Journal of Biomolecular NMR, **19:** 321–329, 2001.
KLUWER/ESCOM

SANE (Structure Assisted NOE Evaluation): An automated model-based approach for NOE assignment

Brendan M. Duggan, Glen B. Legge, H. Jane Dyson & Peter E. Wright*
Department of Molecular Biology MB2 and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd, La Jolla, CA 92037, U.S.A.

Received 10 November 2000; Accepted 17 January 2001

Key words: ambiguous NOE, LFA-1, NOE assignment, structure determination

Abstract

A reliable automated approach for assignment of NOESY spectra would allow more rapid determination of protein structures by NMR. In this paper we describe a semi-automated procedure for complete NOESY assignment (SANE, Structure Assisted NOE Evaluation), coupled to an iterative procedure for NMR structure determination where the user is directly involved. Our method is similar to ARIA [Nilges et al. (1997) *J. Mol. Biol.*, **269**, 408–422], but is compatible with the molecular dynamics suites AMBER and DYANA. The method is ideal for systems where an initial model or crystal structure is available, but has also been used successfully for *ab initio* structure determination. Use of this semi-automated iterative approach assists in the identification of errors in the NOE assignments to short-cut the path to an NMR solution structure.

Introduction

NMR structure determination has rapidly evolved into a technique capable of producing high-resolution structures of molecules of biological importance. Determining a high-resolution structure by NMR requires assigning the resonances of all possible nuclei within the molecule, and using these assignments to identify pairs of protons that give rise to the cross peaks in NOESY spectra. Structure calculations are performed using distance restraints derived from the assigned NOESY cross peaks. In practice, not all NOESY cross peaks can be assigned initially, largely due to resonance overlap. Examination of the initial calculated structures ideally allows more NOESY cross peaks to be assigned, and the whole process can be iterated until acceptable structures are obtained. The entire NMR structure determination process requires careful attention to detail and the determination of an accurate, precise structure using this largely manual approach is time-consuming.

Several automated NOESY cross peak identification protocols have been integrated with structure calculation software (Nilges and O'Donoghue, 1998; Mumenthaler and Braun, 1995). These protocols determine possible assignments by comparing the chemical shifts of the cross peak against a list of the resonance assignments. In some cases, a unique assignment will be found, but in many more cases, multiple assignments will be possible. The protocol NOAH (Mumenthaler and Braun, 1995), which has been implemented in DYANA (Güntert et al., 1997), circumvents this problem either by temporarily ignoring the cross peak or by creating a distance restraint for every possible assignment, even if many of these assignments are incorrect. Another automated NOESY assignment procedure is ARIA (Nilges et al., 1997), which is integrated into Xplor (Brünger, 1992) and treats the problem of multiple possible assignments by creating ambiguous distance restraints. To reduce the number of possible assignments and thus the number of ambiguous distance restraints, ARIA uses an NOE contribution filter to discard possible assignments that

*To whom correspondence should be addressed.

contribute little to the observed NOESY cross peak volume.

NOAH and ARIA automate the iterative nature of the NMR structure determination process, performing successive rounds of restraint generation, structure calculation and violation analysis, with little or no input from the user. Both protocols automatically remove restraints that are consistently violated between successive rounds of structure calculation. The present paper describes an alternative protocol in which ambiguous distance restraints are generated for cross peaks with multiple possible assignments, but unlike NOAH and ARIA, the user is directly involved in violation analysis after each round of structure calculation. This approach has the advantage that user intervention throughout the structure calculation provides input that can help to circumvent erroneous local structures and reduce the number of iterations required to reach acceptable structures. In addition, the program described herein incorporates a distance filter that is based on an initial search model structure, which may be an X-ray structure, an ensemble of solution structures, or even a homology-modeled structure. This approach has been described previously (Breg et al., 1990); here we incorporate it as part of a suite of filters designed to iterate rapidly and completely to an accurately assigned NOE cross peak list, including both unambiguous and ambiguous NOEs, which can be used to calculate accurate and precise solution structures.

Methods

The perl program SANE (Structure Assisted NOE Evaluation) is designed to allow the rapid generation of an accurate set of DYANA or AMBER distance restraints for solution structure determination. To minimize the problem of multiple possible assignments SANE makes use of existing partial assignments, the average distance between protons in one or more structures, relative NOE contributions calculated from the structures, and the expected secondary structure as assignment filters. Any combination of these filters can be used. The code, documentation and sample files are available by contacting the authors.

Input data

SANE works with cross peak lists from both Felix (MSI) and NMRView (Johnson and Blevins, 1994) and is able to analyze 2D, 3D and 4D NOESY spectra including HSQC-NOESY-HSQC spectra, aromatic NOESY spectra and shared time CN NOESY spectra. It will account for 'aliased' chemical shifts (obtained using the TPPI-States method of quadrature detection) but cannot, at present, work with 'folded' spectra (obtained using the TPPI method of quadrature detection). The input files required by SANE are: (i) A parameter file that defines the dimensionality of the spectrum, how the matrix dimensions correspond to the experimental dimensions, chemical shift tolerances, which filters to apply, cutoff values for the filters, how to convert volumes to distance bins, how many possible assignments to accept when writing ambiguous distance restraints and the names of other input files. (ii) A cross peak list generated by either Felix or NMRView containing the positions of the cross peaks and their volumes. Felix cross peak files do not contain volume information, and therefore a volume file must also be supplied when working with Felix. (iii) An assignment list that contains a line for every assignment with residue number, residue name, atom name and chemical shift. For NMRView data the standard 'ppm.out' file can be used, but because this omits the residue type the sequence file used with NMRView must also be supplied. (iv) A file that maps the names used for the cross peaks with the names of their equivalent atoms in the PDB files. This file is also used to generate AMBER format restraints from UPL files. The map file allows great flexibility in the naming of atoms. Problems such as conversion between the HB1/2 and the HB2/3 nomenclature, and discriminating between geminal protons that have been stereo-specifically assigned and those that have not, can both be made much simpler through use of the map file.

Optional input files for SANE are: (i) One or more PDB structures. The PDB structures are required for the distance and NOE contribution filtering, but at the start of a structure determination project, when structures are not yet available, SANE can still determine assignments based on chemical shifts, existing assignments and the secondary structure. (ii) An edited list of the chemical shifts deposited with the BioMagResBank (BMRB). This list can be used to specify likely chemical shift ranges for resonances that have not yet been assigned. The file is a slightly modified version of the statistics on the BMRB web site and is available from the authors.

Output data

The output files created by SANE are: (i) A summary of the analysis for each cross peak. The cross peak number, its chemical shifts, volume, distance bin and existing assignments are listed, followed by the possible assignments for the cross peak after each filtering step. Warnings about potential errors (see later) and the final distance restraint to be created, either unique or ambiguous, are printed at the end of each cross peak entry. (ii) A list of distance restraints including the ambiguous restraints. For use with DYANA the ambiguous restraints must be removed. A comment composed of the cross peak number, the experiment number and the nature of the restraint is attached to each restraint. The nature of the restraint is one of Preassigned (the cross peak was completely assigned before starting SANE), Unique (SANE found a unique assignment), Single Ambig (one side of the restraint is ambiguous) or Double Ambig (both sides of the restraint are ambiguous). If the restraint is Preassigned or Unique then the shortest and average distances in the ensemble are also printed in the comment. (iii) A file defining the groups of atoms in the ambiguous restraints. This file is used when creating ambiguous distance restraints for use with AMBER. Optional output files are: (i) A cross peak file containing all the pre-existing assignments with the additional assignments found by SANE. This file can be read back into Felix or NMRView so that the new assignments need not be manually entered. (ii) A subset of the cross peak file containing only those cross peaks for which it was not possible to obtain assignments. This file can be run through SANE again with different parameters in an attempt to assign 'difficult' cross peaks.

General description of the SANE method

To create distance restraints, SANE considers each cross peak in turn and generates a list of possible assignments for each dimension. It then applies a series of filters, described in detail below, to reduce the number of possible assignments (Figure 1). If only one possible assignment remains after applying the filters, then a unique restraint is generated; an ambiguous distance restraint is written if there is more than one possibility.

Chemical shift filter

Initial identification of possible cross peak assignments is performed by comparing the chemical shift

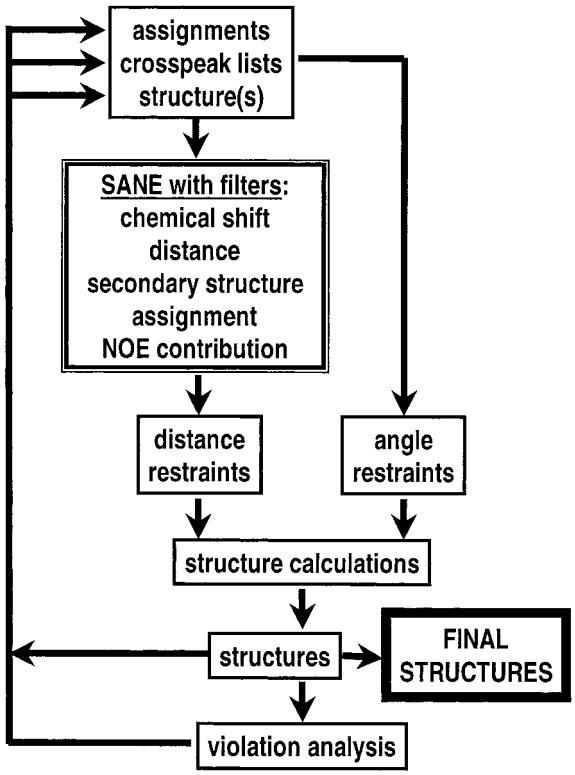

Figure 1. Flow of information through SANE during structure refinement.

in each dimension against the list of resonance assignments. A different chemical shift tolerance can be used for each dimension and the tolerances used are generally slightly larger than the digital resolution. Later examination of the NOESY spectra and the assignments identified by SANE may allow tighter tolerances, which will reduce the number of ambiguities in the assignments. The possible assignments for each dimension are then combined to produce a cross peak assignment consistent with the nature of the spectrum. For example, in a ^{15}N NOESY-HSQC the assignment for the t_3 dimension must be a proton, which is attached to a nitrogen atom, and that nitrogen atom must be one of the possible assignments for the t_2 dimension.

Distance filter

A considerable saving in time and effort results if an initial search model is used at an early stage of the structure calculation. A search model may be the crystal or solution structure of the same or a closely related molecule, or a predicted homology model. Possible assignments can be eliminated if their distance in the

search model is greater than a specified distance cutoff. When a single model, such as a crystal structure, is used as the search model, then only one distance is measured, but when an ensemble of NMR structures is used the mean distance in the ensemble is calculated and compared against the distance cutoff. If a possible assignment involves a pseudoatom (Wüthrich et al., 1983), then the distances to each proton in the pseudoatom are measured and averaged.

Assignment filter

After distance filtering, the possible assignments identified by SANE are compared with any assignments that the user may have already given that cross peak. If the user's assignments are found in the list of possible assignments then only the user's assignments are retained. If none of the possible assignments include the user's assignments, then the user's assignments are ignored and a warning is issued. The assignment filter is particularly useful in crowded regions of ^{13}C NOESY-HSQC spectra, for example, where it is possible to assign the t_2 and t_3 dimensions of a cross peak but not the t_1.

Secondary structure filter

At the start of an NMR structure determination process the secondary structure is usually known and can be used to eliminate some potential NOE assignments. SANE allows the user to define the secondary structure elements in the protein and list the NOEs that one would not expect to observe within those elements. For example, NOEs between residues more than five residues apart in the same α helix or β strand will not be seen and can be eliminated as possible assignments. It is also possible to define potential assignments that are unlikely to be observed except as a result of spin diffusion, such as $d_{N\alpha}(i,i+1)$ NOEs, and eliminate these possibilities from all parts of the protein.

NOE contribution filter

The NOE contribution filter is an implementation of the procedure described by Nilges et al. (1997). For each possible assignment the minimum distance in the search model, or models, is measured. The r^{-6} weighted fractional contribution to cross peak volume (C_i, Equation 1) is then determined using the shortest distance for each possible assignment within the ensemble.

$$C_i = r_i^{-6} \Big/ \left(\sum_{i=1}^{N} r_i^{-6} \right) \quad (1)$$

The contributions are ordered from largest to smallest and summed until a user-defined contribution cutoff is exceeded. Any remaining possible assignments are discarded.

Structure calculation and NOE violation analysis

SANE was developed initially for the structure determination of the leukocyte function-associated antigen-1 (LFA-1) I-domain, using the programs DYANA (Güntert et al., 1997) and AMBER (Case et al., 1999). Preparation of the ^{15}N and ^{15}N/^{13}C isotopically labeled LFA-1 I-domain, collection and processing of the NMR spectra, resonance assignment and structure calculation protocols have been described previously (Legge et al., 1999; Kriwacki et al., 2000). Torsion angle restraints for φ and ψ were derived from the secondary shifts of backbone assignments. Distance restraints were generated from cross peaks in 3D ^{1}H-^{15}N NOESY-HSQC, 3D ^{1}H-^{13}C NOESY-HSQC and 3D ^{1}H-^{15}N HSQC-NOESY-HSQC spectra (Zhang et al., 1994, 1997). The volume-distance calibration was determined manually using unambiguous cross peaks in regions of regular secondary structure. Two-hundred structures were generated by DYANA using the standard protocol (Legge et al., 1999) and unique restraints identified by SANE. The 100 structures with lowest target functions were subjected to two successive 20 ps rounds of simulated annealing using AMBER, as previously described (Legge et al., 1999). Structures with the lowest AMBER energies were selected for analysis of the restraint violations and stereochemical quality.

Results and discussion

Application of SANE in a structure calculation

The calculation of the solution structure of the Inserted (I-)domain of LFA-1 (Legge et al., 1999) provides an excellent example of the application of the SANE approach to NOE cross peak assignment. Examination of the secondary structure, derived from chemical shift indices and medium-range and cross-β-strand NOEs, indicated that the existing crystal structure (PDB# 1LFA) (Qu and Leahy, 1995) would provide a reasonable initial search model for the solution structure. Prior to running the SANE program, only these few secondary structure-based NOEs were assigned, together with a virtually complete set of sequence-specific backbone and side chain assignments (Kriwacki et al., 2000). The remaining NOEs

Table 1. Filtering parameters, number of restraints and RMSDs during LFA-1 I-domain structure calculations

Round	**1**	**2**	**3**	**4**
^{13}C NOESY-HSQC tolerances (ppm)				
t_1 ^{1}H	0.15	0.08	0.08	0.08
t_2 ^{13}C	0.7	0.6	0.5	0.5
t_3 ^{1}H	0.08	0.06	0.06	0.06
^{15}N NOESY-HSQC tolerances (ppm)				
t_1 ^{1}H	0.07	0.06	0.06	0.06
t_2 ^{15}N	0.2	0.15	0.15	0.15
t_3 ^{1}H	0.06	0.05	0.05	0.05
^{15}N HSQC-NOESY-HSQC tolerances (ppm)				
t_1 ^{15}N	0.2	0.15	0.15	0.15
t_2 ^{15}N	0.2	0.15	0.15	0.15
t_3 ^{1}H	0.06	0.05	0.05	0.05
Search model	X-ray	NMR	NMR	NMR
Distance cutoff (Å)	10	7	7	7
Contribution cutoff	none	none	0.95	0.95 and 0.9[a]
Number of peaks[b]	9414	9519	9537	9537
Number of distance restraints				
Intra residue	62	355	393	233
Sequential	231	623	733	677
Medium (2,3,4)	173	507	528	585
Long (5 or more)	218	659	748	804
Total unambiguous	604	2144	2402	2299
Ambiguous[c]	7223	6082	4004	3083
Torsion restraints	275	238	208	285
Total restraints	8182	8464	6614	5656
Percentage of residues in regions of ϕ-ψ space (Laskowski et al., 1996)				
Core	48.4	78.1	78.0	82.9
Allowed	38.1	18.7	19.3	15.4
Generously allowed	10.1	1.6	1.3	1.4
Disallowed	3.4	1.6	1.4	0.4
RMSD (Å)[d]	1.24	0.59	0.44	0.29[e]
Distance violations[f]	336	3452	34	5 (none > 0.4Å)

[a] The contribution cutoff was 0.90 for the ^{15}N NOESY-HSQC and ^{15}N HSQC-NOESY-HSQC experiments and 0.95 for the ^{13}C NOESY-HSQC.

[b] 'Total number of peaks' includes duplicates, but 'total restraints' in the DYANA calculations does not, since the program DYANA removes duplicate restraints.

[c] Ambiguous restraints were not included in the DYANA calculation.

[d] RMSD from the mean of 10 structures with lowest AMBER energies for rounds **1**–**3**. Fitted and calculated on the N, C^{α} and C' atoms in residues 129–161 and 164–306.

[e] Ensemble 4 contains the final 22 structures from Legge et al. (1999).

[f] Average number of distance violations > 0.1 Å per structure.

necessary to define the solution structure to high accuracy and precision were generated iteratively through rounds of SANE, followed by structure calculations and violation analysis.

The changes in the assignment criteria for the NOESY spectra and violation analysis, as well as the chemical shift tolerances and filtering parameters used in a representative set of four iterations of the LFA-1 structure calculation are listed in Table 1. The families of structures derived for each of the conditions (**1**–**4**) listed in Table 1 are shown in Figure 2. The restraints used in the initial round of the LFA-1 structure calculations were generated using the crystal structure as a search model, but in the later rounds the intermediate NMR structures were used to minimize bias. The crystal structure is a good model of the structure in solution; the global fold of LFA-1 was reproduced from the first round of structure calculations, despite the large proportion of ambiguous restraints, which arose as a consequence of the large chemical shift tolerances and distance cutoffs. These large tolerances were used initially to avoid potential local structure bias, where the crystal structure model might differ in detail from that of the final solution structure. Thus, although the initial structures had the correct fold, the RMSD between structures was quite high (Table 1), especially for the C-terminal helix (Figure 2). Short distance and NOE contribution cutoffs could not be used to generate restraints initially, because the 'true' structure in solution is not known at this stage; restraints generated in this way may introduce significant bias from the model. Tighter chemical shift tolerances were introduced in later rounds (**2**–**4**), but the validity of tight chemical shift tolerances is limited by the digital resolution of the spectrum and the quality of both the sequence specific assignments and the peak picking of the NOESY spectrum. Where required, manual repositioning of the automated peak pick center, in combination with selective manual assignment of two of the three dimensions, significantly reduce the ambiguity of the data. In the 3D ^{13}C edited NOESY-HSQC in particular, assignment of cross peaks in the t_2 and t_3 dimensions can be achieved by using line shape to discriminate between assignments that lie within the chemical shift tolerance.

In the course of the structure calculations, the contribution cutoff was reduced to 0.95 for the ^{13}C edited NOESY spectrum and to 0.90 for the ^{15}N edited NOESY spectra (Table 1, rounds **3** and **4**). A more aggressive contribution cutoff, e.g. 0.80, was not used because in some cases this would have discarded possible assignments that would have been retained using manual methods, causing the remaining assignments to be given an overly tight upper distance limit. Table 2 lists the distance above which a possible assignment will be eliminated when a particular contribution cutoff is applied to two possible assignments.

Table 2. Distances excluded by the contribution filter with various shortest distances and contribution cutoffs. Calculated assuming two possible assignments

Shortest distance (Å)	Contribution cutoff				
	0.999	0.99	0.95	0.90	0.80
2.2	6.96	4.73	3.59	3.17	2.77
2.7	8.54	5.81	4.41	3.89	3.40
3.3	10.43	7.10	5.39	4.76	4.16
4.0	12.65	8.60	6.53	5.77	5.04
5.0	15.81	10.75	8.17	7.21	6.30

The point at which the user chooses to halt the cycle shown in Figure 1 and proceed to final statistical analysis and publication of final structures is one of individual choice. In the case of LFA-1, we chose the criterion that all NOE violations should be less than 0.4 Å. This criterion was met in the calculation which we have labeled **4**, and, since the RMSD for structured regions of the protein was acceptably low (0.29 Å), we halted further iterations of the cycle for LFA-1 and proceeded to publication (Legge et al., 1999).

Comparison of SANE with ARIA

In both ARIA and SANE the contribution filter uses the shortest distance in the ensemble. This approach tends to favor assignments between residues distant in the amino acid sequence over assignments between neighboring residues. This is because atoms in nearby residues are separated by fewer bonds than atoms in distant residues, and on average have less variance in their separation in an ensemble of calculated structures. To help eliminate possible bias caused by contribution filtering on the shortest distance of the ensemble, discarded possibilities that have the shortest mean distance in the ensemble were flagged by SANE to allow manual examination of the cross peak and its possible assignments.

Decreasing the distance cutoffs (round **2**) and the introduction of a contribution filter (rounds **3** and **4**) significantly reduce the number of possible assignments for each NOESY cross peak, thereby decreasing the ambiguity of the data (Table 1). However,

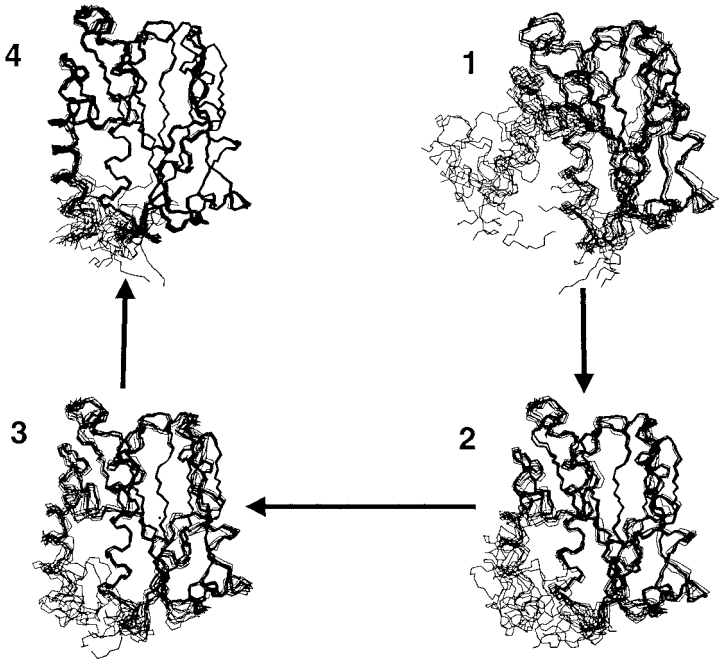

Figure 2. Ensembles of LFA-1 I domain structures from the four rounds of calculations specified in Table 1. Each of ensembles **1**–**3** contains 10 structures superimposed on the N, C^{α} and C' atoms of residues 129–161 and 164–306. Ensemble **4** contains the final 22 structures published by Legge et al. (1999). Numbers refer to the refinement conditions in columns **1**–**4** of Table 1.

this should only be done once violation analysis has shown that the calculated structures are in reasonable agreement with the NOESY assignments.

The major differences between SANE and ARIA are the use of a variable distance filter to eliminate possible assignments and a manual approach to violation analysis. Using a variable distance filter has the advantage that all obviously incorrect possibilities are removed before they enter the structure calculations. An NOE contribution filter alone will not remove incorrect assignments when only one possible assignment is found. The contribution filter compares each possible assignment against the others, whereas the distance filter is applied to each potential assignment individually and is not affected by the other possibilities. Thus, the distance filter is an absolute criterion for acceptance of a possible assignment and we believe it to be a useful safeguard.

The addition of a distance filter provides a considerable time saving in the initial stages of the structure calculation because 'improbable' NOEs that would not otherwise be picked up are removed at relatively early stages. For example, a total of 418 peaks were removed from the restraint list because of the distance filter alone. These were peaks that were spurious, for one of the reasons described in a later section, but, because they only had a single possibility for assignment, they would not have been removed by the contribution filter. They would have been violated in the calculated structures and could possibly have biased the structures incorrectly.

SANE retains duplicated restraints

When distance restraints are created automatically from NOESY cross peaks it is likely that there will be some duplication of the restraints. Duplication can arise from 'mirror' peaks on opposite sides of the diagonal or when using more than one NOESY spectrum. SANE does, however, remove restraints derived from diagonal cross peaks and from cross peaks between geminal protons. Retention of duplicated restraints allows the spectral and peak identifiers for each NOE to be retained throughout all of the calculation cycles, enabling rapid identification and assessment of violated NOES after each round of calculation.

Causes of consistent violations

Violation analysis after each round of LFA-1 structure calculations identified a particular set of problems with the NMR assignments and the cross peak lists.

Similar problems were encountered using SANE in several other structure calculation projects, suggesting that these problems may be intrinsic to this approach for NMR structure calculations. The problems were generally encountered as follows: (i) NOEs to the water resonance being assigned as intra-molecular NOEs; (ii) missing assignments including hydroxyl protons; (iii) mis-assignments; (iv) mis-calibration of the NOESY cross peak volume. To cope with these problems several procedures were implemented in SANE as described below.

NOEs and apparent NOEs to the solvent

NOEs involving the bulk water resonance were handled by defining the chemical shift of the water and a tolerance, then ignoring all cross peaks that fall within this range. Although several real NOEs may lie below the water resonance, removing all of the peaks along this strip is justified due to the high likelihood of an incorrect assignment. The absence of genuine restraints derived from cross peaks near the water resonance was found less troublesome than the introduction of incorrect restraints.

Missing assignments

Missing assignments, including unassigned hydroxyl resonances, were approached in one of two ways. In the first approach, missing assignments were initially ignored and then later identified by the analysis of persistent violations in the calculated structures. This approach was followed for serine and threonine hydroxyls in the LFA-1 structure calculations. In a second approach, useful when resonances other than those of hydroxyls are missing from the assignment list (e.g.: arginine and lysine side chain resonances), SANE uses the mean chemical shift of this type of nucleus from the BMRB as an approximation. The user-defined chemical shift tolerance is replaced for these 'assignments' with the standard deviation of the BMRB shifts multiplied by a user-defined constant. We found that nuclei with fewer than 50 reported chemical shifts have standard deviations too large for reliable use and these were therefore removed. It is also important to make a judicious choice of which unassigned resonances to use the mean BMRB shift for. Nuclei that have a narrow distribution of chemical shifts, such as side-chain methylenes, are the best choice.

An alternative approach for coping with missing assignments is to calculate the chemical shifts from the ensemble of NMR structures (Hare and Wagner, 1999). This approach is only as reliable as the structures used to calculate the chemical shifts of the unassigned nuclei, but will certainly become much more powerful towards the end of the structure determination process when the structures are well converged. The calculation of methylene chemical shifts was reported to be particularly accurate and these are often difficult resonances to assign. We envisage that using the mean BMRB shifts for select nuclei will be useful at the beginning of structure calculations, but once the structures begin to converge then calculation of the unknown chemical shifts will be a better approach.

Mis-assignments

To assist with the identification of mis-assignments, restraint violations were sorted according to the number of times the restraint is violated and the mean size of that violation. This prioritizes the restraint violations for the user to examine manually. SANE labels each restraint with a unique identifier that is carried through the AMBER structure calculations and allows the user to trace a restraint violation back to the spectrum and cross peak from which it was derived. After a round of structure calculations, examination of the cross peak associated with a restraint violation often indicated that the cross peak footprint or chemical shift assignment should be adjusted. Occasionally the chemical shift in the assignment table differed slightly between spectra, in which case a separate chemical shift assignment table was used for each spectrum.

Mis-calibration of the NOESY cross peak volume

During structure calculations, many restraint violations are caused by miscalibration of the cross peak volume and by cross peak overlap. While the former problem is left to the user to correct, the latter problem is handled in SANE by making a list of those cross peaks whose distance bin should be adjusted upwards to the next higher distance bin. Some NOEs were discarded from the intraresidue and sequential categories because they were over-weighted due to a too-high intensity caused by spin diffusion or overlap with neighboring peaks. This accounts for the lower number of (total) unambiguous NOEs between rounds 3 and 4 (Table 1).

As automated methods are applied to NMR structure determination it becomes increasingly obvious that the quality of the NOESY cross peak list is crucial. In the course of the LFA-1 structure determination, the peak lists of all three NOESY spectra were edited manually on several occasions. Manual peak picking

and peak editing of NOESY spectra is a tedious and time-consuming chore. If NMR structure determination is to become more rapid then more sophisticated and reliable automated peak picking algorithms must be developed. In the absence of reliable automated peak-picking protocols, manual analysis of restraint violations is almost mandatory.

Conclusions

The perl program SANE has been developed to automate NOESY spectrum assignment, while allowing the user to control the violation analysis and subsequent input into successive rounds of structure calculation. It was designed to be a semi-automated iterative method, where structure calculations and NOE cross peak assignment progress in parallel.

SANE may be readily applied to any protein for which a closely related structure, or homology model, is available, as was done for the LFA-1 I domain. (Legge et al., 1999). It can also be used in the *de novo* calculation of protein structures for which no initial model is available, such as NAPc2. (Duggan et al., 1999). Several other structures of proteins and domains, ranging in size from 100 to over 200 residues, have been calculated in our laboratory using the SANE method, and are at present nearing completion. SANE uses all available information; namely chemical shift assignments, secondary structure, tertiary structure, existing cross peak assignments, as well as input from the BMRB database, to determine unique NOESY cross peak assignments iteratively as the structure is calculated.

Acknowledgements

We thank John Chung and Gerard Kroon for helpful comments on NMR experiments, David Case for helpful comments, Jaime Pascual, Brian Hudson, Micah Gearhart, Bin Xia, Annette Atkins, Roberto De Guzman, Ouwen Zhang and many other members of the Wright and Dyson labs for helpful suggestions and for testing the SANE method. This work was supported by grants GM36643 (P.E.W.) and GM43238 (H.J.D.) from the National Institutes of Health.

References

Breg, J.N., van Opheusden, J.H.J., Burgering, M.J.M., Boelens, R. and Kaptein, R. (1990) *Nature*, **346**, 586–589.

Brünger, A.T. (1992) *X-PLOR 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1999) *AMBER 6*, University of California, San Francisco, CA.

Duggan, B.M., Dyson, H.J. and Wright, P.E. (1999) *Eur. J. Biochem.*, **265**, 539–548.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Hare, B.J. and Wagner, G. (1999) *J. Biomol. NMR*, **15**, 103–113.

Johnson, B.A. and Blevins, R.A. (1994) *J. Chem. Phys.*, **29**, 1012–1014.

Kriwacki, R.W., Legge, G.B., Hommel, U., Ramage, P., Chung, J., Tennant, L.L., Wright, P.E. and Dyson, H.J. (2000) *J. Biomol. NMR*, **16**, 271–272.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR*, **8**, 477–486.

Legge, G.B., Kriwacki, R.W., Chung, J., Hommel, U., Ramage, P., Case, D.A., Dyson, H.J. and Wright, P.E. (1999) *J. Mol. Biol.*, **295**, 1251–1264.

Mumenthaler, C. and Braun, W. (1995) *J. Mol. Biol.*, **254**, 465–480.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.

Nilges, M. and O'Donoghue, S.I. (1998) *Prog. NMR Spectrosc.*, **32**, 107–139.

Qu, A. and Leahy, D.J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 10277–10281.

Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Zhang, O., Forman-Kay, J.D., Shortle, D. and Kay, L.E. (1997) *J. Biomol. NMR*, **9**, 181–200.

Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR*, **4**, 845–858.