Eur. Phys. J. B **15**, 371–374 (2000)

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

Is there a universality of the helix-coil transition in protein models?

J.P. Kemp[1], U.H.E. Hansmann[2,a], and Zheng Yu Chen[1]

[1] Department of Physics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
[2] Department of Physics, Michigan Technological University, Houghton, MI 49931-1291, USA

Received 8 December 1999

Abstract. The similarity in the thermodynamic properties of two completely different theoretical models for the helix-coil transition is examined critically. The first model is an all-atomic representation for a poly-alanine chain, while the second model is a minimal helix-forming model that contains no system specifics. Key characteristics of the helix-coil transition, in particular, the effective critical exponents of these two models agree with each other, within a finite-size scaling analysis.

PACS. 87.15.He Dynamics and conformational changes – 87.15.-v Biomolecules: structure and physical properties – 64.60.Cn Order disorder transformations; statistical mechanics of model systems

The importance of understanding the statistical physics of the protein-folding problem has been stressed recently [1,2]. For instance, it is now often assumed that the energy landscape of a protein resembles a partially rough funnel. Folding occurs by a multi-pathway kinetics and the particulars of the folding funnel determine the transitions between the different thermodynamic states [1,3]. This "new view" [1] of folding was derived from studies of minimal protein models which capture only a few, but probably dominant parameters (chain connectivity, excluded volume, etc.) in real proteins.

An implicit yet fundamentally crucial assumption is that the basic mechanism of structural transitions in biological molecules depends solely on gross features of the energy function, not on their details, and that a law of corresponding states can be used to explain dynamics and structural properties of real proteins from studies of related minimal models. This assumption needs to be proven. An even stronger notion in statistical physics is the universality hypothesis for critical phenomena. The critical exponents are identical for different theoretical models and realistic systems belonging to the same universality class. Many theoretical concepts in protein folding, such as coil-helix or coil-globular transitions involve phase transition or phase transition-like behavior. Thus, one wonders if physical measurements between two model systems for the same transition would have any "universal" properties.

The purpose of this article is to examine these questions for the helix-coil transition in homopolymers of amino acids [4,5]. Traditionally, the coil-helix transition is described by theories such as the Zimm-Bragg model [6] in which the homopolymers are regarded as one dimensional systems with only local interactions; as such a true thermodynamic phase transition is impossible. However, recently there have been [4,5] indications that the coil-helix transition near the transition temperature displays phase-transition like behavior. We use here finite-size scaling analysis, a common tool in statistical physics, to examine the question of universality of the helix-coil transition in two completely different, illuminating models. On one hand, we have a detailed, all-atomic representation of a homo poly-alanine chain [7]. On the other hand, we have a simple coarse-grained model describing the general features of helix-forming polymers [4]. In this article, our interest lies in finding out how far the similarity of the two models go. If the two models yield the same key physical characteristics, then we at least have one concrete example of the validity of the corresponding state principle or universality hypothesis in biopolymer structures.

Poly-alanine is well-known to have high helix-propensities in proteins, as demonstrated both experimentally and theoretically [5,7]. It has been well tested and generally believed that approximate force fields, such as ECEPP/2 [9] as implemented in the KONF90 program [10], give protein-structure predictions to a surprisingly degree of faithfulness. As our first model, we have "synthesized" poly-alanine with N residues, in which the peptide-bond dihedral angles were fixed at the value 180° for simplicity. Since one can avoid the complications of electrostatic and hydrogen-bond interactions of side chains with the solvent for alanine (a non-polar amino acid), we follow earlier work [7] and neglect explicit solvent molecules in the current study.

Our second model is a minimalistic view of a helix forming polymer [4] without atomic-level specifics. A

[a] e-mail: hansmann@mtu.edu

wormlike chain is used to model the backbone of the molecule, while a general directionalized interaction, in terms of a simple square well form, is used to capture the essence of hydrogen-like bonding. The interaction energy between the residue labeled i and j is modeled by,

$$V_{ij}(\mathbf{r}) = \begin{cases} \infty & r < D \\ -v & D \le r < \sigma \\ 0 & \sigma \le r \end{cases} \tag{1}$$

where $v = \epsilon[\hat{\mathbf{u}}_i \cdot \hat{\mathbf{r}}_{ij}]^6 + \epsilon[\hat{\mathbf{u}}_j \cdot \hat{\mathbf{r}}_{ij}]^6$, $\hat{\mathbf{u}}_i = (\hat{\mathbf{r}}_{i+1,i}) \times (\hat{\mathbf{r}}_{i,i-1})$, $\hat{\mathbf{r}}_{ij}$ is the unit vector between monomer i and j, $D = 3/2a$ is the diameter of a monomer, $\sigma = \sqrt{45/8}a$ is the bonding diameter, and a is the bond length while bond angle is fixed at 60°.

To obtain the thermodynamic properties, we have conducted multicanonical Monte Carlo simulations for both models. In the low-temperature region where most of the structural changes occur, a typical thermal energy of the order $k_{\mathrm{B}}T$ is much less than a typical energy barrier that the polymer has to overcome. Hence, simple canonical Monte Carlo or molecular dynamics simulations cannot sample statistically independent configurations separated by energy barriers within a finite amount of available CPU time, and usually give rise to bias statistics. One way to overcome this problem is the application of generalized ensemble techniques [11], such as the multicanonical algorithm [12] used here, to the protein folding problem, as has recently been utilized and reported [13].

In a multicanonical algorithm [12] conformations with energy E are assigned a weight $w_{\mathrm{mu}}(E) \propto 1/n(E)$, $n(E)$ being the density of states. A simulation with this weight generates a random walk in the energy space; since a large range of energies are sampled, one can use the re-weighting techniques [14] to calculate thermodynamic quantities over a wide range of temperatures by

$$\langle \mathcal{A} \rangle_T = \frac{\int \mathrm{d}x \mathcal{A}(x) w_{\mathrm{mu}}^{-1}(E(x)) \mathrm{e}^{-\beta E(x)}}{\int \mathrm{d}x\, w_{\mathrm{mu}}^{-1}(E(x)) \mathrm{e}^{-\beta E(x)}}, \tag{2}$$

where x stands for configurations and β is the inverse temperature.

In the case of poly-alanine chains, up to $N = 30$ alanine residues were considered. The multicanonical weight factors were determined by the iterative procedure described in reference [12] and we needed between 4×10^5 sweeps (for $N = 10$) and 5×10^5 sweeps (for $N = 30$) for estimating the weight factor approximately. All thermodynamic quantities were measured from a subsequent production run of M Monte Carlo sweeps, where M=4×10^5, 5×10^5, 1×10^6, and 3×10^6 sweeps for $N = 10$, 15, 20, and 30, respectively. In the minimal model, chain lengths up to 39 monomers were considered. In this model a single sweep involves a rotation of a group of monomers *via* the pivot algorithm [15]. For the weight factors a similar number of iterative procedure was used, and for the production run 1×10^8 sweeps was used in all cases.

We obtain the temperature dependence of the specific heat, $C(T)$, by calculating

$$C(T) = \beta^2 \, \frac{\langle E_{\mathrm{tot}}^2 \rangle - \langle E_{\mathrm{tot}} \rangle^2}{N}, \tag{3}$$

where E_{tot} is the total energy of the system. We also analyze the order parameter q which measures the helical content of a polymer conformation and the susceptibility

$$\chi(T) = \frac{1}{N-2}(\langle q^2 \rangle - \langle q \rangle^2). \tag{4}$$

associated with q. For poly-alanine q is defined as

$$q = \tilde{n}_{\mathrm{H}} \tag{5}$$

where $\tilde{n}_{\mathrm{H}}$ is the number of residues (other than the terminal ones) for which the dihedral angles (ϕ, ψ) fall in the range $(-70 \pm 20°, -37 \pm 20°)$. For our worm-like chain model the order parameter q is defined as

$$q = \sum_{i=2}^{N-1} \mathbf{u}_i \cdot \mathbf{u}_{i+1}. \tag{6}$$

In both cases the first and last residues, which can move more freely, are not counted in the procedure.

From a finite-size scaling analysis of the heights and width of specific heat and susceptibility we can extract a set of effective critical exponents which characterize the helix-coil transition in these two models [16]. For instance, with C_{MAX} defined to be the maximum peak in the specific heat, we have

$$C_{\mathrm{MAX}} \propto N^{\frac{\alpha}{d\nu}}. \tag{7}$$

In a similar way, we find for the scaling of the maximum of the susceptibility

$$\chi_{\mathrm{MAX}} \propto N^{\frac{\gamma}{d\nu}}. \tag{8}$$

For both quantities we can also define the temperature gap $\Gamma = T_2 - T_1$ (where $T_1 < T_{\mathrm{MAX}} < T_2$) chosen such that $C(T_1) = bC_{\mathrm{MAX}} = C(T_2)$, and $\chi(T_1) = b\chi(T_{\mathrm{c}}) = \chi(T_2)$ where b is a fraction. The temperature gap obeys

$$\Gamma = T_2 - T_1 \propto N^{-\frac{1}{d\nu}}, \tag{9}$$

as has been suggested in reference [16]. The analysis should be insensitive to the actual fraction, b, of C_{MAX} (χ_{MAX}) considered for defining T_1 and T_2 which was verified from our numerical data fitting of poly-alanine chains.

The scaling exponents, α, ν, and γ, have their usual meaning in critical phenomena; however, the above scaling relations also hold formally for the case of a first-order transition, with effective scaling exponents $d\nu = \alpha = \gamma = 1$ [16,17]. Note that d is the dimensionality of the system, and it always appears in the combination $d\nu$. Without

Table 1. Shown are the location of the specific heat maximum $T_{\rm MAX}$, the maximum of specific heat $C_{\rm MAX}$, susceptibility $\chi_{\rm MAX}$, the width of the half peak in specific heat $\Gamma_{\rm C}$, and width of the half peak of susceptibility Γ_χ for various chain lengths.

N	$T_{\rm MAX}$	$C_{\rm MAX}$	$\Gamma_{\rm C}$	$\chi_{\rm MAX}$	Γ_χ
All-Atomic Model					
10	427(7)	8.9(3)	160(7)	0.49(2)	140(7)
15	492(5)	12.3(4)	119(5)	0.72(3)	110(5)
20	508(5)	16.0(8)	88(5)	1.08(3)	78(5)
30	518(7)	22.8(1.2)	58(4)	1.50(8)	56(3)
Minimal Model					
13	1.25(1)	1.088(2)	1.22(2)	0.232(2)	2.20(2)
19	1.17(1)	1.424(5)	1.12(2)	0.353(3)	0.81(2)
26	1.16(1)	1.789(8)	0.89(2)	0.553(8)	0.57(2)
33	1.13(1)	2.08(1)	0.73(2)	0.78(1)	0.45(2)
39	1.12(1)	2.27(2)	0.61(2)	0.96(2)	0.41(2)

Table 2. Summary of the critical exponents obtained for the two models.

	All-atomic	Minimal
$d\nu$	1.00(9)	0.96(8)
α	0.89(12)	0.70(16)
γ	1.06(14)	1.3(2)

knowing further the effective dimensionality of our systems, we use the combination $d\nu$ as a single parameter in the fit.

It then becomes straightforward to use the above equation and the values given in Table 1 to estimate the critical exponents. We obtain for poly-alanine from the scaling of the width of the specific heat $1/d\nu = 1.02(11)$ with a goodness of fit ($Q = 0.9$) (see Ref. [18] for the definition of Q), for chains of length $N = 15$ to $N = 30$. Inclusion of $N = 10$ leads to $1/d\nu = 0.84(7)$, but with a less acceptable fit ($Q = 0.1$). Similarly, we find from the scaling of the width of the susceptibility $1/d\nu = 0.98(11)$ ($Q = 0.5$) for chains of length $N = 15$ to $N = 30$ and $1/d\nu = 0.81(7)$ ($Q = 0.2$) when the shortest chain $N = 10$ is included in the fit. Hence, we present as our final estimate for the correlation exponent of poly-alanine $d\nu = 1.00(9)$. This value is in good agreement with the estimate $d\nu = 0.93(5)$ obtained from the partition function zero analysis in reference [8].

The results for the exponent α give $\alpha = 0.89(12)$ ($Q = 0.9$) when all chains are considered, and $\alpha = 0.86(10)$ ($Q = 0.9$) when the shortest chain is excluded from the fit. Analyzing the peak in the susceptibility we find $\gamma = 1.06(14)$ ($Q = 0.5$) for chain lengths $N = 15 - 30$ and $\gamma = 1.04(11)$ ($Q = 0.5$) for chain lengths $N = 10 - 30$. We summarize our final estimates for the critical exponents in Table 2. The scaling plot for the susceptibility is shown in Figure 1: curves for all lengths of poly-alanine chains collapse on each other indicating the validity of finite size scaling of our poly-alanine data.

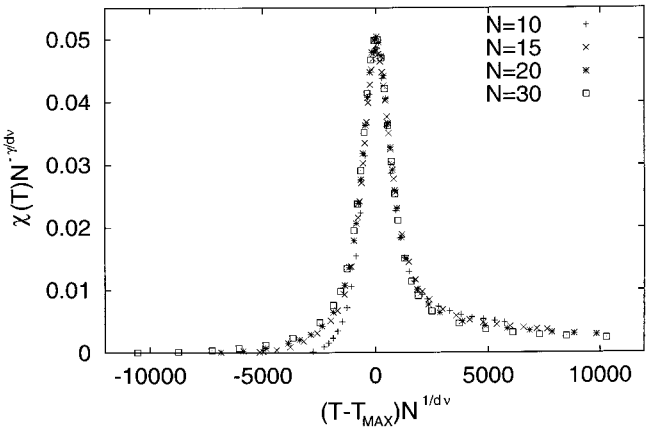

Fig. 1. Scaling plot for the susceptibility $\chi(T)$ as a function of temperature T, for poly-alanine molecules of chain lengths $N = 10$, 15, 20 and 30.

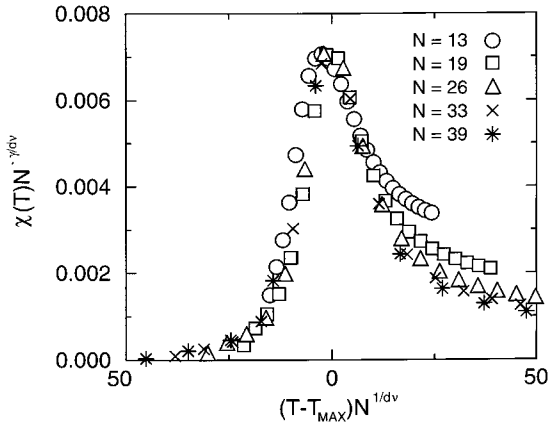

Fig. 2. Scaling plot of $\chi(T)$ as a function of temperature T, for the minimum model of chai n lengths $N = 13$, 19, 26, 33 and 39.

The same procedure can be applied to analyze the data from the minimal model. All calculations have been done with the omission of the shortest chain. Using the widths of the specific heat a $b = 80\%$ of the peak height we obtain $1/d\nu = 1.03(7)$, ($Q = 0.2$). The width of the peak at half maximum is more unreliable in this case as the coil-helix transition is complicated by the additional collapsing transition to a globular state in the vicinity of the coil-helix transition [4]. This exponent agrees with that calculated from the susceptibility widths, $1/d\nu = 0.89(9)$, ($Q = 0.3$). Hence, our final estimate for this critical exponent in our second model is $d\nu = 0.96(8)$. These values are in good agreement with those of the poly-alanine model.

From the $C_{\rm MAX}$ data in Table 1 and using the above given value for the exponent $d\nu$ we find $\alpha = 0.70(16)$ ($Q = 0.3$) which is somewhat smaller than that of the poly-alanine model. The susceptibility exponent as calculated from the data in Table 1 yields a value of $\gamma = 1.3(2)$ ($Q = 0.5$), which agrees with the previous estimation within the error bar. The scaling plot for the susceptibility is shown in Figure 2. While curves corresponding to large polymer sizes collapse into the same curve, the $N = 13$

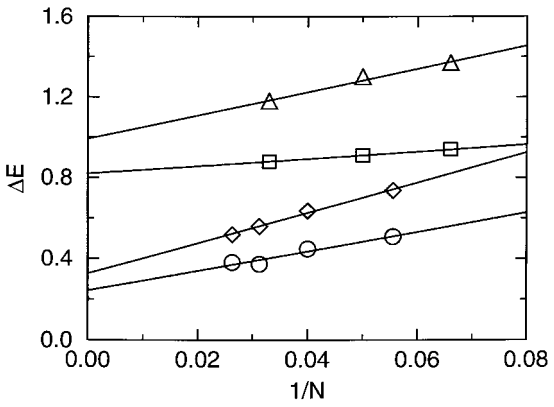

Fig. 3. Scaling of energy gap and transition width at 80% and 90% of C_{MAX}. Here we hav e used $\Delta E_{80\%}$ ($\triangle$ for all-atom model, $\Diamond$ for minimal model), $\Delta E_{90\%}$ ($\square$ for all-atom model, $\bigcirc$ for minimal model).

case shows small disagreement, indicating that the finite size scaling are valid only for longer chain lengths in the minimal model.

Comparing the critical exponents of our two models as summarized in Table 2 we see that the estimates for the correlation exponent $d\nu$ agree well for the two models. Within the error bars, the estimates for the susceptibility exponent γ also agree. The estimates for the specific heat exponent α seem to disagree within the error ranges. However, in view of the fact that both analyses are based on small system size the true error ranges could be actually larger than the ones quoted here. Using these rather crude results, we have already demonstrated a striking similarity in finite-size scalings of the two model. Therefore, we can convincingly make the conjecture that our minimal model can be used to represent the structural behavior of real helix-forming proteins.

Our analysis should tell us also whether the helix-coil transition in our models is of first or second order. In the former case we would expect $d\nu = \alpha = \gamma = 1$ which seems barely supported by our data due to the rather large error bars associated with the estimate of the exponents. We have further explored the nature of the transition from another perspective, by considering the change in energy crossing a small temperature gap (taken to be within 90% of C_{MAX}) from the original data,

$$\Delta E = (E_{\mathrm{tot}}(T_2) - E_{\mathrm{tot}}(T_1))/N. \qquad (10)$$

This value should approach either a finite value or zero as N^{-1} goes to zero. A finite value would indicate a first order transition while a zero value a second order transition. In the case of a first order transitions the intercept would indicate the latent heat. Now, the assumption is that this energy change scales linearly as N^{-1} goes to zero. Figure 3 shows a plot of the data from both the atomic-level and minimal models, where nonzero intercepts can be extrapolated at N^{-1}=0. Hence, our results seem to indicate a finite latent heat and a first-order like helix-coil transition. However, we can not exclude the possibility that the true asymptotic limit of $|E|$ is zero, and some of the results of reference [5] point for the case of poly-alanine rather towards a second-order transition. Further simulations of larger chains seem to be necessary to determine the order of the helix-coil transition without further doubts.

In summary, we conclude that in view of the similarity of the two models examined here, a universality principle can be established for the coil-helix transition. Examining the finite size scaling analysis allows us to calculate estimators for critical exponents in the two models which indicate "universality" of helix-coil transitions.

Financial supports from Natural Science and Engineering Research Council of Canada, the National Science Foundation (CHE-9981874), and a Research Excellence Fund of the State of Michigan are gratefully acknowledged.

References

1. K.A. Dill, H.S. Chan, Nature Struct. Biol. **4**, 10 (1997).
2. E.I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997); T. Veitshans, D. Klimov, D. Thirumalai, Fold. Des. **2**, 1 (1997).
3. J.D. Bryngelson, P.G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 524 (1987); J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Ann. Rev. Phys. Chem. **48**, 545 (1997).
4. J.P. Kemp, Z.Y. Chen, Phys. Rev. Lett. **81**, 3880 (1998).
5. U.H.E. Hansmann, Y. Okamoto, J. Chem. Phys. **110**, 1267 (1999); **111**, 1339(E) (1999).
6. B.H. Zimm, J.K. Bragg, J. Chem. Phys. **31**, 526 (1959).
7. Y. Okamoto, U.H.E. Hansmann, J. Phys. Chem. **99**, 11276 (1995).
8. N.A. Alves, U.H.E. Hansmann, Phys. Rev. Lett. **84**, 1836 (2000).
9. M.J. Sippl, G. Némethy, H.A. Scheraga, J. Phys. Chem. **88**, 6231 (1984), and references therein.
10. H. Kawai *et al.*, Chem. Lett. **1991**, 213 (1991); Y. Okamoto *et al.*, Protein Engineering **4**, 639 (1991).
11. U.H.E. Hansmann, Y. Okamoto, in *Annual Reviews in Computational Physics VI*, edited by D. Stauffer (Singapore, World Scientific, 1998), p. 129.
12. B.A. Berg, T. Neuhaus, Phys. Lett. **267**, 249 (1991).
13. U.H.E. Hansmann, Y. Okamoto, J. Comp. Chem. **14**, 1333 (1993); **18**, 920 (1997).
14. A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); Phys. Rev. Lett. **63**, 1658(E) (1989), and references given in the erratum.
15. N. Madras, A.D. Sokal, J. Stat. Phys. **50**, 109 (1988).
16. M. Fukugita, H. Mino, M. Okawa, A. Ukawa, J. Stat. Phys. **59**, 1397 (1990), and references given therein.
17. K. Binder, D.W. Heermann, *Monte Carlo Simulation in Statistical Physics* (Springer-Verlag, Berlin, 1988).
18. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes*, 2nd edn. (Cambridge University Press, New York, 1992), p. 657.