Biochemical and Biophysical Research Communications 452 (2014) 588–592

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Low-mass molecular dynamics simulation: A simple and generic technique to enhance configurational sampling

Yuan-Ping Pang *

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

ARTICLE INFO

Article history:
Received 14 August 2014
Available online 30 August 2014

Keywords:
Configurational sampling enhancement
Isothermal–isobaric ensemble
Protein folding
Fast-folding miniature protein
Chignolin analogue
β-Hairpin

ABSTRACT

CLN025 is one of the smallest fast-folding proteins. Until now it has not been reported that CLN025 can autonomously fold to its native conformation in a classical, all-atom, and isothermal–isobaric molecular dynamics (MD) simulation. This article reports the autonomous and repeated folding of CLN025 from a fully extended backbone conformation to its native conformation in explicit solvent in multiple 500-ns MD simulations at 277 K and 1 atm with the first folding event occurring as early as 66.1 ns. These simulations were accomplished by using AMBER forcefield derivatives with atomic masses reduced by 10-fold on Apple Mac Pros. By contrast, no folding event was observed when the simulations were repeated using the original AMBER forcefields of FF12SB and FF14SB. The results demonstrate that low-mass MD simulation is a simple and generic technique to enhance configurational sampling. This technique may propel autonomous folding of a wide range of miniature proteins in classical, all-atom, and isothermal–isobaric MD simulations performed on commodity computers—an important step forward in quantitative biology.

1. Introduction

CLN025 is one of the smallest fast-folding proteins that folds into a β-hairpin with a sequence of YYDPETGTWY according to both its X-ray crystal and nuclear magnetic resonance (NMR) structures (Fig. 1A and B) [1]. Autonomous and repeated folding of CLN025 has been observed in a classical, all-atom, canonical, and 106-μs molecular dynamics (MD) simulation performed on a one-of-a-kind, proprietary, special-purpose computer [2]. This article reports the use of low atomic masses to speed up configurational sampling in MD simulations and its successful application to autonomous folding of CLN025 in classical, all-atom, isothermal–isobaric, and 500-ns MD simulations with TIP3P water [3] at the same temperature and pressure conditions (277 K and 1 atm) as those used for the experimental folding study of CLN025 [1].

2. Materials and methods

2.1. Folding simulation protocol

A fully extended backbone conformation of CLN025 was generated by MacPyMOL Version 1.5.0 (Schrödinger LLC, Portland, OR) and solvated with 1532 TIP3P water molecules to keep the closest distance between any atom of CLN025 and the edge of the periodic solvent box at 8.2 Å using LEAP of AmberTools 1.5 (University of California, San Francisco). Because pH 5.7 was used for the NMR structure determination for CLN025 [1], two sodium ions were added for neutrality of the protein. Four sodium chloride molecules were also added to keep the ionic strength of the system at ~143 mM NaCl. Although the ionic strength used for the NMR study was 20 mM Na_3PO_4 [1], in this study ~143 mM NaCl was used to minimize the intermolecular interaction of CLN025 with its periodic image during the MD simulations. Using FF12SB or its derivatives as described in Section 3, the solvated and slightly brined CLN025 was 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), heated from 0 to 277 K at a rate of 10 K/ps under constant temperature and volume, and finally simulated in ten independent MD

Abbreviations: MD, molecular dynamics; LMD, low-mass molecular dynamics; NMR, nuclear magnetic resonance; CαβRMSD, Cα and Cβ root mean square deviation; CRMSD, all carbon atom root mean square deviation.

* Address: Stabile 12-26, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA.

E-mail address: pang@mayo.edu

http://dx.doi.org/10.1016/j.bbrc.2014.08.119
0006-291X/

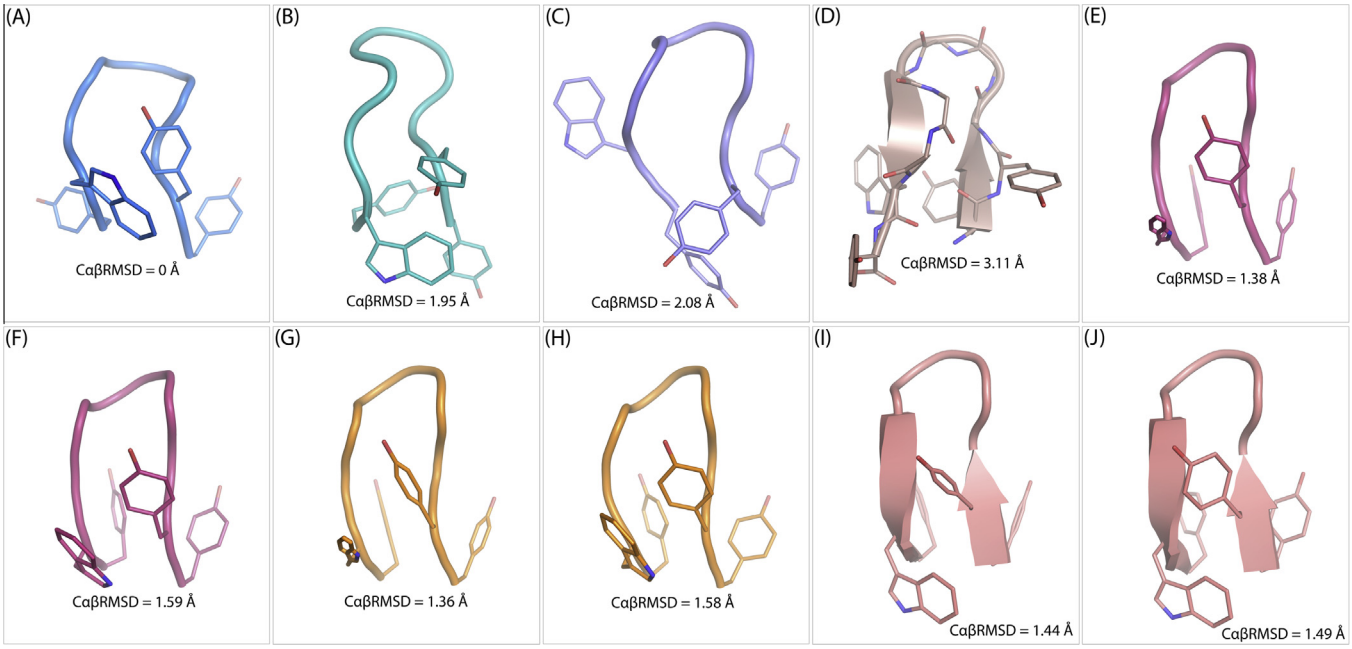

Fig. 1. Native and native-like conformations of CLN025 obtained from experimental and computational studies and their Cα and Cβ root mean square deviations (CαβRMSDs). (A) The NMR structure [1]. (B) The crystal structure [1]. (C) A native-like conformation obtained from simulations using FF12SBlm. (D) Another native-like conformation obtained from simulations using FF12SBlm. (E) The time-averaged conformation of the largest cluster of conformations obtained from the simulations using FF12SBlm. (F) The representative conformation of the largest cluster of conformations obtained from simulations using FF12SBlm. (G) The time-averaged conformation of the largest cluster of conformations obtained from the simulations using FF14SBlm. (H) The representative conformation of the largest cluster of conformations obtained from the simulations using FF14SBlm. (I): The time-averaged conformation of the second largest cluster of conformations obtained from the simulations using FF14SBlm. (J) The representative conformation of the second largest cluster of conformations obtained from the simulations using FF14SBlm.

simulations using PMEMD of AMBER 11 with a periodic boundary condition at a constant temperature of 277 K and a constant pressure of 1 atm with isotropic molecule-based scaling. The ten unique seed numbers for initial velocities of Simulations 1–10 are 1804289383, 846930886, 1681692777, 1714636915, 1957747793, 424238335, 719885386, 1649760492, 596516649, and 1189641421, respectively. All simulations used (1) a dielectric constant of 1.0, (2) the Berendsen coupling algorithm [4], (3) the Particle Mesh Ewald method to calculate long-range electrostatic interactions [5], (4) a time step of 1.0 fs, (5) SHAKE-bond-length constraints applied to all the bonds involving the H atom, (6) a protocol to save the image closest to the middle of the "primary box" to the restart and trajectory files, (7) a formatted restart file, and (8) default values of all other inputs of PMEMD. Each simulation was performed on a 12-core Apple Mac Pro with Intel Westmere (2.40/2.93 GHz). Trajectories were saved at 100-ps intervals in all simulations.

2.2. Folding simulation analysis

The native conformations of CLN025 in the NMR and crystal structures have Tyr2 and Trp9 on one side of the β-sheet and Tyr1 and Tyr10 on the other side [1] (Fig. 1A and B). Interestingly, analysis of the trajectories obtained from MD simulations showed that CLN025 could fold to native-like β-hairpins with Tyr1, Trp9, and Tyr10 on one side of the β-sheet and Tyr2 on the other side (Fig. 1C) or with Tyr1 and Trp9 on one side and Tyr2 and Tyr10 on the other side (Fig. 1D). The lowest Cα and Cβ root mean square deviation (CαβRMSD) between one of the native-like β-hairpins and the NMR structure is 2.08 Å, whereas the corresponding Cα root mean square deviation is 1.33 Å. The CαβRMSD between the NMR and crystal structures is 1.95 Å. Therefore, to distinguish the native β-hairpins from the native-like ones, conformations with CαβRMSDs of $\leqslant$1.96 Å relative to the NMR structure are considered to be at the native or folded state.

A folding event of CLN025 in an MD simulation is delineated by (1) a smoothed curve of CαβRMSD over simulation time (Fig. 2), (2) an average folding time (Table 1), and (3) an overall native-state population (Table 1). The smoothed curves were generated by the PRISM program from GraphPad Software (La Jolla, California) using 32 neighbors on each size and 6th order of the smoothing polynomial. The native-state population of a simulation is the number of conformations with CαβRMSDs of $\leqslant$1.96 Å divided by all conformations obtained from the simulation. If the native-state population of a simulation is $\leqslant$1.5%, the conformations with CαβRMSDs of $\leqslant$1.96 Å are considered to be too transient to constitute a folding event. In other words, if a simulation captures a folding event, the native-state population of this simulation must be >1.5%. The overall native-state population is the number of conformations with CαβRMSDs of $\leqslant$1.96 Å obtained from a set of simulations divided by all conformations of the set. A folding time—defined as the time to fold a protein from a fully extended backbone conformation to its native conformation—is obtained from the first time when CαβRMSD reaches $\leqslant$1.96 Å in a simulation that must have its native-state population of >1.5%. If a simulation does not capture a folding event, the folding time is considered to be greater than the entire simulation time.

The average folding time and the overall native-state population of a set of simulations are provided to complement the smoothed CαβRMSD-vs-time curve that does not display short folding events. Although the smoothed CαβRMSD-vs-time curve is informative with regard to unfolding and refolding events in a simulation, its limitation is that display of CαβRMSD-vs-time curves for a set of simulations requires too much page space. By definition, the overall native-state population is less stochastic than the average folding time. Therefore, the overall native-state population is a better descriptor of protein folding in multiple MD simulations than the CαβRMSD-vs-time curve or the average folding time.