An improved structural characterisation of reduced French bean plastocyanin based on NMR data and local-elevation molecular dynamics simulation

Eur Biophys J (2012) 41:579–595
DOI 10.1007/s00249-012-0824-6

ORIGINAL PAPER

EBSA
Biophysics in Europe

An improved structural characterisation of reduced French bean plastocyanin based on NMR data and local-elevation molecular dynamics simulation

Denise Steiner · Wilfred F. van Gunsteren

Received: 18 January 2012 / Revised: 27 April 2012 / Accepted: 10 May 2012 / Published online: 16 June 2012

Abstract Deriving structural information about a protein from NMR experimental data is still a non-trivial challenge to computational biochemistry. This is because of the low ratio of the number of independent observables to the number of molecular degrees of freedom, the approximations involved in the different relationships between particular observable quantities and molecular conformation, and the averaged character of the experimental data. For example, protein ^3J-coupling data are seldom used for structure refinement because of the multiple-valuedness and limited accuracy of the Karplus relationship linking a ^3J-coupling to a torsional angle. Moreover, sampling of the large conformational space is still problematic. Using the 99-residue protein plastocyanin as an example we investigated whether use of a thermodynamically calibrated force field, inclusion of solvent degrees of freedom, and application of adaptive local-elevation sampling that accounts for conformational averaging produces a more realistic representation of the ensemble of protein conformations than standard single-structure refinement in a non-explicit solvent using restraints that do not account for averaging and are partly based on non-observed data. Yielding better agreement with observed experimental data, the protein conformational ensemble is less restricted than when using standard single-structure refinement techniques, which are likely to yield a picture of the protein which is too rigid.

Keywords Structure refinement · Molecular dynamics simulation · Local-elevation sampling · ^3J-coupling · NMR · Time-averaging

Electronic supplementary material The online version of this article (doi:10.1007/s00249-012-0824-6) contains supplementary material, which is available to authorized users.

D. Steiner · W. F. van Gunsteren (✉)
Laboratory of Physical Chemistry, ETH, Swiss Federal Institute of Technology, 8093 Zurich, Switzerland
e-mail: wilfred.vangunsteren@igc.phys.chem.ethz.ch

Introduction

Structural information about biomolecules such as proteins, DNA, RNA, carbohydrates, and lipids is essential to understanding their involvement in biomolecular processes in the cell, but is not very easy to obtain with high accuracy for any particular biomolecule. This is due to a variety of reasons: their size, their heterogeneity of composition, the relatively small free energy differences that characterise different molecular conformations and mixtures, and the atomic dimensions combined with the great variety of time scales governing their dynamics. X-ray, electron, or neutron diffraction techniques are able to produce representations at the atomic level of biomolecules in the solid state, and spectroscopic techniques such as NMR, CD, IR, Raman, and fluorescence spectroscopy can be used to obtain, albeit less extensive, structural information under more physiological, i.e. relevant conditions. Such techniques measure one or more particular observable quantities Q which depend on the molecular coordinates $\mathbf{r}^N \equiv (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ and momenta $\mathbf{p}^N$ of the N atoms of the molecule. Because of the conformational variability that is governed by the laws of statistical mechanics, any observable $Q(\mathbf{r}^N)$ that is a function of conformation $\mathbf{r}^N$ will also have a distribution $P(Q(\mathbf{r}^N))$ of Q-values. In general, experimental techniques only measure an average over space and time, $\langle Q \rangle_{\exp}$, over this distribution, not the distribution itself.

The challenge of deriving structural information about biomolecules, or, in the ideal case, deriving biomolecular

structures, from experimental data is not trivial, for the following six reasons:

1. The function $Q(\mathbf{r}^N)$ that yields values of the observable Q as function of molecular conformation $\mathbf{r}^N$ may not be precisely known. For example, an accurate calculation of NMR chemical shifts as a function of molecular conformation requires sophisticated quantum-chemical methodology, and still does not reach the precision obtained experimentally. Or, the relationship between a ^3J-coupling constant and the corresponding torsional angle θ is generally approximated by use of the Karplus relationship (Karplus 1959, 1963) with empirically derived coefficients *a, b*, and *c*, which render this function $^3J(\theta)$ rather inaccurate. However, for particular observables Q, e.g. X-ray diffraction intensities, the relationship $Q(\mathbf{r}^N)$ is relatively well known and not too expensive to evaluate.
2. To derive a molecular conformation $\mathbf{r}^N$ from a measured Q-value one needs the inverse function $\mathbf{r}^N(Q)$ of the function $Q(\mathbf{r}^N)$. For X-ray diffraction, this poses no problem, because the structure factors are related by Fourier transform to the electron density. For inversion of a chemical shift calculation, however, one would need to invert the quantum-chemical calculation, a clearly impossible task.
3. Even if the inverse function $\mathbf{r}^N(Q)$ of the function $Q(\mathbf{r}^N)$ is known, it may be multiple-valued, i.e. more than one $\mathbf{r}^N$-value corresponds to one Q-value. This is, e.g., the case when calculating torsional-angle θ-values from NMR ^3J-coupling constants using the Karplus relationship $^3J(\theta)$. Its inverse $\theta(^3J)$ is multiple-valued.
4. Because of the averaging inherent in the measurement it is usually not possible to determine the Q-distribution $P(Q(\mathbf{r}^N))$ or the underlying conformational distribution $P(\mathbf{r}^N)$ from $\langle Q(\mathbf{r}^N)\rangle_{\text{exp}}$. If the conformational distribution $P(\mathbf{r}^N)$ is characterised by a single conformation, as is approximately the case for proteins in crystalline environment, a single conformation $\mathbf{r}^N(\langle Q\rangle)$ may serve as a useful approximation to the conformational distribution $P(\mathbf{r}^N)$. However, if different molecular conformations $\mathbf{r}^N$ contribute significantly to the average $\langle Q(\mathbf{r}^N)\rangle$, as is often the case for observables Q measurable by NMR, the conformation $\mathbf{r}^N(\langle Q\rangle)$ derived from the measured averages $\langle Q\rangle$ may be very unphysical, i.e. may have a negligible Boltzmann weight in the conformational ensemble, and thus will not be representative of it.
5. Experimentally measured $\langle Q\rangle$-values are of finite accuracy, i.e. the accuracy of X-ray or NMR signal intensities may vary depending on a variety of experimental parameters.
6. The number N_Q^{exp} of observable quantities that can be measured for a biomolecular system is generally very much smaller than the number N_{dof} of (atomic) degrees of freedom of the system, e.g. the number of torsional angles of the macromolecule. This makes the problem of determining the conformational distribution from a set of $\langle Q(\mathbf{r}^N)\rangle_{\text{exp}}$-values highly underdetermined. Combining alternative sets of experimental data pertaining to one system could improve the situation, provided these data do not represent inconsistent information, e.g. because of measurement under different thermodynamic conditions or over different time scales (van Gunsteren et al. 2008).

These issues have been discussed in the literature for as long as information on protein structure has been derived from experimental NMR data (e.g. Hoch et al. 1991; Jardetzky and Finucane 2001). Jardetzky (1980) discussed the difference between the average $\langle Q(\mathbf{r}^N)\rangle$ and $Q(\langle \mathbf{r}^N\rangle)$ from quantities Q observable by NMR that have a non-linear dependence $Q(\mathbf{r}^N)$ on $\mathbf{r}^N$. The issue of using a conformational Boltzmann-weighted ensemble when averaging Q in protein structure refinement was raised in 1989 for time-averaged refinement (Torda et al. 1989) and a few years later for ensemble-averaging refinement based on NMR NOE data (Scheek et al. 1991; Fennen et al. 1995). The approximate nature of the Karplus relationship between a ^3J-coupling and the corresponding torsional angle became the focus of investigations in the 1990s (Torda et al. 1993; Brüschweiler and Case 1994; Schmidt et al. 1999), while during the past decade approximations involved in the use of residual dipolar couplings (RDCs) in structure refinement have been investigated, as discussed by Lange et al. (2008) and Salmon et al. (2011), and the use of chemical shifts in structure determination was considered again (Harvey and van Gunsteren 1993; Cavalli et al. 2007).

Some of the six challenges can be met by the use of molecular dynamics (MD) simulation techniques, which enable Boltzmann sampling of conformational space based on a force field that mimics the atomic interactions at the molecular level. Use of MD simulation allows for appropriate averaging and enhances the ratio of the number of values of observable quantities over the number of degrees of freedom, because the (bio)molecular force fields are based on, i.e. are parameterised against, a wide range of experimental data. The use of, be it primitive, force fields has always been a necessary ingredient of methodology to derive biomolecular structure from experimental data (Havel et al. 1979; Hendrickson 1985). Since the 1980s MD simulation has been used to search conformational space for low-energy conformers, first using a non-physical force field energy term that represents NMR observables (Kaptein et al. 1985), and later using one that represents X-ray

diffraction intensities (Brünger et al. 1987). The sampling of conformational space can be biassed towards obtaining a particular $\langle Q \rangle_{\mathrm{exp}}$-value by restraining the (running) simulated average $\langle Q \rangle_{\mathrm{sim}}$-value towards the given $\langle Q \rangle_{\mathrm{exp}}$-value (Torda et al. 1989). In this way an ensemble compatible with the $\langle Q \rangle_{\mathrm{exp}}$-values can be generated.

Note, however, that $\langle Q \rangle_{\mathrm{exp}}$ denotes an observable quantity, that is, a property that can be measured directly. Such primary experimental data should not be confused with secondary, non-observed experimental data, Q^{NO}, that is, data derived from $\langle Q \rangle_{\mathrm{exp}}$ by applying a given procedure, f, based on a variety of assumptions and approximations: $Q^{\mathrm{NO}} = f(\langle Q \rangle_{\mathrm{exp}})$. For example, peak location and intensity from X-ray diffraction or NMR spectroscopic measurements are primary, observed data, whereas molecular structures, NMR order parameters, etc., are secondary, i.e. derived, quantities. Such secondary, non-observed, "experimental" quantities reflect, at least partly, the approximations and assumptions associated with the conversion procedure f and may, in reality, contain little experimental information (Glättli and van Gunsteren 2004). Clearly, when coupling or restraining a simulation to a set of $\langle Q \rangle_{\mathrm{exp}}$-values to ensure that the conformational distribution satisfies $\langle Q \rangle_{\mathrm{sim}} = \langle Q \rangle_{\mathrm{exp}}$, only primary experimental data should be used. Use of secondary data, for example hydrogen-bond or torsional-angle restraints, may restrict the sampling artificially and distort the proper Boltzmann weighting of the conformational ensemble. Yet, because of the low ratio of the number of observables to the number of degrees of freedom in protein structure determination based on NMR data, such secondary, non-observed, "experimental" data are often used in protein structure refinement, which leads, inevitably, to reduced accuracy of the protein structures obtained.

The use of MD simulation based on atomic level force fields also has its caveats. First, a force field, no matter how sophisticated or well calibrated by use of theoretical and experimental data, is of limited accuracy. Second, available computing power still severely limits the extent of sampling of conformational space for a macromolecule. Yet, the progress made with both these over recent decades has made it possible to enhance significantly the accuracy with which protein structure can be derived.

In this study we investigated this progress by applying a recently proposed technique for protein structure refinement based on NMR data to the 99-residue protein reduced French bean plastocyanin (Fig. 1). Its structure was determined almost two decades ago on the basis of NMR data: 1120 NOE intensities, 59 backbone $^3J_{H_N H_\alpha}$- and 108 side-chain $^3J_{\alpha\beta}$-couplings (Moore et al. 1991) (Fig. 1). The NOE intensities were represented as NOE atom–atom distance bounds. For the determination of torsional-angle restraints, the ^3J-coupling constants were converted to secondary (non-observed) data by specifying allowed ranges for 103 ϕ- and χ_1-torsional angles. Of the 108 measured $^3J_{\alpha\beta}$-couplings, 37 were not used in the structure determination because of a lack of indication of the preferred χ_1 rotamer conformations. In addition, hydrogen-exchange data were converted to secondary (non-observed) data by specifying 21 backbone–backbone hydrogen-bond restraints. The structure calculations involved distance geometry calculations to generate a set of structures, which were consecutively refined by use of molecular dynamics temperature annealing without explicit solvent, on the basis of a modified AMBER force field (Weiner et al. 1986; Lee et al. 1989; Gippert et al. 1990). This resulted in a set of 16 NMR model structures that largely satisfied the imposed restraints, but did not wholly comply with all the measured (primary) data. This was attributed to inadequate sampling of conformational space in relatively unconstrained regions of the protein, and to the inadequate representation of $^3J_{\alpha\beta}$-couplings in the conformational sampling of χ_1-torsional angles and possible artifacts arising from the force field used (Moore et al. 1991).

Because of the ample availability of NMR data and the careful description of the determination of the structure of plastocyanin by Moore et al. (1991), this molecule is an appropriate test case for investigating the accuracy that can be achieved by use of more recently developed force fields and sampling methodology:

1. Instead of the AMBER force field (Weiner et al. 1986; Lee et al. 1989; Gippert et al. 1990) developed in the 1980s we use the relatively recent GROMOS force field parameter sets 45B3 (Schuler et al. 2001) and 53A6 (Oostenbrink et al. 2004) for the vacuum and water simulations, respectively, which were obtained by calibrating against thermodynamic (free energy, enthalpy, density) data for small molecules (Oostenbrink et al. 2004).
2. Instead of structure refinement without explicit solvent, which ignores some solvent effects, we use explicit water molecules and periodic boundary conditions, which also enable constant pressure simulation.
3. Instead of simulated temperature annealing we use local-elevation biassing to enhance the sampling of side-chain conformations when both the instantaneous and the averaged $^3J_{\alpha\beta}$-coupling constants calculated from the MD simulation do not match the experimentally measured values.
4. Instead of (instantaneously) applying restraints to every molecular configuration, thereby ignoring the averaged nature of the measured observables, we use averaged quantities $\langle Q \rangle_{\mathrm{sim}}$ for restraining or biassing.

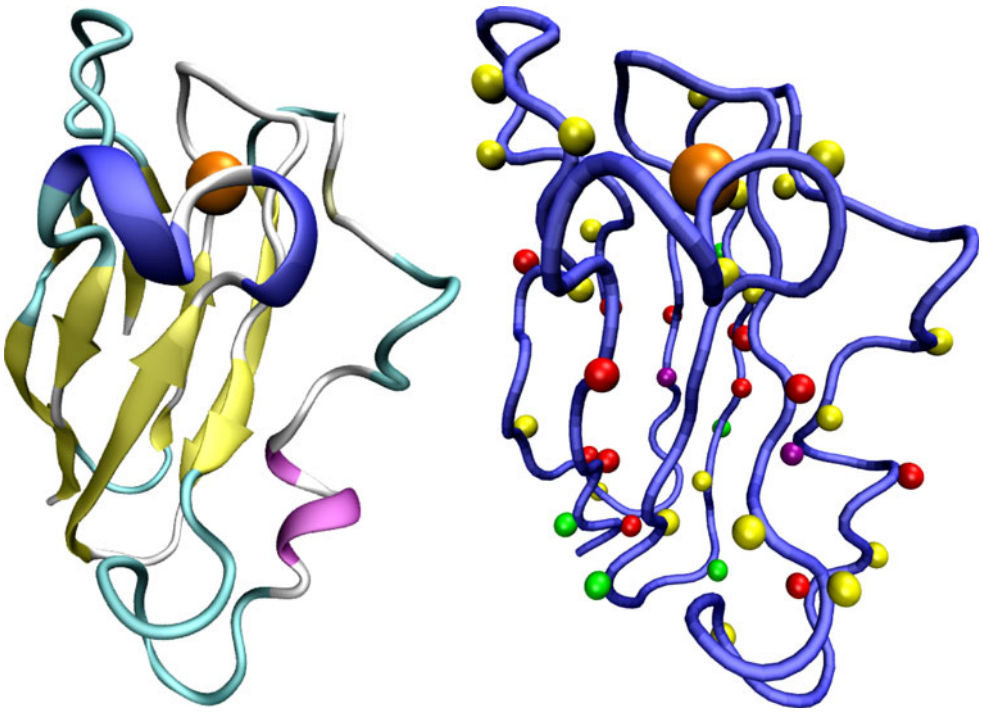

Fig. 1 *Left* schematic representation of the last of the 16 NMR model structures of plastocyanin (Moore et al. 1991), with secondary structure (*purple* α-helix, *blue* 3_{10}-helix, *yellow* β-strand) and the Cu ion in *orange*. *Right* tube representation of the backbone of plastocyanin. The residues for which $^3J_{\alpha\beta}$-couplings are used for restraint are shown as *balls*, VAL in *red*, ILE in *purple*, THR in *green*, and all the amino acid residues with two stereospecifically assigned H_β in *yellow*. The Cu ion is indicated in *orange*

5. Instead of using, apart from primary NOE data, secondary (non-observed) data, for example hydrogen bonds and torsional-angle value ranges in restraints, we use primary data only, i.e. 957 NOE distance bounds and 62 $^3J_{\alpha\beta}$-coupling constants for restraining or biassing. Some of the 1120 NOE bounds involve non-stereospecifically assigned H_{β_2} and H_{β_3} atoms with the same value for the NOE bound. These pairs are represented by one restraint to the pseudo-atom position between H_{β_2} and H_{β_3}. The 59 $^3J_{H_NH_\alpha}$-couplings had only been classified as larger than 9 Hz or smaller than 6 Hz (Moore et al. 1991) and are, therefore, not used as restraints. Of the 108 measured $^3J_{\alpha\beta}$-values, 46 are not used in the structure determination because of a lack of stereospecific assignment. The distribution of the $^3J_{\alpha\beta}$-couplings used as restraints over the protein is shown in Fig. 1.

These differences reflect the development of computational methodology with regard to force-field accuracy and sampling efficiency for refinement of the structure of a protein and of computing power to enable inclusion of solvent degrees of freedom and conformational ensembles. The focus of the analysis is on the use of $^3J_{\alpha\beta}$-couplings in the structure refinement based on local-elevation sampling (Huber et al. 1994) of the χ_1 dihedral angle degrees of freedom (Christen et al. 2007). Recently, Markwick et al. (2009) applied accelerated MD (Hamelberg et al. 2004), a method based on the same idea as local-elevation MD (Huber et al. 1994), to analyse backbone torsional-angle distributions of the proteins GB3 and ubiquitin using ^3J-couplings pertaining to the backbone φ-angle.

Method

The simulations were carried out with the GROMOS biomolecular simulation software (Schmid et al. 2011). For the simulations in vacuo, the 45B3 GROMOS force field (Schuler et al. 2001) was used, and for the simulations in explicit solvent, the 53A6 GROMOS force field (Oostenbrink et al. 2004) was used with the SPC (simple point charge; Berendsen et al. 1981) water model. The lysines and histidines present in the molecule were protonated. The resulting charge of the Cu(I)–plastocyanin was −8.5 e (with the half charge originating from one cysteine), thus eight Na^+ counterions were added to the water simulations to obtain a nearly neutral solution. The vacuum simulation was performed without any counterions, because in the 45B3 force field the charged side chains (Glu, Asp, Lys, Cys) and chain termini are neutralised.

As starting structure for the simulations, the last of the 16 NMR model structures described by Moore et al. (1991) was taken from the Protein Data Bank (Berman et al. 2000, 2003) (PDB ID:9PCY). In the vacuum simulations, the structure was energy minimised followed by thermalisation, which involved position-restraining the protein atoms. Initial velocities were generated from a Maxwell–Boltzmann distribution. The simulation temperature was increased from 50 K in steps of 50 K up to 298 K while

simultaneously reducing the position-restraining coupling constant from 25,000 to 0 kJ mol^{-1} nm^{-2} in logarithmic steps—25,000, 2,500, 250, 25, 2.5, and 0 kJ mol^{-1} nm^{-2}. For every step, simulations of 10 ps were performed. The simulations were continued for 1 ns at 298 K and the trajectories were used for analysis.

For the water simulations, the energy-minimised PDB structure was introduced into a truncated octahedron SPC water box of 6.3 nm edge length containing 3,553 water molecules. Periodic boundary conditions were used for the simulations in solvent. After another energy minimisation and thermalisation as described above, the starting structure for the unrestrained MD simulation in water was obtained. The simulations were conducted at a constant temperature of 298 K using the weak coupling method (Berendsen et al. 1984) and coupling the solute (protein and Cu ion) and solvent degrees of freedom separately to the heat bath with a coupling time $\tau_T = 0.1$ ps and at a constant pressure of 1 atm using $\tau_P = 0.5$ ps. In both types of simulation a triple-range cutoff scheme was used for non-bonded interactions in which, at every time step, interactions within a short-range cutoff of 0.8 nm were calculated from a pair list generated every 5th time step. At every 5th time step interactions between 0.8 and 1.4 nm were updated. A reaction field approach (Barker and Watts 1973; Tironi et al. 1995) and a dielectric permittivity of 61 (Heinz et al. 2001) for water were used for electrostatic interactions outside a 1.4 nm cutoff distance. The equations of motion were integrated with a step-size of 2 fs by applying the leap-frog scheme (Hockney and Eastwood 1981). The SHAKE algorithm (Ryckaert et al. 1977) was used for constraining all bonds of the protein and water and the bond angle of the water molecules. Different restraining functions were used for the NOEs and the $^3J_{\alpha\beta}$-couplings.

$^3J_{\alpha\beta}$-couplings depend on torsional angles θ between H_α–C_α–C_β–H_β via the Karplus relationship (Fig. S1):

$$J(\theta(t)) = a\cos^2\theta(t) + b\cos\theta(t) + c. \tag{1}$$

Since aliphatic hydrogens are not explicitly represented in the GROMOS force fields, the χ_1 torsional angle N–C_α–C_β–C_γ is used, which differs by a phase shift δ from the angle θ (van Gunsteren et al. 1996):

$$\chi_1 = \theta + \delta. \tag{2}$$

The value of δ is either $-120°$ or $0°$, depending on whether the hydrogen is H_{β_2} or H_{β_3} (Fig. S1).

To bias the sampling in a MD simulation toward a particular measured value $\langle {}^3J \rangle_{\text{exp}} = {}^3J^0$, a penalty function V^{restr} can be added to the physical force field term V^{phys} for the potential energy:

$$V\left(\mathbf{r}^N(t)\right) = V^{\text{phys}}\left(\mathbf{r}^N(t)\right) + V^{\text{restr}}\left(\mathbf{r}^N(t)\right). \tag{3}$$

In the case of $^3J_{\alpha\beta}$-coupling restraining, a time-averaging and local-elevation biassing method proposed by Christen et al. (2007) was applied. The restraining potential energy function V_k^{Jres} for the kth $^3J_{\alpha\beta}$-value related to the torsional angle χ_k is built up by N_{le} (here 36) local-elevation terms (Huber et al. 1994):

$$V_k^{\text{Jres}}\left(\chi_k\left(\mathbf{r}^N(t)\right)\right) = \sum_{i=1}^{N_{\text{le}}} V_{ki}^{\text{le}}\left(\chi_k\left(\mathbf{r}^N(t)\right)\right), \tag{4}$$

in which the penalty terms are Gaussian functions centred around χ_{ki}^0:

$$V_{ki}^{\text{le}}\left(\chi_k\left(\mathbf{r}^N(t)\right)\right) = K_k^{\text{Jres}}\omega_{\chi_{ki}}(t)e^{\left(-\left(\chi_k(t)-\chi_{ki}^0\right)^2/2\left(\Delta\chi^0\right)^2\right)}, \tag{5}$$

with $\Delta\chi^0 = 360/N_{\text{le}}$. K_k^{Jres} is the overall penalty function force constant (0.005 kJ mol^{-1} Hz^{-4} here). $\omega_{\chi_{ki}}(t)$ is the weight function of the ith Gaussian penalty function:

$$\omega_{\chi_{ki}}(t) = t^{-1}\int_0^t \delta_{\chi_k(\mathbf{r}^N(t'))\chi_{ki}^0} V^{fb}\left({}^3J\left(\chi_k\left(\mathbf{r}^N(t')\right)\right)\right)V^{fb} \times \left(\overline{{}^3J\left(\chi_k\left(\mathbf{r}^N(t')\right)\right)}\right)\mathrm{d}t', \tag{6}$$

which is non-zero if the instantaneous $\chi_k(\mathbf{r}^N(t'))$ value is in the bin of χ_{ki}^0:

$$\delta_{\chi_k(\mathbf{r}^N(t'))\chi_{ki}^0} = \begin{cases} 1 & \text{if } \chi_{ki}^0 - \Delta\chi^0/2 < \chi_k(\mathbf{r}^N(t')) < \chi_{ki}^0 + \Delta\chi^0/2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and both the instantaneous $^3J(\chi_k(\mathbf{r}^N(t)))$ and the time-averaged $\overline{{}^3J(\chi_k(\mathbf{r}^N(t)))}$ deviate more than ΔJ^0 (1 Hz in this study) from the experimental value $^3J_k^0$:

$$V^{\text{fb}}\left({}^3J\left(\chi_k\left(\mathbf{r}^N(t)\right)\right)\right) = \begin{cases} \left({}^3J(\chi_k(\mathbf{r}^N(t))) - {}^3J_k^0 - \Delta J^0\right)^2 & \text{if } {}^3J(\chi_k(\mathbf{r}^N(t))) > {}^3J_k^0 + \Delta J^0 \\ \left({}^3J(\chi_k(\mathbf{r}^N(t))) - {}^3J_k^0 + \Delta J^0\right)^2 & \text{if } {}^3J(\chi_k(\mathbf{r}^N(t))) < {}^3J_k^0 - \Delta J^0 \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

In $V^{\mathrm{fb}}\left(\overline{{}^3J(\chi_k(\mathbf{r}^N(t)))}\right), {}^3J(\chi_k(\mathbf{r}^N(t)))$ is replaced by $\overline{{}^3J(\chi_k(\mathbf{r}^N(t)))}$, which is the exponentially damped temporal average over the course of an MD simulation:

$$\overline{{}^3J(\chi_k(\mathbf{r}^N(t)))} = \frac{1}{\tau_J(1-\exp(-t/\tau_J))} \times \int_0^t \exp\left(\frac{t'-t}{\tau_J}\right) {}^3J\left(\chi_k\left(\mathbf{r}^N(t')\right)\right)\mathrm{d}t' \quad (9)$$

with memory relaxation time τ_J, here 5 ps (Nanzer et al. 1995; Bürgi et al. 2001). Of the 108 ${}^3J_{\alpha\beta}$-couplings, 62 had been assigned to H_β or stereospecifically to H_{β_2} or H_{β_3}. These values were used for ${}^3J$-restraining (Table S1). The remaining 46 ${}^3J_{\alpha\beta}$-couplings (Table S2) were only used in the analysis. For the side-chain ${}^3J_{\alpha\beta}$-couplings the values $a = 9.5$ Hz, $b = -1.6$ Hz, and $c = 1.8$ Hz (de Marco et al. 1978) were used in the Karplus relationship (Fig. S1). The 59 ${}^3J_{H_NH_\alpha}$-couplings had been categorised as larger than 9 Hz or smaller than 6 Hz. These ${}^3J$-couplings were only used in the analysis. For these backbone ${}^3J_{H_NH_\alpha}$-couplings the values $a = 6.4$ Hz, $b = -1.4$ Hz, and $c = 1.9$ Hz (Pardi et al. 1984) were used in the Karplus relationship (Fig. S1).

For distance restraining, NOE data were used. The NOE distance bounds derived (Moore et al. 1991) from the measured NOE intensities were used as upper bounds. The distance restraining potential energy function is attractive half-harmonic:

$$V^{\mathrm{dr}}\left(\mathbf{r}^N(t)\right) = \begin{cases} 1/2 \sum_{m=1}^{N_{\mathrm{dr}}} K_m^{\mathrm{dr}}\left[r_{nn'} - r_m^0\right]^2 & \text{if} \quad r_{nn'} > r_m^0 \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

in which the sum is over the N_{dr} distance restraints and the force constant K_m^{dr} is 1,000 kJ mol^{-1} nm^{-2}. $r_{nn'}$ is the mth atom–atom distance restraint between atoms n and n' with NOE upper distance bound r_m^0. To take into account the averaged character of the measured NOE intensity, time-averaged (TAR) restraining was performed using the weighted temporal average

$$\left[\overline{r_{nn'}^{-6}(t)}\right]^{-1/6} = \left[\frac{1}{\tau_{\mathrm{NOE}}(1-\exp(-t/\tau_{\mathrm{NOE}}))} \times \int_0^t \exp\left(\frac{t'-t}{\tau_{\mathrm{NOE}}}\right) r_{nn'}^{-6}(t)\mathrm{d}t'\right]^{-1/6} \quad (11)$$

instead of $r_{nn'}$ in Eq. 10, with a coupling time $\tau_{\mathrm{NOE}} = 5$ ps (Nanzer et al. 1995; Bürgi et al. 2001). The NOE violations were calculated as:

$$\left\langle r_{nn'}^{-6}\right\rangle^{-1/6} - r_m^0. \quad (12)$$

where $\langle \ldots \rangle$ denotes an average over the MD ensembles or set of NMR model structures. For some NOE distance bounds the hydrogen atoms could not be stereospecifically assigned. In this case a pseudo-atom or averaging correction (Wüthrich et al. 1983) was added to the bound and a single pseudo-atom position between the two or more hydrogen atoms was used in the restraint (van Gunsteren et al. 1996). These pseudo-atom positions are denoted in Tables S4, S5, S12, and S13 as Q instead of H. A, B, C, D, E, and Z stand for α, β, γ, δ, ϵ and ζ respectively, indicating the position of the carbon (C) or hydrogen (H) in the amino acid. This reduced the number of NOE restraints to 957, 414 being "long-range" NOEs between residues separated by at least three other residues along the polypeptide chain.

Six different MD simulations were performed:

1. UNR_VAC: simulation of the protein in vacuo without restraints;
2. UNR_WAT: simulation of the protein in water without restraints;
3. ${}^3J$_LE_VAC: simulation of the protein in vacuo with ${}^3J$-coupling restraining using local elevation for the 62 ${}^3J_{\alpha\beta}$-couplings of Table S1;
4. ${}^3J$_LE_WAT: simulation of the protein in water with ${}^3J$-coupling restraining using local elevation for the 62 ${}^3J_{\alpha\beta}$-couplings of Table S1;
5. ${}^3J$_LE_NOE_WAT: simulation of the protein in water with ${}^3J$-coupling restraining using local elevation for the 62 ${}^3J_{\alpha\beta}$-couplings of Table S1 and with instantaneous NOE distance restraining for the 957 NOE atom pairs of Table S13; and
6. ${}^3J$_LE_NOE_TAR_WAT: simulation of the protein in water with ${}^3J$-coupling restraining using local elevation for the 62 ${}^3J_{\alpha\beta}$-couplings of Table S1 and with time-averaged NOE distance restraining for the 957 NOE atom pairs of Table S13.

The averaged quantities, NOE atom–atom distances $\left\langle r^{-6}\right\rangle^{-1/6}$ and ${}^3J$-couplings $\left\langle {}^3J\right\rangle$ calculated from the trajectories of these simulations were compared with the averages obtained from the set of 16 NMR model structures. In addition, atom-positional root-mean-square deviations (RMSD) of the trajectory structures from the initial structure, root-mean-square fluctuations of atoms, and the secondary structure content according to the program dssp (Kabsch and Sander 1983) were used to analyse the ensembles.

Results

Figures 2, 3, 4, and 5 enable comparison of the ${}^3J$-coupling and NOE data as calculated and averaged over the six simulated conformational ensembles and over the set of 16

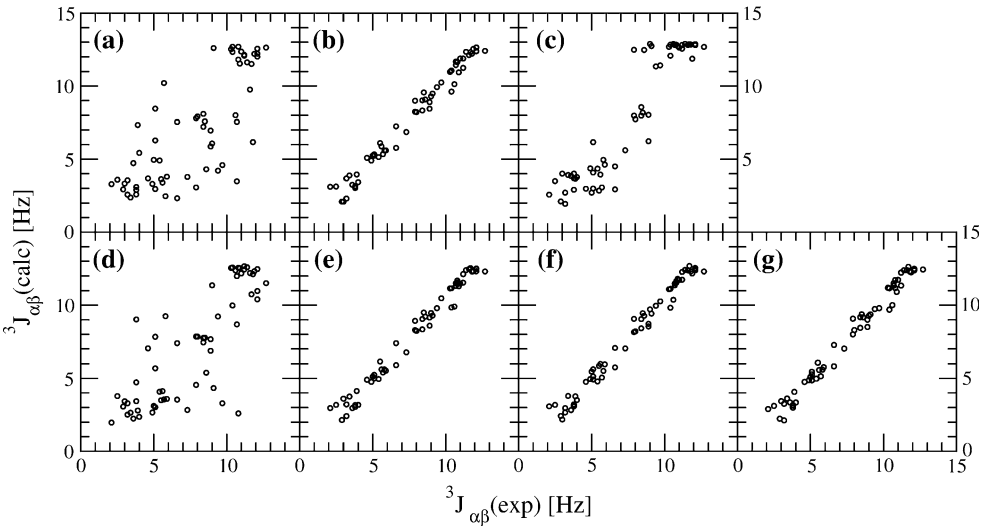

Fig. 2 Comparison of the 62 $^3J_{\alpha\beta}$-couplings that were stereospecifically assigned and could be used as restraints calculated from and averaged over each of the six different conformational MD ensembles or the set of 16 NMR model structures with those measured experimentally. *a* UNR_VAC simulation, *b* ^3J_LE_VAC simulation, *c* NMR set, *d* UNR_WAT simulation, *e* ^3J_LE_WAT simulation, *f* ^3J_LE_NOE_WAT simulation, *g* ^3J_LE_NOE_WAT simulation

NMR model structures with the corresponding measured values. In panels *a* of these figures, the results from MD simulation of the protein in vacuo without application of any restraints (UNR_VAC) are shown. For the stereospecifically assigned $^3J_{\alpha\beta}$-couplings (Fig. 2) poor correlation between simulation and experiment is observed with deviations up to 7 Hz (Table S6). For the other $^3J_{\alpha\beta}$-couplings (Fig. 3) almost no correlation is found, again with sizable deviations (Table S7). All but one of the $^3J_{H_NH_\alpha}$-couplings smaller than 6 Hz are indeed smaller than 6 Hz (Fig. 4 and Table S8), but only a few of the $^3J_{H_NH_\alpha}$-couplings that were measured to be larger than 9 Hz satisfy this lower bound in the simulation. This is not very surprising, in view of the maximum of approximately 9.7 Hz of the corresponding Karplus curve (Fig. S1). Of the 414 "long-range" NOEs, for 32 NOEs the violation is larger than 0.1 nm in the simulation (Fig. 5 and Table S12). The discrepancies between simulated and experimental data could be because of force-field deficiencies, insufficient sampling, or performing the simulation in vacuo.

In panel *d* of Figs. 2, 3, 4, and 5 the results of the MD simulation of the protein in water without application of any restraints (UNR_WAT) are shown. For the ^3J-couplings the agreement between simulation and experiment is not significantly improved by inclusion of the water degrees of freedom in the simulation, but the NOE distance bound violations are much reduced. Of the 414 "long-range" NOEs, only for eight NOEs is the violation larger than 0.1 nm in the simulation (Fig. 5 and Table S12). The discrepancies between the simulated and experimental ^3J-coupling data could be because of force-field deficiencies or insufficient sampling of the torsional-angle degrees of freedom that determine the ^3J-couplings. For the χ_1 side-chain torsional angles, in particular, energy barriers due to non-bonded repulsive interactions hindering side-chain rotation may lead to insufficiently sampled $\langle {}^3J_{\alpha\beta} \rangle$-values.

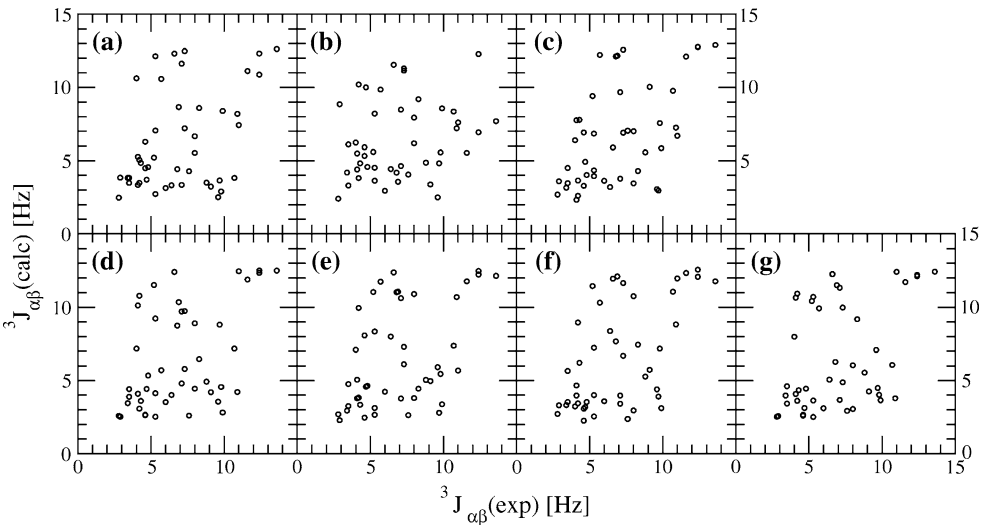

Fig. 3 Comparison of the 46 $^3J_{\alpha\beta}$-couplings that were not part of the set of $^3J_{\alpha\beta}$-coupling restraints calculated from and averaged over each of the six different conformational MD ensembles or the set of 16 NMR model structures with those measured experimentally. *a* UNR_VAC simulation, *b* ^3J_LE_VAC simulation, *c* NMR set, *d* UNR_WAT simulation, *e* ^3J_LE_WAT simulation, *f* ^3J_LE_NOE_WAT simulation, *g* ^3J_LE_NOE_TAR_WAT simulation

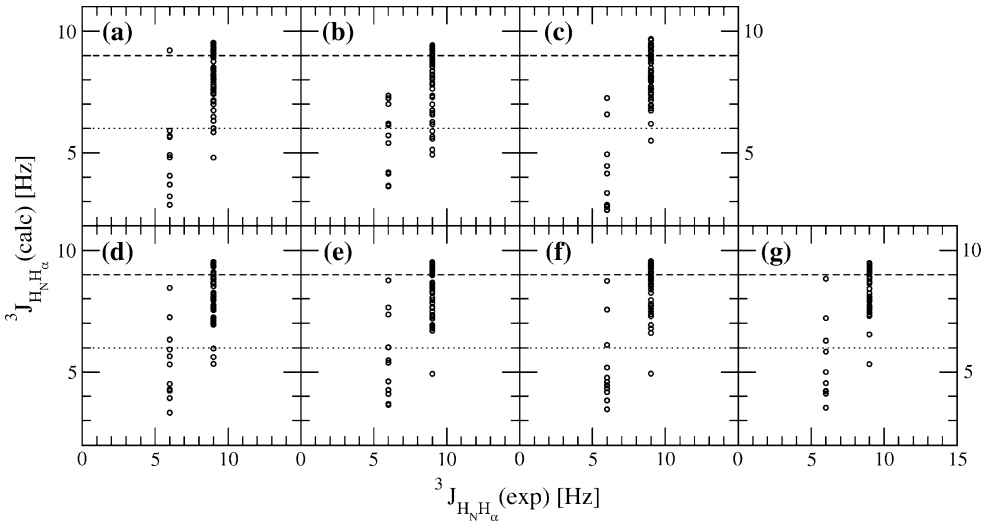

Fig. 4 Comparison with experimental bounds (smaller than 6 Hz, *dotted line*; larger than 9 Hz, *dashed line*) for the 59 $^3J_{H_NH_\alpha}$-couplings calculated from and averaged over each of the six different conformational MD ensembles or the set of 16 NMR model structures. *a* UNR_VAC simulation, *b* ^3J_LE_VAC simulation, *c* NMR set, d UNR_WAT simulation, *e* ^3J_LE_WAT simulation, *f* ^3J_LE_NOE_WAT simulation, *g* ^3J_LE_NOE_TAR_WAT simulation. The experimental $^3J_{H_NH_\alpha}$-value was set to the bounds 6 or 9 Hz

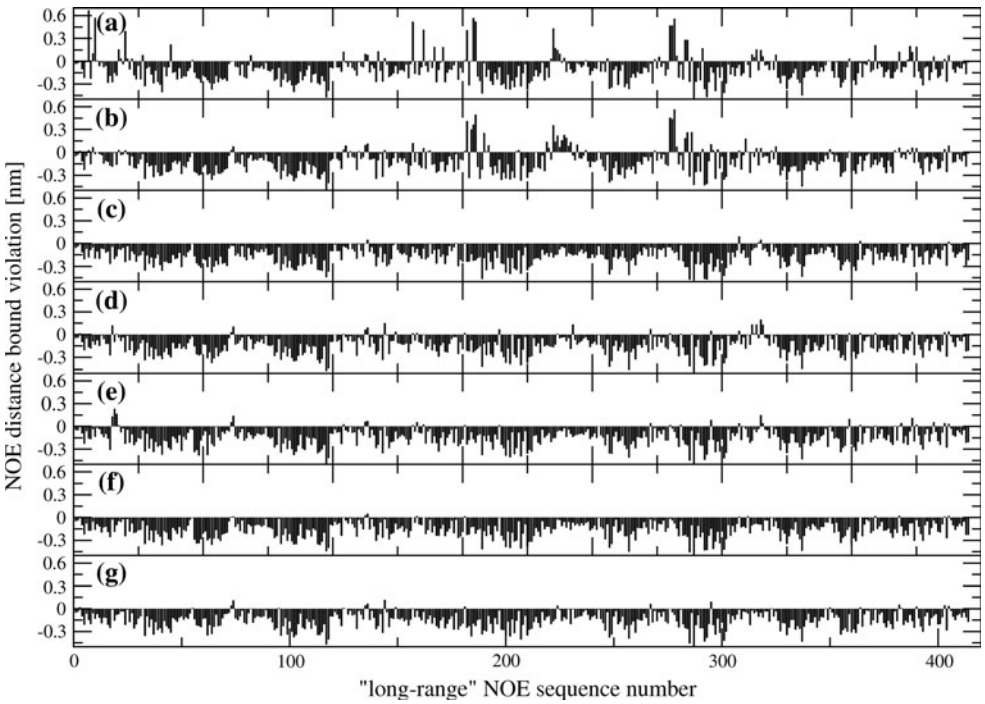

Fig. 5 Difference between the r^{-6} averaged distances and the NOE distance bounds for 414 pairs of hydrogen atoms that are "long-range" in terms of sequence separation, in each of the six different conformational MD ensembles or the set of 16 NMR model structures. *a* UNR_VAC simulation, *b* ^3J_LE_VAC simulation, *c* NMR set, *d* UNR_WAT simulation, *e* ^3J_LE_WAT simulation, *f* ^3J_LE_NOE_WAT simulation, *g* ^3J_LE_NOE_TAR_WAT simulation

The sampling of the χ_1 angles that determine the 62 stereospecifically assigned $^3J_{\alpha\beta}$-couplings can be biassed towards producing, on average, the measured ^3J-couplings by using the technique of local-elevation biassing based on adaptive ^3J-coupling restraints.

Panels *b* and *e* of Figs. 2, 3, 4, and 5 show the results of the MD simulations of the protein with application of the 62 ^3J-coupling restraints in vacuo (^{3}J_LE_VAC) and in water (^{3}J_LE_WAT), respectively. Because the restraints are applied with a flat-bottom potential energy restraining function, with a flat bottom of 2 Hz, the measured $^3J_{\alpha\beta}$-couplings are reproduced within ±1 Hz (Fig. 2 and Table S9). For the other ^3J-couplings no improvement of the deviations between simulations and experiment can be observed (Figs. 3 and 4 and Tables S10 and S11). Comparison of the NOE distance bound violations (Fig. 5) shows that, as observed before without $^3J_{\alpha\beta}$-coupling restraints, inclusion of water in the simulation reduces the discrepancies with experiment significantly.

Panels *c* of Figs. 2, 3, 4, and 5 show the ^3J-couplings and NOE distance bound violations as obtained by averaging over the set of 16 NMR model structures that were derived, by use of these data, as described in the "Introduction". The set of 62 measured and stereospecifically assigned $^3J_{\alpha\beta}$-couplings is reproduced rather approximately (Fig. 2 and Table S9), with deviations up to 4 Hz. The

other measured ^3J-coupling data are reproduced as poorly as in the simulations (Figs. 3 and 4 and Tables S10 and S11). The NOE distance bounds are basically satisfied with only a few small violations (Fig. 2 and Table S4). Compared with the simulations in which the 62 $^3J_{\alpha\beta}$-couplings were restrained, the set of 16 NMR model structures shows slightly worse agreement with experiment for the ^3J-couplings and better agreement with the NOE distance bounds. This is no surprise, because the latter were used as restraints in the determination of the set of NMR model structures, whereas they were not used as such in the simulations discussed so far. Thus the next step is to consider the MD simulations in which NOE distance restraints were used in addition to the $^3J_{\alpha\beta}$-coupling restraints.

Panels *f* and *g* of Figs. 2, 3, 4, and 5 show the results of the MD simulations of the protein in water with application of the 62 $^3J_{\alpha\beta}$-coupling restraints and the 957 NOE distance restraints either using instantaneous restraining (^{3}J_LE_NOE_WAT) or using time-averaged restraining (^{3}J_LE_NOE_TAR_WAT), respectively. The additional NOE distance restraining does not affect the agreement of the ^3J-couplings with experiment (Figs. 2, 3, and 4 and Tables S1–S3) and slightly improves the agreement with the NOE distance bounds (Fig. 5 and Tables S4, S5). Because the measurement of observables such as ^3J-couplings and NOE intensities involves averaging over time and space, we consider the simulation that involves time-averaged, instead of instantaneous, restraints as the better representation of reality. Therefore, we analyse and compare in more detail only the MD simulation ^{3}J_LE_NOE_TAR_WAT and compare its ensemble of conformations with the set of 16 NMR model structures and with the experimental $^3J_{\alpha\beta}$-coupling data.

Figure 6 shows that all MD simulations except that using $^3J_{\alpha\beta}$-coupling restraints in vacuo stay reasonably close to the initial structure, one of the 16 NMR model structures. Not surprisingly, the simulation ^{3}J_LE_NOE_WAT stays closest to the NMR model structure because its restraints are most similar to those used to derive the NMR model structure.

The secondary structure analysis shown in Figs. 7, 8, and 9 indicates that the β-strands (Sheet I: residues Leu 1 to Gly 6, Val 13 to Val 15, Glu 25 to Asn 32, and Gly 67 to Leu 74; Sheet II: residues Ser 17 to Val 21, His 37 to Asp 42, Gly 78 to Cys 84, and Met 92 to Asn 99) and two short helical elements (residues Asp 51 to Ser 56 and Cys 84 to Gly 91) are preserved in the ^{3}J_LE_NOE_WAT and ^{3}J_LE_NOE_TAR_WAT simulations and in the set of 16 NMR model structures. Thus the different types of restraint do not distort the overall structure of the protein significantly.

Global comparison of the set of 16 NMR model structures and the MD simulation ^{3}J_LE_NOE_TAR_WAT

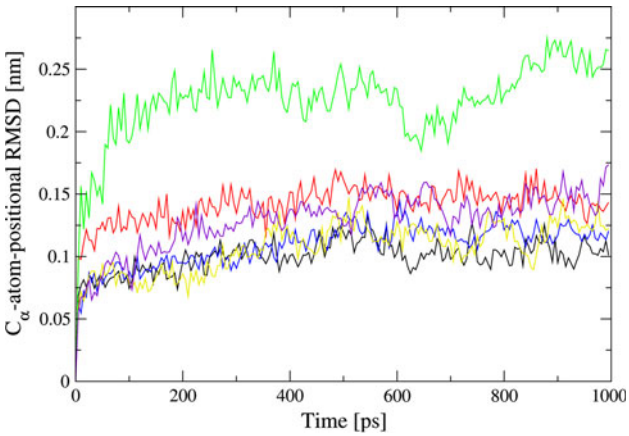

Fig. 6 C_α-atom-positional root-mean-square deviation (RMSD) from the initial structure in the MD simulations. *Red* UNR_VAC simulation, *green* ^{3}J_LE_VAC simulation, *blue* UNR_WAT simulation, *yellow* ^{3}J_LE_WAT simulation, *black* ^{3}J_LE_NOE_WAT simulation, *magenta* ^{3}J_LE_NOE_TAR_WAT simulation

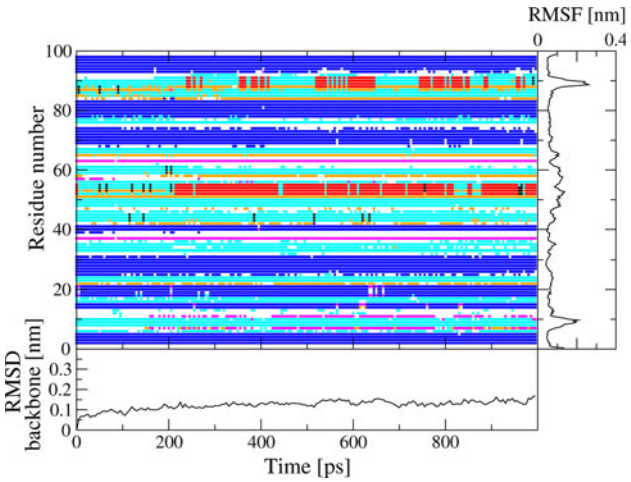

Fig. 7 Secondary structure analysis (Kabsch and Sander 1983) of the ^{3}J_LE_NOE_TAR_WAT simulation, *Black* 3_{10}-helix, *red* α-helix, *cyan* bend, *magenta* β-bridge, *blue* β-strand, *orange* turn. The *right hand panel* shows the root-mean-square fluctuation (RMSF) of the backbone (*N*, C_α, *C*) atoms. The *lower panel* shows the root-mean-square distance between the instantaneous positions of the C_α, *N*, and *C* atoms of the backbone and their positions in the initial structure

with the measured NMR data shows that both sets of conformations agree on average equally well with the experimental data, which is no surprise, because these data were used as restraints in both cases. However, comparison of individual side-chain χ_1-angle distributions and the corresponding averaged $^3J_{\alpha\beta}$-couplings reveals interesting differences. Below we analyse these for nine different side-chains that serve as examples of particular types of side-chain behaviour in protein structure refinement.

Figures 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and 20 and S2–S4 show the behaviour of the χ_1 torsional angle, the

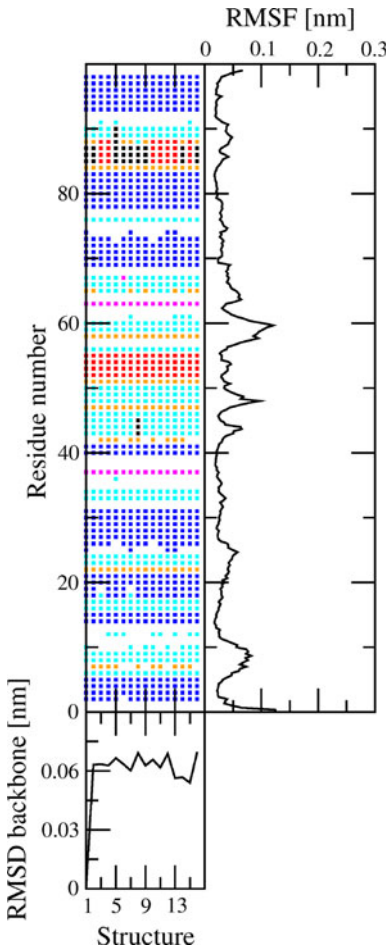

Fig. 8 Secondary structure analysis (Kabsch and Sander 1983) of the set of 16 NMR model structures. *Black* 3_{10}-helix, *red* α-helix, *cyan* bend, *magenta* β-bridge, *blue* β-strand, *orange* turn. The *right hand panel* shows the root-mean-square fluctuation (RMSF) of the backbone (*N*, C_α, *C*) atoms. The *lower panel* shows the root-mean-square distance between the instantaneous positions of the C_α, *N*, and *C* atoms of the backbone and their positions in the initial structure

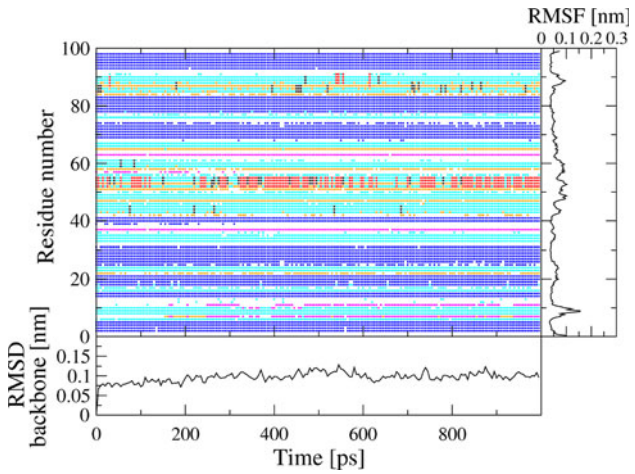

Fig. 9 Secondary structure analysis (Kabsch and Sander 1983) of the ^3J_LE_NOE_WAT simulation. *Black* 3_{10}-helix, *red* α-helix, *cyan* bend, *magenta* β-bridge, *blue* β-strand, *orange* turn. The *right hand panel* shows the root-mean-square fluctuation (RMSF) of the backbone (*N*, C_α, *C*) atoms. The *lower panel* shows the root-mean-square distance between the instantaneous positions of the C_α, *N* and *C* atoms of the backbone and their positions in the initial structure

corresponding $^3J_{\alpha\beta}$-coupling and the biassing local-elevation potential energy $V_{\mathrm{le}}(\chi_1)$ as a function of time during a simulation, together with the resulting χ_1-angle and $^3J_{\alpha\beta}$-coupling distributions and local-elevation biassing potential energy function $V_{\mathrm{le}}^{\mathrm{final}}(\chi_1)$ for the nine side-chains used as examples.

In Fig. 10, the χ_1-angle of Phe 14 serves as an example of the case in which the initial structure is such that the $^3J_{\alpha\beta_2}$-coupling agrees with the measured value of 11.9 Hz. Thus no local-elevation biassing energy function is built up which means that the simulation ^{3}J_LE_NOE_TAR_WAT (circles) yields the same distribution of χ_1-angles and $^3J_{\alpha\beta_2}$-couplings as the unrestrained simulation UNR_WAT (triangles).

In Fig. 11, the χ_1-angle of Val 50 shows, however, different behaviour for these two simulations. The unrestrained simulation yields an incorrect $^3J_{\alpha\beta}$-coupling which can be easily corrected in the biassed simulation by the build-up of a local-elevation energy function around $\chi_1 = 290°$ which drives the dihedral angle value to approximately 190° yielding a $\langle {}^3J_{\alpha\beta}\rangle$-value in better agreement with the $\langle {}^3J_{\alpha\beta}\rangle_{\mathrm{exp}}$.

Figures 12 and S2 show an example, the χ_1-angle of Glu 43, in which averaging over a wide range of χ_1-angles is needed. For both H atoms, H_{β_2} (Fig. 12) and H_{β_3} (Fig. S2), the 16 NMR model structures (squares) also show a substantial spread in $^3J_{\alpha\beta}$-couplings, but reproduce the $\langle {}^3J_{\alpha\beta}\rangle_{\mathrm{exp}}$ less well than the simulation. The averaged χ_1-angle values are quite different in each case, 195° in the simulation and 72° in the set of NMR model structures. The non-linear character of the Karplus relationship between $^3J_{\alpha\beta}$ and χ_1 is illustrated by the different shapes of the respective distributions. For cases such as this, the biassing energy function serves to enhance the sampling.

Figure 13 shows an example, the χ_1-angle of Val 3, in which the biassing energy function provides a small correction of 15° to the χ_1-angle value that is preferred by the force field used. Compared with the values of approximately 184° observed in the set of NMR model structures (squares), a slightly larger χ_1-angle of 199° leads to a reduction of 1.6 Hz in the $\langle {}^3J_{\alpha\beta}\rangle$-value and better agreement with experiment.

Figure 14 shows an example, the χ_1-angle of Val 53, in which the set of NMR model structures also predicts a too large $^3J_{\alpha\beta}$-coupling of 12.9 Hz for a χ_1-angle of 182°. In this case the GROMOS force field and the local-elevation biassing not only shift the distribution of χ_1-angle values

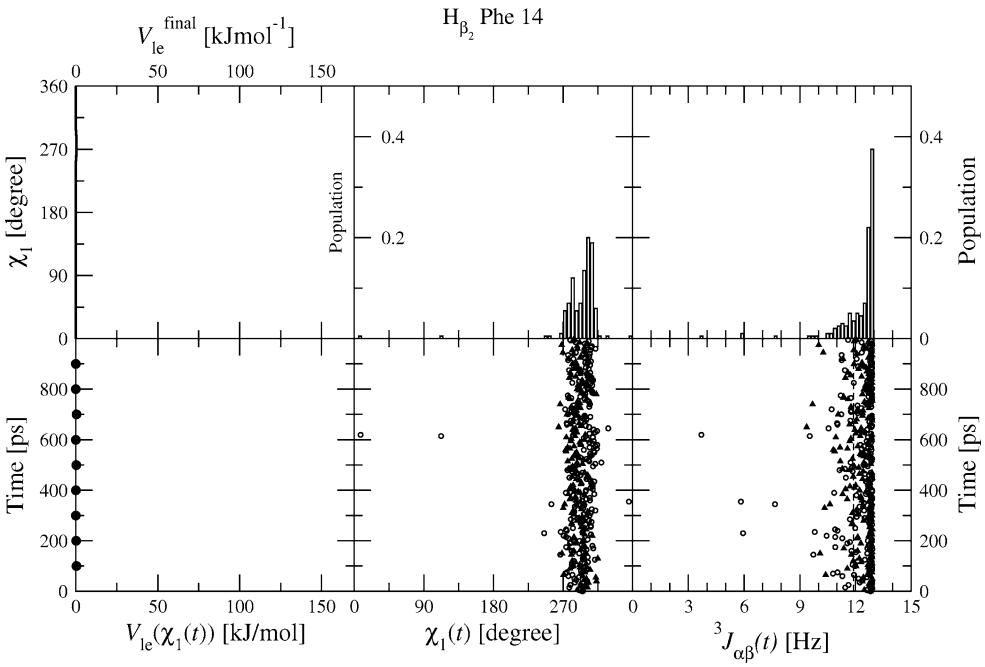

Fig. 10 Properties of the χ_1 torsional angle of Phe 14 and the corresponding $^3J_{H_\alpha H_{\beta_2}}$ in the ^{3}*J*_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* final built-up local-elevation restraining potential energy $V_{le}^{final}(\chi_1)$. *Middle* distribution of χ angles. *Right* distribution of ^{3}*J*-couplings. *Lower row*: *Left* local-elevation potential energy $V_{le}(t)$ acting on χ_1 at specific time points. *Middle* evolution of χ_1 angle in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and the UNR_WAT simulation (*triangles*). *Right* evolution of ^{3}*J*-value in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and the UNR_WAT simulation (*triangles*). The *dashed line* shows the experimental ^{3}*J*-value

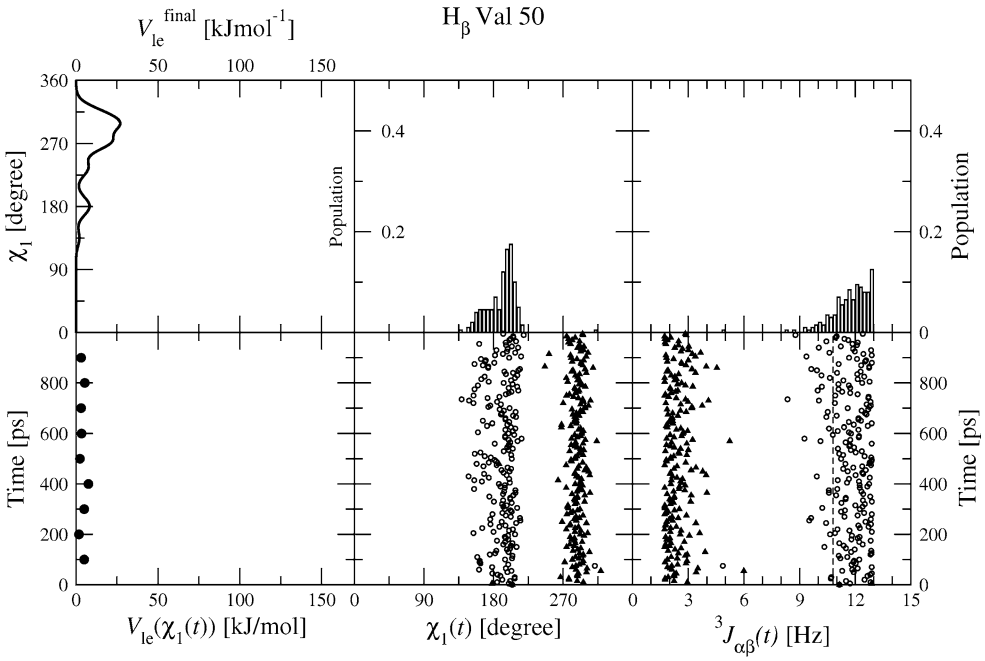

Fig. 11 Properties of the χ_1 torsional angle of Val 50 and the corresponding $^3J_{H_\alpha H_\beta}$ in the ^{3}*J*_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* final built-up local-elevation restraining potential energy $V_{le}^{final}(\chi_1)$. *Middle* distribution of χ_1 angles. *Right* distribution of ^{3}*J*-couplings. *Lower row*: *Left* local-elevation potential energy $V_{le}(t)$ acting on χ_1 at specific time points. *Middle* evolution of χ_1 angle in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and the UNR_WAT simulation (*triangles*). *Right* evolution of ^{3}*J*-value in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and the UNR_WAT simulation (*triangles*). The *dashed line* shows the experimental ^{3}*J*-value

but also induce transitions between two χ_1-angle ranges on either side of 180°. Thus the sampling is enhanced and a slight force-field deficiency is compensated for.

Until now we have considered examples of side-chain χ_1-angles that were members of the list of 62 χ_1-angles that feel a biassing local-elevation force when the discrepancy with the $\langle {}^3J_{\alpha\beta} \rangle_{exp}$ becomes too large. It comes as no surprise that for these angles the experimental $^3J_{\alpha\beta}$-couplings are well reproduced (Fig. 2). However, the behaviour of χ_1-angles that could not be restrained because of a lack of stereospecific assignment also matches the experimental data better in the local-elevation biassing simulation, as the following examples show. Because no stereospecific assignment was available for these $^3J_{\alpha\beta}$-couplings, we chose in Figs. 15, 16, 17, 18, 19, and 20 and S2–S4 to arbitrarily assign H_{β_2} to the lower $^3J_{\alpha\beta}$-coupling.

Figures 15 and 16 show an example, the χ_1-angle of Lys 54, in which the $\langle {}^3J_{\alpha\beta} \rangle$-values calculated from the set of NMR model structures show a large deviation from $\langle {}^3J_{\alpha\beta} \rangle_{exp}$ that can be greatly reduced by averaging over

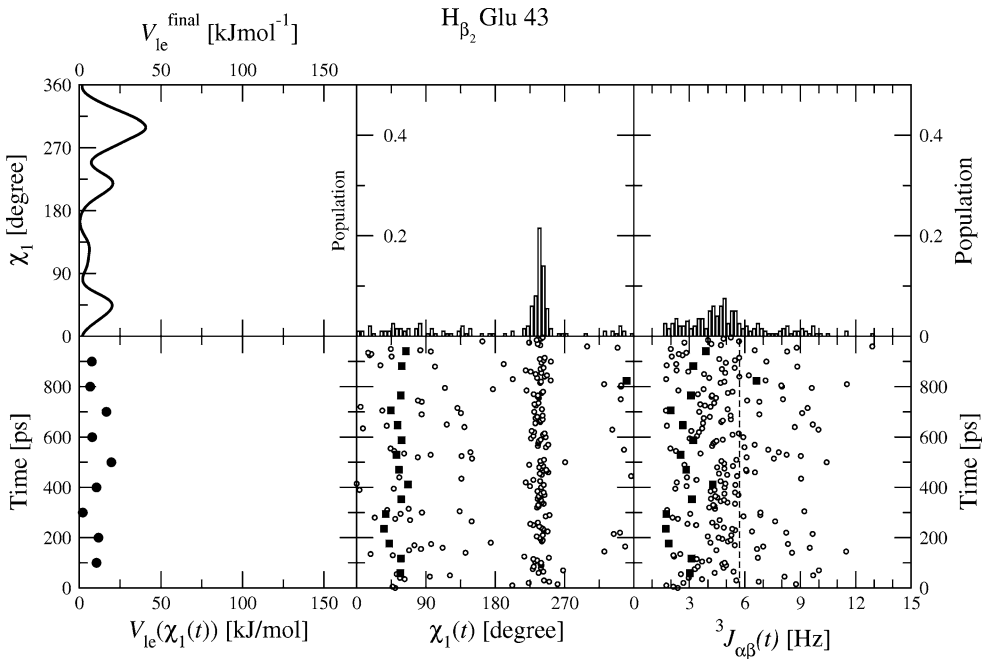

Fig. 12 Properties of the χ_1 torsional angle of Glu 43 and the corresponding $^3J_{H_\alpha H_{\beta_2}}$ in the ^3J_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* final built-up local-elevation restraining potential energy $V_{le}^{final}(\chi_1)$. *Middle* distribution of χ_1 angles. *Right* distribution of ^3J-couplings. *Lower row*: *Left* local-elevation potential energy $V_{le}(t)$ acting on χ_1 at specific time points. *Middle* evolution of χ_1 angle in the ^3J_LE_NOE_TAR_WAT simulation (*circles*) and the χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ^3J-value in the ^3J_LE_NOE_TAR_WAT simulation (*circles*) and ^3J-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ^3J-value

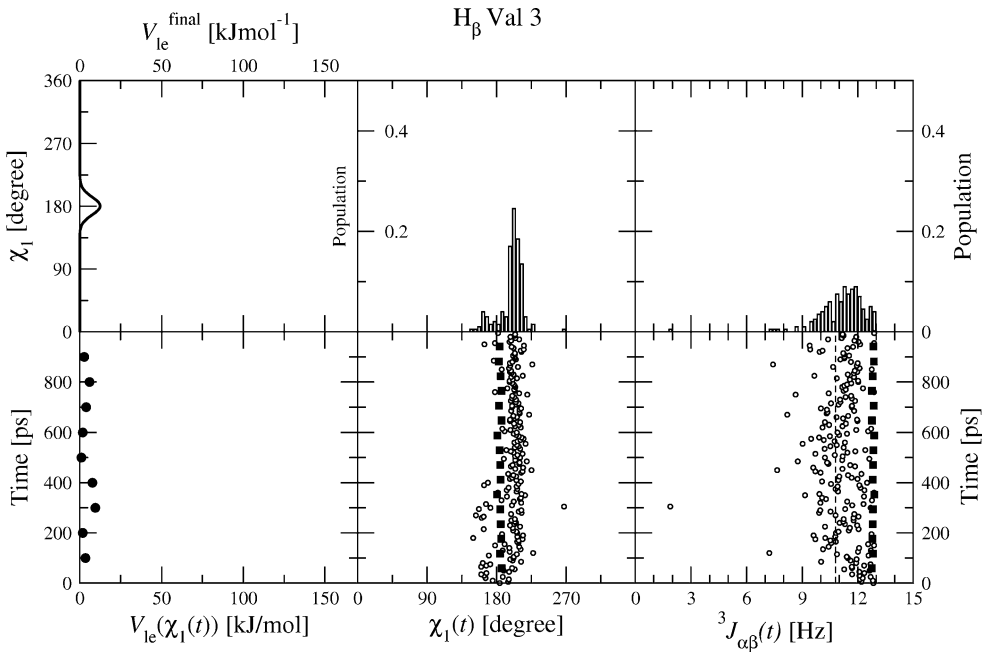

Fig. 13 Properties of the χ_1 torsional angle of Val 3 and the corresponding $^3J_{H_\alpha H_\beta}$ in the ^3J_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* final built-up local-elevation restraining potential energy $V_{le}^{final}(\chi_1)$. *Middle* distribution of χ_1 angles. *Right* distribution of ^3J-couplings. *Lower row*: *Left* local-elevation potential energy $V_{le}(t)$ acting on χ_1 at specific time points. *Middle* evolution of χ_1 angle in the ^3J_LE_NOE_TAR_WAT simulation (*circles*) and the χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ^3J-value in the ^3J_LE_NOE_TAR_WAT simulation (*circles*) and ^3J-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ^3J-value

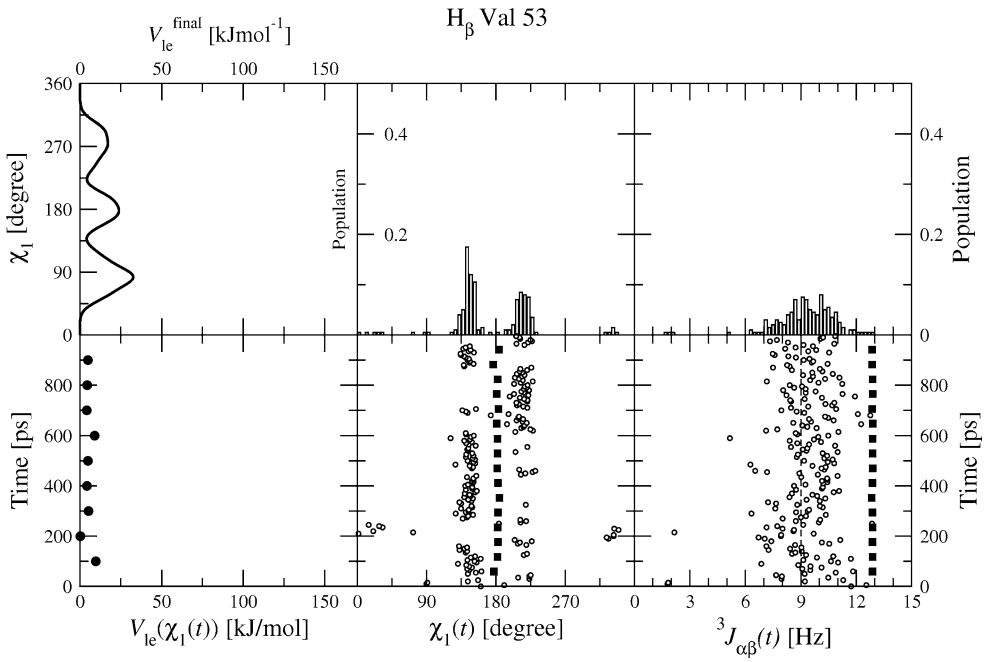

Fig. 14 Properties of the χ_1 torsional angle of Val 53 and the corresponding $^3J_{H_\alpha H_\beta}$ in the ^{3}*J*_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* final built-up local-elevation restraining potential energy $V_{le}^{final}(\chi_1)$. *Middle* distribution of χ_1 angles. *Right* distribution of ^{3}*J*-couplings. *Lower row*: *Left* local-elevation potential energy $V_{le}(t)$ acting on χ_1 at specific time points. *Middle* evolution of χ_1 angle in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and the χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ^{3}*J*-value in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and ^{3}*J*-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ^{3}*J*-value

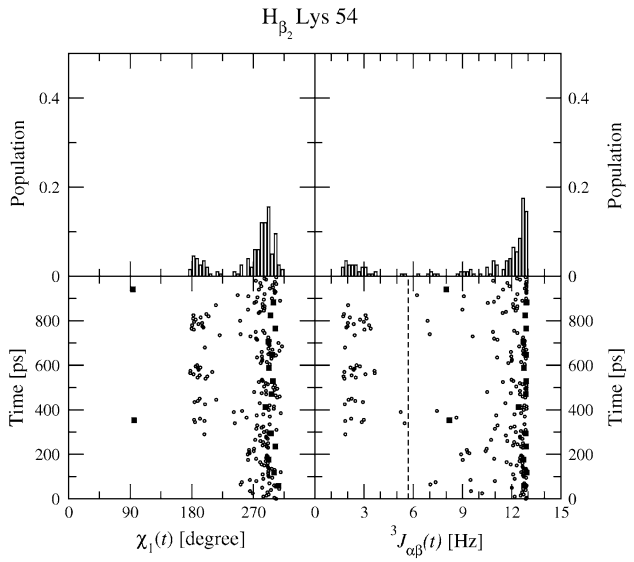

Fig. 15 Properties of the χ_1 torsional angle of Lys 54 and the corresponding $^3J_{H_\alpha H_{\beta_2}}$ in the ^{3}*J*_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* distribution of χ_1 angles. *Right* distribution of ^{3}*J*-couplings. *Lower row*: *Left* evolution of χ_1 angle in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ^{3}*J*-value in ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and ^{3}*J*-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ^{3}*J*-value

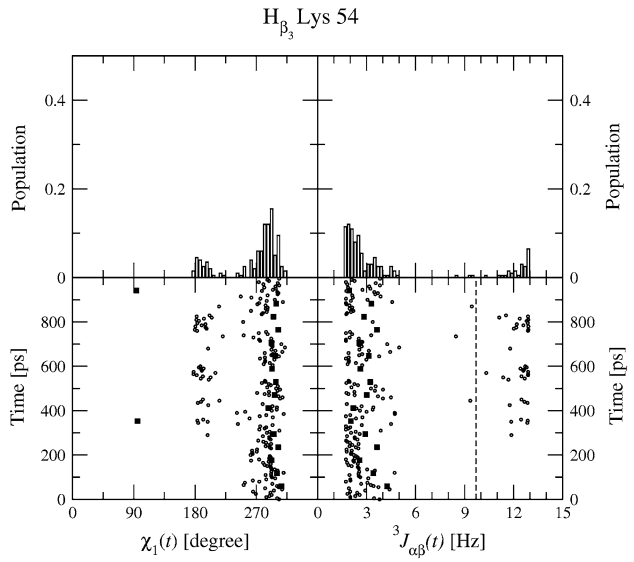

Fig. 16 Properties of the χ_1 torsional angle of Lys 54 and the corresponding $^3J_{H_\alpha H_{\beta_3}}$ in the ^{3}*J*_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* distribution of χ_1 angles. *Right* distribution of ^{3}*J*-couplings. *Lower row*: *Left* evolution of χ_1 angle in the ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ^{3}*J*-value in ^{3}*J*_LE_NOE_TAR_WAT simulation (*circles*) and ^{3}*J*-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ^{3}*J*-value

different χ_1-angles, as observed in the simulation. Comparison of Figs. 15 and 16 also gives an indication of a better stereospecific assignment than that chosen in these figures: a choice of 9.7 Hz for H_{β_2} and 5.7 Hz for H_{β_3} would improve the agreement between the simulated and experimental data.

Figures 17 and 18 show an example, the χ_1-angle of Asn 31, in which the set of NMR model structures reproduces the experimental values by averaging over two ranges of χ_1-values, whereas the simulation yields poor agreement with $\langle {}^3J_{\alpha\beta}\rangle_{\exp}$ because it only samples one range of χ_1-angle values (Table S2). Inversion of the chosen H_{β_2} versus H_{β_3} assignment would improve the agreement for the set of NMR model structures while worsening it for the simulation.

Figures 19 and S3 show an example, the χ_1-angle of Ser 81, in which the averaging in the MD simulation leads to a reproduction of the observed $\langle {}^3J_{\alpha\beta}\rangle_{\exp}$-couplings, while the NMR model structures fail to do so (Table S2).

Finally, the example of Glu 45 in Figs. 20 and S4 also shows the importance of conformational averaging and the ${}^3J_{\alpha\beta}$-value distributions suggest inversion of the chosen assignment.

These examples of the various effects of time-averaged local-elevation biassing based on ${}^3J$-coupling constants

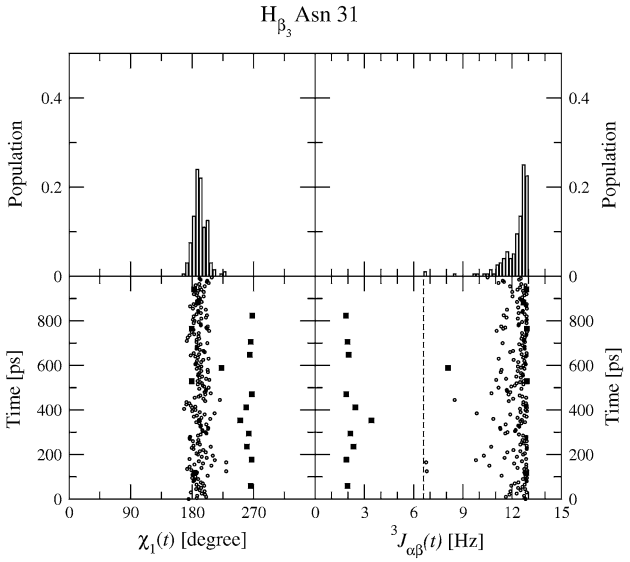

Fig. 18 Properties of the χ_1 torsional angle of Asn 31 and the corresponding ${}^3J_{H_\alpha H_{\beta_3}}$ in the ${}^3J$_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* distribution of χ_1 angles. *Right* distribution of ${}^3J$-couplings. *Lower row*: *Left* evolution of χ_1 angle in the ${}^3J$_LE_NOE_TAR_WAT simulation (*circles*) and χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ${}^3J$-value in ${}^3J$_LE_NOE_TAR_WAT simulation (*circles*) and ${}^3J$-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ${}^3J$-value

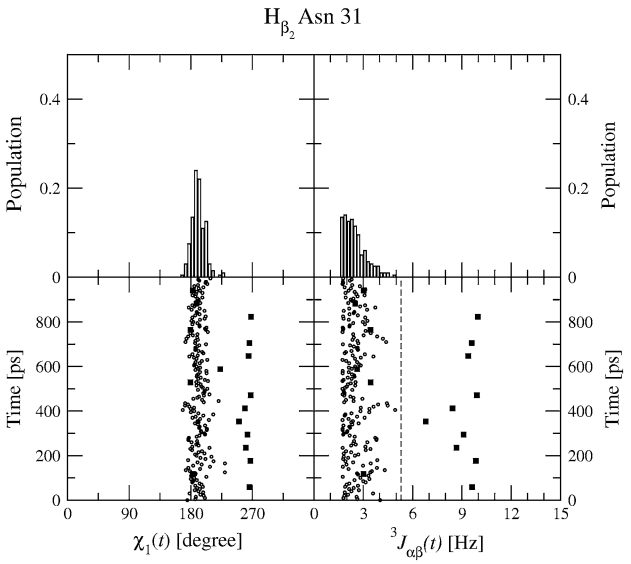

Fig. 17 Properties of the χ_1 torsional angle of Asn 31 and the corresponding ${}^3J_{H_\alpha H_{\beta_2}}$ in the ${}^3J$_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* distribution of χ_1 angles. *Right* distribution of ${}^3J$-couplings. *Lower row*: *Left* evolution of χ_1 angle in the ${}^3J$_LE_NOE_TAR_WAT simulation (*circles*) and χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ${}^3J$-value in ${}^3J$_LE_NOE_TAR_WAT simulation (*circles*) and ${}^3J$-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ${}^3J$-value

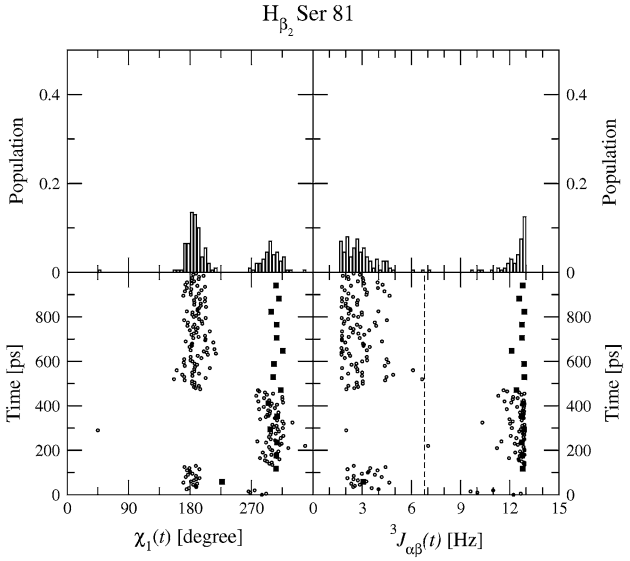

Fig. 19 Properties of the χ_1 torsional angle of Ser 81 and the corresponding ${}^3J_{H_\alpha H_{\beta_2}}$ in the ${}^3J$_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* distribution of χ_1 angles. *Right* distribution of ${}^3J$-couplings. *Lower row*: *Left* evolution of χ_1 angle in the ${}^3J$_LE_NOE_TAR_WAT simulation (*circles*) and χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ${}^3J$-value in ${}^3J$_LE_NOE_TAR_WAT simulation (*circles*) and ${}^3J$-values in the set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ${}^3J$-value

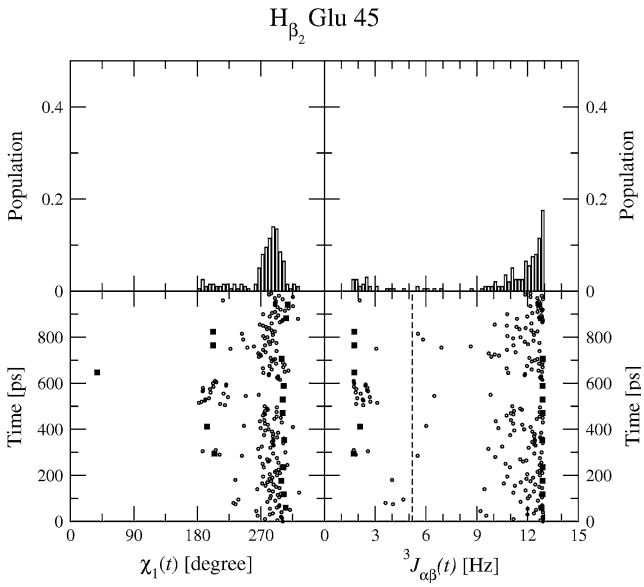

Fig. 20 Properties of the χ_1 torsional angle of Glu 45 and the corresponding $^3J_{H_\alpha H_{\beta_2}}$ in the ^3J_LE_NOE_TAR_WAT simulation. *Upper row*: *Left* distribution of χ_1 angles. *Right* distribution of ^3J-couplings. *Lower row*: *Left* evolution of χ_1 angle in the ^3J_LE_NOE_TAR_WAT simulation (*circles*) and χ_1 angles in set of 16 NMR model structures (*squares*, from *bottom* to *top* structures 1–16). *Right* evolution of ^3J-value in ^3J_LE_NOE_TAR_WAT simulation (*circles*) and ^3J-values in the set of 16 NMR model structures (squares, from *bottom* to *top* structures 1–16). The *dashed line* shows the experimental ^3J-value

show that the technique enhances the search for the appropriate rotamer when needed, extends the sampling when needed, and compensates for force-field deficiencies when needed, based on comparison of time-averaged with measured ^3J-coupling values.

Conclusions

Refinement of a protein structure on the basis of NMR data is still a challenge because of the low ratio of independent observables to molecular degrees of freedom, the approximations involved in the different relationships between particular observable quantities and molecular conformation, and the averaged character of the experimental data which may even, if stemming from different measurements, represent different thermodynamic state points. The recent literature and the Protein Data Bank still contain structures obtained from single-structure refinement in non-explicit solvent using non-observed data as geometric restraints in addition to a low-accuracy force field. Such a procedure may easily result in a set of protein structures which are conformationally too restricted, as is illustrated in Fig. 21. Application of time-averaged restraints and use of enhanced sampling techniques yield a conformationally more diverse ensemble of protein structures while satisfying the experimentally measured ^3J-couplings and NOE distance bounds better than the conformationally restricted set of structures resulting from single-structure refinement.

Regarding the use of ^3J-couplings in structure refinement it is clear that the accuracy of the parametrisation of the Karplus relationship between torsional angle and

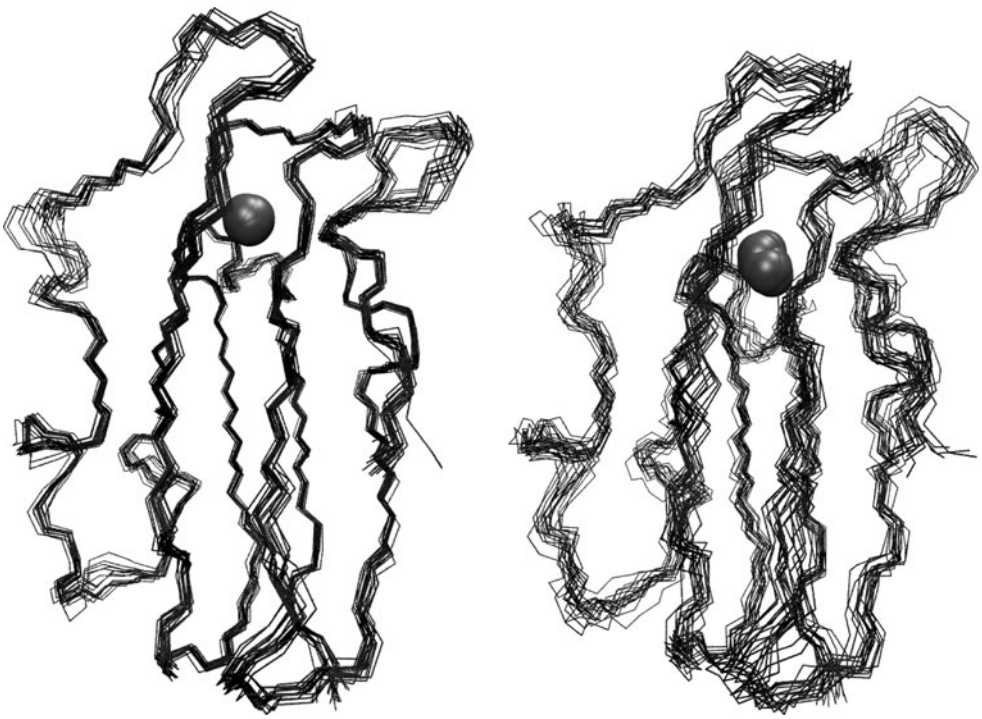

Fig. 21 Best-fit superposition of the backbone N, C^α, C, and O atoms with regard to the last structure of the set of 16 NMR model structures. The positions of the N, C^α and C atoms of the backbone and the Cu ion of the 16 NMR model structures (*left*) and of the 16 structures from the second half of the ^3J_LE_NOE_TAR_WAT simulation (*right*) are shown

^3J-coupling or even the relationship itself must be improved (Steiner et al. 2012). Second, current force fields for proteins do not yet seem accurate enough to predict protein structures in atomic detail without additional restraining or biassing terms representing data measured for the particular proteins. Third, the barriers for conformational changes, e.g. side-chain rotation, are often too high to be observed in nanosecond MD simulations, which makes the use of sampling enhancement techniques mandatory. Regarding all three aspects, progress is expected in the coming decade.

Acknowledgments This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-137827 of the Swiss National Science Foundation, and by grant number 228076 of the European Research Council, which is gratefully acknowledged.

References

Barker JA, Watts RO (1973) Monte Carlo studies of the dielectric properties of water-like models. Mol Phys 26:789–792

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak J (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank, http://www.pdb.org. Nucleic Acids Res 28:235–242

Berman HM, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank, http://www.wwpdb.org. Nat Struct Biol 10:980

Brünger AT, Kuriyan J, Karplus M (1987) Crystallographic R factor refinement by molecular dynamics. Science 235:458–460

Brüschweiler R, Case DA (1994) Adding harmonic motion to the Karplus relation for spin-spin coupling. J Am Chem Soc 116: 11199–11200

Bürgi R, Pitera J, van Gunsteren WF (2001) Assessing the effect of conformational averaging on the measured values of observables. J Biomol NMR 19:305–320

Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. PNAS 104: 9615–9620

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

de Marco A, Llinás M, Wüthrich K (1978) Analysis of the 1H-NMR spectra of ferrichrome peptides. I. The non-amide protons. Biopolymers 17:617–636

Fennen J, Torda AE, van Gunsteren WF (1995) Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble. J Biomol NMR 6:163–170

Gippert GP, Yip PF, Wright PE, Case DA (1990) Computational methods for determining protein structures from NMR data. Biochem Pharm 40:15–22

Glättli A, van Gunsteren WF (2004) Are NMR-derived model structures for peptides representative for the ensemble of structures adopted in solution? Probing the fourth helical secondary structure of β-peptides by molecular dynamics simulation. Angew Chem Int Ed Engl 43:6312–6316

Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 120:11919–11929

Harvey TS, van Gunsteren WF (1993) The application of chemical shift calculation to protein structure determination by NMR techniques, In: Techniques in protein chemistry IV. Academic Press, London, pp 615–622

Havel TF, Crippen GM, Kuntz ID (1979) Effects of distance constraints on macromolecular conformation. II. Simulation of experimental results and theoretical predictions. Biopolymers 18:73–81

Heinz TN, van Gunsteren WF, Hünenberger PH (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. J Chem Phys 115:1125–1136

Hendrickson WA (1985) Stereochemically restrained refinement of macromolecular structures. Methods Enzymol 115:252–270

Hoch JC, Poulsen FM, Redfield C (1991) Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy. NATO ASI Series A225. Plenum Press, New York

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

Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamic simulations. J Comput Aided Mol Des 8:695–708

Jardetzky O (1980) On the nature of molecular conformations inferred from high-resolution NMR. Biochim Biophys Acta 621:227–232

Jardetzky O, Finucane M (2001) Dynamics, structure and function of biological macromolecules. NATO Science Series A315. IOS Press, Amsterdam

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Kaptein R, Zuiderweg ERP, Scheek RM, Boelens R, van Gunsteren WF (1985) A protein structure from nuclear magnetic resonance data: lac repressor headpiece. J Mol Biol 182:179–182

Karplus M (1959) Contact electron-spin coupling of nuclear magnetic moments. J Chem Phys 30:11–15

Karplus M (1963) Vicinal proton coupling in nuclear magnetic resonance. J Am Chem Soc 85:2870

Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived Ubiquitin ensemble in solution. Science 320: 1471–1475

Lee MS, Gippert GP, Soman KV, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science 245:635–637

Markwick PRL, Showalter SA, Bouvignies G, Brüschweiler R, Blackledge M (2009) Structural dynamics of protein backbone φ angles: extended molecular dynamics simulations versus experimental ^3J scalar couplings. J Biomol NMR 45:17–21

Moore JM, Lepre CA, Gippler GP, Chazin WJ, Case DA, Wright PE (1991) High-resolution solution structure of reduced French bean plastocyanin and comparison with the crystal structure of poplar plastocyanin. J Mol Biol 221:533–555

Nanzer AP, van Gunsteren WF, Torda AE (1995) Parametrisation of time-averaged distance restraints in MD simulations. J Biomol NMR 6:313–320

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 Comput Chem 25:1656–1676

Pardi A, Billeter M, Wüthrich K (1984) Calibration of the angular dependence of the amide proton-C_α proton coupling constants, $^3J_{HN_\alpha}$, in a globular protein: use of $^3J_{HN_\alpha}$ for identification of helical secondary structure. J Mol Biol 180:741–751

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of Cartesian equations of motion of a system with constraints—molecular dynamics of *n*-alkanes. J Comp Phys 23:327–341

Salmon L, Bouvignies G, Markwick P, Blackledge M (2011) Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. Biochemistry 50:2735–2747

Scheek RM, Torda AE, Kemmink J, van Gunsteren WF (1991) Structure determination by NMR: the modelling of NMR parameters as ensemble averages. In: Hoch JC, Poulsen FM, Redfield C (eds) Computational aspects of the study of biological macromolecules by nuclear magnetic resonance spectroscopy. NATO ASI Series A225, Plenum Press, New York, pp 209–217

Schmid N, Allison J, Dolenc J, Eichenberger AP, Kunz AP, van Gunsteren WF (2011) Biomolecular structure refinement using the GROMOS simulation software. J Biomol NMR 51:265–281

Schmidt JM, Blümel M, Löhr F, Rüterjans H (1999) Self-consistent ^3J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsional angles. J Biomol NMR 14:1–12

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

Steiner D, Allison JR, Eichenberger AP, von Gunsteren WF (2012) On the calculation of $^3J_{\alpha\beta}$-coupling constants for side chains in proteins. J Biomol NMR (accepted)

Tironi IG, Sperb G, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular-dynamics simulations. J Chem Phys 102:5451–5459

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 3:55–56

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. Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland

van Gunsteren WF, Dolenc J, Mark AE (2008) Molecular simulation as an aid to experimentalists. Curr Opin Struct Biol 18:149–153

Weiner SJ, Kollmann PA, Nguyen DT, Case DA (1986) An all-atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252

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