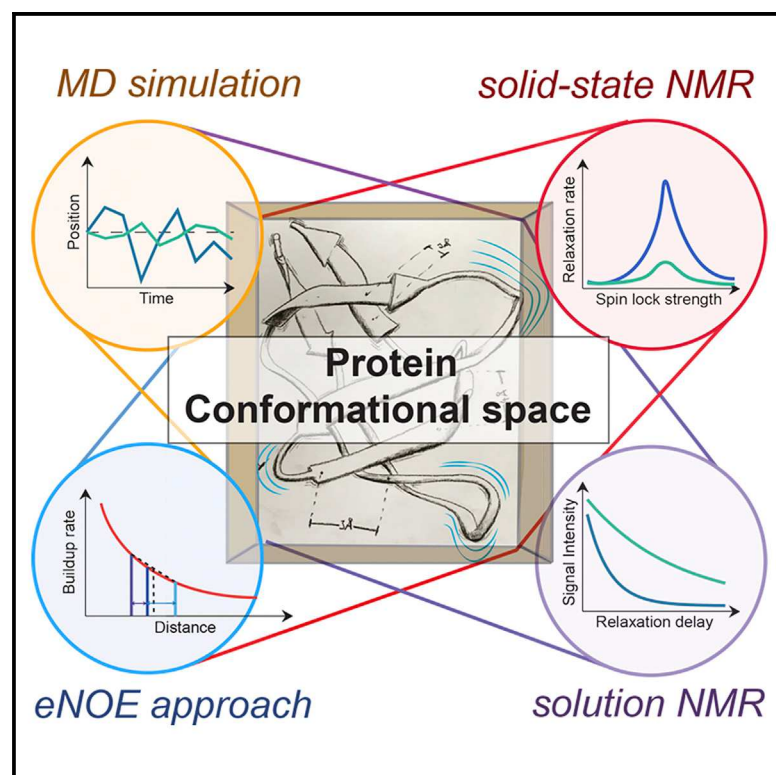


Protein Motional Details Revealed by Complementary Structural Biology Techniques

Graphical Abstract



Authors

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

Correspondence

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

In Brief

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

Highlights

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

Article

Protein Motional Details Revealed by Complementary Structural Biology Techniques

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

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

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

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

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

⁵Lead Contact

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

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

SUMMARY

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

INTRODUCTION

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

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

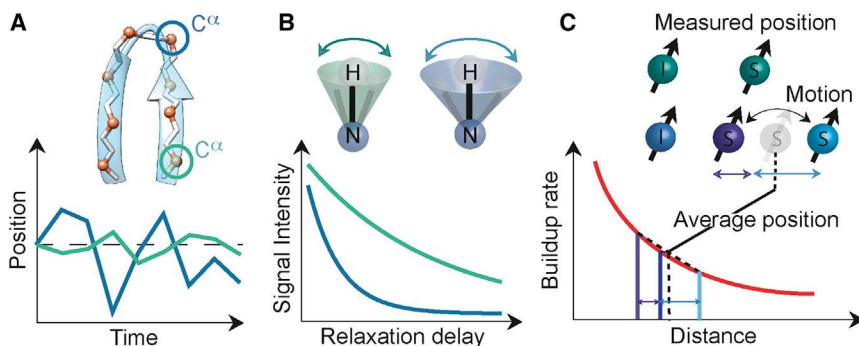


Figure 1. Complementary Methods Assessing Dynamics

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

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

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

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

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

A different recent concept assessing conformational space is based on the accurate measurement of (conflicting) proton-proton internuclear distances (exact NOEs or eNOEs) (Vögeli, 2014; Vögeli et al., 2009, 2014) and relies mainly on the exponential nature of the relationship between magnetization transfer efficiency and intermolecular distance. This over-proportional relation leads to the shortening of distance restraints, and thus inconsistencies with other restraints, in the presence of conformational motions (Figure 1C). The difference between the internuclear distances of the single average structure and the precise restraints can now be exploited to yield information about the conformational space of the protein (Vögeli et al., 2016). Therefore, a num-

ber of static structures (structural states) are calculated simultaneously using the eNOE restraints averaged over these states with an r^{-6} weighting (according to the Solomon equation [Solomon, 1955]). For instance, if a non-averaged restraint would lead to an underestimated distance as depicted in green in Figure 1C, the exponentially weighted restraints resolving the conflict would lead to the two different structural states shown in purple and blue. Importantly, even though “directional dynamics” information is obtained, the structural “states” determined by eNOE-based multi-state structure calculation are governed by a reduction of distance restraint violation via averaging. Therefore, the states are not whole-protein conformers representing a local energy minimum, but local landmarks of proton residences describing the direction and amplitude of motion at given sites. Also, the timescale of the dynamics is elusive from the kind of data acquired.

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

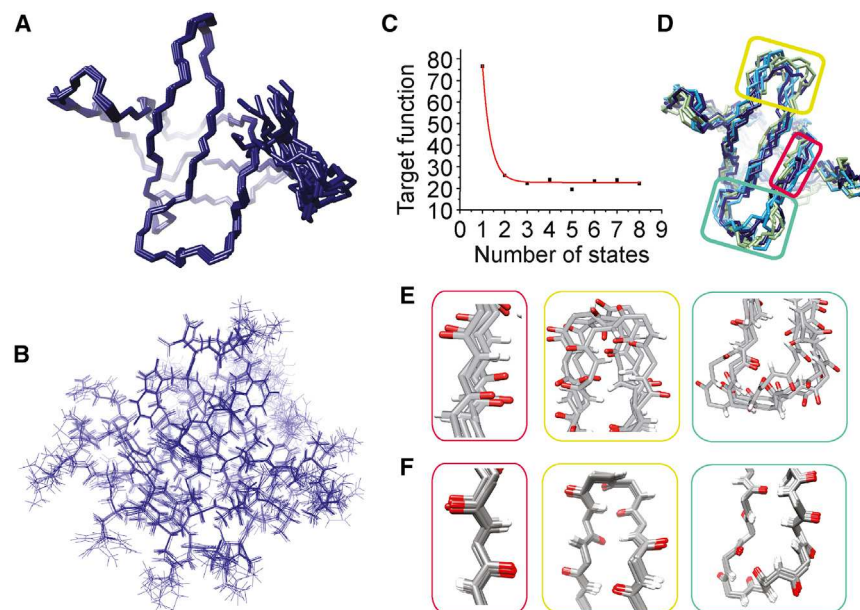


Figure 2. Single-State and Multi-State Structure Calculation of the SH3 Domain Determined via Solution NMR eNOEs

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

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

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

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

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

RESULTS

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

Single-State Protein Structure Calculation from Solution NMR

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

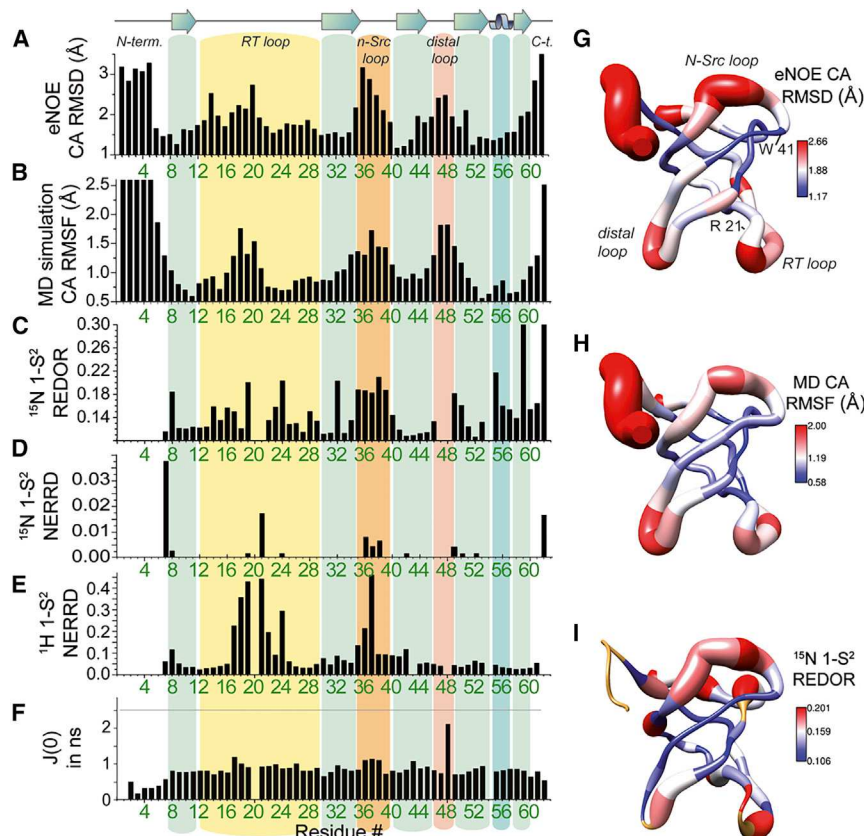


Figure 3. Complementary and Cross-Validating Assessment of Backbone Dynamics from MD Simulations, eNOE-Based Multi-State Structure Determination, Model-Free Order Parameters from Fast Magic Angle Spinning Solid-State NMR, and Solution NMR Relaxation

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

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

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

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

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

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

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

Detailed Mechanistic Views from MD Simulations

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

To break down the complex picture of the full atomistic dynamics, we computed the RMSD of every C α atom from its average position over time (i.e., the RMSF). Site-specific C α RMSF from the MD simulation with the ff15ipq force field at pH 3.5 are shown in Figure 3B. The results obtained with the alternative force fields and at the two different pH conditions simulated (pH 3.5 and 7) are all in close agreement with each other, see Figure S1. In line with previous NMR studies (Chevelkov et al., 2010; Rovó and Linser, 2018), the tips of the RT (17–22), n-Src (36–40),

and distal loop (47–48) show higher flexibility of the backbone, while β sheet regions forming the β barrel core are more rigid. The most flexible residues, in terms of C α atoms in the backbone, are Lys18, Thr37, Asn47, and Asp48, excluding the N and C termini.

Timescale-Specific Information from NMR Relaxation

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

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

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

Site-specific REDOR dephasing curves for an amide proton-back-exchanged ^2H , ^{13}C , ^{15}N sample of micro-crystalline SH3 are shown in Figure 3C. The very flexible residues, e.g., the N terminus, escape the solid-state NMR assessment based on cross-polarization magnetization transfer. The qualitative trends of the dipolar order parameters are consistent with the MD simulation. Exceptions are the residues Leu8, Thr32, Ala55, and Lys59, where a relatively high motional amplitude is measured, while those residues seem comparably rigid in the MD simulation, possibly due to the limited simulation timescale. Figures 3D

and 3E show results from ^{15}N and ^1H NERRD measurements, respectively (see details in the STAR Methods). In both NERRD measurements the data match the MD data; however, in contrast to faster timescale motion, microsecond timescale motion is present only at residue Arg21, the specificity pocket site structurally adjacent to Arg21 that is affected by RT loop motion, and the N and C termini interacting with the latter via crystal-crystal contacts. Whereas strong ^{15}N dispersion is largely limited to these sites, proton NERRD as a longer-range reporter of microsecond timescale motion shows strong dispersion also for residues more distant from Arg21. On one hand, this reflects the stronger (longer-range) changes in the proton dipolar coupling network sensed by ^1H NERRD compared with amide bond fluctuations, the measure of ^{15}N NERRD, in highly mobile regions. On the other hand, as assessed in detail previously (Rovó et al., 2019), the RT loop is dominated by slow μs motion, the distal loop moves on a faster timescale, and the n-Src-loop covers motions on intermediate timescales. Residues 47 and 48 in the distal loop (compare the solution relaxation data) are too mobile to be detected in cross-polarization-based experiments, and residues 20 and 54 are prolines. The faster fluctuations of the distal loop clearly exceed the timescale window assessed in both NERRD experiments.

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

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

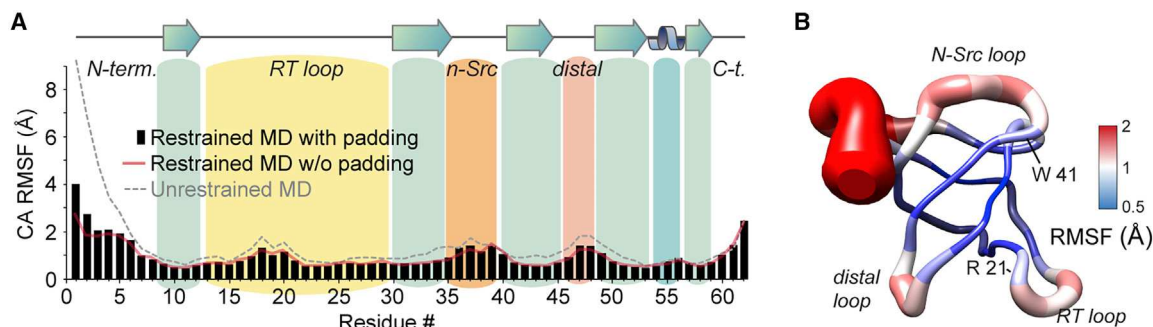


Figure 4. Comparison of C α RMSF in the SH3 Domain in Restrained Ensemble MD Simulations with Unbiased MD Simulations

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

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

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

RMSDs read out from the very well-defined single-state eNOE structure determination (Figures 2A and 2B) do not show the systematic increase in loop regions (Figures S2A and S2L), ruling out a data insufficiency for the described trends. However, it is important to point out that the substructures of the three-state calculation are sorted into the three-state bundles, shown in Figures 2D and 2E, by spatial similarity only. Likewise, angular information is not included the eNOE data, and TALOS dihedral angle restraints are purposely switched off in the final calculation step to reflect assessment of spatial plasticity in a way as unbiased as possible.

Ensemble MD Simulations Based on eNOE Restraints

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

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

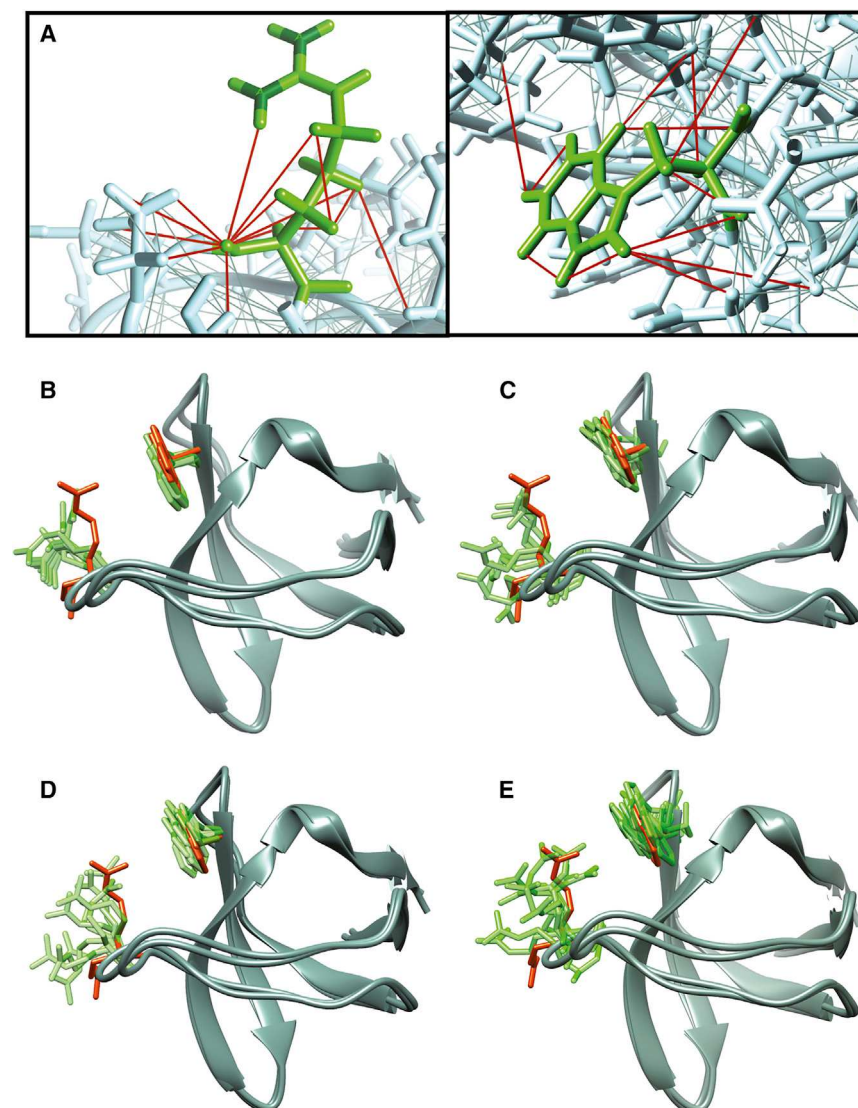


Figure 5. Experimental and *In Silico* Characterization of Side-Chain Conformational Spaces Elucidated by Ensemble-Based eNOE Assessment and MD Simulations

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

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

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

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

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

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

The Conformational Space of Side-Chain Motion

Besides information about mechanistic details of motion for the backbone, MD simulations as well as ^{13}C - and ^{15}N -edited eNOEs provide information about the conformational space sampled by side chains. This is particularly important since the side chains are responsible for biological functionality of various

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

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

DISCUSSION

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

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

tively), (3) backbone motion only versus all-atom assessment (relaxation data and multi-state structure/MD, respectively), as well as (4) the difference between simulated detail and experimental verification.

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

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

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

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

STAR★METHODS

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

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

SUPPLEMENTAL INFORMATION

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

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 5, 2019

Revised: May 5, 2020

Accepted: June 3, 2020

Published: June 23, 2020

REFERENCES

- Abraham, A. (1961). *The Principles of Nuclear Magnetism* (Oxford University Press).
- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25.
- Asami, S., Porter, J.R., Lange, O.F., and Reif, B. (2015). Access to Cα backbone dynamics of biological solids by ¹³C T1 relaxation and molecular dynamics simulation. *J. Am. Chem. Soc.* 137, 1094–1100.
- Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide protein data bank. *Nat. Struct. Mol. Biol.* 10, 980.
- Bishop, J.M. (1985). Viral oncogenes. *Cell* 42, 23–38.
- Blanco, F.J., Ortiz, Á.R., and Serrano, L. (1997). 1H and 15N NMR assignment and solution structure of the SH3 domain of spectrin: comparison of unrefined and refined structure sets with the crystal structure. *J. Biomol. NMR* 9, 347–357.
- Bonomi, M., Heller, G.T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* 42, 106–116.
- Brunger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R., and Nilges, M. (1993). Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* 261, 328.
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101.
- Camilloni, C., Cavalli, A., and Vendruscolo, M. (2013). Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins. *J. Phys. Chem. B* 117, 1838–1843.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.
- Chevelkov, V., Faelber, K., Schrey, A., Rehbein, K., Diehl, A., and Reif, B. (2007). Differential line broadening in MAS solid-state NMR due to dynamic interference. *J. Am. Chem. Soc.* 129, 10195–10200.
- Chevelkov, V., Fink, U., and Reif, B. (2009). Accurate determination of order parameters from 1H,15N dipolar couplings in MAS solid-state NMR experiments. *J. Am. Chem. Soc.* 131, 14018–14022.
- Chevelkov, V., Xue, Y., Linser, R., Skrynnikov, N., and Reif, B. (2010). Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. *J. Am. Chem. Soc.* 132, 5015–5017.
- Chi, C.N., Strotz, D., Riek, R., and Vögeli, B. (2015). Extending the eNOE data set of large proteins by evaluation of NOEs with unresolved diagonals. *J. Biomol. NMR* 62, 63–69.
- Clore, G.M., and Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* 109, 4108–4139.
- Clore, G.M., and Schwieters, C.D. (2004). How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J. Am. Chem. Soc.* 126, 2923–2938.

- Debiec, K.T., Cerutti, D.S., Baker, L.R., Gronenborn, A.M., Case, D.A., and Chong, L.T. (2016). Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. *J. Chem. Theory Comput.* **12**, 3926–3947.
- Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H., and Shaw, D.E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41**, 429–452.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8592.
- Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A., and Kay, L.E. (1995). Spectral density function mapping using ^{15}N relaxation data exclusively. *J. Biomol. NMR* **6**, 153–162.
- Fenwick, R.B., Schwieters, C.D., and Vögeli, B. (2016). Direct investigation of slow correlated dynamics in proteins via dipolar interactions. *J. Am. Chem. Soc.* **138**, 8412–8421.
- Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329.
- Grohe, K., Nimerovsky, E., Singh, H., Vasa, S.K., Söldner, B., Vögeli, B.H., Rienstra, C.M., and Linser, R. (2019). Exact distance measurements for structure and dynamics in solid proteins by fast magic angle spinning NMR. *Chem. Commun.* **55**, 7899–7902.
- Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in $^{13}\text{C}/^{15}\text{N}$ -enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magn. Reson. Ser. B* **101**, 114–119.
- Gullion, T., and Schaefer, J. (1989). Rotational-echo double-resonance NMR. *J. Magn. Reson.* **81**, 196–200.
- Güntert, P., and Buchner, L. (2015). Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62**, 453–471.
- Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
- Hoffmann, F., Mulder, F.A.A., and Schäfer, L.V. (2018a). Accurate methyl group dynamics in protein simulations with AMBER force fields. *J. Phys. Chem. B* **122**, 5038–5048.
- Hoffmann, F., Xue, M., Schäfer, L.V., and Mulder, F.A.A. (2018b). Narrowing the gap between experimental and computational determination of methyl group dynamics in proteins. *Phys. Chem. Chem. Phys.* **20**, 24577–24590.
- Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M., and Kay, L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590.
- Lange, O.F., Lakomek, N.-A., Fares, C., Schröder, G.F., Walter, K.F.A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
- Lindorff-Larsen, K., Best, R.B., DePristo, M.A., Dobson, C.M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–132.
- Linser, R., Chevelkov, V., Diehl, A., and Reif, B. (2007). Sensitivity enhancement using paramagnetic relaxation in MAS solid-state NMR of perdeuterated proteins. *J. Magn. Reson.* **189**, 209–216.
- Linser, R., Fink, U., and Reif, B. (2010). Assignment of dynamic regions in biological solids enabled by spin-state selective NMR experiments. *J. Am. Chem. Soc.* **132**, 8891–8893.
- Lipari, G., and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559.
- Logan, T.M., Olejniczak, E.T., Xu, R.X., and Fesik, S.W. (1992). Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett.* **314**, 413–418.
- Massenet, C., Chenavas, S., Cohen-Addad, C., Dagher, M.-C., Brandolin, G., Pebay-Peyroula, E., and Fieschi, F. (2005). Effects of p47phox C terminus phosphorylations on binding interactions with p40phox and p67phox: structural and functional comparison of p40phox and p67phox SH3 domains. *J. Biol. Chem.* **280**, 13752–13761.
- Palmer, A.G., and Massi, F. (2006). Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106**, 1700–1719.
- Papaleo, E., Camilloni, C., Teilmann, K., Vendruscolo, M., and Lindorff-Larsen, K. (2018). Molecular dynamics ensemble refinement of the heterogeneous native state of NCBD using chemical shifts and NOEs. *PeerJ* **6**, e5125.
- Peng, J.W., and Wagner, G. (1992). Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *J. Magn. Reson.* **98**, 308–332.
- Pratihari, S., Sabo, T.M., Ban, D., Fenwick, R.B., Becker, S., Salvatella, X., Griesinger, C., and Lee, D. (2016). Kinetics of the antibody recognition site in the third IgG-binding domain of protein G. *Angew. Chem. Int. Ed.* **55**, 9567–9570.
- Ravera, E., Sgheri, L., Parigi, G., and Luchinat, C. (2016). A critical assessment of methods to recover information from averaged data. *Phys. Chem. Chem. Phys.* **18**, 5686–5701.
- Reichel, K., Fiset, O., Braun, T., Lange, O.F., Hummer, G., and Schäfer, L.V. (2017). Systematic evaluation of CS-Rosetta for membrane protein structure prediction with sparse NOE restraints. *Proteins* **85**, 812–826.
- Richter, B., Gsponer, J., Várnai, P., Salvatella, X., and Vendruscolo, M. (2007). The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J. Biomol. NMR* **37**, 117–135.
- Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. *Science* **309**, 303–306.
- Robustelli, P., Kohlhoff, K., Cavalli, A., and Vendruscolo, M. (2010). Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* **18**, 923–933.
- Robustelli, P., Piana, S., and Shaw, D.E. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U S A* **115**, E4758.
- Rovó, P., and Linser, R. (2017). Microsecond timescale proton rotating-frame relaxation under magic angle spinning. *J. Phys. Chem. B* **121**, 6117–6130.
- Rovó, P., and Linser, R. (2018). Microsecond timescale protein dynamics: a combined solid-state NMR approach. *ChemPhysChem* **19**, 34–39.
- Rovó, P., Smith, C.A., Gauto, D., de Groot, B.L., Schanda, P., and Linser, R. (2019). Mechanistic insights into microsecond time-scale motion of solid proteins using complementary ^{15}N and ^1H relaxation dispersion techniques. *J. Am. Chem. Soc.* **141**, 858–869.
- Russo, L., Maestre-Martinez, M., Wolff, S., Becker, S., and Griesinger, C. (2013). Interdomain dynamics explored by paramagnetic NMR. *J. Am. Chem. Soc.* **135**, 17111–17120.
- Saksela, K., and Permi, P. (2012). SH3 domain ligand binding: what's the consensus and where's the specificity? *FEBS Lett.* **586**, 2609–2614.
- Schanda, P., and Ernst, M. (2016a). Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: principles and applications to biomolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 1–46.
- Schanda, P., and Ernst, M. (2016b). Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: principles and applications to biomolecules. *Prog. Nucl. Mag. Res. Sp* **96**, 1–46.
- Schanda, P., Meier, B.H., and Ernst, M. (2010). Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 15957–15967.
- Schrödinger, L. (2015). The PyMOL Molecular Graphics System, Version-1.8.

- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223.
- Solomon, I. (1955). Relaxation processes in a system of two spins. *Phys. Rev.* **99**, 559–565.
- Strotz, D., Orts, J., Chi, C.N., Riek, R., and Vögeli, B. (2017). eNORA2 exact NOE analysis program. *J. Chem. Theory Comput.* **13**, 4336–4346.
- Takemura, K., and Kitao, A. (2012). Water model tuning for improved reproduction of rotational diffusion and NMR spectral density. *J. Phys. Chem. B* **116**, 6279–6287.
- Teyra, J., Huang, H., Jain, S., Guan, X., Dong, A., Liu, Y., Tempel, W., Min, J., Tong, Y., Kim, P.M., et al. (2017). Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities. *Structure* **25**, 1598–1610.e3.
- Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., et al. (2008). BioMagResBank. *Nucleic Acids Res.* **36**, D402–D408.
- Vallat, B., Webb, B., Westbrook, J.D., Sali, A., and Berman, H.M. (2018). Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure* **26**, 894–904.e2.
- van Rossum, B.-J., Castellani, F., Rehbein, K., Pauli, J., and Oschkinat, H. (2001). Assignment of the nonexchanging protons of the a-Spectrin SH3 domain by two- and three-dimensional ^1H - ^{13}C solid-state magic-angle spinning NMR and comparison of solution and solid-state proton chemical shifts. *ChemBioChem* **2**, 906–914.
- Vitalini, F., Mey, A.S.J.S., Noé, F., and Keller, B.G. (2015). Dynamic properties of force fields. *J. Chem. Phys.* **142**, 084101.
- Vögeli, B. (2014). The nuclear Overhauser effect from a quantitative perspective. *Prog. Nucl. Magn. Reson. Spectrosc.* **78**, 1–46.
- Vögeli, B., Güntert, P., and Riek, R. (2013). Multiple-state ensemble structure determination from eNOE spectroscopy. *Mol. Phys.* **111**, 437–454.
- Vögeli, B., Kazemi, S., Güntert, P., and Riek, R. (2012). Spatial elucidation of motion in proteins by ensemble-based structure calculation using exact NOEs. *Nat. Struct. Mol. Biol.* **19**, 1053.
- Vögeli, B., Olsson, S., Güntert, P., and Riek, R. (2016). The exact NOE as an alternative in ensemble structure determination. *Biophys. J.* **110**, 113–126.
- Vögeli, B., Olsson, S., Riek, R., and Güntert, P. (2015a). Compiled data set of exact NOE distance limits, residual dipolar couplings and scalar couplings for the protein GB3. *Data Brief* **5**, 99–106.
- Vögeli, B., Olsson, S., Riek, R., and Güntert, P. (2015b). Complementarity and congruence between exact NOEs and traditional NMR probes for spatial decoding of protein dynamics. *J. Struct. Biol.* **191**, 306–317.
- Vögeli, B., Orts, J., Strotz, D., Chi, C., Minges, M., Wälti, M.A., Güntert, P., and Riek, R. (2014). Towards a true protein movie: a perspective on the potential impact of the ensemble-based structure determination using exact NOEs. *J. Magn. Reson.* **241**, 53–59.
- Vögeli, B., Segawa, T.F., Leitz, D., Sobol, A., Choutko, A., Trzesniak, D., van Gunsteren, W., and Riek, R. (2009). Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. *J. Am. Chem. Soc.* **131**, 17215–17225.
- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696.
- Vuister, G.W., and Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in ^{15}N -enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.
- Wang, L.-P., McKiernan, K.A., Gomes, J., Beauchamp, K.A., Head-Gordon, T., Rice, J.E., Swope, W.C., Martínez, T.J., and Pande, V.S. (2017). Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. *J. Phys. Chem. B* **121**, 4023–4039.
- Zhou, D.H., and Rienstra, C.M. (2008). High-performance solvent suppression for proton-detected solid-state NMR. *J. Magn. Reson.* **192**, 167–172.

STAR★METHODS

KEY RESOURCES TABLE

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

RESOURCE AVAILABILITY

Lead Contact

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

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

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

METHOD DETAILS

Protein Expression and NMR-Spectroscopy

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

cDNA of the protein was transformed into competent *E. coli* BL21 (DE3) cells and grown in M9 medium containing $^{13}\text{C}_6$ -glucose (2 g/l) and ^{15}N NH_4Cl (1 g/l) at a temperature of 37 °C to an OD of 0.6, when expression was induced using 1 mM isopropyl β -D-1-thiogalactopyranoside. Cells were harvested after over-night expression at 22 °C and lysed using an Emulsiflex homogenizer. After purification via anion exchange on a Q-SEPH FF column, the protein-containing fractions were pH-adjusted to 3.5, concentrated, and subjected to a Superdex 75 size exclusion column in a citric acid buffer of pH 3.5. After concentration to around 500 μM , the protein was filled into a 5 mm Shigemitsu tube and used for NMR studies. NMR spectra were recorded at a temperature of around 25 °C on a Bruker 800 MHz Avance III spectrometer equipped with a triple-channel cryo-probe or a 1.3 mm MAS probe, respectively. For solution assignment of sidechain chemical shifts, TOCSY-based C(CC)CONH (Grzesiek et al., 1993) and H(CC)CONH (Logan et al., 1992) experiments were recorded on a ^{15}N , ^{13}C -labeled sample at pH 3.5, connecting sidechain and amide chemical shifts. In case of the H(CC)CONH experiment, MLEV17-mixing was employed, with a mixing time of 80 ms. In case of the C(CC)CONH experiment, DIPSI2 mixing was used, with a mixing time of 12 ms (assignments are shown in Table S1 and also deposited in the BMRB under accession code 34420). In total, 620 chemical shifts could be assigned of which 370 are sidechain shifts. A series of ^{15}N - and ^{13}C -resolved 3D [^1H - ^1H]-NOESY-HSQC experiments were recorded using 20, 30, 40, 50, 60, and 70 ms and 20, 40, 50, 60, 70, and 80 ms of NOESY mixing time, respectively. A formula for the maximal mixing time optimal for eNOE is available on the homepage of the Vögeli group. NOE peak assignment was performed using the spectrum with longest mixing time via the semi-automated NOESY assignment function of CCPNmr (Vranken et al., 2005). In total, 689 distance restraints were obtained, from which 530 are eNOEs and 159 generic eNOEs (generic eNOEs lack a resolved diagonal signal, and a generic diagonal decay (average of the resolved diagonal signals) was employed) (Chi et al., 2015). In addition, 119 TALOS ϕ and ψ angles as well as 44 3J couplings were obtained. 3J -coupling constants between H^{N} and H^{α} were determined via a quantitative J -coupling experiment (HNHA) (Vuister and Bax, 1993) with the coupling constants read out from the ratio of cross- and diagonal-peak intensities. Solution NMR T_1 and T_2 times (in the presence of CPMG pulse trains) were determined using relaxation delays of 10, 50, 100, 200 (2x), 300, 400, 500, 600 (2x), 800, 1020 ms and 10, 30, 50 (2x), 70, 90, 110 (2x), 130, 150, 170, 190 ms, respectively. hetNOE spectra were recorded with 8 s of ^{15}N d1 and no or 3 s of saturation alternatingly. ^1H - ^{15}N REDOR data were recorded on a perdeuterated and 100% back-exchanged sample of SH3 of pH 7.0 at 40 kHz MAS, 800 MHz proton Larmor frequency, and roughly 30 °C in a 1.3 mm rotor via recoupling-edited MISSISSIPPI correlations (Zhou and Rienstra, 2008), using rotor-synchronized proton 180 ° pulses of 71 kHz B_1 for up to 850 μs recoupling duration, incremented from 6 to 34 rotor periods in steps of 2, and fitted individually assuming a fixed H-N distance. ^{15}N and ^1H NERRD data were recorded at a temperature of approximately 25 °C. Full experimental details of $^1\text{H}/^{15}\text{N}$ NERRD are described in Rovó et al. (Rovó et al., 2019). All spectra were processed using the program Topspin 3.6 and analyzed using CCPNmr (Vranken et al., 2005). The eNOE analysis was performed using eNORA2 (Strotz et al., 2017).

Structure Calculation

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

In brief, averaged distances D are defined as:

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

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

where k runs over all members of the ensemble.

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

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

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

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

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

Iteration #: 1 2 3 4)

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

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

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

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

MD Simulations

All MD simulations were carried out with Gromacs version 2019.2 (Abraham et al., 2015). The 1.85 Å resolution X-ray crystal structure (PDB 2NUZ) was used as a starting structure of the simulations, after adding the six missing N-terminal residues (Met1-Asp2-Glu3-Thr4-Gly5-Lys6) and the C-terminal residue Asp62, which were not resolved in the X-ray structure; the coordinates of these residues were taken from the solution structure (PDB 6SCW). After adding hydrogens to the protein and crystal water heavy atoms, the system was solvated with ca. 4500 water molecules in a periodic dodecahedron simulation box. The overall charge of the box was neutralized by Cl⁻ ions. The following three Amber protein force fields were used: ff15ipq (Debiec et al., 2016), ff99SB-disp (Robustelli et al., 2018), and FB15 (Wang et al., 2017). In the ff15ipq simulations, the SPC/E_b water model was used (Takemura and Kitao, 2012), whereas the TIP4PD-derived ff99SB-disp 4-site water model (Robustelli et al., 2018) was used in the ff99SB-disp simulations. In the FB15 simulations, the TIP3P-FB water model (Wang et al., 2017) was used. Lennard-Jones 6,12 interactions were smoothly shifted to zero at a cut-off distance of 1.0 nm; this distance was also used for switching between short-range and long-range Coulomb interactions, which were treated with the particle mesh Ewald method (Essmann et al., 1995) with a 0.12 nm grid spacing. After steepest-descent energy minimization, the system was equilibrated for 100 ps in the NpT ensemble with harmonic position restraints on all protein heavy atoms, with force constants of 1000 kJ/mol/nm². Temperature and pressure were kept constant at 300 K and 1 bar, respectively, using the thermostat of Bussi and coworkers (Bussi et al., 2007) and the Berendsen barostat, respectively. The use of the SETTLE and LINCS constraint algorithms to constrain all internal degrees of freedom of the water molecules and all protein bonds, respectively, allowed to integrate the equations of motion with 2 fs time steps. Finally, the position restraints were switched off, and four individual 500 ns MD simulations were initiated from different random seeds to generate the initial atomic velocity distribution at 300 K.

For the restrained ensemble MD simulations, back-calculation of the proton-proton distances included r^{-6} averaging over the ensemble, and restraining was only applied when the calculated ensemble averages deviated from the experimentally derived ensemble averages. Thus, instead of modeling the eNOE distances as static, distance restraints were included in the force field such that the distance restraint needed only to be satisfied as an ensemble average. An MD ensemble comprised of 20 replicas was run for 100 ns in the NpT ensemble at 300 K and 1 bar (with the same thermostat and barostat as described above for the unbiased MD simulations). The eNOE-derived distance bounds were converted into distance restraints and included in the Gromacs protein topology. Flat-bottom harmonic distance restraints with force constants of 1000 kJ/mol/nm² were used as biasing potentials. The onset of the harmonic energy penalty was shifted by +1 Å and -1 Å for the upper and lower distance bounds, respectively, relative to the eNOE-derived distances; this “padding” resolves some possible ambiguities with the precise distance values and was applied previously, e.g., to determine membrane protein structures from sparse NOE restraints (Reichel et al., 2017). We repeated the restrained ensemble MD simulations also without applying this 1 Å padding. PROCHECK (Laskowski et al., 1993) and MolProbit (Chen et al., 2010) evaluation results are shown in Table S1C.

QUANTIFICATION AND STATISTICAL ANALYSIS

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