Spatial Elucidation of Motion in Proteins by Ensemble-based Structure Calculation using exact NOEs

Beat Vögeli¹, Sina Kazemi², Peter Güntert²,³ and Roland Riek¹,*

¹ Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH-Hönggerberg, CH-8093 Zürich, Switzerland.

² Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, and Frankfurt Institute for Advanced Studies, J.W. Goethe-Universität, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany.

³ Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan.

*Corresponding author: Roland Riek, roland.riek@phys.chem.ethz.ch

Keywords: NMR, biological macromolecules, proteins, dynamics, correlated dynamics, NOE, nuclear Overhauser effect, structure calculation, structure ensemble
ABSTRACT

Proteins are inherently dynamic systems. Their motions cover both a large range in distance magnitude and time scale. Due to their omnipresence it is likely that dynamics plays an important role in the function of biomolecules. For a detailed understanding of a protein’s function, its 3D structure and the description of its dynamics are therefore required. Structure determination is well established and NMR relaxation phenomena provide insights into local molecular dynamics while only recently several attempts have been introduced for the detection of concerted motion. Here, we present an ensemble-based structure determination protocol using ensemble-averaged distance restraints obtained from exact NOE rates. Application to the model protein GB3 establishes an ensemble of structures that reveals amongst others correlated motion across the β-sheet, concerted conformational exchange between side chains and backbone, side chain rotamer states and absence of correlated motion between the β-sheet and the α-helix.
At physiologically relevant temperatures proteins inter-change between structural states covering both a large range in magnitude from $10^{-11}$ to $10^{-6}$ meters as well as spanning time scales from $10^{-12}$ to $10^{5}$ seconds and beyond. Due to this inherently dynamical nature of proteins it is likely that evolution took advantage thereof and thus dynamics may play important roles in the activity of proteins. Towards a detailed description and understanding of a protein’s function, its 3D atomic-resolution structure and an accurate description of its entire dynamics are therefore required. NMR relaxation phenomena provide a great deal of insight into local molecular dynamics. However, the dynamic picture is still largely incomplete because no adequate method is available to detect either through-space motion or concerted motion. Attempts towards a comprehensive description of motion are currently under development including residual dipolar couplings (RDC), relaxation dispersion NMR experiments, paramagnetic relaxation enhancement (PRE), cross-correlated relaxation (CCR), and conventional Nuclear Overhauser Enhancement (NOE) data in combination with molecular dynamics simulations or structure prediction software\textsuperscript{1-14}. While the description of the motion of a protein is still in its enfancy, the structure determination of biomolecules by either x-ray crystallography or NMR is well established. NMR structure determination of biomolecules is mainly based on a large collection of nuclear Overhauser enhancements (NOEs) (Fig. 1A)\textsuperscript{15}. For the structure calculation the obtained semi-quantitative NOEs are translated semi-quantitatively into upper limit distances, following the proportionality of the NOE cross-relaxation rate with the inverse 6\textsuperscript{th} power of the distance between two (isolated) interacting spins\textsuperscript{15}. Such an approach results usually in a well-defined 3D NMR structure as exemplified here for the 56-residues model protein GB3 (Fig. 1A). The input of 1956 NOE-based upper limit distance restraints combined with 147 dihedral angle restraints derived from scalar couplings and 54 dihedral angle restraints from $^{13}$C\textsuperscript{a} chemical shifts, and 90 RDCs yielded the 3D NMR structure of GB3 represented by a bundle of nine conformers reflecting the precision of the structure (Fig. 1A and a
backbone r.m.s.d. of 0.47 Å. This practice dates back to the 1980s when it proved difficult to determine NOE rates and convert them into exact distances.\textsuperscript{15} In the following, we present an ensemble-based structure determination protocol using ensemble-averaged distance restraints obtained from recently introduced exact NOE rates.\textsuperscript{16-20} Application of this approach to the model protein GB3 generates a structural ensemble which describes its conformationally occupied space reflecting both its 3D structure and its dynamics.

Results

We recently demonstrated the collection of exact $^1\text{H}^\text{N}$$^1\text{H}^\text{N}$ NOE rates (eNOE) by using state-of-the-art NMR spectrometers, optimized 3D-resolved [$^1\text{H},^1\text{H}$]-NOESY experiments, short NOESY mixing times, an optimized protocol for extracting NOE rates from a series of NOESY spectra, and spin diffusion correction.\textsuperscript{16,17} Here, we extended the eNOE to aliphatic protons establishing a data set of 844 eNOEs for the model protein GB3 (Supplementary Table S1, see also Vögeli et al., to be submitted). The eNOEs were converted into distances by the notion that the NOE is mostly sensitive to slow motion, which was shown to be valid for H-H spin pairs if the local H-X order parameters, which are a measure of fast motion, are larger than 0.5.\textsuperscript{20,21} This is most often the case in folded proteins. Even for lower order parameters down to 0.2 the fast motion alters the eNOE-derived distances by less than 10%.\textsuperscript{21} Overall, the translation from eNOEs to distances is very robust because the $1/r^6$ dependency between NOE and distance reduces the relative distance error six-fold compared to the relative inaccuracies of the NOE measurements.\textsuperscript{16,17,21} The eNOEs of GB3 yielded three classes of distance restraints (Table 1): (i) exact distances with an estimated accuracy of 5% obtained from the eNOEs determined for two symmetric pathways (i.e. from spin $I_1$ to $I_2$ and vice versa from $I_2$ to $I_1$), (ii) distances with an estimated accuracy of 15% derived from single eNOEs pathways, and (iii) distances
with an estimated accuracy of 20% derived from methyl-methyl eNOEs for which additional corrections had to be added to the distance restraint following established arguments\textsuperscript{15,22}.

The eNOE-derived distance restraints in combination with the small set of RDCs and angle restraints derived from scalar couplings and $^{13}$C\textsuperscript{\alpha} chemical shifts were used for a structure calculation of GB3 following standard protocols using the software package CYANA\textsuperscript{23,24}. The bundle of 9 conformers shown in Figure 1B represents the calculated structure. The input data results in an extremely tight, albeit not correct (see below) structure with a small backbone r.m.s.d. of 0.11 \AA and an all-heavy-atom r.m.s.d. of 0.60 \AA only. When compared to the traditional structure calculation using semi-quantitative NOEs, the high precision is particularly striking (compare Fig. 1A with 1B). Furthermore, the eNOE-based single-state NMR structure coincides closely with the RDC-optimized x-ray structure of GB3\textsuperscript{25,26} with an r.m.s.d. of 0.57 \AA for the backbone and 1.17 \AA for all heavy atoms (Supplementary Fig. S1). However, the large target function value of 27.5 \AA$^2$ (Supplementary Table S1, Fig. 2) resulting from many distance restraint violations indicates that the structure does not agree with the experimental data. The large number of violations of experimental restraints can be attributed to the motion-averaged nature of the measured NOE, while the structure calculation protocol is based on a single static structure, which obviously does not take into account any motion\textsuperscript{27}.

Following the ergodic hypothesis the influence of motion of a protein on the NOE probe can be described by an ensemble of structural states. To include the motional dependence of the eNOEs in the structure calculation, an ensemble-based protocol was established within the software package CYANA. The protocol requests that the experimental restraints are fulfilled by a set of structural states rather than by a single structure\textsuperscript{1,3}. 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. Following this protocol structures of GB3 represented by ensembles ranging from two to nine states have been obtained. Figure 2 illustrates the decrease of the target function with the increase of the number of states included in the structure calculation. Most prominent is its decrease from one state, which corresponds to the conventional structure calculation protocol (yet with eNOEs), to two and three states after which a plateau is reached. This observation indicates that in contrast to the single structure a three-state ensemble as represented in Figure 1C describes the experimental data well (backbone r.m.s.d. of 0.46 Å and an all-heavy-atom r.m.s.d. of 0.82 Å). Higher-state ensembles also fulfill the experimental data covering a very similar conformation space as the three-state ensemble (Supplementary Figs. S2 and S3).

In the following the three-state ensemble is used as a representative of the GB3 structure in solution because it is the smallest ensemble that fulfills the experimental data well. The input data represent a self-consistent set and the restraints are well satisfied in the three-state ensemble (Supplementary Table S2). To strengthen this finding all eNOE-derived distances were arbitrarily changed by up to 15% yielding considerable increases of the target functions (Fig. 2). This observation indicates that the eNOE-derived restraints have an accuracy much better than 15% as demonstrated earlier\textsuperscript{16,17}. In addition, it shows that the experimental data set is self-consistent and at least in part over-determined although the free parameter space was extensively enlarged with the introduction of the ensemble-based structure calculation. The need for multi-state ensembles is further supported by a cross-validation test which consists of the arbitrary deletion of 10% of all the eNOEs\textsuperscript{28}. The test shows that the violations of the non-included eNOE-derived distances summed over 10 structure calculations (overall deleting every eNOE exactly once) decrease with the size of the ensemble resulting in a drop of the target function by up to 40% when compared with the single-state structure (Fig. 2). These findings confirm the
prediction based on theoretical considerations that very tight distance restraints (less than 25% error) are required to calculate a multi-state structure. Finally, the close resemblance between local order parameters derived from the three-state ensemble and RDC-derived order parameters measured by Yao et al. indicates that the three-state ensemble represents well the structural space covered by internal motions of GB3 (Fig. 3 and Vögeli et al., the manuscript ready for submission to JMR).

Following the arguments above the three-state structural ensemble is a compact experiment-based representation of GB3 in solution. In contrast to the standard structure determination protocol it takes into account that the NOE is a time/ensemble-averaged parameter yielding an ensemble representation of the structure covering the conformational space of GB3. Because of the bundling restraints present in the calculation, the multi-state structure ensembles cover 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. Individual-state sub-bundle representations are obtained by grouping the most similar structures from each three-state conformer (Figs. 1 and 4). The sub-bundle for each structural state is thereby a measure for the precision of the individual structural states similar to the conventional bundle representation. Interestingly, the same bundle representation can be used for the entire β-sheet and some of the loops indicating that this entire structural segment undergoes conformational exchange between the three states in a concerted fashion. The central parallel β-sheet (strands β1 and β4) moves parallel to the entire β-sheet architecture (i.e. vertically to the polypeptide backbone), while the loops between β1 and β2, and β3 and β4 as well as accompanied segments within the β-strands counteract this motion in an anti-correlated manner (Fig. 4 and Supplementary Fig. S4). A principal component analysis (PCA) of the ensemble visualizes these findings further. In this analysis the major part of the spatial sampling is covered by the first
two PCA modes (Supplementary Fig. S5). This analysis describes concerted motion across the β-sheet indicated by the arrows in Figure 4 that show the directions and amplitudes of the principle motions. In contrast, the α-helix appears to be decoupled from the conformational exchange of the β-sheet as the bundle representation of the β-sheet does not enable to distinguish several structural states of the α-helix (Fig. 4). However, another set of conformers selected to describe structural states of the α-helix indicates that the backbone of the α-helix shows also distinct structural states that interconvert between each other, but the correlation appears to be weaker than for the β-sheet and is localized to the residues that face the hydrophobic core (Supplementary Fig. S4). The timescale of exchange between the conformational states both in the β-sheet as well as in the α-helix is most likely on the sub-ms time scale, because slower motion would result in line broadening or resonance doubling not observed in the spectra, and because the three-state ensemble is consistent with the RDC-derived order parameters\textsuperscript{26} sensitive to motion faster than ms (Fig. 3).

Discussion

The presented eNOE-based ensemble is in good agreement with previously obtained multiple-state ensembles\textsuperscript{2,5} calculated from relaxation order parameters, slow motion-sensitive residual dipolar couplings (RDCs) and (in one case\textsuperscript{6}) crystallographic B-factors although these ensembles appear to be optimally represented by 4 to 8 states rather than the three states derived from the eNOE analysis. All three ensemble representations show a similar amplitude of structural variations described by rms differences between the states within the ensemble of ca 0.5 Å. In all three ensembles crankshaft motions (anticorrelated φ and ψ\textsubscript{-1} angles) along the entire backbone and concerted structural variations in the β-sheet with the largest correlated structural variations in the loops connecting strands β1 and β2, and the α-
helix and β3 are observed (for a more detailed discussion we refer to Vögele et al, manuscript to be submitted to JMR). The observations of concerted motions are further supported by other previous studies on GB3 or the similarly folded ubiquitin using RDCs7,14 in combination with hydrogen-bond scalar couplings3 and cross-correlated relaxation (CCR) rates12. For example, local anti-correlated motion between the Φ and Ψ backbone torsion angles within part of the β-sheet had also been observed14. Our three-state ensemble also shows such anti-correlated properties for these backbone angles (Top panel in Fig. 3).

The observation of the distinct states of the backbone of GB3 can in part be extended to the side chains as visualized for example for residues 3, 12, and 43, for which the side chains show distinct structural states that correlate well with the distinct backbone states (Fig. 4). For most of the side chains in the hydrophobic core such structurally distinct states can be seen, while some of the surface-faced side chains show arbitrary conformations as exemplified by Lys 11 (Fig. 4) indicative of either free motion or lack of sufficient experimental data. Similar findings can be visualized by the χ1 angles that also describe the rotamer states of the side chains (Fig. 3). All but one of the side chains in the hydrophobic core have a single rotamer state that is in good agreement with the x-ray structure. Only residue 54 close to the C-terminus has two rotamer states for the χ1 angle. Interestingly, the rotamers of residue 54 correlate well with the distinct backbone structural states of the last β-strand indicating that the rotamers interchange in a concerted manner with the conformational exchange of the backbone. Outside the hydrophobic core many residues have multiple χ1 rotamer states. By correlating the χ1 rotamer states with the distinct structural states of the backbone, a coupling between them is observed within the secondary structural elements, while most of the side-chain rotamer states in the loops appear to be decoupled from their (local) backbone states.
In the following the side chain rotamers are discussed in more detail. For residues 15, 21, and 35, the same two rotamer states as in the crystal structure are observed. Actually, the NMR ensemble includes all the rotamer states observed in the crystal structures with the exception of residues 7 and 47, for which \( \chi^1 \) deviates by roughly 40º (Fig. 3). Also scalar and residual dipolar couplings as well as cross-correlated relaxation rates measured under liquid-state conditions are in very good agreement with the rotamer states of the structural ensemble including residues 7 and 47 (for a detailed discussion we refer to Vögeli et al, manuscript to be submitted to JMR). For example, all but one rotamer states obtained from scalar couplings are in line with the structural ensemble. The only inconsistent rotamer state is \( \chi^1 \) of residue 8 for which the ensemble has the same rotamer as the x-ray structure. Furthermore, the structural ensemble is in accord with measurements by Chou et al., who determined the rotamer populations of \( \chi^1 \) angles of Val, Ile, and Thr by RDCs, \( ^3J_{C'\gamma} \), and \( ^3J_{NC'\gamma} \) scalar couplings.\(^{30}\) For residues 3, 6, 7, 17, 18, 33, 39, 44 and 52 the single predicted rotamer states, for residues 21, 25, and 54 the two predicted rotamers and for residue 42 all three predicted rotamers are in line with the structural ensemble. Discrepancies are observed for residues 11, 49, 51, and 53. Overall, there is good agreement between the sampling in the ensemble and previously analyzed X-ray or NMR data. Some inconsistencies may be due to different sample conditions such as crystalline/liquid state or different buffer conditions.

In conclusion, by taking into account the motional dependence of the eNOE, an ensemble-based NMR structure determination protocol has been established that results in a description of the conformational space of the protein of interest if a sufficient number of eNOEs is collected. The application to the protein GB3 shows that it comprises distinct albeit similar structural states within the \( \beta \)-sheet and the accompanying loops that interchange most likely on the \( \mu s \) timescale while the \( \alpha \)-helix is decoupled from this motion. This interpretation of the data is visualized in
Supplementary Movie 1 showing the three states back to back suggestive of dynamics. These findings indicate that the measurement of eNOEs opens an avenue towards a comprehensive spatial description of both the 3D structure and the motion of biomolecules with the potential to uncover communication pathways between remote sites of a protein.

Acknowledgment

We thank Dr. Lei Wang for the preparation of the GB3 NMR sample. P. G. gratefully acknowledges financial support by the Lichtenberg program of the Volkswagen Foundation.

References:


determination of NOE rates in perdeuterated and protonated proteins: Practical

18. Olejniczak, E.T., Dobson, C.M., Karplus, M. & Levy, R.M. Motional averaging of
proton nuclear overhauser effects in proteins. Predictions from a molecular


20. Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M. & Ernst,
R.R. Influence of rapid intramolecular motion on NMR cross-relaxation rates. A
molecular dynamics study of antamanide in solution. J. Am. Chem. Soc. 114,

21. Leitz, D., Vögeli, B., Greenwald, J. & Riek, R. Temperature dependence of \(^1\)H\(^N\)-
\(^1\)H\(^N\) distances in ubiquitin as studied by exact measurements of NOEs. J. Phys.

22. Koning, T.M.G., Boelens, R. & Kaptein, R. Calculation of the nuclear overhauser
effect and the determination of proton-proton distances in the presence of

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


protein-G - An analysis by x-ray crystallography of the structure alone and in

26. Yao, L., Vögeli, B., Torchia, D.A. & Bax, A. Simultaneous NMR study of

27. Brüschweiler, R., Blackledge, M. & Ernst, R.R. Multi-conformational peptide
dynamics derived from NMR data: A new search algorithm and its application

Assessing the quality of solution nuclear magnetic resonance structures by

Figure Captions

Figure 1: Heavy-atom structural representations of GB3 following either the classical protocol with NOEs as experimental input, the classical protocol with eNOEs, or the ensemble-based protocol with eNOEs. Left: Bundle calculated with a classical 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 in gold, red and blue.
Figure 2: Target function values of various ensemble-based structure calculations of GB3 highlighting the importance of the ensemble-based structure calculation (left) and the self-consistency of the data by cross-validations (right).

The CYANA target function is the (weighted) sum of the squared violations of the conformational restraints. Left panel: Target function values versus number of simultaneously calculated states. The green graph shows the target function for all violations, the red graph for eNOEs only. The decrease of the target function with the number of states indicates that at least three states are required to describe the experimental data well. Right panel: Cross-validation tests highlighted by target functions. Shown in blue is the normalized target function obtained from a jackknife procedure that deletes 10 times 10% of all the experimental input data randomly (see text). The decrease of the normalized target function with increasing number of states indicates that the data set is requesting an ensemble of states. Colored in yellow, orange, and red are the target functions upon random alteration of the distances obtained from both cross peaks according to normal distributions with standard deviations of 5%, 10%, and 15%, respectively, and upon random alteration of all distances by 10% in pink. All these target function values are substantially larger than those in the left panel indicating that the original experimental data set is self-consistent.
Figure 3: Structural space coverage of the ensemble-based structure of GB3. Top panel: Distribution of $\Phi$ (top) and $\Psi$ (bottom) backbone angles versus the amino acid sequence. The angles are shown for the 9 three-state conformers. The most similar structures from each three-state conformer are grouped in gold, red and blue. Anti-correlated behavior between $\Phi$ (top) and $\Psi$ (bottom) is observed for many residues.
such as residues two and four. Bottom right panel: Backbone H^\text{N}-N order parameters versus amino acid sequence. Red values are computed from the three-state ensemble and blue values are RDC-derived values\textsuperscript{26}. Bottom left panel: Circle diagrams of the $\chi^1$ angles obtained from a three-state ensemble represented by 20 conformers each. The individual diagrams are labeled by the number of the corresponding residue. The corresponding angles from the x-ray structure 1IGD\textsuperscript{25} are indicated in red. If the anisotropically evaluated x-ray structure 2IGD\textsuperscript{25} shows a second one in addition to the former it is indicated in green, and if it exhibits a different single state in yellow.
Figure 4: Backbone structural representation of a three-state ensemble of GB3.

Top panel: The most similar structures from each three-state conformer are grouped in gold, red and blue (same as in Fig. 3). For each ensemble 9 conformers were selected. In addition to the backbone the side chains of hydrophobic core residues and the two solvent exposed residues Lys 10 and Thr 11 are also shown and labeled. The termini of the protein are also labeled. Bottom panel: First (left) and second (right) mode of principal component analysis. The arrows indicate directions and amplitudes (5 times enlarged for better visibility).