Thermodynamics and structural properties of a confined HP protein determined by Wang-Landau simulation

Busara Pattanasiri¹,², Ying Wai Li³,⁷, David P. Landau³, Thomas Wüst⁴ and Wannapong Triampo¹,⁵,⁶
¹ R&D Group of Biological and Environmental Physics, Department of Physics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
² Department of Chemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
³ Center for Simulational Physics, Department of Physics and Astronomy, University of Georgia, Athens, GA 30602, USA
⁴ Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
⁵ ThEP Center, CHE, 328 Si Ayutthaya Road, Bangkok 10400, Thailand
⁶ Institute for Innovative Learning, Mahidol University, 999, Phuttamonthon 4 Road, Salaya, Nakorn Pathom 73170, Thailand
⁷ National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
E-mail: bairng@physast.uga.edu, ywli@physast.uga.edu, dlandau@physast.uga.edu, thomas.wuest@wsl.ch, wtriampo@gmail.com

Abstract. Understanding protein folding confined by surfaces is important for both biological sciences and the development of nanomaterials. In this work, we study the properties of a confined HP model protein by three different types of surfaces, namely, surfaces that attract: (a) all monomers; (b) only P monomers; and (c) only H monomers. The thermodynamic and structural quantities, such as the specific heat, number of surface contacts, and number of hydrophobic pairs, are obtained by using Wang-Landau sampling. The conformational “transitions”, specifically, the debridging process and hydrophobic core formation, can be identified based on an analysis of these quantities. We found that these transitions take place at different temperatures, and the ground state configurations show variations in structural properties when different surface type is used. These scenarios are confirmed by snapshots of typical states of the systems. From our study, we conclude that the thermodynamics of these transitions and the structural changes depend on the combined actions of both the composition of the H monomers and the P monomers in the HP chain and the surface types.

1. Introduction
Proteins are important biological molecules that perform most of the vital functions in living cells. Some proteins provide structural support, while others are involved in defending the body from antigens, or in muscle contraction [1]. In order to perform their specific functions, proteins must fold to a correct three dimensional shape or conformation, known as the native state. Incorrect protein folding are shown to be associated with a number of diseases, including Alzheimers, Huntington’s, Parkinsons, prion and other degenerative diseases [2–4]. The native
conformation of a protein is determined by its amino acid sequence; however, understanding how proteins find their own native structure remains one of the most fundamental problems in the biological sciences.

It is known that within the living cell, the environment is extraordinarily complex, many other molecules exist and folding spaces are restricted [5, 6]. Therefore, the effects of the surrounding environment on protein folding has drawn much attention from the research community; one of which being the confinement effect. This includes not only protein folding in a cell but also its practical applications, such as protein crystallization [7], biotechnology [8–10] and the food industry [11, 12].

Experimental studies and simulations have suggested that protein stability and folding dynamics are strongly correlated with the geometry and degree of confinement [13–16]. The effect of confinement can be absent if the confinement space is too large, whereas the native state is very restricted if the confinement space is too small [17]. Furthermore, hydrophobicity of the confining surfaces also plays an important role in protein folding. For example, weak hydrophobic confining surfaces can accelerate the folding of proteins [18, 19].

In this work, we investigate the effect of surface properties on protein folding under confinement. We focus on the thermodynamics and structural behavior using the hydrophobic-polar (HP) lattice protein model [20]. This simple, coarse-grained protein model is chosen because it is difficult to obtain the “big picture” of this problem using, for instance, atomistic simulations, force-field or molecular dynamics of specific proteins where too many degrees of freedom have to be accounted for. Monte Carlo methods are in general amenable tools to study equilibrium properties. However, the seemingly simple HP model is unexpectedly challenging to study due to the rough energy landscapes. On the contrary, Wang-Landau sampling [21–24], coupled with inventive Monte Carlo moves, is shown to be robust to overcome this problem. It is, therefore, used to estimate the density of states from which the thermodynamic and structural properties can be calculated. The conformational “transitions” are also discussed in this paper.

2. Methods

2.1. Hydrophobic-polar (HP) protein folding model

The HP model is a simplified lattice model which is used to simulate protein folding to its native state. It was first proposed by K. A. Dill in 1985 [20]. The protein chain is defined as a sequence of two types of amino acid residues, hydrophobic (H) or polar (P), and is modeled as a self-avoiding walk in a 3-dimensional cubic lattice. The driving force for protein folding is imitated by assigning a favorable interaction $\varepsilon_{HH}$ between two non-bonded, nearest-neighbor H residues.

To study protein folding under confinement, we construct two parallel surfaces in the $xy$ plane at $z = 0$ and $z = h_w$. The boundary conditions are periodic only in the $x$ and $y$ directions. The energy can then be calculated by:

$$E = -\varepsilon_{H}n_{HH} - \varepsilon_{SH}n_{SH} - \varepsilon_{SP}n_{SP},$$

where $n_{HH}$ is the number of hydrophobic pairs; $n_{SH}$ and $n_{SP}$ are the number of H monomers and P monomers adjacent to the surfaces; $\varepsilon_{SH}$ and $\varepsilon_{SP}$ are the surface attractive strengths for the H and P monomers, respectively. We have used a 36mer ($P_3P_2P_2P_2P_2P_2P_2P_2P_2HP_2$) that consists of 16 H monomers and 20 P monomers. The 36mer is one of the HP benchmark sequences for testing a genetic algorithm [25]. This short sequence is already able to exhibit features of protein folding and it has been used to investigate protein-surface interaction [26, 27].

The distance between the two confining surfaces was set to be smaller than the chain length ($h_w = 10$). A schematic diagram of this model is shown in figure 1.
2.2. Wang-Landau sampling

Wang-Landau (WL) sampling [21–24] is a Monte Carlo algorithm to calculate the relative density of states for a physical system. In the first step of the simulation, we applied WL sampling to determine the density of states as a function of the energy of the system, \( g(E) \). The partition function at a particular temperature \( T \) can be obtained by summing the density of states over all energies weighted by the Boltzmann factor, i.e.,

\[
Z(T) = \sum_{E} g(E) e^{-E/kT},
\]

(2)

where \( E \) is the total energy of the system described by equation (1), and \( k \) is the Boltzmann constant. Since \( g(E) \) is independent of the temperature of the system, we can then compute any desired thermodynamic quantities at all temperatures after obtaining \( g(E) \). For example, the average energy \( \langle E \rangle(T) \) and the heat capacity \( C_V(T) \) are calculated as:

\[
\langle E \rangle(T) = \frac{1}{Z(T)} \sum_{E} E g(E) e^{-E/kT},
\]

(3)

\[
C_V(T) = \frac{1}{kT^2} (\langle E^2 \rangle - \langle E \rangle^2).
\]

(4)

The specific heat is then defined by \( C_V/N \), where \( N \) is the number of monomers in the chain. Traditionally speaking, the specific heat is an indicator of phase transitions as they are usually associated with large energy fluctuations. However, as we will show from our results, it is no longer adequate to signal all the phase transitions for our system as some structural transitions do not necessarily affect the total energy of the system.

In the WL method, \( g(E) \) is iteratively refined during the simulation. The algorithm starts with an initial guess of \( g(E) = 1 \) for all possible energies \( E \). The histogram in energy, \( H(E) \), which keeps track of the number of visits to an energy within an iteration, is set to \( H(E) = 0 \) at the beginning. During a Monte Carlo step, a trial configuration is generated and accepted with probability
Figure 2. (Color online) Examples of Monte Carlo trial moves used in this study (illustrated on a two-dimensional square lattice): (a) a pull move, and (b) a bond-rebridging move.

\[ p(E_{\text{old}} \rightarrow E_{\text{new}}) = \min \left( \frac{g(E_{\text{old}})}{g(E_{\text{new}})}, 1 \right). \quad (5) \]

If the trial move is rejected, the old configuration is retained and it also enters into the statistics. Once the configuration is determined, the density of states is modified by

\[ g(E) \rightarrow f \times g(E), \]

where \( f \) is a modification factor which is slowly reduced to unity during the simulation. In this work, the initial and the final modification factor are \( f_0 = e^{\frac{1}{2}} \approx 2.71828, \) and \( f_{\text{final}} = \exp(10^{-8}) \), respectively. Every time \( g(E) \) is updated, \( H(E) \) is also accumulated by \( H(E) \rightarrow H(E) + 1 \).

This procedure repeats until a flat histogram is obtained, i.e., the simulation time spent at each energy is approximately the same. The histogram is then reset \( H(E) = 0 \), the density of states \( g(E) \) is kept, and a new iteration begins with a reduced modification factor, \( f_{i+1} = \sqrt{f_i} \). A “flat” histogram is attained when all the entries of \( H(E) \) are greater than 80% of the average number of total visits: \( \frac{1}{N_0} \sum_E H(E) \), where \( N_0 \) is the total number of energy bins.

We have adopted two non-traditional Monte Carlo trial moves, pull moves [28] and bond-rebridging moves [29], to generate new configurations. Examples of these moves are shown in figure 2. The combination of WL sampling with these moves has been proven to be an efficient iterative method to estimate an accurate \( g(E) \) [30, 31]. The acceptance rate of the bond-rebridging moves is rather low because it is not always possible to propose this move from a configuration (the configuration needs to be dense enough to have at least a pair of parallel bonds; sometimes two pairs of parallel bonds are necessary for this move to be feasible). As a consequence, we assigned a calling probability for pull move and bond-rebridging move to be 0.2 and 0.8, respectively.

2.3. Production run

The second step of the simulation is making a production run from the \( g(E) \) that we obtained from the first step. Here, we perform a multicanonical simulation [32] to estimate a two-dimensional density of states \( g(E, Q) \) in energy \( E \) and structural quantity \( Q \). In this process, the acceptance probability is the same as equation (5), except that \( g(E) \) is held fixed during this stage. Every time a configuration is selected, the structural quantities \( Qs \) are calculated while the corresponding two-dimensional histograms \( H(E, Q) \) are updated. This has an advantage that a number of structural quantities can be calculated within a single run. Examples of structural quantities are \( n_{HH}, n_{SP}, \) and \( n_{EH} \). At the end of the simulation, the \( g(E, Q) \) is obtained by re-weighting \( g(E) \) with \( H(E, Q) \):

\[ g(E, Q) = g(E) H(E, Q). \quad (6) \]

The partition function and the average of the structural quantity \( Q \) are therefore calculated as
\[
Z(T) = \sum_{E,Q} g(E,Q) e^{-E/kT},
\]
\[
\langle Q \rangle(T) = \frac{1}{Z_Q(T)} \sum_{E,Q} Qg(E,Q)e^{-E/kT}.
\]

3. Results and Discussion

To investigate the effects of surface characteristics on the physical behavior of a confined HP protein, we considered three types of confining surfaces:

(i) non-specific surfaces that attract all monomers \((\varepsilon_{SH} = \varepsilon_{SP} = \varepsilon_{HH} = 1)\);
(ii) polar surfaces \((\varepsilon_{SH} = 0, \varepsilon_{SP} = \varepsilon_{HH} = 1)\);
(iii) hydrophobic surfaces \((\varepsilon_{SP} = 0, \varepsilon_{SH} = \varepsilon_{HH} = 1)\).

Here, 15 independent runs were performed in each case to obtain statistical errors. On an AMD Opteron 2.2 GHz CPU, a single simulation took around 3-21 hours, 5-20 hours, and 1-12 hours for the 36mer confined between non-specific surfaces, polar surfaces, and hydrophobic surfaces, respectively.

Generally, the conformational “phase transitions” can be identified by locating the peak of the specific heat \(C_V/N\). However, it is very important to investigate the structural quantities together with \(C_V/N\) so as to identify individual transition processes, since structural transitions do not necessarily cause large energy fluctuations. In these cases, the transition does not incur a pronounced peak in \(C_V/N\) but merges it into another transition peak caused by other processes, or it occurs as a subtle shoulder.

Figure 3 shows the thermodynamics for the 36mer confined between non-specific surfaces which attract all monomers \((\varepsilon_{SH} = \varepsilon_{SP} = \varepsilon_{HH} = 1)\). At high temperature, the protein tends to stretch out. Since the distance between two confining surfaces was set to be smaller than the chain length, parts of the chain form contacts with both surfaces and the chain acts as a “bridge” between the two surfaces. To maximize the surface contacts as the temperature lowers, the chain detaches from one surface to adsorb on the other until most of the adsorbents sit on the same surface. We call the destruction of the bridge a “debridging” process. This process causes a small amount of energy fluctuations, leading to a shoulder observed in the specific heat at \(kT/\varepsilon_{HH} \approx 1.6\). Further decrease in temperature induces the hydrophobic core (H-core) formation, where the adsorbed, extended chain collapses to form a compact structure, in which the hydrophobic monomers cluster together inside and the polar monomers reside on the outside. Unlike the debridging process, the H-core formation causes a large decrease in energy. A peak in specific heat is then observed at \(kT/\varepsilon_{HH} \approx 0.3\). More examples of these behaviors can be found in [33]. The debridging process and the H-core formation can be identified more clearly by the thermal derivatives of \(\langle n_{SH} \rangle\), \(\langle n_{SP} \rangle\), and \(\langle n_{HH} \rangle\). The negative peaks in \(d\langle n_{SH} \rangle/dT\) and \(d\langle n_{SP} \rangle/dT\) are associated with the formation of surface contacts, i.e., debridging and adsorption; while the negative peak in \(d\langle n_{HH} \rangle/dT\) signifies the construction of the hydrophobic interactions. See [27, 34, 35] for more details of the identification of phase transitions.

Figure 4 and figure 5 show the thermodynamics for the 36mer confined between polar surfaces \((\varepsilon_{SH} = 0, \varepsilon_{SP} = \varepsilon_{HH} = 1)\) and hydrophobic surfaces \((\varepsilon_{SP} = 0, \varepsilon_{SH} = \varepsilon_{HH} = 1)\), respectively. The same basic transitions, debridging at a higher temperature and the H-core formation at a lower temperature, are also observed as in the previous case. However, they occur at different temperatures. With the same surface attractive strength, having a larger number of monomers attracted by the surface makes adsorption to occur easier, resulting in a higher debridging temperature. For this reason, the debridging process of the 36mer confined between surfaces that attract all monomers occurs at a higher temperature \((kT/\varepsilon_{HH} \approx 1.6)\). It takes place at a
Figure 3. (Color online) Thermodynamic and structural quantities of the 36mer confined between two attractive surfaces that attract all monomers ($\varepsilon_{SH} = \varepsilon_{SP} = \varepsilon_{HH} = 1$, $h_w = 10$). Typical structures are displayed with smaller black spheres and larger orange spheres for the H monomers and the P monomers, respectively. The horizontal arrows beside the labels indicate the axes to which the quantities refer. Errors are smaller than the size of the data points.

lower temperature with the polar surfaces ($kT/\varepsilon_{HH} \approx 1.0$), and at an even lower temperature with the hydrophobic surfaces ($kT/\varepsilon_{HH} \approx 0.75$).

The details of forming an H-core are different from case to case, resulting in different ground state structures. Typical configurations of the ground states are shown at $kT/\varepsilon_{HH} = 0$. For the case that the confining surfaces attract all types of monomers, the ground state is a two-dimensional structure with a planar H-core, and it is formed after the chain adsorbed to the surface. For the other two cases, the ground states are three-dimensional structures, but the hydrophobic cores are three-(or two-)dimensional structures that adsorb on a polar (or hydrophobic) surface.

The lower panel of figure 4 reveals a negative peak in $d\langle n_{SH} \rangle/dT$ and a positive peak in $d\langle n_{SP} \rangle/dT$ when the H-core is formed at a very low temperature ($kT/\varepsilon_{HH} \approx 0.25$). This suggests that during the H-core formation process, some surface contacts with the P monomers need to be broken to facilitate the construction of the H-H contacts. Similar behavior is observed in figure 5 for the case of hydrophobic confining surfaces, except that there is no negative peak in $d\langle n_{SH} \rangle/dT$ at low temperature. It is because all the H monomers have already adsorbed
on the surface before H-core formation in the latter case. A subtle shoulder in both $C_V$ and $d\langle n_{HH}\rangle/dT$ in the very low temperature regime ($kT/\varepsilon_{HH} \approx 0.25$) signals a two-stage acquisition of the ground state, where the chain maximizes the number of H-H contacts by uncoiling the dense, but still slightly disordered, globule before collapsing again to form a highly ordered, rectangular H-core.

4. Conclusion
We studied the physical behavior of a protein confined between two attractive surfaces using a coarse-grained lattice model, the HP model. Wang-Landau sampling with inventive Monte Carlo moves, pull moves and bond-rebridging moves, were employed. In this work we considered three different types of surfaces with different affinities to the H or P monomers. The thermodynamic and structural quantities were used to identify conformational “transitions”. It was found that at high temperature the chain can touch both surfaces. As temperature decreases, two basic transitions, debridging and H-core formation, were observed. We further discovered the number of adsorbates, i.e., monomers that are attracted by the surfaces, played a major role in these
transition processes. This, in turn, depends on both the surface types and the proportion of the H and P monomers in the chain. As a consequence, the debridging and H-core formation temperatures varied as these systems parameters change, and different structural properties for the ground states were resulted. All these imply that the materials of the confining surfaces, as well as the protein chain composition, are the dominant factors in determining the physical properties of this kind of composites.

Acknowledgments

This work was supported by the National Science Foundation under Grants DMR-0810223. B. Pattanasiri is supported by The Royal Thai Government Scholarship under the Development and Promotion of Science and Technology Talent Project (DPST), and the grant from The Institute for the Promotion of Teaching Science and Technology (IPST). Y.W. Li was partly sponsored by the Office of Advanced Scientific Computing Research; U.S. Department of Energy. Part of the work was performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. De-AC05-00OR22725.
References
[1] Creighton T E 1993 *Proteins: Structures and Molecular Properties* (New York: W. H. Freeman)
[2] Taubes G 1996 *Science* 271 1493–95
[3] Selkoe D J 2004 *Nat. Cell. Biol.* 6 1054–61
[4] Prusiner S B 1998 *Proc. Natl. Acad. Sci. USA* 95 13363–83
[5] Zhou H -X 2008 *Arch. Biochem. Biophys.* 469 76–82
[6] Ebbinghaus S, Dhar A, McDonald D J and Gruebele M 2010 *Nature Methods* 7 319–23
[7] Ng J D, Stevens R C and Kuhn P 2008 *Methods Mol. Biol.* 426 363–76
[8] Birrbaum S 1993 *Immmobilised Macromolecules: Application Potentials* (Springer Series in Applied Biology) ed U B Sleytr *et al* (London: Springer)
[9] Martin C R and Kohli P 2003 *Nat. Rev. Drug Discov.* 2 29–37
[10] Oh K J, Cash K J and Plaxco K W 2009 *Chemistry* 15 2244–51
[11] Foegeding E A, Luck P J and Davis J P 2006 *Food Hydrocolloids* 20 284–92
[12] Schmidt I, Novales B, Boué F and Axelos M 2010 *J. Colloid Interface Sci.* 345 316–24
[13] Campanini B, Bologna S, Cannone F, Chirico G, Mozzarelli A and Bettati S 2005 *Protein Sci.* 14 1125–33
[14] Peterson R W, Anbalagan K, Tomson C and Wand A J 2004 *J. Am. Chem. Soc.* 126 9498–99
[15] Ping G, Yuan J M, Vallieres M, Dong H, Sun Z, Wei Y, Li F Y and Lin S H 2003 *J. Chem. Phys.* 118 8042
[16] Takagi F, Koga N and Takada S 2003 *Proc. Natl. Acad. Sci. USA.* 100 11367
[17] Rathore N, Knotts T A and Pablo J D 2006 *Biophys. J.* 90 1767–73
[18] Jewett A I, Baumketner A and Shea J -E 2004 *Proc. Natl. Acad. Sci. U.S.A.* 101 13192–97
[19] Jewett A I, Baumketner A and Shea J -E 2006 *J. Mol. Biol.* 363 945–57
[20] Dill K A 1985 *Biochemistry* 24 1501–9
[21] Wang F and Landau D P 2001 *Phys. Rev. Lett.* 86 2050–53
[22] Wang F and Landau D P 2001 *Phys. Rev. E* 64 056101
[23] Wang F and Landau D P 2002 *Comput. Phys. Commun.* 147 674–77
[24] Landau D P, Tsai S -H and Exler M 2004 *Am. J. Phys.* 72 1294
[25] Unger R and Moult J 1993 *J. Mol. Biol.* 231 75–81
[26] Swetnam A D and Allen M P 2009 *Phys. Chem. Chem. Phys.* 11 2046–55
[27] Li Y W, Wüst T and Landau D P 2011 *Comput. Phys. Commun.* 182 1896–99
[28] Lesh N, Mitzenmacher M and Whitesides S 2003 *Proc. Seventh Annual Int. Conf. on Computational Molecular Biology*, ed T Lengauer *et al* (Berlin: RECOMB) p 188
[29] Deutsch J M 1997 *J. Chem. Phys.* 106 8849
[30] Wüst T and Landau D P 2009 *Phys. Rev. Lett.* 102 178101
[31] Wüst T and Landau D P 2012 *J. Chem. Phys.* 137 064903
[32] Berg B A and Neuhaus T 1992 *Phys. Rev. Lett.* 68 9–12
[33] Pattanasiri B, Li Y W, Landau D P and Wüst T 2012 *Int. J. Modern Phys. C* 23 1240008
[34] Wüst T, Li Y W and Landau D P 2011 *J. Stat. Phys.* 144 638–51
[35] Li Y W, Wüst T and Landau D P 2013 *Phys. Rev. E* 87 012706