Monte Carlo simulations of the HP model (the "Ising model" of protein folding)

Ying Wai Li [a,*], Thomas Wüst [b], David P. Landau [a]

[a] *Center for Simulational Physics, The University of Georgia, Athens, GA 30602, USA*
[b] *Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland*

ARTICLE INFO	ABSTRACT
	Using Wang–Landau sampling with suitable Monte Carlo trial moves (pull moves and bond-rebridging moves combined) we have determined the density of states and thermodynamic properties for a short sequence of the HP protein model. For free chains these proteins are known to first undergo a collapse "transition" to a globule state followed by a second "transition" into a native state. When placed in the proximity of an attractive surface, there is a competition between surface adsorption and folding that leads to an intriguing sequence of "transitions". These transitions depend upon the relative interaction strengths and are largely inaccessible to "standard" Monte Carlo methods. © 2011 Elsevier B.V. All rights reserved.

1. Introduction

Over the last few decades, considerable effort has been devoted to the protein folding problem and yet major challenges remain. The complexity arises from enormous combinatorics of the 20 amino acids, forming multitudinous possible protein sequences to which the structure is closely related, and hence its biological function [1]. Complicated interactions among residues also result in rough free energy landscapes [2]. Nevertheless, the underlying principles for protein structure prediction from the sequence are still unclear [3]. On the other hand, technical limitations also hinder experimental progress in obtaining protein structures using X-ray diffraction or NMR. Computer simulation can shed light on the problem where both theoretical and experimental studies are facing bottlenecks. Similar to the importance of the Ising model in the study of magnetism, simple protein models that embrace merely essential features are amenable to attack by a wide range of numerical methods, including computer simulations. This also allows understanding the folding problem from a macroscopic perspective without distraction from unnecessary details [4].

2. Model and methods

The HP model is a simple, prototypical lattice protein model. It consists of only two types of monomers, hydrophobic (H) and polar (P), in a sequence chosen to mimic a real protein [5]. Interactions are restricted to an attractive coupling, ϵ_{HH}, between non-bonded

hydrophobic monomers occupying nearest-neighbor sites. These are introduced to capture the hydrophobic effect, which is considered as the "driving force" of protein folding in forming tertiary structures. Despite its simplicity, the HP model shows unexpectedly rich thermodynamic behaviors and is challenging to study. Finding the ground state of an HP sequence is itself an NP-hard problem [6,7].

It should be pointed out that since the HP chain is constructed specifically to represent an individual protein, thermodynamic and structural properties of the chain depend on its chain length and the sequence of H and P monomers uniquely. In this case, there does not exist finite size scaling for this model, and thus there are no real phases and transitions in the thermodynamic limit.

To understand how a protein interacts with an attractive substrate, a model surface which attracts both types of monomers with strength ϵ_s is introduced to the 3-dimensional lattice as an xy-plane placed at $z = 0$. A second, non-attractive wall is placed at $z = h_w$ to bound the HP chain from above. Periodic boundaries are used for the x- and y-directions. The Hamiltonian of the model is then represented as:

$$\mathcal{H} = -\epsilon_{HH}n_{HH} - \epsilon_s n_s, \tag{1}$$

n_{HH} being the number of H–H interacting pairs and n_s being the number of monomers adjacent to the bottom surface. Such a system has first been introduced and studied in Ref. [8].

We have implemented two inventive Monte Carlo trial moves, namely, pull moves [9] and bond-rebridging moves [10], which are efficient both in compact and elongated configurations [11]. Employing Wang–Landau sampling [12] we have obtained the density

* Corresponding author.
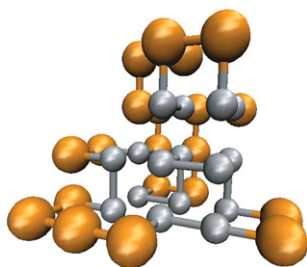*E-mail address:* ywli@physast.uga.edu (Y.W. Li).

Fig. 1. (Color online.) Ground state of the 36mer in free space ($E = -18$, $n_{HH} = 18$). The HP chain is displayed with larger polar and smaller hydrophobic residues.

of states in energy $g(E)$, from which the partition function Z can be calculated:

$$Z = \sum_i g(E_i) e^{-E_i/kT}, \qquad (2)$$

where k is the Boltzmann constant, T is temperature and the sum runs over all possible energies. The average energy $\langle E \rangle$ then follows:

$$\langle E \rangle = \frac{1}{Z} \sum_i E_i g(E_i) e^{-E_i/kT}, \qquad (3)$$

and the specific heat can be calculated from the fluctuation of energy:

$$C_v = \frac{1}{kT^2} \left(\langle E^2 \rangle - \langle E \rangle^2 \right). \qquad (4)$$

In Wang–Landau sampling, $g(E)$ is estimated iteratively. During the simulation, a multiplicative modification factor f is used to update $g(E)$ with an initial value $f_{init} = e^1$. A histogram in energy $H(E)$ is accumulated besides $g(E)$. When a "flat" histogram is attained, the simulation is brought to the next iteration: $H(E)$ is reset and f is reduced, $f_{new} = \sqrt{f}$. A histogram is said to be "flat" when all entries in $H(E)$ are greater than $p \times H_{ave}$, where p is
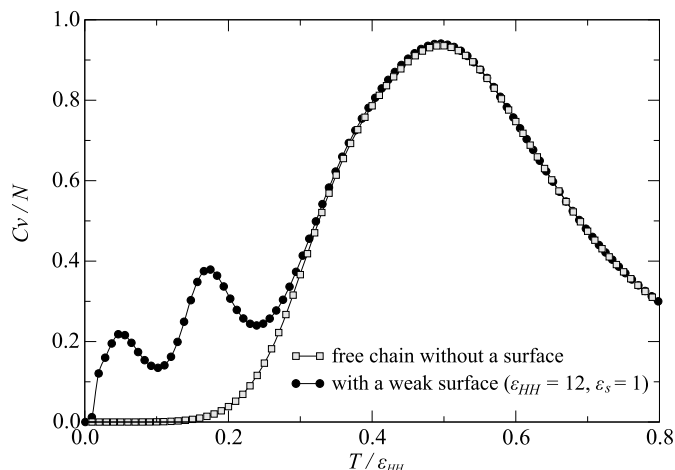

Fig. 2. Specific heat of an HP 36mer: (i) in a 3D free space, (ii) interacting with a weakly attractive surface ($\epsilon_{HH} = 12$, $\epsilon_s = 1$, $h_w = 37$). Error bars are smaller than the data points.

the flatness criterion and H_{ave} is the average of all entries in $H(E)$. To yield reliable estimates for $g(E)$, all results presented in this work are obtained by using a flatness criterion $p = 0.8$. The simulation stops when the natural log of the final modification factor, $\ln(f_{final})$, reaches a preset minimum value of 10^{-8}.

3. Results

As a first step, we have simulated an HP sequence with 36 beads (PPPHHPPHHPPPPPHHHHHHHPPHHPPPPHHPPHPP), which has first been studied in Ref. [13]. In free space, the HP chain undergoes a coil–globule transition as temperature decreases, resulting in a compact hydrophobic core with polar residues residing on the exterior to screen the core from the polar solvents. See Fig. 1. Similar behaviors have also been observed for other lengths and sequences [14–16].
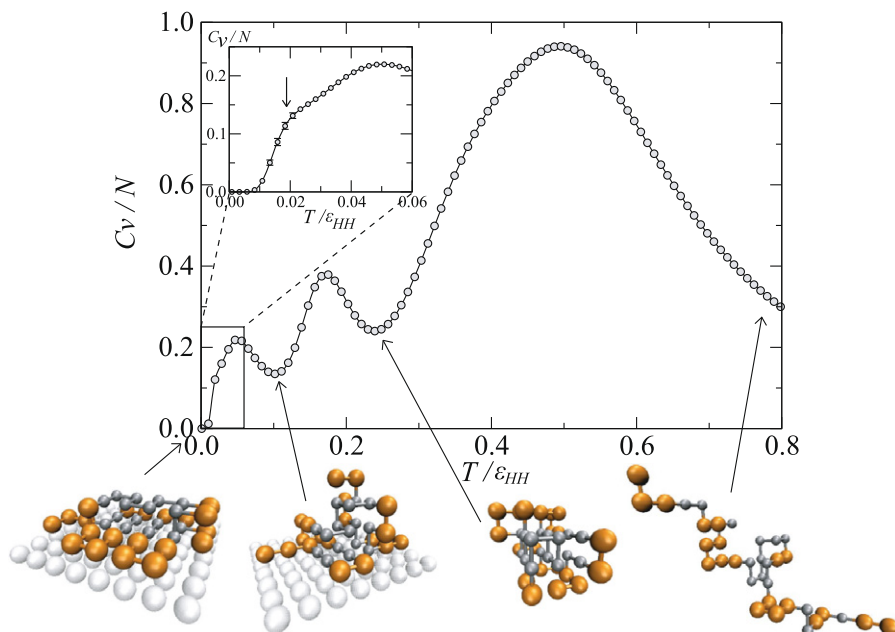

Fig. 3. (Color online.) Specific heat of an HP 36mer interacting with a weak attractive surface ($\epsilon_{HH} = 12$, $\epsilon_s = 1$, $h_w = 37$). Error bars smaller than the data points are not shown. Typical structures are displayed with larger polar and smaller hydrophobic residues.
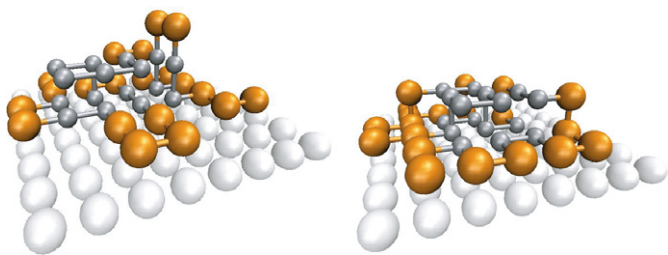
Fig. 4. (Color online.) First excited states of the 36mer interacting with a weak attractive surface, $E = -240$ ($n_{HH} = 18$, $n_s = 24$). The HP chains are displayed with larger polar and smaller hydrophobic residues.

With a surface that attracts both H and P monomers, the HP chain exhibits a much richer "transition" hierarchy due to the competition between surface adsorption and attraction within the polymer. To illustrate the effects of the surface strength on the "transition" behaviors, both a weak surface ($\epsilon_{HH} = 12$, $\epsilon_s = 1$) and a strong surface ($\epsilon_{HH} = 1$, $\epsilon_s = 1$) are considered. 15 independent runs were performed to obtain statistical errors in each case. CPU time grows with the size of the energy range of the systems: it takes about 15 minutes to finish a simulation on an AMD Opteron dual-core 2.2 GHz processor for the surface-free case using 19 energy bins, 5–10 hours for the strong attractive surface case with 51 energy bins, but more than 10 days are generally needed for the weak attractive surface where there are 242 energy bins in the full energy range.

Fig. 2 shows a typical specific heat of an HP chain in three-dimensional free space, as well as that when it interacts with a weakly attractive surface. In the latter case the height of the non-attractive wall is set to be $h_w = N + 1 = 37$, i.e., there are 36 layers between the two horizontal surfaces, the 36mer can touch both surfaces with its ends only when it is a fully stretched, vertical chain. While there is only a single peak corresponding to the collapse "transition" at $T/\epsilon_{HH} \approx 0.5$ for a free chain, three peaks are observed in the case with a weakly attractive surface. From this

comparison it is obvious that the two peaks in the low temperature regime are due to the influence of the attractive substrate.

The largest peak which overlaps with the free chain's specific heat represents the same coil–globule transition: during this stage, the HP polymer transforms from an extended coil-like structure to a desorbed, but compact, globule. Typical resulting structures are principally the same as the ground states of the free chain in the absence of the surface, as shown in Fig. 3.

The middle peak at $T/\epsilon_{HH} \approx 0.18$ signals adsorption during which the compact HP globule "docks" at the surface with the hydrophobic core remaining intact. The conformation spans several layers vertically and the total energy of the system is lowered slightly due to contact with the surface, by an amount dependent upon the number of surface contacts.

Further decrease in temperature brings the system to the third transition at $T/\epsilon_{HH} \approx 0.05$ where it strives for a maximum number of surface interactions without sacrificing an intact, energetically minimized hydrophobic core. With a large value of ϵ_{HH}, forming H–H contacts is immensely more energetically favorable than forming surface contacts.

A closer look at the low temperature regime reveals a subtle shoulder at $T/\epsilon_{HH} \approx 0.02$, which has not been found in Ref. [13]. This shoulder is a crossover signal owing to a transition from the ground state to the first few excited states at low temperature, that is similar to the case reported in the investigation of freezing and collapse of homopolymers [11,17]. In our system, this excitation is an effect purely due to the existence of the surface, which is supported by the analyses on other structural parameters (not presented here) and also by observing the typical states at $T/\epsilon_{HH} = 0$ and $T/\epsilon_{HH} \approx 0.02$. At $T/\epsilon_{HH} = 0$ where the polymer occupies just ground states, only rectangular cores are able to maximize the number of surface contacts. This gives the ground state energy $E = -241$ ($n_{HH} = 18$, $n_s = 25$). An example of such a structure is shown in the bottom left-hand corner of Fig. 3. At $T/\epsilon_{HH} \approx 0.02$ where first and second excited states dominate, two core shapes are observed: rectangular and "L-shape" as shown in Fig. 4. Since these two cores have the same energetic contribution
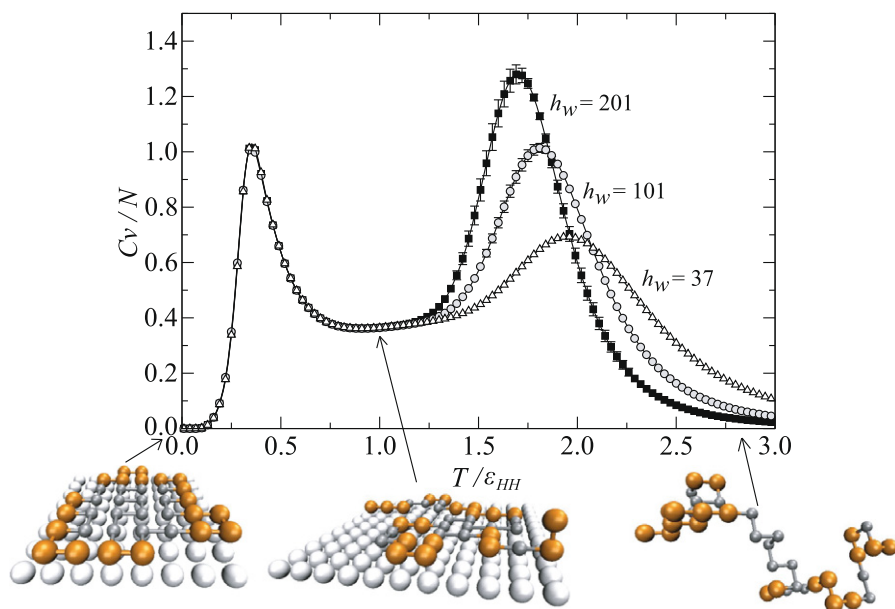

Fig. 5. (Color online.) Specific heat of an HP 36mer interacting with a strong attractive surface ($\epsilon_{HH} = 1$, $\epsilon_s = 1$). Error bars smaller than the data points are not shown. Typical structures are displayed with larger polar and smaller hydrophobic residues.
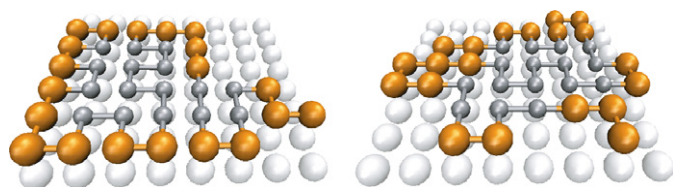
Fig. 6. (Color online.) First excited states of the 36mer interacting with a strong attractive surface, $E = -49$ ($n_{HH} = 13$, $n_s = 36$). The HP chains are displayed with larger polar and smaller hydrophobic residues.

($n_{HH} = 18$, with 8 H monomers interacting with the surface for both cases), the excitation from the ground state to the first few excited states can only be due to the decrease in surface contacts of the P monomers.

On the other hand, when the HP chain is brought near a strong surface field, the transition behavior is quite different from the previous case. We observe only two peaks in the specific heat, see Fig. 5. The one at higher temperature corresponds to an adsorption transition, and the other one, at $T/\epsilon_{HH} \approx 0.3$, corresponds to the collapse transition that takes place on the surface. At high temperature, the extended HP chain first attaches to the surface until most monomers touch it. This adsorbed, yet expanded structure then undergoes a second transition to maximize the number of H–H contacts and finally achieves a film-like, two-dimensional compact structure.

Unlike the previous case with a weakly attractive surface, there is no shoulder in the very low temperature regime. This is because the excitation from the ground state ($E = -50$, $n_{HH} = 14$, $n_s = 36$, as shown in the bottom left-hand corner of Fig. 5) to the first excited states ($E = -49$, $n_{HH} = 13$, $n_s = 36$) in this case is mainly achieved by the destruction of the H–H contacts, while the entire chain still remains completely attached to the substrate. Fig. 6 shows two examples of first excited states having different hydrophobic cores from that of the ground state. Therefore, the effect of the excitation cannot be singled out to give an individual signal, but is rather included within the collapse transition peak.

Fig. 5 also shows the influence of the entropic effects on the transition behavior due to the height h_w of the non-interacting wall bounding the chain from above. Obviously the low-temperature collapse is not affected at all, as it takes place on the attractive surface which is not related to the steric upper surface. The desorption–adsorption transition, however, shows systematic dependence on h_w. A smaller h_w restricts the vertical movements of the HP chain to a larger extent, resulting in a smaller entropy gain (and thus a less pronounced peak in the specific heat) as less translational variations of the same configuration are allowed. The chain is also more likely to be on the attractive surface, resulting in a higher adsorption temperature. This dependence of the

adsorption transition peak on h_w is also reported recently [18], where an off-lattice homopolymer model is used to study polymer adsorption.

4. Summary and outlook

To summarize, we studied protein folding using Wang–Landau sampling with a minimalistic lattice model, the HP model. Two Monte Carlo trial moves, pull moves and bond-rebridging moves, which work particularly well with our algorithm, were implemented. In this work, we have focused on investigating protein adsorption. With a weak surface, three-stage folding is identified, while two-stage folding is found with a strong surface. Varying the height of the upper non-interacting surface regulates the total entropy of the system, which affects the higher temperature "transition". Further work in progress includes longer benchmark HP sequences and the investigation of the thermodynamics of different structural quantities and their relationship with different "transitions".

Acknowledgement

References

[1] C. Branden, J. Tooze, Introduction to Protein Structure, Garland, New York, 1991.
[2] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Protein Struct. Funct. Genet. 21 (1995) 167.
[3] K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl, Annu. Rev. Biophys. Biomol. Struct. 37 (2008) 289.
[4] J.R. Banavar, A. Maritan, Annu. Rev. Biophys. Biomol. Struct. 36 (2007) 261.
[5] K.A. Dill, Biochemistry 24 (1985) 1501;
    K.F. Lau, K.A. Dill, Macromolecules 22 (1989) 3986.
[6] B. Berger, T. Leighton, J. Comput. Biol. 5 (1998) 27.
[7] P. Crescenzi, et al., J. Comput. Biol. 5 (1998) 409.
[8] M. Bachmann, W. Janke, Phys. Rev. E 73 (2006) 020901(R).
[9] N. Lesh, M. Mitzenmacher, S. Whitesides, in: RECOMB, 2003, p. 188.
[10] J.M. Deutsch, J. Chem. Phys. 106 (1997) 8849.
[11] T. Wüst, D.P. Landau, Phys. Rev. Lett. 102 (2009) 178101.
[12] F. Wang, D.P. Landau, Phys. Rev. Lett. 86 (2001) 2050;
    F. Wang, D.P. Landau, Phys. Rev. E 64 (2001) 056101;
    F. Wang, D.P. Landau, Comput. Phys. Comm. 147 (2002) 674.
[13] A.D. Swetnam, M.P. Allen, Phys. Chem. Chem. Phys. 11 (2009) 2046.
[14] M. Bachmann, W. Janke, J. Chem. Phys. 120 (2004) 6779.
[15] T. Wüst, D.P. Landau, Comput. Phys. Comm. 179 (2008) 124.
[16] Y.W. Li, T. Wuest, D.P. Landau, Nanophenomena at Surfaces, in: M. Michailov (Ed.), Springer Series in Surface Sciences, vol. 47, Springer-Verlag, Berlin, Heidelberg, 2011, Chapter 7, doi:10.1007/978-3-642-16510-8_7, in press.
[17] T. Vogel, M. Bachmann, W. Janke, Phys. Rev. E 76 (2007) 061803.
[18] M. Möddel, W. Janke, M. Bachmann, Phys. Chem. Chem. Phys. 12 (2010) 11548.