Journal of Biomolecular NMR, 2 (1992) 373–388
ESCOM

J-Bio NMR 068

A systematic comparison of three structure determination methods from NMR data: Dependence upon quality and quantity of data

Yajun Liu, Daqing Zhao, Russ Altman and Oleg Jardetzky*

Stanford Magnetic Resonance Laboratory, Stanford University, Stanford, CA 94305-5055, U.S.A.

Received 23 March 1992
Accepted 13 May 1992

Keywords: Distance geometry; Optimized filtering; Kalman filter; Simulated annealing; NMR protein structures

SUMMARY

We have systematically examined how the quality of NMR protein structures depends on (1) the number of NOE distance constraints, (2) their assumed precision, (3) the method of structure calculation and (4) the size of the protein. The test sets of distance constraints have been derived from the crystal structures of crambin (5 kDa) and staphylococcal nuclease (17 kDa). Three methods of structure calculation have been compared: Distance Geometry (DGEOM), Restrained Molecular Dynamics (XPLOR) and the Double Iterated Kalman Filter (DIKF). All three methods can reproduce the general features of the starting structure under all conditions tested. In many instances the apparent precision of the calculated structure (as measured by the RMS dispersion from the average) is greater than its accuracy (as measured by the RMS deviation of the average structure from the starting crystal structure). The global RMS deviations from the reference structures decrease exponentially as the number of constraints is increased, and after using about 30% of all potential constraints, the errors asymptotically approach a limiting value. Increasing the assumed precision of the constraints has the same qualitative effect as increasing the number of constraints. For comparable numbers of constraints/residue, the precision of the calculated structure is less for the larger than for the smaller protein, regardless of the method of calculation. The accuracy of the average structure calculated by Restrained Molecular Dynamics is greater than that of structures obtained by purely geometric methods (DGEOM and DIKF).

INTRODUCTION

High-resolution NMR has emerged as a useful method for the definition of 3D solution structures of proteins (Wüthrich, 1986, Jardetzky and Lane, 1988; Clore and Gronenborn, 1989; Gronenborn et al., 1989). A prerequisite for the use of NMR data for this purpose is prior knowledge of the primary sequence of the protein, i.e. of all constraints implied by chemical bond lengths,

* To whom correspondence should be addressed.

0925-2738 $ 5.00

bond angles and van der Waals radii. The primary source of additional constraints derived from NMR are the short-range (< 6 Å) interproton distances estimated from Nuclear Overhauser Enhancement (NOE) measurements. For smaller molecules, it is also possible to obtain ranges of allowed backbone torsion angles from vicinal spin–spin coupling constants (Bystrov, 1976). In cases, where the secondary structure is well defined by a qualitative analysis of the assignment data, backbone NH exchange data can be used to assign additional hydrogen-bonding constraints. In general, the sum total of the available constraints represents only a subset of the total number of constraints which would be necessary to uniquely define the 3D structure. Consequently, a large conformational space has to be searched in order to locate a global minimum – or minima – consistent with all available stereochemical and NMR constraints.

In recent years a number of different computational methods for the calculation of protein solution structures from NMR data has been developed. In principle these fall into three classes: (1) Distance Geometry (DG) (Kuntz et al., 1979, 1989), (2) Restrained Molecular Dynamics (RMD) (van Gunsteren et al., 1983; Scheek et al., 1989), and (3) Double-Iterated Kalman Filter (DIKF) or optimal filtering (Altman and Jardetzky, 1986, 1989; Koehl et al., 1992). Procedures using two or more of these in combination, and also in combination with simulated annealing protocols (Nilges et al., 1988) or back-calculation of the spectra by solution of the relaxation matrix (Keepers and James, 1984; Lefèvre et al., 1987) are coming into increasing use. Given that both the interpretation of NMR data in terms of distance constraints and each method of structure calculation involve subjective judgments and specific assumptions at several steps, it is important to establish the exact conditions under which, and the extent to which, each method can reproduce known structures. Some studies along these lines have been reported (Braun and Go, 1985; Clore et al., 1987; Lichtarge et al., 1987; Altman et al., 1989; Pachter et al., 1990; Altman et al., 1992), showing that all known methods will correctly reproduce the topology of the 3D protein fold within reasonable error limits. The continuing quest for ever greater precision in reporting NMR protein structures makes it necessary however to carefully reexamine the dependence of each method on (1) the number and (2) the (always subjectively determined) precision of the input constraints and on (3) the size of the protein. It must be borne in mind that the subjective judgments on such matters as the precision of NMR data constraints, the shape and magnitude of energy potentials used in some methods (RMD), and inadequate protocols for sampling conformational space (Metzler et al., 1989) can easily lead to calculated structures that appear to be precise, but are in reality inaccurate (Jardetzky, 1991). The present paper reports a performance evaluation of three specific programs – DGEOM (Blaney et al., 1990), XPLOR (Brünger, 1990) and PROTEAN II (Altman et al., 1990) – in their dependence on the number and assumed quality of the constraints and on the size of the protein under investigation.

METHODS

Definition of the problem and calculational strategy

The basic questions to be asked are: (1) How closely can each of the existing methods of structure determination reproduce a known structure, given an identical set of input data and (2) How reproducible are the structures obtained in successive calculations – or how large is the spread of different structures compatible with the data? The first question, one of accuracy, is usually answered by comparing the root mean square deviations (RMSD) of all atomic positions between the

calculated and the starting structure. The second, one of precision, is answered by comparing the RMSD between different members of a calculated family of structures and their average structure. The global average RMSD is frequently taken as an overall measure of either accuracy or precision, but it must be borne in mind that it carries only a fraction of the information obtained from the test calculations: the global RMSD will not discriminate between a structure that is accurately reproduced except for one ill-defined segment, and a structure that is poorly reproduced on the average. These two types of results will generally have different causes, and to distinguish between them it is necessary to examine the entire RMSD profile for the test molecule. Given that NOE constraints are not uniformly distributed throughout a protein structure and that the evaluation of their precision involves subjective judgments, we use the global RMSD only as a rough guide, relying largely on segment-by-segment comparisons of the RMSD.

The test systems chosen were the crystal structures of crambin [~5 kDa, 46 residues, (Hendrichson and Teeter, 1981)] and staphylococcal nuclease [~17 kDa, 149 residues, (Loll and Lattman, 1989)]. Protons were added to the crystal structures using the HBUILD algorithm provided in the XPLOR package of programs. The data sets used for determination of structure were based on the full set of interproton distances less than 6.0 Å extracted from each crystal structure (4099 distances for crambin, and 17233 distances for staphylococcal nuclease). Since experimental NMR measurements cannot routinely distinguish between the two α-protons in the glycyl residue, the two protons in methylene groups, and the three protons in the methyl groups within the protein, redundant distance constraints which require this distinction were removed. The remaining 'pseudoatomic' distance constraints numbered 1960 for crambin and 6597 for staphylococcal nuclease (Wüthrich et al., 1983). These figures correspond to 42 and 44 NOE constraints/residue – numbers substantially larger than the 8–16 NOE constraints/residue typically obtained in NMR experiments. The constraints were divided into three classes: (1) those between 1.8 and 2.7 Å, (2) those between 2.7 and 4.1 Å, and (3) those between 4.1 and 6.0 Å. These classes can often be distinguished experimentally, based on the magnitude of the NOE, although this classification has been shown to be prone to a high degree of error (Madrid and Jardetzky, 1988; Madrid et al., 1989). For distance constraints involving pseudoatoms, appropriate corrections were made as described in an earlier study (Wüthrich, 1986).

The data sets used for the experiments testing the dependence of each structure determination method on quantity of data were created by randomly selecting different percentages of the total data sets as defined above. For crambin, data sets were constructed which contained 10%, 30%, 50%, 70% and 100% of the total data set (corresponding to 4, 13, 21, 30 and 43 constraints/residue). These will be referred to as the '10% data set' or '70% data set' for the remainder of this report. For staphylococcal nuclease, a 30% and 70% data set from the total were constructed.

In order to test the dependence of each method on the precision of the constraints, we constructed perturbed versions of the 30% and 70% data sets for crambin. In addition to the 'normal precision' data set, a 'precise' version of each data set was created with very tight bounds on the mean distances provided to each program. Similarly, an 'imprecise' data set was created with more generous bounds around the mean distances. Since the form of the input for each of these three programs is not exactly the same, we attempted to produce consistent data sets of each precision in the proper format, as described in the next section.

In order to test the effect of stereo-specific assignments and the use of pseudoatoms during these calculations, a comparison was also performed using full atomic representations as well as

pseudoatomic representations. This was done with the 30% and 70% data sets generated from crambin.

Use of PROTEAN

The DIKF calculations were carried out in a two step fashion: definition of gross conformational topology followed by refinement. The first step involves using the hierarchical model building program, PROTEAN1 (Carrara et al., 1990), in which starting structures are determined by a coarse systematic sampling of conformational space available to the secondary structures (which are assigned by analysis of the interproton distances between adjacent residues in the sequence). For crambin, the two helices (residues 7–19 and 23–30) were used and their locations were sampled at intervals of 0.1 Å for position and 10 degrees for orientation. For staphylococcal nuclease, the two anti-parallel β-sheets (residues 7–10, 13–19, 21–27, 29–35, 71–75, and 91–95) and the three helices (residues 54–68, 98–106, and 121–135) were used for the tertiary folding. The position was sampled at 0.5 Å and orientation at 10 degrees for β-sheets and 30 degrees for helices. The result of these calculations were large distributions of possible locations for each α-carbon in the backbone of these secondary structures, which was then refined using PROTEAN2 (Altman et al., 1990) with random placement of the backbone carbons not part of an identified secondary structure. The first round of refinement proceeded with NOE constraints between backbone atoms only (α-carbons and amide protons) using the results of PROTEAN1. The side chains were then added to the structure and the whole structure was subjected to further refinement with all the NOE constraints contained in the data set, as well as covalent bond, bond angle, and dihedral angle constraints. The PROTEAN2 program updated the structure for three iterations with each constraint, with a threshold of 0.1 standard deviation for termination of each iteration. After introducing all constraints, the cycle is repeated until each constrained distance in the structure is less than 0.5 standard deviation from the mean distance provided to the program as input.

For the DIKF, a distance constraint is represented as a mean and variance. For each of the three classes of constraint, the mean of the range of distances was taken: 2.25, 3.40, and 5.05, respectively. The variances for the 'normal precision' data sets, estimated from previously published validations of the DIKF and designed to reflect the experimental observation of uncertainty in the translation of NOE measurements as simple distances, were set to 0.20 $Å^2$, 0.49 $Å^2$, and 0.90 $Å^2$. The variances for the precise data set were 0.10 $Å^2$ and for the imprecise data set were made double the value of the normal precision data set.

Use of DGEOM

Since DGEOM uses only the bond lengths and angles from the starting conformation, the precise fold of the starting structure need only provide this information. In the two stages of conjugate gradient minimization of the error function, default convergence limits of 0.3 and 0.2 were used, along with weights on the sum of squares of the fourth dimensional coordinates set to 0.0 and 0.5 for the two stages, respectively. The final structures are accepted if all distance violations are less than 0.5 Å and all chiral violations are less than 0.5 Å. For the normal, precise, and imprecise data sets, the upper and lower bounds used for DGEOM were calculated from the variances used for the Kalman filter. Since the variance, V, of a uniform distribution from lower bound (l) to upper bound (u) is given by:

$$V = \frac{(u-l)^2}{12}$$

and the mean, m, is given by:

$$m = \frac{(u+l)}{2}$$

Then u and l can be derived: $l = m - \sqrt{3V}$ and $u = m + \sqrt{3V}$.

These values were used for the lower and upper bounds for DGEOM and XPLOR. They were constrained, however, by an absolute lower bound of 1.8 Å (by van der Waals considerations) and 6.0 Å (maximum distance used in creating data sets).

Use of XPLOR

In the XPLOR calculations the same mean distances were used as for the DIKF, with lower and upper bounds set to 1.8 and 6.0 Å (and adjusted for precise and imprecise data sets in the same way as for DGEOM). Random coils were generated as starting structures by assigning random values to the φ and ψ angles. These extended chains were folded and then refined by a four-stage simulated annealing protocol (Brünger et al., 1986; Brünger, 1990): (1) 100 cycles of restrained minimization to ensure acceptable geometry of the initial structure; (2) iteration of dynamics with a time step of 2 ps at 1000 K. Through the process, the van der Waals interaction is suppressed (vdW weight = 0.002) so that atoms can pass by and through each other easily. The bond angle and dihedral angle force constants were set to 0.5 and 0.2 times their equilibrium values. For each iteration, the force constants of the NOE constraints were increased by a factor of 1.5 until the calculation reached equilibrium; (3) the bond angle and dihedral angle force constants were set to their equilibrium values. The dynamics were performed with a time step of 1 ps, while cooling the temperature to 300 K (by 50 K steps), and the van der Waals force constants were raised to their equilibrium value by a factor of 1.5 per step; finally, (4) 300 cycles of restrained minimization were performed. The RMS deviation from ideality in bond lengths is 0.04 Å, while the RMS deviation from ideality for bond angles is 3.6°.

RESULTS AND DISCUSSION

Fifteen structures were calculated from each data set for DGEOM and XPLOR, and one structure by PROTEAN*. As a measure of accuracy, the atomic RMSDs between the averages of the calculated structures and the crystal structure (from which all distances were derived) are listed in Table 1. All values for atomic positions are reported in Ångstroms (Å) and values for dihedral angles in degrees. For the DGEOM and XPLOR calculations, the value reported in Table 1 is the distance of the average structure from the crystal. For DIKF, the value reported is the distance between the crystal and the mean locations as calculated by the program. As a measure of precision, the average atomic RMSDs between each calculated structure and the average structure are listed in Table 2. The average standard deviations of atomic positions are given for DIKF. The three methods reproduce the crystal structure under all testing conditions, with XPLOR consistently being closer to the crystal structure. As expected, the quality of the solutions improves as

---

*Three of the DGEOM structures generated with the 10% data set did not converge. The average atomic RMS difference between these three structures and the crystal structure was 6.4 ± 0.5. They were excluded from the remainder of the analysis.

more data are provided (i.e., moving from the 10% to the 100% data set). Figure 1 shows the relationship between RMS error from crystal and abundance of data for each of the three methods. It is apparent that the structural features of crambin become fairly well defined by all methods, even with the 10% data set, although the errors are substantially larger than with the 30% data set. The RMS of all heavy atoms for the 30% data set as compared with the 10% data set are improved from 2.9 to 1.7 Å, 2.5 to 1.5 Å, and 2.4 to 1.5 Å for the DGEOM, XPLOR, and DIKF, respectively. There is relatively less improvement for data sets of higher abundance.

This result is of major importance. It shows that (1) the general features of a small protein can be reproduced from NOE constraints, even with a sparse data set of about 4 NOE constraints/re-

TABLE I
ACCURACY OF STRUCTURE CALCULATION: RMS DIFFERENCE BETWEEN AVERAGE STRUCTURES AND THE CRYSTAL STRUCTURE

NOEs	DGEOM				XPLOR				DIKF			
	Atomic RMS		Angular RMS		Atomic RMS		Angular RMS		Atomic RMS		Angular RMS	
	Back-bone	Heavy atoms	BB	HA	BB	HA	BB	HA	BB	HA	BB	HA
Crambin												
Conventional constraints												
10%	2.0[a]	2.3[a]	31(10)	36(14)	2.0	2.3	21(6)	28(9)	2.0	2.4	39(10)	41(13)
30%	1.2	1.9	24(7)	24(10)	0.6	1.1	14(2)	19(3)	1.4	1.5	30(7)	35(5)
50%	1.1	1.3	20(7)	13(7)	0.4	0.6	14(2)	16(3)	0.9	1.4	26(8)	32(7)
70%	0.9	1.2	22(5)	11(9)	0.33	0.53	13(2)	16(3)	1.0	1.2	28(6)	34(7)
100%	0.6	0.8	16(1)	17(0)	0.31	0.46	12(2)	20(2)	0.7	1.1	27(5)	33(3)
Loose NOE constraints[b]												
30%	1.5	1.7	21(11)	23(5)	0.8	1.1	20(5)	21(8)	1.7	1.1	38(9)	34(10)
70%	1.2	1.4	20(8)	26(10)	0.5	0.8	15(2)	17(2)	1.1	1.5	39(4)	34(4)
Tight NOE constraints[c]												
30%	1.0	1.2	18(4)	22(1)	0.5	0.6	11(2)	18(2)	1.0	1.4	33(5)	37(4)
70%	0.7	0.9	10(3)	9(2)	0.34	0.47	14(2)	17(2)	0.8	1.2	20(3)	23(2)
Full atomic representation												
30%	1.2	1.6	16(7)	14(6)	0.7	1.0[d]	14(1)	15(2)	1.5	2.1	28(9)	31(8)
70%	1.0	1.3	19(1)	20(2)	0.6	0.9[d]	16(1)	16(1)	1.1	1.5	29(6)	34(6)
The effect of size: staphylococcal nuclease												
30%	1.6	2.3	33(15)	34(19)	0.9	1.3	21(4)	20(7)	2.3	3.0	34(24)	37(25)
70%	1.5	2.0	32(19)	33(18)	0.6	0.8	24(3)	23(5)	1.9	2.7	35(22)	37(27)

[a] For 12 converged structures only.
[b] Doubled the variances.
[c] Exact distance with variances of 0.1 Å^2.
[d] H atoms were added to the structures.

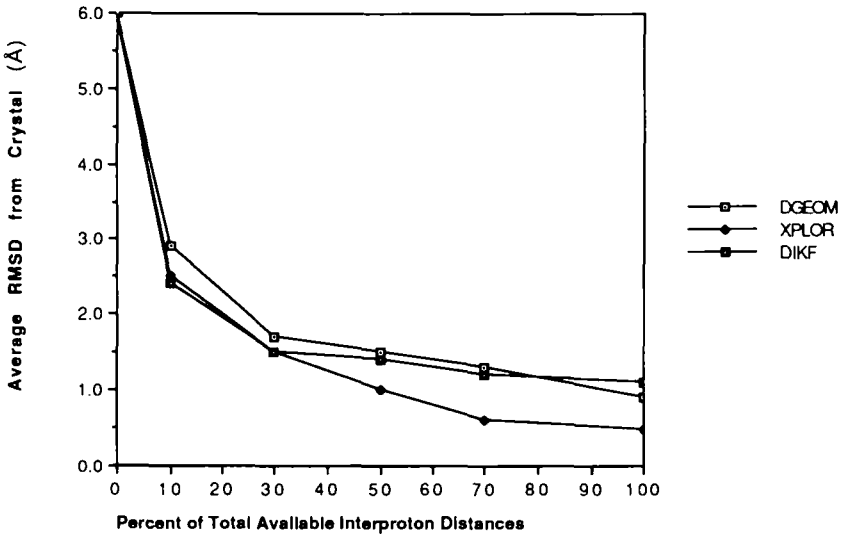

Fig. 1. The performance of the three methods of structure calculation (DGEOM, XPLOR and DIKF) is shown here as a function of the percent of total available interproton distances supplied by each method as compared to the crystal structure. As epxected, the errors decrease exponentially as more data are provided.

sidue, although the accuracy is only of the order of 2.5–3.0 Å, and (2) the limit to the accuracy attainable by any method of calculation is of the order of 0.7–1.0 Å, with relatively little gain in accuracy as the data set is increased above the 30% value of 12–14 NOE constraints/residue. In general, as has been noted before, it is *not* the average number of constraints/residue, but the distribution of strategic constraints within the structure that makes a major contribution to the accuracy of the structure (Jardetzky, 1991). Increasing the number of constraints has a greater effect when energy potentials are also used, as in XPLOR.

PROTEAN seems to do the best with poor quality data sets: in the case of the 10% data set as well as with the imprecise data sets PROTEAN produces the closest match to the crystal structure. Conversely, XPLOR seems to perform the best when precise, high abundance data sets are provided. PROTEAN and DGEOM perform equally as well with pseudoatomic and full atomic representations. XPLOR, however, performs much better with full atomic representations. This is not unexpected, as XPLOR relies on additional energy constraints between atoms which more faithfully model the behavior of real atoms.

Figure 2 shows that the assumed precision of the data is an important determinant of accuracy. It is more critical for high abundance data sets (such as 70%) than for lower abundance data sets (such as 30%). For all three methods, the use of imprecise constraints at 30% causes the results to be slightly less accurate. However, the same decrease in data precision has a relatively larger effect on the 70% data set. For example, XPLOR suffers only a 10% decrease in accuracy (from an error of 1.4 Å to 1.3 Å) when the imprecise 30% data set is compared to the regular 30% data set. However, it suffers a 100% decrease in accuracy (from 0.5 Å to 1.1 Å) with the imprecise 70% data set compared to the regular 70% data set. This probably reflects the fact that the 30% data set produces imprecisions that are roughly on the scale of the imprecision in the data, so the imprecise data do not substantially change the quality of the structure. On the other hand, the 70% data set im-

TABLE 2A
PRECISION OF STRUCTURE CALCULATION (ATOMIC POSITIONS): AVERAGE OF THE RMS DIFFERENCE BETWEEN INDIVIDUAL CALCULATED 'NMR' STRUCTURES AND THE AVERAGE 'NMR' STRUCTURES[a]

NOEs possible	DGEOM		XPLOR		DIKF	
	Atomic RMS		Atomic RMS		Atomic RMS	
	Backbone	Heavy atoms	Backbone	Heavy atoms	Backbone	Heavy atoms
Crambin						
Conventional constraints						
10%	1.6±0.8	2.0±1	1.1±0.5	1.3±0.6	0.6	0.7
30%	0.5±0.1	0.8±0.2	1.0±0.3	1.4±0.3	0.7	0.9
50%	0.5±0.1	0.8±0.1	0.5±0.1	0.6±0.2	0.8	1.0
70%	0.45±0.07	0.7±0.1	0.4±0.1	0.5±0.1	0.8	1.1
100%	0.37±0.05	0.5±0.1	0.4±0.1	0.45±0.1	0.5	0.7
Loose NOE constraints						
30%	0.7±0.1	1.3±0.8	1.1±0.3	1.5±0.3	0.7	0.8
70%	0.5±0.1	0.8±0.1	1.0±0.1	1.2±0.2	0.6	0.8
Tight NOE constraints						
30%	0.4±0.1	0.6±0.1	0.6±0.1	0.7±0.2	0.7	0.9
70%	0.36±0.1	0.65±0.1	0.3±0.1	0.4±0.1	0.6	0.7
Full atomic representation						
30%	0.4±0.1	0.7±0.1	0.6±0.1	0.7±0.1	0.5	0.6
70%	0.3±0.1	0.5±0.1	0.5±0.1	0.8±0.1	0.5	0.5
Staphylococcal nuclease						
30%	1.0±0.2	1.4±0.2	1.0±0.1	1.4±0.2	1.4	2.0
70%	0.6±0.1	1.0±0.1	0.5±2	0.9±0.2	0.8	1.1

[a] See footnotes to Table 1.

plies a relatively precise solution which is adversely affected by the relatively larger imprecisions in the data set. This phenomenon may not be generally appreciated, and demonstrates that absolute abundance of data may not produce high quality structures if the precision of the data does not increase with the abundance. Unfortunately, for many NMR experiments the precision of the data is difficult to assess, and conservative interpretation is often necessary. When precise data are available, however, it appears that a data set of 70% abundance can reproduce the results obtained with 100% of the data in less precise form. The comparison with the staphylococcal nuclease data sets (Fig. 3) shows that for the same percentage abundance of constraints, larger structures are less well defined than smaller ones. The 30% data sets for crambin produce results in global RMSD values within about 1.7 Å of the crystal, whereas the 30% data sets for staphylococcal nuclease produce RMSDs within approximately 2.6 Å of the crystal. This trend is seen for the 70%

TABLE 2B
PRECISION OF STRUCTURE CALCULATION (DIHEDRAL ANGLES): AVERAGE OF THE RMS DIFFERENCE BETWEEN INDIVIDUAL CALCULATED 'NMR' STRUCTURES AND THE AVERAGE 'NMR' STRUCTURES[a]

NOEs possible	DGEOM		XPLOR	
	Angular RMS		Angular RMS	
	φ	ψ	φ	ψ
Crambin				
Conventional constraints				
10%	30±7 (5)	29±4 (6)	23±6 (1)	22±5 (1)
30%	28±5 (6)	24±4 (6)	26±3 (2)	25±3 (1)
50%	30±4 (5)	30±4 (5)	18±4 (0)	19±4 (0)
70%	27±3 (5)	25±3 (6)	20±2 (0)	21±2 (0)
100%	23±4 (2)	23±4 (3)	17±3 (0)	18±4 (0)
Loose NOE constraints				
30%	30±5 (8)	29±5 (8)	18±3 (2)	19±4 (3)
70%	30±4 (5)	29±5 (6)	21±3 (1)	20±3 (1)
Tight NOE constraints				
30%	25±5 (5)	24±4 (5)	21±3 (0)	21±3 (0)
70%	210±6 (2)	18±4 (3)	15±5 (0)	15±5 (0)
Full atomic representation				
30%	23±4 (2)	17±3 (2)	21±3 (1)	22±2 (1)
70%	19±4 (1)	13±3 (0).	19±3 (0)	20±2 (0)
Staphylococcal nuclease				
30%	30±2 (21)	30±2 (22)	21±3 (3)	20±3 (4)
70%	25±2 (16)	23±2 (18)	19±3 (5)	19±3 (5)

[a] See footnotes to Table 1.

data sets as well. Although one might expect that similar numbers of constraints/residue should yield similar errors, this is not the case. As the molecular weight of a molecule increases and its overall dimensions increase, the number of interproton distances less than 6.0 Å will increase linearly with volume – or the number of residues – and hence with N. The total number of interproton distances increases roughly with N^2. Therefore, a 70% data set drawn from the short set of distances will contain relatively less implicit information about the total set of distances, and therefore will produce structures with less precision.

An obvious reason for the generally better results obtained for all data sets with the XPLOR program is the fact that the energies used by the program contain implicit information about the limitations on structural dihedral angles based on the energetics of their conformation. This is supported by Table 1, listing the errors in φ and ψ angles (as compared with the crystal) for each

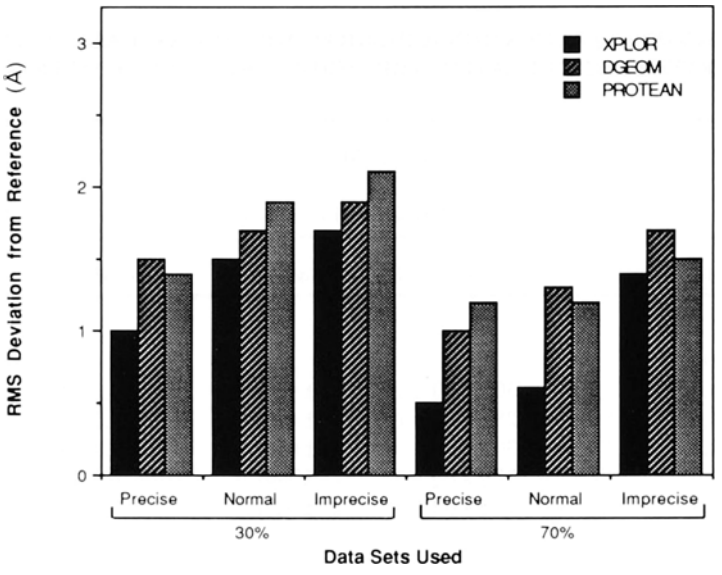

Fig. 2. The performance of the three methods for the 30% and 70% data sets is shown with respect to the precision of the data. XPLOR is consistently the most accurate (as discussed in text). All three methods respond as expected to increases in data imprecision, with increased errors.

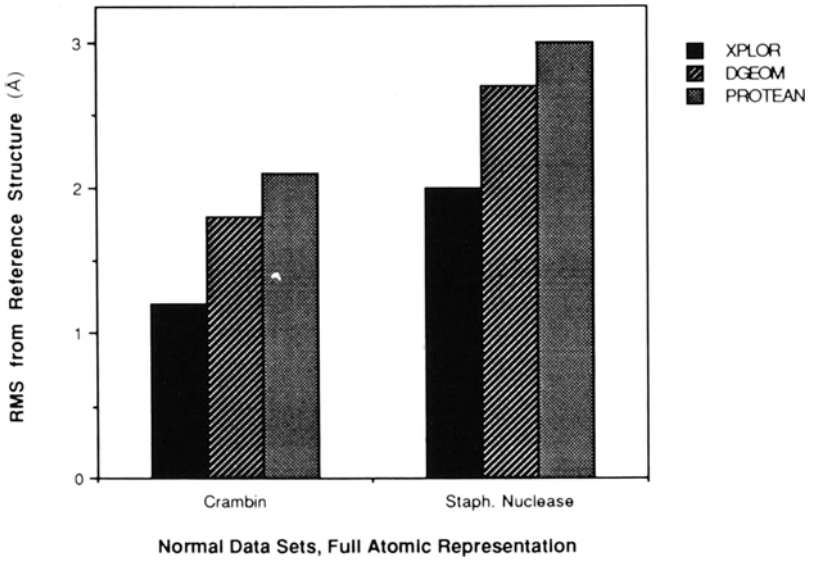

Fig. 3. The performance of the three methods with 30% data sets from a large and a small molecule (staphyloccocal nuclease and crambin, respectively). With the same percent-abundance, the larger molecule is less accurately determined by all three methods. This is a phenomenon related to the relatively decreasing abundance of short-range interproton distances as molecular size increases.

method. The table shows that XPLOR produced consistently lower angular errors than the other two methods. It is interesting, however, that the errors still remain in the region of 25–30 degrees. The reasons for this phenomenon, common to all three methods, can be attributed to the lack of angular constraints and the limited number of distance constraints and their precision. It may also represent the inherent limits of NMR measurements, which may provide dihedral angle information for smaller structures (Jardetzky and Roberts, 1981) but do not in general provide strong – i.e. very precise – information on these angles. This second possibility is supported by the fact that XPLOR, which does in fact provide additional information about dihedral angles, is apparently able to minimize these errors. The global RMS differences between the individual structures calculated by the DGEOM and XPLOR methods and the starting crystal structure are given in Table 3. The average RMSD between individual structures and the crystal structure is larger than

TABLE 3
AVERAGE ACCURACY OF INDIVIDUAL CALCULATED NMR STRUCTURES: AVERAGES OF THE ATOMIC AND ANGULAR RMS DIFFERENCES BETWEEN INDIVIDUAL CALCULATED 'NMR' STRUCTURES AND THE CRYSTAL STRUCTURE[a]

NOEs possible	DGEOM				XPLOR			
	Back-bone	Heavy atoms	φ	ψ	Back bone	Heavy atoms	φ	ψ
Crambin								
Conventional constraints								
10%	2.4±0.5	2.9±0.9	37±4 (12)	33±3 (12)	2.3±0.2	2.5±0.3	31±2 (4)	31±4 (5)
30%	1.3±0.2	1.7±0.2	35±4 (11)	30±5 (13)	1.3±0.2	1.5±0.2	27±3 (3)	28±4 (4)
50%	1.2±0.2	1.5±0.2	35±3 (10)	31±4 (10)	0.7±0.1	1.0±0.1	22±4 (2)	22±3 (2)
70%	1.0±0.1	1.3±0.1	30±3 (8)	29±3 (9)	0.5±0.1	0.6±0.1	22±2 (1)	20±2 (0)
100%	0.7±0.1	0.9±0.1	27±2 (6)	27±3 (6)	0.41±0.05	0.48±0.07	21±3 (0)	21±3 (0)
Loose NOE constraints								
30%	1.4±0.3	1.9±0.5	35±3 (10)	30±4 (10)	1.4±0.2	1.7±0.2	30±4 (3)	30±4 (3)
70%	1.3±0.2	1.7±0.7	34±2 (10)	30±5 (10)	1.1±0.1	1.4±0.1	25±3 (1)	25±3 (1)
Tight NOE constraints								
30%	1.1±0.1	1.5±0.1	32±4 (8)	28±2 (8)	0.8±0.1	1.0±0.1	23±2 (2)	24±3 (2)
70%	0.8±0.1	1.0±0.1	28±3 (6)	27±3 (6)	0.43±0.06	0.50±0.1	20±3 (0)	22±2 (0)
Full atomic representation								
30%	1.3±0.1	1.8±0.2	30±3 (3)	23±3 (3)	0.8±0.2	1.2±0.1	25±2 (1)	26±3 (1)
70%	1.1±0.1	1.4±0.1	23±4 (0)	19±2 (0)	0.7±0.1	1.0±0.1	27±4 (1)	24±3 (1)
Staphylococcal nuclease								
30%	1.9±0.3	2.7±0.4	29±3 (30)	29±2 (31)	1.3±0.4	2.0±0.2	26±2 (7)	27±2 (7)
70%	1.6±0.2	2.2±0.2	32±3 (28)	30±3 (29)	0.8±0.1	1.2±0.1	25±2 (8)	25±3 (9)

[a] See footnotes to Table 1.

the RMSD between the average structure and the crystal structure, because for the latter the positions are averaged among the individual structures before taking the RMS averaging.

In order to test how the number of sample structures used for calculating the average affects the accuracy of the average, we have calculated the RMSD of the average from the crystal structure for 7, 11 and 15 sample structures. For DGEOM the RMSDs are 1.2, 1.0 and 0.9, respectively. For XPLOR they are 0.32, 0.34 and 0.33, respectively. These numbers can be compared to the RMSD values reported in Table 1 for 12 structures each. Thus, with DGEOM the accuracy improves somewhat with the number of sample structures. For XPLOR it does not. This indicates that the use of energy potentials in the calculation makes the distribution of structures in successive runs more uniform. We also note that accuracy of the calculated average structure may show a dependence on the number of calculations only for methods using a target function, such as DGEOM or XPLOR. For optimal filtering methods, such as DIKF, which sample the conformational space in a single calculation, the issue does not arise.

Whereas Tables 1 and 3 illustrate the accuracy of each method (i.e., its ability to accurately reproduce the crystal structure from which all constraints were derived), Tables 2A and 2B address the precision of the methods. It shows the average RMSD between individual calculated structures and the average of these individuals for each data set (for DGEOM and XPLOR). For PROTEAN, it reports the standard deviations of the atomic positions as explicitly calculated by the program. As such, it provides information about the spread of structures around the mean – the precision of each method. The XPLOR calculations are the most precise, while the PROTEAN calculations are the least precise. For the 10% data set, all three methods have precisions that are greater than their accuracy. That is, the distance of the calculated structures from one another is less than the distance of the calculated structures from the crystal structure (from which all data were derived). Thus, for example, for a 10% data set XPLOR calculates a set of structures which are, on average, 2.3 Å from the crystal structure. They are, however, only 1.3 Å from each other. For the geometric methods the same is seen with a larger number of constants, and each of the methods can produce a set of structures that are more precise than the data imply.

The cross comparison between the mean structures calculated by the different methods is shown in Table 4. The extremely important result apparent from this table is that different methods of structure calculation do not converge on precisely the same structure. The global RMSD between structures calculated by different methods is at least of the order of 1 Å and can be as much as 2 Å, using exactly the same initial data set, with identical subjectively determined, preci-

TABLE 4
RMS DISTANCES IN Å BETWEEN THE AVERAGE DIKF, DGEOM, XPLOR AND CRYSTAL STRUCTURES OF CRAMBIN[a]

	DIKF	DGEOM	XPLOR	Crystal
DIKF	0	2.13	1.58	1.49
DGEOM		0	1.71	1.97
XPLOR	-	-	0	1.06
Crystal	-	-		0

[a] 30% data set, heavy atoms.

sion of the constraints. The inherent *accuracy* of NMR structures determined by any method of calculation has, therefore, to be taken to be of this order of magnitude *and no better*. The use of measures of precision (spread in the family of structures calculated by any one method) as a measure of accuracy is not justified in any case (Jardetzky, 1991).

A very instructive comparison of the RMSD from the crystal for the average structures calculated by each method of the α-carbons as a function of residue number is shown in Fig. 4. The figure shows that *within* a structure the RMSD may vary by as much as a factor of 4, probably reflecting the local variations in the density of short-range constraints. In addition, while the local RMSDs of structures calculated by different methods tend to follow the same pattern, they are not always the same, suggesting that for a *given* density of local constraints each method introduces its own bias in the sampling of conformational space. Particularly noteworthy are the three large spikes in the RMSD values of the structure calculated by distance geometry. This is a clear indication that the sampling of conformational space by this specific DG algorithm is, for some reason, not uniform. The results in this figure also underscore again the fact that the inherent *accuracy* of an NMR structure, calculated by any method, is limited by a combination of factors to RMSD values of the order of 1 Å or more. Conversely, however, it is reassuring to find that the accuracy of NMR structures is as high as that, even with relatively modest data sets. A further study of RMSD profiles is clearly indicated, especially as they can be used to diagnose poorly defined regions of mobile protein molecules (Arrowsmith et al., 1991a, b; Hua et al., 1991), *provided* that no artifacts are introduced by the method of calculation.

There do not seem to be any serious systematic errors introduced by the use of only short-range distance constraints per se; the average structures produced by each method do not appear to be any closer to each other than they are to the crystal. If they were systematically closer to one another than to the crystal, then one might believe that the categorical use of short-range distances

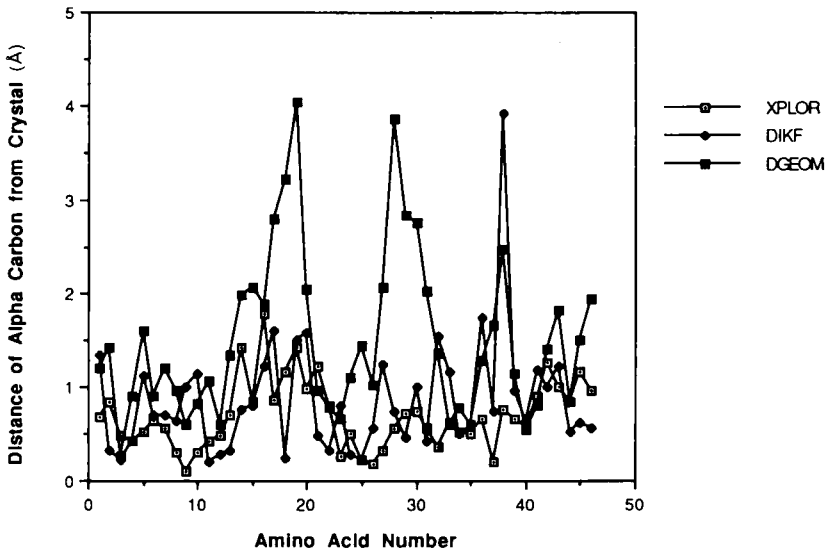

Fig. 4. The errors for individual α-carbons in crambin are shown here for each of the three structure calculation methods. The 30% data set was used. The errors are randomly distributed, for the most part, with a few spikes associated with particularly inaccurate sections of the protein as calculated by DGEOM.

might have systematically affected the ability of these methods to extract the correct structure. There have been demonstrations in the literature that short-range distances tend to produce 'contracted' structures (Pardi et al., 1984) with systematically smaller overall volume compared to crystal structure controls. We have not addressed this issue in our study, although our results are compatible with this observation. We can, however, show that the magnitude of these types of errors is not likely to result in global errors more than about 0.5 Å.

In order to compare the relative computational expense of each of these methods, we ran each of them with the 10% data set on crambin on the Cray/YMP at the Pittsburgh Supercomputing Center. The CPU times to calculate the crambin structure at the pseudoatomic level were 42 s, 384 s and 4775 s per structure for DGEOM, XPLOR and DIKF, respectively. The DIKF structure is, of course, more expensive because only one calculation is needed to quantify the positional uncertainty for each atom and so represents the equivalent of many runs of DGEOM and XPLOR. Specifically, in the time required to calculate the DIKF average structure and uncertainty estimates, one can obtain 114 DGEOM structures or 12 XPLOR structures.

CONCLUSIONS

In the last five years, a number of methods have been used to determine protein structures from NMR data. The goal of the experiments reported here was to provide a check of these methods and their consistency. It is quite reassuring that three methods based on extremely different mathematical formalisms are able to take the given data sets and produce remarkably good protein structures with a reasonable amount of computation. At the same time, however, there are certainly small differences in the performance of these programs that should require their users to analyze the results with a fair amount of circumspection. These differences may be exacerbated in the future as we push these calculational technologies to their limits with larger proteins, larger data sets, and larger uncertainties. We believe that the experiments described in this paper justify 10 conclusions about the methods for structure calculation:

(1) All of the tested methods reproduce the 'gold standard' structure reasonably faithfully. It is unlikely that any of these methods will produce categorically incorrect structures, given reasonable data sets (i.e., greater than 10% of all possible short-range interproton distances or 4 NOE constraints/residue).

(2) For all methods, the essential structural features are very well defined at 30% data abundance (10–14 NOE constraints/residue) for molecules the size of crambin (5 kDa). They are also well defined at comparable data abundance for molecules the size of staphylococcal nuclease (17 kDa), although the accuracy of such calculations deteriorates slightly.

(3) Restrained molecular dynamics produces structures that are *both* more *accurate* and more *precise* than the purely geometric methods, such as DGEOM and PROTEAN. The improved satisfaction of dihedral angular constraints results from the inclusion of explicit theoretical constraints on these angles.

(4) All the tested methods respond to perturbations in the data sets in the same (and expected) qualitative manner: (i) increased data abundance and increased precision of the data produce better structures, (ii) decreased data abundance and decreased precision produce larger errors, (iii) full atomic representations allow more accurate structures, and (iv) larger structures take longer and require relatively more data to solve to the same precision.

(5) All three methods may show a precision that is not fully justified by their accuracy; the calculated structures can be more closely related to each other than to the reference structure. These differences are not dramatic, but may be important when ultrastructural questions are at issue.

(6) PROTEAN and DGEOM have a comparable accuracy in these experiments, but PROTEAN is less precise, given the same precision in the initial data set. There is an infinite number of structures compatible with the means and variances produced by PROTEAN and some of them will be more compatible with the data than others.

(7) The *subjective* judgment on the precision of the constraints is a major determinant of the accuracy and precision of calculated structures, regardless of the method of calculation. Imprecise constraints are generally more damaging to the performance of all three methods in situations of high data abundance. This is a result of the fact that imprecision limits the ability to satisfy more abundant data in an internally consistent manner.

(8) Larger structures are determined less accurately and less precisely than smaller structures given the same data set abundance. This results from the fact that the abundance of short-distance constraints does not rise as rapidly as the abundance of total distance constraints, yet NMR can only capture the short distances.

(9) The global RMSD as usually reported is only a rough measure of the quality of the structure. A residue-by-residue comparison of the RMSD may reveal a contrast between well defined and poorly defined segments. However, it may also reveal unexpected bias in the method of calculation.

(10) The elapsed time for all methods is reasonable for proteins the size of crambin and staphylococcal nuclease. PROTEAN calculates one mean and set of variances in the same time required to calculate roughly 114 DGEOM or 12 XPLOR structures.

The choice of methodology for the calculation of a new NMR structure will, in general, depend on the specific questions being asked by the investigators. The data provided here help quantify both the relative and absolute performance of three methods for structure calculation. Of course, there are also likely to be numerous logistical considerations in the choice of method, including previous use of these algorithms, access to computational platforms and quality of data set.

The question as to whether the commonly used combinations of methods (e.g. DIKF or DG followed by RMD) will produce a more accurate, rather than simply a more precise structure, has not been addressed in this paper because of the computational expense involved in running all the appropriate controls. It must be borne in mind that any 'refinement' calculation starting from a mean determined by another method can pull the mean *away*, as well as *toward* the initial 'gold standard' structure, depending on the additional information (e.g. energy potentials) provided.

ACKNOWLEDGEMENTS

This research was funded by NIH grants RR-02300 and RR-07558. We would like to thank the Pittsburgh Supercomputing Center for time on their Cray/YMP (grant DMB-910057) and the Cornell National Supercomputer Facility (which receives funding from the NSF, IBM, New York State and members of the Corporate Research Institute) for time on their IBM 3090.

REFERENCES

Altman, R.B. and Jardetzky, O. (1986) *J. Biochem.*, **100**, 1403–1423.

Altman, R.B. and Jardetzky, O. (1989) *Methods Enzymol.*, **177**, 218–246.

Altman, R.B., Pachter, R. and Jardetzky, O. (1989) In *Protein Structure and Engineering* (Ed, Jardetzky, O.) Plenum Press, New York, pp. 79–95.

Altman, R.B., Pachter, R., Carrara, E.A. and Jardetzky, O. (1990) *QCPE Bull.*, **10** (4), Program 596.

Altman, R.B., Pachter, R. and Jardetzky, O. (1992) *J. Appl. Magn. Reson.*, in press.

Arrowsmith, C.H., Czaplicki, J., Iyer, S.B. and Jardetzky, O. (1991a) *J. Am. Chem. Soc.*, **113**, 4020–4022.

Arrowsmith, C., Pachter, R., Altman, R. and Jardetzky, O. (1991b) *Eur. J. Biochem.*, **202** (2), 53–66.

Blaney, J.M., Crippen, G.M., Dearing, A. and Dixon, J.S. (1990) *QCPE*, Program 590, Indiana University, Bloomington, IN 47405.

Braun, W. and Gō, N. (1985) *J. Mol. Biol.*, **186**, 611–626.

Brünger, A.T. (1990) *XPLOR Manual*, Yale University, New Haven, CT.

Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3801–3805.

Bystrov, V.F. (1976) *Prog. NMR Spectrosc.*, **10**, 41–81.

Carrara, E.A., Brinkley, J.F., Cornelius, C.C., Altman, R.B., Brugge, J., Pachter, R., Buchanan, B. and Jardetzky, O. (1990) *QCPE Bull.*, **10** (4), Program 596.

Clore, G.M. and Gronenborn, A.M. (1989) *J. Magn. Reson.*, **84**, 398–409.

Clore, G.M., Nilges, M., Brunger, A.T., Karplus, M. and Gronenborn, A.M. (1987) *FEBS Lett.*, **213**, 269.

Gronenborn, A.G., Wingfield, P.T. and Clore, G.M. (1989) *Biochemistry*, **28**, 5081.

Hendrichson, W.A. and Teeter, M.A. (1981) *Nature*, **290**, 107.

Hua, Q.X., Shoelson, S.E., Kochoyan, M. and Weiss, M.A. (1991) *Nature*, **354**, 238–241.

Jardetzky, O. (1991) In *Computational Aspects of the Study of Biological Macromolecules by NMR Spectroscopy* (Eds, Hoch, J.C., Poulson, F.M. and Redfield, C.) Plenum Publ. Corp., New York, pp. 375–390.

Jardetzky, O. and Roberts, G.C.K. (1981) In *NMR in Molecular Biology*, Academic Press, New York, NY, Chapter 5.

Jardetzky, O. and Lane, A.N. (1988) In *Physics of NMR Spectroscopy in Biology and Medicine* (Ed, Maraviglia, B.) Ital. Physical Society, North-Holland Physics Publishing, Amsterdam, pp. 267–300.

Keepers, J.W. and James, T. (1984) *J. Magn. Reson.*, **57**, 404–426.

Koehl, P., Lefèvre, J.-F. and Jardetzky, O. (1992) *J. Mol. Biol.*, **223**, 299–315.

Kuntz, I.D., Crippen, G.M. and Kollman, P.A. (1979) *Biopolymers*, **18**, 939–957.

Kuntz, I.D., Thomason, J.F. and Oshiro, C.M. (1989) *Methods Enzymol.*, **177**, 159–204.

Lefèvre, J.-F., Lane, A.N. and Jardetzky, O. (1987) *Biochemistry*, **26**, 5076–5090.

Lichtarge, O., Jardetzky, O. and Li, C.H. (1987) *Biochemistry*, **26**, 5916–5925.

Loll, P.J. and Lattman, E.E. (1989) *Proteins: Structure, Function, and Genetics*, **5**, 183–201.

Madrid, M. and Jardetzky, O. (1988) *Biochim. Biophys. Acta*, **953**, 61–69.

Madrid, M., Mace, J.E. and Jardetzky, O. (1989) *J. Magn. Reson.*, **83**, 267–278.

Metzler, W.J., Hare, D. and Pardi, A. (1989) *Biochemistry*, **28**, 7045–7052.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Pachter, R., Altman, R.B. and Jardetzky, O. (1990) *J. Magn. Reson.*, **89**, 578–584.

Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.*, **181**, 741.

Scheek, R.M., van Gunsteren, W.F. and Kaptein, R. (1989) *Methods Enzymol.*, **177**, 204–218.

Van Gunsteren, W.F., Kaptein, R. and Zuiderweg, E.R.P. (1983) In *Nucleic Acid Conformation and Dynamics* (Ed, Olson, W.K.) Report of the NATO/CECAM Workshop, Orsay.

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

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