Journal of Biomolecular NMR, 6 (1995) 163–170
ESCOM

J-Bio NMR 291

Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble

Jens Fennen*, Andrew E. Torda** and Wilfred F. van Gunsteren***

Physical Chemistry, ETH Zentrum, CH-8092 Zürich, Switzerland

Received 23 March 1995
Accepted 12 May 1995

Keywords: Ensemble averaging; Time averaging; Structure refinement; Molecular dynamics; Somatostatin

Summary

Time-averaging restraints in molecular dynamics simulations were introduced to account for the averaging implicit in spectroscopic data. Space- or molecule-averaging restraints have been used to overcome the fact that not all molecular conformations can be visited during the finite time of a simulation of a single molecule. In this work we address the issue of using the correct Boltzmann weighting for each member of an ensemble, both in time and in space. It is shown that the molecular- or space-averaging method is simple in theory, but requires a priori knowledge of the behaviour of a system. This is illustrated using a five-atom model system and the small cyclic peptide analogue somatostatin. When different molecular conformers that are separated by energy barriers insurmountable on the time scale of a simulation contribute significantly to a measured NOE intensity, the use of space- or molecule-averaged distance restraints yields a more appropriate description of the measured data than conventional single-molecule refinement with or without application of time averaging.

Introduction

The data measured in an NMR experiment is an average over time and space, the acquisition time and the different conformers that are assumed by the molecules in the sample during that time. Therefore, the measured data often cannot be matched by a single conformation. The distribution of conformers in time and space follows Boltzmann's law:

$$\frac{N_a}{N} = \frac{e^{-(E_a - TS_a)/k_BT}}{\sum_a e^{-(E_a - TS_a)/k_BT}} \quad (1)$$

where the conformer a is characterized by an energy E_a and an entropy S_a. k_B is Boltzmann's constant and T is the temperature.

In a molecular dynamics trajectory, the conformations occur with a relative probability that is given by the Boltzmann distribution. Averaging over a sufficiently long trajectory, therefore, gives a correct picture of the averaging in the experiment. This fact is used by the time-averaging methods in structure refinement, like the use of time-averaging distance restraints (Torda et al., 1989, 1990) and time-averaging J-coupling constant restraints (Torda et al., 1993). These simple averaging methods, however, rely on the assumption that, during the simulation, the molecule will visit all conformers which contribute to the measurement. Unfortunately, simulations are of a finite length and barriers which are readily crossed on the NMR time scale may be insurmountable on the simulation time scale (typically a nanosecond).

A possible alternative to averaging over time is to average over a set of conformations that exists simultaneously (Scheek et al., 1991; Kemmink et al., 1993; Bonvin et al., 1994; Mierke et al., 1994). To match the averaging process in the experiment, this set has to be a Boltzmann-weighted ensemble of conformations, with the probability of finding conformer a given by the Boltzmann distribution (Eq. 1) (Van Gunsteren et al., 1994).

In an ideal situation, this requirement can be fulfilled

*Present address: Ciba-Geigy, GE Leder/Pelz, CH-4002 Basel, Switzerland.
**Present address: Research School of Chemistry, Australian National University, Canberra 0200, Australia.
***To whom correspondence should be addressed.

by simultaneously running a (very) large number of MD trajectories starting from random conformations of the same molecule and by averaging over a snapshot of all trajectories at the same time. In this context, an ideal situation means that the trajectories are run sufficiently long to produce a Boltzmann-weighted ensemble in time and that, starting from any conformation, the entire phase space must be accessible. If this is not true, and some areas of the potential energy hypersurface can only be accessed from some particular starting points, then the set of conformations obtained by a snapshot of all trajectories is not Boltzmann distributed. Whereas the conformations within each single trajectory are still Boltzmann distributed, the trajectory does not cover the whole hypersurface and the snapshot contains conformations with probabilities that do not correspond to a Boltzmann distribution. This situation is always found if high-energy barriers separate different low-energy regions of the potential energy hypersurface. It is this situation that the present study will focus on.

Theory

Neglecting the influence of angular fluctuations (Tropp, 1980), the space or molecule average at time t of the distance r_{ij} between nuclei i and j that determines the NOE intensity, denoted by angle brackets, can be defined as:

$$\langle r_{ij}(t)\rangle \equiv \left[\sum_{\alpha=1}^{N_e} p_\alpha(t)\, r_{ij,\alpha}(t)^{-q}\right]^{-1/q} \tag{2}$$

where q is some integer, $r_{ij,\alpha} \equiv |\mathbf{r}_{i,\alpha}(t) - \mathbf{r}_{j,\alpha}(t)|$ is the distance between atoms i and j in molecule α and p_α is the probability of a conformation or molecule α in the ensemble of N_e conformations or molecules at time t, which is given by the Boltzmann distribution:

$$p_\alpha(t) = \frac{e^{-E(\mathbf{r}_\alpha(t))/k_B T}}{\sum_{\alpha'=1}^{N_e} e^{-E(\mathbf{r}_{\alpha'}(t))/k_B T}} \tag{3}$$

in which $E(\mathbf{r}_\alpha(t)) = V_{phys}(\mathbf{r}_\alpha(t))$ is the potential energy of molecule α containing N atoms in the conformation $\mathbf{r}_\alpha \equiv (\mathbf{r}_{1,\alpha}, \mathbf{r}_{2,\alpha}, \ldots, \mathbf{r}_{N,\alpha})$ as calculated by a molecular interaction function V_{phys} at each time step t of the simulation. In this and previous work, we have set q = 3 in Eq. 2. This is valid when the simulation is short relative to the correlation time of angular fluctuations. Averaging over space, however, may be equivalent to a longer simulation, where angular fluctuations cannot be ignored. In this situation, one should set q = 6 in Eq. 2. The space- or molecule-averaged distances could then be used to define a distance restraining term in the potential energy function (Van Gunsteren et al., 1984):

$$\sum_{\text{NOE pairs}(i,j)} V_{dr}(\langle r_{ij}(t)\rangle) \tag{4}$$

with (Van Gunsteren et al., 1985):

$$V_{dr}(\langle r_{ij}(t)\rangle) = \begin{cases} 0 & \langle r_{ij}(t)\rangle \le r_{ij}^0 \\ \frac{1}{2}k^{dr}\left[\langle r_{ij}(t)\rangle - r_{ij}^0\right]^2 & r_{ij}^0 \le \langle r_{ij}(t)\rangle \le r_{ij}^0 + \Delta r \\ k^{dr}\left[\langle r_{ij}(t)\rangle - r_{ij}^0 - \frac{\Delta r}{2}\right]\Delta r & r_{ij}^0 + \Delta r \le \langle r_{ij}(t)\rangle \end{cases} \tag{5}$$

which is harmonic between r_{ij}^0 and $r_{ij}^0 + \Delta r$ and linear beyond the latter distance. The force on atom i exerted by atom j in molecule α resulting from Eq. 5 is:

$$\mathbf{F}_{i,j,\alpha}^{dr} = -\frac{\partial V_{dr}(\langle r_{ij}\rangle)}{\partial \mathbf{r}_{i,\alpha}} = -\frac{\partial V_{dr}(\langle r_{ij}\rangle)}{\partial \langle r_{ij}\rangle} \cdot \frac{\partial \langle r_{ij}\rangle}{\partial r_{ij,\alpha}} \cdot \frac{\partial r_{ij,\alpha}}{\partial \mathbf{r}_{i,\alpha}} \tag{6}$$

Treating the probability p_α as a parameter that adiabatically follows the variation in $\mathbf{r}$ through Eq. 3, the derivative of the molecule-averaged distance (Eq. 2), which is the second factor in Eq. 6, becomes:

$$\frac{\partial \langle r_{ij}\rangle}{\partial r_{ij,\alpha}} = \left[\sum_{\alpha'=1}^{N_e} p_{\alpha'} r_{ij,\alpha'}^{-q}\right]^{-(q+1)/q} p_\alpha r_{ij,\alpha}^{-q-1} = p_\alpha \left[\frac{\langle r_{ij}\rangle}{r_{ij,\alpha}}\right]^{q+1} \tag{7}$$

As in time-averaging restrained refinement (Torda et al., 1989,1990), the (q+1)th power in Eq. 7 causes large fluctuations in the forces, which can be avoided by using the approximation in analogy to time-averaging refinement:

$$\frac{\partial \langle r_{ij}\rangle}{\partial r_{ij,\alpha}} = p_\alpha \tag{8}$$

The Lagrangian L for the system of N_e identical molecules, each containing N atoms, becomes:

$$L = \sum_{\alpha=1}^{N_e} p_\alpha \sum_{i=1}^{N} \frac{1}{2} m_i \left(\frac{d\mathbf{r}_{i,\alpha}}{dt}\right)^2 - \sum_{\alpha=1}^{N_e} p_\alpha V_{phys}(\mathbf{r}_\alpha) - \sum_{\text{NOE pairs}(i,j)} V_{dr}(\langle r_{ij}\rangle) \tag{9}$$

The coordinates of atom i in molecule α are indicated by $\mathbf{r}_{i,\alpha}$ and the set of all atomic coordinates of molecule α by $\mathbf{r}_\alpha \equiv (\mathbf{r}_{1,\alpha}, \mathbf{r}_{2,\alpha}, \ldots, \mathbf{r}_{N,\alpha})$. Treating the probability p_α defined by Eq. 3 as a parameter, the Lagrangian equations of motion for atom i in molecule α become:

$$p_\alpha m_i \frac{d^2 \mathbf{r}_{i,\alpha}}{dt^2} = -p_\alpha \frac{\partial V_{phys}(\mathbf{r}_\alpha)}{\partial \mathbf{r}_{i,\alpha}} - \sum_{\substack{\text{NOE pairs in-} \\ \text{volving atom i}}} \frac{\partial V_{dr}(\langle r_{ij}\rangle)}{\partial \mathbf{r}_{i,\alpha}} \tag{10}$$

Using the approximation of Eq. 8 in Eq. 6, the equations of motion become:

$$m_i \frac{d^2 \mathbf{r}_{i,\alpha}}{dt^2} = \mathbf{F}_{i,\alpha}^{phys} + \mathbf{F}_{i,\alpha}^{dr} \quad (11)$$

with the distance restraining force on atom i of molecule α given by:

$$\mathbf{F}_{i,\alpha}^{dr} = \sum_{\substack{\text{NOE pairs in-}\\ \text{volving atom i}}} \frac{\partial V_{dr}(\langle r_{ij} \rangle)}{\partial \langle r_{ij} \rangle} \cdot \frac{\partial r_{ij,\alpha}}{\partial \mathbf{r}_{i,\alpha}} \quad (12)$$

Because the contribution of each ensemble member α depends strongly on its instantaneous energy $E(\mathbf{r}_\alpha(t)) = V_{phys}(\mathbf{r}_\alpha(t))$ (Eq. 3), it may show large fluctuations over very short periods of time. These would be instantaneously reflected in the calculated average distance. In practice, one is not interested in these short-time fluctuations, so a mean energy can be used, calculated for each molecule averaged over parts of the trajectory. This was done using the expression given in Eq. 13:

$$\bar{E}_\alpha^t = \left[\tau_E\left(1 - e^{-t/\tau_E}\right)\right]^{-1} \int_0^t e^{-t'/\tau_E} E(\mathbf{r}_\alpha(t - t'))dt' \quad (13)$$

instead of $E(\mathbf{r}_\alpha(t))$ in Eq. 3, where τ_E gives the characteristic time for the exponential decay used in the averaging of the energy. This value need only be large enough to remove the worst fluctuations in the calculated energy.

If one wishes to take advantage of the history of a simulation for averaging distances through time as well as space, this is now simply done by using the following expression in Eqs. 4, 5 and 9–12 instead of $\langle r_{ij}(t) \rangle$:

$$\overline{\langle r_{ij} \rangle}^t = \left(\left[\tau_{dr}\left(1 - e^{-t/\tau_{dr}}\right)\right]^{-1} \int_0^t e^{-t'/\tau_{dr}} \langle r_{ij}(t - t') \rangle^{-q} dt' \right)^{-1/q} \quad (14)$$

where τ_{dr} is the characteristic time used for the exponential decay of the memory function used in averaging over distances, which has the same form as in previous work (Torda et al., 1990,1993) using only time averaging. In Eq. 14, q = 3 should be used. Because $\langle r_{ij}(t) \rangle$ is the space-averaged distance, Eq. 14 provides the machinery for combined time and space averaging.

TABLE I

RESTRAINED C-C DISTANCES IN AN SD SIMULATION OF PENTANE

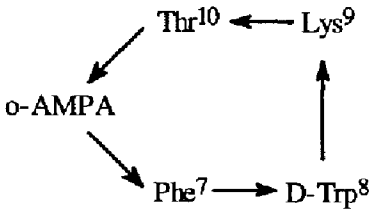
Atom pair (i,j) for distance restraining	Restraining value r_{ij}^0 (Å)	Distance $\langle r_{ij} \rangle$ (Å) (weighted average)[a]
C1-C4	3.80	3.20
C2-C5	3.20	3.80
C1-C5	4.40	4.89

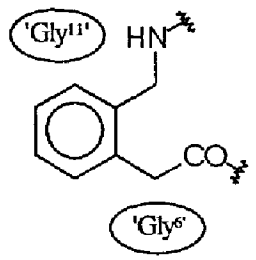
[a] The values are for an equilibrium mixture of conformers.

Methods

All simulations were performed using the stochastic dynamics algorithm (Van Gunsteren and Berendsen, 1988) and the GROMOS suite of programs with the GROMOS force field (Van Gunsteren and Berendsen, 1987) using the 37D4 parameter set for in vacuo simulations. The friction coefficient γ_i for each atom was set to $\gamma_i = \omega_i \gamma$ with $\gamma = 20$ ps^{-1}, where the atomic solvent-accessible surface area ω_i is approximately calculated as proposed by Shi Yun-yu et al. (1988) and updated every 500 steps. The system was weakly coupled ($\tau_T = 0.1$ ps) to a heat bath at 300 K (Berendsen et al., 1984) and the SHAKE algorithm was used to constrain bond lengths with a geometrical precision of 10^{-4} (Ryckaert et al., 1977). No cutoffs were used for long-range interactions.

A first series of test calculations with a weighted space average over $N_e = 3$ molecules was carried out using a modified united atom pentane molecule as a model. The force field parameters of 'normal' pentane were changed in order to mimic the presence of high-energy barriers, which require the use of the space-averaging approach. For this purpose, the force constant for the C1-C2-C3-C4 torsional angle potential energy term was increased from 5.8576 kJ mol^{-1} to 50.0 kJ mol^{-1}; the C2-C3-C4-C5 torsional angle force constant, however, was retained. This resulted in three classes of conformations that could and did not convert into each other during the simulation: class 1, with a C1-C2-C3-C4 torsional angle of 180°, and classes 2 and 3 with this angle being 60° and −60°, respectively. The distance restraints used for this five-atom model were arbitrarily chosen C-C distances. They are different from those





o-AMPA
(*o*-aminomethylphenylacetic acid)

Fig. 1. Cyclic peptide analogue 1 of somatostatin (Van der Elst et al., 1987; Pepermans et al., 1988).

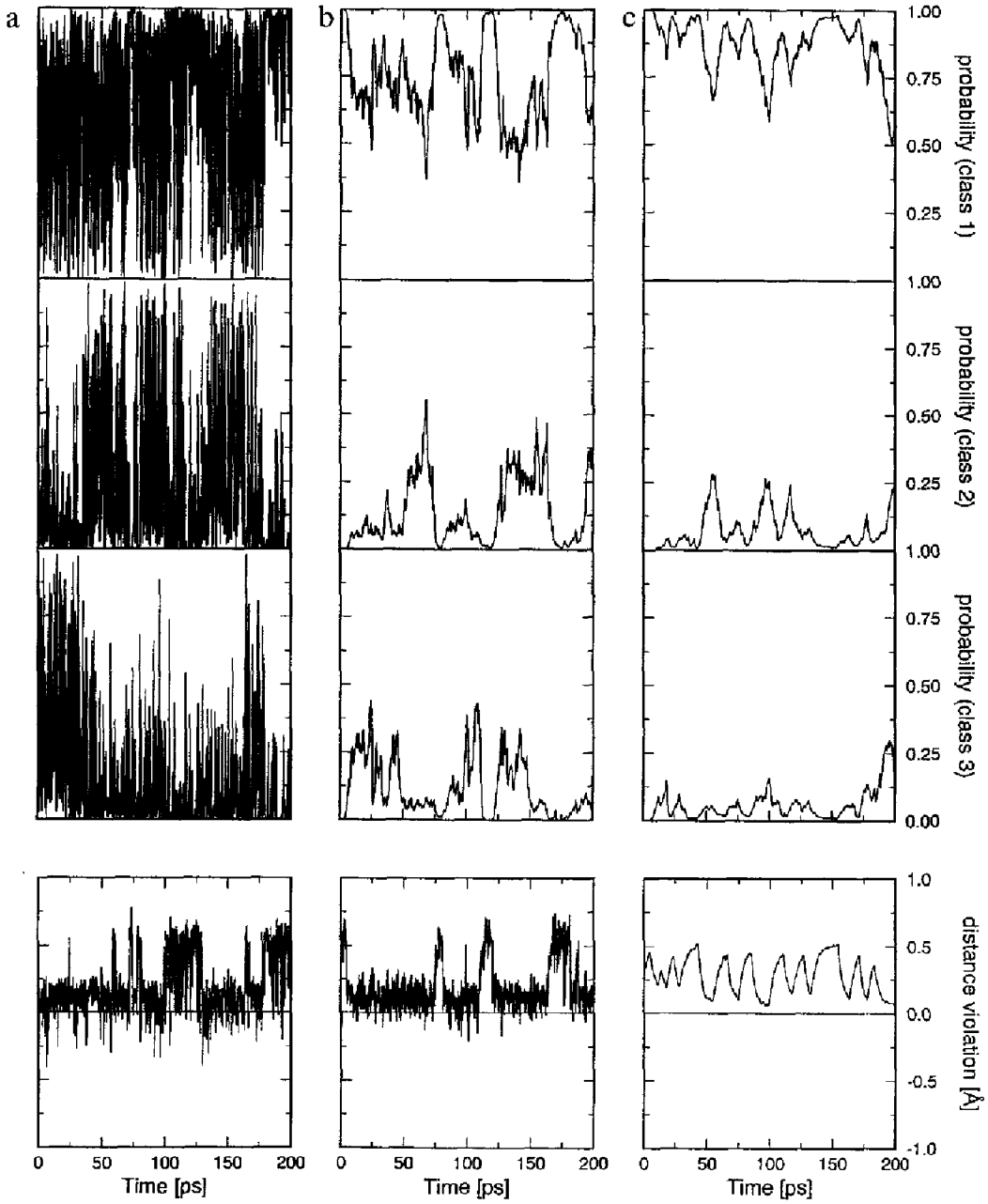


Fig. 2. Conformer probability (upper three rows) and distance violation (lower row) during simulations of pentane using different types of distance restraining. Each plot shows the calculated relative Boltzmann probability of one contributing pentane conformer as a function of time during a stochastic dynamics simulation using molecule-averaging distance restraints ($N_e = 3$). Upper: C1-C2-C3-C4 torsional angle = 180°; middle: 60°; lower: −60°. From left to right: (a) without time averaging of the potential energy ($\tau_E = 0$ ps, $\tau_{dr} = 0$ ps); (b) with time averaging of the potential energy ($\tau_E = 2$ ps, $\tau_{dr} = 0$ ps); (c) combined molecule- and time-averaging distance restraining ($\tau_E = \tau_{dr} = 5$ ps).

calculated as a weighted average from the molecular geometries for an equilibrium mixture of conformers (Table 1). Both upper and lower distance restraint bounds were set to the same value. A force constant of $k^{dr} = 20$ kJ mol^{-1} Å^{-2} was used for the restraining force, which allows violations of up to 0.5 Å at room temperature. A time step of 0.002 ps was used in the integrator and the overall simulation time was 200 ps for all simulations with this model. The contributions of the ensemble members were calculated by using either instantaneous energies (Eq. 3), energies averaged with a characteristic decay time τ_E of 2 ps (Eq. 13), or combined space and time averaging with characteristic decay times τ_E and τ_{dr} of 5 ps (Eq. 14).

As a more realistic test case for simulations using combined space- and time-averaging distance restraints, the cyclic peptide analogue **1** (Van der Elst et al., 1987) of somatostatin was used. The nomenclature and numbering adopted here have been used before (Pepermans et al., 1988) with the o-AMPA spacer as a replacement for two glycine residues. All simulations with this model used

$N_e = 2$, a time step in the integrator of 0.001 ps and simulation times of 100 ps. The force constant for the distance restraining force was set to $k^{dr} = 50$ kJ mol^{-1} Å^{-2} and characteristic decay times τ_E and τ_{dr} of 20 ps were used.

Results

Our calculations with the five-atom model described above clearly show the large fluctuations in the contribution of the ensemble members to $\langle r_{ij}(t) \rangle$ mentioned earlier. A first simulation was carried out using the instantaneous energy to calculate these contributions (Eq. 3). As can be seen in the left-hand side of Fig. 2, very large fluctuations occur in the probabilities of the three conformational classes ($\alpha = 1,2,3$), which in turn cause fluctuations of the instantaneous ensemble-averaged distances $\langle r_{ij}(t) \rangle$ (bottom graph). These fluctuations decreased drastically when the probabilities p_α were calculated from energies that were averaged with a characteristic decay time τ_E of 2 ps according to Eq. 13 (middle column of Fig. 2). Finally, combined ensemble and time averaging (Eq. 14) with characteristic decay times τ_E and τ_{dr} of 5 ps led to a further reduction of these fluctuations (Fig. 2, right-hand side).

A previous NMR study of the somatostatin analogue **1** showed that it exists as a mixture of conformers in fast exchange on the NMR time scale (Pepermans et al., 1988). Using distance geometry (DG) methods as well as restrained energy minimization (EM) and molecular dynamics (MD), it was found that some of the NOE distances could not be reproduced by any single conforma-

TABLE 2
INTERPROTON DISTANCES (Å) FOR SOMATOSTATIN ANALOGUE **1**[a]

	Proton pair				Distance from NOE	Upper bound used in SD	Assignment 1			Assignment 4		
							A	D	A+D	A	D	A+D
Set 1												
1	Phe^7 NH	Phe^7 C$^\alpha$H			2.4	2.6	2.9	2.9	2.8	2.9	2.8	2.8
2	D-Trp^8 NH	Phe^7 C$^\alpha$H			2.1	2.3	2.3	2.3	2.2	2.3	2.3	2.2
3	D-Trp^8 NH	D-Trp^8 C$^\alpha$H			2.5	2.7	2.7	2.7	2.7	2.7	2.8	2.7
4	Lys^9 NH	D-Trp^8 C$^\alpha$H			2.1	2.3	2.1	2.1	2.1	2.1	2.1	2.1
5	Lys^9 NH	Lys^9 C$^\alpha$H			2.8	3.0	2.7	2.7	2.7	2.7	2.7	2.7
6	Thr^{10} NH	Lys^9 C$^\alpha$H			3.0	3.2	3.3	3.3	3.2	3.3	3.3	3.2
7	Thr^{10} NH	Thr^{10} C$^\alpha$H			2.4	2.6	2.8	2.8	2.8	2.8	2.7	2.8
8	Lys^9 NH	Thr^{10} NH			2.3	2.5	2.6	2.6	2.6	2.6	2.6	2.6
9	Thr^{10} NH	Gly^{11} NH			2.4	2.6	3.0	2.9	2.9	2.9	2.9	2.9
Set 2												
	Assignment 1		Assignment 4									
	Gly^6 C$^\alpha$H	Gly^{11} C$^\alpha$H	Gly^6 C$^\alpha$H	Gly^{11} C$^\alpha$H								
10	H_R	H_R	H_S	H_S	2.9	3.1	3.1	3.2	3.3	3.2	3.1	3.4
11	H_R	H_S	H_S	H_R	2.3	2.5	1.7	4.4	2.1	4.4	1.8	2.5
12	H_S	H_R	H_R	H_S	2.5	2.7	4.4	1.7	2.7	1.8	4.4	2.3
13	H_S	H_S	H_R	H_R	2.9	3.1	3.2	3.1	3.4	3.1	3.2	3.4
	Gly^6 C$^\alpha$H	AMPA H_6	Gly^6 C$^\alpha$H	AMPA H_6								
14	H_R		H_S		2.6	2.8	3.5	2.3	2.9	2.3	3.5	2.6
15	H_S		H_R		2.5	2.7	2.3	3.5	2.5	3.5	2.3	2.7
	Gly^6 C$^\alpha$H	AMPA H_3	Gly^6 C$^\alpha$H	AMPA H_3								
16	H_R		H_S		2.5	2.7	2.3	3.5	2.5	3.5	2.4	2.7
17	H_S		H_R		2.5	2.7	3.5	2.4	2.9	2.4	3.5	2.6
	Phe^7 NH	Gly^6 C$^\alpha$H	Phe^7 NH	Gly^6 C$^\alpha$H								
18		H_R		H_S	2.3	2.5	2.6	3.1	2.7	3.2	2.0	2.5
19		H_S		H_R	2.8	3.0	3.3	1.9	2.7	2.4	3.1	2.9
	Phe^7 NH	Gly^{11} C$^\alpha$H	Phe^7 NH	Gly^{11} C$^\alpha$H								
20		H_S		H_R	2.9	3.1	2.9	3.5	3.0	3.6	2.4	3.1
	Gly^{11} NH	Gly^{11} C$^\alpha$H	Gly^{11} NH	Gly^{11} C$^\alpha$H								
21		H_S		H_R	2.4	2.6	2.3	2.5	2.3	2.1	2.6	2.4

[a] Column 1 contains the distances determined from NOE buildup rates. These were used as upper bounds with an error margin of 0.2 Å (column 2). Columns 3–8 list the distances averaged (using Eq. 14) with $\tau_{dr} = \infty$ over a 10–100 ps period of a 100-ps SD simulation. Columns 3, 4, 6 and 7 were calculated from simulations with single molecules (no time or space averaging), starting from either conformer A or conformer D. Columns 5 and 8 derive from simulations using combined ensemble and time averaging with $N_e = 2$, one molecule starting from conformer A, the other from conformer D. Violations of 0.5 Å or more are underlined.

tion. Striking examples were the $Gly^6C^\alpha H$-$Gly^{11}C^\alpha H$ and $Gly^6C^\alpha H'$-$Gly^{11}C^\alpha H'$ distances, which were measured to be 2.3 and 2.5 Å, respectively (Table 2). Using only the NOE distances that were not inconsistent with a single conformation (Set 1 in Table 2), Pepermans et al. (1988) generated a collection of 92 structures using DG and EM methods. These were grouped into four classes (A–D) according to which distances were sufficiently short to account for any NOE that was not used to generate the structures. It was found that these classes could also be defined by the sign of the $Gly^6\phi$ and $Gly^{11}\psi$ torsional angles (Table 3). After performing restrained MD on a representative structure of each class, weighted averages of these NOE distances were calculated for the four possible stereochemical assignments of the Gly^6C^α and $Gly^{11}C^\alpha$ protons. Only assignments 1 and 4 (Table 4) were in agreement with the experimental data.

In the present study, two sets of restraints were used, one assuming assignment 1 and the other assuming assignment 4. First, separate SD simulations with representative structures for each class as starting points were performed. Preliminary calculations showed some differences when compared to those of Pepermans et al. (1988). Whereas these authors saw no conversions between conformational classes, such interconversions did occur in the present work. The reasons for this are the following: (i) the current simulations are three times as long; (ii) stochastic dynamics rather than Newtonian molecular dynamics is used here; and (iii) in the present work the complete set of distance restraints from Table 1 is used. Because we were interested in assessing the space-averaging refinement method in the presence of insurmountable energy barriers, we added an artificial torsional angle potential energy term for the $Gly^6\phi$ and $Gly^{11}\psi$ angles. This potential energy term, having a force constant of 50 kJ mol^{-1} and minima at 90° and −90°, served to stop interconversion between classes A and D. It was used in all subsequent simulations of molecule **1**. Classes B and C still disappeared after a few steps and were not visited again during the simulation time.

The geometries at the end of the separate trajectories for classes A and D were taken as starting points for an SD simulation with space- and time-averaged distance

TABLE 3
DEFINITION OF CLASSES A–D BY $Gly^6\phi$ AND $Gly^{11}\psi$ TORSIONAL ANGLE SIGNS[a]

Class	$Gly^6\phi$	$Gly^{11}\psi$
A	−	+
B	+	+
C	−	−
D	+	−

[a] This definition corresponds to the one given by Pepermans et al. (1988) in Fig. 3 and Table 1; not to the one given in Table 2 of the same publication.

TABLE 4
NUMBERING OF POSSIBLE STEREOCHEMICAL ASSIGNMENTS OF MOLECULE 1

Assignment	Gly^6 $C^\alpha H$	Gly^{11} $C^\alpha H$
1	pro-*R*	pro-*S*
2	pro-*R*	pro-*R*
3	pro-*S*	pro-*S*
4	pro-*S*	pro-*R*

restraints and simulations were performed for both of the possible stereochemical assignments 1 and 4. It turned out that the potential energies as given by the force field are not suitable for direct use in the calculation of Boltzmann probabilities (Eq. 3). In the force field, the structures of class A were extremely stable compared to those of class D, resulting in a probability ratio of almost 1:0. This force field inaccuracy necessitates the introduction of an energy scaling factor k_E, which reduces the dependence of the ensemble members' contributions to $\langle r_{ij}(t)\rangle$ on the calculated potential energy $E(\mathbf{r}_\alpha(t)) = V_{phys}(\mathbf{r}_\alpha(t))$. Equation 3 then becomes

$$p_\alpha = \frac{e^{-k_E E(r_\alpha)/k_B T}}{\sum_{\alpha'=1}^{N_e} e^{-k_E E(r_{\alpha'})/k_B T}} \qquad (15)$$

For the present study, k_E was set to 0.2, which reduces an error in the calculated potential energy of about 10 kJ mol^{-1} to 2 kJ mol^{-1}, which is of the order of magnitude of $k_B T$ at room temperature. In this way, conformations with a very high potential energy still do not contribute to the simulation, but conformations with a relative energy within the margin of error of the interaction function V_{phys} are not strictly ruled out. Using this approach, the restrained distances were averaged over a 10–100 ps time interval for both the space-averaging restraining simulations and the separate runs (Table 2), using Eq. 14 with $\tau_{dr} = \infty$. From Fig. 3, which compares the distance violations of the space and time averaging with the non-averaging simulations, it can easily be seen that there are no violations larger than about 0.3 Å in the space-averaging simulations, whereas the non-averaging simulations show violations of up to 1.9 Å.

Discussion and Conclusions

When time-averaged restraints were first introduced, it was noted that they would only be reliable if a simulation was able to visit all the conformations which contributed to the spectroscopic observations. It was also clear that this will not usually be the case.

Conformations which might be rapidly exchanging on the milli- or microsecond time scale may be indistinguishable in an NMR experiment, but act as if they were different systems on the simulation time scale (typically

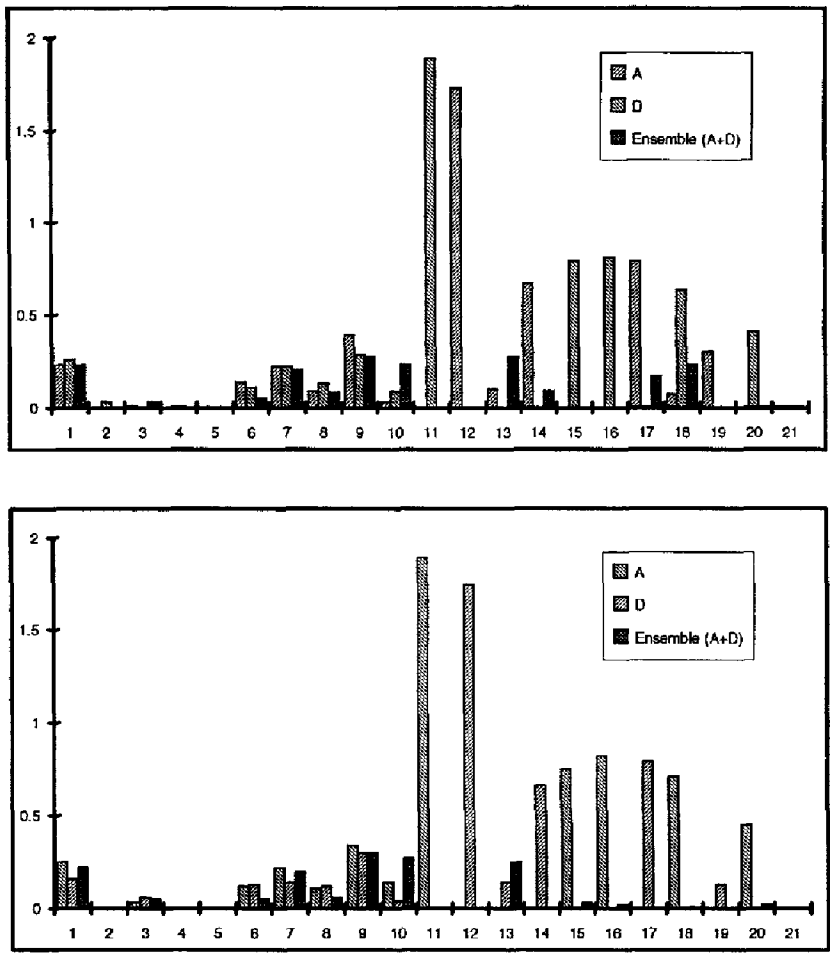


Fig. 3. Comparison of NOE distance constraint violations (Å) for conventional single molecule and molecule-averaging distance restraints (somatostatin analogue 1, top: stereochemical assignment 1; bottom: stereochemical assignment 4). The results A and D are obtained by single molecule refinement, and the result (A+D) by time- and molecule-averaging distance restraining refinement, as described in Table 2.

10^{-14}–10^{-10} s). Naïvely, one might expect that the use of a sufficient number of conformations, scattered on all sides of the conformational barriers, can overcome this limitation. In fact, this is not the case and it is interesting to compare our results with those of previous studies to see the limitations of the different models.

First, our work can be contrasted with that of Bonvin et al. (1994) who used a sophisticated restraining function based on NOE intensities, rather than the derived distances. These authors employed averaging over molecules, but with $p_\alpha = N_c^{-1}$ ($\alpha = 1,2,...,N_c$) and with $N_c = 8$ identical molecules, each having the same starting coordinates. From the point of view of barrier crossing, this does not offer an advantage over a single longer simulation. On the other hand, this approach also does not introduce any new approximations in terms of the distribution of structures on the energy surface. Because any of the eight molecules are equivalent, and their conformations may interconvert to each other, they are as close to a Boltzmann distribution as a single molecule simulation that is eight times as long.

The same type of space or molecule averaging with $p_\alpha = N_c^{-1}$ ($\alpha = 1,2,...,N_c$) and use of identical starting structures has been applied to J-value restraining in a tripeptide with $N_c = 500$ by Mierke et al. (1994). Again, from the viewpoint of barrier crossing, this does not offer an advantage over a single long simulation.

Some of the problems encountered in the work presented here had been noticed previously in the case of a two-state model. Scarsdale et al. (1988) used molecular mechanics energies to estimate population fractions p_α. Although these authors did not perform MD simulations, they did notice that molecular mechanics-based energies tend to yield almost insignificant occupation of certain (higher energy) conformations. Kim and Prestegard (1989, 1990) used a two-state model ($N_c = 2$) and assigned weights (p_α) to each conformer based on a fit to the experimental data, thereby avoiding the problem of relying on molecular mechanics energies to assign population fractions p_α. The most appealing aspect of this approach is that, to the extent that the two-state model is correct, the populations should also be correct. The clear limitation here is that the fitting process would not easily generalise to many conformations.

In contrast to the work described above, one may note that the approach described here generalises to any number of conformations and does not require crossing of high-energy barriers. This, however, has come at the expense of introducing the nonphysical parameter k_E. Clearly, as Scarsdale et al. (1988) noted, relying on a possibly not very accurate molecular mechanics energy to estimate a probability p_α introduces a source of error. Another drawback of the current approach is the fact that it is sensitive to the choice of starting structures. The set of initial conformations should include a member from every conformational class. Furthermore, the energy barriers between the different classes must not be crossed, since the Boltzmann weighting (Eq. 3) would not be correct in that case.

Probably the worst approximation is that, as in all quoted studies, we have used the potential energy rather than the free energy to calculate probabilities p_α. There is no justification for this beyond expediency. Methods do exist to estimate free energies during a simulation, but for this work we had no desire to introduce any more approximations. It is certainly interesting to speculate whether use of free energies in Eq. 3 would show fluctuations in $\langle r_{ij}(t) \rangle$ as large as seen here and would require the use of the scaling parameter k_E.

While molecule or space averaging introduces complications, it is interesting to see that at least one problem associated with time-averaging restraints may be alleviated by it. It has been noted that time-averaging restraints lead to a non-conservative force field and possibly to slight heating of the simulated system (Torda et al., 1990; Pearlman, 1994). The problem is worst when a part of the molecule has to shuttle between different conformers and is artificially driven to do so by the restraints. In contrast, molecule-averaging restraints allow a system to simultaneously occupy several conformers, rather than having to shuttle between them.

In summary, space- or molecule-averaging restraints can be used, but do require the use of at least one ad hoc weighting factor. This is probably the most inelegant aspect of the whole approach. It might then be wise to use only time averaging as a first approach and space or molecule averaging when justified by evidence of the presence of high barriers separating conformers that contribute significantly to the NOE intensities.

Acknowledgements

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

References

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

Bonvin, A.M.J.J., Boelens, R. and Kaptein, R. (1994) *J. Biomol. NMR*, **4**, 143–149.

Kemmink, J., Van Mierlo, C.P.M., Scheek, R.M. and Creighton, T.E. (1993) *J. Mol. Biol.*, **230**, 312–322.

Kim, Y. and Prestegard, J.H. (1989) *Biochemistry*, **28**, 8792–8797.

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

Mierke, D.F., Scheek, R.M. and Kessler, H. (1994) *Biopolymers*, **34**, 559–563.

Pearlman, D.A. (1994) *J. Biomol. NMR*, **4**, 1–16.

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

Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) *J. Comput. Chem.*, **23**, 327–341.

Scarsdale, J.N., Ram, P., Yu, R.K. and Prestegard, J.H. (1988) *J. Comput. Chem.*, **9**, 133 147.

Scheek, R.M., Torda, A.E., Kemmink, J. and Van Gunsteren, W.F. (1991) In *Computational Aspects of the Study of Biological Macromolecules by NMR* (Eds, Hoch, J.C., Poulsen, F.M. and Redfield, C.), Plenum Press, New York, NY, pp. 209–217.

Shi Yun-yu, Wang Lu and Van Gunsteren, W.F. (1988) *Mol. Simul.*, **1**, 369–388.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1989) *Chem. Phys. Lett.*, **157**, 289–294.

Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.

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

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

Van der Elst, P., Van den Berg, E.M., Van der Auwera, L., Pepermans, H., Zeeuws, R., Tourwé, D. and Van Binst, G. (1987) *Int. J. Pept. Protein Res.*, **29**, 319–330.

Van Gunsteren, W.F., Kaptein, R. and Zuiderweg, E.R.P. (1984) In *Proceedings of the NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics* (Ed., Olson, W.K.), CECAM, Orsay, pp. 79–92.

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

Van Gunsteren, W.F. and Berendsen, H.J.C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen.

Van Gunsteren, W.F. and Berendsen, H.J.C. (1988) *Mol. Simul.*, **1**, 173–185.

Van Gunsteren, W.F., Brunne, R.M., Gros, P., Van Schaik, R.C., Schiffer, C.A. and Torda, A.E. (1994) *Methods Enzymol.*, **239**, 619 654.