



ELSEVIER

Contents lists available at ScienceDirect

## Biochemistry and Biophysics Reports

journal homepage: [www.elsevier.com/locate/bbrep](http://www.elsevier.com/locate/bbrep)

## Low-mass molecular dynamics simulation for configurational sampling enhancement: More evidence and theoretical explanation



Yuan-Ping Pang\*

Computer-Aided Molecular Design Laboratory, Mayo Clinic, Stabile 12-26, 200 First Street SW, Rochester, MN 55905, USA

## ARTICLE INFO

## Article history:

Received 10 June 2015

Received in revised form

30 July 2015

Accepted 31 August 2015

Available online 2 September 2015

## Keywords:

Molecular dynamics simulation

Protein folding

Folding rate

Folding time

Chignolin

CLN025

## ABSTRACT

It has been reported recently that classical, isothermal–isobaric molecular dynamics (NTP MD) simulations at a time step of 1.00 fs of the standard-mass time ( $\Delta t = 1.00 \text{ fs}^{\text{smt}}$ ) and a temperature of  $\leq 340 \text{ K}$  using uniformly reduced atomic masses by tenfold offers better configurational sampling than standard-mass NTP MD simulations at the same time step. However, it has long been reported that atomic masses can also be increased to improve configurational sampling because higher atomic masses permit the use of a longer time step. It is worth investigating whether standard-mass NTP MD simulations at  $\Delta t = 2.00$  or  $3.16 \text{ fs}^{\text{smt}}$  can offer better or comparable configurational sampling than low-mass NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$ . This article reports folding simulations of two  $\beta$ -hairpins showing that the configurational sampling efficiency of NTP MD simulations using atomic masses uniformly reduced by tenfold at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  is statistically equivalent to and better than those using standard masses at  $\Delta t = 3.16$  and  $2.00 \text{ fs}^{\text{smt}}$ , respectively. The results confirm that, relative to those using standard masses at routine  $\Delta t = 2.00 \text{ fs}^{\text{smt}}$ , the low-mass NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  are a simple and generic technique to enhance configurational sampling at temperatures of  $\leq 340 \text{ K}$ .

© 2015 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It has been reported recently that use of uniformly reduced atomic masses by tenfold (hereafter abbreviated as low masses) can enhance configurational sampling in classical, isothermal–isobaric molecular dynamics (NTP MD) simulations at a time step ( $\Delta t$ ) of 1.00 fs of the standard-mass time ( $\text{fs}^{\text{smt}}$ ) [1]. The reported sampling enhancement was determined by the smaller number of time steps required for the low-mass simulations to capture the folding of  $\beta$ -hairpin CLN025 [2] than that of the standard-mass NTP MD simulations performed at the same time step [1]. The folding of CLN025 in the low-mass simulations [1] was verified by the native-state conformation of CLN025 that was independently determined by an NMR spectroscopic study [2].

It has long been reported that increasing atomic masses can improve configurational sampling because higher atomic masses can reduce the fastest motions of the system in a molecular dynamics (MD) simulation to lengthen the  $\Delta t$  of the simulation and can gain larger momenta [3–26]. Downscaling the solvent mass can reportedly enhance the configurational sampling of

oligopeptides in MD simulations through reduction of solvent viscosity [27,28]. Differentially downscaling the solvent and side-chain masses also reportedly improved configurational sampling of a nonapeptide in MD simulations through reduction of solvent viscosity and adiabatic decoupling of motions of solvent, backbone, and side-chain [29].

Further, it has been reported that scaling the total mass by a factor of 10 for an MD simulation scales the time of the new system by a factor of  $\sqrt{10}$  [15,30]. In theory, as explained in Section 2.1, low-mass MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  are equivalent to standard-mass MD simulations at  $\Delta t = \sqrt{10} \text{ fs}^{\text{smt}}$ . In practice, it is reasonable to expect that low-mass NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  can be numerically equivalent to a standard-mass NTP MD simulations at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  since MD simulations are typically performed with the relatively high double-precision floating-point format. On one hand, because the SHAKE algorithm for bond constraint tends to fail when using  $\Delta t \geq 2.00 \text{ fs}^{\text{smt}}$  [15], the routine  $\Delta t$  of an MD simulation using AMBER MD programs with the SHAKE algorithm has long been  $2.00 \text{ fs}^{\text{smt}}$  at a temperature ( $T$ ) of  $\leq 300 \text{ K}$  (AMBER 14 reference manual, p289). On the other hand, according to the results reported in Ref. [1], standard-mass NTP MD simulations with the SHAKE algorithm could be performed at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  and  $T \leq 340 \text{ K}$  without the use of the hydrogen mass repartitioning scheme that is developed to avoid system instability caused by the use of  $\Delta t \geq 2.00 \text{ fs}^{\text{smt}}$  [15,19,26].

In this context, three aims were set to investigate (i) whether

Abbreviations:  $\text{C}\alpha\beta\text{RMSD}$ ,  $\text{C}\alpha$  and  $\text{C}\beta$  root mean square deviation;  $\Delta t$ , Time step; LMT, Low-mass time; MD, Molecular dynamics; NTP, Isothermal–isobaric; SD, Standard deviation; SE, Standard error; SMT, Standard-mass time;  $T$ , Temperature

\* Corresponding author.

E-mail address: [pang@mayo.edu](mailto:pang@mayo.edu)

standard-mass NTP MD simulations at  $\Delta t > 2.00 \text{ fs}^{\text{smt}}$  can actually be performed without the use of the hydrogen mass repartitioning scheme, (ii) whether the configuration sampling efficiency of standard-mass NTP MD simulations at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  is statistically equivalent to that of low-mass NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$ , and (iii) whether low-mass NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  can statistically offer better configurational sampling than standard-mass NTP MD simulations at  $\Delta t = 2.00 \text{ fs}^{\text{smt}}$ .

This article reports a comparative study of the three aims using 160 unique, independent, all-atom, and classical NTP MD simulations—each of which was performed with the SHAKE algorithm for  $500 \times 10^6$  time steps—to determine relative configurational sampling efficiencies of using mass scaling factors of 1.0 and 0.1 and time steps of 1.00, 2.00, 3.16, and  $3.50 \text{ fs}^{\text{smt}}$ . CLN025 and chignolin [31] were used as model systems for NTP MD simulations of miniprotein folding. To investigate the configurational sampling efficiency in a forcefield independent manner, two forcefields were used in this study—an up-to-date general-purpose AMBER forcefield FF14SB [32] (for either explicit or implicit solvation) and a special-purpose AMBER forcefield FF12MC [33] (for explicit solvation only).

## 2. Theory and methods

### 2.1. Equivalence of mass scaling and time-step scaling for sampling enhancement

Scaling the total mass by a factor of  $\lambda$  for an MD simulation scales the time of the new system by a factor of  $\sqrt{\lambda}$  [15,30]. The reason is because the sole purpose to scale total mass is to improve configurational sampling. Therefore, the units of distance  $[l]$  and energy  $[m]([l]/[t])^2$  of low-mass simulations are kept identical to those of standard-mass simulations. This is so that the structure and energy of the low-mass simulations can be compared to those of the standard-mass simulations in order to determine relative configurational sampling efficiencies. Let superscripts  $^{\text{lmt}}$  and  $^{\text{smt}}$  denote the times for the low-mass and standard-mass systems, respectively. Then  $[m^{\text{lmt}}] = 0.1[m^{\text{smt}}]$ ,  $[l^{\text{lmt}}] = [l^{\text{smt}}]$ , and  $[m^{\text{lmt}}]([l^{\text{lmt}}]/[t^{\text{lmt}}])^2 = [m^{\text{smt}}]([l^{\text{smt}}]/[t^{\text{smt}}])^2$  lead to  $\sqrt{10}[t^{\text{lmt}}] = [t^{\text{smt}}]$ . It is worth noting that a conventional MD simulation program takes a time step in the standard-mass time rather than the low-mass time. Therefore, low-mass MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  (*viz.*,  $\sqrt{10} \text{ fs}$  of the low-mass time) are theoretically equivalent to standard-mass MD simulations at  $\Delta t = \sqrt{10} \text{ fs}^{\text{smt}}$  (*viz.*,  $\sqrt{10} \text{ fs}$  of the standard-mass time), if both standard-mass and low-mass simulations are carried out for the same number of time steps and if there are no precision issues in performing these simulations.

### 2.2. Molecular dynamics simulations to autonomously fold chignolin and CLN025

Chignolin or CLN025 in a fully extended backbone conformation solvated with the TIP3P water [34] with surrounding counter ions and NaCl molecules was 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 energy-minimized system was then heated from 0 to 277, 300 or 340 K at a rate of 10 K/ps under constant temperature and constant volume, and finally simulated in 20 unique, independent, all-atom, and classical NTP MD simulations using PMEMD of AMBER 11 with a periodic boundary condition at 277, 300, or 340 K and 1 atm employing isotropic molecule-based scaling. The fully extended backbone conformations (*viz.*, anti-parallel  $\beta$ -strand conformations)

of chignolin and CLN025 were generated by MacPyMOL Version 1.5.0 (Schrödinger LLC, Portland, OR). The numbers of TIP3P waters and surrounding ions, initial solvation box size, ionizable residues, and computers used for the NTP MD simulations are provided in Table S1. The 20 unique seed numbers for initial velocities of Simulations 1–20 were taken from Ref. [35]. All simulations used (i) a dielectric constant of 1.0, (ii) the Berendsen coupling algorithm [36], (iii) the Particle Mesh Ewald method to calculate long-range electrostatic interactions [37], (iv)  $\Delta t = 1.00, 2.00, 3.16,$  or  $3.50 \text{ fs}^{\text{smt}}$ , (v) the SHAKE-bond-length constraints applied to all the 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 [38], (ix) a cutoff of  $8.0 \text{ \AA}$  for nonbonded interactions, and (x) default values of all other inputs of the PMEMD module. Instantaneous conformations of each simulation were saved at every  $10^5$  time steps. The forcefield parameter file of FF12MC is provided in Appendix A.

### 2.3. Aggregated native state population calculation

The native conformations of CLN025 in the NMR and crystal structures have Tyr2 and Trp9 on one side of the  $\beta$ -sheet and Tyr1 and Tyr10 on the other (see Ref. [1, Fig. 1A and B]). However, the reported NTP MD simulations showed that CLN025 could fold to native-like  $\beta$ -hairpins with Tyr1, Trp9, and Tyr10 on one side of the  $\beta$ -sheet and Tyr2 on the other (see Ref. [1 Fig. 1C]), or with Tyr1 and Trp9 on one side and Tyr2 and Tyr10 on the other (see Ref. [1 Fig. 1D]). Similarly, the reported NTP MD simulations also showed that chignolin could fold to the native  $\beta$ -hairpin with Tyr2 and Trp9 on the same side of the hairpin [31] and to native-like  $\beta$ -hairpins with Tyr2 on one side of the hairpin and Trp9 on the other [33].

The lowest  $\text{C}\alpha$  and  $\text{C}\beta$  root mean square deviation ( $\text{C}\alpha\beta\text{RMSD}$ ) between one of the native-like  $\beta$ -hairpins and the corresponding NMR structure of CLN025 is  $2.08 \text{ \AA}$ , whereas the corresponding  $\text{C}\alpha$  root mean square deviation is  $1.33 \text{ \AA}$ . The  $\text{C}\alpha\beta\text{RMSD}$  between one of the native-like  $\beta$ -hairpins and the NMR structure of chignolin is  $1.99 \text{ \AA}$ , but the corresponding  $\text{C}\alpha$  root mean square deviation is  $1.58 \text{ \AA}$ . In this study  $\text{C}\alpha\beta\text{RMSDs}$  and  $\text{C}\alpha$  root mean square deviations were calculated using PTRAJ of AmberTools 1.5 with root mean square fit all  $\alpha$  and  $\beta$  carbon atoms to the corresponding ones in the  $\beta$ -hairpin NMR structure without mass weighing. To distinguish the native  $\beta$ -hairpins from the native-like ones, in this study conformations of chignolin and CLN025 with  $\text{C}\alpha\beta\text{RMSDs}$  of  $\leq 1.96 \text{ \AA}$  relative to their NMR structures were considered to be at the native or folded state. The time series of  $\text{C}\alpha\beta\text{RMSD}$  from native conformations for chignolin and CLN025 revealed that these  $\beta$ -hairpins could fold into conformations with  $\text{C}\alpha\beta\text{RMSDs}$  of  $\sim 1.5 \text{ \AA}$  (Fig. S1). However, the  $\text{C}\alpha\beta\text{RMSD}$  cutoff for the native state was set at  $1.96 \text{ \AA}$  rather than  $1.50 \text{ \AA}$  because the  $\text{C}\alpha\beta\text{RMSD}$  between the NMR and crystal structures of CLN025 is  $1.95 \text{ \AA}$ . Otherwise, the use of a  $\text{C}\alpha\beta\text{RMSD}$  cutoff of  $\leq 1.50 \text{ \AA}$  would preclude conformations determined by crystallographic analysis that are commonly considered at the native state.

The individual native state population of chignolin or CLN025 in an MD simulation was calculated as the number of the  $\beta$ -hairpin conformations with  $\text{C}\alpha\beta\text{RMSDs}$  of  $\leq 1.96 \text{ \AA}$  divided by the number of all conformations saved at every  $10^5$  time steps. Averaging the individual native state populations of a set of 20 unique and independent MD simulations gave rise to the aggregated native state population for the set. The standard deviation (SD) and standard error (SE) of the aggregated native state population were calculated according to Eqs. (1) and (2) of Ref. [33], respectively, wherein  $N$  is the number of all simulations,  $P_i$  is the individual native state population of the  $i$ th simulation, and  $\bar{P}$  is the aggregated native state population.

## 2.4. Folding time estimation using survival analysis

The folding time ( $\tau_f$ ) or the reciprocal of the folding rate ( $k_f$ ) of a  $\beta$ -hairpin was estimated by the mean time of the  $\beta$ -hairpin to fold from a fully extended backbone conformation to the native conformation in 20 unique and independent NTP MD simulations, using survival analysis methods [39] from the R survival package (Therneau T.M., A Package for Survival Analysis in S, 2015, Version 2.38-3, <http://CRAN.R-project.org/package=survival>). The  $C\alpha\beta$ RMSD cutoff of  $\leq 1.96 \text{ \AA}$  was used to identify conformations at the native state. For each simulation with instantaneous conformations saved at every  $10^5$  time steps, the first time instant at which  $C\alpha\beta$ RMSD reached  $\leq 1.96 \text{ \AA}$  was recorded as an individual folding time (Fig. S1). If a set of 20 full simulations (each performed for 500 million time steps) all captured a folding event, this set was used to calculate the mean time-to-folding using a two-step procedure. The first step used the Kaplan–Meier estimator [40,41] with the Surv() function in the R survival package. The second used parametric survival functions—that mostly fell within the 95% confidence bounds of the Kaplan–Meier estimator—with the Surrge() function. If the mean of the Kaplan–Meier estimator was identical to the mean derived from a parametric survival function, then this survival function was used to calculate the time-course of the mean time-to-folding of the same set of simulations. If half or more than half of 20 shortened simulations (each performed for  $< 500$  million time steps) did not capture a folding event, the  $\tau_f$  of these simulations was discarded because of their overly large 95% confidence interval.

## 3. Results and discussion

### 3.1. Equal sampling of simulations using low-mass at $1.00 \text{ fs}^{\text{smt}}$ and standard-mass at $3.16 \text{ fs}^{\text{smt}}$

It was reported previously that CLN025 did not fold from a fully extended backbone conformation to its native conformation in 10 unique, independent, all-atom, and classical 500-million-time-step NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  using FF14SB at 277 K and 1 atm [1]. When 20 such simulations were performed under the same simulation conditions, except that the forcefield was changed from FF14SB to FF14SBlm, CLN025 folded in 13 of the 20 simulations (Fig. S1A) with an aggregated native state population with SE of  $17 \pm 6\%$  (Table 1). FF14SBlm has all the parameters of FF14SB, except that atomic masses are reduced uniformly by tenfold. When the FF14SBlm simulations were repeated under the same simulation conditions, except that FF14SBlm and  $\Delta t$  were changed to FF14SB and  $3.16 \text{ fs}^{\text{smt}}$ , respectively, CLN025 folded in 16 of the 20 simulations (Fig. S1B), with an aggregated native state population including SE of  $26 \pm 6\%$  (Table 1). Plotting the aggregated native state population as a function of the number of time steps shows no significant separation between the curves of the simulations using FF14SBlm at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  and those using FF14SB at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  (Fig. 1), according to the unpaired  $t$ -test two-tailed  $P$  value of 0.2465 for the two curves. Both sets of simulations using FF14SB and FF14SBlm did not converge well, according to (i) the large SDs relative to the means for FF14SB and FF14SBlm listed in Table 1 and (ii) the result that some simulations failed to capture a folding event (Fig. S1A and B). Therefore, the  $\tau_f$ s of CLN025 were not calculated for the two sets of simulations. Nevertheless, the aggregated native state populations do show that the configurational sampling of the NTP MD simulations of CLN025 using FF14SBlm at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  is statistically equivalent to that using FF14SB at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$ .

Next, 20 unique, independent, all-atom, and classical NTP MD simulations of chignolin at 300 K and 1 atm were carried out for

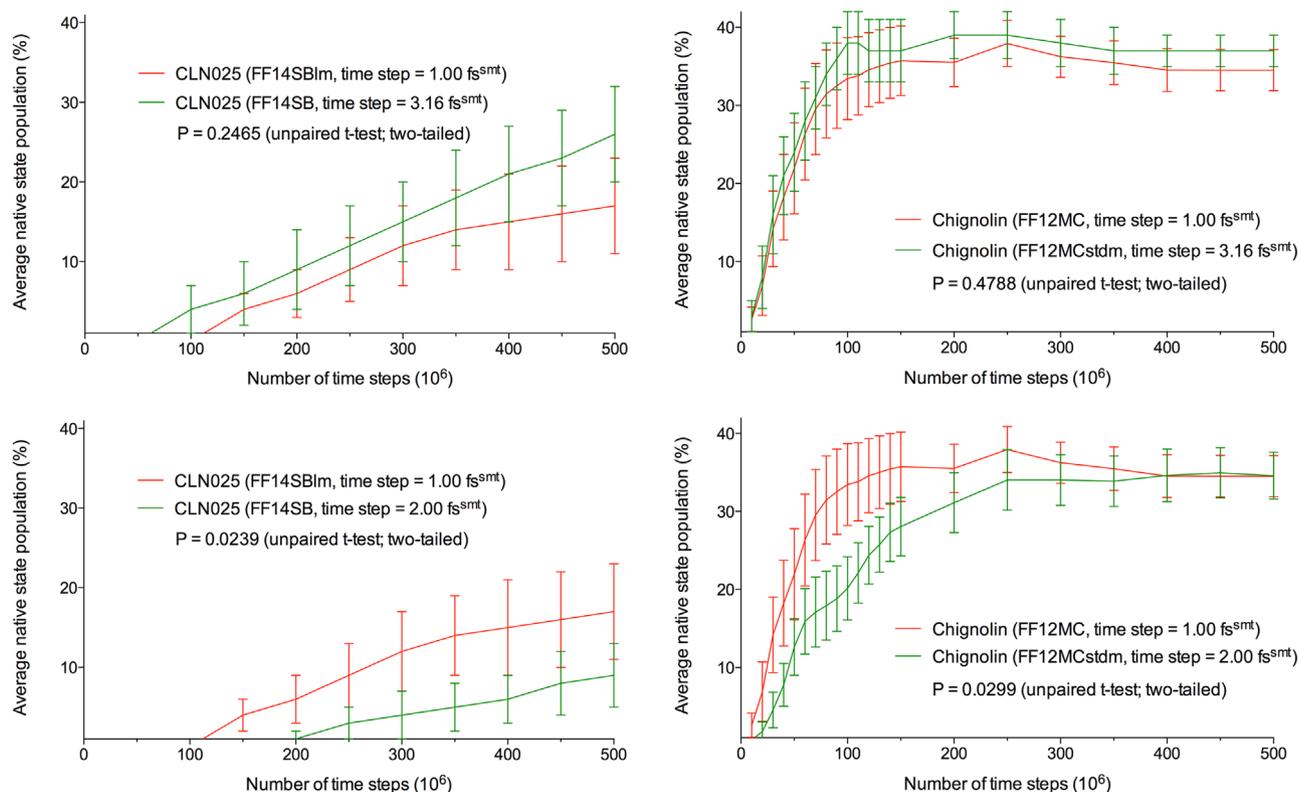
**Table 1**

Folding of CLN025 in 20 NTP MD simulations using different forcefields, temperature, and time steps.

Forcefield	Temp (K)	Time step ( $\text{fs}^{\text{smt}}$ )	Aggregated simulation time ( $\mu\text{s}^{\text{smt}}$ )	Aggregated native state population (%)		
				Mean	SD	SE
FF14SBlm	277	1.00	3.16	0	1	0
FF14SBlm	277	1.00	6.32	0	1	0
FF14SBlm	277	1.00	9.48	4	10	2
FF14SBlm	277	1.00	12.64	6	15	3
FF14SBlm	277	1.00	15.80	9	19	4
FF14SBlm	277	1.00	18.96	12	22	5
FF14SBlm	277	1.00	22.12	14	24	5
FF14SBlm	277	1.00	25.28	15	26	6
FF14SBlm	277	1.00	28.44	16	27	6
FF14SBlm	277	1.00	31.60	17	28	6
FF14SB	277	3.16	3.16	0	0	0
FF14SB	277	3.16	6.32	4	11	3
FF14SB	277	3.16	9.48	6	17	4
FF14SB	277	3.16	12.64	9	21	5
FF14SB	277	3.16	15.80	12	24	5
FF14SB	277	3.16	18.96	15	24	5
FF14SB	277	3.16	22.12	18	25	6
FF14SB	277	3.16	25.28	21	26	6
FF14SB	277	3.16	28.44	23	28	6
FF14SB	277	3.16	31.60	26	29	6
FF14SB	277	2.00	2.00	0	0	0
FF14SB	277	2.00	4.00	0	0	0
FF14SB	277	2.00	6.00	0	1	0
FF14SB	277	2.00	8.00	1	4	1
FF14SB	277	2.00	10.00	3	8	2
FF14SB	277	2.00	12.00	4	11	3
FF14SB	277	2.00	14.00	5	14	3
FF14SB	277	2.00	16.00	6	15	3
FF14SB	277	2.00	18.00	8	17	4
FF14SB	277	2.00	20.00	9	19	4
FF12MCstdm	340	3.16	3.16	29	26	6
FF12MCstdm	340	3.16	6.32	36	16	4
FF12MCstdm	340	3.16	9.48	41	14	3
FF12MCstdm	340	3.16	12.64	41	13	3
FF12MCstdm	340	3.16	15.80	42	12	3
FF12MCstdm	340	3.16	18.96	42	10	2
FF12MCstdm	340	3.16	22.12	41	9	2
FF12MCstdm	340	3.16	25.28	41	9	2
FF12MCstdm	340	3.16	28.44	42	9	2
FF12MCstdm	340	3.16	31.60	41	9	2

Aggregated native state population: number of conformations from 20 simulations with  $C\alpha\beta$ RMSDs of  $\leq 1.96 \text{ \AA}$  divided by number of all conformations from the 20 simulations. All MD simulations were performed for 500 million time steps with conditions described in Section 2 and Table S1. SD: Standard deviation. SE: Standard error.

$500 \times 10^6$  time steps under two conditions. One used FF12MC with  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$ , and the other used FF12MCstdm with  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$ . FF12MCstdm has all the parameters of FF12MC, except that all atomic masses are changed to standard values. Consistent with the report that FF12MC can reduce the number of time steps of all-atom and classical NTP MD simulation required to capture of the folding of chignolin [33], in this study chignolin folded at 300 K and 1 atm from a fully extended backbone conformation to its native conformation in all 20 NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  using FF12MC (Fig. S1C) or at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  using FF12MCstdm (Fig. S1D), respectively. The aggregated native state populations with SEs of the simulations using FF12MC and FF12MCstdm are converged to  $35 \pm 3\%$  and  $37 \pm 2\%$ , respectively (Tables 2A and 2B). The convergences of the two sets of simulations are further supported by the small SDs relative to the means of the populations (Tables 2A and 2B). Using the Kaplan–Meier estimator [40,41] (see Section 2.4), the  $\tau_f$ s of chignolin were estimated to  $79 \text{ ns}^{\text{smt}}$  (95% confidence interval of  $51\text{--}123 \text{ ns}^{\text{smt}}$ ) for FF12MC and  $72 \text{ ns}^{\text{smt}}$  (95% confidence interval of  $47\text{--}112 \text{ ns}^{\text{smt}}$ ) for



**Fig. 1.** Time series of aggregated native state population for chignolin and CLN025. The two-tailed  $P$  values were obtained from unpaired  $t$  tests using the PRISM 5 program.

**Table 2A**

Folding of chignolin in 20 NTP MD simulations at 300 K using FF12MC.

Time step (fs <sup>smt</sup> )	Aggregated simulation time (μs <sup>smt</sup> )	Aggregated native state population (%)			Estimated folding time (ns <sup>smt</sup> )			
		Mean	SD	SE	Mean	LCL	UCL	Event
1.00	0.632	3	7	2	–	–	–	4
1.00	1.264	7	17	4	–	–	–	7
1.00	1.896	14	22	5	87	53	145	15
1.00	2.528	18	25	5	92	56	150	16
1.00	3.160	22	26	6	82	53	129	19
1.00	3.792	26	26	6	79	51	123	20
1.00	4.424	30	26	6	79	51	123	20
1.00	5.056	31	25	6	79	51	123	20
1.00	5.688	33	24	5	79	51	123	20
1.00	6.320	33	23	5	79	51	123	20
1.00	6.952	34	22	5	79	51	123	20
1.00	7.584	35	21	5	79	51	123	20
1.00	8.216	35	21	5	79	51	123	20
1.00	8.848	35	20	5	79	51	123	20
1.00	9.480	36	20	4	79	51	123	20
1.00	12.640	36	14	3	79	51	123	20
1.00	15.800	38	13	3	79	51	123	20
1.00	18.960	36	12	3	79	51	123	20
1.00	22.120	35	13	3	79	51	123	20
1.00	25.280	35	12	3	79	51	123	20
1.00	28.440	35	12	3	79	51	123	20
1.00	31.600	35	12	3	79	51	123	20

Aggregated native state population: number of conformations from 20 simulations with  $C_{\alpha}\beta$ RMSDs of  $\leq 1.96$  Å divided by number of all conformations from the 20 simulations. All MD simulations were performed for 500 million time steps with conditions described in Methods and Table S1. SD: Standard deviation. SE: Standard error. LCL: Lower 95% confidence limit. UCL: Upper 95% confidence limit. Event: number of simulations that captured a folding event.

**Table 2B**

Folding of chignolin in 20 NTP MD simulations at 300 K using FF12MCstdm.

Time step (fs <sup>smt</sup> )	Aggregated simulation time (μs <sup>smt</sup> )	Aggregated native state population (%)			Estimated folding time (ns <sup>smt</sup> )			
		Mean	SD	SE	Mean	LCL	UCL	Event
3.16	0.632	3	8	2	–	–	–	4
3.16	1.264	8	17	4	–	–	–	10
3.16	1.896	16	23	5	95	55	163	13
3.16	2.528	21	22	5	74	47	118	18
3.16	3.160	24	21	5	73	47	114	19
3.16	3.792	28	21	5	74	48	117	19
3.16	4.424	31	20	4	72	47	112	20
3.16	5.056	34	20	4	72	47	112	20
3.16	5.688	36	20	4	72	47	112	20
3.16	6.320	38	19	4	72	47	112	20
3.16	6.952	38	18	4	72	47	112	20
3.16	7.584	37	17	4	72	47	112	20
3.16	8.216	37	17	4	72	47	112	20
3.16	8.848	37	17	4	72	47	112	20
3.16	9.480	37	17	4	72	47	112	20
3.16	12.640	39	15	3	72	47	112	20
3.16	15.800	39	15	3	72	47	112	20
3.16	18.960	38	13	3	72	47	112	20
3.16	22.120	37	12	3	72	47	112	20
3.16	25.280	37	11	2	72	47	112	20
3.16	28.440	37	9	2	72	47	112	20
3.16	31.600	37	7	2	72	47	112	20

Aggregated native state population: number of conformations from 20 simulations with  $C_{\alpha}\beta$ RMSDs of  $\leq 1.96$  Å divided by number of all conformations from the 20 simulations. All MD simulations were performed for 500 million time steps with conditions described in Methods and Table S1. SD: Standard deviation. SE: Standard error. LCL: Lower 95% confidence limit. UCL: Upper 95% confidence limit. Event: number of simulations that captured a folding event.

FF12MCstdm (Tables 2A and 2B).

As indicated by the unpaired  $t$ -test two-tailed  $P$  value of 0.4788

(Fig. 1), there is no significant separation between the two time series of aggregated native state population for chignolin at 300 K using FF12MC with  $\Delta t=1.00 \text{ fs}^{\text{smt}}$  and FF12MCstdm with  $\Delta t=3.16 \text{ fs}^{\text{smt}}$ , respectively. The  $\tau_f$  of chignolin of the FF12MC simulations is statistically equivalent to that of the FF12MCstdm simulations in terms of both mean and 95% confidence interval. These results confirm that the configurational sampling of the NTP MD simulations of chignolin using FF12MC at  $\Delta t=1.00 \text{ fs}^{\text{smt}}$  is statistically equivalent to that using FF12MCstdm at  $\Delta t=3.16 \text{ fs}^{\text{smt}}$ . As to the first two aims described in Section 1, these results suggest that standard-mass NTP MD simulations at  $\Delta t=3.16 \text{ fs}^{\text{smt}}$  can actually be performed without employing the hydrogen mass repartitioning scheme. The results also suggest that, regardless of which forcefield is used, configuration sampling efficiency of standard-mass NTP MD simulations at  $\Delta t=3.16 \text{ fs}^{\text{smt}}$  is statistically equivalent to that of low-mass NTP MD simulations at  $\Delta t=1.00 \text{ fs}^{\text{smt}}$ .

### 3.2. Better sampling of simulations using low-mass at $1.00 \text{ fs}^{\text{smt}}$ than standard-mass at $2.00 \text{ fs}^{\text{smt}}$

Repeating the above simulations of CLN025 at 277 K using FF14SB at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$  showed that CLN025 folded in ten of the 20 simulations (Fig. S1E) with an aggregated native state population including SE of  $9 \pm 4\%$  (Table 1). As indicated by the unpaired  $t$ -test two-tailed  $P$  value of 0.0239 (Fig. 1), there is a significant separation between the two time series of aggregated native state population for CLN025 at 277 K using FF14SBIm and FF14SB. These series are upward and have not reached plateau (Fig. 1). In addition, the SDs are larger than the means of the FF14SBIm and FF14SB simulations. These results indicate that both sets of simulations are not well converged. Nevertheless, the unpaired  $t$ -test two-tailed  $P$  value of 0.0239 indicates that the configurational sampling of the MD simulations of CLN025 using FF14SBIm at  $\Delta t=1.00 \text{ fs}^{\text{smt}}$  is significantly better than that using FF14SB at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$ .

Repeating the above simulations of chignolin at 300 K using FF12MCstdm at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$  revealed that chignolin folded in 19 of the 20 simulations (Fig. S1F), with an aggregated native state population including SE of  $35 \pm 3\%$  (Table 2C) and a  $\tau_f$  of  $147 \text{ ns}^{\text{smt}}$  (95% confidence interval of 94–230  $\text{ns}^{\text{smt}}$ ) that was estimated using the exponential survival function (see Section 2.4). As indicated by the unpaired  $t$ -test two-tailed  $P$  value of 0.0299 (Fig. 1), there is a significant separation between the two time series of aggregated native state population for chignolin at 300 K using FF12MC and FF12MCstdm. Both the mean and 95% confidence interval for the  $\tau_f$  of chignolin estimated from the FF12MCstdm simulations at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$  are nearly twice those obtained from the FF12MC simulations at  $\Delta t=1.00 \text{ fs}^{\text{smt}}$  (Tables 2A and 2C). These results confirm that the configurational sampling of the NTP MD simulations of chignolin using FF12MC at  $\Delta t=1.00 \text{ fs}^{\text{smt}}$  is significantly better than that using FF12MCstdm at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$ . As to the last aim described in Section 1, the results suggest that the low-mass NTP MD simulations at  $\Delta t=1.00 \text{ fs}^{\text{smt}}$  offer significantly better configurational sampling efficiency than the standard-mass NTP MD simulations at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$ .

### 3.3. Low-mass NTP MD simulation for configuration sampling enhancement

According to the survival analysis of the folding simulations, the  $\tau_f$  of chignolin estimated from the Kaplan–Meier estimator [40,41] using the simulation data of FF12MC at  $\Delta t=3.16 \text{ fs}^{\text{lmt}}$  (*viz.*,  $\Delta t=1.00 \text{ fs}^{\text{smt}}$ ) was identical to the one obtained from the exponential survival function. This is evident from the linear relationship between simulation time and natural logarithm of the nonnative state population for chignolin shown in Fig. 2. The  $\tau_f$  of

**Table 2C**

Folding of chignolin in 20 NTP MD simulations at 300 K using FF12MCstdm.

Time step ( $\text{fs}^{\text{smt}}$ )	Aggregated simulation time ( $\mu\text{s}^{\text{smt}}$ )	Aggregated native state population (%)			Estimated folding time ( $\text{ns}^{\text{smt}}$ )			
		Mean	SD	SE	Mean	LCL	UCL	Event
2.00	0.400	1	3	1	–	–	–	2
2.00	0.800	2	6	1	–	–	–	4
2.00	1.200	5	10	2	–	–	–	9
2.00	1.600	8	12	3	–	–	–	10
2.00	2.000	13	16	4	128	71	231	11
2.00	2.400	16	19	4	122	71	209	13
2.00	2.800	17	20	4	132	76	227	13
2.00	3.200	18	20	4	132	78	222	14
2.00	3.600	19	19	4	114	71	184	17
2.00	4.000	20	18	4	105	67	164	19
2.00	4.400	22	17	4	106	67	165	19
2.00	4.800	24	16	4	107	68	167	19
2.00	5.200	26	16	4	108	69	169	19
2.00	5.600	27	17	4	109	69	171	19
2.00	6.000	28	17	4	110	70	172	19
2.00	8.000	31	17	4	114	73	179	19
2.00	10.000	34	17	4	120	77	188	19
2.00	12.000	34	15	3	125	80	196	19
2.00	14.000	34	14	3	130	83	204	19
2.00	16.000	35	15	3	136	87	212	19
2.00	18.000	35	14	3	141	90	221	19
2.00	20.000	35	13	3	147	94	230	19

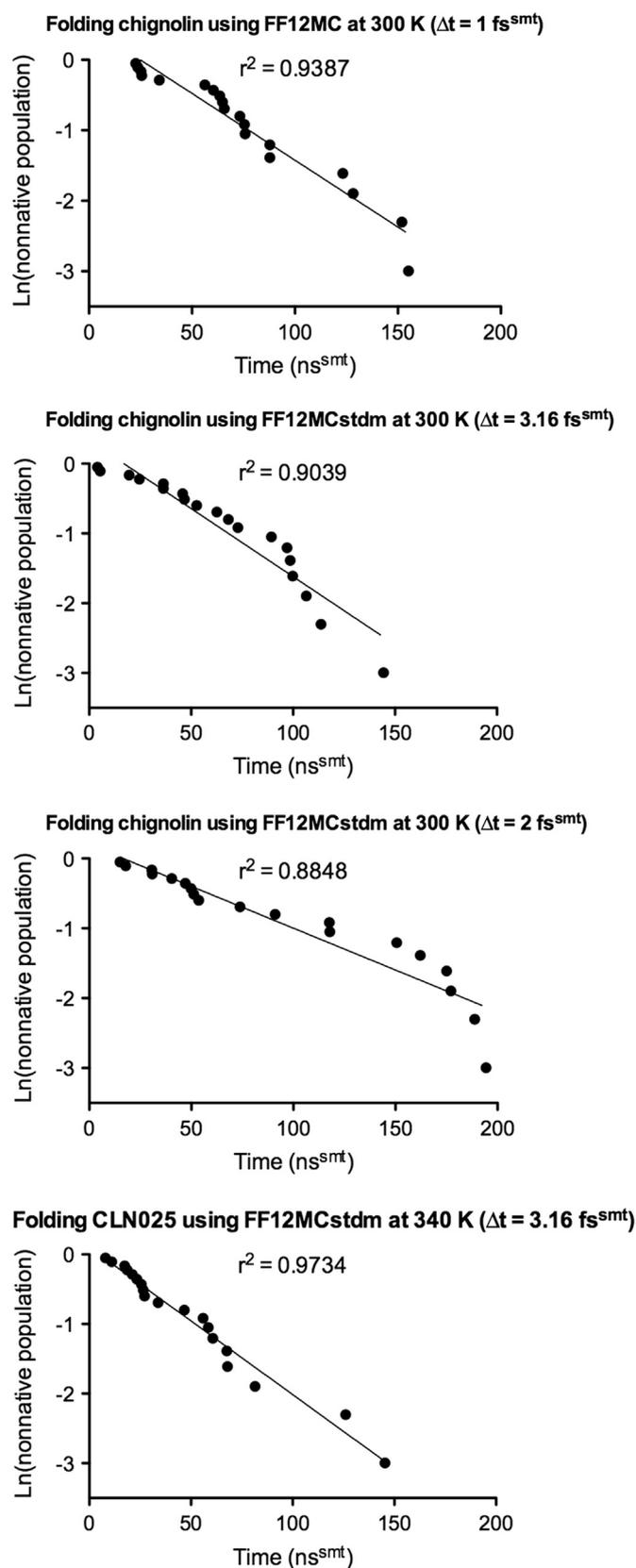
Aggregated native state population: number of conformations from 20 simulations with  $C_{\alpha}\beta$ RMSDs of  $\leq 1.96 \text{ \AA}$  divided by number of all conformations from the 20 simulations. All MD simulations were performed for 500 million time steps with conditions described in Methods and Table S1. SD: Standard deviation. SE: Standard error. LCL: Lower 95% confidence limit. UCL: Upper 95% confidence limit. Event: number of simulations that captured a folding event.

chignolin estimated from the Kaplan–Meier estimator using the data of FF12MCstdm at  $\Delta t=3.16 \text{ fs}^{\text{smt}}$  was also the same as the one obtained from the exponential model (Fig. 2). These observations of the exponential decay of the nonnative state population over simulation time indicate that the folding of chignolin observed in the NTP MD simulations followed a simple two-state kinetics scheme of Eq. (1), wherein D and N denote the nonnative and native conformations, respectively. These results are in excellent agreement with the reported two-state folding kinetics deduced from experimental studies of chignolin, a ten-residue  $\beta$ -hairpin [31]. The results also agree with the generalization that a mini-protein (with residues of  $< 100$ ) folds according to a two-state kinetics scheme [42]. This implies that the folding rate ( $k_f$ ) of such a miniprotein follows the first-order rate law, namely, Eq. (2) [43]. Most importantly, the results demonstrate that the folding simulations of chignolin using  $\Delta t=3.16 \text{ fs}^{\text{lmt}}$  and  $\Delta t=3.16 \text{ fs}^{\text{smt}}$  are realistic.

$$D \xrightleftharpoons[k_u]{k_f} N \quad (1)$$

$$\ln \left( \frac{[D]}{[D]_0} \right) = -k_f t = -\frac{1}{\tau_f} t \quad (2)$$

By contrast, the  $\tau_f$  of chignolin simulated with FF12MCstdm at  $\Delta t=2.00 \text{ fs}^{\text{smt}}$  could not be estimated from the Kaplan–Meier estimator because one of the 20 simulations failed to capture a folding event (Table 2C); the relationship between simulation time and natural logarithm of the nonnative population of chignolin simulated with FF12MCstdm at  $\Delta t=2.00 \text{ fs}^{\text{lmt}}$  ( $r^2=0.8848$ , Fig. 2) is not as linear as those with FF12MC at  $\Delta t=3.16 \text{ fs}^{\text{lmt}}$  and



**Fig. 2.** Plots of natural logarithm of the nonnative population versus simulation time. The individual folding times were taken from the data provided in Fig. S1. The linear regression analysis was performed using the PRISM 5 program.

FF12MCstdm at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  ( $r^2 = 0.9387$  and  $0.9039$ , respectively; Fig. 2A and B). Consequently, the  $\tau_f$  of chignolin simulated with FF12MCstdm at  $\Delta t = 2.00 \text{ fs}^{\text{smt}}$  has to be estimated using the

exponential survival function and has a very large 95% confidence interval relative to those with  $\Delta t = 3.16 \text{ fs}^{\text{lmt}}$  and  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  (Table 2).

In addition, the number of time steps required to capture folding events in the simulations with  $\Delta t = 3.16 \text{ fs}^{\text{lmt}}$  or  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  can be reduced substantially relative to that with  $\Delta t = 2.00 \text{ fs}^{\text{smt}}$ . To fold chignolin at 300 K in 20 NTP MD simulations at  $\Delta t = 3.16 \text{ fs}^{\text{lmt}}$ , only  $120 \times 10^6$  time steps are needed for each of the 20 simulations to obtain a converged aggregated native state population. By contrast,  $400 \times 10^6$  time steps are needed to obtain the converged population for the 20 simulations of chignolin at  $\Delta t = 2.00 \text{ fs}^{\text{smt}}$  using FF12MCstdm (Fig. 1). To run an NTP MD simulation of chignolin at  $\Delta t = 2.00 \text{ fs}^{\text{smt}}$ , 300 K, and 1 atm using FF12MCstdm and other conditions specified in Section 2.2 on a 12-core Apple Mac Pro with Intel Westmere (2.93 GHz), the average wall-clock timing is  $69.52 \text{ ns}^{\text{smt}}$  per day. The difference in number of time steps between the two sets of simulations ( $280 \times 10^6$  time steps) translates to a saving of 193.33 hours of wall-clock computing time if the 20 simulations are performed in parallel or a saving of 3866.51 hours of wall-clock computing time if the 20 simulations are done in serial. To fold CLN025 at 277 K in 20 NTP MD simulations, changing  $\Delta t$  from  $2.00 \text{ fs}^{\text{smt}}$  to  $3.16 \text{ fs}^{\text{lmt}}$  can save 210.67 or 4213.48 hours of computing time when the 20 simulations are performed in parallel or serial, respectively, according to the average timing of  $56.96 \text{ ns}^{\text{smt}}$  per day to simulate CLN025 at 277 K under other identical conditions as those for chignolin at 300 K.

There is a clear incentive from the present work to increase  $\Delta t$  from  $2.00 \text{ fs}^{\text{smt}}$  to  $3.16 \text{ fs}^{\text{lmt}}$  or  $3.16 \text{ fs}^{\text{smt}}$  or perhaps  $\geq 3.16 \text{ fs}^{\text{smt}}$ . However, the undesired reduction of integration accuracy can outweigh the desired reduction of computing time when  $\Delta t$  is  $\geq 3.16 \text{ fs}^{\text{smt}}$  and when the hydrogen mass repartitioning scheme is not used in an MD simulation. Indeed, additional studies show that four of 20 unique and independent NTP MD simulations of chignolin at  $\Delta t = 3.50 \text{ fs}^{\text{smt}}$  using FF12MCstdm at 340 K failed after  $7 \times 10^7$  time steps. But, without employing the hydrogen mass repartitioning scheme, all 20 unique and independent standard-mass NTP MD simulations of CLN025 at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  using FF12MCstdm at 340 K could be carried out successfully for  $5 \times 10^8$  time steps to capture folding events (Fig. S1G), with an aggregated native state population including SE of  $41 \pm 2\%$  and a  $\tau_f$  of  $55 \text{ ns}^{\text{smt}}$  at 340 K with 95% confidence interval of 35–85 ns (Table 1 and Fig. 2). No studies have been reported to show that in practice standard-mass NTP MD simulations of regular and large proteins do not fail at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$  and  $T = 300\text{--}340 \text{ K}$ . On the other hand, years of low-mass NTP MD simulations of both regular and large proteins with biological relevance [44–56] have already been done by this author to confirm that various proteins can be simulated successfully at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  and  $T \leq 340 \text{ K}$ . Lastly, it should be pointed out that uniformly reducing atomic masses of the original system does not change the physical composition of the new system. The kinetic properties derived from low-mass NTP MD simulations can be scaled back to the properties of the original system by a factor of  $\sqrt{10}$ . For these reasons, low-mass NTP MD simulations at  $\Delta t = 1.00 \text{ fs}^{\text{smt}}$  are presently preferred over standard-mass NTP MD simulations at  $\Delta t = 3.16 \text{ fs}^{\text{smt}}$ . Such low-mass NTP MD simulations can be used as a simple and generic technique to enhance configurational sampling for both kinetic and thermodynamic investigations of proteins at  $T \leq 340 \text{ K}$ .

#### 4. Conclusions

The present work demonstrates that the configurational sampling efficiency of low-mass NTP MD simulations of  $\beta$ -hairpin

folding at  $\Delta t=1.00\text{ fs}^{\text{smt}}$  is statistically equivalent to and higher than those using standard masses at  $\Delta t=3.16$  and  $2.00\text{ fs}^{\text{smt}}$ , respectively. This work also shows that, without employing the hydrogen mass repartitioning scheme, standard-mass NTP MD simulations at  $\Delta t=3.16\text{ fs}^{\text{smt}}$  and  $T=277, 300,$  and  $340\text{ K}$  can be performed to capture the autonomous  $\beta$ -hairpin folding. While further studies may show that with sufficient computational precision standard-mass NTP MD simulations at  $\Delta t=3.16\text{ fs}^{\text{smt}}$  might be simpler and equally effective for configurational sampling enhancement at  $T \leq 340\text{ K}$ , the present results together with the previous results reported by this author confirm that low-mass NTP MD simulations are a simple and generic technique to enhance configurational sampling for both kinetic and thermodynamic investigations of proteins at  $T \leq 340\text{ K}$ .

## 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 Materiel Command (W81XWH-04-2-0001), the US Army Research Office (DAAD19-03-1-0318 and W911NF-09-1-0095), the US Department of Defense High Performance Computing Modernization Office, and the Mayo Foundation for Medical Education and Research. The author thanks two anonymous reviewers for their comments and suggestions. He also thanks Professor Terry M. Therneau for developing and maintaining the open-source R Survival package (Version 2.38-3) and for guidance on using the package to estimate the mean time-to-event. The contents of this article are the sole responsibility of the author and do not necessarily represent the official views of the funders.

## Appendix A. Supplementary material

Supplementary data (Table S1, Fig. S1, and the forcefield parameter file of FF12MC) associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.bbrep.2015.08.023>.

## Transparency document

The author declares no conflict of interest.

## References

- [1] 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.
- [2] S. Honda, T. Akiba, Y.S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Oishi, H. Watanabe, T. Odahara, K. Harata, Crystal structure of a ten-amino acid protein, *J. Am. Chem. Soc.* 130 (2008) 15327–15331.
- [3] G. Jacucci, A. Rahman, The possibility of using a larger timestep in molecular dynamics studies of water, in: C. Moser (Ed.), *Report on Workshop Methods in Molecular Dynamics-Long Timescale Events*, CECAM, Orsay, France, 1974, pp. 32–40.
- [4] C.H. Bennett, Mass tensor molecular dynamics, in: C. Moser (Ed.), *Report on Workshop Methods in Molecular Dynamics – Long Timescale Events*, CECAM, Orsay, French, 1974, pp. 41–63.
- [5] C.H. Bennett, Mass tensor molecular dynamics, *J. Comput. Phys.* 19 (1975) 267–279.
- [6] D.W. Wood, Computer simulation of water and aqueous solutions, in: F. Franks (Ed.), *Water: A Comprehensive Treatise*, Plenum Press, New York, 1979, pp. 279–436.
- [7] R. Pomes, J.A. McCammon, Mass and step length optimization for the calculation of equilibrium properties by molecular dynamics simulation, *Chem. Phys. Lett.* 166 (1990) 425–428.
- [8] B. Mao, A.R. Friedman, Molecular dynamics simulation by atomic mass weighting, *Biophys. J.* 58 (1990) 803–805.
- [9] B. Mao, Mass-weighted molecular dynamics simulation and conformational analysis of polypeptide, *Biophys. J.* 60 (1991) 611–622.
- [10] B. Mao, Mass-weighted molecular dynamics simulation of the protein-ligand complex of rhizopuspepsin and inhibitor, *Biophys. J.* 60 (1991) 966–973.
- [11] B. Mao, G.M. Maggiora, K.C. Chou, Mass-weighted molecular dynamics simulation of cyclic polypeptides, *Biopolymers* 31 (1991) 1077–1086.
- [12] E. Egberts, S.-J. Marrink, H.J.C. Berendsen, Molecular dynamics simulation of a phospholipid membrane, *Eur. Biophys. J.* 22 (1994) 423–436.
- [13] S. Elamrani, M.B. Berry, G.N. Phillips, J.A. McCammon, Study of global motions in proteins by weighted masses molecular dynamics: adenylate kinase as a test case, *Proteins* 25 (1996) 79–88.
- [14] P. Guntert, C. Mumenthaler, K. Wuthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol.* 273 (1997) 283–298.
- [15] K.A. Feenstra, B. Hess, H.J.C. Berendsen, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems, *J. Comput. Chem.* 20 (1999) 786–798.
- [16] U. Stocker, D. Juchli, W.F. van Gunsteren, Increasing the time step and efficiency of molecular dynamics simulations: Optimal solutions for equilibrium simulations or structure refinement of large biomolecules, *Mol. Simul.* 29 (2003) 123–138.
- [17] J. Franklin, S. Doniach, Adaptive time stepping in biomolecular dynamics, *J. Chem. Phys.* 123 (2005) 124909.
- [18] W. Han, Y.D. Wu, Coarse-grained protein model coupled with a coarse-grained water model: molecular dynamics study of polyalanine-based peptides, *J. Chem. Theory Comput.* 3 (2007) 2146–2161.
- [19] M.J. Harvey, G. Giupponi, G. De Fabritiis, ACEMD: accelerating biomolecular dynamics in the microsecond time scale, *J. Chem. Theory Comput.* 5 (2009) 1632–1639.
- [20] H. Zheng, S.L. Wang, Y.K. Zhang, Increasing the time step with mass scaling in Born-Oppenheimer ab initio QM/MM molecular dynamics simulations, *J. Comput. Chem.* 30 (2009) 2706–2711.
- [21] W. Han, C.K. Wan, F. Jiang, Y.D. Wu, PACE force field for protein simulations. 1. Full parameterization of version 1 and verification, *J. Chem. Theory Comput.* 6 (2010) 3373–3389.
- [22] E. Tsuchida, Ab initio mass tensor molecular dynamics, *J. Chem. Phys.* 134 (2011) 044112.
- [23] F. Rao, M. Spichty, Thermodynamics and kinetics of large-time-step molecular dynamics, *J. Comput. Chem.* 33 (2012) 475–483.
- [24] S. Michielssens, T.S. van Erp, C. Kutzner, A. Ceulemans, B.L. de Groot, Molecular dynamics in principal component space, *J. Phys. Chem. B* 116 (2012) 8350–8354.
- [25] T. Nagai, T. Takahashi, Mass-scaling replica-exchange molecular dynamics optimizes computational resources with simpler algorithm, *J. Chem. Phys.* 141 (2014) 114111.
- [26] C.W. Hopkins, S. Le Grand, R.C. Walker, A.E. Roitberg, Long-time-step molecular dynamics through hydrogen mass repartitioning, *J. Chem. Theory Comput.* 11 (2015) 1864–1874.
- [27] P.J. Gee, W.F. van Gunsteren, Numerical simulation of the effect of solvent viscosity on the motions of a beta-peptide heptamer, *Chem. Eur. J.* 12 (2006) 72–75.
- [28] P.H. Nguyen, Replica exchange simulation method using temperature and solvent viscosity, *J. Chem. Phys.* 132 (2010) 144109.
- [29] I.C. Lin, M.E. Tuckerman, Enhanced conformational sampling of peptides via reduced side-chain and solvent masses, *J. Phys. Chem. B* 114 (2010) 15935–15940.
- [30] R. Walsler, B. Hess, A.E. Mark, W.F. van Gunsteren, Further investigation on the validity of Stokes–Einstein behaviour at the molecular level, *Chem. Phys. Lett.* 334 (2001) 337–342.
- [31] S. Honda, K. Yamasaki, Y. Sawada, H. Morii, 10 residue folded peptide designed by segment statistics, *Structure* 12 (2004) 1507–1518.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] Y.-P. Pang, Use of 1–4 interaction scaling factors to control the conformational equilibrium between  $\alpha$ -helix and  $\beta$ -strand, *Biochem. Biophys. Res. Commun.* 457 (2015) 183–186.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] T.M. Therneau, P.M. Grambsch, *Modeling Survival Data: Extending the Cox Model*, Springer-Verlag, New York, 2000.

- [40] E.L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, *J. Am. Stat. Assoc.* 53 (1958) 457–481.
- [41] J.T. Rich, J.G. Neely, R.C. Paniello, C.C. Voelker, B. Nussenbaum, E.W. Wang, A practical guide to understanding Kaplan–Meier curves, *Otolaryngol. Head Neck Surg.* 143 (2010) 331–336.
- [42] S.E. Jackson, How do small single-domain proteins fold? *Fold. Des.* 3 (1998) R81–R91.
- [43] P. Atkins, *Physical Chemistry*, W. H. Freeman and Company, New York, 1994.
- [44] Y.-P. Pang, Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment, *J. Mol. Model.* 5 (1999) 196–202.
- [45] Y.-P. Pang, K. Xu, J. El Yazal, F.G. Prendergast, Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach, *Protein Sci.* 9 (2000) 1857–1865.
- [46] Y.-P. Pang, Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method, *Proteins* 45 (2001) 183–189.
- [47] H. Sun, J. El Yazal, O. Lockridge, L.M. Schopfer, S. Brimijoin, Y.-P. Pang, Predicted Michaelis–Menten complexes of cocaine-butrylcholinesterase-engineering effective butrylcholinesterase mutants for cocaine detoxication, *J. Biol. Chem.* 276 (2001) 9330–9336.
- [48] Y.-P. Pang, Three-dimensional model of a substrate-bound SARS chymotrypsin-like cysteine proteinase predicted by multiple molecular dynamics simulations: catalytic efficiency regulated by substrate binding, *Proteins* 57 (2004) 747–757.
- [49] J.G. Park, P.C. Sill, E.F. Makiyi, A.T. Garcia-Sosa, C.B. Millard, J.J. Schmidt, Y.-P. Pang, Serotype-selective, small-molecule inhibitors of the zinc endopeptidase of botulinum neurotoxin serotype A, *Bioorg. Med. Chem.* 14 (2006) 395–408.
- [50] Y.-P. Pang, Novel acetylcholinesterase target site for malaria mosquito control, *PLoS One* 1 (2006) e58.
- [51] Y.-P. Pang, Species marker for developing novel and safe pesticides, *Bioorg. Med. Chem. Lett.* 17 (2007) 197–199.
- [52] B.J. Killian, J.Y. Kravitz, S. Somani, P. Dasgupta, Y.-P. Pang, M.K. Gilson, Configurational entropy in protein–peptide binding: computational study of TSG101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide, *J. Mol. Biol.* 389 (2009) 315–335.
- [53] F. Ekström, A. Hörnberg, E. Artursson, L.-G. Hammarström, G. Schneider, Y.-P. Pang, Structure of HI-6-sarin-acetylcholinesterase determined by X-ray crystallography and molecular dynamics simulation: reactivator mechanism and design, *PLoS ONE* 4 (2009) e5957.
- [54] Y.-P. Pang, H. Dai, A. Smith, X.W. Meng, P.A. Schneider, S.H. Kaufmann, Bak conformational changes induced by ligand binding: insight into BH3 domain binding and Bak homo-oligomerization, *Sci. Rep.* 2 (2012) 257.
- [55] V.P. Chen, Y. Gao, L.Y. Geng, R.J. Parks, Y.-P. Pang, S. Brimijoin, Plasma butrylcholinesterase regulates ghrelin to control aggression, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) 2251–2256.
- [56] L. Zhang, S. Zhang, I. Maezawa, S. Trushin, P. Minhas, M. Pinto, L.-W. Jin, K. Prasain, T.D.T. Nguyen, Y. Yamazaki, T. Kanekiyo, G. Bu, B. Gateno, K.-O. Chang, K.A. Nath, E. Nemetlu, P. Dzeja, Y.-P. Pang, D.H. Hua, E. Trushina, Modulation of mitochondrial complex I activity averts cognitive decline in multiple animal models of familial Alzheimer’s disease, *EBioMedicine* 2 (2015) 294–305.