

# Chain-growth simulations of the HP model for proteins

Fabricio S H Santos<sup>1</sup>, Paulo H L Martins<sup>1</sup>, and Michael Bachmann<sup>1,2,3</sup>

<sup>1</sup> Instituto de Física, Universidade Federal de Mato Grosso, Av. Fernando Corrêa da Costa, 2367, 78060-900, Cuiabá, MT, Brazil

<sup>2</sup> Soft Matter Systems Research Group, Center for Simulational Physics, Department of Physics and Astronomy, The University of Georgia, Athens, Georgia 30602, USA

<sup>3</sup> Departamento de Física, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, C.P. 702, 30123-970, Belo Horizonte, MG, Brazil

E-mail: [fabricio@fisica.ufmt.br](mailto:fabricio@fisica.ufmt.br), [pmartins@fisica.ufmt.br](mailto:pmartins@fisica.ufmt.br), [bachmann@smsyslab.org](mailto:bachmann@smsyslab.org)

**Abstract.** We consider a generalized hydrophobic-polar model for proteins on square lattices. Besides the attraction between nonbonded hydrophobic monomers, the present model also takes into account an interaction between hydrophobic and polar units. By using the pruned-enriched Rosenbluth method (PERM), we investigate a specific polymer sequence composed of 42 monomers that has been proposed to simulate the physical properties of the parallel  $\beta$ -helix of pectate lyase *C*. For each temperature, the total number of generated chains varies from  $10^6$  to  $10^7$ . Physical observables such as specific heat, total energy, end-to-end distance, radius of gyration, and the average number of hydrophobic-hydrophobic and hydrophobic-polar contacts are evaluated for different values of the ratio between the hydrophobic-hydrophobic and the hydrophobic-polar contact energies. Eventually, a pseudo-phase diagram in the space of temperature and the ratio of contact energy scales is constructed.

## 1. Introduction

Proteins are responsible for a wide range of biological functions, including transport and storage of molecules, coordinated movements, enzymatic catalysis, mechanical support, and others [1]. A protein is a linear molecule composed of a particular sequence of monomer units (amino acids) [2]. Since there exist 20 different amino acids in nature, a certain protein with  $N$  monomers is only one among  $20^N$  possible polypeptides with this chain length [3]. The complexity of the system arises from this huge number of combinations. Due to the enormous number of tasks to be necessarily fulfilled to ensure the stability of a biological system, a large number of proteins exists [4], although it is vanishingly small compared to the number of possible sequences.

The biological function of a protein chain is closely related to its sequence and, hence, to its native conformation [5]. In this sense, it is very useful to identify the structure which a specific sequence will assume according to its environmental conditions. It is well known that, as the temperature increases, a heteropolymer chain undergoes changes in its shape, from the native fold to globules, and random coils [6].

In this work, we study a simple two-dimensional lattice model for proteins and systematically investigate the dependence of the structural behavior on the environmental temperature and the ratio of energy scales associated with contacts between nonbonded hydrophobic monomers and



mixed hydrophobic-polar contacts. For this purpose, we employ a chain-growth algorithm for the simulation of this system. We calculate quantities like the specific heat, mean energy, end-to-end distance, and radius of gyration. Different temperatures and solvent conditions are considered to determine the phase diagram of a specific sequence composed by 42 monomers [7, 8, 9].

## 2. Model and Method

Despite the variety of monomer units, simplified models have been proposed in order to better understand the conformational transitions of a polymer chain. A linear chain of identical monomers is called a homopolymer, whereas heteropolymers are composed of different units [7, 10].

The hydrophobic-polar (*HP*) model is one of the simplest lattice models used to describe, very qualitatively, the folding behavior of proteins [7]. It considers two types of monomers, hydrophobic (*H*) and polar (*P*), and the conformations are modeled as interacting self-avoiding walks on the lattice [11]. It also assumes that the hydrophobic interaction is the essential driving force towards the native fold [12]. The chain studied here is a 42mer with the sequence:  $PH_2PHPH_2PHPH_2H_3PHPH_2PHPH_3P_2HHPHPH_2PHPH_2P$ , which was designed to simulate some properties of the parallel  $\beta$ -helix of *pectate lyase C* [13]. The monomers can occupy sites on a two-dimensional square lattice [14]. Adjacent monomers in the polymer sequence have a fixed unity bond length (one lattice unit). Detailed studies of this sequence on three-dimensional lattices, including only hydrophobic-hydrophobic (*HH*) interactions, have been performed in the past (see, e.g., Ref. [4] and references therein). Here, besides the *HH* attractive interaction, we also consider a hydrophobic-polar (*HP*) attraction. Therefore, the total energy of the system reads:

$$E = -\epsilon_{HH}n_{HH} - \epsilon_{HP}n_{HP} , \quad (1)$$

where  $n_{HH}$  and  $n_{HP}$  are the total number of non-bonded *HH* and *HP* pairs of neighboring monomers, respectively. Defining  $s \equiv \epsilon_{HP}/\epsilon_{HH}$  and setting  $\epsilon_{HH} = 1$ , one has

$$E = -(n_{HH} + s n_{HP}) . \quad (2)$$

Note that the limit  $s = 0$  corresponds to the original *HP* model, with *HH* contacts only. The case  $s = 1$  means that *HH* and *HP* contacts have equal strength.

We have performed chain-growth simulations based on the *pruned-enriched Rosenbluth method* (PERM) [6]. This algorithm has proven to be particularly useful for studying the  $\theta$  collapse transition between random-coil and globular structures of homopolymers, but it has also been successfully applied for lattice heteropolymers. It uses the *go with the winners* strategy in the sense that it favors chains with small entropy as reflected by large Rosenbluth weights, which in PERM are introduced as [6]:

$$W_n^{PERM} = \prod_{l=2}^n m_l \exp^{-(E_l - E_{l-1})/k_B T} . \quad (3)$$

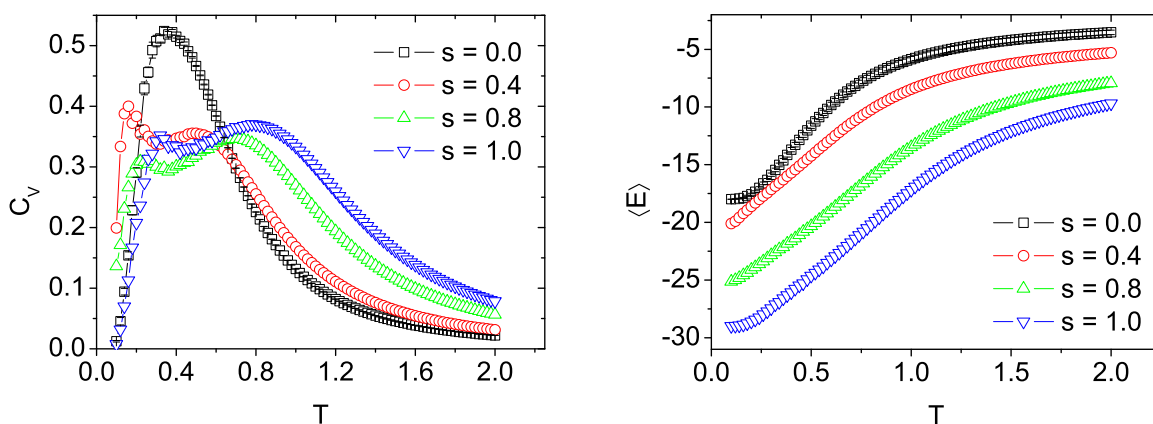
Here,  $n$  is the total length of the chain,  $E_l$  is the energy of the partial chain with size  $l$ ,  $k_B$  is the Boltzmann constant, and  $T$  is the temperature. The average value of a physical observable  $Q$  can be calculated through

$$\langle Q \rangle = \lim_{M \rightarrow \infty} \frac{\sum_{\alpha=1}^M Q_\alpha W_{n,\alpha}^{PERM}}{\sum_{\alpha=1}^M W_{n,\alpha}^{PERM}} , \quad (4)$$

where  $M$  is the number of generated chains,  $Q_\alpha$  is the value of quantity  $Q$  for the chain  $\alpha$ , and  $W_{n,\alpha}^{PERM}$  is the weight associated with that chain. For the subsequent discussion of structural phases in our model, the most interesting quantities are the specific heat and the average number of hydrophobic-hydrophobic and hydrophobic-polar contacts.

### 3. Results

Considering the *HP* chain embedded in an implicit solvent, we have simulated and analyzed the structural behavior of the heteropolymer for several values of *HP/HH* energy scale ratio  $s$  in the range  $0 \leq s \leq 1$ . Temperatures ranged from 0.1 to 2.0. For given  $(s, T)$  parameters, the total number of generated chains varied from  $10^6$  to  $10^7$ . Figure 1 shows, on the left side, the specific heat  $C_V$  as a function of temperature  $T$  for selected values of  $s$ . One can note that, for  $s = 0.0$ , a single peak in the low-temperature region is observed, while for  $s \geq 0.4$ , two distinct peaks appear, corresponding to two conformational transitions. The mean energy  $\langle E \rangle$  is plotted for the same  $s$  values in Fig. 1 (right). As expected, as the temperature increases, the energy approaches zero, indicating that the system has reached the random coil phase.

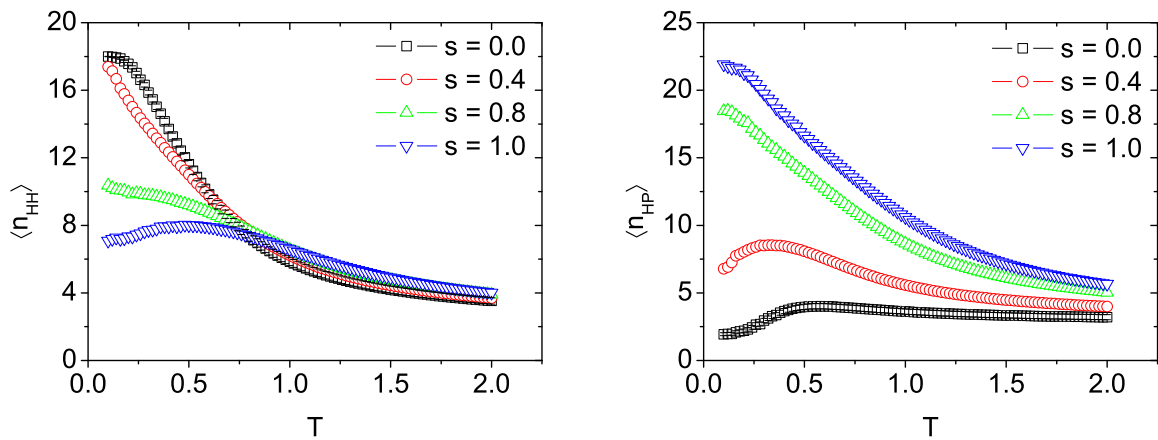


**Figure 1.** Specific heat (*left*) and mean energy (*right*) as functions of temperature for selected values of the parameter  $s$ . Error bars, if not visible, are smaller than the symbols.

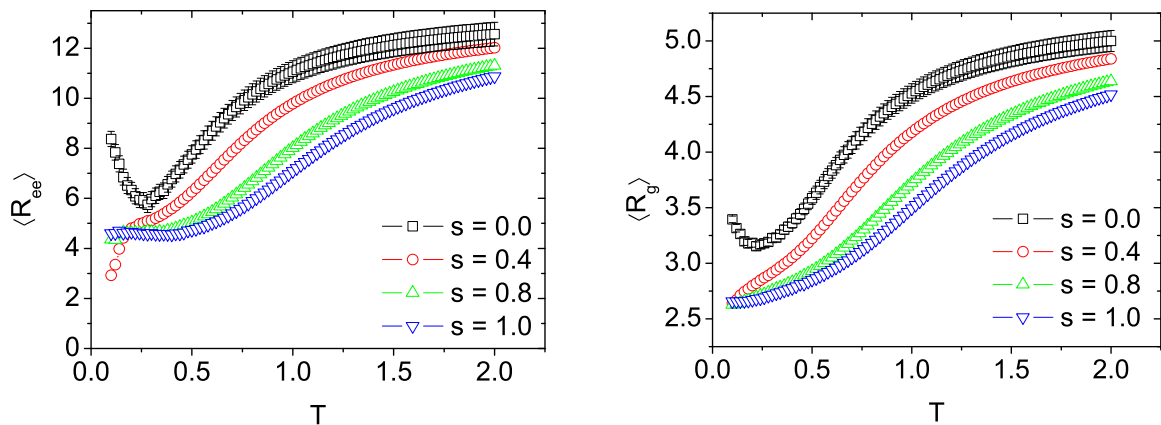
In Fig. 2, we show the average number of *HH* and *HP* contacts ( $n_{HH}$  and  $n_{HP}$ , respectively). For small values of  $s$  (0.0 and 0.4, for instance), one can observe that the low-temperature conformations possess a significant number of *HH* contacts, while the number of *HP* contacts is small. As the parameter  $s$  approaches 1.0 (this limit means that *HH* and *HP* interactions are equal),  $n_{HP}$  gets larger while  $n_{HH}$  decreases.

To have an idea of how conformations look like under different thermal conditions, it is helpful to analyze structural quantities such as end-to-end distance,  $R_{ee}$ , and radius of gyration,  $R_g$ . Mean values of both quantities are shown in Fig. 3. An interesting behavior can be noticed for  $s = 0.0$ . In this case, the lowest energy we have found is -19. In contrast with the three-dimensional case [15], the 2D ground state seems to be highly degenerated with rather large variations of compactness. To maximize  $n_{HH}$  (and consequently minimize the total energy), the polymer forms an *H*-core that also includes some *P* monomers in a compact shape. Other ground-state conformations possess a hydrophobic backbone surrounded by polar monomers, but these structures have a much larger radius of gyration. As the temperature increases,  $\langle R_{ee} \rangle$  and  $\langle R_g \rangle$  decrease and reach a minimum value at  $T \approx 0.25$ . For higher values of  $s$  (especially for  $s \geq 0.3$ ), as the *HP* contacts become more relevant, this behavior is not observed.

Figure 4 summarizes our results. It depicts the phase diagram in the  $s - T$  space and includes some typical conformations in each region of the phase diagram. The transition points were obtained from the peak positions of the specific heat. In the low-temperature region, we find compact conformations dominating for large values of  $s$ , whereas conformations are less



**Figure 2.** *Left:* average number of hydrophobic-hydrophobic contacts  $\langle n_{HH} \rangle$ . *Right:* average number of hydrophobic-polar contacts  $\langle n_{HP} \rangle$ . Error bars are of the order of the symbols size.

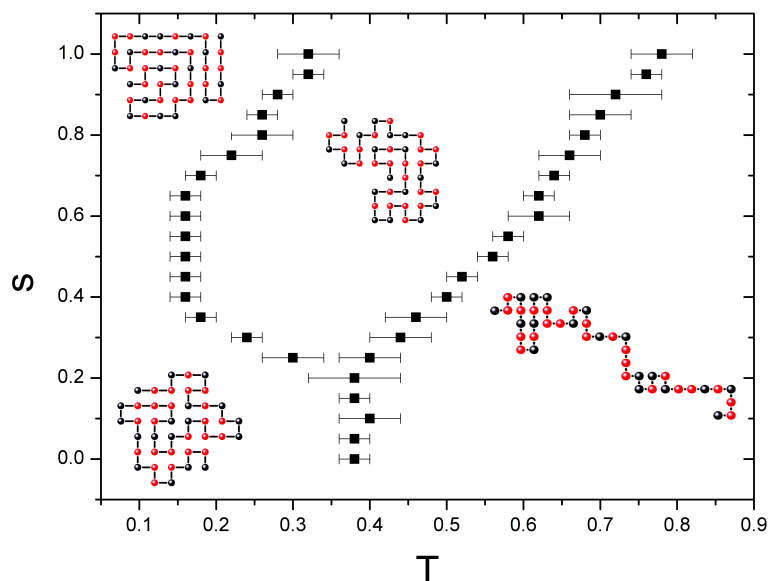


**Figure 3.** Average values of the end-to-end distance (*left*) and the radius of gyration (*right*). For better visualization, error bars are shown for  $s = 0.0$  only. For the other values of  $s$ , error bars have similar sizes.

compact for small values of  $s$ , in which case  $HH$  contacts are favored. For a fixed value of  $s$  ( $s \lesssim 0.20$ ) and increasing temperature, random coil structures form at  $T \approx 0.38$ , characterized by large values of the end-to-end distance and the radius of gyration. For  $s \gtrsim 0.25$ , there is an intermediate globular phase, in which both  $HH$  and  $HP$  contacts are relevant.

#### 4. Conclusion

By using the pruned-enriched Rosenbluth method (PERM), we have analyzed a generalized version of the  $HP$  model for proteins on square lattices, including both hydrophobic-hydrophobic and hydrophobic-polar interactions. Several values of  $s$  (the strength of the  $HP$  attraction, relative to the  $HH$  interaction) were considered and results compared. The maximum values of the specific heat of a particular 42mer were used to determine the pseudophase diagram in the



**Figure 4.** Pseudo-phase diagram in  $s - T$  space. In the conformations shown, an  $H$  monomer is represented by a red bead while  $P$  monomers are black.

$s - T$  plane. The most interesting result of our study is that for  $HP/HH$  energy scale ratios  $s \geq 0.2$ , an intermediate phase of compact, but less ordered conformations forms (“liquid”). Studies of the present model in three dimensions are in progress.

### Acknowledgments

Financial support was provided by FAPEMAT, CAPES, and CNPq (Grant No. 402091/2012-4). MB also acknowledges support by the NSF under Grant No. DMR-1207437.

### References

- [1] Stryer L 1995 *Biochemistry* (New York: W. H. Freeman & Company)
- [2] Lau K F and Dill K A 1989 *Macromolecules* **22** 3986
- [3] Dill K A, Bromberg S, Yue K, Fiebig K M, Yee D P, Thomas P D, and Chan H S 1989 *Protein Sci.* **4** 561
- [4] Bachmann M 2014 *Thermodynamics and Statistical Mechanics of Macromolecular Systems* (Cambridge: Cambridge University Press)
- [5] Branden C and Tooze J 1991 *Introduction to Protein Structure* (New York: Garland)
- [6] Grassberger P 1997 *Phys. Rev. E* **56** 3682
- [7] Dill K A 1985 *Biochemistry* **24** 1501
- [8] Bachmann M and Janke W 2003 *Phys. Rev. Lett.* **91** 208105
- [9] Bachmann M and Janke W 2004 *J. Chem. Phys.* **120** 6779
- [10] Flory P J 1971 *Principles of Polymer Chemistry* (Ithaca: Cornell University Press)
- [11] Chan H S and Dill K A 1993 *Phys. Today* **46** 24
- [12] Dill K A 1990 *Biochemistry* **29** 31
- [13] Yoder M D, Keen N T, and Jurnak F 1993 *Science* **260** 1503
- [14] Orr W J C 1947 *Trans. Faraday Soc.* **43** 12
- [15] Bachmann M and Janke W 2003 *Acta Phys. Pol. B* **34** 4689