Journal of Biomolecular NMR (2006) 34: 153–166
DOI 10.1007/s10858-006-6889-8

Article

Comparison of different torsion angle approaches for NMR structure determination

Benjamin Bardiaux[b], Thérèse E. Malliavin[a,*], Michael Nilges[b] & Alexey K. Mazur[a]
[a]*Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-Chimique, 13 rue P. et M. Curie, 75 005, Paris, France;* [b]*Institut Pasteur Unité de Bioinformatique Structurale, CNRS URA 2185, Institut Pasteur 25-28 rue du Dr Roux, F-75724, Paris Cedex 15, France*

Received 21 November 2005; Accepted 17 January 2006

Key words: BMRB, *cis* peptides, CNS, CYANA, ICMD, NMR, RECOORD, ribosomal protein, structure calculation, torsion angle space, XPLOR-NIH

Abstract

A new procedure for NMR structure determination, based on the Internal Coordinate Molecular Dynamics (ICMD) approach, is presented. The method finds biopolymer conformations that satisfy usual NMR-derived restraints by using high temperature dynamics in torsion angle space. A variable target function algorithm gradually increases the number of NOE-based restraints applied, with the treatment of ambiguous and floating restraints included. This soft procedure allows combining artificially high temperature with a general purpose force-field including Coulombic and Lennard-Jones non-bonded interactions, which improves the quality of the ensemble of conformations obtained in the gas-phase. The new method is compared to existing algorithms by using the structures of eight ribosomal proteins earlier obtained with state-of-the-art procedures and included into the RECOORD database [Nederveen, A., Doreleijers, J., Vranken, W., Miller, Z., Spronk, C., Nabuurs, S., Guntert, P., Livny, M., Markley, M., Nilges, M., Ulrich, E., Kaptein, R. and Bonvin, A.M. (2005) *Proteins*, **59**, 662–672]. For the majority of tested proteins, the ICMD algorithm shows similar convergence and somewhat better quality Z scores for the ϕ, ψ distributions. The new method is more computationally demanding although the overall load is reasonable.

Introduction

The NMR structure determination is usually performed using a simulated annealing protocol to minimize an energy function, which involves a restraint energy based on the NMR measurement of geometric parameters and a chemical energy describing the physico-chemical properties of the studied biomolecules (Brunger et al., 1998a). The relationship between the NMR data and the atomic coordinates is not straightforward: NMR measures local geometrical parameters (distances and angles), which may be easily computed from the molecular structure, but not vice versa. Therefore, NMR biomolecular structures often exhibit defects, which may hamper molecular modeling studies (Fan and Mark, 2003).

During the last 10 years, many successful methods have been proposed to improve the convergence of the NMR structure refinement and the quality of the structures obtained. The parameters of the non-bonded interactions were improved (Linge and Nilges, 1999; Linge et al., 2003) to

*To whom correspondence should be addressed: E-mail: therese.malliavin@ibpc.fr

optimize the quality of the obtained conformations. The integration of motion in the torsion angle space (Güntert et al., 1997; Stein et al., 1997; Schwieters and Clore, 2001) was shown to improve the structure convergence as well as the structure quality. This is because the use of internal coordinates along with the rigid standard geometry of chemical groups generally improves the sampling of the conformational space due to the reduced complexity of the potential energy landscape (Mazur, 2001). Moreover, the fixed standard geometry of polypeptides allows for a significant increase of the temperature and the time step in MD-based simulated annealing protocols. On the other hand, the final refinement of structure in water was shown (Kordel et al., 1997; Linge et al., 2003) to further improve the structure quality because a force-field including the Coulombic and Lennard-Jones non-bonded interactions is used during this refinement. Indeed, because of technical reasons, the high-temperature stages of the simulated annealing protocol are run using simplified force-fields.

The Internal Coordinate Molecular Dynamics (ICMD) approach was one of the first approaches (Abagyan and Mazur, 1989; Mazur and Abagyan, 1989) for modeling polymers with partially fixed spatial structure. ICMD is not limited to dynamics with a single choice of variables corresponding to the torsion angle space. In contrast, it can vary the degree of molecular flexibility by fixing and/or unfixing torsions, bond angles, and bond lengths. In the last 10 years, this approach was significantly enhanced by making possible propagation of MD trajectories in the space of canonical variables with new symplectic numerical integrators and the recursive mass matrix inversion algorithm (Jain et al., 1993; Mazur, 1997). These techniques were adapted for all-atom simulations of proteins and nucleic acids in explicit aqueous environments (Mazur, 1998a, b, 2002). All these advances are potentially useful for NMR-based refinement applications since they greatly improve the quality and stability of MD trajectories and also allow easy modeling of multimolecular complexes. A new algorithm for NOE-based determination of biomolecular structures has been recently introduced in ICMD that has been successfully applied in studies of peptides and peptide-ion complexes (Kozin et al., 2001; Zirah et al., 2006). This algorithm applies the idea of the variable target function (Braun and Go, 1985) in the context of torsion angle MD and also involves a simplified treatment of floating and ambiguous restraints.

An important advantage of the variable target function approach consists in the possibility of using a general purpose force-field at all stages of structural refinement. In standard MD based protocols, NMR restraints are applied simultaneously from the very beginning. Under high temperature, this results in stiff atom–atom collisions, which would require a drastic reduction of the time step if the full Coulombic and Lennard-Jones non-bonded interactions were used. Therefore simplified force-fields with softened atom–atom repulsive terms have to be employed. Further improvement of the structures obtained is achieved by additional in-water MD simulations under normal temperature with standard force-fields (Linge et al., 2003). In contrast, in the variable target function protocol, the NMR restraints are introduced gradually in a certain order that allows the refinement to proceed via valleys of potential energy landscapes, and makes this procedure compatible with standard all-atom force-fields.

Here we present an adaptation of this approach to the classical problem of NOE-based determination of protein structures, as well as a thorough comparative test on a representative group of proteins earlier studied by the state-of-the-art methods. For such comparison, a set of eight proteins was selected from the database RECOORD (Nederveen et al., 2005). This database resulted from an effort of protein structure recalculation, with methods corresponding to the state-of-the-art of the NMR structure determination. Our results show that the ICMD algorithm converges to the same target structures, with a slightly larger RMSD between the conformations produced in repeated runs. For the majority of tested proteins the new method also shows a somewhat better quality Z scores for ϕ, ψ combinations. At the same time, it is more computationally demanding.

Materials and methods

Molecular dynamics simulations

The numerical integration of ICMD equations of motion employed an implicit leapfrog integrator in

torsion angle space, which makes it possible to use a time-step of 10 fs step (Mazur, 1997). This time step is optimal for torsion MD of biopolymers with standard atom parameters (Mazur, 1998b). To assure stability of numerical integration under high temperature, the moments of inertia of all rigid bodies are additionally increased by 300 amu $Å^2$ (Mazur, 1997). The accurate treatment of flexibility of five-member rings by using the projector operator approach (Mazur, 1999) requires bond angle degrees of freedom. Therefore, a simpler approximation was used here where proline rings were left flexible with only torsion degrees of freedom whereas the N–C_δ bonds were closed by using standard bond length and bond angle potentials.

The NMR-based molecular dynamics simulations were performed in the gas-phase. The force-field parameters were taken from the AMBER all-atom parameter set (Cornell et al., 1995), and the covalent geometry of the amino-acids was derived from the CNS paramallhdg5.2 parameters (Linge and Nilges, 1999). The non-bonded van der Waals energy was modeled through the Lennard-Jones potential with simple spherical cutoff. The Coulomb electrostatic energy was computed with a force-shift truncation method (Levitt et al., 1997). The cutoff distance was 6 Å for all non-bonded interactions. The temperature bath coupling was performed using the Berendsen algorithm (Berendsen et al., 1984).

In the AMBER force-field, the *cis* and *trans* peptide bond conformations are characterized by a modest energy difference (ca 10 kcal/mol) and a torsion barrier of 10 kcal/mol. Therefore the high temperature phase of refinement produces a non-zero population of *cis* conformers if they are compatible with NMR restraints. The *cis*/*trans* barrier is raised to 100 kcal/mol before cooling to remove intermediate conformations and reveal the sites of high *cis*-peptide population. These populations are usually low and the *cis* peptide conformers are removed from the final structures by a short MD run with a dihedral restraint that eliminates *cis*-conformers. The same restraint can be applied all through the variable target function protocol if the *cis*/*trans* statistics is not interesting.

The variable target function protocol is implemented as set of scripts in ICMD command language that involve calls to ICMD standard functions for energy minimization, MD runs etc. These scripts as well as binary versions of ICMD compiled on Linux, are freely available at: http://www.ibpc.fr/~terez/ICMD. The structures analyzed here can be also downloaded from the same Web address.

NMR restraints

The dihedral angle restraints were applied as usual, that is, by using a harmonic potential, calculated from the list of the four atoms concerned, the energy constant value, the target value to which the dihedral angle is restrained, and the interval around this target value. The dihedral angle potential was equal to zero in the interval defined by the target value and the interval range, and was harmonic outside this interval. All dihedral angle restraints were active permanently throughout the calculations.

The NOE restraints take as input the upper bound of the distance interval, and two groups of atoms between which the restraint is applied. The NOE restraint potential was equal to zero for distances smaller than the upper bound value, and harmonic for larger distances. For groups of atoms including more than one element two alternative procedures were tested. In the first procedure (swapping: floating restraint) all distances between all possible atom pairs were calculated, and the restraint was applied to the smallest distance. The swapping was repeated each time the restraint list and/or forces are changed (see below). This simple approach is considered as standard in our variable target function protocol. In the second procedure (ADR: ARIA ambiguous restraint (Nilges, 1993)), the restraint was applied to the pseudo-distance D, corresponding to the sum of the inverse sixth power distances d_i between all possible atom pairs:

$$\frac{1}{D^6} = \sum_i \frac{1}{d_i^6}. \tag{1}$$

This second method was introduced for better comparison because the reference structures from the RECORD database were obtained with the ambiguous form of restraints. We note that the floating restraint can be considerably more stringent than the ambiguous restraint.

Distance restraints were applied using the variable target function algorithm (Braun and Go,

1985), modified as follows. The torsion angle separation (TAS) is defined as the number of torsion angles connecting a given pair of atoms. The corresponding TAS values were assigned to all atom pairs potentially involved in distance restraints. The variable target function calculations started from TAS = 1 and the following TAS levels were added consecutively. In the case of floating restraints, the swapping procedure was applied before every cycle of dynamics or minimization, with all NOE restraints checked and the atom pair corresponding to the smallest distance, chosen; the TAS of this atom pair was next used to decide if the corresponding restraint should be applied at the current stage of the protocol. In the case of ambiguous restraints, the contributions to each restraint were introduced successively, according to their corresponding TAS value.

Simulated annealing protocol

The parameters of the simulated annealing protocol are given in Table 1. The starting protein conformation was the extended polypeptide chain. The protocol started with 100 steps of minimization, followed by the high-temperature unfolding phase (Table 1: lines 1–2), i.e., 100 steps of molecular dynamics at 500 K, and 100 steps of molecular dynamics at 7000 K. The upper limits of the temperature were 1000 and 10,000 K, respectively. The relaxation time for the coupling to the temperature bath was 10 ps. At the end of this unfolding phase, the polypeptide conformation was stored to be used as a starting point for the next conformation generation.

The second step of the simulated annealing protocol was the variable target function annealing phase (Table 1). It started with seven runs of molecular dynamics at 3000 K (Table 1: lines 3–9) with a maximum temperature value of 10,000 K and a relaxation time of 0.1 ps for the coupling to the temperature bath. For each level of torsion angle separation (TAS), 200 steps of molecular dynamics were performed followed by six runs of 100 steps. During these simulations, the energy constant of the NOE restraints was increased from 0.1 up to 10 kcal/mol/Å^2 and the energy constant of the dihedral restraints was increased from 1 up to 50 kcal/mol/rad^2. At the end of the round seven, 300 additional steps of molecular dynamics were run (Table 1: line 10). The following condition *C* was then checked: the value *R* of the restrained energy per restraint is larger than 1.2, and the TAS level was not processed more than three times. The larger the ratio *R*, the larger is the number of unsatisfied restraints. If the condition *C* was true, the variable target annealing phase was again performed with the same TAS level. If the condition is false, the next TAS level was processed.

When all restraints have been applied, the NOE regrouping phase (Table 1: lines 11–18) was performed. Eight MD runs were carried out at a temperature of 3000 K, with a maximum temperature of 10,000 K, and a relaxation time of 0.1 ps. The NOE energy constant was increased from 0.1 up to 50 kcal/mol/Å^2 and the dihedral angle energy constant from 1 up to 50 kcal/mol/rad^2. The number of steps was 300 for the first run and 100 for the others.

The cooling phase was then taking place with a relaxation time of 1 ps for the temperature coupling bath (Table 1: lines 19–22). The energy barrier between *cis* and *trans* peptide conformers is raised as explained above. Three runs of 600 steps (Table 1: lines 19–21) were performed at mean temperature values of 200, 100 and 50 K, with energy constants, respectively of 5 and 50 kcal/mol/rad^2 for the NOE and dihedral restraints. Then, 100 steps of MD simulation (Table 1: line 22) were run with the target temperature of 0 K. Finally, two runs of energy minimization are carried out each one preceded by the NOE restraint swapping.

A statistical analysis of non-proline peptide bonds is then performed in order to detect sites of persistent *cis* conformers. This situation is rare, but it cannot be completely neglected. A high population (> 50%) of a specific *cis* conformer can arise from false NMR restraints as well as from a true protein conformation. Analysis of high-resolution X-ray structures shows (Jabs et al., 1999) that 0.028% of non-proline peptide bonds are in *cis* conformation, and detection of such cases by improved NMR techniques may be anticipated. In the calculations presented here, populations of *cis* conformers never approached 50%, which means that they can be neglected under normal temperature. Moreover, the RECORD database structures used here for comparisons were all refined with non-proline peptide bonds restrained to the *trans* conformation. It was preferable, therefore, to

Table 1. Parameters of the simulated annealing protocol used in the ICMD methodology

	Number of steps	Mean temperature (K)	Maximum temperature (K)	Relaxation time (ps)	NOE energy constant (kcal/mol/Å^2)	Dihedral energy constant (kcal/mol/rad^2)
High T unfolding phase						
1	100	500	1000	10	0	0
2	100	7000	10000	10	0	0
Variable target phase						
3	200	3000	10000	0.1	0.1	1
4	100	3000	10000	0.1	0.3	5
5	100	3000	10000	0.1	0.9	10
6	100	3000	10000	0.1	2.7	20
7	100	3000	10000	0.1	5.4	30
8	100	3000	10000	0.1	8.1	40
9	100	3000	10000	0.1	10	50
10	300	3000	10000	0.1	10	50
NOE regrouping phase						
11	300	3000	10000	0.1	0.1	1
12	100	3000	10000	0.1	0.5	5
13	100	3000	10000	0.1	1	10
14	100	3000	10000	0.1	5	20
15	100	3000	10000	0.1	10	30
16	100	3000	10000	0.1	20	40
17	100	3000	10000	0.1	40	50
18	100	200	10000	0.1	50	50
Cooling phase						
19	600	200	10000	1	5	50
20	600	100	10000	1	5	50
21	600	50	10000	1	5	50
22	100	0	10000	1	5	50
Repeated NOE regrouping phase						
23	300	1000	10000	0.1	0.1	1
24	300	3000	10000	0.1	0.1	1
25	100	3000	10000	0.1	0.5	5
26	100	3000	10000	0.1	1	10
27	100	3000	10000	0.1	5	20
28	100	3000	10000	0.1	10	30
29	100	3000	10000	0.1	20	40
30	100	3000	10000	0.1	40	50
31	100	200	10000	0.1	50	50
Repeated cooling phase						
32	600	200	10000	1	5	50
33	600	100	10000	1	5	50
34	600	50	10000	1	5	50
35	100	0	10000	1	5	50

carry out our comparisons with *cis*-peptide conformers eliminated completely. To this end their remaining population was further reduced by repeating NOE regrouping and cooling phases (Table 1: lines 23–35) with an additional dihedral restraint that raised the energy of the *cis* peptide bonds by about 30 kcal/mol. As a result, the population of *cis*-peptide conformers reduced from 1.8 to 0.1% and the fraction of protein structures with at least one *cis*-peptide bond

reduced from 71.1 to 6.9%, and this last group of conformations was excluded from further analysis.

For each protein, about 100 conformations were calculated. The exact numbers of conformations are given in Table 2.

Structure analysis

Comparison of different calculation protocols was performed by using parameter values averaged over the corresponding ensembles of conformations. Two parameters describe the convergence of calculations and the fitting of the NMR restraints, namely, the RMSD between selected protein regions (Table 2), and the number of consistent violations. The second set of parameters describes the quality of the structures obtained, namely, the percentage of residues in the core region of the PROCHECK Ramachandran diagram, three WHATIF Z scores, the Ramachandran score, the 2nd-generation packing score and the backbone conformation score, and the number of bumps detected by WHATIF. The analysis was performed using PROCHECK v.3.5.4 (Laskowski et al., 1993) and WHATIF (Hooft et al., 1996).

The violations of the distances restraints were calculated in two different ways according to the type of restraints used in calculations (ambiguous or floating). For floating restraints, the violations were calculated by taking the distances of all possible pairs and checking if they were larger than $U+0.5$ Å, where U is the upper bound value of the restraint. If all distances were larger than $U+0.5$ Å, then the restraint was violated. For ambiguous restraints, the distance D (Equation 1) was calculated from the set of distances d_i between the two groups of considered atoms. The restraint was violated if D was larger than $U+0.5$ Å. If not otherwise stated, the violations in RECOORD structures were determined for ambiguous restraints. The consistent violations, i.e., the restraints violated in more than 50% of the analyzed conformations (Nederveen et al., 2005) were included in the analysis, as well as the RMS of violations, calculated on all violations larger than U.

The RMSD between the conformations was calculated by determining the mean coordinates of all conformations, and by calculating the RMSD of each conformation with respect to the mean conformation. The mean RMSD value gives the RMSD between the conformations ($RMSD_{conf}$). The average CNW structure were compared to other RECOORD and ICMD average structures by calculating the RMSD values ($RMSD_{diff}$) between them. Only the regions of conformations corresponding to well-defined structures were superimposed: these regions are given in Table 2. The RMSD values analyzed here were calculated on the backbone atoms.

Results and discussion

The set of proteins used

The set of eight proteins used to analyze the results obtained with ICMD and to compare the ICMD conformations with those stored in

Table 2. Set of the proteins used for the comparison

Name	Size (a)	α (%) (b)	β (%) (c)	NOEc (d)	PDB (id)	Reference	Blup (e)	Adr2 (f)	RMSD (g)	NOEn (h)
L25	94	9.0	43.5	0.54	1b75	Stoldt et al. (1998)	106	99	8–76	16.4
L30	104	37.6	19.5	0.31	1ck2	Mao and Willamson (1999)	101	111	10–101	14.2
L11	76	45.9	5.9	0.49	1fow	Markus et al. (1997)	110	100	8–76	12.0
L20	60	57.1	0.0	0.42	1gyz	Raibaud et al. (2002)	112	109	1–60	13.1
L23	96	20.6	27.1	0.50	1n88	Ohman et al. (2003)	104	100	7–60, 75–96	17.7
S28e	82	0.5	33.8	0.64	1ny4	Aramini et al. (2003)	114	109	7–56	21.1
S19	92	11.3	9.4	0.36	1qkh	Helgstrand et al. (1999)	107	137	3–67	15.1
S27e	66	0.2	30.0	0.37	1qxf	Herve du Penhoat et al. (2004)	134	139	4–53	11.5

(a) The size of the protein is expressed as the number of residues. (b) The percentage of α secondary structures. (c) The percentage of β secondary structures. The percentages of α and β secondary structures were calculated by DSSP (Kabsch and Sander, 1983). (d) The NOE completeness taken from the database RECOORD (Nederveen et al., 2005). (e) Number of conformations calculated using the floating restraints. (f) Number of conformations calculated using the ambiguous restraints. (g) Residue numbers used in the conformation superposition for the RMSD calculation. (h) Number of NOE restraints by residue.

the RECOORD database are ribosomal proteins with sizes in the range of 60–104 residues (Table 2). They are thus medium-size proteins with respect to the sizes which are now attainable in NMR structural studies (Miclet et al., 2003). Most of the proteins are α–β proteins (L25, L30, L11, L23 and S19), and there are one α protein (L20) and two β proteins (S28e and S27e). The NOE completeness are in the 0.31–0.64 range, but half of the proteins (L30, S19, S27e and L20) display a NOE completeness smaller than 0.45, and are thus in the region of low completeness (Doreleijers et al., 1999). The number of NOE restraints by residue is in the 11.5–21.1 range (Table 2).

Proteins L11, L23, S28e and S19 contain long (about or more than 10 residues) loops and/or N or C terminal tails. These regions may display larger internal flexibility. The structures of the majority of these proteins, except L25 and L30, were determined only by NMR. X-ray (Lu and Steitz, 2000) and NMR (Stoldt et al., 1999) structures of the L25–RNA complex were determined. The structure of the protein L30 isolated (Chen et al., 2003) was determined by X-ray crystallography, and the structure of the L30–RNA complex was determined from a joint X-ray and NMR refinement (Chao and Williamson, 2004).

Backbone dihedral angle restraints are available only for the proteins L30, L11, L23 and S28e. Side chain χ_1 angle restraints were also measured on L30. The dihedral angles given for L23 were automatically determined using TALOS (Cornilescu et al., 1999).

Analysis of the ICMD structures

The results presented in this and the following subsections are obtained with the floating form of restraints. The corresponding data for the alternative form (ambiguous) are similar and they will be discussed in subsection 'Ambiguous and floating form of restraints.' The following analysis involves the complete set of ICMD conformations as well as the 25 lowest energy conformations, obtained after the second NOE regrouping and cooling phases (Table 1: lines 23–35). The total energy is the sum of the ICMD potential energy, and of the restraint energies.

For all proteins analyzed, the $RMSD_{conf}$ values calculated on the set of the 25 best conformations are smaller than the corresponding values calculated on the total set of conformations (Figure 1a). The RMSD change during the selection of the 25 best conformations (Figure 1a) is smaller than 0.5 Å, for all proteins except for S19 (3 Å). This difference for the S19 structure may come from the presence of the long C terminal tail (residues 63–80) which exhibits a non-canonical secondary structure, formed by consecutive loops. Indeed, two structure clusters, exhibiting different energies, were observed in the total set of conformations. All proteins, except L11 and S19, have RMSD values over the complete set of conformations smaller than 1.7 Å, indicating that the protocol used by ICMD converged to a well-defined 3D structure. The protein L11 displays a large conformational variability in the loop 19–33: this variability increases the corresponding RMSD.

The number of consistent violations (Figure 1b) varies significantly between different proteins. L25 and L20 show the largest number of violations whereas for L30, L11, S28e and S27e the number of violations is low. The presence and the quality of dihedral angle restraints influence strongly the number of consistent violations. In fact, no dihedral angle restraints were available for all proteins (L25, L20 and S19) exhibiting the largest number of violations. Moreover, the dihedral angle restraints for L23 were automatically

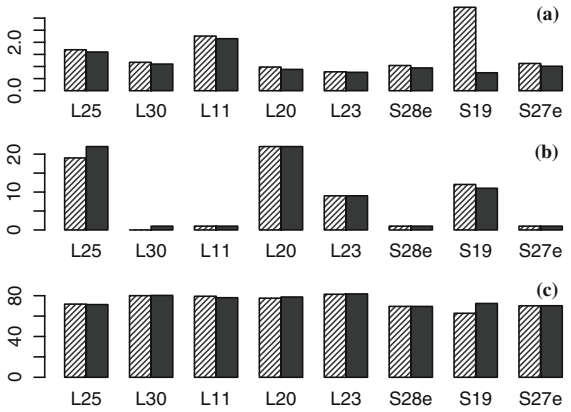

Figure 1. Comparison of the RMSD (Å) (a), of the number of consistent violations (b) and of the percentage of residues in the core Ramachandran regions (c), for calculations run with ICMD on the set of proteins from Table 2. The results shown are those obtained using all floating restraints conformations (shaded dark blue) and the best 25 floating restraints conformations (dark blue). The compared values are mean values calculated on the set of the analyzed conformations.

derived from the TALOS protocol, and may contain incorrect predictions (Cornilescu et al., 1999). Thus, it seems that the application of reliable restraints on dihedral angles reduces the number of violations. This tendency is probably intensified by the use of larger energy constants for dihedral angles than for distances (Table 1).

The number of consistent violations does not change much over all generated conformations or over the best 25 conformations. This is in agreement with the small improvement of convergence obtained by the selection.

The majority of violated restraints involve methyl or methylene protons. In general, one-half of them concerns long-range NOEs. The majority of violations in S19 concerns the C-terminal β strand (residues 56–62) and the C-terminal tail (residues 63–80) of the protein. The majority of other proteins (L25, L30, L23) exhibits violations involving residues located in β strands. Protein L11 shows violations for long-range NOEs between α-helices.

The percentage of residues in the core Ramachandran regions, determined by PROCHECK, is displayed in Figure 1c. All percentage values for the best 25 conformations are larger than 69%. The variation of percentage between all conformations and the 25 best ones is smaller than 2%, except for S19, and is parallel to the variation of RMSD. Proteins S28e, S19 and S27e display a percentage smaller than 75%. The residues located by PROCHECK in the non-allowed Ramachandran regions were analyzed in detail for the best 25 conformations. These residues are mainly located in loops, or in the C-terminal tails of S19 and S28e.

The conformations obtained with ICMD were also analyzed by calculating WHATIF quality Z scores (Figure 2). For the majority of proteins analyzed here, the Z scores are larger than −4 and the conformations generated by ICMD are thus of good quality. Also, the Z scores increase when the 25 best conformations are selected. Z scores smaller than −4 are observed for L25 (Ramachandran and backbone conformation scores), for L20 (Ramachandran score), for S28e (backbone conformation score) and for S19 (Ramachandran, 2nd-generation packing and backbone conformation scores). For S19, L20 and S28e, the results of WHATIF are in agreement with the results previously described for PROCHECK.

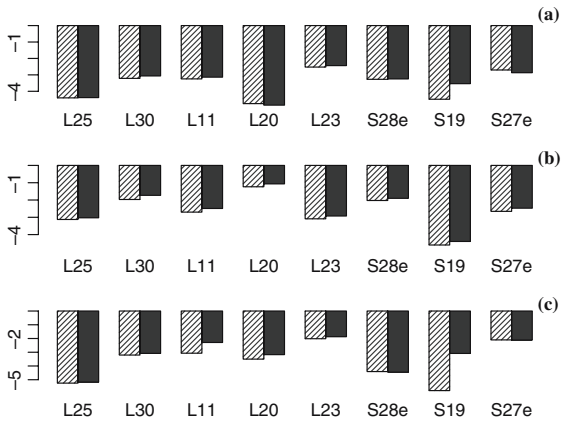

Figure 2. Comparison of the WHATIF Ramachandran Z score (a), of the 2nd-generation packing Z score (b), and of the backbone conformation Z score (c) for calculations run on the set of proteins from Table 2. The results shown are those obtained using all floating restraints conformations (shaded dark blue), and the best 25 floating restraints conformations (dark blue). The compared values are mean values calculated on the set of the analyzed conformations.

The WHATIF results were analyzed in detail for the 25 best conformations, to detect which residues are responsible for lowering the backbone conformation Z score. Similarly to PROCHECK, these residues are mainly located in loops and in the N- or C-terminal tails. Almost the same set of residues was detected whatever the restraint type (floating or ambiguous) was used.

The selection of the 25 best conformations, based on their total energy, improves the convergence of the calculation, as the RMSD between conformations decreases, and also improves the structural quality, as well as the restraint fit. The convergence and the quality of the obtained conformations are generally good before the selection, except for S19. A larger number of violated NOE restraints are observed for proteins exhibiting no dihedral angle restraints, or automatically predicted dihedral angle restraints.

Ambiguous and floating form of restraints

Comparison of these two different treatments of the restraints presents a non-negligible methodological interest. It should be clear that they were used separately, in different refinement series. With the floating restraints, one always assigns a restraint to a definite pair of protons. This proce-

dure is simple and transparent and it can be employed only in the framework of the variable target function protocol where re-attribution of restraints is made frequently anyway. With ambiguous restraints, many atoms contribute to the restraint simultaneously, which in some cases is equivalent to the first approach, and in some cases not. The latter method is standard in the ARIA-CNS refinement algorithm used for producing the reference structures borrowed from the RECORD database, and we should check if any large difference in the final structures can arise just from the different type of NOE restraints.

The data assembled in Table 3 demonstrate that the results obtained with ICMD using the two types of restraints are not significantly different. This concerns the quality scores, the radii of gyration, the $RMSD_{conf}$ values, as well as the

Table 3. Comparison of the RECOORD conformations from the CNS and CNW sets with the 25 best ICMD conformations calculated using floating (blup) and ambiguous restraints (adr)

Name	set	bph (a)	PROCHECK % core (b)	RMS of violations (c)	WHATIF bump (d)	% Sidechains (e)	$RMSD_{conf}$ (f)	$RMSD_{diff}$ (g)	R_{gyr} (h)
L25	cns	8	48.8	0.03 (0.07)	20.7	56.4	1.5	0.8	13.3 ± 0.1
L25	cnw	7	66.1	0.03 (0.07)	6.6	28.7	1.6	0.0	12.9 ± 0.1
L25	blup	0	71.3	0.13 (0.12)	16.2	79.2	1.6	1.2	13.1 ± 0.1
L25	adr	1	75.8	0.09	15.4	82.7	1.6	1.5	13.1 ± 0.1
L30	cns	4	78.1	0.01 (0.02)	4.8	69.5	0.9	0.7	13.5 ± 0.1
L30	cnw	0	80.4	0.01 (0.02)	2.5	44.7	1.1	0.0	13.0 ± 0.1
L30	blup	1	80.3	0.05 (0.07)	9.0	89.3	1.1	0.7	13.1 ± 0.2
L30	adr	0	81.5	0.05	9.2	92.1	1.3	0.6	13.1 ± 0.1
L11	cns	1	83.5	0.01 (0.03)	6.7	90.7	1.8	0.9	14.1 ± 0.3
L11	cnw	0	78.8	0.01 (0.03)	3.6	36.7	1.8	0.0	13.3 ± 0.4
L11	blup	1	78.0	0.06 (0.06)	8.4	85.8	2.1	1.1	13.2 ± 0.5
L11	adr	1	81.1	0.05	8.8	85.6	2.4	1.1	13.2 ± 0.6
L20	cns	24	71.6	0.04 (0.08)	25.3	44.1	0.9	0.5	10.7 ± 0.1
L20	cnw	1	77.3	0.04 (0.09)	7.6	41.1	0.8	0.0	10.6 ± 0.1
L20	blup	1	78.7	0.19 (0.17)	10.9	68.9	0.9	0.6	10.7 ± 0.1
L20	adr	0	78.8	0.12	7.8	75.2	0.9	0.5	10.8 ± 0.1
L23	cns	20	73.6	0.06 (0.04)	23.2	58.0	0.5	1.0	14.6 ± 0.2
L23	cnw	11	81.6	0.06 (0.03)	10.8	48.0	0.6	0.0	14.1 ± 0.3
L23	blup	1	81.8	0.09 (0.09)	10.3	85.0	0.8	1.8	14.6 ± 0.5
L23	adr	0	81.5	0.05	8.2	85.8	0.8	1.0	14.7 ± 0.5
S28e	cns	6	70.0	0.01 (0.04)	7.1	61.2	0.9	1.8	14.1 ± 0.6
S28e	cnw	5	72.7	0.01 (0.04)	5.0	24.2	1.1	0.0	13.0 ± 0.6
S28e	blup	1	69.5	0.06 (0.05)	9.6	79.8	0.9	1.6	13.2 ± 0.7
S28e	adr	3	69.4	0.05	9.4	78.8	1.0	2.0	13.4 ± 0.7
S19	cns	4	64.4	0.02 (0.06)	16.0	62.9	0.7	0.6	12.2 ± 0.2
S19	cnw	0	71.2	0.03 (0.07)	7.9	52.2	0.8	0.0	11.8 ± 0.2
S19	blup	4	72.4	0.13 (0.11)	12.5	87.5	0.7	0.7	11.9 ± 0.3
S19	adr	0	71.7	0.12	11.8	86.6	0.9	1.0	12.2 ± 0.3
S27e	cns	15	70.8	0.05 (0.04)	4.7	75.7	1.0	0.8	11.5 ± 0.1
S27e	cnw	1	77.0	0.04 (0.04)	3.4	62.0	1.2	0.0	11.0 ± 0.1
S27e	blup	1	70.2	0.05 (0.04)	6.8	89.6	1.0	1.2	11.1 ± 0.2
S27e	adr	1	71.2	0.01	7.6	86.9	1.2	1.4	11.3 ± 0.2

The calculations were performed on the set of proteins described in Table 2. The analyzed parameters are: (a) the total number of hydrogen pairs closer than 1.5 Å, (b) the mean % of residues in the PROCHECK Ramachandran core, (c) the RMS of NOE violations (Å). Numbers in parentheses are calculated on the minimized conformations described in Figure 4. (d) The mean number of WHATIF bumps, (e) the percentage of WHATIF bumps involving one or two sidechain atoms. (f) The RMSD value ($RMSD_{conf}$: Å) calculated between the conformations. (g) The RMSD value ($RMSD_{diff}$: Å) of the average structures (cns, blup, adr) with respect to the cnw average structure. (h) Mean values and standard deviations (Å) of the protein radius of gyration.

population of *cis*-peptide conformers at different stages of the refinement (not shown). The number of consistent violations is sometimes different, but one should keep in mind that this number is evaluated differently for ambiguous and floating restraints (see Methods). Using a single universal definition would be misleading because structures obtained with ambiguous restraints exhibit an exaggerated apparent number of floating violations and *vice versa*. We will show below that the apparently different number of NOE violations in structures produced with floating and ambiguous restraints disappears when the corresponding conformations are re-minimized in identical conditions.

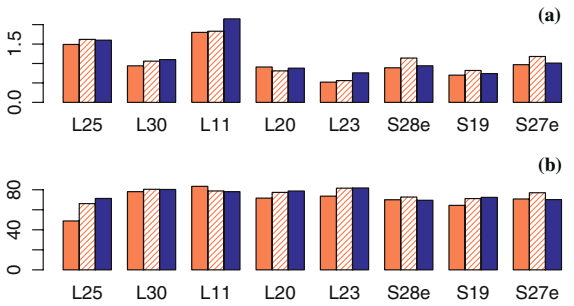

Figure 3. Comparison of the $RMSD_{conf}$ value (Å) (a), and of the percentage of residues in the core Ramachandran regions (b), for calculations run on the set of proteins from Table 2. The results shown are those obtained using CNS (red), CNW (shaded red) and ICMD with floating restraints (dark blue). The compared values are mean values calculated on the set of the analyzed conformations.

Comparison between ICMD and RECOORD

In order to compare the ICMD methodology with the state-of-the-art methods in NMR structure determination, we compared the 25 lowest energy conformations obtained with ICMD with the conformations of the same proteins, stored in the database RECOORD (Nederveen et al., 2005). RECOORD is an effort of protein structure recalculation, based on the most recent protocols and softwares available (CYANA (Herrmann et al., 2002), CNS (Brunger et al., 1998b)) using ARIA simulated annealing scripts. The comparison presented here was performed with the conformations obtained in the gas-phase using CNS (the CNS set) and with the conformations obtained after a further refinement step in water (the CNW set) (Linge et al., 2003). Both comparisons are important because the ICMD approach can be considered as an intermediate between the CNS and CNW protocols. A comparison with the conformations calculated with CYANA (CYA and CYW sets) gives results similar to those obtained in the ICMD/CNS comparison.

Within each set of conformations (CNS, CNW, and ICMD), the RMSD value $RMSD_{conf}$ between the best 25 conformations are shown in Figure 3a and in Table 3(f, g). The variation of the $RMSD_{conf}$ values between the different sets is small (0.1–0.6 Å): there is no large difference of convergence, whatever the method used. One can notice that larger $RMSD_{conf}$ values are generally obtained within the ICMD conformations than within RECOORD conformations. The largest differences of $RMSD_{conf}$ values are observed for proteins L23 and L11. To compare the structures obtained in each set of calculation, the RMSD values $RMSD_{diff}$ between the average CNW structure and the three average structures of the sets CNS and ICMD, are shown in Table 3. The $RMSD_{diff}$ values are smaller than the $RMSD_{conf}$ values, except for L23, S28e and S27e. For the majority of proteins, the difference between the structures is within the conformational variability of each structure, and the structures are thus very similar. The large $RMSD_{diff}$ observed for L23, S28e and S27e arise from local conformational variability in loops and in helix 15–23 of L23.

For the majority of cases studied here, the largest $RMSD_{conf}$ values are obtained in the ICMD structures. The use of the ICMD protocol thus produces a slightly better exploration of the conformational space than the other methods. In RECOORD, the refinement of the protein structure by a short molecular dynamics simulation in water (CNW set) increases the RMSD values with respect to the CNS set. The use of ICMD has thus an effect similar to those produced by a molecular dynamics simulation in water: this may come from the general purpose force-field used for the ICMD calculations.

The largest percentages of residues in the core Ramachandran region are observed (Figure 3b and Table 3(b)) for the ICMD structures of L25, L30, L20, L23 and S19.

The quality of NMR structures depends upon the relative weight attributed to violations of experimental data with respect to the conformational energy. By changing this balance, the

number of violated restraints can always be reduced at the expense of degraded chemical geometry, and *vice versa*. In order to compare equivalent objects, we decided to energy-minimize the CNS, CNW and ICMD floating conformations in identical conditions. To this end, the internal biopolymer geometry was set free and all ICMD floating conformations were re-minimized. Simultaneously, the corresponding structures were imported from RECORD database to ICMD and energy minimized in identical conditions. The results of these calculations are compared in Figure 4 .

Similar numbers of consistent violations are obtained for the three sets of conformations (Figure 4a), which proves that the differences in these numbers previously observed did not arise from large differences in protein structures. On the other hand, the quality of minimized conformations was evaluated through the percentage of residues in the core Ramachandran regions (Figure 4b). The comparison of the percentage of residues in the core Ramachandran diagram shows that these percentages were not sensibly modified with respect to Figure 3b. The conformation quality is thus not significantly modified by the minimization.

The RMS of NOE violations, averaged over all the violations (Table 3(c)) are larger in ICMD than in RECOORD structures, except for the proteins L23 and S27e. The largest RMS are obtained for structures calculated with floating restraints. Nevertheless, the RMS of NOE violations were calculated on the re-minimized RECORD and ICMD conformations (Table 3(c)) and they increase, usually by 50%, in all RECOORD structures except L23. These numbers are thus not very stable with small structure variations, and their variations are even larger than 0.04 Å for L25 and L20. On the other hand, the RMS of NOE violations decrease in RECOORD structures re-minimized in ICMD. The differences between the RMS of NOE violations in ICMD and in RECOORD structures become smaller than 0.05 Å after the ICMD minimization, in half of the structures considered. The difference observed in ICMD and RECOORD structures for the RMS of the NOE violations are not significant.

The WHATIF quality Z scores (Figure 5) display different trends when the ICMD and the RECOORD structures are compared. The worst Ramachandran (Figure 5a) and 2nd-generation packing (Figure 5b) Z scores are observed for the CNS structures. The best Ramachandran Z scores (Figure 5a) are generally observed in the ICMD structures, whereas the best 2nd-generation packing Z scores (Figure 5b) are generally obtained in the CNW structures. Exceptions are observed for

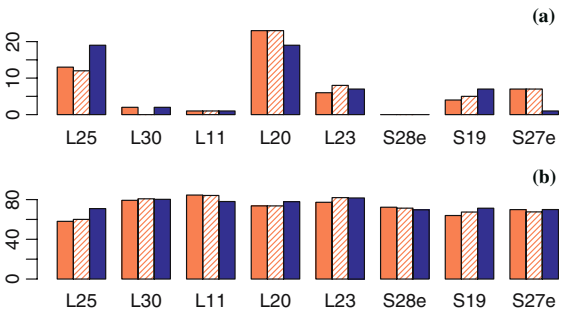

Figure 4. Comparison of the number of consistent violations (a) and of the percentage of residues in the core Ramachandran regions (b) for calculations run on the set of proteins from Table 2. The results shown are those obtained on minimized ICMD conformations obtained with floating restraints (dark blue), obtained on CNS conformations minimized with floating restraints (red) and obtained on CNW conformations minimized with floating restraints (shaded red). All minimizations were performed with ICMD. The compared values are mean values calculated on the set of the analyzed conformations.

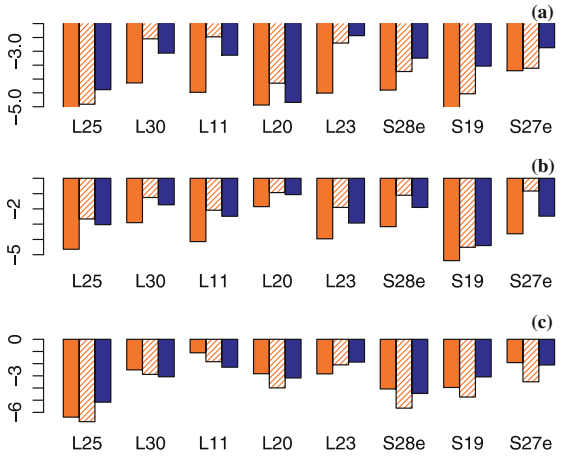

Figure 5. Comparison of the WHATIF Ramachandran Z score (a), of the 2nd-generation packing Z score (b), and of the backbone conformation Z score (c) for calculations run on the set of proteins from Table 2. The results shown are those obtained using CNS (red), CNW (shaded red), ICMD with floating restraints (dark blue). The compared values are mean values calculated on the set of the analyzed conformations.

the proteins L30, L11 and L20, for which the CNW structure exhibit the best Ramachandran Z scores, and for protein S19, for which ICMD structure displays the best 2nd-generation packing Z score.

Concerning the backbone conformation Z score (Figure 5c), the worst values are mainly exhibited by the RECOORD structures, except for the proteins L30 and L11. For all calculations, the residues showing most of the poor ϕ, ψ values are mainly located in loops, C- and N-terminal tails.

The number of hydrogen pairs closer than 1.5 Å, are smaller in ICMD than in RECOORD structures (Table 3(a)). On the other hand, the mean numbers of bumps detected by WHATIF (Table 3(d)) are in the 2.5–25.3 range for the RECOORD structures, and in the 7.6–16.2 range for the ICMD structures. A closer inspection reveals that the bumps in ICMD and in RECOORD conformations are somewhat different on average. In RECOORD structures, for all proteins except L11 (CNW set) the percentage of bumps where one or two sidechain atoms are involved is in the 24.2–75.7 range (Table 3(e)). For the ICMD structures, the fraction of such bumps is significantly higher (68.9–92.1%). As the sidechain atoms have a higher probability to be located in more mobile regions, the bumps between such atoms are less severe. The decrease of the percentage of main chain atoms involved in the bumps is also in agreement with the improvement of the Ramachandran Z scores in ICMD structures. Furthermore, visual inspection of the residues involved in the bumps, shows that their sidechains are mainly directed towards the protein exterior. The bumps observed in ICMD structures should thus have a more limited impact on the molecular dynamics trajectories starting from these structures (Fan and Mark, 2003).

The mean radii of gyration of the ICMD structures were calculated and compared to those of the CNS and CNW sets (Table 3(h)). For proteins S28e and S27e the polypeptide chain considered in the ICMD calculation is larger than the ones in the PDB and RECOORD databases. Therefore, some residues are unrestrained, and their conformational variability increases significantly the radius of gyration. To remove this bias, the radii of gyration of S28e and S27e were calculated by selecting only the sub-sequence present in the RECOORD structures. The mean radii of gyration are smaller in ICMD structures than in CNW one, for protein L11. The other proteins the mean radii of gyration are larger in ICMD structures than in the CNW structures, but these differences are smaller than or equal to the sum of the corresponding standard deviations. One can conclude, therefore, that the use of the all atom force-field in the gas-phase calculations in ICMD introduces no undesirable general bias towards more compact or more extended protein structures.

The CPU time required to produce one conformation with ICMD was in the 1–3 h range on a processor operating at 2.4 GHz for the proteins analyzed here. On the other hand, the CPU time needed for the production of one conformation using the CNS (respectively CNW) protocol is 15 min (respectively 40 min). The ICMD protocol is thus more time-consuming than the CNS and CNW protocols, which is not surprising since the number of integration steps is 20-fold larger. The number of steps in the CNS protocol is 20,000, whereas the number of steps for ICMD is $3100 + (100 n_{\max})$, where $n_{\max}$ is the number of TAS levels contained in the set of restraints. Although there are many possibilities for improving this timing we did not work on that because, with the CPU speed currently available, this factor becomes essential only for large proteins not considered here. The CPU time used for running the ICMD protocol varies much within the set of proteins studied, as the variable target function algorithm (Braun and Go, 1985) may behave in a very different way for α and β secondary structures, and depends on the protein size. The protocol is also certainly sensitive to the consistency of the restraints, as too many inconsistent restraints produce repeated cycles of variable target phase (Table 1: lines 3–9), and thus delay the conformation calculation.

In summary, the comparison between the RECOORD and the ICMD structures shows that the ICMD methodology exhibits an efficiency similar to those of the state-of-the-art methods, as far as the convergence of the calculation and the quality of the obtained structures are concerned. The quality of the ϕ, ψ distributions is slightly better in the ICMD than in the RECOORD structures. The bumps detected by WHATIF in the ICMD structures concern mainly sidechain atoms, in contrast to the observations made in the RECOORD structures. The radii of gyration

display similar values in ICMD and in RECOORD structures. The $RMSD_{conf}$ values are slightly larger, which is the sign of a better exploration of the conformational space. Thus, the ICMD approach permits a slight improvement of the quality of the obtained structures as well as of the exploration of the conformational space, justifying the more important computational effort.

Conclusion

With the ICMD methodology, structures of a quality similar to that observed in RECOORD, are obtained, and ICMD is thus competing well with the current NMR refinement methods. The calculations performed here also indicate that it may be important to use a general purpose force-field, including the Coulombic and Lennard-Jones non-bonded interactions, from the beginning of the refinement. Indeed, ICMD, which uses such a force-field, produces conformations, with generally better Ramachandran Z scores than the RECOORD structures. The use of a relatively short (6 Å) cutoff value for the non-bonded interactions reduces the influence of long-range interactions during the high-temperature phase of the refinement, and probably improves the calculation stability. The use of a slower cooling protocol in ICMD can be another reason for obtaining better Ramachandran Z scores in the structures. This tendency is in agreement with the observation recently made (Fossi et al., 2005) that a slow cooling in structure determination is more productive.

The ICMD conformations produced with the floating restraints often exhibit a larger number of consistently violated restraints, but this apparent difference results from the definition of restraints and is canceled by minimization of the conformations in identical conditions.

The residues lowering the structure quality are mainly located in the less-defined parts of the molecules, in long loops and tails. However, the methods to check structure quality rely on the knowledge of the X-ray crystallographic structures, and the flexible parts (loops and tails) are usually invisible in the electronic densities (Kwasigroch et al., 1997). Thus, the poor quality Z scores due to residues located in such protein regions may arise from the protein structure itself as well as from the lack of database knowledge from X-ray structures.

A larger computational time is required to perform a refinement using the ICMD protocol. Nevertheless, as the ICMD methodology permits to run molecular dynamics in torsion space, with only one evaluation of the energy per step, it has a relatively fast integration scheme, and is particularly well-adapted to the protocol proposed here.

The protocol presented here was developed in order to demonstrate the possibility of a structure calculation completely performed with all terms of a general purpose force-field, including the Coulombic and Lennard-Jones interactions. The authors believe that such an approach can modify the further analysis of structures, as well as their use in molecular modeling studies. In a certain sense, the ICMD approach can be considered as intermediate between the CNS and CNW protocols. Compared to the latter two methods, the $RMSD_{conf}$ values observed for the ICMD results are between those of the CNS and CNW. At the same time, the percentage of residues in the core Ramachandran diagram as well as the WHATIF Z scores for ICMD structures are most often slightly superior to both CNS and CNW sets. The number of WHATIF bumps and the RMS of violations obtained with ICMD are worse than or intermediate between the CNS and CNW results.

The difference between ICMD and the approaches classically used for NMR structure determination reside mainly in the simulated annealing protocol. Indeed, the algorithm used for the integration in the torsion angle space is not very different from others published in the literature (Schwieters and Clore, 2001). However, the simultaneous use of TAD and of a general purpose force-field to describe the interactions between atoms during an NMR structure determination, was not, to our knowledge, proposed until now.

Acknowledgements

Benjamin Bardiaux thanks the European Union Grant SPINE (QLG2-CT-2002-00988) and the ''Ministère de l'Enseignement Sup'erieur'' for the financial support in the frame of the ACI IMPBio (project ICMD-RMN). The authors thank the ACI IMPBio, the CNRS and the Institut Pasteur for funding.

References

Abagyan, R. and Mazur, A. (1989) *J. Biomol. Struct. Dyn.* **6**, 833–845.

Aramini, J., Huang, Y., Cort, J., Goldsmith-Fischman, S., Xiao, R., Shih, L., Ho, C., Liu, J., Rost, B., Honig, B., Kennedy, M., Acton, T. and Montelione, G. (2003) *Protein Sci.* **12**, 2823–2830.

Berendsen, H., Postma, J., Gunsteren, W.van and DiNola, A. (1984) *J. Chem. Phys.* **81**, 3684–3690.

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

Brunger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T. and Warren, G. (1998a) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.

Brunger, A., PD, A., Clore, G., DeLano, W., Gross, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T. and Warren, G. (1998b) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.

Chao, J. and Williamson, J. (2004) *Structure* **12**, 1165–1176.

Chen, Y., Bycroft, M. and Wong, K. (2003) *Biochemistry* **42**, 2857–2865.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) *J. Am. Chem. Soc.* **117**, 5179–5197.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR* **13**, 289–302.

Doreleijers, J., Raves, M., Rullmann, T. and Kaptein, R. (1999) *J. Biomol. NMR* **14**, 123–132.

Fan, H. and Mark, A. (2003) *Proteins* **53**, 111–120.

Fossi, M., Oschkinat, H., Nilges, M. and Ball, L. (2005) *J. Magn. Reson.* **175**, 92–102.

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

Helgstrand, M., Rak, A., Allard, P., Davydova, N., Garber, M. and Hard, T. (1999) *J. Mol. Biol.* **292**, 1071–1081.

Herrmann, T., Güntert, P. and Wüthrich, K. (2002) *J. Mol. Biol.* **319**, 209–227.

Herve du Penhoat, C., Atreya, H., Shen, Y., Liu, G., Acton, T., Xiao, R., Li, Z., Murray, D., Montelione, G. and Szyperski, T. (2004) *Prot. Sci.* **13**, 1407–1416.

Hooft, R., Vriend, G., Sander, C. and Abola, E. (1996) *Nature* **381**, 272–272.

Jabs, A., Weiss, M. and Hilgenfeld, R. (1999) *J. Mol. Biol.* **286**, 291–304.

Jain, A., Vaidehi, N. and Rodriguez, G. (1993) *J. Comput. Phys.* **106**, 258–268.

Kordel, J., Pearlman, D. and Chazin, W. (1997) *J. Biomol. NMR* **10**, 231–243.

Kozin, S.A., Bertho, G., Mazur, A.K., Rabesona, H., Girault, J.P., Haertle, T., Takahashi, M., Debey, P. and Hoa, G.H. (2001) *J. Biol. Chem.* **276**, 46364–46370.

Kwasigroch, J., Chomilier, J. and Mornon, J. (1997) *J. Mol. Biol.* **259**, 855–872.

Laskowski, R., MacArthur, M., Moss, D. and Thornton, J. (1993) *J. Appl. Cryst.* **26**, 283–291.

Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. and Daggett, V. (1997) *J. Phys. Chem.* **101**, 5051–5061.

Linge, J. and Nilges, M. (1999) *J. Biomol. NMR* **13**, 51–59.

Linge, J., Williams, M., Spronk, C., Bonvin, A. and Nilges, M. (2003) *Proteins* **50**, 496–506.

Lu, M. and Steitz, T. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2023–2028.

Mao, H. and Willamson, J. (1999) *J. Mol. Biol.* **292**, 345–359.

Markus, M., Hinck, A., Huang, S., Draper, D. and Torchia, D. (1997) *Nat. Struct. Biol.* **4**, 70–77.

Mazur, A. and Abagyan, R. (1989) *J. Biomol. Struct. Dyn.* **6**, 815–832.

Mazur, A.K. (1997) *J. Comput. Chem.* **18**, 1354–1364.

Mazur, A.K. (1998a) *J. Am. Chem. Soc.* **120**, 10928–10937.

Mazur, A.K. (1998b) *J. Phys. Chem. B* **102**, 473–479.

Mazur, A.K. (1999) *J. Chem. Phys.* **111**, 1407–1414.

Mazur, A.K. (2001) In *Computational Biochemistry and Biophysics*, Becker, O.M., MacKerell, A.D. Jr., Roux, B., Watanabe, M. (Eds.), Marcel Dekker, New York, pp. 115–131.

Mazur, A.K. (2002) *J. Am. Chem. Soc.* **124**, 14707–14715.

Miclet, E., Duffieux, F., Lallemand, J. and Stoven, V. (2003) *J. Biomol. NMR* **25**, 249–250.

Nederveen, A., Doreleijers, J., Vranken, W., Miller, Z., Spronk, C., Nabuurs, S., Guntert, P., Livny, M., Markley, J., Nilges, M., Ulrich, E., Kaptein, R. and Bonvin, A.M. (2005) *Proteins* **59**, 662–672.

Nilges, M. (1993) *Proteins* **17**, 297–309.

Ohman, A., Rak, A., Dontsova, M., Garber, M. and Hard, T. (2003) *J. Biomol. NMR* **26**, 131–137.

Raibaud, S., Lebars, I., Guillier, M., Chiaruttini, C., Bontems, F., Rak, A., Garber, M., Allemand, F., Springer, M. and Dardel, F. (2002) *J. Mol. Biol.* **323**, 143–151.

Schwieters, C. and Clore, G. (2001) *J. Magn. Reson.* **152**, 288–302.

Stein, E., Rice, L. and Brunger, A. (1997) *J. Magn. Reson.* **124**, 154–164.

Stoldt, M., Woehnert, J., Goerlach, M. and Brown, L. (1998) *EMBO J.* **17**, 6377–6384.

Stoldt, M., Wohnert, J., Ohlenschlager, O., Gorlach, M. and Brown, L. (1999) *EMBO J.* **18**, 6508–6521.

Zirah, S., Kozin, S.A., Mazur, A., Blond, A., Cheminant, M., S'egalas-Milazzo, I., Debey, P. and Rebuffat, S. (2006) *J. Biol. Chem.*, in press.