Use of multiple picosecond high-mass molecular dynamics simulations to predict crystallographic B-factors of folded globular proteins

Yuan-Ping Pang

Computer-Aided Molecular Design Laboratory, Mayo Clinic, Rochester, MN 55905, USA

Corresponding author: Stabile 12-26, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA; E-mail address: pang@mayo.edu; Telephone: 1-507-284-7868

ABSTRACT

It is challenging to predict crystallographic B-factors of a protein from a conventional molecular dynamics (MD) simulation. This is partly because the B-factors calculated through sampling the atomic positional fluctuations in a picosecond MD simulation are unreliable, and longer samplings often yield substantially large root mean square deviations (RMSDs) between calculated and experimental B-factors. This article reports the use of uniformly increased atomic masses by 100-fold to increase the time resolution of an MD simulation so that sampling the atomic positional fluctuations in multiple picosecond MD simulations with such high masses can improve the B-factor prediction. Using the third immunoglobulin-binding domain of protein G, bovine pancreatic trypsin inhibitor, ubiquitin, and lysozyme as model systems, the Cα and Cγ B-factor RMSDs of these proteins were ranging from 3.1±0.2 Å^2 to 9.2±0.8 Å^2 or from 3.6±0.1 Å^2 to 9.6±0.2 Å^2, respectively, when the sampling was done, for each of these proteins, in 20 distinct, independent, and 50-picosecond high-mass MD simulations using AMBER forcefield FF12MC or FF14SB. These results suggest that sampling the atomic positional fluctuations in multiple picosecond high-mass MD simulations may be conducive to *a priori* prediction of crystallographic B-factors of a folded protein.

Keywords: B-factor; Thermal motion; GB3; BPTI; Ubiquitin; Lysozyme.

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; GB3, the third immunoglobulin-binding domain of protein G; MD, molecular dynamics; NPT, isobaric-isothermal; PDB, Protein Data Bank; SE, standard error.

1. Introduction

As a measure of the uncertainty of the atomic mean position, the crystallographic B-factor of a given atom reflects the displacement of the atom from its mean position in a crystal structure and this displacement attenuates X-ray scattering and is caused by both thermal motion of the atom and static disorder of the atom in a crystal lattice [1-6]. Despite the challenges of separating the thermal motion in time from the static disorder in space [7], B-factors can be used to quantitatively identify *less mobile* regions of a crystal structure as long as the structure is determined without substantial crystal lattice defects, rigid-body motions, and refinement errors [8,9]. A low B-factor indicates a small degree of thermal motion, while a high B-factor may imply a large degree of thermal motion. Such B-factor or mobility information is useful to structure-based design of protein modulators. As more comparative models of folded globular proteins are used in the protein modulator design, methods to predict the B-factors of a folded globular protein from molecular dynamics (MD) simulations become more desirable.

However, due to the use of different protein environments, different timescales to detect thermal motions, and different methods to determine the B-factors, predicting B-factors of a folded protein by sampling of the atomic positional fluctuations of the protein in a conventional MD simulation with solvation might not be feasible [10]. For example, a reported MD simulation study showed that the B-factors derived on the picosecond timescale were unreliable and that the simulated B-factors on the nanosecond timescale were considerably larger than the experimental values [10]. Although simulations of proteins in their crystalline state [11,12] can avoid the difference in protein environment, such simulations are inapplicable to *a priori* prediction of B-factors of comparative models of proteins.

This article reports a study using numerous sets of 20 distinct, independent, and isobaric-isothermal (NPT) MD simulations with atomic masses that were uniformly increased or decreased to investigate simulation conditions that may offer B-factor prediction useful for

structure-based design of protein modulators. The third immunoglobulin-binding domain of protein G (GB3) [13], bovine pancreatic trypsin inhibitor (BPTI) [14], ubiquitin [15], and lysozyme [16] were used in this study as model systems of folded globular proteins. AMBER forcefields FF12MC [17] and FF14SB [18] were used to examine the simulation conditions in a forcefield independent manner.

2. Theory and Methods

2.1. Theory of using uniformly scaled atomic masses to compress or expand MD simulation time

Reducing atomic masses uniformly by tenfold (hereafter referred to as low masses) can enhance configurational sampling in NPT MD simulations [19]. The effectiveness of the low-mass NPT MD simulation technique can be explained as follows: To determine the relative configurational sampling efficiencies of two simulation systems—one with standard masses and another with low masses, the units of distance $[l]$ and energy $[m]([l]/[t])^2$ of the low-mass simulations are purposefully kept identical to those of the standard-mass simulations. This is so that the structure and energy of the low-mass simulation system can be compared to those of the standard-mass simulation system. Let superscripts ^{lmt} and ^{smt} denote the times for the low-mass and standard-mass systems, respectively. Then $[m^{\mathrm{lmt}}] = 0.1\ [m^{\mathrm{smt}}]$, $[l^{\mathrm{lmt}}] = [l^{\mathrm{smt}}]$, and $[m^{\mathrm{lmt}}]([l^{\mathrm{lmt}}]/[t^{\mathrm{lmt}}])^2 = [m^{\mathrm{smt}}]([l^{\mathrm{smt}}]/[t^{\mathrm{smt}}])^2$ lead to $\sqrt{10}\ [t^{\mathrm{lmt}}] = [t^{\mathrm{smt}}]$. A conventional MD simulation program takes the timestep size (Δt) of the standard-mass time rather than that of the low-mass time. Therefore, low-mass NPT MD simulations at $\Delta t = 1.00\ \mathrm{fs}^{\mathrm{smt}}$ (*viz.*, $\sqrt{10}\ \mathrm{fs}^{\mathrm{lmt}}$) are theoretically equivalent to standard-mass NPT MD simulations at $\Delta t = \sqrt{10}\ \mathrm{fs}^{\mathrm{smt}}$, as long as both standard-mass and low-mass simulations are carried out for the same number of timesteps and there are no precision issues in performing these simulations. This equivalence of mass *downscaling* and timestep-size *upscaling* explains why uniform mass reduction can compress

MD simulation time and why low-mass NPT MD simulations at $\Delta t = 1.00\ fs^{smt}$ offer better configurational sampling efficacy than conventional standard-mass NPT MD simulations at $\Delta t = 1.00\ fs^{smt}$ or $\Delta t = 2.00\ fs^{smt}$. It also explains why the kinetics of the low-mass simulation system can be converted to the kinetics of the standard-mass simulation system simply by scaling the low-mass time with a factor of $\sqrt{10}$ [17]. In this context, to efficiently sample alternative conformations from a crystallographically determined conformation, low-mass NPT MD simulations at $\Delta t = 1.00\ fs^{smt}$ and temperature of <340 K were used for GB3, BPTI, ubiquitin, and lysozyme in this study, although standard-mass simulations at $\Delta t = 3.16\ fs^{smt}$ can achieve the same sampling efficiency.

In the same vein, let superscript ^{hmt} denote the time for the system with uniformly increased atomic masses by 100-fold (hereafter referred to as high masses), then $[m^{hmt}] = 100\ [m^{smt}]$, $[l^{hmt}] = [l^{smt}]$, and $[m^{hmt}]([l^{hmt}]/[t^{hmt}])^2 = [m^{smt}]([l^{smt}]/[t^{smt}])^2$ lead to $[t^{hmt}] = 10\ [t^{smt}]$. This equivalence of mass *upscaling* and timestep-size *downscaling* explains why uniform mass increase can expand MD simulation time and why high-mass NPT MD simulations at $\Delta t = 1.00\ fs^{smt}$ can increase their time resolution by tenfold. Therefore, to adequately sample the atomic positional fluctuations in a short simulation, high-mass NPT MD simulations at $\Delta t = 1.00\ fs^{smt}$ were used for GB3, BPTI, ubiquitin, and lysozyme in the present study.

2.2. *MD simulations of proteins*

A folded globular protein was solvated with the TIP3P water [20] with or without surrounding counter ions and then energy-minimized for 100 cycles of steepest-descent minimization followed by 900 cycles of conjugate-gradient minimization to remove close van der Waals contacts using SANDER of AMBER 11 (University of California, San Francisco). The resulting system was heated—in 20 distinct, independent, unrestricted, unbiased, and classical

MD simulations with a periodic boundary condition and unique seed numbers for initial velocities—from 0 to 295 or 297 K at a rate of 10 K/ps under constant temperature and constant volume, then equilibrated with a periodic boundary condition for 10^6 timesteps under constant temperature and constant pressure of 1 atm employing isotropic molecule-based scaling, and lastly simulated under the NPT condition at 1 atm and a constant temperature of <340 K using PMEMD of AMBER 11.

The initial conformations of GB3, BPTI, ubiquitin, and lysozyme for the simulations were taken from the crystal structures of Protein Data Bank (PDB) IDs of 1IGD, 5PTI, 1UBQ, and 4LZT, respectively. A truncated 1IGD structure (residues 6–61) was used for the GB3 simulations. Four interior water molecules (WAT111, WAT112, WAT113, and WAT122) were included in the initial 5PTI conformation. The simulations for GB3, BPTI, and ubiquitin were done at 297 K as the exact data-collection temperatures of these proteins had not been reported. The lysozyme simulations were done at the reported data-collection temperature of 295 K [16].

The numbers of TIP3P waters and surrounding ions, initial solvation box size, ionizable residues, and computers used for the NPT MD simulations are provided in Table S1. The 20 unique seed numbers for initial velocities of Simulations 1–20 were taken from Ref. [21]. All simulations used (*i*) a dielectric constant of 1.0, (*ii*) the Berendsen coupling algorithm [22], (*iii*) the Particle Mesh Ewald method to calculate electrostatic interactions of two atoms at a separation of >8 Å [23], (*iv*) $\Delta t = 1.00\ fs^{smt}$, (*v*) the SHAKE-bond-length constraints applied to all bonds involving hydrogen, (*vi*) a protocol to save the image closest to the middle of the "primary box" to the restart and trajectory files, (*vii*) a formatted restart file, (*viii*) the revised alkali and halide ions parameters [24], (*ix*) a cutoff of 8.0 Å for nonbonded interactions, (*x*) atomic masses that were uniformly increased by 100-fold or decreased by tenfold relative to the standard atomic masses, and (*xi*) default values of all other inputs of the PMEMD module. The forcefield parameters of FF12MC are available in the Supporting Information of Ref. [17].

2.3. *Crystallographic B-factor prediction*

Using a two-step procedure with PTRAJ of AmberTools 1.5, the B-factors of Cα and Cγ atoms in a folded globular protein were predicted from all conformations saved at every 10^3 timesteps of 20 simulations of the protein using the simulation conditions described above. The first step was to align all saved conformations onto the first saved one to obtain an average conformation using root mean square fit of all CA atoms (for Cα B-factors) or all CG and CG2 atoms (for Cγ B-factors). The second step was to root mean square fit all CA atoms (or all CG and CG2 atoms) in all saved conformations onto the corresponding atoms of the average conformation and then calculate the Cα (or Cγ) B-factors using the "atomicfluct" command in PTRAJ. For each protein, the calculated B-factors in Table S2 and Fig. 1 are the average of all B-factors derived from 20 simulations of the protein. The standard error (SE) of a B-factor was calculated according to Eq. 2 of Ref. [25]. The SE of an RMSD between computed and experimental B-factors was calculated using the same method for the SE of a B-factor. The experimental B-factors of GB3, BPTI, ubiquitin, and lysozyme were taken from the crystal structures of PDB IDs of 1IGD, 4PTI, 1UBQ, and 4LZT, respectively.

2.4. *Correlation analysis*

The correlation analysis was performed using PRISM 5 for Mac OS X of GraphPad Software (La Jolla, California) with the assumption that data were sampled from Gaussian populations.

3. Results and discussion

3.1. *Using high–time-resolution picosecond simulations to retrospectively predict B-factors*

The internal motions—such as the motions of backbone N–H bonds of a folded globular protein at the solution state—are on the order of tens or hundreds of ps^{smt} [26]. Therefore, the timescale of the thermal motions in the B-factors of a protein at the crystalline state is unlikely greater than a nanosecond. As explained in Section 1, the B-factor of a given atom reflects both the thermal motion of the atom and the static disorder of the atom in a crystal lattice [1-6]. In this context, 20 distinct, independent, and picosecond high-mass NPT MD simulations of a folded globular protein were carried out to investigate whether combining the sampling of the atomic positional fluctuations of the protein on a picosecond timescale with the sampling of such fluctuations over conformations derived from the 20 distinct and independent NPT MD simulations could approximate the experimental B-factors of the protein. The use of high-mass NPT MD simulations was to increase the time resolution of the simulations so that the B-factor prediction could be done in a statistically relevant manner. These high-mass simulations were performed with FF12MChm and FF14SBhm, which denote the AMBER forcefields FF12MC [17] and FF14SB [18] with uniformly increased atomic masses by 100-fold relative to the standard atomic masses.

As listed in Table 1, regardless of which forcefield was used, the RMSDs between computed and experimental B-factors were <10 $Å^2$ for all four proteins when the atomic positional fluctuations of these proteins were sampled over 25 ps^{smt} or 50 ps^{smt}. When FF12MChm was used, longer samplings led to B-factor RMSDs of >10 $Å^2$ for all four proteins. When FF14SBhm was used, the RMSDs were also >10 $Å^2$ for GB3, ubiquitin, and BPTI. The FF14SBhm-derived RMSDs of lysozyme were ≤9.7±0.4 $Å^2$ and >10 $Å^2$ when the atomic positional fluctuations were sampled over 1 ns^{smt} (Table 1) and 20 ns^{smt} (Table S3), respectively. FF12MChm best reproduced most of the experimental B-factors on the timescale of 50 ps^{smt} with RMSDs ranging from 3.1±0.2 to 9±1 $Å^2$ for Cα and from 7.3±0.9 to 9.6±0.2 $Å^2$ for Cγ. FF14SBhm also best reproduced most of the experimental B-factors on the timescale of 50 ps^{smt} with RMSDs ranging from

3.6±0.1 to 8.2±0.6 Å^2 for Cα and from 8.4±0.3 to 9.6±0.2 Å^2 for Cγ. Regardless of which forcefield was used, the means and SEs of the B-factor RMSDs of ubiquitin were larger than those of the other proteins (Table 1). This suggested that the conformational variations resulting from 20 distinct, independent, and picosecond NPT MD simulations might be insufficient to mimic the static disorders of the ubiquitin crystals. However, extending the number of the simulations of ubiquitin from 20 to 40 or 80 reduced the SEs but not the mean (Table S4).

For all four proteins, a good agreement of the experimental values with the calculated Cα and Cγ B-factors on the timescale of 50 ps^{smt} is shown in Fig. 1, and the SEs of the predicted B-factors are listed in Table S2. The Pearson correlation coefficients are 0.6, 0.7, 0.8, and 0.9 for the predicted Cα B-factors of GB3, ubiquitin, BPTI, and lysozyme using FF12MChm, respectively. The respective coefficients are 0.6, 0.7, 0.7, and 0.9 for FF14SBhm. The Pearson correlation coefficients of the predicted Cγ B-factors using FF12MChm or FF14SBhm are 0.4–0.6 or 0.5–0.6 for the four proteins, respectively (Fig. 1). These results suggest that combining the sampling of the atomic positional fluctuations of a folded protein over the ~50-ps^{smt} timescale with the sampling of such fluctuations over conformations derived from 20 distinct ~50-ps^{smt} NPT MD simulations can approximate the experimental B-factors with RMSDs of <10 Å^2 and the Pearson correlation coefficients of 0.6–0.9 for Cα and 0.4–0.6 for Cγ.

3.2. *Using alternative conformers as mimics of static disorders to improve B-factor prediction*

In all the B-factor calculations described above, the conformational variations—as mimics of the static disorders of a protein in its crystal lattice—stemmed from 20 distinct, independent, and picosecond NPT MD simulations. Each of these simulations used a unique seed number for initial velocities and a common initial conformation that was taken from a respective crystal structure. These simulations were performed sequentially for 30 ps^{smt} to set the system

temperature at a desired value, for 100 ps^{smt} to equilibrate the system at the desired temperature, and for a period of time (such as 50 ps^{smt}) to sample the atomic positional fluctuations of the protein. It is not unreasonable to suspect that the conformational heterogeneity resulting from the heating and equilibration over a combined period of 130 ps^{smt} might be insufficient to present the static disorders in a crystal lattice of the protein.

Therefore, 20 distinct, independent, and 948-ns^{smt} low-mass NPT MD simulations using FF12MC were carried out for each of the four proteins to effectively sample conformations varied from the crystallographically determined conformation. Each of the 20 low-mass simulations of a protein used a unique seed number for initial velocities and the crystallographically determined conformation as the initial conformation. Three instantaneous conformations were saved at 316-ns^{smt} intervals for each of the 20 low-mass simulations, resulting in three sets of 20 instantaneous conformations saved at 316 ns^{smt}, 632 ns^{smt}, and 948 ns^{smt}. The 20 distinct, independent, and 50-ps^{smt} high-mass NPT MD simulations using FF12MChm described in Section 3.1 were then repeated three times as follows: Each of the 20 high-mass simulations used a unique seed number for initial velocities and an initial conformation that was taken from one of the 20 instantaneous conformations in each of the three sets.

As listed in Table 2, the differences among the RMSDs derived from the initial conformations at 316 ns^{smt}, 632 ns^{smt}, and 948 ns^{smt} are marginal. Consistent with the observation described in Section 3.1, most of the RMSDs that were sampled over the 50-ps^{smt} timescale are smaller than those sampled over the shorter or longer timescale. For BPTI and lysozyme, the RMSDs of the multiple conformations derived from the low-mass simulations were larger than those of the single conformation taken from the respective crystal structure, but the difference is $\leq$1.3 $Å^2$. For GB3 and ubiquitin, the RMSDs of the multiple conformations were smaller than those of the single conformation, and the difference is $\leq$2.9 $Å^2$. These results suggest that use of alternative conformations sampled in 20, distinct, independent, and submicrosecond low-mass

NPT MD simulation can slightly improve the prediction of B-factors of proteins that are devoid of disulfide bonds but slightly impair the prediction for proteins with their conformations restrained by disulfide bonds.

3.3. *Twenty ~50-ps^{smt} simulations might be conducive to prediction of B-factors*

Using atomic masses that are purposefully scaled up by 100-fold to expand the MD simulation time, the present study demonstrates that the atomic positional fluctuations of a folded globular protein sampled over a period of ~50 ps^{smt} in 20 distinct and independent high-mass NPT MD simulations approximate the experimental B-factors better than the fluctuations sampled over a shorter or longer timescale. This observation is in agreement with the report that the internal motions such as the motions of backbone N–H bonds of a folded globular protein in the solution state are on the order of tens or hundreds of ps^{smt} [26]. Using the same notion as described above to increase the time resolution of the simulations, the Lipari-Szabo order parameters [27] of backbone N–H bonds—extracted from ^{15}N spin relaxation data of GB3 [28], BPTI [29], ubiquitin [30], and lysozyme [31]—also were found to be best reproduced at the timescale of ~50 ps^{smt} (unpublished result of YPP).

This study compared two B-factor prediction methods. One uses the conformational variations resulting from the heating and equilibration of a respective crystal structure over a combined period of 130 ps^{smt}. The other uses the conformational heterogeneity resulting from multiple instantaneous conformations saved at 316-ns^{smt} intervals of 20 distinct, independent, and 948-ns^{smt} low-mass NPT MD simulations. The result of the comparative study shows that sampling the atomic positional fluctuations over the multiple instantaneous conformations approximates the experimental B-factors of GB3 and ubiquitin better than sampling the fluctuations over the conformational variations from a crystal structure, and *vise versa* for BPTI

and lysozyme. This observation seemed puzzling at first, but, with the explanation below, it correlates well with the structures of the four proteins.

Unlike BPTI and lysozyme, GB3 and ubiquitin do not have any disulfide bonds to restrain their folded conformations. There is no structural difference between the solution and solid states for GB3 or ubiquitin [13,15,32,33]. However, the C14–C38 disulfide bond in BPTI flips between left- and right-handed configurations [34] in the NMR structure (PDB ID: 1PIT). This bond is locked entirely at the right-handed configuration in the crystal structure (PDB ID: 4PTI). For lysozyme, its C64–C80 disulfide bond adopts both configurations in the NMR structure (PDB ID: 1E8L) and the left-handed configuration in the crystal structure (PDB ID: 4LZT). Sampling the conformation of BPTI or lysozyme in solution for >130 ps^{smt} inevitably incorporates the conformations resulting from the flipping of the disulfide bond, but such conformations are absent at the crystalline state. This explains why sampling the atomic positional fluctuations over the multiple instantaneous conformations in solution impairs the B-factors of BPTI and lysozyme, but improves those of GB3 and ubiquitin. This also helps explain why the B-factor RMSDs generally progress in time (Table 1) and underscores the necessity to confine the sampling to the timescale of ~50 ps^{smt}.

Taking all the results together, the present study suggests that sampling the atomic positional fluctuations in 20 distinct, independent, and ~50-ps^{smt} high-mass NPT MD simulations may be conducive to *a priori* prediction of crystallographic B-factors of a folded protein for structure-based protein modulator design. These high-mass simulations may use initial conformations taken from the conformations sampled in 20 distinct and independent nanosecond low-mass NPT MD simulations for *a priori* prediction or use a common initial conformation take from the crystallographically determined conformation for retrospective prediction of B-factors. The retrospective prediction may offer insight into relative contributions of the thermal motions in time and the static disorders in space to the experimental B-factors.

Conflict of interest

The author declares no conflict of interest.

Acknowledgments

Yuan-Ping Pang acknowledges the support of this work from the US Defense Advanced Research Projects Agency (DAAD19-01-1-0322), the US Army Medical Research Material Command (W81XWH-04-2-0001), the US Army Research Office (DAAD19-03-1-0318, W911NF-09-1-0095, and W911NF-16-1-0264), the US Department of Defense High Performance Computing Modernization Office, and the Mayo Foundation for Medical Education and Research. The contents of this article are the sole responsibility of the author and do not necessarily represent the official views of the funders. The author is most grateful to the organizers of the RapiData course at the US National Synchrotron Light Source of the Brookhaven National Laboratory, which offered him hands-on training in macromolecular X-ray diffraction measurement and inspired this work.

References

[1] P. Debye, Interference of x rays and heat movement, Ann. Phys. 43 (1913) 49–95.

[2] I. Waller, On the effect of thermal motion on the interference of X-rays, Z. Phys. 17 (1923) 398–408.

[3] B.T.M. Willis, A.W. Pryor, Thermal vibrations in crystallography, Cambridge University Press, London, 1975.

[4] A. Kidera, N. Go, Normal mode refinement: Crystallographic refinement of protein dynamic structure. 1. Theory and test by simulated diffraction data, J. Mol. Biol. 225 (1992) 457–475.

[5] K.N. Trueblood, H.B. Burgi, H. Burzlaff, J.D. Dunitz, C.M. Gramaccioli, H.H. Schulz, U. Shmueli, S.C. Abrahams, Atomic displacement parameter nomenclature: Report of a subcommittee on atomic displacement parameter nomenclature, Act Crystallogr., Sect. A 52 (1996) 770–781.

[6] A.E. Garcia, J.A. Krumhansl, H. Frauenfelder, Variations on a theme by Debye and Waller: From simple crystals to proteins, Proteins 29 (1997) 153–160.

[7] L. Meinhold, J.C. Smith, Fluctuations and correlations in crystalline protein dynamics: A simulation analysis of Staphylococcal nuclease, Biophys. J. 88 (2005) 2554–2563.

[8] J. Kuriyan, W.I. Weis, Rigid protein motion as a model for crystallographic temperature factors, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 2773–2777.

[9] J. Drenth, Principles of protein X-ray crystallography, 3rd ed., Springer, 2007.

[10] P.H. Hünenberger, A.E. Mark, W.F. van Gunsteren, Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations, J. Mol. Biol. 252 (1995) 492–503.

[11] Z.Q. Hu, J.W. Jiang, Assessment of biomolecular force fields for molecular dynamics simulations in a protein crystal, J. Comput. Chem. 31 (2010) 371–380.

[12] P.A. Janowski, C. Liu, J. Deckman, D.A. Case, Molecular dynamics simulation of triclinic lysozyme in a crystal lattice, Protein Sci. 25 (2016) 87–102.

[13] J.P. Derrick, D.B. Wigley, The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab, J. Mol. Biol. 243 (1994) 906–918.

[14] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber, The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors, Acta Crystallogr., Sect. B: Struct. Sci. 39 (1983) 480–490.

[15] S. Vijaykumar, C.E. Bugg, W.J. Cook, Structure of ubiquitin refined at 1.8 Å resolution, J. Mol. Biol. 194 (1987) 531–544.

[16] M.A. Walsh, T.R. Schneider, L.C. Sieker, Z. Dauter, V.S. Lamzin, K.S. Wilson, Refinement of triclinic hen egg-white lysozyme at atomic resolution, Acta Crystallogr., Sect. D: Biol. Crystallogr. 54 (1998) 522–546.

[17] Y.-P. Pang, Low-mass molecular dynamics simulation for configurational sampling enhancement: More evidence and theoretical explanation, Biochem. Biophys. Rep. 4 (2015) 126–133.

[18] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser, C. Simmerling, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theory Comput. 11 (2015) 3696–3713.

[19] Y.-P. Pang, Low-mass molecular dynamics simulation: A simple and generic technique to enhance configurational sampling, Biochem. Biophys. Res. Commun. 452 (2014) 588–592.

[20] W.L. Jorgensen, J. Chandreskhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[21] Y.-P. Pang, Use of 1–4 interaction scaling factors to control the conformational equilibrium between α-helix and β-strand, Biochem. Biophys. Res. Commun. 457 (2015) 183–186.

[22] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.

[23] T.A. Darden, D.M. York, L.G. Pedersen, Particle mesh Ewald: An N log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[24] I.S. Joung, T.E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, J. Phys. Chem. B 112 (2008) 9020–9041.

[25] Y.-P. Pang, At least 10% shorter C–H bonds in cryogenic protein crystal structures than in current AMBER forcefields, Biochem. Biophys. Res. Commun. 458 (2015) 352–355.

[26] S. Morin, A practical guide to protein dynamics from ^{15}N spin relaxation in solution, Prog. Nucl. Magn. Reson. Spectrosc. 59 (2011) 245–262.

[27] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, J. Am. Chem. Soc. 104 (1982) 4546–4559.

[28] J.B. Hall, D. Fushman, Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G, J. Biomol. NMR 27 (2003) 261–275.

[29] S.A. Beeser, T.G. Oas, D.P. Goldenberg, Determinants of backbone dynamics in native BPTI: Cooperative influence of the 14–38 disulfide and the Tyr35 side-chain, J. Mol. Biol. 284 (1998) 1581–1596.

[30] N. Tjandra, S.E. Feller, R.W. Pastor, A. Bax, Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation, J. Am. Chem. Soc. 117 (1995) 12562–12566.

[31] M. Buck, J. Boyd, C. Redfield, D.A. Mackenzie, D.J. Jeenes, D.B. Archer, C.M. Dobson, Structural determinants of protein dynamics: Analysis of ^{15}N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg-white lysozyme, Biochemistry 34 (1995) 4041–4055.

[32] T.S. Ulmer, B.E. Ramirez, F. Delaglio, A. Bax, Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy, J. Am. Chem. Soc. 125 (2003) 9179–9191.

[33] G. Cornilescu, J.L. Marquardt, M. Ottiger, A. Bax, Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase, J. Am. Chem. Soc. 120 (1998) 6836–6837.

[34] J.S. Richardson, The anatomy and taxonomy of protein structure, Adv. Protein Chem. 34 (1981) 167–339.

Fig. 1. Experimental and calculated B-factors of GB3, BPTI, ubiquitin, and lysozyme. The B-factors were calculated from 20 50-ps^{smt} MD simulations using FF12MChm or FF14SBhm. The letter "r" is the abbreviation for the Pearson correlation coefficient.

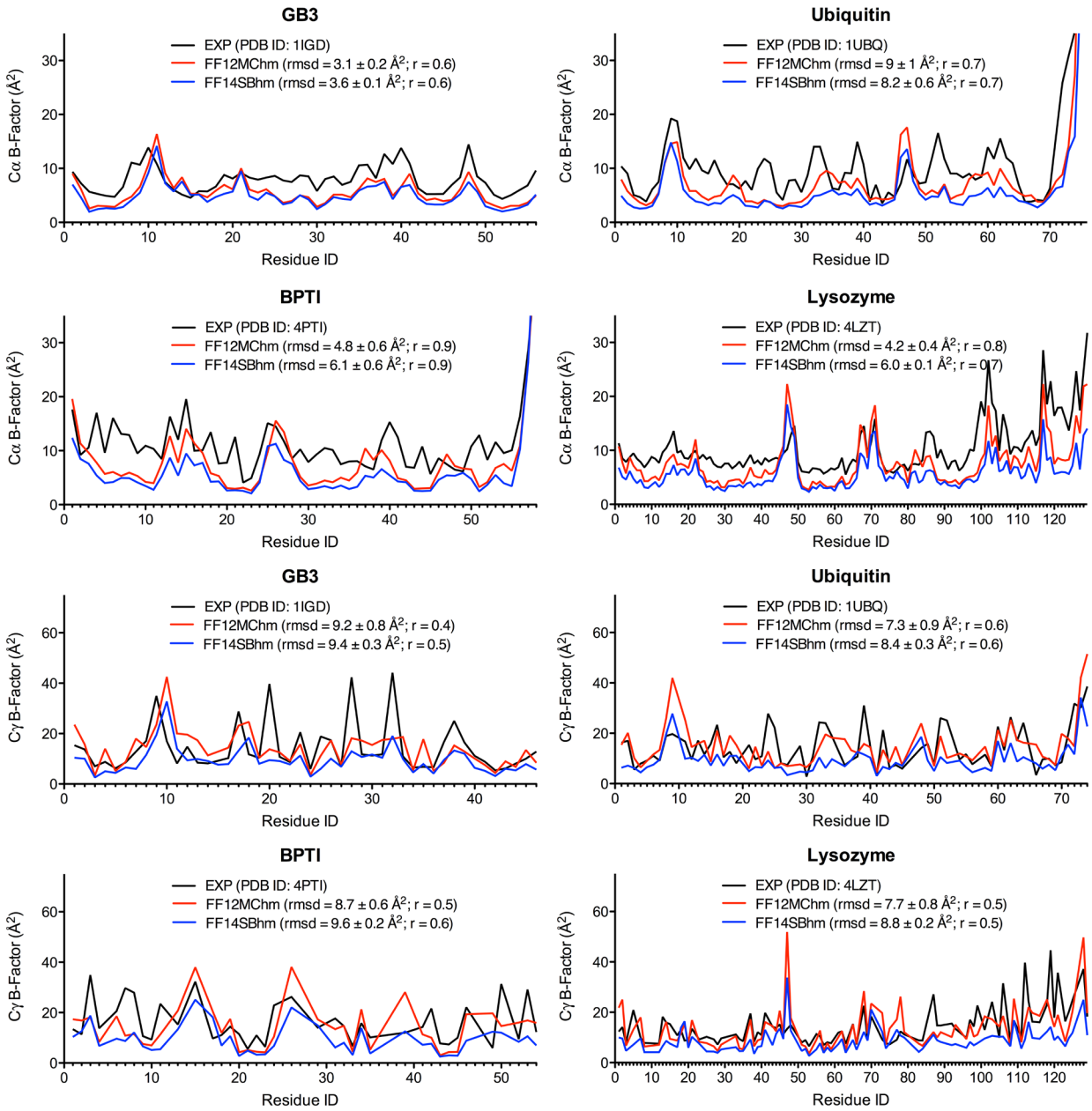

Table 1. Root mean square deviations and standard errors between experimental and calculated B-factors of GB3, BPTI, ubiquitin, and lysozyme.

Time (ps^{smt})	RMSD ± SE (Å^2)			
	FF12MChm	FF14SBhm	FF12MChm	FF14SBhm
	GB3 (297 K)		Lysozyme (295K)	
		Cα		
25	3.7±0.1	4.2±0.1	5.2±0.3	6.7±0.1
50	3.1±0.2	3.6±0.1	4.2±0.4	6.0±0.1
100	3.7±0.7	3.4±0.2	3.5±0.6	5.5±0.1
200	5.3±0.9	3.3±0.2	4.0±0.6	5.1±0.1
300	5.9±0.8	3.2±0.2	5.2±0.6	5.1±0.1
400	8±1	3.3±0.2	6.9±0.8	5.0±0.1
500	9±1	3.6±0.3	8±1	4.9±0.1
600	9±1	4.0±0.5	9±1	4.9±0.1
700	10±1	4.3±0.5	10±1	4.9±0.1
800	10±1	4.6±0.6	11±2	4.8±0.1
900	10±1	4.9±0.7	11±2	4.8±0.1
1000	10±1	5.2±0.7	12±2	4.8±0.1
		Cγ		
25	9.3±0.5	10.3±0.2	7.4±0.5	9.5±0.1
50	9.2±0.8	9.4±0.3	7.7±0.7	8.8±0.2
100	12±2	8.8±0.6	10±1	8.4±0.2
200	17±2	8.4±0.7	13±1	8.3±0.3
300	19±2	8.0±0.6	17±1	8.6±0.4
400	23±2	8.4±0.6	20±1	8.9±0.4
500	25±2	9.5±0.9	22±2	9.0±0.4
600	26±2	11±1	24±2	9.2±0.4
700	27±2	12±1	26±3	9.4±0.4
800	28±2	12±1	27±3	9.5±0.4
900	28±2	13±2	28±3	9.6±0.4
1000	29±2	13±2	29±3	9.7±0.4
	Ubiquitin (297 K)		BPTI (297 K)	
		Cα		
25	6.2±0.3	7.1±0.2	5.9±0.3	6.8±0.3
50	9±1	8.2±0.6	4.8±0.6	6.1±0.6
100	16±2	12±1	5.2±0.8	7.3±0.9
200	32±3	21±2	8±1	10±1
300	37±4	28±3	13±2	14±2
400	40±4	32±3	15±2	16±2
500	43±4	36±3	17±2	18±2
		Cγ		
25	7.0±0.9	9.3±0.2	8.6±0.4	10.7±0.2
50	7.3±0.9	8.4±0.3	8.7±0.6	9.6±0.2
100	12±1	7.8±0.6	11±1	9.1±0.3
200	20±2	9±1	13±1	8.8±0.4
300	25±3	10±1	15±1	8.8±0.5
400	27±3	11±1	17±1	8.9±0.6
500	29±3	12±2	19±1	8.9±0.6

Time: the duration of 20 different and independent molecular dynamics simulations over which the B-factors were calculated. RMSD: root mean square deviation. SE: standard error.

Table 2. Root mean square deviations and standard errors between experimental and calculated B-factors of GB3, BPTI, ubiquitin, and lysozyme.

Time (ps^{smt})	RMSD ± SE (Å^2)			
	IC at 130 ps^{smt}	IC at 316 ns^{smt}	IC at 632 ns^{smt}	IC at 948 ns^{smt}
	GB3 (297 K)			
	Cα			
25	3.7±0.1	3.2±0.2	3.3±0.2	3.3±0.2
50	3.1±0.2	3.0±0.4	2.9±0.2	3.1±0.4
100	3.7±0.7	3.8±0.8	2.9±0.4	3.4±0.4
	Cγ			
25	9.3±0.5	8.8±0.6	8.3±0.5	8.8±0.6
50	9.2±0.8	10±1	8.5±0.6	9±1
100	12±2	13±2	11±1	12±1
	Ubiquitin (297 K)			
	Cα			
25	6.2±0.3	6.9±0.6	6.6±0.4	6.3±0.5
50	9±1	7±1	6.1±0.8	6.4±0.9
100	16±2	9±2	9±1	9±1
	Cγ			
25	7.0±0.9	8.2±0.5	7.9±0.6	8.1±0.6
50	7.3±0.9	8±1	7±1	9±1
100	12±1	9±2	9±2	10±1
	BPTI (297 K)			
	Cα			
25	5.9±0.3	7.1±0.2	6.9±0.2	6.4±0.3
50	4.8±0.6	6.0±0.3	6.0±0.3	5.2±0.5
100	5.2±0.8	4.9±0.5	4.7±0.8	4.6±0.9
	Cγ			
25	8.6±0.4	9.4±0.6	9.0±0.5	8.3±0.6
50	8.7±0.6	9.2±0.9	9.4±0.8	9±1
100	11±1	10±1	11±1	10±1
	Lysozyme (295K)			
	Cα			
25	5.2±0.3	5.8±0.2	5.8±0.3	5.5±0.3
50	4.2±0.4	5.1±0.4	5.2±0.7	4.7±0.9
100	3.5±0.6	4.8±0.7	6±1	6±2
	Cγ			
25	7.4±0.5	7.9±0.7	7.7±0.8	7.9±0.7
50	7.7±0.8	8±1	9±1	10±1
100	10±1	10±1	12±2	14±3

Time: the duration of 20 different and independent molecular dynamics simulations over which the B-factors were calculated. IC: the initial conformation that was taken from the instantaneous conformation saved at 130 ps^{smt}, 316 ns^{smt}, 632 ns^{smt}, or 948 ns^{smt}. RMSD: root mean square deviation. SE: standard error.