Elastic Energy and Phase Structure of a Continuous Spin Ising Chain with Applications to the Protein Folding Problem

Martin Lundgren[1], Antti Niemi[1,2], Maxim Chernodub[2,3]

[1]Department of Physics and Astronomy, Uppsala University, Box 803, S-75108, Uppsala, Sweden
[2]Laboratoire de Mathématiques et Physique Théorique, Université François-Rabelais Tours, Fédération Denis Poisson - CNRS, Parc de Grandmont, 37200 Tours, France
[3]Department of Mathematical Physics and Astronomy, Krijgslaan 281, 59, Gent, B-9000, Belgium

based on arXiv:1008.4552

Introduction

We have created a model of a continuous spin Ising chain that shares several similarities with real proteins. Even though the model at this stage cannot be used to model a particular folding protein we would like to test if it can describe the statistical properties of folding proteins, in particular when it comes to their dependence on temperature.

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 of an entire class of physical systems using only a single representative model. In the case of polymers one expects there are three different nontrivial phases and these correspond to the universality class of self-avoiding random walk (SARW), ordinary random walk (RW) and the class of collapsed polymers[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 random walk phase takes place at the theta-temperature that separates the SARW phase from the collapsed phase.

Compactness index

The most videly 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[4]

$$R_g^2 = \frac{1}{2N^2}\sum (r_i - r_j)^2 \approx R_0^2 N^{2\nu}\left(1 + \beta_1 N^{-\triangle_1} + ...\right)$$

Here r_i (i=1..N) are the locations of the N monomers. The critical exponents ν and Δ are universal quantities, but the form factor R_0 that characterizes the effective distance between the monomers in the large N limit , and the amplitude β_1 is not.

For a polymer the compactness index has the following mean field values[3]: $\nu_{mf} = \begin{cases} 3/5\ SARW \\ 1/2\ RW \\ 1/3\ collapsed \end{cases}$

The Model

The model is defined by the following internal energy[5],

$$E = \sum a_{ij}\left\{1 - \cos\left[\omega_{ij}\left(\kappa_i - \kappa_j\right)\right]\right\} + \sum\left\{b_i\kappa_i^2\tau_i^2 + c\left(\kappa_i^2 - m_i^2\right)^2 + d_i\tau_i\right\}$$

Here i,j = 1,...,N label the N monomers of a polygonal chain. The geometry is determined by the order parameter κi, that is a discrete version of the Frenet curvature, and by the order parameter τi, that is the lattice version of the Frenet torsion[5]. Once the values of (κ_i, τ_i) for each i=1,...,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.

The aij,ωij, bi,..di are parameters. The first sum in the energy describes long-distance interactions, here limited to only nearest neighbour interactions. The first two terms in the second sum are the interaction between κi and τi and the self-interaction of κ_i. The final term is a discretized version of the one-dimensional Chern-Simons functional. It is the origin of the chirality in the polymer chain.[5] The parameters in these simulations are independent of site.

We have used a standard Metropolis algorithm with the initial polymer being shaped as a straight rod. Each step consists of a small shift in the curvature and torsion. This shift is accepted with probability

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

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

Results

Elastic Energy

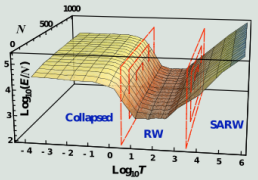

Figure 1: A three dimensional plot of a logarithm of the specific energy as a function of temperature and the number of monomers N. The three phases are identified and the critical temperatures are denoted by the vertical planes.

For fixed parameter values the free energy 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 figure 1 we display a three dimensional plot of the specific elastic energy, i.e. the energy per monomer, as a function of these two parameters.

In the figure we can clearly identify three different phases that are separated from each other by clearly identifiable transition temperatures T_{c1} and T_{c2}. Both the low temperature ($T < T_{c1}$) collapsed phase and the medium temperature RW phase are characterized by having a more or less temperature independent specific energy. The high temperature SARW phase on the other hand is characterized by a power law increase in the specific energy as a function of the temperature.

We have found that the dependence of the specific energy on temperature can be approximated with very good accuracy by a function

$$\log\left(\frac{E}{N}\right) = F^{fit}\left(\log T\right) \quad F^{fit}\left(x\right) = h_1 + h_2\arctan\left[h_3\left(x - x_1\right)\right] + h_4 x\arctan\left[h_5\left(x - x_2\right)\right] - h_6 x \quad (1)$$

The parameters h_1...h_6 and $x_{1,2}$ are determined by fitting to the numerical data at fixed value of the monomer number N. In figure 2 we display several examples of fits with this function. The fitted functional form (1) allows us to pinpoint the two critical temperatures T_{c1} and T_{c2}. For this we can look at the following property which is known to have its maxima at the location of critical temperatures,

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

The result is shown in figure 3. Using the averages of the resulting critical temperatures for different values of N we estimate the critical temperatures to be T_{c1}=3.38±15 and T_{c2}=3306±716. Note also that the peak at T_{c1} is much sharper, indicating a true phase-transition while the behavior at T_{c2} indicates a more smooth transition.

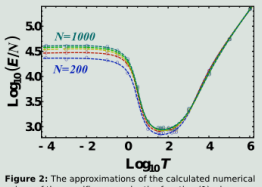

Figure 2: The approximations of the calculated numerical values of the specific energy by the function (1) when N=200, 400, 600, 800, 1000

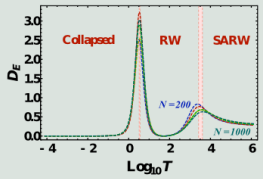

Figure 3: The quantity (2) obtained from the best fit of the function (1) for various values of polymer length N. The vertical red lines correspond to the critical temperatures.

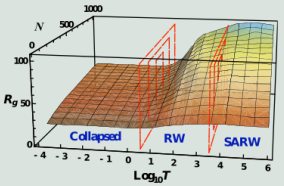

Figure 7: The radius of gyration R_g as a function of temperature T and the number of monomers N. The three phases are identified with the position of the ensuing critical temperatures denoted by the vertical planes.

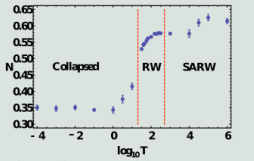

Figure 8: The compactness index ν vs. temperature T. The vertical lines correspond to temperature values where ν reaches the mean field values.

Radius of Gyration

In figure 7 we show how the radius of gyration depends on the (Metropolis) temperature T and the number of monomers N. In figure 8 we show the T dependence of ν. In these figures we try to identify three phases corresponding to the three mean field values of ν.

There is clearly a low-temperature phase where the dependence of T is very weak. From figure 8 we estimate the value of ν in this phase to ν=0.348(7). This is so close to the mean field value 1/3 that we obviously are in that phase. When the temperature increases to close to the critical temperature, ν starts to increase and for T≈4, ν≈0.38 which is also the estimated value of the dependence of the radius of gyration on the number N of alpha carbons for all single stranded proteins in PDB[6].

The transition from the collapsed phase to the RW phase is very visible in our figures. There is a clear rapid transition in ν, reminiscent of a phase transition. The transition point is very close to the mean field value 1/2. For larger T ν is a slowly increasing function of temperature that eventually plateaus around the value ν≈0.58. This is slightly below the values reported for ν in the SARW phase and since ν appears to have a tendency to approach its large-N limit from above we conclude that we are in the RW phase.

The transition to the SARW phase has the characteristics of a smooth cross-over. We find that for very high temperatures ν approaches ν≈0.62. This is a higher value than those obtained by more careful studies of random walk, but considering the short lengths we are using that is an acceptable result.

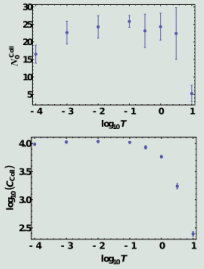

Figure 4: The parameters of the fit (3): N_0 (upper plot) and C (lower plot)

Collapsed phase

In the collapsed phase the dependence of the free energy on the number of monomers can be described by the following, temperature independent, logarithmically corrected linear law:

$$E\left(N\right) = C_{coll}N\ln\frac{N}{N_0^{coll}} \quad (3)$$

The parameters N_0 and C can be calculated using a fitting procedure and the result is shown in figure 4. N_0 is essentially temperature independent with value

$$N_0^{coll} \approx 22$$

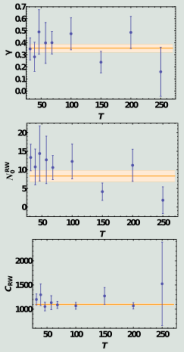

Figure 5: The best fit parameters γ, N_0 and C of the function (4). The horizontal lines mark the central values and the width describe the corresponding errors.

RW phase

In the RW phase we have found that the energy obeys the following scaling law

$$E\left(T, N\right) = C_{RW}(T)N\cdot \cdot\left[1 - \left(\frac{N}{N_0^{RW}(T)}\right)^{-\gamma(T)}\right] \quad (4)$$

The best fits of the parameters can be seen in figure 5, as functions of temperature. We find that all of these parameters are more or less temperature independent, with the following average central values

$$\gamma = 0.355(33)$$
$$N_0^{RW} = 8.3(1.5)$$
$$C_{RW} = 1098(22)$$

SARW phase

In the SARW phase we conclude that the energy is a linear function of the monomer number

$$E\left(T\right) = C_{SARW}(T)N = C_0 T^{\alpha} N \quad (5)$$

The fitting parameter C_{SARW} is shown in figure 6 as a function of temperature. We find that it can be described by a power law (RHS of (5)) where the coefficients are

$$C_0 = 12(4)$$
$$\alpha = 0.76(6)$$

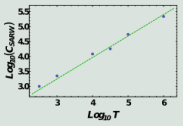

Figure 6: The coefficient of the linear law (5) as a function of temperature. The dashed line illustrates the best fit.

Conclusion

We have investigated the statistical properties of a homopolymer model that has been introduced to describe the properties of collapsed proteins in the PDB. We have found that the model indeed realize the three known phases of polymers. Furthermore we have found that the model predicts that the transition between the collapsed and RW phases is a phase transition while the RW and SARW phases are separated by a smooth cross-over. These findings are in line with general arguments on the phase structure of polymers.[3]

In the full paper we have in addition to this made a comparison with an expression, introduced by Huang and Lei for the elastic energy of collapsed proteins[6]. We found that their formula describes our model well in the temperature region corresponding to the proteins in the PDB.

Acknowledgements

This work was supported by a STINT Institutional grant IG2004-2 025.

References

[1] K.G. Wilson, Phys. Rev. **B4** 3174 (1971), *ibid* **B4** 3184 (1971)
[2] L.P. Kadanoff, in *Phase Transitions and Critical Phenomena*, C. Domb and M.S. Green Eds. (Academic Press, London 2976) 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] U.H. Danielsson, M. Lundgren and A.J. Niemi, Phys. Rev. **E** (accepted)
[6] J. Lei and K. Huang, e-print arXiv:1002.5013 [cond-mat.stat-mech], e-print arXiv:1002.5024 [cond-mat.stat-mech]