J Mol Model (2012) 18:2785–2794
DOI 10.1007/s00894-011-1282-2

ORIGINAL PAPER

Effect of the thermostat in the molecular dynamics simulation on the folding of the model protein chignolin

Carlos A. Fuzo · Léo Degrève

Received: 17 May 2011 / Accepted: 13 October 2011 / Published online: 25 November 2011

Abstract Molecular dynamics simulations of the model protein chignolin with explicit solvent were carried out, in order to analyze the influence of the Berendsen thermostat on the evolution and folding of the peptide. The dependence of the peptide behavior on temperature was tested with the commonly employed thermostat scheme consisting of one thermostat for the protein and another for the solvent. The thermostat coupling time of the protein was increased to infinity, when the protein is not in direct contact with the thermal bath, a situation known as minimally invasive thermostat. In agreement with other works, it was observed that only in the last situation the instantaneous temperature of the model protein obeys a canonical distribution. As for the folding studies, it was shown that, in the applications of the commonly utilized thermostat schemes, the systems are trapped in local minima regions from which it has difficulty escaping. With the minimally invasive thermostat the time that the protein needs to fold was reduced by two to three times. These results show that the obstacles to the evolution of the extended peptide to the folded structure can be overcome when the temperature of the peptide is not directly controlled.

Keywords Berendsen thermostat · Chignolin · Minimally invasive thermostat · Molecular dynamics · Protein folding

C. A. Fuzo (✉) · L. Degrève
Grupo de Simulação Molecular, Departamento de Química,
Faculdade de Filosofia Ciências e Letras de Ribeirão Preto,
Universidade de São Paulo,
Av. Bandeirantes 3900,
14040-901, Ribeirão Preto, SP, Brazil
e-mail: cafuzo@usp.br

Introduction

Molecular dynamics (MD) simulations have been widely employed as an attempt to understand biological systems, especially the dynamics of peptides and proteins [1–3]. MD studies involving proteins and peptides have usually aimed at better understanding of experimental observations. For this reason, a proper treatment of the thermodynamic conditions used in the computer experiments is necessary. One of the most important thermodynamic properties concerning the control of equilibrium conditions is temperature. Moreover, because many properties of proteins and peptides are strictly and sensitively dependent on temperature, it is of utmost importance to use an appropriate temperature control during a MD study, so that the thermodynamic equilibrium conditions can be met.

In MD simulations, the temperature is commonly controlled by algorithms known as thermostats [4–8]. In these algorithms, the system's molecules, or subsets of molecules, are in contact with an infinite external thermal bath at the reference temperature, thus allowing kinetic energy exchange. The most frequently used thermostat algorithms are based on the rescaling of the atomic velocities, and the Berendsen thermostat [8] is the most widely employed for that purpose. The advantage of the Berendsen thermostat is that the system is weakly coupled to the external thermal bath [8], which leads to its slow relaxation to the reference temperature. However, the Berendsen thermostat has the shortcoming of not generating a canonical temperature distribution [9].

MD simulations involving solutes such as peptides and proteins are generally conducted with molecules embedded into a solvent. Utilization of a single thermostat for temperature control in these solute-solvent systems leads

to differences between the solvent and solute temperatures, an effect called "hot solvent – cold solute problem" [10]. This problem is normally overcome by employing one thermostat for the proteins and another for the solvent [11]. However, it has recently been [12] reported that the small differences observed between the solute and solvent temperatures upon application of a sole thermostat are due to the way the temperature is calculated and a large difference in temperature indicates a problem in the methodology. The application of a single thermostat coupled to the solvent has been proposed as remediation for these drawbacks [12, 13]. This strategy, denominated "minimally invasive thermostat" [13], was proposed and tested in a study on polyalanine octapeptide embedded in water medium, and it was shown that it leads to a canonical distribution of the instantaneous peptide temperature, thereby showing that the explicit solvent environment is an ideal thermostat for the solute.

Among the great diversity of biological processes studied at the atomic level, understanding protein folding is considered to be one of the most important and challenging problems [14, 15]. For this reason, mini-proteins such as the chignolin peptide [16] have been designed for use as folding models and have been a useful tool for folding studies involving long MD simulations [17–19]. Chignolin is a good folding model because it is constituted by only 10 residues (GYDPETGTWG) and because its tertiary structure folded in water is unique [16].

In the present study, the influence of the Berendsen thermostat on the MD of the folding of the short peptide chignolin is examined. To this end, the behavior of the instantaneous temperature of the native structure was investigated using different coupling time constants for the thermostat, starting with a widely employed value of 0.1 ps, and increasing it until the minimally invasive thermostat condition. Then, folding MD studies were conducted starting from the extended structure of chignolin using the common thermostat scheme and the minimally invasive thermostat, so as to gain insight into the behavior of the conformation of the peptide as a function of time.

Models and methods

Temperature coupling using the Berendsen thermostat [8] is accomplished by scaling the velocities of the atoms at each time step with the following scaling factor:

$$\lambda = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T} - 1\right)\right]^{\frac{1}{2}}, \quad (1)$$

where Δt is the time step, τ is the coupling time constant, T_0 is the temperature of the thermal bath, and T is the instantaneous temperature.

Studies on a set of six simulations were first performed by increasing the coupling time constant for the peptide (τ_P) employed in the Berendsen thermostat, while the coupling time constant for the solvent molecules (τ_S) was maintained at 0.1 ps. These simulations, described as N_1 *to* N_6 in Table 1, were carried out with the native structure of chignolin. This set of calculations was used to assess the behavior of the instantaneous temperature of the peptide as a consequence of the variations in τ_P from 0.1 ps to $\tau_P \rightarrow \infty$. In the latter condition, the minimally invasive thermostat is achieved, so the peptide is not in contact with the thermostat and its temperature is not regulated by the thermal bath.

A second set of simulations was conducted beginning with the chignolin peptide in an extended structure (Table 1). In the first group of five experiments, designated $F_{\tau,1}$ to $F_{\tau,5}$, the temperature coupling constant τ_P was set to 0.1 ps, while in the second group, denominated $F_{\infty,1}$ to $F_{\infty,5}$, the coupling constant was $\tau_P \rightarrow \infty$. In both groups $F_{\tau,1}$ to $F_{\tau,5}$ and $F_{\infty,1}$ to $F_{\infty,5}$, different initial distributions of the atomic velocities were applied. These simulations were carried out in order to study the influence of the temperature control algorithm on the evolution of the conformation of the peptide during the folding process and, to observe the behavior of the systems by beginning with different

Table 1 Identification of the simulation conditions

Simulation	Conformation	τ_P (ps)	Time (ns)
N_1	Native	0.1	10
N_2	Native	0.5	10
N_3	Native	2.0	10
N_4	Native	5.0	10
N_5	Native	20.0	10
N_6	Native	∞	10
$F_{\tau 1}$	Extended	0.1	1000
$F_{\tau 2}$	Extended	0.1	1000
$F_{\tau 3}$	Extended	0.1	1000
$F_{\tau 4}$	Extended	0.1	1000
$F_{\tau 5}$	Extended	0.1	1000
$F_{\infty 1}$	Extended	∞	150
$F_{\infty 2}$	Extended	∞	400
$F_{\infty 3}$	Extended	∞	150
$F_{\infty 4}$	Extended	∞	150
$F_{\infty 5}$	Extended	∞	250

distributions of velocities. The initial conformations and the initial distribution of the velocities were the same for the $F_{\tau,i}$ and $F_{\infty,i}$ systems.

The structure of the chignolin peptide determined by NMR spectroscopy at 277 K, *Protein Data Bank* (PDB) code 1UAO [16], was used as initial guess in the N_1 *to* N_6 simulations, while the F simulations were initiated with the extended structure built by the PyMol package [20]. All the systems were constructed by locating the peptide at the center of a cubic water box with edges equal to 5.45 nm hydrated by 5246 water molecules. Cl^- and Na^+ counterions were inserted into electrostatically favorable positions, in order to preserve the local electroneutrality of the systems.

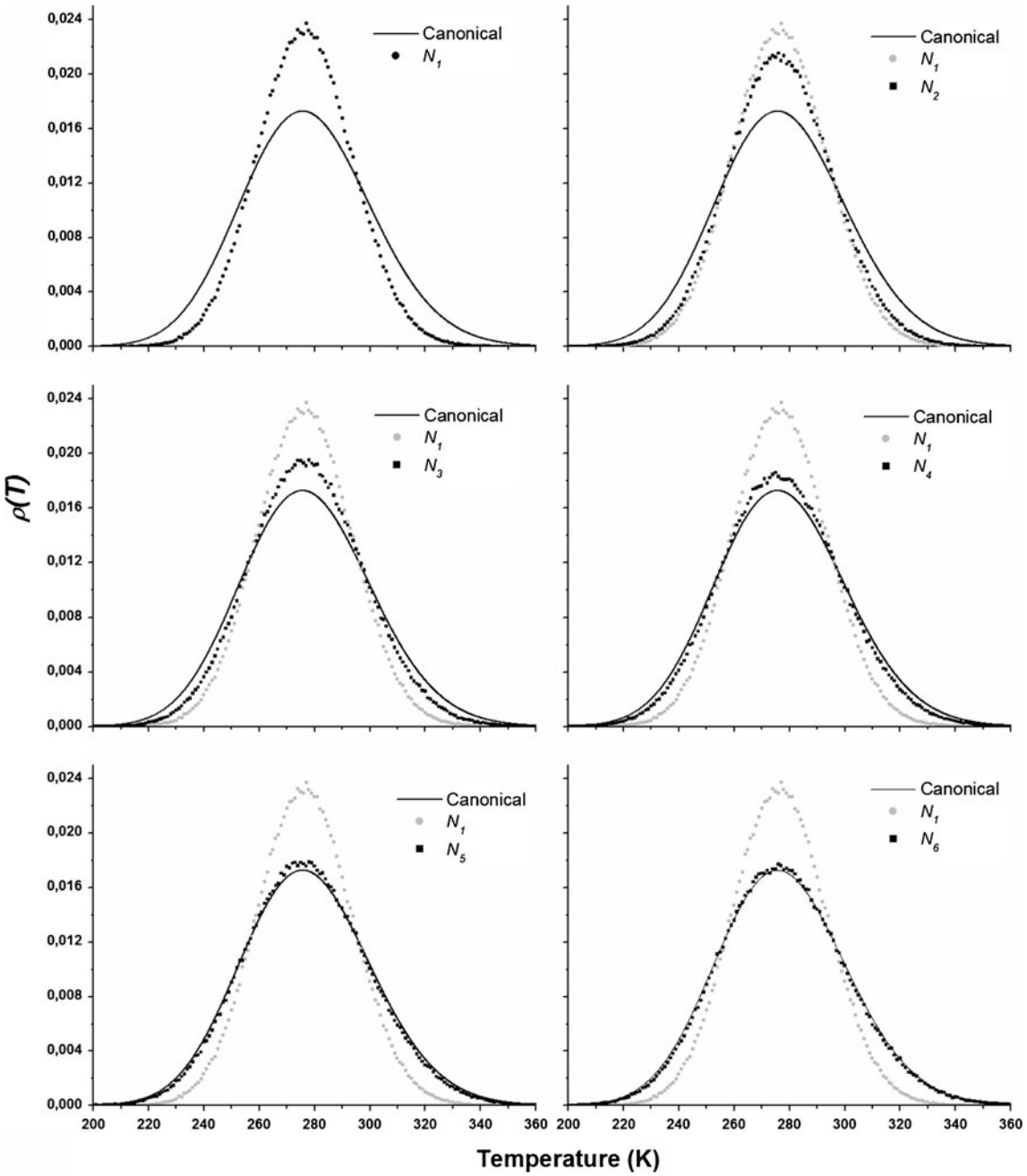

Fig. 1 Distributions of the instantaneous temperatures of chignolin during the N simulations. The canonical distribution is shown as a solid line. The data for N_1 simulation are presented in all panels for comparison

The simulations were carried out with the GROMACS 4.5.3 simulation package [1]. The GROMOS96 43A1 force field [21] was utilized for the peptide, and the SPC model was employed for water molecules [22]. The peptide covalent bonds involving hydrogen atoms were constrained by the LINCS algorithm [23], and the SETTLE algorithm [24] was applied to maintain the water molecules structurally stable. In all cases, the pressure (1 atm) was regulated by the Berendsen's algorithm [8] using a time coupling constant of 0.5 ps. The same temperature of 277 K employed for the determination of the chignolin structure [16] was used as reference temperature in the simulations. Cut-offs on the van der Waals and real space electrostatic interactions were applied at 1.0 nm with the long range electrostatic interactions calculated by the particle mesh Ewald summation method [25]. The excesses of interaction energies in these systems were eliminated by an energy minimization phase using a steepest-descent algorithm. The initial conformations used in the *N* and *F* simulations were obtained by simulating the last energy minimized systems by 1 ns with restrictions on the peptide atomic coordinates by application of a force of 1000 kJ/mol/nm^2 in each direction. Finally, the simulations were conducted without restrictions. The MD motion equations were integrated by the *leap-frog* algorithm [26] with a time step of 2 fs.

Results and discussion

The distributions of instantaneous temperatures for the peptides in the *N* simulations were compared to the canonical distribution [13]

$$\rho(T) = \frac{\left(N_{DoF}T/2T_{ref}\right)^{N_{DoF}/2}}{\Gamma(N_{DoF}/2)T} \exp\left[-\frac{N_{DoF}T}{2T}\right], \quad (2)$$

where T_{ref} is the reference temperature used in the simulations (277 K) and N_{DoF} is the number of degrees of freedom of the peptide given by the following equation [27]:

$$N_{DoF} = (3N^p - N_c^p)\frac{3N^s - N_c^s - N_{com}}{3N^s - N_c^s} \quad (3)$$

where N^p and N_c^p are the numbers of atoms and constraints of the peptide; N^s and N_c^s are the total numbers of atoms and constraints in the system and N_{com} is the number of degrees of freedom of the system center-of-mass. The N^p in the chignolin peptide modeled by the GROMOS96 43A1 force field is 101, the N_c^p of the covalent bonds with hydrogen atoms is 24, the number of water molecules included in the system is 5246 (10,492 constraints), the number of ions is 4 and N_{com}=3, so that N_{DoF} is equal to 278.97.

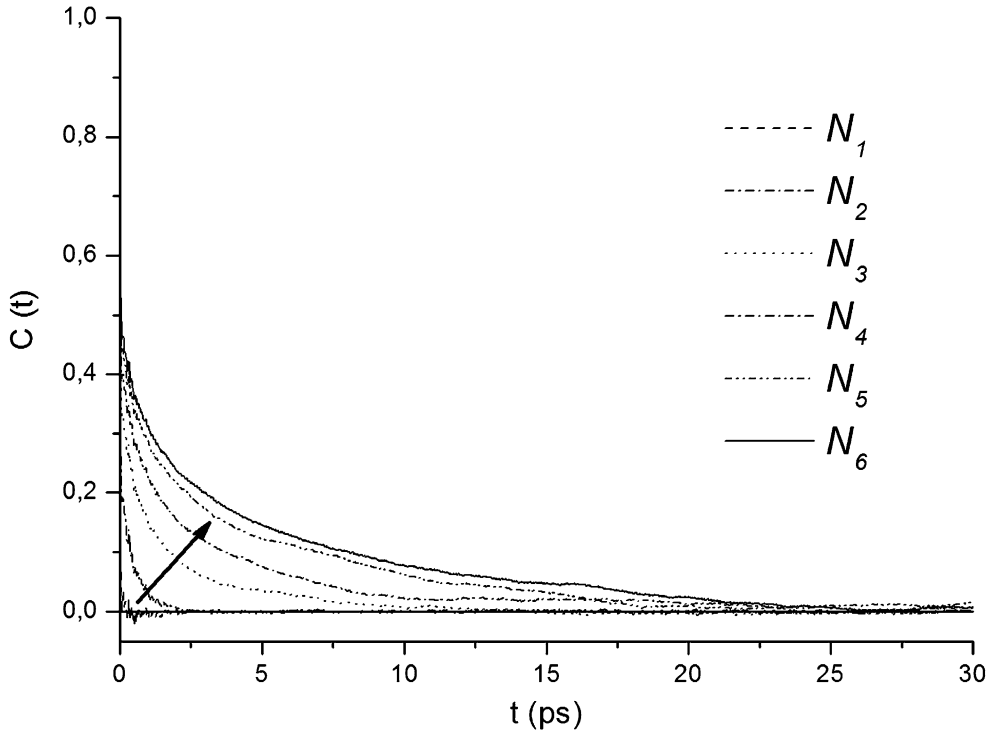

Fig. 2 Temperature autocorrelation functions obtained from *N* simulations. The arrow indicates the direction in which τ_P increases

Table 2 Correlation time, τ_{sim}, obtained from the temperature autocorrelation functions of the N simulations

Simulation	τ_{sim} (ps)
N_1	0.22
N_2	0.50
N_3	2.95
N_4	4.41
N_5	5.88
N_6	6.93

The first 1 ns of each simulation was considered as equilibration run and was excluded in the determination of the instantaneous temperature distribution. The results indicate that the distributions of the instantaneous temperatures become closer to the canonical distribution as τ_P is raised, Fig. 1. With increasing τ_P, a wider population of temperatures is visited as a direct consequence of the decrease in temperature control imposed by the thermostat. The distributions of the instantaneous temperatures confirm that the canonical distribution is attained with $\tau_P \rightarrow \infty$, as already observed in a study on a polyalanine octapeptide [13]. The fundamental point is that the canonical instantaneous temperature distribution is practically recovered when coupling of the peptide temperature with the thermal bath is removed. It is noteworthy that even if there is no direct coupling of the temperature of the peptide with the thermal bath, there still is indirect coupling through coupling of the temperature of the solvent. The result is that the solvent acts as a thermostat by heat flow with the solute.

The behavior of the instantaneous temperature as a function of time can be examined by the temperature autocorrelation functions, $C(t)$, plotted in Fig. 2, which depicts the increase in the correlation time, τ_{sim}, Table 2, as τ_P is augmented. The $C(t)$ functions indicate that the minimally invasive thermostat, namely simulation N_6, allows the peptide to visit higher and lower temperatures states for longer periods compared to what is observed when a stronger coupling to the thermal bath is applied. This feature can be noticed in the 200 ps temporal evolutions of temperatures shown in Fig. 3 for the N_1 and N_6 simulations. The longer the τ_{sim} correlation times, the longer the periods allowed for the system to assume high temperatures, *i.e.*, high kinetic energy states that enable the peptide to overcome the potential energy barriers more easily and to escape from local energy wells. This ability can lead the peptide to interchange minima of the potential energy landscape more easily. On the other hand, the short time periods the peptides

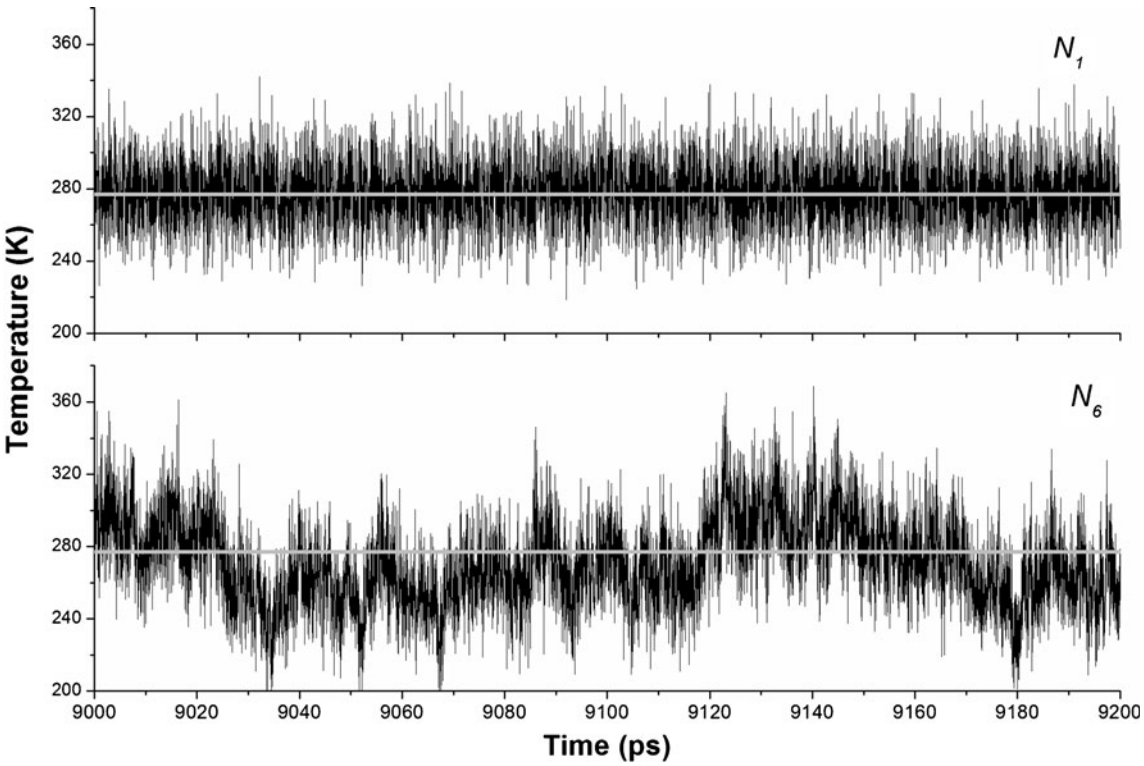

Fig. 3 Instantaneous peptide temperatures for the N_1 and N_6 simulations. The gray line indicates the temperature of 277 K

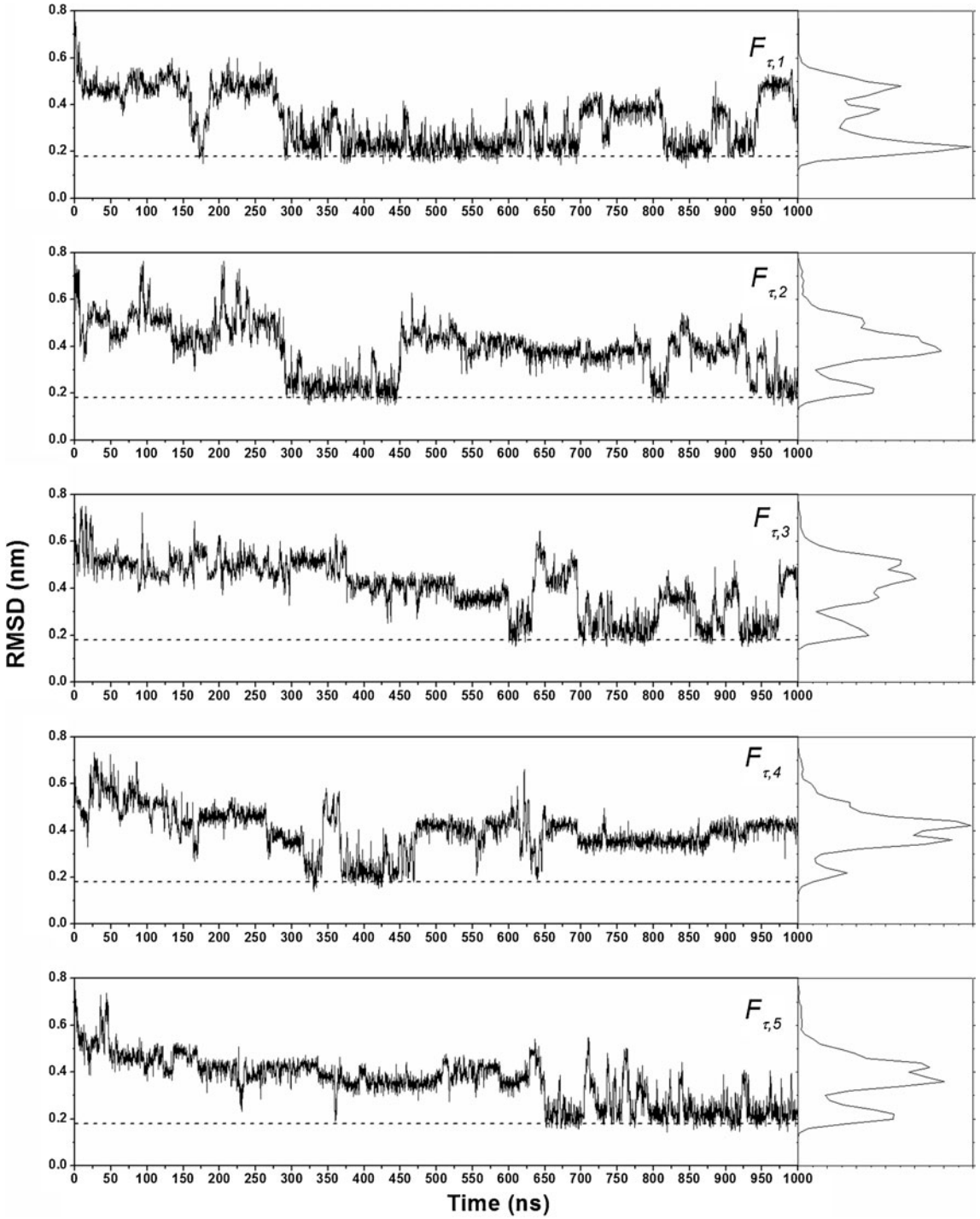

Fig. 4 All atoms' RMSDs for F_τ simulation systems and, distributions in arbitrary units. The dashed line is relative to the RMSD value of 0.18 nm, used as reference for the folding

remain in the higher temperature in simulation N_1 must reduce or abolish the interchange between the energy minima states via kinetic traps, thereby resulting in slow dynamics of peptide. For this reason, alternative methods have been proposed for the solute, such as the widely used replica exchange method [28].

In the N simulations, it is observed that the "natural" peptide temperature evolution along the simulation is

affected by the temperature control scheme. Examining *C(t)*, it has also been noted that the peptide stays at higher and low temperatures for long periods with the minimally invasive thermostat. These observations have led to the supposition in this work that greater ability for the peptide to skip from a minimum of energy to another can be reached by application of the minimally invasive thermostat simulation. These parameters can be employed for investigation of systems that are not in equilibrium, particularly the folding processes of proteins. Consequently, this paper aims to verify whether the utilization of minimally invasive thermostat can accelerate the folding of a model protein. With this objective in mind, the F_τ and F_∞ simulations will be compared in the two extreme cases of temperature coupling by applying correlation times of τ_P=0.1 ps and $\tau_P \rightarrow \infty$, respectively.

The folding success of the chignolin along the *F* simulations were evaluated by the root mean-square deviations (RMSD) calculated with reference to the first structure of the 1UAO file. The mean of all atoms RMSD between this 1UAO structure and the other 17 structures is 0.18± 0.06 nm. Therefore, it will be considered that RMSD values in this range will be a reliable criterion for identification of chignolin peptide folding in its native conformation.

All atoms' RMSDs for the F_τ are illustrated as functions of time in Fig. 4. Different RMSD profiles are observed however they fluctuate within very similar limits. The histograms of the RMSD values show that they are distributed mainly in two peaks: one around 0.4 nm which corresponding to the unfolded states, and the second around 0.2 nm involving the folded structures. The first fold events observed in F_τ simulations occur within the interval from 177 to 652 ns. This range is below the folding time of 1.0± 0.3 μs predicted by van der Spoel and Seibert from molecular dynamics studies [18]. However, only five F_τ simulations are not sufficient to predict the folding time. Superposition of the folded structures on the first PDB structure is shown in Fig. 5 together with the fit of all the first folded conformations. The RMSD between the first folded conformations observed in the F_τ simulations is equal to 0.12±0.02 nm, thereby suggesting that the folded

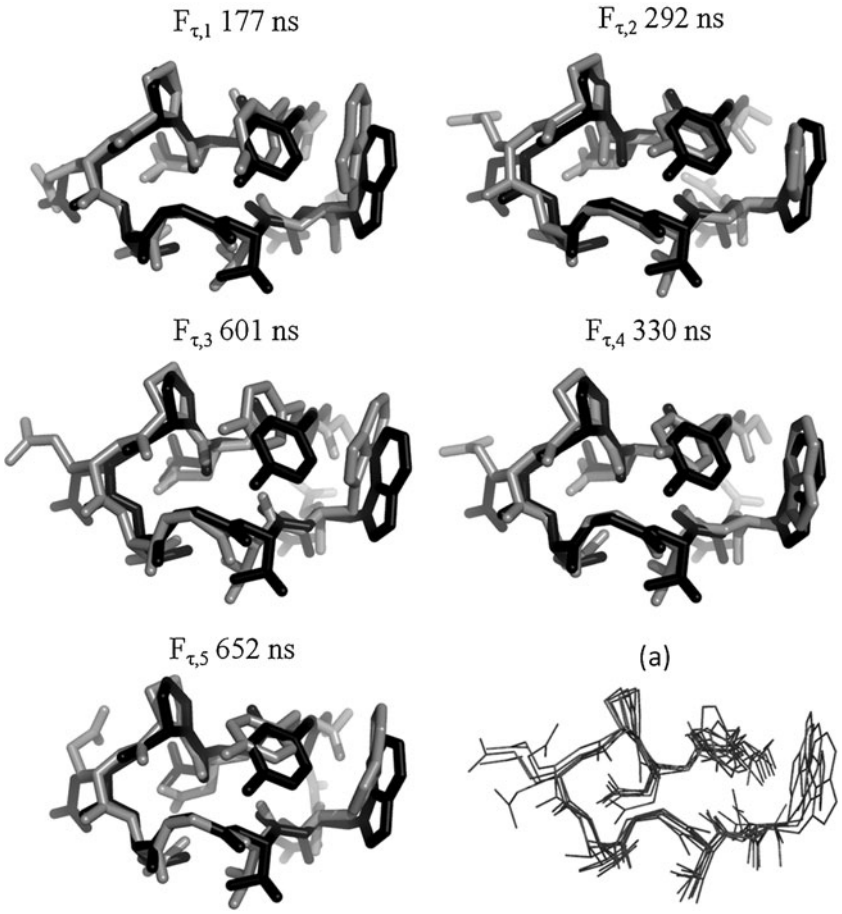

Fig. 5 First folded structures obtained in $F_{\tau,1}$ and $F_{\tau,2}$ simulations. The black colored structure is the PDB structure and the structures of the simulations are in gray. The fit of the folded structures is shown in panel (**a**)

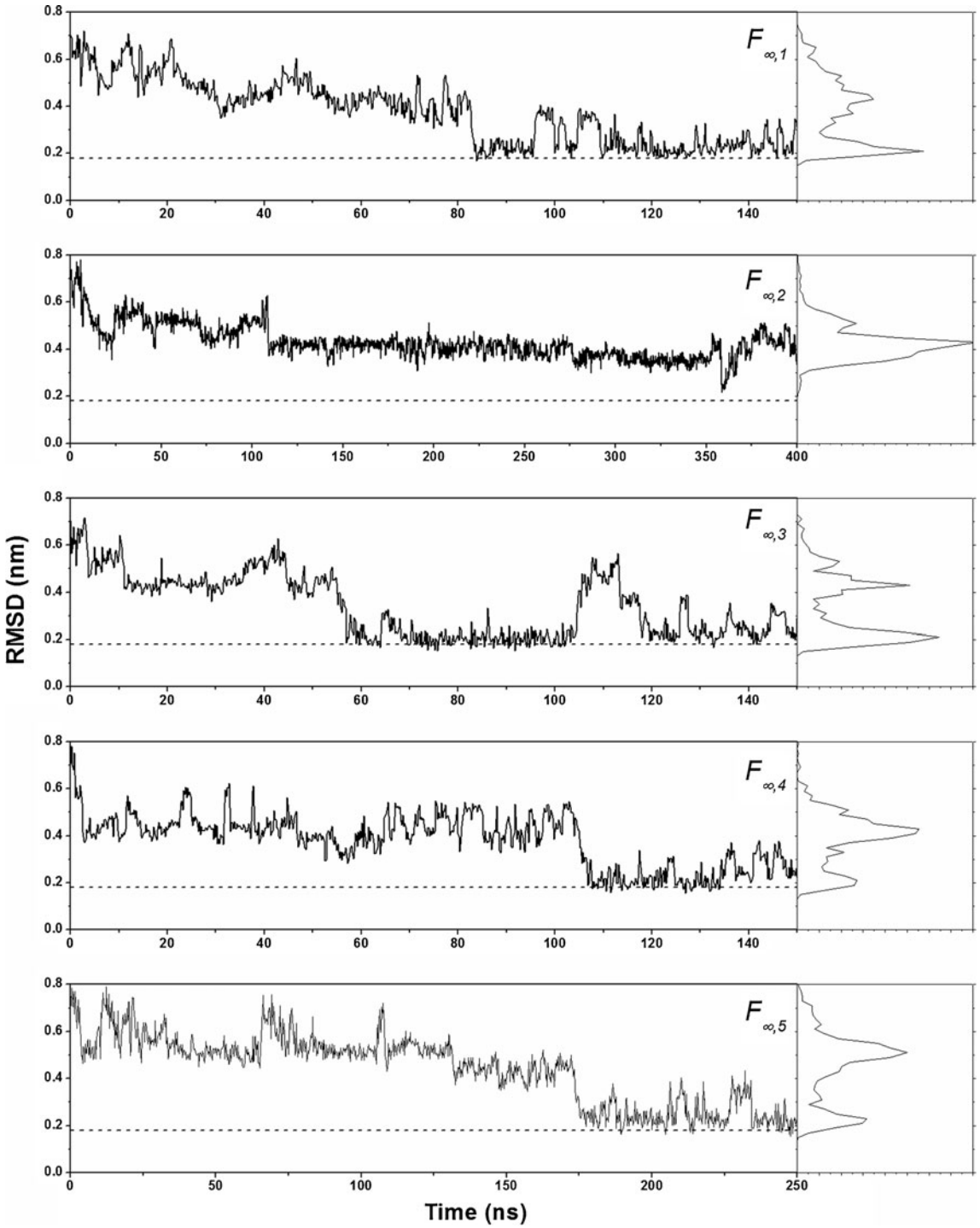

Fig. 6 All atoms' RMSDs for F_∞ simulation systems and, on the right, their distribution in arbitrary units. The dashed line is relative to the RMSD value of 0.18 nm, used as reference for the folding

conformations are equivalent. All atoms' RMSDs obtained in the F_∞ simulations, using the experimental structure as the reference, are displayed in Fig. 6. The mean value of the RMSDs of the first folded structures, presented in Fig. 7, is 0.12±0.05 nm. This value is smaller than the one observed between the NMR structures, showing that equivalent conformations are also reached in the F_∞ simulations. In these simulations, the first folded structures are observed

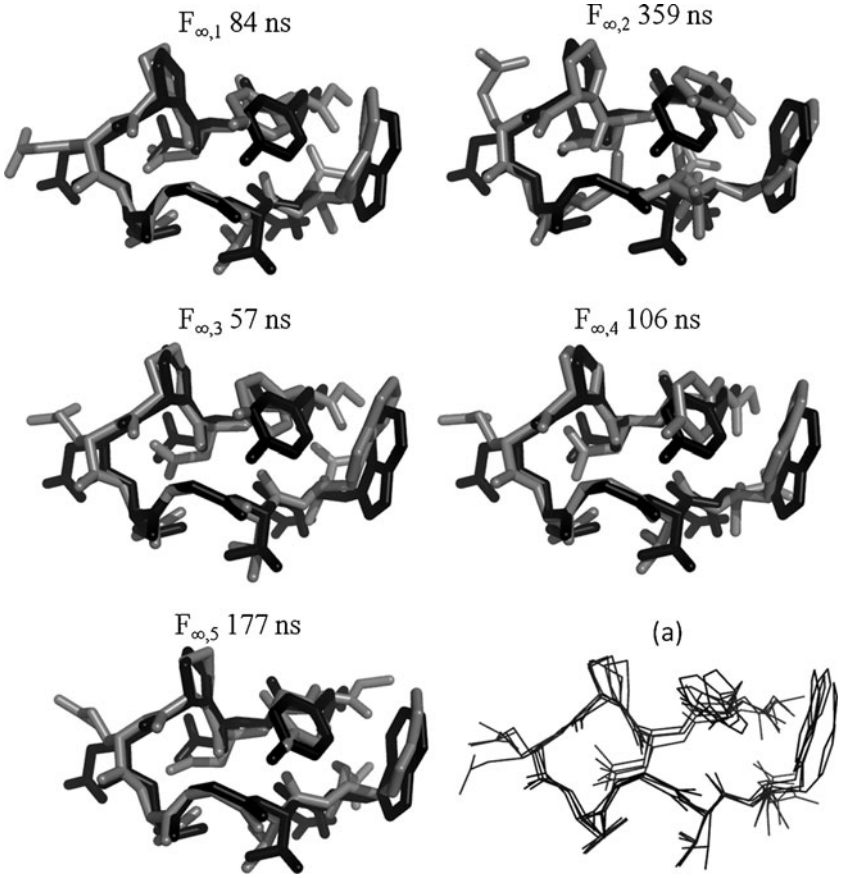

Fig. 7 First folded structures obtained in F_∞ simulations. The black colored structure is the PDB structure, and the structures of the simulations are in gray. The fit of the folded structures is shown in panel (**a**)

within 57 to 359 ns, *i.e.*, 32 to 55% of the time intervals observed in the F_τ simulations, thereby demonstrating that the utilization of minimally invasive thermostat accelerates the folding of chignolin.

Taking into account that the RMSD relative to the native structure can be understood as a coordinate in the folding free-energy landscape, where the minimal energy is relative to a RMSD equivalent to that of the native structure, the diverse RMSD behaviors detect different pathways in the folding free-energy landscape, or in the folding process. It was observed in the F_τ simulations that the systems are trapped in local minima regions with evident difficulties to escape, as seen by the long periods of small RMSD oscillations. In some cases of the F_τ simulations, it was observed that the peptide escapes, after the folding, from the native structure adopting, for long time intervals, structures with very high RMSD values. On the other hand, the behavior along the F_∞ simulations resembles to a search over the free-energy landscape in direction of lower free energies, i.e., in the direction of the native conformation. The fast recovery of the chignolin peptide native structure in the F_∞ simulations implies that the obstacles to the natural evolution from the extended peptide to its folded structure can be overcome avoiding the direct control of the peptide temperature that introduces unphysical forces.

Conclusions

The application of temperature control algorithms that allow for exchange of energy in the system under study with an infinite external thermal bath results in direct control of the distribution of kinetic energies. By examining the behavior of systems out of equilibrium to achieve lower energy states, it becomes clear that restrictions on the kinetic energies should result in slower kinetics than the one that would be observed without the application of restrictions. This idea was tested and applied to the problem of protein folding. The application of the minimally invasive thermostat employed in the molecular dynamics study of the folding of the protein model, the chinoglin, proved to be highly productive, since the time that the protein needs to fold was reduced by comparing with the

folding time when strict restrictions on the distributions of kinetic energy are applied. The present results open new possibilities for studies on protein folding with all its implications, and for the design of simulation methods that would lead to effective observation of folding.

Acknowledgments The authors are thankful to the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), and to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial supports.

References

1. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718
2. Sanbonmatsu KY, Tung CS (2007) High performance computing in biology: multimillion atom simulations of nanoscale systems. J Struct Biol 157:470–480
3. Klein ML, Shinoda W (2008) Large-scale molecular dynamics simulations of self-assembling systems. Science 321:798–800
4. Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. J ChemPhys 72:2384–2393
5. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 52:255–268
6. Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. Phys Rev A 31:1695–1697
7. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101
8. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
9. Morishita TJ (2000) Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath. Chem Phys 113:2976–2982
10. Cheng A, Merz KM (1996) Application of the nosé−hoover chain algorithm to the study of protein dynamics. J Phys Chem 100:1927–1937
11. Hünenberger PH (2005) Thermostat algorithms for molecular dynamics simulations. Adv Polym Sci 173:105–149
12. Eastwood MP, Stafford KA, Lippert RA, Jensen MO, Maragakis P, Predescu C, Dror RO, Shaw DE (2010) Equipartiotion and the calculation of temperature in biomolecular simulations. J Chem Theor Comput 6:2045–2058
13. Lingenheil M, Denschlag R, Reichold R, Tavan P (2008) The "hot-solvent/cold-solute" problem revisited. J Chem Theor Comput 4:1293–1306
14. Kennedy D, Norman C (2005) So much more to know. Science 309:78–102
15. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys 37:289–316
16. Honda S, Yamasaki K, Sawada Y, Morii H (2004) 10 Residue folded peptide designed by segment statistics. Structure 12:1507–1518
17. Seibert MM, Patriksson A, Hess B, van der Spoel D (2005) Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. J Mol Biol 354:173-183
18. van der Spoel E, Seibert MM (2006) Protein folding kinetics and thermodynamics from atomistic simulations. Phys Rev Lett 96:238102
19. Suenaga A, Narumi T, Futatsugi N, Yanai R, Ohno Y, Okimoto N, Taiji M (2007) Folding dynamics of 10-residue β-hairpin peptide chignolin. Chem Asian J 2:591–598
20. DeLano WL (2002) The PyMOL Molecular Graphics System (http://www.pymol.org)
21. Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, kruger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Hochschul-verlag AG an der ETH Zürich, Zürich
22. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Intermolecular Forces (ed) Interactions models for water in relation to protein hydration. Reidel, Dordrecht, the Netherlands, pp 331–342
23. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
24. Miyamoto S, Kollman PA (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models. J Comput Chem 13:952–962
25. Darden T, York D, Pedersen L (1993) Particle mesh ewald: an N-log(N) method for ewald sums in large systems. J Chem Phys 98:10089–10092
26. Hockney RW, Goel SP, Eastwood J (1974) Quiet highresolution computer models of plasma. J Comput Phys 14:148–158
27. Van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC (2005) Gromacs User Manual version 4.5, www.gromacs.org
28. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151