Biomolecular structure refinement using the GROMOS simulation software

J Biomol NMR (2011) 51:265–281
DOI 10.1007/s10858-011-9534-0

ARTICLE

Biomolecular structure refinement using the GROMOS simulation software

Nathan Schmid · Jane R. Allison · Jožica Dolenc · Andreas P. Eichenberger · Anna-Pitschna E. Kunz · Wilfred F. van Gunsteren

Received: 22 March 2011 / Accepted: 14 July 2011 / Published online: 20 August 2011

Abstract For the understanding of cellular processes the molecular structure of biomolecules has to be accurately determined. Initial models can be significantly improved by structure refinement techniques. Here, we present the refinement methods and analysis techniques implemented in the GROMOS software for biomolecular simulation. The methodology and some implementation details of the computation of NMR NOE data, ^3J-couplings and residual dipolar couplings, X-ray scattering intensities from crystals and solutions and neutron scattering intensities used in GROMOS is described and refinement strategies and concepts are discussed using example applications. The GROMOS software allows structure refinement combining different types of experimental data with different types of restraining functions, while using a variety of methods to enhance conformational searching and sampling and the thermodynamically calibrated GROMOS force field for biomolecular simulation.

Keywords GROMOS · Structure refinement · Crystallography · NMR

Introduction

The accurate determination of the structure of biomolecules such as proteins, nucleic acids and membranes is crucial for the understanding of their biological function. However, structural information is not easily obtained with a high degree of accuracy for any particular biomolecule. X-ray, electron or neutron diffraction techniques can be used to produce models at the atomic level of biomolecules in the solid state. On the other hand, spectroscopic measurements such as NMR, CD, IR, Raman and fluorescence spectroscopy can be used to obtain, albeit less detailed, structural information of molecules in solution, i.e. under more physiological conditions. Experimentally, the positions of the atoms cannot be directly observed and some other quantity such as X-ray scattering intensities or nuclear Overhauser enhancements are determined. Subsequently, measured values for these observables are used to derive an (atomic) model. Experimental quantities are then computed from the model structure and used to judge the quality of it.

In the process of structure determination four major challenges are faced:

1. A quantity $Q(\mathbf{r}^N, \mathbf{p}^N)$ is measured. In classical mechanics this quantity depends on the Cartesian (or internal, generalised) coordinates of the N atoms $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ and their momenta $\mathbf{p}^N = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)$. However, for most of the quantities of interest in structure determination the quantity Q only depends on the coordinates $\mathbf{r}^N$. The measurement has a limited accuracy and in structure determination and refinement this should be accounted for. In general, the number of experimental observables is much smaller than the number of degrees of freedom of the system. This makes the problem of structure determination and refinement highly underdetermined. In structure determination of proteins by X-ray crystallography or NMR spectroscopy, for example, high resolution structural

N. Schmid · J. R. Allison · J. Dolenc · A. P. Eichenberger · A.-P. E. Kunz · W. F. van Gunsteren (✉)
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology ETH, 8093 Zürich, Switzerland
e-mail: wfvgn@igc.phys.chem.ethz.ch

J. Dolenc
Faculty of Chemistry and Chemical Technology, University of Ljubljana, 1000 Ljubljana, Slovenia

parameters of the amino acids such as bond-lengths and -angles are incorporated to address this problem.

2. Generally, a measurement cannot probe the instantaneous value of a quantity $Q(\mathbf{r}^N)$ but only an average $\langle Q \rangle$ over multiple molecules and time which is denoted as the experimental value Q^0. From statistical mechanics it is known that the averaging is carried out using Boltzmann weighting of the different conformers. This has the implication that the averaging is irreversible, i.e. the probability distribution of $Q(\mathbf{r}^N)$ and the conformational probability distribution $P(\mathbf{r}^N)$ are inaccessible from the average $\langle Q \rangle$. In the case of a protein in a crystal environment, where the X-ray scattering data was collected at low temperature, a single structure may serve as a rather good approximation of the conformational ensemble. In solution, however, this is not the case. Multiple different conformations contribute to the average $\langle Q \rangle$. A single conformer derived from such an average can have a negligible Boltzmann weight and thus constitute a very bad approximation of the conformational ensemble.
3. The dependence of Q on $\mathbf{r}^N$ has to be known. The functional form can either be physically correct or an approximation and in the latter case the accuracy of the approximation plays an important role. For example, the accurate computation of chemical shifts involves costly, high-level quantum-chemical methodology, whereas the methodology to compute X-ray diffraction intensities is well known and not too expensive to execute. The function $Q(\mathbf{r}^N)$ has to be invertible in order to derive the atomic positions $\mathbf{r}^N$ from Q. Often, as in the example of the chemical shifts, this condition is not met.
4. If the function $\mathbf{r}^N(Q)$, the inverse of $Q(\mathbf{r}^N)$, exists, it can be multiple-valued. For example in NMR measurements, the same ^3J-coupling is related to several different dihedral-angle conformers through the Karplus relation. Sampling of conformations has to be biased towards the multiple solutions of $\mathbf{r}^N(Q)$. In the case of a ^3J-coupling, for example, it is not possible to judge the quality of the different solutions from the experimental value alone. The measured value can either be an average of multiple conformers or come from only a few individual conformations. Additional modelling is required to judge the quality of the multiple solutions.

Quantities and averaging

Using the GROMOS software a variety of quantities can be used in structure refinement. Many of these quantities can be used to bias sampling in an MD simulation towards the experimentally measured values Q^0. Other quantities can only be calculated from the ensemble generated and can be used to judge the final model or interpret simulation results. Over the years theories and methods for the computation of such quantities have been developed. The ones listed below were used in GROMOS:

NOE intensities	$\langle I^{NOE} \rangle$
NOE distances	$\langle r^{-p} \rangle$ with $p = 3$ or 6
^3J-coupling constants	$\langle {}^3J \rangle$
Residual dipolar couplings	$\langle D \rangle$
Crystallographic structure-factor amplitudes	$\langle \mid F \mid \rangle$
Crystallographic electron density	$\langle \rho \rangle$
Neutron scattering intensities	$\langle I^{neu} \rangle$
X-ray scatting intensities	$\langle I^{xray} \rangle$

Experimental data can be classified as two kinds: observable, measured, primary data and non-observable, derived, secondary data. For example an NOE intensity can be measured directly, whereas an NOE distance bound is derived from the actual intensity. In general, the more assumptions flow into the determination of the secondary, derived data, the more problematic is the use of such data in structure determination or refinement.

Experiments generally probe the ensemble average of a quantity, thus the quantity is averaged according to Boltzmann's distribution,

$$\langle Q(\mathbf{r}^N) \rangle_{\mathbf{r}^N} = \frac{\int Q(\mathbf{r}^N) e^{-V(\mathbf{r}^N)/k_B T} \mathrm{d}\mathbf{r}^N}{\int e^{-V(\mathbf{r}^N)/k_B T} \mathrm{d}\mathbf{r}^N}, \quad (1)$$

where $V(\mathbf{r}^N)$ is the energy of conformation $\mathbf{r}^N$. Methods used to sample conformations have to reproduce such an ensemble. As molecular dynamics simulations do so, time averaging (Torda et al. 1989) can be applied using the equation

$$\langle Q(\mathbf{r}^N) \rangle_t = \frac{1}{\tau_Q (1 - e^{-t/\tau_Q})} \int_0^t e^{-(t-t')/\tau_Q} Q(\mathbf{r}^N(t')) \mathrm{d}t', \quad (2)$$

where τ_Q is known as the time-averaging memory-relaxation time of the quantity Q.

Biasing potential energy terms

Because the accuracy of the physical force field V^{phys} is limited, biasing potential energy terms V^Q are added to drive the simulation towards the experimental data,

$$V(\mathbf{r}^N) = V^{\mathrm{phys}}(\mathbf{r}^N) + V^Q(\mathbf{r}^N; Q^0). \quad (3)$$

The biasing (or restraining) potential energy function V^Q can have various functional forms. Generally, a harmonic potential energy function is chosen for simplicity,

$$V^{Q}(\mathbf{r}^N;Q^0)=\sum_{i=1}^{N_Q}\frac{1}{2}K_i^{Q}\left(Q_i(\mathbf{r}^N)-Q_i^0\right)^2, \tag{4}$$

where the sum is carried out over the N_Q experimental observables Q_i^0 taken into account, and K_i^Q is a force constant which can be weighted according to the accuracy or importance of an individual value Q_i^0. This functional form can be easily adapted to incorporate averaging by replacing $Q(\mathbf{r}^N)$ by its time-average $\langle Q(\mathbf{r}^N)\rangle_t$ (Torda et al. 1989) resulting in

$$V^{Q}(\mathbf{r}^N;Q^0)=\sum_{i=1}^{N_Q}\frac{1}{2}K_i^{Q}\left(\langle Q_i(\mathbf{r}^N)\rangle_t-Q_i^0\right)^2. \tag{5}$$

There are quantities for which a long memory-relaxation time is required in order to get meaningfully averaged quantities. Thus a penalty function based on such an average will only react very slowly to instantaneous changes of the quantities Q_i. For this reason, a combined instantaneous and time-averaged, biquadratic functional form can be used

$$V^{Q}(\mathbf{r}^N;Q^0)=\sum_{i=1}^{N_Q}\frac{1}{2}K_i^{Q}\left(Q_i(\mathbf{r}^N)-Q_i^0\right)^2\left(\langle Q_i(\mathbf{r}^N)\rangle_t-Q_i^0\right)^2, \tag{6}$$

which depends on both the instantaneous quantity $Q(\mathbf{r}^N)$ and the time-averaged quantity $\langle Q(\mathbf{r}^N)\rangle_t$. The energy will be low if either the instantaneous or the averaged quantity or both are in agreement with the experimental value. In the former case, it can be avoided that the slowly changing average is driving the system away from conformations that are in agreement with the experimental data. If the average is in agreement with Q^0, the energy is also low and additional conformations can be explored.

Sampling and its enhancement

Because V^Q is generally a sum over many observables which have a complicated dependence on the conformation, V^Q is a complex and rugged potential energy hypersurface which contains many local minima separated by high barriers. In order to overcome these barriers, sampling enhancement techniques (Christen and van Gunsteren 2007) have to be applied.

Simulated annealing In the simulated annealing approach (Kirkpatrick et al. 1983; Černy 1985), the temperature is raised to a high value, e.g. 5,000 K, so that the thermal energy $k_B\ T$ is large enough to overcome most barriers present on the V^Q surface. The system is then cooled down as slowly as possible to the experimental temperature exploring the multiple valleys of the potential energy surface. This search method is applicable for all kinds of force fields, restraining methods and different functional forms.

Temperature replica exchange The principle of temperature replica exchange (Irbäck and Potthast 1995; Hukushima and Nemoto 1996) is similar to that of simulated annealing: sampling is enhanced using multiple parallel simulations of replicas of the system, some running at an elevated temperature. After a fixed number of MD integration steps, a Monte Carlo exchange between two replicas i and j is attempted. By using the detailed balance condition

$$p_{ij}=\begin{cases}1 & \text{for } \Delta\le 0\\ \exp(-\Delta) & \text{for } \Delta>0\end{cases} \tag{7}$$

with

$$\Delta=(\beta_i-\beta_j)(V(\mathbf{r}_j^N)-V(\mathbf{r}_i^N)) \tag{8}$$

as the switching probability, a canonical, *NVT*, ensemble for every replica is obtained. Here $\beta_i=(k_BT_i)^{-1}$. The method can also be generalised to generate other ensembles (Okabe et al. 2001). Because the switching probability is decreasing with increasing system size, many replicas are required for large systems using explicit solvent, which makes the method rather inefficient in the latter cases.

Hamiltonian replica exchange As in temperature replica exchange, multiple simulations of the same system are used, but instead of being at different temperatures, the replicas differ in their interactions, thus each has a different Hamiltonian $H(\mathbf{r}^N,\mathbf{p}^N)$ (Sugita et al. 2000). A coupling parameter λ is introduced into H such that $\lambda=0$ corresponds to state A and $\lambda=1$ to state B. For example, state A can be a fully restrained system and state B a fully unrestrained system. The replicas are sampling using Hamiltonians at different values of λ between 0 and 1. In order to obtain a canonical ensemble for every replica, the switching probability given in (7) is used with Δ given by

$$\Delta=\beta\left[V_i(\mathbf{r}_j^N)-V_i(\mathbf{r}_i^N)\right]-\beta\left[V_j(\mathbf{r}_j^N)-V_j(\mathbf{r}_i^N)\right] \tag{9}$$

with V_i being the potential energy function at λ_i.

Local elevation The local elevation (LE) method (Huber et al. 1994) is a more specific way to overcome barriers along a defined set of reduced coordinates $q_j(\mathbf{r}^N)$. Biasing potential energy terms

$$V^{\mathrm{LE}}(\mathbf{r}^N)=K^{\mathrm{LE}}\sum_{i=1}^{N_G}w_i(t)\prod_{j=1}^{N_q}g_j\left(q_j(\mathbf{r}^N);q_{ij}^0\right) \tag{10}$$

are added to $V^{\mathrm{phys}}(\mathbf{r}^N)$, with K^{LE} being a force constant or weight factor, N_G the number of grid points along the selected degrees of freedom, $w_i(t)$ a time-dependent weight function, N_q the dimensionality of the set of coordinates chosen and g_j a local, repulsive function such as a Gaussian

or a truncated polynomial. In the ordinary local elevation method the weight function $w_i(t)$ is chosen as

$$w_i(t) = w_i(t - \Delta t) + 1. \tag{11}$$

The weight is increased by one every time a grid point is visited. In this way all configurations visited are uniformly penalised. For structure refinement this is not very useful. The weight function has be chosen such that configurations that are in agreement with the experimental data are not penalised and the weight increase is largest for the conformations having the worst agreement. This can be achieved by increasing the weight function by the energy of a restraining function leading to

$$w_i(t) = w_i(t - \Delta t) + V^Q(\mathbf{r}^N; Q^0), \tag{12}$$

which makes K^{LE} a unitless weight factor. Here, the function $V^Q(\mathbf{r}^N;Q^0)$ should be a flat-bottom function, i.e. zero for small deviations from the experimental data (Christen et al. 2007).

The new GROMOS simulation software, which is written in an object-oriented manner in C++, allows for the use of the mentioned searching and sampling enhancement techniques in structure refinement based on spectroscopic (NMR) or diffraction (X-ray) data. In this article their implementation is summarised and examples of their use are presented.

Methods

Nuclear magnetic resonance spectroscopy

Using the GROMOS software Nuclear Magnetic Resonance (NMR) data such as NOE intensities or bounds, ^3J-couplings and residual dipolar couplings (RDCs) can be calculated from the simulation trajectories. In addition, biasing potential energy functions V^Q based on NOE bounds and ^3J-couplings are available.

NOE intensities The GROMOS software can be used to compute NOESY or ROESY intensities I^{NOE} from a trajectory (Peter et al. 2001). For every inter-proton pair of interest time correlation functions

$$C(\tau) = \left\langle \frac{P_2\left(\cos\left(\chi_{t,t+\tau}\right)\right)}{r^3(t)r^3(t+\tau)} \right\rangle \tag{13}$$

are calculated, where P_2 is the second-order Legendre polynomial, $\chi_{t,t+\tau}$ is the angle between the inter-spin vector at the two time points t and $t + \tau$, and r is the length of the inter-spin vector. The corresponding spectral densities $J(\omega)$ are computed by Fourier transformation of $C(\tau)$. The intensities of a multi-spin system can then be derived by solving the following equation

$$\frac{d\underline{M}(t)}{dt} + \underline{R}\underline{M}(t) = \underline{0}. \tag{14}$$

Here, $\underline{M}(t)$ is the matrix containing the intensities as diagonal elements and $\underline{R}$ the relaxation matrix containing the direct, dipolar relaxation rates as diagonal elements and the cross-relaxation rates as off-diagonal elements. The latter two can be computed from the spectral densities. This matrix equation is solved by diagonalisation of $\underline{R}$, leading to the determination of $\underline{M}(t_m)$ which contains the NOE intensities at mixing time t_m.

This approach is computationally very demanding, because time correlation functions and spectral densities have to be computed for all proton pairs and a long simulation time is required to obtain a converged spectrum. For the latter reasons it is not possible to use this approach for structure refinement using the GROMOS software.

NOE distance bounds NOE distances can be computed from a trajectory by evaluating

$$\overline{r_i} = \langle r_i^{-p} \rangle^{-1/p}, \tag{15}$$

where r_i is the length of the inter-nuclear vector corresponding to an NOE signal. The constant p is chosen as 3 when averaging over a relatively short time, when the angular correlation in r_i should be neglected, and as 6 when this is not the case (Tropp 1980; van Gunsteren et al. 1996). Positions of protons that are not treated explicitly by the force field are calculated from standard configurations (van Gunsteren et al. 1996). In cases where the NOE distance bound corresponds to more than one proton, a pseudo-atom approach (Wüthrich et al. 1983) is applied and the NOE distance bound is corrected (Oostenbrink et al. 2005) accordingly. Subtraction of the NOE distance bound distance from $\overline{r}_i$ leads to the NOE violation.

In GROMOS an MD simulation can be biased towards the experimental NOE distance bound r^0 using the distance restraining machinery. For each of the N_{NOE} NOE distance upper bounds r_i^0, a half-harmonic attractive potential energy term (Zuiderweg et al. 1985) is added

$$V^{\mathrm{NOE}}(\mathbf{r}^N; r^0) = \sum_{i=1}^{N_{NOE}} \begin{cases} 0 & \text{for } r_i(\mathbf{r}^N) \le r_i^0 \\ \frac{1}{2}K_i^{\mathrm{NOE}}\left(r_i(\mathbf{r}^N) - r_i^0\right)^2 & \text{for } r_i(\mathbf{r}^N) > r_i^0, \end{cases} \tag{16}$$

where K_i^{NOE} is a force constant and r_i is the length of the inter-nuclear vector for this NOE distance bound.

In order to observe an NOE signal it is not necessary that the distance upper bound is fulfilled at all times in the MD simulation, thus appropriate averaging should be used in (16). The instantaneous distance r_i is replaced by a time-averaged distance $\langle r_i^{-p} \rangle_t^{-1/p}$ (Torda et al. 1989). As the memory relaxation time in the average is generally chosen to be short, p is taken as 3. Because of the minus third

power, the average is dominated by short distances, allowing for more movement and fluctuations than when using (16) with the instantaneous distances.

In the same way as described above, a half-harmonic repulsive term, an NOE lower bound, can be added (de Vlieg et al. 1986) representing non-observed NOE signals. However, this is not recommended as a large distance average $\overline{r_{ij}}$ of two nuclei is only one of the possible reasons for not observing a signal.

^{3} J-Couplings The ^{3} *J*-coupling constant $^3J_{m,m'}$ depends on the value of the dihedral angle $\varphi(m-j-k-m')$ involving the three covalent bonds connecting the atoms m and m' through atoms j and k according to the Karplus relation (Karplus 1959)

$$^3J_{m,m'} = A\cos^2\left(\varphi_{m,m'}\right) + B\cos\left(\varphi_{m,m'}\right) + C, \tag{17}$$

where the coefficients A, B and C depend on the types of the atoms j, k, m and m' (Pardi et al. 1984). The average over the trajectory of an MD simulation can be compared to the measured ^3J-coupling constant.

^3J-couplings can also be used for structure refinement. The ^3J-values are denoted as J for ease of notation. In this case N_J harmonic potential energy terms

$$V^J(\mathbf{r}^N; J^0) = \sum_{i=1}^{N_J} \frac{1}{2} K_i^J \left(J_i(\mathbf{r}^N) - J_i^0\right)^2 \tag{18}$$

are added. It is also possible to use half-harmonic attractive or repulsive functions similar to (16). Time-averaging is introduced by replacing J_i by its time-average $\langle J_i \rangle_t$ in (18). A further option with time-averaging is to use a biquadratic penalty function

$$V(\mathbf{r}^N; J^0) = \sum_{i=1}^{N_J} \frac{1}{2} K_i^J \left(J_i(\mathbf{r}^N) - J_i^0\right)^2 \left(\langle J_i(\mathbf{r}^N)\rangle_t - J_i^0\right)^2. \tag{19}$$

This functional form avoids the large structural fluctuations that occur when standard time-averaging is used with ^3J-value restraints (Torda et al. 1993; Nanzer et al. 1997). The Karplus curve has multiple minima and in order to overcome the barriers between them sampling has to be enhanced. This can be achieved by using the local elevation method. For every ^3J-coupling constant i of interest a local elevation term (10) is added. The dihedral angle that determines the ^3J-coupling is chosen as the local-elevation variable. The weight function is given by

$$w_i(t) = w_i(t-\Delta t) + \begin{cases} \frac{1}{2}K_i\left(J_i - J_i^0 - \Delta J_i^0\right)^2 & \text{for} \quad J_i > J_i^0 + \Delta J_i^0 \\ \frac{1}{2}K_i\left(J_i - J_i^0 + \Delta J_i^0\right)^2 & \text{for} \quad J_i < J_i^0 - \Delta J_i^0 \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

where ΔJ_i^0 is a tolerance threshold accounting for inaccuracy in the J_i^0 value and in the Karplus relation. $\Delta J_i^0 > 0$ gives a flat-bottom weight function. The local, repulsive functions are only built up if the ^3J-value is not in agreement with the experimental value. Similar weight functions using time-averaging and in analogy to (19) are implemented and the latter should be used in practical applications (Christen et al. 2007).

Residual dipolar couplings The residual dipolar coupling D_{ij} between two spins i and j can be determined if the molecules in the sample are partially aligned with respect to a static magnetic field, leading to incomplete averaging of the dipolar coupling. Residual dipolar couplings are averages over all inter-spin orientations of the dipolar interaction

$$D_{ij} = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi^3}\left\langle \frac{P_2\left(\cos\left(\theta_{ij}\right)\right)}{r_{ij}^3}\right\rangle. \tag{21}$$

Here, γ_i and γ_j are the gyromagnetic ratios of the two spins, μ_0 is the magnetic permittivity of vacuum and h is Planck's constant. P_2 denotes the second-order Legendre polynomial and θ_{ij} is the angle between the inter-nuclear vector $\mathbf{r}_{ij}$ and the static magnetic field. In practice, (21) is generally not evaluated directly because the angle θ_{ij} is not known. Assuming a rigid molecule, it is possible to reformulate (21) using an alignment tensor $\underline{A}$ (Blackledge 2005), leading to

$$D_{ij} = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi^3 r_{ij}^3}\sum_{k\in x,y,z}\sum_{l\in x,y,z}\underline{A}_{kl}\cos(\zeta_k)\cos(\zeta_l), \tag{22}$$

where ζ_x is the angle between the inter-nuclear vector and the x-axis and ζ_y, ζ_z are the angles for the y- and z-axes, respectively. Because the molecule is assumed to be rigid, for intra-molecular RDCs r_{ij} can be moved to the prefactor. Further, the maximal RDC $D_i^{\max}$ is introduced as

$$D_{ij}^{\max} = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi^3 r_{ij}^3}. \tag{23}$$

If the molecule is not rigid, for instance, if the RDCs are to be calculated for a trajectory rather than a single structure, it is necessary to incorporate averaging in (22), leading either to

$$D_{ij} = D_{ij}^{\max}\sum_{k\in x,y,z}\sum_{l\in x,y,z}\underline{A}_{kl}\langle\cos(\zeta_k)\cos(\zeta_l)\rangle, \tag{24}$$

where it is assumed that the alignment tensor is the same for all configurations, or to use of (22) for each structure in the trajectory. In (23), the distance can either be provided as a parameter or, in the case of inter-molecular RDCs or flexible bonds, computed as $\langle r_{ij}\rangle$ from the trajectory coordinates.

In GROMOS, RDCs can be calculated from a single structure using (22) or from a trajectory using (22) or (24). The alignment tensor $\underline{A}$ is determined by a singular-value decomposition (SVD) fit by solving the equation

$$\underline{C}\mathbf{a} = \mathbf{R}, \tag{25}$$

where the 5-dimensional vector $\mathbf{a} = (A_{xx}, A_{yy}, A_{xy}, A_{xz}, A_{yz})$ contains the five independent elements of the 3×3 alignment tensor. The other 4 elements of the alignment tensor are determined by the condition that the alignment tensor is symmetric and traceless. The N_D-dimensional vector $\mathbf{R}$ contains the N_D experimental RDCs used for the fit scaled by $1/D^{\max}$. The N_D rows of the $N_D \times 5$ coefficient matrix $\underline{C}$ are defined by the $\cos(\zeta)$ functions ($i = 1, \ldots, N_D$)

$$C_i = \left[\langle\cos^2(\zeta_x)\rangle - \langle\cos^2(\zeta_z)\rangle\right. \tag{26}$$

$$\langle\cos^2\left(\zeta_y\right)\rangle - \langle\cos^2(\zeta_z)\rangle, \tag{27}$$

$$2\langle\cos(\zeta_x)\cos\left(\zeta_y\right)\rangle, \tag{28}$$

$$2\langle\cos(\zeta_x)\cos(\zeta_z)\rangle \tag{29}$$

$$\left.2\langle\cos\left(\zeta_y\right)\cos(\zeta_z)\right]. \tag{30}$$

In (26) and (27), $\langle\cos^2(\zeta_z)\rangle$ is subtracted to fulfil the condition that the alignment tensor is traceless. The factor 2 in (28), (29) and (30) reflects the symmetry of the 3×3 alignment tensor.

The quality of the fit is assessed using R- and Q-values,

$$R = \frac{\sum_{i=1}^{N_D} |D_i^0 - D_i|}{\sum_{i=1}^{N_D} |D_i^0|}, \tag{31}$$

$$Q = \frac{\left[\frac{1}{N_D}\sum_{i=1}^{N_D}\left(D_i^0 - D_i\right)^2\right]^{\frac{1}{2}}}{\left[\frac{1}{N_D}\sum_{i=1}^{N_D}\left(D_i^0\right)^2\right]^{\frac{1}{2}}}. \tag{32}$$

Implementation details The SVD fitting procedure from the *Gnu Scientific Library* (Galassi et al. 2003) is used. The alignment tensor determination is implemented in analogy to a C++ program written by Robert Best, University of Cambridge, UK.

X-ray crystallography

The GROMOS software can be used to compute two kinds of crystallographic quantities, structure factors and electron density.

Structure factors In GROMOS a five Gaussian parametrisation (Waasmaier and Kirfel 1995) of the atomic scattering factor f_i is used,

$$f_i(\mathbf{k}) = O_i \exp\left(-\frac{B_i}{4k^2}\right)\left[\sum_{j=1}^{5}\left[a_{ij}\exp\left(-\frac{b_{ij}}{4k^2}\right)\right] + c_i\right]. \tag{33}$$

Here, O_i and B_i are the atomic occupancy and the atomic B-factor, a, b and c are coefficients which depend on the type and charge of the atom i (Waasmaier and Kirfel 1995). The atomic electron density ρ_i is given by the analytical Fourier backwards transform of the scattering factor f_i. The global electron density is then computed as the sum of the atomic densities

$$\rho(\mathbf{r}; \mathbf{r}^N) = \sum_{i=1}^{N} \rho_i(\mathbf{r}; \mathbf{r}^N) \tag{34}$$

on a grid whose resolution is an adjustable parameter. The structure factor is obtained by Fourier transform of the electron density of the system

$$F(\mathbf{k}; \mathbf{r}^N) = \mathcal{F}\{\rho(\mathbf{r}; \mathbf{r}^N)\}. \tag{35}$$

The structure factors obtained by the procedure described above are denoted as the *calculated* structure factors. Their amplitudes $|F|$ can be compared to the observed structure-factor amplitudes $|F^0|$, derived from the scattering intensities, through the R-factor which is given by

$$R = \frac{\sum_{i=1}^{N_F} \| F_i^0| - s|F_i\|}{\sum_{i=1}^{N_F} |F_i^0|}, \tag{36}$$

where N_F is the number of observed structure-factor amplitudes. As $|F^0|$ is generally on an arbitrary scale, the calculated amplitudes are fitted to the observed ones by means of a weighted linear regression, leading to the scaling factor

$$s = \frac{\sum_{i=1}^{N_F} w_i |F_i^0||F_i|}{\sum_{i=1}^{N_F} w_i |F_i|^2}. \tag{37}$$

The weight w_i of an individual reflection is generally taken as the inverse of the variance of the structure-factor amplitude.

As for ^3J-couplings, a harmonic potential energy term for the structure-factor amplitude can be used to restrain the computed structure-factor amplitudes to the observed ones

$$V^F(\mathbf{r}^N; |F^0|) = \frac{1}{2}\frac{K^F}{\sum_{i=1}^{N_F} w_i |F_i^0|^2}\sum_{i=1}^{N_F} w_i\left(|F_i^0| - s|F_i(\mathbf{r}^N)|\right)^2. \tag{38}$$

The weight factor or force constant K^F is made resolution independent by the factor $\frac{1}{\sum_{i=1}^{N_F} w_i |F_i^0|}^2$. Time-averaging is introduced by substitution of $|F_i|$ by $\langle|F_i|\rangle_t$ in (38) (Gros

et al. 1990). In addition to the instantaneous and time-averaging restraining term, a biquadratic term is available

$$V^F(\mathbf{r}^N;|F^0|) = \frac{1}{2}\frac{K^F}{\sum_{i=1}^{N_F} w_i |F_i^0|^4} \times \sum_{i=1}^{N_F} w_i \left(|F_i^0| - s_{\text{inst}}|F_i(\mathbf{r}^N)|\right)^2 \left(|F_i^0| - s_{\text{avg}}\langle |F_i(\mathbf{r}^N)|\rangle_t\right)^2. \quad (39)$$

We note that two individual scaling constants s_{inst} and s_{avg} are used. s_{inst} is computed using (37), while s_{avg} is calculated using the same relation, where the calculated amplitudes $|F_i|$ were replaced by the time-averaged ones $\langle |F_i| \rangle_t$.

In structure factor refinement sampling enhancement is of importance. GROMOS has the ability to scale parameters during the course of the simulation using predefined relations or curves. Different simulated annealing protocols can be easily implemented using these curves. It is also possible to adjust the resolution using such a curve. For example, all structure-factor amplitudes with a higher resolution than the current one can be omitted in the evaluation of (38). Using the relation

$$Res(\lambda) = Res_{\min} + \lambda(Res_{\max} - Res_{\min}) \quad (40)$$

the resolution can also be made λ-dependent and the resulting λ-dependent Hamiltonian can be used for Hamiltonian replica exchange. Application of the local elevation method is not possible when using structure factors. Through (35), every structure factor F_i depends on all the atomic coordinates $\mathbf{r}^N$. Thus it is not possible to define a reduced set of coordinates of low dimensionality required by the local elevation method.

Electron density In X-ray crystallography, the electron density ρ^0 is not an experimentally observable quantity. As the phase of F^0 is generally not observable, it is taken from the structure factor F computed from an atomic model. The electron density ρ^0 is then computed as

$$\rho^0(\mathbf{r}) = \mathcal{F}^{-1}\{\left(2|F^0(\mathbf{k})| - s|F(\mathbf{k})|\right)\exp[\text{i}\arg(F(\mathbf{k}))]\}, \quad (41)$$

where $\arg(F(\mathbf{k}))$ is the phase computed from the model. In analogy to (36), a real-space R-factor (Jones et al. 1991) can be defined

$$R = \frac{\sum_{\mathbf{r}} |\beta\rho^0(\mathbf{r}) + \alpha - \rho(\mathbf{r})|}{\sum_{\mathbf{r}} |\beta\rho^0(\mathbf{r}) + \alpha + \rho(\mathbf{r})|}. \quad (42)$$

Summation is only carried out over the extent of the atoms of interest. α and β are scaling constants computed using a linear regression of

$$\rho(\mathbf{r}) = \beta\rho^0(\mathbf{r}) + \alpha. \quad (43)$$

These values can be calculated for an arbitrary set of atoms. Fitting is carried out using only the grid points lying within spheres with a radius R_{cut} centered at the positions of these atoms.

The electron density computed using (41) can be used for restraining. The penalty function (Chapman 1995) is defined as

$$V^\rho(\mathbf{r}^N;\rho^0) = \frac{1}{2}K^\rho \sum_{\mathbf{r}} \left(\beta\rho^0 + \alpha - \rho(\mathbf{r}^N)\right)^2. \quad (44)$$

The summation is carried out over all grid points of the unit cell. Time-averaging is not available for electron density restraints, neither is a biquadratic penalty function.

As the electron density is a rather local quantity, a weight function for local elevation using a reduced set of coordinates i can be defined as

$$w_i(t) = w_i(t - \Delta t) + \begin{cases} \frac{1}{2}K_i\left(R_i - R_i^0\right)^2 & \text{for} \quad R_i > R_i^0 \\ 0 & \text{otherwise,} \end{cases} \quad (45)$$

where R_i is the real-space R-factor and R_i^0 is some tolerance constant. The summation in (42) should be carried out over the atoms affected by a configurational change in the reduced set of coordinates and not only over the atoms defining this set. For example, if the local elevation is carried out in the χ_1 angle of a phenylalanine residue, all atomic positions of the side chain are altered by a conformational change in χ_1 and thus should be used in the computation of R_i. In order to incorporate more electron density perturbation induced by the movement of these atoms, the cutoff for the summation is an adjustable parameter and can be extended to allow consideration of grid points farther away from the atomic centres.

Symmetry In a crystal, having a space group different from P1, multiple identical asymmetric units (ASUs) are assembled to construct the unit cell. The relationship between the ASUs is fully defined by the space group and the size of the unit cell. A space group S contains N_{sym} number of symmetry operations $\mathbb{S}_i$. Every atom position in the first asymmetric unit $\mathbf{r}_i$ has $N_{sym} - 1$ images $\mathbf{r}_i^{(j)}$ defined as

$$\mathbf{r}_i^{(j)} = \mathbb{S}_j\mathbf{r}_i. \quad (46)$$

Note that $\mathbb{S}_1$ is the identity symmetry operation and thus $\mathbf{r}_i^{(1)}$ and $\mathbf{r}_i$ are identical. Symmetry operations can be described in terms of a rotation followed by a translation,

$$\mathbb{S}_j\mathbf{r}_i = \underline{R}_j\mathbf{r}_i + t_j. \quad (47)$$

Here, the rotation matrix $\underline{R}_j$ of the symmetry operation j is, in contrast to the translation vector t_j, independent of the

size of the unit cell. The inverse symmetry operation $\mathbb{S}^{-1}$ can be formulated as

$$\mathbb{S}_j^{-1}\mathbf{r}_i = \underline{R}_j^{-1}\left(\mathbf{r}_i - \mathbf{t}_j\right). \tag{48}$$

When simulating a unit cell, one may want to impose restraints on the symmetry of the system as, in many cases, insufficient experimental data for refinement of the whole unit cell is available. An easy way of achieving this is by adding harmonic potential energy terms for every symmetry-related pair. For every atom i in the first asymmetric unit, the term

$$V^{\text{sym}}(\mathbf{r}_i; K^{\text{sym}}) = \frac{1}{2}K^{\text{sym}} \sum_{i'=1}^{N_{\text{sym}}-1} \sum_{j'=i'+1}^{N_{\text{sym}}} \left[\mathbb{S}_{j'}^{-1}\mathbf{r}_i^{(j')} - \mathbb{S}_{i'}^{-1}\mathbf{r}_i^{(i')}\right]^2 \tag{49}$$

is added. The term can be interpreted as follows: all possible symmetry pairs are transformed to their position in the first asymmetric unit and a set of springs between corresponding atoms is introduced for each pair of molecules.

Atomic B factors In addition to the position of the atoms, the size of their electron cloud is an important parameter of the atomic model. The B_i factor is used in (33) for the computation of the electron density and finally the structure factors. The atomic B_i factors can be optimised every requested time step by steepest-descent energy minimisation of (38). The termination criteria, a chosen gradient size or number of minimisation steps, are adjustable parameters of the method. It is not recommended to optimise the B-factors during an MD simulation as discontinuities are introduced in the potential energy surface which can lead to artifacts such as heating or large forces.

Implementation details The crystallographic refinement features of GROMOS are implemented using the *Clipper* framework (Cowtan 2003). The electron density is computed on a crystallographic map (`Xmap`) using the atomic electron-density implementation of the class `AtomShapeFn` in the same way as in the `EDcalc_iso` function object. The code was reorganised for a parallel evaluation of the sum in (34). The structure factors are computed using the standard FFT functionality of the `Xmap` class. The observed electron density is computed by the aid of the same classes. For fitting and scaling of the structure-factor amplitudes and the computation of the R-value, particular routines were written. For structure factor restraints, forces are computed using a difference map which was implemented as an FFT grid for the P1 space group (class `FFTmap_p1`). The symmetry operations of the space group are applied if this is necessary. The forces are computed using a convolution of the difference map with the atomic electron density gradient from class `AtomShapeFn`. As every convolution can be computed independently, the force loop was parallelised. Forces for the electron density restraints are computed using the atomic electron density gradients in a similar way. The routine for the structure factor restraint forces can also be used for the computation of the derivative of the B factor. The B factor is optimised using the multi-dimensional function minimiser of the *Gnu Scientific Library* (Galassi et al. 2003). The steepest-descent algorithm was chosen. Particular routines were written for the scaling of electron densities and the computation of the real-space R-value.

Neutron scattering

Neutron scattering data is a valuable source of observed, primary experimental data. It can be used to describe properties of liquids in terms of the radial distribution function $g(r)$. In an isotropic liquid, the distinct scattering intensities $I(Q)$ are computed from an intra- and an intermolecular contribution (Debye 1913; van Hove 1954; Windsor 1973)

$$I(Q) = I_{\text{intra}}(Q) + I_{\text{inter}}(Q), \tag{50}$$

where $\mathbf{Q} = 2\pi\mathbf{q}$ and $\mathbf{Q} = \mathbf{k}_{\text{in}} - \mathbf{k}_{\text{out}}$, the difference of the inverse wave-length vectors of the incident and scattered beams. One has

$$I_{\text{intra}}(Q) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\langle b_i\rangle\langle b_j\rangle F_{ij}^{\text{intra}}(Q) \tag{51}$$

$$I_{\text{inter}}(Q) = \frac{1}{N}\sum_{\alpha=1}^{N_{\text{types}}}\sum_{\beta\geq\alpha}^{N_{\text{types}}}(2-\delta_{\alpha\beta})c_\alpha c_\beta\langle b_\alpha\rangle\langle b_\beta\rangle F_{\alpha\beta}^{\text{inter}}(Q), \tag{52}$$

where the summation in (51) is only carried out over intra-molecular pairs and in (52) only over all combinations of atom types N_{types}. The coefficient b_i is the scattering length of an atom i and $c_\alpha = \frac{N_\alpha}{N}$ is the fraction of atoms of type α. In (52) the Kronecker delta $\delta_{\alpha\beta}$ is used to avoid double counting. The structure factors F for the intra- and inter-molecular contributions are given by

$$F_{ij}^{\text{intra}}(Q) = \frac{\sin(Qr_{ij})}{Qr_{ij}}\exp\left[-\frac{(\sigma_{ij}Q)^2}{2}\right] \tag{53}$$

$$F_{\alpha\beta}^{\text{inter}}(Q) = 4\pi\rho_a\int_0^\infty r^2(g_{\alpha\beta}(r)-1)\frac{\sin(Qr)}{Qr}dr, \tag{54}$$

where σ_{ij} is the standard deviation of the intra-molecular distance between atoms i and j and ρ_a is the atomic number density. The function $g_{\alpha\beta}(r)$ denotes the radial distribution function for the two atom types α and β which can be computed from an MD trajectory. In a classical MD simulation it is generally not practical to determine σ_{ij} from the

simulation, as the bond-length fluctuations are of quantum-mechanical nature and time scale. In a rigid solvent model, like those standardly used in GROMOS, these fluctuations are not present and thus $\sigma_{ij} = 0$. For this reason, the parameters σ_{ij} are usually determined by a least-squares fit of $I(Q)$ on $I^0(Q)$, the measured scattering intensity.

Implementation details The computation of the radial distribution functions was parallelised.

X-ray scattering

In GROMOS, it is possible to compute small- and wide-angle X-ray scattering (SAXS/WAXS) intensities. The structure factor $F(\mathbf{q})$ is given by

$$F(\mathbf{q};\mathbf{r}^N) = \sum_{i=1}^{N} f_i \exp(2\pi \mathrm{i}\mathbf{r}_i\mathbf{q}), \tag{55}$$

where the atomic form factor f_i is set to the number of electrons of the atom i. The intensities $I(\mathbf{q})$ are given as the product of the structure factor with its complex conjugate $\overline{F}(\mathbf{q})F(\mathbf{q})$, resulting in

$$I(\mathbf{q}) = \sum_{i=1}^{N}\sum_{j\neq i}^{N} f_i f_j \exp\left[2\pi \mathrm{i}(\mathbf{r}_i - \mathbf{r}_j)\mathbf{q}\right]. \tag{56}$$

In practice, (56) is evaluated using a fast Fourier transform. The product of the atomic form factors $f_i f_j$ is assigned on a mesh at the position of the inter-atomic distances $\mathbf{r}_i - \mathbf{r}_j$ using assignment functions of first or second order (see Table 7 of Christen et al. 2005). If requested, intra-molecular distances are neglected. The mesh is averaged over the trajectory and a fast Fourier transform is applied, yielding the three-dimensional scattering intensities $I(\mathbf{q})$. One- and two-dimensional spectra are generated by projection of the three-dimensional intensities.

Implementation details The FFT routines of the *Gnu Scientific Library* (Galassi et al. 2003) are used.

Example applications

The following examples draw upon the novel combination of recently developed techniques for conformational searching and sampling using biasing penalty functions that allow for conformational averaging. The examples regarding the use of NMR data concern a 16-residue peptide for which single-structure refinement in vacuo could not produce a set of model structures reproducing the NOE and ^3J-coupling data, and the 129-residue protein lysozyme for which it is shown that averaging is of importance when calculating residual dipolar couplings. The examples regarding the use of X-ray data concern the 72-residue protein ubiquitin for which different conformational sampling methods were investigated: replica-exchange and damped oscillatory variation of the biasing weight factors. The local-elevation conformational sampling method is investigated using the 56-residue bovine pancreatic trypsin inhibitor. In addition, the significance of using sufficient data compared to degrees of freedom is illustrated. Finally, two examples of the calculation of X-ray or neutron scattering intensities are presented for a membrane and a liquid.

NMR refinement

The structure of the C-terminal trigger peptide of GCN4-p1 has been determined using 2D NOESY experiments by Steinmetz et al. (Steinmetz et al. 2007), and had been refined using the X-PLOR software and a standard simulated-annealing protocol (Gronenborn et al. 1991). The obtained set of NMR model structures was deposited in the protein data bank with the entry code 2OVN. In Dolenc et al. (2010), analysed the set of NMR model structures with respect to NOE distance bound violations and compared the computed ^3J-couplings with the experimental ones (denoted $^3J^0$, top panel of Fig. 1). There were a few,

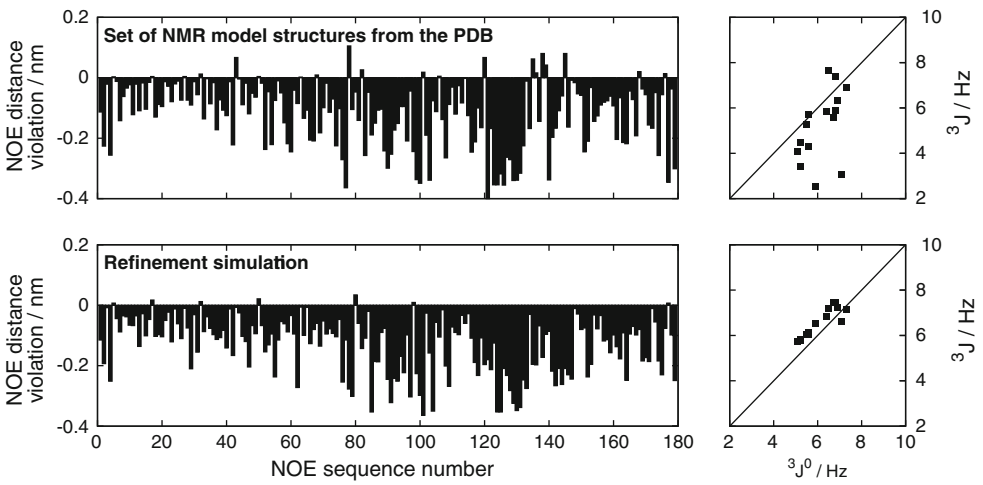

Fig. 1 Peptide GCN4-p1 in water: NOE distance bound violations and ^3J-coupling deviations from experiment. Only positive NOE distances $\langle r_i^{-3}\rangle^{1/3} - r_i^0$ are true upper-bound violations. Top panels: values obtained from the deposited set of 20 NMR model structures from Steinmetz et al. (2007). Bottom panels: values obtained from a refinement simulation by Dolenc et al. (2010) using time-averaged NOE restraining and the ^3J-coupling weighted local-elevation method

rather small, positive NOE upper-bound violations (negative ones are not true violations) but the computed ^3J-couplings deviated significantly from the experimental ones. For this reason, several 10 ns refinement simulations using the GROMOS software were carried out. The time-averaged NOE restraining method in conjunction with the ^3J deviation weighted local elevation method yielded an ensemble which was very well in agreement with both experimental data sets (bottom panel of Fig. 1). The NOE violations are significantly lower than the ones obtained from the deposited set of NMR model structures derived by single-structure refinement in vacuo and in addition, the ^3J-coupling deviations from experiment were well below 1 Hz, which is the estimated accuracy of the Karplus relation. Because of averaging of the NOE distances through (15), the inter-nuclear distance only has to be within the NOE bound for a rather small fraction of time in order to fulfil the bound. Using the local elevation method, multiple conformations representing different solutions of the Karplus relation (17) are visited. Because the time-averaging and local elevation methods were used, the resulting ensemble shows much more conformational variation than the set of NMR model structures, while being in better agreement with the experimental data.

Residual dipolar couplings

The protein hen egg-white lysozyme (HEWL) was used as an example for back calculation of RDC data. As the crystallographic model deposited in the protein data bank (entry code 1AKI (Artymiuk et al. 1982)) does not contain any hydrogens, they were added using the GROMOS++ program *gch*. An energy minimisation in vacuo was carried out. The N-H RDCs were computed using (22) (single structure data in Fig. 2). After a standard equilibration scheme for an MD simulation of a protein (see e.g. ref. Schmid et al. 2008), a 20 ns simulation was carried out. Conformations were saved every 5 ps. For the energy minimisations and the simulation the 53A6 force field (Oostenbrink et al. 2004) was used. The resulting trajectory was used to evaluate (24) (trajectory data in Fig. 2). The overall agreement with the experimentally observed D^0 data can be expressed as R and Q values. They were $R = 28\%$ and $Q = 41\%$ for the starting structure and $R = 49\%$ and $Q = 67\%$ for the trajectory. The agreement with the experimental data in the starting structure is, with a root-mean-square deviation of 2.7 Hz, not very good. In the free MD simulation the experimental data is even less well reproduced, with an RMSD of 4.5 Hz. As the quality of the force field is limited and long sampling times are required to converge the averages in (24), additional restraining techniques for RDCs are needed. Various new approaches for back calculation and refinement of RDCs that allow for conformational averaging are under development.

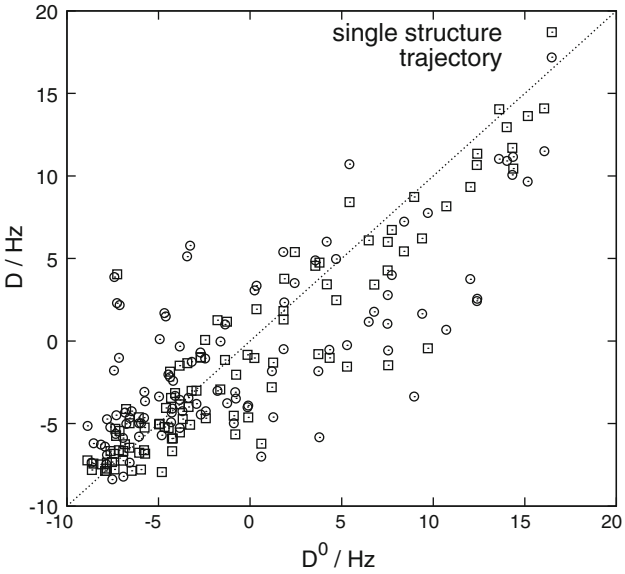

Fig. 2 Lysozyme in water: N–H residual dipolar couplings D calculated from a single structure or an MD simulation (D) in comparison with experimentally observed values (D^0) of Higman et al. (2010)

X-ray crystallographic refinement

Evaluation of the different residuals In GROMOS three different penalty functions for reciprocal space crystallographic structure refinement were implemented, namely instantaneous, time-averaged and biquadratic structure-factor amplitude restraining. These methods were evaluated by refinement of the protein ubiquitin. Here, the aim was only to illustrate the differences between the refinement methods using a well characterised model protein and not to propose a new refined structure of ubiquitin. The initial coordinates, B-factors, occupancies and structure-factor amplitudes were taken from the protein data bank (entry code 1UBQ (Vijay-Kumar et al. 1987)). The reported R-factor was 17.6%. If the R-factor is computed using GROMOS, a value of 21.2% is obtained due to the lack of solvent corrections, data filtering and because a single scaling constant for the whole resolution range was used. The unit cell was constructed using the parameters given in the PDB file and the symmetry operations of the P 2_1 2_1 2_1 space group. For all simulations and energy minimisations the 53A6 force field (Oostenbrink et al. 2004) was used. Electrostatic interactions of the crystalline system were evaluated using the Particle-Particle, Particle-Mesh (P^3M) method (Hockney and Eastwood 1981). A truncated polynomial charge shaping function ($N_\gamma = 2$) with a width of 0.8 nm, a grid of dimensions $32 \times 32 \times 32$, an assignment function and finite difference operator of

third order, and tinfoil boundary conditions were used. The energy of the unit cell was minimised, yielding an *R*-factor of 26%. This configuration was taken as the starting point of all the refinement simulations (Figs. 3, 4, 5). The symmetry was restrained using the methodology described above and a force constant of 2 MJ mol^{-1} unless indicated otherwise. The equations of motion were integrated with the leap-frog method and a time step of 1 fs. Bond lengths and the bond angle of the water molecules were constrained using the SHAKE algorithm (Ryckaert et al. 1977). 10% of the structure-factor amplitudes were randomly removed from the restraining set and used to compute a free *R*-factor (Brünger 1992). The free *R*-factor computed from the initial configuration was 28%.

Three 200–400 ps refinement simulations of ubiquitin at 300 K, using either the instantaneous, time-averaged or biquadratic structure-factor amplitude restraining methods, were carried out (top panels of Fig. 3). The force constant was chosen as 1 GJ mol^{-1} yielding a root-mean-square force due to the restraints being half of the root-mean-square force due the force field. The time-averaging memory-relaxation time was 1 ps. Using the instantaneous restraining method, the free *R*-factor is reduced by 2% compared to the one of the initial model. However, the atom positional root-mean-square deviations of the backbone atoms from the X-ray structure (Vijay-Kumar et al. 1987) are very low. The conformation is locked in a single energy minimum. The time-averaging restraining method yields significantly lower free *R*-factors than the instantaneous restraining simulation. The lowest *R*-factor found is 18%, which is very close to the value reported in the protein data bank. Note that this value was achieved without any solvent corrections or similar techniques. With an RMSD reaching up to 0.15 nm, sampling is significantly enhanced with respect to the instantaneous simulation. By allowing these structural fluctuations, not only is a better *R*-value achieved, but additional insight into the flexibility of the protein is gained. The biquadratic restraining is not able to improve the initial model but makes it worse. This can be explained by the factorisation of the scaling constants in (39)

$$V^F(\mathbf{r}^N;|F^0|) = \frac{1}{2}\frac{K^F}{\sum_{i=1}^{N_F} w_i |F_i^0|^4} \times \sum_{i=1}^{N_F} w_i s_{\text{inst}}^2 s_{\text{avg}}^2 \left(\frac{|F_i^0|}{s_{\text{inst}}} - |F_i(\mathbf{r}^N)|\right)^2 \left(\frac{|F_i^0|}{s_{\text{avg}}} - \langle |F_i(\mathbf{r}^N)| \rangle_t\right)^2. \quad (57)$$

Because the instantaneous and time-average structure-factor amplitudes are not exactly on the same scale, the minima in the two harmonic oscillator factors in (57) are different. So instead of improving the overshooting problem, these two harmonic oscillators are working against each other. This can only be avoided by choosing a very short memory relaxation time for averaging, such that it is not physically meaningful anymore (data not shown).

In addition, a simulation using the electron density restraining methodology was carried out. Because the potential energy surface was very steep, the integration time step was reduced to 0.5 fs, resulting in an overall simulation length of 100 ps. The trajectory obtained does not show large structural fluctuations and the RMSD values are rather low. However, the free *R*-factor is high and the obtained structures are in poor agreement with the experimental data. This can be explained by the model bias found in the "observed" electron density.

An additional instantaneous restraining simulation of 200 ps with the symmetry restraints being removed was carried out to investigate their effect. The *R*-factor is reduced by 0.5% in the simulation in which the symmetry was not restrained (time series not shown). However, the free *R*-factor increased by 2% (middle panels of Fig. 3). There are not enough experimental structure-factor amplitudes available to refine the four asymmetric units independently from each other. For this reason, the symmetry was restrained in all further refinement simulations.

Three 400 ps time-averaging refinement simulations with memory-relaxation times of 1, 2 and 5 ps were carried out. The longest memory-relaxation time of 5 ps yields the lowest free *R*-values (bottom panel of Fig. 3) and the lowest *R* values found in this run were the same as the one reported in the PDB. However, after 150 ps the free *R*-factor and RMSD were growing in the simulations using $\tau = 2$ and 5 ps. Because the averaged structure factor amplitudes were satisfying the observed ones so well, the conformation was allowed to explore different regions in phase space, which is reflected in the very large RMSD values obtained in these runs. Using these relaxation times, the average $\langle |F| \rangle_t$ became too insensitive to configurational changes and as the conformation left the convergence region, the free *R*-factor and RMSD were steadily increasing.

Modifying refinement parameters In GROMOS, many parameters such as the temperature, restraining force constant and resolution can be modified during the course of a simulation. The instantaneous value is given by a relation or curve defined by control points. Between these control points either linear or cubic spline interpolation can be used. In addition, it is possible to use the control points to define a Bézier curve (de Casteljau 1959). Finally, a harmonic oscillation with defined amplitude, phase and frequency can also be used. In all examples, the instantaneous structure-factor amplitude restraining method was used.

A temperature simulated annealing protocol can easily be implemented using these curves. Generally, the refinement is started at elevated temperature (here 1,000 K) and reduced resolution. Three 50 ps refinement simulations

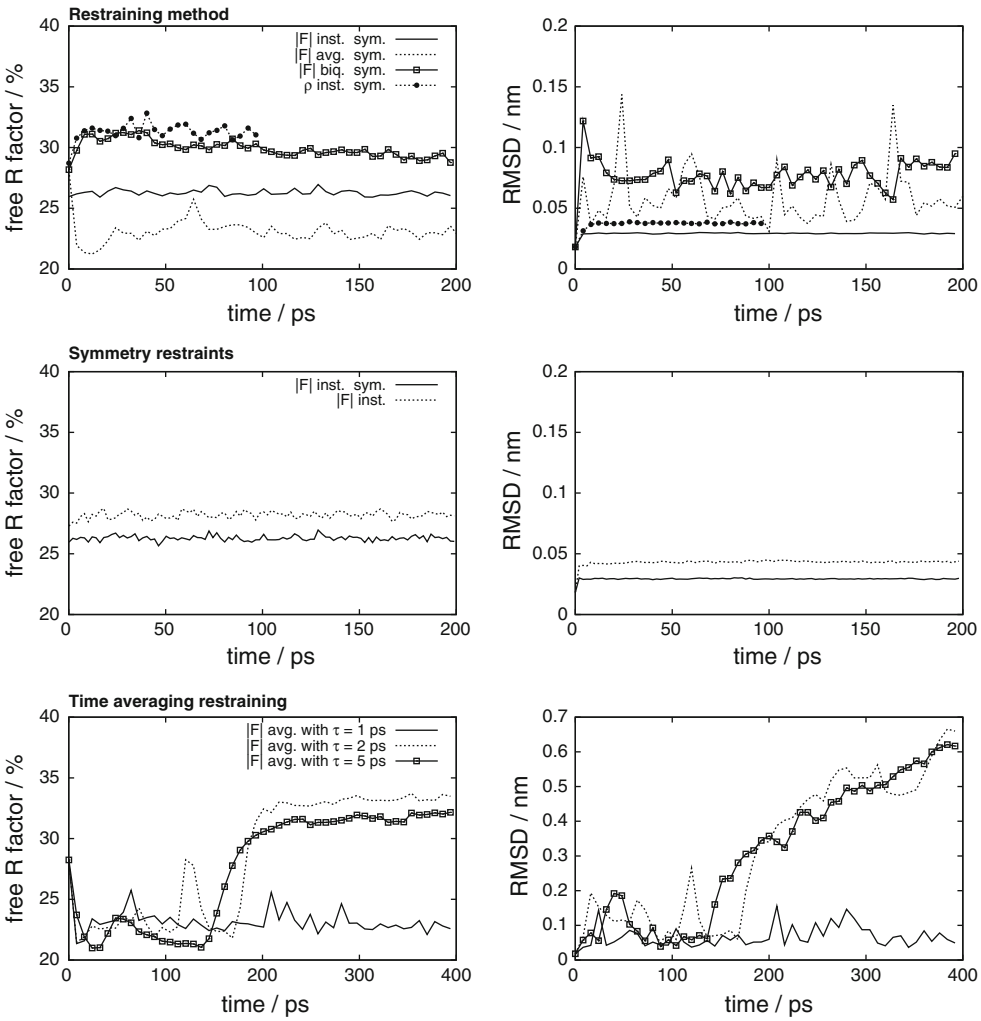

Fig. 3 Crystalline ubiquitin: variation of structure refinement methods, restraints and parameters using X-ray diffraction data. Restraining methods used: instantaneous structure-factor amplitude restraining (|*F*| inst.), time-averaged structure-factor amplitude restraining (|*F*| avg.), biquadratic structure-factor amplitude restraining (|*F*| biq.), instantaneous electron density restraining (ρ inst.), symmetry restraining (sym.). *Left panels* free *R*-factor. *Right panels* The atom-positional RMSD of the backbone atoms of the four asymmetric units of the unit cell with respect to the X-ray structure (Vijay-Kumar et al. 1987)

were carried out starting with a resolution of 1, 0.5 or 0.18 nm (top panels of Fig. 4). During the course of the simulation the resolution was raised to 0.18 nm. At the same time, the temperature was lowered from an initial value of 1000 K to 300 K. The obtained free *R*-values do not differ much, but the run where the resolution was always kept at 0.18 nm yields the lowest values.

In additional simulations, the resolution was modified in an oscillatory way at a constant temperature of 300 K. As the potential energy surface is smoother at lower resolution, sampling should be enhanced. Two 200 ps simulations were carried out, one with a harmonically oscillating resolution and one where a trapezoidal step function of the same frequency was used (middle panels of Fig. 4). The free *R*-factors obtained in these runs are comparable to the ones obtained in the simulated annealing run. However, sampling was not significantly enhanced.

In an additional set of simulations, the force constant of the restraining energy term was scaled in a similar way as the resolution (bottom panels of Fig. 4). The maximum force constant was the one used in all the refinement simulations and the minimum was a tenth of it. By reducing the weight of the restraints, strain in the protein is released and sampling is enhanced. Using a step function, where the weight of the restraints was reduced by a tenth for 5 ps per oscillation, fewer low free *R* conformations were found than by using a harmonic oscillation scheme.

The parameter changing feature of GROMOS makes it very flexible software. Different refinement protocols and strategies can be employed without having to write additional modules or scripts.

Replica exchange The replica exchange method is a well known sampling-enhancement technique. In the Hamiltonian replica-exchange method proposed for X-ray crystallographic structure refinement, a parameter λ is introduced which makes the Hamiltonian dependent on the resolution through (40). A refinement simulation of ubiquitin using instantaneous structure-factor amplitude restraints and a resolution range of 0.18–1.2 nm was carried out. 42 parallel refinement simulations (replicas) with resolutions at constant intervals of 0.024 nm were used. Every 200 MD steps (0.2 ps), an exchange trial between neighbouring

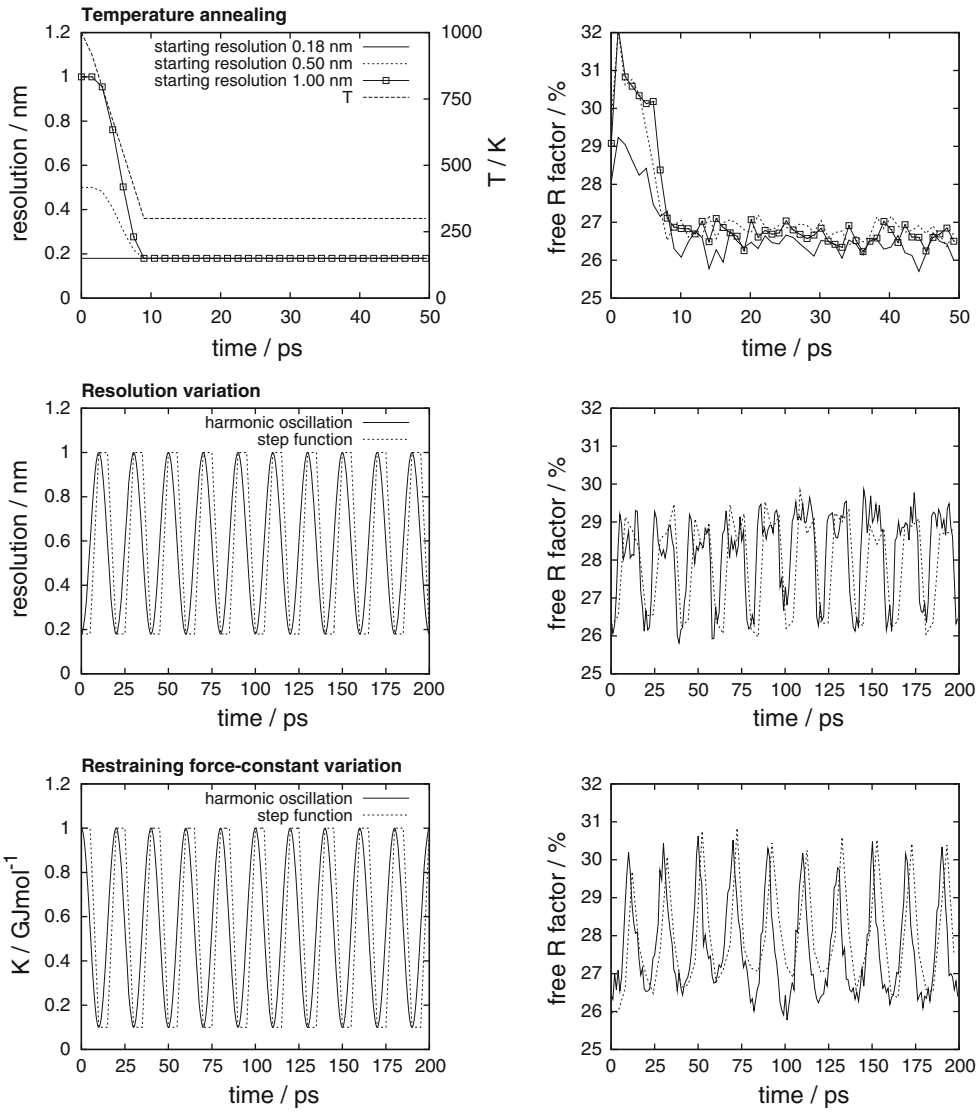

Fig. 4 Crystalline ubiquitin: examples of different structure-refinement strategies. *Top panel* The temperature T is reduced from 1,000 to 300 K while the resolution is either kept constant at 0.18 nm or increased from an initial value of 0.50 or 1.00–0.18 nm. *Middle panels* The resolution is changed in the range from 0.18 to 1.00 nm according to a harmonic oscillation function or a step function of the same frequency. *Bottom panel* The restraining force constant is changed in the range from 1.0 to 0.1 GJ mol^{-1} according to a harmonic oscillation function or a step function of the same frequency

replicas was attempted. In total, 800 such exchange trials were attempted per simulation, resulting in a total simulation length of 160 ps. A time-series of the resolution of the individual refinement simulations is displayed in Fig. 5. In the very low resolution range (0.8–1.0 nm) there were many switching events and the exchange probabilities were high. In the higher resolution replicas, however, the exchange probability was almost zero and only a few exchange events were observed. The restraining potential energy function (38) contains a normalisation factor to bring the energies on the same scale regardless of the number of experimental amplitudes or resolution in use. However, the total potential energies of the replicas at a switching step are on different scales (bottom panel of Fig. 5). This is due to the different conformations of the replicas: even though the restraining energies are on the same scale, the physical energies need not be. Because there is almost no overlap between a few of these curves,

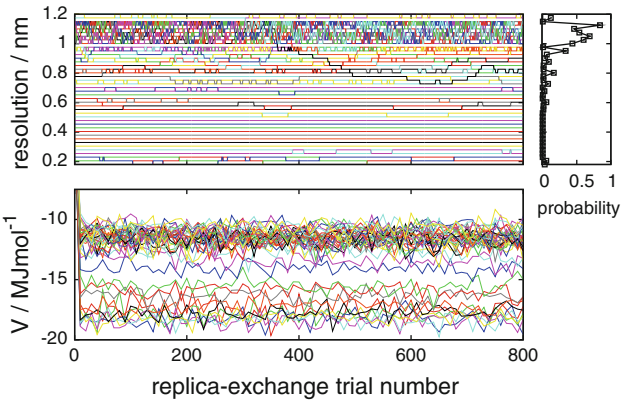

Fig. 5 Crystalline ubiquitin: structure refinement using resolution dependent Hamiltonian replica-exchange. *Top panels* resolution of the individual replicas, their switching behaviour and switching probability averaged over all attempts. *Bottom panel* total potential energy V, i.e. the force-field plus restraining energy, of the individual replicas

the switching probability is very low for these replica pairs. In the low resolution range, where only a few restraints are in use, the perturbation between the replicas is much smaller than in the high resolution range, so the problem could be partly redressed by increasing the number of replicas. Because the calculation of the structure factors and their gradients is computationally demanding and has to be carried out in parallel, an increase of the number of replicas is not justified with respect to the sampling enhancement gained. In addition, in order to significantly enhance sampling, the resolution range should be increased, which leads to an additional increase in the number of replicas. For these reasons, it is recommended to instead use an oscillating resolution or force constant to obtain similar results in a single simulation.

Local elevation The protein bovine pancreatic trypsin inhibitor (BPTI) was used to evaluate sampling enhancement based on local elevation and the real-space *R*-factor. The atomic coordinates, B-factors and occupancies, the unit-cell parameters and structure-factor amplitudes were taken from the protein data bank (entry code 6PTI) (Wlodawer et al. 1987). In this case, the initial coordinates were generated from those in the PDB by changing the N-CA-CB-OG torsional angle of the serine 46 residue from −119° to 50°. Only one asymmetric unit was simulated in vacuum. All atoms but the OG and HG atoms of serine 46 were positionally constrained to their position in the initial structure. In this way, sampling of the serine 46 side-chain conformations could be investigated. Several 100 ps refinement simulations at a temperature of 300 K were carried out, in which the local elevation method on the dihedral angle N-CA-CB-OG of serine 46 with a force constant of 1 MJ mol^{-1} was used. Repulsive functions of truncated-polynomial form were built on 36 grid points depending on the weight function (45) with a force constant of 0.5 kJ mol^{-1} and threshold values of $R_0 = 0.21$, $R_0 = 0.22$, $R_0 = 0.23$ or $R_0 = 0.24$. In (42), the real-space *R*-factor was only computed over the space covered by the atoms CA, CB, OB and HG of serine 46 defined by a atom/grid-point cutoff distance of 0.3 nm. For the computation of the free *R*-factor 10% of the structure-factor amplitudes were randomly selected from the experimental set.

The *R*-factors and dihedral-angle distributions obtained strongly depend on the threshold parameter R_0 (Fig. 6). In the local-elevation method the definition of a flat-bottom weighting function is crucial. In the case of $R_0 = 0.21$ and 0.22, the threshold parameters were too low and all conformations were penalised. This is reflected in the rather flat dihedral-angle distributions obtained for those values (right panel of Fig. 6). All grid points were visited and the free *R*-factor was sampling a 16.5–17.2% range. In an additional simulation, the threshold value was chosen was $R_0 = 0.23$. Immediately, the wrong conformation of 50°

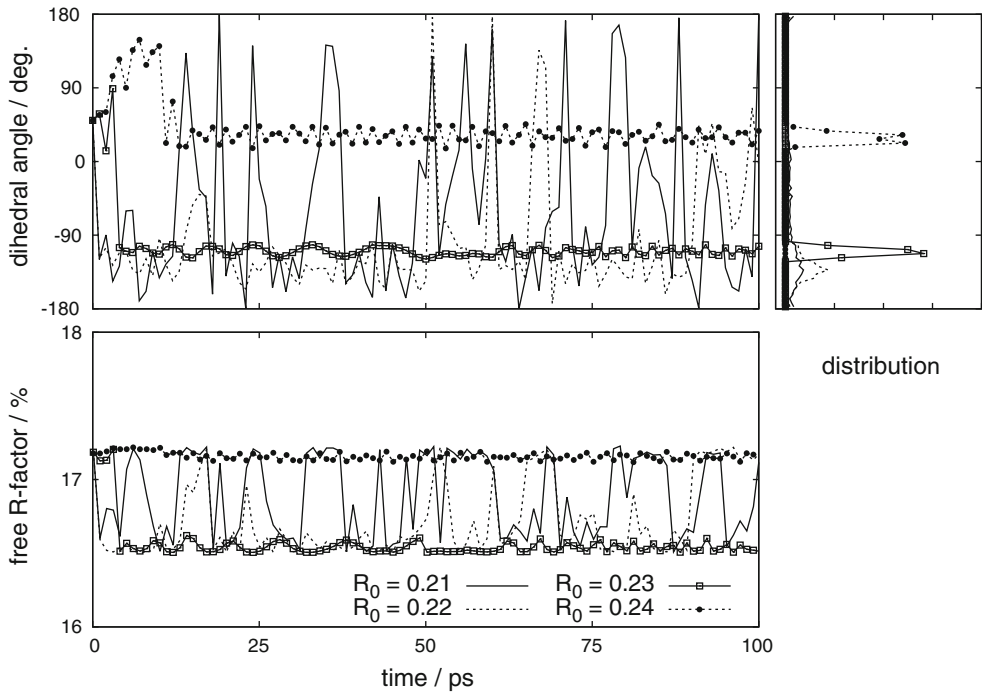

Fig. 6 Asymmetric unit of BPTI in vacuo: structure refinement of the side chain of serine 46 using the real-space *R*-factor weighted local-elevation method on the dihedral angle N-CA-CB-OG. Different threshold parameters R_0 (45) of 0.21–0.24 were used. *Top panels* Instantaneous value of the dihedral angle N-CA-CB-OG of serine 46 and its distribution obtained in the simulation. *Bottom panels* free *R*-factors

was penalised and the hydroxyl group rotated to the correct conformation of −119°. As in this conformational bin, the real-space *R*-factor was lower than the threshold value of 0.23, no repulsive functions were built and the correct conformation of the serine residue was retained during the remaining simulation. This is also reflected in the sharp peak in the dihedral angle distribution. Finally, a simulation with a too large threshold value of 0.24 was carried out. Using this value the initial, wrong conformation is not penalised and for this reason the serine side chain was only moving by a few degrees and was subsequently locked in the wrong conformation.

The threshold value R_0 must be chosen for every side chain to be refined. An optimal value is slightly larger than the real-space *R*-factor of the best conformation. Unfortunately, this value is a priori not known and has to be estimated. This becomes increasingly difficult for simultaneous refinement of multiple side chains. A method with a growing R_0 value was implemented. This approach appeared to be inefficient as all possible conformations were uniformly sampled multiple times, like in the $R_0 = 0.22$ simulation in order to find the correct minimum. Generalising the local-elevation method to multi-dimensional local elevation and multiple residues, the sampling method becomes impractical. The conformations of nearby side chains are correlated and so are their real-space *R*-factors. Therefore, a decoupling of the local-elevation coordinates for nearby residues from each other is a poor approximation. Without decoupling, the refinement of the neighbouring serine 46 and asparate 49 side chains requires a four dimensional local elevation, in the χ_1 and χ_2 angles of both residues, which is computationally very hard to carry out. For larger side chains like those of lysine or arginine the method becomes not applicable anymore.

X-ray scattering

X-ray scattering intensities, SAXS and WAXS spectra, can be used to derive small and large scale structural information. These experimental methods are often applied to membrane systems as the latter cannot be crystallised. The methodology for X-ray scattering (56) was applied to trajectories obtained from MD simulations of a DPPC bilayer, consisting of 2 × 8 × 8 DPPC molecules, in SPC water at two different temperatures. The simulation at 293.15 K was in the gel phase while the one at 313.15 K was in the liquid crystalline phase. The two 1 ns simulations were carried out in the $NA_{xy}P_zT$ ensemble (constant number of particles *N*, area in the *xy*-plane A_{xy}, pressure P_z along the *z*-axis, perpendicular to the membrane, and temperature *T*) using a protocol which is described elsewhere (Horta et al. 2010). For the computation of the spectra, a grid size of 512^3 and cutoff of 40 nm were used. As this cutoff is beyond the size of the computational box, the remaining space of computational grid was filled with zeros in order to smooth the spectra.

The quantitative agreement of the calculated and experimental (Frühwirth et al. 2004) spectra is rather poor (Fig. 7). Nevertheless, the main qualitative features of the spectra are reproduced for both the gel and liquid crystal data. In particular, the peak found in the gel phase data due to enhanced order in the bilayer is clearly seen.

Neutron scattering

Neutron scattering intensities $I(Q)$ can be used to obtain a description of the structure of liquids in terms of the radial distribution functions $g(r)$. In molecular simulation, the determination of $g(r)$ is straightforward and standardly carried out. However, its value for validation of simulation data is limited by the methodology and assumptions used to compute the $g(r)$ from the neutron scattering intensities. The latter are the observed, primary experimental data, whereas a $g(r)$ constitutes derived, secondary data.

For the validation of a new charge-on-spring (COS) polarisable model of CCl_4 (Kunz et al. 2011), the methodology described above was used to determine the neutron scattering intensities. A 1 ns production simulation was carried out. The scattering lengths of $b_C = 6.646$ fm and $b_{Cl} = 9.58$ fm were used (Lide 2007). The root-mean-square displacements σ_{CCl} and σ_{ClCl} were estimated from those of water, see (Kunz et al. 2011). The neutron scattering intensities determined in that way and those measured (Misawa 1989; Pusztai and McGreevy 1997) are displayed in Fig. 8. The simulation data reproduces the experimental scattering intensities very well on the whole resolution range. In conjunction with physically meaningful $g(r)$ curves, the structural properties of the new CCl_4 model could be validated.

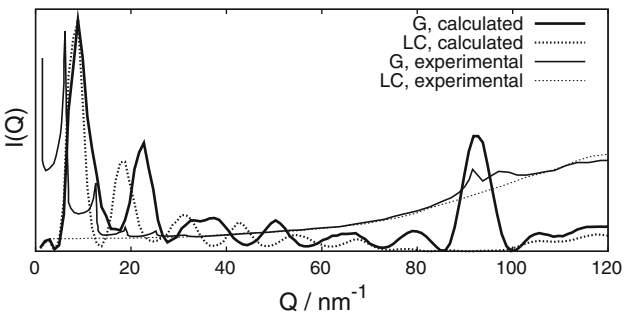

Fig. 7 DPPC bilayer in water: WAXS intensities $I(Q)$ computed from a 1 ns simulation of a DPPC bilayer at 293.15 K (G, gel phase) and at 313.15 K (LC, liquid crystalline phase) and experimental data (Frühwirth et al. 2004)

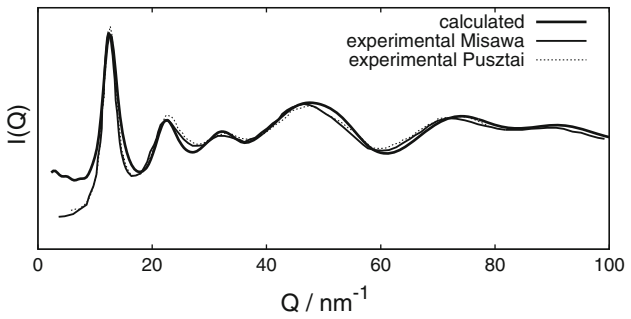

Fig. 8 CCl_4: Neutron scattering intensities $I(Q)$ of liquid CCl_4: computed from a 1 ns simulation (Kunz et al. 2011), *bold line*. Experimental data from (Misawa 1989), solid line, and from (Pusztai and McGreevy 1997), *dotted line*

Conclusions

The functionality of the GROMOS software for biomolecular simulation in regard to structure refinement of biomolecules based on a variety of spectroscopic and diffraction data was described. It includes time-averaging of various observable quantities, such as ^3J-couplings and NOE related distances derived from NMR spectroscopic data and structure factor amplitudes and electron density derived from X-ray crystallographic diffraction data. The software allows combination of different types of experimental data in a single refinement MD simulation. It also allows the use of a variety of techniques to enhance conformational searching or sampling such as local-elevation search or Hamiltonian replica-exchange sampling in structure refinement. Finally, it allows the use of the thermodynamically calibrated GROMOS force fields (Oostenbrink et al. 2004; Schuler et al. 2001; Schmid et al. 2011) in structure refinement.

The use of local-elevation search and Hamiltonian replica-exchange in structure refinement based on X-ray crystallographic data for the proteins ubiquitin and bovine pancreatic trypsin inhibitor was investigated, but turned out to yield no enhanced efficiency.

Acknowledgments We thank Robert Best for providing his RDC singular value decomposition code and Kevin Cowtan for advice on how to compute gradients using the Clipper libraries. In addition, contributions from Markus Christen, Fabian Freitag, Bruno Horta, Bettina Keller, Chris Oostenbrink, Christine Peter, Maxime Richard and Alex de Vries are gratefully acknowledged. This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-121913 of the Swiss National Science Foundation, and by grant number 228076 of the European Research Council, which is gratefully acknowledged.

References

Artymiuk PJ, Blake CCF, Rice DW, Wilson KS (1982) The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 angstroms resolution. Acta Crystallogr Sect B 38:778–783

Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog NMR Spec 46:23–61

Brünger AT (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355:472–474

Černy V (1985) A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm. J Optim Theory Applic 45:41–51

Chapman MS (1995) Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. Acta Cryst A51:69–80

Christen M, van Gunsteren WF (2007) On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review. J Comput Chem 29:157–166

Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, Geerke DP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem 26:1719–1751

Christen M, Keller B, van Gunsteren WF (2007) Biomolecular structure refinement based on adaptive restraints using local-elevation simulation. J Biomol NMR 39:265–273

Cowtan K (2003) The clipper C++ libraries for X-ray crystallography. IUCr Comput Comm Newslett 2:4–9

de Casteljau P (1959) Courbes à poles. INPI

de Vlieg J, Boelens R, Scheek RM, Kaptein R, van Gunsteren WF (1986) Restrained molecular dynamics procedure for protein tertiary structure determination from NMR data: a lac repressor headpiece structure based on information on J-coupling and from presence and absence of NOE's. Isr J Chem 27:181–188

Debye P (1913) Interference of X-rays and heat movement. Annalen der Physik 43(1):49–95

Dolenc J, Missimer JH, Steinmetz MO, van Gunsteren WF (2010) Methods of NMR structure refinement: molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1. J Biomol NMR 47:221–235

Frühwirth T, Fritz G, Freiberger N, Glatter O (2004) Structure and order in lamellar phases determined by small-angle scattering. J Appl Cryst 37:703–710

Galassi M, Davies J, Theiler J, Gough B, Jungman G, Booth M, Rossi F (2003) Gnu scientific library: reference manual. Network Theory Ltd, UK

Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. 253(5020):657–661

Gros P, van Gunsteren WF, Hol WGJ (1990) Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. Science 249:1149–1152

Higman VA, Boyd J, Smith LJ, Redfield C (2010) Residual dipolar couplings: are multiple independent alignments always possible?. J Biomol NMR 49:53–60

Hockney RW, Eastwood JW (1981) Computer simulation using particles. McGraw-Hill, New York

Horta B, Peric-Hassler L, Hünenberger PH (2010) Interaction of the disaccharides trehalose and gentiobiose with lipid bilayers: a

comparative molecular dynamics study. J Mol Graph Model 29:331–346
Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. J Comput Aided Mol Design 8:695–708
Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Jpn 65: 1604–1608
Irbäck A, Potthast F (1995) Studies of an off-lattice model for protein folding: sequence dependence and improved sampling at finite temperature. J Chem Phys 103:10298–10305
Jones TA, Zuo JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst 47A: 110–119
Karplus M (1959) Interpretation of the electron-spin resonance spectrum of the methyl radical. J Chem Phys 30:11–15
Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680
Kunz A-PE, Eichenberger AP, van Gunsteren WF (2011) A simple, efficient, polarisable molecular model for liquid carbon tetrachloride. Mol Phys 109:365–372
Lide DR (2007) CRC handbook of chemistry and physics, 88th ed. CRC, Boca Raton
Misawa M (1989) Temperature-dependence of structure of liquid carbon-tetrachloride measured by pulsed neutron scattering. J Chem Phys 91:5648–5654
Nanzer AP, Torda AE, Bisang C, Weber C, Robinson JA, van Gunsteren WF (1997) Dynamical studies of peptide motifs in the plasmodium falciparum circumsporozoite surface protein by restrained and unrestrained MD simulations. J Mol Biol 267: 1012–1025
Okabe T, Kawata M, Okamoto Y, Mikami M (2001) Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. Chem Phys Lett 335:435–439
Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comp Chem. 25:1656
Oostenbrink C, Soares TA, van der Vegt NFA, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. Eur Biophys J 34:273–284
Pardi A, Billeter M, Wüthrich K (1984) Calibration of the angular dependence of the amide proton-C alpha proton coupling constants, $^3J_{HN} - H_\alpha$, in a globular protein. Use of $^3J_{HN} - H_\alpha$ for identification of helical secondary structure. J Mol Biol 741–751
Peter C, Daura X, van Gunsteren WF (2001) Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. J Biomol NMR 20:297–310
Pusztai L, McGreevy RL (1997) The structure of liquid CCl4. Mol Phys 90:533–539
Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys 23:327–341
Schmid N, Bolliger C, Smith LJ, van Gunsteren WF (2008) Disulfide bond shuffling in bovine α-lactalbumin: MD simulation confirms experiment. Biochemistry 47(46):12104–12107
Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 40(7):843–856. doi:10.1007/s00249-011-0700-9
Schuler LD, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comput Chem 22:1205–1218
Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, Frank S, Alexandrescu AT, Kammerer RA (2007) Molecular basis of coiled-coil formation. PNAS 104:7062–7067
Sugita Y, Kitao A, Okamoto Y (2000) Multidimensional replica-exchange method for free-energy calculations. J Chem Phys 113: 6042–6052
Torda AE, Scheek RM, van Gunsteren WF (1989) Time-dependent distance restraints in molecular dynamics simulations. Chem Phys Lett 157:289–294
Torda AE, Brunne RM, Huber T, Kessler H, van Gunsteren WF (1993) Structure refinement using time-averaged J-coupling constant restraints. J Biomol NMR 3:55–66
Tropp J (1980) Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: the effect of fluctuating internuclear distances. J Chem Phys 72:6035–6043
van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Hochschulverlag AG, ETH Zurich
van Hove L (1954) Correlations in space and time and Born approximation scattering in systems of interacting particles. Phys Rev 95:249
Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 194:531–544
Waasmaier D, Kirfel A (1995) New analytical scattering-factor functions for free atoms and ions. Acta Cryst A51:416–431
Windsor CG (1973) Chemical applications of thermal neutron scattering. Oxford University Press, Oxford
Wlodawer A, Nachman J, Gilliland GL, Gallagher W, Woodward C (1987) Structure of form III crystals of bovine pancreatic trypsin inhibitor. J Mol Biol 198:469–480
Wüthrich K, Billeter M, Braun W (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J Mol Biol 169:949–961
Zuiderweg ERP, Scheek RM, Boelens R, van Gunsteren WF, Kaptein R (1985) Determination of protein structures from nuclear magnetic resonance data using a restrained molecular dynamics approach: the lac repressor DNA binding domain. Biochimie 67:707–715