

**Phase diagram and the pseudogap state in a linear chiral homopolymer model**A. Sinelnikova,<sup>1,\*</sup> A. J. Niemi,<sup>2,3,4,†</sup> and M. Ulybyshev<sup>1,5,6,‡</sup><sup>1</sup>*Institute for Theoretical Problems of Microphysics, Moscow State University, Moscow 119899, Russia*<sup>2</sup>*Department of Physics and Astronomy, Uppsala University, P.O. Box 803, S-75108, Uppsala, Sweden*<sup>3</sup>*Laboratoire de Mathématiques et Physique Théorique CNRS UMR 6083, Fédération Denis Poisson, Université de Tours, Parc de Grandmont, F37200 Tours, France*<sup>4</sup>*Department of Physics, Beijing Institute of Technology, Haidian District, Beijing 100081, P. R. China*<sup>5</sup>*Institute of Theoretical Physics, University of Regensburg, D-93053 Germany, Regensburg, Universitätsstrasse 31*<sup>6</sup>*ITEP, B. Chermushkinskaya str. 25, Moscow 117218, Russia*

(Received 21 January 2015; revised manuscript received 23 June 2015; published 17 September 2015)

The phase structure of a single self-interacting homopolymer chain is investigated in terms of a universal theoretical model, designed to describe the chain in the infrared limit of slow spatial variations. The effects of chirality are studied and compared with the influence of a short-range attractive interaction between monomers, at various ambient temperature values. In the high-temperature limit the homopolymer chain is in the self-avoiding random walk phase. At very low temperatures two different phases are possible: When short-range attractive interactions dominate over chirality, the chain collapses into a space-filling conformation. But when the attractive interactions weaken, there is a low-temperature unfolding transition and the chain becomes like a straight rod. Between the high- and low-temperature limits, several intermediate states are observed, including the  $\theta$  regime and pseudogap state, which is a novel form of phase state in the context of polymer chains. Applications to polymers and proteins, in particular collagen, are suggested.

DOI: [10.1103/PhysRevE.92.032602](https://doi.org/10.1103/PhysRevE.92.032602)

PACS number(s): 36.20.-r, 05.70.-a, 05.10.Ln, 87.15.A-

**I. INTRODUCTION**

A linear homopolymer is made of a single type of a repeat unit. An important example is polyacetylene, an organic conductive polymer which is the paradigm material for the fractional fermion number [1,2]. Additional examples, among many others, are poly-L-lysine and poly-L-glutamic. The former is a food preservative with potential for wider, even pharmaceutically relevant, antimicrobial effects [3] while the latter is used for drug delivery against cancer [4].

From a theoretical point of view, the concept of a homopolymer chain is a useful coarse-grained approximation, even in the case of a heteropolymer that exhibits only approximately repeating patterns: For a sufficiently long chain the distinct monomers are simply combined into appropriate subunits, to dispose of the inhomogeneities in the monomer species. For example, collagen, which is the most abundant protein in mammals, displays a repeated glycine-proline-X pattern, where X is any amino acid other than glycine and proline. DNA, RNA, and the C $\alpha$  backbone of a protein chain are additional examples where a homopolymer approximation is occasionally profitably introduced [5–7].

Here the phase structure of a *chiral* linear homopolymer is investigated, in terms of a universal energy function [8–13]; a chiral polymer is one where parity is broken, and the mirror image of a stable chiral polymer conformation is in general not stable. A chiral polymer often has a tendency to form helical structures. For example, in the case of proteins right-handed helical structures are more common than left-handed ones. There are also chiral proteins that can form different right-

and left-handed structures. An important example is collagen for which the right-handed polyproline I conformation is more compact than the left-handed polyproline II.

It is found that the phase structure of a chiral homopolymer is more complex than that of a nonchiral one. In particular, a chiral homopolymer can be in a *pseudogap state*. Here we give a brief description of this phase, following Refs. [14–18] for some abstract statistical system which displays a continuous symmetry that is associated with a complex order parameter

$$\Psi = \rho e^{i\varphi}. \quad (1)$$

Here  $\rho$  and  $\varphi$  are just the modulus and phase of this abstract order parameter. Such an order parameter is often present, e.g., in models of superconductivity. Further in Sec. II E [see Eq. (30)] we give a description of this complex order parameter in terms of polymer degrees of freedom. The restoration of the continuous symmetry commonly takes place so that the free energy of the symmetry breaking state with nonvanishing condensate  $\rho \neq 0$  becomes larger than the free energy of the symmetric state where  $\rho$  vanishes. But the symmetry can also be restored by phase decoherence, even when  $\rho \neq 0$ . This occurs when the phase  $\varphi$  in the order parameter becomes disordered so that  $\langle e^{i\varphi} \rangle = 0$ . This implies that the expectation value of the order parameter also vanishes,  $\langle \Psi \rangle = 0$ . The system is then in the pseudogap state [14–18]. The pseudogap state is a symmetric phase precursor state in the broken symmetry phase. In particular, the transition between the broken symmetry phase and the pseudogap state is not a phase transition but a crossover prelude to the fully symmetric state that the system enters when the lowest energy state of the effective potential is one where the modulus vanishes.

Note that the general arguments due to Kadanoff and Wilson [19–23] imply that in the thermodynamical limit the phase transition properties of a material system are commonly

\*ann.sinelnikova@gmail.com

†Antti.Niemi@physics.uu.se; <http://www.folding-protein.org>

‡ulybyshev@goa.bog.msu.ru

*universal*, i.e., independent of the atomic-level details. From this perspective, the construction of the phase diagram of a linear and chiral structureless homopolymer presented here should be relevant for the understanding of the phase diagram of more elaborate linear chiral homopolymers, maybe even that of certain heteropolymers [5–7]. Indeed, linear polymers are presumed to have a very similar phase structure, quite independently of their chemical composition [5–7] even though the phase where a particular polymer resides depends on many factors such as concentration, the quality of solvent, ambient temperature, and pressure.

The article is organized as follows: The next section describes the background and the methods. The standard phase structure of linear polymers is first reviewed. The geometrical order parameter variables that are used to model the free energy of a homopolymer are defined, followed by an outline how the ensuing universal Hamiltonian emerges in the limit of low wavelength deformations. The order parameter that detects the presence of a pseudogap state is then defined. The zero temperature ground state is identified in the case of pure steric repulsion, and a universal attractive long-range interaction is introduced to model the effect of hydrophobic forces in the homopolymer chain. A detailed analysis of various Monte Carlo algorithms is presented, to identify one that is computationally most effective in the case of a single homopolymer chain. In the subsequent section the results are then described. The effect of the various parameters to the phase structure is revealed, and in particular the pseudogap state is identified. Finally, the phase diagram is constructed as a function of the various parameters. It is found that in the case of a chiral homopolymer the phase diagram has a much richer structure than in the case of a nonchiral homopolymer.

## II. METHODS

### A. Phases

A review of the known phase structure of linear, nonchiral homopolymers is now presented, as a background and motivation for the subsequent study.

Three different, universal phases are commonly identified, and these phases are categorized by the way how the polymer structure fills the space [5–7]: Under poor solvent conditions or at low temperatures, when the attractive interactions between the monomers dominate, a single polymer chain is presumed to collapse into a configuration which is space filling. On the other hand, in a good solvent or at high temperatures, when the repulsive interactions dominate and cause the chain to effectively swell, its geometric structure resembles that of a self-avoiding random walk (SAW). Between the two, there is a  $\theta$  regime (possibly a tricritical  $\theta$  point) where the attractive and repulsive interactions cancel each other. In the  $\theta$  regime the polymer chain is presumed to have the characteristics of an ordinary random walk (RW). Finally, some polymers such as collagen are more like straight, rigid rods. Each of these four phases, rigid rod, SAW, RW, and the space-filling one, can be characterized by the inverse of the Hausdorff dimension of the chain, called the scaling exponent  $\nu$  [5–7]. This quantity is

defined by the radius of gyration

$$R_{gyr} = \sqrt{\frac{1}{2N^2} \sum_{i,j} (\mathbf{r}_i - \mathbf{r}_j)^2}, \quad (2)$$

where  $\mathbf{r}_i$  are the coordinates of the individual monomers. When the number of monomers  $N$  becomes very large, the radius of gyration has the asymptotic expansion [24–27]

$$R_{gyr}^2 \xrightarrow{N \text{ large}} R_0^2 N^{2\nu} (1 + R_1 N^{-\delta_1} + \dots) \sim R_0^2 N^{2\nu}. \quad (3)$$

Here the length scale  $R_0$ , the Kuhn length, is the effective distance between the monomers in the large- $N$  limit. The Kuhn length is not a universal quantity; its value can in principle be computed from the atomic-level details of the polymer and environment including pressure, temperature, and chemical microstructure of the solvent. The dimensionless scaling exponent  $\nu$ , i.e., the inverse Hausdorff dimension that governs the large- $N$  asymptotic form of Eq. (3), is presumed to be a universal quantity. Its numerical value is independent of the local atomic-level structure of the polymer [5,7,24–26]. Here  $\delta_1$  are critical exponents, and  $R_1$  are the corresponding amplitudes, and together they constitute the finite-size corrections. Also,  $\delta_1$  are universal quantities [26], but  $R_1$  are not universal [26].

The following mean field values are conventionally assigned to  $\nu$  [5–7]:

$$\nu = \begin{cases} 1/3 & \text{collapsed} \\ 1/2 & \text{RW} \\ 3/5 & \text{SAW} \\ 1 & \text{rod} \end{cases}. \quad (4)$$

Under poor solvent conditions or at low temperatures, the polymer collapses into a space-filling conformation [28–30] with the mean field exponent  $\nu = 1/3$ . Folded proteins are commonly found in this phase. For an ordinary RW the mean field value is  $\nu = 1/2$ . This corresponds to the  $\theta$  regime that separates the collapsed phase from the high-temperature self-avoiding random walk phase for which the Flory value  $\nu = 3/5$  is found. Finally, when  $\nu = 1$ , the polymer loses its inherently fractal structure and behaves like a straight rod.

The transition between the collapsed phase and the SAW phase has been studied extensively, and it involves the RW phase as a tricritical  $\theta$  point or more generally as a transitional  $\theta$  regime. But the transitions between the rigid rod phase and the other three are less studied. However, there is a physically and biologically very important scenario where such a transition could have a role: that of cold denaturation of a protein chain. The presence of all four phases (4) opens the possibility of a four-critical point, under proper conditions [31].

Finally, the  $\theta$ -point value  $\nu = 1/2$  is exact for a polymer with no long-range interactions [5]. For a space-filling structure the value  $\nu = 1/3$  is also exact, and similarly  $\nu = 1$  is exact for a straight, linear rodlike structure. But in the case of SAW the mean field value is corrected by fluctuations. A numerical Monte Carlo evaluation, computed directly by using the self-avoiding random-walk model on a square lattice, gives the estimate [26]

$$\nu = 0.5877 \pm 0.0006. \quad (5)$$

In the sequel the scaling exponent  $\nu$  in (4) is evaluated as a function of various parameters in different phases of the homopolymer, using numerical simulations in the context of a universal off-lattice energy function. Of particular interest is the effect of the parameter that characterizes the chirality and the parameter that characterizes the strength of the attractive (hydrophobic) forces.

### B. Geometry

The order parameters that determine the free energy of the homopolymer chain in terms of its geometry are now identified, following Ref. [32]. For this a homopolymer chain with  $i = 1, \dots, N$  monomers is considered, with  $\mathbf{r}_i$  the three-dimensional space coordinates. The unit tangent vectors along the lines that connect two consecutive monomers are

$$\mathbf{t}_i = \frac{\mathbf{r}_{i+1} - \mathbf{r}_i}{|\mathbf{r}_{i+1} - \mathbf{r}_i|}. \quad (6)$$

The unit binormal vectors are defined by

$$\mathbf{b}_i = \frac{\mathbf{t}_{i-1} \times \mathbf{t}_i}{|\mathbf{t}_{i-1} \times \mathbf{t}_i|}. \quad (7)$$

The unit normal vectors are defined by

$$\mathbf{n}_i = \mathbf{b}_i \times \mathbf{t}_i. \quad (8)$$

The three vectors  $(\mathbf{n}_i, \mathbf{b}_i, \mathbf{t}_i)$  determine an orthonormal frame at the monomer position  $\mathbf{r}_i$ . The discrete bond angles are

$$\kappa_i \equiv \kappa_{i+1,i} = \arccos(\mathbf{t}_{i+1} \cdot \mathbf{t}_i), \quad (9)$$

and the discrete torsion angles are

$$\tau_i \equiv \tau_{i+1,i} = \text{sgn}[(\mathbf{b}_{i-1} \times \mathbf{b}_i) \cdot \mathbf{t}_i] \times \arccos(\mathbf{b}_{i+1} \cdot \mathbf{b}_i). \quad (10)$$

Conversely, when the angles  $(\kappa_i, \tau_i)$  are known the discrete Frenet equation [32]

$$\begin{pmatrix} \mathbf{n}_{i+1} \\ \mathbf{b}_{i+1} \\ \mathbf{t}_{i+1} \end{pmatrix} = \begin{pmatrix} \cos \kappa \cos \tau & \cos \kappa \sin \tau & -\sin \kappa \\ -\sin \tau & \cos \tau & 0 \\ \sin \kappa \cos \tau & \sin \kappa \sin \tau & \cos \kappa \end{pmatrix}_{i+1,i} \cdot \begin{pmatrix} \mathbf{n}_i \\ \mathbf{b}_i \\ \mathbf{t}_i \end{pmatrix} \quad (11)$$

determines the frames iteratively, by computing the frame at the position of the  $(i+1)$ -th monomer from the frame at the position of the  $i$ th monomer. Once all the frames have been constructed, the entire chain is obtained as follows:

$$\mathbf{r}_k = \sum_{i=0}^{k-1} |\mathbf{r}_{i+1} - \mathbf{r}_i| \cdot \mathbf{t}_i. \quad (12)$$

With no loss of generality one can set  $\mathbf{r}_0 = 0$  and orient  $\mathbf{t}_0$  to point into the direction of the positive  $z$  axis.

A framing is necessary for the construction of the chain from the bond and torsion angles. But Eq. (12) does not involve the vectors  $\mathbf{n}_i$  and  $\mathbf{b}_i$ . Thus any linear combination of these two vectors could be chosen to define a framing, to construct the chain from the angles. This also determines the symmetry that enables the identification of the pertinent order parameter (1): Consider a local SO(2) transformation that rotates the frame  $(\mathbf{n}_i, \mathbf{b}_i)$  by an angle  $\Delta_i$  leaving  $\mathbf{t}_i$  intact,

$$\begin{pmatrix} \mathbf{n} \\ \mathbf{b} \\ \mathbf{t} \end{pmatrix}_i \rightarrow \begin{pmatrix} \cos \Delta_i & \sin \Delta_i & 0 \\ -\sin \Delta_i & \cos \Delta_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{n} \\ \mathbf{b} \\ \mathbf{t} \end{pmatrix}_i. \quad (13)$$

On the Frenet frame bond and torsion angles in (11), this has the following effect:

$$\kappa_i T^2 \rightarrow e^{\Delta_i T^3} (\kappa_i T^2) e^{-\Delta_i T^3}, \quad (14)$$

$$\tau_i \rightarrow \tau_i + \Delta_{i-1} - \Delta_i, \quad (15)$$

where the  $(T^a)_{bc} = \epsilon_{abc}$  are the SO(3) generators  $[T^a, T^b] = \epsilon_{abc} T^c$ . The range of  $\tau_i$  is  $[-\pi, \pi) \bmod(2\pi)$ . Equations (14) and (15) may be used to extend the range of the bond angle from  $[0, \pi)$  to  $\kappa_i$  into  $[-\pi, \pi) \bmod(2\pi)$ . The extension is compensated for by the following discrete  $\mathbb{Z}_2$  symmetry

$$\begin{aligned} \kappa_k &\rightarrow -\kappa_k & \text{for all } k \geq i \\ \tau_i &\rightarrow \tau_i - \pi \end{aligned} \quad (16)$$

which leaves the chain intact.

In the numerical simulations presented here, all the distances between nearest neighbor monomers are fixed to the uniform constant value

$$|\mathbf{r}_{i+1} - \mathbf{r}_i| = \delta = 3.8 \text{ \AA}. \quad (17)$$

This equals the average distance between two consecutive C $\alpha$  atoms along a protein backbone, measured in Ångströms.

A polymer is subject to steric constraints, due to overlapping electron clouds and various short-range Born repulsions. Accordingly the following forbidden volume constraint is introduced:

$$|\mathbf{r}_i - \mathbf{r}_k| \geq \delta \equiv 3.8 \text{ \AA} \quad \text{for } |i - k| \geq 2. \quad (18)$$

This is in line with the minimum distance observed between any two C $\alpha$  atoms, in folded protein structures. The numerical values (17) and (18) can both be independently modified, with no effect on conclusions.

In the sequel only scaled dimensionless units are used, and in particular the dimensionless unit of length is 1 Å.

### C. Free energy in the infrared limit

The bond and torsion angles constitute a complete set of geometric variables, to describe protein C $\alpha$  backbones [33]. Furthermore, according to Eq. (3) and Eq. (4) the structure-dependent phase diagram of a homopolymer is determined by the three-dimensional chain geometry. Thus the bond and torsion angles are a complete set of *order parameters*, in the sense of Kadanoff and Wilson [19–23]: The geometrically defined, structural phase diagram of a homopolymer can be fully determined by a thermodynamical free energy which is constructed from these order parameters only.

A detailed derivation of the free energy used here is now presented. For this a homopolymer chain in thermal equilibrium is considered. Let  $F$  be the ensuing thermodynamical Helmholtz free energy. Thus, the minimum of  $F$  describes the chain configuration, under thermodynamical equilibrium conditions. The free energy is the sum of the internal energy  $U$  and the entropy  $S$ , at temperature  $T$

$$F = U - TS. \quad (19)$$

It is a function of all the interatomic distances

$$F = F(r_{\alpha\beta}); \quad r_{\alpha\beta} = |\mathbf{r}_\alpha - \mathbf{r}_\beta|, \quad (20)$$

where the indices  $\alpha, \beta, \dots$  extend over all the atoms in the homopolymer system, including those of the solvent environment. Consider the infrared, long distance limit where the characteristic length scales of spatial deformations along the homopolymer chain around its thermal equilibrium configuration are large in comparison to the distance (17) between neighboring monomers. This is synonymous to an assumption that there are no abrupt wrenches and buckles along the chain, that there are only gradual long wavelength bends, which is the limit of adiabatic deformations. The completeness of the bond and torsion angles to describe a protein structure [33] implies that, in order to determine the thermodynamical phase state of the homopolymer chain, it is sufficient to consider the response of all the distances between all the atoms to the variations in the bond and torsion angles only:

$$r_{\alpha\beta} = r_{\alpha\beta}(\kappa, \tau). \quad (21)$$

Here, and in the sequel,  $(\kappa, \tau)$  denotes collectively all the variables  $\kappa_i$  and  $\tau_i$ .

Suppose that at a local extremum of the free energy, the bond and torsion angles along the homopolymer chain have the values

$$(\kappa_i, \tau_i) = (\kappa_{i0}, \tau_{i0}). \quad (22)$$

Consider a conformation where the  $(\kappa_i, \tau_i)$  deviate from these extremum values. The deviations are

$$\begin{aligned} \Delta\kappa_i &= \kappa_i - \kappa_{i0}, \\ \Delta\tau_i &= \tau_i - \tau_{i0}. \end{aligned} \quad (23)$$

Start by Taylor expanding the infrared limit Helmholtz free energy (19) around the extremum,

$$\begin{aligned} F[r_{\alpha\beta} = r_{\alpha\beta}(\kappa_i, \tau_i)] &\equiv F(\kappa, \tau) = F(\kappa_0, \tau_0) \\ &+ \sum_k \left\{ \frac{\partial F}{\partial \kappa_k|_0} \Delta\kappa_k + \frac{\partial F}{\partial \tau_k|_0} \Delta\tau_k \right\} \\ &+ \sum_{k,l} \left\{ \frac{1}{2} \frac{\partial^2 F}{\partial \kappa_k \partial \kappa_l|_0} \Delta\kappa_k \Delta\kappa_l + \frac{\partial^2 F}{\partial \kappa_k \partial \tau_l|_0} \Delta\kappa_k \Delta\tau_l \right. \\ &\left. + \frac{1}{2} \frac{\partial^2 F}{\partial \tau_k \partial \tau_l|_0} \Delta\tau_k \Delta\tau_l \right\} + O(\Delta^3). \end{aligned} \quad (24)$$

The first term in the expansion evaluates the free energy at the extremum. Since  $(\kappa_{i0}, \tau_{i0})$  correspond to the extremum, the second term vanishes. Denote in the sequel  $(\kappa_i, \tau_i)$  collectively, as the variable  $\rho_i$ . Then,

$$\begin{aligned} F(\kappa, \tau) &\equiv F(\rho) \\ &= F(\rho_0) + \frac{1}{2} \sum_{k,l} \frac{\partial^2 F}{\partial \rho_k \partial \rho_l|_0} \Delta\rho_k \Delta\rho_l + O(\Delta^3). \end{aligned} \quad (25)$$

Following Ref. [34] the expansion (25) is *rearranged* in terms of of the differences in the angles  $\rho_i \sim (\kappa_i, \tau_i)$ , as follows:

$$\begin{aligned} F(\rho) &= \sum_k \{ V_k(\rho_k; \rho_{0k}) \\ &+ Z_k(\rho_k; \rho_{0k})(\rho_k \rho_{k+1} + \rho_k \rho_{k-1}) + \dots \}. \end{aligned} \quad (26)$$

Here  $\rho_{0k}$  denotes a combination of the various parameters  $(\kappa_{i0}, \tau_{i0})$  along the chain. But  $V_k(\rho_k; \rho_{0k})$ ,  $Z_k(\rho_k; \rho_{0k})$  and so forth depend on the variable  $\rho_k$  only on the site  $k$ ; these functions are *ultralocal*. The terms that are not shown explicitly consist of higher order differences  $\rho_k \rho_{k+i}$  with  $i \geq 2$ , and higher powers of the differences. The local terms  $V_k(\rho_k)$  constitute the *effective potential*

$$V_{\text{eff}} = \sum_k V_k(\rho_k). \quad (27)$$

The structure of the effective potential is commonly used to conclude whether a spontaneous symmetry breaking takes place [34].

The transition from (25) to (26) involves, *a priori*, an *infinite* rearrangement of the terms in the Taylor expansion (25). In particular, the expansion (26) has been designed so that in the continuum limit where distance between neighboring monomers vanishes, i.e.,  $\delta \rightarrow 0$  in (17), it becomes, at least *naively*, an expansion of the free energy in powers of momentum about the point where momentum vanishes: For a single scalar variable  $\rho_k$  with continuum limit  $\rho_k \rightarrow \phi(x)$  the corresponding continuum limit of (26) is the derivative expansion [34]

$$F(\phi) = \int \left[ V(\phi) + \frac{1}{2} (\partial_\mu \phi)^2 Z(\phi) + \dots \right]. \quad (28)$$

#### D. Effective Hamiltonian

Clearly, the free energy must remain invariant under the local frame rotations (14) and (15); the physical properties of the chain do not depend on the choice of framing. Accordingly, it has been concluded [8–13] that, in the unitary gauge, to the leading nontrivial order the free energy has the form

$$\begin{aligned} H &= - \sum_{i=1}^{N-1} 2\kappa_{i+1}\kappa_i + \sum_{i=1}^{i=N} \{ 2\kappa_i^2 + q(\kappa_i^2 - m^2)^2 + cd\kappa_i^2\tau_i^2 \} \\ &+ \sum_{i=1}^{i=N} \left\{ \frac{1}{2} c\tau_i^2 - a\tau_i - b\kappa_i^2\tau_i \right\}. \end{aligned} \quad (29)$$

This is adopted as the (effective) Hamiltonian, in the sequel. In (29)  $q, m, a, b, c, d$  depend on the atomic level physical properties and the chemical microstructure of the homopolymer chain and its environment. In principle, these parameters can be computed from this knowledge. (Note the combination  $cd$  in the last term of second sum; this choice is made for later convenience.)

It can be shown [8–13] that (29) is the most general, universal and gauge, i.e., frame rotation (13) invariant Hamiltonian, that models a homopolymer in the limit where the characteristic length scales of spatial deformations around the minimum energy configuration become large in comparison to the distance (17) between consecutive monomers. The effective Hamiltonian (29) coincides with the *naively* discretized continuum *Abelian Higgs model Hamiltonian* with one complex scalar field, when expressed in the unitary gauge and the U(1) gauge transformation is identified with the frame rotation (13); the term with parameter  $a$  is the Chern-Simons term, commonly introduced in gauge theories to break parity.

TABLE I. The parameters in (29) that are kept fixed during our simulations.

$q$	$m$	$b$	$c$	$d$
3.5	1.5	0	$10^{-4}$	$10^{-4}$

In particular, (29) is *unique* in the sense of Kadanoff and Wilson.

Implicit in (29) is the assumption that there are no abrupt wrenches and buckles along the polymer chain. Only small, gradual bends are present in the deviations around the energy minimum configuration, which is obtained by minimizing the Hamiltonian (29). This defines the limit of adiabatic deformations.

In line with the Abelian Higgs model (see Ref. [35] in the present context), the Hamiltonian (29) displays the discrete symmetry  $\kappa_i \rightarrow -\kappa_i$ . As in the case of the Abelian Higgs model, this symmetry may become spontaneously broken by the ground state. It should be noted that (29) is *not* invariant under the local  $\mathbb{Z}_2$  gauge symmetry (16), as it coincides with the leading nontrivial contribution to an expansion of the Helmholtz free energy around a fixed background. To recuperate the  $\mathbb{Z}_2$  symmetry one may replace  $\tau$  in Hamiltonian by  $\frac{1}{2} \sin 2\tau$ . Alternatively, the Hamiltonian (29) can be interpreted as a deformation of the standard energy function of the discrete nonlinear Schrödinger equation (DNLS) [36,37]. The first two sums coincide with the energy of the standard DNLS equation, in terms of the discrete Hasimoto variable of Ref. [10]. The first ( $c$ ) term in the third sum is the Proca mass that has a claim of gauge invariance; here the Proca mass is a “regulator,” as explained in Ref. [13]. The second ( $a$ ) term is the helicity, and the last ( $b$ ) term is the conserved momentum. The last two terms break the  $\mathbb{Z}_2$  parity symmetry; these two terms are responsible for helicity of the homopolymer chain.

The simulations that are described in this article have been performed by keeping some of the parameter values fixed. In Table I the parameter values that are kept fixed have been listed.

The numerical values of these parameters have been chosen in conformity with those that are commonly encountered in the case of proteins [27,38]; for a protein, the torsion angles are much more flexible than bond angles. The value of  $m$  in Table I corresponds to an  $\alpha$ -helical structure.

The parameter  $a$ , which is not fixed, is of particular interest in the sequel. This is the parameter that breaks chirality; note that the momentum of the DNLS hierarchy is not considered here, i.e.,  $b = 0$  for simplicity. This term lacks a direct interpretation in the context of the Abelian Higgs model. It turns out that the effects of this term are largely accounted for by the  $a$ -dependent helicity, in any case.

### E. Pseudogap

In Ref. [10] the following combination of the bond and torsion angles has been considered:

$$\psi_i = \sigma_i e^{i\vartheta_i} \equiv \tan \frac{\kappa_i}{2} e^{i\vartheta_i}, \quad (30)$$

where the phase is

$$\vartheta_i = \frac{1}{2} \left( \sum_{k=1}^i \tau_k - \sum_{k=i+1}^N \tau_k \right). \quad (31)$$

The variable (30) is essentially the discrete version of the Hasimoto variable [10] in terms of the Frenet frame coordinates: It is the complex variable that relates (29) into a generalized version of the discrete nonlinear Schrödinger equation [36,37]. It is also the present version of the complex order parameter (1). Thus the pseudogap state can be identified with a state where the bond angles are nonvanishing and ordered,

$$\langle \kappa_i \rangle = \kappa_0 \neq 0, \quad (32)$$

for some site-independent  $\kappa_0$ , while torsion angles are essentially randomly fluctuating to the effect that

$$\langle e^{i\vartheta_i} \rangle \approx 0. \quad (33)$$

Accordingly in the sequel the pseudogap state is detected by monitoring both  $\kappa_i$  and  $\tau_i$  simultaneously. It should be noted that since the effective potential (27) is insensitive to the phase in (30), the pseudogap state can be difficult to detect in terms of the minima of the effective potential alone. A dynamical computation that engages fluctuations is needed, to detect the presence of the pseudogap.

It should be kept in mind, that a relation such as (32) is commonly deduced by inspection of the effective potential. In the full theory there are always corrections, due to fluctuations. In particular, in the full theory, at a finite temperature, (32) never vanishes identically. The modulus of the order parameter (30) is a positive definite quantity and thus, due to fluctuations, it always acquires a nonvanishing value in the full theory, as also shown in the simulations presented here.

### F. Zero temperature

The zero temperature ground state of the Hamiltonian (29) is a solution to the equations of motion,

$$\tau_i = \frac{a}{c} \frac{1}{d\kappa_i^2 + 1}, \quad (34)$$

$$\kappa_{i+1} = 2\kappa_i - \kappa_{i-1} + 2q(\kappa_i^2 - m^2)\kappa_i + cd\tau_i^2\kappa_i. \quad (35)$$

Accordingly the minimum energy ground state of (29) is

$$\kappa_i = \pm \sqrt{m^2 - \frac{cd}{2q}\tau_i^2} \approx \pm m, \quad (36)$$

$$\tau_i = \frac{a}{2c} \frac{1}{dm^2 + 1} \approx \frac{a}{2c}.$$

In the sequel the parameter  $m$  has the fixed value, given in Table I, throughout. As a consequence the ground state is controlled by the ratio  $a/c$ , and in the sequel the phase structure is investigated by varying the ratio  $a/c$  within the range  $a/c \in [0, 4\pi]$ .

It should be noted that the configuration (36) does not necessarily describe the minimum energy homopolymer: For some parameter values there can be a conflict between the values of  $(\kappa_i, \tau_i)$  given by (36) and the forbidden volume constraint (18). A configuration (36) which satisfies the

forbidden volume constraint is a helix. But if the forbidden volume constraint is not obeyed, the lowest energy ground state configuration is the one that minimizes (29), subject to the constraint (18).

The range of the bond angle become extended to negative values, by the  $\mathbb{Z}_2$  symmetry (16). The two ground states  $\kappa_i = \pm m$  have the same energy. In addition of these two ground states, there can also be local minima of (29) that have the profile of a kink [38,39], i.e., a domain wall that interpolates between the two ground states  $\kappa_i = \pm m$ . The energy of a kink is higher than the energy of the ground state helix (36). Two kinks can annihilate each other, thus any pair of kinks can be removed by continuous deformations of the chain. A single kink can be translated, so that it becomes removed through the ends of the chain. However, on a discrete lattice the translation invariance is commonly broken by the Peierls-Nabarro barrier [40–43]. Thus, in general it costs (thermal) energy to translate a kink along the chain.

The present scenario is different from the one that appears in the case of kinks in folded proteins [38,39]. There the parameter values in (29) are different for different super-secondary structures, i.e., helix-loop-helix motifs. A folded protein is described by a heteropolymer generalization of (29), and the ground state is not a straight helix such as (36). A short analysis of a simple heteropolymer is presented in the sequel, in Sec. III D.

### G. Attractive interaction

The constraint (18) is a purely repulsive interaction, and in the case of a homopolymer it models forbidden volume constraints which have a short spatial range. It could be generalized to include an attractive component, with a short range but one that exceeds the extent of forbidden volume constraints. Accordingly, the following rudimental extension of (18) to model both short-range forbidden volume constraints and attractions is introduced:

$$U(r) = \begin{cases} +\infty & 0 < r < \delta \\ U_0\{\tanh(r - R_0) - 1\} & \delta < r < +\infty \end{cases} \quad (37)$$

Here  $\delta$  is the radius of the self-avoiding condition (18). For  $r < \delta$  the forbidden volume condition (18) persists. But for  $r > \delta$  there is a short-range attractive interaction with strength determined by the parameter  $U_0$ ; when the parameter  $U_0$  vanishes (37) reduces to (18). In the sequel this parameter will be varied, jointly with the ratio  $a/c$  that characterizes helicity.

The parameter  $R_0$  which determines the range of the attractive interaction will have the following value  $R_0 = 5.0 \text{ \AA}$  throughout; this choice is in line with all-atom molecular dynamics simulations where any long-range interactions between atoms are commonly cut off, sharply, beyond distances around  $10 \text{ \AA}$  or so.

The attractive interaction can be given a physical interpretation, in terms of “hydrophobic” forces: In the case of, e.g., a protein chain under physiological conditions, there is an effective attractive interaction between those amino acids which are considered “hydrophobic” [5–7]. Thus in the presence of the attractive interaction (37) the energy function (29) models a chain made of “hydrophobic” residues, with “hydrophobicity” that depends on the value of  $U_0$ .

It is noted that qualitatively, the present results have been found to be quite insensitive to the details of the profile of the potential  $U(r)$ . Accordingly (37) can be considered “universal.”

### H. Monte Carlo algorithms

The protein-folding problem [44–46] is notorious for its computational complexity. A comprehensive all-atom simulation with *Anton* [47,48], which is by far the fastest molecular dynamics machine available, can produce no more than a few microseconds of a folding trajectory per day *in silico*, in the case of proteins with fewer than 100 amino acids. Since many proteins take seconds, even days, to fold into their native state starting from an initial random conformation, it could take thousands of years to fold such a protein with presently available computer resources. Even in the case of effective off-lattice models such as (29) and (18) the simulation of a full folding trajectory is a formidable computational challenge, even with the most powerful computers available. Accordingly, to identify an effective computational scenario, the performance of three different Markovian Monte Carlo algorithms [49] have been tested. The aim has been to identify a Monte Carlo algorithm that has the fastest rate of convergence towards a thermal equilibrium state, in the case of a single polymer chain. In these tests *only* the forbidden volume constraint (18) has been used; the effect of the attractive (“hydrophobic”) interaction has not been included. The three algorithms are the following:

- (1) The Heat Bath algorithm
- (2) The Metropolis algorithm
- (3) The Mixed algorithm.

The curvature and torsion angles are updated according to a probability distribution that satisfies the detailed balance condition

$$P(\{\kappa_{\text{new}}, \tau_{\text{new}}\}, \{\kappa_{\text{old}}, \tau_{\text{old}}\}) \exp[-\beta H(\{\kappa_{\text{old}}, \tau_{\text{old}}\})] \\ = P(\{\kappa_{\text{old}}, \tau_{\text{old}}\}, \{\kappa_{\text{new}}, \tau_{\text{new}}\}) \exp[-\beta H(\{\kappa_{\text{new}}, \tau_{\text{new}}\})].$$

Here  $\beta$  is the inverse Monte Carlo temperature. The equilibrium distribution

$$\exp\{-\beta H(\kappa, \tau)\} \quad (38)$$

of a canonical ensemble is obtained in the limit of an infinite number of updates. Each update consists of a “walk” through the entire chain with a provisional revision of each value  $(\kappa_i, \tau_i)$  which is subject to the requirement that the forbidden volume constraint (18) is preserved; the three algorithms differ from each other only in the manner how the new values  $(\kappa_i^{\text{new}}, \tau_i^{\text{new}})$  are generated.

It should be kept in mind in the sequel that the Monte Carlo temperature  $T = \beta^{-1}$  is *not* equal to the physical temperature factor  $k_B\theta$  where  $k_B$  is the Boltzmann constant and the temperature  $\theta$  is measured on the Kelvin scale.  $T$  is a dimensionless quantity like energy  $H$  [see Eq. (29)] and all parameters in it (see Table I). Instead, in the low-temperature collapsed regime general renormalization group arguments [50] propose that the dimensionless Monte Carlo temperature  $T$  is connected with real physical temperature in

the following way:

$$\ln T = k_B \theta + \dots \quad (39)$$

### 1. Heat Bath algorithm

In the Heat Bath algorithm, new values  $(\kappa_i^{\text{new}}, \tau_i^{\text{new}})$  are generated randomly, according to probability distributions

$$P(\kappa_i^{\text{new}}) = \frac{1}{Z_{i,\kappa}} \exp\{-\beta H_{i,\kappa}(\kappa_i^{\text{new}})\} \quad (40)$$

and

$$P(\tau_i^{\text{new}}) = \frac{1}{Z_{i,\tau}} \exp\{-\beta H_{i,\tau}(\tau_i^{\text{new}})\}. \quad (41)$$

Here  $H_{i,\kappa}$  and  $H_{i,\tau}$  are the sum of all those terms in the Hamiltonian (29) that contain  $\kappa_i$  and  $\tau_i$ , respectively, with the given index  $i$ . The  $Z_{i,\kappa}$  and  $Z_{i,\tau}$  are normalization factors. The updated values of  $\kappa_i^{\text{new}}$  and  $\tau_i^{\text{new}}$  do not depend on the previous values of  $\kappa_i$  and  $\tau_i$ .

The probability density for  $\kappa_i^{\text{new}}$  has the form

$$P(\kappa_i) \sim \exp\{-c_1 \kappa_i^4 - c_2 \kappa_i^2 - c_3 \kappa_i\}, \quad (42)$$

where

$$\begin{aligned} c_1 &= \beta q, \\ c_2 &= \beta \left( 2 - 2qm^2 + \frac{c}{2} d\tau_i^2 - ab\tau_i \right), \\ c_3 &= \beta [-2(\kappa_{i+1} + \kappa_{i-1})]. \end{aligned} \quad (43)$$

Thus (42) is non-Gaussian. On the other hand, the probability density  $P(\tau_i^{\text{new}})$  has the Gaussian profile

$$P(\tau_i) \sim \exp\left\{-\beta \left( \frac{c}{2} [d\kappa_i^2 + 1] \tau_i^2 - a\tau_i \right)\right\}. \quad (44)$$

Rejection sampling has been used to generate random numbers according to these probability distribution: After generating  $\kappa_i^{\text{new}}$  and  $\tau_i^{\text{new}}$  the forbidden volume condition (18) is checked, and the update is rejected when the condition is violated.

### 2. Metropolis algorithm

New values of  $\kappa$  and  $\tau$  are generated according to Gaussian probability distributions, which is centered at the old values. The dispersion of each Gaussian can be adjusted, to enhance the convergence of the algorithm. The new values of  $\kappa$  and  $\tau$  are accepted or rejected, in the same manner as in the conventional Metropolis algorithm. For example, in the case of  $\tau_i$  the probability of acceptance of a new value is

$$P(\tau_i) = \min\{1, \exp(-\beta \Delta H)\}, \quad (45)$$

where  $\Delta H$  is the difference of the energy between the new and the old configurations. In addition, the self-avoidance condition (18) is also verified at each step.

### 3. Mixed algorithm

The values of  $\kappa_i^{\text{new}}$  are generated in the same manner as in the Heat Bath algorithm, while for  $\tau_i^{\text{new}}$  the Metropolis algorithm is used. The convergence of the algorithm can be adjusted, by changing the dispersion of the Gaussian distribution in the  $\tau$  update.

### 4. Algorithm comparison

*A priori*, each of the three Monte Carlo algorithms should converge towards the same equilibrium distribution albeit at a different speed. Thus, when the Markovian length is not sufficient and the equilibrium distribution is not yet reached, the result depends on the algorithm and the number of Monte Carlo steps. Accordingly, the three algorithms have been tested and compared, to identify the one with the fastest convergence rate towards a known equilibrium state; for this, the zero temperature ground state described in Sec. II E is utilized. In these tests, the following parameter values have been used in the Hamiltonian (29),

$$\begin{aligned} a &= -1.0 \times 10^{-4}, \\ c &= 1.0 \times 10^{-4}. \end{aligned} \quad (46)$$

These values have been chosen to reproduce a monotonous  $\alpha$ -helical structure in a manner which is consistent with the forbidden volume constraint (18), as the lowest energy conformation; the choice is not unique. The simulations have been performed with varying chain lengths, from  $N = 100$  to  $N = 900$ . In each case, simulated annealing has been used and the initial configuration is always a linear straight rod with  $\kappa_i = \tau_i = 0$ . The initial configuration is first heated to very high-temperature values (up to  $T_{\text{max}} = 100$ ), where the structure is fully thermally randomized. This is followed by a slow cooling period, to the target temperature. The cooling takes place with small temperature steps, with each step equal to  $\Delta T = 0.05$ – $0.5$  in logarithmic scale. After each step,  $10^3$ – $5 \times 10^4$  Monte Carlo updates are performed along the whole chain, to ensure that it becomes thermalized to the ambient temperature. Here the term “thermalized” means that the Markovian chain reaches its equilibrium distribution. We will call “thermalization length” the number of Monte Carlo updates per one step of cooling process. The longer the chain, the longer the thermalization. For each temperature step only the last configuration is used for calculation of observables. Thus during one cooling process one final (thermalized) chain configuration is obtained for each temperature value. The final phase diagrams have been calculated using 5000 Monte Carlo updates per one step of cooling, equal to  $\Delta T = 0.05$  in logarithmic scale. One single cooling procedure takes 6 CPU-hours, producing one configuration for each temperature value. Final statistics is compiled from 128 thermalized chain configurations, for each set of parameter values.

In the sequel the radius of gyration (2) is utilized as the principal observable, to characterize the geometry of the homopolymer chain. Its value is calculated as an average over statistically independent chain configurations, produced during a Monte Carlo process. In Fig. 1 the value of the radius of gyration (2) is compared, as a function of Monte Carlo temperature for a chain with  $N = 100$  monomers. The results obtained using the Metropolis algorithm are found to be very different from those obtained using the Heat Bath and Mixed algorithms; for small dispersion  $\Delta\tau$  the Mixed algorithm coincides with the Metropolis algorithm. But when  $\Delta\tau$  increases, the Mixed algorithm approaches the Heat Bath algorithm as shown in the figure.

There is no *a priori* reason why the results for the three algorithms should be different: The stationary distribution is

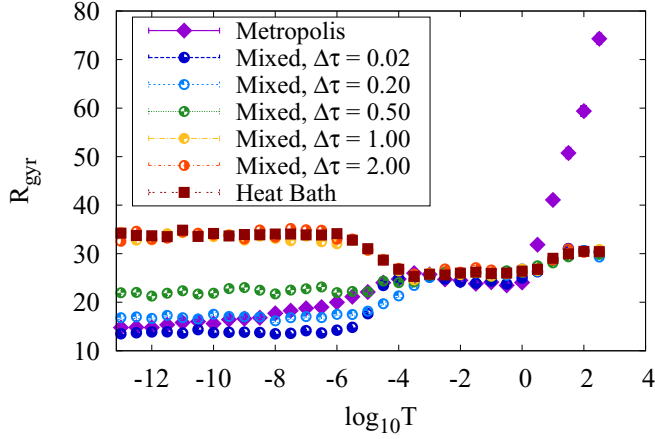


FIG. 1. (Color online) Dependence of  $R_{gyr}$  on temperature for identical homopolymer chains simulated with different algorithms. Here  $\Delta\tau$  is dispersion of Gaussian distribution for generation of  $\tau_{new}$  in a mixed algorithm. In the conventional Metropolis algorithm the dispersions are  $\Delta\kappa = \Delta\tau = 0.01$ . Length of polymer chain is  $N = 100$ , parameters of the Hamiltonian are from Table I.

the same in each of the three algorithms. But it is found that the Metropolis algorithm converges *very* slowly towards the equilibrium distribution. For this, the dependence of the result on the length of simulation has been analyzed. In the case of the Metropolis algorithm, the results are shown in Fig. 2, for  $T = 10^{-13}$  which is the lowest Monte Carlo temperature value that has been used in the present simulations. As shown in the figure,  $R_{gyr}$  continues to increase with increasing length of thermalization. The Metropolis algorithm approaches the Heat Bath algorithm very slowly, as a function of the simulation time, and even at the present, relatively long simulation times the chain is still far from equilibrium distribution.

On the other hand, the results shown in Fig. 3 demonstrate that in case of the mixed algorithm the radius of gyration  $R_{gyr}$

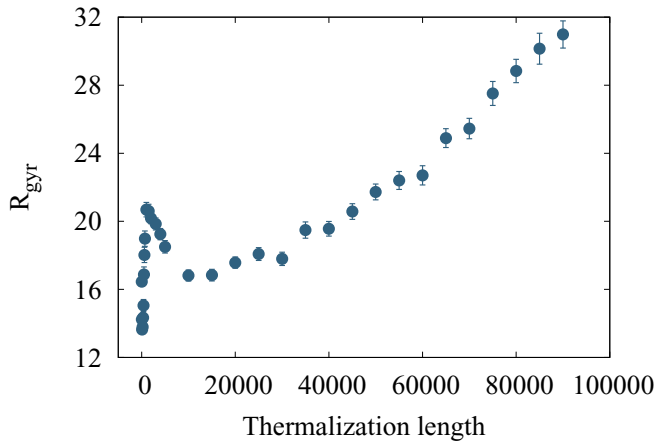


FIG. 2. (Color online) Dependence of  $R_{gyr}$  on the thermalization length in the Metropolis algorithm. The number of updates, per one step in simulated annealing process (thermalization length), is shown along the horizontal axis. Temperature is  $T = 10^{-12}$ , length of polymer  $N = 100$ , and parameters of the Hamiltonian are from Table I.

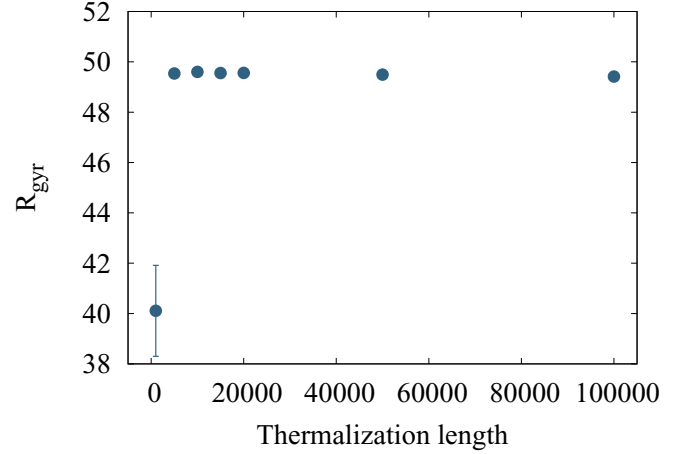


FIG. 3. (Color online) Dependence of  $R_{gyr}$  on the thermalization length in Mixed algorithm. The number of updates, per one step in simulated annealing process (thermalization length), is shown along the horizontal axis. Parameters of this run are the same as for Fig. 2.

approaches a fixed value with increasing of the thermalization length. It is concluded that using the present simulation times, the Markovian homopolymer reaches a stationary distribution. Either the Heat Bath algorithm, or alternatively the Mixed algorithm with sufficiently large  $\Delta\tau$ , should be used to try and describe the thermal equilibrium configurations.

The compactness index  $\nu$ , i.e., the inverse of the Hausdorff dimension has also been inspected, using the three different algorithms. The results for the Heat Bath algorithm are shown in Fig. 4. Between  $T = 100$  and  $T = 1000$  the value of  $\nu$  is essentially temperature independent and apparently corresponds to the SAW phase. At very low temperatures a transition to the rigid rod state, with  $\nu \approx 1$ , is observed. This result is in line with the general arguments that are presented in Sec. II G, on the expected phase structure of the homopolymer model (29) and (18).

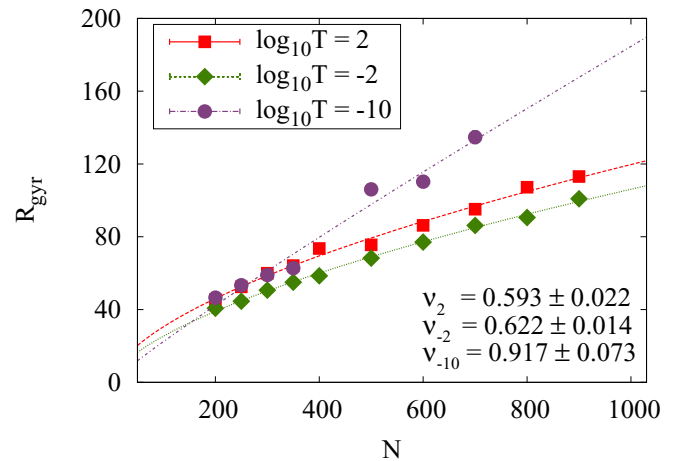


FIG. 4. (Color online) Compactness index  $\nu$  at different temperatures for a homopolymer chain. Simulation performed with Heat Bath algorithm. Parameters of Hamiltonian are from Table I.



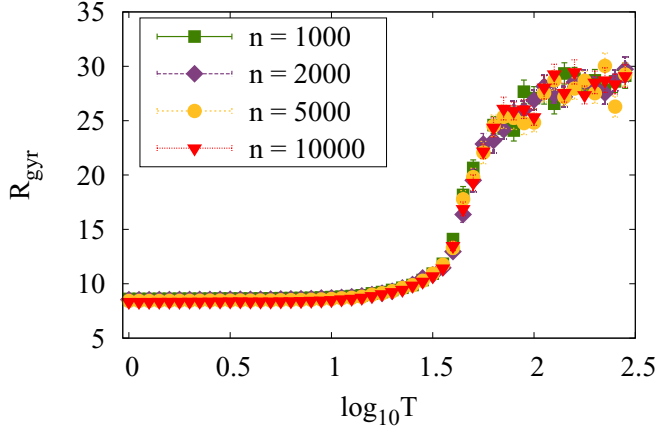


FIG. 5. (Color online) Dependence of  $R_{gyr}$  on temperature  $T$  for a chain with 100 monomers, in the truncated model. Here  $n$  is the number of Monte Carlo updates per one simulated annealing step in temperature (thermalization length). The final algorithm described in Sec. II is used.

### 5. The final algorithm

On the basis of the results from the test runs, the following improved Heat Bath algorithm is employed in the sequel. The final probability distribution is

$$P = \frac{1}{Z} \exp \left\{ -\beta \left[ H + \sum_{i < j} U(\vec{r}_i - \vec{r}_j) \right] \right\}. \quad (47)$$

Here  $H$  is the Hamiltonian (29) and  $U$  is the potential (37). The Metropolis algorithm is used for acceptance, but with a proposal distribution that coincides with the Heat Bath algorithm: The new values of  $\kappa_i$  and  $\tau_i$  are generated using the distributions (40) and (41). The ensuing homopolymer configuration is then accepted, provided it satisfies both the self-avoidance condition (18) and the Metropolis accept-reject

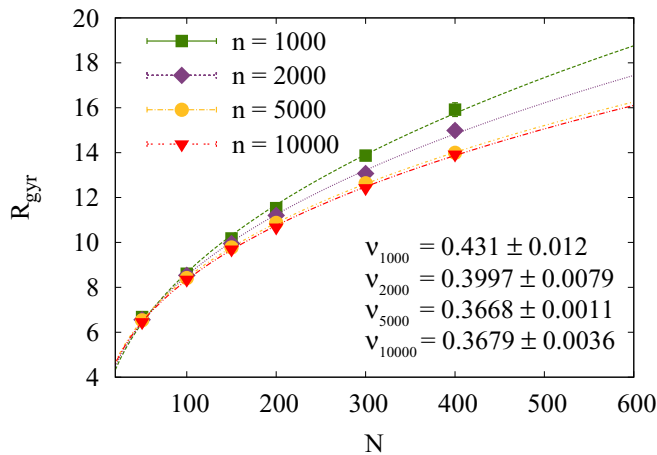


FIG. 6. (Color online) Dependence of  $R_{gyr}$  on the length of polymer chain  $N$  for different thermalization lengths with  $T = 1$ , in the truncated model. The final algorithm is used.

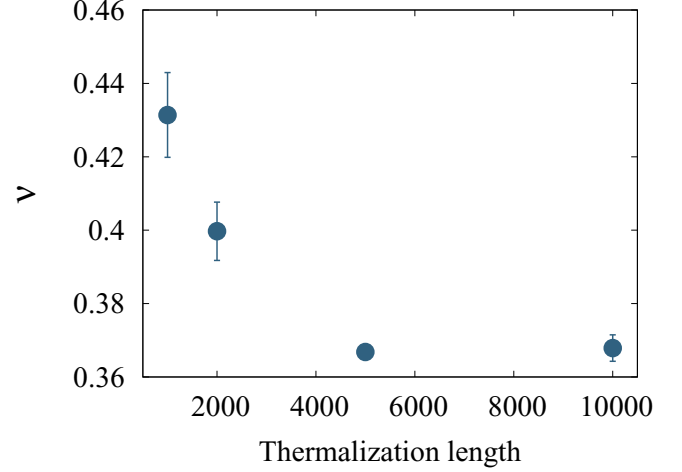


FIG. 7. (Color online) Dependence of compactness index  $\nu$  on thermalization length when  $T = 1$ , in the truncated model.

condition that utilizes the residual energy

$$E_U = \sum_{i < j} U(\vec{r}_i - \vec{r}_j). \quad (48)$$

The acceptance criterion is

$$\exp(-\beta \Delta E_U) > \lambda, \quad (49)$$

where  $\lambda$  is a random number which is uniformly distributed between 0 and 1, and  $\Delta E_U$  is the change in  $E_U$  under the update of  $\kappa_i$  and  $\tau_i$ .

Finally, the algorithm has been calibrated by considering the limit of a *truncated* Hamiltonian, where the Hamiltonian  $H$  is removed and only the attractive potential (37) together with the forbidden volume constraint (18) are retained. The numerical value of  $U_0$  determines solely the scale of the Monte Carlo temperature  $T$ , in the present simulations the value  $U_0 = 15$  is used. The results are shown in Fig. 5. A smooth transition is observed in the radius of gyration, from larger values at high temperatures to smaller values at low temperatures.

Figure 6 shows how  $R_{gyr}$  depends on the length of polymer chain  $N$  and fitted to the leading order contribution in (3).

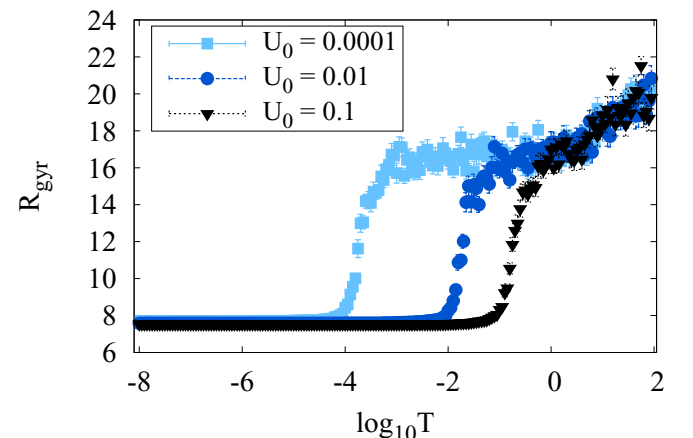


FIG. 8. (Color online) Dependence of  $R_{gyr}$  on temperature for a chain with 50 monomers. Parameters of the Hamiltonian from Table I.

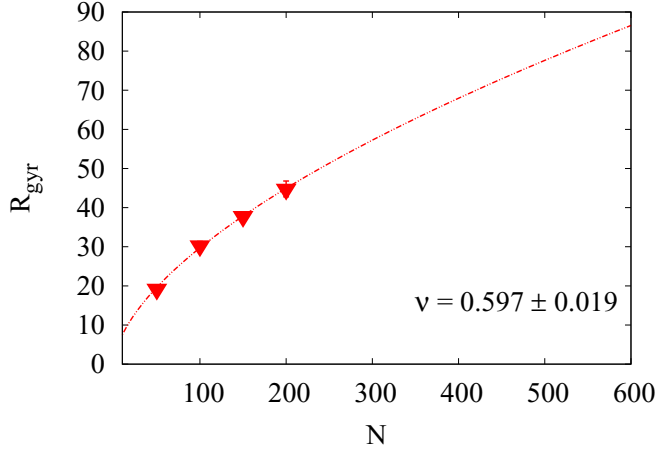


FIG. 9. (Color online) Dependence of  $R_{gyr}$  on the length of chain for high-temperature  $T = 10^{1.7}$  and with  $U_0 = 0.01$ . Parameters of the Hamiltonian from Table I.

The results demonstrate how the equilibrium distribution is reached with a sufficiently long thermalization length, even in the region of low temperatures where the convergence is at its lowest. In particular it is found that the value of the compactness index  $\nu$  converges towards the mean field value  $1/3$  of the collapsed phase, as shown in Fig. 7; the difference between the space filling  $\nu = 1/3$  and the numerically deduced  $\nu \approx 0.36$  is attributed to the finite size corrections in (3); the available computer power does not enable an identification of the amplitudes  $R_1, \dots$  or the critical exponents  $\delta_1, \dots$  in (3).

### III. RESULTS

The present variant of the Heat Bath algorithm has been used in extensive numerical simulations to investigate the phase structure of the homopolymer model (29) and (37). The results are summarized in Figs. 8–12.

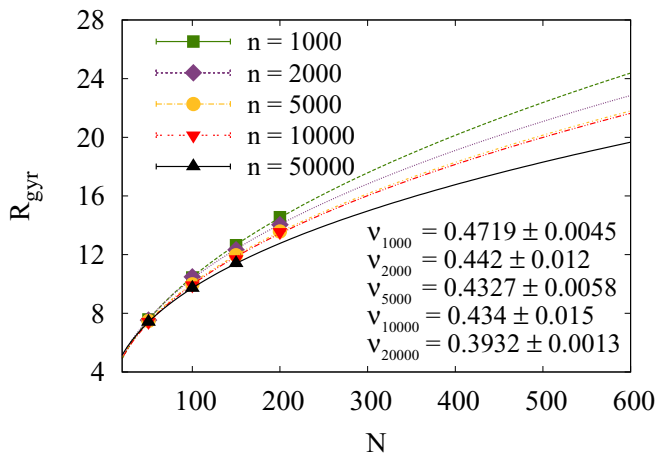


FIG. 10. (Color online) Dependence of  $R_{gyr}$  on the length of chain for low-temperature  $T = 10^{-6}$  and various thermalization lengths, and with  $U_0 = 0.01$ . Parameters of the Hamiltonian from Table I.

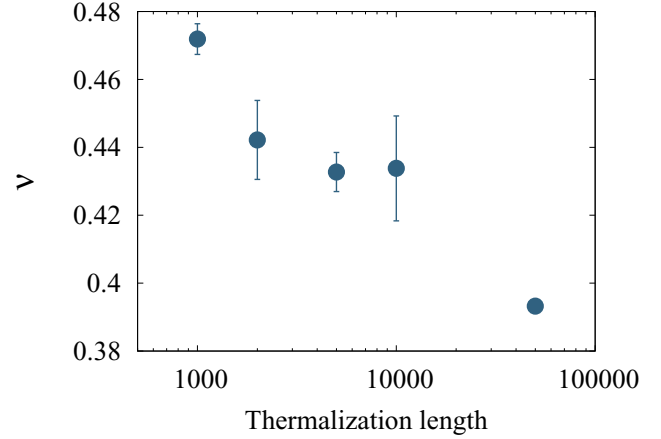


FIG. 11. (Color online) Dependence of compactness index  $\nu$  on thermalization length. Temperature is equal to  $T = 10^{-6}$  and  $U_0 = 0.01$ . Parameters of the Hamiltonian from Table I.

#### A. Effect of parameters

##### 1. Parameter $U_0$

Figures 8–10 describe the properties of the radius of gyration for three representative values of the strength parameter  $U_0$  in (37):

$$U_0 = \begin{cases} 10^{-4} \\ 10^{-2} \\ 10^{-1} \end{cases} \quad (50)$$

For the other parameters the values (46) are used. For each value of  $U_0$ , three different characteristic regimes are observed. In the high-temperature limit the homopolymer is found in the SAW phase; this is confirmed in Fig. 9. When the temperature decreases, the homopolymer enters a regime of decreasing  $R_{gyr}$ . Finally, there is the low-temperature regime where the radius of gyration  $R_{gyr}$  has a small value; see Fig. 10. The compactness index  $\nu$  shows that when the thermalization increases the value of  $\nu$  converges towards the values close to  $\nu \approx 0.39$  which is indicative of the mean field value  $1/3$ . This is shown in Fig. 11. Again, the difference between the mean field value  $\nu = 1/3$  and the measured value  $\nu \approx 0.36$  is

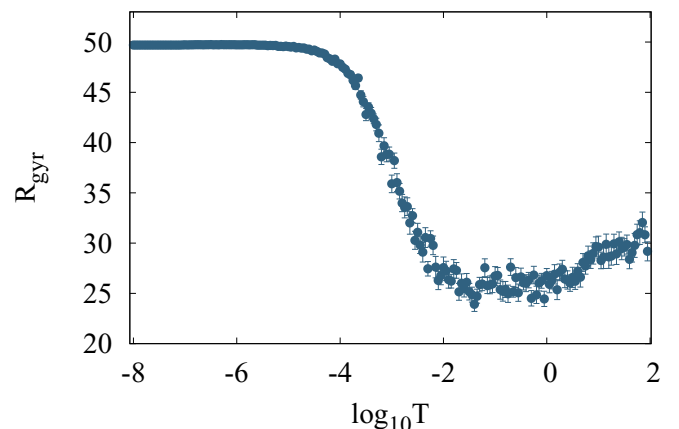


FIG. 12. (Color online) Dependence of  $R_{gyr}$  on temperature. The chain has 100 monomers, and parameters  $(a, c, U_0)$  are given in (51).

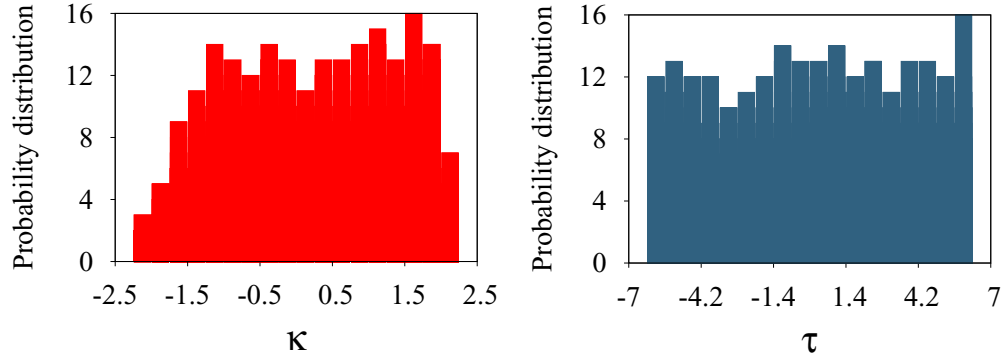


FIG. 13. (Color online) Distribution of  $(\kappa_i, \tau_i)$  in the high-temperature SAW phase. Simulation parameters are the same as for Fig. 8 ( $U_0 = 0.01$ ), temperature is  $T = 100$ .

allocated to finite size corrections in (3); the available computer power is not sufficient to deduce the detailed form of these corrections.

It is concluded that when the scale for  $U_0$  exceeds that of  $a$  and  $c$  the low-temperature state is a collapsed configuration.

2. Parameters  $a$  and  $c$

According to (36) the classical ground state profile of  $\tau_i$  remains intact when the parameters  $a$  and  $c$  are changed in such a manner that the ratio  $a/c$  is constant. To study the effect of such a change in  $a$  and  $c$  at finite temperature, in particular how it conspires with the parameter  $U_0$ , simulations have first been performed with

$$a = c = 10^{-2} \quad \text{and} \quad U_0 = 10^{-4} \quad (51)$$

with the values (46) for the remaining parameters. Thus, unlike in the previous simulations now the characteristic scale of the attractive interaction is smaller than that of the torsion angle-dependent terms in the Hamiltonian.

The results for the radius of gyration are presented in Fig. 12. At high temperatures the chain is again in the SAW phase. Then, as temperature decreases, there is a transition to a regime akin to the intermediate regime shown in Fig. 8. Finally, there is a low-temperature regime where the chain fluctuates around the classical solution (36). The scale of transition to the low-temperature regime is controlled by parameters  $a$  and  $c$  in Hamiltonian.

It is concluded that when the scale  $U_0$  of the short-range attractive interaction is smaller than the scale of the parameters  $a$  and  $c$ , the low-temperature limit is described by helical structures.

B. Analysis of different phase regimes

The bond and torsion angles form the complete set of local order parameters to probe the phase structure (4), in the case of the present homopolymer model. These order parameters have the following characteristics, in the different regimes that have been analyzed in Figs. 8–12; the results are summarized in Figs. 13–16.

In the very high-temperature SAW phase, both the bond angle and the torsion angle are subject to large fluctuations; the simulation results are shown in Fig. 13. For the bond angles, the values are distributed in the range  $0 \leq \kappa \leq \kappa_{\max} \sim 2.2$ . The upper limit reflects the forbidden volume constraint (18). As temperature increases, the values of  $\kappa$  become increasingly evenly distributed over this range so that in the  $T \rightarrow \infty$  limit the distribution is fully uniform; Fig. 13 shows the bond and torsion angle distribution at a generic but high-temperature value. It is apparent from this figure that both angles are *disordered*.

It is concluded that the SAW phase is a *disordered phase*.

Next, we observe the intermediate regime takes place between temperature values within the range  $10^{-3} < T < 10^1$  as can be seen in Figs. 8 and 12. In this intermediate region the

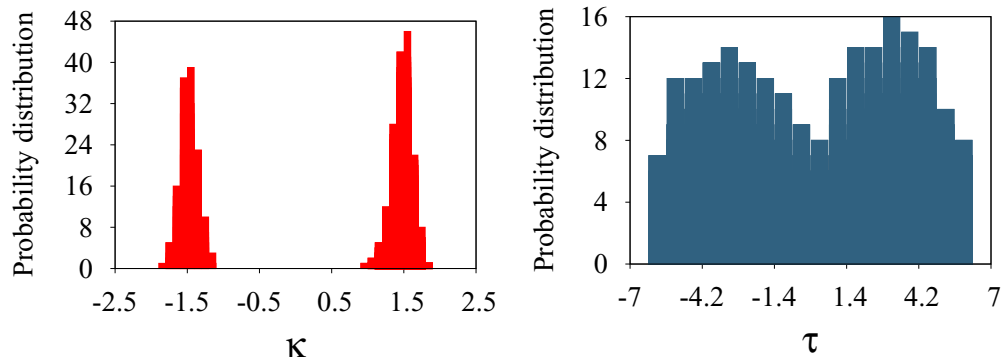


FIG. 14. (Color online) Distribution of  $(\kappa_i, \tau_i)$  in the intermediate transition regime: an example of pseudogap state (see description in Sec. II E). Simulation parameters are the same as for Fig. 8 ( $U_0 = 0.01$ ), temperature is  $T = 1$ .

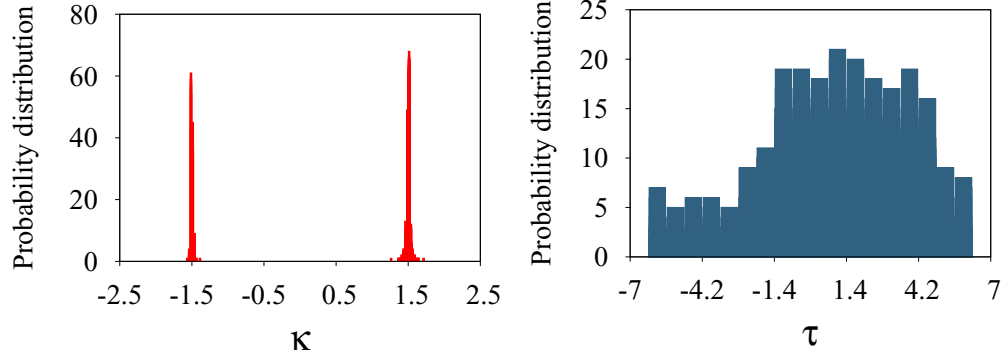


FIG. 15. (Color online) Distribution of  $(\kappa_i, \tau_i)$  in the collapsed phase. The distribution of  $\tau_i$  is asymmetric, corresponding to broken chirality. Simulation parameters are the same as for Fig. 8 ( $U_0 = 0.01$ ), temperature is  $T = 10^{-7}$ .

values of the bond angle are found to become *ordered*. This is shown in Fig. 14: The values of  $\kappa_i$  are thermally fluctuating at around  $|\kappa| \approx 1.5$ . The values of the torsion angle remain largely *disordered*. The intermediate region is identified as a *pseudogap state*, as described in the Introduction.

Finally, there are two different low-temperature phases: The collapsed phase shown in Fig. 8 where the attractive short-distance interaction dominates and the helical rodlike phase shown in Fig. 12 where the attractive short-distance interaction becomes weak.

The distribution of bond and torsion angles in the collapsed phase are displayed in Fig. 15. The bond angle is highly *ordered* around the classical value (36), but the torsion angle remains *disordered*. However, there is an apparent spontaneous symmetry breaking that has taken place; the double-well structure seen in the  $\tau$  distribution of Fig. 14 has been removed, in a way that resembles the familiar spontaneous symmetry breaking in a  $\mathbb{Z}_2$  symmetric potential well.

In the helical rodlike phase, both the bond and torsion angles become peaked around the classical values (36). The configurations are akin straight helical rods; a little like, e.g., collagen when biologically active. The  $\kappa$  distribution reflects the discrete  $\mathbb{Z}_2$  gauge symmetry. But the  $\mathbb{Z}_2$  symmetry in the  $\tau$  distribution observed in Fig. 14 is fully broken.

The transition between the collapsed phase and the helical rodlike phase entails a transition where the torsion angles become ordered. Due to the very low-temperature values involved, the fluctuations are strongly suppressed and a

simulation becomes tedious. It is conjectured that, when the temperature is kept in the low-temperature regime, an initial helical rodlike structure but with parameter values corresponding to the collapsed phase is in a *glassy* phase. *Vice versa*, an initial collapsed configuration with parameter values in the helical rodlike phase will eventually become subject to *cold denaturation*.

### C. Phase diagram

The homopolymer phase is found to depend on three relevant scales:

(1) First, there is the extrinsic temperature scale where the values of  $\kappa_i$  become ordered. This scale can also be controlled intrinsically, by the parameter  $q$  in (29), but the details have not been addressed here.

(2) Then there is temperature scale where the values of  $\tau_i$  become ordered. This scale can be controlled intrinsically, by the parameter ratio  $a/c$  in (29).

(3) Finally, the effects of the scale  $U_0$  for the short-range attractive interactions have been investigated. This parameter determines an intrinsic scale that controls the transition temperature alternatively to the collapsed phase or to the helical rodlike phase.

Thus, by changing the relations between the three scales the phase diagram of the homopolymer can be identified; the phase diagram is constructed here in terms of  $(T, a, U_0)$ . All

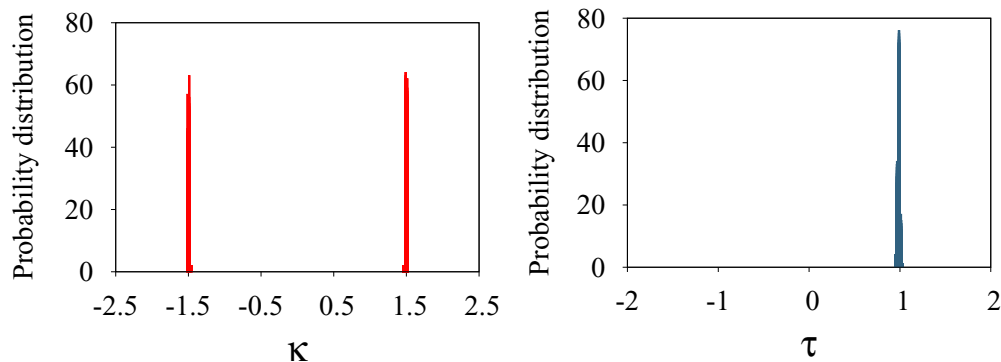


FIG. 16. (Color online) Distribution of  $(\kappa_i, \tau_i)$  in the helical rodlike phase. Simulation parameters are the same as for Fig. 12, temperature is  $T = 10^{-7}$ .

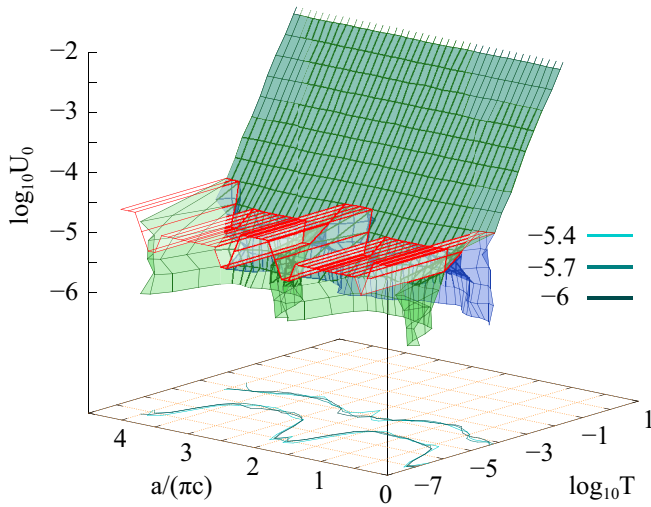


FIG. 17. (Color online) The phase diagram on the  $(U_0, T, a)$  space. It was obtained for the polymer length  $N = 100$  using our final algorithm. All parameters in the Hamiltonian except  $a$  are fixed according to Table I.

the remaining parameters are fixed and given by the values in Table I.

Figure 17 shows the three-dimensional phase diagram in the  $(T, a, U_0)$  space. Figures 18–21 show various cross sections, taken at selected values of  $U_0$ ; i.e., these figures show the phase diagram in Fig. 17 on the  $(a, T)$  plane, with different values of  $U_0$ .

Figure 18 shows the phase diagram, for a quite large value of  $U_0$  (strong coupling) At high temperatures there is the SARW phase. When temperature decreases there is the pseudogap state, which becomes the collapsed phase at low temperatures. Between the pseudogap state and the low-temperature collapsed phase there is a  $\theta$  regime, or rather a  $\theta$  point as it is observed only over a very narrow temperature range.

(1) Figure 19 displays the phase diagram, when the value of  $U_0$  is lowered but still relatively large (intermediate but not weak coupling). At high temperatures there is again the SARW phase, followed by the pseudogap state as the temperature decreases. At low temperatures the pseudogap state becomes converted either to the collapsed phase or to the straight rod

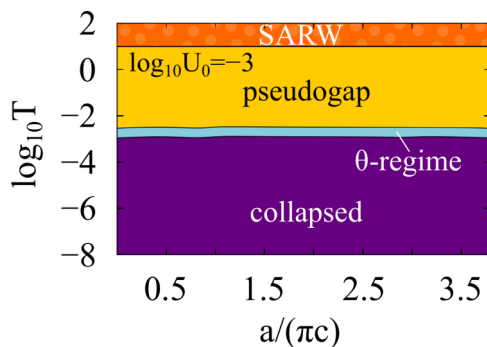


FIG. 18. (Color online) A cross section of the phase diagram in Fig. 17 at  $U_0 = 10^{-3}$ .

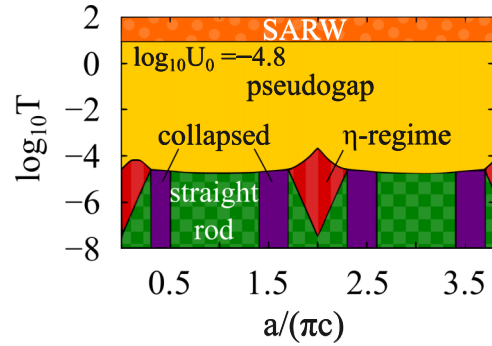


FIG. 19. (Color online) A cross section of the phase diagram in Fig. 17 at  $U_0 = 10^{-4.8}$ .

phase, depending on the value of the helicity parameter  $a$ . In addition, there is a range of values of  $a$ , when a intermediate similar to the  $\theta$  regime is observed between the pseudogap state and the straight rod phase. This is the  $\eta$  regime. It is notable, that there is a possibility of a four-critical point involving the pseudogap state, the  $\eta$  regime, and the collapsed and straight rod phases.

(2) Figure 20 shows the phase diagram, as the value of  $U_0$  becomes further decreased (intermediate but not strong coupling). The collapsed state has entire disappeared and replaced by the straight rod phase at very low temperatures. The  $\eta$  regime displays a periodic structure in the parameter  $a$ . There appears to be a tricritical point involving the pseudogap state, the  $\eta$  regime, and the straight rod phase.

(3) Finally, in Fig. 21 the weak coupling  $U_0$  phase diagram is displayed. The overall topology of the phase diagram is similar to the one in Fig. 18, but with the straight rod phase as the low-temperature phase instead of the collapsed phase.

Figure 22 shows the phase diagram on the  $(U_0, T)$  plane, with helicity fixed. It is notable that there might be a five-criticality involving the  $\eta$  and  $\theta$  regimes, the pseudogap state and the collapsed and straight rod phases. However, the detailed investigation of this region of the phase diagram is beyond the capacity of the computer power which is presently available to us.

In summary, it is remarkable how the phase diagram (Figs. 17–21) is periodic in the helicity parameter  $a$  (when  $c$  is fixed). Moreover, there is a rapid transition between the helical

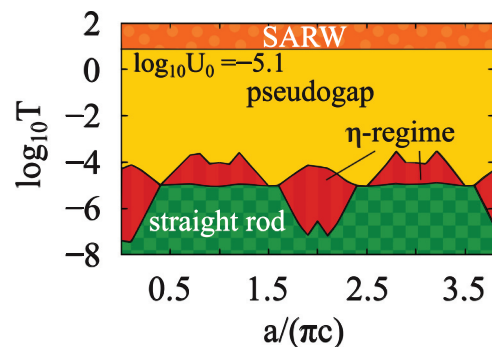


FIG. 20. (Color online) A cross section of the phase diagram in Fig. 17 at  $U_0 = 10^{-5.1}$ .

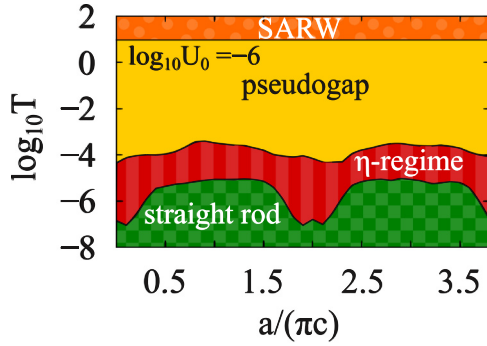


FIG. 21. (Color online) A cross section of the phase diagram in Fig. 17 at  $U_0 = 10^{-6}$ .

phase where the compactness index  $\nu = 1$  and the collapsed phase where  $\nu \approx 1/3$  at low temperatures, as shown in Fig. 19. There is a pseudogap state that appears as a transition regime, akin to the conventional  $\theta$  regime, between the collapsed and the SARW phases. For the transition regime between SARW and helical phases shown in Fig. 20 there is a pseudogap state that should essentially coincide in its properties with the  $\theta$  regime pseudogap state. But due to computational limitations the present analysis is not sufficient to confirm this. There is an apparent four-critical, even five-critical point as shown in Figs. 19 and 22, but the detailed analysis of this region in the phase diagram needs to be performed using more extensive simulations, which is postponed to a future project. However, it is observed that the potential presence of a four-critical point in a theoretical context *very* similar to the present one (i.e., Abelian Higgs model) has been reported in Ref. [31].

#### D. Heteropolymer and proteins

Finally, it is inquired how the present results could be extended to heteropolymers, to draw conclusions on the potential phase structure of proteins. For this, the effect of perturbations that break the homogeneity of the homopolymer model have been investigated as follows: A collapse has been found to take place when the parameter  $U_0$  that characterizes the strength of self-interaction is larger than the parameters  $a$  and  $c$  that characterize the torsion angle-dependent terms. Thus, a short segment is introduced along the chain, where  $U_0$  is less than  $a$  and  $c$ . Accordingly, a simulation is performed where a heteropolymer is constructed so that for a short

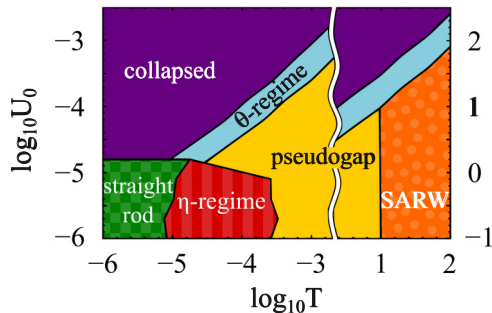


FIG. 22. (Color online) The phase diagram on the  $(U_0, T)$  plane, with  $a = 10^{-4}$ .

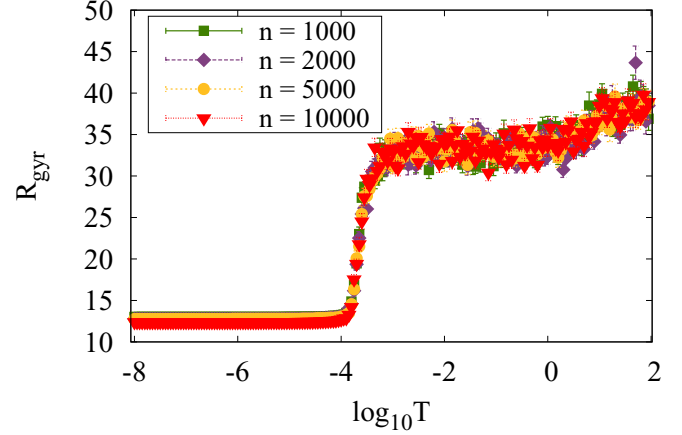


FIG. 23. (Color online) Dependence of  $R_{gyr}$  on temperature for a heteropolymer chain with 150 monomers. Thermalization length  $n$  is number of updates per one step of simulated annealing. Parameters of the Hamiltonian and attraction potential are described in Sec. III D.

subchain of 12 monomers, the values of the parameters  $a$  and  $c$  is increased from  $a = c = 10^{-6}$  to  $a = c = 10^{-2}$ , while  $U_0 = 10^{-4}$  along the entire chain.

Results of simulations are presented in Figs. 23 and 24, for a chain with 150 monomers. The dependence of  $R_{gyr}$  on temperature is shown in Fig. 23; the phase diagram is very similar to eight. For example, at low temperatures it is found that the compactness index is close to the mean field value  $\nu = 1/3$  of the collapsed phase. However, the geometry of a configuration in the collapsed phase is different: As shown in Fig. 25 a helical structure appears only in that subchain where the parameter values  $a$  and  $c$  have been increased.

#### IV. SUMMARY

The phase structure of chiral homopolymers under varying ambient temperature values have been investigated theoretically, in terms of a universal infrared limit energy function in combination with forbidden volume constraints (self-avoidance) and a short-range attractive interaction between

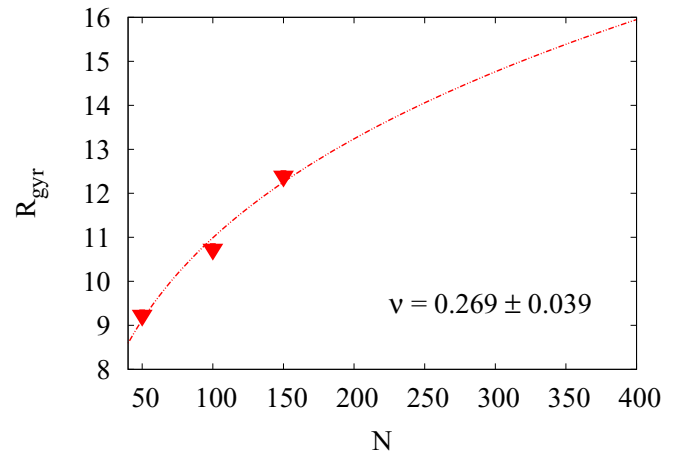


FIG. 24. (Color online) Dependence of  $R_{gyr}$  on the length of polymer chain for low-temperature  $T = 10^{-7.5}$ . Simulation parameters are the same as for Fig. 23.

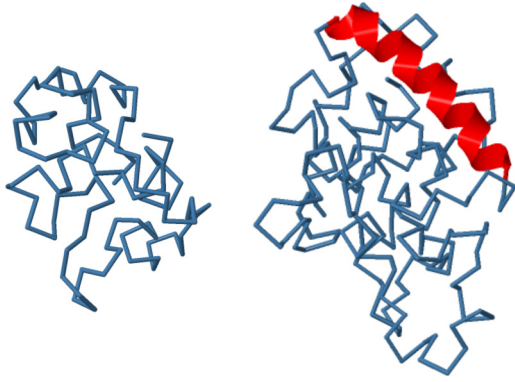


FIG. 25. (Color online) (Left) Homopolymer does not display helices in the ground state. (Right) Addition of a subchain that breaks homogeneity, gives rise to a helix in the subchain region. Homopolymer configuration was obtained in simulation under the parameters inside the collapsed region of the phase diagram (see Fig. 17). In case of heteropolymer the simulation parameters are the same as for Fig. 23, dimensionless temperature  $T$  (see Sec. II H) is equal to  $10^{-7.5}$ .

residues. As such, the model should provide a realistic description even in the case of heteropolymers that display an approximately repeating monomer pattern, provided the scale of the repeat can be considered small in comparison to the chain length. A biologically important example is given by collagen, the most prevalent protein in a human body, in which case the short-range attractive interaction models weak hydrophobicity of the amino acids.

It is found that the phase diagram displays a high level of complexity, in terms of the parameters that control the helicity, and the strength of the attractive interaction. In particular, the low-energy phase is either like a linear one-dimensional straight rod, or a space-filling collapsed configuration. At intermediate temperatures, there is a state which can be identified as an example of the pseudogap state, and there are also intermediates that are more like the conventional  $\theta$  regime. It is possible that these regimes merge, in the thermodynamical limit, at least for some range of parameter values. However, the possibility of the existence of four-critical, even five-critical points in the phase diagram is also proposed, but cannot be confirmed with presently available computer power.

The extension of the present approach to investigate properties of proteins and other heteropolymers remains a challenge to future research.

#### ACKNOWLEDGMENTS

A.J.N. acknowledges support from Region Centre Recherche d'Initiative Academique grant, Vetenskapsrådet, Carl Trygger's Stiftelse för vetenskaplig forskning, and Qian Ren Grant at BIT. The work of M.U. was supported by the DFG grant SFB/TR-55 and by Grant RFBR-14-02-01261-a. Numerical calculations were performed at the ITEP computer systems "Graphyn" and "Stakan" and at the supercomputer center of Moscow State University. A.J.N. thanks A. Sieradzan and Nevena Ilieva for discussions. We are grateful to the late M.I. Polikarpov for playing an important role in establishing our collaboration and for participating in fruitful discussions in the early stages of our work.

- 
- [1] A. J. Heeger, *Rev. Mod. Phys.* **73**, 681 (2001).
  - [2] A. J. Niemi and G. W. Semenoff, *Phys. Rep.* **135**, 99 (1986).
  - [3] J. Hiraki, *J. Antib. Antif. Ag.* **23**, 349 (1995).
  - [4] C. Li, D. F. Yu, A. Newman, F. Cabral, C. Stephens, N. R. Hunter, L. Milas, and S. Wallace, *Cancer Res.* **58**, 2404 (1998).
  - [5] P. G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).
  - [6] A. Yu. Grosberg and A. R. Khokhlov *Statistical Physics of Macromolecules*, AIP Series in Polymers and Complex Materials (AIP, Woodbury, NY, 1994).
  - [7] L. Schäfer, *Excluded Volume Effects in Polymer Solutions, as Explained by the Renormalization Group* (Springer-Verlag, Berlin, 1999).
  - [8] A. J. Niemi, *Phys. Rev. D* **67**, 106004 (2003).
  - [9] U. H. Danielsson, M. Lundgren, and A. J. Niemi, *Phys. Rev. E* **82**, 021910 (2010).
  - [10] S. Hu, Y. Jiang, and A. J. Niemi, *Phys. Rev. D* **87**, 105011 (2013).
  - [11] T. Ioannidou, Y. Jiang, and A. J. Niemi, *Phys. Rev. D* **90**, 025012 (2014).
  - [12] A. J. Niemi, *Theor. Math. Phys.* **181**, 1235 (2014).
  - [13] A. J. Niemi, [arXiv:1412.8321](https://arxiv.org/abs/1412.8321) [cond-mat.soft].
  - [14] V. J. Emery and S. A. Kivelson, *Nature (London)* **374**, 434 (1995).
  - [15] J. Corson, R. Mallozzi, J. Orenstein, J. N. Eckstein, and I. Bozovic, *Nature (London)* **398**, 221 (1999).
  - [16] N. Mannella *et al.*, *Nature (London)* **438**, 474 (2005).
  - [17] H. Kleinert and E. Babaev, *Phys. Lett. B* **438**, 311 (1998).
  - [18] K. Zarembo, *JETP Lett.* **75**, 59 (2002).
  - [19] L. P. Kadanoff, *Physics* **2**, 263 (1966).
  - [20] K. G. Wilson, *Phys. Rev. B* **4**, 3174 (1971).
  - [21] K. G. Wilson and J. Kogut, *Phys. Rep.* **12**, 75 (1974).
  - [22] B. Widom, *J. Chem. Phys.* **43**, 3892 (1965).
  - [23] M. E. Fisher, *Rev. Mod. Phys.* **46**, 597 (1974).
  - [24] P. G. De Gennes, *Phys. Lett. A* **38**, 339 (1972).
  - [25] J. C. LeGuillou and J. Zinn-Justin, *Phys. Rev. B* **21**, 3976 (1980).
  - [26] B. Li, N. Madras, and A. Sokal, *J. Stat. Phys.* **80**, 661 (1995).
  - [27] A. Krokhotin, A. Liwo, A. J. Niemi, and H. A. Scheraga, *J. Chem. Phys.* **137**, 035101 (2012).
  - [28] M. L. Huggins, *J. Chem. Phys.* **9**, 440 (1941).
  - [29] P. J. Flory, *J. Chem. Phys.* **9**, 660 (1941).
  - [30] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY, 1953).
  - [31] M. Bock, M. N. Chernodub, E.-M. Ilgenfritz, and A. Schiller, *Phys. Rev. B* **76**, 184502 (2007).
  - [32] S. Hu, M. Lundgren, and A. J. Niemi, *Phys. Rev. E* **83**, 061908 (2011).
  - [33] K. Hinsen, S. Hu, G. R. Kneller, and A. J. Niemi, *J. Chem. Phys.* **139**, 124115 (2013).

- [34] S. Coleman and E. Weinberg, *Phys. Rev. D* **7**, 1888 (1973).
- [35] M. N. Chernodub, L. Faddeev, and A. J. Niemi, *J. High Energy Phys.* **12** (2008) 014.
- [36] L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer-Verlag, Berlin, 1987).
- [37] M. J. Ablowitz, B. Prinari, and A. D. Trubatch, *Discrete and Continuous Nonlinear Schrödinger Systems*, London Mat. Soc. Lect. Note Series 302 (Cambridge University Press, London, 2003).
- [38] N. Molkenhain, S. Hu, and A. J. Niemi, *Phys. Rev. Lett.* **106**, 078102 (2011).
- [39] M. Chernodub, S. Hu, and A. J. Niemi, *Phys. Rev. E* **82**, 011916 (2010).
- [40] R. Peierls, *Proc. Phys. Soc.* **52**, 34 (1940).
- [41] F. Nabarro, *Proc. Phys. Soc.* **59**, 256 (1947).
- [42] F. Nabarro, *Mat. Sci. Eng. A* **234**, 67 (1997).
- [43] A. K. Sieradzan, A. J. Niemi, and X. Peng, *Phys. Rev. E* **90**, 062717 (2014).
- [44] P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, *Nat. Phys.* **6**, 751 (2010).
- [45] K. A. Dill and J. K. MacCallum, *Science* **338**, 1042 (2012).
- [46] S. Piana, J. L. Klepeis, and D. E. Shaw, *Curr. Opin. Struc. Biol.* **24**, 98 (2014).
- [47] D. E. Shaw *et al.*, *Commun. ACM* **51**, 91 (2008).
- [48] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, and B. Towles, in *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)* (ACM Press, New York, 2009).
- [49] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).
- [50] A. Krokhotin, M. Lundgren, A. J. Niemi, and X. Peng, *J. Phys.: Condens. Matter* **25**, 325103 (2013).