J Biomol NMR (2013) 57:211–218
DOI 10.1007/s10858-013-9780-4

ARTICLE

Stereospecific assignments in proteins using exact NOEs

Julien Orts · Beat Vögeli · Roland Riek · Peter Güntert

Received: 17 July 2013 / Accepted: 4 September 2013 / Published online: 18 October 2013

Abstract Recently developed methods to measure distances in proteins with high accuracy by "exact" nuclear Overhauser effects (eNOEs) make it possible to determine stereospecific assignments, which are particularly important to fully exploit the accuracy of the eNOE distance measurements. Stereospecific assignments are determined by comparing the eNOE-derived distances to protein structure bundles calculated without stereospecific assignments, or an independently determined crystal structure. The absolute and relative CYANA target function difference upon swapping the stereospecific assignment of a diastereotopic group yields the respective stereospecific assignment. We applied the method to the eNOE data set that has recently been obtained for the third immunoglobulin-binding domain of protein G (GB3). The 884 eNOEs provide relevant data for 47 of the total of 75 diastereotopic groups. Stereospecific assignments could be established for 45 diastereotopic groups (96 %) using the X-ray structure, or for 27 diastereotopic groups (57 %) using structures calculated with the eNOE data set without stereospecific assignments, all of which are in agreement with those determined previously. The latter case is relevant for structure determinations based on eNOEs. The accuracy of the eNOE distance measurements is crucial for making stereospecific assignments because applying the same method to the traditional NOE data set for GB3 with imprecise upper distance bounds yields only 13 correct stereospecific assignments using the X-ray structure or 2 correct stereospecific assignments using NMR structures calculated without stereospecific assignments.

Keywords Stereospecific assignment · NOE · Distance restraint · Protein structure · CYANA

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9780-4) contains supplementary material, which is available to authorized users.

J. Orts · B. Vögeli · R. Riek (✉)
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, 8093 Zurich, Switzerland
e-mail: roland.riek@phys.chem.ethz.ch

P. Güntert (✉)
Center for Biomolecular Magnetic Resonance, Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
e-mail: guentert@em.uni-frankfurt.de

P. Güntert
Frankfurt Institute for Advanced Studies, Goethe University Frankfurt am Main, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

P. Güntert
Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, Tokyo 192-0397, Japan

Introduction

Nuclear Overhauser effect (NOE) measurements yield the most important structural data that can be obtained from NMR with proteins. For the purpose of structure determination NOEs are traditionally interpreted in a conservative way as loose upper distance bounds. This approach, that has been used successfully to determine more than 8,500 protein solution structures, takes implicitly into account that proteins are dynamic molecules and that NOEs do not fulfill the independent spin pair approximation, in addition to the experimental difficulties to determine NOE rates with high accuracy. However, using imprecise upper

distance bounds entails a significant loss of information. We have recently shown that this loss of information can be avoided largely by a quantitative determination of NOE rates, resulting in "exact" NOEs (eNOEs) that can yield distances with accuracy better than 5 % (Vögeli et al. 2009, 2010). This made it possible to measure the temperature dependence of $^{1}H^{N}$–$^{1}H^{N}$ distances in ubiquitin (Leitz et al. 2011), and to elucidate motion in proteins by ensemble-based structure calculation on the basis of eNOEs (Orts et al. 2012; Vögeli et al. 2012, 2013).

Stereospecific assignments are important to fully exploit the potential of eNOEs, lest part of the increased accuracy be lost to account for the lack of stereospecific assignments. Stereospecific assignments are therefore more relevant in the context of eNOEs than with traditional upper distance bounds. Fortunately, the high accuracy of eNOEs also opens up new ways to determine stereospecific assignments, as will be shown in this work.

The standard NMR resonance assignment methods do not yield stereospecific assignments for diastereotopic groups. Thus, there have been a variety of studies on the impact of the presence, or absence, of stereospecific assignments on NMR structure determinations of proteins (Driscoll et al. 1989; Fletcher et al. 1996; Güntert 1998; Güntert et al. 1989; Havel 1991), and a variety of methods for determining stereospecific assignments, mostly from the early 1990s, including approaches based on stereospecific isotope labeling (Kainosho and Güntert 2009; Kainosho et al. 2006; Neri et al. 1989; Plevin et al. 2011; Senn et al. 1989), and computational algorithms based on systematic searches of the local conformation space (Güntert et al. 1989; Hyberts et al. 1987; Nilges et al. 1990; Polshakov et al. 1995; Tejero et al. 1999) or analyses of preliminary three-dimensional structures (Beckman et al. 1993; Folmer et al. 1997; Güntert et al. 1991a, b; Pristovšek and Franzoni 2006; Weber et al. 1988). Stereospecific assignment methods based on isotope labeling are reliable and are widely used for the methyl groups of valine and leucine (Senn et al. 1989), for which stereospecific assignments have the largest impact on the structure. The computational methods, on the other hand, have a certain potential for errors, especially when internal dynamics is present (Folmer et al. 1997; Havel 1991). For this reason, and because methods have been developed that reduce the loss of structural information in the absence of stereospecific assignments (Fletcher et al. 1996), the use of computational approaches for determining stereospecific assignments has decreased during the last decade.

In this paper we introduce a computational method based on eNOEs that can provide stereospecific assignments for a large number of methylene and isopropyl methyl groups in a straightforward and reliable way without need for additional experiments.

Materials and methods

NMR measurement and evaluation of eNOEs

NMR measurements of the protein GB3 were performed with 350 µl of a 4 mM uniformly ^{13}C,^{15}N-labeled protein solution in 97 % H_2O, 3 % D_2O, 50 mM potassium phosphate buffer, pH 6.5, and 0.5 mg/ml sodium azide on a Bruker 700 MHz spectrometer equipped with a triple resonance cryoprobe at 298 K. A series of 3D ^{15}N- or ^{13}C-resolved [^{1}H,^{1}H]-NOESY spectra with mixing times $\tau_m = 20$, 30, 40, 50, and 60 ms was recorded for the measurement of NOE buildups. Cross-relaxation rates were extracted following the previously established protocol (Vögeli et al. 2010). Details and experimental data have been presented elsewhere (Vögeli et al. 2013).

A total of 823 distances were measured based on eNOEs. Of these, 324 were obtained from two pathways (two symmetrically related peaks in the spectrum) and were used as exact distance restraints (upper bound = lower bound), 481 were obtained from one pathway and were used with ±15 % distance error, and 18 were between two methyl groups and were used with ±20 % distance error (Vögeli et al. 2010). In addition, there were 61 NOEs with aromatics that were used conventionally with an upper distance bound of 8 Å. For comparison, NOEs were also interpreted in the traditional, semi-quantitative way, yielding 1,956 upper distance bounds. Of these, 1,041 were non-redundant conformation-restricting restraints. In addition to the NOE distance restraints, the NMR data for GB3 comprised, in both cases, also 54 torsion angle restraints obtained from $^{13}C^{\alpha}$ chemical shifts, 147 $^{3}J_{HNH\alpha}$, $^{3}J_{HNC'}$, and $^{3}J_{HNC\beta}$ scalar coupling restraints, as well as 90 ^{15}N–$^{1}H^{N}$ and ^{13}C–$^{1}H^{\alpha}$ residual dipolar coupling (RDC) restraints (Vögeli et al. 2012). The stereospecific assignments for βCH_2 of amino acid residues 3, 5, 8, 22, 30, 35, 36, 37, 40, 43, 45, 46, 47, 52 and 54 were also confirmed independently using a set of scalar couplings and RDCs reported in the literature (Lian et al. 1992; Miclet et al. 2005; Vögeli et al. 2013).

Stereospecific assignment based on eNOEs

The original eNOE restraints are given arbitrary stereospecific assignments. Stereospecific assignments are determined by comparing the eNOE-derived distance restraints to structures that were calculated with the program CYANA in the absence of any stereospecific assignments using the same eNOE data, and possibly other conformational restraints such as torsion angle restraints, residual dipolar couplings, etc. The absence of stereospecific assignments is handled by symmetrizing the restraint list (Güntert et al. 1991a, b, using the CYANA command 'distances modify'.

In short, a pair of distance restraints $d(A, B_1) < u_1$ and $d(A, B_2) < u_2$ from an atom A to the two atoms B_1 and B_2 of a diastereotopic group are replaced by (in general) three restraints that are invariant under exchange of the stereospecific assignment, i.e. a restraint to a pseudoatom Q located centrally with respect to the positions of atoms B_1 and B_2, $d(A, Q) < u_Q$, and two restraints with identical upper bound $u = \max(u_1, u_2)$ for the individual distances, $d(A, B_1) < u$ and $d(A, B_2) < u$ (Güntert et al. 1991a). Structures are calculated using the standard torsion angle dynamics simulated annealing protocol of the program CYANA. Starting from 250 conformers with random torsion angles, 25,000 torsion angle dynamics steps were applied per conformer, and the 50 conformers with lowest final target function values were selected for analysis. Structures obtained in this way are strictly independent of the arbitrary stereospecific assignments assumed in the input restraints.

The following algorithm can also be applied to structures obtained in other ways, for example to an X-ray crystal structure. In this paper, this was done with the RDC-refined X-ray structure of GB3 (Derrick and Wigley 1994; Yao et al. 2008).

The algorithm then calculates for each diastereotopic group the weighted target function difference upon exchanging its stereospecific assignment, $\Delta f = (f_R - f_I) \times |f_R - f_I|/\max(f_I, f_R)$, where f_I and f_R are the CYANA target function values with the stereospecific assignment of the group under consideration as in the input or reversed, respectively, calculated only for the distance restraints that involve the given diastereotopic group. This definition of the weighted target function difference upon exchanging its stereospecific assignment, Δf, captures the idea that a stereospecific assignment should be safer when the target function difference $f_R - f_I$ is higher, and that the difference is the more significant when the relative difference between the two target function values is larger. Therefore we combine these two concepts into a single formula by multiplying them. For instance, the stereospecific assignment of a diastereotopic group with $f_I = 0$ Å^2 and $f_R = 2$ Å^2, yielding $\Delta f = 2$ Å^2, is considered more significant than one with $f_I = 8$ Å^2 and $f_R = 10$ Å^2, yielding $\Delta f = 0.4$ Å^2. Since the stereospecific assignment of one diastereotopic group can in principle have an influence on the target function values f_I and f_R of another diastereotopic group, the optimal swapping of the entire set of all stereospecific assignments is iterated multiple times until no further change occurs for any of the diastereotopic groups. If there are multiple structures, the minimal absolute value of Δf over the ensemble is taken, and the maximal fraction q of conformers with either $\Delta f \geq 0$ (i.e., the input stereospecific assignment is preferred) or $\Delta f < 0$ (the reversed stereospecific assignment is preferred) is computed. If the same stereospecific assignment yields consistently a lower target function value for all conformers of the structure ensemble, then $q = 1$, and in all cases $q \geq 0.5$, because we divide the set of target function difference values into two groups ($\Delta f \geq 0$ or $\Delta f < 0$). A stereospecific assignment is considered as reliable if Δf and q exceed given thresholds, $|\Delta f| \geq \Delta f_{\min}$ and $q \geq q_{\min}$.

The calculations of this paper were performed with $\Delta f_{\min} = 0.1$ Å^2 when using an input X-ray structure or $\Delta f_{\min} = 0.2$ Å^2 when using only the NMR data. In the latter case it was required that all conformers yielded a consistent stereospecific assignment, i.e. $q_{\min} = 1$.

The algorithm can be applied to the side-chain NH_2 groups of Asn and Gln in exactly the same way as to diastereotopic groups. In this paper, we therefore include these side-chain NH_2 groups among the diastereotopic groups.

Structure calculations with three-state ensemble-averaged restraints

Structure calculations were performed with the ensemble-based structure determination protocol using ensemble-averaged distance restraints obtained from eNOE rates, as described recently (Vögeli et al. 2012, 2013). CYANA structure calculations were started from 100 conformers with random torsion angle values, simulated annealing with 50,000 torsion angle dynamics steps was applied, and the 20 conformers with the lowest final target function values were analyzed. For the ensemble-averaged calculations 3 structural states of the entire protein were calculated simultaneously, excluding steric repulsion between atoms of different states, and applying the eNOE distance restraints to the $1/r^6$ averages of the corresponding distances in the individual states. The absence of stereospecific assignments was handled as described above. Similarly, the ^3J coupling restraints and the RDC restraints were applied to the arithmetic mean of the corresponding quantities in the individual states. Bundling restraints were applied in order to keep the individual structural states together in space as far as permitted by the experimental restraints. To this end weak upper distance bounds of 1.2 Å were imposed on all distances between the same nitrogen and carbon atoms in different states. The weight of these bundling restraints was 100 times lower than for NOE upper distance bounds, except for the backbone atoms N, C^α, C', and C^β, for which a 10 times lower weight than for NOEs was used.

Results and discussion

The protein GB3 contains 75 diastereotopic groups. The eNOE distance restraint set (Vögeli et al. 2012) provides distance restraints for 47 diastereotopic groups (Table 1).

Table 1 Stereospecific assignments for GB3 using eNOEs

Diastereotopic group	Accessibility (%)	Distance restraints	Using X-ray structure		Using only NMR data		
			Δf (Å^2)	Assignment	Δf (Å^2)	q (%)	Assignment
Gln2 βCH2	41	22	0.44	Correct	−0.01	98	Ambiguous
Tyr3 βCH2	7	22	2.24	Correct	1.95	100	Correct
Lys4 βCH2	28	26	8.42	Correct	7.65	100	Correct
Lys4 γCH2	28	14	0.53	Correct	0.04	84	Ambiguous
Leu5 βCH2	0	15	0.96	Correct	0.98	100	Correct
Leu5 δ(CH3)2	0	42	44.91	Correct	45.38	100	Correct
Ile7 γ1CH2	1	26	2.18	Correct	3.79	100	Correct
Asn8 βCH2	26	30	7.69	Correct	0.55	100	Correct
Asn8 δ2NH2	26	6	0.73	Correct	0.77	100	Correct
Gly9 αCH2	3	34	15.91	Correct	9.92	100	Correct
Leu12 βCH2	22	36	2.22	Correct	0.30	100	Correct
Leu12 δ(CH3)2	22	18	5.26	Correct	0.07	84	Ambiguous
Lys13 βCH2	48	22	2.24	Correct	1.75	100	Correct
Lys13 γCH2	48	6	−0.04	Ambiguous	−0.11	80	Ambiguous
Lys13 δCH2	48	10	1.86	Correct	0.00	58	Ambiguous
Gly14 αCH2	19	34	0.14	Correct	0.04	100	Ambiguous
Glu15 βCH2	38	24	6.20	Correct	2.46	100	Correct
Lys19 γCH2	52	8	0.80	Correct	0.13	80	Ambiguous
Asp22 βCH2	23	2	0.11	Correct	0.00	74	Ambiguous
Glu24 βCH2	38	8	1.07	Correct	−0.26	96	Ambiguous
Glu27 βCH2	21	3	1.01	Correct	2.06	100	Correct
Lys28 γCH2	42	6	−0.02	Ambiguous	0.04	60	Ambiguous
Phe30 βCH2	1	39	10.97	Correct	11.12	100	Correct
Lys31 βCH2	26	24	5.11	Correct	4.30	100	Correct
Lys31 γCH2	26	18	0.92	Correct	1.19	100	Correct
Lys31 δCH2	26	10	1.63	Correct	0.00	74	Ambiguous
Gln32 γCH2	34	12	0.66	Correct	−0.06	74	Ambiguous
Gln32 ε2NH2	34	10	3.51	Correct	0.55	100	Correct
Asn35 βCH2	40	12	2.42	Correct	2.60	100	Correct
Asn35 δ2NH2	40	10	1.36	Correct	1.34	100	Correct
Asp36 βCH2	43	16	1.61	Correct	3.16	100	Correct
Asn37 βCH2	27	18	0.13	Correct	0.01	100	Ambiguous
Asn37 δ2NH2	27	24	11.20	Correct	10.21	100	Correct
Val39 γ(CH3)2	3	50	25.77	Correct	28.02	100	Correct
Asp40 βCH2	47	12	0.10	Correct	0.33	98	Ambiguous
Gly41 αCH2	9	26	2.11	Correct	6.50	100	Correct
Val42 γ(CH3)2	40	16	0.20	Correct	0.00	76	Ambiguous
Trp43 βCH2	16	22	5.25	Correct	5.39	100	Correct
Tyr45 βCH2	24	12	0.29	Correct	0.38	100	Correct
Asp46 βCH2	28	32	5.24	Correct	6.35	100	Correct
Asp47 βCH2	34	16	2.00	Correct	0.14	96	Ambiguous
Lys50 βCH2	18	12	0.32	Correct	−0.16	100	Ambiguous
Lys50 γCH2	18	8	1.17	Correct	0.00	56	Ambiguous
Lys50 δCH2	18	6	1.54	Correct	0.00	72	Ambiguous
Phe52 βCH2	6	18	1.66	Correct	1.39	100	Correct
Val54 γ(CH3)2	0	82	55.31	Correct	69.22	100	Correct

Table 1 continued

Diastereotopic group	Accessibility (%)	Distance restraints	Using X-ray structure		Using only NMR data		
			Δf (Å^2)	Assignment	Δf (Å^2)	q (%)	Assignment
Glu56 βCH2	22	16	0.32	Correct	0.25	74	Ambiguous

The solvent accessibility was calculated with the program MOLMOL (Koradi et al. 1996) for the entire residue using a solvent radius of 1.4 Å. Numbers of distance restraints refer to the distance restraints that involve atoms of the given diastereotopic group. Upper and lower distance bounds are counted separately and added. The table lists all diastereotopic groups for which the eNOE data set contains relevant distance restraints. Results obtained using the X-ray structure were determined by comparing the distance restraints in the eNOE data set to the (single) RDC refined X-ray crystal structure of GB3 (Derrick and Wigley 1994; Yao et al. 2008). Results obtained using only NMR data were determined by comparing the distance restraints in the eNOE data set to a bundle of 50 conformers calculated with CYANA with undefined stereospecific assignment. Δf denotes the weighted target function difference upon exchanging the stereospecific assignment. See "Materials and methods" for details. A negative sign indicates that the smallest absolute value of Δf is found with reversed stereospecific assignment. The maximal percentage q of conformers with either $\Delta f \geq 0$ (i.e., the input stereospecific assignment is preferred) or $\Delta f < 0$ (the reversed stereospecific assignment is preferred) is computed. Stereospecific assignments are indicated as 'correct' if $|\Delta f| \geq 0.1$ Å^2 when using the X-ray structure, or $|\Delta f| \geq 0.2$ Å^2 and $q = 100$ % when using only NMR data. Otherwise, they are classified as 'ambiguous'

No distance restraints are available for the remaining 28 diastereotopic groups.

Stereospecific assignments based on the RDC-refined X-ray structure

Using the RDC-refined X-ray structure (Derrick and Wigley 1994; Yao et al. 2008) and the eNOE data set for GB3 (Vögeli et al. 2012), our algorithm yielded stereospecific assignments for 45 out of the 47 diastereotopic groups for which the eNOE data set provided relevant information (Table 1). All 45 stereospecific assignments were in agreement with those reported earlier (Vögeli et al. 2012). There are significant differences with regard to the unambiguousness of the stereospecific assignments: 6 diastereotopic groups show a weighted target function difference upon exchanging the stereospecific assignment $\Delta f > 10$ Å^2, 24 have $1 < \Delta f \leq 10$ Å^2, and 15 have $0.1 < \Delta f \leq 1$ Å^2. Two diastereotopic groups have $\Delta f \leq 0.1$ Å^2, and are therefore not stereospecifically assigned. The stereospecific assignments with $\Delta f > 0.1$ Å^2 include 3 out of 3 Gly αCH_2 groups, 24 out of 24 βCH_2 groups, 6 out of 8 γCH_2 groups, 3 out of 3 δCH_2 groups, 5 out of 5 Val and Leu isopropyl $(CH_3)_2$ groups, and 4 out of 4 Asn and Gln side-chain NH_2 groups, for which eNOE data is available (Fig. 1). This shows that eNOEs in conjunction with a high-resolution X-ray structure of GB3 enable our algorithm to determine unambiguous stereospecific assignments for the large majority of diastereotopic methylene and isopropyl methyl groups, as well as for the planar side-chain amide groups of Asn and Gln.

It should be noted that a comparable result could not be achieved with traditional, semi-quantitative NOE distance restraints. Applying the same algorithm to the conventional NMR data set of 1,956 upper distance bounds (and no lower distance bounds), yielded correct stereospecific assignments only for 13 diastereotopic groups, instead of 45 when using eNOEs.

Stereospecific assignments based on NMR data alone

In the absence of an input (e.g. X-ray) structure, a bundle of 50 conformers was generated with CYANA using the NMR data set for GB3 (Vögeli et al. 2012) after "stereosymmetrization", as described in the "Materials and methods" section. A larger number than the usual 20 NMR conformers was generated to increase the statistical significance and thus the reliability of the stereospecific assignment. Applying the present stereospecific assignment algorithm to the eNOE data set with these 50 NMR conformers yielded stereospecific assignments for 27 out of the 47 diastereotopic groups with relevant experimental data. The stereospecific assignments with $\Delta f > 0.2$ Å^2 include 2 out of 3 Gly αCH_2 groups, 16 out of 24 βCH_2 groups, 2 out of 8 γCH_2 groups, 0 out of 3 δCH_2 groups (all Lys), 3 out of 5 Val and Leu isopropyl $(CH_3)_2$ groups, and 4 out of 4 Asn and Gln side-chain NH_2 groups, for which eNOE data is available (Fig. 1).

The choice of the cutoff value Δf for the weighted target function difference is to some extent arbitrary. The numbers of correct/wrong stereospecific assignments vary with increasing Δf values (and $q = 100$ %) as follows: $\Delta f = 0.0$ Å^2, 31 correct/2 wrong; $\Delta f = 0.1$ Å^2, 27 correct/2 wrong; $\Delta f = 0.2$ Å^2, 27 correct/0 wrong; $\Delta f = 0.3$ Å^2, 27 correct/0 wrong; $\Delta f = 0.4$ Å^2, 25 correct/0 wrong; $\Delta f = 0.5$ Å^2, 25 correct/0 wrong. We have chosen the Δf cutoff value as the lowest "round" number that excluded any erroneous stereospecific assignments. However, as the above numbers show, one could also choose significantly higher (safer) cutoffs without loosing a significant number of stereospecific assignments.

There is a correlation between the solvent accessibility of a residue and its stereospecific assignments. All 10 diastereotopic groups with eNOE data in buried residues with <10 % solvent accessibility could be assigned correctly, whereas the stereospecific assignments remained

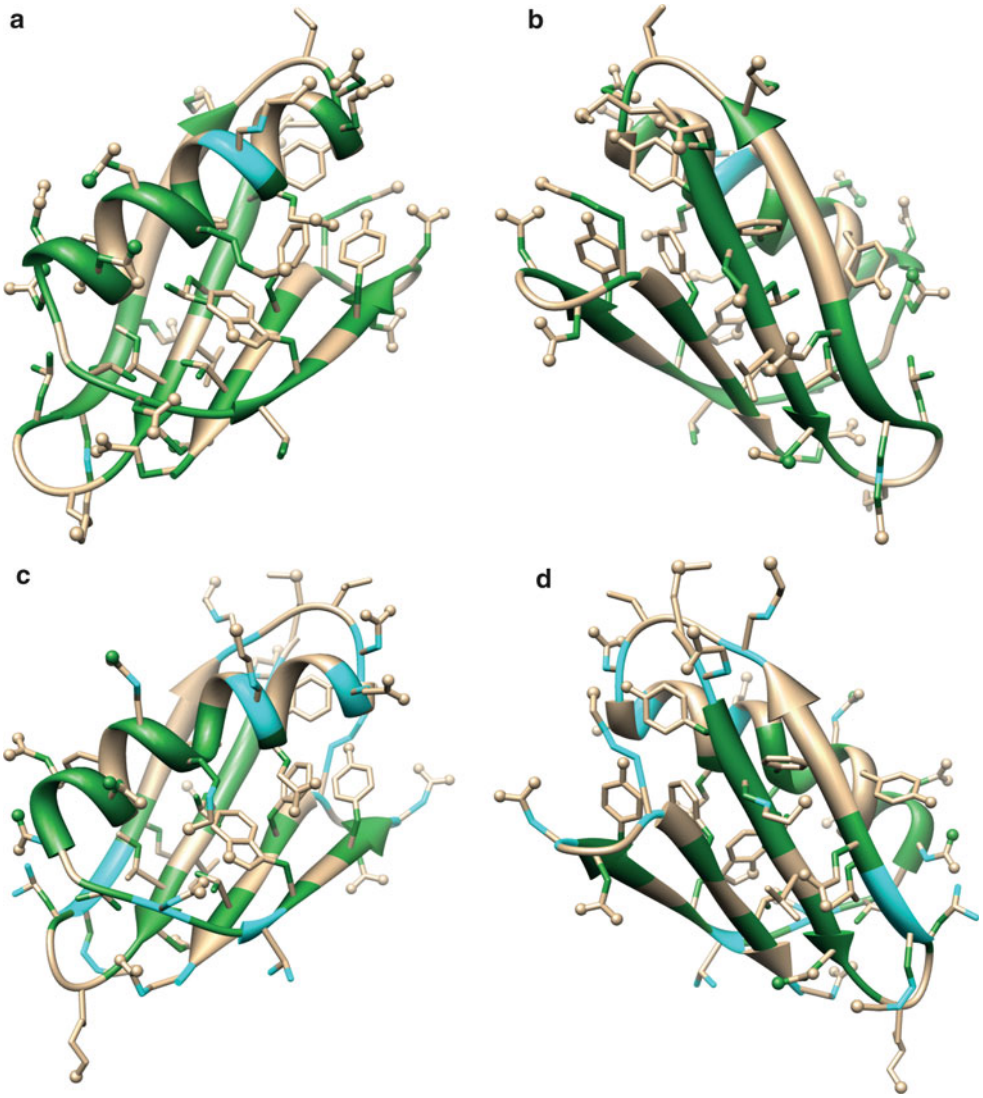

Fig. 1 Stereospecific assignments for GB3. Diastereotopic groups with correct and ambiguous stereospecific assignments are colored in *green* and *cyan*, respectively. The ribbon is colored in *green* or *cyan* if the majority of the stereospecific assignments of a residue is correct or ambiguous, respectively. **a** Stereospecific assignments determined using eNOEs and the X-ray structure, mapped onto the X-ray structure. **b** Same as **a**; structure rotated by 180° around a *vertical axis*. **c** Stereospecific assignments determined using eNOEs and the NMR structure bundle calculated without stereospecific assignments, mapped on the structure with the lowest target function value. **d** Same as **c**; structure rotated by 180° around a vertical axis. The 10 threonines and the 6 alanines are not shown, as they do not have diastereotopic groups. *Spheres* represent oxygen, nitrogen, or sulfur atoms

ambiguous for 7 out of the 11 diastereotopic groups with eNOE data in the highly solvent exposed residues with more than 40 % solvent accessibility. Most of the 20 ambiguous stereospecific assignments occur in charged or hydrophilic residues, i.e. 17 in Lys, Asp, Glu, but only 3 in other residues (Gly, Val, Leu).

Comparing the results obtained with the X-ray structure and on the basis of the NMR data alone, it is apparent that in the clear-cut cases of stereospecific assignments with high Δf values, the latter are very similar with X-ray and NMR structures. In general the X-ray structure yields slightly higher Δf values. On the other hand, there are 10 diastereotopic groups with $\Delta f > 0.5$ Å^2 when using the X-ray structure but insignificant $\Delta f < 0.1$ Å^2 when using the NMR structure. The eNOE restraints and the NMR structure calculated from them without any assumptions on the stereospecific assignments can thus serve to determine many but not all of the stereospecific assignments that are possible by knowledge of a high-resolution X-ray structure. The eNOE restraints provide significantly more stereospecific assignments than the set of conventional semi-quantitative upper distance limits, which yield only 2 reliable stereospecific assignments with $\Delta f > 0.2$ Å^2.

Using only the NMR data, the stereospecific assignments of 20 out of 47 relevant diastereotopic groups remain ambiguous. This may appear to be a significant number. However, it should be noted that of the total of 1,707 distance restraints in the eNOE data set, only 254 (15 %) involve atoms without stereospecific assignment. Thus, the present stereospecific assignment method serves well its principal purpose to enable the accurate interpretation of the large majority of the eNOEs.

This finding is corroborated by comparing the results of CYANA structure calculations of three-state ensembles

(Vögeli et al. 2012) using the 27 stereospecific assignments that can be made from NMR data alone with those obtained with complete stereospecific assignments (Fig. S1). Overall there is little difference between the two calculations that exhibit very similar heavy atom RMSDs to the mean coordinates of 0.85 and 0.87 Å, respectively. Also the summaries of PSVS Protein Structure Validation Suite (Bhattacharya et al. 2007) structure quality factors in Tables S1 and S2 show similar values for the two structures.

It is conceivable that the extent of stereospecific assignments could be increased by an iterative procedure that uses the stereospecific assignments determined by a first application of the algorithm to the structure obtained in the absence of any stereospecific assignments as input for the calculation of a new NMR structure bundle that incorporates the stereospecific assignments that have been made so far. The stereospecific assignment algorithm is then run again with this new NMR structure bundle as input, etc. We applied this approach for ten iterative cycles of NMR structure calculation and stereospecific assignment determination for GB3. The results showed that virtually no additional stereospecific assignments could be determined compared to the first, non-iterative cycle of the procedure, and that occasionally incorrect stereospecific assignments appeared in later cycles. This indicates that the stereospecific assignments of different diastereotopic groups are essentially independent from each other. We therefore conclude that it is sufficient and more reliable to run the structure calculation and the stereospecific assignment algorithm only once for a given eNOE data set.

Conclusions

In this paper we have presented an algorithm for the determination of stereospecific assignments on the basis of "exact" eNOEs. Application of the algorithm to the protein GB3 shows that a significant number of stereospecific assignments can be obtained, and that all these stereospecific assignments are in agreement with those determined earlier by other methods. The use of eNOEs is thereby essential, as corresponding calculations with traditional, semi-quantitative NOE distance restraints resulted in far less stereospecific assignments. The stereospecific assignment algorithm is automatic and fast, requiring less than 1 s of CPU time on a laptop computer for GB3.

The GB3 protein sample used for this study was of exceptionally good quality in terms of sample concentration and stability. It is possible that for more demanding proteins the eNOE analysis could not be carried out to the same degree of completeness as for GB3, resulting in a smaller number of unambiguous stereospecific assignments. Nevertheless, the approach presented in this paper will remain valid. We have initiated eNOE measurements of several other proteins, including cyclophilin A, for which results will be reported in the future.

Stereospecific assignments are of particular importance for the optimal use of eNOE data, for example to elucidate motions in proteins (Vögeli et al. 2012). Much of the accuracy of the eNOE-based distance measurements is otherwise lost by corrections that have to be made to account for the absence of the stereospecific assignments (Fletcher et al. 1996; Güntert 1998). This effect is illustrated in Fig. S2 by the distributions of the χ^1 torsion angle values in three-state ensembles of GB3 obtained from eNOEs with either no stereospecific assignments or complete stereospecific assignments. This figure clearly shows that stereospecific assignments for the βCH_2 groups lead in many cases to significantly narrower χ^1 distributions. The present stereospecific assignment method is therefore a crucial complement of the eNOE methodology (Orts et al. 2012; Vögeli et al. 2009, 2010, 2012, 2013).

Acknowledgments This work was financially supported by the Swiss National Science Foundation (Grant 140214 to B.V.), the Federation of the European Biochemical Societies (FEBS long-term fellowship to J.O.), and the Eidgenössische Technische Hochschule Zürich. P.G. gratefully acknowledges financial support by the Lichtenberg program of the Volkswagen Foundation, the Japan Society for the Promotion of Science (JSPS), and the Deutsche Forschungsgemeinschaft (DFG).

References

Beckman RA, Litwin S, Wand AJ (1993) Statistical strategy for stereospecific hydrogen NMR assignments: validation procedures for the floating prochirality method. J Biomol NMR 3:675–700

Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66:778–795

Derrick JP, Wigley DB (1994) The third IgG-binding domain from streptococcal proteinG: an analysis by X-ray crystallography of the structure alone and in a complex with Fab. J Mol Biol 243:906–918

Driscoll PC, Gronenborn AM, Clore GM (1989) The influence of stereospecific assignments on the determination of three-dimensional structures of proteins by nuclear magnetic resonance spectroscopy: application to the sea anemone protein BDS-I. FEBS Lett 243:223–233

Fletcher CM, Jones DNM, Diamond R, Neuhaus D (1996) Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. J Biomol NMR 8:292–310

Folmer RHA, Hilbers CW, Konings RNH, Nilges M (1997) Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data. J Biomol NMR 9:245–258

Güntert P (1998) Structure calculation of biological macromolecules from NMR data. Q Rev Biophys 31:145–237

Güntert P, Braun W, Billeter M, Wüthrich K (1989) Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. J Am Chem Soc 111:3997–4004

Güntert P, Braun W, Wüthrich K (1991a) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J Mol Biol 217:517–530

Güntert P, Qian YQ, Otting G, Müller M, Gehring W, Wüthrich K (1991b) Structure determination of the Antp(C39S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. J Mol Biol 217:531–540

Havel TF (1991) An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. Prog Biophys Mol Biol 56:43–78

Hyberts SG, Märki W, Wagner G (1987) Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin c. Eur J Biochem 164:625–635

Kainosho M, Güntert P (2009) SAIL—stereo-array isotope labeling. Q Rev Biophys 42:247–300

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 14:51–55

Leitz D, Vögeli B, Greenwald J, Riek R (2011) Temperature dependence of $^{1}H_{N}$-$^{1}H_{N}$ distances in ubiquitin as studied by exact measurements of NOEs. J Phys Chem B 115:7648–7660

Lian LY, Derrick JP, Sutcliffe MJ, Yang JC, Roberts GCK (1992) Determination of the solution structures of domains II and III of protein G from *Streptococcus* by ^{1}H nuclear magnetic resonance. J Mol Biol 228:1219–1234

Miclet E, Boisbouvier J, Bax A (2005) Measurement of eight scalar and dipolar couplings for methine-methylene pairs in proteins and nucleic acids. J Biomol NMR 31:201–216

Neri D, Szyperski T, Otting G, Senn H, Wüthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry 28:7510–7516

Nilges M, Clore GM, Gronenborn AM (1990) ^{1}H-NMR stereospecific assignments by conformational data-base searches. Biopolymers 29:813–822

Orts J, Vögeli B, Riek R (2012) Relaxation matrix analysis of spin diffusion for the NMR structure calculation with eNOEs. J Chem Theory Comput 8:3483–3492

Plevin MJ, Hamelin O, Boisbouvier J, Gans P (2011) A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl groups in proteins. J Biomol NMR 49:61–67

Polshakov VI, Frenkiel TA, Birdsall B, Soteriou A, Feeney J (1995) Determination of stereospecific assignments, torsion-angle constraints, and rotamer populations in proteins using the program AngleSearch. J Magn Reson B 108:31–43

Pristovšek P, Franzoni L (2006) Stereospecific assignments of protein NMR resonances based on the tertiary structure and 2D/3D NOE data. J Comput Chem 27:791–797

Senn H, Werner B, Messerle BA, Weber C, Traber R, Wüthrich K (1989) Stereospecific assignment of the methyl ^{1}H NMR lines of valine and leucine in polypeptides by nonrandom ^{13}C labelling. FEBS Lett 249:113–118

Tejero R, Monleon D, Celda B, Powers R, Montelione GT (1999) HYPER: a hierarchical algorithm for automatic determination of protein dihedral-angle constraints and stereospecific $C^{\beta}H_{2}$ resonance assignments from NMR data. J Biomol NMR 15: 251–264

Vögeli B, Segawa TF, Leitz D, Sobol A, Choutko A, Trzesniak D, van Gunsteren W, Riek R (2009) Exact distances and internal dynamics of perdeuterated ubiquitin from NOE buildups. J Am Chem Soc 131:17215–17225

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

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

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

Weber PL, Morrison R, Hare D (1988) Determining stereo-specific ^{1}H nuclear magnetic resonance assignments from distance geometry calculations. J Mol Biol 204:483–487

Yao L, Vögeli B, Torchia DA, Bax A (2008) Simultaneous NMR study of protein structure and dynamics using conservative mutagenesis. J Phys Chem B 112:6045–6056

Supplementary Material

Stereospecific assignments in proteins using exact NOEs

Julien Orts · Beat Vögeli · Roland Riek · Peter Güntert

Julien Orts · Beat Vögeli · Roland Riek (✉)
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, 8093 Zürich, Switzerland
e-mail: roland.riek@phys.chem.ethz.ch

Peter Güntert (✉)
Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, and Frankfurt Institute for Advanced Studies, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
e-mail: guentert@em.uni-frankfurt.de

Peter Güntert
Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, Tokyo 192-0397, Japan

Fig. S1 Heavy-atom structural representations of three-state ensembles of GB3 (Vögeli et al. 2012) obtained from eNOEs with (**a**) complete stereospecific assignments (red), or (**b**) 27 stereospecific assignments that can be determined from the NMR data (blue).

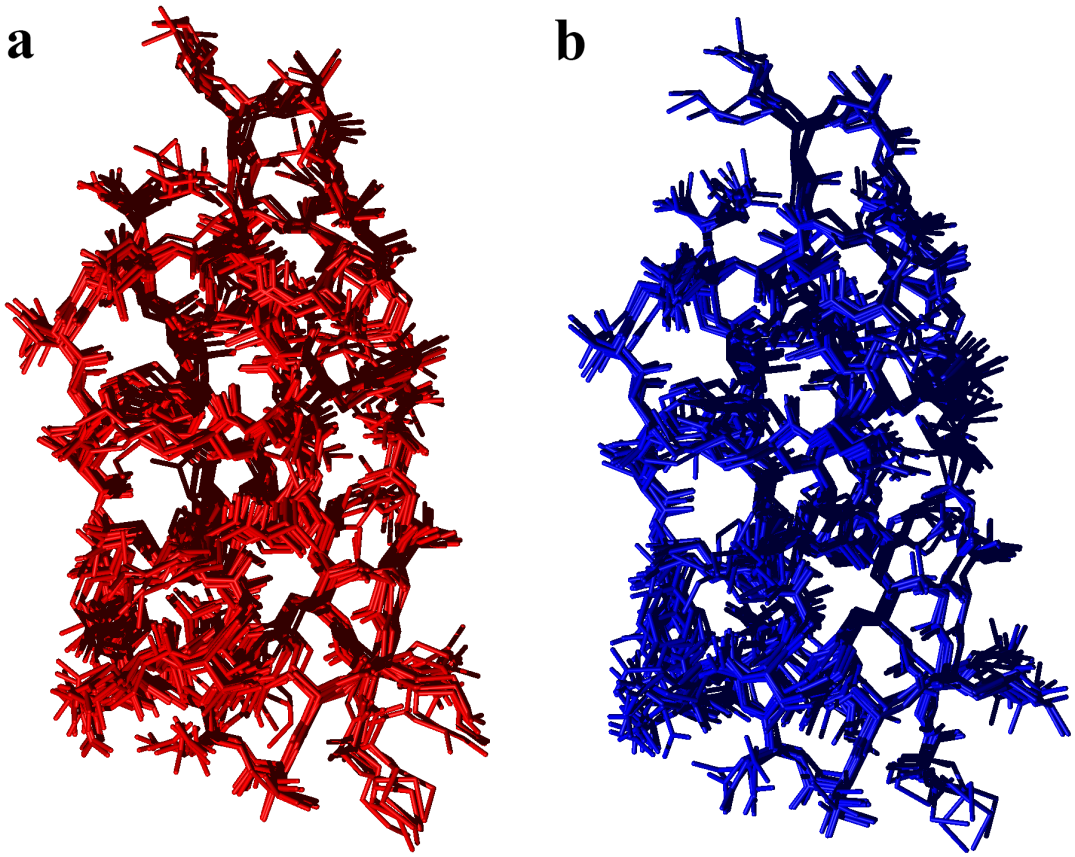

Fig. S2 Distribution of χ^1 torsion angle values for the 47 residues with eNOE data for the βCH_2 groups in three-state ensembles of GB3 obtained from eNOEs with (**a**) no stereospecific assignments, or (**b**) complete stereospecific assignments for these residues. Rectangles are drawn to groups values that are less than 60º separated from each other.

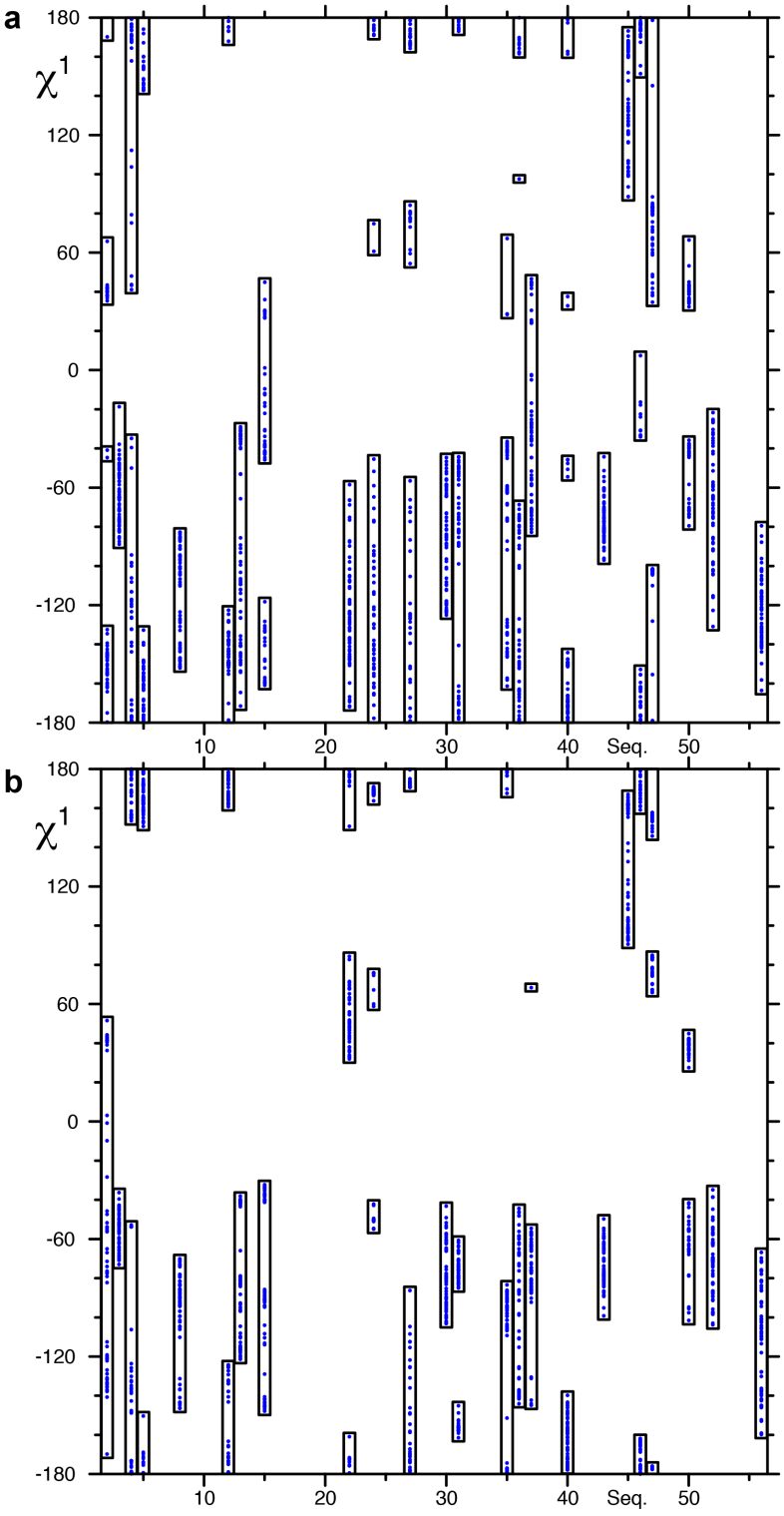

Table S1 Protein Structure Validation Suite (PSVS) summary of structure quality factors for three-state ensembles of GB3 obtained from eNOEs with complete stereospecific assignments (Fig. S1A).

Analyses performed for order residues.

Total structures computed	600		
Number of structures used	60		
RMSD Values			
	all	ordered[a]	Selected[b]
All backbone atoms	0.5 Å	0.5 Å	0.5 Å
All heavy atoms	0.9 Å	0.8 Å	0.8 Å
Structure Quality Factors - overall statistics			
	Mean score	SD	Z-score [c]
Procheck G-factor [a] (phi / psi only)	-0.46	N/A	-1.49
Procheck G-factor [a] (all dihedral angles)	-0.61	N/A	-3.61
Verify3D	0.42	0.0327	-0.64
ProsaII (-ve)	0.74	0.0692	0.37
MolProbity clashscore	16.76	3.6826	-1.35
Ramachandran Plot Summary from Procheck [b]			
Most favoured regions	85.4%		
Additionally allowed regions	14.5%		
Generously allowed regions	0.0%		
Disallowed regions	0.1%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured regions	93.5%		
Allowed regions	6.4%		
Disallowed regions	0.1%		

[a] Residues with sum of phi and psi order parameters > 1.8

Ordered residue ranges: 3A-15A,17A-37A,40A-48A,54A-56A

[b] Residues selected based on: Dihedral angle order parameter, with S(phi)+S(psi)>=1.8

Selected residue ranges: 3A-15A,17A-37A,40A-48A,54A-56A

[c] With respect to mean and standard deviation for for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

Generated using PSVS 1.5

Table S2 Protein Structure Validation Suite (PSVS) summary of structure quality factors for three-state ensembles of GB3 obtained from eNOEs with 27 stereospecific assignments that can be determined from the NMR data (Fig. S1B).

Analyses performed for order residues.

Total structures computed	600		
Number of structures used	60		
RMSD Values			
	all	ordered[a]	Selected[b]
All backbone atoms	0.5 Å	0.5 Å	0.5 Å
All heavy atoms	0.9 Å	0.8 Å	0.8 Å
Structure Quality Factors - overall statistics			
	Mean score	SD	Z-score[c]
Procheck G-factor [a] (phi / psi only)	-0.42	N/A	-1.34
Procheck G-factor [a] (all dihedral angles)	-0.49	N/A	-2.90
Verify3D	0.41	0.0294	-0.80
ProsaII (-ve)	0.72	0.0717	0.29
MolProbity clashscore	18.04	4.4263	-1.57
Ramachandran Plot Summary from Procheck [b]			
Most favoured regions	87.7%		
Additionally allowed regions	12.1%		
Generously allowed regions	0.0%		
Disallowed regions	0.1%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured regions	94.5%		
Allowed regions	5.4%		
Disallowed regions	0.1%		

[a] Residues with sum of phi and psi order parameters > 1.8

Ordered residue ranges: 2A-15A,17A-37A,40A-48A,51A-56A

[b] Residues selected based on: Dihedral angle order parameter, with S(phi)+S(psi)>=1.8

Selected residue ranges: 2A-15A,17A-37A,40A-48A,51A-56A

[c] With respect to mean and standard deviation for for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

Generated using PSVS 1.5