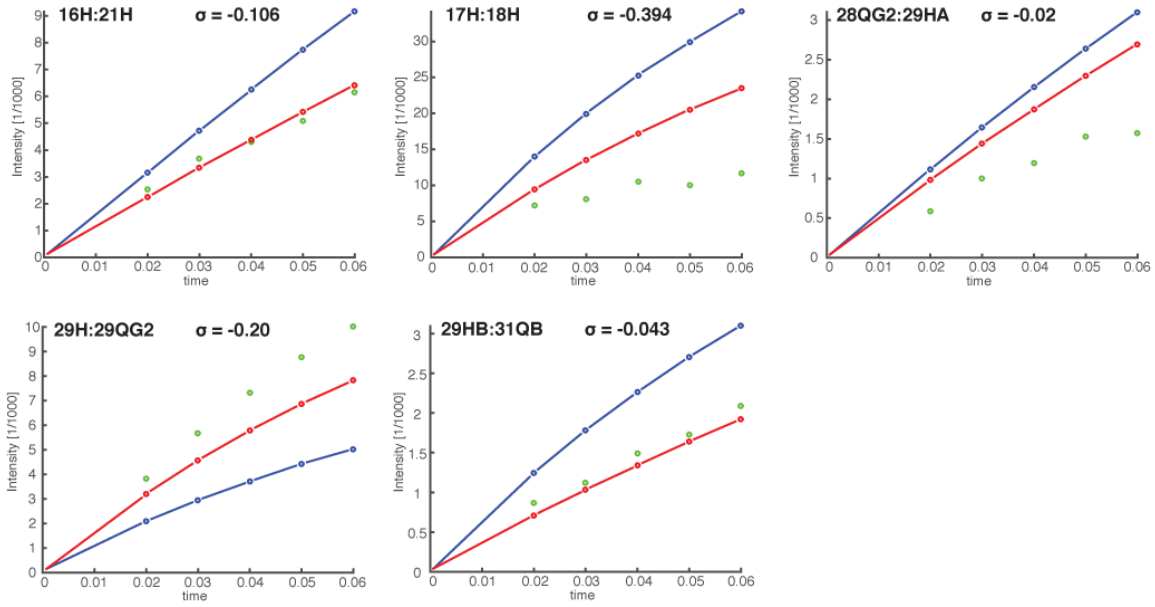


Fig. S9: Experimental bidirectional buildups (green) versus back-predicted buildups of representative NOEs for single-state (blue) and two-states (red) ensemble. The connecting lines are drawn to guide the eye. The two-states ensemble fulfills the data better than the single-state structure. Only when the difference between single-state and two-states buildups is large enough have the eNOEs an effect on the conformational space of the two-states bundle, since the cross relaxation rate is proportional $1/r^6$. The back-predicted buildups were calculated using eNORA2 [9, 10].

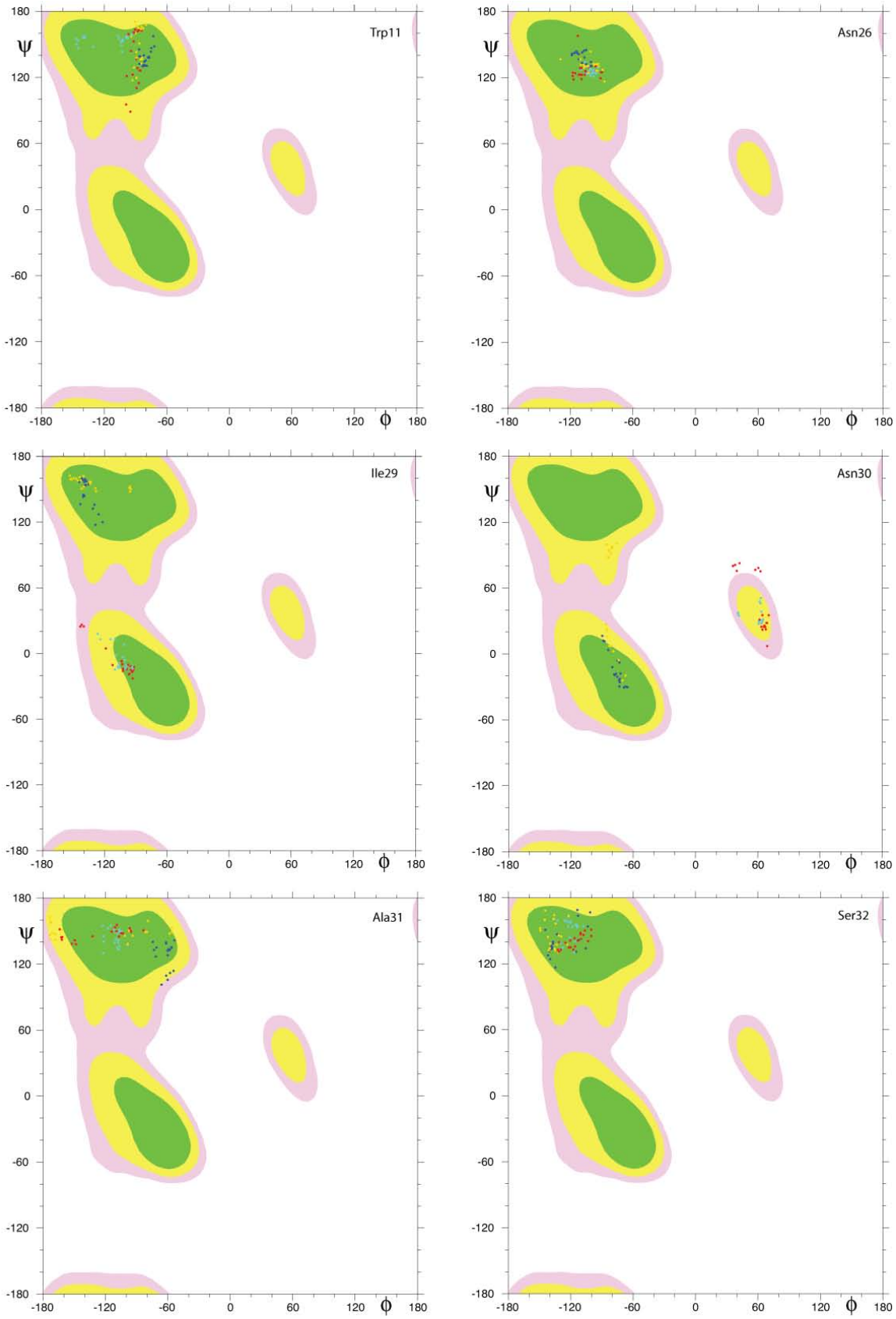


Fig. S10: Ramachandran plots of backbone Ψ and Φ angles for select residues, illustrating the loss of correlation upon ligand binding in residues following loop2 (28-30) at the beginning of β-strand3 (31-33). The dots are following the apo (cyan/blue) and holo (orange/red) color scheme. All Figures were created with molmol [11].

Ensemble analysis

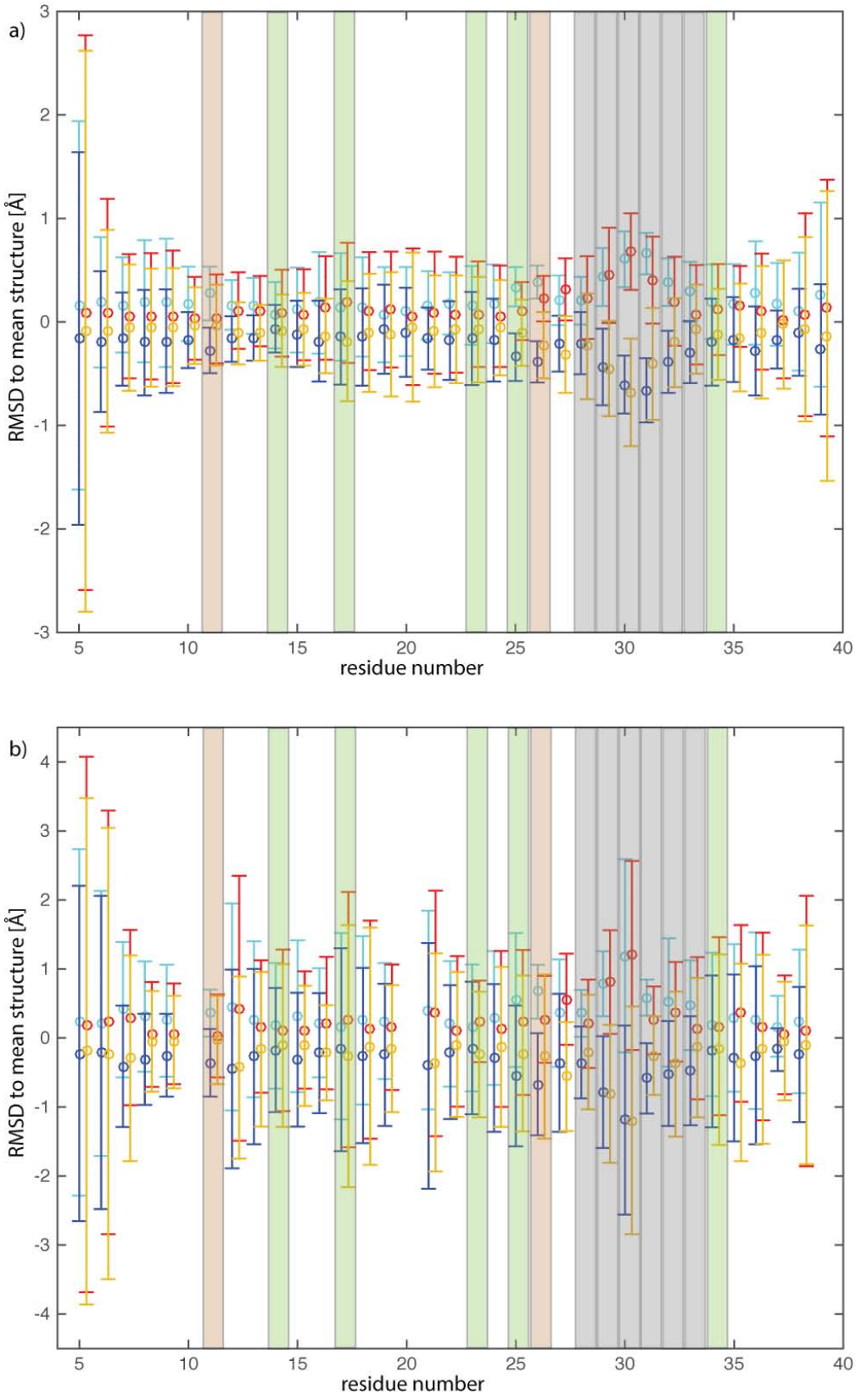


Fig. S11: RMSD to mean structure per residue illustrating the overall changes in conformational sampling space. Ligand binding site (green bars), Pin1-WW:PPIase interaction interface (grey bars) and binding event allosteric relay residues (red bars) are indicated. The apo (cyan/blue) are offset from the holo (orange /red) so both forms can be seen. a) Backbone RMSD and b) side-chain RMSD. The error bars are a measure for the precision of the states.

Structural input and statistics

Table S1: Structural statistics and Cyana input data for the apo Pin1-WW.

NMR distance and dihedral constraints						
Distance constraints						
Total eNOEs	729					
eNOEs from one pathways	450 ($NORM_{orig}$ and $NORM_{dest}$ [8])					
eNOEs from two pathway	279					
Intra-residue, $\| i - j \| = 0$	273					
Sequential, $\|i - j\| = 1$	188					
Short-range, $\| i - j \| <= 1$	461					
Medium-range, $1 < \|i - j\| < 5$	73					
Long-range, $\|i - j \| >= 5$	195					
Dihedral angle restraints						
$3J$HNα scalar couplings	26					
$3J$HαHβ scalar couplings	24					
$3J$HNCG scalar couplings (aromatic)	6					
$3J$HNCOCG scalar couplings (aromatic)	6					
13Cα chemical shifts	10					
	One-state ensemble	Two-states ensemble		Three-states ensemble		
Structure statistics						
Average CYANA target function value ($Å^2$)	15.43 ± 0.03	9.34 ± 0.08		8.10 ± 0.13		
Violations						
Distance constraints (> 0.5Å)	5	0		1		
Dihedral angle constraints (> 5°)	0	0		0		
Deviations from idealized geometry						
RMSD (Å)						
Backbone to mean	0.05 ± 0.03	0.55 ± 0.18		0.62 ± 0.13		
Heavy atoms to mean	0.67 ± 0.15	1.22 ± 0.18		1.23 ± 0.15		
		Cyan state	Blue state	Cyan state	Green state	Blue state
Backbone to mean		0.40 ± 0.16	0.42 ± 0.15	0.54 ± 0.11	0.57 ± 0.13	0.48 ± 0.12
Heavy atoms to mean		1.09 ± 0.18	1.17 ± 0.21	1.19 ± 0.17	1.18 ± 0.17	1.15 ± 0.16
RMSD to X-ray structure (Å)						
Backbone						
1Pin.pdb	0.62	0.73	0.98	0.85	0.87	0.97
2ZQT.pdb (M130A)	0.55	0.65	0.93	0.79	0.81	0.91
Heavy atoms						
1Pin.pdb	1.32	1.21	1.39	1.35	1.30	1.37
2ZQT.pdb (M130A)	1.26	1.19	1.40	1.34	1.30	1.37

Table S2: Structural statistics and Cyana input data for the holo Pin1-WW.

NMR distance and dihedral constraints						
Distance constraints						
Total eNOEs	727					
eNOEs from one pathways	458 ($NORM_{orig}$ and $NORM_{dest}$ [8])					
eNOEs from two pathway	269					
Intra-residue, $\| i - j \| = 0$	265					
Sequential, $\|i - j\| = 1$	171					
Short-range, $\| i - j \| <= 1$	436					
Medium-range, $1 < \|i - j\| < 5$	88					
Long-range, $\|i - j \| >= 5$	203					
Dihedral angle restraints						
^3JHNα scalar couplings	24					
^3JHαHβ scalar couplings	23					
^3JHNCG scalar couplings (aromatic)	6					
^3JHNCOCG scalar couplings (aromatic)	6					
13Cα chemical shifts	9					
	One-state ensemble	Two-states ensemble		Three-states ensemble		
Structure statistics						
Average CYANA target function value ($Å^2$)	22.24 ± 0.20	11.37 ± 0.23		9.44 ± 0.21		
Violations						
Distance constraints (> 0.5Å)	9	0		1		
Dihedral angle constraints (> 5°)	0	0		0		
Deviations from idealized geometry						
RMSD (Å)						
Backbone to mean	0.07 ± 0.06	0.59 ± 0.15		0.66 ± 0.15		
Heavy atoms to mean	0.80 ± 0.20	1.28 ± 0.19		1.32 ± 0.18		
		Red state	Orange state	Red state	Pink state	Orange state
Backbone to mean		0.57 ± 0.16	0.53 ± 0.14	0.54 ± 0.13	0.61 ± 0.15	0.54 ± 0.14
Heavy atoms to mean		1.26 ± 0.21	1.23 ± 0.21	1.23 ± 0.20	1.27 ± 0.19	1.21 ± 0.21
RMSD to X-ray structure (Å)						
Backbone						
1Pin.pdb	0.44	0.56	0.62	0.64	0.61	0.70
2ZQT.pdb (M130A)	0.44	0.54	0.60	0.64	0.60	0.69
Heavy atoms						
1Pin.pdb	1.51	1.23	1.31	1.26	1.28	1.39
2ZQT.pdb (M130A)	1.56	1.27	1.36	1.34	1.35	1.44

Experimental section

Expression and purification of Pin1

The gene of full-length Pin1 protein (S18N-W34F mutation) was bought from genescript, sub-cloned into a pET28a vector containing an N-terminal His-tag with a thrombin cleavage site (MHHHHHHLVPRGS). For expression the cDNA was transformed into *E. coli* BL21 cells and plated on a kanamycin-containing plate (50ug/ml). The cells were grown over-night at 37˚ C and then used to inoculate a 10 ml pre-culture. The culture was grown at 37˚ C for three hours and thereafter used to inoculate a 1-liter culture (kanamycin 50ug/ml) of M9 medium for $^{15}N/^{13}C$- or ^{15}N-labeling. Cells were grown to an OD_{600} of 0.9. Protein expression was initiated by adding 1 mM IPTG (isopropyl β-D-thiogalactopyranoside). The Cells were then allowed to express over night at 18˚C. The Cells were then harvested by spinning at 5,000 *g* for 15 minutes and re-suspended in purification buffer (10 mM Tris/HCl, 200mM NaCl). The Cells were lyophilized and spun at 40,000 *g* for 20 minutes. The supernatant was filtered (0.4 μm and 0.2 μm filters) and loaded onto a nickel (II)-charged chelating sepharose FF column (Amersham Biosciences), equilibrated with purification buffer as above and washed with 400 ml of the same buffer. The bound protein was eluted with 250 mM imidazole at pH 7.9, in aliquots of 10 ml. Fractions containing partially pure proteins were pooled, desalted and passed through a DEAE column equilibrated with purification buffer. Pure Pin1 was collected as flow-through. The purity was checked on SDS PAGE stained with coomassie brilliant blue. The pure protein preparation was concentrated to experimental concentration of 0.4mM. The concentration was determined by absorption measurements using the molar absorption coefficient.

ITC measurements

Isothermal Titration Calorimetry (ITC) experiments were performed in 20 mM sodium phosphate, pH 6.5 at 5˚ C on an iTC200 (Malvern Instruments) according to the instructions of the manufacturer. Protein concentrations were measured using the absorbance at 280 nm. Prior to the measurement, the peptide and Pin1-WW domain were dialyzed against the same buffer to minimize artifacts due to buffer mismatch.

Chemical shift perturbations of full length Pin1

The [^{15}N, ^{1}H]- HSQC for full length Pin1 was recorded with 128(t_1) x 1024(t_2) real points on a Bruker 700 MHz spectrometer. The maximal evolution times were $t_{1max,1H}$ = 51.5 ms, $t_{2max,15N}$ = 81.1 ms. The time-domain data were multiplied with a squared cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to 1024 x 256 real points.

NMR experiments were done at 25˚ C in a 20 mM sodium phosphate, pH 6.5, 5 mM DTT, 0.01 % $NaNO_3$ in 10 % D_2O. Binding experiments were done as follows; for a fixed amount of labeled ^{15}N/^{13}C Pin1 (300 μM), increasing amount of 2.4 mM peptide was titrated into the protein solution and the extent of binding determined by measuring the chemical shift change in the HSQC spectrum. Full saturation was obtained when the perturbation ($(\Delta\delta^{15}N^2 + \Delta\delta\ ^{1}H_N^2)^{1/2}$) was close to zero.

STD experiments

The 2D STD [15N, 1H]-HSQC were recorded at 5˚ C with 128(t1) x 4096(t2) real points on a Bruker 700 MHz spectrometer. On-resonance spectra were irradiated at 10.3 ppm for 1 s at 77 Hz power using gaussian pulses train of 50 ms each, while off-resonance spectra were irradiated identically but with an offset of 100 ppm. The maximal evolution times were t1max, 1H = 51.2ms ms, t2max,15N = 28.8 ms. Exponential (EM, with LB 25 Hz) window functions were applied in both dimensions to the fid and zero-filled to 1024 x 256 real points. The final 2D STD spectrum is the difference between the off-resonance and the on-resonance irradiated [15N, 1H]-HSQC spectra.

References

[1] Ranganathan R, Lu KP, Hunter T, Noel JP, Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent, Cell 1997, 89(6), 875-86.

[2] Wilson KA, Bouchard J, Peng J, Interdomain interactions support interdomain communication in human Pin1, Biochemistry 2013, 52, 6968-81.

[3] Peng J, Investigating Dynamic Interdomain Allostery in Pin1, J Biophys Rev 2015, 7, 239-249.

[4] Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel J, Structural basis for phosphoserine-proline recognition by group IV WW domains, Nature 2000, 7, 639-643.

[5] Vögeli B, Güntert P, Riek R, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol Phys 2013, 111(3), 437-454.

[6] Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G, Simultaneous determination of Ψ and Φ angles in proteins from measurements of cross-correlated relaxation effects, J Biomol NMR 1999, 14(3), 277-280.

[7] Fenwick RB, Schwieters CD, Vögeli B, Direct Investigation of Slow Correlated Dynamics in Proteins via Dipolar Interactions, JACS 2016, 138(27), 8412-21.

[8] Jobichen C, Liou YC, Sivaraman J, Structural studies on PIN1 mutants, structure deposition: 2ZQT.pdb.

[9] Orts J, Vögeli B, Riek R, Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs, J Chem Theory Comput 2012, 8(10), 3483-3492.

[10] Strotz D, Orts J, Chi C, Riek R, Vögeli B, The eNORA2 exact NOE analysis program, to be submitted.

[11] Koradi R, Billeter M, Wüthrich K, MOLMOL: A program for display and analysis of macromolecular structures, J Mol Graphics 1996, 14, 29–32.

10.5. Appendix to chapter 8: A structural ensemble of the enzyme cyclophilin reveals an orchestrated mode of action at atomic resolution

Materials and methods

Peptides

The unlabeled peptides Dansyl-AAPF and Dansyl-AAP$_{(trans)}$F (analogue proline in trans conformation) were bought from GL Biochem Shanghai.

Sample preparation

A detailed description of the sample preparation can be found in Chi CN, Vögeli B, Bibow S, Strotz D, Orts J, Güntert P, Riek R, A Structural Ensemble of the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution, Angew Chem Int Ed 2015, doi: 10.1002/anie.201503698.

Data collection and analysis

All NMR experiments were acquired on Bruker 600, 700, and 900 MHz spectrometers equipped with triple resonance cryogenic probes at 299 K (except otherwise stated) in 10 mM sodium phosphate, pH 6.5, 5 mM DTT and 0.01 % $NaNO_3$. Protein samples were dissolved in 3% D_2O. For assignment purposes 3D HNCACB, ^{15}N-resolved [^{1}H, ^{1}H]-NOESY-HSQC, ^{15}N-resolved HMQC-[^{1}H, ^{1}H]-NOESY, ^{15}N-resolved [^{1}H, ^{1}H]-TOCSY-HSQC and ^{13}C-resolved HCCH-TOCSY 3D spectra were recorded for apo cyclophilin [1]. Four sets of $^1D_{HN}$ residual dipolar couplings (RDC) were determined from pairs of 3D HNCO-based or IPAP-HSQC experiments in the presence and absence of an alignment medium [2]. We measured the RDCs in four different alignment media (i.e. pf1 phages, C*n*E*m*/hexanol, C*n*E*m*/hexanol doped with CTAB, and neutral gels) and obtained a total of 396 H^N-N RDCs from which we determined the alignment tensors. $^3J_{HNHA}$, $^3J_{HNCO}$, and $^3J_{HNCB}$ scalar couplings were obtained from a 2D ct-HMQC-J frequency modulated experiment [1,3] and 3D HNCA-E.COSY-type experiments [4] in protonated and perdeuterated samples, respectively. NOE buildup rates were followed with 3D ^{15}N/^{13}C-resolved [^{1}H, ^{1}H]-NOESY-HMQC experiments (protein concentration range 1.0-1.5 mM) using mixing times of 15, 20, 25, 30, 40, and 50 ms. The upper and

lower distance restraints were determined with the use of the eNORA software package [5] following five established steps: i) the diagonal intensity I_{diag} of each residue was fitted mono- exponentially to back-predict the intensity at zero mixing time $I_{diag(0)}$ and to get the auto-relaxation rate. ii) The initial intensity $I_{diag(0)}$ was then used to normalize the cross peak intensity I_{cross}. iii) These normalized I_{cross} values measured at various mixing times were fitted to determine the cross-relaxation rate σ_{ij} (Fig. S6) iv) A correction for spin diffusion effects was estimated from simulations of apparent cross-relaxation rates from an input structure assuming the full relaxation matrix method [5]. The input structure used was the lowest energy conformer from ref [6] pdb; 1oca. v) The upper and lower distance restraints were then set following the established protocol [7] (Supplementary Table S1). Magnetically equivalent protons were treated by r^{-6}-summation rather than a pseudo-atom approach. R_1, $R_{1\rho}$, and R_2 relaxation rates (needed for the determination of τ_c) were determined for backbone H^N. Data processing and analysis were done with NMRpipe [8] and CCPnmr [9].

Equilibrium binding experiments

A detailed description of equilibrium binding experiments can be found in Chi CN, Vögeli B, Bibow S, Strotz D, Orts J, Güntert P, Riek R, A Structural Ensemble of the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution, Angew Chem Int Ed 2015, doi: 10.1002/anie.201503698.

Structure calculation and evaluation

The following input files were used for structure determination: Upper and lower distance restraints, $^3J_{HNHA}$, $^3J_{HNCO}$ and $^3J_{HNCB}$ scalar couplings [11], four sets of $^1D_{HN}$ RDCs, and conservative phi dihedral angle restraints from Cα chemical shifts with zero weight in the final part of the calculation. Structural coordinates were determined with version 3.96 of the program CYANA [12]. Calculations were done with 200'000 torsion angle dynamics steps for 200 conformers with random torsion angles by simulated annealing. The 20 conformers with the lowest final target function values were selected and analyzed. For ensemble-averaged calculations, 2–6 structural states were calculated simultaneously. A weak harmonic well potential with bottom width of 1.2 Å was used to keep identical heavy atoms from the different states together following the proposal by Vögeli et al. [12b]. However, for residues 64-74 and the residues 54-55 closest in space to residues 64-75 this potential was not used because its use resulted in a too constrained conformational space and significant violations of experimental restraints.

To test for self-consistency of the experimental data, a cross validation test was performed with a jackknife procedure that repeats structure calculation ten times with 10% of the experimental input data randomly deleted such that each distance restraint is omitted exactly once. Furthermore, calculations were repeated omitting the 3 RDC data sets from pf1, neutral polyacrylamide gel and neutral PEG alignment media for reconstruction of the RDC target function, and omitting the $^{3}J_{\mathrm{HNHA}}$ couplings for reconstruction of the $^{3}J_{\mathrm{HNHA}}$ coupling target function, respectively.

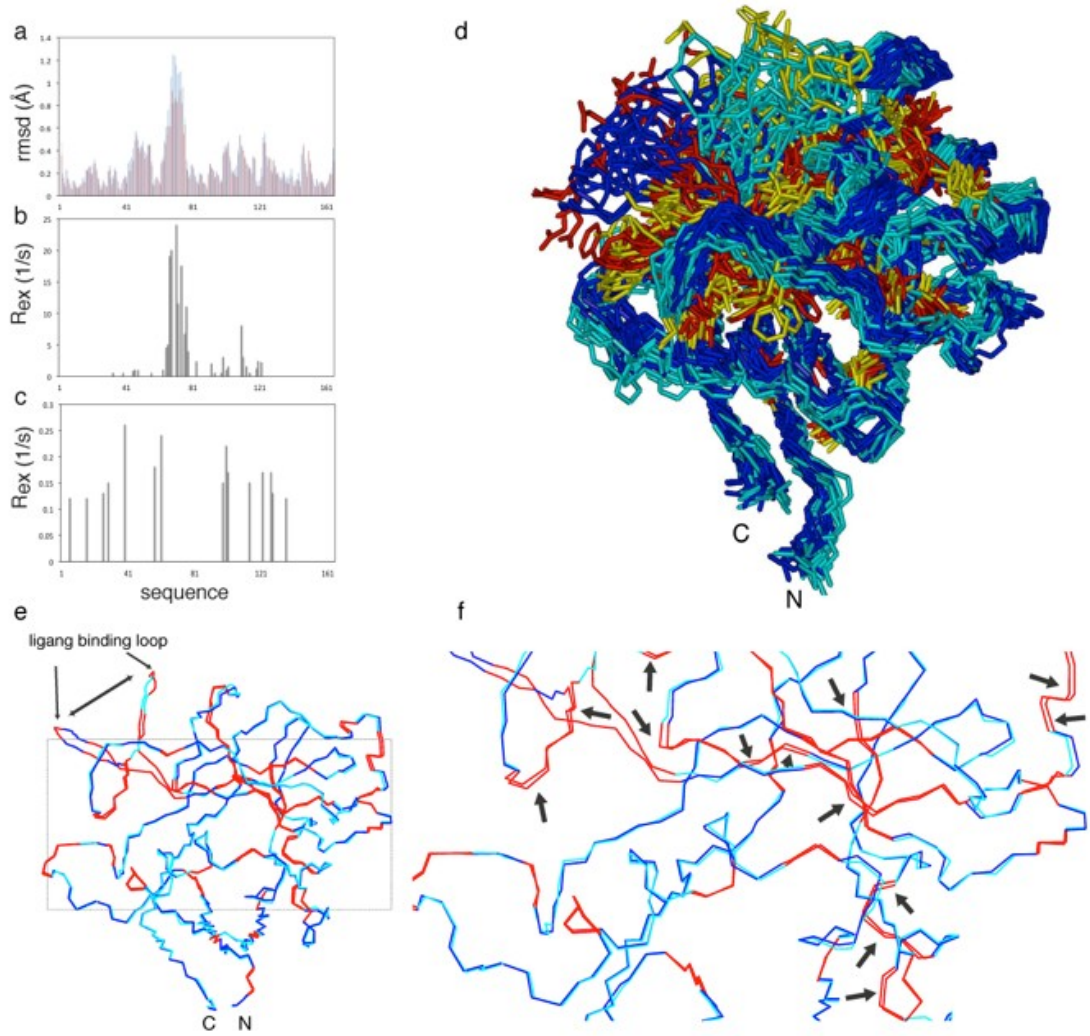


Figure S1. Correlation between the relaxation data of cyclophilin by the Kern group [13] with the structural perturbation of the presented two-states ensemble. (a) Root mean square deviation (r.m.s.d) between the two-state ensemble is plotted as a function of residue number highlighting in particular the large conformational difference of the loop comprising residues 64-74. Red bars are the r.m.s.d for the backbone atoms and blue bars r.m.s.d for the heavy side-chain atoms b) (amides) and c) methyls CPMG relaxation data (R_{ex}) for both (b) the backbone ^{15}N-^{1}H and (c) methyl moieties obtained by the Kern group [13]. (d) Side chains in the two states ensemble are shown if the residues undergo slow conformational exchange dynamics determined by CPMG relaxation data [13] shown in (b) and (c). The same color code for the two states ensemble is used as in Figure 2 with cyan for the closed and blue for the open state, respectively. (e) and (f) the CPMG data are mapped onto the backbone of the 2 states ensemble. In (e) the average backbone trace from each state is color-coded in accordance to Figure (d) and the backbone trace of residues that show slow conformational exchange dynamics from (b) and (c) are marked in red. (f) Expansion of (e) with arrows showing concerted patterns that match the relaxation data.

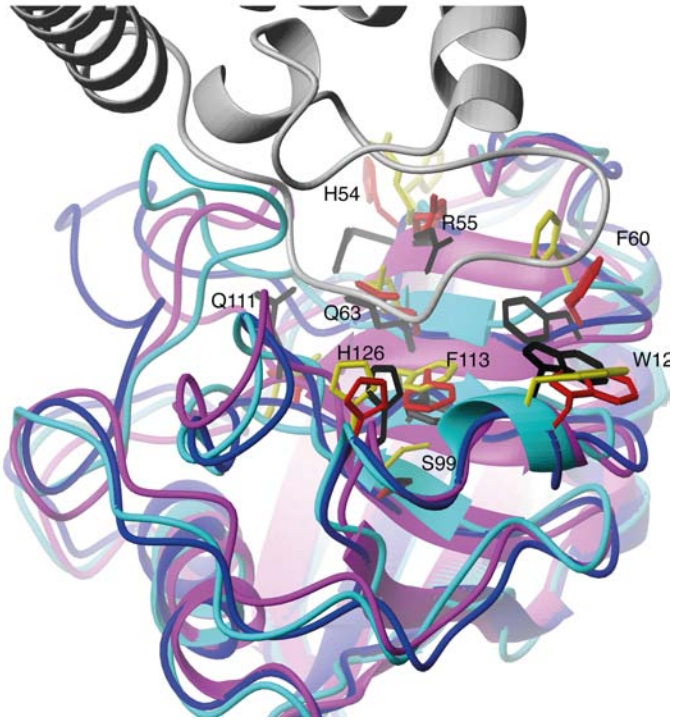


Figure S2. The closed state cannot accommodate the ligand as shown by the virtual conformational clash in the superposition of the two-states ensemble with the X-ray structure of cyclophilin in complex with the HIV-1 capsid protein. The X-ray structure from Cyclophilin A in complex with the HIV-1 capsid protein (pdb: 1ak4) was superimposed with the presented two-state ensemble. As a representative of the NMR two states structure the lowest energy two states conformers were selected. Following the color code of Figure 3 (chapter 8), the backbone of the closed state is color coded with cyan, the open state is color coded with blue, and the x-ray structure (pdb:1ak4) is color coded with magenta for the backbone ribbon and grey for the HIV capsid protein, respectively.

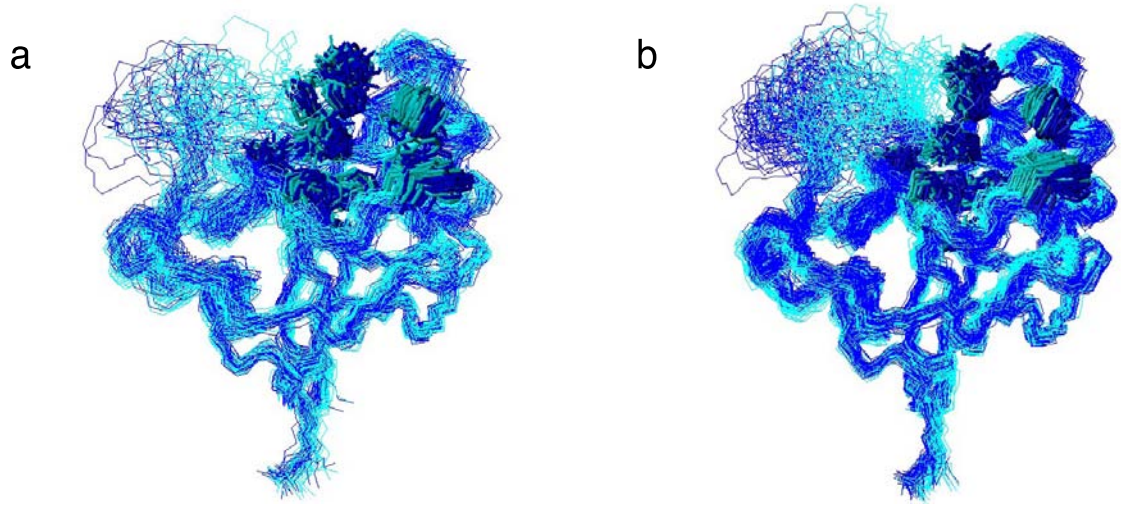


Figure S3. The two states are also present in the three and the six states ensembles of cyclophilin. To illustrate the presence of the open and closed states in the three (a) and six (b) states ensemble calculations, the backbone of the closed and open states are color coded in accordance with Figure 3 (chapter 8) with cyan for the closed and blue for the open state. Therefore, every state is matched to one of the states of the two-states structure. Side-chains of a few active site residues are shown in addition to the backbone. The finding that the two states observed in the two states ensemble structure calculation (Figure 1-3 in chapter 8) are also present in the three and six states structure calculations exclude the potential problem that the presence of two states and their correlated conformational pattern throughout the protein structure are an artifact of constraining the number of states to two.

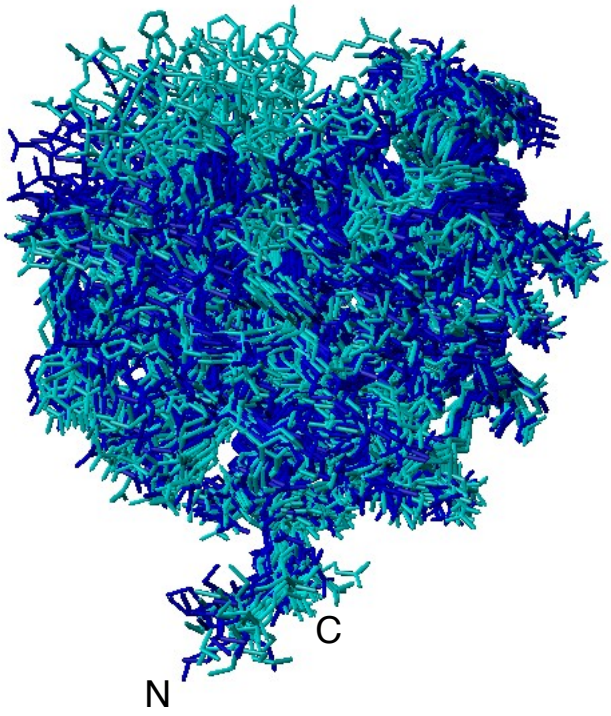


Figure S4. Widespread distinction between the open and closed states of the cyclophilin two-states ensemble. To illustrate the that the two distinct states are observed throughout most of the structure, the backbone and side chains of both states are shown on all the 20 conformers of the two states structure and color coded with blue for the open state and cyan for the closed, respectively.

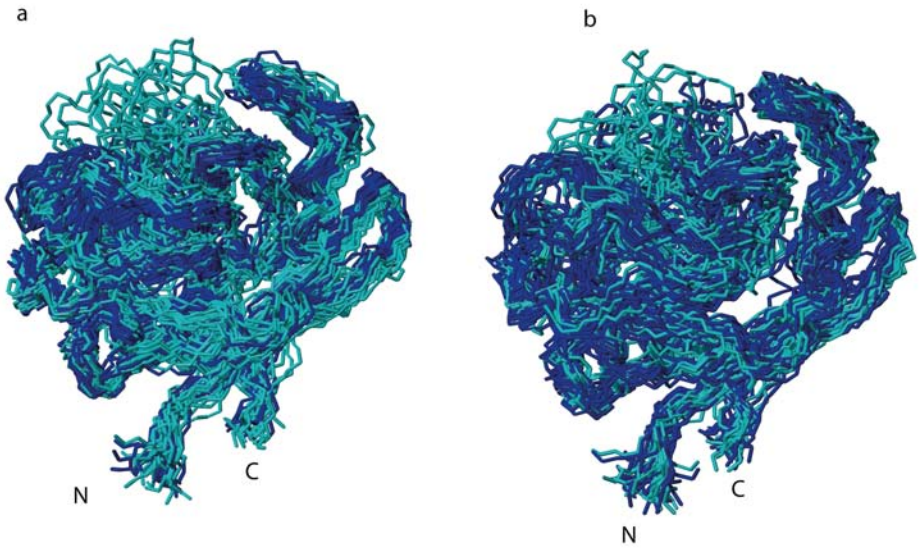


Figure S5. Two-states structure calculations with a decreased number of distance restraints. (a) Two-states structure ensemble of cyclophilin A without the distance restraints often violated in the two-states structure calculation listed in Table S4. Since the two states are preserved the violated restraints appear to have no influence. (b) Two-states structure ensemble of cyclophilin A without the distance restraints often violated in the single state structure calculation listed in Table S3. Since the two-states separation is lost the distance restraints listed in Table S3 are critical. The backbone of the 2 x 20 conformers shown were color coded with blue for the open state and cyan for the closed, respectively.

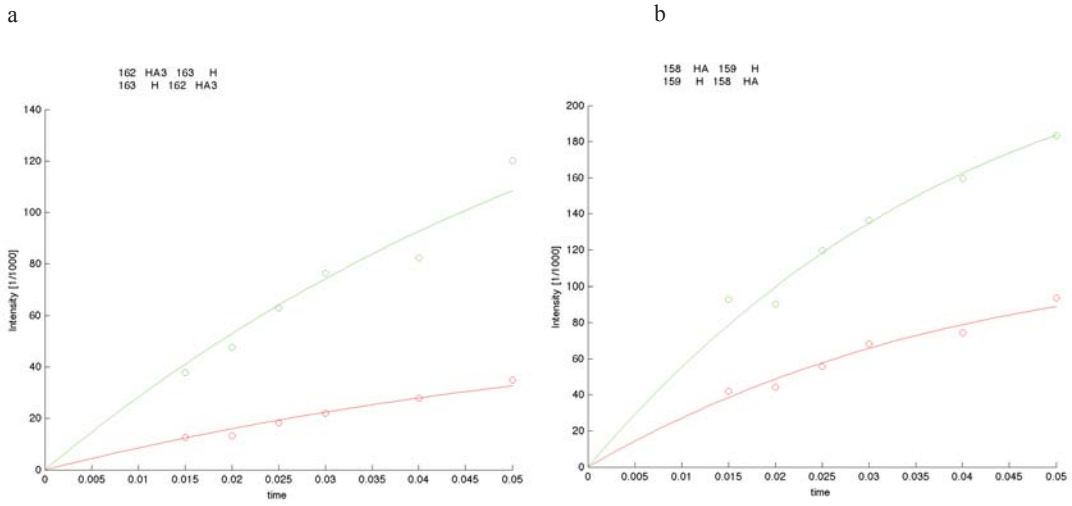


Figure S6. NOE buildup curves. For two examples indicated on top, the NOE buildup curves (i.e. NOE cross peak intensity versus mixing time) from both NOE cross peaks from both sides of the diagonal (color coded green and red) of the ^{15}N, ^{13}C-resolved [^{1}H, ^{1}H]-NOESY spectra are shown. The intensity scales are arbitrarily chosen but identical for both buildups. The figure is an automatic output from the eNORA2 software [14].

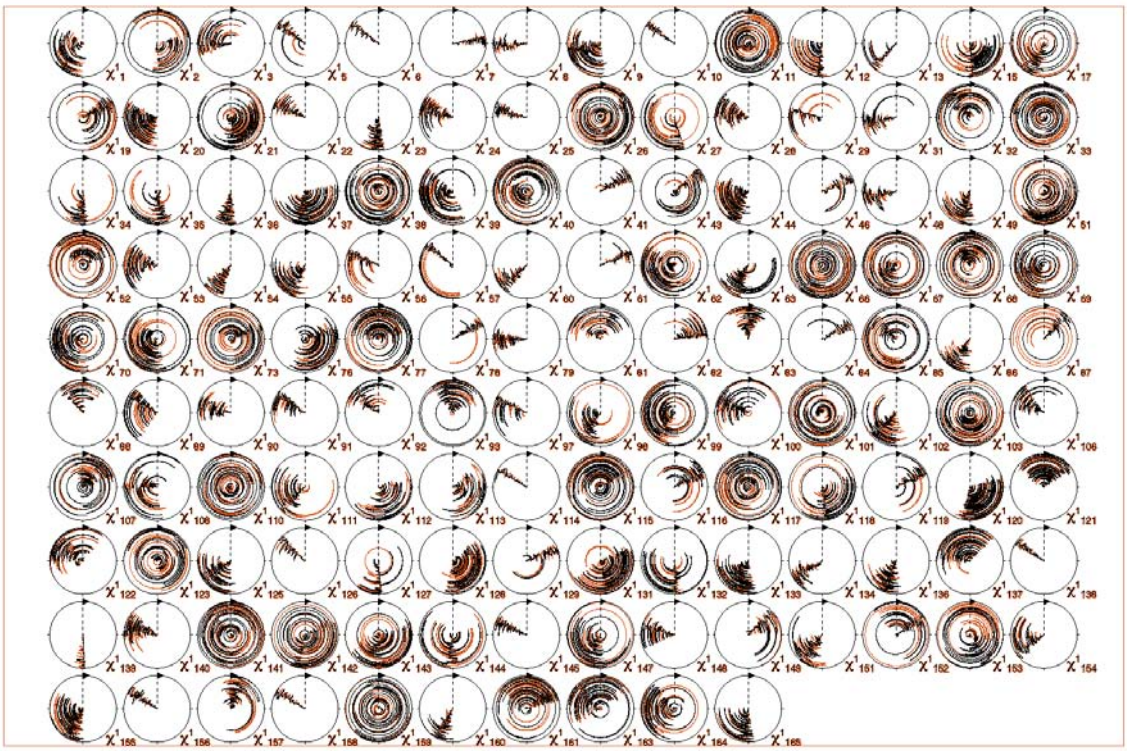


Figure S7. Comparison of the distribution of the X1 angles for the open (black) and closed (red) states. For each states the angles of 20 conformers are plotted from the center to the outer circle in the circle diagrams. For most residues, the rotamer distribution between the open and closed state are similar. The figure was generated in MolMol [15].

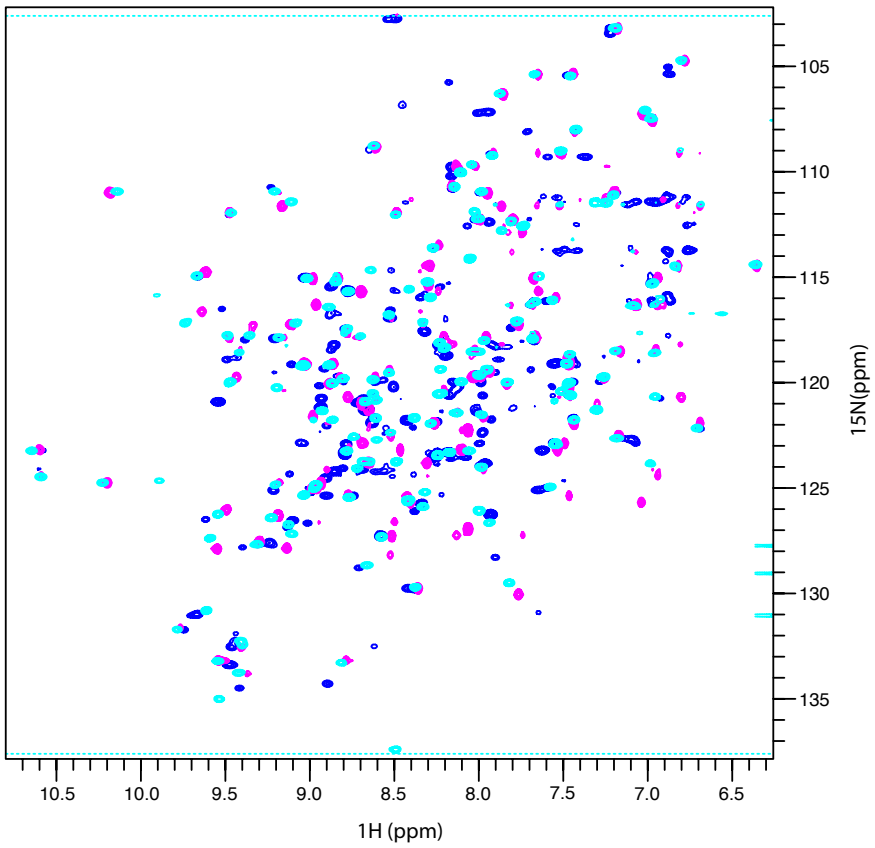


Figure S8. [^{15}N, ^{1}H]-TROSY spectrum of the free cyclophilin A (cyan), G64A_G65A mutant (blue) and cyclophilin in complex with the Trans_peptide (purple). The overall similarity between the [^{15}N, ^{1}H]-TROSY spectra of the free cyclophilin A (cyan) and the G64A_G65A mutant (blue) indicates that the mutant is similarly folded as wild-type.

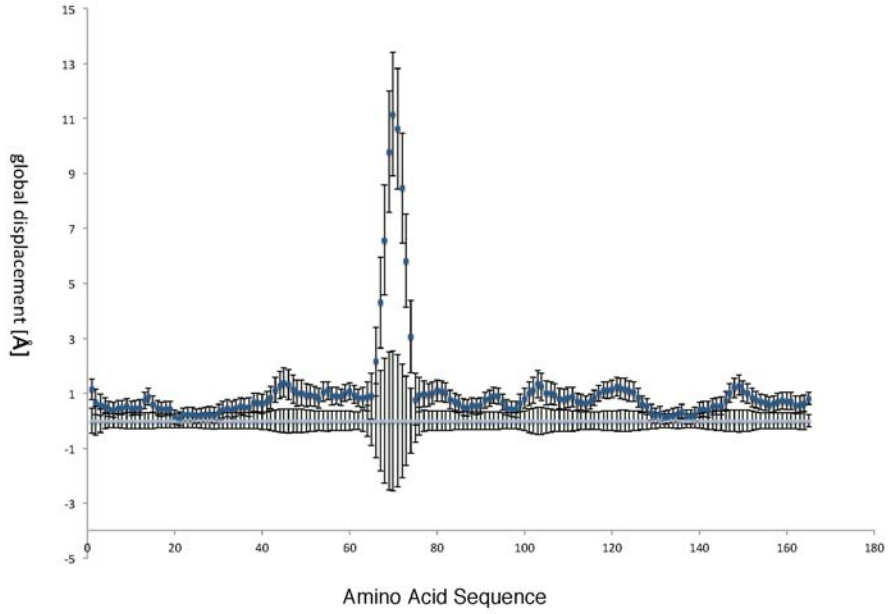


Figure S9. For the six states ensemble the backbone global displacement between the two mean structures of the two substates identified in the two states structure calculation are plotted against the amino acid sequence. The backbone global displacement between the mean structures of the two substates are plotted against the amino acid sequence. The error bars are the root mean square deviations (r.m.s.d.) of the individual sub-states over the backbone atoms over the residue of interest and its neigbouring residues. The r.m.s.d. was calculated with the software package MolMol [15].

Tables

Table S1. Structural statistics and NMR distance and dihedral constraints

	Two-state ensenble	One-state ensemble
Distance constraints		
Total eNOEs	3629	3629
eNOEs from two pathways	326	326
eNOEs from one pathways	1159	1159
Generic normalized eNOEs (correct upper limit)	2144	2144
Total residual dipolar coupling restraints (^{15}N-$^{1}H_N$)	396	396
Total dihedral angle restraints	407	407
$^{3}J_{H_{N\alpha}}$ scalar couplings	82	82
$^{3}J_{H_{NC'}}$ scalar couplings	100	100
$^{3}J_{HNC\beta}$ scalar couplings	97	97
$^{13}C_{\alpha}$ chemical shifts	128	128
Structure statistics		
Average CYANA target function value (Å^2)	66.76 +/- 1.95	137.85 +/- 4.04
Violations		
Distance constraints (>0.5 Å)	10	130
Dihedral angle constraints (>5°)	0	0
Deviations from idealized geometry		
Bond length (Å)	0.001	0.001
Bond angles (°) Impropers (°)	0.2	0.2
	0.0	0.0
Average pairwise r.m.s. deviation (Å)		
Heavy atoms	1.29 +/- 0.06	1.11 +/- 0.07
Backbone	0.85 +/- 0.06	0.71 +/- 0.07
R.m.s. deviation to RDC-refined X-ray structure (Å)		
Heavy atoms	1.84	1.75
Backbone	1.29	1.21

Table S2. Root mean square deviation (r.m.s.d.) of the individual and combined states in the two states ensemble

Region/state	Residues 60-80 (backbone r.m.s.d. (Å))	Residues 60-80 and active site residues* (backbone r.m.s.d. (Å))	Active site residues* (heavy atom r.m.s.d. (Å))
Closed State (color coded cyan in the Figures)	2.5	2.45	1.5
Open State (color coded blue in the Figures)	2.28	2.19	1.58
Both States	4.93	5.13	2.91

* residues 54,55,60,63,99,111,113,121,126

Table S3. eNOE-derived distance restraint violations in the one-state structure calculation

Atom	Res	Res no	Atom	Res	Res no	Distance (Å)	no[a]	Avg(Å)[b]	Max (Å)[c]
HA	THR	5	H	LEU	24	4.25	6	0.28	0.96
HB	THR	5	QB	LEU	164	2.83	16	0.31	0.86
QQG	VAL	6	HB2	PHE	22	2.63	9	0.19	0.54
QG1	VAL	6	QB	ALA	26	5.58	20	0.58	0.77
QG2	VAL	6	QD	PHE	22	2.68	14	0.26	0.6
QG2	VAL	6	HZ	PHE	22	5.43	15	0.37	0.67
HB3	PHE	7	QD1	LEU	164	2.96	12	0.25	0.53
HZ	PHE	7	H	GLY	162	4.83	11	0.23	0.51
H	ASP	9	QG2	ILE	158	3.01	15	0.28	0.56
HB3	ASP	9	H	ASP	160	2.6	20	0.42	0.57
HB3	ASP	9	HB3	ASP	160	3.66	18	0.35	0.5
HB	ILE	10	QG2	ILE	138	2.65	15	0.33	0.57
HB	ILE	10	QB	ALA	141	5.7	20	0.66	0.98
HG12	ILE	10	QG2	VAL	20	2.78	18	0.32	0.62

QB	ALA	11	HA2	GLY	14	3.35	20	0.39	0.62
QG1	VAL	12	QD1	ILE	138	6.85	20	0.58	0.73
HB3	PRO	16	H	ALA	159	5.72	11	0.24	0.51
HG2	PRO	16	HD2	PRO	16	2.93	20	0.63	0.63
H	LEU	17	HG	LEU	17	3.89	10	0.26	0.62
HB2	LEU	17	H	GLY	18	2.99	10	0.37	0.79
HB2	LEU	17	H	GLY	18	2.99	9	0.17	0.54
H	ARG	19	QG2	ILE	138	4.31	14	0.28	0.53
H	VAL	20	HB	VAL	20	2.89	14	0.42	0.77
H	VAL	20	QG2	VAL	20	2.94	20	0.74	0.95
HA	VAL	20	QG1	VAL	20	2.63	20	0.43	0.5
QG1	VAL	20	QE	PHE	22	3.82	20	0.43	0.72
QG1	VAL	20	QD1	ILE	56	6.75	20	0.82	1.07
HA	SER	21	HB3	SER	21	2.83	14	0.33	0.55
HB3	PHE	22	QD	PHE	36	5.79	6	0.16	0.5
QE	PHE	22	QD1	LEU	98	4.19	10	0.25	0.59
HA	LEU	24	QD2	LEU	24	2.75	20	0.6	0.74
QD2	LEU	24	QD1	LEU	98	3.77	7	0.14	0.5
QD2	LEU	24	QD2	LEU	98	4.36	7	0.16	0.67
QD2	LEU	24	QG2	VAL	128	6.25	20	0.76	1.05
QB	ALA	26	HB3	ARG	37	5.15	20	0.73	0.98
QG1	VAL	29	HA	ILE	89	6.19	18	0.32	0.54
QG1	VAL	29	HA	SER	99	5.85	20	1.15	1.47
QG2	VAL	29	QG2	ILE	89	5.26	18	0.41	0.67
H	THR	32	HB	THR	32	2.49	16	0.35	0.56
H	LEU	39	HG	LEU	39	2.56	13	0.31	1.21
HB2	LEU	39	HB3	LYS	44	4.84	20	0.82	1.27
QD2	LEU	39	QD	TYR	48	5.55	7	0.2	0.9
H	SER	40	HB3	SER	40	2.92	18	0.38	0.7
HB3	SER	40	H	GLY	162	3.67	17	0.27	0.51
HA	THR	41	HA3	GLY	162	3.91	12	0.28	1.01
HB2	LYS	44	HA2	GLY	47	6.19	18	0.53	0.97
H	TYR	48	H	LYS	49	3.3	10	0.34	1.28
QD	TYR	48	H	SER	51	3.07	16	0.53	1.02
HA	LYS	49	H	GLY	50	2.22	20	1.26	1.35
HD2	LYS	49	HE2	LYS	49	2.94	17	0.34	0.64
HE3	LYS	49	HB2	CYS	161	3.16	19	0.54	0.85
H	GLY	50	HA3	GLY	50	2.39	8	0.16	0.55
H	GLY	50	H	SER	51	2.64	13	0.35	1.1
H	GLY	50	HA	ALA	159	5.97	19	0.68	1.22
HB2	PHE	53	QE	MET	142	6.08	11	0.22	0.65
QE	PHE	53	QE	MET	142	4.25	18	0.5	0.79
H	ARG	55	HG12	ILE	56	5.22	10	0.29	0.8
HA	ARG	55	H	ILE	56	2.32	7	0.26	0.84

H	ILE	56	HA3	GLY	150	4.43	12	0.25	0.52
QG2	ILE	56	HA	GLU	143	2.59	8	0.2	0.53
QG2	ILE	56	HG2	ARG	148	6.49	18	0.52	0.81
QD1	ILE	56	HD3	LYS	155	5.77	20	0.51	0.8
H	ILE	57	H	MET	61	3.17	20	0.76	1.05
QG2	ILE	57	HD2	PRO	58	3.24	19	0.38	0.66
HG2	PRO	58	H	GLY	146	5.1	17	0.44	0.74
HA	PHE	60	QG2	THR	119	3.5	18	0.43	0.66
HB2	MET	61	QG	MET	61	2.7	14	0.29	0.57
HB2	MET	61	H	CYS	62	2.04	17	0.37	0.56
H	CYS	62	H	ILE	114	3.48	16	0.39	0.89
HB3	CYS	62	QE	MET	142	3.49	19	0.45	0.85
H	GLY	64	H	PHE	112	4.9	8	0.18	0.52
HB2	ARG	69	H	GLY	74	5.33	7	0.19	0.7
H	THR	73	H	GLY	74	2.39	13	0.33	0.81
HB3	LYS	76	QD	LYS	76	3.01	16	0.32	0.53
HG3	LYS	76	HA	ILE	78	4.28	18	0.38	0.73
QD	LYS	76	HE2	LYS	76	3.03	20	0.6	0.72
H	SER	77	HA	LYS	82	3.67	7	0.17	0.51
H	ILE	78	H	TYR	79	2.71	15	0.3	0.5
H	LYS	82	QG	LYS	82	3.48	20	0.63	1.31
HA	LYS	82	QG	LYS	82	3.72	20	0.34	0.77
HB2	LYS	82	HD2	LYS	82	2.51	8	0.19	0.96
HB2	LYS	82	QG2	THR	107	3.54	9	0.22	0.55
QG	LYS	82	HB	THR	107	3.31	14	0.32	0.58
QE	PHE	83	HA3	GLY	109	4.55	11	0.28	1.07
H	PHE	88	HA	ILE	89	3.68	20	0.68	0.78
H	LEU	90	H	HIS	92	6.61	20	0.59	0.88
H	THR	93	QG2	ILE	97	3.97	11	0.25	0.6
HA	PRO	95	HA	THR	116	3.19	19	0.39	0.71
HA	PRO	95	HB3	LYS	118	5.63	19	0.74	1.04
HA	ILE	97	HA	LYS	131	2.88	12	0.24	0.57
H	LEU	98	HB3	LEU	98	2.84	15	0.34	0.81
H	LEU	98	H	GLY	130	3.19	19	0.45	0.68
HB2	LEU	98	QD2	LEU	98	3.07	8	0.18	0.51
HB2	LEU	98	H	GLY	130	3.76	10	0.28	1.05
QD2	LEU	98	QG1	VAL	132	6.83	18	0.53	1.08
H	SER	99	H	PHE	112	4.96	20	1.2	1.62
H	MET	100	HB	VAL	127	5.13	10	0.41	1.16
QB	ALA	101	QD	PHE	113	2.8	6	0.15	0.52
HA	ASN	102	H	GLY	109	5.81	12	0.26	0.8
HB2	ASN	102	H	THR	107	6.08	7	0.17	0.7
HB3	ASN	102	HA	ASN	108	4.45	9	0.26	0.91
HA3	GLY	104	HD1	HIS	126	3.09	18	0.58	1.03
H	SER	110	H	GLN	111	2.22	11	0.27	0.54
QE	PHE	112	QG2	ILE	156	7.34	8	0.18	0.56

HZ	PHE	112	H	PHE	113	6.11	18	0.64	0.96
QD	PHE	113	QD2	LEU	122	4.94	19	0.42	0.64
QE	PHE	113	QD1	LEU	122	3.42	11	0.21	0.52
H	THR	116	QG2	THR	119	4.08	17	0.34	0.85
QG2	THR	116	HG3	GLU	143	3.09	18	0.44	0.87
HA	THR	119	H	TRP	121	3.31	20	0.39	0.41
QG2	THR	119	H	LEU	122	4.98	10	0.26	0.62
HB3	LEU	122	QD2	LEU	122	2.75	20	0.52	0.61
HA	ASP	123	H	GLY	124	2.17	6	0.42	1.36
H	LYS	125	HB3	LYS	125	3.11	7	0.14	0.53
QG1	VAL	128	H	PHE	129	2.77	15	0.27	0.51
H	PHE	129	QD	PHE	129	2.76	10	0.22	0.63
QG2	VAL	132	H	ASN	137	6.35	15	0.26	0.51
HG2	LYS	133	QD2	LEU	164	5.74	20	0.79	1.05
HB3	GLU	134	QD1	LEU	164	6.42	12	0.21	0.68
HA	MET	136	HG2	MET	136	2.72	14	0.3	0.55
QB	ASN	137	H	MET	142	4.43	20	1.97	2.05
H	ALA	141	QD	PHE	145	5.74	20	0.54	0.76
HA	MET	142	HG2	MET	142	2.83	20	0.68	0.81
HA	MET	142	HB3	ARG	144	4.66	15	0.24	0.41
HB3	MET	142	QE	MET	142	2.79	16	0.3	0.53
HA	GLU	143	HG3	GLU	143	4.04	20	0.5	0.65
HA2	GLY	150	H	LYS	151	3.22	6	0.18	0.94
H	LYS	155	HG3	LYS	155	3.67	20	0.46	0.6
H	ILE	156	HG12	ILE	156	3.8	20	0.37	0.5
QG2	ILE	158	HB3	CYS	161	2.81	12	0.32	0.84
HB2	CYS	161	H	GLY	162	2.77	15	0.28	0.49
H	GLY	162	HG	LEU	164	6.64	19	0.62	0.97
H	GLY	162	QD2	LEU	164	2.74	20	1.46	1.69
H	GLN	163	HB3	GLN	163	3.44	10	0.34	0.71
HG3	GLN	163	HE21	GLN	163	3.47	9	0.21	0.57
HA	GLU	165	HB3	GLU	165	3.7	20	0.92	1.08

a The number of conformers in which the distance restraint was violated

b The average distance restraint violation

c The maximal distance restraint violation

Table S4. Often occuring eNOE-derived distance restraint violations of the two-states ensemble

Atom	Res	Res no	Atom	Res	Res no	Distance (Å)	no	Avg(Å)	Max (Å)
H	VAL	20	QG2	VAL	20	2.62	20	0.43	0.56
H	ILE	57	H	MET	61	2.82	20	0.57	0.73
H	LYS	82	QG	LYS	82	3.1	20	0.4	0.54
H	PHE	88	HA	ILE	89	3.28	20	0.62	0.73
H	SER	99	H	PHE	112	4.42	20	0.61	0.81
HA	ASP	123	H	GLY	124	1.93	20	0.51	0.99
QB	ASN	137	H	MET	142	3.95	20	0.97	1.29
H	ALA	141	QD	PHE	145	5.11	20	0.55	0.73
H	GLY	162	QD2	LEU	164	2.44	20	0.71	0.98

References

[1] Cavanagh J, Fairbrother WJ, Palmer AG III. , Skelton NJ, Rance M, Protein NMR Spectroscopy Principles and Practice, Academic Press, 2006.

[2] Permi P, Rosevear PR, Annila A, A set of HNCO-based experiments for measurement of residual dipolar couplings in 15N, 13C, (2H)-labeled proteins, J Biomol NMR 2000, 17, 43-54.

[3] Kuboniwa H, Grzesiek S, Delaglio F, Bax A, A set of HNCO-based experiments for measurement of residual dipolar couplings in 15N, 13C, (2H)-labeled proteins, J Biomol NMR 1994, 4, 871-878.

[4] Bax A, Dazzling world of spins, Chem Biol 1996, 3, 93-94.

[5] Orts J, Vögeli B, Riek R, Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs, J Chem Theory Comput 2012, 8, 3483-3492.

[6] Ottiger M, Zerbe O, Güntert P, Wüthrich K, The NMR solution conformation of unligated human cyclophilin A, J Mol Biol 1997, 272, 64-81.

[7] a) Vögeli B, Güntert P, Riek R, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol Phys 2013, 111, 437-454;
b) Chi CN, Strotz D, Riek R, Vögeli B, Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals, J Biomol NMR 2015, 62, 63-69.

[8] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, J Biomol NMR 1995, 6, 277-293.

[9] Vranken WF, Boucher W, Stevens TJ, R. Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 2005, 59, 687-696.

[10] Chi CN, Engström A, Gianni S, Larsson M, Jemth P, Two conserved residues govern the salt and pH dependencies of the binding reaction of a PDZ domain, J Biol Chem 2006, 281, 36811-36818.

[11] Vögeli B, Ying J, Grishaev A, Bax A, Limits on variations in protein backbone dynamics from precise measurements of scalar couplings, J Am Chem Soc 2007, 129, 9377-9385.

[12] a) Güntert P, Dötsch V, Wider G, Wüthrich K, Processing of multi-dimensional NMR data with the new software PROSA, J Biomol NMR 1992, 2, 619-629;
b) Vögeli B, Kazemi S, Güntert P, Riek R, Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs, Nat Struct Mol Biol 2012, 19, 1053-1057.

[13] Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D, Intrinsic dynamics of an enzyme underlies catalysis, Nature 2005, 438, 117-121.

[14] Strotz D, Orts J, Chi C, Riek R, Vögeli B, The eNORA2 exact NOE analysis program, to be submitted.

[15] Koradi R, Billeter M, Wüthrich K, MOLMOL: a program for display and analysis of macromolecular structures, J Mol Graphics 1996, 14, 29–32.

Publications

[7] Strotz D, Orts J, Chi CN, Riek R, Vögeli B, The eNORA2 exact NOE analysis program (to be submitted).

[6] Olsson S, Strotz D, Vögeli B, Riek R, Cavalli A, The dynamic basis for signal propagation in human Pin1-WW, Structure 2016, pii: S0969-2126(16)30159-9.

[5] Strotz D, Orts J, Minges M, Vögeli B, The experimental accuracy of the uni-directional exact NOE, J Magn Reson 2015, 259, 32-46.

[4] Chi CN, Vögeli B, Bibow S, Strotz D, Orts J, Güntert P, Riek R, A Structural Ensemble for the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution, Angew Chem Int Ed Engl 2015, 54, 11657-61.

[3] Chi CN, Strotz D, Riek R, Vögeli B, Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals, J Biomol NMR 2015, 62, 63-9.

[2] Vögeli B, Orts J, Strotz D, Chi C, Minges M, Wälti MA, Güntert P, Riek R, Towards a true protein movie: a perspective on the potential impact of the ensemble-based structure determination using exact NOEs, J Magn Reson 2014, 241, 53-9.

[1] Vögeli B, Orts J, Strotz D, Güntert P, Riek R, Discrete three-dimensional representation of macromolecular motion from eNOE-based ensemble calculation, Chimia 2012, 66, 787-90.