STRUCTURAL SUSCEPTIBILITIES IN TOY MODELS OF PROTEINS

Mai Suan Li

Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland
and Institut für Theoretische Physik, Universität zu Köln, Zülpicher Straße 77, D-50937 Köln, Germany

New definitions of the structural susceptibilities based on the fluctuations of distances to the native state of toy protein models are proposed. The calculation of such susceptibilities does not require the basin of native state and the folding temperature can be defined from their peak if the native conformation is compact. The number of peaks in the derivatives of distances to the native state with respect to temperature, when plotted versus temperature, may serve as a criterion for foldability. The thermodynamics quantities are obtained by Monte Carlo and molecular dynamics simulations.

PACS Nos. 71.28.+d, 71.27.+a

Understanding of many aspects of protein folding has been recently advanced through studies of toy lattice models \[\[\text{[1,2]}\]. A more realistic modelling, however, requires considering off-lattice systems. In lattice models, the native state is usually non-degenerate and it coincides with the ground state of the systems. In the case of off-lattice models the native state has a zero measure and delineating boundaries of the native basin in off-lattice systems is vital for studies of almost all equilibrium and dynamical properties. For instance, stability of a protein requires considering off-lattice systems. In lattice models, however, ambiguous because it depends on a choice of contacts. The definition of the native contacts remains, however, ambiguous because it depends on a choice of the cutoff distance. We have developed two systematic approaches to delineate the native basin \[\[\text{[3]}\]. One of them is based on exploring the saddle points on selected trajectories emerging from the native state. In the second approach, the basin is determined by monitoring random distortions in the shape of the protein around the native state. It should be noted that the implementation of these methods becomes difficult in the case of long chains. The question we ask in the present paper is what one can learn about the folding thermodynamics and the foldability of the off-lattice sequences without the knowledge of \[\[\text{δ}_c\].

We start our discussion by introducing the following distances to the native state

\[
\delta_d = \sqrt{\frac{2}{N^2 - 3N + 2} \sum_{i \neq j, j \pm 1} (d_{ij} - d^{N\text{AT}}_{ij})^2},
\delta_{ba} = \sqrt{\frac{1}{N - 2} \sum_{i=1}^{N-2} (\theta_i - \theta_i^{N\text{AT}})^2},
\delta_{da} = \sqrt{\frac{1}{N - 3} \sum_{i=1}^{N-3} (\phi_i - \phi_i^{N\text{AT}})^2}.
\]

Here \(d_{ij} = |\vec{r}_i - \vec{r}_j|\) are the monomer to monomer distances in the given structure, \(N\) is a number of beads. The subscripts \(d, ba\) and \(da\) refer to the distances, the bond angles and the dihedral angles, respectively. The superscript \(N\text{AT}\) corresponds to the native state. The bond angle, \(\theta_i\), is defined as the angle between two successive vectors \(\vec{v}_i\) and \(\vec{v}_{i+1}\), where \(\vec{v}_i = \vec{r}_{i+1} - \vec{r}_i\). The dihedral angle, \(\phi_i\), is the angle between two vector products \(\vec{v}_{i-1} \times \vec{v}_i\) and \(\vec{v}_i \times \vec{v}_{i+1}\). The angular distances to the native state have not been studied in previous works.

We define the structural susceptibilities corresponding to the distances (1) as follows

\[
\chi_d = <\delta_d^2> - <\delta_d>^2,
\chi_{ba} = <\delta_{ba}^2> - <\delta_{ba}>^2,
\chi_{da} = <\delta_{da}^2> - <\delta_{da}>^2,
\]

(2)
where the angular brackets indicate a thermodynamic average. As one can see below, these three susceptibilities behave qualitatively in the same way. The sharpness of their peaks may be, however, different (see, for instance, Fig. 4) and it is useful to calculate all of them.

In the case of an off-lattice model the departures of the sequence geometry from its native conformation are usually described through the structural overlap function \[\[\text{[8,9]}\] as

\[
\delta_o = 1 - \frac{2}{N^2 - 3N + 2} \sum_{i \neq j, j \pm 1} \Theta(\delta_c - |d_{ij} - d^{N\text{AT}}_{ij}|),
\]

(3)
where \(\Theta(x)\) is the Heaviside function. The overlap structural susceptibility, \(\chi_o\), is then defined as the thermal fluctuation of \(\chi_o\)

\[
\chi_o = <\delta_o^2> - <\delta_o>^2.
\]

(4)
The maximum in \(\chi_o\), when plotted against \(T\), may be interpreted as a signature of the folding temperature \(T_f\) \[\[\text{[8]}\]. The advantage of the new definitions of the structural susceptibilities (2) compared to \(\chi_o\) is that the native basin \(\delta_c\) is not involved in their computation.
We have also studied the following derivatives of distances with respect to $T$

\[
D_d = \frac{d < \delta_d >}{dT}, D_{ba} = \frac{d < \delta_{ba} >}{dT},
\]
\[
D_{da} = \frac{d < \delta_{da} >}{dT}, D_o = \frac{d < \delta_o >}{dT},
\]
\[
D_g = \frac{d < R_g >}{dT},
\]

(5)

where $R_g$ is the gyration radius. Naively one can expect that the peaks of the derivatives $D$, when plotted against $T$, would coincide with those of the corresponding susceptibilities $\chi$. It is, however, true only when the native conformations are compact.

Using the Monte Carlo and the molecular dynamics simulations we have demonstrated that $T_f$ locates at the peaks of $\chi_d (D_d), \chi_{ba} (D_{ba})$ or $\chi_{da} (D_{da})$ provided the native conformations are compact. Thus, the determination of $T_f$ does not require the native basin $\delta_c$. This is the main advantage of the new quantities given by Eqs. (2) and (5).

The situation becomes more complicated when the native conformations are not compact. In this case the native basin is necessary for the accurate estimate of $T_f$. The information about the foldability may be, however, obtained without $\delta_c$ monitoring the temperature dependence of $D_d, D_{ba}$ and $D_{da}$. Namely, a good folder would have only one peak in $D_d, D_{ba}$ or $D_{da}$, when plotted against temperature, whereas a bad folder would have two peaks. This may serve as a criterion to distinguish the good folders from the bad ones.

We focus on four sequences whose native conformations are shown in Fig.1. The 27-monomer lattice chain, $L_{27}$, is a Go sequence [10]. Its Hamiltonian is as follows

\[
H = \sum_{i<j} \alpha_{ij} \Delta_{ij},
\]

(6)

where $\Delta_{ij} = 1$ if monomers $i$ and $j$ are in contact and $\Delta_{ij} = 0$ otherwise. The quantity $\alpha_{ij} = -1$ if monomers $i$ and $j$ are in contact in the native conformation and $\alpha_{ij} = 0$ otherwise. We use $L_{27}$ to check the behavior of the new quantities $\delta_{ba} (D_{ba})$ and $\delta_{ba} (D_{ba})$ for the lattice models.

The sequences denoted by $G$ and $R'$ are two-dimensional versions of the model introduced by Iori, Marinari and Parisi [1]. The Hamiltonian is given by

\[
H = \sum_{i\neq j} \{k(d_{ij} - d^0_i)^2 \delta_{i,j+1} + 4\epsilon \big( \frac{C}{d_{ij}} - \frac{A_{ij}}{d_{ij}^6} \big) \},
\]

(7)

where $i$ and $j$ range from 1 to $N=16$. $d_{ij}$ is measured in units of $\sigma$, the typical value of which is $\sigma = 5\AA$. We take $d^0_i$ to be equal to $21/6\sigma$ and $1.16\sigma$ for $G$ and $R'$, respectively [12]. The harmonic term in the Hamiltonian, with the spring constant $k$, couples the beads that are adjacent along the chain. The remaining terms represent the Lennard-Jones potential. Random values of $A_{ij}$ describe the quenched disorder. In Eq. (7) $\epsilon$ is the typical Lennard-Jones energy parameter. We adopt the units in which $C=1$ and consider $k$ to be equal to $25\epsilon$. Smaller values of $k$ may violate the self-avoidance of the chain. The coupling constants $A_{ij}$ for system $R'$ are listed in Ref. [12]. These are shifted Gaussian-distributed numbers with the strongest attracting couplings assigned to the native contacts. For system $G$, $A_{ij}$ is taken to be 1 or 0 for the native and non-native contacts respectively. System $R'$ has been shown to be structurally overconstrained and hard to fold.

![FIG. 1. Native conformations of four sequences studied in this work.](Image)

The helical system $H$ has a native state that mimics typical $\alpha$-helix secondary structures. In this case the distances between beads are assumed to have the length $d_0 = 3.8\AA$. As one proceeds along the helix axis from one bead to another, the bead’s azimuthal angle is rotated by $100^\circ$ and the azimuthal length is displaced by 1.5 $\AA$. The Hamiltonian used to describe the helix is a Go-like modification of Eq. (7) and it reads

\[
H = V^{BB} + V^{NAT} + V^{NON}
\]

(8)

The first term is a backbone potential which includes the harmonic and anharmonic interactions

\[
V^{BB} = \sum_{i=1}^{N-1} [k_1(d_{i,i+1} - d_0)^2 + k_2(d_{i,i+1} - d_0)^4].
\]

(9)

We take $d_0 = 3.8\AA$, $k_1 = \epsilon$ and $k_2 = 100\epsilon$. The interaction between residues which form native contacts in the target conformation is chosen to be of the Lennard-Jones form

\[
V^{NAT} = \sum_{i<j} 4\epsilon \left[ \frac{\sigma_{ij}}{d_{ij}} \right]^{12} - \left( \frac{\sigma_{ij}}{d_{ij}} \right)^6.
\]

(10)
We choose $\sigma_{ij}$ so that each contact in the native structure is stabilized at the minimum of the potential, i.e., $\sigma_{ij} = 2^{-1/6} d_N^{ij}$, where $d_N^{ij}$ is the length of the corresponding native contact. Residues that do not form the native contacts interact via a repulsive soft core potential $V_{\text{NON}}$, where

$$V_{ij}^{\text{NON}} = \sum_{i<j}^{\text{NON}} V_{ij}^{\text{NON}},$$

$$V_{ij}^{\text{NON}} = \begin{cases} 4\epsilon[(\sigma_0/\sigma_{ij})^{12} - (\sigma_0/\sigma_{ij})^6] + \epsilon, & d_{ij} < d_{\text{cut}} \\ 0, & d_{ij} > d_{\text{cut}}. \end{cases}$$

Here $\sigma_0 = 2^{-1/6} d_{\text{cut}}$, $d_{\text{cut}} = 5.5\text{Å}$. The difference between a Go and Go-like sequences is in the choice of the non-native contact interaction energy which is taken to be zero for the Go sequence and nonzero for the latter one.

FIG. 2. The temperature dependence of $C_v$, $\chi$ and $D$ for sequence L27. The arrow corresponds to $T_f$. $T_f$ is defined as a temperature at which the probability of being in the native state is $1/2$. $C_v$ and $D_v$ are denoted by dotted lines whereas $\chi_o$ and $D_o$ - by thick lines. The results are averaged over 50 starting conformations. The error bars are smaller than the symbol sizes.

The thermodynamics of L27 is studied by a Monte Carlo procedure that satisfies the detailed balance condition. The dynamics allows for single and two-monomer (crankshaft) moves. For each conformation of the chain one has $A$ possible moves and the maximum value of $A$, $A_{\text{max}}$, is equal to $A_{\text{max}} = N + 2$. In our 27-monomer case $A_{\text{max}} = 29$. For a conformation with $A$ possible moves, the probability to attempt any move is taken to be $A/A_{\text{max}}$ and the probability not to do any attempt is $1-A/A_{\text{max}}$. In addition, probability to do a single move is reduced by the factor of 0.2 and to do the double move - by 0.8. The attempts are rejected or accepted as in the standard Metropolis method. The equilibration is checked by monitoring the stability of data against at least three-times longer runs. We have used typically $10^6$ Monte Carlo steps (the first $5 \times 10^5$ steps are not taken into account when averaging).

FIG. 3. The same as in Fig. 2 but for sequence G. The native basin defined by the shape distortion approach is equal to $\delta_o = 0.2\sigma$. The results are averaged over 100 molecular dynamics trajectories.

Fig. 2 shows the temperature dependence of $C_v$, $\chi$ and $D$ for L27, where $\chi$ ($D$) is a common notation for $\chi_d$ ($D_d$), $\chi_{ba}$ ($D_{ba}$), $\chi_{da}$ ($D_{da}$) and $\chi_o$ ($D_o$). In this on-lattice case the overlap structural susceptibility $\chi_o$ is also given by Eq. (4) but $\delta_o$ reads as follows

$$\delta_o = 1 - \frac{2}{N^2 - 3N + 2} \sum_{i\neq j, j \geq 1} \delta(d_{ij} - d_{ij}^{\text{NAT}}).$$

For sequence L27 the peaks of all quantities are located at $T = T_f$. The fact that the maxima of $\chi_o$ and $D_o$ are located at the same position has been also observed for some on-lattice sequences. Our new result is that, similar to $\chi_o$, the susceptibilities based on the fluctuations of the distances to the native conformation and $D$ also give a correct position for $T_f$. According to the thermodynamic criterion, L27 should be a good folder because $T_f$ coincides with the collapse temperature $T_{\theta}$ ($T_{\theta}$ is defined as a temperature where $C_v$ develops a peak).
In order to study the time evolution of the off-lattice sequences $G, R'$ and $H$, we use the equations of motion by the Langevin uncorrelated noise terms:

$$m \ddot{x} = -\Gamma \dot{x} + F_c + \eta,$$

where $F_c = -\nabla \epsilon E_p$ and

$$\langle \eta(0) \eta(t) \rangle = 2 \Gamma k_B T \delta(t).$$

(14)

Here $k_B$ and $\Gamma$ are the Boltzmann constant and the kinetic coefficient, respectively. Equation (14) is integrated by the fifth order predictor-corrector method. The integration step is chosen to be 0.005$\tau$, where $\tau = m \sigma^2 / \epsilon$ is the characteristic time unit and $m$ is the mass of a bead. We take $\Gamma$ equal to 2. In the following, the temperature will be measured in the reduced units of $\epsilon/k_B$.

![Graph](image)

**FIG. 4.** The same as in Fig. 2 but for sequence $R'$. The native basin is equal to $\delta_c = 0.09\sigma$. The results are averaged over 170 molecular dynamics trajectories. The folding properties of $G$, $R'$ and $H$ were characterized in detail previously [2][3]. One of them, $R'$, is a bad folder and two others are good folders. We calculate the thermodynamic quantities of $G$, $R'$ and $H$ by averaging over many molecular dynamics trajectories using the native state as the starting configuration to make sure that the evolution takes place in the right part of the phase space [2]. For all of these sequences, the time used for averaging in each trajectory is 4000$\tau$ for each temperature. The first 2000$\tau$ are discarded.

Fig. 3 and 4 show the results for $G$ and $R'$. Since these sequences are two-dimensional, $\chi_{ba}$ and $D_{ba}$ corresponding to the dihedral angles do not appear. The basin was obtained by the shape distortion approach [4] and is equal to $\delta_c = 0.2\sigma$ and $\delta_o = 0.09\sigma$ for $G$ and $R'$, respectively. Within the error bars of 0.02 all of the maxima of $\chi, D$ and $C_v$ are located at the folding temperature $T_f$ ($T_f = 0.24 \pm 0.02$ and $T_f = 0.10 \pm 0.02$ for $G$ and $R'$, respectively). Therefore, the determination of $T_f$ does not require the native basin because it is enough to find the peak of $\chi$ (or of $D$) in which $\delta_c$ is not involved.

For $R'$, $\chi_o$ has only one peak at $T_f$ whereas $\chi_d$ and $\chi_{ba}$ have an additional maximum at $T = T_\theta$. Therefore, the advantage of $\chi_d$ and $\chi_{ba}$ compared to $\chi_o$ is that they allow to find not only $T_f$ but also $T_\theta$. Since the maximum of $D_g$ is broad around the folding temperature, it is better to locate $T_f$ as a second peak of $\chi_d$ ($D_d$) or $\chi_{ba}$ ($D_{ba}$). It demonstrates the another advantage of the new quantities compared to the standard quantity $D_g$.

It should be noted that the behavior of $\chi_d$ and $\chi_{ba}$ is qualitatively the same but there is a quantitative difference in the sharpness of their peaks. It is clear from Fig. 4 that at $T = T_f$ the maximum of $\chi_{ba}$ is more pronounced compared to that of $\chi_d$. An opposite situation takes place at $T = T_\theta$. So, the study of all of susceptibilities would help us to isolate peaks better.

The fact that $\chi_o$ has only one peak, but the others have two may be explained in the following way. Since $\chi_o$ is a fluctuation of the overlap with the native state it reflects the behavior of the system in the vicinity of the native basin and it should have, therefore, only one peak at $T_f$. The remaining susceptibilities related to the chain compactness would have two maxima at $T_f$ and $T_\theta$ where the topology changes abruptly.

The temperature dependence of $\chi, D$ and $C_v$ for the three-dimensional sequence $H$ is shown in Fig. 5. In this case we have the basin $\delta_c = 0.12\sigma$ and $T_f = 0.24 \pm 0.02$ [3][4]. Since $\chi_d, \chi_{ba}$ and $\chi_{da}$ do not display any peak in the relevant temperature interval, they cannot be used to determine $T_f$. It is also true for $D$ (their extremal points are located at $T = 0.325 \pm 0.025$ which is far from $T_f$). The overlap susceptibility $\chi_o$ has the maximum at $T_{\chi_o} = 0.275 \pm 0.025$. Within the error bars $T_{\chi_o}$ may be identified as $T_f$ but such an estimate is less accurate compared to the case of $G$ and $R'$. Furthermore its computation involves the native basin $\delta_c$.

From the results presented in Figs. 2 - 5 we propose the following criterion for foldability: a good folder would have only one peak in the derivatives of distances to the native state with respect to temperature whereas a bad folder has two. Our criterion is compatible with the fact that for the good folders the folding takes place just after the collapse transition. A three state scenario of folding is, however, more suitable for the bad folders [1]. Thus, one can still gain information about the foldability for $H$ without the native basin $\delta_c$.

The question we ask now is why $H$ is so different from the other sequences. The main difference is that its native conformation is not compact. It results in the non-trivial dependence of $R_g$ on $T$: $D_g$ does not develop a maximum but rather a minimum around the the collapse transition. This leads to the anomalous behavior of $\chi_d, \chi_{ba}$
and $\chi_{ha}$. 

![Graph](image_url)

**FIG. 5.** The same as in Fig. 2 but for sequence H. The native basin is equal to $\delta_c = 0.12\sigma$. The results are averaged over 200 molecular dynamics trajectories.

In conclusion, we have introduced several new structural susceptibilities as fluctuations of distances to the native conformation. If the native conformation of proteins is compact $T_f$ may be obtained from the peak of $\chi$ and the native basin is not required. For sequences with non-compact native conformation $\delta_c$ is not needed to establish the foldability but the accurate estimate of $T_f$ should involve it. The number of peaks in the derivatives of distances to the native state with respect to temperature, when plotted against $T$, may serve as a tool to distinguish between good and bad folders. The question of why the susceptibilities $\chi$ and the derivatives $D$ behave in the same way if the native conformations are compact remains to be elucidated. Nevertheless, in agreement with other studies (see, for instance, Ref. [20] and references there), our results indicate that the topology of the native state plays a crucial role in the folding process.

We thank M. Cieplak for useful discussions. This work was supported by Komitet Badan Naukowych (Poland; Grant number 2P03B-146 18).

[1] K. A. Dill, S. Bromberg, S. Yue, K. Fiebig, K. M. Yee, D. P. Thomas, and H. S. Chan, Protein Science 4, 561 (1995).
[2] V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Rev. Mod. Phys. 72, 259 (2000).
[3] A. Irback, C. Peterson and F. Pottast, Proc. Natl. Acad. Sci. USA 93, 9533 (1996); A. Irback, C. Peterson and F. Pottast, Phys. Rev. E 55, 860 (1997); A. Irback, C. Peterson and F. Pottast, and O. Sommelius, J. Chem. Phys. 107, 273 (1997)
[4] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. 79, 317 (1997).
[5] T. Veitshans, D. K. Klimov, and D. Thirumalai, Folding and Design 2, 1 (1997).
[6] C. Clementi, H. Nymeyer, and J. N. Onuchic, cond-mat/0003460; J. N. Onuchic, H. Nymeyer, H. García, A. E. Chaikine, and N. D. Socci, Adv. Protein Chem. 53, 87 (2000).
[7] M. S. Li and M. Cieplak, J. Phys. A 32, 5577 (1999).
[8] C. J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. USA 90, 6369 (1993).
[9] D. Thirumalai, J. Physique (Paris) I 5, 1457 (1995).
[10] N. Go and H. Abe, Biopolymers 20, 991 (1981).
[11] G. Iori, E. Marinari, and G. Parisi, J. Phys.A 24, 5349 (1991).
[12] M. S. Li and M. Cieplak, Phys. Rev. E 59, 970 (1999).
[13] T. X. Hoang and M. Cieplak, J. Chem. Phys. 112, 6851 (2000).
[14] M. Cieplak, M. Henkel, J. Karbowski, and J. R. Banavar, Phys. Rev. Lett. 80, 3654 (1998); M. Cieplak, M. Henkel, and J. R. Banavar, Cond. Matt. Phys. (Ukraine) 2, 369 (1999).
[15] H. S. Chan and K. A. Dill, J. Chem. Phys. 99, 2116 (1994); H. S. Chan and K. A. Dill, J. Chem. Phys. 100, 9238 (1994).
[16] D. K. Klimov and D. Thirumalai, Phys. Rev. Lett. 76, 4070 (1996).
[17] D. K. Klimov and D. Thirumalai, J. Chem. Phys. 109, 4119 (1998).
[18] M. P. Allen and D. J. Tildesley, Computer simulation of liquids, Oxford University Press, New York (1987).
[19] M. S. Li, M. Cieplak, and N. Sushko, Phys. Rev. E 63, 2000 (in press).
[20] D. Baker, Nature 405, 39 (2000).