PHYSICAL REVIEW E, VOLUME 63, 032901

Thermodynamically important contacts in folding of model proteins

Antonio Scala,[1,*] Nikolay V. Dokholyan,[1,2,†] Sergey V. Buldyrev,[1] and H. Eugene Stanley[1]
[1]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215*
[2]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138*
(Received 13 October 2000; published 23 February 2001)

We introduce a quantity, the entropic susceptibility, that measures the thermodynamic importance—for the folding transition—of the contacts between amino acids in model proteins. Using this quantity, we find that only one equilibrium run of a computer simulation of a model protein is sufficient to select a subset of contacts that give rise to the peak in the specific heat observed at the folding transition. To illustrate the method, we identify thermodynamically important contacts in a model 46-mer. We show that only about 50% of all contacts present in the protein native state are responsible for the sharp peak in the specific heat at the folding transition temperature, while the remaining 50% of contacts do not affect the specific heat.

DOI: 10.1103/PhysRevE.63.032901 PACS number(s): 87.15.Aa, 87.15.Cc

Proteins are heteropolymers, composed of 20 types of amino acids, that perform specific functions. The amino acid composition of proteins determines their unique structure, function, and folding kinetics. Understanding the relevance of the interactions between amino acids to protein folding is a complex task that has been the subject of a number of theoretical and experimental studies [1–13]. The transition from the unfolded to the folded state of a protein is accompanied by a drastic reduction of the entropy. In one popular scenario, the folding transition for short proteins is analogous to the nucleation process at a first-order transition [2,6,8,12], with competition between two free-energy minima: the folded state with low energy and entropy and the unfolded state with high energy and entropy. These two minima are separated by a free-energy barrier corresponding to the transition states.

At the folding transition temperature T_f, there is an abrupt change in the energy of the system resulting in a pronounced peak in the specific heat. At T_f, a small increase in interaction energy ϵ_{ij} between amino acids i and j ''contact strength'' results in rapid transition to the folding state, while a small decrease in contact strength results in transition to the unfolded state. Temperature is measured in units of k_B/ϵ [10]. However, different amino acids have a different contribution to the folding transition. Small variation in ϵ_{ij} for different pairs i and j has a different effect on the folding transition. Here, we study the thermodynamic importance of each interaction during folding by computing the entropic susceptibility—the response function to a small perturbation of ϵ_{ij}.

We assume that the protein potential energy is additive in the pair potentials (contacts)

$$U \equiv \frac{1}{2}\sum_{i,j} U_{ij} \equiv \frac{1}{2}\sum_{i,j} \epsilon_{ij}\phi(\vec{r}_i,\vec{r}_j), \tag{1}$$

where U_{ij} is the energy of a single pair, $\phi(\vec{r}_i,\vec{r}_j)$ models the shape of the potential and protein at positions $\vec{r}_i$ and $\vec{r}_j$. We define the entropic susceptibility χ_{ij} of a contact between amino acids i and j as

$$\chi_{ij} \equiv -\epsilon_{ij}\frac{\partial S}{\partial \epsilon_{ij}} \equiv \beta^2(\langle U \quad U_{ij}\rangle - \langle U\rangle\langle U_{ij}\rangle) \equiv \beta^2\langle \delta U \quad \delta U_{ij}\rangle, \tag{2}$$

where $\delta U \equiv U - \langle U\rangle$, $\delta U_{ij} = U_{ij} - \langle U_{ij}\rangle$, and $\langle \dots \rangle$ is the Boltzmann average, $\beta = 1/T$ [14].

The entropic susceptibility measures the effect of a contact strength perturbation on the folding transition of the protein, thus identifying the thermodynamic relevance of such contact for the folding transition. Next, we demonstrate how this measure can be used to study contributions of the various contacts between amino acids in the protein for the folding transition. We simulate the ''beads on a string'' protein model [12], where the amino acids are hard spheres of unit mass, with the centers at the positions of the corresponding α carbons. The potentials of interaction between amino acids are square wells of depth ϵ_{ij}. We study the 46-mer (the folding transition temperature is at $T_f \approx 1.44$) that has been examined in [12]. We use Gō model for the contact potential, U_{ij}: U_{ij} is attractive ($\epsilon_{ij} = -1$) if the contact exists in the native (ground) state, otherwise the contact potential is repulsive ($\epsilon_{ij} = +1$) [15,16]. Our simulations employ the discrete molecular dynamics (MD) algorithm and are performed using methods described in [10,12,17]. The matrix of native contacts of the 46-mer is shown in Fig. 1. This particular 46-mer is known to have a stable native state and to undergo first-order-like folding $\rightleftharpoons$ unfolding transitions without stable intermediates [12].

We calculate χ_{ij} at different temperatures below and above T_f. A histogram of the values of χ_{ij} for various T is shown in Fig. 2. For $T \approx T_f$ the distribution has a pronounced peak at large values of χ_{ij}, which indicates that there is a separation of all contacts in two distinct sets with large and small values of χ_{ij}. The set of contacts with large values of χ_{ij} are ''thermodynamically important contacts,'' since for these contacts a small variation in their strength is correlated with a drastic change in the entropy of the model protein. To select the thermodynamically important contacts, we define a

---

[*] Also affiliated with INFM, Udr Rome ''La Sapienza,'' Piazzale Aldo Moro 2, I-00185, Rome, Italy.

[†] Corresponding author. Email address: dokh@wild.harvard.edu

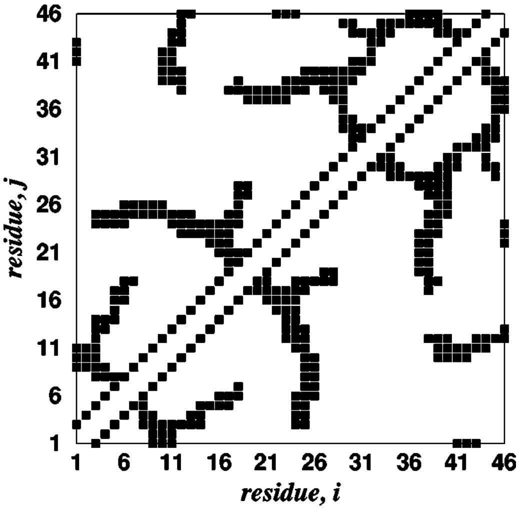

FIG. 1. Contact map of the native state of the 46-mer: dark squares denote residues that have contacts in the native state. Interactions are assigned according to Gō model [2]: all pairs of residues that have a contact in the native state are assigned attractive potential ($\epsilon_{ij}=-1$), while remaining pairs of residues are assigned to a repulsive potential ($\epsilon_{ij}=+1$).

temperature-dependent threshold $\chi_{\text{th}}(T)$ corresponding to the value of χ_{ij} where the distribution has a maximum in the space of all contacts.

Interestingly, thermodynamically important contacts are not randomly distributed in 3d space but are rather concentrated within well-defined structural regions in a model protein. Figure 3 represents the intensity map of the values χ_{ij}. In the upper part of Fig. 3, we show only the values of χ_{ij} that are above the threshold $\chi_{\text{th}}(T_f)=3.2$ that, according to our definition, corresponds to the thermodynamically important contacts. Although 50% of the contacts are above threshold, the filtered map of Fig. 3 shows that they are clustered together and are among well-defined regions of the model protein. Further, we find that the regions of thermodynamically important interactions $[\chi_{ij}(T)>\chi_{\text{th}}(T)]$ in the filtered map remain qualitatively the same as the ones shown in Fig. 3 for temperatures in the range $T=T_f\pm5\%$.

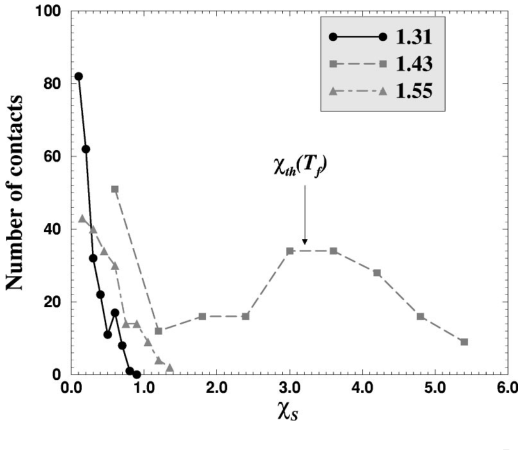

FIG. 2. Histogram of the entropic susceptibility χ_s [where s indicates arbitrary pairs (i,j)] for the 46-mer at temperatures $T=1.31$, 1.44, and 1.55. At T_f the distribution of χ_s has a pronounced peak, centered at $\chi_s=3.2$. Accordingly, we choose $\chi_{\text{th}}=3.2$.

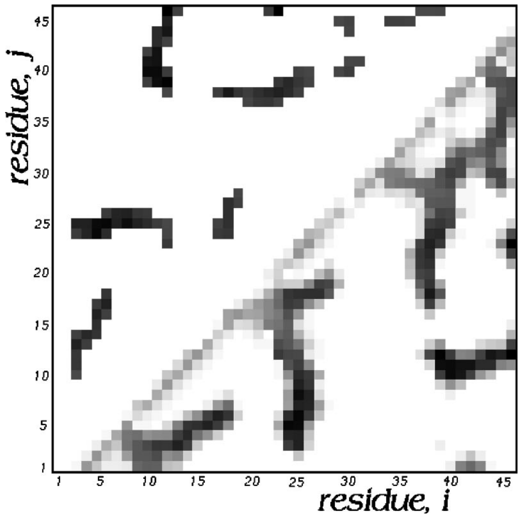

FIG. 3. The role of the thermodynamically important contacts: the lower corner is the intensity map of the entropic susceptibility χ_{ij} obtained from the simulations of the 46-mer at T_f. Darker colors correspond to higher values of χ_{ij}. The upper corner is the "filtered" map, where only values of χ_{ij} above the threshold $\chi_{\text{th}}=3.2$ defined in Fig. 2 are presented. Note that short-ranged contacts ($i\approx j$, corresponding to near-to-the-diagonal elements of the matrix χ_{ij}) do not contribute significantly to the change of entropy at the folding transition, while the relevant long-ranged contacts ($|i-j|\gg1$) are clustered in the islands in the filtered map. Specifically, nucleic contacts determined in [12] belong to the cluster in the top left corner with $i\approx10$ and $j\approx40$.

To verify that the thermodynamically important contacts are indeed thermodynamically the most relevant to the folding of our 46-mer, we measure the contribution of thermodynamically important contacts to the specific heat

$$C_V\equiv\frac{1}{2}\sum_{ij}\chi_{ij}. \tag{3}$$

Thus, we can interpret χ_{ij} as the contribution to C_V of a single contact. It is then possible to partition C_V as

$$C_V=C_V^{\text{TIC}}+C_V^{\text{others}}, \tag{4}$$

where C_V^{TIC} arises from the thermodynamically important contacts, and C_V^{others} from contacts below the threshold $\chi_{\text{th}}(T)$. Figure 4 shows that the thermodynamically important contacts give a sharp contribution to the specific heat around T_f. We find the number of contacts above threshold $\chi_{\text{th}}(T_f)$ is about 50% of the number of contacts in the native state, in agreement with Flory-type arguments [10].

It is natural to inquire whether the thermodynamically important contacts could be determined by analyzing the average contact energies $\langle U_{ij}\rangle$, which are related to the contact frequency map [11]. For square well potentials, $\langle U_{ij}\rangle=\epsilon_{ij}\langle f_{ij}\rangle$ where f_{ij} is the contact frequency for amino acids i and j. We find that the contacts with the largest values of f_{ij}

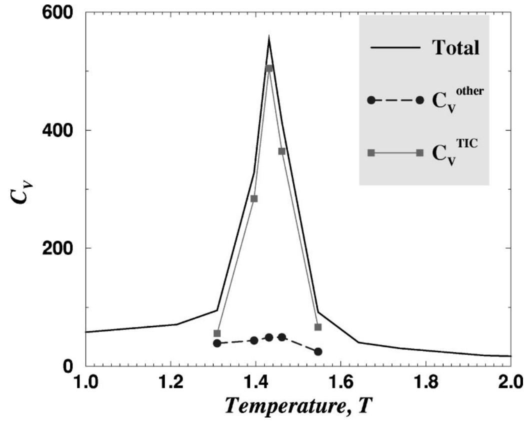

FIG. 4. Dimensionless specific heat as a function of dimensionless temperature of the 46-mer (solid line). There is a pronounced peak in the distribution of χ_{ij} in the region $1.31<T<1.55$. We separate the contribution to the specific heat from the thermodynamically important contacts $C_V^{\rm TIC}$ (squares) and the remaining ones $C_V^{\rm other}$ (circles). The number of contacts above threshold is always found to be approximatively 50% of the number of native contacts.

are the nearest and the next nearest neighbors. Thus, in order to account for the long-range contacts we have to go beyond the estimation of frequencies. An alternative way of computing the entropic susceptibility is to note that the contact frequencies are related to the change in free energy F

$$\epsilon_{ij}\frac{\partial F}{\partial \epsilon_{ij}} \equiv \langle U_{ij}\rangle \tag{5}$$

and therefore the entropic susceptibility can be rewritten as

$$\chi_{ij} \equiv -\epsilon_{ij}\frac{\partial}{\partial \epsilon_{ij}}\left(-\frac{\partial F}{\partial T}\right) = \frac{\partial \langle U_{ij}\rangle}{\partial T} = \epsilon_{ij}\frac{\partial f_{ij}}{\partial T}. \tag{6}$$

Thus, the information about the thermodynamically important contacts can be inferred from the temperature *derivative* of the frequency map.

''Core contacts'' were defined in [10] as those that form most stable elements of the protein three-dimensional structure that remains intact at folding transition temperature. Specifically, they were defined as contacts that are present with frequency above 0.5 at T_f. Molecular dynamics simulations performed at $T=T_f$ (see Fig. 5) show that these contacts are mostly short range. (The range for the contact between residues i and j is defined as $|i-j|$.) This result is not surprising since local contacts can form with high probability even in the unfolded state at $T\approx T_f$.

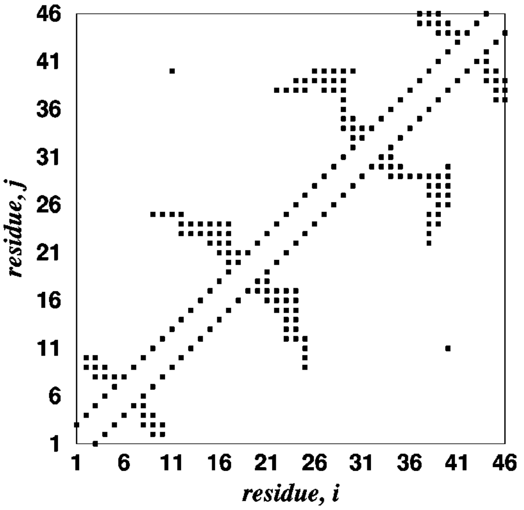

FIG. 5. The map of core contacts of the 46-mer: the dark squares denote contacts between any two residues i and j with frequencies $f_{ij}>0.5$. The core is comprised mostly of local contacts.

In contrast, we find that the thermodynamically important contacts are mostly long range, for which $|i-j|\gg 1$ (see Fig. 3). According to our definition, the thermodynamically important contacts correlate with the potential energy, thus, they are likely to be present in the folded state with the low potential energy and are likely to be absent in the unfolded states with high potential energy. Therefore, we believe that they are important for stabilization of the native structure. This hypothesis is in agreement with the general observation [18,19] that long-range interactions are important for protein stabilization.

We also find that the set of the thermodynamically important contacts contains all five nucleic contacts [(11,39), (10,40), (11,40), (10,41), and (11,41)] discovered in [12], indicating the dual role some amino acids play in protein folding: the nucleic residues, which play crucial role in the kinetics of folding transition, may also be important for stabilizing proteins in their native state. The evidence for the existence of such residues is supported by evolutionary [13,20] and phenomenological studies [21–24].

In conclusion, we demonstrate that by calculating the cross correlations between the potential energy of a single contact and the total potential energy of a model protein, it is possible to identify the set of contacts that are thermodynamically most relevant to the folding process. The tool of identifying thermodynamically important contacts is simple and can be implemented in the molecular dynamics studies of model proteins. The computational effort can be directed to aid experimental studies of real proteins.

We thank F. Sciortino, E. I. Shakhnovich, and M. Vendruscolo for very useful discussions. N.V.D. is supported by NIH. The Center for Polymer Studies acknowledges the support of the NSF.

---

[1] C. B. Anifsen, Science **181**, 223 (1973).

[2] N. Gō, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).

[3] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).

[4] K. A. Dill, Biochemistry **29**, 7133 (1990).

[5] M. Karplus and E. I. Shakhnovich, in *Protein Folding*, edited by T. Creighton (Freeman, New York, 1994), pp. 127–196.

[6] E. I. Shakhnovich, V. I. Abkevich, and O. Ptitsyn, Nature (London) **379**, 96 (1996).

[7] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).

[8] E. I. Shakhnovich, Curr. Opin. Struct. Biol. **7**, 29 (1997).

[9] H. Li, C. Tang, and N. S. Wingreen, Phys. Rev. Lett. **79**, 765 (1997).

[10] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, Folding Des. **3**, 577 (1998).

[11] C. Micheletti, J. R. Banavar, A. Maritan, and F. Seno, Phys. Rev. Lett. **82**, 3372 (1998).

[12] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, J. Mol. Biol. **296**, 1183 (2000).

[13] N. V. Dokholyan, L. A. Mirny, and E. I. Shakhnovich, e-print cond-mat/0007084.

[14] In general, the potential energy can take the form $U \equiv \Sigma_n \Sigma_{i_1,\ldots,i_n} U^{(n)}_{i_1,\ldots,i_n}$, where $U^{(n)}_{i_1,\ldots,i_n}$ is the n-body interaction between amino acids $i_1,\ldots,i_n$. The resulting expression for the entropic susceptibility is defined then as $\chi^{(n)}_{i_1,\ldots,i_n} = \beta^2 \langle \delta U \quad \delta U^{(n)}_{i_1,\ldots,i_n} \rangle$.

[15] The repulsive potential is necessary to ensure the existence of a nondegenerate ground (folded) state.

[16] N. Gō and H. Abe, Biopolymers **20**, 991 (1981); H. Abe and N. Gō, *ibid.* **20**, 1013 (1981).

[17] B. J. Alder and T. E. Wainwright, J. Chem. Phys. **31**, 459 (1959); A. Yu. Grosberg and A. R. Khokhlov, *Giant Molecules* (Academic Press, Boston, 1997), see Appendix; M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987), Chap. 3; D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge University Press, Cambridge, 1997), Chap. 12.

[18] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, J. Mol. Biol. **252**, 460 (1995).

[19] S. Govindarajan and R. A. Goldstein, Biopolymers **36**, 43 (1995).

[20] N. V. Dokholyan and E. I. Shakhnovich (unpublished).

[21] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).

[22] A. R. Fersht, Proc. Natl. Acad. Sci. U.S.A. **97**, 1525 (2000).

[23] N. V. Dokholyan, L. Li and E. I. Shakhnovich (unpublished).

[24] F. Cecconi, C. Micheletti, P. Carloni, and A. Maritan (unpublished).