Globule-coil transition in the dynamic HP model

Kamilla Faizullina, Evgeni Burovski
National Research University Higher School of Economics, 101000 Moscow, Russia

Abstract. We consider a dynamic version of the HP model of a linear polymer: a self-avoiding walk on the square lattice, with monomers being either hydrophobic (H) or polar (P). We simulate the model in two dimensions in the grand canonical ensemble via the Berretti-Sokal algorithm across the globule-coil transition and map out the phase diagram. Our results are consistent with the universality class of the transition being the same as the universality class of the theta-transition of an interacting self-avoiding walk.

1. Introduction
Predicting physical properties of macromolecules from first principles—via computer simulations using atomistic potentials—is a challenging problem of an enormous computational complexity. To this end, it is useful to study coarse-grained models where chemical monomers are integrated into a subunit up a certain lengthscale, and further modeling proceeds in terms of these subunits [1]. The simplest model of this class is the self-avoiding random walk (SAW) on a lattice, which captures the excluded volume effect [2]. Conceptually, this corresponds to a polymer chain in a good solvent. To capture the effects due to van der Waals attraction between monomers, the model is augmented by assigning negative interaction energy to monomers which occupy nearest neighbor lattice sites. The resulting interacting SAW model contains the globular phase (which corresponds to a collapsed polymer in a poor solvent) and captures the physics of the globule-coil transition, the so-called theta-point [3].

To model heteropolymers, one considers SAWs with several kinds of monomers. The simplest — yet highly non-trivial — model is the so-called HP model [4], originally devised to simulate the protein folding, and later used for protein-like polymers [5, 6]. This model features monomers of two kinds: hydrophobic (H) and polar (P). In this model, only neighboring H monomers are interacting, while P-P and H-P contacts are not. In the “protein folding problem”, the sequence of monomer types is fixed, and the conformation of the chain is dynamic. The opposite limiting regime, where the internal states of a monomer are represented by Ising-type spins located on the chain of a fixed conformation, has been studied for both swollen and globular regimes via mean-field calculations and Monte Carlo simulations [7, 8, 9, 10]. Notably, previous work considers ensembles of SAWs in a given regime (globular or swollen). Surprisingly little is known for the case where both conformations and local variables are dynamic.

In this paper, we consider a dynamic version of the HP model (DHP), where the state of each monomer is a dynamic variable. The numbers of monomers of each type is not fixed and is defined by the conditions of thermal equilibrium. We simulate the model in two spatial dimensions with Monte Carlo method using a variant of the Berretti-Sokal algorithm in the grand canonical ensemble.
canonical ensemble. We map out the phase diagram of the model, focusing on the boundary between the collapsed and swollen phases.

The rest of the paper is organized as follows. In Section 2 we define the DHP model, Sec. 3 describes the Monte Carlo method we use for simulations. In Sec. 4 we discuss results of numerical simulations, followed by conclusions and outlook.

2. Dynamic HP model
Consider a self-avoiding walk (SAW) of \(N\) steps (hence \(N + 1\) sites) on a simple square lattice. Each site of the walk can be either hydrophobic (H) or polar (P). Define a topological contact as a pair of monomers which occupy nearest neighboring sites of the lattice and are not nearest neighbors along the walk (see Fig. 1 for an illustration). Each topological contact of H-H is assigned an interaction energy \(-\epsilon_b < 0\). Other contacts, P-P and P-H, have zero associated interaction energy.

For a fixed conformation \(u\) of \(N\) steps, the grand canonical energy function is

\[
E_u = -\epsilon_b m - \epsilon_a N,
\]

where \(\epsilon_a\) is the chemical potential and \(m\) is the number of H-H topological contacts.

Figure 1: A sample conformation of the length \(N = 10\) and \(m = 2\) topological H-H contacts. Here the sequence of monomers is HPHHHPPPPHH.

The grand canonical partition function, corresponding to Eq. (1) is

\[
Z = \sum_{N=0}^{\infty} \sum_u e^{-E_u/kT} = \sum_{N=0}^{\infty} x^m \beta^N,
\]

where \(\beta = e^{\epsilon_a/kT}\) is the monomer fugacity and \(x = e^{\epsilon_b/kT}\) is conjugate to the number of contacts, \(m\). To fix the energy scale we set the temperature \(kT = 1\) henceforth. The summation over \(u\) in Eq. (2) runs over all possible SAW conformations of \(N\) steps and \(2^{N+1}\) sequences of H and P states of \(N + 1\) monomers along the chain.

It is instructive to compare the DHP model (2) to the standard SAW model of a homopolymer (see, e.g. [11]). In the limit of \(\epsilon_b = 0\), all conformations with \(N\) steps have equal contribution to Eq. (2), which reduces to

\[
Z(x = 1, \beta) = \sum_{N=0}^{\infty} \sum_{|u|=N} \beta^N = \sum_{N=0}^{\infty} 2^{(N+1)c_N} \beta^N,
\]

where \(c_N\) is the number of SAWs of length \(N\). This way, Eq. (3) is nothing but a partition function of a non-interacting SAW with a fugacity of \(2\beta\).
In the limit of $N \to \infty$, the number of SAWs of length $N$ scales as $c_N \sim \mu^N N^{\gamma - 1}$, where $\mu$ is a connectivity constant and $\gamma$ is an entropic exponent [11, 12]. In two dimensions, the values of $\mu$ and $\gamma$ are known from exact enumeration [13], and field-theoretic calculations [14], respectively. The scaling form of $c_N$ implies that the partition function (3) stays finite in the $N \to \infty$ limit provided $\beta < \beta_c = 1/\mu$. The opposite regime, $\beta > \beta_c$, corresponds to an unphysical regime where mean length of the walk, $\langle N \rangle$, is infinite.

For the DHP model, Eq. (2), the critical value of $\beta$ is $x$-dependent, $\beta < \beta_c(x)$. We use Monte Carlo simulations to map out the values of $\beta_c(x)$.

3. Berretti-Sokal algorithm for the DHP model
To simulate the model (1)–(2), we use a straightforward generalization of the Berretti-Sokal algorithm, also known as Slithering-tortoise algorithm, [15]. The algorithm is based on the Markov chain Monte Carlo process with two types of elementary updates: the positive move appends a new step to the end of the SAW, and the negative move deletes the last edge. For the dynamic model, the positive move also selects the state of the new site (H or P) at random.

Following Ref. [15], the positive move consists of (i) proposing to chose a neighbor of the endpoint of the SAW at random and (ii) checking the self-avoidance condition. If the proposed configuration crosses itself, the update is rejected. To satisfy the detailed balance condition for the Metropolis algorithm [16], we select the positive move with the probability $P_+ = 8/9$ and $p_- = 1/9$ [15].

We note that the autocorrelation time of the Berretti-Sokal algorithm is at least $\tau \sim \langle N \rangle^2$ [15]. Furthermore, sampling becomes progressively less efficient as $\epsilon_b$ increases [17], especially in the globular phase (where the SAW is essentially a Hamiltonian walk on a finite size lattice). This way, our simulations are limited to relatively short walks (up to $N \lesssim 10^3$).

4. Numerical results
4.1. The grand canonical phase diagram

![Figure 2: Phase diagram for rescaled critical values $\beta_c$ with errorbars. Simulations for DHP were done using $4 \times 10^9$ Monte Carlo steps. In these simulations, the typical lengths of the chains are 50 to 200. Also shown are the critical values $\beta_c$ for a homopolymer, i.e., a SAW with the HH···H sequence, taken from Ref. [17].](image)

To construct the phase diagram of the DHP model, (1)–(2), we perform a series of simulations varying both parameters $\beta$ and $x$ and measuring the mean value of length of the walk, $\langle N \rangle$. 

"
To extract the critical value of the fugacity from the Monte Carlo data, we follow Ref. [15] which, for a non-interacting SAW, uses

$$\langle N \rangle \approx \frac{\beta \gamma}{\beta_c - \beta},$$

which is supposed to hold for $|\beta - \beta_c| \ll 1$. Here $\gamma$ is the value of the entropic exponent. At a given value of $x$, we use a similar Ansatz with additional $x$-dependence,

$$\langle N \rangle \approx \frac{\beta \gamma(x)}{\beta_c(x) - \beta}$$

We fit the Monte Carlo data at a fixed value of $x$ using Eq. (4). While the scaling form (4) is only an Ansatz, this procedure allows us to obtain first crude approximations for $\beta_c$. Fig. 2 shows the result of this procedure. The lengths of the chains are defined by the simulation parameters $x$ and $\beta$, the typical values varying from 50 to 200. For comparison, Fig. 2 also shows the critical fugacity for the interacting SAW model of a heteropolymer [17] — which corresponds to a fixed sequence HH····H. To compare the critical values for the two models, we rescale the values of $\beta_c$ for the DHP (squares) by two, i.e. for the DHP the vertical axis is $2\beta_c$. This rescaling is motivated by Eq. (3). The dependence of $\beta_c$ on the coupling $\epsilon_b$ is qualitatively similar for both models. For a fixed value of $\epsilon_b$, the critical fugacity $\beta_c$ for the DHP model is larger, because there is no interaction energy between topological P-P and H-P pairs.

4.2. End-to-end distance scaling

![Figure 3: Estimates of the exponent $\nu$ versus $\epsilon_b$. Errorbars are show at all points. Simulations are performed with $\sim 10^{11}$ MC steps. The lengths of the SAW chains vary from $N = 100$ to 300. See text for discussion.](image)

In order to characterize the geometric properties, we measure the end-to-end distance of a walk as a function of the length, which is expected to scale with the metric exponent $\nu$,

$$\langle R_N^2 \rangle \sim N^{2\nu} \quad N \to \infty.$$
For a non-interacting SAW the exact value is \( \nu = \frac{3}{4} \) [18]. Exactly at the theta-point, \( \nu_\theta = \frac{4}{7} \) [19].

Since the accessible values of \( N \) are limited (for too short chains, corrections to scaling are too large; for too long chains we lack Monte Carlo statistics), we use the following phenomenological Ansatz [15]

\[
\langle \log(R_N^2 + k_1) \rangle = 2\nu \log(N + k_2) + b, \tag{5}
\]

where \( k_1, k_2 \) and \( b \) are phenomenological non-universal fit parameters meant to mimic corrections-to-scaling.

Fig. 3 shows the result of this fitting procedure based on simulations of \( \sim 10^{11} \) Monte Carlo steps and chain lengths from 100 to 300. We also perform similar simulations for the homopolymer HH\cdots H. It is clear from Fig. 3 that for small \( \epsilon_b \) the metric exponent agrees with the non-interacting SAW limit for both models.

We also show in Fig. 3 the value \( \nu = 4/7 \), which is the conformal field theory result for the metric exponent at the theta-point of the interacting SAW model of a homopolymer [19]. One of the recent estimates for the homopolymer \( \epsilon_{\theta \theta} \approx 0.6673(5) \) [20] agrees with \( \nu = 4/7 \). Assuming that the DHP model belongs to the same universality class, we obtain from Fig. 3 an estimate for the theta-point of the DHP model: \( 1.2 \leq \epsilon_{\theta \theta} \leq 1.35 \).

### 4.3. Location of the theta-point

![Graph showing simulation results for the heat capacity, Eq. (6) for several lengths. In these simulations we use up to \( 2 \times 10^{11} \) MC steps. See text for discussion.](image)

As an independent probe of the location of the theta-point, we consider the “heat capacity” \( C \), defined as

\[
C = \epsilon_b^2 \langle (m^2) - \langle m \rangle^2 \rangle, \tag{6}
\]

at a fixed value of \( N \). We expect the heat capacity to have a peak at the transition point in the thermodynamic limit \( N \to \infty \). For finite lengths, the peak is broadened and shifted by the finite-size corrections. Fig. 4 shows results of simulations with lengths \( N = 100 \) to \( N = 250 \). A peak is clearly seen developing with increasing \( N \) precisely at the region where the metric
exponent is close to the Coulomb gas value $\nu = 4/7$. Given the resolution of the simulations, we read off Fig. 4 the location of the theta-point to be

$$1.25 \leq \epsilon_{b\theta} \leq 1.27.$$  \hfill (7)

It is well known that the peak of the heat capacity is rather strongly affected by the finite size corrections. An improved estimate for $\epsilon_{b\theta}$ and a check of the universality class can be done by analyzing the scaling form of the end-to-end distance in the critical region around the theta-point [20]:

$$\langle R^2 \rangle_N = a N^{2\nu_{\theta}} + b N^{2\nu_{\theta}} + \phi(\epsilon_b - \epsilon_{b\theta}) + \cdots,$$  \hfill (8)

where $\nu_{\theta}$ and $\phi$ are universal critical exponents, $a$ and $b$ are non-universal amplitudes, and dots represent higher-order corrections to scaling. However significantly larger lengths $N$ are required to reliably improve the estimate (7), and we postpone this study to a future work.

Conclusions and outlook

In this work, we study a DHP model of a lattice heteropolymer. We simulate the model using a variant of the Berretti-Sokal Monte Carlo algorithm in the grand canonical ensemble. Using numerical simulations, we construct the grand canonical phase diagram of the model, and obtain an estimate of the location of the theta-transition between a swollen and a globular phase.

The grand canonical nature of the algorithm, and the singularity structure of the partition function of the model make it challenging to obtain precise estimates of the critical parameters. For future work, we plan to improve the numerical algorithm to be able to simulate longer chains, and to extend the technique to models of magnetic polymers, where the internal states of a SAW are modeled as Ising or Heisenberg spins.

This research was supported in part through computational resources of HPC facilities at NRU HSE. We acknowledge partial support of RFBR project No 19-07-01117.

References

[1] Binder K 1995 Introduction: General Aspects of Computer Simulation Techniques and their Applications in Polymer Physics (Oxford University Press) p 3–46
[2] Vanderzande C 1998 Lattice models of polymers (Cambridge University Press)
[3] De Gennes P G 1979 Scaling concepts in polymer physics (Cornell university press)
[4] Lau K F and Dill K A 1989 Macromolecules 22 3986–3997
[5] Khokhlov A R and Khalatur P G 1998 Physica A 249 253–261
[6] Murnen H K, Khokhlov A R, Khalatur P G, Segalman R A, and Zuckermann R N 2012 Macromolecules 45(12) 5529–5246
[7] Aertsens M and Vanderzande C 1992 Journal of Physics A: Mathematical and General 25 735–747
[8] Chakrabarti B K, Maggs A C and Stinchcombe R B 1985 Journal of Physics A: Mathematical and General 18 L373–L376
[9] Garel T, Orland H and Orlandini E 1999 The European Physical Journal B-Condensed Matter and Complex Systems 12 261–268
[10] Papale A and Rosa A 2018 The European Physical Journal E 41 144
[11] Vanderzande C 1998 From polymers to random walks (Cambridge University Press) p 1–18 Cambridge Lecture Notes in Physics
[12] van Rensburg E 2015 The Statistical Mechanics of Interacting Walks, Polygons, Animals and Vesicles Oxford lecture series in mathematics and its applications (Oxford University Press) ISBN 9780199666577
[13] Jensen I 2004 Journal of Physics A: Mathematical and General 37 5503–5524
[14] Duplantier B 1988 Phys. Rev. Lett. 57(8) 941–944
[15] Berretti A and Sokal A D 1985 Journal of Statistical Physics 40 483–531
[16] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 The Journal of Chemical Physics 21 1087–1092 (Preprint https://doi.org/10.1063/1.1699114)
[17] Nidras P P 1996 Journal of Physics A: Mathematical and General 29 7929–7942
[18] Li B, Madras N and Sokal A 1994 Journal of Statistical Physics 80
[19] Duplantier B and Saleur H 1987 Phys. Rev. Lett. 59(5) 539–542
[20] Caracciolo S, Gherardi M, Papinutto M and Pelissetto A 2011 *Journal of Physics A: Mathematical and Theoretical* **44** 115004