



ELSEVIER

Physica A 249 (1998) 374–379

PHYSICA A

The collapse transition in the HP model

Rafael J. Najmanovich^{a,*}, Jorge L. deLyra^b, Vera B. Henriques^a

^aDepartment of General Physics, University of Sao Paulo, Sao Paulo, Brazil

^bDepartment of Mathematical Physics, University of Sao Paulo, Sao Paulo, Brazil

Abstract

We studied the collapse transition on small homopolymeric and heteropolymeric sequences in the HP model on a two-dimensional square lattice both with exact enumerations and Monte Carlo simulations. For heteropolymers, our analysis is restricted to native sequences (sequences which present one single minimal energy configuration) with a fixed fraction of hydrophobic residues ($\rho = N_H/N = 0.5$). In the homopolymer limit ($\rho = 1$) we were able to reproduce several known results such as the exponent ν and the critical temperature for the collapse transition. For the heteropolymer case, we devised a combinatorial extrapolation scheme based on the method used for homopolymers which enabled us to study the dependence of the collapse critical temperature on the monomer sequence in the thermodynamic limit. For the small chain lengths studied we find that the θ -point is strongly dependent on the monomer sequence, our extrapolation scheme suggests that even in the thermodynamic limit, the critical temperature of the collapse transition in the HP model in the square lattice depends on the specific location of the hydrophobic residues in the chain. We have also studied the folding transition, our preliminary results for the chain lengths studied do not indicate the existence of an *equilibrium* folding transition occurring at a temperature different from that of the collapse transition. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Collapse transition; Heteropolymers; HP model

1. Introduction

The specific amino acid sequences of proteins were selected evolutively to present stable, kinetically accesible native states, that is, *proteins are not random heteropolymers*. Following this idea we study sequences which present one single minimal energy configuration, and we do not average over random sequences. Among the several questions that compose the protein folding problem we are interested in what is known as the Levinthal paradox, namely how can the folding process occur so fast?

* Correspondence address: Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: rafj@elect1.weizmann.ac.il.

Simplified lattice models [1,2] may be seen as coarse grained representations of proteins where a single minimal energy configuration is associated with the native state. In studying the Levinthal paradox, one idea is that the existence of phase transitions could effectively decrease the time to find the native state by restricting a random search to a more confined region of the configurational space containing the native state configuration. It is argued in the literature that simplified lattice models present at least two phase transitions (see for example Ref. [3]): (i) a collapse transition from the set of extended configurations to a set of collapsed configurations and (ii) a folding transition from the set of collapsed configurations to the native state. These findings are mostly based on dynamical studies.

2. The HP model

The basic simplification introduced by Dill [2] (embodied in the HP model) is that of using two amino acid types (monomers), hydrophobic (H) and hydrophilic or polar (P), instead of the twenty naturally occurring amino acids to define a protein sequence. This simplification is related to the fact that the hydrophobic effect can be considered as the driving force of the protein folding process. The hamiltonian of the HP model can be written as

$$\mathcal{H} = - \sum_{i=1}^N \sum_{j=1}^{i-3} C(i,j) S_i S_j$$

where N is the chain length, $C(i,j)$ is the contact map and S_x determines the type of monomer. The contact map is an $N \times N$ matrix whose elements are unitary whenever monomers i and j are topological neighbors and zero otherwise, while S_x determines the type of monomer, with $S_x = \delta_{H,x}$.

3. The collapse transition in homopolymers

The collapse transition in homopolymers have been extensively studied both analytically and numerically (see for example Refs. [4–7]). The square radius of gyration is a measure of compactness. In the infinite chain limit it behaves as a function of temperature as $\langle R_g^2 \rangle \sim N^{2\nu}$ with $\nu = \nu_c = 1/d$ for $T < T_\theta$, $\nu = \nu_\theta(d)$ for $T = T_\theta$ and $\nu = \nu_0 = 3/(2+d)$ for $T > T_\theta$. To estimate the value of the exponent ν we used a method known as the ratio method [7]:

$$\frac{\langle R_g^2(N) \rangle}{\langle R_g^2(N-k) \rangle} \sim \left(\frac{N}{N-k} \right)^{2\nu_0}$$

The value $k=2$ was used to prevent the large oscillations that occur for even and odd values of the chain length N on the square lattice [7]. In the high-temperature limit

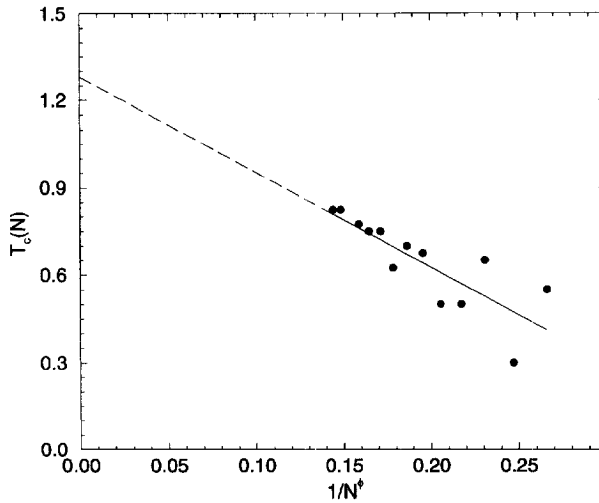


Fig. 1. Extrapolation of $T_{\theta}(N)$ from the specific heat peaks of the finite systems as a function of $1/N^{\phi} \rightarrow 0$ for $N = \{8, \dots, 21\}$.

($T = 50$) we obtain upon extrapolation the value $2\nu_0 = 1.522 \pm 0.005$. We consider this value to be in good agreement with the expected value $2\nu_0 = 1.5$ given the small chain lengths used. We obtain the expected behavior for ν for $d = 2$ both for high and low temperatures. We cannot predict the θ -point due to the small chain lengths used.

We reproduced the critical temperature T_{θ} for the collapse transition on the square lattice. The pseudo critical temperatures (referred simply as critical) for the finite systems $T_{\theta}(N)$ are obtained from the peaks of the specific heat. To determine the collapse critical temperature in the thermodynamic limit we extrapolated (linear extrapolation) the values of $T_{\theta}(N)$ as a function of $1/N^{\phi} \rightarrow 0$ as $T_{\theta}^{(\infty)} \sim \lim_{n \rightarrow \infty} T_{\theta}(N)$, where $n = N - 1$ is the linear size of the chain and $\phi_{\theta} \sim 0.6364$ [5]. We obtain $T_{\theta}^{(\infty)} \sim 1.28 \pm 0.13$ (Fig. 1). This value is in good accordance with the values presented by Baumgartner [5] ($T_{\theta}^{\infty} \sim 1.31 \pm 0.06$) and Ishinabe [6] ($T_{\theta}^{\infty} \sim 1.33$).

4. The collapse transition in heteropolymers

We determined by exact enumeration all native sequences for chains with 8, 10, 12 and 14 monomers with a fixed fraction of hydrophobic residues ($\rho = 0.5$). For chains 16 monomers long we studied 44 of the 1539 existing native sequences which comprises about 3% of the total. For $N = 18$ we found 3 native sequences and 5 native sequences for $N = 26$. The latter were analyzed with the Monte Carlo method with Metropolis dynamics implemented by the Pivot Algorithm [8]. To improve computational efficiency [9] we included some local moves, namely: one bead internal move and crankshaft move.

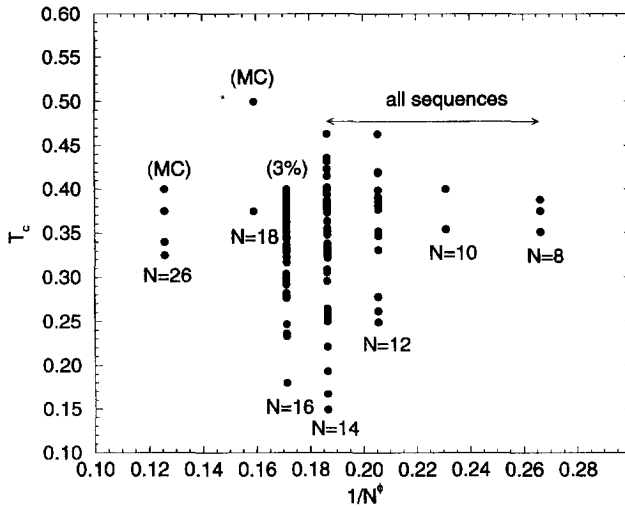


Fig. 2. Critical temperatures of the several native sequences studied for different chain lengths as a function of $1/N^{\phi_0}$.

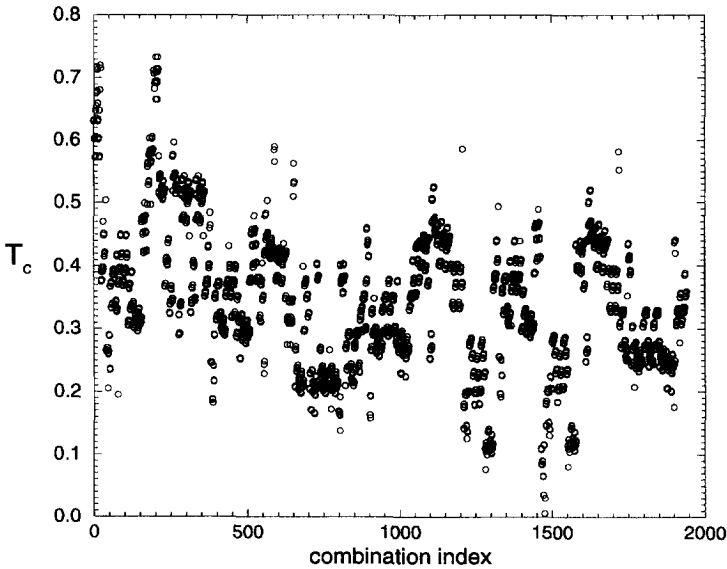


Fig. 3. Extrapolated critical temperatures for the limit of infinite size from different combinations of T_c of different sequences and chain lengths.

The location of the peak in the specific heat curves presents large variations (see Fig. 2, note that this plot is the same as Fig. 1 for homopolymers). Clearly there is no sign of convergence as the chain length increases. Hence, for small chains, the location of the θ -point is extremely dependent on the specific location of monomers

in the chain. Since there is no obvious way in which to increase the chain length of heteropolymers without changing the sequence or even maintaining the non-degeneracy of the ground state, we selected sets of sequences of increasing lengths (one for each chain length) as the combinations of such (native) sequences that showed the best linear fitting of their critical temperatures. We used all such combinations to extrapolate the critical temperature for the heteropolymer collapse transition using the same method of homopolymers. The resulting values for T_c^∞ for all combinations can be seen in Fig. 3. Clearly there is no indication for a single critical temperature, therefore we suggest that *the critical temperature for the collapse transition depends on the specific location of the residues in the chain even in the thermodynamic limit.*

5. The folding transition

The folding transition is supposed to take the heteropolymer from a set of compact conformations to the native state. With this behavior in mind we calculated the average fraction of contacts, $\langle C \rangle$ and the average fraction of native contacts, $\langle Q_0 \rangle$. If there were a folding transition occurring at a temperature different from the collapse transition (T_c) we would expect to find a temperature range δT (with $T \in \delta T$, $T < T_c$) on which there would be a plateau in $\langle C \rangle$ indicating that the average compactness remains constant

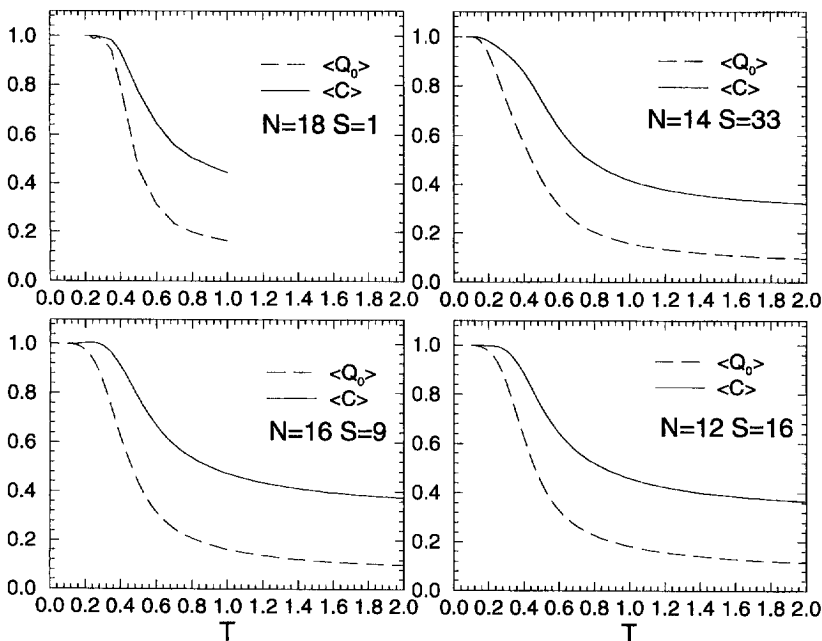


Fig. 4. Behaviour of $\langle Q_0 \rangle$ and $\langle C \rangle$ as a function of temperature. The curves for $S = 1$, $N = 18$ were obtained using the Monte Carlo method.

while $\langle Q_0 \rangle$ would increase very fast indicating a rearrangement of the chain towards the native state. In Fig. 4 we show these quantities for some sequences of different sizes. None of the native sequences studied presented the above behavior, therefore not indicating the existence of an *equilibrium* folding transition occurring at a temperature different from that of the collapse transition.

6. Conclusions

We studied the HP model on the (two-dimensional) square lattice for sequences which present one single minimal energy configuration with a fixed ratio of hydrophobic residues ($\rho = N_H/N = 0.5$). In this particular situation, the location of the θ -point is extremely dependent on the monomer sequence. We found evidences supporting the idea that even in the thermodynamic limit the critical temperature of the collapse transition depends on the specific location of residues in the chain. The study of small heteropolymeric chains may not be sufficient to extrapolate the behaviour of infinite chains. Yet in the homopolymer case it is possible to derive result for the thermodynamic limit from the study of small chains. Preliminary results failed to show any evidences supporting the existence of an *equilibrium* folding transition occurring at a critical temperature different from that of the collapse transition.

Our conclusions may not be valid for the three-dimensional case or if we were able to study larger native sequences. But it is increasingly difficult to determine native sequences as the chain length is increased.

References

- [1] T. Creighton (Ed.), Protein Folding, W.H. Freeman, New York, 1992.
- [2] K.A. Dill et al., Prot. Sci. 4 (4) (1995) 561–601.
- [3] A. Dinner et al., J. Chem. Phys. 101 (2) (1994) 1444–1451.
- [4] M. Wittkop et al., J. Chem. Phys. 104 (9) (1996) 3373–3385.
- [5] A. Baumgartner, J. de Physique 43 (1982) 1407–1411.
- [6] T. Ishinabe, J. Phys. A 18 (1985) 3181–3187.
- [7] V. Privman, J. Phys. A 19 (1986) 3287–3297.
- [8] N. Madras, A.D. Sokal, J. Stat. Phys. 50 (1988) 109.
- [9] M.C. Tesi et al., J. Stat. Phys. 82 (1/2) (1996) 155–181.