Journal of Biomolecular NMR, **12:** 501–508, 1998.
KLUWER/ESCOM

On using time-averaging restraints in molecular dynamics simulation

Walter R.P. Scott, Alan E. Mark and Wilfred F. van Gunsteren*
Laboratory of Physical Chemistry, ETH Zentrum, CH-8092 Zürich, Switzerland

Received 11 March 1998; Accepted 3 July 1998

Key words: computer simulation, J-coupling constants, molecular dynamics, structure refinement, time-averaging restraints

Abstract

Introducing experimental values as restraints into molecular dynamics (MD) simulations to bias the values of particular molecular properties, such as nuclear Overhauser effect intensities or distances, ^3J coupling constants, chemical shifts or crystallographic structure factors, towards experimental values is a widely used structure refinement method. To account for the averaging of experimentally derived quantities inherent in the experimental techniques, time-averaging restraining methods may be used. In the case of structure refinement using ^3J coupling constants from NMR experiments, time-averaging methods previously proposed can suffer from large artificially induced structural fluctuations. A modified time-averaged restraining potential energy function is proposed which overcomes this problem. The different possible approaches are compared using stochastic dynamics simulations of antamanide, a cyclic peptide of ten residues.

Introduction

Experimental techniques such as X-ray diffraction and NMR spectroscopy are widely used to derive structural information from molecules in solution, solid state or in crystal form. A molecular structure, in the form of Cartesian coordinates of the N_a atoms $\vec{r} = (\vec{r}_1, \ldots, \vec{r}_{N_a})$ cannot be directly observed in the experiment. Instead, N_{obs} configuration dependent parameters $\vec{q}(\vec{r}) = (q_1(\vec{r}), \ldots, q_{N_{obs}}(\vec{r}))$ are observed, which are subsequently used as input to a refinement procedure, the result of which is a structure or set of structures which best satisfies the experimental data. This contribution concentrates on the underlying assumptions and models that flow into the refinement procedure and, as a consequence, influence the results obtained.

The experimental methods considered here have in common that the observed values are averages *over time* and over *an ensemble of molecules*, i.e.

$$\vec{q}^{obs} = \langle\{\vec{q}(\vec{r}(t))\}\rangle, \tag{1}$$

where $\{\}$ denotes an average over the molecules in the system at any given point in time and $\langle\rangle$ denotes an average over time. Deriving information about $\vec{r}$ from $\vec{q}^{obs}$ is hampered by the following considerations:

- The number of observables N_{obs} is typically too small to uniquely derive $\vec{r}$.
- As a result of the effect of averaging inherent in the experimental techniques, the observables can contain conflicting data which cannot be reconciled with one single configuration.
- Generally, it is not clear how to invert the averages in Equation 1 to obtain $\vec{r}$.
- The relation $\vec{q}(\vec{r})$ is often empirical, as is, for example, the Karplus equation (Karplus, 1959).
- The relation $\vec{q}(\vec{r})$ is itself often not invertible, i.e. $\vec{r}(\vec{q})$ is not uniquely defined, e.g. by the Karplus equation

In the molecular modelling refinement approach, an empirical model of the molecule under study is constructed. Averaging is introduced by means of combining the experimental data with the empirical model in a molecular dynamics simulation from which a trajectory of coordinates results. An average of the observable can then be calculated from this trajectory

*To whom correspondence should be addressed.

and compared to the experimental value. In general, only the *time* average is considered, as the averages over time and over molecules are the same under the assumptions of infinite dilution and ergodicity. In certain cases these assumptions do not hold, notably due to crystal packing effects.

The most widely used method of obtaining an ensemble of configurations consistent with a set of experimental data is the restraining or penalty function approach. In this approach, the physical potential energy function $V^{phys}(\vec{r}(t))$ is combined with a penalty function $V^{qr}(\vec{r}(t); \vec{q}^{obs})$ to produce $V(\vec{r}(t))$

$$V(\vec{r}(t)) = V^{phys}(\vec{r}(t)) + V^{qr}(\vec{r}(t); \vec{q}^{obs}) \qquad (2)$$

which is then used in the coordinate generation procedure. The choice of $V^{qr}(\vec{r}(t); \vec{q}^{obs})$ is important. The means by which the experimental data is introduced into the simulation must be consistent with the manner in which the experimental data was collected if serious artefacts are to be avoided.

In conventional instantaneous restraining (denoted by the symbol *inst*), a functional form harmonic in $q_i(\vec{r}(t))$ is generally chosen

$$V_i^{inst}(\vec{r}(t); K_i^{qr}, q_i^{obs}) = \frac{K_i^{qr}}{2}(q_i(\vec{r}(t)) - q_i^{obs})^2 \qquad (3)$$

for the *i*-th observable. A better approach is to take the averaging inherent to the experimental technique into account. In conventional time-averaged restraining (denoted by the symbol *ctave*), the potential energy function is harmonic in the average of $q_i(\vec{r}(t))$:

$$V_i^{ctave}(\vec{r}(t); K_i^{qr}, q_i^{obs}) = \frac{K_i^{qr}}{2}\left(\overline{q_i(\vec{r}(t))} - q_i^{obs}\right)^2, \qquad (4)$$

where

$$\overline{q_i(\vec{r}(t))} = \frac{1}{[\tau_{qr}(1 - \exp(-\frac{t}{\tau_{qr}}))]} \int_0^t \exp(-\frac{t'}{\tau_{qr}}) q_i(\vec{r}(t - t'))dt' \qquad (5)$$

is the weighted average used during the simulation. The true time average value over the course of a simulation is

$$\langle q_i(\vec{r}(t))\rangle = \frac{1}{t}\int_0^t q_i(\vec{r}(t'))dt' \qquad (6)$$

which may be calculated from the trajectory and compared to the experimental value after the simulation. During the course of a simulation, however, $\langle q_i(\vec{r}(t))\rangle$ becomes increasingly insensitive to instantaneous fluctuations for increasing values of *t*. To avoid this, a decay with characteristic decay time τ_{qr} may be built into the averaging process during the simulation. This approach has been successfully applied to refinement with experimental data originating from Nuclear Overhauser Effect (NOE) measurements with the quantity q taken as the inverse third power of an atom-atom distance d, $q \equiv d^{-3}$ (Torda et al., 1990), ^3J coupling constant experiments with the quantity q taken as the ^3J coupling constant, $q \equiv J$ (Torda et al., 1993) and X-ray diffraction experiments with the quantity q taken as the crystallographic structure factor F, $q \equiv F$ (Schiffer et al., 1995).

However, it has been observed in ^3J coupling constant restraining that, although the experimental average $\langle q_i(\vec{r}(t))\rangle$ is achieved, the fluctuations can be much larger than observed for unrestrained simulations (Pearlman, 1994; Nanzer et al., 1997). This is due to the fact that overshooting occurs during the simulation. As the average value lags behind the instantaneous value in time, a force due to the restraining function continues to be applied for some time after $q_i(\vec{r}(t))$ equals q_i^{obs}. The problem is inherent to all standard time-averaging restraining applications, but is less manifest in NOE distance restraining applications for the following reasons:

- In most cases, only the upper bound violations of atom-atom distances are penalized.
- Overshooting towards low distances is blocked by the van der Waals repulsive interactions.
- The average $\langle d^{-3}\rangle$ is dominated by lower values of *d*, so $\langle d^{-3}\rangle$ quickly follows d^{-3} when the upper bound is satisfied.

In order to address this problem which is particularly important in the case of ^3J-value restraining, we propose alternative functional forms to Equation 4 for $V^{qr}(\vec{r}(t); \vec{q}^{obs})$ in the next Section. The test results are presented in the Section Results and discussed in the Section Discussion.

Materials and methods

In conventional time-averaged ^3J-value restraining (Torda et al., 1993), the potential energy function for a single upperbound ^3J restraint *i* is

$$V_i^{ctave}(\vec{r}(t); K_i^{qr}, q_i^0) = \frac{K_i^{qr}}{2}\left[\overline{q_i(\vec{r}(t))} - q_i^0\right]^2 H(\overline{q_i(\vec{r}(t))}; q_i^0), \qquad (7)$$

where $H(x; x_o)$ is the Heavyside step function

$$H(x; x_o) = \begin{cases} 0 & \text{if } x < x_o \\ & \\ 1 & \text{otherwise} \end{cases}, \quad (8)$$

$q_i(\vec{r}(t))$ is the instantaneous coupling constant calculated from coordinates using the Karplus equation

$$q_i(\vec{r}(t)) = A_K \cos^2(\phi_i(\vec{r}(t))) + B_K \cos(\phi_i(\vec{r}(t))) + C_K \quad (9)$$

with the empirically derived Karplus constants A_K, B_K and C_K, $\overline{q_i(\vec{r}(t))}$ is its time average with an exponential memory function (Equation 5) and q_i^0 is the experimentally determined coupling constant. The corresponding lower bound restraint is obtained by interchanging the arguments of the Heavyside function in Equation 7.

In practice, $\overline{q_i(\vec{r}(t))}$ is not calculated during the simulation by applying Equation 5; instead, use is made of the equivalence

$$\begin{aligned} &\overline{q_i(\vec{r}(t))} \\ &= \overline{A_K \cos^2(\phi_i(\vec{r}(t))) + B_K \cos(\phi_i(\vec{r}(t))) + C_K} \\ &= A_K \overline{\cos^2(\phi_i(\vec{r}(t)))} + B_K \overline{\cos(\phi_i(\vec{r}(t)))} + C_K. \end{aligned} \quad (10)$$

Equation 5 is then applied to the individual cosine terms and discretized to obtain expressions for $\overline{\cos^2(\phi_i(\vec{r}(t)))}$ and $\overline{\cos(\phi_i(\vec{r}(t)))}$. For an arbitrary power of m, the equivalent of Equation 5 is

$$\overline{\cos^m(\phi(\vec{r}(t)))} = \frac{1}{[\tau_{qr}(1 - \exp(-\frac{t}{\tau_{qr}}))]} \int_0^t \exp(-\tfrac{t-t'}{\tau_{qr}}) \cos^m(\phi_i(\vec{r}(t')))dt'. \quad (11)$$

During a simulation using discrete time steps of size Δt, the average at time t_n can be approximated by the sum

$$\begin{aligned} &\overline{\cos^m(\phi_i(\vec{r}(t_n)))} \\ &= \frac{1}{[\tau_{qr}(1 - \exp(-\frac{t_n}{\tau_{qr}}))]} \sum_{i=0}^{n} \exp(-\frac{t_n - t_i}{\tau_{qr}}) \cos^m(\phi_i(\vec{r}(t_i)))\Delta t \\ &= \frac{1}{[\tau_{qr}(1 - \exp(-\frac{t_n}{\tau_{qr}}))]} \sum_{i=0}^{n-1} \exp(-\frac{t_n - t_i}{\tau_{qr}}) \cos^m(\phi_i(\vec{r}(t_i)))\Delta t \\ &+ \frac{1}{[\tau_{qr}(1 - \exp(-\frac{t_n}{\tau_{qr}}))]} \cos^m(\phi(\vec{r}(t_n)))\Delta t \\ &= \exp(\frac{-\Delta t}{\tau_{qr}}) \frac{[\tau_{qr}(1 - \exp(-\frac{t_n - 1}{\tau_{qr}}]}{[\tau_{qr}(1 - \exp(-\frac{t_n}{\tau_{qr}}]} \overline{\cos^m(\phi_i(\vec{r}(t_{n-1})))} \\ &+ \frac{1}{[\tau_{qr}(1 - \exp(-\frac{t_n}{\tau_{qr}}))]} \cos^m(\phi_i(\vec{r}(t_n)))\Delta t \end{aligned} \quad (12)$$

For $t_n >> \tau_{qr}$ this reduces to

$$\overline{cos^m(\phi_i(\vec{r}(t_n)))} \approx \exp(-\frac{\Delta t}{\tau_{qr}}) \overline{\cos^m(\phi_i(\vec{r}(t_{n-1})))} + \frac{\Delta t}{\tau_{qr}} \cos^m(\phi_i(\vec{r}(t_n))). \quad (13)$$

Equation 13 is equivalent to Equation 11 in (Torda et al., 1993):

$$\overline{cos^m(\phi_i(\vec{r}(t)))} \approx \exp(-\frac{\Delta t}{\tau_{qr}}) \overline{\cos^m(\phi_i(\vec{r}(t - \Delta t)))} + \left(1 - \exp(-\frac{\Delta t}{\tau_{qr}})\right) \cos^m(\phi_i(\vec{r}(t))), \quad (14)$$

using $\left(1 - \exp(-\frac{\Delta t}{\tau_{qr}})\right) \approx \frac{\Delta t}{\tau_{qr}}$ for $\Delta t << \tau_{qr}$. Thus, the potential energy function, Equation 7, reads

$$\begin{aligned} &V_i^{ctave}(\vec{r}(t); K_i^{qr}, q_i^0) \\ &= \frac{K_i^{qr}}{2} [\overline{q_i(\vec{r}(t))} - q_i^0]^2 H(\overline{q_i(\vec{r}(t))}; q_i^0) \\ &= \frac{K_i^{qr}}{2} [A_K \overline{\cos^2(\phi_i(\vec{r}(t)))} \\ &\quad + B_K \overline{\cos(\phi_i(\vec{r}(t)))} + C_K - q_i^0]^2 \\ &\times \; H(\overline{q_i(\vec{r}(t))}; q_i^0) \end{aligned} \quad (15)$$

with the corresponding force

$$\begin{aligned} \vec{f}_i^{qr}(\vec{r}, t) &= -K_i^{qr} [\overline{q_i(\vec{r}(t))} - q_i^0] \\ &\quad H(\overline{q_i(\vec{r}(t))}; q_i^0) \frac{\partial \overline{q_i(\vec{r}(t))}}{\partial \vec{r}(t)} \\ &= +K_i^{qr} [\overline{q_i(\vec{r}(t))} - q_i^0] H(\overline{q_i(\vec{r}(t))}; q_i^0) \\ &\times \left[2A_K \cos(\phi_i(\vec{r}(t))) \frac{\partial \overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial \cos^2(\phi_i(\vec{r}(t)))} \right. \\ &\quad \left. + B_K \frac{\partial \overline{\cos(\phi_i(\vec{r}(t)))}}{\partial \cos(\phi_i(\vec{r}(t)))} \right] \\ &\times \; \sin(\phi_i(\vec{r}(t))) \frac{\partial \phi_i(\vec{r}(t))}{\partial \vec{r}(t)} \end{aligned} \quad (16)$$

Setting $X = \frac{\partial \overline{\cos(\phi_i(\vec{r}(t)))}}{\partial \cos(\phi_i(\vec{r}(t)))} = \frac{\partial \overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial \cos^2(\phi_i(\vec{r}(t)))}$, the force reads

$$\begin{aligned} \vec{f}_i^{qr}(\vec{r}, t) &= K_i^{qr} [\overline{q_i(\vec{r}(t))} - q_i^0] \\ &\quad H(\overline{q_i(\vec{r}(t))}; q_i^0) X \end{aligned}$$

$$\times\ [2A_K \cos(\phi_i(\vec{r}(t))) + B_K]$$
$$\sin(\phi_i(\vec{r}(t)))\frac{\partial\phi_i(\vec{r}(t))}{\partial\vec{r}(t)} \qquad (17)$$

Note that $X = \frac{\Delta t}{\tau_{qr}}$ if Equation 13 is applied, or $X = \left(1 - \exp(-\frac{\Delta t}{\tau_{qr}})\right)$ if Equation 14 is applied. This means that the restraining force is inversely proportional to the exponential memory relaxation time τ_{qr} (for $\Delta t << \tau_{qr}$). In order to obtain comparable restraining forces for different values of τ_{qr}, one would have to choose the values of the force constant K_i^{qr} proportional to $[1-\exp(-\frac{\Delta t}{\tau_{qr}})]^{-1}$. Since the choice of the value for the force constant K_i^{qr} is empirical anyway, the factor X is set to 1 in Equation 17 by default in the current implementation of time averaging in the GROMOS96 code (Scott et al., 1998; van Gunsteren et al., 1996). The values in Torda et al. (1993) were produced using $X = 1$.

The problem of generating too large fluctuations of $q_i(\vec{r}(t))$ when using conventional time averaging restraining is due to the fact that the restraining function 15 only depends indirectly on the instantaneous value $q_i(\vec{r}(t))$ at time t through the average $\overline{q_i(\vec{r}(t))}$. The overshooting effect can be reduced by increasing the weight of the instantaneous value compared to that of the average value in the restraining function. Below, we propose two different simple modifications of the restraining function 7 that lead to a larger weight of $q_i(\vec{r}(t))$ relative to $\overline{q_i(\vec{r}(t))}$ when determining the restraining energy and force, while keeping these quantities a continuous function of $q_i(\vec{r}(t))$ and $\overline{q_i(\vec{r}(t))}$.

In the biquadratic restraining function (symbol *biquad* or *bq*), the instantaneous value $q_i(\vec{r}(t))$ and the average value $\overline{q_i(\vec{r}(t))}$ are equally weighted:

$$\begin{aligned} &V_i^{biquad}(\overline{q_i(\vec{r}(t))}, q_i(\vec{r}(t)); K_i^{qrbq}, q_i^0) \\ &= \frac{K_i^{qrbq}}{2}[\overline{q_i(\vec{r}(t))} - q_i^0]^2[q_i(\vec{r}(t)) - q_i^0]^2 \\ &\times\ H(q_i(\vec{r}(t)); q_i^0)H(\overline{q_i(\vec{r}(t))}; q_i^0). \end{aligned} \qquad (18)$$

The corresponding force is

$$\begin{aligned} \vec{f}_i^{qr}(\vec{r},t) = &-K_i^{qrbq}\Big([\overline{q_i(\vec{r}(t))} - q_i^0] \\ &[q_i(\vec{r}(t)) - q_i^0]^2\ \frac{\partial\overline{q_i(\vec{r}(t))}}{\partial\vec{r}(t)} \\ &+\ [\overline{q_i(\vec{r}(t))} - q_0^i]^2[q_i(\vec{r}(t)) - q_i^0] \\ &\frac{\partial q_i(\vec{r}(t))}{\partial\vec{r}(t)}\Big) \\ &\times\ H(\overline{q_i(\vec{r}(t))}; q_i^0)H(q_i(\vec{r}(t); q_i^0) \\ = &+K_i^{qrbq}\Big(\Big[\overline{q_i(\vec{r}(t))} - q_i^0\Big][q_i(\vec{r}(t)) - q_i^0]^2 \\ &\times\ \Big[2A_K\cos(\phi_i(\vec{r}(t)))\frac{\partial\overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial\cos^2(\phi_i(\vec{r}(t)))} \\ &+\ B_K\frac{\partial\overline{\cos(\phi_i(\vec{r}(t)))}}{\partial\cos(\phi_i(\vec{r}(t)))}\Big] \\ &+\ \Big[\overline{q_i(\vec{r}(t))} - q_i^0\Big]^2\Big[q_i(\vec{r}(t)) - q_i^0\Big] \\ &\Big[2A_K\cos(\phi_i(\vec{r}(t))) + B_K\Big]\Big) \\ &\times\ \sin(\phi_i(\vec{r}(t)))\frac{\partial\phi_i(\vec{r}(t))}{\partial\vec{r}(t)}H(\overline{q_i(\vec{r}(t))}; q_i^0) \\ &H(q_i(\vec{r}(t)); q_i^0) \\ = &+K_i^{qrbq}\Big(\Big[\overline{q_i(\vec{r}(t))} - q_i^0\Big]\Big[q_i(\vec{r}(t)) - q_i^0\Big]^2 \\ &X + \Big[\overline{q_i(\vec{r}(t))} - q_i^0\Big]^2\Big[q_i(\vec{r}(t)) - q_i^0\Big]\Big) \\ &\times\ \Big[2A_K\cos(\phi_i(\vec{r}(t))) + B_K\Big]\sin(\phi_i(\vec{r}(t))) \\ &\frac{\partial\phi_i(\vec{r}(t))}{\partial\vec{r}(t)} \\ &\times\ H(\overline{q_i(\vec{r}(t))}; q_i^0)H(q_i(\vec{r}(t)); q_i^0). \end{aligned} \qquad (19)$$

The elliptic restraining function

$$\begin{aligned} &V_i^{ell}(\overline{q_i(\vec{r}(t))}, q_i(\vec{r}(t)); K_i^{qr}, q_i^0, A_{ell}) \\ &= \frac{K_i^{qr}}{2}\Big[A_{ell}\overline{q_i(\vec{r}(t))} + (1-A_{ell})q_i(\vec{r}(t)) - q_i^0\Big]^2 \\ &\times H(A_{ell}\overline{q_i(\vec{r}(t))} + (1-A_{ell})q_i(\vec{r}(t); q_i^0) \end{aligned} \qquad (20)$$

is obtained from Equation 7 by replacing the average $\overline{q_i(\vec{r}(t))}$ by a linear combination of the average $\overline{q_i(\vec{r}(t))}$ and the instantaneous value $q_i(\vec{r}(t))$. The corresponding force is

$$\begin{aligned} &\vec{f}_i^{qr}(\vec{r},t) \\ &= -K_i^{qr}\Big[A_{ell}\overline{q_i(\vec{r}(t))} + (1-A_{ell})q_i(\vec{r}(t)) - q_i^0\Big] \\ &\times\left(A_{ell}\frac{\partial\overline{q_i(\vec{r}(t))}}{\partial\vec{r}(t)} + (1-A_{ell})\frac{\partial q_i(\vec{r}(t))}{\partial\vec{r}(t)}\right) \\ &H(A_{ell}\overline{q_i(\vec{r}(t))} + (1-A_{ell})q_i(\vec{r}(t)); q_i^0) \\ &= +K_i^{qr}\Big[A_{ell}\overline{q_i(\vec{r}(t))} + (1-A_{ell})q_i(\vec{r}(t)) - q_i^0\Big] \\ &\times\left(A_{ell}\left[2A_K\cos(\phi_i(\vec{r}(t)))\frac{\partial\overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial\cos^2(\phi_i(\vec{r}(t)))}\right.\right. \end{aligned}$$

$$
\begin{aligned}
&+ B_K \frac{\partial \overline{\cos(\phi_i(\vec{r}(t)))}}{\partial \cos(\phi_i(\vec{r}(t)))}\Bigg] \\
&+ (1 - A_{ell})\left[2A_K \cos(\phi_i(\vec{r}(t))) + B_K\right] \Bigg) \\
&\times \sin(\phi_i(\vec{r}(t)))\frac{\partial \phi_i(\vec{r}(t))}{\partial \vec{r}(t)} H(A_{ell}\overline{q_i(\vec{r}(t))} \\
&\quad +(1 - A_{ell})q_i(\vec{r}(t)); q_i^0) \\
&= +K_i^{qr}\left[A_{ell}\overline{q_i(\vec{r}(t))} + (1 - A_{ell})q_i(\vec{r}(t)) - q_i^0\right] \\
&\times (A_{ell}X + (1 - A_{ell}))[2A_K \cos(\phi_i(\vec{r}(t))) + B_K] \\
&\times \sin(\phi_i(\vec{r}(t)))\frac{\partial \phi_i(\vec{r}(t))}{\partial \vec{r}(t)} H(A_{ell}\overline{q_i(\vec{r}(t))} \\
&\quad +(1 - A_{ell})q_i(\vec{r}(t)); q_i^0). \qquad (21)
\end{aligned}
$$

The parameter $A_{ell} \in [0 \ldots 1]$ is a mixing parameter. For $A_{ell} = 0$ and using q_i^0 both as upper and lower bound, Equation 20 reduces to the instantaneous restraining potential energy function in Equation 3. For $A_{ell} = 1$ it reduces to the conventional time-averaged restraining potential energy function in Equation 7.

Results

In order to assess the three time-averaging methods of restraining, a series of simulations for a cyclic decapeptide, antamanide (Karle et al., 1979) was conducted. This is an interesting system, as previous studies have found no single conformation which could explain the NOE values measured (Kessler et al., 1988, 1989; Brüschweiler et al., 1991). All simulations were performed using the GROMOS96 biomolecular simulation package (Scott et al., 1998; van Gunsteren et al., 1996) and the GROMOS96 43A1 force field (van Gunsteren et al., 1996). Starting from the X-ray structure (Karle et al., 1979), an energy minimisation was performed with the six backbone HN–H_α ^3J value restraints in place (Table 1). Simulations ensued from this minimised structure, with velocities generated from a Maxwell distribution at 300 K. All simulations, each of 1 ns in length, were performed in vacuo using the GROMOS96 stochastic dynamics simulation algorithm (van Gunsteren and Berendsen, 1988) with the SHAKE procedure to constrain bond lengths (Ryckaert et al., 1977) at a temperature of 300 K using a time step of 0.002 ps. The atomic friction coefficients were set to 19 ps^{-1} for atoms. No temperature coupling was employed and no cut-offs were used for long-range interactions. K_i^{qr} was set either to 10 $kJmol^{-1}$ Hz^{-2} or 50 000 $kJmol^{-1}$ Hz^{-2} in all cases, with K_i^{qrbq} set to 10 $kJmol^{-1}$ Hz^{-4} or 50 000 $kJmol^{-1}$ Hz^{-4} for the biquadratic time-averaged functional form. The Karplus constants $A_K = 9.4$ Hz, $B_K = -1.1$ Hz and $C_K = 0.4$ Hz were taken from Bystrov (1976) despite the existence of newer values (Pardi et al., 1984) in order to compare to the literature (Kessler et al., 1988; Torda et al., 1993). A value of $\tau_{qr} = 10$ ps was chosen for the averaging simulations. Coordinates and energies were saved to file every ps, yielding 1000 values from which the averages $\langle q_i(\vec{r}(t))\rangle$ in Tables 1, 2 and 3 were calculated using Equation 6. In addition, the average and root mean square (rms) fluctuations of the potential energy of the system *excluding* the restraining term, and the average rms atomic position fluctuation were calculated

$$\Delta X = \left\langle \frac{1}{N_p} \sum_{i=1}^{N_p} [\vec{r}_i(t) - \langle \vec{r}_i(t) \rangle]^2 \right\rangle^{\frac{1}{2}} \qquad (22)$$

both for all atoms ($N_p = 110$) and for just the C_α atoms ($N_p = 10$). The averages involving positions are calculated after a superposition of the structures taking all atoms into account. In Equation 22, the square of a vector is defined as its scalar or dot product.

Discussion

Table 1 summarises the results obtained using a force constant of $K_i^{qr} = 10$ $kJmol^{-1}$ Hz^{-2} ($K_i^{qrbq} = 10$ $kJmol^{-1}$ Hz^{-4} in the case of biquadratic restraining) and setting $\frac{\partial \overline{\cos(\phi_i(\vec{r}(t)))}}{\partial \cos \phi_i(\vec{r}(t)))}$ and $\frac{\partial \overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial \cos^2 \phi_i(\vec{r}(t)))}$ both to 1. The unrestrained simulation fails to reproduce the experimental results, especially for residues 1 and 6. Clearly, the physical force field alone is unable to reproduce all experimental ^3J values on the time scale simulated. The conventional instantaneous restraining method produces better averages, although the ^3J values for residues 5 and 10 are not satisfactorily reproduced. The fluctuations are small when compared to the unrestrained simulation, an indication that the method is restricting the molecule's motion. The conventional time-averaging method achieves the correct averages, however, the fluctuations are increased for residues 1 and 6, the ones that needed the most restraining. Time averaged restraining using the biquadratic functional form results in worse averages (residues 4, 6 and 9) and reduced fluctuations

Table 1. Comparison of simulation results (averages and rms fluctuations) of the antamanide test system with $\frac{\partial\overline{\cos(\phi_i(\vec{r}(t)))}}{\partial\cos(\phi_i(\vec{r}(t)))}$, $\frac{\partial\overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial\cos^2(\phi_i(\vec{r}(t)))} = 1$ and $K_i^{qr} = 10$ kJmol^{-1} Hz^{-2} and $K_i^{qrbq} = 10$ kJmol^{-1} Hz^{-4} (see text). Experimental values are taken from Kessler et al. (1988). Calculated ^3J values in Hz and average positional rms fluctuations in nanometres are full trajectory averages calculated using Equations 6 and 22. Entries labelled 'potential energy' are averages and rms fluctuations in kJmol^{-1} of the potential energy of the system excluding the contribution of the restraining term

	^3J coupling constant (Hz)												potential		pos. fluc. (nm)	
Residue	1Val		4Ala		5Phe		6Phe		9Phe		10Phe		energy			
Restraint	1		2		3		4		5		6		(kJmol^{-1})		C_α	all
Experiment	7.3	–	8.6	–	6.8	–	6.6	–	8.3	–	6.7	–	–	–	–	–
Unrestrained	10.0	1.8	7.4	2.3	8.0	2.2	9.9	1.4	7.1	2.5	8.0	2.0	136	26	0.06	0.11
Instantaneous	7.1	0.5	8.6	0.5	7.4	0.5	6.4	0.5	8.3	0.5	7.3	0.5	178	27	0.05	0.10
$A_{ell} = 0.1$	7.1	0.5	8.6	0.5	7.4	0.6	6.4	0.5	8.3	0.5	7.3	0.5	184	29	0.05	0.09
$A_{ell} = 0.4$	7.2	0.6	8.7	0.7	7.0	0.7	6.5	0.6	8.4	0.7	6.9	0.7	208	27	0.10	0.17
$A_{ell} = 0.6$	7.2	0.8	8.7	0.8	6.9	0.9	6.5	0.8	8.3	0.8	6.8	0.9	205	28	0.10	0.16
$A_{ell} = 0.8$	7.1	1.1	8.6	1.1	7.3	1.1	6.4	1.1	8.3	1.1	7.2	1.1	187	30	0.05	0.09
$A_{ell} = 0.9$	7.6	1.6	8.6	1.5	6.9	1.4	6.7	1.6	8.3	1.5	6.8	1.3	166	34	0.05	0.09
$A_{ell} = 0.95$	7.7	1.9	8.7	1.6	6.8	1.6	6.8	1.8	8.3	1.7	6.8	1.7	168	33	0.06	0.11
Conv. t-ave	7.5	3.0	8.6	2.3	6.9	2.6	6.8	2.8	8.2	2.3	6.8	2.4	188	32	0.08	0.15
Biquadratic	7.6	0.8	4.6	2.2	7.2	0.5	2.6	1.5	6.9	2.8	7.0	0.6	213	27	0.07	0.13

Table 2. Comparison of simulation results (averages and rms fluctuations) of the antamanide test system with $\frac{\partial\overline{\cos(\phi_i(\vec{r}(t)))}}{\partial\cos(\phi_i(\vec{r}(t)))}$, $\frac{\partial\overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial\cos^2(\phi_i(\vec{r}(t)))} = \frac{\Delta t}{\tau_{qr}}$, $K_i^{qr} = 10$ kJmol^{-1} Hz^{-2} and $K_i^{qrbq} = 10$ kJmol^{-1} Hz^{-4} (see text). Experimental values are taken from Kessler et al. (1988). Calculated ^3J values in Hz and average positional rms fluctuations in nanometres are full trajectory averages calculated using Equations 6 and 22. Entries labelled 'potential energy' are averages and rms fluctuations in kJmol^{-1} of the potential energy of the system excluding the contribution of the restraining term

	^3J coupling constant (Hz)												potential		pos. fluc. (nm)	
Residue	1Val		4Ala		5Phe		6Phe		9Phe		10Phe		energy			
Restraint	1		2		3		4		5		6		(kJmol^{-1})		C_α	all
Experiment	7.3	–	8.6	–	6.8	–	6.6	–	8.3	–	6.7	–	–	–	–	–
Unrestrained	10.0	1.8	7.4	2.3	8.0	2.2	9.9	1.4	7.1	2.5	8.0	2.0	136	26	0.06	0.11
Instantaneous	7.1	0.5	8.6	0.5	7.4	0.5	6.4	0.5	8.3	0.5	7.3	0.5	178	27	0.05	0.10
$A_{ell} = 0.1$	7.2	0.6	8.7	0.6	6.9	0.6	6.5	0.6	8.3	0.6	6.8	0.6	203	27	0.08	0.14
$A_{ell} = 0.4$	6.9	0.8	8.6	0.9	7.7	0.8	6.3	0.8	8.3	0.9	7.6	0.8	181	29	0.05	0.09
$A_{ell} = 0.6$	7.3	1.4	8.6	1.3	7.7	0.9	6.3	1.3	8.4	1.2	6.8	1.2	181	29	0.07	0.12
$A_{ell} = 0.8$	8.5	2.0	8.5	1.7	7.0	1.8	7.6	2.2	8.2	1.8	6.9	1.7	151	29	0.06	0.10
$A_{ell} = 0.9$	8.5	2.6	8.4	2.0	6.9	2.0	7.9	2.6	8.0	2.3	7.4	1.9	151	32	0.06	0.11
$A_{ell} = 0.95$	8.5	2.9	8.0	2.2	7.4	2.0	7.8	3.0	7.7	2.4	7.7	1.8	148	31	0.06	0.11
Conv. t-ave	2.1	1.8	8.3	2.1	6.7	2.0	8.6	2.5	5.1	3.1	8.4	0.4	148	26	0.06	0.11
Biquadratic	2.6	1.7	6.1	2.8	7.3	0.6	3.3	1.7	6.6	3.0	7.3	0.6	172	26	0.05	0.10

Table 3. Comparison of simulation results (averages and rms fluctuations) of the antamanide test system with $\frac{\partial\overline{\cos(\phi_i(\vec{r}(t)))}}{\partial\cos(\phi_i(\vec{r}(t)))}$, $\frac{\partial\overline{\cos^2(\phi_i(\vec{r}(t)))}}{\partial\cos^2(\phi_i(\vec{r}(t)))} = \frac{\Delta t}{\tau_{qr}}$, $K_i^{qr} = 50\ 000$ kJmol^{-1} Hz^{-2} and $K_i^{qrbq} = 50\ 000$ kJmol^{-1} Hz^{-4} (see text). Experimental values are taken from Kessler et al. (1988). Calculated ^3J values in Hz and average positional rms fluctuations in nanometres are full trajectory averages calculated using Equations 6 and 22. Entries labelled 'potential energy' are averages and rms fluctuations in kJmol^{-1} of the potential energy of the system excluding the contribution of the restraining term

	^3J coupling constant (Hz)												potential energy		pos. fluc. (nm)	
Residue	1Val		4Ala		5Phe		6Phe		9Phe		10Phe					
Restraint	1		2		3		4		5		6		(kJmol^{-1})		C_α	all
Experiment	7.3	–	8.6	–	6.8	–	6.6	–	8.3	–	6.7	–	–	–	–	–
Unrestrained	10.0	1.8	7.4	2.3	8.0	2.2	9.9	1.4	7.1	2.5	8.0	2.0	136	26	0.06	0.11
Conv. t-ave	7.5	3.0	8.7	2.1	6.9	2.6	6.7	3.0	8.3	2.2	6.9	2.3	196	38	0.09	0.15
Biquadratic	7.2	2.8	6.3	3.4	6.9	2.4	5.7	3.1	7.2	3.2	6.6	2.5	651	247	0.24	0.34

(residues 1, 5 and 10), even when compared to the unrestrained simulations. The elliptic time-averaged restraining method was applied using different values of A_{ell}. For small values, the method shows the same behaviour as instantaneous restraining which corresponds to $A_{ell} = 0$. For larger values of A_{ell}, the ^3J values are better reproduced overall. Note that the averages for residues 5 and 10 generally improve with increasing A_{ell}. The fluctuations also increase, but do not become excessive when compared to the unrestrained run. In these simulations, the molecule moves between conformations to produce the correct averages overall. Furthermore, the average potential energy (excluding the restraining term) decreases with increasing A_{ell} beyond 0.4. Values around $A_{ell} = 0.95$ produce a low average potential energy and a good compromise between correct ^3J value averages and fluctuations.

Table 2 was produced with $X = \frac{\Delta t}{\tau_{qr}}$ and a force constant of $K_i^{qr} = 10$ kJmol^{-1} Hz^{-2} ($K_i^{qrbq} = 10$ kJmol^{-1} Hz^{-4} in the case of biquadratic restraining). The various methods are affected to varying degrees by the choice of X. In conventional time-averaged restraining, the force that is applied when setting $X = \frac{\Delta t}{\tau_{qr}}$ is typically much smaller than is the case with $X = 1$. In the simulations discussed here, it is $\frac{\tau_{qr}}{\Delta t} = 5000$ times smaller. With this much smaller force, conventional time-averaging fails to achieve the correct experimental averages. As Table 3 shows, comparable averages and fluctuations in the ^3J values as in Table 1 can be achieved using conventional time-averaging by choosing $X = \frac{\Delta t}{\tau_{qr}}$ and a correspondingly high force constant. As expected, the average potential energy is higher in Table 3. For the elliptic restraining method the effect of the choice of X depends on the value of A_{ell}, as a comparison of Tables 1 and 2 reveals. In Table 1 the correct averages are largely maintained with increasing A_{ell} (only the fluctuations increase as expected). In Table 2 the quality of some (residues 1 and 6) of the calculated averages deteriorates with increasing A_{ell}. In this case the method behaves progressively like conventional time-averaging which fails for $X = \frac{\Delta t}{\tau_{qr}}$ in combination with a low K_i^{qr}. The effect of the choice of X on the biquadratic restraining method is difficult to ascertain, either from Equation 19 or from the comparison of Tables 1 and 2. In either case, the method performs poorly. With the much higher force constant (Table 3), the quality of the averages improves, but the fluctuations are large.

The results of the various simulations of antamanide illustrate that both proposed modifications of the conventional time-averaging restraining function (Equation 7), the biquadratic restraining function (Equation 18) and the elliptic restraining function (Equation 20) do achieve their goal of reduced fluctuations in the restrained quantities, in this case ^3J values, when compared to conventional time-averaging restraining as used in the literature (Torda et al., 1990; Schiffer et al., 1995; Nanzer et al., 1997). Use of the biquadratic restraining function has the undesired side-effect of a larger discrepancy between the averaged ^3J values and the target (restraint) ^3J values, and of a higher potential energy of the molecule.

The elliptic restraining function (Equation 20) with $X = 1$ and A_{ell} around 0.95 best fulfills the combined requirements of i) low potential energy; ii)

agreement with the experimental values; while iii) producing fluctuations comparable to those produced by an unrestrained simulation for the molecular model of antamanide considered here.

In conclusion, one may say that the overshooting problem inherent in the use of time-averaging restraining in molecular dynamics refinement of molecular structure can be alleviated by using an elliptic restraining function in which the relative weight of the time-averaged and the instantaneous restrained quantity (parameter A_{ell}) can be adapted to tune averages and fluctuations to desired values. However, the proper choice of the values of the parameters of the restraining function, the force constant K_i^{qr}, the memory relaxation time τ_{qr} and the weight A_{ell} will depend on the type of molecule and the number and types of restraints. A wholly satisfactory solution to the problem of structure refinement may have to wait for improved force fields and much longer simulations, which may reproduce the experimentally measured values without the use of any restraining function.

References

Brüschweiler, R., Blackledge, M. and Ernst, R.R. (1991) *J. Biomol. NMR*, **1**, 3–11.

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

Karle, I.L., Wieland, T., Schermer, D. and Ottenheym, H.C.J. (1979) *Proc. Natl. Acad. Sci. USA*, **76** 1532–1536.

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

Kessler, H., Bats, J.W., Lautz, J. and Müller, A. (1989) *Liebigs Ann. Chem.* , **9**, 913–928.

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

Nanzer, A.P., Torda, A.E., Bisang, C., Weber, C., Robinson, J.A. and van Gunsteren, W.F. (1997) *J. Mol. Biol.*, **267**, 1012–1025.

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

Pearlman, D.A. (1994) *J. Biomol. NMR*, **4**, 279–299.

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

Schiffer, C.A., Gros, P. and van Gunsteren, W.F. (1995) *Acta Cryst.*, **D51**, 85–92.

Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. and van Gunsteren, W.F. (1998) *J. Phys. Chem. A.* to be submitted.

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

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

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

van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996) *Biomolecular Simulation : The GROMOS96 Manual and User Guide.* VdF: Hochschulverlag AG an der ETH Zürich and BIOMOS b.v., Zürich, Groningen.