J. Mol. Biol. **(1990) 214, 223–235**

Time-averaged Nuclear Overhauser Effect Distance Restraints Applied to Tendamistat

Andrew E. Torda, Ruud M. Scheek and Wilfred F. van Gunsteren

Laboratory of Physical Chemistry
University of Groningen, Nijenborgh 16
9747 AG Groningen, The Netherlands

(Received 12 December 1989; accepted 25 February 1990)

A penalty function is introduced into molecular dynamics simulations that improves on current methods for enforcing nuclear magnetic resonance-based distance restraints. Rather than treating nuclear Overhauser effects as static distance bounds, they are considered as quantities that must be satisfied on average over the course of a simulation trajectory.

The efficacy of the method is demonstrated on the previously determined structure of tendamistat. The molecular dynamics simulations show that the time-averaged constraints increase the mobility allowed to molecules, produce better agreement with distance bounds, improve searching properties and give a better estimate of the conformational space occupied by the molecule in solution.

1. Introduction

Historically, one of the strengths of nuclear magnetic resonance (n.m.r.†) spectroscopy has been its ability to provide information about dynamic processes within molecules (Wuthrich, 1986). More recently, however, the use of n.m.r. as a tool for structure determination has brought about some change of emphasis. Instead of looking to see what motions are occurring in a molecule, current distance geometry and restrained molecular dynamics simulation techniques are all geared towards generating single structures that are consistent with all the experimental information. Given the known dynamic properties of molecules, we would suggest that this may not be a generally correct treatment.

In the extreme case of a small molecule hopping between two distinct conformations, the approach is demonstrably inappropriate. For example, the small peptide antamanide appears to exist in four confortnations related by the flip of two peptide planes (Kessler *et al.*, 1988). Nuclear Overhauser effect (NOE) measurements, however, reflect both conformations simultaneously. In this case, attempting to satisfy all the experimental constraints simultaneously is equivalent to trying to generate some practically non-existent transition state. In larger molecules, it may not be possible to identify such distinct conformational states, but measured NOEs are still average values reflecting the whole region of configurational space visited by molecules on the n.m.r. time scale.

The problem with current methods is that they use a simple penalty function that forces each single structure to agree with all the experimental data. **A** more physically realistic penalty function should allow for some averaging of the NOE through time. This inclusion of time leads naturally to the use of such a term in molecular dynamics (MD) simulations rather than a distance geometry method.

Recently, we proposed such a term and demonstrated its application on a simple, artificial model system (Torda *et al.*, 1989). Each particle was allotted a memory of its history with respect to distance constraint violations and was only required to satisfy the distance constraints over the course of a trajectory. Here, we apply the procedure to the realistic case of refining a large polypeptide using published n.m.r. data. These results are then contrasted with those obtained by a more conventional refinement.

For assessing the use of these time-averaged distance constraints, we used the published distance geometry structures and data from tendamistat (Kline *et al.*, 1988). This is a polypeptide of 74 residues, M, 8000, and has been referred to as the most precisely determined solution structure of a protein (Wuthrich, 1989). This was a particularly interesting test case as the large volume of experimental data resulted in a family of very tightly clustered distance geometry structures with no evidence of conformational averaging.

† Abbreviations used: n.m.r., nuclear magnetic resonance; NOE, nuclear Overhauser effect; MD, molecular dynamics: r.m.s., root-mean-square.

0022–2836/90/130223–13 $03.00/0

2. Theory

MD simulations are usually used for refining n.m.r.-based structures by the construction of an artificial energy term that 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 (van Gunsteren & Berendsen, 1987) this term is quadratic with respect to violations of distance constraints, so:

$$V_{dc}(r)\begin{cases}=\frac{1}{2}K_{dc}(r-r^0)^2, & \text{if } r>r^0,\\ =0, & \text{if } r\le r^0,\end{cases} \quad (1)$$

where $V_{dc}(r)$ is the potential 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_{dc}, is used to weight this term relative to the rest of the force field. In practice, restraints may also be repulsive and the force can be made constant beyond a defined cut-off (van Gunsteren *et al.*, 1985).

In our previous work, we replaced the instantaneous distance r in equation (1) with a time-averaged distance, denoted $\bar{r}$, so that the distance restraint energy term became:

$$V_{dc}[\bar{r}(t)]\begin{cases}=\dfrac{K_{dc}}{2}(\bar{r}-r^0)^2, & \text{if } \bar{r}(t)>r^0,\\ =0, & \text{if } \bar{r}(t)\le r^0.\end{cases} \quad (2)$$

It is then necessary to define $\bar{r}$ so that it reflects the quantity measured in an experiment. The NOE is due to dipolar interactions between nuclei, so its intensity is proportional to r^{-6}, and the averaging should reflect this. The timescale of a MD simulation, however, is short compared to the correlation time for molecular tumbling, so the influence of angular fluctuations should be ignored (Tropp, 1980; Kessler *et al.*, 1988). The NOE should then be treated as a function of r^{-3}, so we define:

$$\bar{r}(t)=\left(\frac{1}{t}\int_0^t r(t')^{-3}\,dt'\right)^{-1/3}. \quad (3)$$

Equation (3) would be suitable for the definition of $\bar{r}(t)$ if simulations were of infinite length. It is also the correct way to calculate $\bar{r}(t)$ when analysing trajectories. In practice, however, as the total time for a simulation increased, so would the time used for averaging $\bar{r}(t)$. This would result in $\bar{r}(t)$ becoming progressively less sensitive to the instantaneous state of a molecule. To avoid this problem, a memory function can be built into equation (3) so that $\bar{r}(t)$ is calculated as a running average with recent history weighted most heavily. An exponential form for this weighting leads to:

$$\bar{r}(t)=\left(\frac{1}{\tau}\int_0^t e^{-t'/\tau}[r(t-t')]^{-3}\,dt'\right)^{-1/3}, \quad (4)$$

where τ is the characteristic time for the exponential decay.

Although this form of the potential worked well for a simple system with Lennard–Jones particles, it led to a force with fourth-power terms with respect to $\bar{r}(t)/r(t)$. This gave rise to occasional large forces in simulations of macromolecules. To avoid this problem, an alternative approach was used so that the potential was not formally defined at all. Instead, only a force was constructed so that:

$$\vec{F}_i(t)\begin{cases}=-K_{dc}[\bar{r}_{ij}(t)-r^0]\dfrac{\vec{r}_{ij}(t)}{r_{ij}(t)}, & \text{if } \bar{r}(t)>r^0,\\ =0, & \text{if } \bar{r}(t)\le r^0.\end{cases} \quad (5)$$

where $\vec{F}_i(t)$ is the force on atom i due to atom j and $\vec{r}_{ij}=\vec{r}_i-\vec{r}_j$. Integration of equation (5) 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.

From equation (4), it can be seen that, at zero time, $\bar{r}(t)$ is undefined. In practice, this turns out to be an advantage. If $\bar{r}(t)$ is initially set to be slightly smaller than r^0, then no force will be applied at the start of a simulation. If the initial $r(t)$ is actually smaller than r^0, then $\bar{r}(t)$ will remain smaller and no force will be applied at all. If, however, the initial $r(t)$ is larger than r^0, then $\bar{r}(t)$ will rise and a force will gradually be applied. This procedure results in the artificial force being turned on as necessary during a simulation and being turned on most rapidly for those distances most severely violating their restraints.

3. Methods

All simulations were carried out using software from the GROMOS suite of programs (van Gunsteren & Berendsen, 1987). Distance restraints were imposed using either eqn (1) or (5). Dihedral angle restraints were enforced as described by de Vlieg *et al.* (1986). So, for each restrained dihedral angle, a potential, V_{dlr} is defined:

$$V_{dlr}(\Phi)=K_{dlr}[1+\cos(\Phi-\delta)], \quad (7)$$

where K_{dlr} is a force constant controlling the relative weight of the term, Φ is the dihedral angle and δ is a phase angle, chosen so that $V_{dlr}=0$ when Φ is at the value chosen for the restraint.

Simulations were carried out *in vacuo* with a cut-off for non-bonded interactions of 8 Å (1 Å = 0·1 nm). Weak coupling to a temperature bath was used to maintain the temperature near 300 K for all simulations (Berendsen *et al.*, 1984). The SHAKE algorithm was used to maintain all bond lengths and the integrator time step was 0·001 ps (Ryckaert *et al.*, 1977). As described in Theory, simulations with time-averaged distance restraints require the choice of an initial value for $\bar{r}(t)$. For all the runs using time-averaged distance restraints, $\bar{r}(0)$ was set for each distance restraint to be 0·2 Å less than r^0. Other parameters were varied over the course of the simulations and are shown in Table 1.

Initial co-ordinates, distance restraint data and dihedral angle restraint information for tendamistat were

all obtained from W. Braun (ETH, Zürich), although the co-ordinates have now been deposited with the Brookhaven Protein Data Bank. The nomenclature for structures in this work follows that of the original publication, so structure I refers to structure I of Table 7 of Kline *et al.* (1988).

Distance restraint information was used as given in Table 6 of Kline *et al.* (1988). Restraints involving stereospecific or pseudo atoms (Wüthrich *et al.*, 1983) were referred to appropriate sites in terms of GROMOS atoms (van Gunsteren *et al.*, 1985). The set of restraints was unchanged with the following exceptions. Firstly, the disulphide bridges were treated as covalent rather than distance restraints. Secondly, 34 restraints based on 17 implied hydrogen bonds were not used. Kline *et al.* (1988) had assessed the likely location of hydrogen bonds on the basis of determined secondary structure and the presence of characteristic NOEs. They then restrained appropriate O–H and N–O distances so as to reproduce correct hydrogen bond geometry. This is equivalent to adding a simple hydrogen-bonding term to the distance geometry calculations and would be inappropriate for a MD simulation where the electrostatic and Lennard–Jones terms already reproduce the correct geometry (Hermans *et al.*, 1984). The result of this culling of data was a set of 842 distance restraints.

The dihedral angle information of Kline *et al.* (1988) was used after conversion to a form suitable for eqn (7). Values for the force constant K_{dlr} are given in Table 1, but an additional weighting was used for each restraint individually, to reflect the range that had been allowed in the original distance geometry calculations. This weighting was inversely proportional to the range allowed to the angle by Kline *et al.* (1988); so, for example, a range of 50° (strong restraint) corresponded to a weight of 1. A weaker dihedral restraint with a range of 100° was then given a weight of 0·5.

4. Results

(a) *Selection of structures and the course of the molecular dynamics refinement*

In order to test the use of time-averaged distance restraints, it was decided to subject two of the nine distance geometry structures of Kline *et al.* (1988) to MD refinement. Before selection, all nine were first subjected to 300 steps of restrained steepest-descents energy minimization. The results of this are shown in Table 2. After minimization, the potential energies were all in a narrow range from −2834 to −3140 kJ mol^{-1}, so this was not used as a criterion for selection. The sum of violations, however, spanned the range 32·5 to 47·3 Å. Two structures were then selected for more extensive refinement. Firstly, structure III was selected, as it had the lowest sum of violations after energy minimization. Secondly, structure I was selected, since it was judged to be the best of the structures in the original work of Kline *et al.* (1988). No further use was made of the remaining seven structures.

The two structures were then subjected to the simulation protocol outlined in Table 1. An initial 2·5 picoseconds of MD was performed on each structure without using time-averaged restraints. Then for both structures I and III, parallel simulations were performed with and without the use of time-averaged distance restraints. These runs consisted of a ten picosecond preparation simulation and then a final 20 picosecond simulation, which was used for analysis and assessment of the procedure. The second energy minimization (EM2) was performed only because comparisons of energies would otherwise have suffered from fluctuations in the kinetic and potential energy terms. Otherwise, structures that were the endpoints of trajectories were not analysed. Table 3 shows the potential energy after each stage of the refinement.

In Table 4, we provide another measure of the progress of the refinement procedure, root-mean-

Table 1

Protocol for refinement of distance geometry structures

Time (ps)	K_{dc} [a] (kJ mol^{-1} Å^{-2})	τ_{dc} [b] (ps)	K_{dlr} [c] (kJ mol^{-1})	τ_T [d] (ps)
EM1 [e]	5	—	0	—
0–2·5	5	0	70	0·01
2·5–12·5	15	0 or 0·5	20	0·05
12·5–32·5	25	0 or 1·25	10	0·04
EM2 [f]	25	—	10	—

[a] K_{dc} was the force constant used for distance restraints.
[b] τ_{dc} was the time constant for the memory function for distance restraints. Between 2·5 and 32·5 ps, parallel runs were performed with either time-averaged distance restraints or without ($\tau_{dc} = 0$ ps).
[c] K_{dlr} was the force constant used for dihedral angle restraints. For each restraint it was then multiplied by an individual weighting factor, described in Methods, so as to reflect the range allowed by Kline *et al.* (1988).
[d] τ_T was the time constant for coupling to the temperature bath.
[e] Three hundred steps steepest-descents energy minimization (EM).
[f] Four hundred steps conjugate gradients energy minimization.

Table 2

Distance geometry structures before and after energy minimization

	Before EM		After EM	
Structure[a]	Energy (kJ mol^{-1})	Σ violations (Å)	Energy (kJ mol^{-1})	Σ violations (Å)
I	548	16·1	−2834	34·1
II	848	16·0	−3140	33·1
III	704	16·7	−2931	32·5
IV	584	18·7	−3038	36·1
V	978	21·1	−3028	38·2
VI	865	23·5	−2946	41·5
VII	1107	20·3	−3095	41·4
VIII	1556	24·2	−2890	47·3
IX	1148	21·4	−2876	45·3

EM, energy minimization.
[a] The numbering for structures corresponds to that of Kline *et al.* (1988).
[b] Σ refers to the sum of violations of NOE bounds. Note that the sum is not identical with that of Table 7 of Kline *et al.* (1988), since disulphide bridge and implied hydrogen bonds were part of the force field and not included as distance restraints.

Table 3
Energies of structures during refinement

Time (ps)	Potential energy (kJ mol^{-1}) Structure I: MD without averaging[a]	Potential energy (kJ mol^{-1}) Structure I: MD with averaging	Potential energy (kJ mol^{-1}) Structure III: MD without averaging	Potential energy (kJ mol^{-1}) Structure III: MD with averaging
Initial	548		704	
EM1	−2834		−2931	
2·5	−2080		−2092	
12·5	−2240	−2121	−2113	−2098
32·5	−2325	−2368	−2111	−2293
EM2	−3965	−3642	−3954	−3543

[a] Without averaging refers to a conventional MD refinement. With averaging refers to the use of time-averaged NOEs.

Table 4
Comparison of initial and trajectory average structures

Structure	r.m.s. difference[a] (Å) Structure I: Without averaging[b]	r.m.s. difference[a] (Å) Structure I: With averaging	r.m.s. difference (Å) Structure III: Without averaging	r.m.s. difference (Å) Structure III: With averaging
EM1	0·20		0·22	
MD[c] (average)	1·28	1·11	1·20	1·09

[a] r.m.s. differences based on the C^{α} atoms of residues 5 to 73 and after fitting on the same subset of atoms.
[b] Without averaging refers to a conventional MD refinement. With averaging refers to the use of time-averaged distance constraints.
[c] Average structure from 20 ps trajectory.

square (r.m.s.) positional comparisons of the initial structures with those after the first energy minimization, and after refinement. We avoid the influence of dynamic fluctuations by using average structures from each 20 picosecond trajectory, rather than static end structures. Co-ordinates were fitted based on the C^{α} atoms of residues 5 to 73, so as to allow comparison with Table 10 of Kline *et al.* (1988). The calculated values also reflect only the differences in this subset of atoms. As would be expected, the energy minimization produces only small shifts, although Table 2 shows that these alone were sufficient to bring about a large decrease in potential energy. Comparison of the initial and trajectory average structures, however, shows r.m.s. positional shifts between 1·1 and 1·3 Å. These do not constitute large changes in structural terms, but it is surprising that they are of a size similar to the spread amongst the entire family of distance geometry structures proposed by Kline *et al.* (1988).

The simulation protocol was chosen so as to provide a thorough refinement with a reasonable computational expense. The first 2·5 picoseconds were used to relax the molecules in the GROMOS force field. On the basis of the potential energies, most of the strain within the molecule is removed during this period and only a small improvement is achieved over the next 30 picoseconds. The period from 2·5 to 12·5 picoseconds served to set the system up for the analysis trajectories by ensuring that the molecules were adapted to the raised force constant for restraints and were also relaxed around the restrained dihedral angles. This was to allow the use of a small force constant for dihedral angle restraints in the final simulations. By the last 20 picoseconds of simulation, there was little strain in the molecule due to either the physical or artificial terms in the force field.

(b) *Agreement with experimental data*

For analysing the agreement of trajectories with experimental data, $\bar{r}(t)$ was calculated according to equation (3); that is, the averaging was conducted over the length of the trajectory without the use of a memory function. While the memory function is necessary as a computational device because of the finite time of a simulation, it is the average over the whole simulation that should reproduce the measured data. Table 5 shows the sum of violations and largest single violation in the analysis trajectories for the two structures, run with and without time averaging of NOEs.

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 tendamistat trajectories, with or without time-averaged NOEs, there are no violations larger than 0·7 Å. This result is in agreement with Table 7 of Kline *et al.* (1988) and suggests that no major disruption occurs in either of the structures when they are put into a MD force field.

Considering the largest single violations shows that no trajectory produces unsatisfactory structures, but the sum of violations in Table 5 shows a more pronounced effect from the application of time-averaged NOEs. For structure I, the sum of violations is 31% smaller when the distance

Table 5
Distance restraint violations over 20 picosecond trajectories

Structure	Σ violation[a] (Å): MD without averaging	Σ violation[a] (Å): MD with averaging	Largest violation[b] (Å): MD without averaging	Largest violation[b] (Å): MD with averaging
I	20·6	14·2	0·69	0·66
III	18·9	15·1	0·60	0·66

[a] Σ, the sum of violations is an average calculated over the 20 ps trajectory using eqn (4).
[b] The largest single average violation over the 20 ps trajectory.

restraints are enforced as a time average. For structure III, this figure is 20%. Trial runs with tendamistat and simulations on smaller peptides have shown that trajectories using the time-average NOEs typically have violations 60 to 80% of those refined by conventional means (our unpublished results). This in itself should be considered to be a substantial benefit from the technique.

Although Table 2 contains results from the initial static structures, these can be cautiously compared with the trajectory averages in Table 5 in order to view the effect of the MD refinement. When time-averaged distance restraints were used, the original violations were reduced from 16·1 to 14·2 Å and 16·7 to 15·1 Å. Without using time-averaged NOEs, the violations actually increased (although accompanied by a massive decrease in potential energy). This largely reflects the parameters used in the refinement protocol. The relatively small force constant for the artificial distance restraint term, $K_{dc} = 25$ kJ mol^{-1} Å^{-2} (less than 6 kcal mol^{-1} Å^{-2}; 1 cal = 4·184 J) was chosen so as to produce only a small disturbance of the physical terms in the force field. Violations in all trajectories could have been made to appear smaller by using a larger value for K_{dc}, but this would have been at the expense of higher potential energies. These results then show another major advantage to the use of time-averaged constraints. Structures can be refined and experimental data reproduced using much smaller artificial forces.

Distance restraints provided the bulk of the structural information on tendamistat, but the experimental dataset also provided restraints on dihedral angles. Although no averaging was used in the artificial dihedral restraint term, the four trajectories were checked to ensure that, on average, they agreed with the ranges set by Kline *et al.* (1988). Considering all of the runs, the largest single violation occurred in the trajectory of structure I without time-averaged constraints. The ϕ angle of Ser63 averaged 15° (±22°, 1 standard deviation), whereas the allowed range was −90° to −40°. There was no other dihedral that was, on average, more than 11° out of its allowed range. Considering the accuracy of such restraints (Pardi *et al.*, 1984), this constitutes satisfactory agreement.

(c) *Mobility of structures*

Usually, after generating a family of distance geometry structures or even static endpoints from MD simulations, r.m.s. differences are calculated between the conformations to provide some measure of the size of conformational space that they span. Because we are examining trajectories, we can calculate the space spanned, over the simulation, for individual structures. Positional r.m.s. fluctuations for α-carbons of each residue were calculated for each trajectory and the results are shown in Figure 1.

The difference between simulations with and without time averaged NOEs is quite striking.

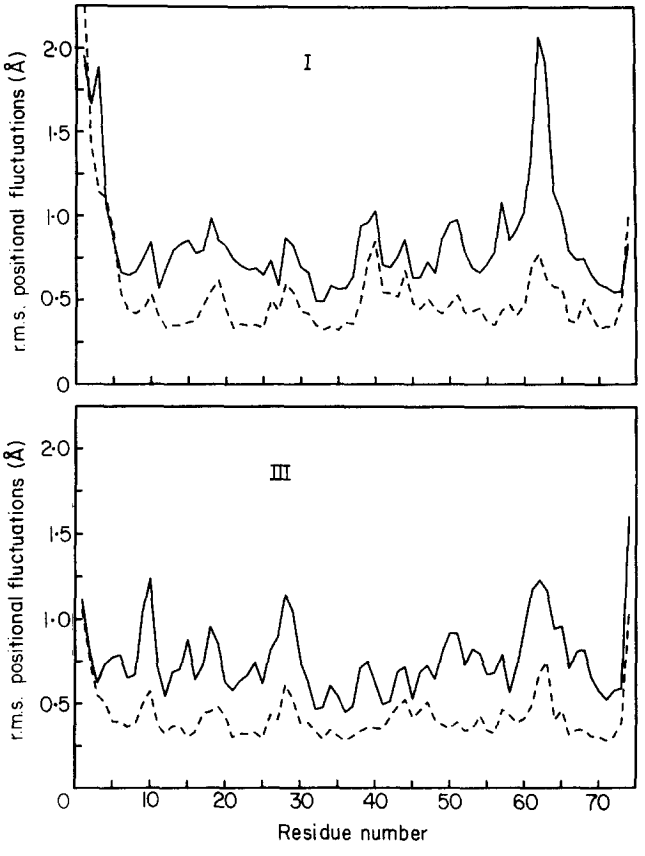

Figure 1. r.m.s. positional fluctuations of α-carbons in tendamistat over 20 ps trajectories. Continuous lines are from runs using time-averaged NOEs, broken lines from runs using conventional MD refinement. I and III refer to the structure names as given in text.

When the time-averaged force is used, fluctuations are consistently larger, and often double those in the more conventional refinement simulations. Again, the small value for K_{dc} should be noted. The difference between the trajectories would have appeared even greater if a larger force constant were used and the mobility of the conventionally refined structures further reduced.

The plots also show that the two structures retain their identities, despite the relatively long simulation times. That is, there is no plot of fluctuations characteristic of this molecule and this dataset. Instead, the plot for structure I without averaging of NOEs is closest to that of structure I with averaging and similarly for structure III. This means that on the short timescale of the simulations, the molecules are distinct and, for example, structure I does not spontaneously transform itself into structure III. This fact is also reflected by pairwise comparisons of r.m.s. positional differences within the set of four trajectory average structures, shown in Table 6. Although each of the four structures is centred about different averages, the structure I averages are closer to each other than to those for the other structure, and similarly for structure III.

Table 6
r.m.s. differences between trajectory average structures

		I		III	
		r.m.s. (Å) without averaging	r.m.s. (Å) with averaging	r.m.s. (Å) without averaging	r.m.s. (Å) with averaging
I	Without averaging	—	0·62	0·78	0·92
	With averaging	1·99	—	0·86	0·86
III	Without averaging	2·89	2·80		0·45
	With averaging	2·88	2·69	0·88	—

Each of the structures for comparison is an average over a 20 ps trajectory. Calculations were performed after fitting on the C^{α} atoms from residues 5 to 73. Values up and right of the diagonal are the r.m.s. differences from this subset of atoms. Values below and left of the diagonal are calculated based on all atoms.

It is also of interest to compare the plots of Figure 1 here with Figure 5 of Kline *et al.* (1988), where residual violations were plotted as a function of residue number. There is little obvious correlation between the two measures, except for the large peak in both mobility (Fig. 1 here) and violations (Fig. 5 of Kline *et al.*, 1988) near residues 62 and 63. Possibly this is an example where, indeed, the NOEs are reflecting some movement between conformations in solution.

A last indication of the relative mobilities is provided by counting the number of dihedral transitions that occur over the 20 picosecond trajectories. Here, a dihedral transition is defined as follows: after going through the minimum of a well of the dihedral angle potential, a transition is considered to have occurred when the value of the angle crosses the minimum of an adjacent well of the dihedral term in the potential. Simply totalling the number of transitions is a primitive, but, in this case, dramatic indication of mobility. In structure I, there were 795 dihedral transitions over the 20 picoseconds when no averaging was applied to distance restraints. In contrast, there were 1951 transitions when averaging was used. The corresponding figures for structure III were 592 and 1768.

5. Discussion

(a) *Effects of mobility on residual violations*

From Results, sections (b) and (c), above, it can be seen that the use of time-averaged NOEs has several major benefits. Firstly, the size of residual violations is decreased. Secondly, smaller force constants can be used for the artificial terms in the force field. Thirdly, the fluctuations of structures are less hindered by the artificial forces that are used. Obviously, these effects are closely connected. This is best shown by considering some examples from the trajectories.

Figure 2 shows the distance between stereospecifically assigned methyl groups from Val35 and Leu70. In the top panel, the distance is plotted as a

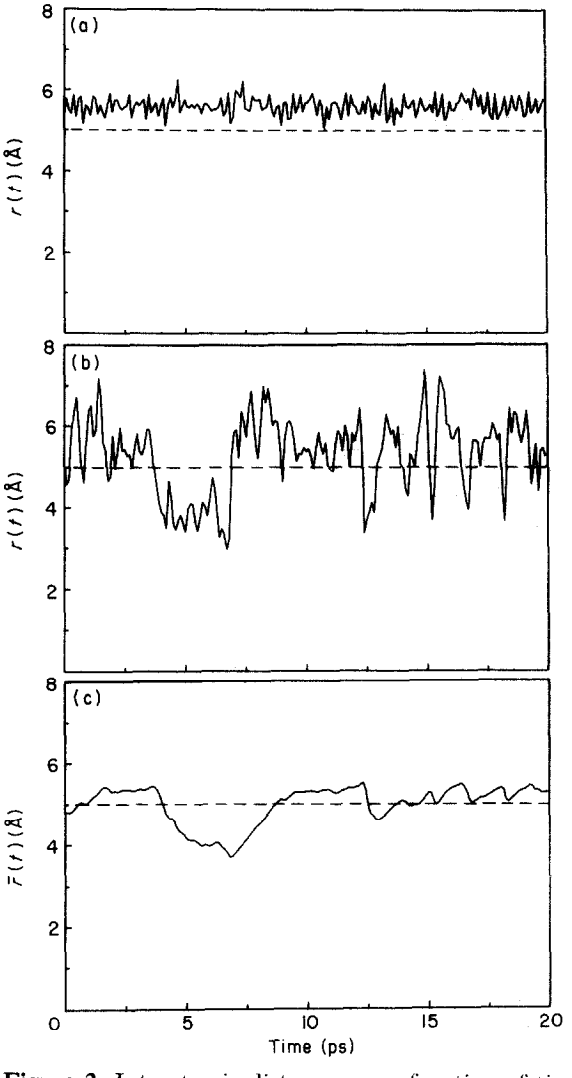

Figure 2. Interatomic distances as a function of time during 20 ps trajectories of structure I. All plots show the distance from the CG2 methyl group of Val35 to the CD1 methyl group of Leu70. The broken lines indicate the 5·0 Å distance constraint. (a) Distance, $r(t)$, in the simulation with distance restraints enforced as static bounds. (b) $r(t)$ in the simulation with time-averaged distance restraints. (c) $\bar{r}(t)$ from the same simulation as (b), calculated with the memory function according to eqn (4).

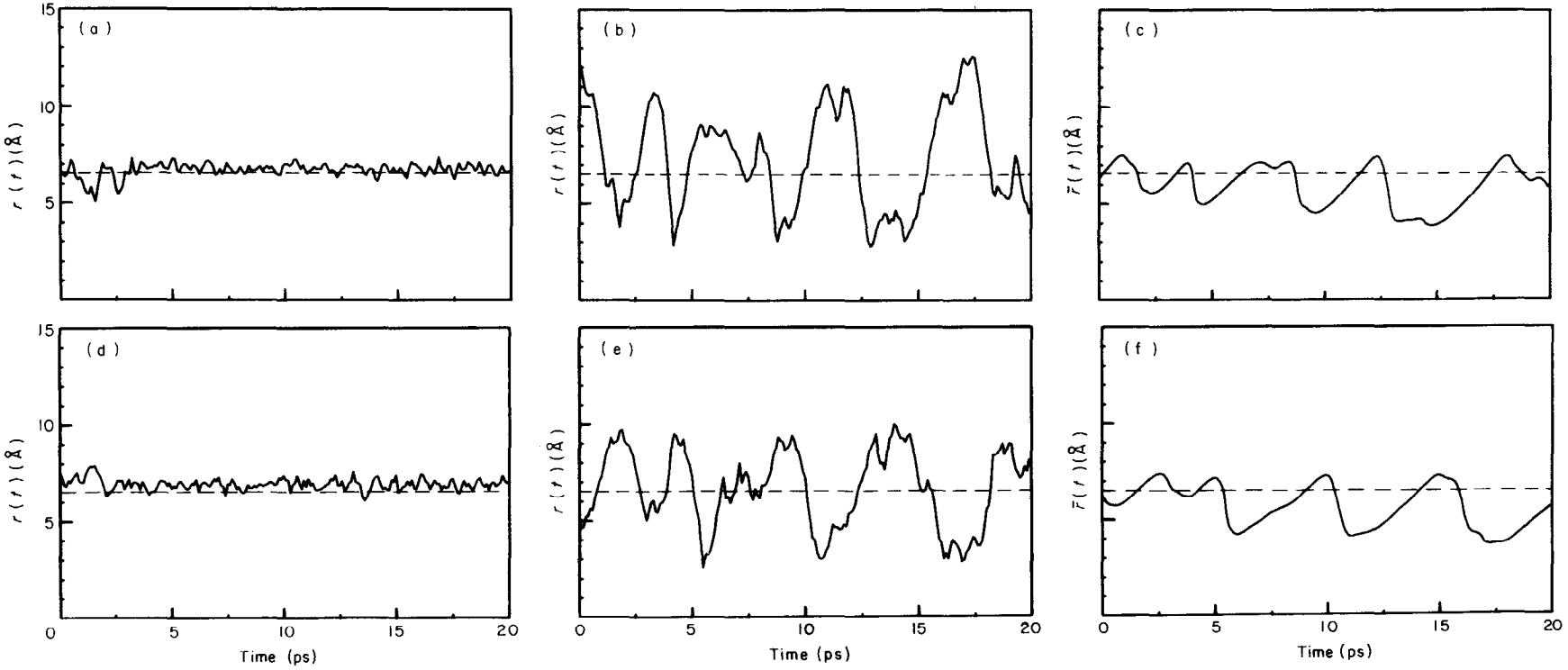

Figure 3. Interatomic distances as a function of time during 20 ps trajectories of structure I. Top plots ((a), (b), (c)), show the distance from the CZ pseudo atom of Tyr15 to the β-methine proton of Thr13. Bottom plots ((d), (e), (f)), show the distance from the same Tyr15 site to a stereospecifically assigned β-methylene proton on Ser17. Broken lines mark the 6·5 Å distance restraint. Left panels ((a), (d)), show distance, $r(t)$, in the simulation with distance restraints enforced as static bounds; centre panels ((b), (e)), $r(t)$ in the simulation with time-averaged distance restraints; right panels ((c), (f)), $\bar{r}(t)$ from the same simulation as (b), calculated with the memory function according to eqn (4).

function of time during the simulation of structure I when no averaging was used for the calculation of the distance restraint force. The NOE restraint of 5 Å was, on average, violated by 0·6 Å over this run. Because of this violation, a relatively constant force was applied throughout the trajectory, preventing the methyl groups from moving apart significantly. The centre panel shows the same distance from the same structure, but from the simulation when time-averaged restraints were used. In this trajectory, the NOE restraint was, on average, satisfied by 0·1 Å. From the plot, it can be seen that the methyl groups are at times more than 7 Å apart and that the NOE bound is violated for many picoseconds at a time. This is a very desirable result, since, by averaging over the trajectory, the experimental data are well reproduced. It is also interesting to see that the protons are able to spend some time as close as 4 Å to each other. This probably reflects the transient violation of some other NOE in the molecule, allowing this one to be satisfied.

The result can be further explained by considering the value of $\bar{r}(t)$ calculated with the memory function according to equation (4) as shown in Figure 2(c). This is the value that actually determines the size of the artificial force applied (eqn (5)). Despite the instantaneous fluctuations in the distance, the running average shows a relatively flat shape, as one would want. The force should not be applied so as to prevent small or transient violations of the NOE. The bottom panel also highlights the importance of the r^{-3} averaging employed. Although the distance between the methyl groups is often quite large, they only have to spend a relatively short time close to each other to reduce the size of the average distance and thus reduce the applied force.

The distance plot of Figure 2(b) shows large fluctuations that are hard to interpret in conformational terms. A distance between points far apart in the sequence will inevitably reflect many different internal motions and the interplay of many NOE restraints. A different situation can be seen by considering distances between sites on residues closer in the polypeptide sequence. The top panels of Figure 3 show the fluctuations in distance from a pseudo atom on Tyr15, centred between the δ hydrogen atoms, to the β-methine proton from Thr13. The bottom panels show the distance from the same site on Tyr15 to a stereospecifically assigned β-methylene proton on Ser17. Both distances were restrained by NOEs of 6·5 Å.

Without the use of time-averaged restraints (left-hand plots), the NOEs were violated by 0·2 and 0·4 Å. Because of these violations, the tyrosine ring would have been continually pulled towards the side-chains of both Thr13 and Ser17. On average, it cannot come close enough to either residue to satisfy the NOEs. In contrast, the centre plots show the same distances from the trajectory with time-averaged distance restraints. The distances show a remarkable oscillatory behaviour with a period close to five picoseconds. The physical nature of the motions can be seen in Figure 4. The aromatic ring of Tyr15 actually moves between positions where it alternately satisfies the NOEs to residues 13 and 17. The diagram can be contrasted with Figure 7 from Kline *et al.* (1988) and Figure 5(c) from Billeter *et al.* (1989) where the ring adopted a nearly identical conformation in all distance geometry structures. Furthermore, it had appeared that this residue adopted different conformations in solution and X-ray crystallographic structures. It might be more correct to state that the conformational space of the n.m.r. structure includes that of the X-ray structure.

These results do not mean that, in solution, the ring regularly moves with a period of five pico-

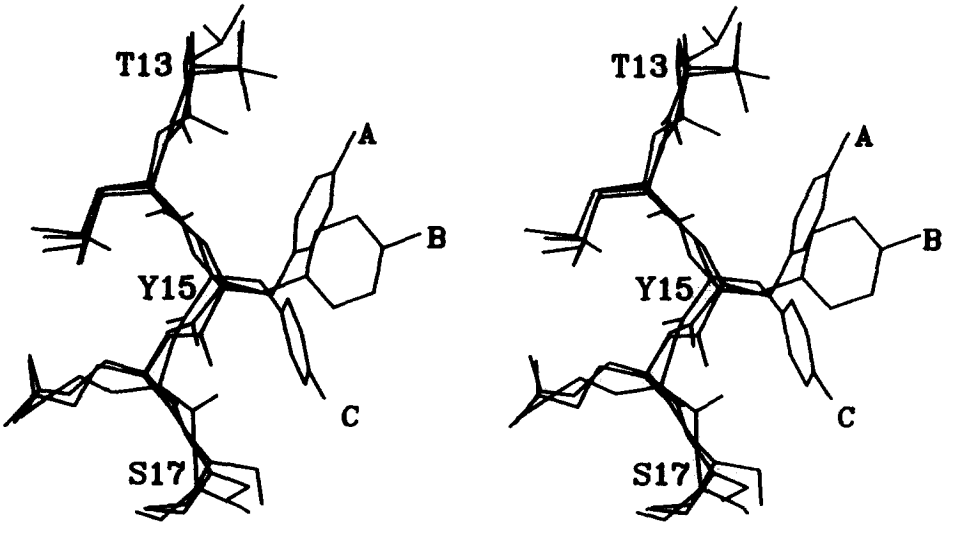

Figure 4. Mobility of Tyr15 in structure I over a 20 ps trajectory in stereo. The peptide segment from residues 13 to 17 is shown from (A) 9 ps into trajectory. (B) Distance geometry starting structure. (C) 16·2 ps into trajectory. Both MD structures were least squares fitted to (B) on the basis of all backbone heavy atoms.

seconds. The timescale of a simulation *in vacuo* is much faster than solution behaviour and the timescale of these motions is influenced by the time constant, τ, for the decay of the distance restraint memory function in equation (4). The simulations also neglect the stochastic influence of the solvent. The results do, however, suggest that on the basis of the n.m.r. data, the tyrosine ring spends some time close to the serine and some time close to the threonine. Whether these represent different conformations is a matter of arbitrary definition. Certainly, they are two distinct positions, but they are separated by energetic barriers that are readily surmounted both in the simulation (shown in the plots) and in solution (since the NOE data produced only 1 set of peaks). This oscillatory behaviour was seen for the equivalent data from the simulation of structure III (data not shown).

Before assuming the existence of conformational averaging involving Tyr15, it is necessary to check if it is consistent with all the available experimental information. Firstly, Thr13, Tyr15 and Ser17 are all surface residues that one would expect to be mobile. It was therefore surprising that they appeared to adopt such a well-defined conformation in the distance geometry structures (Kline *et al.*, 1988). This would occur if there was either a single well-defined conformation or conformational averaging gave rise to many NOEs.

Next, one can examine the plots of residual violations in the distance geometry structures (Fig. 5 of Kline *et al.*, 1988). Every distance geometry structure shows a small peak in violations at this residue, as would be expected if the NOEs did not result from a single structure fixed in time. Of course, this peak in violations could simply reflect the large number (26) of NOEs involving Tyr15.

Lastly, one should consider the comparison of the X-ray and solution structures for tendamistat (Billeter *et al.*, 1989). In the X-ray structure, no electron density was observed for the side-chain of this residue. This is exactly what one would expect for a site undergoing rapid motions.

The possibility of motional averaging, suggested here, is actually in agreement with the conclusions of Billeter *et al.* (1989). They stated that multiple local conformations separated by a sizeable energy barrier could be excluded. Indeed, it seems to be the case that there are multiple conformations in fast exchange, separated by only small energetic barriers. These then give rise to many NOEs and single averaged peaks in the n.m.r. spectrum, and vanishing electron density in the X-ray crystallographic maps.

The right-hand panels of Figure 3 again show the value of $\bar{r}(t)$, calculated with the memory function, over the course of the trajectory when time-averaged constraints were used. As would be expected, the plots show the importance of using r^{-3} averaging when calculating forces. Although the instantaneous distances often violate the restraints by 3 to 5 Å, the averaged distance never reaches these peaks. The periods when the instantaneous

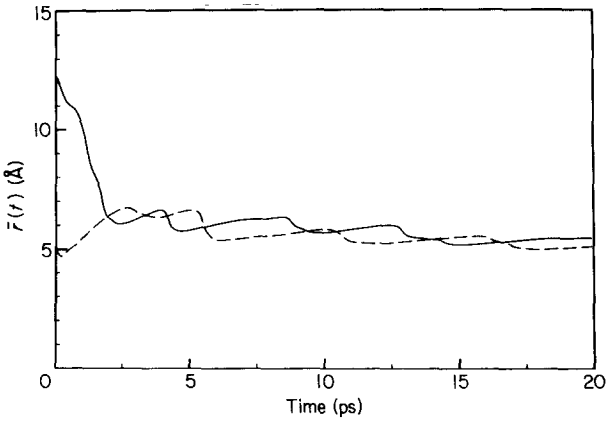

Figure 5. Convergence of average distances over trajectory of structure I with time-averaged distance restraints. Both lines show $\bar{r}(t)$, calculated without the use of a memory function according to eqn (3). The continuous line is calculated from the distance between Tyr15 to Thr13 sites corresponding to the centre top panel of Fig. 3. The broken line corresponds to the Tyr15 to Ser17 distance shown in the centre bottom panel of Fig. 3.

distances are well satisfied bring the average down very rapidly. This also gives rise to the asymmetrical or sawtooth appearance of the oscillations of $\bar{r}(t)$.

The plots also show the effect of one of the other parameters of the simulation, the initial value for $\bar{r}(t)$. As discussed in Theory, this parameter must be assigned some initial value. In these simulations, it was always 0·2 Å less than the size of the distance restraint. So, the top centre panel of Figure 3 shows that at the start of the simulation, the particular NOE was violated by more than 5 Å. The corresponding plot on the right-hand side shows that, initially, $\bar{r}(t)$ was 0·2 Å less than the restraint, so at the start of the simulation, no force was applied. Over the next two picoseconds, however, $\bar{r}(t)$ began to catch up with the instantaneous distance, so the artificial force was gradually turned on. For this particular NOE, this was appropriate behaviour, since, at the start of the simulation, even without the use of an artificial force, the restrained distance was beginning to decrease. From the plots, one can also see the time lag of $\bar{r}(t)$ with respect to the instantaneous r. This is what gives rise to the method's ability to tolerate short transient violations of NOE restraints. As shown in Figure 2(c), an interproton distance may well rise and fall before the average distance has risen and turned on the artificial force.

(b) *Technical points*

When one is setting up a simulation with time-averaged NOEs, values for several parameters must be chosen. Firstly, τ, the time constant for the decay of the memory function used in distance averaging, should be long enough so as not to hinder motions in the system. In the limiting case of $\tau = 0$, there is no

averaging of NOEs and mobility is very restricted. For this work, the value of $\tau = 1{\cdot}25$ picoseconds was chosen after test runs with both tendamistat and small peptides (our unpublished results).

A very important effect of τ is less obvious. It directly affects the rate of motions in the simulation and thus the length of simulation necessary to observe proper averaging. This is explained by Figure 5. The plots show the average $\bar{r}(t)$ calculated up to each point in the trajectory according to equation (3), without the use of a memory function. It is this average that must ultimately satisfy the distance constraints. The particular distances plotted correspond to those in the centre panels of Figure 3. At the start of the simulation, this average follows the instantaneous distances, but it becomes less sensitive as the averaging is conducted over a longer time. Even after 20 picoseconds, it has not completely converged. As a compromise with computational expense, it seems that a simulation should be at about an order of magnitude longer than the value for τ. This value is also what was found in our previous work with time-averaged constraints (see Fig. 3 of Torda *et al.*, 1989).

A much more subtle phenomenon is also occurring in the simulations. The artificial force described by equation (5) at time t is a function not only of the configuration at time t but also of the previous configurations. This means that the force is not conservative. Such a force will act like a stochastic force in that it causes a heating of the system (van Gunsteren *et al.*, 1981). This phenomenon, although small, could be seen in the simulations discussed here. Without time-averaged distance constraints, both structures I and III ran at an average temperature of 299 K, with fluctuations of 7 and 8 K, respectively (1 standard deviation). With time-averaged distance constraints, the systems ran at temperatures of 307($\pm$8) and 306($\pm$8) K. Continuing the analogy with stochastic dynamics, it may prove feasible to replace the current overall temperature coupling (Berendsen *et al.*, 1984) with an individual atom coupling through the Langevin equation (van Gunsteren & Berendsen, 1988) where atomic friction coefficients are replaced by a factor related to the artificial restraint force. We are currently investigating this possibility.

The heating of structures has an unusual property in that it depends on the number of NOE bounds affecting an atom and the extent to which the structure satisfies the bounds. This results in an uneven temperature throughout the molecule. This can be seen by considering the kinetic energy of each atom, averaged over every time step in the 20 picosecond trajectories, plotted in Figure 6. The atomic sequence number used in the plots is an arbitrary scheme, but does follow the numbering of residues. The top plots, from the conventional refinement, show a fairly even distribution of temperature. If the simulations were of infinite length, the plots would be flat lines. The bottom plots show the kinetic energy of each atom during the trajectories with time-averaged distance bounds. Firstly, the variation in kinetic energies is much larger, and, secondly, the general shape of the curve shows the regions of the protein undergoing the most heating. The peaks in the plots near atom 127 reflect the motions of Tyr15, discussed in section (a), above. The peaks near atom 549 are from the side-chain of Ile61, another residue that showed large motions in attempting to satisfy the experimental data. One can also see that, because the temperature coupling is based on the overall temperature, certain parts of the molecule stay relatively cold in order to maintain the average. This explains the generally cool region near the N terminus (low atom sequence numbers) where there were relatively few NOEs. In structural terms, these fluctuations are not a problem and only reflect the rapid motions of a few atoms, usually at the ends of side-chains, driven to satisfy NOE bounds.

This uneven distribution of heating also accounts for the results in Table 3, which showed slightly higher potential energies for structures generated using time-averaged distance constraints. The warm parts of the molecule, on average, spend more time in high-energy parts of conformational space. When the structures are energy minimized and brought to their nearest local energetic minima, the warm regions move, on average, to correspondingly higher energy conformations. Lower energy static conformations can certainly be achieved by a gradual cooling of the system, rather than the rapid quenching of an energy minimization step.

(c) *Implications of time-averaged distance constraints*

It has been previously stated that solution structures based on n.m.r. data represent some kind of average over various dynamic processes (Wüthrich, 1989). To a large extent, this is the justification for using only upper-bound distance constraints and the addition of correction factors to account for motional averaging (Braun *et al.*, 1981). At the same time, the nature of this average is not simple. Structures do not distribute themselves in a simple geometric manner about symmetrical energetic minima. Instead, conformations are distributed according to a Boltzmann distribution over a very uneven and intricate energetic surface. Perhaps this is made clearest not by the example of the oscillations of Tyr15 (Fig. 3), but instead by the apparently random fluctuations in distance shown by the distance plots of Figure 2. That plot does not reflect movement between a number of identifiable conformations, but rather reflects numerous transitions between states accessible to the system at the temperature of the simulation.

Similarly, measured quantities such as the NOE or J-coupling constants are not linear with respect to cartesian or even dihedral co-ordinates and this must be accounted for when generating structures or trajectories to explain the data (Jardetzky, 1980). The physical nature of average structures

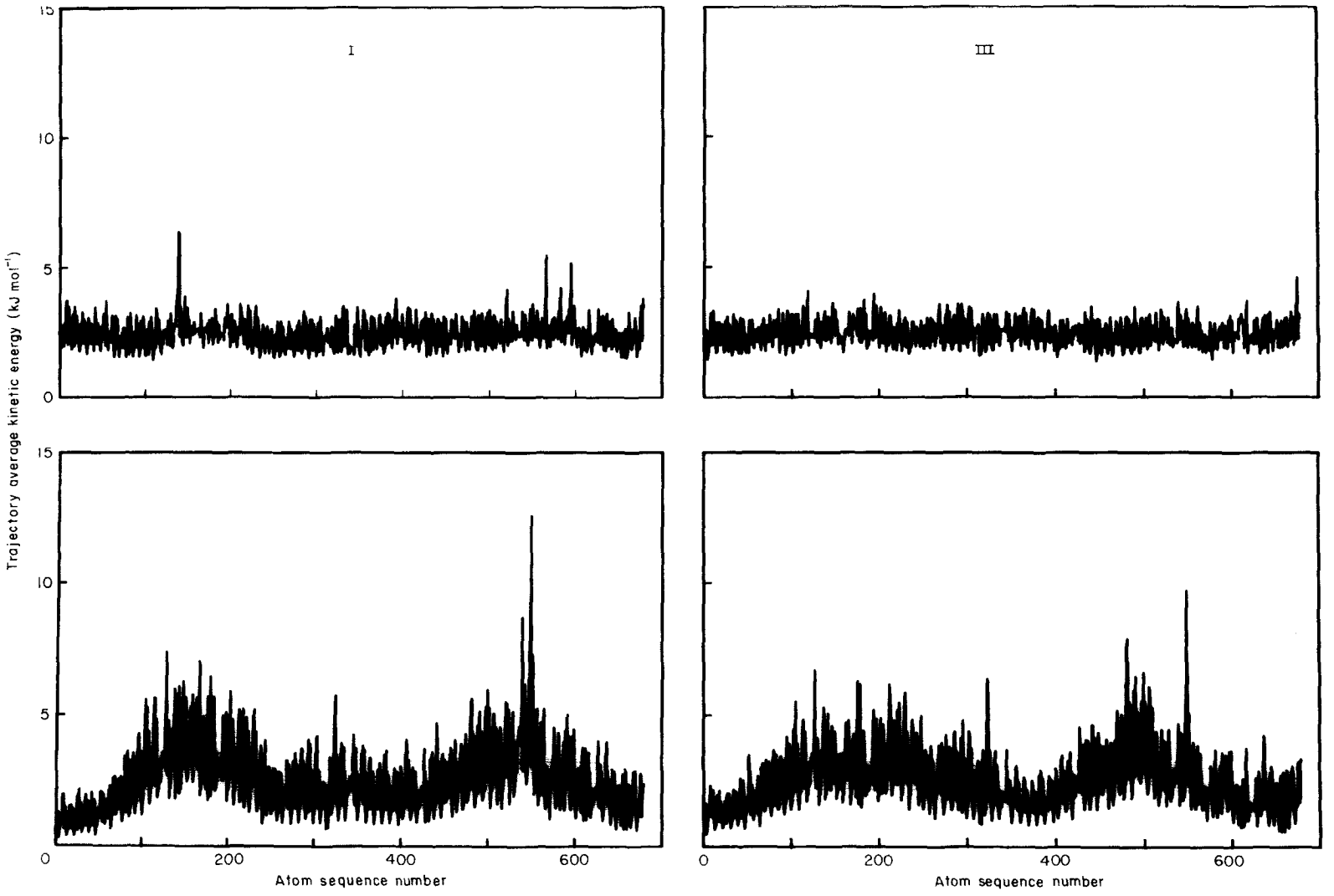

Figure 6. Trajectory average kinetic energy as a function of atomic sequence number. Top plots from simulations using conventional MD refinement; bottom simulations using time-averaged distance constraints. I and III refer to the structure names.

may also be unsatisfactory. If a molecule is shifting between two or more low energy conformations, the average or mid-point conformation may actually be a high-energy structure. It will only be the minimum with respect to a penalty function enforcing static distance bounds. With these considerations, it would appear that only methods such as molecular dynamics simulations or Monte Carlo techniques can achieve the ensemble averaging necessary to reflect the physical properties of the molecules and experimental measurements.

The use of time-averaged distance constraints has implications also for the accuracy of published structures and the methods used to estimate this accuracy. Certainly, tendamistat is one of, if not the most, precisely determined solution structure so far published, but it is not clear how its degree of determination should be estimated. In this work, we have not stressed the r.m.s. differences between structures I and III, since our emphasis is on the conformational space spanned by a molecule through time. The backbone fluctuations in this work in one structure (Fig. 1), however, are actually comparable to the r.m.s. deviations shown by the original distance geometry structures from their mean (Fig. 10 of Kline *et al.*, 1988).

The quantity of experimental constraints for tendamistat also highlights an interesting property of the procedure described here. It is actually the comprehensive nature of the tendamistat dataset that makes some of the internal motions become clear. Using the example of the apparently mobile Tyr15 side-chain, it can be seen that, if this residue's NOEs to either residue 13 or 17 were not observed, then the motion of the side-chain probably would not have been detected. If one of the NOEs was missing, the aromatic ring would have simply been pulled towards the other residue. If neither NOE had been detected, the ring would have simply wobbled around its lowest energy position with respect to the force field and any other NOEs. A case such as this, where extra NOEs lead to extra observed motions, further complicates any attempt to arrive at an acceptable method for estimating the precision of structural determination.

The last implication for structural determination concerns the possibility of experimental errors. Currently, it is presumed that there are some checks on the internal consistency of experimental data. If a dataset is sufficiently complete, then, hopefully, an incorrectly assigned NOE should manifest itself by being violated or causing violations of neighbouring NOEs. With the use of time-averaged NOEs, one could, conceivably, have an error in the dataset and interpret it as evidence of conformational averaging. Obviously, one is much more reliant upon the quality of raw data. In practice, this should not prove to be much of a problem. A distance geometry step is still necessary to generate starting structures and the presence of residual violations should serve at least to identify regions that require further investigation.

(d) *Future applications of time-averaged distance restraints*

As mentioned in Introduction, results from small peptides originally prompted consideration of the possible effects of conformational averaging in structure determination (Kessler *et al.*, 1988; Pepermans *et al.*, 1988). Motional averaging, however, is not unique to small peptides, it is just easier to detect there. One must then consider the use of time-averaged NOEs for protein structure determinations in general. There is already evidence that other structures in the literature would benefit from the method. For example, in the recent publication of the structure of desulphatohirudin, it was stated that some NOESY cross-peaks might be due to transient near approach of proton pairs associated with flexible parts of the molecule (Haruyama & Wüthrich, 1989).

More generally, one must remember that, on the basis of the original tendamistat structures, it was considered that there was no evidence for conformational averaging. By the nature of the method, however, all structures will display additional motions if distance constraints are enforced as averages through time. These additional motions may be especially important if protein structure is not an end in itself. If, for example, a structure is to be used as the basis for designing other molecules, it is essential not merely to know about its geometric average conformation but also to know likely conformations it may adopt in solution.

Another reason for using time-averaged distance constraints is the improvement in conformational searching properties. This is shown, in these results, by the size of the r.m.s. positional fluctuations. Temporarily relaxing the distance restraint penalty function greatly increases the region of space that the molecule is able to enter. Another way to view this is by comparison with other minimization techniques. Methods such as molecular dynamics simulations or simulated annealing (Kirkpatrick *et al.*, 1983) are effective because they can take steps that are unfavourable with respect to the penalty function. The time-averaged distance restraint method modifies the distance restraint term so that transient increases are not penalized. In the particular case of tendamistat, the distance geometry structures were so thoroughly refined that no obvious benefit could be seen in this regard. In the future, we intend to examine the method's searching properties on less thoroughly refined structures.

Two other aspects of this method make it attractive for future structural determinations. Firstly, the artificial force is applied less often and for shorter periods than in a conventional refinement. This reduces the disturbance to the force field. Secondly, the extra mobility and ability to temporarily violate some NOEs mean that it is easier to satisfy other NOEs. One can then use smaller force constants for the artificial terms in the force field, further reducing the disturbance to trajectories.

Despite the advantages, there are also limitations to what can be achieved using time-dependent distance constraints. Most notably, a set of NOEs may result from multiple conformations rapidly exchanging on the n.m.r. timescale. The barriers between these conformations may, however, be large enough to be insurmountable on the much shorter picosecond timescale of a MD simulation. In this case, time-averaged NOEs will still not allow one to see all possible conformations. Another disadvantage with the procedure is that one must simulate for long enough to sample conformational space reasonably and achieve convergence of the average distances. This may be computationally expensive.

In conclusion, we have presented a penalty function for incorporation into molecular dynamics simulations that improves the method's ability to model and reproduce experimental distance constraints. The method reduces disturbance to the normal force field and improves searching properties. Although it was originally designed to address problems in the refinement of small peptides, it is equally important for the refinement of large molecules. In the future, we intend to apply similar procedures to dihedral angle constraints and further refine the determination of protein structures and the assessment of the conformational space that they occupy in solution.

These investigations were supported in part by the Netherlands' Foundation for Chemical Research (SON) with financial aid from the Netherlands' Technology Foundation. We also thank ICI for their financial support and interest in the project. We are grateful to Professors I. D. Kuntz and H. J. C. Berendsen for their suggestions and stimulating discussions. We also thank Werner Braun and Professor Wüthrich for readily providing all of the initial data used in this work and their interest in the methods.

References

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

Billeter, M., Kline, A. D., Braun, W., Huber, R. & Wüthrich, K. (1989). *J. Mol. Biol.* **206**, 677–687.

Braun, W., Bösch, C., Brown, L. R., Gō, N. & Wüthrich, K. (1981). *Biochim. Biophys. Acta*, **667**, 377–396.

de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R. & van Gunsteren, W. F. (1986). *Isr. J. Chem.* **27**, 181–188.

Haruyama, H. & Wüthrich, K. (1989). *Biochemistry*, **28**, 4301–4312.

Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F. & Postma, J. P. M. (1984). *Biopolymers*, **23**, 1513–1518.

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.

Kessler, H., Griesinger, C., Lautz, J., Müller, A., van Gunsteren, W. F. & Berendsen, H. J. C. (1988). *J. Amer. Chem. Soc.* **110**, 3393–3396.

Kirkpatrick, S., Gelatt, C. D., Jr & Vecchi, M. P. (1983). *Science*, **220**, 671–680.

Kline, A. D., Braun, W. & Wüthrich, K. (1988). *J. Mol. Biol.* **204**, 675–724.

Pardi, A., Billeter, M. & Wüthrich, K. (1984). *J. Mol. Biol.* **180**, 741–751.

Pepermans, H., Tourwé, D., van Binst, G., Boelens, R., Scheek, R. M., van Gunsteren, W. F. & Kaptein, R. (1988). *Biopolymers*, **27**, 323–338.

Ryckaert, J.-P., Cicotti, 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. Letters*, **157**, 289–294.

Tropp, J. (1980). *J. Chem. Phys.* **72**, 6035–6043.

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. (1988). *Mol. Sim.* **1**, 173–185.

van Gunsteren, W. F., Berendsen, H. J. C. & Rullmann, J. A. C. (1981). *Mol. Phys.* **44**, 69–95.

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

van Gunsteren, W. F., Boelens, R., Kaptein, R., Scheek, R. M. & Zuiderweg, E. R. P. (1985). In *Molecular Dynamics and Protein Structure* (Hermans, J., ed.), pp. 92–99, Polycrystal Book Service, Western Springs, IL.

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

Wüthrich, K. (1989). *Science*, **243**, 45–50.

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

Edited by R. Huber