A Reassessment of the Structure of Chymotrypsin Inhibitor 2 (CI-2) Using Time-Averaged NMR Restraints†

Alain P. Nanzer,‡ Flemming M. Poulsen,§ Wilfred F. van Gunsteren,‡ and Andrew E. Torda*,‡

Department of Physical Chemistry, Swiss Federal Institute of Technology Zürich, 8092 Zürich, Switzerland, and Kemisk Afdeling, Carlsberg Laboratorium, Gamle Carlsberg Vej 10, DK-2500 Valby, Copenhagen, Denmark

Received May 10, 1994; Revised Manuscript Received September 9, 1994®

ABSTRACT: Chymotrypsin inhibitor 2 (CI-2) is one of the growing family of proteins for which well-defined solution and crystal structures have been published and for which small, but distinct differences between these were found. It presents an ideal case to address the question of whether a structural difference is physically real or due to the simplifying approximations with respect to averaging that are used in the conventional methods for structure refinement. NOE distance and ^3J coupling constant restrained molecular dynamics simulations were performed using conventional and time-averaged restraints, both in vacuo and in aqueous solution, and the trajectories were compared with structural properties of published structures. The time-averaged restrained molecular dynamics simulations sampled more conformations at various times and visited states consistent with both previously published solution and crystal structures. It was found that the difference between these structures is due to the refinement methodology used. Application of time-averaged restraints in structure refinement yields a physically different picture of the molecular mobility.

Although NMR[1] spectroscopy continues to provide ever better determined structures of proteins, the way dynamics is accounted for remains an issue which is not fully resolved. All NMR spectroscopic parameters are the result of averaging on the NMR time scale, and the kind of averaging in each case is distinctly nonlinear (Jardetzky, 1980). For example, neglecting spin diffusion, nuclear Overhauser effect distances are an r^{-6} average, a value which is dominated by small distances. Similarly, ^3J scalar coupling constants are related to angles via the Karplus curve, a quadratic function with respect to the cosine of the included torsion angle (Karplus, 1959). Clearly, viewing NMR structures as some kind of geometric average over a molecule's allowed space is not strictly correct.

At the same time, X-ray crystallographic measurements are subject to their own (longer time scale) averaging, and the derived structures are the result of a series of assumptions. Most protein structures are refined with isotropic rather than anisotropic temperature factors, and most atom positions are probably not accurately represented by the gaussian distribution functions employed.

Given this series of approximations, one might expect that even if a protein adopts identical structures in solution and crystal, the biases of the different methods would result in different structures. Furthermore, one might expect the differences to be more pronounced the better and more complete each data set is. These thoughts lead to an interesting question when comparing structures determined by different means. If one is faced with apparent differences between solution and crystal structures, can they be reconciled by better modeling one or both forms of experimental data? One step in this direction would be to model the NMR distance and J-coupling restraints directly as time averages (Torda *et al.*, 1989, 1993). The comparison of structures then becomes an issue of whether the ensemble of structures used to model the NMR data now reasonably includes that static structure regarded as the crystallographic structure.

The 64-residue structured domain of chymotrypsin inhibitor 2 (CI-2) is an excellent candidate for addressing some of these questions. CI-2 is one of the few proteins, for which well-resolved structures for both crystal and the solution phases have been published and deposited with the Protein Data Bank. It is an 83-residue protein inhibiting serine proteinases although no structure of the first 20 N-terminal residues was observed in either the solution or in the crystal studies.

The CI-2 crystal structure was resolved and refined by McPhalen and co-workers (McPhalen *et al.*, 1983, 1985; McPhalen & James, 1987) in several steps. The three-dimensional structure has been determined at 2-Å resolution by the molecular replacement method and was refined afterward by restrained-parameter least-squares methods to a crystallographic R factor of 0.198.

The NMR structure has been published (Clore *et al.*, 1987a; Ludvigsen *et al.*, 1991a,b) and compared with the X-ray structure (Clore *et al.*, 1987b; Ludvigsen *et al.*, 1991b; Billeter, 1992) several times, each time with quantitatively more and qualitatively better experimental data.

Not surprisingly, crystal and solution forms of the structure are essentially the same. One unexpected result, however, was the additional pair of antiparallel β-strands, consisting

† Financial support was obtained from the Schweizerischer Nationalfond (project 21-35909.92), which is gratefully acknowledged.

* Author to whom correspondence should be addressed. E-mail: ana@igc.chem.ethz.ch, carlfmp@unidhp.uni-c.dk, wfvgn@igc.chem.ethz.ch, and torda@igc.chem.ethz.ch.

‡ Swiss Federal Institute of Technology Zürich.

§ Carlsberg Laboratorium.

® Abstract published in *Advance ACS Abstracts,* November 1, 1994.

[1] Abbreviations: CI-2, chymotrypsin inhibitor 2; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; MD, molecular dynamics.

0006-2960/94/0433-14503$04.50/0

of residues "(65−67) and (81−82) that was not seen in either the crystal structure nor the previous solution structure" (Ludvigsen et al., 1991b). In contrast, the crystallographic structure contains a series of water molecules between the two strands. Here, we would like to investigate if a long time molecular dynamics simulation could satisfy both views. To this end, we conducted a series of molecular dynamics simulations using conventional and time averaged restraints, both *in vacuo* and in solution and starting both from crystal and solution structures.

METHODS

(a) Theory. Molecular dynamics simulations are usually used for refining NMR based structures by constructing an artificial energy term which raises the energy of the system as violations of the experimental data increase (van Gunsteren *et al.*, 1984; Kaptein *et al.*, 1985). In the case of the GROMOS force field, this term is quadratic with respect to violations of distance constraints, so

$$V_{dr}(r) = \frac{1}{2}K_{dr}(r - r^0)^2 \quad \text{if } r > r^0 \qquad = 0 \quad \text{if } r \le r^0 \tag{1}$$

where $V_{dr}(r)$ is the potential energy due to the distance restraint term for a given pair of atoms, r is the instantaneous distance between the cross-relaxing nuclei, and r^0 is the distance calculated from the measured NOE. A force constant, K_{dr}, is used to weight this term relative to the rest of the force field.

Torda *et al.* (1990) proposed an alternative approach in which the potential energy was not formally defined at all. Instead only a force was constructed so that

$$\vec{F}_i(t) = -K_{dr}(\bar{r}_{ij}(t) - r^0)\frac{\vec{r}_{ij}(t)}{r_{ij}(t)} \quad \text{if } \bar{r}(t) > r^0 \qquad = 0 \quad \text{if } \bar{r}(t) \le r^0 \tag{2}$$

where $\vec{F}_i(t)$ is the force on atom i due to atom j, at time t, and $r_{ij} = r_i - r_j$, the exponentially weighted time average, is

$$\bar{r}(t) = \left(\frac{1}{\tau_{dr}(1 - e^{-t/\tau_{dr}})}\int_0^t e^{-t'/\tau_{dr}}[r(t - t')]^{-3}\,dt'\right)^{-1/3} \tag{3}$$

where τ_{dr} is the characteristic time for the exponential decay. Integration of eq 2 with respect to the instantaneous distance would lead to some expression for the instantaneous potential energy, but the inherent time dependence of the term means that this loses its normal physical meaning. Effectively, this is no longer a conservative force field, and it would be inappropriate to treat it as such. For this reason, we do not refer to a restraint energy.

For each measured ^3J value, a restraining potential V_J was calculated so as to directly restrain the coupling constants rather than the calculated dihedral angle. Static restraints were enforced according to eq 4 (Kim & Prestegard, 1990), and time-averaged restraints were imposed according to eq 5 (Torda *et al.*, 1993)

$$V_J = \frac{K_J}{2}[J(\theta) - J_0]^2 \tag{4}$$

$$V_J = \frac{K_J}{2}[\bar{J}[\theta(t)] - J_0]^2 \tag{5}$$

where $J(\theta)$ is the coupling constant calculated from the dihedral angle θ, J_0 is the measured value, K_J is a force constant, and $\bar{J}[\theta(t)]$ is the time averaged coupling constant, calculated according to

$$\bar{J}[\theta(t)] = \frac{1}{\tau_j(1 - e^{-t/\tau_j})}\int_0^t e^{(-t'/\tau_j)}J[\theta(t - t')]\,dt' \tag{6}$$

Again τ_j is the characteristic time for the exponential decay. Note that, as in our previous work, we assume that $t \gg \tau$ so

$$1 - e^{-t/\tau} \approx 1 \tag{7}$$

and this term is dropped from both eq 3 and eq 6.

(b) Molecular Model and Simulation Set Up. All simulations were carried out using software from the GROMOS suite of programs and united atom force field (van Gunsteren & Berendsen, 1987). The temperature was held constant by weak coupling to an external bath. For simulations with solvent, the pressure was similarly coupled (Berendsen *et al.*, 1984). Nonbonded interactions were treated using a twin range method (van Gunsteren & Berendsen, 1990). In this approach all interactions within a short-range cutoff radius (8 Å) are determined every time step, while longer range interactions within a second cutoff radius (12 Å) are only calculated during the update of the pair list every 10 time steps. The SHAKE algorithm was used to maintain all bond lengths (Ryckaert *et al.*, 1977) and the integrator time step was 0.002 ps.

Simulations with time-averaged distance restraints require the choice of an initial value for $\bar{r}(t)$ and time intervals τ_{dr} and τ_j for averaging J-coupling and distance restraints. For all the runs using time-averaged distance restraints, $\bar{r}(0)$ was set, for each distance restraint, to 0.2 Å less than r_0, and the time intervals to average J-coupling and distance restraints were set to $\tau_{dr} = \tau_j = 20$ ps. Each simulation was carried out once with conventional J-coupling ($\tau_j = 0$) and distance restraints ($\tau_{dr} = 0$) and once time averaged.

As starting coordinates of the CI-2 solution structure, we used published data (Ludvigsen *et al.*, 1991b), obtained from the Protein Data Bank at Brookhaven National Laboratory (Bernstein *et al.*, 1977; Abola *et al.*, 1987), entry 3CI2. Similarly the 961 experimental distance and the 39 J-coupling restraints were also as previously published (Ludvigsen *et al.*, 1991a,b).[2] Starting coordinates of CI-2 crystal structure were from the same origin, entry 2CI2, published by McPhalen and James (1987).

To quantify differences between conformations, we used the root mean square (rms) difference between intramolecular distances (Rooman *et al.*, 1990) and (Levitt, 1983). The difference D_{ab} between two conformations, a and b is defined as

[2] Experimental data were submitted to the Protein Data Bank and are available from the original authors.

$$D_{ab} = \left(\frac{2}{N(N-1)} \sum_{i>j}^{N} (d_{ij}^{a} - d_{ij}^{b})^2 \right)^{1/2} \quad (8)$$

where the summation runs over all pairs ij of the N atoms being considered in the configurations a and b. d_{ij} is the three dimensional distance between atoms i and j. This measure has the advantage that it obeys the triangle inequality and D_{ab} is well correlated with the difference between structures a and b. It has the clear disadvantage that it is not sensitive to chirality, so in the extreme case, $D_{ab} = 0$ for mirror image structures. We do not expect chiral inversions or changes in overall fold during a MD simulation.

(c) Protocol of MD Simulations. The 20 previously published solution structures were first subjected to 1000 steps of restrained steepest descents energy minimization. After minimization, two MD simulations were performed (10 and 20 ps) to equilibrate the system in the GROMOS force field. The force constant for distance restraints was $K_{dr} = 60$ kJ mol^{-1} Å^{-2}, and the force constant for ^3J-value restraints was in all runs $K_J = 20$ kJ mol^{-1} Hz^{-2}. Afterward, the structures were again energy minimized with the steepest descents method to have comparable structures for choosing the "best" one, for further examination. The potential energies varied from −69 to −74 kJ mol^{-1}, the sum of violations spanned a range from 96 to 114 Å. Both of these parameters were used as a criterion for selection. The chosen structure had a potential energy of −74 kJ mol^{-1} and a sum of violations of 98 Å. No further use was made of the remaining 19 structures.

This structure, after the initial 30-ps equilibration MD, was then used as the starting point for two long (1 ns) parallel MD simulations. One calculation used conventional distance restraints ($\tau_{dr} = \tau_J = 0$ ps), and one employed time-averaged distance and J-coupling restraints ($\tau_{dr} = \tau_J = 20$ ps). This resulted in two long *in vacuo* trajectories analyzed in some detail below.

Four solution simulations were then conducted. The first started from the same NMR-based structure as the *in vacuo* simulations. Since we were interested in apparent differences between the solution and crystal structures, an identical set of simulations were conducted starting from a crystal structure. Again, parallel runs were conducted using either conventional or time averaged distance and J-coupling restraints.

For the simulations in solution, both starting structures were placed in a truncated octahedron and surrounded by 3132 water molecules, which were modeled by a simple rigid three point charge model (SPC) (Berendsen et al., 1981). Each simulation in water began with a 10-ps period in which the protein atoms were harmonically restrained to their starting positions. The force constant of the position restraining potential energy term was 90 kJ mol^{-1} Å^{-2} and only water atoms were allowed to move freely. A further 30 ps of equilibration MD was conducted with all atoms free to move (τ_{dr}, τ_J, K_{dr}, and K_J as in the *in vacuo* runs), followed by the production MD simulations of 500 ps length.

(d) Comparison with the Experimental NMR Data. For analyzing the agreement of the simulations with the experimental data, some parameters of the calculation were averaged over the whole trajectory and compared with experimental data. Table 1 shows the sum of violations and the averaged total potential energy for the 30-ps equilibration MD and for the six long 1-ns *in vacuo* and 0.5-ns in solution simulation trajectories.

Table 1: Sum of Violations and Total Potential Energy for Equilibration and Long MD Simulations[a]

	NMR ta *in vacuo*	NMR cr *in vacuo*	NMR ta in H_2O	NMR cr in H_2O	X-ray ta in H_2O	X-ray cr in H_2O
Σviol (Å)						
equil	40	40	42	42	36	36
long	32	10	24	10	24	10
E_{pot} (kJ mol^{-1})						
equil	89.6	89.6	−92703	−92703	−91247	−91247
long	−127.3	−37.2	−95356	−94863	−95100	−94957

[a] All parameters averaged over the trajectory. Columns: NMR ta, time-averaged restrained MD simulation starting from a NMR structure; NMR cr, conventionally restrained MD simulation starting from a NMR structure; X-ray ta, time-averaged restrained MD simulation starting from an X-ray structure; X-ray cr, conventionally restrained MD simulation starting from an X-ray structure.

Table 2: Number of Distance Restraint Violations of Different MD Simulations Sorted in Size[a]

no. of violations	NMR ta in vacuo	NMR cr in vacuo	NMR ta in H_2O	NMR cr in H_2O	X-ray ta in H_2O	X-ray cr in H_2O
>0.75 Å	0	0	0	0	0	0
0.50−0.75 Å	1	3	2	2	2	4
0.25−0.50 Å	3	10	15	10	15	14
0.00−0.25 Å	52	32	45	44	47	28

[a] Number of violations averaged over the whole trajectory. See Table 1 footnotes.

The size of the largest single violation is a necessary test of the quality of a structure, since any one big violation would result in local stress in a molecule. In the case of the CI-2 trajectories, with or without time-averaged NOE's, there are no violations larger than 0.7 Å, and in all six simulations there were less than 5 violations between 0.5 and 0.7 Å. The distribution of the violated distances for the six MD runs is shown in Table 2.

Although the distance restraints constitute the bulk of the experimental data, it is also necessary to ensure that the simulation reproduces the measured J-coupling information. The agreement with the experimental J-coupling data is summarized in Figure 1. It shows four Karplus curves, $J(\theta)$, using calibration constants from Ludvigsen *et al.* (1991a). The y value of each point ◇ is a ^3J value calculated by averaging over the whole trajectory. The x value of the point ◇ is an angle calculated from the experimentally measured ^3J value using the same calibration curve. If a trajectory average were to agree perfectly with the measured value, the plotted point would lie directly on the curve. The vertical bar on each point shows the root mean square fluctuation of the calculated ^3J value. Figure 1a shows the calculated ^3J values for the solution MD simulation starting from an NMR structure using time averaged restraints, whereas Figure 1b represents the same for conventional, instantaneous restraining. Figure 1 panels c and d depict the same for the solution MD simulations starting from an X-ray structure.

RESULTS AND DISCUSSION

(a) Agreement with the Experimental NMR Data. The first question arising from a restrained molecular dynamics simulation is how well are the experimental data reproduced?

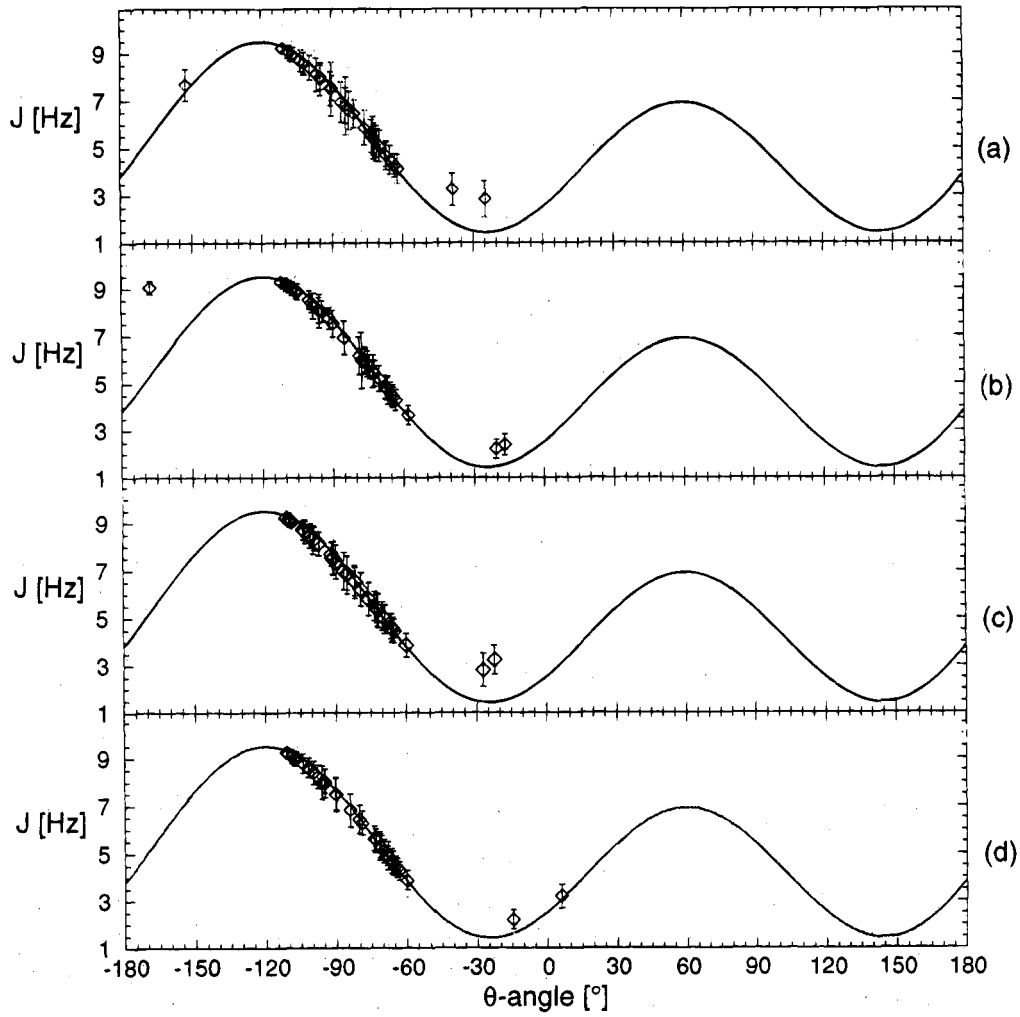

FIGURE 1: Karplus curves $J(\theta)$ (solid line), using the calibration constants of Ludvigsen *et al.* (1991a), $a = 6.7$, $b = -1.3$, and $c = 1.5$. The y value of each point indicated by ◇ is a ^3J value calculated by averaging over the whole MD trajectory. The x value of the point ◇ is an angle calculated from the experimentally measured ^3J value using the same calibration curve. A 60° phase shift is used so the angle is expressed according to conventional protein nomenclature. The vertical line on each point shows the root mean square fluctuation of the calculated ^3J value. (a) Time-averaged restrained MD simulation starting from a NMR structure. (b) Conventionally restrained MD simulation starting from a NMR structure. (c) Time-averaged restrained MD simulation starting from an X-ray structure. (d) Conventionally-restrained MD simulation starting from an X-ray structure.

We first consider distance restraints. The sum of violations, averaged over the whole trajectory, is reduced in all six simulations compared to the values obtained in the equilibration simulations (Table 1). Table 2 lists the violations according to their size and clearly indicates the small divergence of the ensembles of structures compared to the experimental data. In no case is there more than four violations greater than 0.5 Å, and all of the large violations can be explained by two sets of interactions.

First, there is a distance bound from the $C^{\delta 1}$ atom of Trp 24 to both of the stereospecifically assigned β-protons of the same residue. One of these bounds is violated by more than 0.5 Å in all calculations except the time-averaged simulation *in vacuo*. Simulations with explicit solvent do show slower, more damped motions than those *in vacuo*, so it is not surprising that only in the *in vacuo* simulation is the bulky side chain able to reorient often enough to satisfy all the restraints.

Secondly, all of the remaining large violations can be explained by considering the δ protons of Ile 39. This residue is involved in 63 NOEs and the δ methyl group in 14 bounds. The experimental data require the methyl group to be close to the hydrophobic side chains of residues Leu 27, Ile 48, and Val 66. It is possible that either the bounds are calibrated somewhat too tightly or that the force field in the simulations does not allow enough movement of this side chain, even when using time-averaged distance restraints. It is also possible that the time constant used for the averaging of the bounds is too short. This last possibility will be investigated in future calculations. Keeping in mind that the measured NOEs are an average in time, one can claim that the experimental distance restraints are well satisfied.

Although it would appear that all the trajectories agree with the experimental data, Table 1 shows a small but systematic difference between conventional and time-averaged refinement. For all three pairs of simulations, the sum of violations is more than twice as large when using time-averaged restraints. This is in contrast to previous results on comparable systems (Torda *et al.*, 1990; Pearlman & Kollman, 1991). There are several possible reasons for this behaviour. First, if the restraint data really do not suffer from the influence of dynamics, then a static model of the structure and a time-averaged model should both agree equally well with the data. In practice, the time-averaged

Table 3: Distance Matrix Error for Starting Structures and the Average Conformations of the Six MD Simulations Calculated from Similarity Matrices Based on the Rms Difference of Intramolecular Distances[a]

D_{ab} (Å)	NMR start	X-ray start	NMR ta *in vacuo*	NMR cr *in vacuo*	NMR ta in H_2O	NMR cr in H_2O	X-ray ta in H_2O	X-ray cr in H_2O
NMR start		2	1.9	2	2.2	2.2	2.1	2
X-ray start	1.9		2.2	2.4	2.1	2.4	1.8	1.7
NMR ta in vacuo	1	1.9		0.9	1.4	1.3	1.5	1.6
NMR cr in vacuo	1	1.9	0.4		1.5	1.2	1.7	1.7
NMR ta in H_2O	1.2	1.8	0.6	0.7		0.9	0.7	1.1
NMR cr in H_2O	1.1	1.8	0.5	0.5	0.4		1.4	1.5
X-ray ta in H_2O	1.2	1.8	0.6	0.7	0.3	0.5		0.9
X-ray cr in H_2O	1	1.8	0.6	0.6	0.4	0.5	0.4	

[a] Lower triangle: only backbone atoms, excluding those of residues 54−62 are used. Upper triangle: all atoms are used. NMR start: NMR starting structure. X-ray start: X-ray starting structure. See also the footnotes of Table 1.

restraints will allow the system to temporarily violate the restraints, and this may cause small, but systematic violations. It is unlikely that any data set will be totally immune to the influence of motions, but it is a matter of degree. If the calibration of these distance bounds is sufficiently loose, the influence of molecular motion will be reduced. Next, there is the issue of force constants for the nonphysical restraint terms. The higher the force constant for the pseudopotential energy terms, the greater their (positive) contribution to the system's total energy. Table 1 shows that, in each case, the system's average potential energy over the whole trajectory is lower using time averaged restraints.

The simulations were also assessed with respect to their agreement with measured ^3J values. Figure 1 shows a series of Karplus curves (see Methods). All averaged values lie near the curve, and the error bars indicate the larger fluctuation in the time-averaged runs. As with the distance restraints, the trajectory averages agree well with the experimental data.

Thus, in summary, all six long MD simulations agree with the experimental data very well, and the agreement is improved when compared to the initial equilibration period of 30 ps.

(b) The Influence of the Aqueous Environment. The calculations in vacuo served to provide a quick initial assessment of the calculational problem. They demonstrated that there would be no difficulty agreeing with experimental data and, not surprisingly, showed differences in mobility when run with and without time-averaged restraints. These relatively quick calculations also allowed us to look at relative mobility within the molecule and, aside from the N-terminal and binding loop regions, suggested the presence of an additional somewhat mobile region near residues 71−74. With hindsight, this can be seen to agree with a peak in structural fluctuation in the NMR structures (Ludvigsen et al., 1991b) and in the crystallographic *B*-factors (McPhalen & James, 1987).

The calculation *in vacuo* did suggest one very surprising pattern of behavior. The two β-strands 64−68 and 80−83 appeared to periodically move apart from each other (data not shown). This made it unclear whether they should be regarded as part of one four-stranded β-sheet or parts of two separate β-sheets. At the same time, this naturally led to concern over whether an *in vacuo* simulation would be valid in a situation where water molecules might be required to stabilize certain conformations.

(c) Effects of Treating the Experimental Restraints as Averages in Time Instead of as Instantaneous Restraints. A measure to compare structures is given in Table 3. Here each value represents a difference D_{ab} between two structures (see Methods), where the first two columns are the starting structures of the MD simulations and the six others represent the averaged structures of the six long time MD trajectories. The upper triangle is the difference calculated for all protein atoms, whereas the lower triangle is the difference based on only the backbone atoms (carbonyl, carbon, α-carbon, and nitrogen), without the loop region, residues 54−62, the so called globular backbone (Ludvigsen *et al.*, 1991b). We do not consider the loop region because those residues are poorly defined by the experimental data.

Comparing first the two starting structures, one can see that they differ by almost 2 Å from each other. This is in contrast to all other cases where the backbone difference is always significantly smaller.

Probably the clearest result from the table is the degree to which all trajectory averaged structures essentially converge. Considering backbone atoms only, no two average structures differ by more than 0.7 Å. The table also suggests that the use of time-averaged restraints has very little effect on the trajectory average, the change never being larger than 0.4 Å.

Figure 2 gives another view of the extent of space sampled by the different CI-2 simulations, by showing the positional fluctuation of α-carbon atoms during the whole trajectory for each residue. Figure 2a shows results from the time-averaged (solid line) and the conventionally (dashed) restrained MD simulations *in vacuo* starting from the solution structure, Figure 2b the same for the simulations starting from an NMR structure in H_2O and Figure 2c again the same for the two MD simulations in H_2O starting from the X-ray structure. The fluctuations for all α-carbons in the time-averaged NMR simulation *in vacuo* are much larger than those in the conventional run. The largest fluctuations are observed for residues 54−60, which were previously marked as having the greatest flexibility (McPhalen & James, 1987). This is not surprising, since this region is exposed to the solvent, there is no evidence of secondary structure and there are no disulfide bridges. The only likely stabilizing influence is the proposed pair of salt bridges involving Arg 65 and Arg 67 (McPhalen & James, 1987). The reactive site loop fluctuations are observed for the two simulations in H_2O, but to a lesser extent. This comes about partly because the trajectories in solution are half as long and partly because the motion of protein atoms is slowed down by the water molecules.

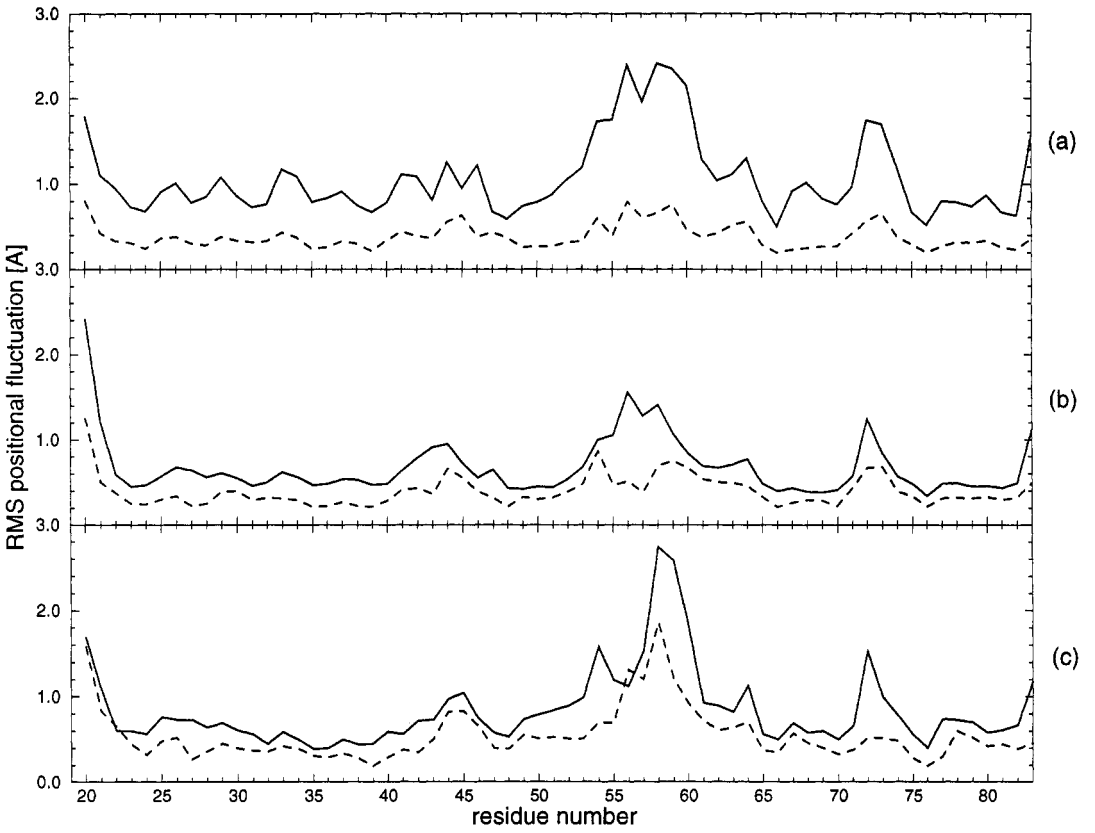

FIGURE 2: Rms positional fluctuations of α-carbons of CI-2 over long MD trajectories. Solid lines are from simulations involving time-averaged restraints, dashed lines from using conventionally restrained MD simulations. (a) 1-ns trajectories *in vacuo* starting from an NMR structure. (b) 0.5 ns trajectories starting from an NMR structure in solution. (c) As in panel b but starting from an X-ray structure.

Summarizing the effects of treating the experimental data as time-averaged instead of instantaneous restraints, one can conclude that both restraining methods make the MD simulations fulfil the experimental restraints. Nevertheless, time-averaged restraints allow the molecule to sample many more conformers. Given that all six simulations agree with the experimental data and result in similar average structures, one may examine further properties. We now wish to compare the time behavior of the simulations, examine the influence of solvent, and compare with the previously published structures.

(d) Comparison of the in Solution Simulations. When looking at the simulations in solution, we generally wish to make two comparisons. First, we are interested in the difference between time-averaged and conventional, instantaneous restraints. Secondly, we would also want to see if any phenomena are present in both simulations, whether they started from the previously refined NMR structure or from the crystal structure.

Table 3 also shows that simulations starting from different structures converge. The starting NMR structure had previously been refined with respect to the NMR data, so one would not expect it to change much, even over these long simulations. Trajectories starting from the crystal structure, however, converge to the same structure as runs starting from the NMR-based structure. Initially, the NMR and crystal starting structures differ by almost 2 Å. For the solution simulation periods, the trajectory averaged structures differ by less than 0.5 Å in every case, independent of whether one has used conventional or time-averaged restraints.

As described above, we wanted to address the issue of the extent of the β-sheets in the protein, so our analysis began by monitoring the size of certain characteristic distances between the two strands 64−67 and 80−84. Figure 3 shows the calculated distance between the α-protons of Val 65 and Val 82 for (a) the time-averaged (solid) and the conventionally (dashed) restrained runs starting from the NMR structure and below (b) the same for the simulations starting from the crystal structure. The experimentally measured NOE distance restraint for the H_α protons is 3 Å. For either starting structure, the difference clearly depends on how the restraints are imposed. When restraining the system at each time step to the original value, the distance remains more or less constant around the initial value (dashed). If the restraint is enforced as a time average, the distance tends to enlarge and be sometimes pushed back. Over time, the tendency to open this gap between the two β-sheets gets stronger and the distance remains large, that is, around 8 Å. This effect is observable for both runs.

The solution runs starting from the NMR structure and from the crystal structure show similar behavior in other respects. Figure 2 shows that the rms fluctuations of the α-carbon backbone are very similar, except for the loop region (residues 54−61). The difference can be explained by net movement of the backbone of this flexible region in the MD simulations starting from the X-ray structure (data not shown). Secondly, the distance matrix errors D_{ab} in Table 3 show very small differences between both solvent runs. The value 0.3 Å for the difference between the averaged globular backbone structures of the time-averaged restrained MD simulations starting from the NMR and the X-ray structure is very small considering that the two starting structures differ by 1.9 Å. Comparing only the globular backbone, all six averaged long time MD simulation

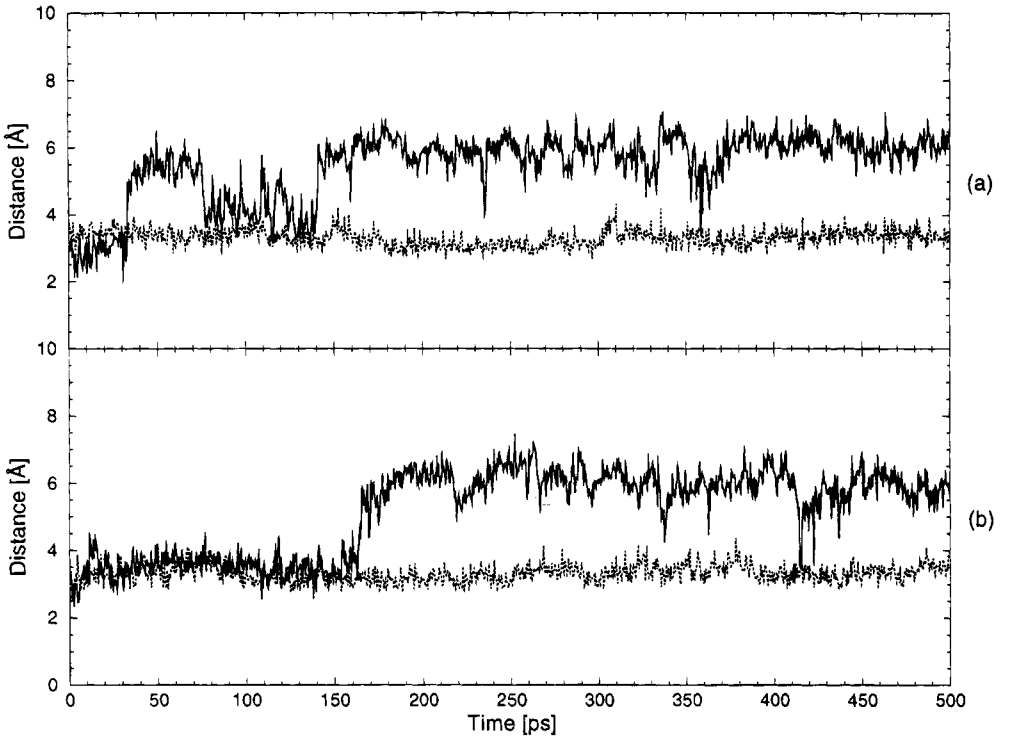

FIGURE 3: Interatomic distance as a function of time during the in solution simulations. The plots show the distance between the α-protons of Val 66 and Val 82. The experimental distance restraint for the H_α atoms is 3 Å. Solid lines are from runs using time-averaged restraints, dashed lines from the conventionally restrained MD simulations. (a) Starting from the NMR structure and (b) from the X-ray structure.

structures differ about 1 Å from the NMR starting structure and about 2 Å from the X-ray starting structure.

Summarizing this section, one can state that, despite the different starting structures, the averaged structures of long time MD simulations result in average protein structures that differ only slightly from each other, this independent of restraining method and protein environment.

(e) Differences in Secondary Structures. The ostensible justification for these MD simulations was to obtain a deeper insight into the differences between the solution and the crystal structure of CI-2 during a long time simulation. In the previous section we concentrated on the overall properties of the protein.

The main difference between the NMR and the X-ray structure, as already mentioned, is the core region with the two β-sheets. Figure 3 shows an experimentally measured backbone NOE distance between the two β-sheets. The jump from a lower distance to a higher value indicates the flexibility of the backbone of the protein in this region.

Given the dynamic nature of internal distances in the region, one would expect corresponding behavior in the hydrogen-bonding pattern. To judge the presence of hydrogen bonds, strict criteria were used. The angle subtended by N, H, and O atoms should be greater than 135°, and the H to O distance should be less than 2.5 Å. By looking at a small number of critical donors and acceptors, one can assess some of the dynamic and structural properties of the β-strands.

Table 4 shows how often hydrogen bonds occur between atoms of the strands 47−49 and 64−67 forming a first β-sheet, and the strands 22−24 and 80−82 forming a second β-sheet. In both time-averaged restrained MD simulations, the hydrogen bonds between the β-sheets were present for 63 to 83% of the time, whereas in the conventionally restrained runs, the percentage is between 68 and 99%.

Table 4: Percentage of Hydrogen Bonding between β-Strands of CI-2 in the Long MD Simulations in Solution[a]

residue ↔ residue	NMR ta	NMR cr	X-ray ta	X-ray cr
68 N−H ↔ C=O 49	78	97	77	93
49 N−H ↔ C=O 66	63	68	65	89
66 N−H ↔ C=O 47	69	96	83	90
67 N−H ↔ C=O 81	5	15	6	14
83 N−H ↔ C=O 65	9	45	31	42
24 N−H ↔ C=O 80	79	88	71	86
82 N−H ↔ C=O 22	82	99	75	86

[a] See Table 1 footnotes.

Considering these hydrogen bonds, it is clear that one sees at least two pairs of antiparallel β-strands joined into small β-sheet regions. The fact that the hydrogen bonds are not always present reflects the strict criteria applied to recognize hydrogen bonds. In the case of the simulations with time-averaged restraints, one can also see the greater internal flexibility.

Ludvigsen *et al.* (1991b) describe two other hydrogen bonds between these two β-sheets, namely, from 67 N−H to C=O 81 and from 83 N−H to C=O 65. The first one is seen only in the NMR structure and not in the crystal structure. Here, the dynamic movements in the core region are more pronounced. The two hydrogen bonds were observed in the conventionally restrained runs for about 15 and 45% of the time and in the time-averaged MD runs only for 5−31%. This is explicable in terms of experimental data for this area and the influence on the MD simulation. In the instantaneously restrained MD simulations, the penalty function forces the system to agree with the experimental data at each time step. In contrast, the time-averaged restrained runs impart each particle with a memory of its history with respect to distance restraint violations, and the molecule is only required to satisfy the distance restraints

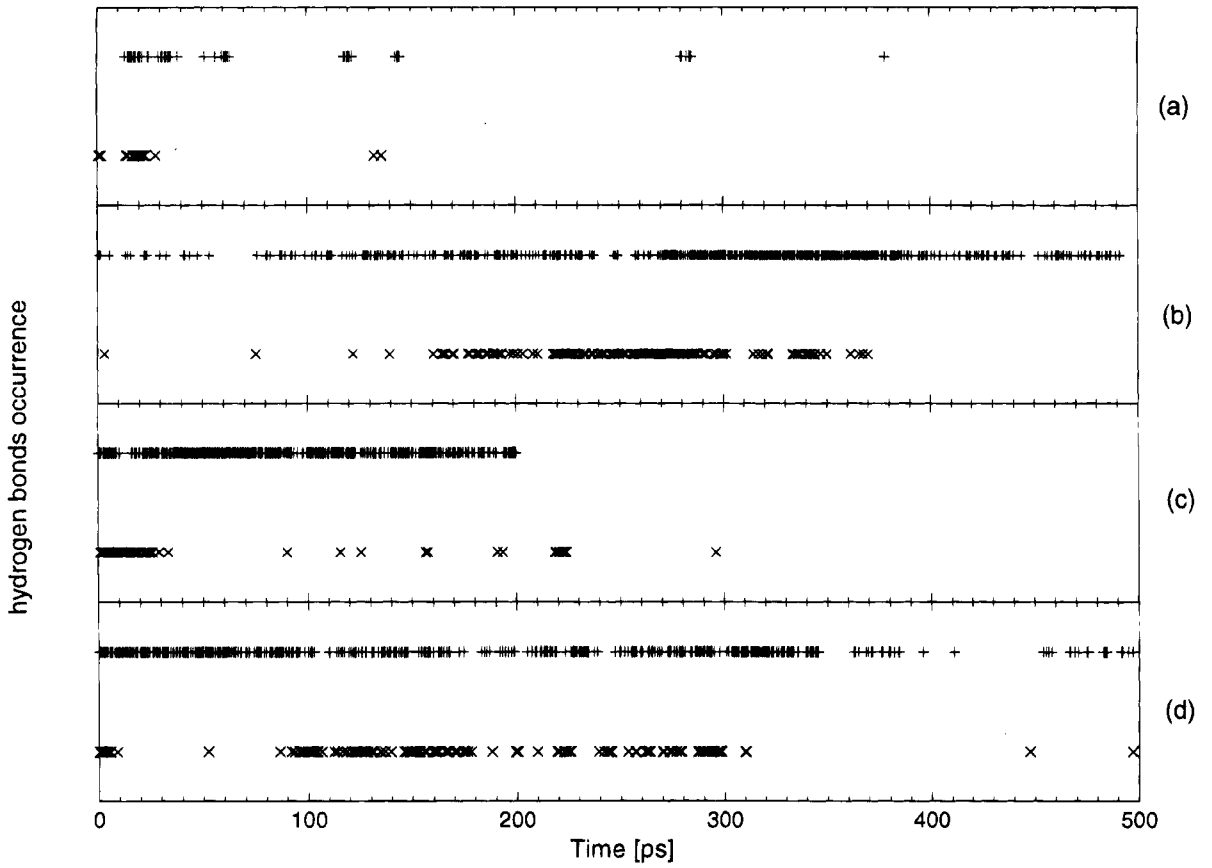

FIGURE 4: Occurrence of hydrogen bonds between 67 NH and 81 CO (+) and between 83 NH and 65 CO (×) during the MD trajectory. See caption of Figure 1 for the meaning of a–d.

on average. That is, the β-sheets move apart, water can penetrate, and the internal hydrogen bond is often broken. This is supported by the plot of the distance between the two strands (Figure 3) over the trajectory, where the distance jumps up to a much higher value. The same conclusion can be made by plotting the appearance of these two hydrogen bonds during the MD simulation. In Figure 4 each plus sign marks the appearance of hydrogen bond 67 N–H ↔ O=C 81 and each cross the hydrogen bond 83 N–H ↔ O=C 65. The four different graphs represent the four solution simulations. Figures 3 and 4 show the same time dependent behavior. If the distance restraint in Figure 3 fluctuates around the measured value of 3 Å, the hydrogen bonds in Figure 4 are regularly formed. As soon as it jumps up to the higher value, hydrogen bond criteria are hardly ever fulfilled.

To give an idea of this highly mobile region around the two β-sheets, Figure 5 shows two snapshots from the NMR time averaged restrained trajectory starting from the NMR structure—one from the very beginning, the other from the end of the run. The hydrogen bond (dashed line), which connects the strands in the first snapshot, is clearly broken in the overlaid second snapshot.

Independent of the restraining method used, all MD simulations in H_2O support the results of X-ray diffraction or NMR spectroscopy experiments during certain time intervals. That is, part of the time the two β-sheets form hydrogen bonds resulting in a four stranded β-sheet in the core of the protein, and part of the time the two sheets separate and form hydrogen bonds with nearby water molecules.

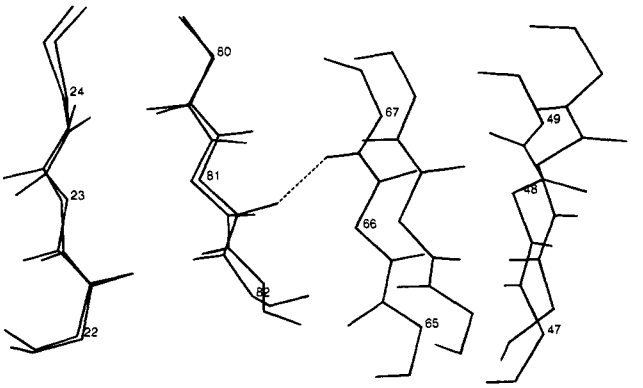

FIGURE 5: Mobility of the four β-strands in a time-averaged restrained MD simulation in solution starting from the NMR structure. The two structures show snapshots from the trajectory after about 100 ps and about 450 ps. The structures were least-squares fitted using the backbone atoms of the first two β-strands.

Another remarkable fact is that the specified hydrogen bonds need to be present for less than 10% of the time in order to explain the NOE data set. The original structures were refined using an *in vacuo* force field (Ludvigsen *et al.*, 1991b), where hydrogen bond donors have little alternative but to find acceptors within the protein. Our own preliminary calculations *in vacuo* showed the two interstrand hydrogen bonds to be present for more than 20% of the time in the time-averaged restrained MD simulation.

CONCLUSIONS

Not surprisingly, the time-averaged restrained MD simulations show a much more dynamic behavior when compared

to the conventional instantaneously restrained ones. This is independent of starting structure and protein environment. Although the experimental data were still well reproduced, the total potential energy is lower, and, in contrast, the system samples larger areas in space through the course of the trajectory.

The choice of restraining method depends on the kind of question one tries to address with a MD simulation. If the aim is to get a static structure, which fulfils the experimental data in all detail, conventionally restrained MD simulations with a final energy minimization calculation resulting in a single structure are probably the better choice.

On the other hand, if one wants to have an idea of the dynamic behavior of a protein, the analysis of a time-averaged restrained MD simulation trajectory gives a better insight of the flexibility and the motions such a protein undergoes. At the same time it demands the analysis of a long time trajectory with averaged quantities and possibly thousands of different configurations, instead of one single structure. A static structure of a protein is surely the only possibility to give a picture of its shape, but the natural behavior is more dynamic and not representable by either a rigid structure or averaged values of time-dependent properties.

The clearest result to come from these calculations is that using conventional distance and ^3J-coupling restraints, one reproduces the previously refined and published NMR structure. Using time-averaged restraints, one generates an ensemble of structures which include properties of both the crystal and previous NMR structures. Most notably, we can say that for part of the time, the structure behaves as if it had a single four-stranded antiparallel β-sheet and part of the time it behaves as if it had two separate two stranded antiparallel β-sheets. This statement is more remarkable as it does not matter if we start the calculations from the X-ray or previous NMR structures.

Finally, we should note that although the calculations here were somewhat expensive, they are still not nearly as comprehensive as one would desire. Although the solution simulations starting from two different structures (X-ray, NMR) appear to have converged to one average conformation, we cannot state that this is a general result. With unlimited computational time, one would wish to see if the convergence was repeated with many starting structures, for example, other members of the set of 20 distance geometry structures. Even for the one NMR structure which was subjected to extensive calculations, it is not clear that the solution simulations were as long as would be desired. The plots of intramolecular distances suggest the presence of infrequent transitions which are not properly sampled over 500 ps.

Given these caveats, the results do suggest that with an appropriate calculational strategy, it may be possible to reconcile even the small apparent differences between X-ray and NMR derived structures.

REFERENCES

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., & Weng, J. (1987) in *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G., & Sievers, R., Eds.) pp 107−132, Data Commission of the Internationall Union of Crystallography, Bonn/Cambridge/Chester.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981) in *Intermolecular Forces* (Pullman, B., Ed.) pp 331−342, Reidel.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684−3690.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.

Billeter, M. (1992) *Q. Rev. Biophys. 25*, 346−348.

Clore, G. M., Gronenborn, A. M., Kjær, M., & Poulsen, F. M. (1987a) *Protein Eng. 1*, 305−311.

Clore, G. M., Gronenborn, A. M., Kjær, M., James, M. N. G., McPhalen, C. A., & Poulsen, F. M. (1987b) *Prot. Eng. 1*, 313−318.

Jardetzky, O. (1980) *Biochim. Biophys. Acta. 621*, 227−232.

Kaptein, R., Zuiderweg, E. P. R., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179−182.

Karplus, M. (1959) *J. Chem. Phys. 30*, 11−15.

Kim, Y., & Prestegard, J. H. (1990) *Proteins 8*, 377−385.

Levitt, M. (1983) *J. Mol. Biol. 168*, 621.

Ludvigsen, S., Andersen, K. V., & Poulsen, F. M. (1991a) *J. Mol. Biol. 217*, 731−736.

Ludvigsen, S., Shen, H., Kjær, M., Madsen, J. Chr., & Poulsen, F. M. (1991b) *J. Mol. Biol. 222*, 621−635.

McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261−269.

McPhalen, C. A., Evans, C., Hayakawa, K., Jonassen, I., Svendsen, I., & James, M. N. G. (1983) *J. Mol. Biol. 168*, 445−447.

McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7242−7246.

Pearlman, D. A., & Kollman, P. A. (1991) *J. Mol. Biol. 220*, 457−479.

Rooman, M. J., Rodriguez, J., & Wodak, S. J. (1990) *J. Mol. Biol. 213*, 327.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.

Torda, A. E., Scheek, R. M., & van Gunsteren, W. F. (1989) *Chem. Phys. Lett. 157*, 289−294.

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

Torda, A. E., Brunne, R. M., Huber, T., Kessler, H., & van Gunsteren, W. F. (1993) *J. Biomol. NMR 3*, 55−66.

van Gunsteren, W. F., & Berendsen, H. J. C. (1987) in *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, The Netherlands.

van Gunsteren, W. F., & Berendsen, H. J. C. (1990) *Angew. Chem., Int. Ed. Engl. 29*, 992−1023.

van Gunsteren, W. F., Kaptein, R., & Zuiderweg, E. R. P. (1984) in *Proceedings of the NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics* (Olsen, W. K., Ed.) pp 79−82, CECAM, Orsay.