

Probing the energy landscape of alanine dipeptide and decalanine using temperature as a tunable parameter in molecular dynamics

A Chatterjee² and S Bhattacharya¹

¹Department of Physics, Indian Institute of Technology Guwahati, Guwahati, Assam, India 781039

²Department of Chemical Engineering, Indian Institute of Technology, Bombay, Maharashtra.

Abstract. We perform several molecular dynamics (MD) calculations of solvated alanine dipeptide and decalanine in vacuum with temperature as a tunable parameter and in the process, generate Markov state models (MSMs) at each temperature. An interesting observation that the kinetic rates appear to obey the Arrhenius rate law allows us to predict the dynamics of alanine dipeptide at 300 K at the microsecond timescales using the nanoseconds long high temperature calculations without actually performing MD simulations at 300 K. We conclude that the energy landscape of alanine dipeptide contains superbasins deeper than $k_B T$ and determine the energy barriers associated with the moves from the Arrhenius rate expression. Similar insights regarding the energy landscape associated with folding/unfolding pathways of a deca-alanine molecule are obtained using kinetic rates calculated at different temperatures.

1. Introduction

A protein molecule can exist in different configurations that determine its biological activity. Consequently, significant efforts have been made over the last three decades to understand the stable configurations, free energy differences between the stable configurations, and the mechanisms for transformations from one configuration to another[1–6]. In recent years, a large number of studies have focussed on the application of Markov State Models to describe the dynamics of solvated biomolecular systems. An MSM enables one to study the dynamics of protein systems in terms of state-to-state transitions thereby providing a systematic approach for understanding the nature and timescales of evolution of the system. Pande and co-workers have pioneered the development of an automated kinetic clustering procedure for building of MSMs by post-analyzing long room temperature MD trajectories[5]. Generating an accurate MSM requires extensive MD calculations especially when the system remains trapped in certain states for extended periods of time, a situation that is frequently encountered when large free energy barriers are present. Several accelerated MD methods have been proposed to overcome the issue of long-lived metastable states leading to poor sampling[7–12].

Our interest lies in methods employing high temperature simulations to overcome energy barriers. Although high temperature simulations are often used for seeding standard simulations and REMD has been widely used for enhanced conformational sampling, very few studies have directly applied high temperature simulations to generate MSMs of biomolecular systems[13]. Establishing the connection between high temperature and low temperature dynamics opens up the possibility of applying



temperature accelerated schemes[14,15] to enhance study of rare events in proteins and other biomolecular systems while recovering the correct dynamics at room temperature. We show that high temperature MD calculations can efficiently provide information about the states and the kinetic rates of the solvated alanine dipeptide with a speed-up of almost 500 times over room-temperature MD calculations[16]. In this process, additional insights about the energy landscape are obtained compared to single temperature MD calculations of similar duration.

We employ a swarm of independent MD calculations in parallel at high temperatures to accelerate conformational sampling. The states are determined *on-the-fly* in terms of energy-minimized configurations of the biomolecule using a distance-metric. Once the states have been identified at a high temperature, we perform shorter MD simulations starting from selected states at different temperatures to obtain the rate constants associated with specific moves originating from these states. A consistent picture of the states and kinetic pathways is obtained across different temperatures. A similar method is applied to the decaalanine system in a force spectroscopy setup in vacuum to obtain the energy barriers along a folding/unfolding path.

2. Methodology

2.1 System Preparation

A single molecule of blocked N-acetyl-N'-methyl-L-alanylamide (or alanine dipeptide), was immersed in a rectangular box of 390 pre-equilibrated TIP3P water molecules bringing the total system size to 1192 atoms. The final system was in the form of a cubical cell with dimensions $2.3 \times 2.3 \times 2.3 \text{ nm}^3$. The CHARMM27 force-field[17] was used. We constructed MSMs for the capped deca-alanine (Ala_{10}) with acetylated N-terminus and amidated C-terminus in vacuum, stretched under a force-spectroscopy setup. The C_α atoms at the two ends (residues 1 and 10) are tethered to two anchor points by harmonic restraints with a spring constant of $0.86 \text{ kcal/mol/\AA}^2$. The initial configuration was obtained from the 104-atom helical model of Ref.[18].

2.2 Simulation Protocols

Prior to the main simulation, the system was minimized using the conjugate gradient method for 1000 steps following which it was equilibrated for 3 ns at 300 K and constant pressure of 1 atm. Subsequent simulations were performed at constant volume and temperature (canonical ensemble) with NAMD[19] with periodic boundary conditions and particle mesh Ewald electrostatics[20] in the canonical ensemble using Langevin thermostat[21] at 600, 450 and 750 K. An integration time-step of 2 fs was used with RATTLE[22] and SETTLE[23] algorithms applied to covalent bonds involving hydrogen in the peptide and water respectively. A swarm of independent MD trajectories are generated that share the state information. The state of the system in each trajectory is monitored on-the-fly. State-matching is performed based on the internal coordinates of the molecule with a tolerance of 1.5 \AA following alignment using the Kabsch algorithm. A match is said to occur when respective heavy (non-hydrogen) atoms of the alanine dipeptide molecule are within the tolerance value after the molecule has been aligned to one of the energy-minimized structures via rotation and translation operations. The MD calculation is stopped once the system moves to a new state. The waiting time for each escape is recorded and a new calculation is started from the new state using the same procedure.

3. Results

3.1 Kinetics of alanine dipeptide

A total of 16 states of alanine dipeptide were discovered using the state-identification procedure described above. Of these, five major states (labelled A-E) were identified that accounted for >99% occupation probability of the molecule in the range of temperatures studied. States A and C are associated with the α_R conformation of the molecule. States B and D are located in the region commonly associated with the $\beta/\text{PII}/\text{C7}_{\text{eq}}$ structures of the system. State E can be identified with the C7_{ax} conformation. The rate constants were obtained using the maximum likelihood estimate (MLE)[24] *i.e.*,

$k_\alpha = n/t_{MD}$ where n denotes the number of times a move has been observed and t_{MD} is the total MD time elapsed for this move. The Arrhenius plot for the move A-B is shown in Figure 1(b). The rate constants for the move from state A to B from MLE are given by 0.045, 0.102, 0.150 and 0.188 ps^{-1} at 300, 450, 600 and 750 K when plotted against inverse temperature, fall on a straight line in semi-logarithmic coordinates demonstrating the validity of the Arrhenius expression $k_\alpha = \nu_\alpha \exp(-E_\alpha / k_B T)$. The pre-exponential factor and activation barrier for the A-B move were estimated to be 0.53 ps^{-1} and 0.064 eV. An interesting observation is that the pre-exponential factor is not temperature dependent for the range of temperatures studied. The fact that the same pre-exponential factor and activation barrier can be used at different temperatures indicates that the move between two states of alanine dipeptide proceeds via a single mechanism.

The rates estimated at 300 K for the moves from state A using the Arrhenius expression were compared to the ones obtained directly from regular MD using MLE at 300 K to validate our procedure. The estimated rates for the move from state A to state B, C and D are 0.048, 0.0014, 0.0011 ps^{-1} , respectively, while the rate calculated using MLE were 0.045, 0.0015, 0.0011 ps^{-1} , respectively. Although a direct comparison of the kinetic rates with previous studies is not possible due to differences in the system

setup and force fields used, we found that the kinetic rates predicted at 300 K from the high temperature simulations are consistent with results from experiments as well as previous simulation studies. The rate for moves from A to B (alpha-helix to beta strand transitions) is estimated to be 0.045 ps^{-1} corresponding to a mean escape time of ~ 22 ps which is comparable to previous studies[25].

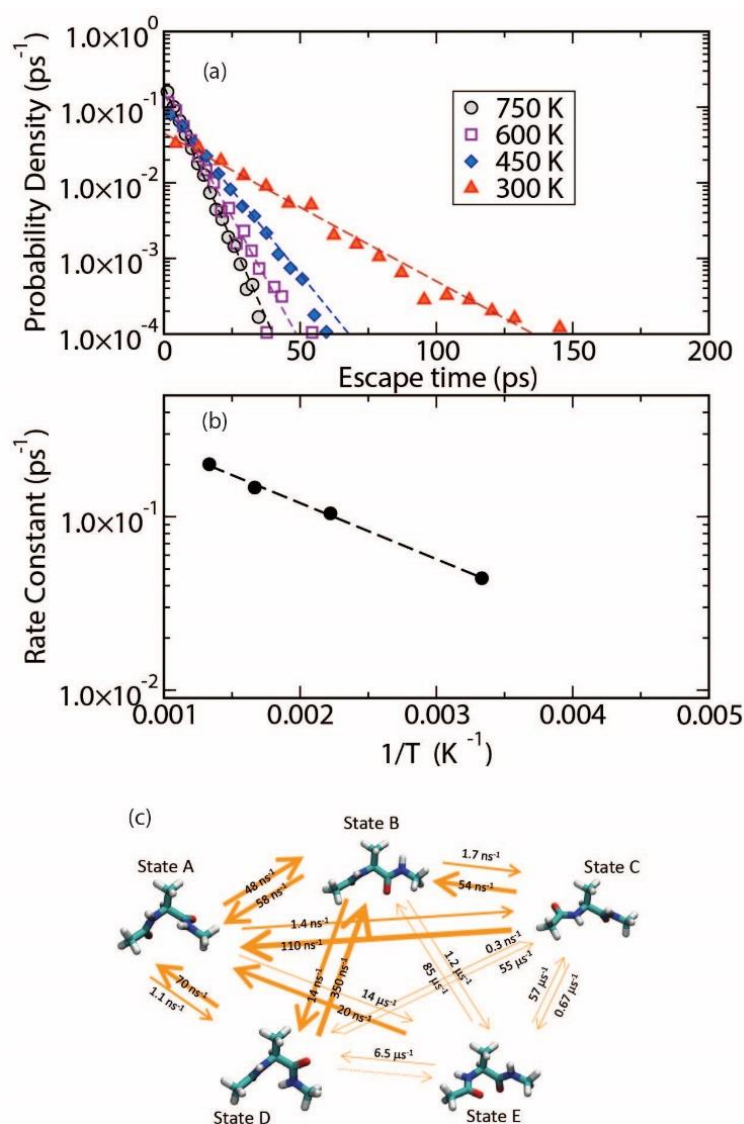


Fig. 1. (a) Probability density for escape (or waiting) times distributions obtained for the move from state A to B at 750, 600, 450 and 300 K. Dashed lines denotes the exponential distribution of escape times using the rate constant calculated by maximum likelihood estimation. (b) Arrhenius plot showing the variation of the rate constants with temperature for the move A-B. The rate constants fall on the dashed line showing the Arrhenius behaviour. (c) Markov State Model (MSM) constructed for solvated alanine dipeptide. The rate constants at 300 K are the obtained from high temperature simulations.

The energy barriers and pre-exponential factors estimated using the Arrhenius expression showed the pre-factors to be temperature independent indicating that each move proceeds via a single mechanism. The largest energy barrier was 8.3 kcal/mol for the move C-E, while the move from C-A had a negligible barrier of 0.054 kcal/mol. The size of the observed energy

barriers can be explained by making comparisons to the corresponding barriers for alanine dipeptide in vacuum from previous studies[26,27]. The barriers for the moves *B-E*, *E-B*, and *E-D* are 6.55, 5.7 and 5 kcal/mol, respectively, in the solvated system. The corresponding barriers in vacuum[27] are less than 7, 5 and 5 kcal/mol, respectively, i.e., they are similar to the ones for solvated alanine dipeptide. The location of the states B, D and E in the ϕ - ψ map are slightly altered due to the presence of solvent molecules. The large activation barriers point to the depth of the superbins being significantly more than $k_B T$. The barriers involved in these moves are energetic in nature. The origin of these barriers can be attributed to the large energy changes associated with the backbone torsional changes in alanine dipeptide that are also witnessed in vacuum. The presence of solvent molecules introduces ripples in the potential energy landscape

3.2 Kinetics of Decaalanine

We consider a single deca-alanine molecule that is stretched at two ends as shown in Fig. 2a. The C_α atoms (residues 1 and 10) are tethered to two anchor points by harmonic restraints (spring constant of $0.86 \text{ kcal/mol/\AA}^2$). The anchors are separated by a distance of 24 \AA . The result of the applied force is that the α -helical structure (Fig. 2a, state 1) is no longer the preferred configuration, unlike the behaviour observed in the absence of an applied force. Instead the molecule can stretch dramatically as shown in Fig.2a (state 17).

State-constrained MD calculations were performed at 300, 400 and 500 K starting from the α -helical structure to reveal other states of the system using a procedure analogous to the one for alanine dipeptide. The pathway involving

transitions between states 17-6-2-1 can be regarded as the folding pathway. As evident from Table 1, the α -helicity decreases while 3_{10} -helicity increases along this pathway. In addition, the α -helical structure can form a deformed helical state denoted as state 3, which in fact is the dominant state at anchor separation of 24 \AA . Unfolding of the deca-alanine entails breaking of the hydrogen bonds, so the question arises how large is the energy change.

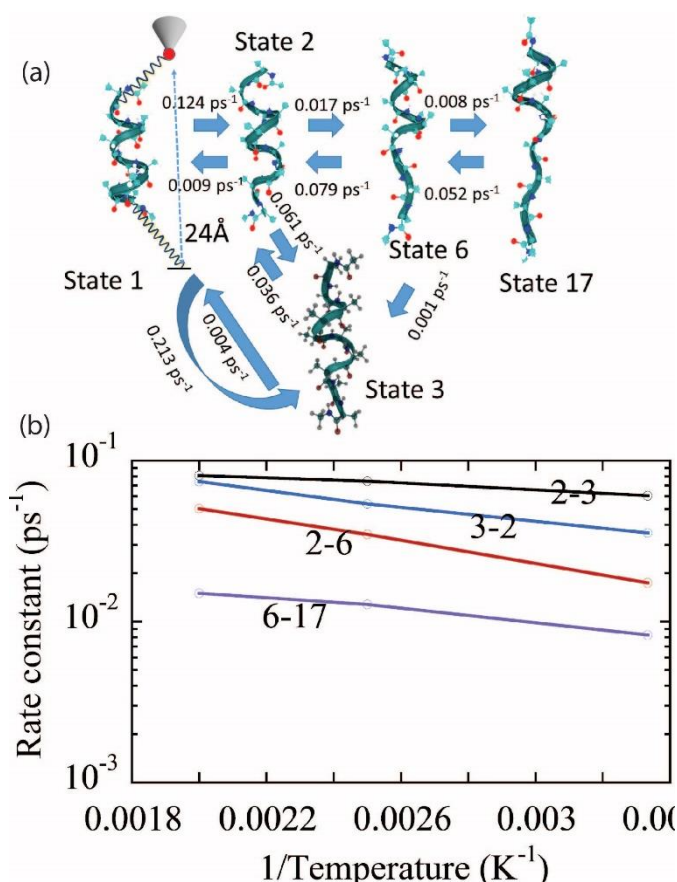


Figure 2. (a) Snapshots showing the states 1,2,3,6 and 17 along the chosen folding/unfolding pathway with the kinetic rates calculated at 300 K with anchor separation 24 \AA . (b) The Arrhenius plot showing the variation of the kinetic rates of conversion between the states with inverse temperature in semi-logarithmic scale.

We answer this question by analyzing the temperature dependent kinetics from states 2, 3 and 6 (see Fig. 2). The correlation coefficient between the data and the Arrhenius fit exceeds 0.96. This permits us to calculate the Arrhenius parameters. Surprisingly, the activation barriers involved in the pathways are small. For instance, the barrier for the move from state 2-1, 2-6

and 6-17 are found to be 0.01, 0.069 and 0.039 eV, respectively. These small barriers indicate the folding or unfolding proceeds as the molecules slithers into a compact or elongated form. Minor rearrangements in the process could explain the small barriers. This conclusion is also supported by the α - and 3_{10} -helicity shown in Table 1.

Table 1. Average α -helicity and 3_{10} -helicity of states along a folding pathway.

State	α -Helicity	3_{10} -Helicity
1	0.801	0.08
2	0.201	0.21
6	0.021	0.19
17	0.013	0.15

4. Discussions

We use molecular dynamics (MD) temperature as a probe to explore the overall features of the potential energy landscape of the solvated alanine dipeptide system and the decaalanine system in a force spectroscopy setup in vacuum. In alanine dipeptide, the moves between the states are found to be first order processes and the rate constants were found to obey the Arrhenius expression. Some of the activation barriers extracted from the Arrhenius fits were found to be significantly larger than $k_B T$ showing that the landscape can be characterized in terms of deep energy superbasins. The method was applied to the decaalanine system to obtain the energy barriers along a folding/unfolding pathway in a force spectroscopy setup. The surprisingly small barriers along the pathway chosen in the study shows that the molecule unfolds/folds by small rearrangements, i.e., wriggles into compact or elongated form.

Acknowledgments

SB acknowledges support from the Department of Biotechnology Grant No. BT/PR8311/BID/7/451/2013.

References:

- [1] Gnanakaran S, Nymeyer H and Portman J 2003 Peptide folding simulations Kevin Y Sanbonmatsu and Angel E Garcı *Curr. Opin. Struct. Biol.* **13** 168–74
- [2] Dodson G G, Lane D P and Verma C S 2008 new windows on mechanisms in biology *EMBO Rep.* **9** 144–50
- [3] Oliveira C A F De, Hamelberg D and Mccammon J A 2007 Estimating kinetic rates from accelerated molecular dynamics simulations : Alanine dipeptide in explicit solvent as a case study Estimating kinetic rates from accelerated molecular dynamics simulations : Alanine dipeptide in explicit solvent as a case stu *J. Chem. Phys.* **127** 175105
- [4] Hamelberg D, Shen T and Mccammon J A 2005 Relating kinetic rates and local energetic roughness by accelerated molecular- dynamics simulations Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations *J. Chem. Phys.* **122** 241103
- [5] Bowman G R, Huang X and Pande V S 2009 Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49** 197–201
- [6] Freddolino P L, Harrison C B, Liu Y and Schulten K 2010 Challenges in protein-folding simulations *Nat. Phys.* **6** 751–8
- [7] de Oliveira C A F, Hamelberg D and McCammon J A 2007 Estimating kinetic rates from

- accelerated molecular dynamics simulations: alanine dipeptide in explicit solvent as a case study. *J. Chem. Phys.* **127** 175105
- [8] Wang Y, Harrison C B, Schulten K and McCammon J A 2011 Implementation of Accelerated Molecular Dynamics in NAMD. *Comput. Sci. Discov.* **4** 015002
- [9] Voter A F 1997 Hyperdynamics : Accelerated Molecular Dynamics of Infrequent Events *Phys. Rev. Lett.* **78** 3908
- [10] Laio A and Gervasio F L 2008 Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science *Reports Prog. Phys.* **71** 126601
- [11] Tribello G A, Ceriotti M and Parrinello M 2010 A self-learning algorithm for biased molecular dynamics *Proc. Natl. Acad. Sci.* **107** 17509–14
- [12] Sugita Y and Okamoto Y 1999 Replica exchange molecular dynamics method for protein folding simulation. *Chem. Phys. Lett.* **314** 141–51
- [13] Buchete N V and Hummer G 2008 Peptide folding kinetics from replica exchange molecular dynamics *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **77** 1–4
- [14] Sorensen M R and Voter A F 2000 Temperature-accelerated dynamics for simulation of infrequent events *J. Chem. Phys.* **112** 9599
- [15] Divi S and Chatterjee A 2014 Accelerating rare events while overcoming the low-barrier problem using a temperature program. *J. Chem. Phys.* **140** 184115
- [16] Chatterjee A and Bhattacharya S 2015 Uncertainty in a Markov state model with missing states and rates: Application to a room temperature kinetic model obtained using high temperature molecular dynamics. *J. Chem. Phys.* **143** 114109
- [17] Mackerell A D, Bashford D, Bellott M, Dunbrack R L, Evanseck J D, Field M J, Fischer S, Gao J, Guo H, Ha S, Kuchnir L, Kuczera K, Lau F T K, Mattos C, Michnick S, Ngo T, Nguyen D T, Prodhom B, Reiher W E, Roux B, Schlenkrich M, Smith J C, Stote R, Straub J, Watanabe M, Wio J, Yin D and Karplus M 1998 All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins † *J. Phys. Chem. B* **5647** 3586–616
- [18] Park S, Khalili-Araghi F, Tajkhorshid E and Schulten K 2003 Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality *J. Chem. Phys.* **119** 3559–66
- [19] Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D and Kalé L 2005 Scalable Molecular Dynamics with NAMD *J. Comput. Chem.* **26** 1781–802
- [20] Darden T, York D and Pedersen L 1993 Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems *J. Chem. Phys.* **98** 10089
- [21] Allen P and Tildesley D J 1991 *Computer Simulation of Liquids* (Oxford University Press)
- [22] Andersen C 1983 Rattle : A “ Velocity ” Molecular Version of the Shake Dynamics Calculations for *J. Comput. Phys.* **52** 24–34
- [23] Miyamoto S and Kollman P a. 1992 Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models *J. Comput. Chem.* **13** 952–62
- [24] Ensing D E and Pande V S 2009 Bayesian Single-Exponential Kinetics in Single-Molecule Experiments and Simulations *J. Phys. Chem. B* **113** 12410–23
- [25] Chekmarev D S, Ishida T and Levy R M 2004 Long-Time Conformational Transitions of Alanine Dipeptide in Aqueous Solution : Continuous and Discrete-State Kinetic Models *J. Phys. Chem. B* **108** 19487–95
- [26] Tobias D J and Brooks III C L 1992 Conformational Equilibrium in the Alanine Dipeptide in the Gas Phase and Aqueous Solution: A Comparison of Theoretical Results *J. Phys. Chem.* **96** 3864–70
- [27] Ren W, Vanden-Eijnden E, Maragakis P and E W 2005 Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.* **123** 134109