

Discrete Three-dimensional Representation of Macromolecular Motion from eNOE-based Ensemble Calculation

Beat Vögeli^{*a}, Julien Orts^a, Dean Strotz^a, Peter Güntert^b, and Roland Riek^a

Abstract: Three-dimensional structural data and description of dynamics are fundamental to infer and understand protein function. Structure determination by NMR follows well-established protocols while NMR relaxation phenomena provide insights into local molecular dynamics. However, methods to detect concerted motion were not pursued until very recently. Here, we present an ensemble-based structure determination protocol using ensemble-averaged distance restraints obtained from exact NOE (eNOE) rate constants. An application of our protocol to the model protein GB3 established an ensemble of structures that reveals correlated motion across the β -sheet and concerted motion between the backbone and side chains localized in the core. Furthermore, the data repudiate concerted conformational exchange between the β -sheet and the α -helix.

Keywords: Correlated dynamics · Dynamics · eNOE · NMR · NOE · Nuclear Overhauser effect · Proteins · Structure calculation · Structure ensemble

1. Introduction

The average conformation, as obtained from NMR spectroscopy by the established structure determination protocols, has been described as representing “... the shape of the molecule as it would be seen on a photograph taken at low shutter speeds”,^[1] and furthermore, “...the average derived from spectroscopic data represents a virtual structure devoid of physical meaning”.^[1] This view may be overly pessimistic, but certainly reflects the fact that a NMR observable is an averaged property rather than a function of an averaged structure. For example, if a molecule fluctuates between two states, the observed Nuclear Overhauser Enhancement (NOE) is the average of the two individual NOEs. The extracted distance then would yield a value that is between those of the two states.

In reality, the extracted distance is never sampled at all.

Proteins in particular are highly dynamic systems. Their internal motions cover many orders of magnitude on the spatial and time scales. As these motions are an intrinsic property of biomolecules evolution allocated them a key role in their functions. For a detailed understanding of protein function at atomic level not only the 3D atomic-resolution structures are a prerequisite but also an accurate description of dynamic properties.^[2–5] The well-established standard structure determination protocols used in NMR spectroscopy make use of the overabundant number of experimentally readily accessible Nuclear Overhauser Enhancement (NOE) rate constants – typically up to 20 per residue in small proteins.^[6] Although the NOE rate constant is proportional to the inverse 6th power of the distance between two dipolar interacting spins^[7] these rate constants are employed in a semi-quantitative manner at most because the measurement of NOEs is flawed by spin diffusion, low signal-to-noise ratio and technical limitations.

On the other hand, even though NMR relaxation phenomena provide a great deal of insight into local motion the dynamic picture is still largely incomplete because it is difficult to detect translational or concerted motion.

We have recently developed an ensemble-based structure determination protocol using ensemble-averaged distance restraints obtained from exact NOE (eNOE) rate constants and applied it to the model protein GB3 (third immunoglobulin binding domain of protein G).^[8,9]

2. Extraction of Distance Restraints

With our protocol, we first demonstrated that it is possible to obtain very precise and accurate H^N – H^N NOEs in both deuterated and protonated protein samples.^[10,11] We measured NOE buildups with optimized 3D-resolved $[^1H, ^1H]$ -NOESY experiments on cryogenic probes and converted them into precise distances. For example, distances up to 5 Å obtained from a perdeuterated ubiquitin sample have a random error of only ≈ 0.07 Å.^[10] This is considerably smaller than the 0.24 Å pairwise rms deviations from distances extracted from corresponding high-resolution NMR or X-ray structures. The extraction of exact eNOEs between spins i and j is severely hampered by spin diffusion, that is relayed magnetization transfer *via* neighboring spins.^[12,13] We have developed two related approaches to extract exact distances that are distinguished mainly by the methods to calculate the effects of spin diffusion. The experimental procedure encoded in the MATLAB program eNORA (exact NOE by Relaxation matrix Analysis) is depicted in Fig. 1.^[14] In the eNORA routine, the diagonal peak intensities derived from the NOESY spectra are fitted to mono-exponential decay functions to extract the auto-relaxation rate constants, ρ_i , ρ_j , and the initial magnetizations, $\Delta M_{ii}(0)$ and $\Delta M_{jj}(0)$. Then, cross peak build-up curves are simulated with the full relaxation matrix approach^[12] applied to a conventionally determined NMR structure or an X-ray structure. This approach takes into account spin diffusion in that the magnetization transfer pathways between all spins

^{*}Correspondence: Dr. B. R. Vögeli^a
Tel.: +41 44 633 4405

E-mail: beat.voegeli@phys.chem.ethz.ch

^aLaboratory of Physical Chemistry
Swiss Federal Institute of Technology
ETH-Hönggerberg, CH-8093 Zürich

^bInstitute of Biophysical Chemistry and Center
for Biomolecular Magnetic Resonance
Goethe University
Max-von-Laue-Str. 9
60438 Frankfurt am Main, Germany

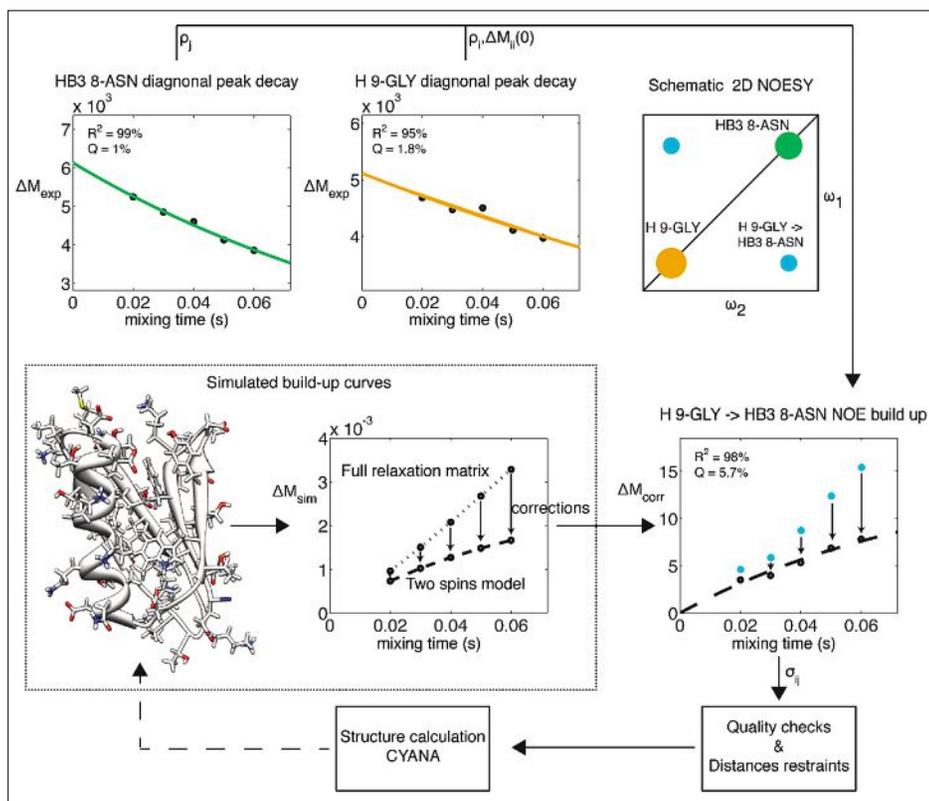


Fig. 1. Flow chart representing the method for the determination of eNOEs and the structure calculation with eNOEs. As an example, the eNOE originating from the amide H of Gly9 (spin i , orange) and enhancing H^{β} of Asn8 (spin j , green) of GB3 is shown. (1) The diagonal peak intensities derived from the NOESY spectra are fitted to mono-exponential decay functions to extract the auto-relaxation rate constants, ρ_i , ρ_j , and the initial magnetization on spin i , $\Delta M_i(0)$. (2) A build-up curve taking into account all magnetization pathways is simulated with the full relaxation matrix approach. This simulation requires a 3D structure as input, which may be based on a conventionally determined NMR structure or an X-ray structure. (3) Corrections for the intensities at each mixing time are applied to the experimental NOE build-up. (4) The NOE build-up is fitted, the quality of the fit is evaluated and distance restraints are created. (5) A structure calculation is performed with the new distance restraints using established software packages such as CYANA.^[15,16] This structure may be used as an input for (2) in a new cycle as indicated by the broken arrow.

are active simultaneously. Corrections for the intensities at each mixing time are derived from the simulation and applied to the experimental intensities. The corrected cross-peak build-up curves are fitted by using ρ_i , ρ_j and $\Delta M_i(0)$ and $\Delta M_j(0)$ as fixed input parameters and the cross relaxation rate constants σ_{ij} and σ_{ji} as free variables. The quality of the fit is evaluated and σ_{ij} and σ_{ji} are converted into distance restraints r through the relationship $\sigma \propto r^{-6}$. A structure calculation is then performed with the new distance restraints using software packages such as CYANA.^[15,16] This new structure may be used again as an input for the next correction simulation and refinement of the structure.

The second approach to derive exact distances is centered on the MATLAB script DOMINO.^[10,11] Here, instead of calculating the simulated intensities *via* the full relaxation matrix approach, individual correction contributions from each neighboring spin k obtained from the exact solution of three-spin systems ijk are summed

up. This approach is particularly well suited for partially deuterated proteins. We have shown that this approach is in agreement with the eNORA approach for a large range of molecular overall tumbling times and NOESY mixing times where the spin diffusion is well traceable.^[14]

In a first application of the method, we made use of the unprecedented accuracy in distance extraction. We measured exact effective distances between amide protons in ubiquitin over a temperature range of 50 K.^[17] The average eNOE increases translate into effective distance changes of 4%. Such a minute effect is virtually impossible to detect with other methods.

We then extended the method to aliphatic protons.^[9] It became possible to map a dense eNOE network throughout the entire 56-residue protein GB3^[18] for which we collected 823 eNOEs (Fig. 2, top panel).^[8] The eNOEs were converted into distances by the notion that the NOE is mostly sensitive to slow motion. This was shown to be valid for H–H spin pairs

if the local H–X squared order parameters, which are a measure of fast motion, are larger than 0.5.^[17]

3. Structure Calculation with eNOEs

The eNOE-derived distance restraints in combination with a small set of residual dipolar couplings (RDCs) and angle restraints derived from scalar couplings and C^{α} chemical shifts were used for a structure calculation of GB3 following standard protocols using the software package CYANA.^[15,16] The input data results in an extremely tight structure with a backbone rmsd of 0.11 Å and an all-heavy-atom rmsd of 0.60 Å only. When compared to the rms deviations obtained from a traditional structure calculation using semi-quantitative NOEs (backbone 0.47 Å; all-heavy-atom 0.87 Å) the high precision is particularly striking. Furthermore, the eNOE-based structure coincides closely with the highly accurate RDC-optimized X-ray structure of GB3^[19,20] with an rmsd of 0.57 Å for the backbone and 1.17 Å for all heavy atoms. However, the large CYANA target function value of 27.5 Å² resulting from many distance restraint violations indicates that this static structure does not agree with the experimental data. The large number of violations of experimental restraints is caused by the motion-averaged nature of the measured NOEs. It is obvious that exact NOEs reveal a wealth of information about internal motion that is sacrificed in routine structure calculation protocols. The resulting effect on the structure calculation when using traditional methods is depicted in Fig. 3. It is clear that an ensemble of structural states of which the averaged NOEs have to match the experimental eNOEs would be a better representation of the structure.^[21]

4. Multiple-state Structure Calculation with eNOEs

To obtain structural ensembles consistent with the eNOEs we have modified CYANA.^[15,16] A comparison between GB3 structures obtained with the eNOE-based routine protocol and the multiple-state ensembles protocol is shown in Fig. 2. A normalized target function decreases when changing from the single-state structure, which corresponds to the conventional protocol (yet with eNOEs), to the two- and three-state ensembles after which a plateau is reached. We chose the three-state ensemble as a representative of the GB3 structure in solution because it is the smallest ensemble that fits the experimental data well. To strengthen this

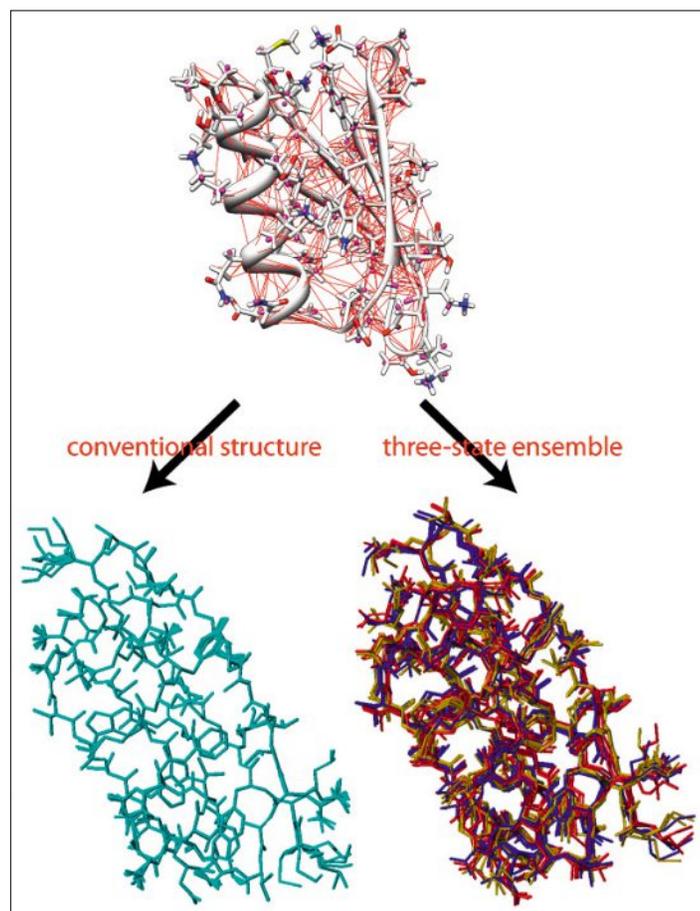


Fig. 2. Structure calculation of GB3 using eNOEs. The top panel shows the 823 eNOEs as red lines superimposed on a cartoon representation of the heavy atoms. The eNOEs were converted into distance restraints for structure calculation with CYANA. A structural bundle of nine conformers obtained with a conventional protocol (left bottom panel) is compared to a three-state ensemble (right bottom panel). Three conformers each exhibiting three states are shown. Rms deviations of the backbone are 0.11 Å and 0.47 Å, and for all heavy atoms 0.60 Å and 0.86 Å, respectively, for the conventional bundle and the three-state ensemble.

finding all eNOE-derived distances were arbitrarily changed by up to 15% yielding considerable increases of the target functions. This observation indicates that the eNOE-derived restraints have an accuracy much better than 15%. In addition, this shows that the experimental data set is self-consistent and at least in part over-determined, even though the free parameter space was enlarged extensively 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.^[22] The test shows that the violations of the omitted eNOE-derived distances summed over ten structure calculations (overall deleting every eNOE exactly once) decrease with the number of included states and result in a drop of the target function by up to 40% when compared with the single-state structure.

The three structural states in the ensemble are distinct from each other. Individual-state sub-bundle representations are ob-

tained by grouping the most similar states from each three-state conformer as shown in Fig. 4. The sub-bundle for each structural state is a measure for the precision of the individual structural states. The same bundle representation can be used for the entire β -sheet and some of the loops. Thus, this entire structural segment undergoes conformational exchange between the three states in a concerted manner. The loops connecting β_1/β_2 and β_3/β_4 and accompanying segments of the β -strands counteract the motion of the central parallel β -sheet in an anti-correlated manner. In contrast, the α -helix is decoupled from the conformational exchange of the β -sheet as the bundle representation of the β -sheet does not allow one to distinguish several structural states of the α -helix. However, a selection applied to the α -helix indicates that the helix backbone exhibits also distinct structural states that interconvert between each other. The correlation appears to be weaker than for the β -sheet and is localized in the residues that face the hydrophobic core.

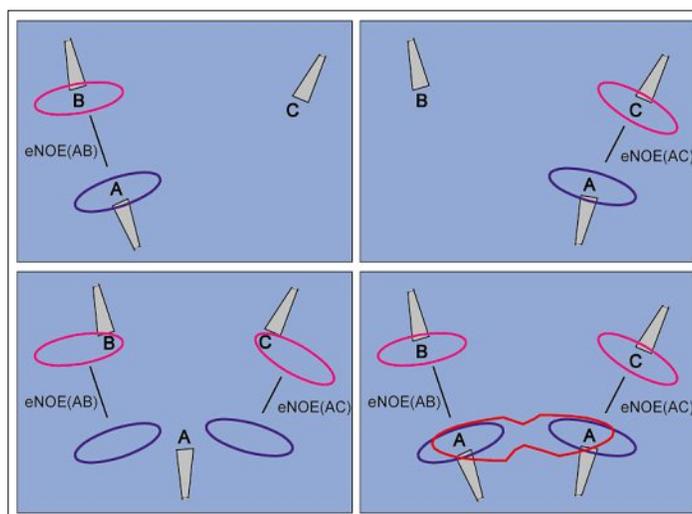


Fig. 3. Schematic representation of the effect of multiple eNOEs on structure calculation. eNOEs between spins A and B, and A and C are measured. Use of only the A-B eNOE leads to the situation shown in the left top panel and similarly for the A-C eNOE in the right top panel. The blue ellipsoid shows the expected position of spin A and the pink ellipsoid of spins B and C. Simultaneous use of both eNOEs in a single-state calculation would place spin A between the two positions and pull spins B and C towards each other (left bottom panel). In a two-state ensemble, spin A is restricted to the red area, where it is once located in the left and once in the right ellipsoid (right bottom panel).

5. Conclusion and Future Direction

The eNOE offers advantages over other novel methods introduced into NMR to extend its application range. Since eNOEs do not require any modification of standard NMR samples, such as the addition of alignment media (as in the case of RDCs) or paramagnetic labels (PRE measurements), it is straight-forward to measure them and merely additional measurement time is required. No additional (and often time consuming) spectral assignment efforts are required because NOESY assignments are indispensable in structural NMR work. In addition, the NOE ranks among the observables that are measurable even for high molecular weight samples such as large proteins, protein complexes or membrane proteins substituted in membrane-mimicking environments. We are currently exploring the use of the eNOE technique on highly deuterated, methyl-reprotonated samples, for the characterization of intermolecular interactions and for quick stereospecific assignments of prochiral groups.

By taking into account the motional dependence of the eNOE, we established an ensemble-based NMR structure determination protocol resulting in a description of the conformational space occupied by the protein of interest (if a sufficient number of eNOEs can be collected). The demonstrated use of eNOEs opens an avenue towards a discrete spatial description of three-dimensional structures and inter-

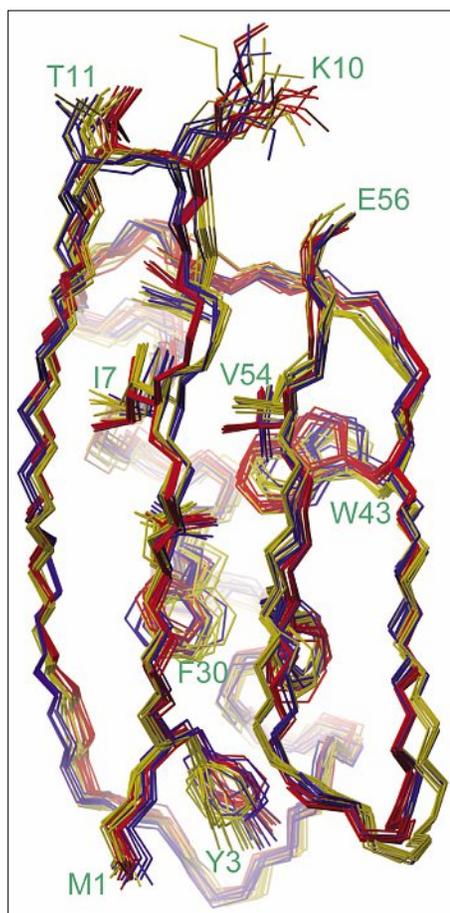


Fig. 4. Representation of three states of GB3 obtained from the ensemble-based protocol using eNOEs. The most similar structures from each three-state conformer are grouped in gold, red and blue. For each ensemble 9 conformers were selected. The backbone, the side chains of the hydrophobic core and the two solvent-exposed residues L10 and T11 are also shown. The termini and the side chains are labeled with the residue number.

nal motions of biomolecules. The method holds promise to allow the detection of small structural differences between molecules under different conditions, trace folding/unfolding events close to the melting temperature or to uncover communication pathways between remote sites of a protein.

Acknowledgement

We gratefully acknowledge support of the Eidgenössische Technische Hochschule in Zürich, the Federation of the European Biochemical Society (FEBS long-term fellowship to J.O.) and the Lichtenberg program of the Volkswagen Foundation (financial support to P.G.).

Received: August 7, 2012

- [1] O. Jardetzky, *Biochim. Biophys. Acta* **1980**, *621*, 227.
- [2] A. Fersht, 'Structure and Mechanism in Protein Science', **1998**, Freeman, San Francisco.
- [3] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, *Science* **2010**, *330*, 341.
- [4] P. I. Zhuravlev, G. A. Papoian, *Quart. Rev. Biophys.* **2010**, *43*, 295.
- [5] J. H. Chill, F. Naider, *Curr. Opin. Struct. Biol.* **2011**, *21*, 627.
- [6] K. Wüthrich, 'NMR of Proteins and Nucleic Acid', **1986**, Wiley, New York.
- [7] I. Solomon, *Phys. Rev.* **1955**, *99*, 559.
- [8] B. Vögeli, S. Kazemi, P. Güntert, R. Riek, *Nat. Struct. Mol. Biol.* **2012**, doi: 10.1038/nsmb.2355.
- [9] B. Vögeli, P. Güntert, R. Riek, *Mol. Phys.*, in press.
- [10] B. Vögeli, T. F. Segawa, D. Leitz, A. Sobol, A. Choutko, D. Trzesniak, W. van Gunsteren, R. Riek, *J. Am. Chem. Soc.* **2009**, *131*, 17215.
- [11] B. Vögeli, M. Friedmann, D. Leitz, A. Sobol, R. Riek, *J. Magn. Reson.* **2010**, *204*, 290.
- [12] J. W. Keepers, T. L. James, *J. Magn. Reson.* **1984**, *57*, 404.
- [13] A. Kalk, H. J. C. Berendsen, *J. Magn. Reson.* **1976**, *24*, 343.
- [14] J. Orts, B. Vögeli, R. Riek, *J. Chem. Theory Comput.* **2012**, doi:10.1012/ct3002249.
- [15] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283.
- [16] P. Güntert, *Eur. Biophys. J.* **2009**, *38*, 129.
- [17] D. Leitz, B. Vögeli, J. Greenwald, R. Riek, *J. Phys. Chem. B* **2011**, *115*, 7648.
- [18] J. P. Derrick, D. B. Wigley, *J. Mol. Biol.* **1994**, *243*, 906.
- [19] T. S. Ulmer, B. E. Ramirez, F. Delaglio, A. Bax, *J. Am. Chem. Soc.* **2003**, *125*, 9179.
- [20] L. Yao, B. Vögeli, D. A. Torchia, A. Bax, *J. Phys. Chem. B* **2008**, *112*, 6045.
- [21] R. Brüschweiler, M. Blackledge, R. R. Ernst, *J. Biomol. NMR* **1991**, *1*, 3.
- [22] A. T. Brünger, G. M. Clore, A. M. Gronenborn, R. Saffrich, M. Nilges, *Science* **1993**, *261*, 328.