J Biomol NMR (2010) 47:221–235
DOI 10.1007/s10858-010-9425-9

ARTICLE

Methods of NMR structure refinement: molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1

Jožica Dolenc · John H. Missimer · Michel O. Steinmetz · Wilfred F. van Gunsteren

Received: 6 January 2010 / Accepted: 21 April 2010 / Published online: 4 June 2010

Abstract The C-terminal trigger sequence is essential in the coiled-coil formation of GCN4-p1; its conformational properties are thus of importance for understanding this process at the atomic level. A solution NMR model structure of a peptide, GCN4p16–31, encompassing the GCN4-p1 trigger sequence was proposed a few years ago. Derived using a standard single-structure refinement protocol based on 172 nuclear Overhauser effect (NOE) distance restraints, 14 hydrogen-bond and 11 ϕ torsional-angle restraints, the resulting set of 20 NMR model structures exhibits regular α-helical structure. However, the set slightly violates some measured NOE bounds and does not reproduce all 15 measured $^3J(H_N\text{-}H_{C\alpha})$-coupling constants, indicating that different conformers of GCN4p16–31 might be present in solution. With the aim to resolve structures compatible with all NOE upper distance bounds and ^{3}J-coupling constants, we executed several structure refinement protocols employing unrestrained and restrained molecular dynamics (MD) simulations with two force fields. We find that only configurational ensembles obtained by applying simultaneously time-averaged NOE distance and ^{3}J-coupling constant restraining with either force field reproduce all the experimental data. Additionally, analyses of the simulated ensembles show that the conformational variability of GCN4p16–31 in solution admitted by the available set of 187 measured NMR data is larger than represented by the set of the NMR model structures. The conformations of GCN4p16–31 in solution differ in the orientation not only of the side-chains but also of the backbone. The inconsistencies between the NMR model structures and the measured NMR data are due to the neglect of averaging effects and the inclusion of hydrogen-bond and torsional-angle restraints that have little basis in the primary, i.e. measured NMR data.

Keywords NMR structure determination · Time averaging · Local elevation · NOE upper bounds · ^{3}J-coupling constants

Introduction

Structural investigations of biomolecules are essential to understand the role they play in biological processes. However, since biomolecules in solution exist as ensembles of different conformers rather than as single conformers, neglecting the dynamic nature of biomolecules may lead to misunderstanding their biological function (Jardetzky 1980; Karplus and McCammon 1983; Bonvin and Brünger 1995; Bax and Tjandra 1997; Best et al. 2006; Vendruscolo 2007).

In solution nuclear magnetic resonance (NMR) spectroscopy the primary, measured data are collected as temporal and spatial averages of molecular conformations. The interpretation of NMR observables, therefore, requires

Electronic supplementary material The online version of this article (doi:10.1007/s10858-010-9425-9) contains supplementary material, which is available to authorized users.

J. Dolenc · W. F. van Gunsteren (✉)
Laboratory of Physical Chemistry, ETH, Swiss Federal Institute of Technology, 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

J. H. Missimer · M. O. Steinmetz
Biomolecular Research, Paul Scherrer Institut, 5232 Villigen, Switzerland

accounting for the conformational averaging in the NMR structure refinement protocol (Kessler et al. 1988; Torda et al. 1990; Pearlman and Kollman 1991; Bonvin et al. 1994; Daura et al. 1999; Bürgi et al. 2001). The primary data obtained from the solution NMR experiment are usually a set of distance restraints between pairs of hydrogen atoms extracted from nuclear Overhauser effects (NOEs) and a set of ^{3}J-coupling constants, which can be related to the torsional angles by the Karplus relation (Karplus 1963),

$$J(\phi(\mathbf{r}(t))) = a\cos^2\phi(t) + b\cos\phi(t) + c \quad (1)$$

where ϕ is the torsional angle defined by the four covalently bound atoms that determine a particular ^{3}J-coupling constant, $\mathbf{r}(t)$ denotes a molecular conformation as function of time, and a, b and c are empirical coefficients. Figure 1 shows the Karplus curves for $^3J(H_N\text{-}H_{C\alpha})$-coupling constants determined by four different sets of parameters obtained for different molecules under different conditions (Pardi et al. 1984; Brüschweiler and Case 1994; Wang and Bax 1996; Schmidt et al. 1999).

Since molecular dynamics (MD) simulations provide representations of the dynamics of molecules in solution, yielding trajectories appropriate for averaging, they have become a well established tool in NMR structure refinement (van Gunsteren and Berendsen 1990; Schmitz et al. 1992; Mierke et al. 1994; van Gunsteren et al. 1994; Berndt et al. 1996; Cuniasse et al. 1997; Glättli and van Gunsteren 2004; Trzesniak et al. 2005; Beckman et al. 2006; Zagrovic and van Gunsteren 2006; Fawzi et al. 2008; Zagrovic et al. 2008). However, the utility of unrestrained MD simulations in determining biomolecular structure can be limited by insufficient sampling of conformational space (Fig. 2, examples D, E and F) and by the finite accuracy of the force field used (van Gunsteren and Mark 1998; van Gunsteren et al. 2008) (Fig. 2, examples B, D and E). In order to bias the sampling towards the relevant regions of the configurational space, primary experimental data, such as measured NOE upper distance bounds and ^{3}J-coupling constants, can be introduced as restraints in MD simulations by adding a penalty function V^{restr} to the physical force field V^{phys} (Kaptein et al. 1985):

$$V(\mathbf{r}(t)) = V^{phys}(\mathbf{r}(t)) + V^{restr}(\mathbf{r}(t)) \quad (2)$$

Various functional forms of V^{restr} have been developed, each restricting the sampled conformational space differently. In NMR structure refinement based on instantaneous restraints (IR) the penalty function V^{restr} usually has a harmonic functional form,

$$V^{restr}(\mathbf{r}(t)) = 1/2\sum_{i=1}^{N_{restr}} K_i^{qr}\left(q_i(\mathbf{r}(t)) - q_i^0\right)^2 \quad (3)$$

or in case of NOEs the corresponding half-harmonic attractive form, which raises the energy of the system as the deviation of the actual value of an observable $q_i(t)$ from the experimentally measured value q_i^0 increases (Kaptein et al. 1985). Since instantaneous restraints only allow conformations that satisfy the observed averaged $<q_i>$, MD simulations applying instantaneous restraints predict an ensemble of structures that either agrees (Fig. 2, examples A, B, D and E) or disagrees with the real ensemble of structures (Fig. 2, examples C and F), depending on the real potential energy surface. A more tolerant approach to impose restraints is to treat the NMR data as quantities satisfied only on average over the course of a restrained MD simulation (Torda et al. 1989; Torda et al. 1993; Fennen et al. 1995; Nanzer et al. 1995; Nanzer et al. 1996; Keller et al. 2007). This can be achieved by using the weighted temporal average during the simulation,

$$\overline{q_i(\mathbf{r}(t))} = \frac{1}{\tau_{qr}\left(1-\exp\left(-t/\tau_{qr}\right)\right)}\int_0^t \exp\left(\frac{t'-t}{\tau_{qr}}\right) q(\mathbf{r}(t'))dt' \quad (4)$$

in Eq. (3) or its half-harmonic equivalent. Here τ_{qr} is a characteristic memory relaxation (or averaging) time. The structure refinement protocols based on MD simulations with time-averaged restraints (TAR) result in ensembles of conformations rather than in single structures. MD simulations with time-averaged distance restraints based on NOE upper bounds have successfully been applied in a number of cases (Torda et al. 1990; Nanzer et al. 1994; Nanzer et al. 1997; Gattin et al. 2009). However, their application to ^{3}J-coupling constant refinement may be

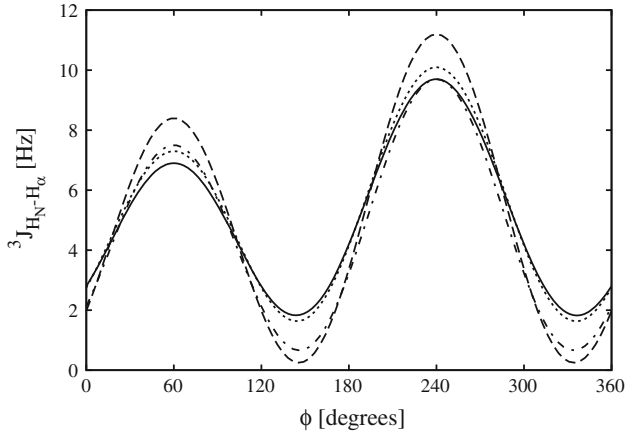

Fig. 1 Karplus curves for the $^3J(H_N\text{-}H_{C\alpha})$ couplings with the calibration constants from (Pardi et al. 1984) (*solid line*), (Brüschweiler and Case 1994) (*dashed line*), (Wang and Bax 1996) (*dotted line*) and (Schmidt et al. 1999) (*dash-dotted line*). A phase shift of 60° was applied

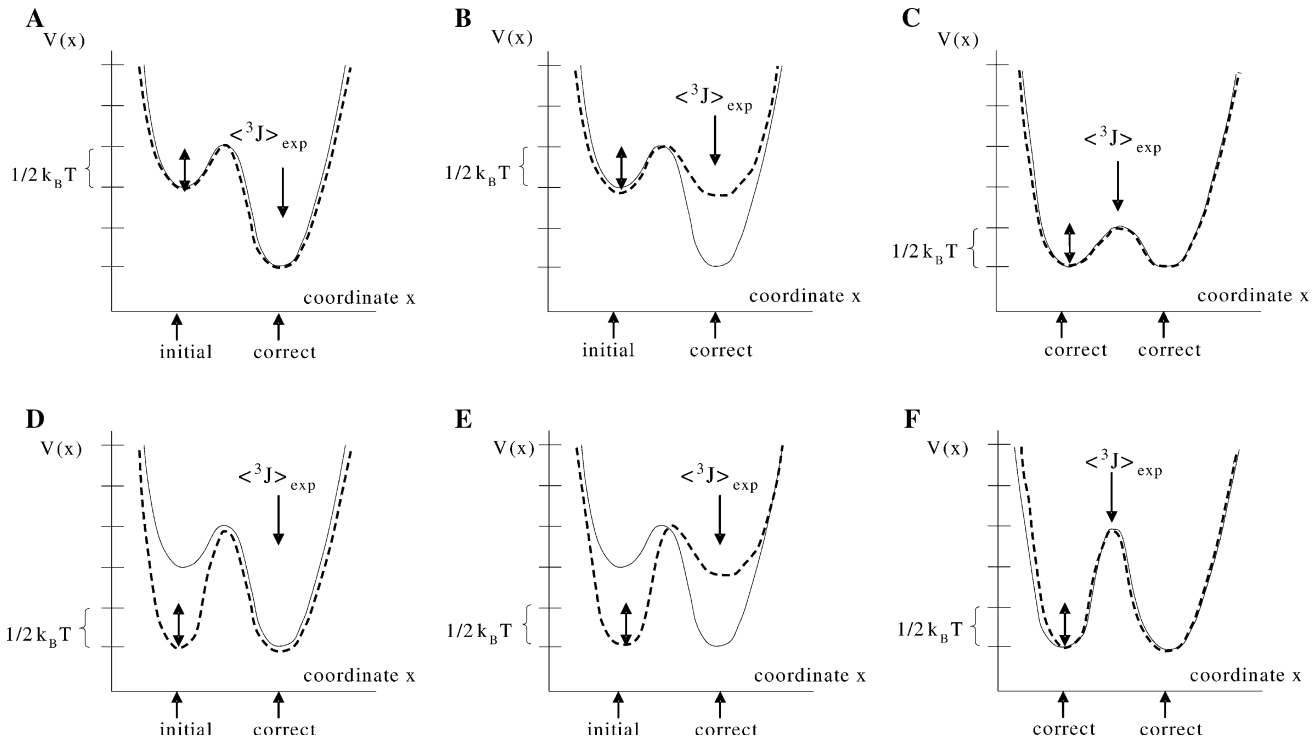

Fig. 2 Schematic representation of six different real (*solid line*) and model (*dashed line*) potential energy functions illustrating the force-field problem of unrestrained MD simulations (examples B, D, E), the sampling problem of MD simulations which occurs when instantaneous restraints are applied (examples C and F), and the search problem of MD simulations due to high-energy barriers between different conformations (examples D, E, F). The double arrow indicates the thermal energy (1/2 k_BT) associated with the degree of freedom x. If the thermal energy is comparable to the barrier height, transitions are easy, whereas a higher barrier leads to rare transitions. If the measured ^{3}J-value, $< {}^3\mathrm{J} >_{\mathrm{exp}}$, corresponds according to the non-linear Karplus relation to a torsional-angle coordinate x for which the potential energy is greater (examples E and F), instantaneous restraining will lead to an unrealistic configuration x

problematic due to i) the multi-valuedness of the inverse of the Karplus relation, implying that a particular value of a ^{3}J-coupling constant can correspond to several different torsional-angle values (Fig. 1) and ii) high-energy barriers between conformations with different torsional-angle values that may prevent the sampling of the entire range of torsional-angle values contributing to the measured (average) ^{3}J-coupling constants. Recently a method using time-averaging and local-elevation (LE) biasing of the torsional-angle conformational search has been proposed to enforce ^{3}J-coupling constant restraints (Christen et al. 2007). The restraining potential energy function associated with the k-th ^{3}J-coupling constant is a sum of N_{le} LE terms (Huber et al. 1994)

$$V_{\mathrm{k}}^{\mathrm{Jres}}(\phi_{\mathrm{k}}(\mathbf{r}(t))) = \sum_{\mathrm{i}=1}^{\mathrm{N_{le}}} V_{\mathrm{ki}}^{\mathrm{le}}(\phi_{\mathrm{k}}(\mathbf{r}(t))) \tag{5}$$

in which the penalty terms are Gaussian functions centered at ϕ_{ki}^0,

$$V_{\mathrm{ki}}^{\mathrm{le}}(\phi_{\mathrm{k}}(\mathbf{r}(t))) = K_{\mathrm{k}}^{\mathrm{Jres}} w_{\phi_{\mathrm{ki}}}(t) \times \exp\left(-\left(\phi_{\mathrm{k}}(t) - \phi_{\mathrm{ki}}^0\right)^2 \Big/ 2\left(\Delta\phi^0\right)^2\right) \tag{6}$$

where $K_{\mathrm{k}}^{\mathrm{Jres}}$is the overall penalty function force constant and $w_{\phi_{ki}}$is the weight of the i-th penalty function. The latter is calculated using a product of two flat-bottom (fb) terms, one for ^{3}J(t) and one for $^3\bar{J}(t)$ in order to determine if the instantaneous or time-averaged ^{3}J-value deviates from the experimental one:

$$\omega_{\phi_{\mathrm{ki}}}(t) = t^{-1} \int_0^{\mathrm{t}} \delta_{\phi_{\mathrm{k}}(r(t'))\phi_{\mathrm{ki}}^0} V^{\mathrm{fb}}\left({}^3J(\phi_{\mathrm{k}}(\mathbf{r}(t')))\right) \times V^{\mathrm{fb}}\left(\overline{{}^3J(\phi_{\mathrm{k}}(\mathbf{r}(t')))}\right) dt' \tag{7}$$

$$\delta_{\phi_{\mathrm{k}}(r(t))\phi_{\mathrm{ki}}^0} = \begin{cases} 1 \; if \quad \phi_{\mathrm{ki}}^0 - \Delta\phi^0/2 < \phi_{\mathrm{k}}(\mathbf{r}(t)) < \phi_{\mathrm{ki}}^0 + \Delta\phi^0/2 \\ 0 \quad \text{otherwise} \end{cases} \tag{8}$$

$$V^{\mathrm{fb}}(J_{\mathrm{k}}(t)) = \begin{cases} \left(J(\phi_{\mathrm{k}}(\mathbf{r}(t))) - J^0_{\mathrm{k}} - \Delta J^0\right)^2 & \text{if } J(\phi_{\mathrm{k}}(\mathbf{r}(t))) > J^0_{\mathrm{k}} + \Delta J^0 \\ \left(J(\phi_{\mathrm{k}}(\mathbf{r}(t))) - J^0_{\mathrm{k}} + \Delta J^0\right)^2 & \text{if } J(\phi_{\mathrm{k}}(\mathbf{r}(t))) < J^0_{\mathrm{k}} - \Delta J^0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Unless the time-averaged value ${}^3\bar{J}_k(\phi_k(t))$ or the current value ${}^3J_k(\phi_k(t))$ are close to the experimental one ${}^3J^0_k$, the conformation is pushed away from ϕ^0_{ki} resulting finally in an average close to the experimental ${}^3J_k$-value (Christen et al. 2007).

Here we assess seven different structure refinement procedures with two different GROMOS force fields applied to a peptide corresponding to the C-terminal coiled-coil trigger sequence of the yeast transcriptional activator GCN4, denoted GCN4p16–31, for which the current PDB model structures (PDB entry 2ovn) do not reproduce all 15 experimentally determined ${}^3J(H_N\text{-}H_{C\alpha})$-coupling constants (Fig. 3, panel B). The NMR solution structures of this peptide (Steinmetz et al. 2007) were derived using the simulated annealing approach (Kirkpatrick et al. 1983; Nilges et al. 1988) based on 172 distance restraints derived from measured NOEs, and assuming 25 standard α-helical restraints suggested by the measured ${}^3J(H_N\text{-}H_{C\alpha})$-coupling constants and secondary Cα and Hα chemical shifts; these included 14 hydrogen-bond restraints between the N and H atoms of residues 22–28 and the O atoms of residues 18–24 as well as 11 ϕ torsional-angle restraints for the residues 18–28 (Steinmetz et al. 2007) (*Note*: The supplementary material of this reference cites 11 ϕ torsional-angle restraints whereas the reference refers to only 8). In addition to the use of the derived α-helical restraints, the structure refinement protocol relied on MD calculations performed at very high temperature with a simplified force field and without explicit consideration of the solvent degrees of freedom. The deficiencies of this refinement protocol prompted us to perform several structure determination procedures based on MD simulations using the thermodynamically calibrated GROMOS force fields at room temperature and explicit solvation. In order to generate an ensemble of configurations in complete agreement with all of the primary NMR data, i.e. 172 NOEs and 15 ${}^3J$-couplings, a series of MD simulations of the GCN4p16–31 involving proton-proton distance or ${}^3J(H_N\text{-}H_{C\alpha})$-coupling constant restraints as well as two unrestrained MD simulations were carried out (Table 1). Comparison of the NOE distances and ${}^3J(H_N\text{-}H_{C\alpha})$-coupling constants calculated from the simulated MD trajectories with the primary, measured NMR data provided the quality criteria.

Methods

Molecular dynamics simulations

All MD simulations reported in this paper were carried out using the GROMOS biomolecular simulation package (van Gunsteren et al. 1996; Scott et al. 1999; Christen et al. 2005) and the 43A1 (van Gunsteren et al. 1996; Daura et al. 1998) and 53A6 (Oostenbrink et al. 2004; Oostenbrink et al. 2005) GROMOS force-field parameter sets. The

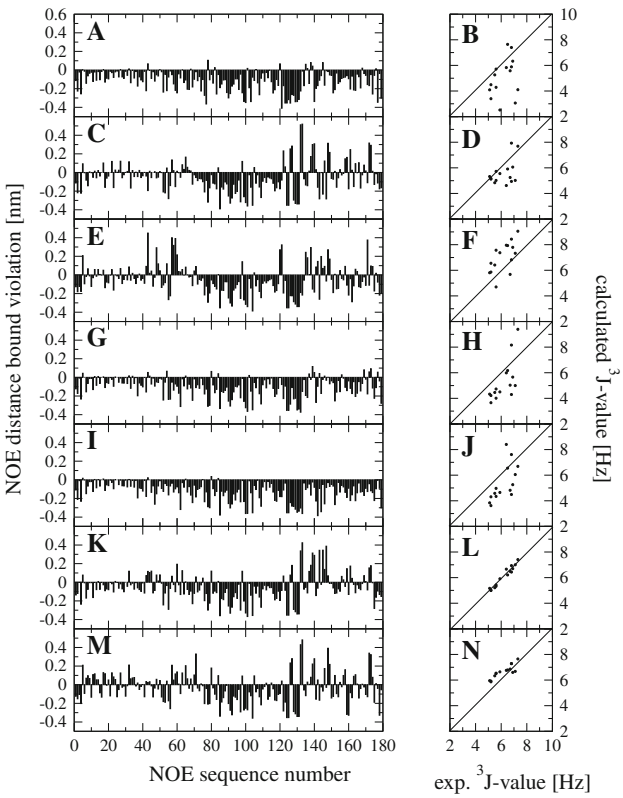

Fig. 3 Violations of the experimental NOE upper distance bounds as a function of the NOE sequence number (*left-hand panels*) and comparison of the experimental and calculated ${}^3J(H_N\text{-}H_{C\alpha})$-coupling constants (*right-hand panels*) for the 20 NMR model structures (*panels A and* B) and for the following six simulations: unrestrained_43A1 (*panels C and D*), unrestrained_53A6 (panels E and F), NOE_IR_43A1 (*panels G and H*), NOE_TAR_43A1 (*panels I and J*), ${}^3J$_IR_43A1 (*panels K and L*) and ${}^3J$_LE_43A1 (*panels M and N*). Simulation nomenclature is given Table 1 and NOE sequence numbers in Table 2 as well as in Table S1 (Online Resource)

Table 1 Overview of the MD simulations

MD simulation	Name	Starting coordinates	Simulation time [ns]
Unrestrained, 43A1 force field	unrestrained_43A1	NMR model 1	50
Unrestrained, 53A6 force field	unrestrained_53A6	NMR model 1	50
Instantaneous NOE distance restraining, 43A1 force field	NOE_IR_43A1	NMR model 1	10
Time-averaged NOE distance restraining, 43A1 force field	NOE_TAR_43A1	coordinates after 1 ns of NOE_IR_43A1	10
Instantaneous ^{3}J-coupling restraining, 43A1 force field	^{3}J_IR_43A1	NMR model 1	10
Local-elevation biased ^{3}J-coupling restraining, 43A1 force field	^{3}J_LE_43A1	NMR model 1	10
Time-averaged NOE distance restraining and instantaneous ^{3}J-coupling restraining, 43A1 force field	NOE_TAR + ^{3}J_IR_43A1	coordinates after 1 ns of NOE_IR + ^{3}J_IR_43A1	10
Time-averaged NOE distance restraining and instantaneous ^{3}J-coupling restraining, 53A6 force field	NOE_TAR + ^{3}J_IR_53A6	coordinates after 1 ns of NOE_IR + ^{3}J_IR_53A6	10
Time-averaged NOE distance restraining and local-elevation biased ^{3}J-coupling restraining, 43A1 force field	NOE_TAR + ^{3}J_LE_43A1	coordinates after 1 ns of NOE_IR + ^{3}J_IR_43A1	10
Time-averaged NOE distance restraining and local-elevation biased ^{3}J-coupling restraining, 53A6 force field	NOE_TAR + ^{3}J_LE_53A6	coordinates after 1 ns of NOE_IR + ^{3}J_IR_53A6	10
Time-averaged NOE distance restraining after 10 ns of NOE_TAR + ^{3}J_LE_43A1 simulation, 43A1 force field	NOE_TAR + -^{3}J_LE_43A1	coordinates after 10 ns of NOE_TAR + ^{3}J_LE_43A1	5

GCN4p16–31 peptide comprises the sequence: Ac-16Asn-17Tyr-18His-19Leu-20Glu-21Asn-22Glu-23Val-24Ala-25Arg-26Leu-27Lys-28Lys-29Leu-30Val-31Gly-NH_2. The His residue is protonated at NE2, the Arg and Lys side chains are protonated with charge +e. Coordinates of the first model structure of the NMR set of structures (PDB entry 2ovn) were taken as the starting coordinates for MD simulations (Table 1). The last residue (32Glu) of the model structure was removed because it was not present in the NMR experiment (Steinmetz et al. 2007). After steepest descent energy minimisation, the structure was solvated in a rectangular box of approximately 3000 pre-equilibrated simple point charge (SPC) water molecules (Berendsen et al. 1981) with a minimal solute-to-wall distance of 1.0 nm. The system was relaxed by performing a steepest-descent energy minimisation with harmonic positional restraints on all solute atoms (force constant 2.5×10^4 kJmol^{-1} nm^{-2}) followed by a 100 ps long equilibration, in which the positional restraints were gradually released reducing the force constant to 0.0 kJmol^{-1} nm^{-2} and the temperature was raised from 60 to 278 K. The initial atomic velocities were taken from a Maxwell distribution at 60 K. All MD simulations were performed using periodic boundary conditions. The equations of motion were integrated using the leap-frog algorithm with a time step of 2 fs. Centre of mass motion was stopped every 2 ps. Bond lengths of the peptide and the geometry of the water molecules were constrained by applying the SHAKE algorithm with a relative geometric tolerance of 10^{-4} (Ryckaert et al. 1977). The temperature and pressure were maintained at 278 K and 1 atm using the Berendsen thermostat with a coupling time $\tau_T = 0.1$ ps and barostat with a coupling time $\tau_P = 0.5$ ps and an isothermal compressibility of 4.575×10^{-4} (kJmol^{-1}nm^{-3})$^{-1}$ (Berendsen et al. 1984). A reaction-field approach was used to treat the electrostatics employing a triple-range cutoff scheme, with cutoffs of 0.8 and 1.4 nm, and a dielectric permittivity of 66.6 (Glättli et al. 2002). The pairlist was updated every five steps. The 179 NOE distance restraints in which GROMOS pseudo atom corrections (van Gunsteren et al. 1996) were included and 15 ^{3}J(H_N-$H_{C\alpha}$)-coupling constant restraints were deduced from the corresponding measurements (Steinmetz et al. 2007). The ^{3}J-coupling constants are listed in Table S2 (Online Resource).

The set of NOE distance restraints listed in Tables 2 and S1 contained 7 more restraints than used in (Steinmetz et al. 2007). Two proton pairs omitted in the X-PLOR refinement, H_N(18His)-H_N(16Asn) and H_N(17Tyr)-H_N(16Asn), were included, and five ambiguous assignments: H_N(21Asn)-H_α(18His) or H_α(17Tyr); H_N(20Glu)-H_α(18His) or H_α(17Tyr); $H_{\delta 2}$(18His)-H_γ(20Glu) or H_γ(22Glu); H_δ(17Tyr)-H_γ(20Glu) or H_γ(22Glu); H_ε(17Tyr)-H_γ(20Glu) or H_γ(22Glu), were incorporated as pairs of distance bounds. (*Note*: The number of NOE distance restraints used in computing the NMR model structures is not wholly clear due to a discrepancy between the supplementary material of Steinmetz et al., which quotes 175 NOE distance restraints, and an X-PLOR input file obtained from one of the authors (AA) of the publication, which lists 174 distance restraints of which "… two were commented out in the course of the refinement due to violations or incorrect assignments".) We considered that the two deleted restraints were probably primary data and

decided to include them. Inclusion of the five additional restraints was based on the consideration that the ambiguous assignments could reflect the averaging inherent to the measurement, suggesting that Boltzmann distributed ensemble generated in the MD simulation should reveal whether one or both of the NOE restraint pairs can be satisfied. The instantaneous NOE distance restraints were imposed with a force constant of 2000 kJmol^{-1}nm^{-2}, time-averaged NOE distance restraints with a force constant of 6000 kJmol^{-1}nm^{-2} and the ^{3}J(H_N-$H_{C\alpha}$)-coupling constant instantaneous restraints were imposed with a force constant of 10 kJmol^{-1} Hz^{-2}. The time-averaged NOE distance restraints used a memory relaxation time τ_{qr} of 20 ps (Torda et al. 1993; Nanzer et al. 1997). The LE ^{3}J-coupling biasing used a memory relaxation time τ_{qr} of 5 ps (Allison and van Gunsteren 2009). Additionally, due to the uncertainty in the ^{3}J(H_N-$H_{C\alpha}$)-coupling constants calculated from the corresponding ϕ angles via the Karplus relation a flat-bottom restraining energy term with a 2 Hz wide well ($\Delta J^{\circ} = 1$ Hz) was used in the LE ^{3}J-coupling biasing simulations (Allison and van Gunsteren 2009). The number of LE Gaussian functions per dihedral angle was set to $N_{le} = 36$ and the restraints were imposed with a force constant $K^{Jres} =$ 0.005 kJmol^{-1} Hz^{-4}.

Table 2 Proton pairs corresponding to the sequence numbers of the experimental NOE upper distance bounds for GCN4p16–31 in Figs. 3 and 4

NOE sequence no.	Proton pairs
1–56	H_N(i)–H_N(i + 1), H_N(i + 2)
	H_α(i)–H_N(i + 1), H_N(i + 2), H_N(i + 3), H_N(i + 4)
57–71	H_α(i)–H_β(i + 3)
72–118	H_N(i)–H_β(i-3), H_β(i-1), H_β(i), H_β(i + 1)
	H_N(i)–H_γ(i-1), H_γ(i), H_γ(i + 1), H_γ(i + 4)
	H_N(i)–H_δ(i-1), H_δ(i),
119–147	H_δ(Y17)–H_α(Y17), H_β(Y17), H_ε(H18), H_δ(L19), H_β(E20), H_γ(E20), H_γ(E22)
	H_ε(Y17)–H_α(Y17), H_β(Y17), H_ε(H18), H_δ(L19), H_γ(E20), H_β(E22), H_γ(E22)
	H_δ(H18)–H_α(H18), H_β(H18), H_δ(L19), H_β(E20), H_γ(E20), H_γ(E22)
	H_ε(H18)–H_δ(Y17), H_ε(Y17), H_β(H18), H_δ(L19), H_γ(E22)
148–155	H_γ(E22)–H_δ(R25)
	H_δ(R25)–H_β(N21), H_α(E22), H_α(R25)
156–165	H_δ(L26)–H_β(R25)
	H_α(E22)–H_γ(R25)
	H_β(E22)–H_α(L19), H_β(L19), H_β(L26), H_δ(R25)
	H_γ(E22)–H_β(L19), H_γ(V23), H_δ(L19)
166–174	H_β(N21)–H_β(A24)
	H_α(N21)–H_β(E20), H_γ(E20)
	H_α(H18)–H_β(L19)
	H_α(Y17)–H_γ(E20)
	H_β(Y17)–H_β(L19), H_δ(L19)
	H_α(N16)–H_δ(L19)
175–179	H_α(V23)–H_β(L26), H_γ(L26), H_γ(E22)
	H_γ(V23)–H_α(E20), H_γ(E22)

The residue sequence numbers are given within parentheses. See also Table S2 (Online Resource)

From the equilibrated structure two 50 ns long unrestrained MD simulations using the GROMOS 43A1 and 53A6 force fields (unrestrained_43A1, unrestrained_53A6) were started. In addition, the following 9 restrained MD simulations were performed: two 10 ns long MD simulations using the 43A1 force field in which NOE distance restraints were imposed either as instantaneous or time-averaged restraints (NOE_IR_43A1 and NOE_TAR_43A1); two 10 ns long MD simulations using the 43A1 force field in which ^{3}J(H_N-$H_{C\alpha}$)-coupling constant restraints were imposed either as instantaneous restraints or restraints using time-averaging together with LE biasing of the conformational search (^{3}J_IR_43A1, ^{3}J_LE_43A1); four 10 ns long MD simulations using the 43A1 and 53A6 force fields in which NOE distance restraints were imposed as time-averaged restraints and ^{3}J(H_N-$H_{C\alpha}$)-coupling constant restraints were imposed either as instantaneous restraints or restraints using time averaging together with LE biasing of the conformational search (NOE_TAR + ^{3}J_IR_43A1, NOE_TAR + ^{3}J_LE_43A1, NOE_TAR + ^{3}J_IR_53A6, NOE_TAR + ^{3}J_LE_53A6); one 5 ns long MD simulation extending the NOE_TAR + ^{3}J_LE_43A1 simulation but with the LE potential energy term removed (NOE_TAR + -^{3}J_LE_43A1). A list of all MD simulations together with the nomenclature used in this paper is given in Table 1. MD simulations applying time-averaged NOE distance restraining were started after a 1 ns long MD simulation in which the distance restraints were imposed as instantaneous restraints.

Analysis

The trajectory configurations were saved every 0.5 ps. The NMR model structures and the trajectory configurations of all MD simulations were analysed in terms of atom-positional root-mean-square-deviation (RMSD) from the energy minimised initial structure. The RMSD values were calculated for the heavy atoms of the backbone (C_α, N, C) and side chains of all the residues using the backbone atoms C_α, N and C to perform the superposition of centres of mass and rotational least-squares fit superposition (Kearsley 1989) of the successive structures onto the reference one. Additionally, for the set of NMR model structures and for the MD simulation trajectories atom-positional root-mean-square fluctuations (RMSF) were

calculated for the heavy atoms of the backbone (C_α, N and C) and side-chains of all the residues. Interproton distances derived from the NOE cross-peak intensities were compared with the average interproton distances calculated from the simulated and model structures using $\langle r^{-6} \rangle^{-1/6}$ averaging. The results are presented as distance bound violations, i.e., as a difference between the distances averaged over the simulation and the corresponding NMR derived upper distance bounds. Because the GROMOS force fields make use of united atoms, positions of aliphatic hydrogen atoms of interest were constructed based on standard geometries (van Gunsteren et al. 1996). If a NOE upper bound involved non stereo-specifically assigned protons, a pseudo atom was constructed (van Gunsteren et al. 1996). The pseudo-atom bound corrections used in the original NOE list were subtracted from the upper bounds and the GROMOS pseudo-atom bound corrections were applied. The list of NOE hydrogen-atom pairs, the corresponding NOE upper bounds and the violations calculated for the 20 NMR model structures and the NOE_TAR + ^{3}J_IR_43A1, NOE_TAR + ^{3}J_LE_43A1, NOE_TAR + ^{3}J_IR_53A6, and NOE_TAR + ^{3}J_LE_53A6 trajectories are given in Table S1 (Online Resource). Additionally, Table S2 provides the complete list of the $^3J(H_N\text{-}H_{C\alpha})$-coupling constants and the violations calculated for the 20 NMR model structures and the trajectories listed above. The $^3J(H_N\text{-}H_{C\alpha})$-coupling constants were calculated for the simulated and model structures using the Karplus relation (Eq. 1) with the parameters $a = 6.4$ Hz, $b = -1.4$ Hz and $c = 1.9$ Hz (Pardi et al. 1984). The secondary structure assignment was done with the program DSSP, based on the Kabsch-Sander rules (Kabsch and Sander 1983). For the visual analysis the VMD program was used (Humphrey et al. 1996).

Results and discussion

NMR model structures

The solution NMR structure of the C-terminal peptide of GCN4-p1, denoted GCN4p16–31 (Steinmetz et al. 2007) is represented (Berman et al. 2000) as a set of 20 model structures, which were obtained using a simulated annealing approach with the program X-PLOR (Schwieters et al. 2003). Analysis of the NOE distances and $^3J(H_N\text{-}H_{C\alpha})$-coupling constants performed on the NMR model structures shows that these satisfy the set of NOE distance bounds with minor violations associated with the following proton pairs: H_N-H_α of the residues 21Asn and 17Tyr, H_N-H_γ of the residues 21Asn and 25Arg, H_ε-H_δ and H_δ-H_δ of the residues 18His and 19Leu, H_δ-H_β and H_δ-H_γ of the residues 18His and 20Glu and H_ε-H_γ of the residues 17Tyr and 20Glu. The violations do not exceed 0.1 nm and are thus not very significant (Fig. 3, panel A and Online Resource, Table S1). However, a comparison of the $^3J(H_N\text{-}H_{C\alpha})$-coupling constants that were back-calculated from the set of 20 NMR model structures with the corresponding experimental values shows that the calculated $^3J(H_N\text{-}H_{C\alpha})$-coupling constants for the residues 18His, 19Leu, and 23Val deviate from the measured ones by more than 1.5 Hz, i.e. by 4.0, 3.4 and 1.8 Hz, respectively (Fig. 3, panel B and Online Resource, Table S2). The very poor agreement of these ^{3}J-values with the experimental ones is most likely due to the assumption of standard α-helical hydrogen bonds and ϕ-angle restraints. These assumed restraints bias the sampling of the GCN4p16–31 conformational space towards the α-helical region, producing a set of closely related structures which violate the primary, i.e. measured data. In addition, these restraints restrict the structural heterogeneity of the peptide. The RMSF values which range from 0.01 to 0.29 nm for the heavy atoms of the backbone and from 0.03 to 0.35 nm for the heavy atoms of the side chains (Online Resource, Figure S2) indicate a restricted conformational variability.

Unrestrained MD simulations

Unrestrained MD simulations were carried out to test the performance of the GROMOS force fields 43A1 and 53A6 regarding the GCN4p16-31 peptide. From panels C, D, E and F of Fig. 3 it is evident that the unrestrained trajectories do not satisfy all experimental NOE upper bounds and nor do they reproduce the $^3J(H_N\text{-}H_{C\alpha})$-coupling constants well. Of the 179 NOE distance bounds, the unrestrained_43A1 simulations violated 24 by more than 0.1 nm; the largest violation of 0.52 nm arises from the proton pair H_ε-H_δ in the side chains of the residues 17Tyr and 19Leu (Fig. 3, panel C). Despite the NOE distance bound violations the helical structure of the peptide is preserved. However, a transition from an α- to a π-helix, which had already been observed in the previously reported MD simulations of GCN4p16–31 (Missimer et al. 2005), occurred in the first 7 ns (Online Resource, Figure S3). Large violations of NOE distance bounds are also a prominent result of the unrestrained_53A6 simulation. 20 NOE distances are violated by more than 0.1 nm, with the largest violation of 0.45 nm arising from the protons H_N and H_α of the residues 21Asn and 17Tyr (Fig. 3, panel E). We note, however, that this NOE assignment was ambiguous. The helical structure of the peptide is only preserved for the central residues where transitions from an α- into a π-helix and back can be observed (Online Resource, Figure S3). The calculated $^3J(H_N\text{-}H_{C\alpha})$-coupling constants do not agree with the measured ones although they show an improvement relative to the set of NMR model structures.

In the unrestrained_43A1 simulation 5 ^{3}J-coupling constants deviate from the measured ones by more than 1 Hz, and in the unrestrained_53A6 simulation deviations greater than 1 Hz occur for 7 ^{3}J-coupling constants. These observations indicate that both unrestrained MD simulations sample regions of conformational space which are not compatible with the primary experimental data. This may be due to the limited accuracy of the force fields as in examples B, D and E of Fig. 2 or due to insufficient sampling of the conformational space as in examples D, E and F.

NOE_IR and NOE_TAR simulations

In the restrained MD simulations NOE_IR_43A1 and NOE_TAR_43A1, 179 experimental NOE upper distance bounds were imposed either as instantaneous or as time-averaged distance restraints. As expected, the agreement with the experimental NOE data improved significantly (Fig. 3, panels G and I). However, the attempt to satisfy all the NOE distances using instantaneous distance restraining resulted in 10 NOE distances violated by more than 0.05 nm with the largest violation of 0.12 nm coming from the protons H_δ and H_β of the residues 18His and 20Glu. Interestingly, these violations disappear when time-averaged distance restraints are applied (Fig. 3, panel I), which shows that if the interproton distance bounds for GCN4p16–31 derived from experiment are correct, they do not correspond to a single structure but represent an average over several different conformations. As indication of the conformational variability of the NOE_IR_43A1 and NOE_TAR_43A1 ensembles, atom-positional RMS fluctuations of the backbone and side-chain heavy atoms of all residues were calculated. The RMS fluctuations of the backbone vary from 0.04 to 0.31 nm for the NOE_IR_43A1 ensemble and from 0.06 to 0.30 nm for the NOE_TAR_43A1 ensemble; the RMS fluctuations of the side-chain atoms vary from 0.05 to 0.29 nm for the NOE_IR_43A1 ensemble and from 0.08 to 0.44 for the NOE_TAR_43A1 ensemble. The comparison indicates a greater conformational variability of the side chains at the N-terminal end in the NOE_TAR_43A1 ensemble (Online Resource, Figure S2). Similar observations apply to the RMS deviations from the energy minimized starting structure which lie roughly between 0.04 and 0.23 nm for the backbone atoms of the NOE_IR_43A1 ensemble and between 0.05 and 0.38 nm for the backbone atoms of the NOE_TAR_43A1 ensemble. The RMSD of the side chains are between 0.16 and 0.40 nm and between 0.21 and 0.60 nm for the NOE_IR_43A1 and for the NOE_TAR_43A1 ensembles, respectively (Online Resource, Figure S1). Despite the increased conformational heterogeneity in the NOE_TAR_43A1 simulations, no significant variability in terms of secondary structure assignment is observed (Online Resource, Figure S3). The improved agreement with the experimental NOE upper distance bounds yielded by the NOE_TAR_43A1 trajectories did not significantly improve the agreement of the calculated $^3J(H_N\text{-}H_{C\alpha})$-coupling constants with the experimental ones. The NOE_IR_43A1 simulation yielded 9 ^{3}J-coupling constants violating the measured ones by more than 1 Hz, the NOE_TAR_43A1 simulation 8 such violations (Fig. 3, panels H and J), indicating that the entire set of the experimental NMR NOE and ^{3}J-coupling data can only be satisfied by also imposing $^3J(H_N\text{-}H_{C\alpha})$-coupling constant restraints in the NOE_TAR simulations.

^{3}J_IR and ^{3}J_LE simulations

MD simulations applying $^3J(H_N\text{-}H_{C\alpha})$-coupling constant restraints either as instantaneous restraints or by using LE potential energy terms yielded calculated ^{3}J-coupling constants close to the experimental values (Fig. 3, panels L and N). The ^{3}J_IR_43A1 simulation, yielding an average deviation of 0.2 Hz, reproduced the experimentally measured $^3J(H_N\text{-}H_{C\alpha})$-coupling constants better than the ^{3}J_LE_43A1 simulation, for which the average deviation was 0.6 Hz. Although ^{3}J_IR_43A1 and ^{3}J_LE_43A1 succeeded in satisfying the experimental $^3J(H_N\text{-}H_{C\alpha})$-coupling constants, neither of the restraining methods succeeded in satisfying the experimental NOE distance bounds (Fig. 3, panels K and M). The ^{3}J_IR_43A1 simulation violated 23 NOE distance bounds by more than 0.1 nm with the largest violation of 0.42 nm coming from the protons H_ε and H_δ of the 17Tyr and 19Leu side chains. The ^{3}J_LE_43A1 simulations produced even more pronounced NOE distance bound violations; 29 NOE distances were violated by more than 0.1 nm with the largest violation of 0.48 nm again coming from the H_ε and H_δ protons of the 17Tyr and 19Leu side chains. Moreover, in contrast to the secondary structure assignments derived from the ^{3}J_IR_43A1 simulation, the assignments derived from the ^{3}J_LE_43A1 simulation reveal a major loss of the α-helical structure in the peptide (Online Resource, Figure S3). Evidently, ^{3}J-coupling constant restraining using LE biasing accesses regions of GCN4p16-31 conformational space featuring relatively great configurational freedom while compatible with the measured ^{3}J-coupling constants, but inaccessible to the simulation using instantaneous ^{3}J-coupling constants restraints. Examples C and F of Fig. 2 illustrate this phenomenon. The ^{3}J_IR simulations, although reproducing all the experimental ^{3}J-coupling constant values, do not sample configurations separated by high energy barriers, thereby restraining the molecule to an unrealistic average conformation. In contrast, the search enhancement techniques such as local elevation allow the system to escape

from the local minima, leading to an improved searching efficiency. Exploration of more extensive regions of configurational space by the ^{3}J_LE_43A1 simulation is also indicated by increased atom-positional RMS fluctuations and RMSD of the backbone and side-chain atoms. The range of the backbone RMSF increases from between 0.06 and 0.32 nm for the ^{3}J_IR_43A1 simulations to between 0.15 and 0.40 nm for the ^{3}J_LE_43A1 simulations; the range of side-chain RMSF increases from between 0.10 and 0.70 nm to between 0.24 and 0.77 nm. The range of backbone RMSD increases from between 0.05 and 0.37 nm for the ^{3}J_IR_43A1 simulations to between 0.06 and 0.45 nm for the ^{3}J_LE_43A1 simulations, and the range of side-chain RMSD from between 0.23 and 0.72 nm to between 0.21 and 0.92 nm, respectively (Online Resource, Figures S1 and S2).

NOE_TAR + ^{3}J_IR and NOE_TAR + ^{3}J_LE simulations

The results discussed above indicate that neither time-averaged NOE restraints nor ^{3}J-coupling constant restraints are alone sufficient in MD structure refinement protocols to reproduce the entire set of 194 experimental NMR data. As is evident from the panels A to D and G to J in Fig. 4, both refinement protocols imposing the two sets of restraints, NOE_TAR + ^{3}J_IR and NOE_TAR + ^{3}J_LE, successfully reproduce all experimental data. The similar results obtained for the NOE_TAR + ^{3}J_IR_43A1 and NOE_TAR + ^{3}J_IR_53A6 simulations as well as for the NOE_TAR + ^{3}J_LE_43A1 and NOE_TAR + ^{3}J_LE_53A6 simulations show that the results of the two refinement protocols are insensitive to the differences between two recent GROMOS force fields.

In order to illustrate the conformational differences among the set of NMR model structures and the ensembles of the NOE_TAR + ^{3}J_IR_53A6 and NOE_TAR + ^{3}J_LE_53A6 simulations, a superposition of first 10 NMR model structures and superpositions of 10 conformations taken at intervals of 1 ns from the NOE_TAR + ^{3}J_IR_53A6 and NOE_TAR + ^{3}J_LE_53A6 simulations are presented in Fig. 5. The conformational space of the peptidic backbone generated by the NOE_TAR + ^{3}J_LE refinement is larger than those of the NOE_TAR + ^{3}J_IR or conventional NMR refinements. These two refinement protocols restrict the sampled configuration space, preventing the structure from deviating markedly from the initial α-helical conformation, either by imposing instantaneous ^{3}J-coupling constant restraints or by explicitly imposing α-helical hydrogen-bond and torsional-angle restraints. The differences in the conformational space sampled by NOE_TAR + ^{3}J_IR and NOE_TAR + ^{3}J_LE simulations using the 43A1 and 53A6 force fields are also

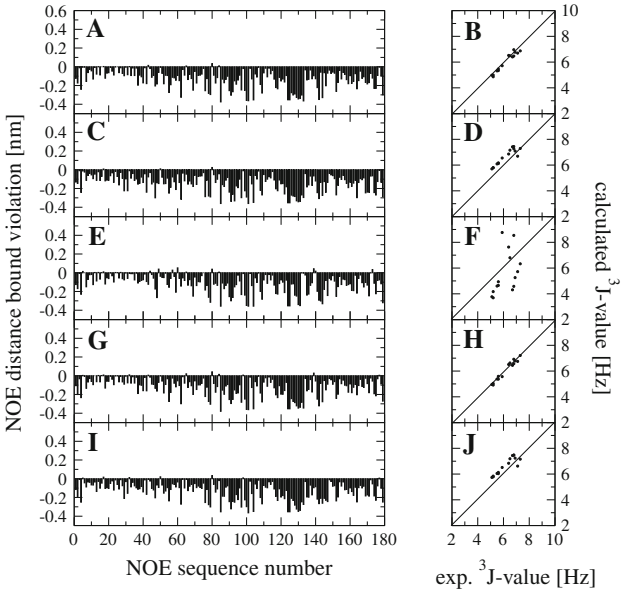

Fig. 4 Violations of the experimental NOE upper distance bounds as a function of the NOE sequence number (*left-hand panels*) and comparison of the experimental and calculated $^3J(H_N\text{-}H_{C\alpha})$-coupling constants (*right-hand panels*) for the following 5 simulations: NOE_TAR + ^{3}J_IR_43A1 (*panels A and B*), NOE_TAR + ^{3}J_LE_43A1 (*panels C and D*), NOE_TAR + ^{3}J_LE_43A1 (panels E and F), NOE_TAR + ^{3}J_IR_53A6 (*panels G and H*) and NOE_TAR + ^{3}J_LE_53A6 (*panels I and J*). Simulation nomenclature is given in Table 1 and NOE sequence numbers in Table 2 as well as in Table S1 (Online Resource)

reflected in the secondary structure analysis presented in Fig. 6. When ^{3}J-coupling constant restraints are applied as instantaneous restraints, GCN4p16-31 remains stable as an α-helix. On the other hand, the variation in secondary structure is larger in the NOE_TAR + ^{3}J_LE simulation demonstrating that the measured NOE distance bounds and ^{3}J-coupling constants permit substantial flexibility in the backbone of GCN4p16–31 and do not restrict it to a rigid α-helical conformation. The differences in the ensembles simulated by the two protocols are, particularly in the case of the 43A1 force field, also evident from the RMS fluctuations of the backbone and side-chain atoms as well as from the RMS deviations of the backbone and the side-chains from the energy minimised starting structure of the GCN4p16–31. The range of the backbone RMSF increases from between 0.07 and 0.35 nm for the NOE_TAR + ^{3}J_IR_43A1 simulation to between 0.12 and 0.45 nm for the NOE_TAR + ^{3}J_LE_43A1 simulation; the range of the side-chain RMSF increases from between 0.09 and 0.54 nm to between 0.17 and 0.65 nm. The range of the backbone RMSD increases from between 0.06 and 0.35 for the NOE_TAR + ^{3}J_IR_43A1 simulation to between 0.09 and 0.43 nm for NOE_TAR + ^{3}J_LE_43A1 simulation; the range of the side-chain RMSD increases from between

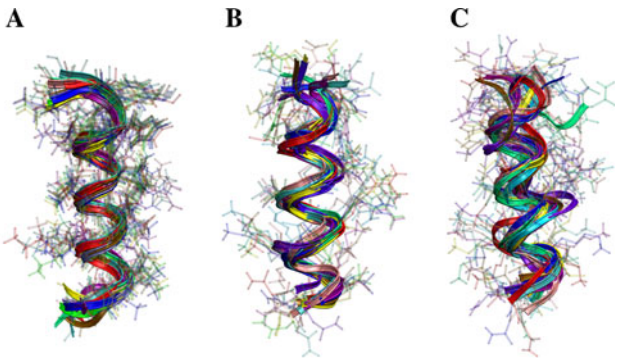

Fig. 5 Superposition of 10 NMR model structures of GCN4p16–31 (**a**), 10 conformations taken from the NOE_TAR + ^{3}J_IR_53A6 trajectories (**b**) and from the NOE_TAR + ^{3}J_LE_53A6 trajectories (**c**) at regular intervals of 1 ns. The structures are superimposed using the heavy atoms of the backbone of the first model or trajectory structure

0.18 and 0.63 to between 0.19 and 0.94, respectively (Online Resource, Figures S1 and S2).

The application of LE biasing may serve a dual function in a simulation: (1) enhancing sampling by enabling the transition of barriers much larger than k_BT; (2) compensating force-field deficiencies by building up Gaussian potential energy hills. In order to investigate these effects of LE biasing we performed a 5 ns long restrained MD simulation (NOE_TAR + -^{3}J_LE_43A1) using only time averaged NOE distance restraints starting from the final coordinates of the NOE_TAR + ^{3}J_LE_43A1 simulation. The results presented in panels E and F of Fig. 4 show that removal of the LE biasing from the NOE_TAR + ^{3}J_LE_43A1 simulation yielded deviations of the calculated ^{3}J-coupling constants from the experimentally measured ones comparable to the deviations of NOE_TAR_43A1, demonstrating that either the force field used does not favour the real conformations (Fig. 2, cases B, D, E) or that LE biasing of the conformational search is needed to enable sampling over a high barrier (Fig. 2, case F).

In Figs. 7, 8, 9 we present the time evolution of the 15 ϕ dihedral angles and $^3J(H_N\text{-}H_{C\alpha})$-coupling constants as well as the build-up of LE biasing potential energy during the NOE_TAR + ^{3}J_LE_53A6 simulation. The experimentally determined $^3J(H_N\text{-}H_{C\alpha})$-coupling constants, J^0, are displayed in the bottom panels of the Figures. Two bands of dihedral angles between 200° and 300° are evident in all but ϕ_{17}; the distance between them tends to decrease with increasing J^0, as the Karplus curve in Fig. 1 would imply. For the dihedral angle ϕ of the first four residues, a third band centered about 60° is evident, corresponding to the lowest maximum of the Karplus curve; a weaker, variable band at lesser dihedral-angle values, characteristic of smaller J^0, is visible for most of the other residues. The distribution of the dihedral angle ϕ is clearly related for each residue to the sampling of the corresponding $^3J(H_N\text{-}H_{C\alpha})$-coupling constant. In the case of the residues 16–18, 21, and 27–30 the experimentally determined $^3J(H_N\text{-}H_{C\alpha})$-coupling constants, J^0, are relatively large, corresponding to the upper part of the Karplus curve, while the $^3J(H_N\text{-}H_{C\alpha})$-coupling constants for the central residues

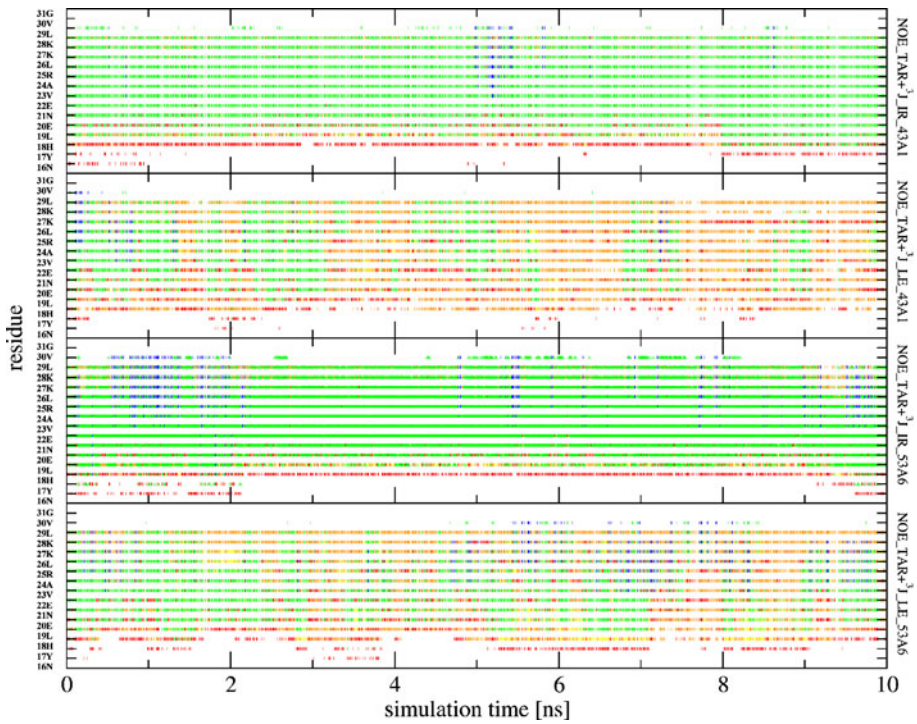

Fig. 6 Time series of secondary structure elements for NOE_TAR + ^{3}J_IR_43A1, NOE_TAR + ^{3}J_LE_43A1, NOE_TAR + ^{3}J_IR_53A6 and NOE_TAR + ^{3}J_LE_53A6 simulations of GCN4p16-31. α-helix is displayed in *green*, 3_{10}-helix in *yellow*, π-helix in *blue*, bend in *orange* and turn in *red*. Simulation nomenclature is given in Table 1

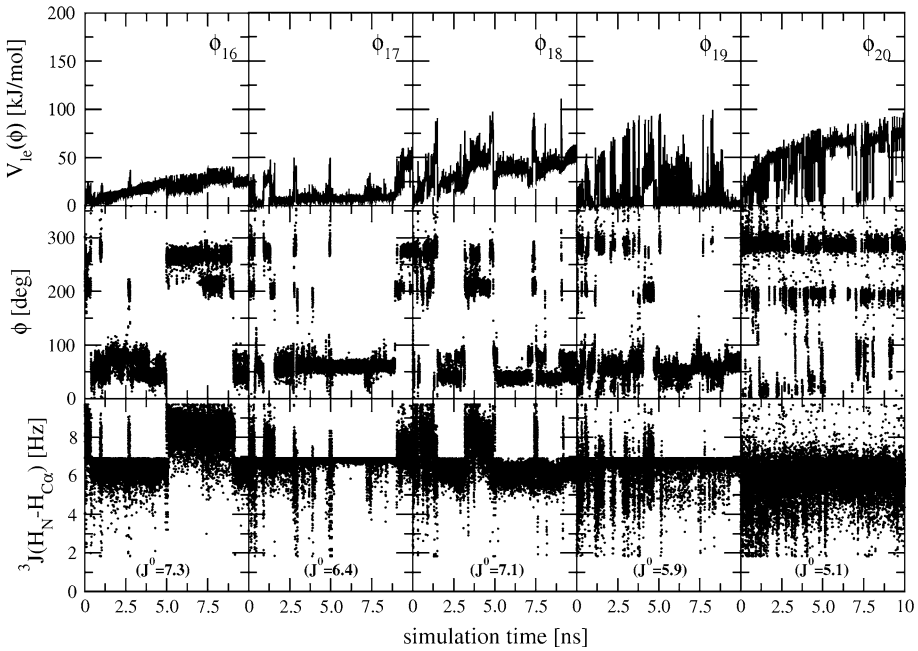

Fig. 7 Time series of the local elevation potential energy (*upper panels*), dihedral angle ϕ (*middle panels*) and $^3J(H_N\text{-}H_{C\alpha})$-coupling constants (*bottom panels*) for residues 16–20 in the NOE_TAR + ^{3}J_LE_53A6 simulation. In the *bottom panels* the experimental $^3J^0(H_N\text{-}H_{C\alpha})$-coupling constants are given for each of the angles

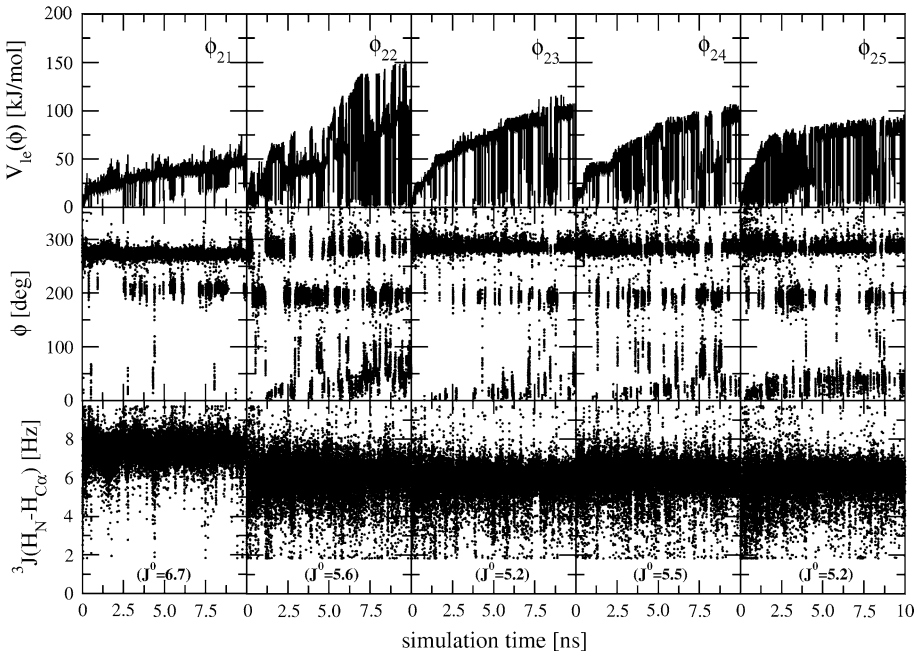

Fig. 8 Time series of the local elevation potential energy (*upper panels*), dihedral angle ϕ (*middle panels*) and $^3J(H_N\text{-}H_{C\alpha})$-coupling constants (*bottom panels*) for residues 21–25 in the NOE_TAR + ^{3}J_LE_53A6 simulation. In the bottom panels the experimental $^3J^0(H_N\text{-}H_{C\alpha})$-coupling constants are given for each of the angles

19, 20, and 22–26 are smaller, corresponding to the middle part of the Karplus curve (Fig. 1). The $^3J(H_N\text{-}H_{C\alpha})$-coupling constants and dihedral angles of the residues 16–18 indicate two exclusive sets of configurations, whereas the ^{3}J-coupling constants of the residues 20–30 vary widely about their mean values, suggesting a broad continuum of configurations compatible with the experimental value J^0. The build-up of the biasing potential energy function indicates that the residues 19, 20 and 22–26, evidencing smaller J^0 corresponding to broader regions under the Karplus curve, require enhanced sampling of ^{3}J-values in order to satisfy the experimental data. If the ^{3}J-value is close to the experimental one from the beginning of the simulation, as for residues 21, and 27–30, the build-up of the biasing potential function is small and the corresponding dihedral angle remains close to its starting value.

Figure 10 presents the time evolution of the dihedral angles ϕ and ^{3}J-values derived from the NOE_TAR + ^{3}J_IR _53A6 simulation showing the effect of instantaneous $^3J(H_N\text{-}H_{C\alpha})$-coupling constant restraints on the sampling of the ϕ torsional-angle degrees of freedom. Despite the instantaneous restraining, the first five dihedral

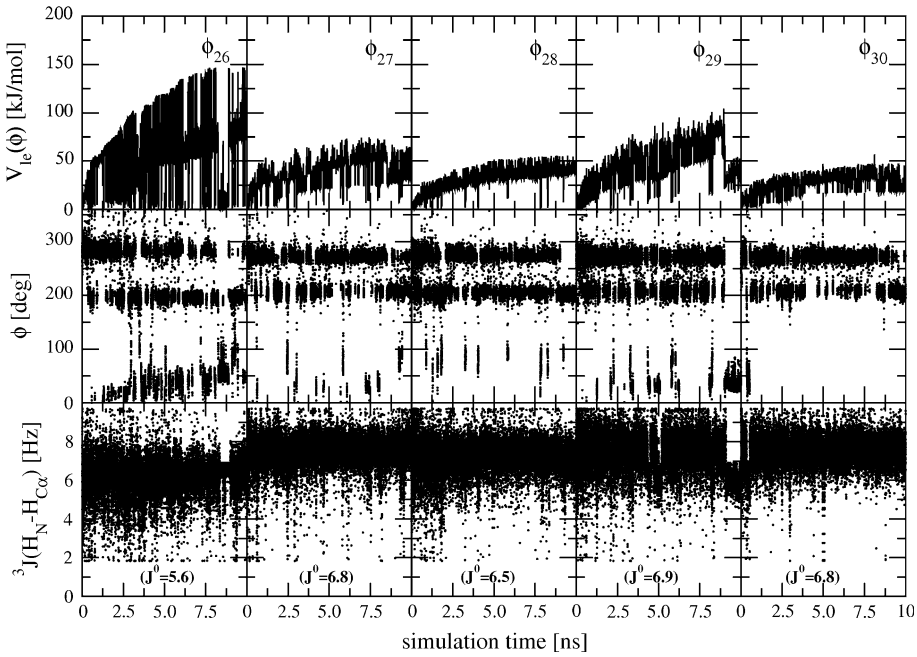

Fig. 9 Time series of the local elevation potential energy (*upper panels*), dihedral angle ϕ (*middle panels*) and $^3J(H_N\text{-}H_{C\alpha})$-coupling constants (*bottom panels*) for residues 26–30 in the NOE_TAR + ^{3}J_LE_53A6 simulation. In the *bottom panels* the experimental $^3J^0(H_N\text{-}H_{C\alpha})$-coupling constants are given for each of the angles

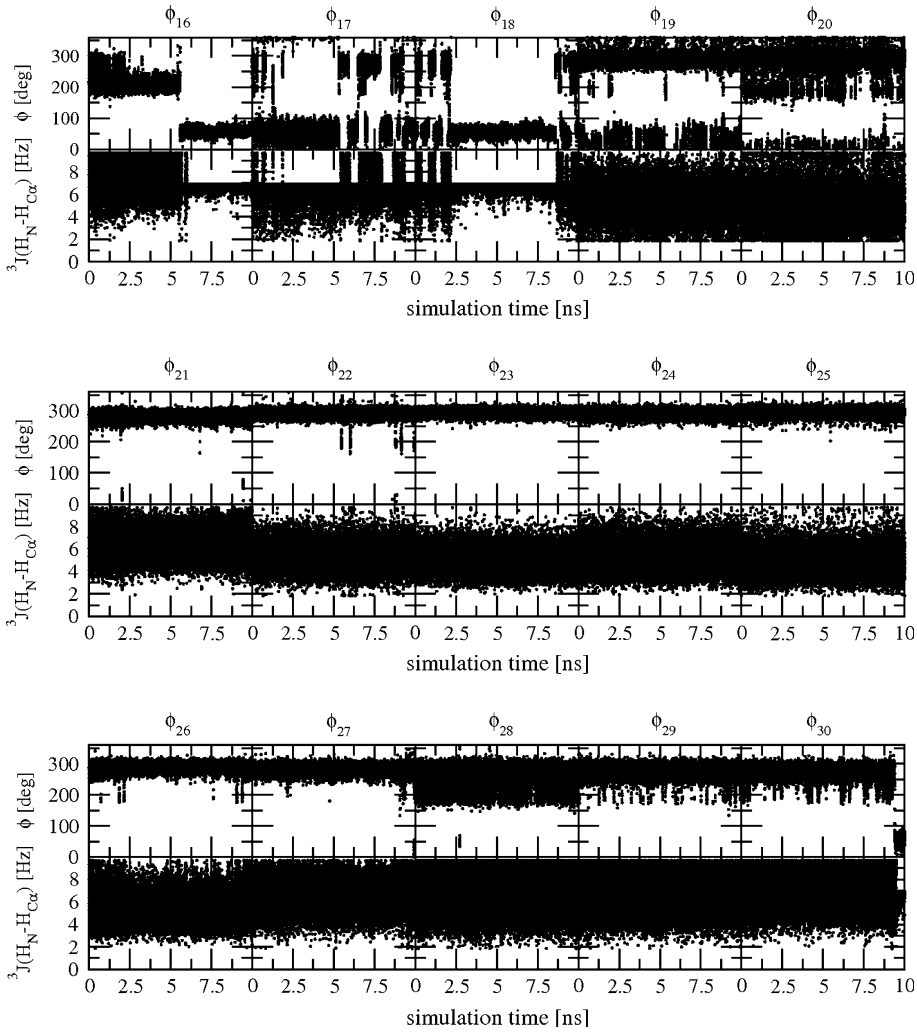

Fig. 10 Time series of the dihedral angle ϕ (*upper panels*) and $^3J(H_N\text{-}H_{C\alpha})$-coupling constants (*bottom panels*) for all residues of GCN4p16–31 in the NOE_TAR + ^{3}J_IR_53A6 simulation

angles ϕ of GCN4p16–31 show larger fluctuations. However, the ϕ angles for all other residues (21–30) sample relatively narrow ranges of the torsional angle space. Yet, the fluctuations of the corresponding ^{3}J-coupling constants are still large.

Conclusions

We have investigated the effect of MD structure refinement protocols on the conformational heterogeneity of the calculated ensembles using the C-terminal peptide of GCN4p1, GCN4p16–31, as an example. The agreement of the simulated with the primary, measured NMR data was used as criterion for success. The choice of GCN4p16–31 was motivated by the observation that the set of 20 NMR model structures deposited in the protein data bank did not completely agree with the measured NMR data on which the single structure refinement was based (Steinmetz et al. 2007). Six NOE upper distance bounds were slightly violated and three ^{3}J-coupling constants disagreed with the experimental values by 1.8–4 Hz. Using restrained MD simulations we could significantly improve the agreement of the conformational ensemble with the measured experimental data. In addition, the following observations emerged from the analysis. i) The 179 NOE upper distance bounds for GCN4p16–31 can only be fully satisfied if the NMR data are included as time-averaged distance restraints in the MD simulations. Thus the NOE signals are averages that cannot be described by a single structure. ii) The 15 experimental ^{3}J-coupling constants are not well reproduced by applying only the NOE distance restraints in the structure refinement, which is due to the limited sampling of the corresponding torsional-angle degrees of freedom. iii) In order to enable the peptide to cross high-energy barriers and to enforce agreement with the experimental ^{3}J-coupling constant values, the sampling of the corresponding ϕ torsional angle degrees of freedom can be enhanced using LE biasing of the conformational search. We find that the 15 $^3J(H_N\text{-}H_{C\alpha})$-coupling constants, which depend only on the torsional angles between H_N and $H_{C\alpha}$ protons, are not sufficient to define the overall structure of GCN4p16–31. iv) Using time-averaged NOE distance restraints in combination with instantaneous ^{3}J-coupling constant restraining in the MD simulation results in a stable α-helical peptide conformation. However, restraining ^{3}J-coupling constants instantaneously, i.e. excluding averaging effects, neglects the basic fact that the results of the NMR measurements are averages over time and space. This is in particular true for ^{3}J-couplings, which depend in a highly non-linear manner on the local conformation. This prompted us to apply time-averaged NOE distance restraints in combination with LE biased ^{3}J-value restraining in the MD simulation. The ensemble of structures calculated in this simulation satisfies all experimental data while including conformations not predicted by the standard single-structure refinement protocol. This result shows that single-structure refinement involving assumptions, such as hydrogen-bond and torsional angle restraints, suggested only indirectly by the measured data, may lead to biomolecular structures not representative of the conformational variability of a biomolecule in aqueous solution. Proper accounting for the average nature of measured observables, avoiding the use of assumed data in the restraint set and implementing proper sampling of the relevant degrees of freedom, are essential ingredients of any procedure to derive biomolecular structure on the basis of measured data.

Acknowledgments Financial support by the National Centre of Competence in Research (NCCR) in structural biology and by grant number 200020-121913 of the Swiss National Science Foundation (SNSF) and by grant number 228076 of the European Research Council (ERC) to W. F. van G., and by the Slovenian Research Agency (ARRS), grant number Z1-9576 to J. D., is gratefully acknowledged. We would like to thank Jane R. Allison for help with the local-elevation biased ^{3}J-coupling restraining, and Andrei Alexandrescu and Wolfgang Jahnke for their constructive criticism of the manuscript.

References

Allison JR, van Gunsteren WF (2009) A method to explore protein side chain conformational variability using experimental data. Chem Phys Chem 10:3213–3228

Bax A, Tjandra N (1997) Are proteins even floppier than we thought? Nat Struct Biol 4:254–256

Beckman RA, Moreland D, Louise-May S, Humblet C (2006) RNA unrestrained molecular dynamics ensemble improves agreement with experimental NMR data compared to single static structure: a test case. J Comput Aided Mol Des 20:263–279

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, The Netherlands, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (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. Nucleic Acids Res 28:235–242

Berndt KD, Güntert P, Wüthrich K (1996) Conformational sampling by NMR solution structures calculated with the program DIANA evaluated by comparison with long-time molecular dynamics calculations in explicit water. Proteins Struct Funct Genet 24: 304–313

Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M (2006) Relation between native ensembles and experimental structures of proteins. Proc Natl Acad Sci USA 103:10901–10906

Bonvin A, Brünger AT (1995) Conformational variability of solution nuclear magnetic resonance structures. J Mol Biol 250:80–93

Bonvin A, Boelens R, Kaptein R (1994) Time-averaged and ensemble-averaged direct NOE restraints. J Biomol NMR 4:143–149

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

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

Cuniasse P, Raynal I, Yiotakis A, Dive V (1997) Accounting for conformational variability in NMR structure of cyclopeptides: ensemble averaging of interproton distance and coupling constant restraints. J Am Chem Soc 119:5239–5248

Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CH_n united atoms of GROMOS96 force field. J Comput Chem 19:535–547

Daura X, Antes I, van Gunsteren WF, Thiel W, Mark AE (1999) The effect of motional averaging on the calculation of NMR-derived structural properties. Proteins Struct Funct and Genet 36:542–555

Fawzi NL, Phillips AH, Ruscio JZ, Doucleff M, Wemmer DE, Head-Gordon T (2008) Structure and dynamics of the Aβ21–30 peptide from the interplay of NMR experiments and molecular simulations. J Am Chem Soc 130:6145–6158

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

Gattin Z, Schwartz J, Mathad RI, Jaun B, van Gunsteren WF (2009) Interpreting experimental data by using molecular simulation instead of model building. Chem Eur J 15:6389–6398

Glättli A, van Gunsteren WF (2004) Are NMR-derived model structures for beta-peptides representative for the ensemble of structures adopted in solution? Angew Chem Int Edit 43:6312–6316

Glättli A, Daura X, van Gunsteren WF (2002) Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. J Chem Phys 116:9811–9828

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 Des 8:695–708

Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:33–38

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

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 (1963) Vicinal proton coupling in nuclear magnetic resonance. J Am Chem Soc 85:2870–2871

Karplus M, McCammon JA (1983) Dynamics of proteins: elements and function. Annu Rev Biochem 52:263–300

Kearsley SK (1989) On the orthogonal transformation used for structural comparisons. Acta Crystallogr Sect A 45:208–210

Keller B, Christen M, Oostenbrink C, van Gunsteren WF (2007) On using oscillating time-dependent restraints in MD simulation. J Biomol NMR 37:1–14

Kessler H, Griesinger C, Lautz J, Müller A, van Gunsteren WF, Berendsen HJC (1988) Conformational dynamics detected by nuclear magnetic resonance NOE values and J-coupling constants. J Am Chem Soc 110:3393–3396

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Mierke DF, Kurz M, Kessler H (1994) Peptide flexibility and calculations of an ensemble of molecules. J Am Chem Soc 116:1042–1049

Missimer JH, Steinmetz MO, Jahnke W, Winkler FK, van Gunsteren WF, Daura X (2005) Molecular dynamics simulations of C- and N-terminal peptide derivatives of GCN4–p1 in aqueous solution. Chem Biodivers 2:1086–1104

Nanzer AP, Poulsen FM, van Gunsteren WF, Torda AE (1994) A reassessment of the structure of chymotrypsin inhibitor 2 (CI-2) using time-averaged NMR restraints. Biochemistry 33:14503–14511

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

Nanzer AP, Huber T, Torda AE, van Gunsteren WF (1996) Molecular dynamics simulation using weak-coupling NOE distance restraining. J Biomol NMR 8:285–291

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

Nilges M, Clore GM, Gronenborn AM (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. FEBS Lett 239:129–136

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

Oostenbrink C, Soares TA, van der Vegt NFA, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. Eur Biophys J Biophys Lett 34:273–284

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

Pearlman DA, Kollman PA (1991) Are time-averaged restraints necessary for nuclear magnetic resonance refinement? a model study for DNA. J Mol Biol 220:457–479

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 Comput Phys 23:327–341

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

Schmitz U, Kumar A, James TL (1992) Dynamic interpretation of NMR data: molecular dynamics with weighted time-averaged restraints and ensemble R-factor. J Am Chem Soc 114:10654–10656

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73

Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. J Phys Chem A 103:3596–3607

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. Proc Natl Acad Sci USA 104:7062–7067

Torda AE, Scheek RM, van Gunsteren WF (1989) Time-dependent distance restraints in molecular dynamics simulations. Chem Phys Lett 157:289–294

Torda AE, Scheek RM, van Gunsteren WF (1990) Time-averaged nuclear overhauser effect distance restraints applied to tendamistat. J Mol Biol 214:223–235

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

Trzesniak D, Glättli A, Jaun B, van Gunsteren WF (2005) Interpreting NMR data for beta-peptides using molecular dynamics simulations. J Am Chem Soc 127:14320–14329

van Gunsteren WF, Berendsen HJC (1990) Computer simulation of molecular dynamics methodology: applications and perspectives in chemistry. Angew Chem Int Edit 29:992–1023

van Gunsteren WF, Mark AE (1998) Validation of molecular dynamics simulation. J Chem Phys 108:6109–6116

van Gunsteren WF, Brunne RM, Gros P, van Schaik RC, Schiffer CA, Torda AE (1994) Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data. In: James TL, Oppenheimer NJ (eds) Methods in enzymology: nuclear magnetic resonance. Academic Press, New York, pp 619–654

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

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

Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 17:15–20

Wang AC, Bax A (1996) Determination of the backbone dihedral angles phi in human ubiquitin from reparametrized empirical Karplus equations. J Am Chem Soc 118:2483–2494

Zagrovic B, van Gunsteren WF (2006) Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us? Proteins Struct Funct Bioinf 63:210–218

Zagrovic B, Gattin Z, Lau JKC, Huber M, van Gunsteren WF (2008) Structure and dynamics of two beta-peptides in solution from molecular dynamics simulations validated against experiment. Eur Biophys J Biophys Lett 37:903–912