Elastic energy and phase structure in a continuous spin Ising chain with applications to chiral homopolymers

M. N. Chernodub,[1,2,*] Martin Lundgren,[3,†] and Antti J. Niemi[1,3,‡]

[1]*Laboratoire de Mathématiques et Physique Théorique, Université François-Rabelais Tours, Fédération Denis Poisson-CNRS, Parc de Grandmont, F-37200 Tours, France*

[2]*Department of Mathematical Physics and Astronomy, Krijgslaan 281, 59, Gent, B-9000, Belgium*

[3]*Department of Physics and Astronomy, Uppsala University, P.O. Box 803, S-75108, Uppsala, Sweden*

We present a numerical Monte Carlo analysis of the phase structure in a continuous spin Ising chain that describes chiral homopolymers. We find that depending on the value of the Metropolis temperature, the model displays the three known nontrivial phases of polymers: At low temperatures the model is in a collapsed phase, at medium temperatures it is in a random walk phase, and at high temperatures it enters the self-avoiding random walk phase. By investigating the temperature dependence of the specific energy we confirm that the transition between the collapsed phase and the random walk phase is a phase transition, while the random walk phase and self-avoiding random walk phase are separated from each other by a crossover transition. We propose that the model can be applied to characterize the statistical properties of protein folding. For this we compare the predictions of the model to a phenomenological elastic energy formula, proposed by J. Lei and K. Huang [e-print arXiv:1002.5013; Europhys. Lett. **88**, 68004 (2009)] to describe folded proteins.

PACS number(s): 05.70.−a, 05.10.Ln, 87.15.A−, 87.15.Cc

I. INTRODUCTION

The concept of universality [1,2] divides critical physical systems into universality classes that differ from each other essentially only by their space-time dimensionality and the symmetry group of their order parameter. This enables the computation of critical properties for an entire class of physical systems using only a single representative model. In the case of polymers one expects that there are three different *nontrivial* phases and these correspond to the universality class of self-avoiding random walk (SARW), the universality class of Brownian motion, that is, ordinary random walk (RW), and the universality class of polymer collapse [3]. These phases are each characterized by the different values of certain critical exponents that describe the scaling properties of the polymer in the limit where the number N of monomers becomes large. The most widely used critical exponent, the compactness index ν, computes the inverse of the Hausdorff dimension of the polymer. It can be introduced by considering how the polymer's radius of gyration R_g increases in the number of monomers, asymptotically for large values of N [4]:

$$R_g^2 = \frac{1}{2N^2} \sum_{i,j} (\mathbf{r}_i - \mathbf{r}_j)^2 \approx R_0^2 N^{2\nu} (1 + \beta_1 N^{-\Delta_1} + \cdots). \quad (1)$$

Here $\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$ are the locations of the N monomers in $\mathbb{R}^3$. The critical exponents ν and Δ_1 are universal quantities. However, the form factor R_0 that characterizes the effective distance between the monomers in the large N limit and the amplitude β_1 that parametrizes the leading finite size corrections are not. The asymptotic expansion (1) is an example of a general result [5,6] that states, that when N becomes large the mean value of any global observable $\mathcal{O}$ of a polymer should behave like

$$\langle \mathcal{O} \rangle_N = A N^p \Bigg[1 + \frac{\alpha_1}{N} + \frac{\alpha_2}{N^2} + \cdots + \frac{\beta_1}{N^{\Delta_1}} + \frac{\beta_2}{N^{\Delta_1+1}} + \frac{\beta_3}{N^{\Delta_1+2}} + \cdots \Bigg], \quad (2)$$

where the exponents are universal, but the prefactor and the various amplitudes are all nonuniversal.

For a polymer the compactness index has the following mean field (mf) values [3]:

$$\nu_{\mathrm{mf}} = \begin{cases} 3/5 & \text{SARW}, \\ 1/2 & \text{RW}, \\ 1/3 & \text{collapsed}. \end{cases} \quad (3)$$

As a function of temperature, the collapsed phase occurs at low temperatures (bad solvent) while the SARW describes the high-temperature (good solvent) behavior of polymers. The RW phase takes place at the Θ temperature that separates the SARW phase from the collapsed phase. In general, the mean field values of the critical exponents acquire corrections due to fluctuations, and for the universality class of the SARW the improved values are $\nu = 0.5880 \pm 0.0015$ and $\Delta_1 = 0.47 \pm 0.03$. These values were obtained in [7] by utilizing the concept of universality that relates the SARW with the $n \to 0$ component ϕ^4 field theory [8]. The subsequent direct Monte Carlo evaluation reported in [6] gave the very similar values $\nu = 0.5877 \pm 0.0006$ and $\Delta_1 = 0.56 \pm 0.03$, in line with the concept of universality.

Qualitatively, at the level of a mean field theory the phase structure of a polymer can be described in terms of the Flory-Huggins theory [3]. For this we characterize the polymer concentration by an order parameter $\phi(x)$, with $0 \leqslant \phi(x) \leqslant 1$.

———————

*maxim.chernodub@lmpt.univ-tours.fr; On leave from ITEP, Moscow, Russia.

†martin.lundgren@physics.uu.se

‡antti.niemi@physics.uu.se

At low concentrations the polymer free energy density (per temperature) has the Landau expansion

$$\frac{1}{T}E[\phi] = \eta(\nabla\phi)^2 + \gamma \cdot \phi \ln\phi + \frac{1}{2}(1 - 2\chi)\phi^2 + \frac{g}{3!}\phi^3 + \cdots.$$

(4)

Here η, γ, χ, and g are parameters. The first term is a stiffness term. The second term describes entropy contributions. The third term describes monomer-monomer interactions; the (Flory) interaction parameter χ is generically a decreasing function of temperature. The last term characterizes the three-body (and higher-order contributions) monomer interactions. The phase structure can be exposed by ignoring the stiffness term and by minimizing the remaining potential energy contribution to free energy. With proper relative values of the parameters the potential has a form that is familiar from spontaneous symmetry breaking: When the ground state expectation value $\langle\phi\rangle$ is nonvanishing we are in the collapsed phase while the vanishing value $\langle\phi\rangle \approx 0$ implies that we are in the universality class of SARW. The borderline that separates these two phases determines the Θ temperature where the polymer is in the universality class of RW. It occurs at that value of temperature (or denaturant concentration) for which the excluded volume parameter vanishes, and to first order

$$1 - 2\chi(T_\Theta) = 0.$$

Thus, for $\chi(T) > 1/2$ we are in the collapsed phase, while for $\chi(T) < 1/2$ we enter the SARW phase and, in particular, at the Θ point the ϕ^2 (i.e., mass) contribution to the free energy is absent.

Here we present results of a comprehensive numerical analysis of polymer phase structure in the context of the continuous spin Ising model introduced in [9] to describe chiral homopolymers. The applicability of the model to analyze the properties of (chiral) polymers in all three phases can be justified by the concept of universality. Indeed, the derivation of the model in [9] is very much based on the universality concept: The model should account for the monomer complexity such as the presence of amino acid side chains in polymers such as proteins and polymer-solvent interactions in an effective manner for the purpose of statistical analyses. In particular, it has been shown [9] that the model indeed appears to describe certain universal properties of the folded proteins [10] in the Protein Data Bank (PDB) [11] with very high accuracy. This also motivates us to compare our results to a recently presented phenomenological model of protein folding [12].

II. THE MODEL

The model introduced in [9] is based on a discretized version of the standard Abelian Higgs model. It has the following internal energy:

$$E = \sum_{ij} a_{ij}\{1 - \cos[\omega_{ij}(\kappa_i - \kappa_j)]\}$$

$$+ \sum_i \left\{ b_i\kappa_i^2\tau_i^2 + c_i\left(\kappa_i^2 - m_i^2\right)^2 + d_i\tau_i \right\}.$$

(5)

Here $i, j = 1, \ldots, N$ label the N monomers of a (chiral) polygonal chain in $\mathbb{R}^3$. These monomers are located at the

vertices of the polygon, and the chain geometry changes when the polymer fluctuates in $\mathbb{R}^3$. The geometry is determined by the order parameter κ_i that is a discrete lattice version of the Frenet curvature and by the order parameter τ_i that is the lattice version of the Frenet torsion [9]. Notice that there is no quadratic term in τ_i; in the underlying Abelian Higgs model [9] such a term would correspond to the Proca mass of a gauge vector.

Once the values of (κ_i, τ_i) for each $i = 1, \ldots, N$ are given, the actual shape of the polymer as a polygonal chain in the three-dimensional space can be computed by integrating the appropriate discrete version of the Frenet equations. This integration introduces parameters δ_i, the three-dimensional distances between the monomers.

The $a_{ij}, \omega_{ij}, b_i, \ldots, d_i$ in (5) are parameters. The first sum in the free energy describes long-distance interactions; we have introduced the cosine function to tame excessive fluctuations in κ_i in the numerical simulations. In the second sum the first term describes the interaction between κ_i and τ_i, and the second term describes the self-interaction of κ_i. Finally, the last term is a discretized version of the one-dimensional Chern-Simons functional; it is the origin of chirality in the polymer chain [9], with handedness that depends on the sign of d_i.

For a general polymer the quantities $(a_{ij}, \omega_{ij}, b_i, c_i, \mu_i, d_i)$ are *a priori* site-dependent parameters, and different values of these parameters can be used to describe different kinds of monomer structures, amino acids in the case of proteins. For generic a_{ij} (5) is a spin-class model. Here we are interested in the limiting case of a homopolymer where we restrict ourselves to only the nearest neighbor interactions with

$$a_{ij} = \begin{cases} a & i = 1, \ldots, N-1, \quad j = i + 1, \\ 0 & \text{otherwise} \end{cases}$$

(6)

and we also select *all* the remaining parameters to be *independent* of the site index i. Thus, the model in the form studied here reads

$$E = \sum_{n.n.} a\{1 - \cos[\omega(\kappa_i - \kappa_j)]\}$$

$$+ \sum_i \left\{ b\kappa_i^2\tau_i^2 + c\left(\kappa_i^2 - m^2\right)^2 + d\tau_i \right\},$$

(7)

where the first sum extends over the nearest neighbors. Notice that since the overall scale of the parameters a, b, c, and d can be absorbed into the definition of the scale of the Metropolis temperature T, as it stands there are five independent intrinsic parameters. Consequently, the scale of energy, say in electron volts, remains indeterminate and should be defined by (re)normalization at some convenient value of T. We also note that, classically, the model (7) has a ground state which is a helix, with $\kappa_i \approx \pm m$. With the present energy functional, such a configuration is subject to Peierl's instability.

For concreteness, we select the numerical values of the parameters following [9]. There, the parameters were selected so that model reproduces the average values of PDB proteins such as the average numbers of central carbons in a single full (2π) turn in an α helix, the average length of turns between loops, correct chirality, etc. However, we also emphasize that universal quantities are expected to be independent of the actual parameter values. The parameter values that we use are

Parameter	Value
a	4
ω	4.25
b	5.488×10^{-4}
c	0.5
m	24.7
d	-20

shown in Table I. Furthermore, we assume that the distances δ_i between the monomers that we need to introduce when we integrate the discrete Frenet equations to construct the polygonal chain in $\mathbb{R}^3$ have the fixed value

$$|\mathbf{r}_i - \mathbf{r}_{i-1}| = \delta = 3.8 \text{ (Å)}, \quad i = 2, \ldots, N. \tag{8}$$

This value (in Å) is chosen to coincide with the average distance between C_α carbons in the backbone of PDB proteins. Finally, we exclude steric clashes by demanding that the distance between any two monomers satisfies the bound

$$|\mathbf{r}_i - \mathbf{r}_j| \geqslant z = 3.7 \text{ (Å)} \quad \text{for} \quad |i - j| \geqslant 2. \tag{9}$$

Again, this numerical value has been chosen to match the protein data in PDB.

We have used the standard Metropolis algorithm to simulate the model (5). The initial configuration is a straight rod with $\kappa_i = \tau_i = 0$, a configuration that we expect to have a very large conformational entropy, due to Peierl's instability. Each Monte Carlo step consists of a shift of the curvature and torsion by a typical value of $\Delta\kappa_i = \Delta\tau_i = 0.05$. This shift is accepted with the probability

$$P = \min\left[1, \exp\left(-\frac{\Delta E}{T}\right)\right],$$

where T is the Metropolis temperature. We use this temperature as an external parameter that allows us to probe the different phases of the polymer.

The simulations proceeded as follows: For each temperature value, between 10 and 16 different polymer lengths were selected. The number of the Monte Carlo iterations of each chain was chosen to be 11 000 multiplied by the number N of monomers in the polymer. We created around 200 or more polymers for each individual temperature value T and monomer number N, with less for the extremely long and the highest-temperature curves. The shortest polymers in our simulations had 50 monomers, and the longest ones had 1800 monomers. These values were chosen to be representative of the single domain proteins in PDB.

Finally, since the free energy [Eqs. (5) and (7)] is quadratic in τ_i and furthermore since τ_i only interacts locally, we can eliminate it by using its equation of motion

$$\frac{\partial E}{\partial \tau_i} = 2b_i\kappa_i^2\tau_i + d_i = 0$$

$$\Rightarrow \tau_i[\kappa_i] = -\frac{d_i}{2b_i\kappa_i^2}. \tag{10}$$

This gives us

$$E = \sum_{i,j} a_{ij}\{1 - \cos[\omega_{ij}(\kappa_i - \kappa_j)]\}$$

$$+ \sum_i \left\{ c_i \left(\kappa_i^2 - m_i^2\right)^2 - \frac{d_i^2}{2b_i\kappa_i^2} \right\},$$

and in the limit of uniform chain and small ω we get (after we add boundary contributions and choose $\kappa_0 = \kappa_{N+1} = 0$)

$$E \approx -a\omega^2 \sum_i \kappa_i\kappa_{i+1}$$

$$+ \sum_i \left\{ a\omega^2\kappa_i^2 + c\left(\kappa_i^2 - m^2\right)^2 - \frac{d^2}{4b\kappa_i^2} \right\}. \tag{11}$$

We recognize here a version of the continuos spin Ising chain [13]: Indeed, the *only* difference between (11) and the conventional continuous spin Ising chain is in the presence of the last term in (11). We note that this last term, which has its origin in (10), is quite reminiscent of the potential term that appears in the widely studied Calogero model [14] for the relative coordinate in the two-body case. Furthermore, if we absorb the parameter combination $a\omega^2$ into the definition of overall scale of temperature, in (10) and (11) there are only four independent parameter combinations.

It has been a commonly held point of view [15] that the lattice version of the ϕ^4 model is always in the same universality class with the pure Ising model. However, this has been disputed in the one-dimensional case by explicit computations, for example, in [16]. Here we have an additional interaction term, the last Calogero-type term in (11), and we shall explicitly show that the ensuing phase structure is highly nontrivial.

III. THE RADIUS OF GYRATION

We shall first investigate the radius of gyration (1) with the goal to confirm that the model [9] does indeed describe the three different polymer phases characterized by the mean field values (3) of the critical exponent ν. In Fig. 1 we show how the radius of gyration R_g depends on the (Metropolis) temperature
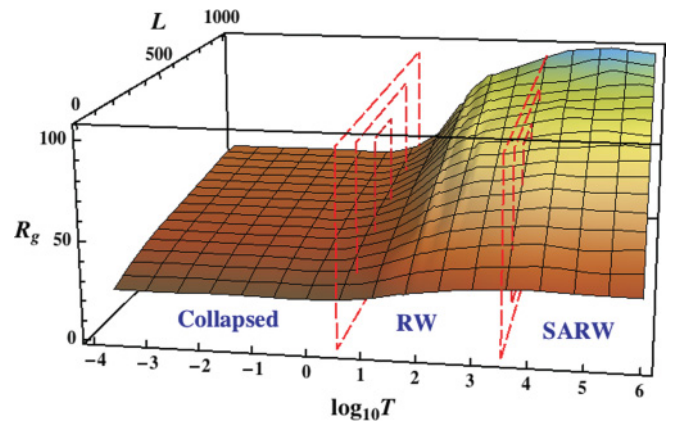

FIG. 1. (Color online) The radius of R_g as a function of temperature T and the number of monomers N. The three putative phases are identified with the putative position of the ensuing critical temperatures $T_{c1} < T_{c2}$, denoted by the vertical planes.
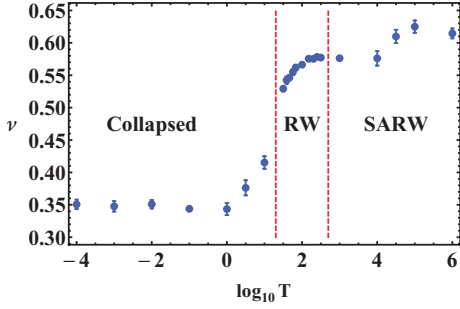
FIG. 2. (Color online) The compactness index ν vs temperature T. The vertical lines correspond to temperature values where $\nu(T)$ reaches the mean field values (3).

T and the number of monomers N. In Fig. 2 we depict the T dependence of ν, and Fig. 3 shows the T dependence of the prefactor R_0 in (1).

A. Collapsed phase

In the Fig. 1 we clearly identify a low-temperature phase which is the putative collapsed phase. In this phase R_g is constant, or has only very weak T dependence, and we can fit the data with very high accuracy using the following relation:

$$R_g = R_0 N^\nu, \tag{12}$$

where R_0 and ν are the fitting parameters. From the data in Fig. 2 we estimate in the low-temperature limit

$$\nu = 0.348 \pm 0.007. \tag{13}$$

This is so close to the mean field value $\nu = 1/3$ of the collapsed phase that obviously we are in that phase.

The parameter R_0 that we present in Fig. 3 describes the *effective* distance between the monomers. In the collapsed phase we estimate in the low-temperature limit

$$R_0 = 3.25 \pm 0.15 \text{ (Å)}.$$

This is clearly smaller than the bare value (8) in our model, proposing that in the collapsed phase the monomers have the tendency to become more densely packed also *along* the polymer chain.

From our data we are not able to deduct any nonvanishing value for the subleading critical exponent Δ_1 in (1).
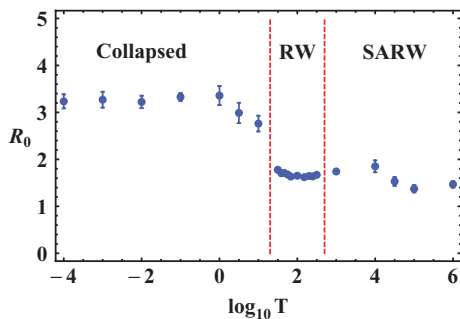

FIG. 3. (Color online) The prefactor R_0 in (1) vs temperature T. The vertical lines correspond to the temperature values where $\nu(T)$ in Fig. 2 reaches the mean field values (3).

When the temperature increases beyond $\log_{10} T \approx 0$, $\nu(T)$ starts increasing and we enter a transition region between the collapsed phase and the putative RW phase. At the same time as the value of ν starts increasing, the value of R_0 decreases, and when temperature approaches the value [17]

$$T_{c1} \approx 3.81 \pm 1.52, \tag{14}$$

we obtain the fit

$$R_g \approx 2.8 N^{0.38} \tag{15}$$

for $50 \leqslant N \leqslant 1.800$, which is very close to the estimate presented in [9]

$$R_g^{PDB} \approx 2.25 N^{0.38}. \tag{16}$$

However, we point out that when $T \approx T_{c1}$ both $\nu(T)$ and $R_0(T)$ have a quite strong temperature dependence, indicative of vicinity of a phase transition that makes the accuracy of our estimates prone to relatively large errors, and for more precise estimates one needs simulations with substantially more computer time.

B. RW and SARW phases

In Fig. 4 we display how the radius of gyration depends on the number of monomers N for a range of values of temperature beyond the collapsed phase and compare the data with a fit of the form (12). As visible in this figure, even beyond the collapsed phase the data can be fitted with very high accuracy by the relation (12). However, unlike in the collapsed phase, where the radius of gyration is practically temperature independent, both in the putative RW phase and in the putative SARW phase the radius of gyration is a slowly but monotonically increasing function of the temperature.

The transition from the collapsed phase to the putative RW phase is very visible in our Figs. 2 and 3. There is a clear, rapid transition in both $\nu(T)$ and $R_0(T)$, reminiscent of a phase transition. From Fig. 2 we estimate that at the transition point ν is very close to the value

$$\nu \approx 0.5,$$

which is the value of compactness index in the case of ideal (noninteracting) RW, that is, the value of the Θ point in mean field theory that also defines the temperature T_Θ [3]. For $T > T_\Theta$, the compactness index $\nu(T)$ is a slowly increasing
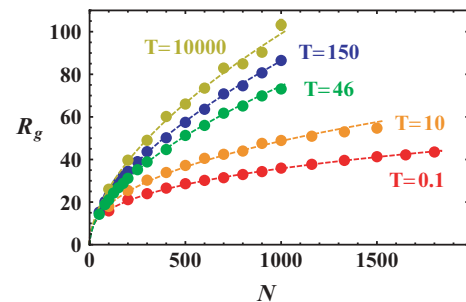

FIG. 4. (Color online) The radius of gyration R_g vs the number of monomers N at different values of temperature T. The error bars are of the same order or smaller than the size of the symbols used. The dashed lines represent the fits by Eq. (12).

function of temperature that eventually plateaus around the value

$$\nu \approx 0.58.$$

This is slightly above the Θ-point value, but slightly below the SARW values reported in [6,7]. Since the compactness index $\nu(N)$ appears to have a tendency to approach its large-N limit from above [6], we conclude that we are in the RW phase.

For the effective monomer distance R_0 we find the value

$$R_0 = 1.67 \pm 0.03 \text{ (Å)},$$

which is *clearly* lower than the bare value (8).

In general one expects that the transition between the collapsed phase and the RW phase is a phase transition, while the transition between the RW and SARW phases is a smooth crossover [3]. The results in Figs. 1–3 are in line with this; the transition between the RW phase and the putative SARW phase is much less dramatic than the transition between the collapsed phase and the RW phase. This also makes the precise identification of the RW and SARW phases more involved.

We find that asymptotically at very high temperatures ν approaches the value

$$\nu \xrightarrow{T \to \infty} 0.62 \pm 0.03.$$

This is slightly above the mean field value and the final values obtained in [6,7] but fully in line with the computations in [6] that revealed that the asymptotic value of ν is reached from above as the number of monomers increases. We note that here we have restricted ourselves to consider only values of N in the range 50–1000 that are relevant for single strand proteins, while [6] considered self-avoiding walks with up to 80 000 steps. Consequently, finite scaling corrections have a much stronger effect on our estimates. We also point out that as $T \to \infty$ only the self-avoiding condition (9) persists. Thus, in this limit we *must* be in the universality class of SARW.

We note that for the effective monomer distance R_0 we find in the high-temperature limit the value

$$R_0 = 1.62 \pm 0.08 \text{ (Å)},$$

that is, essentially the same as in the RW phase.

In summary, the distinction between the collapsed phase and the RW phase appears very clear in our analysis of the compactness index and suggests the presence of either a first- or a second-order phase transition. On the other hand, the transition from RW phase to SARW phase is much more difficult to pinpoint, and it appears to proceed much more like a smooth crossover transition than a phase transition. These observations are fully in line with general expectations [3], and we conclude that the model [9] does indeed correctly describe all the three phases of a polymer.

IV. ELASTIC ENERGY

A. General behavior

For the fixed parameter values that we have given in Table I, the free energy (7) is a function of two extrinsic parameters, the temperature T, and the number of monomers N. Its numerical value can be identified as the *elastic* energy of the polymer chain. In Fig. 5 we display a three-dimensional plot of (a
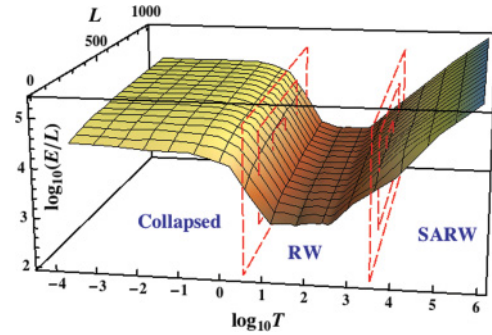

FIG. 5. (Color online) A three-dimensional plot of a logarithm of the specific energy as a function of a logarithm of temperature T and the number of monomers N. As in Fig. 1, the three phases are identified and the putative position of the ensuing two critical temperatures $T_{c1} < T_{c2}$ are denoted by the vertical planes.

logarithm of) the *specific* elastic energy, that is, the elastic energy per monomer as a function of these two parameters:

$$E_{\text{specific}} = \frac{E}{N}. \tag{17}$$

In this figure we clearly identify the presence of three different phases that are separated from each other by clearly identifiable transition (critical) temperatures T_{c1} and T_{c2} (with $T_{c1} < T_{c2}$), and both the low-temperature collapsed phase ($T < T_{c1}$) and the medium-temperature RW phase ($T_{c1} < T < T_{c2}$) are characterized by essentially temperature *independent* specific energy. Notice that in the collapsed phase the specific energy has a value that is more than one order of magnitude larger than in the RW phase. This is understandable as it should indeed take much more energy to extend a polymer that is collapsed and resists being extended than a polymer that behaves like an ideal chain and thus does not really care about its shape. The increase of temperature beyond T_{c2} leads to a transition to the SARW phase, which is characterized by a power-law increase of the specific energy as a function of the temperature: The larger its thermal fluctuations, the more the polymer resists becoming extended. Note also that in the collapsed and RW phases the specific energy in Fig. 5 exhibits a weak dependence on the number of monomers N. However, in the high-temperature SARW phase the specific energy becomes essentially independent of N. This is consistent with the expected behavior of SARW; it is driven solely by the condition (9) and no reference to the details of the free energy survives the infinite temperature limit. In this limit, the polymer is only subject to random thermal fluctuations.

B. Critical temperature

We have found that the dependence of the specific energy on temperature displayed in Fig. 5 can be approximated with a *very* good accuracy by a function,

$$\log_{10}(E/N) = F^{\text{fit}}(\log_{10} T), \tag{18}$$

that has the following explicit form:

$$F^{\text{fit}}(x) = h_1 + h_2 \arctan[h_3(x - x_1)]$$
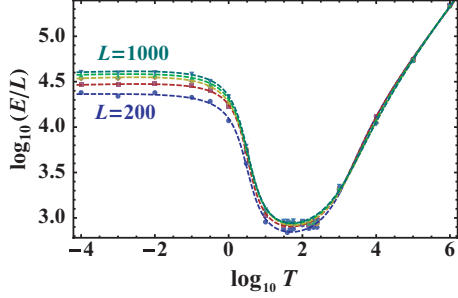$$+ h_4 x \arctan[h_5(x - x_2)] - h_6 x. \tag{19}$$

FIG. 6. (Color online) The approximations (dashed lines) of the calculated numerical values (dots) of the specific energy by the function [Eqs. (18) and (19)] when $N = 200, 400, 600, 800, 1000$. The specific energy is a monotonically rising function of the monomer number. The lowest and the highest sets correspond to $N = 200$ and $N = 1000$, respectively.

The parameters $h_1, \ldots, h_6$ and $x_{1,2}$ are determined by fitting to the numerical data at fixed value of the monomer number N. This explicit form yields an excellent fit whenever there are more than around $N = 100$ monomers. In Fig. 6 we display several examples where we have fitted the functional form [Eqs. (18) and (19)] to polymers as described by our model, where the values of N are between 200 and 1000.

The fitted functional form [Eqs. (18) and (19)] allows us to pinpoint the two critical temperatures T_{c1} and T_{c2}. For this we locate the maxima of the squared logarithmic derivative of the specific energy with respect to the logarithm of temperature,

$$D_E(T,N) = \left[\frac{\partial \log_{10} E(T,N)}{\partial \log_{10} T}\right]^2. \quad (20)$$

This quantity resembles susceptibility that is known to have its maxima at the location of critical temperature(s). The result is shown in Fig. 7.

The maxima of (20) appear as peaks that are clearly visible for all values of N that we have studied and displayed in Fig. 7. From the results in Table II we estimate that the critical temperatures have the following values,

$$\log_{10} T_{c1} = 0.53 \pm 0.02, \quad \text{or} \quad T_{c1} = 3.38 \pm 0.15, \quad (21)$$

$$\log_{10} T_{c2} = 3.52 \pm 0.09, \quad \text{or} \quad T_{c2} = 3306 \pm 716. \quad (22)$$
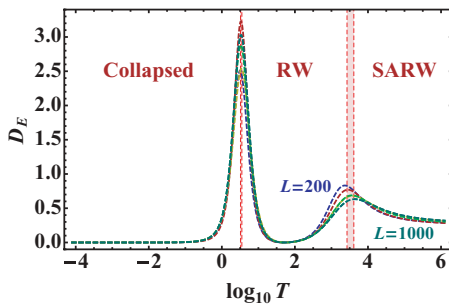

FIG. 7. (Color online) The quantity (20) obtained from the best fits of the functions (18) and (19) for various values of monomer length N. The vertical red lines correspond to the critical temperatures (21) and (22). The width of the lines gives the uncertainty in the definition of the critical temperatures.

TABLE II. The critical temperatures T_{c1} and T_{c2}, determined for each fixed number of monomers N. The averaged value is shown in the bottom row (in bold face) along with respective errors.

N	$\log_{10} T_{c1}(L)$	$\log_{10} T_{c2}(L)$
200	0.5023	3.365
300	0.5114	3.397
400	0.5229	3.450
500	0.5402	3.599
600	0.5379	3.570
700	0.5671	3.552
800	0.5184	3.563
900	0.5360	3.534
1000	0.5254	3.638
Average	**0.53(2)**	**3.52(9)**

Notice that the position of the first maximum is practically the same for all values of N, but the larger the value of N the higher is the height of the maximum. This indicates that the transition between the collapsed phase and the RW phase at $T = T_{c1}$ is indeed phase transition, which is either of the second order or of the first order. Our analysis is not sufficient to determine the order of this transition.

On the other hand, the transition between the RW and SARW phases at $T = T_{c2}$ is likely to be a smooth crossover transition since now both the position of the maximum and its height do not reflect any similar clearly localized profile with increasing values of monomer number N.

V. THE PHASE DEPENDENCE OF THE FREE ENERGY

We have found that in each of the three phases the elastic energy computed from (7) has its distinct, universal dependence on the monomer number N, alternatively radius of gyration R_g.

A. Collapsed phase

In the collapsed phase $T < T_{c1}$ the dependence of the free energy on the number of monomers can be described by the following temperature independent, logarithmically corrected linear law:

$$E(N)/E_0 = C_{\text{Coll}} N \ln \frac{N}{N_0^{\text{Coll}}}. \quad (23)$$

Here E_0 is a parameter that defines the scale of the energy (say) in electron volts and must be obtained by an independent measurement. We find the presence of the logarithmic correction to scaling—as opposed to the analytic corrections proposed by (2)—to be quite notable: We have made a very detailed analysis of the functional form (23) and the logarithmic correction to scaling is consistently exceeding the accuracy of any power-law alternative.

The parameters N_0^{Coll} and C_{Coll} can be calculated using a fitting procedure. The results are shown, respectively, in Fig. 8. The parameter N_0^{Coll} is essentially temperature independent in the low-temperature regime, with value
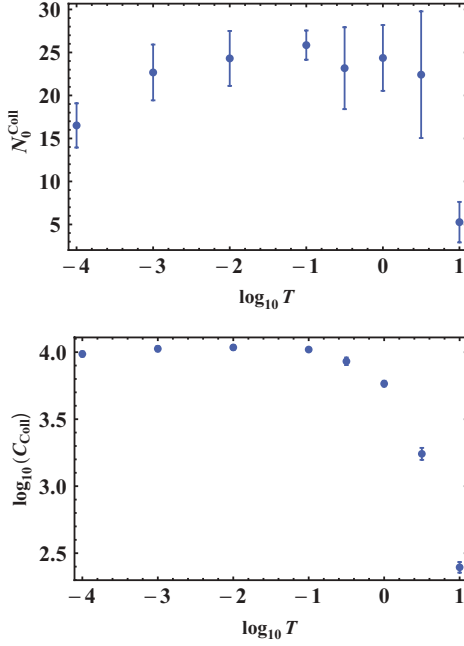
$$N_0^{\text{Coll}} \simeq 22.$$

FIG. 8. (Color online) The parameters of the fit (23): N_0^{Coll} (top panel) and C_{Coll} (bottom panel).

In terms of the radius of gyration we get from (12) and (13) the approximate expression (per units of energy)

$$E(R_g) \approx 971.0 R_g^{2.86} \ln\left[\frac{R_g}{9.53}\right].\tag{24}$$

The relevant aspect of (24) is its dependence on R_g. Since the radius of gyration scales in proportion to the end-to-end distance the result (24) means there is a very rapidly growing elastic force between the end points of the collapsed polymer in our model; in particular, the elastic force is growing clearly more rapidly than in Hooke's law.

Notice that according to the value of the critical temperature (21), the last data point in Fig. 8 (the one with the highest temperature value) is in the RW phase. However, we have found that the two-parameter fit (23) can still be applied to successfully describe this point.

B. RW phase

In the RW phase we have found that the energy obeys the following scaling law (per units of energy):

$$E(T,N) = C_{\mathrm{RW}}(T)N\left[1 - \left(\frac{N}{N_0^{\mathrm{RW}}(T)}\right)^{-\gamma(T)}\right].\tag{25}$$

This is an example of the general form (2). The best fits of the parameters γ, N_0^{RW}, and C_{RW} are shown in Fig. 9 as functions of temperature. We find that all of these parameters are essentially temperature independent with the following average central values:

$$\gamma = 0.355(33),$$
$$N_0^{\mathrm{RW}} = 8.3(1.5),\tag{26}$$
$$C_{\mathrm{RW}} = 1098(22).$$

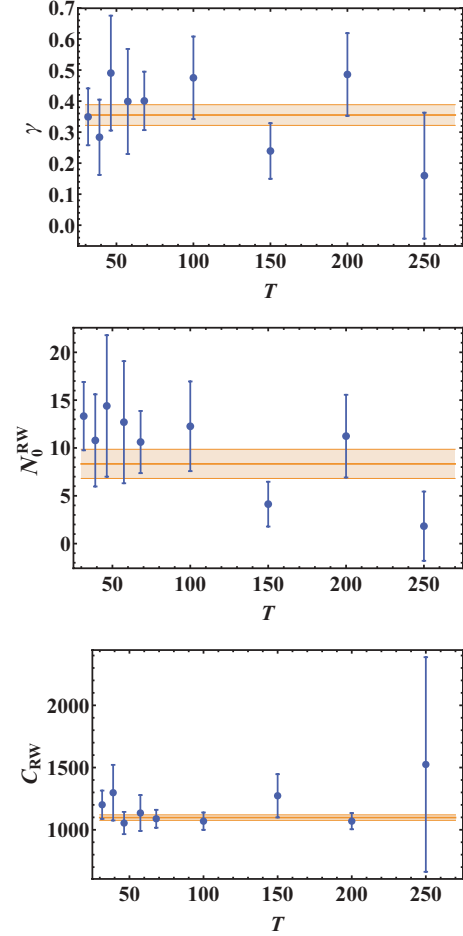These values are shown as horizontal lines in Fig. 9.

FIG. 9. (Color online) The best-fit parameters γ, N_0^{RW}, and C_{RW} of the function (25). The horizontal lines mark the central values (26), and the widths of the lines describe the corresponding errors.

If we use the approximation that $\nu \approx 1/2$ in the RW phase, (26) gives us Hooke's law with a (temperature dependent) correction term (per units of energy):

$$E(R_g, T) \approx C_{\mathrm{RW}}\left(\frac{R_g}{R_0}\right)^2\left[1 - \left(N_0^{\mathrm{RW}}\right)^{\gamma}\left(\frac{R_g}{R_0}\right)^{-2\gamma}\right].\tag{27}$$
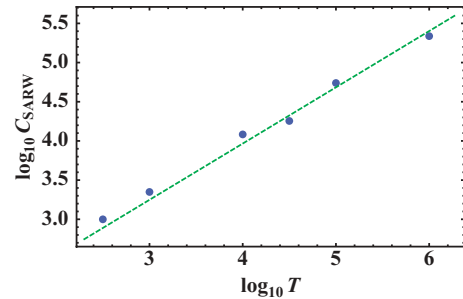

FIG. 10. (Color online) The coefficient of the linear law (28) as a function of temperature. The dashed line illustrates the best fit (30) with the parameters (31).
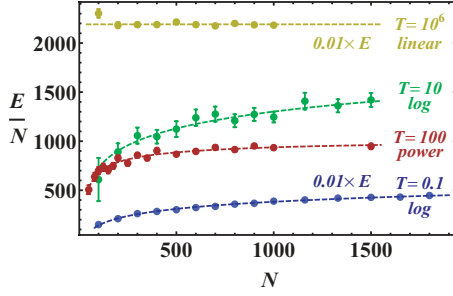
FIG. 11. (Color online) The logarithmic (23), power (25), and linear (28) fits at various temperatures.

C. SARW phase

In the SARW phase we conclude that the energy is a linear function of the monomer number (per units of energy),

$$E(N,T) = C_{\mathrm{SARW}}(T)N, \qquad (28)$$

and with the mean field value of the compactness index $\nu = 3/5$ we get in terms of radius of gyration (per units of energy)

$$E(R_g,T) \approx C_{\mathrm{SARW}}(T)\left(\frac{R_g}{R_0}\right)^{5/3}. \qquad (29)$$
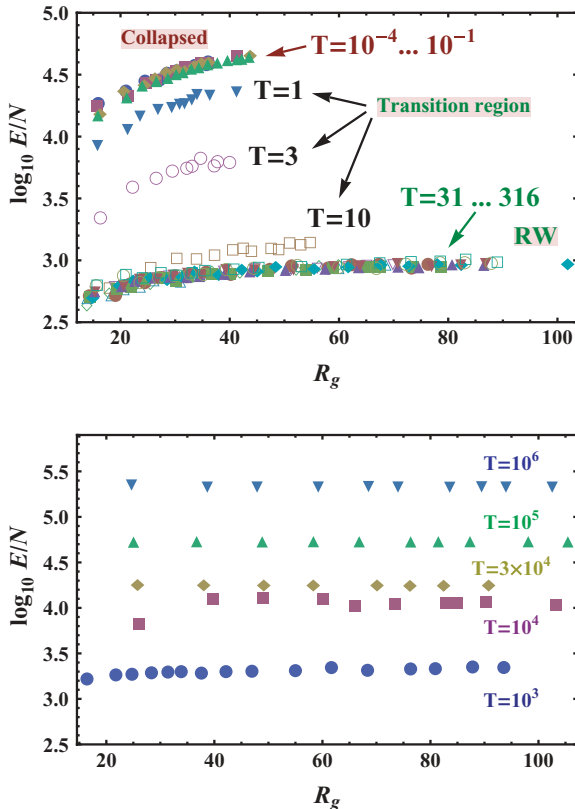

FIG. 12. (Color online) Logarithm of specific elastic energy E/N vs the radius of gyration R_g at different temperatures. The distinct points in the same series correspond to different numbers of monomers N. (Top) The low-temperature collapsed phase and medium-temperature RW phase including the transition region between them. (Bottom) The high-temperature SARW phase. The results are fully in line with the analytic expressions (24), (27), and (29), respectively.
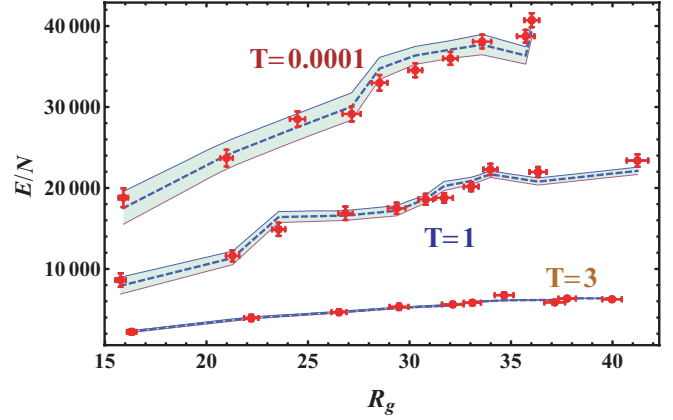

FIG. 13. (Color online) Three examples of the fits of the elastic energy E by the Huang-Lei formula (32). The $T = 0.0001$ line is deep in the collapsed phase while the $T = 1$ and $T = 3$ lines are both in the transition region from collapsed to RW phase, in the vicinity of the critical value $T_{c1} \approx 3.38$.

From our data we are not able to observe any of the correction terms in (2). The only fitting parameter, $C_{\mathrm{SARW}}(T)$, is shown in Fig. 10 as a function of temperature.

We also find that the temperature dependence of the coefficient C_{SARW} can be described by a power law,

$$C_{\mathrm{SARW}}(T) = C_0\, T^{\alpha}, \qquad (30)$$

where the prefactor C_0 and the exponent α are

$$\begin{aligned} C_0 &= 12(4), \\ \alpha &= 0.72(6). \end{aligned} \qquad (31)$$

Note that according to the value of the critical temperature (22), in Fig. 10 the first two points that have the lowest temperature values belong to the RW phase but they can still be described with the present fit. In fact, the N dependence of the free energy at these two temperature values can be fitted *both* by the linear law (28) and by the more general power law (25). However, the power-law fit will lead to very large error bars for the best fit parameters, and therefore we have not shown these points in Fig. 9. Moreover, since we expect that the transition between the RW and SARW phases is a crossover, there should be no clear distinction between these phases in the vicinity of the transition region.

The logarithmic (23), power (25), and linear (28) fits are all shown in Fig. 11.

Finally, we summarize the results in Fig. 12, where we show how the *specific* elastic energy (17) depends on the radius of gyration R_g for various temperatures. The top panel of Fig. 12 corresponds to the collapsed and RW phases. It is very visible that both in the collapsed phase and the RW phase the relation $E_{\mathrm{specific}} = E_{\mathrm{specific}}(R_g)$ is indeed universal: There is no observable temperature dependence. We also note the rapid change from collapsed phase to RW phase. The bottom panel of Fig. 12 describes the high-temperature SARW phase. While an increasing function of temperature, the energy now has only very weak (if any) dependence on the radius of gyration.
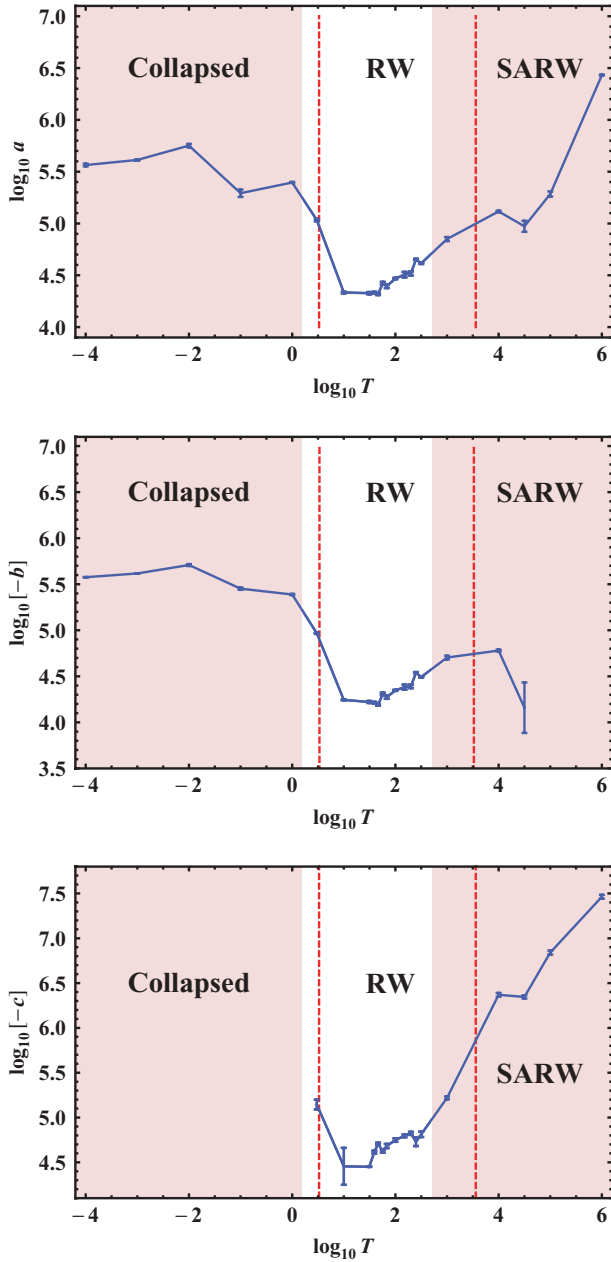
FIG. 14. (Color online) The best-fit parameters of the fits of the elastic energy (32) are shown using the log-log scale. The description is given in the text.

FIG. 15. (Color online) The parameters of the fit (32) in the RW region. The dashed lines give the best fits given by Eqs. (35), (36), and (37).

VI. PROTEINS AND THE HUANG-LEI ELASTIC ENERGY

In [12] the authors propose that the elastic energy of folded proteins in PDB can be described by the following phenomenological (Huang-Lei) formula (per units of energy):

$$E_{HL}(R_g, N) = a\,N^{4/5} + b(N R_g)^{1/2} + c\frac{N^2}{R_g^3}. \quad (32)$$

Here a, b, and c are fitting parameters. By minimizing the energy, the authors [12] compute for the compactness index the value
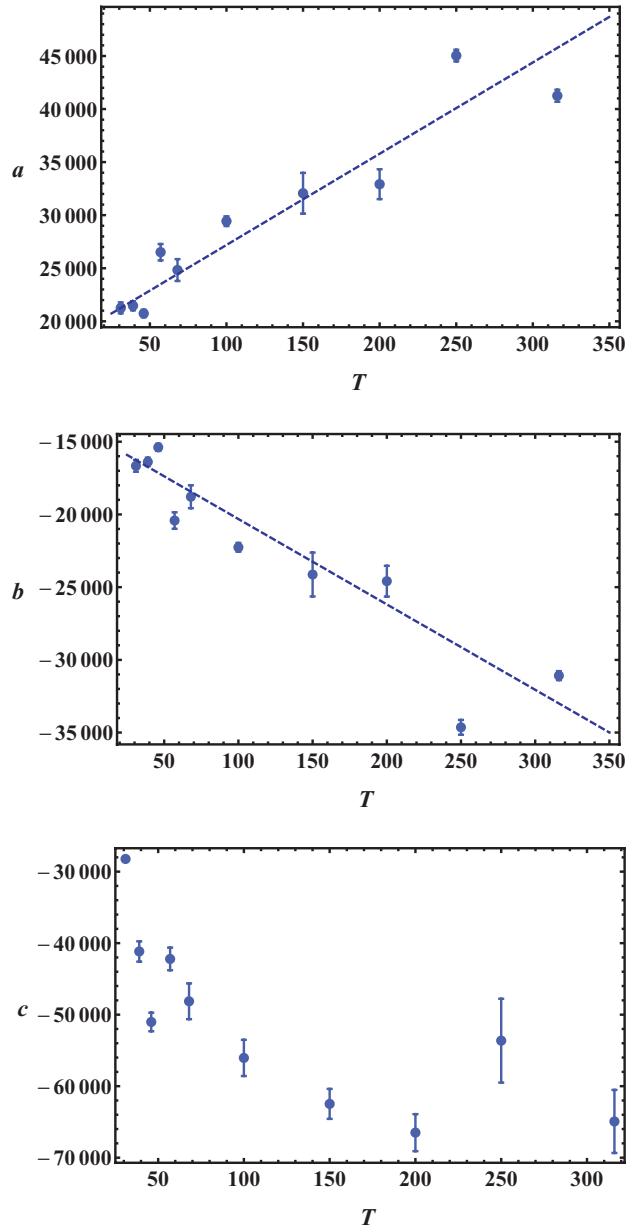
$$\nu_{HL} = 3/7. \quad (33)$$

A priori this suggests [12] that folded proteins could be in a universality class which is different from the known ones (3).

In this section we analyze the formula (32) in the context of our model. We find that it gives an accurate description of data in our model, in particular around the transition point between the collapsed phase and the RW phase where the compactness index grows continuously and monotonically from around $\nu \approx 1/3$ to around $\nu \approx 1/2$ over a finite temperature interval, due to finite scaling effects that are characteristic to a finite length chain: The value (33) corresponds to temperature value

$$T_{HL} \approx 12.5 \pm 1.7 \quad \text{or} \quad \log_{10} T_{HL} = 1.10 \pm 0.06$$

in our model, which suggests that we are (slightly) above the transition temperature T_{c1} between collapsed and RW phases.

In Fig. 13 we show examples where we have fitted (32) to elastic energy computed from our model for three different values of the temperature: deep in the collapsed phase and in the vicinity of the critical temperature T_{c1} that separates the collapsed phase from the RW phase in our model. The widths of the best-fit lines describe the uncertainty in the best-fit parameters, reflecting the statistical errors in our data.

We have found that deep in the collapsed phase the fit is not very good and consequently (32) does not describe fully collapsed proteins, as expected from the value of the compactness index. However, when we enter the transition region between collapsed phase and RW phase and the compactness index starts increasing (continuously as a function of temperature for finite length chains), the quality of the fit becomes increasingly improved and in the vicinity of the critical temperature T_{c1} we find for the statistical χ-square parameter per degree of freedom (dof) a value around

$$\chi^2/(\mathrm{dof}) \approx 1.$$

In Fig. 14 we summarize our findings for the set of best-fit parameters for (32). The red-colored zones correspond to those values of temperature where the $\chi^2/(\mathrm{dof})$ parameter is very large, typically taking values around 10 and higher. In the uncolored (white) zones the $\chi^2/(\mathrm{dof})$ parameter has values that are in the vicinity of unity.

In Fig. 15 we show the behavior of the parameters a, b, and c in the region where the $\chi^2/(\mathrm{dof})$ values are in the vicinity of unity that is near the transition between collapsed phase and RW phase and within the RW phase. We have found that the temperature dependence of the parameters a and b can be fitted by linear functions:

$$a(T) = C_a + \left(1 + \frac{T}{T_a}\right), \tag{34}$$

$$b(T) = C_b + \left(1 + \frac{T}{T_b}\right), \tag{35}$$

where

$$C_a = 8.6(1.4) \times 10^3, \quad T_c = 216(37), \tag{36}$$

$$C_b = -1.4(1) \times 10^4, \quad T_b = 246(44). \tag{37}$$

These fits are shown in Fig. 15 by the dashed lines.

Our conclusion is that the Huang-Lei formula (32) gives a very good description of the elastic energy in our model, in particular when we are very near the transition point between the collapsed and RW phases and slightly inside the RW phase. However, it is not very accurate for temperature values that are deep in the collapsed phase, nor when we approach the crossover to the SARW phase. We note that this is consistent with the behavior of the compactness index in our model as displayed in Fig. 2. When we compare the computed value (33) with Fig. 2 we find that this value corresponds to the transition region.

Together with [12], and the comparison between (15) and (16), as well as (14) and (21), these results suggest that our model should describe the statistical properties of folded proteins in PDB for temperature values that are very close to the critical value $T_{c1} \approx 3.38$.

VII. DISCUSSION

We have investigated the statistical properties of a chiral homopolymer model that has been previously introduced to describe the properties of collapsed proteins in the PDB. We have found that as a function of temperature the model does indeed realize the three known phases of polymers: the collapsed phase, the RW phase, and the SARW phase. Furthermore, we have found that the model predicts that the transition between the collapsed phase and the RW phase is a phase transition, while the RW and SARW phases are separated from each other by a smooth crossover transition. These findings are in line with general arguments on the phase structure of polymers [3].

We have also computed the elastic energy as a function of radius of gyration, that is, end-to-end distance of a polymer. In the collapsed phase we have found that the energy grows faster than in Hooke's law, in the RW phase we find Hooke's law with temperature dependent corrections, and finally in the SARW phase we find that the dependency of energy on the radius of gyrations is weaker than in Hooke's law. It would be interesting to test our predictions experimentally in the case of proteins, for example using atomic force microscopy.

Finally, we have compared our model with a phenomenological expression that has been introduced by Huang and Lei to describe the elastic energy of collapsed proteins. We have found that the Huang-Lei formula gives a good effective description of our model, in particular when we are in the vicinity of the transition region that separates the collapsed phase from the RW phase.

[1] K. G. Wilson, Phys. Rev. B **4**, 3174 (1971); **4**, 3184 (1971).

[2] L. P. Kadanoff, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic Press, London, 1976), Vol. 5A, pp. 1–34.

[3] P. G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).

[4] B. G. Nickel, Macromolecules **24**, 1358 (1991).

[5] F. J. Wegner, Phys. Rev. B **5**, 4529 (1972).

[6] B. Li, N. Madras, and A. Sokal, J. Stat. Phys. **80**, 661 (1995).

[7] J. C. LeGuillou and J. Zinn-Justin, Phys. Rev. B **21**, 3976 (1980).

[8] P. G. De Gennes, Phys. Lett. A **38**, 339 (1972).

[9] U. H. Danielsson, M. Lundgren, and A. J. Niemi, Phys. Rev. E (to be published).

[10] K. A. Dill, O. S. Banu, M. S. Shell, and T. R. Weikl, Annu. Rev. Biophys. **37**, 289 (2008).

[11] H. M. Berman, K. Henrick, H. Nakamura, and J. L. Markley, Nucleic Acids Res. **35** (Database issue) D301 (2007).

[12] J. Lei and K. Huang, e-print arXiv:1002.5013; Europhys. Lett. **88**, 68004 (2009).

[13] C. J. Thompson, J. Math. Phys. **9**, 232 (1968).

[14] F. Calogero, Lett. Nuovo Cimento **13**, 411 (1975).

[15] M. Barma and M. E. Fisher, Phys. Rev. B **31**, 5954 (1985).

[16] G. A. Baker, Phys. Rev. Lett. **60**, 1844 (1988).

[17] The error in the estimate is large because around this point the temperature changes rapidly, indicating that we are near a phase transition.