Research Article

Relaxation Estimation of RMSD in Molecular Dynamics Immunosimulations

Wolfgang Schreiner,[1] Rudolf Karch,[1] Bernhard Knapp,[1] and Nevena Ilieva[2]

[1] *Section of Biosimulation and Bioinformatics, Center for Medical Statistics, Informatics, and Intelligent Systems (CeMSIIS), Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria*
[2] *Institute for Nuclear Research and Nuclear Energy (INRNE), Bulgarian Academy of Sciences, 72, Tzarigradsko Chaussee, 1784 Sofia, Bulgaria*

Correspondence should be addressed to Rudolf Karch, rudolf.karch@meduniwien.ac.at

Molecular dynamics simulations have to be sufficiently long to draw reliable conclusions. However, no method exists to prove that a simulation has converged. We suggest the method of "lagged RMSD-analysis" as a tool to judge if an MD simulation has not yet run long enough. The analysis is based on RMSD values between pairs of configurations separated by variable time intervals Δt. Unless RMSD(Δt) has reached a stationary shape, the simulation has not yet converged.

1. Introduction

Mathematical modeling and computational biology have proved extremely successful in the efforts to understand the mechanisms of immunologic reactions, for example, in modeling the T-cell proliferation dynamics following hepatitis C, HIV infection [1], or the growth of an immunogenic tumor [2]. Methods of statistical mechanics may successfully be applied in order to quantitatively understand the immune system [3, 4], and different modeling approaches are adequate for acquired immunity, as surveyed in a seminal paper by Perelson and Weisbush [5]. Modeling and simulation can be performed on different levels, starting at the top level with agent-based models [6, 7] for the cooperation of numerous large biomolecules in the formation of the immune synapse [8–11] down to more detailed models of antigen binding [12], recognition, and signaling [13–20]. T-cell proliferation modulated by interleukin 2 has been simulated [21]. Applications aiming to predict the reaction to epitopes [22] improved vaccine design [23], vaccination

against tumors [24], and for improved patient care have been devised [25–27], also in personalized medicine [28]. In particular, the mechanisms how T-cell receptors (TCRs) detect antigen peptides (p) presented by major histocompatibility complex (MHC) molecules can be investigated by molecular dynamics (MD) on a molecular or even atomic level [29–32]. However, TCRpMHC complexes are huge protein complexes, which have to be studied in water, which adds an even larger number of atoms to model and simulate. For these reasons, sufficiently long simulations are mandatory in order to obtain realistic results. This is even more so since molecular recognition phenomena may operate on rather long timescales as compared to the length of usual MD trajectories. Hence, efforts to assess sampling quality are mandatory in such simulations.

Given the trajectory of an actual MD simulation, it is clear from first principles that no formal check whatsoever can prove that complete sampling has been achieved. If parts of the phase space have not been visited yet, there is no possibility that this becomes evident from looking
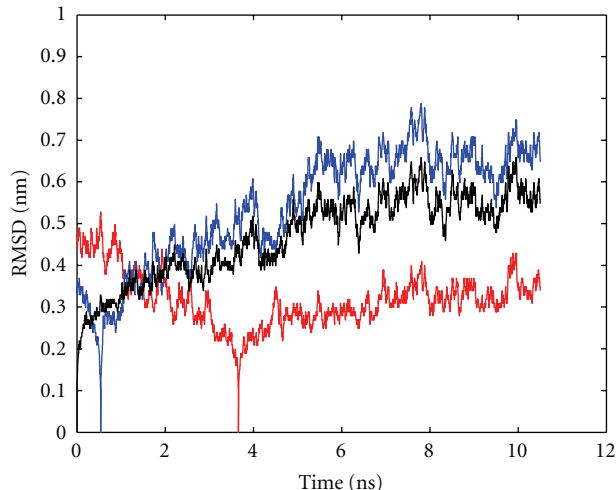
FIGURE 1: RMSD as a function of time for trajectory 1 (10 ns). Black curve: RMSD between first configuration and all successive ones. Blue curve: for a given configuration (t_j), RMSD to each of the other configurations of the same trajectory was computed and averaged, yielding $\overline{\mathrm{RMSD}}(t_j)$. That value of t_j for which $\overline{\mathrm{RMSD}}(t_j)$ is maximum is then adopted as a reference (t_1) to plot RMSD values of the whole trajectory. Red curve: as blue curve, but for minimum $\overline{\mathrm{RMSD}}(t_j)$. Note that the RMSD between the reference configuration and itself is zero by definition.

only at data from those other parts that have been visited. Zhang et al. [33] proposed a formal approach to decompose the phase space into Voronoi polyhedra [34, 35].

At any rate, formal checks of sampling quality can only draw on simulation data already produced and *at the most detect possible shortcomings* of those trajectories. The situation is analogous to tests for the randomness of pseudorandom-number generators [36]. One can never prove randomness as such. It is only possible to detect specific deviations from randomness, such as serial correlations—and any such detection works only with a preselected error rate.

Likewise, MD trajectories may be investigated for specific markers of nonrandomness, the most important type being trends of energy and molecular deformations. This work focuses on the latter, deviations in shape being quantified via the *root mean square deviation* (RMSD) [37] at time t_2 with respect to a given reference structure at time t_1:

$$\mathrm{RMSD}(t_1, t_2) = \left[\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i(t_2) - \mathbf{x}_i(t_1)\|^2\right]^{1/2}, \quad (1)$$

where $\mathbf{x}_i(t)$ is the position of atom i at time t and N is the total number of atoms in the molecule.

Often, the first frame of a trajectory (t_1) is used as a reference, and values of $\mathrm{RMSD}(t_1, t_2)$ are computed for all successive ($t_2 > t_1$) frames; see the black curve in Figure 1. RMSD monitored this way shows large rapid fluctuations on top of long-term variations and jumps. Generally, it is difficult to identify such long-term variations, which may

relate to functional modes. A recent study [38] reflected the insufficiency of visual RMSD inspection alone.

Extensive work has been published on using the RMSD for the characterization of structural changes, drifts, and trends; see [33] and the references cited therein. Grossfield and Zuckerman [39] gives a seminal conceptual discussion of ergodicity, absolute and relative convergence and proposes checks for overall sampling quality. Block averaging was proposed [40] to reduce short-term fluctuations and to obtain more reliable indicators of long-term trends. However, averaging RMSD over all configurations within a block involves many different time intervals $\Delta t = t_j - t_i$, thus reducing the specificity of the resulting average for one particular time interval.

For this reason we consider the RMSD between pairs of configurations separated by a constant time lag Δt as described in Section 2.2.1. In this way we obtain more stable estimates (averages) which are nevertheless perfectly specific for a given time interval Δt.

2. Materials and Methods

2.1. Molecular Dynamics Simulations

2.1.1. Employed Structures. We applied the proposed methodology to a total of 6 MD simulations. For this purpose we used 2 different TCRpMHC complexes: an immunogenic wild-type peptide bound between TCR/MHC and the same TCR/MHC with a less immunogenic mutant peptide. We employed the crystal structure of the LC13 TCR bound to HLA-B*08 : 01 and the Epstein Barr Virus peptide FLRGRAYGL. This structure is available from the Protein Data Bank (PDB) [41] via PDB accession code 1mi5 [42] and is referred to as wild-type. For the mutant complex we substituted the side chain of tyrosine at position 7 of the peptide to alanine (Y7A). This was performed using SCWRL [43] since we could previously show that this tool is most appropriate for mutations in pMHC complexes [44, 45].

The LC13 TCR in complex with the FLRGRAYGL peptide and HLA-B*08 : 01 is an ideal test set for molecular dynamics simulations since this complex was crystallized and described in its parts and as a whole. Initially Kjer-Nielsen et al. crystallized the TCR [46] and the MHC [47] separately while they published a structure of the whole TCRpMHC system [42] afterwards. These available data give substantial insight into the LC13/EBV/HLA-B*08 : 01 system and led to the choice of our wild-type and mutated system for this study. The whole system is illustrated in Figure 2.

2.1.2. Molecular Dynamics Simulation Protocol. The wild-type and the mutant complex were simulated in independent runs for 10, 50, and 200 ns yielding a total of 6 simulations (see Table 1). The following protocol for the simulations was employed. First, we minimized the energy of the systems using a steepest descent method. Then we immersed the complexes in explicit SPC [48] artificial water baths allowing for a minimal distance of 2 nm between the box boundary and the protein. Next, we warmed the complex up to

Table 1: Molecular dynamics simulation runs.

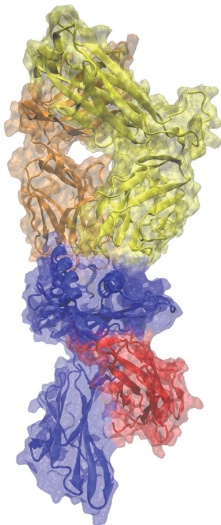|	$t_{\max}$ (ns)	$\Delta t_{\mathrm{config}}$ (ps)	Peptide	n
Trajectory 1	10	3	FLRGRAYGL	3500
Trajectory 2	50	50	FLRGRAYGL	1000
Trajectory 3	200	50	FLRGRAYGL	4322
Trajectory 4	10	3	FLRGRAAGL	3500
Trajectory 5	50	50	FLRGRAAGL	1000
Trajectory 6	200	50	FLRGRAAGL	4322

Figure 2: Illustration of the LC13 TCR in complex with HLA-B*08 : 01. Blue: MHC, red: β2-microglobulin, gray: peptide, orange: TCR alpha-chain, yellow: TCR beta-chain.

310 K using position restraints. Finally, we carried out the simulations using GROMACS 4 [49] and the GROMOS96 force field [50]. All further parameters were set in accordance with [51].

2.1.3. RMSD Calculation. After the simulations were finished, we calculated the RMSD values for each configuration in a given trajectory with respect to every other configuration of the same trajectory using the standard *g_rms* function of GROMACS. This yields an $n \times n$ matrix of RMSD values where n is the number of configurations in the trajectory.

2.2. RMSD between Configurations of Trajectories

2.2.1. Averaging and Modeling of Lagged RMSD. Given one configuration $\mathbf{x}(t_i)$ of an MD simulation as a reference, the RMSD to some other configuration $\mathbf{x}(t_j)$ may be considered a "distance measure" along the time interval $\Delta t = t_j - t_i$. If Δt is short enough, it may be shifted along the whole trajectory and RMSD values be sampled. The average $\overline{\mathrm{RMSD}}(\Delta t)$ is characteristic for the difference between configurations separated by Δt in the particular simulation run considered.

Small values of Δt characterize configurations close to each other in time. Increasing Δt means to compare configurations more distant to each other in time, which

are—intuitively speaking—"less related" to each other. This fact should be reflected in $\overline{\mathrm{RMSD}}(\Delta t)$ for increasing values of Δt.

As Δt increases, dependences should diminish and approach the level of "unrelated" or "independent" configurations. In order to quantify such a saturation trend, we applied the *Hill equation* [52]:

$$\overline{\mathrm{RMSD}}(\Delta t) = \frac{a \cdot \Delta t^{\gamma}}{t^{\gamma} + \Delta t^{\gamma}}, \qquad (2)$$

where the parameter a reflects the maximum value to which the function is asymptotic ("plateau value" $\overline{\mathrm{RMSD}}(\Delta t \to \infty)$), τ is the time lag Δt for which $\overline{\mathrm{RMSD}}(\Delta t) = a/2$ (i.e., the value of Δt for which half-saturation is achieved), and the *Hill coefficient* γ is a parameter that determines the shape, that is, the level of sigmoidicity, of the model functions. The parameters were estimated by fitting the *Hill equation* (2) to the measured values of $\overline{\mathrm{RMSD}}(\Delta t)$ using the *nlinfit* function as implemented in the *Statistics Toolbox* of MATLAB (Mathworks, Natick, MA, USA). The maximum time lag Δt was chosen half the total simulation time of the respective trajectory.

2.2.2. Assessing the Influence of Initial Conditions. The initial phase $t \le t_{\mathrm{offset}}$ of each MD simulation strongly depends on the starting configuration, usually a crystal structure, which can by no means be representative for a configuration obtained from a trajectory. This is even true after energy minimization and warming up. Hence, if the initial phase of an MD trajectory is included in $\overline{\mathrm{RMSD}}(\Delta t)$, a bias will result. This is not only true for individual values of $\overline{\mathrm{RMSD}}(\Delta t)$ but also for the parameters estimated from the fit. In particular, the limiting plateau value $a = \overline{\mathrm{RMSD}}(\Delta t = \infty)$ will also be biased and depend on t_{offset}: in fact, $a = a(t_{\mathrm{offset}})$. We modeled this dependence as

$$a(t_{\mathrm{offset}}) = a_0 + \beta \cdot \exp(-\lambda \cdot t_{\mathrm{offset}}), \qquad (3)$$

where a_0 is the limiting value, $a_0 = a(t_{\mathrm{offset}} = \infty)$, and β and λ are scaling parameters. Note that a_0 is an extrapolated estimate for $\overline{\mathrm{RMSD}}(\Delta t = \infty)$ and $t_{\mathrm{offset}} = \infty$, that is, the estimate for an RMSD between two totally unrelated configurations of a trajectory, independent of initial phase effects.

Values for t_{offset} were selected in the interval $0 \le t_{\mathrm{offset}} \le t_{\max}/2$, where $t_{\max}$ denotes the total simulation time of the respective trajectory.

3. Results and Discussion

3.1. Convergence of RMSD with Increasing Time Lag. Each panel of Figure 3 shows a representative plot (for reasons of conciseness we display results in the figures only for trajectory 1–3) of mean RMSD values (red circles, y-axis) obtained between configurations of one MD trajectory (see figure caption and Table 1), separated by respective time lags Δt (x-axis). As the time-lag between configurations increases, mean RMSD approaches a plateau.
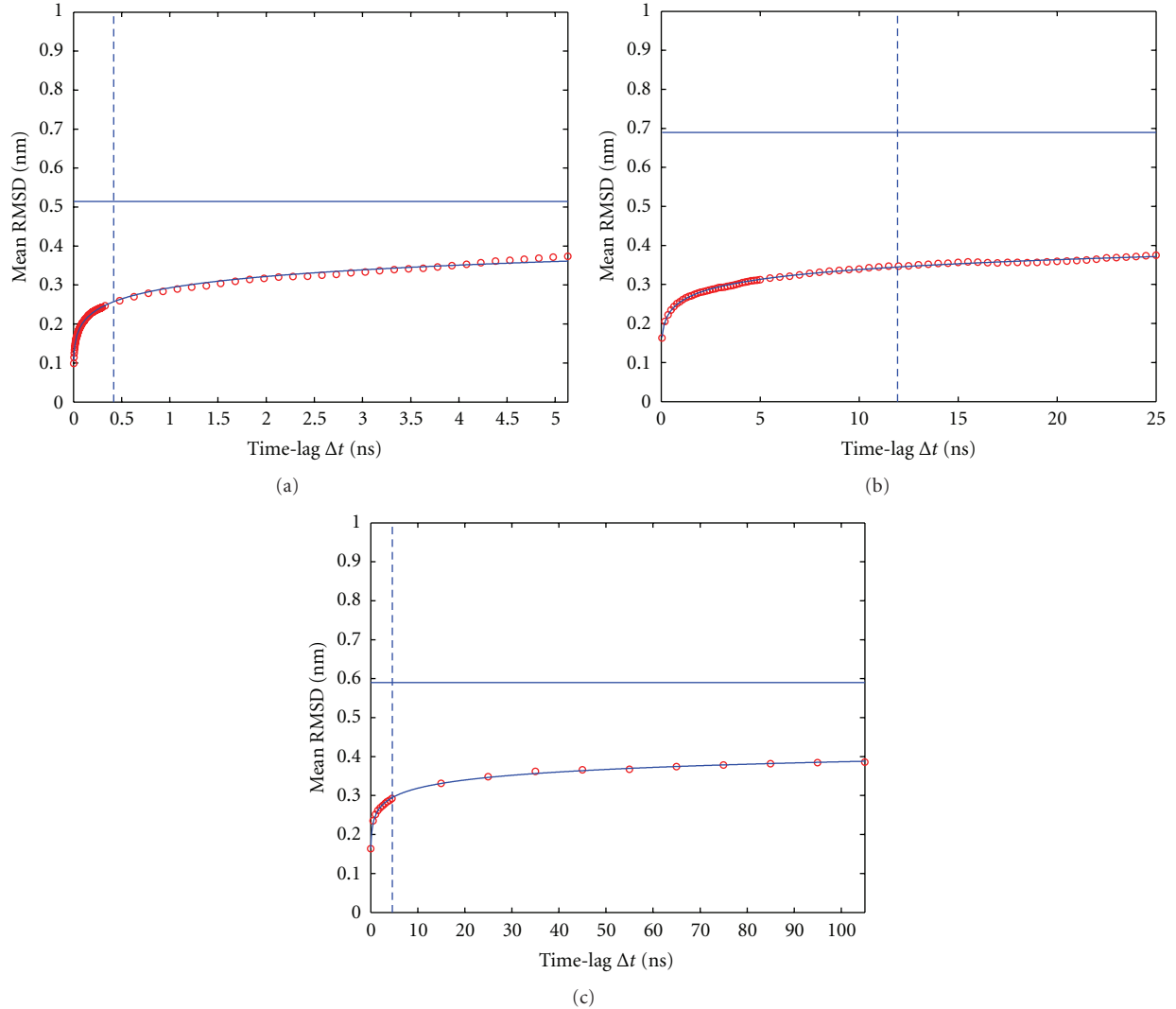
(a)

(b)

(c)

FIGURE 3: Dependence of mean RMSD on the time lag Δt for $t_{\text{offset}} = 0$. (a) 10 ns trajectory 1, (b) 50 ns trajectory 2, (c) 200 ns trajectory 3; see Table 1. Note that due to the different total durations of the simulations, the maximum time lag considered (equal to half the simulation length) varies. Vertical lines indicate the time interval τ, for which half-saturation is achieved; see also (2). The horizontal line indicates the estimated plateau of the mean RMSD, corresponding to the parameter a in (2).

Fits of the model (2) to the values of $\overline{\text{RMSD}}(\Delta t)$ are displayed as solid lines. Parameters obtained from the fits can readily be interpreted as follows. The estimate for parameter a in (2) represents the limiting value of $\overline{\text{RMSD}}(\Delta t)$ and is indicated by the horizontal line. The estimate of the parameter τ corresponds to a "characteristic" ("half-saturation") time interval Δt and is shown as a vertical line in the plots.

Note that the initial phase of each trajectory strongly determines the shape of $\overline{\text{RMSD}}(\Delta t)$ which should be properly represented by the fitted model. In contrast, the remainder of each trajectory is characteristic for the long-term trend of $\overline{\text{RMSD}}(\Delta t)$. Although it shows much smaller changes in RMSD, the remainder contains naturally by far more data points as compared to the initial phase. To achieve an appropriate balance, we increased the lag length in small steps during an initial phase (cf. the initially dense succession

of red circles in Figure 3) and in larger steps (more loose succession of circles) later on. This procedure puts increased weight on the data points for the initial phase.

3.2. Influence of the Offset.
Applying different temporal offsets t_{offset} before starting to analyze the respective trajectories changes the dependence of $\overline{\text{RMSD}}(\Delta t)$ as a function of the time lag Δt; see the typical results displayed in Figure 4.

The larger the offset t_{offset} from the start configuration, the smaller the time lag Δt necessary for the system to level off to its $\overline{\text{RMSD}}$ plateau. The exemplary display of the dependence of fitted parameters on t_{offset}, as shown previously, is systematically analyzed as follows; see Figure 5. With increasing t_{offset} both the plateau value a and the half-saturation parameter t decrease, whereas the shape
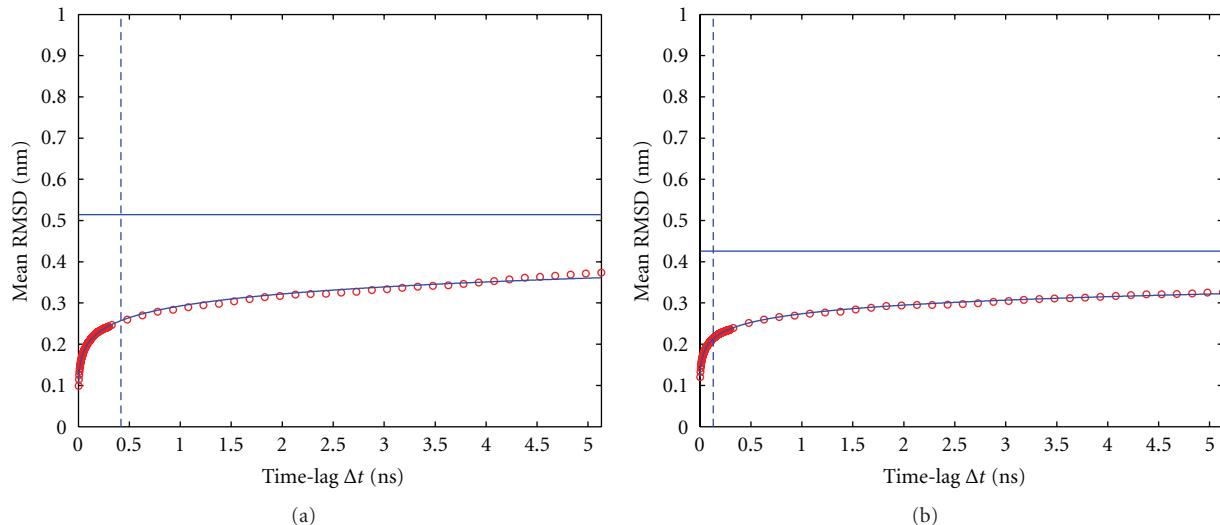
FIGURE 4: Dependence of mean RMSD on t_{offset} for trajectory 1 (10 ns). (a): $t_{offset} = 0$ ps, (b): $t_{offset} = 200$ ps. Vertical lines indicate the time lag τ, for which half-saturation is achieved; see also (2). The horizontal line indicates the estimated plateau of the mean RMSD, corresponding to the parameter a in (2).

parameter γ is fairly constant and almost independent of t_{offset}.

In a next step the systematic dependence of the extrapolated plateau on t_{offset} was fitted via (3); see Figure 6. The monoexponential decay model of (3) represents the most parsimonious choice, given the shape of the simulation results as displayed in Figure 6.

Figures 6(a)–6(c) illustrates the influence of the offset t_{offset} from the start configuration on the respective $\overline{RMSD}$ plateau values as estimated from the parameter a of the model (2). $\overline{RMSD}$ plateau values tend to decrease with increasing t_{offset}.

4. Conclusions

For molecular dynamics simulations of proteins, questions of "convergence" and sampling are important issues if sensible conclusions are to be drawn from such simulations. Since convergence of a particular simulation (in the sense that statistical sampling of the phase space is complete enough with respect to the specific phenomena studied) cannot be judged in advance [39], various techniques have been advised to demonstrate that a simulation has not yet converged [37, 53–56].

Here, we propose a method to assess if a simulation has *not* yet run long enough, based on RMSD analysis of successive configurations of a given MD trajectory, separated by time lags Δt of varying length (up to half the total simulation time). As long as a simulation has not yet converged, the shape of the function $\overline{RMSD}(\Delta t)$ still considerably depends on t_{offset}, that is, the time point along the trajectory, where the analysis interval Δt starts. As t_{offset} is large enough, the shape of $\overline{RMSD}(\Delta t)$ becomes stationary; that is, in Figure 4 the shape of the mean RMSD curve is different in the left and right panel, but further increasing

t_{offset} would not further change its overall appearance. This is also reflected in the respective model parameters, converging to constant values for large values of t_{offset}; see Figure 5.

To describe this Δt dependence of the mean RMSD values, $\overline{RMSD}(\Delta t)$, for a given t_{offset}, the Hill function was used. This function type is frequently applied in enzyme kinetics to model saturation phenomena together with the number of reactive sites on an enzyme. In the present work, the Hill function proved flexible enough to model the functional form of $\overline{RMSD}(\Delta t)$ and to identify the respective parameters (plateau value, shape parameter, and half-saturation time). Contrary to a simple exponential saturation function, the Hill function is able to model different levels of sigmoidicity.

In order to quantify the influence of the initial conditions and the equilibration phase on $\overline{RMSD}(\Delta t)$, we systematically increased the time t_{offset} before starting to analyze each trajectory. With the exception of the shape parameter γ all the other parameters of the Hill function describing $\overline{RMSD}(\Delta t)$ exhibit a distinctive dependence on the initial phase of the trajectories and thus indicate that biased estimates would result if the initial phase would be included in the analysis.

As pointed out in the introduction, the method proposed here combines the smoothing effect of averaging, while retaining the specificity of a precise time interval between configurations being compared. For example, the height of the extrapolated $\overline{RMSD}$ plateau a_0 (see Figure 6) may be interpreted as "configurational distance" between two arbitrarily selected, totally unrelated configurations of the particular part of phase space currently visited. If one chooses an arbitrary configuration out of it as the reference, all others visited by the trajectory will have an average distance a_0. In contrast, Figure 1 shows two limiting cases of reference configurations: (i) the configuration with maximum average distance (i.e., RMSD) to all others (blue curve),
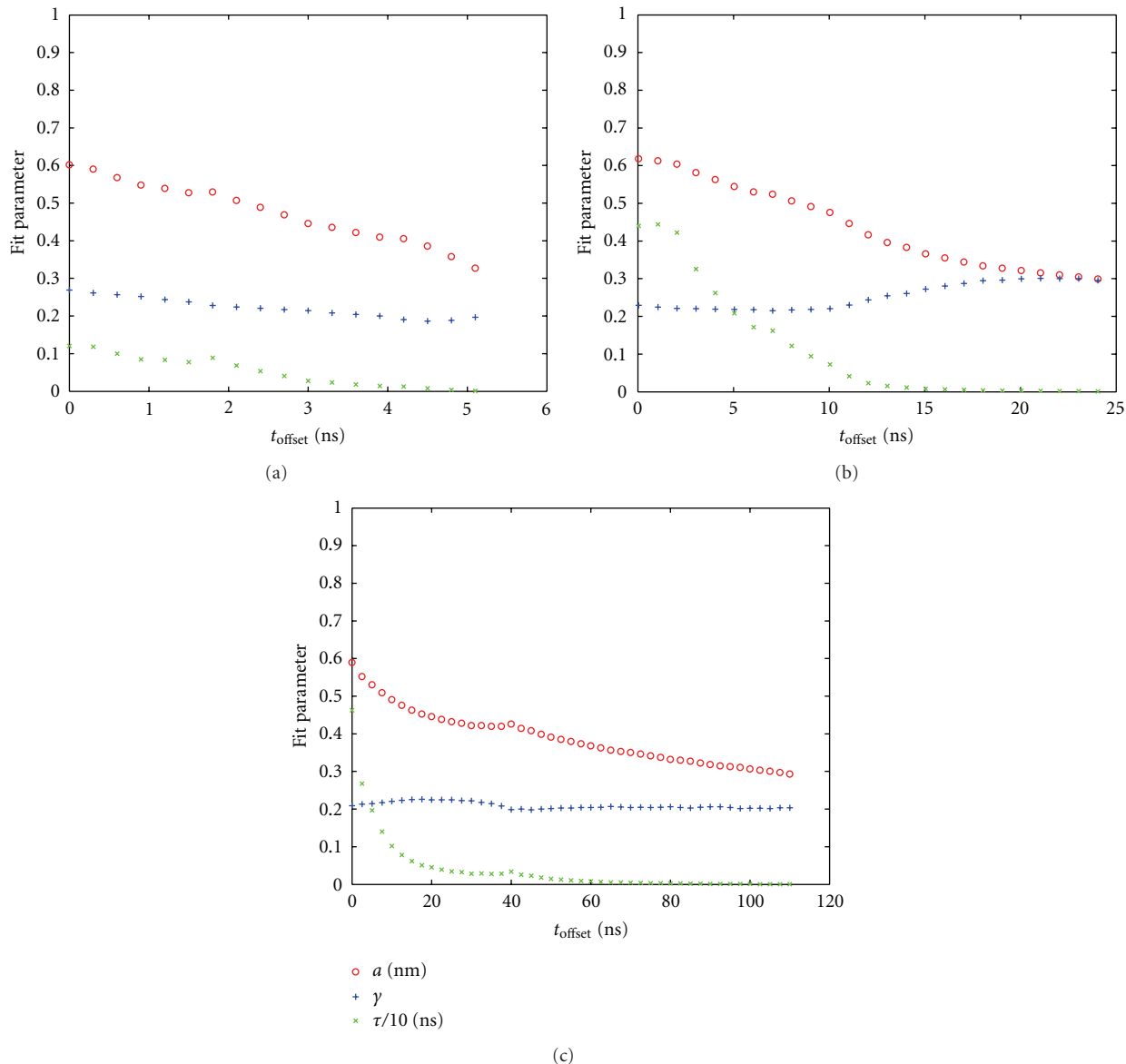
(a)

(b)

   ○ a (nm)
   + γ
   × $\tau/10$ (ns)

(c)

FIGURE 5: Increasing t_{offset} changes fitted parameters of $\overline{RMSD}(\Delta t)$, (2). (a) 10 ns trajectory, (b) 50 ns trajectory, (c) 200 ns trajectory; see Table 1.

which represents the "maximum outlier" ever seen in this trajectory and (ii) the configuration with minimum average distance (i.e., RMSD) to all others (red curve), which is "most central" within the trajectory.

Note that extrapolated plateau values are significantly larger than the $\overline{RMSD}(\Delta t)$ actually reached in the trajectories (see Figure 3). Only if we consider the extrapolated plateau as a function of t_{offset}, we obtain realistic (i.e., lower) estimates (see Figure 6) to aim at during actual simulations. In this sense, we may attribute the proposed fitting procedure some forecast capability regarding the level of RMSD to be finally expected if the trajectory were carried on. This extrapolated mean RMSD corresponds to pairs of configurations separated by time intervals large enough to consider such configurations approximately uncorrelated,

independent representatives of the configuration space of the respective molecule.

Likewise, the "half-saturation time" τ obtained from the fit decreases with increasing t_{offset}. Thus, taking configurations with large enough t_{offset} as a reference, it takes only a short time to get close to the $\overline{RMSD}$ plateau a_0. In each panel of Figure 5, the parameter τ approaches an almost constant level at about half the maximum t_{offset} considered (i.e., 25% of the total simulation time). This might suggest that—regardless of the total simulation length—the fraction of usable (independent) configurations remains the same (final 75% of the trajectory). As the length of the simulation run increases, the criterion derived from the run itself gets increasingly more stringent. From the 200 ns run the first 50 ns should be discarded, which is more than 5 times the
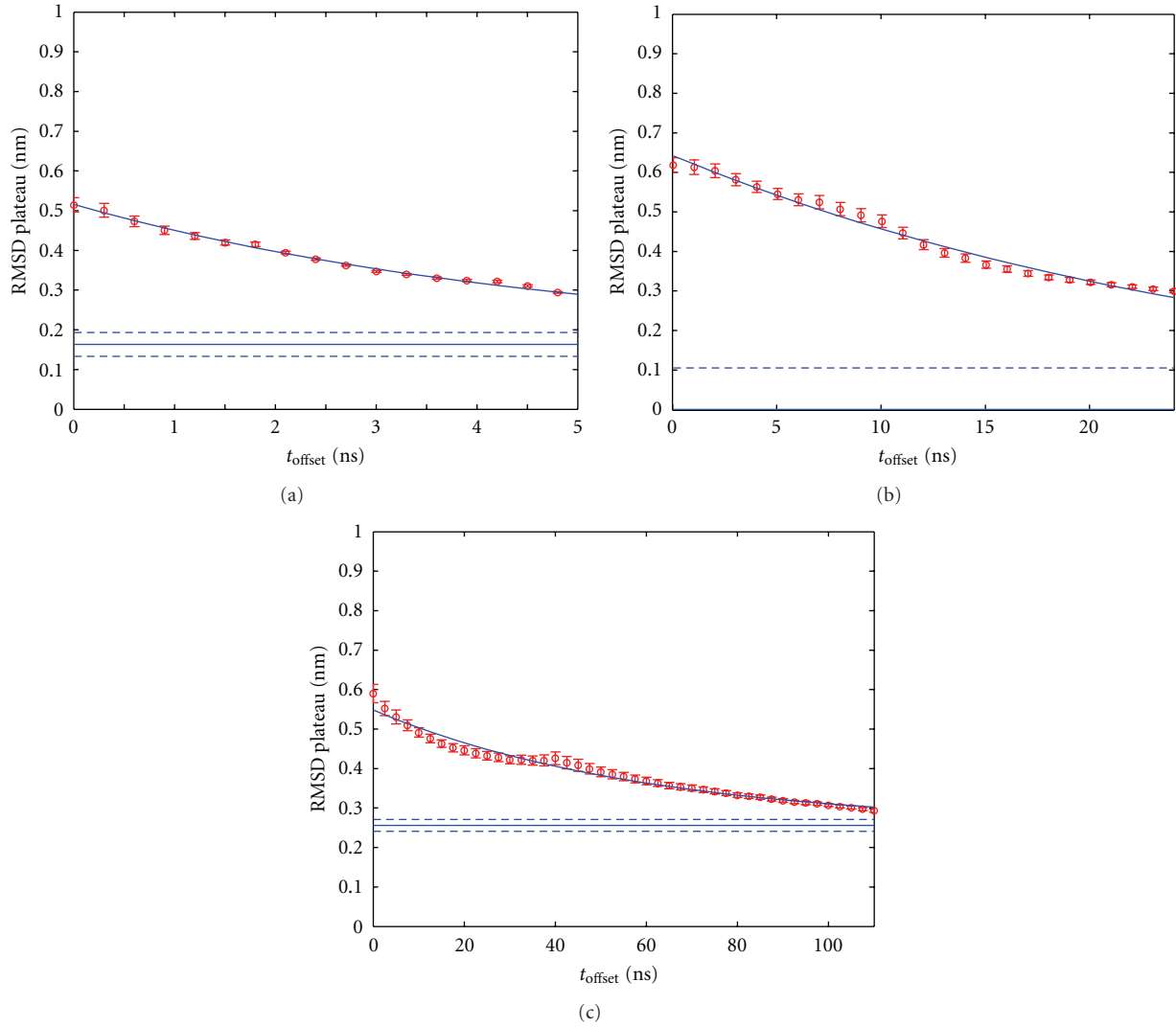
(a)

(b)

(c)

FIGURE 6: Fit and extrapolation of $\overline{\text{RMSD}}$ plateau values for increasing t_{offset}. (a) 10 ns trajectory 1, (b) 50 ns trajectory 2, (c) 200 ns trajectory 3; see Table 1. Red circles represent plateau values $a(t_{\text{offset}})$ the error bars denote their asymptotic standard errors. The solid blue curve shows a nonlinear least-squares fit of (3) to the plateau values $a(t_{\text{offset}})$. From the latter fit the limiting plateau value $a_0 = a(t_{\text{offset}} = \infty)$ was extracted and is shown as a solid horizontal line together with its asymptotic standard error as obtained from the fit of (3).

total length of the 10 ns trajectory. If using the latter as a basis of estimate, however, the last 7.5 ns seem to be trustable.

Although the analysis reported in the present work is specific to TCRpMHC complexes, we expect the method of lagged RMSD analysis to be applicable to similar molecular systems, such as membrane proteins comparable in size and structure. The approach presented here is designed to assess the degree of convergence of MD simulations and hence the statistical quality of conclusions drawn from such simulations.

Acknowledgments

References

[1] A. S. Perelson, "Modelling viral and immune system dynamics," *Nature Reviews Immunology*, vol. 2, no. 1, pp. 28–36, 2002.

[2] V. A. Kuznetsov, I. A. Makalkin, M. A. Taylor, and A. S. Perelson, "Nonlinear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis," *Bulletin of Mathematical Biology*, vol. 56, no. 2, pp. 295–321, 1994.

[3] A. Barra and E. Agliari, "A statistical mechanics approach to autopoietic immune networks," *Journal of Statistical Mechanics*, vol. 2010, no. 7, Article ID P07004, 2010.

[4] A. Barra and E. Agliari, "Stochastic dynamics for idiotypic immune networks," *Physica A*, vol. 389, no. 24, pp. 5903–5911, 2010.

[5] A. S. Perelson and G. Weisbush, "Immunology for physicists," *Reviews of Modern Physics*, vol. 69, no. 4, pp. 1219–1267, 1997.

[6] J. A. Adam, "Effects of vascularization on lymphocyte/tumor cell dynamics: qualitative features," *Mathematical and Computer Modelling*, vol. 23, no. 6, pp. 1–10, 1996.

[7] Y. Vodovotz, G. Clermont, C. Chow, and G. An, "Mathematical models of the acute inflammatory response," *Current Opinion in Critical Care*, vol. 10, no. 5, pp. 383–390, 2004.

[8] N. J. Burroughs and C. Wülfing, "Differential segregation in a cell-cell contact interface: the dynamics of the immunological synapse," *Biophysical Journal*, vol. 83, no. 4, pp. 1784–1796, 2002.

[9] M. L. Dustin, "Stop and go traffic to tune T cell responses," *Immunity*, vol. 21, no. 3, pp. 305–314, 2004.

[10] B. Favier, N. J. Burroughs, L. Wedderburn, and S. Valitutti, "TCR dynamics on the surface of living T cells," *International Immunology*, vol. 13, no. 12, pp. 1525–1532, 2001.

[11] A. Lanzavecchia and F. Sallusto, "Antigen decoding by T lymphocytes: from synapses to fate determination," *Nature Immunology*, vol. 2, no. 6, pp. 487–492, 2001.

[12] H. H. Lin, S. Ray, S. Tongchusak, E. L. Reinherz, and V. Brusic, "Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research," *BMC Immunology*, vol. 9, article 8, 2008.

[13] K. H. Lee, A. D. Holdorf, M. L. Dustin, A. C. Chan, P. M. Allen, and A. S. Shaw, "T cell receptor signaling precedes immunological synapse formation," *Science*, vol. 295, no. 5559, pp. 1539–1542, 2002.

[14] E. Bell, "T-cell activation—T-cell-APC interactions," *Nature Reviews Immunology*, vol. 4, article 930, 2004.

[15] Z. Borovsky, G. Mishan-Eisenberg, E. Yaniv, and J. Rachmilewitz, "Serial triggering of T cell receptors results in incremental accumulation of signaling intermediates," *Journal of Biological Chemistry*, vol. 277, no. 24, pp. 21529–21536, 2002.

[16] K. Choudhuri, D. Wiseman, M. H. Brown, K. Gould, and P. A. van der Merwe, "T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand," *Nature*, vol. 436, no. 7050, pp. 578–582, 2005.

[17] M. L. Dustin, "The cellular context of T cell signaling," *Immunity*, vol. 30, no. 4, pp. 482–492, 2009.

[18] D. Gil, A. G. Schrum, B. Alarcón, and E. Palmer, "T cell receptor engagement by peptide—MHC ligands induces a conformational change in the CD3 complex of thymocytes," *Journal of Experimental Medicine*, vol. 201, no. 4, pp. 517–522, 2005.

[19] L. L. Jones, L. A. Colf, A. J. Bankovich et al., "Different thermodynamic binding mechanisms and peptide fine specificities associated with a panel of structurally similar high-affinity T cell receptors," *Biochemistry*, vol. 47, no. 47, pp. 12398–12408, 2008.

[20] A. Lanzavecchia, G. Iezzi, and A. Viola, "From TCR engagement to T cell activation: a kinetic view of T cell behavior," *Cell*, vol. 96, no. 1, pp. 1–4, 1999.

[21] C. T. H. Baker, G. A. Bocharov, and C. A. H. Paul, "Mathematical modelling of the interleukin-2 T-cell system: a comparative study of approaches based on ordinary and delay direrential equations," *Journal of Theoretical Medicine*, vol. 2, pp. 117–128, 1997.

[22] C. Bianca, "Mathematical modeling for keloid formation triggered by virus: malignant effects and immune system competition," *Mathematical Models and Methods in Applied Sciences*, vol. 21, no. 2, pp. 389–419, 2011.

[23] F. Pappalardo, M. D. Halling-Brown, N. Rapin et al., "ImmunoGrid, an integrative environment for large-scale simulation of the immune system for vaccine discovery, design and optimization," *Briefings in Bioinformatics*, vol. 10, no. 3, pp. 330–340, 2009.

[24] C. Bianca and M. Pennisi, "The triplex vaccine effects in mammary carcinoma: a nonlinear model in tune with SimTriplex," *Nonlinear Analysis: Real World Applications*, vol. 13, pp. 1913–1940, 2012.

[25] F. Pappalardo, P. L. Lollini, F. Castiglione, and S. Motta, "Modeling and simulation of cancer immunoprevention vaccine," *Bioinformatics*, vol. 21, no. 12, pp. 2891–2897, 2005.

[26] M. Pennisi, R. Catanuto, F. Pappalardo, and S. Motta, "Optimal vaccination schedules using simulated annealing," *Bioinformatics*, vol. 24, no. 15, pp. 1740–1742, 2008.

[27] G. L. Zhang, D. S. DeLuca, D. B. Keskin et al., "MULTIPRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles," *Journal of Immunological Methods*, vol. 374, no. 1–2, pp. 53–61, 2011.

[28] F. Pappalardo, P. Zhang, M. Halling-Brown et al., "Computational simulations of the immune system for personalized medicine: state of the art and challenges," *Current Pharmacogenomics and Personalized Medicine*, vol. 6, no. 4, pp. 260–271, 2008.

[29] S. Wan, P. V. Coveney, and D. R. Flower, "Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing," *Journal of Immunology*, vol. 175, no. 3, pp. 1715–1723, 2005.

[30] M. Zacharias and S. Springer, "Conformational flexibility of the MHC Class I $\alpha1$-$\alpha2$ domain in peptide bound and free states: a molecular dynamics simulation study," *Biophysical Journal*, vol. 87, no. 4, pp. 2203–2214, 2004.

[31] B. Knapp, U. Omasits, B. Bohle et al., "3-Layer-based analysis of peptide-MHC interaction: in silico prediction, peptide binding affinity and T cell activation in a relevant allergen-specific model," *Molecular Immunology*, vol. 46, no. 8-9, pp. 1839–1844, 2009.

[32] B. Knapp, U. Omasits, W. Schreiner, and M. M. Epstein, "A comparative approach linking molecular dynamics of altered peptide ligands and MHC with *In Vivo* immune responses," *PLoS ONE*, vol. 5, no. 7, Article ID e11653, 2010.

[33] X. Zhang, D. Bhatt, and D. M. Zuckerman, "Automated sampling assessment for molecular simulations using the effective sample size," *Journal of Chemical Theory and Computation*, vol. 6, no. 10, pp. 3048–3057, 2010.

[34] M. Held, "VRONI: an engineering approach to the reliable and efficient computation of Voronoi diagrams of points and line segments," *Computational Geometry*, vol. 18, no. 2, pp. 95–123, 2001.

[35] A. Okabe, B. Boots, K. Sugihara, S. N. Chiu, and D. G. Kendall, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Chichester John Wiley & Sons, 2nd edition, 2000.

[36] D. Knuth, *The Art of Computer Programming*, vol. 2 of *Seminumerical Algorithms*, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1989.

[37] A. E. García, "Large-amplitude nonlinear motions in proteins," *Physical Review Letters*, vol. 68, no. 17, pp. 2696–2699, 1992.

[38] B. Knapp, S. Frantal, M. Cibena, W. Schreiner, and P. Bauer, "Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible?" *Journal of Computational Biology*, vol. 18, no. 8, pp. 997–1005, 2011.

[39] A. Grossfield and D. M. Zuckerman, "Chapter 2 quantifying uncertainty and sampling quality in biomolecular simulations," *Annual Reports in Computational Chemistry*, vol. 5, pp. 23–48, 2009.

[40] H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data," *The Journal of Chemical Physics*, vol. 91, no. 1, pp. 461–466, 1989.

[41] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[42] L. Kjer-Nielsen, C. S. Clements, A. W. Purcell et al., "A structural basis for the selection of Dominant $\alpha\beta$ T cell receptors in antiviral immunity," *Immunity*, vol. 18, no. 1, pp. 53–64, 2003.

[43] A. A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack, "A graph-theory algorithm for rapid protein side-chain prediction," *Protein Science*, vol. 12, no. 9, pp. 2001–2014, 2003.

[44] B. Knapp, U. Omasits, and W. Schreiner, "Side chain substitution benchmark for peptide/MHC interaction," *Protein Science*, vol. 17, no. 6, pp. 977–982, 2008.

[45] B. Knapp, U. Omasits, S. Frantal, and W. Schreiner, "A critical cross-validation of high throughput structural binding prediction methods for pMHC," *Journal of Computer-Aided Molecular Design*, vol. 23, no. 5, pp. 301–307, 2009.

[46] L. Kjer-Nielsen, C. S. Clements, A. G. Brooks, A. W. Purcell, J. McCluskey, and J. Rossjohn, "The 1.5 Å crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance," *Structure*, vol. 10, no. 11, pp. 1521–1532, 2002.

[47] L. Kjer-Nielsen, C. S. Clements, A. G. Brooks et al., "The structure of HLA-B8 complexed to an immunodominant viral determinant: peptide-induced conformational changes and a mode of MHC class I dimerization," *Journal of Immunology*, vol. 169, no. 9, pp. 5153–5160, 2002.

[48] H. J. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, and J. Hermans, "Interaction models for water in relation to protein hydration," *Intermolecular Forces*, pp. 331–342, 1981.

[49] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GRGMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.

[50] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1656–1676, 2004.

[51] U. Omasits, B. Knapp, M. Neumann et al., "Analysis of key parameters for molecular dynamics of pMHC molecules," *Molecular Simulation*, vol. 34, no. 8, pp. 781–793, 2008.

[52] A. V. Hill, "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves," *Journal of Physiology*, vol. 40, supplement, pp. 4–7, 1910.

[53] J. D. Faraldo-Gómez, L. R. Forrest, M. Baaden et al., "Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations," *Proteins*, vol. 57, no. 4, pp. 783–791, 2004.

[54] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, "Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations," *Journal of Chemical Theory and Computation*, vol. 3, no. 1, pp. 26–41, 2007.

[55] L. J. Smith, X. Daura, and W. F. Van Gunsteren, "Assessing equilibration and convergence in biomolecular simulations," *Proteins*, vol. 48, no. 3, pp. 487–496, 2002.

[56] B. Hess, "Convergence of sampling in protein simulations," *Physical Review E*, vol. 65, no. 3, Article ID 031910, 10 pages, 2002.