Effect of surface attractive strength on structural transitions of a confined HP lattice protein

Busara Pattanasiri¹, Ying Wai Li², Thomas Wüst³, and David P. Landau¹

¹Center for Simulational Physics, Department of Physics and Astronomy, The University of Georgia, Athens, GA 30602, USA
²National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
³Scientific IT Services, ETH Zürich IT Services, 8092 Zürich, Switzerland
E-mail: busara@physast.uga.edu

Abstract.
We investigate the influence of surface attractive strength on structural transitions of a hydrophobic-polar (HP) lattice protein confined in a slit formed by two parallel, attractive walls. We apply Wang-Landau sampling together with efficient Monte Carlo updates to estimate the density of states of the system. The conformational transitions, namely, the debridging process and hydrophobic core formation, can be identified by analyzing the specific heat together with several structural observables, such as the numbers of surface contacts, the number of hydrophobic pairs, and radii of gyration in different directions. As temperature decreases, we find that the occurrence of the debridging process is conditional depending on the surface attractive strength. This, in turn, affects the nature of the hydrophobic core formation that takes place at a lower temperature. We illustrate these observations with the aid of a HP protein chain with 48 monomers.

1. Introduction
Confinement is one of the major effects that impacts protein folding in a cellular environment. In a living cell, there is a highly crowded and complex setting containing multiple biological macromolecules, such as lipids, carbohydrates, and proteins [1]. For example, proteins and nucleic acids inside the cytoplasm occupy about 30 – 40% of the total cell volume [2, 3, 4]. This complexity can hinder the correct folding of a protein to its native structure, which leads to various pathologies. Incorrect protein folding is shown to be associated with a number neurodegenerative disorders, including Alzheimer’s, Huntington’s, and Mad Cow diseases [5, 6].

The behavior of confined proteins that deviates from bulk properties has drawn much attention in a wide variety of research fields, ranging from the understanding of biological processes [7, 8, 9], improving material properties [10, 11], enhancing biological technologies [12, 13], and designing nanodevices and nanomaterials [14, 15]. Chemical properties of the confining surfaces govern the interactions with proteins, which can be exploited to enhance its selectivity. An understanding of the effects of confinement along with the confining surface properties on conformational or structural changes of proteins is thus essential.

In this work, we study the effect of surface attractive strength on structural transitions of
a confined protein with the use of a simplified coarse-grained model - the hydrophobic-polar (HP) lattice protein model [16, 17]. In spite of its simplicity, the ground state search of a HP model has been shown to be NP-complete for both 2- and 3-dimensional lattices [18, 19]. Traditional Metropolis sampling [20] is extremely inefficient in approaching the problem, due to the fact that the random walker is easily trapped in local free energy minima. Obtaining correct low temperature thermodynamics is then impractical within a reasonable time frame. To circumvent the problem, we employed a modern algorithm called Wang-Landau (WL) sampling [21, 22, 23], which allows us to deal with rugged free energy landscapes efficiently. Additionally, two inventive Monte Carlo moves, pull moves [24] and bond-rebridging moves [25], have been utilized in combination with WL sampling. This has been proven to be robust in ground state searches and accurate estimation of the density of states for lattice protein models [26, 27].

2. Model and simulation methods

2.1. Hydrophobic-polar (HP) protein model with surface confinement

The HP model [16, 17] is a well-known coarse-grained lattice model for the study of protein folding problems. This model is motivated by the fact that the hydrophobic effect, i.e., the tendency of hydrophobic (non-polar) amino acids to assemble together in an aqueous environment, is a dominant driving force for tertiary structure formation. Amino acids in a protein are classified as hydrophobic (H) or polar (P) according to the interactions of their side chains with water. The protein chain is modeled as a self-avoiding walk on a 2-dimensional square or a 3-dimensional cubic lattice, where an amino acid is located at each lattice site along the walk without overlaps. The hydrophobic effect is mimicked by assigning an attraction, $\varepsilon_{HH}$, between two non-bonded, nearest-neighbor H monomers. In order to study a confined protein in our system, two attractive walls are constructed in parallel to the $xy$–plane and distance $h_w$ apart. For simplicity, we only considered surfaces that attract all monomers with the same strength, $\varepsilon_S$, and this interaction is identical for both the top and bottom surfaces. The total energy $E$ of a confined protein is thus defined as

$$E = -\varepsilon_{HH}n_{HH} - \varepsilon_S(n_{SH} + n_{SP}),$$  \hspace{1cm} (1)

where $n_{HH}$, $n_{SH}$ and $n_{SP}$ are the numbers of H-H contacts and surface contacts with H and P monomers, respectively.

2.2. Thermodynamics and Wang-Landau sampling

In general, the transitions between conformational phases can be investigated via the temperature dependence of specific heat $C_V/N$, where $N$ is the system size. The heat capacity $C_V$ can be calculated by the energy fluctuation:

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{kT^2}, \hspace{1cm} \text{where} \hspace{1cm} \langle E \rangle = \frac{1}{Z} \sum_E E g(E) e^{-E/kT}. \hspace{1cm} (2)$$

Here, $k$ is the Boltzmann constant, $T$ is the temperature and $Z$ is the partition function, which can be obtained by summing the density of states, $g(E)$, over all energies weighted by the Boltzmann factor:

$$Z = \sum_E g(E) e^{-E/kT}. \hspace{1cm} (3)$$

In our simulations, we employed Wang-Landau (WL) sampling [21, 22, 23] to determine the $g(E)$ of the system. The simulation starts with an initial estimate of $g(E) = 1$ for all possible energies $E$. In each Monte Carlo update, a random configuration has been generated 20% of the
time from pull moves [24] and 80% of the time from bond-rebridging moves [25]. The probability of accepting such an update is given by

\[ p(E_i \rightarrow E_j) = \min \left( \frac{g(E_i)}{g(E_j)}, 1 \right), \tag{4} \]

where \( E_i \) is the initial energy and \( E_j \) is the energy of the trial configuration. Every time a new configuration is accepted, the density of states is adjusted by \( g(E_i) \rightarrow g(E_j) \times f \); otherwise \( g(E_i) \rightarrow g(E_i) \times f \). \( f \) is a modification factor which is initially set to \( f_0 = e^{1/2} \approx 2.71828 \).

With each energy update, the corresponding entry in the histogram \( H(E) \) is incremented by \( H(E) \rightarrow H(E) + 1 \). This procedure repeats until the histogram \( H(E) \) is sufficiently flat, i.e., all the entries of \( H(E) \) are greater than 80% of the average number of total visits. The histogram is then reset to \( H(E) = 0 \) and a new iteration begins with a reduced modification factor, \( f \rightarrow \sqrt{f} \). The simulation continues until \( f \) reaches \( f_{\text{final}} = \exp(10^{-3}) \). Finally, we obtain a good estimate for \( g(E) \) up to an undetermined constant.

### 2.3. Structural observables and production run

Observing \( C_V \) alone is no longer sufficient as it merely gives overall energy fluctuations that could be induced by a combination of coexisting structural transitions. In some cases, this may result in enhanced or suppressed energy fluctuations. Supplementary structural quantities are then necessary for a reliable identification of conformational transitions. The canonical averages of the number of nearest-neighbor H-H contacts, \( \langle n_{HH} \rangle \), and the numbers of monomer-surface contacts, \( \langle n_{SH} \rangle \) and \( \langle n_{SP} \rangle \), are useful quantities for monitoring energy contributing interactions. The average radii of gyration in the \( x \), \( y \), and \( z \) directions, \( \langle R^x_g \rangle \), \( \langle R^y_g \rangle \), and \( \langle R^z_g \rangle \), are informative properties about the HP chain conformations. \( R^x_g \) measures the degree of the spreading of monomers in the \( x \) direction and is defined by

\[ R^x_g = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r^x_i - r^x_{CM})^2}, \quad \text{with} \quad r^x_{CM} = \frac{1}{N} \sum_{i=1}^{N} r^x_i. \tag{5} \]

\( r^x_i \) and \( r^x_{CM} \) are the \( x \) coordinates of the \( i^{\text{th}} \) monomer and the center of mass, respectively. \( R^y_g \) and \( R^z_g \) are defined likewise.

To obtain these structural observables, we perform a multicanonical simulation [28] as a production run using the \( g(E) \) that we obtained from the previous simulation as the sampling weights. We estimate a 2-dimensional joint density of states \( g(E, Q) \) in energy \( E \) and structural quantity \( Q \). In this process, the acceptance probability is the same as in Eq. (4), except that \( g(E) \) is held fixed during this stage. Every time a configuration is selected, the structural quantities \( Q_s \) are calculated while the corresponding 2-dimensional histograms \( H(E, Q)_s \) are updated. At the end of the simulation, \( g(E, Q)_s \) are obtained by reweighting \( H(E, Q)_s \) with \( g(E) \):

\[ g(E, Q) = g(E)H(E, Q). \tag{6} \]

The average of the structural quantity \( Q \) can then be calculated by

\[ \langle Q \rangle = \frac{1}{Z} \sum_{E, Q} Qg(E, Q)e^{-E/kT}. \tag{7} \]

### 3. Results and discussion

We performed simulations on a 48mer (PHPHP3HPHPH3HPH6P2H3PHP2 HPH2P2HPH3P4H), which is one of the ten sequences proposed by Yue et al. for algorithm efficiency testing [29]. In
Figure 1. (color online) Thermodynamic and structural quantities of the 48mer confined between two surfaces with a separation of $h_w = 5$ and different surface attractive strength, $\varepsilon_S$. In all figures, error bars smaller than the data points are not shown. (a) Specific heat; (b)-(d) thermal derivatives of $\langle n_{HH} \rangle$, $\langle n_{SH} \rangle$, and $\langle n_{SP} \rangle$; (e)-(g) average radii of gyration in $x$, $y$, and $z$ directions. The 2D and 3D surface-free cases (i.e., without confinement) are also plotted for comparisons. Typical ground state structures are displayed at the bottom.
this sequence, there are equal numbers of H- and P-monomers. The chain is placed between two attractive surfaces with a small surface separation $h_w = 5$. The effect of surface attraction was studied by varying the ratio between the surface attractive strength and the attraction between the H-monomers, $\varepsilon_S/\varepsilon_{HH}$. Two cases were considered in this study: $\varepsilon_S/\varepsilon_{HH} = 0.5$ and 2.0. 15 independent runs were performed to obtain statistical errors in both cases.

Depending on the surface attractive strength, structural transitions occur at different temperatures. Fig. 1 shows the specific heat and the thermodynamic properties we defined in Sec. 2.2 and 2.3. Since the surface separation is much smaller than the HP chain length, i.e. $h_w \ll N$, parts of the chain form contacts with both surfaces and the chain forms a “bridge” between the two surfaces at high temperature for all cases. As temperature lowers, the HP chain tends to maximize the surface contacts. This process is signaled by negative peaks in the thermal derivatives of $\langle n_{SH} \rangle$ and $\langle n_{SP} \rangle$ due to the increase in the numbers of surface contacts. This occurs at $kT/\varepsilon_{HH} \approx 1.75$ for $\varepsilon_S/\varepsilon_{HH} = 2.0$, and at $kT/\varepsilon_{HH} \approx 0.8$ for $\varepsilon_S/\varepsilon_{HH} = 0.5$.

When the surface attraction is relatively strong, i.e. $\varepsilon_S/\varepsilon_{HH} = 2.0$, the chain in fact detaches from one surface to adsorb onto the other until most of the monomers sit on the same surface. Since most of the monomers need to spread out on the same surface to form a film-like structure as a result of this debridging process, this is accompanied by an increase in $\langle R_S^d \rangle$ and $\langle R_S^g \rangle$, but a decrease in $\langle R_H^z \rangle$ around the transition temperature. In terms of the conformational changes and the signals observed in relevant structural properties, the debridging process resembles an adsorption transition as if there was only one adsorbing surface. In this case, the same analysis techniques as for the HP adsorption model [30] could be applied for the understanding of structural transitions.

Note that this debridging process is completely missing for the weakly attractive surface case $\varepsilon_S/\varepsilon_{HH} = 0.5$. The slight increase in surface contacts (as signaled by small peaks in $d\langle n_{SH} \rangle/dT$ and $d\langle n_{SP} \rangle/dT$ at $kT/\varepsilon_{HH} \approx 0.8$), the relatively constant $\langle R_S^d \rangle$ but the decreasing $\langle R_H^z \rangle$ and $\langle R_S^g \rangle$ as temperature lowers give the hint of a slight unwinding of the chain in preparation for a more energetically favorable hydrophobic core (H-core) formation that follows.

Further decrease in temperature induces H-core formation, where the extended chain collapses to form a compact structure with hydrophobic monomers forming a central core and polar monomers forming the outer shell. Regardless of the surface attraction strength, this occurs at $kT/\varepsilon_{HH} \approx 0.4$ as shown by the peaks in $d\langle n_{HH} \rangle/dT$. However, we stress that these H-core formations are not identical. While the one in the $\varepsilon_S/\varepsilon_{HH} = 0.5$ case is a 3-dimensional process in which the globular structure retains contacts to both surfaces, the one in the $\varepsilon_S/\varepsilon_{HH} = 2.0$ case is a 2-dimensional process where it occurs on a single adsorbing surface.

4. Conclusion
We studied the influence of surface attractions on the structural transitions of a HP lattice protein confined between two attractive surfaces, combining Wang-Landau sampling with efficient Monte Carlo moves. We focused on the case where the surface separation is small compared to the chain length, and depending on the surface attraction strength, the debridging process may or may not be present. Decreasing the surface attraction shifts this process to a lower temperature (as proven by our simulations but results were not shown in this work); until it vanishes when the surface attraction becomes too weak. At low temperature, there is a hydrophobic core formation. With a strongly attractive surface where debridging has taken place at high temperature, H-core formation is a 2-dimensional process that results in a film-like ground state. For a weakly attractive surface, the H-core formation occurs while the HP chain is in contact with both the top and bottom surfaces, leading to a 3-dimensional, globular ground state.

Structural transitions are also be affected by other factors or system parameters as discussed in our previous studies [31, 32, 33]. The debridging process can disappear when the separation
between the two attractive surfaces is relatively large or sufficiently small compared to the chain length. Moreover, both the surface types and the proportion of the H and P monomers in the chain can influence the transitions. It implies that the size of confining space, as well as the chemical properties of the confining surfaces, are the dominant factors of structural transitions.

Acknowledgments
B. Pattanasiri was 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.L. was sponsored by the Oak Ridge Leadership Computing Facility, Office of Advanced Scientific Computing Research, U.S. Department of Energy under Contract DE-AC05-000R22725. Computing time was provided by the Georgia Advanced Computing Resource Center at the University of Georgia.

References
[1] Goodsell D S 1991 Trends Biochem. Sci. 16 203–6
[2] Zimmerman S and Trach S 1991 J. Mol. Biol. 222 599–620
[3] Zimmerman S and Minton A 1993 Annu. Rev. Biophys. Biomol. Struct. 22 27–65
[4] Ellis R 2001 Trends Biochem. Sci. 26 597–604
[5] Taubes G 1996 Science 271 1493–95
[6] Sewell R 2007 Protein Misfolding in Neurodegenerative Diseases: Mechanisms and Therapeutic Strategies Enzyme Inhibitors Series (Taylor & Francis)
[7] Minton A P 2000 Curr. Opin. Struct. Biol. 10 34–9
[8] Ellis R J 2001 Curr. Opin. Struct. Biol. 11 114–9
[9] Lucent D, Vishal V and Pande V S 2007 Proc. Natl. Acad. Sci. U.S.A. 104 10430–4
[10] Ruckenstein E and Byungip Chung D 1988 J. Colloid Interface Sci. 123 170–85
[11] Martins D C, Chu V and Conde J P 2013 Biomicrofluidics 7 034111
[12] Sang J, Du H, Wang W, Chu M, Wang Y, Li H, Zhang H A, Wu W and Li Z 2013 Biomicrofluidics 7 024112
[13] Griesshauer D, MacKenzie R, Vörös J and Reimhult E 2008 Sensors 8 1400–58
[14] Kim S J and Han J 2008 Anal. Chem. 80 3507–11
[15] Duan C, Wang W and Xie Q 2013 Biomicrofluidics 7 026501
[16] Dill K A 1985 Biochemistry 24 1501–9
[17] Dill K 1990 Science 250 297–8
[18] Berger B and Leighton T 1998 J. Comp. Biol. 5 27–40
[19] Crescenzi P, Goldman D, Papadimitriou C, Piccolboni A and Yannakakis M 1998 J. Comp. Biol. 5 423–65
[20] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 J. Chem. Phys 21 1087
[21] Wang F and Landau D 2001 Phys. Rev. Lett. 86(10) 2050–2053
[22] Landau D P and Wang F 2002 Comput. Phys. Commun. 147 674–7
[23] Landau D P, Tsai S H and Exler M 2004 Am J Phys 72 1294–1302
[24] Lesh N, Mitzenmacher M and Whitesides S 2003 A complete and effective move set for simplified protein folding Proc. 7th Annual Inter. Conf. on Research in Computational Molecular Biology (RECOMB) (Berlin: ACM Press) pp 188–95
[25] Deutsch J M 1997 J. Chem. Phys 106 8849–54
[26] Wüst T and Landau D 2008 Comput. Phys. Commun. 179 124–7
[27] Wüst T and Landau D P 2009 Phys. Rev. Lett. 102 178101
[28] Berg B A and Neuhaus T 1992 Phys. Rev. Lett. 68 9–12
[29] Yue K, Fiebig K M, Thomas P D, Chan H S, Shakhnovich E I and Dill K A 1995 Proc. Natl. Acad. Sci. U.S.A. 92 325–9
[30] Li Y W, Wüst T and Landau D P 2013 Phys. Rev. E 87 012706
[31] Pattanasiri B, Li Y W, Landau D P and Wüst T 2012 Int. J. Mod. Phys. C 23 1240008
[32] Pattanasiri B, Li Y W, Landau D P, Wüst T and Triampo W 2012 J. Phys.: Conf. Series 402 012048
[33] Pattanasiri B, Li Y W, Landau D P, Wüst T and Triampo W 2013 J. Phys.: Conf. Series 454 012071