Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 350 (2005) 38–44
PHYSICA A
www.elsevier.com/locate/physa

Finite size effects on calorimetric cooperativity of two-state proteins

Mai Suan Li[a,*], D.K. Klimov[b], D. Thirumalai[b]

[a]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*
[b]*Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

Available online 8 December 2004

Abstract

Finite size effects on the calorimetric cooperatity of the folding-unfolding transition in two-state proteins are considered using the Go lattice models with and without side chains. We show that for models without side chains a dimensionless measure of calorimetric cooperativity κ_2 defined as the ratio of the van't Hoff to calorimetric enthalpy does not depend on the number of amino acids N. The average value $\overline{\kappa_2} \approx \frac{3}{4}$ is lower than the experimental value $\kappa_2 \approx 1$. For models with side chains κ_2 approaches unity as $\kappa_2 \sim N^{\mu}$, where $\mu \approx 0.17$. Above the critical chain length $N_c \approx 135$ these models can mimic the truly all-or-non folding–unfolding transition.

PACS: 61.42.+h; 87.15.Da; 64.60.Cn; 64.60.Kw

Keywords: Protein folding; Calorimetric cooperativity; Lattice model; Lattice model with side chain; Monte Carlo simulation

1. Introduction

Single domain globular proteins, which are finite sized objects, undergo remarkably cooperative transitions from an ensemble of unfolded states to

*Corresponding author.
E-mail address: masli@ifpan.edu.pl (M.S. Li).

0378-4371/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2004.11.029

well-ordered folded (or native) states as the temperature is lowered [1]. In many cases, the transition to the native state takes place in an apparent two-state manner, i.e., the only detectable species are the native (more precisely, the ensemble of conformations belonging to the native basin of attraction [2]) or unfolded states [3]. In order to characterize the two-state folding one can use the dimensionless quantity κ_2 [4]

$$\kappa_2 = \Delta H_{vh}/\Delta H_{cal}\,, \tag{1}$$

where $\Delta H_{vh} = 2T_{\max}\sqrt{k_B C_P(T_{\max})}$ and $\Delta H_{cal} = \int_0^\infty C_P(T)\,\mathrm{d}T$, are the van't Hoff and the calorimetric enthalpy, respectively, $C_P(T)$ is the specific heat. κ_2 may be considered as a measure of the calorimetric cooperativity. Since real globular proteins have κ_2 very close to unity (chymotrypsin inhibitor 2 is a prime example [5]) it was proposed that [6] $\kappa_2 \approx 1$ can serve as one of requirements for realistic models of proteins. There are technical problems in evaluating κ_2 using experiments or computations. Inadequate treatment of baseline subtractions in $C_P(T)$ obscures estimates of κ_2. As a result it is possible that even sequences with $\kappa_2 \approx 1$ may not clearly be two-state folders. Nevertheless, κ_2 or related measures have often been used as a measure of calorimetric cooperativity.

In series of works [4,6–8] Chan et al. have shown that the calorimetric criterion is difficult to satisfy theoretically. Even Go models [9] which are more cooperative than others (2-letter, 3-letter and 20-letter models) have κ_2 notably smaller than 1. The studies of the Chan group are limited to few sequences and it remains, therefore, unclear if the Go modeling can meet the calorimetric requirement. One of our goals is to try to solve this problem by carrying out comprehensive simulations of lattice Go models.

Another dimensionless measure of thermodynamic cooperativity is Ω_c defined as follows [10]:

$$\Omega_c = \frac{T_F^2}{\Delta T}\left(\frac{\mathrm{d}\langle\chi\rangle}{\mathrm{d}T}\right)_{T=T_F}. \tag{2}$$

Here χ is the structural overlap with the native state and it can be identified as the probability of occupation of the native basin of attraction, T_F is the folding temperature and ΔT is the transition width [11]. Ω_c may be referred to as the structural cooperativity. Recently, we have shown that [12] it grows with the chain length as $\Omega_c \sim N^{\zeta}$, where the universal exponent $\zeta \approx 2.22$. This result is supported by experimental data collected for 32 two-state wild-type proteins and by simulations for lattice models. The main goal of this paper is to consider the finite size effects on κ_2 of two-state folders with the help of lattice Go models and Monte Carlo simulations. From the definition of κ_2 it follows that it should be independent of N because both ΔH_{vh} and ΔH_{cal} are extensive variables. However, the approach to the asymptotic behavior is unclear.

We have studied two classes of models: lattice models without side chains (LM) and lattice models with side chain (LMSC). For the first class, in accord with experiments, κ_2 was found to be scale-invariant at least up to $N \leqslant 80$. However, for 78 sequences studied their average value $\overline{\kappa_2} \approx \frac{3}{4}$ which is clearly smaller unity. Thus,

in agreement with the previous results [4,6–8], Go LMs do not satisfy the proteinlike cooperativity principle although they are minimally frustrated.

For Go LMSCs we have found that κ_2 scales with N as

$$\kappa_2 \sim N^{\mu} \tag{3}$$

before reaching the maximal value 1 at the critical value $N_c \approx 135$. Here exponent $\mu = 0.17 \pm 0.02$.These results suggest that κ_2 becomes scale-invariant for $N \gtrsim N_c$ and the LMSCs can meet the strict calorimetric cooperativity criterion only for this range of system sizes. If one assumes that the all-or-non folding takes place at $\kappa_2 \gtrsim 0.9$ then the critical value N_c is reduced to $N^* = 70$ (see below). In this case the LMSC with $N \gtrsim N^*$ can capture the calorimetric behavior of two-state proteins.

2. Models and method

In the coarse grained representation of LM, each amino acid is represented as a single bead confined to the vertices of a cubic lattice [13]. The LMSC is also modeled on a cubic lattice by a backbone (BB) sequence of N beads, to which a "side" bead, representing a side chain, is attached. The peptide bond and the α-carbon are given by a single bead and the system has in total $2N$ beads. Self-avoidance is imposed, i.e., any backbone and side beads cannot occupy the same lattice site more than once.

In the LMSC the energy of a conformation is [14,15]

$$E = \varepsilon_{bb} \sum_{i=1, j>i+1}^{N} \delta_{r_{ij}^{bb},a} + \varepsilon_{bs} \sum_{i=1, j\neq i}^{N} \delta_{r_{ij}^{bs},a} + \varepsilon_{ss} \sum_{i=1, j>i}^{N} \delta_{r_{ij}^{ss},a}\,, \tag{4}$$

where $\varepsilon_{bb}, \varepsilon_{bs}$ and ε_{ss} are BB–BB, BB–SC and SC–SC contact energies. r_{ij}^{bb}, r_{ij}^{bs} and r_{ij}^{ss} are the distances between the ith and jth residues for the BB–BB, BB–SC and SC–SC pairs, respectively, a is lattice spacing. Energies $\varepsilon_{bb}, \varepsilon_{bs}$ and ε_{ss} are chosen to be -1 for native contacts and 0 for non-native ones. For the LM the energy in Eq. (4) has only the BB term.

The specific heat in Eq. (1) is defined as the energy fluctuation. For LMSC the overlap function χ is defined as

$$\chi = \frac{1}{2N^2 - 3N + 1}\left[\sum_{i<j} \delta(r_{ij}^{ss} - r_{ij}^{ss,N}) + \sum_{i<j+1} \delta(r_{ij}^{bb} - r_{ij}^{bb,N}) + \sum_{i\neq j} \delta(r_{ij}^{bs} - r_{ij}^{bs,N})\right], \tag{5}$$

where the upper script N refers to the native state and factor $2N^2 - 3N + 1$ ensures that $\chi = 1$ in the native conformation. The last equation with only the BB term is applied to the LMs.

The Monte Carlo simulations were carried out using the move set MS3 [15–17] which involves single, double and triple bead moves. Because this move set involves multiparticle updates it is much more efficient compared to the standard move set

[18]. The thermodynamic properties are calculated using the multiple histogram method [19]. Sequences are selected as two-state folders if their free energy plotted against the number of native contacts has two well-defined minima.

3. Results

Fig. 1a shows the typical native conformation of the $N = 40$ LMSC sequence. The free energy is calculated as a function of the number of native contacts, which is treated as an approximate reaction coordinate for Go models, and the corresponding results obtained at $T = T_F$ are shown in Fig. 1b. Since the free energy profile has only one local maximum located at the transition state this sequence is a two-state folder. Clearly, for Go models the peaks of C_P and $\mathrm{d}\langle\chi\rangle/\mathrm{d}T$ coincide (Fig. 1c).

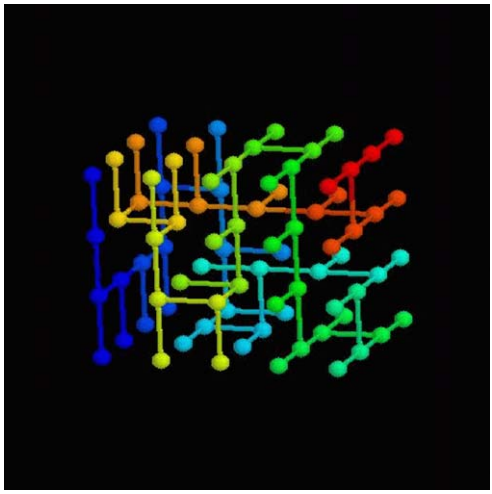
Fig. 2a shows the structural cooperativity against the calorimetric one for a given value of N. As expected, Ω_c grows with κ_2 for both LMs and LMSCs. However, the relation between these quantities becomes non-trivial if we combine the results for all values of N (Fig. 2b and c). The correlation remains strong for LMSCs but surprisingly it almost vanishes for LMs. It is not clear if the absence of correlation for the LMs is intrinsic or it is merely an artifact of the limited set of data. Clarification of this point requires further investigation. From all sequences 176 sequences studied (78 LM sequences and 98 LMSC ones) 10 sequences have $\kappa_2 \gtrsim 0.85$ and only one sequence which has $\kappa_2 \approx 0.9$ nearly satisfies the calorimetric cooperativity principle.

Since κ_2 of the LMs is not sensitive to N we can calculate its averaged value over the whole data set (78 sequences) and obtain $\overline{\kappa_2} \approx \frac{3}{4}$ which is notably smaller than unity. Thus our results, which are in accord with Kaya and Chan, also suggest that it is hard to meet the calorimetric criterion for Go LMs for any chain length. Using the relation $\kappa_2 = \sqrt{1 - 4(T_G/T_F)^2}$ derived from the random energy model [7,20,6], where T_G is interpreted as the temperature below which folding kinetics is dominated by trapping mechanisms [21], we obtain $T_F/T_G = \frac{8}{\sqrt{7}} \approx 3$. This value is far below the proposed $T_F/T_G = 4.6$ [6] required for the two-state melting with $\kappa_2 = 0.9$ but higher than, say, $T_F/T_G = 1.6$ for three-letter models [7].

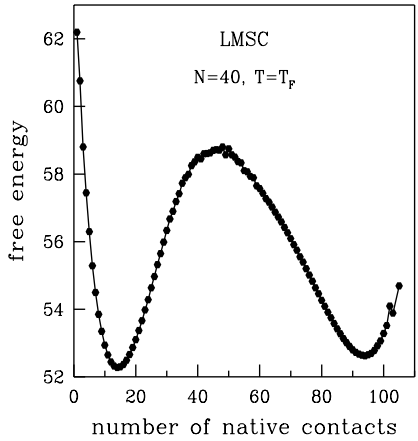
The difference in the scaling behavior of LMs and LMSCs is clearly seen in Fig. 3a where the size effect is visible only for sequences with SC. From the log–log plot (Fig. 3b) we obtain exponent $\mu = 0.17 \pm 0.02$. Interpolating our results to $\kappa_2 = 1$ we find the critical length $N_c \approx 135$ above which LMSCs always satisfy the calorimetric cooperativity requirement. If we assume that the transition is two-state if $\kappa_2 \gtrsim 0.9$ then the calorimetric cooperativity is satisfied for $N \gtrsim N^*$, where $N^* \approx 70$.

4. Conclusion

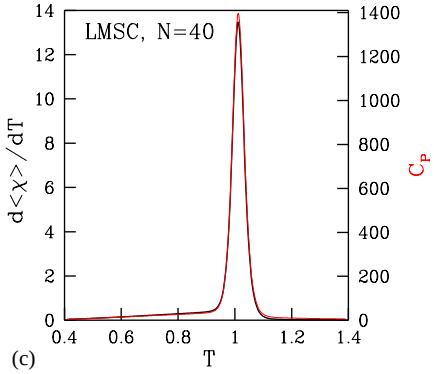
We have shown that for a given system size the structural cooperativity correlates with the calorimetric one. The scaling of the calorimetric cooperativity has been examined for lattice two-state Go models of proteins. The LMs superficially mimic



(a)



(b)



(c)

Fig. 1. (a) Typical native conformation of $N = 40$ of the LMSC. The BB and SC beads occupy sites of the compact $4 \times 4 \times 5$ lattice. (b) Dependence of the free energy (measured in $k_B T$) obtained for the sequence whose the native conformation is shown in a) on the number of native contacts at $T = T_F$. Since the free energy has only one local maximum at the transition state this sequence is a two-state folder. (c) Temperature dependence of $\mathrm{d}\langle\chi\rangle/\mathrm{d}T$ (black) and C_P (red, right-hand scale) for the sequence whose the native conformation is shown in a).

experiments in the sense that κ_2 is almost insensitive to the system sizes. However, they are not able to reproduce the experimental value $\kappa_2 \approx 1$. The rate of success for designing a Go LM which have $\kappa_2 \gtrsim 0.9$ is rather low (about 1%). The lack of scaling of LM folding cooperativity with chain length prevents these models to describe the cooperativity of wild-type proteins. This appears to be an inherent deficiency of LM without side chains.

For the Go LMSCs κ_2 depends on the system size up to the critical size N_c above which the full requirement of the calorimetric cooperativity is satisfied. Their advantage is that the criterion $\kappa_2 \gtrsim 0.9$ may be satisfied for relatively small globular

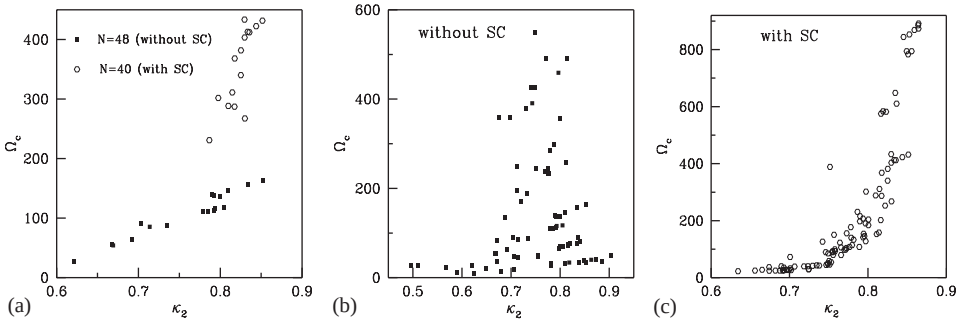


Fig. 2. The dependence of Ω_c on κ_2 for $N = 48$ LMs (solid squares, 18 sequences) and $N = 40$ LMSC (open hexagons, 15 sequences) (a), for all N LMs (b) and for all N LMSCs (c). For LMs we have studied $N = 27$ (17), 36 (17), 48 (18), 64 (15) and 80 (11) and for LMSCs—$N = 18$ (30), 24 (18), 32 (20), 40 (15) and 50 (15). Numbers of studied sequences are indicated in parenthesis.

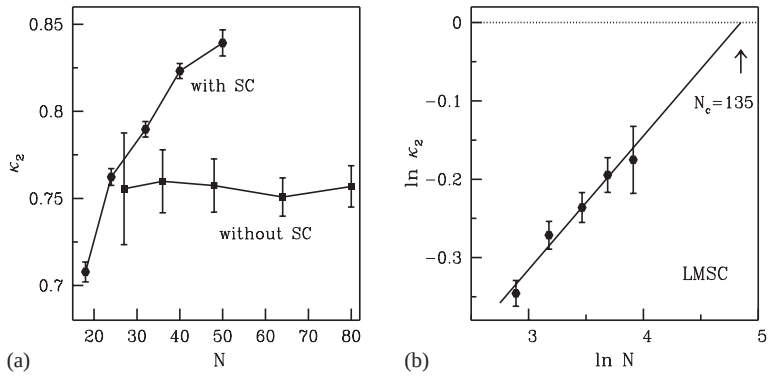


Fig. 3. (a) Dependence of κ_2 on N for LMs (solid squares) and LMSCs (solid hexagons). The sequences are the same as in Fig. 2. (b) The same as for LMSCs in a) but data are shown in the log–log plot. The dotted line refers to $\kappa_2 = 1$. The solid straight line is linear fit $y = -0.809 + 0.165x$ (the correlation coefficient is 0.96). It crosses the $\kappa_2 = 1$ line at the critical value $N_c = 135$.

proteins ($N \sim N^* = 70$). Our study shows that incorporation of side chains in protein LM represents a crucial modification, which makes LMSC protein-like.

It should be noted that we have considered the pairwise interaction (4) for Go models and it may be the reason why the calorimetric criterion is hard to fulfill even for LMSCs. The multiparticle interactions may be required to quantitatively describe cooperativity seen in proteins [22,8].

Acknowledgements

This work was supported by the KBN Grant No 1P03B01827 and the National Science Foundation Grant (NSF CHE-0209340). M.S.L. thanks H.S. Chan for providing Ref. [6].

References

[1] A.V. Finkelstein, O.B. Ptitsyn, Protein Physics: A Course of Lectures, Academic Press, New York, 2002.
[2] M.S. Li, M. Cieplak, J. Phys. A 32 (1999) 5577.
[3] (a) D. Poland, H.A. Scheraga, Theory of Helix-Coil Transitions in Biopolymers, Academic Press, New York,1970
(b) T.E. Creighton, Proteins, Structures and Molecular Principles, W.H. Freeman & Co., New York, 1993
(c) P.L. Privalov, Adv. Phys. Chem. 33 (1979) 167.
[4] H. Kaya, H.S. Chan, Phys. Rev. Lett. 85 (2000) 4823–4826.
[5] S.E. Jackson, A.R. Fersht, Biochemistry 30 (1991) 10428.
[6] H.S. Chan, S. Shimizu, H. Kaya, Meth. Enzymol. 380 (2004) 350.
[7] H. Kaya, H.S. Chan, Proteins Struct. Funct. Genet. 40 (2000) 637.
[8] H. Kaya, H.S. Chan, J. Mol. Biol. 326 (2003) 911.
[9] N. Go, Annu. Rev. Biophys. Bioeng. 12 (1983) 183.
[10] C.J. Camacho, D. Thirumalai, Proc. Natl. Acad. Sci. USA 90 (1993) 6369.
[11] M.S. Li, D.K. Klimov, D. Thirumalai, Polymer 45 (2004) 573.
[12] M.S. Li, D.K. Klimov, D. Thirumalai, Phys. Rev. Lett. (in press), q-bio.BM/0411050.
[13] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, H.S. Chan, Protein Sci. 4 (1995) 561.
[14] D.K. Klimov, D. Thirumalai, Fold. Des. 3 (1998) 127.
[15] M.S. Li, D.K. Klimov, D. Thirumalai, J. Phys. Chem. B 106 (2002) 8302.
[16] M.R. Betancourt, J. Chem. Phys. 109 (1998) 1545.
[17] M.S. Li, D.K. Klimov, D. Thirumalai, Comp. Phys. Commun. 147 (2002) 625.
[18] H.J. Hilhorst, J.M. Deutch, J. Chem. Phys. 63 (1975) 5153.
[19] A.M. Ferrenberg, R.H. Swendsen, Phys. Rev. Lett. 63 (1989) 1195.
[20] H.S. Chan, Proteins Struct. Funct. Genet. 40 (2000) 543.
[21] J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Annu. Rev. Phys. Chem. 48 (1997) 545.
[22] J. Tsai, M. Gerstein, M. Levitt. Prot. Sci. 6 (1997) 2606.