Deviation versus violation plots: A new technique for assessing the self-consistency of NMR data

Marc Adler

Berlex Biosciences, 15049 San Pablo Avenue, Richmond, CA 94804-0099, U.S.A.

Summary

A new method has been developed to test the self-consistency of distance constraints derived from NOESY spectra. The technique is based on the premise that the further the atomic coordinates of any given structure vary from the 'correct' structure, the more NOE violations will occur in that structure. This relationship is quantified by plotting the deviation of each structure against the sum of the residual NOE violations. This type of plot is called a DVplot, which is generated by the following means: first the experimental constraints are used to generate a set of structures, then the amount of deviation and violation is quantified for each structure. The deviation of each structure is derived from the root-mean-squared deviation (rmsd) between each structure and the average structure. Violations are quantified for each structure using the new terms S1, S2, and S3. These terms measure the sum of all residual NOE violations greater than 0.1 Å, 0.2 Å, and 0.3 Å, respectively. DVplots are used to show that for series of structures calculated from a single set of NMR constraints, there is an approximate linear correlation between the rmsd and each of the three sums, S1, S2, and S3. Furthermore, it is proposed that the x-intercepts derived from the three plots of S1, S2, and S3 will converge if the NOE constraints are self-consistent. The new technique is applied to five different proteins using both experimental and simulated constraint sets.

Introduction

The problem treated in this paper, in theory, is simple; there is no reliable method for measuring the accuracy of structures derived from experimental NMR data (reviewed by James, 1994). Our colleagues in X-ray crystallography have a fairly straightforward approach to this problem. The final proposed structure is used to calculate a theoretical set of intensities, F_c. These values are then compared to the experimental intensities, F_o, using the following equation:

$$R = \frac{\sum |F_c - F_o|}{\sum F_o} \qquad (1)$$

where the summation is carried over all experimental data points. This equation measures the difference between the simulated and experimental results. Unfortunately, this approach has less utility when applied to NMR data. In a typical NMR spectrum, a large percentage of the signal may arise from artifacts. Examples of such artifacts include baseline distortions from receiver imperfections, residual signal from H_2O, partial bleaching of exchangeable amide protons, baseline distortions caused by diagonal peaks, T_1 noise ridges, and so on. Most of these problems have systematic effects on the observed intensities. For instance, baseline distortions are often strongest in the center of the spectrum, where they affect the measured intensities of the $C^\alpha H$ resonances. X-ray crystallography does have some similar problems with experimental artifacts, which have been addressed by a variety of techniques (Blundell and Johnson, 1976). However, these

---

problems rarely lead to systematic distortions in a subset of calculated interatomic distances.

There is another, more subtle, reason why it is difficult to compare experimental NMR data with a theoretical data set calculated from a model structure. In X-ray crystallography, each experimental intensity represents the summation of interactions over all interatomic pairs. In cases where the structure of a single region of a protein has been misinterpreted, the errors cause an overall reduction in the value of R. However, this effect is spread throughout the data set and does not cause a systematic distortion in a subset of calculated intensities. In NMR structure analysis, each NOE peak is primarily interpreted as an interaction between two distinct protons. Therefore, errors have a very localized effect, i.e., the misassignment of a single NOE will only directly affect that peak. This author once misinterpreted an NOE peak while solving the structure of kistrin (M. Adler and G. Wagner, unpublished results). This single mistake caused a 10 Å shift in the loop of the protein that contained the most critical residues. The proposed NOE assignment still provided a reasonable fit between calculated and measured NOE peaks. Generally, small mistakes in the interpretation of both the assignment and the intensity of NOE peaks can cause errors that often escape detection.

Determining the precision of NMR structures has been an easier task. Most NMR structure publications list the root-mean-squared deviation (rmsd) of a set of refined structures. This is a measure of the scatter in the atomic position between independently calculated structures. A test structure is first superimposed upon a reference structure of the macromolecule. The rmsd is then measured by calculating the average distance of the atoms in the test structure versus the corresponding atoms in the reference structure. However, this approach alone does not judge the accuracy of the NMR data used to calculate the structures, since it fails to detect any systematic errors in preparing the data set.

This paper proposes a new empirical method for assessing the quality of NMR structures. The method is based on a very simple assumption. Let us suppose that there is a reliable reference structure that has been calculated for a given protein. A good choice for the reference structure is the average structure calculated from an ensemble of superimposed structures. The exact choice is not critical in most cases. What is important is that the reference structure is a reasonable model for the 'correct' structure. The key premise of this paper is that the further the atomic coordinates of any given structure vary from the reference, the more NOE violations will occur in that structure. Therefore, there should be a correlation between the deviation from the reference structure as measured by the rmsd and the sum of all residual NOE violations.

This paper examines the results from different proteins using a new technique called deviation versus violation plots (DVplots). The results show that DVplots can be used to test the self-consistency of a constraint set. Examples are given for five proteins using both experimental and simulated constraint sets.

Methods

Structure calculations

All calculations were performed on an Indigo[2] extreme workstation (Silicon Graphics, Mountain View, CA, U.S.A.). Structure calculations were performed using the DGII module of INSIGHTII, v. 2.3.0 (Biosym Technologies, San Diego, CA, U.S.A.). Default settings were used unless otherwise noted. A matrix was constructed from the initial upper and lower bound constraints, which was first smoothed with the triangle inequality. Then distance geometry was used to generate between 50 and 150 independent starting structures. The structures were subjected to between 10 000 and 40 000 steps of simulated annealing. Each step corresponded to 0.2 ps. The molecular dynamics were constrained using a residual penalty function (Biosym Technologies, 1993). The residual penalty function yields a unitless number that measures the ratio of distances and it goes to zero when all constraints are met. A variable cutoff was used to test the residual penalty function. If the penalty function exceeded this amount, then the mirror image of the embedded structure was subjected to simulated annealing. If both structures failed to converge, then no further calculation was performed and the structure was rejected. Also, if a limited number (<4) of structures clearly had the incorrect fold and had aberrant values of the residual penalty function, then the structures were manually rejected. The number of rejected structures is given in Table 1. The structures that passed the cutoff underwent a maximum of 1000 steps of conjugate gradient minimization.

Preparation of constraint files and structure calculations

Table 1 lists the 12 sets of calculations presented in this paper. Each set is given a code name that is used throughout the paper.

Eglin C

The constraints for eglin C (Hyberts et al., 1992) were taken from the diskette that accompanied the original publication. The first seven residues of eglin C were removed to reduce the time needed to generate new structures. This modification does not alter the structure, because there are no long-range NOEs between these first seven residues and the remaining 63 residues. A program was written to translate the remaining 1115 NOE, 16 hydrogen bond, and 74 dihedral angle constraints into the INSIGHTII format. There were several problems with the

TABLE 1
STRUCTURES USED IN THIS WORK

Protein	Code name[a]	Description	Number of structures[b]	Penalty function[c]	Best rmsd[d] (Å)	Figure[e]
Eglin C	egc.5	5 Å upper bounds	50 (0)	0.1–1.4	0.70	1
Eglin C	egc.exp	Exp.[f]	75 (0)	0.5–1.7	0.39	2
C-loop, TM-EGF4	tm4c.exp	Exp.	100 (0)	0.03–1.0	0.57	3a
C-loop, TM-EGF4	tm4c.120	Upper bound at 20%	100 (0)	0.1–1.2	0.46	
C-loop, TM-EGF4	tm4c.115	Upper bound at 15%	100 (0)	0.3–2.4	0.48	
C-loop, TM-EGF4	tm4c.110	Upper bound at 10%	150 (1)	0.6–2.6	0.41	3b
Heregulin-α	hera.exp	Exp.	60 (7)	0.04–0.6	0.83	4a
Heregulin-α	hera.bad	Eight randomized NOEs	50 (6)	0.4–1.5	1.07	4b
Protein G	gpro.exp	Exp.	75 (0)	0.6–1.2	0.37	5a
Protein G	gpro.fix	Exact distance	75 (0)	0.03–0.3	0.26	
Protein G	gpro.sim	Simulate exp.	75 (0)	0.03–0.3	0.26	5b
Kistrin	kis.5	5 Å upper bounds	50 (2)	0.3–3.0	0.92	6

[a] This code name is used throughout the paper.
[b] The total number of structures calculated with a given set of constraints. The number of rejected structures is listed in parentheses.
[c] The range of the residual penalty values for the structure. Rejected structures are not included.
[d] The rmsd for the best 10% of the structures. In this case, the best structures are defined as having the lowest rmsd values when compared to the average structure.
[e] Figure where the data for this structure appear.
[f] The constraints are derived directly from the published experimental constraints.

original atom names. Stereospecific NOEs to the side-chain amine protons of asparagines and glutamines were lost. The program that converted the NOEs was checked for accuracy. The NOE constraints that were violated in at least 10% of the final structures derived from the original distances were examined by hand. No errors were found in these 28 constraints.

Two sets of NOE constraints were used to calculate new structures. The first, whose code is egc.exp, contained a direct translation of the original NOE constraints. The second set, whose code is egc.5, retained all lower bound distances, but the initial target for the upper bound distances was set to 5 Å. Hyberts et al. (1992) specified an additional length that was added to the upper bounds to compensate for inherent uncertainties with NOEs involving pseudoatoms, methyl groups, and aromatic ring protons. All these adjustments were published and used without modification. Embedded structures were then subjected to 30 000 steps of simulating annealing. All superimpositions of the structures were performed using the N, C^α, and C atoms of residues 8–38 and 50–70.

C-loop of TM-EGF4

The constraints for the 19-residue fragment of the fourth EGF-like domain of thrombomodulin (TM-EGF4), residues 371–389, had been generated previously by this author (Adler et al., 1995). The code name of this data set is tm4c.exp. Structure calculations indicated that there were consistent violations of two intraresidue NOEs between the HN and the HB protons of His[384], which were equal in intensity. This problem may have been caused by conformational mobility of this side chain (see Discussion). The decision was made to drop these constraints

from the data set. The remaining 210 NOE and 8 φ dihedral angle constraints were used without modification. Using 20 000 steps of simulated annealing, 100 structures were generated. Superimpositions were performed using the N, C^α, and C atoms of residues 372–389.

Four other sets of constraints were generated for this fragment, using a different algorithm to translate peak intensities into distance constraints. The initial formula (Adler et al., 1995) assumed that the peak height was accurate to within a factor of two. Furthermore, the formula assumed that the peak volume was proportional to $1/r^5$ instead of the usual $1/r^6$. This had the effect of lengthening the distances for the less intense NOEs. The net effect on the experimental data was that the target distances of 2.2, 2.5, 3.0, 3.5, and 4.0 Å were lengthened to 2.6, 3.1, 3.8, 4.6, and 5.0 Å, respectively (an upper limit of 5.0 Å was used for all observed NOEs). Trail sets of constraints were constructed by first calculating the exact distances from NOE peak intensities, using the standard formula. The upper bound was then set to either 110, 115, or 120% of this distance. This adjustment in the length of the upper bound represents an uncertainty of either 10, 15, or 20%, respectively, in calculating the distance from the NOE peak intensity. Lower bound constraints for all data sets were set to the van der Waals contact radii. The data sets are named tm4c.110, tm4c.115, and tm4c.120. For the tightest set of constraints, tm4c.110, this formula reduced the upper bound distances by roughly 0.4 Å. Only the longest upper bound distances were significantly shortened in the tm4c.120 constraint set. Either 150 structures (for tm4c.110) or 100 structures (for tm4c.115 and tm4c.120) were generated, using the same techniques as described above.

As stated previously, the distance bounds for these calculations were smoothed using triangle inequality. Calculations were also repeated for two of the data sets, tm4c.exp and tm4c.120, using the additional smoothing provided by the tetrangle inequality. The NOE between Ile[379] HN and Ile[379] HB had to be eliminated from both data sets. The upper bound distance for this NOE was set to 2.6 Å and this constraint caused fatal errors during the tetrangle smoothing. The tetrangle smoothing was limited to the constraints between sequential residues. The embedded structures were then subjected to either 10 000, 20 000 or 40 000 rounds of simulated annealing.

EGF domain of heregulin-α

Structural studies of this protein are still in progress; therefore, there is no published reference for the NOE constraints. The code name will be hera.exp. The solution structure of the EGF-like domain of heregulin-α has been published (Jacobsen et al., 1996). In our hands, the folded portion of this protein at pH 3.1 consists of a 49-residue stretch, residues 177–225, which is homologous to murine epidermal growth factor (EGF). There were 418 NOE, 14 hydrogen bond, and 27 ϕ dihedral angle constraints. No attempt was made to quantify the NOE intensities. Uncertainties in the prochiral assignments were handled using the same techniques as employed for the C-loop of TM-EGF4 (Adler et al., 1995). Sixty structures were subjected to 40 000 steps of simulated annealing. Superimpositions were performed using the N, C^α, and C atoms of residues 177–199 and 207–225. The excluded residues all belong to the chain reversal of the B-loop at the end of the β-sheet.

An additional NOE data set was constructed for this protein, which is identified by the name hera.bad. A total of eight NOEs were omitted from the data set because of uncertainties caused by chemical shift degeneracy. One of the protons that gave rise to a NOE could not be assigned. For the data set hera.bad, the assignment was made by random selection between the two choices. This data set was taken as a realistic example of spectra that contain a limited number of misassigned NOEs. Out of the 50 structures, six failed to converge.

Protein G

The constraints for protein G (Gronenborn et al., 1991) were obtained by personal communication (G.M. Clore and M. Whitlow). The code name will be gpro.exp. A C program was written to convert the format. Minor modifications had to be made to the constraints involving diastereo pairs, because DGII has less flexibility in declaring pseudoatoms than has XPLOR. Whenever a pseudoatom could not be used, the nearest heavy atom was selected and an additional distance was added to the upper bound to compensate for the distance between the heavy atom and the original pseudoatom. The dihedral angle

constraints on the side-chain angles and ψ were also ignored. The final data set for the 56-residue protein was constructed from 845 NOEs, 32 hydrogen bond constraints, and 54 constraints on ϕ (note that redundant NOEs were eliminated from the data and the shorter distance was retained). As before, the program that converted the NOEs was checked for accuracy. The NOE constraints that were violated in at least 10% of the final structures were examined by hand. No errors were found in these 45 constraints. Seventy-five structures were calculated using 20 000 steps of simulated annealing, followed by a maximum of 1000 steps of minimization. The N, C^α, and C atoms of all residues were used in the superimpositions.

Two sets of data were based on a set of ideal distances for protein G. The best structure based on the residual penalty function was selected from the structures calculated from the original experimental constraints. IN-SIGHTII was then used to list the actual distance for every interatomic pair present in the NOE constraint set. This included the HN to O and the N to O distances used as hydrogen bond constraints. An AWK program was written to convert these distances into experimental constraints. The program took the ideal distance and added ±0.4 Å to get the upper and lower bound distances. All constraints had a minimum value for the upper and lower bounds of 1.8 and 2.5 Å, respectively. In addition, 1.6 Å was added to all NOEs that included methyl protons. NOEs to methylene and aromatic protons with unknown stereoassignments had 1.1 or 2.1 Å, respectively, added to their upper bounds. The code name for this data set is gpro.fix.

A second data set was generated by adding a random error to the initial ideal distance; the code name for this data set is gpro.sim. The random error was generated by using the Box–Muller approximation of a normal (Gaussian) deviate, taken from the Numerical Recipes (Press et al., 1986). A distance of 0.2 Å was used as a single standard deviation. The program took the new distance and added ±0.5 Å to get the upper and lower bound distances. Seventy-five structures were calculated, using the same techniques as for the original protein G data set.

Kistrin

The original data set (Adler et al., 1991) contained many inconsistencies that were not detected in the original structure calculations using the program Dspace. Thus, the initial upper bound distances for all 562 NOEs were set to 5.0 Å and retained all the original adjustments for methyl, methylene and aromatic protons. Twenty-one constraints for ϕ were also used from the original work. The code name for this set is kis.5. Fifty structures were subjected to 40 000 steps of simulated annealing. Superimpositions were performed with the N, C^α, and C atoms of residues 4–45 and 55–64.
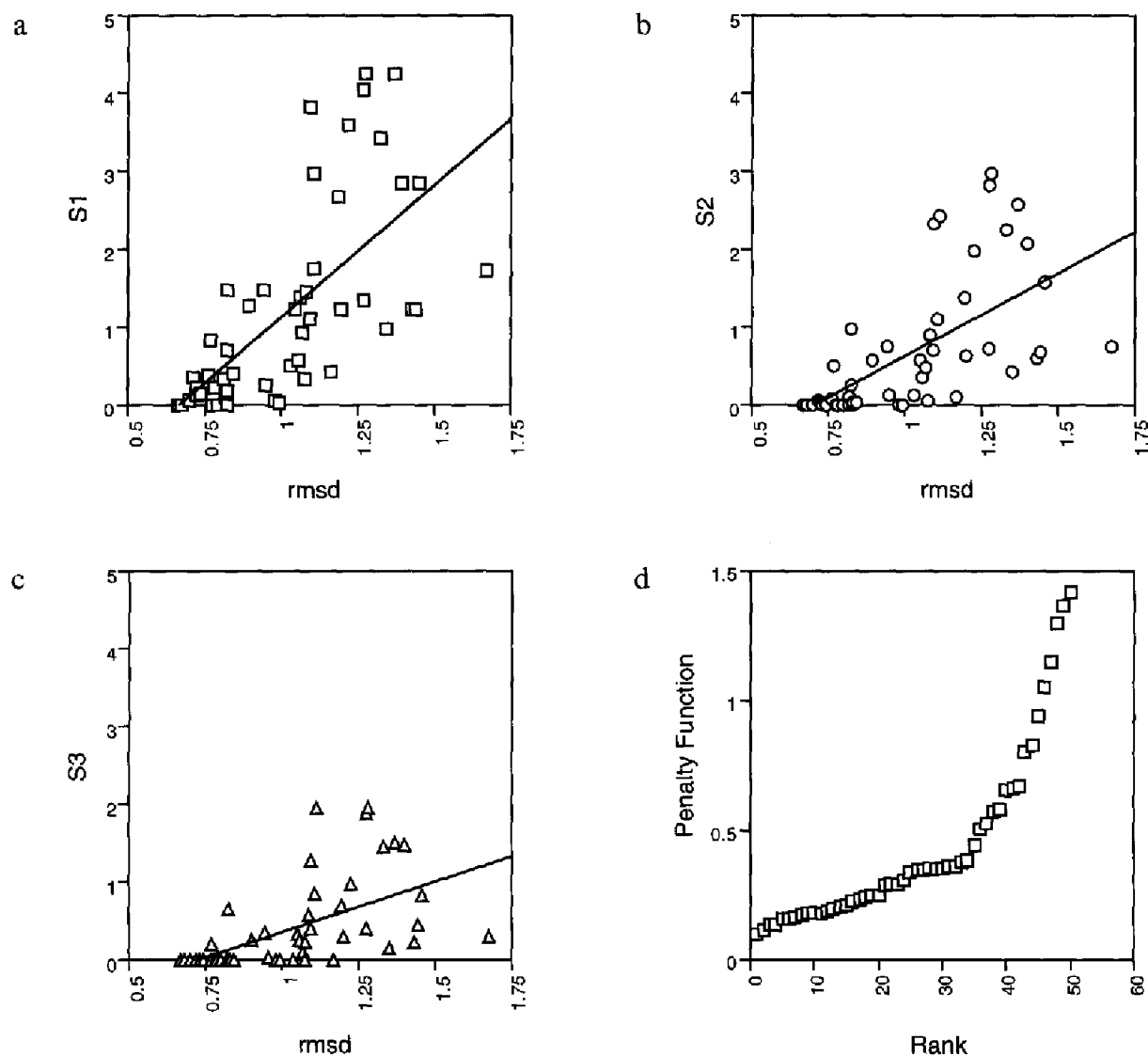
Fig. 1. DVplots for all structures of eglin C, data set egc.5. (a) S1 versus rmsd; (b) S2 versus rmsd; and (c) S3 versus rmsd. Lines represent a linear least-squares fit for each plot. The axes of all plots, except (d), are in Å. (d) Ranking of the residual penalty functions.

Calculating the average structure

An average structure was calculated from a subset of structures using the following means. The best structure was selected based on the value of the residual penalty function as measured by INSIGHTII. Then a subset of all final structures was selected by setting a cutoff for the residual energy function. Care was taken to ensure that the cutoff was set low enough to exclude all structures that did not have the same chain trace. The subset structures included a little more than half of the final structures. The selected structures were then superimposed on the best structure, using the N, C$^\alpha$, and C atoms of selected residues. A subset of residues, listed for each of the proteins, was excluded when the original authors indicated that parts of the structures were poorly defined. Coordinates of the superimposed structures were then arithmetically averaged. No attempt was made to correct the geometry of the covalent structure using restrained molecular dynamics.

Deviation versus violation plots

Deviation versus violation plots (DVplots) were developed in order to establish a correlation between the rmsd and the residual NOE violations. An AWK program, getinserror.a, generated an INSIGHTII macro that performed two tasks. First, it superimposed each structure on the average structure, using the selected backbone atoms listed above. This command also printed the value of the rmsd. Second, it listed all NOE violations that were greater than 0.09 Å and printed the size of the violation. The output of INSIGHTII was directed to a text file as the macro was running. A second AWK program,

dvplot.a, analyzed the text file. For each structure, it kept five running sums based on the size of the distance violations. First, it subtracted 0.1 Å from all the violations and added any residual positive distance to the variable S1. Then 0.2 Å was subtracted from the initial distance violation and the positive residuals were added to S2. The process was repeated for distances of 0.3, 0.4, and 0.5 Å. Once all NOEs had been examined, the program printed out the name of the structure, the value of the rmsd for the backbone atoms, the residual penalty value and the five running sums S1–S5. Columns were delineated by tabs. The results were then transferred to a Macintosh and analyzed using the program Cricket Graph III (Computer Associates International Inc., New York, NY, U.S.A.). It was found by experience that most structures that had the correct fold contained no NOE violations greater than 0.4 Å. Therefore, only the sums S1, S2, and S3 are discussed in the rest of the text.

Statistical analysis

All algorithms for determining the linear least-squares fit and the standard approximation for the correlation coefficient, r, were adapted from the CRC Standard Mathematical Tables and Formulae (1991). The coefficient r can assume values from −1 to 1. A value of 1 means that x and y are linearly correlated, i.e., the points give a good fit to a straight line with a positive slope. A value of 1 would be obtained even if there was random error in the measurement of y. An r of 0 indicates that y is not a function of x. Negative values of r are found when the best fit line has a negative slope.

Results

Figure 1 was derived from the published NOE data set of eglin C (Hyberts et al., 1992). The structure calculations have been explained in the Methods section. An important change was made in the constraint file, i.e., all the initial target distances for the upper bound were set at 5 Å (see Methods). A suite of AWK programs was used to obtain the residual NOE violations for each structure, as explained in the Methods section. These measures are called S1, S2, and S3. S1 is generated by subtracting 0.1 Å from all NOE violations and adding together all the remaining positive values. S2 and S3 are calculated the same way by subtracting 0.2 and 0.3 Å, respectively. Figure 1a shows S1 plotted against the rmsd, with each point representing a single structure. Figures 1b and c show the corresponding plots for S2 and S3, calculated from the same set of structures.

S1, S2, and S3 all measure the size of the residual NOE violations. They are, however, sensitive to different phenomena. A violation of 0.3 Å or greater would indicate that a structure is incompatible with the NOE con-

straints. This problem could arise from incorrect NOE assignments. Alternatively, a high value of S3 occurs when the structure is misfolded. This was found for several structures in this data set with high residual penalty functions (see below). S1 is a more sensitive measure of error than S3. It would detect problems with the interpretation of NOE peak volumes. This would lead to small but systematic errors in the distance bounds. Thus, there would be a large number of violations on the order of 0.2 Å or less, which would increase S1 without affecting S2 and S3. This effect was observed in data sets tm4c.115 and tm4c.120 (see below).

Inspection of Fig. 1 indicates that there is only a loose correlation between the rmsd and the residual NOE violations. Indeed, the correlation coefficient r ranges from a low value of 0.56 for Fig. 1c to a high value of 0.67 for Fig. 1a. A large part of the problem stems from the fact that a single number, either the rmsd, S1, S2, or S3, is used to sum up the conformation of the protein. Clearly, a single value cannot accurately describe the conformation of a protein and it is expected that conformational perturbations will have different effects on different measurements. For instance, rotation of a side chain about χ^1 may produce violations in the NOE constraints without affecting the rmsd value for the structure.

The plots shown in Figs. 1a–c do have some features in common. Noticeably, the x-intercept of the best fit line for plots 1a–c is 0.70 ± 0.03 Å. As explained above, S1, S2, and S3 are different ways of measuring the residual NOE violations. The close agreement in the x-intercepts indicates that the data from each plot reflect the same basic phenomenon, i.e., the further the structure is from the reference structure, the greater the residual NOE violations. If this premise is completely correct, then this intercept represents the limiting accuracy of the data set. If the structure calculations worked 'perfectly', then every structure should reach the point where there are no residual NOE violations and the backbone of all the structures would lie within 0.7 Å of this average structure. As will be shown below, the convergence of the x-intercept (Fig. 1) depends on the accuracy of the constraint set used to calculate the structures. Furthermore, this convergence is a sensitive test as to whether the data are self-consistent.

One more point about the technique should be discussed before moving onward. Figure 1d shows a plot of residual penalty functions of the 50 calculated structures for egc.5. In this case, the values of the penalty function have been presorted by their numerical values. The x-axis represents the ranking of each value amongst the 50 structures. The first 35 points fit on a straight line. However, the error rapidly rises for the last 15 structures. Examination of six of these structures, chosen at random, revealed a common feature. In each structure, at least one side chain was misfolded and appeared on the wrong side of

TABLE 2
STATISTICS FROM DVPLOTS

Code name	All structures		Best structures[a]	
	x-intercept[b]	r[c]	x-intercept	r
Passed test for self-consistency[d]				
egc.5	0.70 ± 0.03	0.62 ± 0.05	0.66 ± 0.06	0.75 ± 0.01
tm4c.exp	0.66 ± 0.06	0.79 ± 0.07	0.52 ± 0.09	0.53 ± 0.10
gpro.sim	0.30 ± 0.04	0.80 ± 0.05	0.19 ± 0.07	0.16 ± 0.11
Marginal results[d]				
tm4c.120	0.59 ± 0.15	0.87 ± 0.04	0.37 ± 0.23	0.71 ± 0.11
tm4c.115	0.31 ± 0.33	0.90 ± 0.02	0.19 ± 0.35	0.72 ± 0.02
hera.exp	0.53 ± 0.07	0.54 ± 0.06	0.36 ± 0.23	0.37 ± 0.08
gpro.fix	0.21 ± 0.04	0.23 ± 0.01	0.05 ± 0.09	0.07 ± 0.01
Failed test for self-consistency[d]				
egc.exp	−0.14 ± 0.14	0.50 ± 0.05	−10.17 ± 10.06	0.13 ± 0.09
tm4c.110	−0.59 ± 0.66	0.86 ± 0.02	−4.14 ± 3.14	0.19 ± 0.04
hera.bad	0.00 ± 0.42	0.60 ± 0.01	−1.19 ± 0.47	0.28 ± 0.05
gpro.exp	−0.33 ± 0.09	0.32 ± 0.02	11.62 ± 19.94	0.00 ± 0.04
kis.5[e]	0.06 ± 0.41	0.91 ± 0.01	−22.21 ± 30.24	0.06 ± 0.05

[a] The best two thirds of the structures. Best is defined as having the lowest value of the residual penalty function.
[b] Mean and standard deviation (in Å) for intercepts for the best fit lines through S1–S3.
[c] Mean and standard deviation for correlation coefficients, r.
[d] See text for a definition of these terms.
[e] Only 27 out of the 48 structures had the correct fold and were used as the best structures.

the backbone. There was a large number of NOE violations in the region surrounding the misfolded side chain. Presumably, the side chain became trapped during simulated annealing.

Plots similar to that shown in Fig. 1d were made for the data sets presented in this paper. All of these had roughly the same shape: a linear increase in the residual penalty function for the first 70–90% of the structures, followed by a rapid rise. Presumably, the structures with the large residual penalty function failed to converge. Perhaps a greater commitment of both intellectual and computational resources would have improved the convergence. However, a simple approach was adopted. It was decided that a parallel analysis would be performed using only the best two thirds of the structures as judged by the residual penalty function. Thus, the results focus on those structures that had the correct fold. As can be seen in Table 2, the x-intercepts remain constant for the data set egc.5 when only the best structures are used in making the DVplots. It will be shown that this behavior is not found for data sets where errors have been purposely introduced.

Figure 2 shows a second DVplot for the protein eglin C. The only difference between the two sets of data shown in Fig. 1 and in Fig. 2 is that the original upper bound distances were used in Fig. 2 (Hyberts et al., 1992). The plot in Fig. 2 indicates that above roughly 0.55 Å rmsd there is still a weak correlation between the rmsd and the

residual NOE violations. This correlation is mostly dependent on the structures with the largest residual penalty functions and may include misfolded structures. However, below 0.55 Å most structures appear to have roughly the same amount of residual error. This is reflected by the large drop in the correlation coefficients. Table 2 lists the statistics for this data set. The x-intercepts for the best structures no longer converge to a single point. The low values of the correlation coefficients indicate that the x-intercepts may not retain any useful information. These results suggest that the original constraints are not self-consistent.

Figure 3a shows the results obtained for the C-loop of the fourth EGF-like domain of thrombomodulin (C-loop of TM-EGF4). The NOE peak intensities were quantified for 179 of the 210 constraints for this 19-residue fragment. The intercept of S1 is displaced downwards by 0.1 Å from those of S2 and S3. One possible explanation for this result stems from the large number of violations of the NOE that connected Ile[379] HN to Ile[379] HB. This upper bound distance was set to 2.57 Å and the average violation for this constraint was 0.12 Å. These violations shifted the value of S1 upwards for each structure, which in turn shifted the x-intercept of S1 downwards. However, these errors were too small to affect S2.

A series of structure calculations was performed using progressively tighter upper bound constraints (see Methods). Internal checks, such as the intraresidue and sequential HN to $C^\alpha H$ distances, were used to verify the accuracy of the constraints. The range of interatomic distances predicted for the tightest set of constraints, tm4c.110, was close to that of the expected distances (Wüthrich, 1986). Therefore, there was little room in the constraints for any experimental errors. Table 3 shows that the process of tightening the constraints has system-
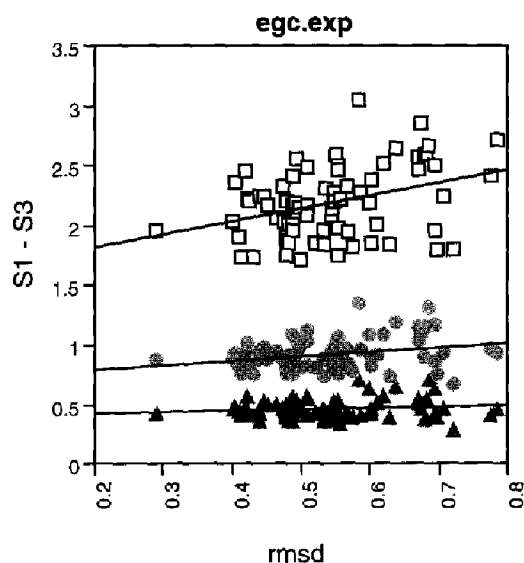

Fig. 2. DVplots for all structures of eglin C, experimental constraints, data set egc.exp. (□): S1; (✳): S2; (▲): S3.
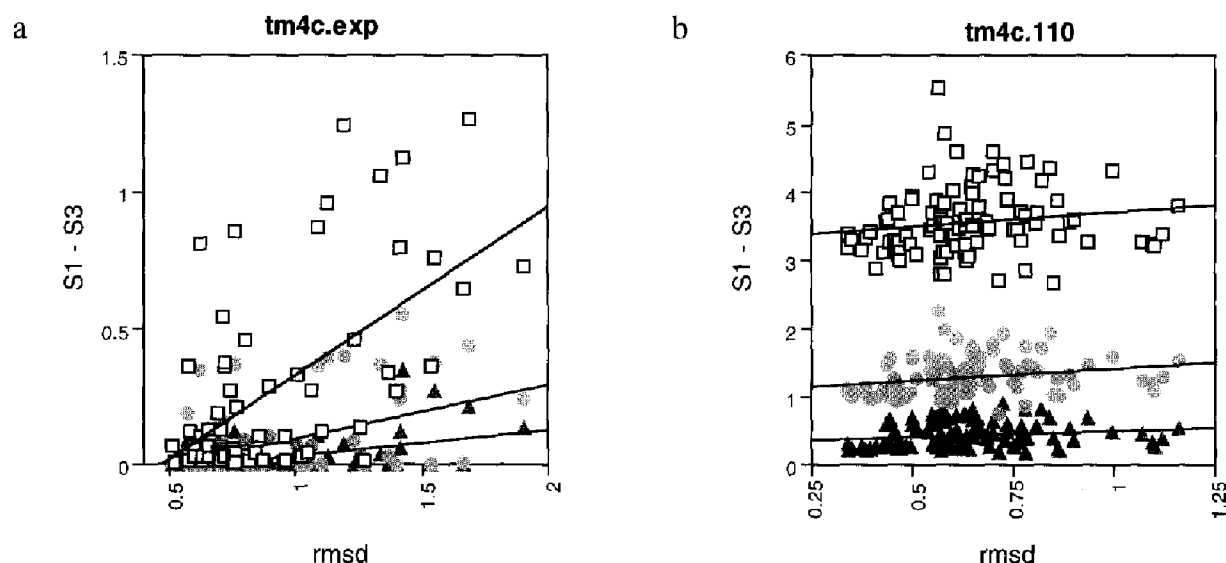
Fig. 3. DVplots for the C-loop of TM-EGF4, showing only the best two thirds of the structures. (□): S1; (*): S2; (▲): S3. (a) Experimental constraints, tm4c.exp; (b) overconstrained data set, tm4c.110.

atic effects on the results. Notably, the x-intercepts of DVplots become steadily smaller as the constraints get tighter (Table 3). This shift is caused by the steady decrease in the intercept of S1 (data not shown). Also, the convergence of the x-intercepts drops considerably. However, the precision of the structures as measured by the backbone rmsd of the best structures is significantly improved for the tightest set of constraints, tm4c.110. Visual inspection of the average structures indicates that systematic distortions are introduced by using the tighter constraints (data not shown). This trend is confirmed by a steady reduction in the radius of gyration. It is unlikely that any structural information is gained by using the tighter constraints.

Tests were performed to measure how the different annealing protocols would affect the results from the DVplots. The structure calculations were repeated for the constraint sets tm4c.exp and tm4c.120, using tetrangle smoothing of the distance bounds (see Methods). A new

set of embedded structures was derived from the modified distance bounds, and simulated annealing was performed for either 10 000, 20 000 or 40 000 iterations. The x-intercepts did converge for data set tm4c.exp using 10 000 and 20 000 rounds (0.62±0.05 Å and 0.50±0.06 Å, respectively, for the best two thirds of the structures). However, the x-intercept did not converge when simulated annealing was continued for 40 000 iterations. These calculations produced 57 structures with no NOE violations greater than 0.1 Å. Under these circumstances, the x-intercepts for S2 and S3 are not reliable. Experience has shown that the x-intercepts often do not converge for data sets that contain a large number of structures with no NOE violations (tm4c.exp, gpro.fix, gpro.sim, and data not shown). However, the existence of these structures by themselves can be taken as sufficient proof that the NOEs are self-consistent.

The x-intercepts did not converge for any of the calculations performed with the tm4c.120 data set. However,

TABLE 3
EFFECT OF TIGHTENED NOE CONSTRAINTS

Code name	Description[a]	Best rmsd[b] (Å)	x-intercept[c]	Radius of gyration[d] (Å)	S1 of average tm4c[e] (Å)
tm4c.exp	Exp.	0.57	0.52±0.09	7.58	0.0
tm4c.120	UB + 20%	0.46	0.37±0.23	7.56	0.5
tm4c.115	UB + 15%	0.48	0.19±0.35	7.53	2.5
tm4c.110	UB + 10%	0.41	−4.14±3.14	7.49	6.0

[a] Exp. in the first row refers to the published data set. For the other three data sets, the upper bound (UB) was calculated directly from the peak intensity and then increased by the percentage shown in this column (see Methods).

[b] The rmsd for the best 10% of the structures. In this case, the best structures are defined as having the lowest values of the rmsd when compared to the average structure.

[c] Mean and standard deviation (in Å) for S1-S3, using the best two thirds of the structures.

[d] Measured using the heavy atoms of residues 372-389.

[e] The residual NOE violations, S1, were calculated for reference structures using the various restraint sets. The average structure calculated from the tm4c.exp constraint set is used as the reference.
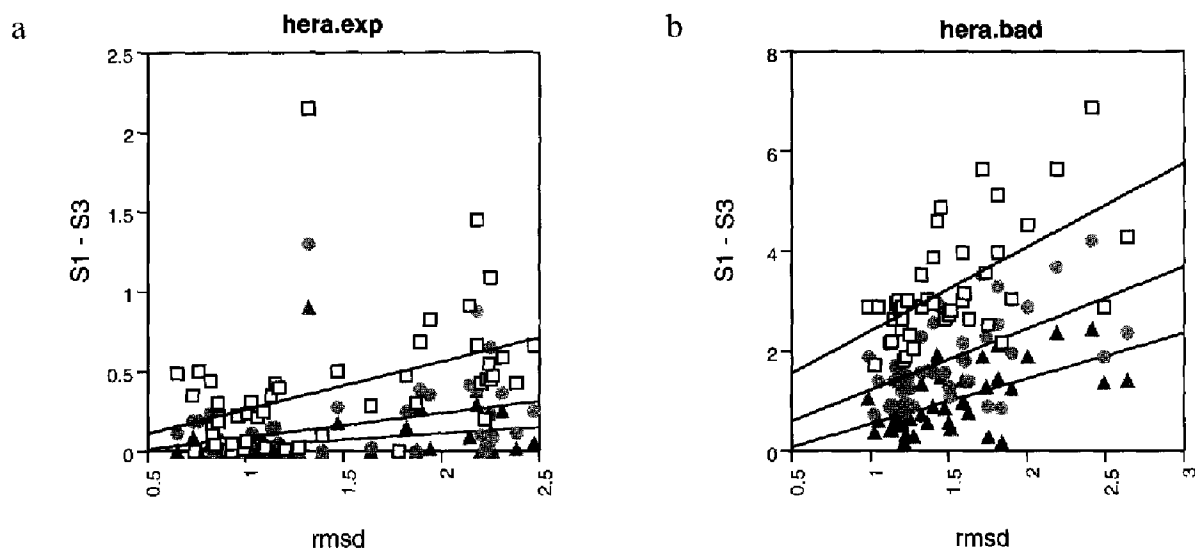
a
hera.exp

b
hera.bad

Fig. 4. DVplots for heregulin-α, showing only the best two thirds of the structures. (□): S1; (●): S2; (▲): S3. (a) Experimental constraints, hera.exp; (b) incorrect NOEs included, hera.bad.

values for both the x-intercepts and the correlation coefficients did vary with the number of iterations. The results showed several trends. As the number of iterations increased, the mean value of the x-intercepts dropped from 0.2 Å to −0.5 Å, the convergence of the x-intercepts fell from ±0.4 Å to ±1.0 Å, and the correlation coefficients went from a high of 0.8 to a low of 0.2. There was also 30% reduction in the average value for the rmsds. These trends indicate that the structures converged on a well-defined fold. However, this convergence did not relieve the residual NOE violations. Therefore, the additional rounds of simulated annealing reduced the correlation between deviation and violation. The results from the DVplots consistently showed that the tm4c.120 data set

was not self-consistent, although the quantitative results did vary with the number of iterations.

The effect of misassigned NOEs can be seen in Fig. 4 and in Table 2. Figure 4a shows the results for the original data set of heregulin-α. Figure 4b shows the comparable results when eight new NOEs were added to the original data set. All eight of these NOEs were initially omitted due to problems with overlapping resonances. In the second data set, the assignments were made by random selection between two possible assignments. As is shown in Table 2, the correlation coefficients for both sets of data are still fairly high. However, the x-intercepts no longer converge to a positive number in the second data set.
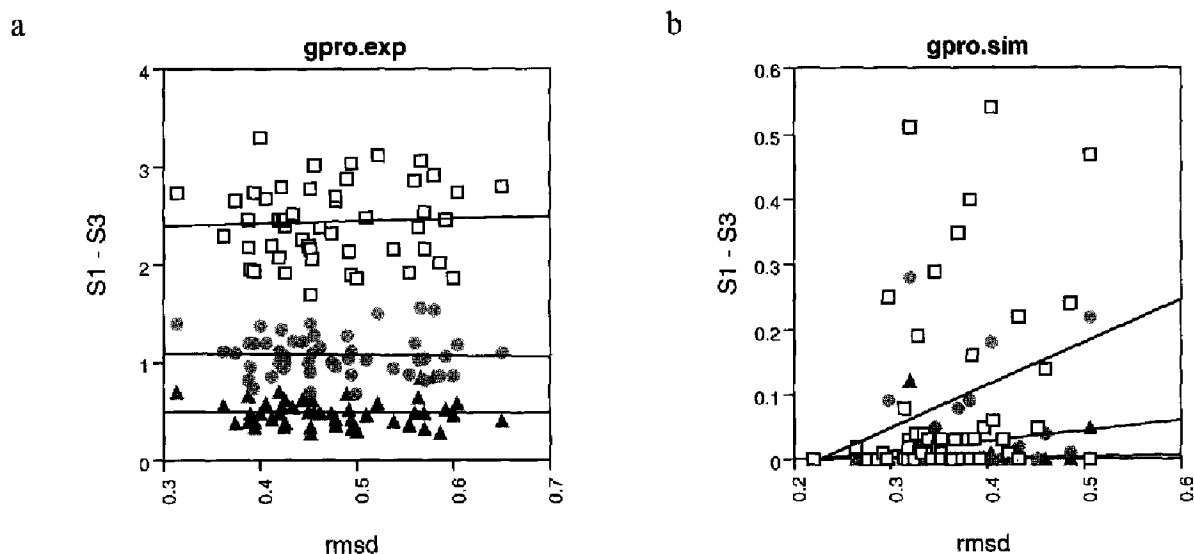
a
gpro.exp

b
gpro.sim

Fig. 5. DVplots for protein G, showing only the best two thirds of the structures. (□): S1; (●): S2; (▲): S3. (a) Experimental constraints, gpro.exp; (b) data set with simulated experimental distances, gpro.sim.

Figure 5a shows the DVplot obtained for protein G based on the published constraint set (Gronenborn et al., 1991). The data were analyzed as previously described. As shown in Fig. 5a, the x-intercepts fail to converge and the value of the correlation coefficient (Table 2) has dropped compared to the egc.5 data set.

The calculated structure for protein G was used to generate an ideal set of NOE constraints for the protein (gpro.fix). The original NOE constraints were used as the basis of the new data set. The pairs of atoms used to define each NOE remained the same. However, the initial target distances were obtained by measuring the interatomic distances directly from the best calculated structure of protein G, using the original constraints (gpro.exp). Upper and lower bounds were then constructed by adding ±0.4 Å to represent the experimental uncertainty. As can be seen in Fig. 5b and in Table 2, the x-intercept still converges, 0.05 ± 0.09 Å. However, the average value of the x-intercepts is unreasonably low and the values of the correlation coefficients are close to zero. This suggests that the method may be less accurate for high-resolution structures.

Roughly 20% of the structures in the data set gpro.fix have no NOE distance violations greater than 0.1 Å; therefore, the S1 values of these structures equal zero. Somewhat surprisingly, the rmsd values for these structures range from 0.2 to 0.5 Å. One reason for the low value of the correlation coefficients for this data set is that there are a large number of points along the baseline. These results are similar to those described for tm4c.exp when 40 000 iterations of simulated annealing were employed.

A second, more realistic, data set was constructed using the same initial target distances, gpro.sim. In this case, each initial distance was modified by adding a random distance (see Methods). The results presented in Table 2 indicate that the x-intercepts converge with reasonable accuracy to 0.19 ± 0.07 Å. However, the best fit line for S3 gave lower values for both the intercept and the correlation coefficient. Again, these problems stem from the large number of points along the baseline.

The final data set is based upon the NOEs derived for the protein kistrin (Adler et al., 1991). The calculation was performed by setting all initial upper bound distances to 5.0 Å. Figure 6 shows the results for 68 structures. As with the data shown for eglin C in Fig. 2, there is correlation between the rmsd and the NOE violations for structures further from the average structure than a given distance (1.7 Å). In this case, the structures that lie outside the 1.7 Å cutoff no longer have the same global fold. Structures within this 1.7 Å envelope show no correlation between error and rmsd (Table 2). The data indicate that the NOE constraints were not self-consistent. Presumably, this problem arose due to mistakes in the assignments of the NOEs.
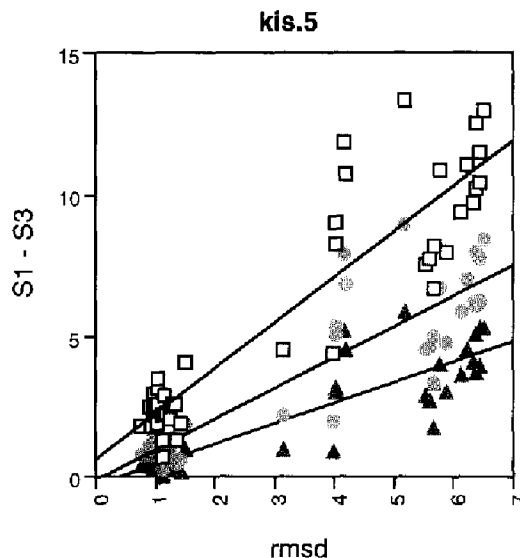

kis.5

Fig. 6. DVplots for all structures of kistrin, data set kis.5. (□): S1; (●): S2; (▲): S3.

The data sets can be roughly broken up into three main groups. The data sets are distinguished by the convergence of the x-intercepts for the best two thirds of the structures; see Table 2, column 4. The first group included egc.5, tm4c.exp, and gpro.sim. The x-intercepts all converge within ±0.1 Å. The x-intercepts themselves are positive numbers that could reasonably represent the limiting accuracy of the data sets. The x-intercepts also remain fairly constant when either the entire data set, or just the best structures, are examined. The second group includes tm4c.120, tm4c.115, hera.exp, and gpro.fix. There is a poor convergence of the intercepts, and the mean values of the intercepts are lower. With the exception of gpro.fix, the results indicate that there were minor inconsistencies in the data, on the order of 0.3 Å or less. The third group of structures includes five data sets (Table 2). The best structures no longer show a strong correlation between deviations and violations.

Technical notes During the development of this technique, several modifications were tried in an effort to improve the results. These modifications generally had little effect on those data sets where the results were straightforward. For example, the sets exp.5 and tm4c.exp always passed the test for self-consistency; kis.5 and tm4c.110 always failed. However, the values obtained from the DVplots for some data sets would change significantly with small variations in the method. In particular, the results from the simulated data sets, gpro.fix and gpro.sim, were dependent on the exact choice of parameters.

The most useful results were obtained by analyzing only the best two thirds of the structure. If the misfolded structures were included in the analysis, then there was a strong correlation between deviation and violation. This can be seen from the high values of the correlation coefficient obtained when all the structures were included in the

analysis (Table 2, column 3). However, these results tell more about the problems with the structure calculations than with the NOE constraints.

The calculations always produced structures that failed to converge. This was true even in a test run with the data set tm4c.exp, where tetrangle smoothing was applied and the number of rounds of simulated annealing was doubled to 40 000 (data not shown). The nonconverged structures would have aberrant values of the residual penalty function, rmsd, or NOE violations, but not necessarily all three. Therefore, the DVplots were sensitive to the means of selecting the best two thirds of the structures. The results presented in Table 2, columns 4 and 5, were measured using the residual penalty function as the criterion for selecting the best structures. The DVplots were also calculated using the rmsd to select the best structures. The two data sets with simulated NOEs, gpro.fix and gpro.sim, behaved differently, depending on the selection criterion. The x-intercepts for gpro.fix converged to 0.22 ± 0.05 Å if the best structures were selected by the rmsds, but failed to converge if the residual penalty function was used (Table 2). The results from gpro.sim showed the opposite behavior based on the selection criterion. Experience has shown that a few structures with a large number of NOE violations could drastically alter the direction of the best fit lines through the data points. The x-intercepts would converge for these data sets, but the convergence would disappear when the structures with large NOE violations were omitted. Therefore, the plots shown in the figures became an integral part of the analysis. Only visual inspection could confirm if the convergence of the x-intercepts represented legitimate trends.

One question raised during the development of this technique was which structure should be selected as the reference structure for measuring the rmsd. The results presented here use the average as a reference. However, the average structure is a mathematical construct, and there are significant distortions in the covalent geometry. A structure derived from the experimental constraints would avoid these problems. To test which structure makes a better reference, the analysis of the deviations versus the violations was repeated for the following data sets: egc.5, tm4c.exp, tm4c.115, hera.exp, gpro.sim, and gpro.fix. In each case, four different structures were used as a reference for measuring the rmsds. A structure was considered to be a suitable reference if the x-intercepts converged for S1, S2, and S3 using only the best two thirds of the structures. The data for the reference structure were not used in the analysis, since its rmsd with itself is 0 by definition. For egc.5 and tm4c.exp, the results were fairly insensitive to the choice of reference structure, as long as the reference structure had the correct fold. The second group of structures, tm4c.115, hera.exp, and gpro.sim, gave roughly the same results when the

best structure was used as the reference or when the reference structure had a small number of violations that were less than 0.2 Å. However, the results degraded when the reference structure had NOE violations greater than 0.3 Å. The final data set, gpro.fix, proved to be very sensitive to the choice of reference. It gave significantly different results, even when the best structure was used. The results indicate that the best structure can act as a suitable reference for the rmsd. However, the average structure seems to be slightly more reliable. (Note that using a real structure as the reference can also affect the selection of the best two thirds of the structures. Poor results were obtained if a real structure was used as the reference and the best two thirds of the structures were selected by the size of the rmsd.)

Finally, new measures of both deviation and violation were tested. It was hoped that better measures of these quantities would demonstrate a closer correlation between the two. In one attempt, the square of the residual NOE violations beyond a certain cutoff was used instead of S1, S2, and S3. It was felt that the square of the distance might better reflect the energy terms used for these constraints during simulated annealing. For example, if there was an NOE violation of 0.25 Å, it then would add 0.15 Å to the value of S1. The same violation would add 0.0225 Å² to the corresponding term S²1. However, there was no consistent improvement in the correlation coefficients (data not shown) and the technique was not pursued further.

Another approach used a new measure of the rmsd. The atoms specified in the NOE constraint file became the reference points for measuring the rmsd. The weight of each atom was proportional to the number of times each atom appeared in the NOE constraint file. This measure was called NOE weighted root-mean-squared deviation (nwrmsd). One advantage of this approach is that the nwrmsd is sensitive to the position of the side chains. The technique also automatically increases the weight for atoms in the best determined parts of the structure.

There were, however, problems with this method. First, there was no reliable way to include information from the angular constraints that were derived from the coupling constants. More importantly, not all the NOEs were of equal importance. Intraresidue and sequential NOEs often contain less structural information than long-range NOEs. A crude filter was used to eliminate the short-range NOEs from the determination of the weight of each atom. A more precise measurement of the importance of each NOE might have improved the results.

Using the nwrmsd, there was a noticeable improvement in convergence of the x-intercepts for the data sets based on simulated NOEs, gpro.fix and gpro.sim (data not shown). However, there was no substantial improvement in the values of the correlation coefficients for most of the data sets.

The method presented in this paper was selected, at least partially, for its simplicity. More complicated approaches failed to deliver significant improvements. The reader, however, is encouraged to explore new techniques to find the one that works best for him or her.

Discussion and Conclusions

In theory, NOEs act as guides that push a structure towards the correct fold. The reverse side of this proposition states that the NOE constraints should also act as faithful reporters that will post warnings if the structure strays from the fold. The method presented here is based on the second premise. The empirical results show that this premise is at least qualitatively true. It has been more difficult to establish a quantitative relationship for some of the examples.

The convergence of the x-intercepts is a fairly sensitive test for self-consistency. Experience has shown that the most reliable results are obtained if the intercepts are measured using only the best two thirds of the structures. In this author's opinion, a constraint set is self-consistent if the average value of the x-intercept for S1–S3 is a reasonable positive number with a standard deviation of less than 0.1 Å. The results shown in Table 2, column 5, correlate well with the preconceived bias that this author got from examining residual NOE violations.

The technique is also fairly easy to use and it does not require any extensive new calculations. For a given set of structures, the user must determine the average structure and construct a list of NOE violations. These steps are fairly routine for most investigators. What is proposed here is a new way to compile the information.

Potential problems

This technique measures the self-consistency of a data set, not its accuracy. It is possible to make errors in the constraints that will only cause minimal perturbations to the structure. For instance, the published structures of both eglin C (Hyberts et al., 1992) and protein G (Gronenborn et al., 1991) closely resemble X-ray structures of the same proteins, even though the results presented here indicate that the data sets are not self-consistent. However, the errors in the distance constraints may have caused subtle distortions in the structure.

The assumption of a single static conformation, by itself, can generate inconsistencies in the data. It was shown that NOE violations in tendamistat resulted from multiple conformations of the side chain of Tyr[15] (Torda et al., 1990). The calculation presented here could be performed by summing the violations from the time-average distance constraints as used by Torda et al. (1990). This should relieve NOE violations that arise from multiple conformations. However, this author does not have the appropriate software to perform this test.

Using the methods presented here, the technique would detect inconsistencies in the initial distance constraints, even if the errors resulted from conformational flexibility. However, this may still present useful information to investigators.

It is hard to judge whether the utility of DVplots will extend to NMR structures that are derived from other refinement protocols. Currently, DGII is the only processing package available in this laboratory. However, the relevant features in DGII are shared by many algorithms. First, an initial set of coordinates is generated by some form of randomization. Restrained molecular dynamics are then used to search the conformational space. This is usually combined with a simulated annealing protocol. As in most packages, a minimum temperature is maintained to ensure that the structures can hop between local minima. Finally, the simulated annealing is stopped and the energy of the structure is reduced through minimization. The structures that retain a large number of NOE violations are presumably trapped in local minima. To a certain extent, this technique does depend on the information provided by these higher energy structures, which have the correct fold but still retain some NOE violations. Therefore, it is difficult to judge the effects of different annealing protocols on the results. The limited results presented here using different DGII protocols indicate that quantitative results may vary, but the qualitative results remain the same.

Empirically, the technique is less reliable for cases where there are a large number of structures with no NOE violations. In these cases the x-intercepts for S2 and S3 can be very sensitive to a small number of outlying points. Simple inspection of the residual NOE violations provides more accurate information for these examples than the DVplots.

It is noteworthy that other tests for accuracy have not reached the same conclusions with the same data sets. Brünger and co-workers (1993) have developed a technique based on randomly deleting approximately 10% of the data set and then recalculating the structure with the reduced data set. This technique was originally adapted from X-ray crystallography (Brünger, 1992) and was used to analyze the constraint set for protein G. A detailed analysis of this technique lies beyond the scope of this paper. The work presented here suggests that there may be significant problems with the original data, which escaped detection by Brünger and co-workers (1993). Perhaps the discrepancies arise because the different techniques are sensitive to different phenomena.

Finally, it should be noted that the technique does not establish any relationship between the NOE constraints and the experimental data. A true test for accuracy must include a comparison between the calculated and predicted spectra. This task lies beyond the scope of this paper.

416

Acknowledgements

I would like to thank Laurie Adler, Jerry Dallas, Doris A. Hollander, Brian Sykes, and Marc Whitlow for helpful discussions and for their assistance in preparing this manuscript. Furthermore, I thank Berlex Biosciences for their generous support of this work.

References

Adler, M., Lazarus, R.A., Dennis, M.S. and Wagner, G. (1991) *Science*, **253**, 445–448.

Adler, M., Seto, M.H., Nitecki, D.E., Lin, J.-H., Light, D.R. and Morser, J. (1995) *J. Biol. Chem.*, **270**, 23366–23372.

Biosym Technologies (1993) *NMRchitect User Guide*, v. 2.3, San Diego, CA, U.S.A., pp. 2-16–2-18.

Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*, Academic Press, New York, NY, U.S.A., pp. 310–333.

Brünger, A.T. (1992) *Nature*, **355**, 472–475.

Brünger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R. and Nilges, M. (1993) *Science*, **261**, 328–331.

Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T. and Clore, G.M. (1991) *Science*, **253**, 657–661.

Hyberts, S.G., Goldberg, M.S., Havel, T.F. and Wagner, G. (1992) *Protein Sci.*, **1**, 736–751.

Jacobsen, N.E., Abadi, N., Sliwkowski, M.X., Reilly, D., Skelton, N.J. and Fairbrother, W.J. (1996) *Biochemistry*, **35**, 3402–3417.

James, T.L. (1994) *Methods Enzymol.*, **239**, 416–439.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986) *Numerical Recipes*, Press Syndicate of the University of Cambridge, Cambridge, U.K., pp. 202–203.

Torda, A.E., Scheek, R.M. and van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY, U.S.A.

Zwillinger, D. (Ed.) (1996) *CRC Standard Mathematical Tables and Formulae*, CRC Press, Boca Raton, FL, U.S.A., pp. 626–627.