# Supporting information for: Probabilistic Determination of Native State Ensembles of Proteins.

Simon Olsson,[*,†,⊥] Beat Rolf Vögeli,[‡] Andrea Cavalli,[¶] Wouter Boomsma,[§] Jesper Ferkinghoff-Borg,[‖] Kresten Lindorff-Larsen,[§] and Thomas Hamelryck[*,†]

*Bioinformatics Centre, Department of Biology, Faculty of Science, University of Copenhagen, Denmark, Laboratory of Physical Chemistry, Eidgenössische Technische Hochschule Zürich, 8093 Zürich, Switzerland, Institute for Research in Biomedicine, CH-6500 Bellinzona, Switzerland, Structural Biology and NMR Laboratory, Department of Biology, Faculty of Science, University of Copenhagen, Denmark, and Cellular Signal Integration Group, Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark*

E-mail: solsson@binf.ku.dk; thamelry@binf.ku.dk

## 1 Materials and Methods

**NMR spectroscopy.** GB3 was expressed and purified as described in reference.[1] The NMR samples contained 350 $\mu$l / 500 $\mu$l of 4 mM / 2 mM 13C,15N- / 2H,13C,15N-labeled protein solution in 97%:3% / 95%:5% $H_2O$:$D_2O$, 50 mM potassium phosphate buffer, pH 6.5 / 7.0, and 0.5

[*]To whom correspondence should be addressed
[†]Bioinformatics Centre, Department of Biology, Faculty of Science, University of Copenhagen, Denmark
[‡]Laboratory of Physical Chemistry, Eidgenössische Technische Hochschule Zürich, 8093 Zürich, Switzerland
[¶]Institute for Research in Biomedicine, CH-6500 Bellinzona, Switzerland
[§]Structural Biology and NMR Laboratory, Department of Biology, Faculty of Science, University of Copenhagen, Denmark
[‖]Cellular Signal Integration Group, Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark
[⊥]Institute for Research in Biomedicine, CH-6500 Bellinzona, Switzerland

mg/mL sodium azide. All experiments were performed on a BRUKER DRX600 MHz spectrometer, equipped with a z-axis gradient cryogenic probe, respectively, at 298 K. $R_{HN_iN_i/HN_{i+1}N_{i+1}}$ were obtained from the triple labeled sample and 3D ct-$^{13}$C'-HN(CA)CON / ct-$^{13}$C$\alpha$-HNCA(CO)N versions of a previously proposed 2D experiment[2] recorded with 36(N, $t_1$) x 40(C', $t_2$) / 20(C$\alpha$, $t_2$) x 512(H$^N$, $t_3$) complex points, $t_{1\max} = 18.0$ms, $t_{2\max} = 26.4\{6.6\}$ms, $t_{3\max} = 51.2$ms, an interscan delay of 1s, $\tau_{MQ} = 43$ms, and typically 32 / 96 scans per increment resulting in a measurement time of 2 / 6 days. Because the 'trans' spectra are considerably less sensitive they were typically recorded twice and added. $R_{HA_iCA_i/HA_{i-1}CA_{i-1}}$ were obtained from the 2D HNCA(CA) experiment put forward in reference[3] and a 3D $^{13}$C$\alpha$-HNCA(CO)CA experiment recorded with 64(N, $t_1$) x 512(H$^N$, $t_2$) and 36(N, $t_1$) x 24(C$\alpha$,$t_2$) x 512(H$^N$, $t_3$) complex points, $t_{1\max} = 32.0$ms, $t_{2\max} = 51.2$ms and $t_{1\max} = 18.0$ms, $t_{2\max} = 7.92$ms, $t_{3\max} = 51.2$ms, an interscan delay of 1s, $\tau_{MQ} = 28$ms, and typically 512 and 64 scans per increment resulting in measurement times of 1 and 4 days, respectively. Because the 'trans' spectra are considerably less sensitive than the 'reference' spectra they were typically recorded twice and added. $R_{HN_iN_i/HA_iCA_i}$ were obtained from three experiments. Two experiments used previously were repeated.[4] In addition, a 3D ct-HNCA MMQ experiment adapted from reference[5] was used. The spectrum was recorded with $\tau_{MQ} = 31.0$ms or $\tau_{MQ} = 33.5$ms, 50(MQ[N,C$\alpha$], $t_1$) or 55(MQ[N,C$\alpha$], $t_1$) x 36(N, $t_2$) x 512(H$^N$, $t_3$) complex points, $t_{1\max} = 25.0$ms or $t_{1\max} = 27.5$ms, $t_{2max} = 18.0$ms, $t_{3max} = 63.28$ms, interscan delays of 1s or 0.92s and 32 or 48 scans per increment resulting in measurement times of 3 or 4 days. $R_{HN_iN_i/HA_{i-1}CA_{i-1}}$ were obtained by repetition of three previously used experiments.[4] All time domain data were multiplied with a square cosine function in the direct dimension and cosine functions in the indirect dimensions and zero-filled to standard numbers. All spectra were processed and analyzed using the software package NMRPipe.[6] Corrections to the apparent CCR rates were calculated with a full matrix analysis as outlined in reference.[7]

**Profasi forcefield** The Profasi forcefield consists of four terms[8]

$$E_{\text{prof}}(\mathbf{x}) = E_{\text{loc}}(\mathbf{x}) + E_{\text{ev}}(\mathbf{x}) + E_{\text{hb}}(\mathbf{x}) + E_{\text{sc}}(\mathbf{x})$$

$E_{\mathrm{loc}}(\mathbf{x})$ describes local interactions, that is, interactions between atoms separated by only a few covalent bonds. The remaining three terms account for *non-local* interactions, such as excluded-volume effects ($E_{\mathrm{ev}}(\mathbf{x})$) and hydrogen-bonds ($E_{\mathrm{hb}}(\mathbf{x})$). Finally, charge-charge and hydrophobic interactions between side chains are contained in the term $E_{\mathrm{sc}}(\mathbf{x})$. There are no explicit terms for bond-lengths and bond-angles. Here, we overcome this by using ideal bond-lengths and using a narrow distribution of bond-angles centered at the idealized values.[9]

**Replica averaged molecular dynamics simulations** Similar to what was previously described,[10] a replica averaged potential ($E_{\mathrm{RDC}} = \kappa(D^{\mathrm{exp}} - D^{\mathrm{calc}})^2$) was implemented into the molecular dynamics framework Almost.[11] $D^{\mathrm{calc}}$ constitutes average residual dipolar couplings back-calculated from $N$ replicas. Briefly, in each step, an average Saupe tensor was calculated using averaged bond vector projections to solve a linear system of equations by singular value decomposition.[12] One tensor was used for each alignment condition. $D^{\mathrm{exp}}$ constitutes the experimental RDC data and $\kappa$ an empirical force constant. The implementation is available for download as part of the Almost 2.2 branch on Sourceforge.

We simulated GB3 in the Amber 03 forcefield[13] restrained by the potential described above yielding the effective potential,

$$E = E_{\mathrm{Amber03}} + E_{\mathrm{RDC}}, \tag{1}$$

for each replica. We initialized the simulation from the native structure (pdb: 2OED) after being subject to 1000 steps of steepest descent energy minimization in the Amber03 forcefield.[13] The number of replicas $N$ and the force constant $\kappa$ were optimized empirically by running a number of short trajectories with $N = \{2,4,8\}$ $\kappa = \{N \cdot 5 \cdot 10^{-2}, N \cdot 10^{-2}, 5 \cdot N \cdot 10^{-3}, N \cdot 10^{-3}\}$. Optimality was assessed as a tradeoff between computational demands, stability and fit of RDC data, also considering previous studies. A production simulation was subsequently carried out with the optimal parameter set $N = 4, \kappa = 4 \cdot 10^{-2}$. Using the optimal parameter set we performed 4 independent simulations of 1.1 ns each, corresponding to 17.6 ns in total. This corresponds to the computational resources spent by the restraining method presented here. Average statistics from

these simulations were used in the analysis. 1 fs time-steps were used, and the average temperature was kept at 300 K using a Berendsens thermostat.[14] Bonds involving protons were constrained using the SHAKE algorithm.[15] The simulations were conducted using a Generalized-Born implicit solvent model.[16]

**Cross-correlated relaxation rates** For flexible macromolecules, assuming separation of angular and radial dynamics and anisotropic overall tumbling,[17] the dipolar cross-correlated relaxation rate may be approximated by,[18,19]

$$R_{X-Y,U-V} = \frac{2}{5}\frac{\gamma_X\gamma_Y}{\langle r_{X-Y}^3\rangle}\frac{\gamma_U\gamma_V}{\langle r_{U-V}^3\rangle}(\frac{\hbar\mu_0}{4\pi})^2\sum_{k=-2}^{2}\langle \mathbf{C}_k\rangle\tau_k \tag{2}$$

where $\gamma_X$ is the gyromagnetic ratio of nuclei $X$, $\hbar$ is Plancks constant divided by $2\pi$, $\mu_0$ is the permeability of free space and $r_{X-Y}$ is the inter-atomic bond length between nuclei $X$ and $Y$. $\tau_k$, which describe the anisotropic overall diffusion are given by,

$$\begin{aligned}
1/\tau_{-2} &= 6D + \sqrt{D^2 + D'^2} \\
1/\tau_{-1} &= D_x + D_y + 4D_z \\
1/\tau_0 &= 6D + \sqrt{D^2 - D'^2} \\
1/\tau_1 &= 4D_x + D_y + D_z \\
1/\tau_2 &= D_x + 4D_y + D_z
\end{aligned}$$

with, $D' = \sqrt{\frac{D_xD_y+D_xD_z+D_yD_z}{3}}$ and $D = \frac{D_x+D_y+D_z}{3}$. $\mathbf{C}_k$, which describe the dependency on the orientation of bonds are given by,

$$\begin{aligned}
\mathbf{C}_2 = \; & \tfrac{3w^2}{4N^2}\sin^2(\theta_A)\sin^2(\theta_B)\cos(2\phi_A)\cos(2\phi_B) + \tfrac{\sqrt{3}\mu w}{4N^2}[\sin^2(\theta_A)\cos(2\phi_A)(3\cos^2(\theta_B)-1) \\
& + \sin^2(\theta_B)\cos(2\phi_B)(3\cos^2(\theta_A)-1)] + \tfrac{\mu^2}{4N^2}(3\cos^2(\theta_A)-1)(3\cos^2(\theta_B)-1) \\
\mathbf{C}_{-2} = \; & \tfrac{3}{4}\sin^2(\theta_A)\sin^2(\theta_B)\sin(2\phi_A)\sin(2\phi_B) \\
\mathbf{C}_0 = \; & \tfrac{3\mu^2}{4N^2}\sin^2(\theta_A)\sin^2(\theta_B)\cos(2\phi_A)\cos(2\phi_B) - \tfrac{\sqrt{3}\mu w}{4N^2}[\sin^2(\theta_A)\cos(2\phi_A)(3\cos^2(\theta_B)-1) \\
& + \sin^2(\theta_B)\cos(2\phi_B)(3\cos^2(\theta_A)-1)] + \tfrac{w^2}{4N^2}(3\cos^2(\theta_A)-1)(3\cos^2(\theta_B)-1) \\
\mathbf{C}_1 = \; & \tfrac{3}{4}\sin(2\theta_A)\sin(2\theta_B)\sin(\phi_A)\sin(\phi_B) \\
\mathbf{C}_{-1} = \; & \tfrac{3}{4}\sin(2\theta_A)\sin(2\theta_B)\cos(\phi_A)\cos(\phi_B)
\end{aligned}$$

with $\mu = \sqrt{3}(D_x - D_y)$, $w = 2D_z - D_x - D_y + 2\Delta$, $\Delta = 3\sqrt{D^2 + D'^2}$ and $N = 2\sqrt{\Delta w}$. $\langle\cdot\rangle$ denotes ensemble averaging. We used the previously reported anisotropic diffusion tensor.[17]

**Hydrogen bond scalar couplings** The through hydrogen bond scalar couplings, ${}^{h3}J_{NC'}$, between $^{15}$N-$^{13}$C nuclei were calculated using the density functional theory derived equations provided by Barfield,[20]

$$\begin{aligned}
{}^{h3}J_{NC'} = (&-1.31\cos^2\theta_2 + [0.62\cos^2\rho + 0.92\cos\rho \\
&+0.14]\sin^2\theta_2)\exp[-3.2(r_{HO'} - r^0_{HO'})] + 0.01\text{Hz}
\end{aligned} \tag{3}$$

where $r_{HO'}$, $\theta_2$ and $\rho$ are the H$\cdots$O hydrogen bond length, HO'C angle and the H-O'-C'-N dihedral angle, respectively. The empirical constant $r^0_{HO'}$ is fixed at 1.760 Å. In the case of ensemble models, the HBC was calculated for all members and the average quantity was used for comparison with experimental data.

**J-Couplings** Back-calculation of J-couplings was performed using the Karplus equation,[21,22]

$$^3J_\theta = A\langle\cos^2(\theta)\rangle + B\langle\cos(\theta)\rangle + C \tag{4}$$

where $\theta$ is the dihedral angle spanned by the four nuclei giving rise to the scalar-coupling, $^3J_\theta$. The Karplus parameters ($A, B$ and $C$) were fitted using ordinary least squares to the structural models. $\langle\cdot\rangle$ denotes an ensemble average.

**Residual dipolar couplings** Residual dipolar couplings were back-calculated as described previously.[23] Quantitative assessment of the fit to experimental data was carried out using the Q-factor as defined by Bax.[24]

**Exact nuclear Overhauser enhancements** Unambiguous exact nuclear Overhauser derived distances were correlated with back-calculated average distances. The averaged distances were back-calculated using power-averaging, justified by the assumption that the conformational dynamics is either much faster ($\text{eNOE}_{-3}$) or much slower ($\text{eNOE}_{-6}$) than the molecular tumbling. The power-averaging was using the relationship,[25]

$$\text{eNOE}_{-n} \propto \langle r^{-n}\rangle, \tag{5}$$

where $r$ is the inter-atomic distance corresponding to a particular eNOE.

**Training dataset details** A total of 413 previously published experimental RDCs were used to restrain the simulations in the EM algorithm. Standard deviations were uniformly assumed to be 1Hz.

**$\mathbf{B_k}$ estimation details** We wish to estimate the parameter $\mathbf{B}_{k+1}$ in order to accomplish $\overline{\mathbf{f}(\mathbf{x})} = \mathbf{e}$. This can be done using an EM algorithm, by minimizing the following expression in the M-step,

$$\arg\min_{\mathbf{B}_{k+1}} \sum_i \left\| \mathbf{e}_i - \overline{\mathbf{f}(\mathbf{x})_{i,\mathbf{B}_{k+1}}} \right\|^2 \tag{6}$$

where $\overline{\mathbf{f}(\mathbf{x})_{i,\mathbf{B}_{k+1}}}$ is the average of $\mathbf{f}(\mathbf{x})$ for a given $\mathbf{e}_i$ and $\mathbf{B}_{k+1}$, and $i$ runs over the number of samples of $\mathbf{e}_i$.

We may obtain $\overline{\mathbf{f}(\mathbf{x})_{i,\mathbf{B}_{k+1}}}$ by importance sampling using the $N$ samples $\mathbf{f}(\mathbf{x})_i, \mathbf{e}_i$ according to the following expression,

$$\overline{\mathbf{f}(\mathbf{x})_{i,\mathbf{B}_{k+1}}} = \frac{1}{N} \sum_j \mathbf{f}(\mathbf{x})_j \frac{\mathscr{N}(\mathbf{d} \mid \mathbf{e}_i, \sigma)\mathscr{G}(\mathbf{f}(\mathbf{x})_j \mid \mathbf{e}_i, \mathbf{B}_{k+1})}{\mathscr{N}(\mathbf{d} \mid \mathbf{e}_j, \sigma)\mathscr{G}(\mathbf{f}(\mathbf{x})_j \mid \mathbf{e}_j, \mathbf{B}_k)} \tag{7}$$

where $\mathbf{B}_k$ is the value for $\mathbf{B}_{k+1}$ used in obtaining the samples $\mathbf{f}(\mathbf{x})_i, \mathbf{e}_i$, and the sum runs over all samples. This corresponds to "replacing" $\mathbf{B}_k$ and $\mathbf{e}_j$ by $\mathbf{B}_{k+1}$ and $\mathbf{e}_i$ in the posterior (Eq. 1, in the main text), according to the principles of importance sampling.

We can approximate and simplify Eq. 6 by assuming that $\mathbf{e}$ does not vary much – in the limit of no experimental uncertainty (Maximum entropy restraining), this is exactly the case. In that case, instead of summing over all samples $\mathbf{e}_i$, in Eq. 6 we can just use the average of the samples $\bar{\mathbf{e}} = \frac{1}{N}\sum_i \mathbf{e}_i$. This causes the $\mathscr{N}(\mathbf{d} \mid \cdot, \sigma)$ factors in Eq. 7 cancel, as $\mathbf{e}_i = \mathbf{e}_j = \bar{\mathbf{e}}$.

As we can't calculate the exact normalization factors $\mathscr{Z}$ for $\mathscr{G}(\cdot)$, we can't use the expression (7) directly. Instead, we simply calculate (unnormalized) importance weights as follows, as this does not require the normalization factors,

$$w_{i,j,\mathbf{B}_{k+1}} \propto \frac{\exp(\mathbf{f}(\mathbf{x})_j^T \mathbf{B}_{k+1} \mathbf{e}_i)}{\exp(\mathbf{f}(\mathbf{x})_j^T \mathbf{B}_k \mathbf{e}_j)}. \tag{8}$$

We subsequently approximate the normalization constant with the expression

$$\hat{\mathscr{Z}} = \sum_{j}^{N} \exp(\mathbf{f}(\mathbf{x})_j^T (\mathbf{B}_{k+1} - \mathbf{B}_k)\mathbf{e}_i), \tag{9}$$

and finally evaluate the expectation

$$\overline{\mathbf{f}(\mathbf{x})_{i,\mathbf{B}_{k+1}}} = \hat{\mathscr{Z}} \sum_{j} \mathbf{f}(\mathbf{x})_j w_{i,j,\mathbf{B}_{k+1}}.$$

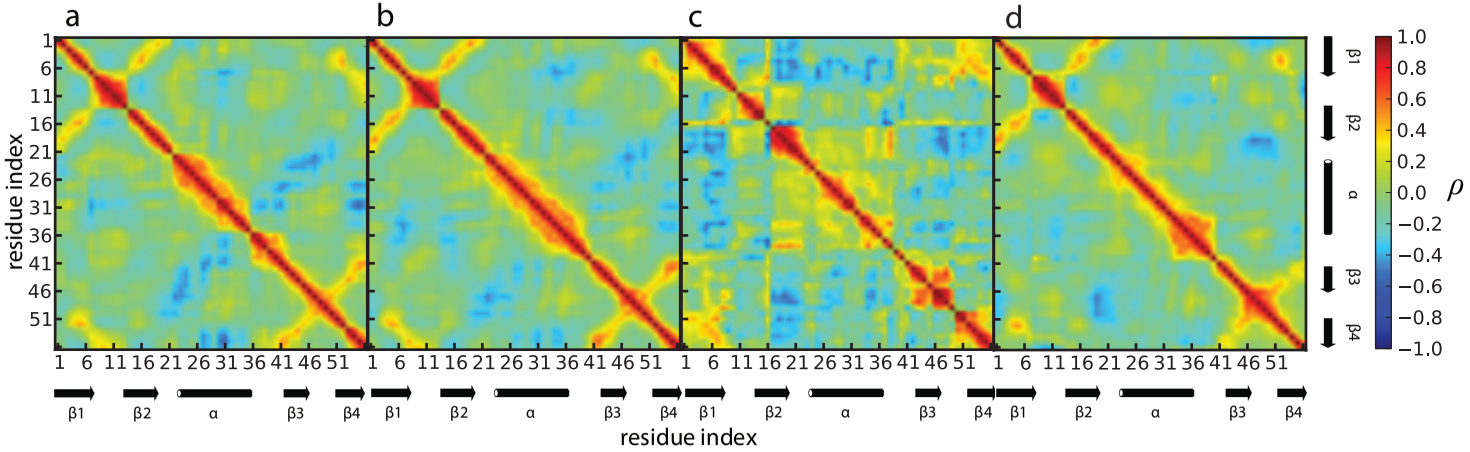


Figure 1: Heatmaps of pairwise correlations, $\rho$, of fluctuations of $C^{\alpha}$ atomic positions along the backbone of GB3. (a) CHARMM22*, (b) Amber ff99SB*-ILDN, (c) 2LUM and (d) the replica averaged molecular dynamics simulations restrained with the same data as $EM_8$ (RAMD). Correlation matrices were calculated using THESEUS.[26]

Table 1: Reproduction of experimental data and other ensemble properties as a function Expectation Maximization step. Model taking into account experimental noise. Pearsons correlation coefficient is denoted by $\rho$.

| EM-step | ENLP | $\langle$RMSD$\rangle$ (Å)[a] | mean Q-factor | $^{h3}J_{NC'}$ (rms, $s^{-1}$) | $\rho(\mathbf{B}_k, \mathbf{B}_{k-1})$[b] | eNOE$_{-3}$ ($\rho$) | eNOE$_{-6}$ ($\rho$) | $^{3}J_{H^N-H^\alpha}$ (rms, $s^{-1}$) | $^{3}J_{H^N-C^\beta}$ (rms, $s^{-1}$) | $^{3}J_{H^N-C'}$ (rms, $s^{-1}$) | $R_{H\alpha C\alpha_i/H\alpha C\alpha_{i+1}}(\rho)$ | $R_{HN_i/HN_{i+1}}(\rho)$ | $R_{HN_i/H\alpha C\alpha_i}(\rho)$ | $R_{HN_i/H\alpha C\alpha_{i-1}}(\rho)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 316.5 | 2.76 | 0.41 | 0.16 | - | 0.69 | 0.66 | 0.97 | 0.51 | 0.66 | 0.96 | 0.63 | 0.90 | 0.94 |
| 1 | 159.1 | 1.66 | 0.17 | 0.13 | 0 | 0.75 | 0.70 | 0.62 | 0.33 | 0.48 | 0.93 | 0.90 | 0.96 | 0.98 |
| 2 | 291.0 | 2.36 | 0.33 | 0.15 | 0.27 | 0.69 | 0.67 | 0.96 | 0.46 | 0.64 | 0.95 | 0.69 | 0.90 | 0.95 |
| 3 | 182.7 | 1.59 | 0.17 | 0.13 | 0.58 | 0.76 | 0.70 | 0.59 | 0.35 | 0.44 | 0.93 | 0.91 | 0.96 | 0.98 |
| 4 | 155.0 | 1.53 | 0.15 | 0.13 | 0.97 | 0.75 | 0.71 | 0.58 | 0.32 | 0.42 | 0.94 | 0.92 | 0.97 | 0.98 |
| 5 | 129.0 | 1.56 | 0.14 | 0.13 | 0.98 | 0.77 | 0.72 | 0.55 | 0.31 | 0.40 | .93 | 0.92 | 0.97 | 0.98 |
| 6 | 112.4 | 1.66 | 0.13 | 0.13 | 0.99 | 0.75 | 0.71 | 0.52 | 0.28 | 0.38 | 0.92 | 0.93 | 0.97 | 0.98 |
| 7 | 90.4 | 1.46 | 0.13 | 0.13 | 0.99 | 0.77 | 0.73 | 0.48 | 0.28 | 0.35 | 0.93 | 0.93 | 0.97 | 0.99 |
| 8 | 72.2 | 1.42 | 0.12 | 0.13 | 1.00 | 0.77 | 0.73 | 0.51 | 0.29 | 0.36 | 0.93 | 0.94 | 0.97 | 0.98 |
| 9 | 56.9 | 1.57 | 0.12 | 0.13 | 1.00 | 0.78 | 0.73 | 0.49 | 0.29 | 0.36 | 0.93 | 0.94 | 0.97 | 0.98 |
| 10 | 40.4 | 1.51 | 0.12 | 0.13 | 1.00 | 0.77 | 0.72 | 0.46 | 0.28 | 0.35 | 0.93 | 0.94 | 0.98 | 0.98 |
| 11 | 26.5 | 1.55 | 0.11 | 0.13 | 1.00 | 0.78 | 0.73 | 0.49 | 0.28 | 0.37 | 0.93 | 0.94 | 0.98 | 0.98 |
| 12 | 12.0 | 1.47 | 0.11 | 0.12 | 1.00 | 0.78 | 0.73 | 0.47 | 0.27 | 0.35 | 0.93 | 0.95 | 0.98 | 0.98 |
| 13 | -1.5 | 1.51 | 0.11 | 0.13 | 1.00 | 0.78 | 0.72 | 0.46 | 0.27 | 0.34 | 0.93 | 0.95 | 0.98 | 0.98 |
| 14 | -14.6 | 1.47 | 0.11 | 0.12 | 1.00 | 0.78 | 0.73 | 0.45 | 0.26 | 0.33 | 0.93 | 0.95 | 0.98 | 0.99 |
| 15 | -30.0 | 1.45 | 0.1 | 0.13 | 1.00 | 0.77 | 0.71 | 0.46 | 0.25 | 0.34 | 0.93 | 0.95 | 0.98 | 0.98 |
| 16 | -42.6 | 1.45 | 0.1 | 0.13 | 1.00 | 0.78 | 0.72 | 0.47 | 0.27 | 0.33 | 0.93 | 0.95 | 0.98 | 0.99 |
| 17 | -51.0 | 1.55 | 0.11 | 0.13 | 1.00 | 0.78 | 0.72 | 0.49 | 0.26 | 0.36 | 0.93 | 0.95 | 0.97 | 0.98 |
| 18 | -66.2 | 1.46 | 0.1 | 0.13 | 1.00 | 0.78 | 0.73 | 0.46 | 0.25 | 0.35 | 0.93 | 0.94 | 0.98 | 0.98 |
| 19 | -75.9 | 1.41 | 0.1 | 0.13 | 1.00 | 0.77 | 0.73 | 0.45 | 0.26 | 0.34 | 0.93 | 0.95 | 0.98 | 0.98 |
| 20 | -88.0 | 1.34 | 0.09 | 0.12 | 1.00 | 0.79 | 0.75 | 0.44 | 0.26 | 0.32 | 0.93 | 0.95 | 0.98 | 0.98 |

[a] Average pair-wise $C^\alpha$ root-mean square deviation.
[b] Correlation of $\mathbf{B}_k$ and $\mathbf{B}_{k-1}$

Table 2: Reproduction of experimental data and other ensemble properties as a function Expectation Maximization step. Model assuming no experimental noise. Pearsons correlation coefficient is denoted by $\rho$.

| EM-step | ENLP | $\langle$RMSD$\rangle$ (Å)[a] | mean Q-factor | $^{h3}J_{NC'}$ (rms, $s^{-1}$) | $^{3}J_{H^N-H^\alpha}$ (rms, $s^{-1}$) | $^{3}J_{H^N-C^\beta}$ (rms, $s^{-1}$) | $^{3}J_{H^N-C'}$ (rms, $s^{-1}$) | eNOE$_{-3}$ ($\rho$) | eNOE$_{-6}$ ($\rho$) | $R_{H\alpha C\alpha_i/H\alpha C\alpha_{i+1}}(\rho)$ | $R_{HN_i/HN_{i+1}}(\rho)$ | $R_{HN_i/H\alpha C\alpha_i}(\rho)$ | $R_{HN_i/H\alpha C\alpha_{i-1}}(\rho)$ | $\rho(\mathbf{B}_k, \mathbf{B}_{k-1})$[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 109.6 | 2.58 | 0.39 | 0.15 | 0.98 | 0.56 | 0.67 | 0.67 | 0.64 | 0.95 | 0.59 | 0.90 | 0.95 | - |
| 1 | -28.4 | 1.65 | 0.18 | 0.14 | 0.64 | 0.35 | 0.47 | 0.73 | 0.68 | 0.94 | 0.91 | 0.96 | 0.98 | 0.0 |
| 2 | 88.2 | 2.36 | 0.32 | 0.15 | 0.92 | 0.47 | 0.62 | 0.69 | 0.66 | 0.94 | 0.67 | 0.91 | 0.95 | 0.34 |
| 3 | -10.9 | 1.65 | 0.17 | 0.13 | 0.59 | 0.34 | 0.45 | 0.75 | 0.70 | 0.94 | 0.91 | 0.96 | 0.97 | 0.63 |
| 4 | -36.6 | 1.78 | 0.15 | 0.13 | 0.58 | 0.31 | 0.41 | 0.75 | 0.72 | 0.94 | 0.92 | 0.97 | 0.98 | 0.97 |
| 5 | -61.7 | 1.56 | 0.14 | 0.13 | 0.56 | 0.30 | 0.41 | 0.77 | 0.72 | 0.93 | 0.93 | 0.97 | 0.98 | 0.98 |
| 6 | -82.5 | 1.59 | 0.13 | 0.13 | 0.56 | 0.30 | 0.41 | 0.78 | 0.72 | 0.93 | 0.92 | 0.97 | 0.98 | 0.99 |
| 7 | -101.3 | 1.58 | 0.13 | 0.13 | 0.52 | 0.28 | 0.38 | 0.73 | 0.71 | 0.93 | 0.93 | 0.97 | 0.98 | 0.99 |
| 8 | -118.1 | 1.45 | 0.13 | 0.13 | 0.54 | 0.28 | 0.38 | 0.76 | 0.72 | 0.93 | 0.93 | 0.97 | 0.98 | 0.99 |
| 9 | -136.5 | 1.54 | 0.12 | 0.13 | 0.52 | 0.29 | 0.37 | 0.77 | 0.73 | 0.93 | 0.95 | 0.97 | 0.99 | 0.99 |
| 10 | -151.9 | 1.56 | 0.12 | 0.13 | 0.50 | 0.29 | 0.36 | 0.77 | 0.73 | 0.93 | 0.94 | 0.97 | 0.98 | 1.00 |
| 11 | -168.5 | 1.67 | 0.11 | 0.12 | 0.51 | 0.27 | 0.37 | 0.77 | 0.73 | 0.93 | 0.94 | 0.97 | 0.98 | 1.00 |
| 12 | -184.0 | 1.50 | 0.11 | 0.13 | 0.48 | 0.27 | 0.34 | 0.78 | 0.74 | 0.93 | 0.94 | 0.98 | 0.98 | 1.00 |
| 13 | -197.7 | 1.41 | 0.11 | 0.12 | 0.46 | 0.27 | 0.34 | 0.77 | 0.73 | 0.93 | 0.95 | 0.98 | 0.98 | 1.00 |
| 14 | -212.2 | 1.36 | 0.11 | 0.13 | 0.48 | 0.27 | 0.36 | 0.77 | 0.73 | 0.93 | 0.95 | 0.97 | 0.98 | 1.00 |
| 15 | -224.2 | 1.33 | 0.11 | 0.13 | 0.48 | 0.27 | 0.36 | 0.78 | 0.73 | 0.93 | 0.94 | 0.98 | 0.98 | 1.00 |
| 16 | -235.9 | 1.43 | 0.1 | 0.13 | 0.47 | 0.25 | 0.34 | 0.78 | 0.73 | 0.93 | 0.94 | 0.98 | 0.99 | 1.00 |
| 17 | -246.4 | 1.46 | 0.1 | 0.13 | 0.50 | 0.26 | 0.33 | 0.76 | 0.72 | 0.93 | 0.95 | 0.97 | 0.98 | 1.00 |
| 18 | -259.3 | 1.44 | 0.1 | 0.13 | 0.46 | 0.26 | 0.33 | 0.78 | 0.73 | 0.93 | 0.95 | 0.98 | 0.98 | 1.00 |
| 19 | -271.2 | 1.41 | 0.1 | 0.12 | 0.46 | 0.25 | 0.34 | 0.76 | 0.72 | 0.93 | 0.95 | 0.98 | 0.99 | 1.00 |
| 20 | -279.9 | 1.37 | 0.1 | 0.13 | 0.45 | 0.27 | 0.35 | 0.79 | 0.74 | 0.93 | 0.95 | 0.98 | 0.98 | 1.00 |

[a] Average pair-wise $C^\alpha$ root-mean square deviation.
[b] Correlation of $\mathbf{B}_k$ and $\mathbf{B}_{k-1}$

Table 3: Reproduction of experimental data and other ensemble properties for the first 100 ns and 1 $\mu$s of previously reported 10 $\mu$s molecular dynamics simulations.[27] Pearsons correlation coefficient is denoted by $\rho$.

| Simulation | $\langle$RMSD$\rangle$ (Å)[a] | mean Q-factor | $^{h3}J_{NC'}$ (rms, $s^{-1}$) | $^{3}J_{H^N-H^\alpha}$ (rms, $s^{-1}$) | $^{3}J_{H^N-C^\beta}$ (rms, $s^{-1}$) | $^{3}J_{H^N-C'}$ (rms, $s^{-1}$) | eNOE$_{-3}$ ($\rho$) | eNOE$_{-6}$ ($\rho$) | $R_{H\alpha C\alpha_i/H\alpha C\alpha_{i+1}}(\rho)$ | $R_{HN_i/HN_{i+1}}(\rho)$ | $R_{HN_i/H\alpha C\alpha_i}(\rho)$ | $R_{HN_i/H\alpha C\alpha_{i-1}}(\rho)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 ns CHARMM22* | 0.99 | 0.29 | 0.15 | 0.91 | 0.41 | 0.55 | 0.74 | 0.71 | 0.97 | 0.78 | 0.93 | 0.92 |
| 1 $\mu s$ CHARMM22* | 0.91 | 0.28 | 0.15 | 0.86 | 0.40 | 0.53 | 0.73 | 0.69 | 0.97 | 0.79 | 0.93 | 0.92 |
| 100 ns ff99SB*-ILDN | 1.06 | 0.20 | 0.11 | 0.56 | 0.28 | 0.4 | 0.75 | 0.71 | 0.98 | 0.91 | 0.96 | 0.96 |
| 1 $\mu s$ ff99SB*-ILDN | 1.06 | 0.20 | 0.12 | 0.59 | 0.26 | 0.39 | 0.75 | 0.69 | 0.98 | 0.90 | 0.96 | 0.96 |

[a] Average pair-wise $C^\alpha$ root-mean square deviation.

Table 4: Experimental values of the cross-correlated relaxation rates (R=$R_{H\alpha_i C\alpha_i - H\alpha_{i+1} C\alpha_{i+1}}$+$R_{H\alpha_i C\alpha_{i+1} - H\alpha_{i+1} C\alpha_i}$ – referred to as $R_{H\alpha C\alpha_i - H\alpha C\alpha_{i+1}}$ above) measured.

| i (Residue index) | R ($s^{-1}$) | i (Residue index) | R ($s^{-1}$) |
|---|---|---|---|
| 3 | 22.9 | 32 | -18.3 |
| 4 | 15.5 | 33 | -16.9 |
| 5 | 22.3 | 34 | -4.4 |
| 6 | 23.4 | 35 | -8.3 |
| 7 | 24.8 | 36 | -15.9 |
| 8 | 22.2 | 37 | -5.8 |
| 9 | 1.5 | 38 | 1.3 |
| 10 | -0.9 | 39 | 2.6 |
| 11 | 4.2 | 40 | 23.1 |
| 12 | -2.6 | 41 | 1.2 |
| 13 | 23.6 | 42 | 1.5 |
| 14 | 1.4 | 43 | 12.0 |
| 15 | 5.2 | 44 | 24.5 |
| 16 | 24.9 | 45 | 20.7 |
| 17 | 11.3 | 46 | 23.7 |
| 18 | 25.6 | 47 | 6.9 |
| 19 | 6.6 | 48 | -9.8 |
| 22 | 9.6 | 49 | 5.0 |
| 23 | -6.6 | 50 | -2.3 |
| 24 | -8.5 | 51 | -6.4 |
| 25 | -16.3 | 52 | 18.9 |
| 26 | -10.0 | 53 | 24.8 |
| 28 | -20.3 | 54 | 24.6 |
| 29 | -9.9 | 55 | 22.2 |
| 30 | -3.4 | 56 | 14.2 |
| 31 | -17.4 | | |

Table 5: Experimental values of the cross-correlated relaxation rates (R=$R_{H_iN_i-H_{i+1}N_{i+1}}$+$R_{H_iN_{i+1}-H_{i+1}N_i}$ – referred to as $R_{HN_i-HN_{i+1}}$ above) measured.

| i (Residue index) | R ($s^{-1}$) | i (Residue index) | R ($s^{-1}$) |
|---|---|---|---|
| 3 | 3.24 | 29 | 4.71 |
| 4 | 3.53 | 30 | 4.49 |
| 5 | 3.97 | 31 | 4.62 |
| 6 | 3.91 | 32 | 4.88 |
| 7 | 3.36 | 33 | 5.07 |
| 8 | 0.86 | 34 | 4.96 |
| 9 | -1.15 | 36 | 2.90 |
| 10 | 3.10 | 37 | -1.66 |
| 11 | 0.37 | 38 | 2.38 |
| 12 | 2.22 | 39 | 2.76 |
| 13 | 3.83 | 40 | 4.02 |
| 14 | 3.74 | 41 | 2.21 |
| 15 | 3.95 | 42 | 2.29 |
| 16 | 3.35 | 43 | 2.11 |
| 17 | 3.97 | 44 | 4.11 |
| 18 | 4.04 | 45 | 4.29 |
| 19 | 3.62 | 46 | 3.55 |
| 20 | 4.71 | 47 | 2.55 |
| 21 | 1.74 | 48 | 3.52 |
| 22 | 4.69 | 49 | -1.43 |
| 23 | 3.49 | 50 | 3.77 |
| 24 | 4.24 | 51 | 3.86 |
| 25 | 4.23 | 52 | 3.21 |
| 26 | 4.92 | 53 | 4.56 |
| 27 | 4.79 | 54 | 4.41 |
| 28 | 4.37 | 55 | 4.12 |

Table 6: Experimental values of the cross-correlated relaxation rates (R=$R_{H_iN_i-H\alpha_iC\alpha_i}$+$R_{H\alpha_iN_i-H_iC\alpha_i}$ − referred to as $R_{HN_i-H\alpha C\alpha_i}$ above) measured.

| i (Residue index) | R ($s^{-1}$) | i (Residue index) | R ($s^{-1}$) |
|---|---|---|---|
| 3 | -11.19 | 31 | -2.94 |
| 4 | -12.19 | 32 | -3.94 |
| 5 | -12.27 | 33 | -0.83 |
| 6 | -11.69 | 34 | -2.23 |
| 7 | -11.36 | 35 | -2.34 |
| 8 | -11.72 | 36 | -1.43 |
| 10 | -3.70 | 37 | -10.91 |
| 11 | -10.14 | 39 | -11.582 |
| 12 | -8.99 | 40 | -9.55 |
| 13 | -10.87 | 42 | -10.14 |
| 15 | -9.40 | 43 | -9.97 |
| 16 | -6.60 | 44 | -10.68 |
| 17 | -10.22 | 45 | -11.37 |
| 18 | -6.82 | 46 | -12.19 |
| 19 | -11.17 | 47 | -2.85 |
| 20 | -6.42 | 48 | -3.34 |
| 21 | -1.28 | 49 | -11.63 |
| 22 | -5.45 | 50 | -3.90 |
| 23 | -1.57 | 51 | -12.44 |
| 24 | -2.52 | 52 | -12.02 |
| 25 | -5.26 | 53 | -11.74 |
| 26 | -2.59 | 54 | -11.91 |
| 27 | -0.04 | 55 | -12.36 |
| 28 | -2.55 | 56 | -8.97 |
| 29 | -3.47 | | |
| 30 | -3.83 | | |

Table 7: Experimental values of the cross-correlated relaxation rates (R=$R_{H_iN_i-H\alpha_{i-1}C\alpha_{i-1}}$+$R_{H\alpha_{i-1}N_i-H_iC\alpha_{i-1}}$ – referred to as $R_{HN_i-H\alpha C\alpha_{i-1}}$ above) measured.

| i (Residue index) | R ($s^{-1}$) | i (Residue index) | R ($s^{-1}$) |
|---|---|---|---|
| 3 | -9.82 | 29 | -0.64 |
| 4 | -7.83 | 30 | -0.54 |
| 5 | -9.81 | 31 | -1.30 |
| 6 | -11.14 | 32 | -1.16 |
| 7 | -10.85 | 33 | -1.40 |
| 8 | -10.98 | 34 | -0.82 |
| 9 | -6.12 | 35 | -1.06 |
| 11 | -0.58 | 36 | -1.55 |
| 12 | 1.28 | 37 | 1.06 |
| 13 | -10.22 | 38 | 4.83 |
| 14 | -10.00 | 40 | -10.16 |
| 16 | -8.57 | 41 | -7.97 |
| 17 | -1.92 | 43 | -9.27 |
| 18 | -6.34 | 44 | -11.00 |
| 19 | -5.27 | 45 | -7.63 |
| 20 | -10.15 | 46 | -10.70 |
| 21 | -6.97 | 47 | -10.33 |
| 22 | 2.00 | 48 | 1.12 |
| 23 | -4.47 | 49 | 0.18 |
| 24 | 0.01 | 50 | 5.18 |
| 25 | -0.88 | 51 | 1.05 |
| 26 | -0.95 | 52 | -9.75 |
| 27 | -0.66 | 53 | -10.29 |
| 28 | -0.69 | 54 | -10.00 |

## References


(1) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179–9191.

(2) Pelupessy, P.; Ravindranathan, S.; Bodenhausen, G. *J. Biomol. NMR* **2003**, *25*, 265–280.

(3) Chiarparin, E.; Pelupessy, P.; Ghose, R.; Bodenhausen, G. *J. Am. Chem. Soc.* **2000**, *122*, 1758–1761.

(4) Vögeli, B.; Yao, L. *J. Am. Chem. Soc.* **2009**, *131*, 3668–3678.

(5) Yang, D.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 9880–9887.

(6) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.

(7) Vögeli, B. *J. Magn. Reson.* **2013**, *226*, 52–63.

(8) Irbäck, A.; Mitternacht, S.; Mohanty, S. *PMC Biophys.* **2009**, *2*, 2.

(9) Bottaro, S.; Boomsma, W.; E. Johansson, K.; Andreetta, C.; Hamelryck, T.; Ferkinghoff-Borg, J. *J. Chem. Theory Comput.* **2012**, *8*, 695–702.

(10) Montalvao, R. W.; De Simone, A.; Vendruscolo, M. *J. Biomol. NMR* **2012**, *53*, 281–292.

(11) Fu, B.; Sahakyan, A. B.; Camilloni, C.; Tartaglia, G. G.; Paci, E.; Caflisch, A.; Vendruscolo, M.; Cavalli, A. *J. Comput. Chem.* **2014**, *35*, 1101–1105.

(12) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158–4159.

(13) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J Comput Chem* **2003**, *24*, 1999–2012.

(14) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(15) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*.

(16) Bashford, D.; Case, D. A. *Annu Rev Phys Chem* **2000**, *51*, 129–152.

(17) Hall, J. B.; Fushman, D. *J. Biomol. NMR* **2003**, *27*, 261–275.

(18) Vögeli, B.; Riek, R. *J. Biomol. NMR* **2010**, *46*, 135–147.

(19) Vögeli, B. *J. Chem. Phys.* **2010**, *133*, 014501.

(20) Barfield, M. *J. Am. Chem. Soc.* **2002**, *124*, 4158–4168.

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

(22) Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *J. Biomol. NMR* **2005**, *32*, 273–280.

(23) Losonczi, J. A.; Andrec, M.; Fischer, M. W.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334 – 342.

(24) Bax, A. *Protein Sci.* **2003**, *12*, 1–16.

(25) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035–6043.

(26) Theobald, D. L.; Wuttke, D. S. *Bioinformatics* **2006**, *22*, 2171–2172.

(27) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PLoS ONE* **2012**, *7*, e32131.