Crambin Homologues in the H0P Lattice Model

Author(s):
Zhang, Zewen; Farris, Alfred C.K.; Shi, Guangjie; Wüst, Thomas; Landau, David P.

Publication date:
2019

Permanent link:
https://doi.org/10.3929/ethz-b-000383678

Originally published in:
Journal of Physics: Conference Series 1290(1), https://doi.org/10.1088/1742-6596/1290/1/012018

Crambin Homologues in the H0P Lattice Model

IOP Conf. Series: Journal of Physics: Conf. Series **1290** (2019) 012018 doi:10.1088/1742-6596/1290/1/012018

Crambin Homologues in the H0P Lattice Model

Zewen Zhang,[1*] Alfred C.K. Farris,[1] Guangjie Shi,[1] Thomas Wüst,[2] David P. Landau[1]

[1]Center for Simulational Physics, University of Georgia, Athens, GA 30602, USA
[2]Scientific IT Services, ETH Zürich, 8092 Zürich, Switzerland

E-mail: *zewenzhang@physast.uga.edu

Abstract. To compare folding behavior among lattice proteins which have similar corresponding structures in nature, Crambin homologues are tested in the semi-flexible H0P lattice model using replica-exchange Wang-Landau sampling. Our simulation shows that, at low temperature, these lattice homologues have two common signals in their specific heat curves, implying similarity in the thermodynamic behaviors; while the structural behaviors are more diverse, showing the different stability of their ground state structures at very low temperature. The ground state structures of different homologues can also vary dramatically.

1. Introduction

The problem of protein folding has been studies for over 80 years. Progress in simulations, as well as experimental and theoretical works, have deepened our understanding of the folding pathway[1], structural prediction[2] and molecular biological function[3]. However, it is still not clear why the folding process is so cooperative[4, 5, 6]. To solve this puzzle, it is usually more important to make use of some coarse-grained models than to attempt to use complex atomic details[7, 8, 9]. A representative model is the original hydrophobic-polar (HP) model[10], in which 20 amino acids are condensed into 2 different categories: hydrophobic (H) and polar (P). In this model, hydrophobic interactions among amino acids are considered as the key driving force during the folding process. The further simplification of the HP model is that it places monomers on lattice sites, and only accounts for interactions between nearest neighbor non-bonded HH contacts. This model turns out to be very helpful to our understanding of the protein folding problem, giving descriptions about collapse transition[11], structural stability[12, 13], cooperativity[14], etc.

Experimental evidence indicates that proteins with over 40% sequence identity often have the same native structures. Correspondingly, some previous research on single-site mutation in HP model[13] are partially consistent with these experimental results. However, even in the H and P "degeneracy" (which have 10 kinds of amino acids for either H or P category), the differences between two sequences can still be much larger than a single site. Do homologues with a difference of more than two monomers still have stable native states in the lattice model? In this paper we test five homologues of Crambin to explore this question.

Meanwhile, to overcome high degeneracies of ground states in the original HP model[15, 16], we introduced a third kind of monomer, "neutral" (0) and a stiffness energy into our lattice protein model. In the resulting H0P model, energy ranges of lattice proteins are usually wider

than those in the HP model. To keep the sampling efficient, here we apply replica-exchange Wang-Landau (REWL)[17] sampling to the simulation. We also extract some structural behaviors from replica-exchange multicanonical (REMUCA) sampling[18].

2. Models and Simulation Methods

2.1. Semi-flexible H0P Lattice Protein Model

In consideration of the important role that hydrophobic driving force plays in folding of globular proteins, the original HP model is designed to reduce the complexity of atomic details in a protein sequence, allowing for exploring the full conformational space extensively and efficiently[19]. Previous research indicates that HP model leads to high degeneracy in native states[13, 15, 20]. Our semi-flexible H0P model shows great improvement on reducing such degeneracy, by adding more detail on the monomeric scale. The model classifies 20 amino acids into three different types of "beads", based on their hydrophobicity levels, and all "beads", or monomers, in a chain are put on different sites of a simple cubic lattice. In this H0P polymer system, only contacts between two non-bonded nearest neighbors are taken into consideration. While in principle there could be interactions between all types of monomers, here we only choose HH contacts and H0 contacts, which are supposed to be the major contributions to free energy in globular polymers, for calculation of system energy. We also add an energetic term for any "bend" in the protein structure to show the internal rigidity of polypeptide chain[22]. Therefore, the Hamiltonian of our semi-flexible H0P lattice model is given by[15]:

$$\mathcal{H} = -\epsilon_{HH} n_{HH} - \epsilon_{H0} n_{H0} - \epsilon_{\theta} n_{\theta}, \tag{1}$$

where n_{HH} is the number of HH contacts, n_{H0} is the number of H0 contacts and n_{θ} is the number of bends; ϵ_{HH} is the HH contact energy, ϵ_{H0} is the H0 contact energy and ϵ_{θ} is the bending energy.

2.2. Crambin Homologues

Crambin is a short protein which is often used to test simulation algorithms[21]. It has also been mapped and studied in the lattice protein model[15, 16, 23]. Meanwhile, sequences and structures of Crambin and its homologues (small plant toxins, such as purothionin and viscotoxin) have long been known by experiment[24, 25]. Therefore, these small proteins are an ideal starting point for us to look into similarities among homologues in the H0P model.

In this paper, one purothionin (PDB entry *1bhp*), one hordothionin (*1wuw*) and three viscotoxins (*1orl*, *1jmn* and *3c8p*) are chosen to be tested as Crambin homologues. Their sequences are mapped into an H0P form, as shown in **Table 1**. Although there are only three kinds of monomers in the H0P representation, these homologues still have significant differences in their sequences. However, comparing to their differences in real sequences, which are around 50%, they are more "alike" in the H0P model.

2.3. Replica-exchange Wang-Landau Sampling

Traditional Monte Carlo methods such as Metropolis sampling are sometimes inefficient at low temperature; for systems with complex free energy landscapes like lattice proteins, these sampling methods are likely to be trapped at some local minima. Instead, Wang-Landau sampling[26] is used to get around these difficulties.

The partition function (Z) at a certain temperature T can be derived from a summation over the energy space:

$$Z(T) = \sum_{\{E_i\}} g(E_i) \exp(-E_i/k_B T), \tag{2}$$

Table 1. Sequences of Crambin homologues in the H0P Model. In the sequences of homologues, monomers different from the corresponding sites in Crambin sequence are marked in red. The dash (-) here represents for one missing monomer, which helps to align the sequences.

Protein	**Sequence in H0P Model**									
Crambin	PPHH0	0HHHP	0PHPH	HPH00	00P0H	H0000	0HHHH	0000H	00P00	P
1bhp	P0HHP	00H0P	PH0PH	HP0P0	0PP-H	H0PHH	PHPH0	00H0H	0PPH0	P
1wuw	P0HHP	00H0P	PH0PH	HPHP0	0PP-H	H0P0H	PHPH0	00HPH	000H0	P
1orl	P0HH0	P000P	PH0P0	HPH00	00PPP	H0PH0	0HPHH	0000H	00P00	P
1jmn	P0HH0	P000P	PH0P0	HPH00	00PPH	H00H0	0HPHH	0000H	00P00	P
3c8p	P0HH0	0000P	PH0P0	HPH00	00PP0	H0PH0	0HPHH	0000H	00P00	P

where g(E) is the temperature-independent density of states (DoS) at a certain energy (E), and k_B is the Boltzmann constant. Wang-Landau sampling provides an unbiased estimation of g(E) by iteratively performing a random walk in energy space. During simulation, the acceptance probability of one MC trial move relating the current configuration A and proposed configuration B is given by:

$$P(A \to B) = min\{1, \frac{g'(E_A)}{g'(E_B)}\}, \tag{3}$$

where $g'(E_A)$ and $g'(E_B)$ are current estimators of the DoS for energy E_A and E_B, respectively. After a trial move, the DoS for the accepted energy level E_n will be updated according to $g'(E_n) \to f \times g'(E_n)$, where f is a modification factor. A histogram H(E) is also kept during the simulation to count all the energy levels visited. After every trial move, it is updated according to $H(E_n) \to H(E_n) + 1$. For a preset number of trial moves, the H(E) is checked for its flatness. If it is flat enough, which indicates all energy levels are almost evenly visited, the modification factor is then adjusted to a lower value such as $\sqrt{f}$, and the histogram is reset to zero.

The replica-exchange Wang-Landau algorithm improves the efficiency of WL sampling by taking advantage of parallel computing[17]. It divides the whole energy range into multiple overlapping energy windows; in each window, there are one or more walkers sampling independently. Replica-exchanges between two walkers in the overlapping area of two neighboring windows i and j are proposed regularly at fixed intervals, with the acceptance probability given as:

$$P_{acc} = min\{1, \frac{g'_i(E_A)}{g'_i(E_B)} \frac{g'_j(E_B)}{g'_j(E_A)}\}, \tag{4}$$

where $g'_k(E_X)$ donates the estimator for the density of states from walker k in configuration X. In this way, REWL allows us to efficiently search the whole complex energy landscapes with consistency among all overlapping energy windows. To further improve the efficiency of sampling, special trial moves are applied in this research[27, 28].

3. Results

3.1. Degeneracy

In this research, the values of contact and bending energy ($\epsilon_\theta = -0.1\epsilon_{H0} = -0.05\epsilon_{HH}$) are chosen specifically to keep the degeneracy of Crambin native states low[16, 22]. With this set of parameters, degeneracies of native states and some low lying excited states of all homologues are searched for with multicanonical sampling. As listed in **Table 2**, the degeneracies of all native states are low, which supports that the parameters chosen here are suitable for simulating this family of lattice proteins. However, some homologues still show high degeneracy in their low lying excited states.

Table 2. Degeneracy of ground states and low excited states. g(E_i) represents the DoS of the i-th excited state from the ground state. E_0 is the ground state. 0 in g(E_2) of Crambin indicates no state has been found at this energy level, which might be a gap in the energy spectrum.

Energy Level	Crambin	*1bhp*	*1wuw*	*1orl*	*1jmn*	*3c8p*
g(E_0)	1	3	2	3	2	5
g(E_1)	6	26	10	24	24	47
g(E_2)	0	84	17	133	118	362
g(E_3)	6	119	24	633	358	1810

3.2. Specific Heat

Once the DoS is well estimated from REWL, we can directly extract the specific heat of the tested lattice protein systems. Specific heat curves of three homologues are plotted in **Fig.1** as examples. For all tested homologues as well as Crambin itself, their specific heat curves show two signals at high temperature (peaks and shoulders), corresponding to the coil-globule transition and folding transition. Meanwhile, at low temperature, there are also two signals at fixed temperature (T = 0.02 and T = 0.05), which indicate that all these homologues start rearranging their structures at the same temperature. Therefore, in the H0P lattice model, these homologues show similar features in their thermodynamic behaviors.

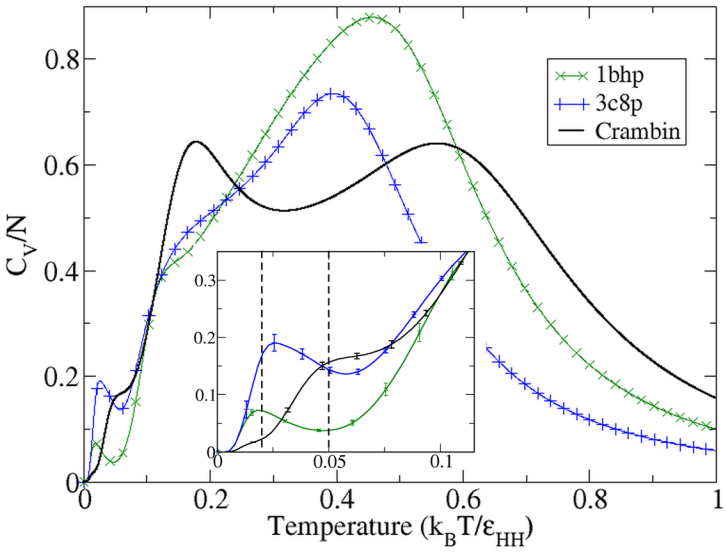

Figure 1: Specific Heat (C_V) curves for Crambin Homologues

Low temperature region is magnified in the bottom. For clarity, only three of the C_V curves for six tested proteins are shown in the graph. Two dashed lines here help to point out signals at T=0.02 and T=0.05. Error bars smaller than the size of the points are not shown.

3.3. Radius of Gyration

The radius of gyration for different homologues is also measured. Their first derivatives are plotted in **Fig.2**. At low temperature, some of the homologues, such as *1wuw*, still keep two signals at T = 0.02 and T = 0.05; however, other homologues, such as *1orl*, do not show signals at these temperatures. Therefore, the structural behaviors of these homologues at low temperature region are dissimilar to each other in the H0P model.

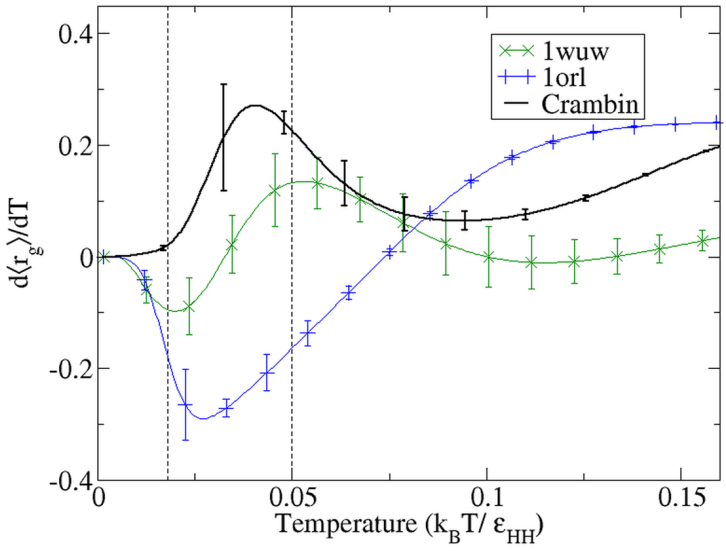

Figure 2: First detivative of the average radius of gyration $\langle r_g \rangle$ for Crambin Homologues

For clarity, only three of the $\langle r_g \rangle$ curves for six tested proteins are shown in the graph. Two dashed lines here help to point out signals at T=0.02 and T=0.05. Error bars smaller than the size of the points are not shown.

3.4. Native Structures

All of the tested lattice proteins in this paper have a small number of ground states, as shown in 3.1. Therefore, it is possible to compare the contact maps of their ground state structures. The results show that, even though their structures in nature are similar[25], in the H0P model the similarity between their native structures are mostly low. The identity between the native structures of different homologues are mostly less than 50%. There is only one case where two homologues have same native structures, as shown in **Fig.3**. In this example, the monomers that are different in these two sequences are all on the surface of the globular structures, which agrees with predictions[7].

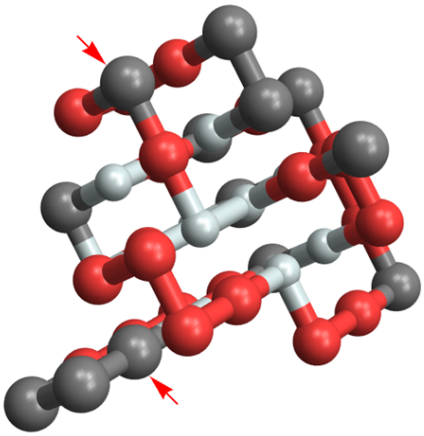

Figure 3.1 Native structure of *1orl*

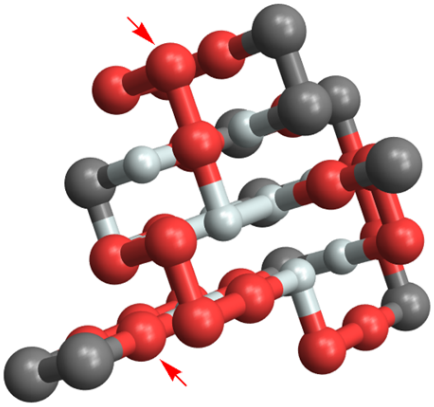

Figure 3.2 Native structure of *3c8p*

The native states for *1orl* and *3c8p* have same ground state structures. Different monomers in sequences are indicated by red arrows. White represents H monomers; red represents 0 monomers; grey represents P monomers.

4. Conclusion

With the REWL sampling method, we tested six similar lattice proteins in the semi-flexible H0P lattice model. The results show that their thermodynamic behaviors are similar, with two signals in the low temperature region. These common signals indicate that the native structures of these lattice homologues start "melting down" at the same temperature. However, their structural behaviors have very little in common. Even though the same structures are observed for different sequences in one case, most of the native structures of these short homologues have very limited similarities in the H0P lattice model, which implies the impact of mutations is much stronger in the H0P model than it in the 20 amino acid model.

Acknowledgements

The computing sources of this work is provided by Georgia Advanced Computing Resource Center, a partnership between the University of Georgias Office of the Vice President for Research and Office of the Vice President for Information Technology; and Texas Advanced Computing Center under XSEDE Grant No. PHY130014.

References

[1] Wallin S and Shakhnovich E I 2008 *J. Phys.: Condens. Matter.* **20** 283101
[2] Hoque T, Chetty M and Sattar A 2009 *J. Comput. Biol.* **16** 85-103
[3] Koga N, Tatsumi-Koga R, Liu G, Xiao R Acton T B, Montelione G T and Baker D 2012 *Nature.* **491** 222
[4] Davidson A R 2008 *Proc. Natl. Acad. Sci.* **105** 2759-2760
[5] Lammert H, Wolynes P G and Onuchic J N 2012 *Proteins* **80** 362-373
[6] Muñoz V, Campos L A and Sadqi M 2016 *Curr. Opin. Struct. Biol.* **36** 58-66
[7] Dill K A 1999 *Protein Sci.* **8** 1166-1180
[8] Kolinski A and Skolnick J 2004 *Polymer* **45** 511-524
[9] Clementi C 2008 *Curr. Opin. Struct. Biol.* **18** 10-15
[10] Dill K A 1985 *BioChemistry* **24** 1501-1509
[11] Shakhnovich E I 1997 *Curr Opin Struct Biol.* **7** 29-40
[12] Shortle D, Chan H S and Dill K A 1992 *Protein Sci.* **1** 201-215
[13] Shi G, Vogel V, Wüst T, Li Y W and Landau D P 2014 *Phys. Rev.* E **90** 033307
[14] Chan H S, Zhang Z, Wallin S and Liu Z 2011 *Annu. Rev. Phys. Chem.* **62** 301-326
[15] Shi G, Farris A CK, Wüst T and Landau D P 2016 *J. Phys. Conf. Ser.* **686** 012001
[16] Farris A CK, Shi G, Wüst T and Landau D P 2018 *J. Phys. Conf. Ser.* **1012** 012008
[17] Li Y W, Vogel T, Wüst T and Landau D P 2014 *J. Phys. Conf. Ser.* **510** 012012
[18] Sugita Y and Okamoto Y 2000 *Chem. Phys. Lett.* **329** 261-270
[19] Dubey S P, Kini N G, Balaji S and Kumar M S 2018 *Crit. Rev. Biomed. Eng.* **46**
[20] Yue K, Fiebig K M, Thomas P D, Chan H S, Shakhnovich E I and Dill K A 1995 *Proc. Natl. Acad. Sci.* **92** 325-329
[21] Shi G, Wüst T, Li Y W and Landau D P 2015 *J. Phys. Conf. Ser.* **640** 012017
[22] Farris A CK, Shi G, Wüst T and Landau D P 2018 *J. Chem. Phys.* **149** 125101
[23] Dill K A, Fiebig K M and Chan H S 1993 *Proc. Natl. Acad. Sci.* **90** 1942-1946
[24] Debreczeni J É, Girmann B, Zeeck A, Krätzner R and Sheldrick G M 2003 *Acta. Crystallogr. D Biol. Crystallogr.* **59** 2125-2132
[25] Schmidt A, Teeter M, Weckert E and Lamzin V S 2011 *Acta. Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67** 424-428
[26] Landau D P, Tsai S H and Exler M 2004 *Am. J. Phys.* **72** 1294-1302
[27] Wüst T and Landau D P 2009 *Phys. Rev. Lett.* **102** 178101
[28] Wüst T and Landau D P 2012 *J. Chem. Phys.* **137** 064903