ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Magnetic Resonance 164 (2003) 19–27

JMR Journal of Magnetic Resonance

www.elsevier.com/locate/jmr

Orientation restraints in molecular dynamics simulations using time and ensemble averaging

B. Hess and R.M. Scheek*

Department of Biophysical Chemistry, State University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 23 December 2002; revised 14 May 2003

Abstract

In this article we present methodology for simulating protein dynamics while imposing restraints derived from NMR measurements on partially ordered molecules. Such measurements may include residual dipolar couplings and chemical-shift anisotropies. We define a restraint potential for use in molecular dynamics and energy minimization. The presented potential is consistent with the simultaneously optimized molecular order tensor. Restraining can be performed with time and ensemble averaging. We performed a large number of molecular dynamics simulations of the histidine containing phosphocarrier protein with restraints on backbone N–H vector orientations derived from residual dipolar couplings. From these simulations it is evident that the use of time- or ensemble-averaged restraints is essential to leave the fluctuations of the restrained vectors unaffected. Without averaging the fluctuations of the restrained vectors are reduced significantly. This also has the effect of decreasing the apparent molecular order-parameter tensor.

Keywords: Residual dipolar couplings; Chemical-shift anisotropies; Restraints; Time and ensemble averaging; Molecular dynamics

1. Introduction

Residual dipolar couplings can provide a wealth of detailed information in addition to information from nuclear Overhauser effects (NOEs). When a molecule tumbles isotropically in solution the dipolar couplings, which are well known from solid-state NMR, are averaged to zero. By slightly orienting molecules in solution the dipolar couplings are partially reintroduced. Dipolar couplings provide very accurate data, as they are measured as frequency splittings and not as intensities, like NOEs. Dipolar couplings can provide valuable long-range information. When the orientations of vectors within two parts of a molecule are known, the relative orientation of the two parts can be derived. Care has to be taken when applying orientation restraints derived from residual dipolar couplings, because such restraints can be (partially) fulfilled by local reorientations of inter-nuclear vectors, while a more global reordering may be required.

It is therefore no surprise that orientation restraints (derived from residual dipolar couplings) have been introduced for structure refinement, complementing distance restraints (derived from NOEs) and dihedral-angle restraints (derived from J-couplings), see for instance Bayer et al. [1], Clore and Garrett [2] and Schwalbe et al. [3]. In a typical protocol for applying such restraints a restraint potential is defined, which has a minimum when the measured value (NOE, J-coupling, dipolar coupling) and the value that can be back-calculated from the model are the same, and which rises, usually quadratically, with increasing difference between the measured and calculated value. The restraining forces that are calculated from such a restraint potential ensure that in the refined model all measured values are in agreement with the model values within experimental error. When the model used consists of only a single molecule, this is too strong a requirement, since, as we argued before, measured values need only be reproduced

* Corresponding author. Fax: +31-50-3634800.
E-mail address: R.Scheek@chem.rug.nl (R.M. Scheek).

1090-7807/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1090-7807(03)00178-2

on average, over time and/or over a larger number of molecules. The 'NMR-structures,' which make up a large part of our protein-structure databases, must be viewed as an appealing visualisation of the datasets used, but not necessarily as an accurate model of a protein molecule in solution.

The case of the designed triple-stranded β-sheet peptide betanova [4] is illuminating: while inter-strand NOEs clearly show the antiparallel alignments of strands 1 and 2, as well as the antiparallel alignment of strands 2 and 3, there is no justification for the additional assumption that these features should be realised simultaneously in a single structure. Therefore it should come as no surprise that the proposed 'NMR-structure' of betanova indeed shows the designed three-stranded anti-parallel β-sheet, whereas molecular-dynamics simulations suggest a more dynamic 'structure,' with frequent formation and breaking of two two-stranded β-sheets. Needless to say that on average the NMR data are reproduced equally well by the dynamic model. The lesson to learn from this is that the simultaneity restraint, which is implicit in single-structure refinement, may lead to well-defined features in a static model, which, however, must be considered as modeling artefacts, being the result of unrecognised dynamic features of the modeled molecule.

To improve this situation we introduced 'ensemble-averaging' protocols, where the back-calculation includes ensemble averaging and restraining is only applied when such calculated ensemble averages are not in agreement with the measured ensemble averages. The present paper extends this idea to orientation restraints that can be derived from residual dipolar couplings or chemical-shift anisotropies. Several groups have shown, using residual dipolar coupling data, that proteins are floppier than we thought based on relaxation data alone [5–7]. This emphasizes the need for ensemble-averaged restraining, since time-averaged restraining only captures fluctuations that occur within the averaging time. We show that time- and ensemble-averaged orientation restraints can be applied during a molecular-dynamics simulation of a single protein molecule or an ensemble of two or more molecules, resulting in a trajectory which on average not only agrees with the experimental values within the true experimental error (typically a fraction of a Hertz for the dipolar couplings used in this work), but which retains most of the ns-timescale motions that unrestrained molecular-dynamics simulations of proteins typically show. The implementation we present allows the user to perform dynamics simulations using restraints derived from measurements on samples with different orders, e.g., because different orienting media were employed. In this way orientation restraints will be useful in combination with other restraints (distance, dihedral angle) to improve the quality of a dynamic protein model.

2. Calculating dipolar couplings

In an NMR experiment orientations of vectors can be measured when a molecule does not tumble completely isotropically in the solvent. Two examples of such orientation measurements are residual dipolar couplings (between two nuclei) or chemical-shift anisotropies. An observable for a vector $\mathbf{r}_i$ can be written as follows:

$$\delta_i = \frac{2}{3}\mathrm{tr}(\mathbf{S}\mathbf{D}_i), \tag{1}$$

where $\mathbf{S}$ is the dimensionless order tensor of the molecule. The tensor $\mathbf{D}_i$ is given by:

$$\mathbf{D}_i = \frac{c_i}{\|\mathbf{r}_i\|^{\alpha}} \begin{pmatrix} 3xx-1 & 3xy & 3xz \\ 3xy & 3yy-1 & 3yz \\ 3xz & 3yz & 3zz-1 \end{pmatrix}, \tag{2}$$

with

$$x = \frac{r_{i,x}}{\|\mathbf{r}_i\|}, \quad y = \frac{r_{i,y}}{\|\mathbf{r}_i\|}, \quad z = \frac{r_{i,z}}{\|\mathbf{r}_i\|}. \tag{3}$$

For a dipolar coupling $\mathbf{r}_i$ is the vector connecting the two nuclei, $\alpha = 3$ and the constant c_i is given by

$$c_i = \frac{\mu_0}{4\pi}\gamma_1^i\gamma_2^i\frac{\hbar}{4\pi}, \tag{4}$$

where γ_1^i and γ_2^i are the gyro-magnetic ratios of the two nuclei. Note that $\|\mathbf{r}_i\|$ can be time dependent, so our approach can be applied to any pair of nuclei with a measurable dipolar coupling between them.

Eq. (2) is also valid for axially symmetric chemical-shift anisotropies. In principle chemical-shift anisotropies are not symmetric, but usually the asymmetry is small. To treat chemical-shift anisotropies, we set α to 0, c_i to one half the anisotropy (in frequency units, or in ppm), and let $\mathbf{r}_i$ point along the main axis of the chemical-shift anisotropy tensor. The vector $\mathbf{r}_i$, with arbitrary length, is defined in a molecule-fixed reference frame. It can be defined using two points: the position of the atom and an interaction site. The interaction site can be constructed from the position of the atom and several atoms in its environment, using linear combinations of the positions and cross products of difference vectors, see [8] for details.

The order tensor is symmetric and has trace zero. Using a rotation matrix $\mathbf{T}$ it can be transformed into the following form:

$$\mathbf{T}^{\mathrm{T}}\mathbf{S}\mathbf{T} = s\begin{pmatrix} -\frac{1}{2}(1-\eta) & 0 & 0 \\ 0 & -\frac{1}{2}(1+\eta) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

where $-1 \leqslant s \leqslant 1$ and $0 \leqslant \eta \leqslant 1$. s is called the order parameter and η the asymmetry of the order tensor $\mathbf{S}$. When the molecule tumbles isotropically in the solvent s is zero and no orientational effects can be observed as all δ_i are zero.

2.1. Calculating dipolar couplings in a simulation

From now on we will use the term dipolar couplings, but all formulas also apply to chemical-shift anisotropies. For reasons which are explained below, the **D** matrices are calculated which respect to a reference orientation of the molecule. The orientation is defined by a rotation matrix **R** which is needed to least-squares fit the current coordinates of a selected set of atoms onto a reference conformation. The reference conformation is the starting conformation of the simulation. In case of ensemble averaging, which will be treated later, the structure is taken from the first subsystem. The calculated $\mathbf{D}_i^{\mathrm{c}}$ matrix is given by:

$$\mathbf{D}_i^{\mathrm{c}}(t) = \mathbf{R}(t)\mathbf{D}_i(t)\mathbf{R}^{\mathrm{T}}(t). \tag{6}$$

When orientations between atoms with fixed distance are used, the tensors $\mathbf{D}_i(t)$ are time independent and the time dependence of the couplings arises only from the orientations of the vectors. The calculated dipolar coupling for vector i is given by

$$\delta_i^{\mathrm{c}}(t) = \frac{2}{3}\mathrm{tr}(\mathbf{S}(t)\mathbf{D}_i^{\mathrm{c}}(t)). \tag{7}$$

The order tensor $\mathbf{S}(t)$ is usually unknown. A reasonable choice for the order tensor is the tensor which minimizes the (weighted) mean-square difference between the calculated and the observed dipolar couplings:

$$\left(\sum_{i=1}^{N} w_i\right)^{-1} \sum_{i=1}^{N} w_i(\delta_i^{\mathrm{c}}(t) - \delta_i^{\mathrm{exp}})^2. \tag{8}$$

2.2. Time averaging

Since the tensors $\mathbf{D}_i$ fluctuate rapidly in time, much faster than can be observed in experiment, they should be time averaged in the simulation. However, in a simulation the time as well as the number of copies of a molecule is limited. Usually one cannot obtain a converged average of the $\mathbf{D}_i$ tensors over all orientations of the molecule. If one assumes that the average orientations of the $\mathbf{r}_i$ vectors in a molecule-fixed frame converge much faster than the tumbling time of the molecule, the tensor can be averaged in an axis system which rotates with the molecule, as expressed by Eq. (6). This means that care must be taken when the protocol is applied to highly flexible molecules. The time-averaged tensors are calculated using an exponentially decaying memory function:

$$\mathbf{D}_i^{\mathrm{a}}(t) = \frac{\int_{u=t_0}^{t} \mathbf{D}_i^{\mathrm{c}}(u)\exp\left(-\frac{t-u}{\tau}\right)\mathrm{d}u}{\int_{u=t_0}^{t}\exp\left(-\frac{t-u}{\tau}\right)\mathrm{d}u}. \tag{9}$$

Assuming that the order tensor **S** fluctuates more slowly than the $\mathbf{D}_i$, the time-averaged dipolar couplings can be calculated as

$$\delta_i^{\mathrm{a}}(t) = \frac{2}{3}\mathrm{tr}(\mathbf{S}(t)\mathbf{D}_i^{\mathrm{a}}(t)), \tag{10}$$

where the order tensor $\mathbf{S}(t)$ is calculated using expression (8) with $\delta_i^{\mathrm{c}}(t)$ replaced by $\delta_i^{\mathrm{a}}(t)$.

3. Restraining

The simulated system can be restrained by applying a force proportional to the difference between the calculated and the experimental couplings. When no time averaging is applied a proper potential can be defined as

$$V = \frac{1}{2}k\sum_{i=1}^{N} w_i(\delta_i^{\mathrm{c}}(t) - \delta_i^{\mathrm{exp}})^2. \tag{11}$$

The definition of the order tensor as the tensor that minimizes expression (8) is consistent with this potential as it minimizes V. The optimization of the order tensor does not perform work on the system as it only influences the system via V, which is exactly the quantity which is minimized. The forces are given by minus the gradient of V. The force $\mathbf{f}_i$ acting on vector $\mathbf{r}_i$ is

$$\begin{aligned}\mathbf{f}_i(t) &= -\frac{\mathrm{d}V}{\mathrm{d}\mathbf{r}_i} = -kw_i(\delta_i^{\mathrm{c}}(t) - \delta_i^{\mathrm{exp}})\frac{\mathrm{d}\delta_i(t)}{\mathrm{d}\mathbf{r}_i} \\ &= -kw_i(\delta_i^{\mathrm{c}}(t) - \delta_i^{\mathrm{exp}})\frac{2c_i}{\|\mathbf{r}\|^{2+\alpha}} \\ &\quad\times\left(2\mathbf{R}^{\mathrm{T}}\mathbf{S}\mathbf{R}\mathbf{r}_i - \frac{2+\alpha}{\|\mathbf{r}\|^2}\mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{S}\mathbf{R}\mathbf{r}_i\mathbf{r}_i^{\mathrm{T}})\mathbf{r}_i\right).\end{aligned} \tag{12}$$

3.1. Ensemble averaging

Ensemble averaging can be applied by simulating a system of M subsystems which each contain an identical set of orientation restraints. The systems only interact via the orientation-restraint potential, which is defined as

$$V = M\frac{1}{2}k\sum_{i=1}^{N} w_i\langle\delta_i^{\mathrm{c}}(t) - \delta_i^{\mathrm{exp}}\rangle^2. \tag{13}$$

The force on vector $\mathbf{r}_{i,m}$ in subsystem m is given by

$$\mathbf{f}_{i,m}(t) = -\frac{\mathrm{d}V}{\mathrm{d}\mathbf{r}_{i,m}} = -kw_i\langle\delta_i^{\mathrm{c}}(t) - \delta_i^{\mathrm{exp}}\rangle\frac{\mathrm{d}\delta_{i,m}^{\mathrm{c}}(t)}{\mathrm{d}\mathbf{r}_{i,m}}. \tag{14}$$

3.2. Time averaging

When using time averaging it is not possible to define a proper potential. We can still define a quantity which gives a rough estimate of the energy stored in the restraints:

$$V = M\frac{1}{2}k^{\mathrm{a}}\sum_{i=1}^{N} w_i\langle\delta_i^{\mathrm{a}}(t) - \delta_i^{\exp}\rangle^2. \tag{15}$$

The force constant k^{a} is switched on slowly to compensate for the lack of history at times close to t_0. For consistency it should be chosen exactly proportional to the amount of average which has been accumulated:

$$k^{\mathrm{a}} = k\frac{1}{\tau}\int_{u=t_0}^{t} \exp\Big(-\frac{t-u}{\tau}\Big)\,\mathrm{d}u. \tag{16}$$

What really matters is the definition of the force. It is chosen to be proportional to the square root of the product of the time-averaged and the instantaneous deviation. The use of only the time-averaged deviations induces large oscillations. The force is given by

$$\mathbf{f}_{i,m}(t) = \begin{cases} 0 & \text{for} \quad ab \leqslant 0, \\ k^{\mathrm{a}} w_i \frac{a}{|a|}\sqrt{a\,b}\,\frac{\mathrm{d}\delta_{i,m}^{\mathrm{c}}(t)}{\mathrm{d}\mathbf{r}_{i,m}} & \text{for} \quad ab > 0, \end{cases} \tag{17}$$

$$a = \langle\delta_i^{\mathrm{a}}(t) - \delta_i^{\exp}\rangle,$$

$$b = \langle\delta_i^{\mathrm{c}}(t) - \delta_i^{\exp}\rangle.$$

3.3. Multiple experiments

When different experiments have been performed to measure the same or different types of residual dipolar couplings, restraints from all measurements can be imposed simultaneously on one simulation. This results in a better ensemble than the combination of ensembles from simulations with different sets of restraints. If one can be sure that the order tensor of the molecule is the same in all experiments, the complete set of restraints can be imposed using a single optimized molecular order tensor. However, the partial alignment of the molecules may well be different from one experiment to the next, either because the orienting medium itself is not perfectly stable or because different orienting media were employed. In these cases the molecular order tensors will differ per experiment and restraining all vector orientations with a single molecular order tensor will introduce errors. In our approach the restraints can be imposed using molecular order tensors which are optimized for each experimental dataset individually. This is accomplished by summing several potentials of type (11), (13) or (15) with independently optimized order tensors. The weight factors should be adjusted to reflect the relative accuracy of the experiments.

All types of restraining discussed above, including restraining using multiple measurements with independently optimized order tensors, have been implemented in the GROMACS 3.1 molecular dynamics package [9].

4. Simulations

We performed several simulations of the histidine containing phosphocarrier protein (HPr). The dipolar coupling were determined by van Lune et al. [10]. We used the measurements with a bicelle sample containing PEG2000-PE in 0.69 M KP_i. In this sample the molecular order parameter was negative, i.e., the main axis of orientation was perpendicular to the magnetic field. The simulations were performed with the Gromos 43a2 force field [11]. using the GROMACS 3.1 molecular-dynamics package [9]. As a starting structure for all simulations we used Protein Data Bank entry 1POH [12]. For this structure the order parameter of the optimal order tensor is -6.6×10^{-4} and the asymmetry factor is 0.46. The protein with crystal waters was solvated in a rhombic dodecahedron unit cell with a periodic image distance of 5.5 nm. The total number of SPC water molecules [13] was 3467. All angular degrees of freedom of the protons were removed by using interaction sites [8]. This prevents local adaptation of the proton angles to the orientation restraints and allows the use of a time step of 4 fs. The cut-off for the non-bonded interactions was 1.4 nm. The interactions within 1.0 nm were updated every step, the interactions between 1.0 and 1.4 nm and the neighbor list were updated every five steps. For the electrostatics a reaction field was used with a dielectric constant of 80. The solvated system was subjected to 25 steps of steepest-descents energy minimization. This structure with one set of random velocities was used as the starting point for all simulations.

We performed nine simulations of 7 ns which differ only in force constant for the restraints (0, 1 and 10 kJ mol^{-1} Hz^{-2}) and averaging time (0, 10 and 100 ps). We left the first nanosecond out of the analysis. The results are shown in Table 1. When we look at the non-restrained simulations ($k = 0$) we can see that the root-mean-square difference between calculated and experimental dipolar couplings is significant: around 2 Hz, with a largest measured value of −8.45 Hz. As expected the time averaging decreases the difference between the calculated and the experimental couplings. With time averaging the order parameter s increases as the calculated magnitude of the tensor elements of $D_i^{\mathrm{a}}(t)$ decreases due to the motional averaging. To monitor the effect of the restraints on the fluctuations of the N–H vector orientations we analyzed the fluctuations of the instantaneous couplings. To remove the effect of slow conformational changes we applied a cosine filter with a width of 10 ps. This filter suppresses oscillations with a period of 5 and 10 ps with factor of 0.21 and 0.67, respectively. To obtain numbers which are proportional to the fluctuations of the N–H vectors, the fluctuations of the couplings were divided by the order parameter s of the molecule, which differs between different runs. When restraining is used without time averaging the

Table 1
Results for pdb entry 1POH and nine MD simulations of HPr

	S^a	k^b	τ^c (ps)	s^d (10^{-3})	Dipolar couplings		Backbone
					rmsd[e] (Hz)	Fluctuations/s[f] (10^3 Hz)	rmsd[g] (nm)
pdb	opt.			−0.66	0.92	0.00	0.00
pdb	best			−0.83	1.40	0.00	0.00
MD	opt.	0	0	−0.56	2.25	1.85	0.16
			10	−0.62	2.00	1.88	0.16
			100	−0.64	1.92	1.87	0.16
MD	best	0	0	−0.83	2.93	1.91	0.16
			10	−0.83	2.47	1.91	0.16
			100	−0.83	2.32	1.91	0.16
MD	opt.	1	0	−0.66	1.05	1.24	0.09
			10	−0.73	0.75	1.69	0.11
			100	−0.77	0.60	1.66	0.11
MD	opt.	10	0	−0.70	0.45	0.59	0.08
			10	−0.80	0.19	1.41	0.14
			100	−0.83	0.15	1.45	0.11

[a] opt. indicates the tensor was optimized, best indicates that the tensor of the ensemble and time averaged simulation (see Table 2) was used.
[b] k (kJ mol^{-1} Hz^{-2}) is the force constant for the restraints.
[c] τ is the averaging time for the restraints.
[d] s is the order parameter, averaged over 1–7 ns.
[e] The RMS deviation of the calculated couplings from the experimental couplings, averaged over 1–7 ns.
[f] The RMS high-frequency fluctuation of the calculated couplings, averaged over 1–7 ns.
[g] The RMS deviation from the starting structure of the backbone atoms of HPr, averaged over 1–7 ns.

fluctuations are reduced significantly, while the differences between the experimental and calculated couplings are still significantly higher than the experimental accuracy of 0.1–0.2 Hz. With time-averaged restraining one can reach the experimental accuracy (see Fig. 1) with only a slight reduction of the fluctuations.

We also calculated the couplings for the average structure of the MD simulation. This structure was calculated by averaging all structures between 1 and 7 ns after superpositioning with the same least-squares fitting procedure as used for the molecular order tensor. Since the N–H vectors shorten due to the averaging we used their standard lengths in the calculation. As can be seen in Fig. 1 the deviations for the average structure, like those calculated for the X-ray structure, are much larger than for the time-averaged simulation.

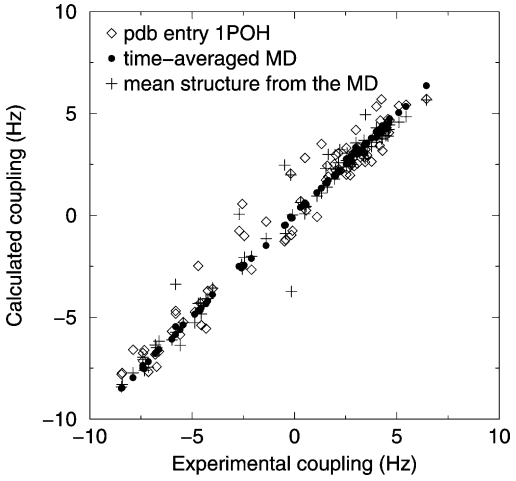

Fig. 1. The calculated residual dipolar couplings for 79 N–H vectors of HPr plotted against the experimentally measured couplings. Shown are the couplings for pdb entry 1POH with backbone amide hydrogens placed in the amide plane in the minimum of the angle potential, the average couplings in the MD simulation with force constant 10 kJ mol^{-1} Hz^{-2} and averaging time 100 ps and the couplings for the mean N–H orientations in the same MD simulation.

4.1. Ensemble averaging

We performed three ensemble-averaged simulations. The ensemble consists of four copies of the system with water which is described above. The four subsystems only interact via the orientation-restraint potential. The starting conformations are the same as for the single-protein simulations. The only difference in initial conditions between the four copies in the ensemble are the initial, randomly chosen, velocities. We performed one unrestrained simulation with time averaging and two restrained simulations, one without and one with time averaging. All simulations had a length of 4 ns. The first nanosecond was left out of the analysis. The results are

Table 2
Results for three simulations of an ensemble of four HPr molecules[a]

k[b]	τ[c] (ps)	Time and ensemble average		Instantaneous and local		
			Dipolar couplings	Dipolar couplings		Backbone
		s[d] (10^{-3})	rmsd[e] (Hz)	rmsd[e] (Hz)	Fluctuations/s[f] (10^3 Hz)	rmsd[g] (nm)
0	100	−0.71	1.83	2.33	1.87	0.11
				2.45	1.93	0.14
				2.28	1.83	0.11
				2.50	1.90	0.14
10	0	−0.78	0.24	1.70	1.57	0.12
				1.45	1.56	0.11
				1.52	1.60	0.11
				1.45	1.57	0.11
10	100	−0.83	0.10	1.58	1.73	0.12
				1.75	1.85	0.12
				1.82	1.85	0.12
				1.61	1.79	0.12

[a] Each of the four proteins in one ensemble is simulated in a separate box of solvent. The restraints are ensemble averaged.
[b] k (kJ mol^{-1} Hz^{-2}) is the force constant for the restraints.
[c] τ is the averaging time for the restraints.
[d] s is the order parameter, averaged over 1–4 ns.
[e] The RMS deviation of the calculated couplings from the experimental couplings, averaged over 1–4 ns.
[f] The RMS high-frequency fluctuation of the calculated couplings, averaged over 1–4 ns.
[g] The RMS deviation from the starting structure of the backbone atoms of HPr, averaged over 1–4 ns.

shown in Table 2. As one would expect the ensemble averaging brings the calculated couplings closer to the experimental values. In the simulation with ensemble averaging and without time-averaged restraining the fluctuations in the couplings are a factor 2.7 larger than in the single-protein simulation. The fluctuations are only 16% smaller than in the unrestrained ensemble. When time averaging is used the fluctuations are almost unaffected. The deviations of instantaneously calculated couplings from the experimental ones are also reported in Table 2 for each structure. These couplings were recalculated with an instantaneously optimized molecular order tensor. These values would be obtained if one would consider the individual structures, neglecting dynamical and ensemble averaging. Although instantaneous restraint violations in the restrained runs are smaller than in the unrestrained runs, the values are still an order of magnitude larger than in the time- and ensemble-averaged run, which are of the order of the experimental error. Clearly, from the restraint violations calculated for an individual structure one can conclude little about the quality of this structure.

4.2. Quality of the ensembles

One feature that can be observed in all the presented simulations is that the molecular order parameter s decreases as the deviation from experiment increases. This is caused by the optimization of the order tensor S. When the orientations of the N–H vectors are not correct the eigenvalues of the optimized S will be smaller than those of the true S. For a set of randomly oriented N–H vectors s would be almost zero. This means that the actual deviation of the residual dipolar couplings is slightly larger than reported in Tables 1 and 2. To obtain more accurate values one should recalculate the couplings for all simulations with the average S tensor of the simulation with the smallest deviations (as was done for the X-ray structure and for the unrestrained simulations reported in Table 1).

It is difficult to assess the quality of an MD ensemble, because there are not many experimental data. A check that one can always apply is cross validation: leaving out a part of the restraints and checking how well the simulations reproduce these missing experimental data. One can also compare results from restrained MD simulations with results from unrestrained MD simulations. However one cannot use structures from MD simulations as a reference, as the current force fields are not accurate enough to produce a reliable ensemble. One property of MD simulations that is reliable is the fluctuation in the picosecond range. The use of restraints during a simulation should not affect these fluctuations significantly.

A graph of the instantaneous coupling for the N–H vector of residue 10 is shown in Fig. 2. There are six curves, with and without time and ensemble averaging. The instantaneous couplings for simulations with averaging were calculated with the time- and/or ensemble-averaged molecular order tensor. The only difference in the two unrestrained simulations shown in Fig. 2 is in the calculation of the molecular order tensor. Instantaneous restraining clearly reduces the fluctuations.

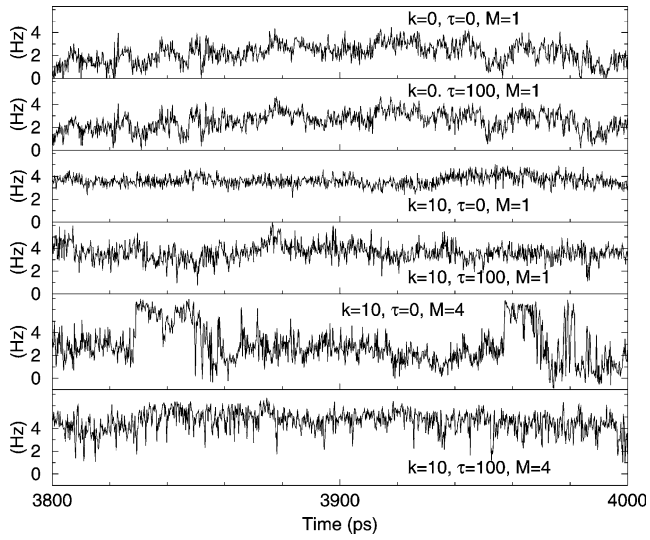

Fig. 2. The calculated instantaneous residual dipolar coupling for the N–H vector of residue 10 of HPr. There are six curves for six simulations with and without time and ensemble averaging. Where applicable the instantaneous coupling was calculated with a time- and/or ensemble-averaged molecular order tensor. k (kJ mol^{-1} Hz^{-2}) is the force constant for the restraints, τ (ps) is the averaging time, and M is the number of HPr molecules in the ensemble.

Time-averaged restraining reduces the low-frequency fluctuations, but preserves the high-frequency fluctuations. Ensemble averaging also preserves the high-frequency fluctuations, both with and without time averaging, but it introduces several spikes on the curve, which are not present in the unrestrained runs. Ensemble averaging can also cause jumps in the dipolar couplings, as is visible in the curve of the instantaneously restrained ensemble at 3830 and 3960 ps. With time averaging such jumps are smaller and less frequent.

To test if restraining improves the average direction of the N–H vectors in ensembles generated by MD, we performed several simulations where we left out 8 of the 79 restraints. We left out every 10th restraint starting at restraint 2, 4, 6, or 8. This resulted in four sets of 71 restraints. We performed four MD simulations with each set of restraints and four MD simulations without restraints, with four different sets of starting velocities. For comparison we analyzed the crystal structure. To calculate the couplings backbone amide hydrogens were placed in the amide plane in the minimum of the angle potential. We also performed energy minimization of the crystal structure in vacuum by subjecting it to 1000 steps of steepest-descent energy minimization with each of the four sets of restraints. Just like in the MD simulations all hydrogens were treated as interaction sites. This means that the N–H vectors can only be reoriented by reorienting the complete peptide plane. To remove the influence of the molecular order tensor all calculations were performed with the order tensor of the time- and ensemble-averaged simulation. The deviations of the couplings are shown in Table 3. The results show

Table 3
Cross validation[a]

	k[b]	τ[c] (ps)		Dipolar couplings rmsd		Backbone
				Restraints[d] (Hz)	Not restraints[e] (Hz)	rmsd[f] (nm)
pdb[g]				1.40	1.54	0.00
EM[h]	10			0.16	1.77	0.04
MD[i]	0	100	1–3 ns	2.17	2.11	0.14
			3–5 ns	2.34	1.94	0.16
			5–7 ns	2.42	2.46	0.15
MD[i]	1	100	1–3 ns	0.67	1.67	0.12
			3–5 ns	0.65	1.81	0.12
			5–7 ns	0.65	1.92	0.12
MD[i]	10	100	1–3 ns	0.14	1.62	0.12
			3–5 ns	0.14	1.65	0.13
			5–7 ns	0.15	1.97	0.14

[a] Results for simulations in which 71 of the 79 N–H vectors for which dipolar couplings were measured were restrained. All calculations were performed with the order tensor of the time- and ensemble-averaged simulation (see Table 2). All numbers are averages of four simulations or structures.

[b] k (kJ mol^{-1} Hz^{-2}) is the force constant for the restraints.

[c] τ is the averaging time for the restraints.

[d] The RMS deviation of the calculated couplings from the experimental couplings for 71 restrained vectors.

[e] The RMS deviation of the calculated couplings from the experimental couplings for eight unrestrained vectors.

[f] The RMS deviation from the starting structure of the backbone atoms of HPr.

[g] pdb is structure from pdb entry 1POH.

[h] EM are four structures after energy minimizing pdb entry 1POH.

[i] The MD rows are results from different parts of MD simulations averaged over four simulations.

that there is a coupling between the different N–H vectors. When performing an energy minimization the restrained N–H vectors are optimized at the expense of the unrestrained N–H vectors. Interestingly the MD simulations show the opposite effect. In the restrained simulation the unrestrained N–H vectors show smaller deviations than in the unrestrained simulations. This is a significant finding, as in systems with many more degrees of freedom than restraints, the optimization of the restrained degrees of freedom typically comes with increased violations in unrestrained degrees of freedom, as was the case for the restrained energy minimization of the crystal structure. This shows that time-averaged restraining improves the quality of the MD ensemble. However, the deviations of the free N–H vectors in the restrained MD ensemble are still slightly larger than in the crystal structure. There are two explanations for this, lack of sampling and the quality of the forcefield. A simulation of 7 ns is certainly too short to sample a reasonable amount of the available phase space, but it is unclear to what extent this lack of sampling will affect the dipolar couplings. The same argument applies to the force field. The force field certainly is not perfect, but how this would affect the sampling is unclear. To further refine the ensemble and to perform a better quality check more experimental data is required, such as dipolar couplings for other atom pairs and NOEs.

5. Conclusions

We have presented methodology for applying weighted time- and ensemble-averaged orientation restraints in MD simulations. For the case without time averaging we have defined a restraining potential which is equivalent to the quantity used for optimization of the order tensor of the molecule. This is a requirement for energy conservation. For time-averaged restraints we have defined a restraining force which ensures the correct average orientation without significantly affecting the fluctuations in the orientation. Note that time averaging precludes the possibility to define a proper potential.

An important conclusion is that the order tensor strongly depends on the ensemble of structures. At room temperature the orientations of the N–H vectors fluctuate significantly. When refining a single structure or when simulating without time and ensemble averaging, the disorder of each N–H vector is transfered to the overall molecular order tensor. In the different simulations the magnitude of this effect was on average 15%. But more importantly, the order can vary significantly between the different N–H vectors. Ignoring this can lead to a misinterpretation of the vector orientations, because the N–H vectors in less ordered regions will be biased towards the magic angle. Moreover, the effect is not proportional to the coupling: angles close to 0° and 90° will be affected more (see Fig. 3). Our present approach alleviates this problem: it allows the use of all orientation restraints and avoids the need to pre-identify less ordered regions. Time averaging captures fluctuations up to time scales of the averaging time (up to 100 ps in this work). Ensemble averaging allows the inclusion of disorder on even longer time scales.

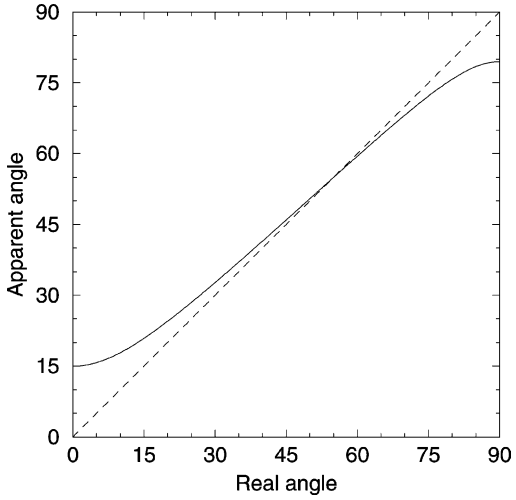

Fig. 3. The angle apparent from a dipolar coupling plotted against the real angle when a disorder characterised by $S = 0.9$ has scaled down the dipolar coupling. The value of 0.9 can be seen as the ratio of disorder between flexible and more rigid parts of a protein.

Our present approach allows the direct use of the experimental values as target values during refinement. The force constants can be chosen such that the violations of the restraints in the final model ensemble reflect the true experimental uncertainties (see Fig. 1). This is in contrast to current restraining protocols, where the user is tempted to leave out data from mobile regions of the molecule, or to use smaller force constants for such regions.

The quality of the generated ensembles improved slightly by restraining 71 of the 79 N–H vectors for which residual dipolar couplings were measured. For a significant improvement in the quality of the ensemble more experimental data are required such as C_α–H_α, C_α–C' and C_α–C_β residual dipolar couplings or NOEs. This will be reported in future papers.

References

[1] P. Bayer, L. Varani, G. Varani, Refinement of the structure of protein–RNA complexes by residual dipolar coupling analysis, J. Biomol. NMR 14 (1999) 149–155.
[2] G.M. Clore, D.S. Garrett, *r*-factor, free *r*, and complete cross-validation for dipolar coupling refinement of NMR structures, J. Am. Chem. Soc. 121 (1999) 9008–9012.
[3] H. Schwalbe, S.B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C.M. Dobson, C. Redfield, L.J. Smith, A refined structure of hen

lysozyme determined using residual dipolar coupling data, Protein Sci. 10 (2001) 677–688.
[4] G. Colombo, D. Roccatano, A.E. Mark, Folding and stability of the three-stranded β-sheet peptide betanova: insights from molecular dynamics, Proteins 46 (2002) 380–392.
[5] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins, J. Am. Chem. Soc. 123 (2001) 1416–1424.
[6] W. Peti, J. Meiler, R. Brüschwieler, C. Griesinger, Model-free analysis of protein backbone motion from residual dipolar couplings, J. Am. Chem. Soc. 124 (2002) 5822–5833.
[7] J.R. Tolman, A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy, J. Am. Chem. Soc. 124 (2002) 12020–12030.
[8] A.K. Feenstra, B. Hess, H.J.C. Berendsen, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems, J. Comput. Chem. 20 (8) (1999) 786–798.
[9] E. Lindahl, B. Hess, D. van der Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Mod. 7 (2001) 306–317.
[10] F. van Lune, L. Manning, K. Dijkstra, H.J.C. Berendsen, R.M. Scheek, Order-parameter tensor description of HPr in a medium of oriented bicelles, J. Biomol. NMR 23 (2002) 169–179.
[11] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hünenberger, P. Krüger, A.E. Mark, W.R.P. Scott, I.G. Tironi, Biomolecular Simulation: GROMOS96 Manual and User Guide, BIOMOS b.v., Zürich, Groningen, 1996.
[12] Z. Jia, J.W. Quail, E.B. Waygood, L.T.J. Delbaere, The 2.0 Å-resolution structure of *Escherichia coli* histidine-containing phosphocarrier protein HPr. A redetermination, J. Biol. Chem. 268 (1993) 22490–22501.
[13] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), Intermolecular Forces, Reidel, Dordrecht, 1981, pp. 331–342.