

Relaxation Matrix Analysis of Spin Diffusion for the NMR Structure Calculation with eNOEs

Julien Orts, Beat Vögeli, and Roland Riek*

ETH, Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, Wolfgang-Pauli-Strasse 10, 8093 Zürich, Switzerland

S Supporting Information

ABSTRACT: NMR structure determination is usually based on distance restraints extracted semiquantitatively from cross peak volumes or intensities in NOESY spectra. The recent introduction of exact NOEs (eNOE) by Vogeli et al. opens an avenue for the ensemble-based structure determination of proteins on the basis of eNOE-derived quantitative distance restraints. We present an approach to extract eNOE from build-up curve intensities. For the determination of eNOEs, spin diffusion is a major source of errors. A full relaxation matrix analysis is used to calculate the spin diffusion contribution to the NOESY cross peaks of each individual spin pair of interest. A software program is written, which requires as input the peak intensities from the various NOESY spectra as well as a 3D structure of the protein. This structure can be either an X-ray structure or an NMR structure determined with the conventional approach. The outputs of the program are the eNOE rates, the autorelaxation rates, as well as graphs and quality factors from the individual NOE build-up curves for semiautomated analysis of the derived rates. The protocol is straightforward, and the program integrates well into the current structure calculation workflow.

INTRODUCTION

The most prominent information for the structure determination of biomolecules by solution state NMR is the distance between protons.^{3–7} For decades, the standard procedure of NMR structure determination relied on deriving a large amount of semiquantitative NOEs (Nuclear Overhauser Enhancement) between protons from NOESY spectra.^{8–12} Although the NOE rate is directly related to the internuclear distance,¹³ only a semiquantitative interpretation is used to convert NOEs to distance restraints, such that NOEs are either classified into few groups only (i.e., weak, medium, strong) or translated into distances with large tolerances, and only upper distance restraints are used. Therefore, current structure determination methods rely on the quantity of the data rather than the quality. The reasons for the semiquantitative translation from NOEs to distance restraints include peak overlap in the NOESY spectrum, the lack of relaxation measurements, uncontrolled relaxation during the NMR experiments, the time and ensemble-averaged nature of the NOE, and very importantly the presence of spin diffusion in the NOESY experiments. Spin diffusion is defined as the transfer of magnetization between two spins of interest via remote spins contributing to the cross peak in the NOESY spectra.¹⁴

Spin diffusion is one of the major causes for inaccuracy in deriving distance restraints.¹⁵ As opposed to other NOE-based applications where spin diffusion is needed,^{16–20} in the standard NMR structure calculation workflow its effect is usually largely suppressed by the use of relatively short mixing times. Several approaches were developed in order to correct NOE restraints for spin diffusion.^{21–29} The complete description of spin diffusion is made by the full relaxation matrix formalism applied to the Solomon equations.^{13,30} When all NOE cross and diagonal peaks can be measured unambiguously, methods for back-transformation of the full NOE intensity matrix into interproton cross-relaxations rates

would be successful. The strength of this approach is that it does in principle not require any (prior) knowledge about the 3D structure of the studied molecule. Applications of this approach to small molecules can be found in the literature.³¹ When the size of the molecules increases, spectral crowding yields an incomplete data set, and the full relaxation matrix approach fails. While partial deuteration of the molecule is a possible solution to crowding problems, it creates difficulties in the full relaxation matrix approach (to be discussed later). Because of the expected incompleteness of the data set, “hybrid” back-transformation methods combining experimental and simulated NOEs, the latter based on a (preliminary) 3D structure, together with an iterative procedure have been developed. This treatment results in a NOE-based refined structure.^{26,32,33} Applications using 2D NOE data can be found in the literature.^{34–36}

In our recent studies on exact NOEs (eNOEs) and their use for ensemble-based structure determination, short mixing times were used to limit the extent of spin diffusion, and a simple three-spin approach was used to calculate spin diffusion contributions.^{1,37} Here, we use the full relaxation matrix approach to calculate the contribution of spin diffusion to the NOE, based on an available 3D structure, in order to derive eNOE rates. The impact of this improved approach is discussed including the case of partially deuterated molecular systems. Furthermore, we established the software program eNORA (exact NOE by Relaxation matrix Analysis) for the eNOE rate determination. eNORA integrates well within and between other software tools for structure determination, such as spectrum analysis software and structure calculation software. Application of the method to GB3,^{38–40} a 56-residue globular

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: March 18, 2012

Published: May 15, 2012



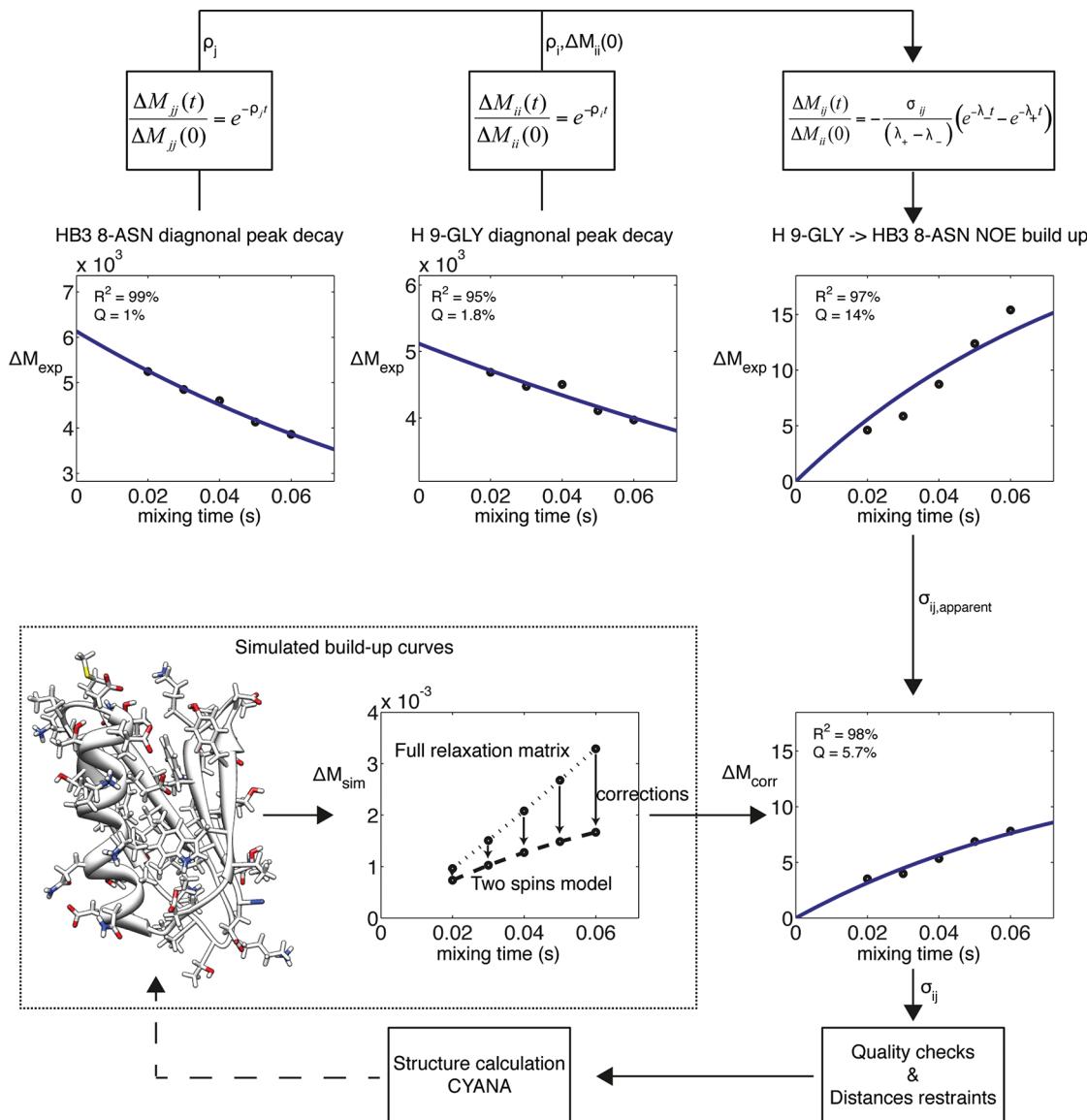


Figure 1. Flowchart representing the method for the determination and correction of the exact NOEs as well as structure calculation with eNOEs. (1) The diagonal peak intensities derived from the NOESY spectra are fitted to monoexponential decay functions to extract the autorelaxation rates, ρ_i and ρ_j , and the initial magnetizations, $\Delta M_{ii}(0)$. (2) All cross peak NOE build-up curves are fitted initially with a two-spin model by using ρ_i , ρ_j , and $\Delta M_{ii}(0)$ as fixed input parameters. (3) 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 an X-ray structure or a conventionally determined NMR structure. (4) Corrections for the intensities at each mixing time are derived, and the experimental build-up is corrected as indicated. (5) The NOE build-up is again fitted. The quality of the fit is evaluated, and distance restraints are created. (6) A structure calculation is performed with the new distance restraints using established software packages such as CYANA.⁴¹ This structure may be used as an input for 3, as indicated by the broken arrow.

protein, is presented. Using an experimental data set the eNOE-derived structure is obtained and compared to a conventionally NMR-derived structure and to a high resolution X-ray structure.

METHODS AND THEORY

Theoretical Basis. NOE peak intensities follow the Solomon equations to a very good approximation:^{13,30}

$$\frac{d}{dt}\mathbf{M}(t) = -\mathbf{R}(\mathbf{M}(t) - \mathbf{M}_0) \quad (1)$$

The solution of eq 1 is

$$\mathbf{M}(\tau_m) = \exp(-\mathbf{R}\tau_m)(\mathbf{M}(0) - \mathbf{M}_0) + \mathbf{M}_0 \quad (2)$$

where \mathbf{R} is the relaxation matrix, \mathbf{M}_0 is the equilibrium magnetization, $\mathbf{M}(0)$ is the starting magnetization, and τ_m is the mixing time. Individual elements of the relaxation matrix are expressed as follows:

$$R_{ii} = \rho_i = \sum_{j \neq i} \frac{b^2}{r_{ij}^6} (J(0) + 3J(\omega) + 6J(2\omega)) \quad (3.1)$$

$$R_{ij} = \sigma_{ij} = \frac{b^2}{r_{ij}^6} (6J(2\omega) - J(0)) \quad (3.2)$$

with

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1 + (\omega\tau_c)^2} \right) \quad (4.1)$$

assuming isotropic tumbling of a rigid protein and with

$$b = \frac{1}{2} \frac{\mu_0}{4\pi} \hbar \gamma_H^2 \quad (4.2)$$

ρ_i is the autorelaxation rate of the proton H_i , σ_{ij} is the cross-relaxation rate between protons H_i and H_j , and r_{ij} is the distance between H_i and H_j .

The exact analytical solution for a two-spin system is described by

$$\frac{\Delta M_{ii}(t)}{\Delta M_{ii}(0)} = \frac{1}{2} \left[\left(1 - \frac{\rho_i - \rho_j}{\lambda_+ - \lambda_-} \right) e^{-\lambda_- t} - \left(1 + \frac{\rho_i - \rho_j}{\lambda_+ - \lambda_-} \right) e^{-\lambda_+ t} \right] \quad (5.1)$$

$$\frac{\Delta M_{ij}(t)}{\Delta M_{ii}(0)} = -\frac{\sigma_{ij}}{(\lambda_+ - \lambda_-)} (e^{-\lambda_- t} - e^{-\lambda_+ t}) \quad (5.2)$$

with

$$\lambda_{\pm} = \frac{(\rho_i + \rho_j)}{2} \pm \sqrt{\left(\frac{\rho_i - \rho_j}{2} \right)^2 + \sigma_{ij}^2} \quad (5.3)$$

where $\Delta M_{ii}(t)$ is the diagonal peak intensity and $\Delta M_{ij}(t)$ is the cross peak intensity at time t . Obviously, the solution for a multispin system is much more elaborate and can be determined only numerically.

Method and Software Workflow. The presented 3D structure determination protocol based on distance restraints from eNOEs can be split up into the following steps: (i) starting material (such as sample preparation, NOESY spectra, sequential assignment, conventional structure calculation), (ii) spectra analysis, (iii) determination of the rates, (iv) assessment of the quality of the data, (v) determination of the distance restraints, and (vi) structure calculation. Step i includes the measurement of a series of NOESY spectra as well as the NMR structure determination using conventional protocols. For ii and vi, established software packages are used, such as nmrPipe⁴² for the spectra analysis and CYANA for the structure calculation, while steps iii–v are carried out using the presented eNOE software eNORA, written in MATLAB.⁴³ Figure 1 shows the workflow of steps iii–v. Figure S1 in the Supporting Information also highlights the formats of the input and output files as well as the various variables to be set by the user.

Starting Material. The eNOE-based structure determination procedure starts in general after a standard NMR structure calculation with common software like CYANA, Aria, or Xplor,^{41,44,45} yielding a 3D structure and an assigned NOESY peak list as well as optionally a chemical shift list. For the structure calculation, our laboratory mainly uses CYANA. Therefore, we will refer to it, but the method does not rely on specific structure calculation software packages. Alternatively, one may use an X-ray structure together with an assigned NOESY peak list. In addition, a set of NOESY spectra with different mixing times is required in order to get time-resolved cross peak build-up curves as well as decays of the diagonal. Furthermore, an accurate determination of the correlation time

of the molecule is necessary, which usually is determined from ¹⁵N T_1 and $T_{1\rho}$ measurements.⁴⁶

Spectra Analysis. An assigned NOESY peak list must be available for the spectra analysis. It can be taken from the standard structure calculation (see above). On the basis of this list, the NOESY peak intensities or volumes must be quantified. Different integration procedures have been described in the literature,^{42,47,48} ranging from integration over boxes/ellipses to direct intensity measurement. We use line shape fitting from NMRDraw⁴² in order to derive the height of the NOESY cross peaks and diagonal peaks. It is evident that the integration has to be consistent over all spectra.

Unambiguous Cross Peaks. The extraction of eNOE rates requests the use of cross peaks that are unambiguously assignable. The eNORA software does select such cross peaks optionally. It requests thereby a list of chemical shifts in addition to the 3D structure and the NOE peak list. The eNORA software classifies a peak as unambiguous if there is only a single possible assignment. The criteria are based on the assumption that a cross peak is observed in the NOESY spectrum only if the two involved spins are closer than 6 Å in the input structure (or by a user defined value) and that such calculated cross peaks differ in their chemical shifts by at least 0.02 ppm in the ¹H dimensions and 0.2 ppm in the heavy atom dimensions, respectively (again the user may define the thresholds).⁴⁹ In the opposite case, the restraint will not be used for structure calculation.

Deriving the Relaxation Rates. Once the input data for the program are prepared, the eNORA software fits the experimentally measured diagonal peak intensities to mono-exponential decay functions to extract the autorelaxation rates ρ_i and ρ_j and the initial magnetizations $\Delta M_{ii}(0)$ (Figure 1). In addition, for each fit, a graph showing the experimental data and the fit are displayed (Figure 1) together with measures that check the quality of the fit (i.e., R^2 and Q , see below). Subsequently, eNORA fits all cross peak NOE build-up curves initially with a two-spin model by using ρ_i , ρ_j , and $\Delta M_{ii}(0)$ as fixed input parameters yielding initial cross-relaxation rates. In addition, for each NOE build-up, a graph with the experimental data points, the fit, and measures for the quality of the fit (i.e., R^2 and Q) is given as output for quality control (Figure 1, see below).

As mentioned in the Introduction, the cross-relaxation rates resulting from these fits are biased by spin diffusion. Therefore, the build-up curves must be corrected for spin diffusion. The corrections are made on the basis of the available structure (either an X-ray or the NMR derived structure is used as input in PDB format, Figure 1). In detail, for each experimental NOE build-up curve, eNORA simulates the same build-up curve, based on the 3D structure. The simulation uses the full relaxation matrix approach and takes into account all magnetization pathways within a definable sphere centered at each of the two protons involved in the NOE. The radius of the sphere is set by the user (the default value is 6 Å). Subsequently, eNORA derives a correction for the experimental intensities at each mixing time by taking the intensity ratio between the full relaxation matrix build-up and the two-spin system build-up as depicted in Figure 1. The experimental build-up is then corrected accordingly. The extraction of an experimental cross-relaxation rate with a two-spin model is now possible and accurate. The use of build-up curves is regarded as a strength of the fitting procedure because the complete build-up curve contains more information than the set of intensities

treated separately for each mixing time and allows for a visual inspection.

Alternatively, it is possible to directly correct the cross-relaxation rate by a correction factor. Even though this is not the default procedure, this approach is also implemented in the eNORA software. In this approach, the program fits the simulated build-up curve using the full relaxation matrix formalism with a two-spin system in order to extract an apparent simulated cross-relaxation rate (the required auto-relaxation rates are also fitted with a monoexponential to the diagonal peak decays simulated by the full relaxation matrix, and the initial magnetization is 1). Then, the ratio between the fitted and the calculated two-spin cross-relaxation rates is the correction factor applied to the experimental cross-relaxation rate. Both methods for deriving the correction factors lead to highly similar results, but the first method provides corrected build-up curves that can be plotted and visualized.

The eNORA software treats further the methyl groups and nonstereospecifically assigned methylenes as pseudo-atoms and corrects automatically for their proton multiplicity. It addresses also the following issue: Since there are larger overlaps for the diagonal peaks than the cross peaks, high quality NOE build-ups may be obtained, but the cross-relaxation rates cannot be extracted because of the lack of the required information from the diagonal fits (i.e., ρ_i). In such cases, eNORA opens the possibility to replace the ill-defined autorelaxation rates with the average value derived from the best fits from other diagonal decays.

The eNORA software output for each NOE build-up is the exact NOE rate as well as a graph showing the fit to the experimentally derived and spin diffusion-corrected intensities together with quality factors as discussed in the following.

Assessing the Quality of the Data. As indicated above, the NOE build-up curves are fitted twice, before and after spin diffusion correction. The quality of the fit can be evaluated using the Pearson's correlation coefficient R^2 as provided in eq 6, the normalized sum of squared errors Q as provided in eq 7, and the visual inspection of the fit and the data by the user. All build-up and decay curves are thereby saved as plots for visual inspection. Atom types, residue numbers of the involved protons, and R^2 and Q coefficients are available in the picture file so that the user can conveniently screen the data related to each restraint and judge them individually (Figure 1).

$$R^2 = \frac{\left(\sum_i (I_{\text{obs},i} - \langle I_{\text{obs}} \rangle)(I_{\text{calc},i} - \langle I_{\text{calc}} \rangle)\right)^2}{\sum_i (I_{\text{obs},i} - \langle I_{\text{obs}} \rangle)^2 \sum_i (I_{\text{calc},i} - \langle I_{\text{calc}} \rangle)^2} \quad (6)$$

$$Q = \sqrt{\frac{\sum_i (I_{\text{obs},i} - I_{\text{calc},i})^2}{\sum_i I_{\text{obs},i}^2}} \quad (7)$$

Sufficient quality fits for both the diagonal peak decay and the cross peak build-ups are defined by the Pearson's correlation coefficients or Q factors. It is important to include the quality of the diagonal decay in the assessment because if $\Delta M_{ii}(0)$ were obtained from an ill-defined data set, the cross-relaxation rate would suffer proportionally, while the extraction of the cross-relaxation rate is relatively insensitive to slight errors of the autorelaxation rates. Unless the operator decides differently upon visual inspection, the restraints are kept for the structure calculation if R^2 or Q is, respectively, above or below a threshold controlled by the user (the default values are 0.8 for R^2 and 0.2 for Q , but they are adjustable for each data set).

Distance Restraints. eNOE extracted distances are converted automatically into upper and lower distance restraints by eNORA for the structure calculation. If the cross peak build-ups on both side of the diagonal as well as both corresponding diagonal decays have good quality fits, the upper and lower distance restraints are set identical and equal to the geometrically averaged distances as described in our previous work,³⁷ $\sigma = (\sigma_{ij}\sigma_{ji})^{1/2}$. This approach does not only cancel the effect of different incomplete longitudinal relaxation of the magnetization during the recycling delay but also removes errors due to different magnetization pathways during the pulse sequence.³⁷ If only one build-up is present, 15% error is assumed, and the upper and lower distance restraints are modified accordingly. If pseudo-atoms are involved in the NOE a penalty of 15% is also applied to the upper and 1 Å is subtracted from the lower distance restraint following established procedures.^{4,50} While these penalties are the default, the user can also choose them. The output of eNORA is a list of distance restraints (both lower and upper level distance restraints), which can be used for the structure calculation and a list of "pseudo-bonds" representing the restraints that can be visualized for example with the software tool Chimera,⁵¹ as shown in Figure 2 for the model protein GB3.

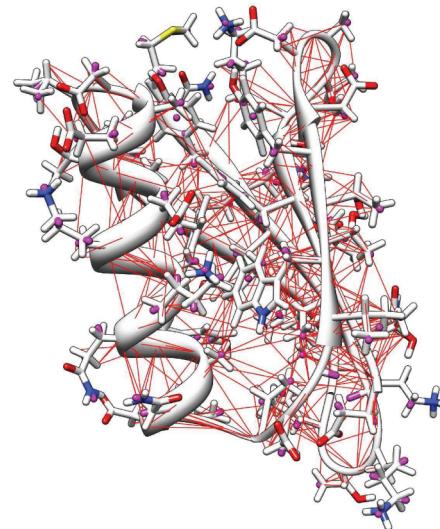


Figure 2. Ribbon representation of the 3D structure of GB3 with 781 eNOE-derived distance restraints shown in red. The side chains are depicted with bonds involving carbons atoms shown in white, involving oxygen atoms colored in red, nitrogen atoms colored in blue, sulfur atom in yellow, and pseudo-atoms colored in pink, respectively.

All scripts are available upon request and on the authors' Web page: <http://www.bionmr.ethz.ch/downloads/index>.

Structure Calculation. The eNORA software generated both a list of 781 upper distance restraints and a list with 781 lower distance restraints for the input of the structure calculation by CYANA.⁴¹ Using this data set in addition to 147 angle restraints from secondary chemical shifts,² a CYANA standard structure calculation was performed. A total of 100 conformers were initially generated by CYANA, and the bundle of 20 conformers with the lowest target functions was used to represent the 3D NMR structure shown in Figure 4B.

Experimental Procedures. The sample³⁷ and the experimental data have been previously used in another

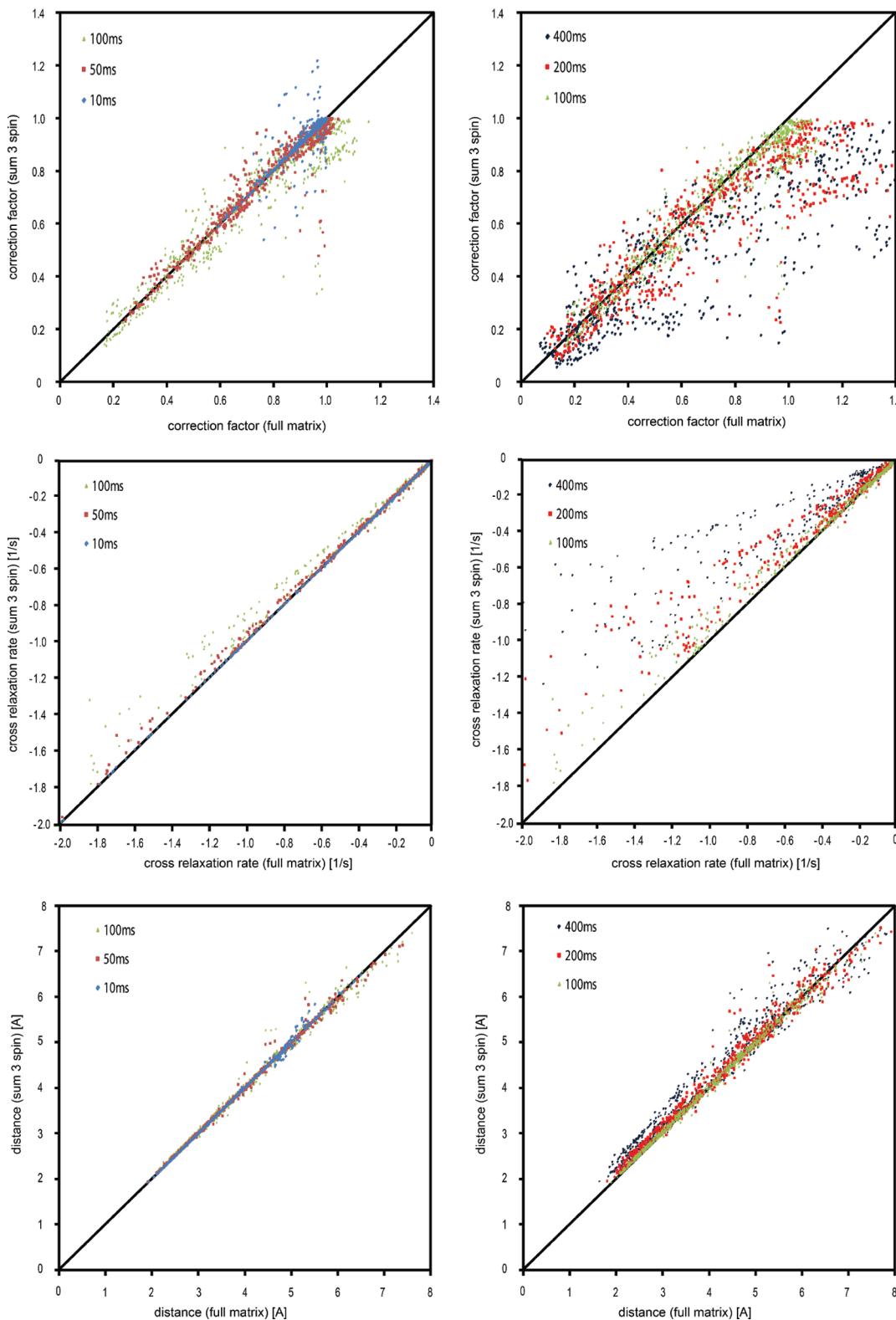


Figure 3. Comparison of the two methods, (i) full matrix relaxation approach and (ii) sum of three-spin systems approach, for calculating the corrections for spin diffusion for the determination of eNOE-derived distances for the protonated protein GB3. The calculations are based on a proton position-optimized X-ray structure of GB3.^{38–40} Data are shown for spin pairs for which a good NOE build-up fit could be obtained experimentally.² In the top graphs, the correction factors derived from the simplified approach using a sum of three spins are plotted versus those from the full relaxation matrix approach. In the middle panel, the comparisons between the correspondingly corrected eNOE rates are shown, while in the bottom panel the derived distances are shown. All of the comparisons are shown for five series of four NOESY spectra with equidistant mixing-time steps simulated with maximal mixing times $t_{\text{mix,max}}$ of 10, 50, and 100 on the left and $t_{\text{mix,max}}$ of 100, 200, and 400 ms in the right panels, respectively. The black lines indicate a slope of 1. The full relaxation matrix approach is indicated as “full matrix”, while the simplified three-spin system approach is indicated as a “sum 3 spins”.

study.² Briefly, the experiments were performed at 298 K on a Bruker 700 MHz spectrometer equipped with a triple resonance cryoprobe. Series of 3D [¹⁵N, ¹³C]-resolved [¹H, ¹H]-NOESY spectra were recorded with 20, 30, 40, 50, 60, and 100 ms mixing times. The spectrum at 100 ms was used for the conventional structure calculation and the series up to 60 ms for the build-up fits. The spectra were recorded with $200(t_1) \times 40(t_2) \times 1024(t_3)$ complex points; maximal evolution times $t_{1\max,1H} = 22.0$ ms, $t_{2\max,15N} = 14.4$ ms, $t_{2\max,13C} = 7.6$ ms, and $t_{3\max,1H} = 102.4$ ms; an interscan delay of 0.6 s; and four scans per increment, resulting in a measurement time of 1 day per spectrum. Spectra were processed with NMRPipe and analyzed with NMRDraw.⁴² The time-domain data were multiplied with a squared cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to $1024 \times 128 \times 512$ complex points. A correlation time τ_c of 4.15 ns was determined from ¹⁵N T_1 and $T_{1\rho}$ relaxation measurements.⁴⁶

RESULTS AND DISCUSSION

The work presented here establishes on the one hand the implementation of the full relaxation matrix approach to calculate correction factors for the subtraction of the contribution of spin diffusion in NOE build-ups in order to extract eNOE cross-relaxation rates. On the other hand, software has been written that enables the determination of eNOE cross-relaxation rates using coordinates from a 3D structure (in PDB format) as input data and a mixing time series of assigned peak intensities or volumes. The output of this software is lists of distance restraints (both upper and lower range distance restraints) for the structure determination as well as graphs and measures to validate the quality of the extracted data.

Full Relaxation Matrix Approach for the Calculation of Spin Diffusion in the Determination of eNOE Cross-Relaxation Rates in a Protonated Protein. The phenomenon of spin diffusion and its effect in determining quantitative cross-relaxation rates and concomitantly exact ensemble-averaged distances is a well-known problem in NMR structure determination.^{15,21,26,28,31,52–54} In our recent investigations to determine eNOE rates, the contribution of spin diffusion to the build-up curves was calculated by a simple approach dealing with three spins.^{1,37} In detail, to evaluate the contribution of spin diffusion, the DOMINO MATLAB program¹ simulates build-up curves assuming either a two-spin system or a sum of three-spin systems with the individual clandestine spins 6 Å in the vicinity of both spins of interest. The NOE build-ups were simulated by solving the corresponding differential equations analytically in the case of the two-spin system or numerically in the three-spin systems. It is obvious that this approach does not take into account either spin diffusion pathways that involve more than two steps of magnetization transfer or the simultaneous and parallel evolution of various spin diffusion pathways. The well-established full relaxation matrix approach used here takes into account all of these points and therefore provides a more rigorous description of the magnetization pathways. The exponential of the relaxation matrix over time, eq 2, is calculated with the function “expm” in Matlab that uses the Padé approximation with scaling and squaring.⁵⁵ Figure 3 shows a comparison between the two methods applied to the protonated protein GB3 with a rotational correlation time of 4.15 ns using the proton position-optimized X-ray structure of GB3^{38–40} for the calculations. Five series of four NOESY

spectra with equidistant mixing-time steps were simulated with maximal mixing times $t_{\text{mix,max}}$ of 10, 50, 100, 200, and 400 ms. Comparative data are shown for spin pairs for which a good NOE build-up fit could be obtained experimentally.² For short mixing times, most of the calculated correction factors of the two methods are similar. Generally, a very high correlation is obtained for mixing times up to 100 ms. The few large outliers for the correction factors are obtained for very small cross-relaxation rates (Figure 3) and thus do not appear in the plots of corrected rates (Figure S2). The largest systematic disagreement is obtained for NOEs which involve methylene protons (Figure S2). For long mixing times (≥ 100 ms), however, the discrepancy between the two methods is apparent. For the majority of data points, the simplified three-spin approach produces lower correction factors as compared to the full relaxation matrix approach (Figure 3). Corresponding observations are also obvious in the comparison of the corrected cross-relaxation rates as well as the eNOE-derived distances (Figure 3). For long mixing times, the three-spin approach underestimates the true eNOE rate and thus overestimates the distance (up to 0.18 Å in average at 400 ms mixing time). For the distances, the effect is smaller since the translation from eNOE rates to distance has a $1/r^6$ dependence. In summary, as expected, the full relaxation matrix formalism calculates the contribution of the spin diffusion in a protonated protein better than the previously used three spin approach. In particular, for long distances and long mixing times, it is the method of choice. While this is true for fully protonated proteins, we shall see below that the full relaxation matrix approach cannot describe partially deuterated systems in a straightforward manner.

The Calculation of Spin Diffusion in Partially Deuterated Molecules. To date, almost all quantitative measurements of NOE build-ups were conducted on protonated samples. In practice, this approach is limited to molecules no larger than ca. 35 kDa. To circumvent this limitation, samples with partial deuteration are used to reduce spectral overlap and relaxation. For NOE applications, these are typically fully deuterated/perdeuterated samples, ca. 50% deuterated samples, or fully deuterated/methyl-reprotonated samples.^{56,57} SAIL labeling is another possibility to partially deuterate the studied protein.⁵⁸ We have recently shown that quantitative observation of NOE build-ups is also possible in perdeuterated samples.^{1,37} The Kainosho group applied this approach to across-disulfide bond NOEs between $\beta 2/3$ protons in an otherwise deuterated sample.⁵⁹ It appears that spin diffusion pathways through residual protons cannot be neglected even in the highest deuterated samples available (98% deuteration).¹ Since the residual protonation level is both of a stochastic nature and also dependent on the synthesis chemistry and the biology of the expressing host, the distribution of the residual protonation is highly complex. In principle, an individual calculation of each possible isotopomer following eq 1 (and summation thereafter) is required. Accounting for nonuniform deuterium levels at individual atom positions by a simple modification of the relaxation matrix appears not to be feasible. However, the problem can be significantly simplified if the calculation of the overall contribution is divided into individual contributions from three-spin systems (that is, the two protons involved in the NOE plus one spin diffusion partner). In this analysis, as implemented in the DOMINO software,¹ the impact of the deuteration levels can be accounted for by scaling the according

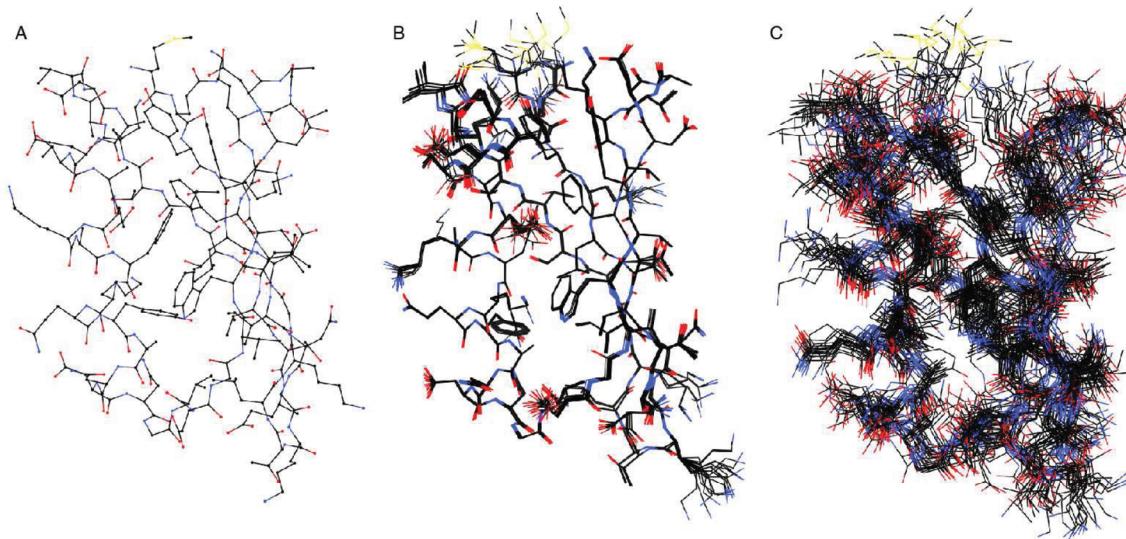


Figure 4. Structures of GB3 (A) from X-ray crystallography, (B) from NMR structure calculations based on eNOEs, and (C) from standard NMR structure calculations. Bundles in panels B and C are the 20 best conformers with the lowest target functions. Black lines represent bonds with an involvement of carbon atoms, red lines with oxygen atoms, blue with nitrogen atoms, and yellow lines with sulfur atoms. Heavy atom RMSD to the mean is 0.7 Å and 1.3 Å for panels B and C, respectively.

contribution by the residual protonation level (without restriction from nonuniformity in the deuteration level). On the other hand, this approach is not exact, as demonstrated by the corresponding comparison with protonated samples discussed above (Figure 3 and Figure S2), but reproduces the rates reasonably well for short mixing times. Following the evaluations above, it is recommended that in the case of perdeuterated GB3, having a rotational correlation time of 4.15 ns, a maximal mixing time of ca. 120 ms should be chosen. For molecules with different tumbling times, it is recommended to maintain a constant product of the mixing and tumbling times of ca. 5×10^{-10} s², while for protonated samples, the value is approximately 2.5×10^{-10} s². These suggestions are based on the approximation of the linear regime for which the cross peak volume in the NOESY is proportional to the product of the NOE rate and the rotational correlation time, while the spin diffusion contribution is proportional to the square of this product.

Full Relaxation Matrix Approach Only for the Spin Diffusion Correction. In the presented eNORA software, the full relaxation matrix approach is only used to correct the NOE for spin diffusion contribution. Methods to directly refine the structure against the NOESY intensities have been developed,^{28,35,60,61} but the refinement algorithm is completely incorporated in the structure calculation software, rendering the method less versatile. Alternatively, as described in the Introduction, “hybrid” back-transformation methods combining experimental and simulated NOEs derived from a 3D structure together with an iterative procedure have been developed.^{26,32} A detailed comparison showed that this approach and the one presented here yield very similar eNOE rates for perdeuterated ubiquitin (Alexander Sobol, Beat Vögeli, Roland Riek, unpublished result) but that the application implemented here in the eNORA software is more practical. In particular, the assessment of the quality of the build-ups and its fits appear to be solid, while the output from reverse transformation of the full relaxation matrix is a list of cross-relaxation rates per mixing time, which may be difficult to judge in their quality, as already mentioned by Brüschweiler and Case.⁶¹ In addition, the

“hybrid” full relaxation matrix approach appears to require a scaling between experimental and theoretical values in order to account for their different origin.^{22,23,25,26,34,35} In return, the presented approach does rely more heavily on the input structure since experimental values are not used for the calculation of the spin diffusion. If however, the spin diffusion contribution is small, which can be established by the use of short mixing times (following the suggestions given in the previous paragraph), the input 3D structure is only used for a small correction of the NOE build-up. Furthermore, the spin diffusion corrections applied to the experimental data can be monitored. Finally, the eNOE-derived structure can be used for the correction of spin diffusion, and this iteration (depicted in Figure 1 as dashed line) will also diminish the influence of a potentially, partially wrong input starting structure.

From the NOE Intensities to the 3D Structure. The eNORA software translates the measured NOE intensities into distance restraints as described in the Methods and Theory section (Figure 1) and Supporting Information Figure S1. This software integrates well into the current structure calculation workflow having interfaces with spectra analysis software packages such as nmrPipe⁴² and structure calculation software packages such as CYANA.⁴¹ To demonstrate its functionality, a 3D structure of GB3, a 56 amino acids residue protein, was determined with the workflow of Figure 1 using experimental data from Vogeli et al.² A total of 1282 NOE build-ups could be used from the spectra² and taken as input for the eNORA software. Only 1143 were suitable according to the quality criteria implemented in the eNORA software, as mentioned above, resulting in 781 lower and upper distance restraints (Figure 2). Diagonal peaks were cured, as described in the section Deriving the Relaxation Rates, if they had an R^2 factor below 80% with the average autorelaxation rate calculated from the fits exhibiting $R^2 > 90\%$. NOE cross peak build-ups were considered if $R^2 > 80\%$ and if the normalizing diagonal peak decays have an $R^2 > 80\%$. Out of the 182 diagonal peak decays, 132 and 164 have an $R^2 > 90\%$ and $R^2 > 80\%$, respectively. Five diagonal peak decays have an $R^2 < 60\%$. A total of 1050 NOE build-up curves have an $R^2 > 90\%$. Similar results can be

obtained using Q as a criterion. For spin diffusion corrections, the RDC refined X-ray structure was used (PDB code: 1IGD).^{38–40} Spin diffusion was simulated with protons up to 6 Å away from either of the two protons involved in the NOE of interest. Protons further away do not participate significantly in spin diffusion (data not shown). The simulations have been completed in 8 min on a Macbook Pro with 2 GHz dual processors. Saving the figures of the fits took an additional 20 min (optional). The eNORA software generated both a list of 781 upper distance restraints and a list with 781 lower distance restraints for the input of the structure calculation by CYANA in accordance with the description in the Methods and Theory section. Using this data set in addition to a small set of angle restraints from secondary chemical shifts,² a standard structure calculation protocol performed by CYANA was used to determine the GB3 structure represented by a bundle of 20 conformers, as shown in Figure 4. The high structural accuracy of the eNOEs obtained from this approach is reflected in the heavy atom RMSD to the X-ray structure of 1.3 Å (435 atoms) and the backbone RMSD to the X-ray structure of 0.7 Å (168 atoms), respectively. The precision of the structure is reflected by the low average backbone RMSD to the mean of 0.2 Å and the low heavy atom RMSD to the mean of 0.7 Å. The eNOE-derived 3D structure is also compared with a conventional structure determination using 1956 upper distance restraints collected from the NOESY spectra recorded at the longest mixing time (100 ms) and the angle restraints from secondary chemical shifts (see above, Figure 4). The structure calculation was performed again with the CYANA software, yielding a bundle of 20 conformers with the lowest target functions. This bundle has a heavy atom RMSD to the mean of 1.5 Å to the X-ray structure and a close average resemblance to the eNOE-derived structural ensemble. A significant difference is that the bundle is considerably less well-defined (RMSD to the mean of 1.3 Å versus 0.7 Å, Figure 4). Although the comparison between the eNOE-derived GB3 structure with the X-ray structure and the conventionally calculated NMR structure strongly indicates that the eNORA software works well, it must be mentioned that there are 116 distance restraint violations in the eNOE-derived ensemble. They are all below 1 Å; 16 of them are above 0.5 Å and 30 within 0.3–0.5 Å. Since the NOE is a time- and ensemble-averaged probe and the estimated accuracy of the eNOE-derived distances is on the order of 0.1 Å or better,¹ these violations are not attributed to miss-assignments or errors in the data but rather reflect the dynamic behavior of the structure studied. The ensemble and time-averaged nature of the (e)NOEs request therefore the establishment of an ensemble-based structure calculation, which is the focus of another study.²

CONCLUSION

Structure determination of biomolecules by solution state NMR was established 30 years ago and mainly uses the distance dependence of the NOE cross peak intensity in the NOESY spectra semiquantitatively. The recent introduction of eNOEs opens a new avenue for the description of the structure and dynamics of biomolecules including ensemble-based structure calculations.¹ In the present work, software is presented that integrates well between the spectra analysis software and the structure calculation software in order to simplify the use of eNOEs for (ensemble-based) structure determination. This software also includes a full relaxation matrix approach to

calculate correctly the contribution of spin diffusion to the cross peak intensity.

ASSOCIATED CONTENT

Supporting Information

Description of the formats of the input and output files of the eNORA software and an additional graphic comparing the two methods (three-spin systems versus the full matrix methods) for calculating the spin diffusion are included in the Supporting Information. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: roland.riek@phys.chem.ethz.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Federation of the European Biochemical Society (FEBS long-term fellowship to Julien Orts) and the Eidgenössische Technische Hochschule in Zürich, who supported this work.

DEDICATION

This work is dedicated to Prof. Dr. Wilfred van Gunsteren on the occasion of his 65th birthday.

REFERENCES

- (1) Vogeli, B.; Segawa, T. F.; Leitz, D.; Sobol, A.; Choutko, A.; Trzesniak, D.; van Gunsteren, W.; Riek, R. Exact Distances and Internal Dynamics of Perdeuterated Ubiquitin from NOE Buildups. *J. Am. Chem. Soc.* **2009**, *131* (47), 17215–17225.
- (2) Vögeli, B.; Kazemi, S.; Güntert, P.; Riek, R. *Submitted 2012*.
- (3) Kaptein, R.; Boelens, R.; Scheek, R. M.; Vangunsteren, W. F. Protein Structures from NMR. *Biochemistry* **1988**, *27* (15), 5389–5395.
- (4) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
- (5) Wuthrich, K. Protein-Structure Determination in Solution by Nuclear Magnetic-Resonance Spectroscopy. *Science* **1989**, *243* (4887), 45–50.
- (6) Clore, G. M.; Gronenborn, A. M. Determination of 3-Dimensional Structures of Proteins in Solution by Nuclear-Magnetic-Resonance Spectroscopy. *Protein Eng.* **1987**, *1* (4), 275–288.
- (7) Wright, P. E. What Can Two-Dimensional Nmr Tell Us About Proteins. *Trends Biochem. Sci.* **1989**, *14* (7), 255–260.
- (8) Kumar, A.; Wagner, G.; Ernst, R. R.; Wuthrich, K. Buildup Rates of the Nuclear Overhauser Effect Measured by Two-Dimensional Proton Magnetic-Resonance Spectroscopy - Implications for Studies of Protein Conformation. *J. Am. Chem. Soc.* **1981**, *103* (13), 3654–3658.
- (9) Hare, D.; Shapiro, L.; Patel, D. J. Extrahelical Adenosine Stacks into Right-Handed DNA - Solution Conformation of the D(C-G-C-A-G-a-G-C-T-C-G-C-G) Duplex Deduced from Distance Geometry Analysis of Nuclear Overhauser Effect Spectra. *Biochemistry* **1986**, *25* (23), 7456–7464.
- (10) Kline, A. D.; Braun, W.; Wuthrich, K. Determination of the Complete 3-Dimensional Structure of the Alpha-Amylase Inhibitor Tendamistat in Aqueous-Solution by Nuclear Magnetic-Resonance and Distance Geometry. *J. Mol. Biol.* **1988**, *204* (3), 675–724.
- (11) Clore, G. M.; Gronenborn, A. M.; Brunger, A. T.; Karplus, M. Solution Conformation of a Heptadecapeptide Comprising the DNA-Binding Helix-F of the Cyclic-Amp Receptor Protein of Escherichia-Coli - Combined Use of H-1 Nuclear Magnetic-Resonance and Restrained Molecular-Dynamics. *J. Mol. Biol.* **1985**, *186* (2), 435–455.

- (12) Clore, G. M.; Gronenborn, A. M.; Kjaer, M.; Poulsen, F. M. The Determination of the 3-Dimensional Structure of Barley Serine Proteinase Inhibitor-2 by Nuclear-Magnetic-Resonance, Distance Geometry and Restrained Molecular-Dynamics. *Protein Eng.* **1987**, *1* (4), 305–311.
- (13) Macura, S.; Ernst, R. R. Elucidation of Cross Relaxation in Liquids by Two-Dimensional NMR-Spectroscopy. *Mol. Phys.* **1980**, *41* (1), 95–117.
- (14) Kalk, A.; Berendsen, H. J. C. Proton Magnetic-Relaxation and Spin Diffusion in Proteins. *J. Magn. Reson.* **1976**, *24* (3), 343–366.
- (15) Keepers, J. W.; James, T. L. A Theoretical-Study of Distance Determinations from NMR - Two-Dimensional Nuclear Overhauser Effect Spectra. *J. Magn. Reson.* **1984**, *57* (3), 404–426.
- (16) Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* **2001**, *21* (4), 349–359.
- (17) Stoesz, J. D.; Redfield, A. G.; Malinowski, D. Cross relaxation and spin diffusion effects on the proton NMR of biopolymers in H₂O: Solvent saturation and chemical exchange in superoxide dismutase. *FEBS Lett.* **1978**, *91* (2), 320–324.
- (18) Orts, J.; Tuma, J.; Reese, M.; Grimm, S. K.; Monecke, P.; Bartoschek, S.; Schiffer, A.; Wendt, K. U.; Griesinger, C.; Carlomagno, T. Crystallography-Independent Determination of Ligand Binding Modes. *Angew. Chem., Int. Ed.* **2008**, *47* (40), 7736–7740.
- (19) Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, *123* (25), 6108–6117.
- (20) Orts, J.; Griesinger, C.; Carlomagno, T. The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method. *J. Magn. Reson.* **2009**, *200* (1), 64–73.
- (21) Clore, G. M.; Gronenborn, A. M. How accurately can interproton distances in macromolecules really be determined by full relaxation matrix analysis of nuclear overhauser enhancement data? *J. Magn. Reson.* **1989**, *84* (2), 398–409.
- (22) Edmondson, S. P. Noe R-Factors and Structural Refinement Using Firm, an Iterative Relaxation Matrix Program. *J. Magn. Reson.* **1992**, *98* (2), 283–298.
- (23) Leeflang, B.; Kroon-Batenburg, L. CROSREL: Full relaxation matrix analysis for NOESY and ROESY NMR spectroscopy. *J. Biomol. NMR* **1992**, *2* (5), 495–518.
- (24) Borgias, B. A.; Gochin, M.; Kerwood, D. J.; James, T. L. Relaxation Matrix Analysis of 2d-NMR Data. *Prog. Nucl. Magn. Reson. Spectrosc.* **1990**, *22*, 83–100.
- (25) Borgias, B. A.; James, T. L. Mardigras - a Procedure for Matrix Analysis of Relaxation for Discerning Geometry of an Aqueous Structure. *J. Magn. Reson.* **1990**, *87* (3), 475–487.
- (26) Boelens, R.; Koning, T. M. G.; Kaptein, R. Determination of Biomolecular Structures from Proton-Proton Noes Using a Relaxation Matrix Approach. *J. Mol. Struct.* **1988**, *173*, 299–311.
- (27) Marion, D.; Genest, M.; Ptak, M. Reconstruction of Noes Maps - a Requirement for a Reliable Conformational-Analysis of Biomolecules Using 2d NMR. *Biophys. Chem.* **1987**, *28* (3), 235–244.
- (28) Yip, P.; Case, D. A. A New Method for Refinement of Macromolecular Structures Based on Nuclear Overhauser Effect Spectra. *J. Magn. Reson.* **1989**, *83* (3), 643–648.
- (29) Post, C. B.; Meadows, R. P.; Gorenstein, D. G. On the evaluation of interproton distances for three-dimensional structure determination by NMR using a relaxation rate matrix analysis. *J. Am. Chem. Soc.* **1990**, *112* (19), 6796–6803.
- (30) Solomon, I. Relaxation Processes in a System of 2 Spins. *Phys. Rev.* **1955**, *99* (2), 559–565.
- (31) Olejniczak, E. T.; Gampe, R. T.; Fesik, S. W. Accounting for Spin Diffusion in the Analysis of 2d Noe Data. *J. Magn. Reson.* **1986**, *67* (1), 28–41.
- (32) Nikonowicz, E.; Meadows, R.; Gorenstein, D. G. A Hybrid Complete Relaxation Matrix Structural Refinement of an Extra-Helical Adenosine Tridecamer D(Cgcagaattcgcg)2 from 2d H-1-NMR. *Bull. Magn. Reson.* **1989**, *11* (3/4), 226–229.
- (33) Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M. Correction of spin diffusion during iterative automated NOE assignment. *J. Magn. Reson.* **2004**, *167* (2), 334–42.
- (34) Boelens, R.; Koning, T. M. G.; Vandermarel, G. A.; Vanboom, J. H.; Kaptein, R. Iterative Procedure for Structure Determination from Proton Proton Noes Using a Full Relaxation Matrix Approach - Application to a DNA Octamer. *J. Magn. Reson.* **1989**, *82* (2), 290–308.
- (35) Nilges, M.; Habazettl, J.; Brunger, A. T.; Holak, T. A. Relaxation Matrix Refinement of the Solution Structure of Squash Trypsin-Inhibitor. *J. Mol. Biol.* **1991**, *219* (3), 499–510.
- (36) Nikonowicz, E. P.; Meadows, R. P.; Gorenstein, D. G. NMR Structural Refinement of an Extrahelical Adenosine Tridecamer D(Cgcagaattcgcg)2 Via a Hybrid Relaxation Matrix Procedure. *Biochemistry* **1990**, *29* (17), 4193–4204.
- (37) Vogeli, B.; Friedmann, M.; Leitz, D.; Sobol, A.; Riek, R. Quantitative determination of NOE rates in perdeuterated and protonated proteins: Practical and theoretical aspects. *J. Magn. Reson.* **2010**, *204* (2), 290–302.
- (38) Derrick, J. P.; Wigley, D. B. The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. *J. Mol. Biol.* **1994**, *243* (5), 906–18.
- (39) Yao, L.; Vogeli, B.; Torchia, D. A.; Bax, A. Simultaneous NMR study of protein structure and dynamics using conservative mutagenesis. *J. Phys. Chem. B* **2008**, *112* (19), 6045–6056.
- (40) Yao, L. S.; Vogeli, B.; Ying, J. F.; Bax, A. NMR Determination of Amide N-H Equilibrium Bond Length from Concerted Dipolar Coupling Measurements. *J. Am. Chem. Soc.* **2008**, *130* (49), 16518–+.
- (41) Guntert, P.; Mummenthaler, C.; Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **1997**, *273* (1), 283–298.
- (42) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277–293.
- (43) MATLAB, 7.7; The MathWorks Inc.: Natick, MA, 2008.
- (44) Brünger, A. T. *X-PLOR Version 2.1, User Manual*; Yale University: New Haven, CT, 1990.
- (45) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **2007**, *23* (3), 381–382.
- (46) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Skelton, N. J. *Protein NMR Spectroscopy, Principles and Practice*; Academic Press: San Diego, CA, 2007.
- (47) Goddard, T. D.; Kneller, D. G. SPARKY 3; University of California: San Francisco, CA, 2008.
- (48) Bartels, C.; Xia, T. H.; Billeter, M.; Guntert, P.; Wuthrich, K. The Program Xeasy for Computer-Supported NMR Spectral-Analysis of Biological Macromolecules. *J. Biomol. NMR* **1995**, *6* (1), 1–10.
- (49) Herrmann, T.; Guntert, P.; Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **2002**, *319* (1), 209–227.
- (50) 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 Internal Motions. *J. Magn. Reson.* **1990**, *90* (1), 111–123.
- (51) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
- (52) Lefevre, J. F.; Lane, A. N.; Jarretzky, O. Solution Structure of the Trp Operator of Escherichia-Coli Determined by NMR. *Biochemistry* **1987**, *26* (16), 5076–5090.

- (53) Bonvin, A. M. J. J.; Brunger, A. T. Do NOE distances contain enough information to assess the relative populations of multi-conformer structures? *J. Biomol. NMR* **1996**, *7* (1), 72–76.
- (54) Bonvin, A. M. J. J.; Rullmann, J. A. C.; Lamerichs, R. M. J. N.; Boelens, R.; Kaptein, R. “Ensemble” iterative relaxation matrix approach: A new NMR refinement protocol applied to the solution structure of crambin. *Proteins: Struct., Funct., Bioinf.* **1993**, *15* (4), 385–400.
- (55) Higham, N. J. The scaling and squaring method for the matrix exponential revisited. *Siam J. Mat. Anal. App.* **2005**, *26* (4), 1179–1193.
- (56) Religa, T. L.; Kay, L. E. Optimal methyl labeling for studies of supra-molecular systems. *J. Biomol. NMR* **2010**, *47* (3), 163–169.
- (57) Orts, J.; Grimm, S. K.; Griesinger, C.; Wendt, K. U.; Bartoschek, S.; Carlonmagno, T. Specific methyl group protonation for the measurement of pharmacophore-specific interligand NOE interactions. *Chem.—Eur. J.* **2008**, *14* (25), 7517–7520.
- (58) Kainosh, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Ono, A. M.; Guntert, P. Optimal isotope labelling for NMR protein structure determinations. *Nature* **2006**, *440* (7080), 52–57.
- (59) Takeda, M.; Terauchi, T.; Kainosh, M. Conformational analysis by quantitative NOE measurements of the β -proton pairs across individual disulfide bonds in proteins. *J. Biomol. NMR* **2012**, 1–13.
- (60) Borgias, B. A.; James, T. L. Comatose, a Method for Constrained Refinement of Macromolecular Structure Based on Two-Dimensional Nuclear Overhauser Effect Spectra. *J. Magn. Reson.* **1988**, *79* (3), 493–512.
- (61) Bruschweiler, R.; Case, D. A. Characterization of Biomolecular Structure and Dynamics by Nmr Cross-Relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26*, 27–58.