Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 350 (2005) 45–51
PHYSICA A
www.elsevier.com/locate/physa

Partition function zeros of the two-dimensional HP model for protein folding

Chi-Ning Chen[a,*], Chai-Yu Lin[b,c]

[a]*Department of Physics, National Dong-Hwa University, Hualien, Taiwan, ROC*
[b]*Department of Physics, National Chung Cheng University, 160 Shang-Shing, Min-Hsiung Hsiang, Chia-Yi 66117, Taiwan, ROC*
[c]*National Center for Theoretical Sciences at Taipei, Physics Division, National Taiwan University, Taipei 10617, Taiwan, ROC*

Available online 9 December 2004

Abstract

The partition function of the two-dimensional lattice HP model for protein folding is computed by exact enumeration. For a protein-like sequence, the distribution of partiton function zeros shows roughly a two-ring pattern, while for a nonprotein-like sequence, the outer ring of zeros is ill-developed and cannot induce the folding transition. By tracing the peak or shoulder of the heat capacity in the complex plane, the phase boundaries can be determined along the positive real axis.

PACS: 05.20.−y; 87.15.Cc

Keywords: Partition function zeros; HP model; ISAW; Protein folding

Linear polymers in good solvents undergo a collapse transition from elongated coils to compact globules when temperature is lowered. The collapse transition point is called the Theta point [1]. One simple model for such phenomena is the interacting self-avoiding walk (ISAW) on regular lattices. In this model, besides the effect of excluded volume, there is also attraction between monomers if they are the nearest

*Corresponding author.
E-mail addresses: cnchen@mail.ndhu.edu.tw (C.-N. Chen), lincy@phy.ccu.edu.tw (C.-Y. Lin).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2004.11.035

neighbors but are not connected by the walk. Proteins are linear heteropolymers composed of 20 different types of monomers (amino acids). The heterogeneity of protein is important and necessary for specific folding leading to an unique ground state, i.e., the native state [2]. Pure homopolymer models such as ISAW cannot serve as protein models since their ground states are highly degenerate.

The HP model is a generalized ISAW with two types of monomers, H and P [3,4]. Here H represents a hydrophobic amino acid, and P polar or hydrophilic animo acid. The hydrophobic interaction is considered as one of the major driving forces in protein folding [5,6]. The Hamiltonian of the HP model is written as

$$H = \sum_{1 \leqslant i < j \leqslant N} \Delta_{ij}\, \varepsilon_{\sigma_i \sigma_j}\,, \tag{1}$$

where N is the number of monomers, $\Delta_{ij} = 1$, if monomer i and monomer j have a nonbonding contact, and 0 otherwise, and σ_i and σ_j are the types of monomer i and monomer j (H or P). There are three kinds of $\varepsilon_{\sigma_i \sigma_j}$, namely ε_{HH}, ε_{HP} $(= \varepsilon_{PH})$ and ε_{PP}, representing the H–H, H–P and P–P interaction energy, respectively. They satisfy the following two relations: (1) $\varepsilon_{HH} < \varepsilon_{HP} < \varepsilon_{PP}$ because H monomers tend to hide from water and stay in the core of a structure; (2) $\varepsilon_{HH} + \varepsilon_{PP} < 2\varepsilon_{HP}$ because monomers of different types tend to segregate. From the MJ matrix whose elements are statistical inter-residue contact energies [7], relative magnitudes of these three energies can be estimated as $\varepsilon_{HH} : \varepsilon_{HP} : \varepsilon_{PP} \approx -3.3 : -2 : -1$ [8]. For simplicity of computation, we adopt integer energies in this paper: $\varepsilon_{HH} = -4$, $\varepsilon_{HP} = -2$ and $\varepsilon_{PP} = -1$, without altering main features of the HP model. The case of noninteger energies can be resolved by dividing the energy range into intervals of a fixed bin size.

Now, the partition function of the HP model can be expressed as a polynomial

$$Z = \sum_{\text{all structures}} \mathrm{e}^{-E/T} = \sum_{E=E_0}^{0} n(E)\, \mathrm{e}^{-E/T} = \sum_{k=0}^{M} n(-k)\, x^k = n_0 \prod_{l=1}^{M} (x - x_l)\,, \tag{2}$$

where the Boltzmann constant is set to 1, structures mean self-avoiding walks on lattice, $x = \mathrm{e}^{1/T}$, E_0 is the ground state energy, k $(= -E)$ a nonnegative integer, $n(E)$ the density of states, M $(= -E_0)$ the degree of the polynomial, n_0 the ground state degeneracy, and x_l the lth root of the polynomial, i.e., the lth partition function zero. Once the partition function is obtained, other thermodynamic quantities can be calculated. For example, the heat capacity can be expressed by the new variable x and partition function zeros as

$$C_V = \frac{\mathrm{d}}{\mathrm{d}T} \langle E \rangle = \frac{\mathrm{d}}{\mathrm{d}T} \left(T^2 \frac{\mathrm{d}}{\mathrm{d}T} \log Z \right) = -x(\log x)^2 \sum_{l=1}^{M} \frac{x_l}{(x - x_l)^2}\,. \tag{3}$$

When x approaches x_l, C_V approaches infinity, which is a signature of phase transition. However, none of the zeros would touch the physical parameter space, here $1 < x < \infty$, until the number of interacting components of the system goes to infinity. This scenario is proposed in Yang and Lee's celebrated papers for partition function zeros and phase transition [9,10]. In the last 50 years, numerous studies of partition function zeros have been performed [11], nevertheless only a few are on