ELSEVIER

Biophysical Chemistry 115 (2005) 145–152

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Topological properties of the energy landscape of small peptides

L. Bongini*

Centro Interdipartimentale per lo Studio delle Dinamiche Complesse, via Sansone, 1-I-50019 Sesto Fiorentino, Italy

Received 30 June 2004; received in revised form 17 November 2004; accepted 10 December 2004

Abstract

The topology of the potential energy landscape (PEL) underlying the dynamics of a two dimensional off-lattice model for a heteropolymer is analyzed for different sequences of amino-acids. A statistical characterization of the metastable minima and first-order saddles of the PEL highlights structural differences in the landscape of good and bad folding sequences and provides insight on the chain dynamics during folding.

Keywords: Protein folding; Energy landscape; Activation dynamics

1. Introduction

The characterization of the topology and of the structure of the potential energy landscapes (PEL) (or of the free energy landscapes (FEL)) is an effective tool for obtaining information about the dynamical and structural properties of many physical systems: supercooled liquids, glasses, atomic clusters and biomolecules [1]. After the pioneering works by Stillinger and Weber on "inherent" states of liquids [2], the direct investigation of the statistics and topology associated with the stationary points of PEL has been used successfully for identifying the structural arrest temperature of glasses [3] and supercooled liquids [4]. More recently, FEL analysis has been applied to the characterization of the energetic funnels associated with protein models [5–8]. In particular in [8] it is shown how the folding propensity and the folding cooperativity of a model depend on a interplay between the energetic bias toward the native configuration (NC) and the roughness of the potential energy surface. In order to reproduce the folding dynamics of a protein, it is necessary to identify the paths followed by the protein during the folding process, from the unfolded state to the NC, and the kind of dynamics performed by the protein to move from one inherent configuration to another. Preliminary results obtained for an off-lattice model suggest that the basic dynamical mechanism (at least within the native valley at the folding temperature) reduces to a combination of oscillations around local minima of the PEL and thermally activated jumps between the basins of attraction of different minima [9]. This jumping behavior was shown to result in a marked stepwise behavior of the reaction coordinate [10]. These results partially contrast with the assumption of a smooth diffusive dynamics in the reaction coordinate made by Bryngelson and Wolynes in their work on random-energy models [11] and the fact that in some on-lattice models of heteropolymers the jumping dynamics in the multi-dimensional configuration space does not show in the reaction coordinate [12]. This apparent contradiction is based on the fact that the overall kinetic behaviour of the system depends in a complex way both on the PEL shape and on the connectivity between different minima. It is therefore not surprising that different models show different behaviors or that the same model [13] behaves differently in different regimes. In order to achieve a better quantitative understanding of the influences of the PEL topology on the folding properties of a model it is in particular by identifying the meta-stable minima and the first-order saddles, which are known to play a major role in the dynamics. In this paper, a statistical characterization of the PEL for a simple mesoscopic off-lattice model of heteropolymers is presented. In particular, I am going to show that the folding propensity of a sequence arises from a delicate balance between energetic and entropic contributions which may be quantitatively accessed by using the appropriate

---

* Tel.: +46 852487250; fax: +46 8311101.

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.12.019

transition theory. This analysis is mainly based on a novel algorithm, which implements an efficient search strategy of the PEL stationary points.

2. The model

Although more realistic three-dimensional models exist [14,15], I choose for this study a minimal two-dimensional off-lattice model that is shown to reproduce many fundamental properties of proteins, namely spontaneous folding in a temperature range above which denaturation occurs and below which the system dynamics is found to undergo a dynamical arrest [10]. It must be stressed that this model would hardly provide a quantitatively precise description of the funnel of any real protein. Nonetheless, due to its compactness and ease of investigation, it proves a very useful tool to investigate limited issues such as the dependence of folding properties from the landscape topology. The model is a modification of the two-dimensional off-lattice model introduced by Stillinger et al. [16] and consists in a one-dimensional chain made of L point-like monomers (mimicking the residues of a real protein). For the sake of simplicity, only two types of residues are considered: hydrophobic (H) and polar (P) ones. Thus, each heteropolymer can be unambiguously identified by a sequence of binary variables $\{\xi_i\}$ (with i=1,$\ldots$, L) along the backbone, where ξ_i=1, if the ith residue is of type H, and ξ_i=−1, otherwise. The intramolecular potential is composed of three terms: a nearest-neighbor harmonic potential V_1 intended to maintain the bond distance almost constant, a three-body interaction V_2 that simulates the cost of local bending and a Lennard–Jones-like V_3 interaction acting among next-to-nearest neighboring monomers (i.e. $|i-j|>1$):

$$V_1(r_{i,i+1}) = \alpha(r_{i,i+1}-r_0)^2, \quad V_2(\theta_i) = \frac{1-\cos\theta_i}{16},$$
$$V_3(r_{i,j}) = \frac{1}{r_{i,j}^{12}} - \frac{c_{i,j}}{r_{i,j}^{6}} \tag{1}$$

Here $r_{i,j}$ denotes the distance between the ith and the jth monomer and θ_i is the bond angle centered on the ith residue. The parameters α and r_0 are fixed to 20 and 1, respectively. V_3 is the only contribution that depends on the nature of the monomers: the parameters $c_{i,j}$=1/8(1+ξ_i+ξ_j+5$\xi_i\xi_j$) are defined in such a way that the interaction is attractive if both residues are either hydrophobic or polar (with $c_{i,j}$=1 and 1/2, respectively), while it is repulsive if the residues belong to different species (with c_{ij}=−1/2). The Hamiltonian of the system writes as

$$H = \sum_{i=1}^{L} \frac{p_{x,i}^2 + p_{y,i}^2}{2} + \sum_{i=1}^{L-1} V_1(r_{i,i+1}) + \sum_{i=2}^{L-1} V_2(\theta_i)$$
$$+ \sum_{i=1}^{L-2} \sum_{j=i+2}^{L} V_3(r_{ij}, \xi_i, \xi_j) \tag{2}$$

where all monomers are assumed to have the same unitary mass and the momenta are defined as ($p_{x,i}$, $p_{y,i}$)=($\dot{x}_i$, $\dot{y}_i$).

Accurate Monte-Carlo (MC) simulations, performed by employing innovative schemes [17], have revealed that, in analogy with real proteins, only a few sequences systematically fold onto the same (native) structure and can be thereby called “good folders” [18,19]. These studies have been confirmed and complemented by direct molecular dynamics simulations [10].

In what follows, I shall limit my investigations to the three following cases, all of length L=20,

- [S0]=[0000 0000 0000 0000 0000] a homopolymer composed of hydrophobic residues only;
- [S1]=[0001 0001 0001 1001 1000] a sequence first studied in [19] (therein indicated with the code number 81) where it was identified as a good folder;
- [S4]=[1110 0100 0000 0001 0010] a randomly generated sequence with 6 P-type residues, identified as a bad folder in [10].

A reasonably accurate characterization of each sequence can be obtained by determining three transition temperatures. The first one, T_θ, denotes the temperature below which the polymer is in a collapsed rather than in a random-coil configuration [20]. The folding temperature T_f is the temperature below which the heteropolymer stays predominantly in the native valley. Here, analogously to [10], I define the native valley as the basins of attraction of the native configuration (NC) and of its neighboring minima. A quantitative estimate of T_f can be then obtained by determining the temperature at which the chain spends 50% of the time within the native valley. Finally, the glass-transition temperature T_g is the temperature below which the structural arrest of the system take place. These temperatures have been reported for the examined sequences in [9].

3. Minima and saddles database

In order to obtain a meaningful reproduction of the heteropolymer energy landscape in terms of a connectivity graph it is necessary to have a sufficiently exhaustive database, containing the potential energy minima and the first-order saddles connecting them. Even if at present I have not yet at disposal a complete description of the landscape, I shall report some preliminary results limited to collapsed configurations, i.e. to minima with an energy smaller than V_θ=V_0+$L \cdot T_\theta$, where V_0 is the potential energy of the NC.

The algorithm employed for obtaining first-order saddles is conceptually similar to the slowest slide method [21]. The basic difference between the two lies in the fact that the starting point for each saddle search, rather than being randomly selected between the highest potential energy configurations of a molecular dynamic trajectory, is chosen starting from previous knowledge about the position of a couple of minima of the potential energy. Care is devoted to ensure that, whenever a first-order saddle exists in-between the two minima, the algorithm will actually find it. The final aim of the proposed method is to provide a tool for a systematic search of saddles where the number of steps to find all possible inter-basin connections only depends on the complexity of the landscape itself and not on parameters such as temperature or integration time step of a molecular dynamics simulation. Another important feature of this algorithm, illustrated in detail in Appendix A, is that it can produce, as a final output of each search, both the coordinates of a first-order saddle or those of a minimum. More specifically if the basins of attraction of the two starting minima are directly neighboring it will find a first-order saddle. If instead the two starting minima are separated by one or more intermediate basins of attractions the final output will be the coordinates of one of intermediate minima. As a result this method stands as a bifunctional tool able to recover informations about saddles while at the same time finding new minima. It must be stressed that in practice the relative percentage of saddles and minima found by the algorithm will depend on the completeness of the minima database. In the beginning of the process when the minima database is still small and incomplete, the probability that the minima, though near according to some metric, are actually neighboring is very low. In this conditions the algorithm will mainly find new minima. On the contrary when the minima database will be rich enough the probability that nearby minima are actually neighboring will increase thus leading in an increase in the percentage of saddles found.

Actually, by starting form quite a limited number of minima (of the order of 50–100), the algorithm is able to rapidly enrich the database both of minima and of first-order saddles. However, since the number of new minima found by the algorithm initially grows much faster than the number of identified saddles, I have not yet been able to complete the analysis of all the possible connections among the minima in the database. The number of minima and saddles so far identified is reported in Table 1.

It must be stressed that in principle one expects the number of minima of the energy landscape of this and similar systems such as Lennard–Jones clusters to grow exponentially with the number L of monomers composing the chain and the number of saddles to be of the order of L-times the number of minima [22]. The number of minima and saddles found so far is then much inferior to the existing estimates. Still some characteristic differences emerge from the comparative study of the databases of different sequences. A first check of the energies of the identified minima shows however that only a small fraction of them lies above V_θ, thus indicating that the proposed analysis will not be able to account for the model behavior above the temperature of hydrophobic collapse.

Table 1
Number of identified minima and saddles for the three analyzed sequences

	S0	S4	S1
Number of minima in the initial database	47	43	89
Number of minima in the final database	195 362	27 484	47 762
Number of saddles in the final database	75 722	41 367	39 598

4. Statistical analysis of the minima

The main interest of this study is the dynamic behavior of the model protein when immersed in a solvent. Under this circumstances the dynamics of the system is described by a Langevin where the effects of the solvent molecules are approximated in terms of a frictional drag on the solute as well as random kicks associated with the thermal motions of the solvent molecules. The two terms must satisfy the fluctuation–dissipation relation in order to fix the correct temperature. For sufficiently low temperatures a Langevin dynamics on a rough energetic landscape consists in a thermally activated hopping dynamics between basins of attraction of different minima. In order for the system to move from minimum i of the potential to minimum j the basins of attraction of the two minima must be separated by a first-order saddle. In the case of a Langevin dynamics the jumping rate from minimum i to minimum j may be computed up to different degrees of approximation, ranging from the simple Arrhenius approximation to estimates involving high order derivatives of the potential both in the saddle and in minimum i. For multi-dimensional potentials however the best available approximation is still the second-order Langer's theory [23]. In this approximation and in the over-damped limit the jumping rate holds:

$$\Gamma_{i \to j} = \frac{\omega_{\exp}}{2\pi\gamma} \frac{\Omega_i}{\Omega_{\text{saddle}}} \mathrm{e}^{-(V_{\text{saddle}} - V_i)/k_{\mathrm{B}}T}. \tag{3}$$

Where Ω_i is the square root of the product of the strictly positive Hessian eigenvalues in the minimum i, Ω_{saddle} is the same quantity computed in the saddle, $\omega_{\exp}$ is the square root of the absolute value of the only negative Hessian eigenvalue in the saddle, γ is the friction coefficient of the dynamics and V_{saddle} and V_i are the potentials in the saddle and in the minimum i, respectively. It is worth noticing that i (which I will often refer to as the *entropic term*) is, in the over-damped limit, a second order estimate of the entropy in the minimum. Actually Langer's transition theory supplies an elegant way to bridge between the PEL and FEL descriptions by naturally providing to connections between minima of the potential energy with weights depending on their free energy difference. The ratio between $\Gamma_{i \to j}$ and

$\Gamma_{j\to i}$, which I call $\eta_{i\to j}$, measures the directionality of the jump. When $\eta_{i\to j}>1$ the system will preferentially jump from i to j, while behaving in the opposite way when $\eta_{i\to j}<1$

$$\eta_{i\to j} = \frac{\Omega_i}{\Omega_j} \mathrm{e}^{\left(V_i - V_j\right)/k_{\mathrm{B}}T}. \tag{4}$$

Notably η does not depend on the saddle details (energy and entropic term). The only significant information left in it is the existence of a saddle connecting i and j because η itself would have no meaning when referring to two unconnected minima. Note also that Eq. (4) is correct only when i and j are separated by just 1 saddle. Would this not be the case η will show a dependence from the energy end the entropic term of all the saddles separating the two basins of attraction.

A first statistical characterization of the minima in the database in terms of the quantities that influence jump directionality can point out some quantitative differences between the landscape of the good folder S1 and those of bad folders (i.e. the homopolymer S0 and the sequence S4). I first focus on the energy of the minima and on their angular distance from the NC which, for this 2-D model, can be considered a good reaction coordinate [10]. By denoting with M a generic minimum and with 0 the NC, this distance can be defined as:

$$\delta(M,0) = \sum_{i=2}^{L-1} \mid \theta_i^M - \theta_i^0 \mid \tag{5}$$

where $\{\theta_i^M\}$ and $\{\theta_i^0\}$ are the bond angles of the configuration of the chain corresponding to the minimum M and to the NC, respectively.

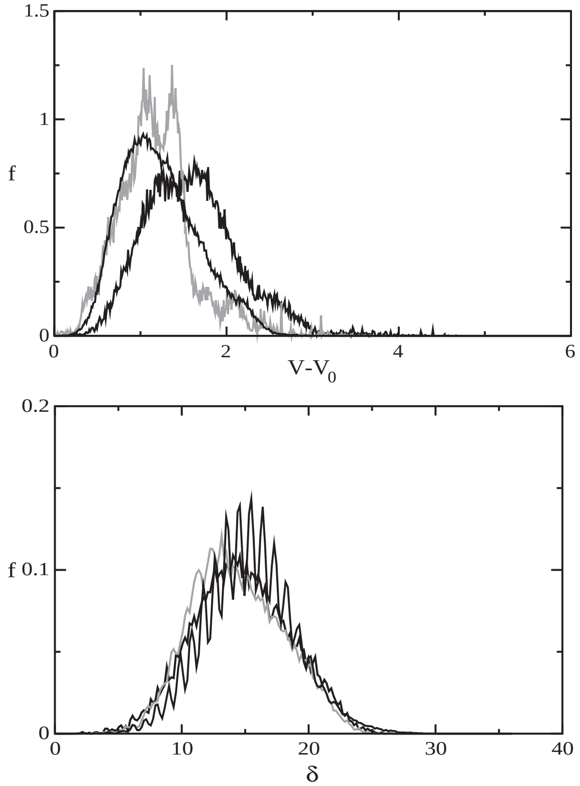

Fig. 1. Energy (top) and angular distance (bottom) distribution of the minima of the potential energy for the three considered sequences. The black thin line refers to the homopolymer S0, the black bold line to the good folder S1 and the grey line to the bad folder S4.

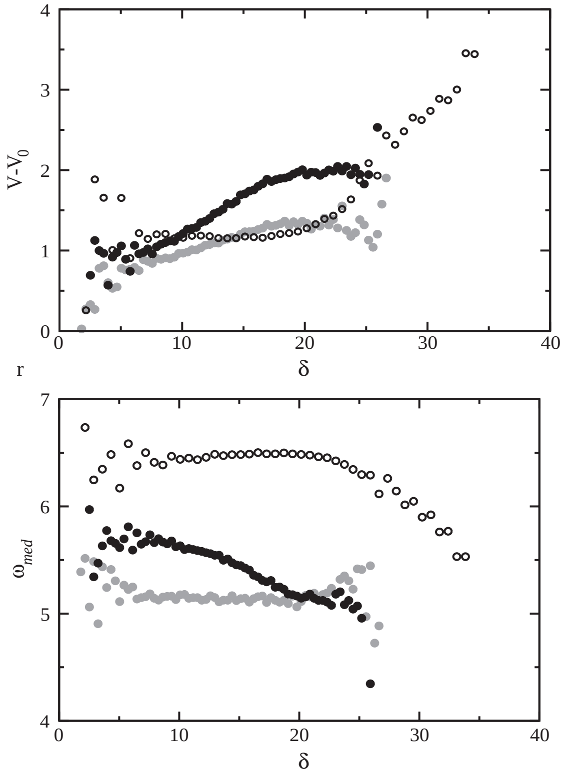

Fig. 2. Top: average energy of the minima of angular distance from the NC comprised between δ and δ+0.2. Bottom: average curvature (geometrical average of the nonzero eigenfrequencies of the potential Hessian) of the minima of angular distance from the NC comprised between δ and δ+0.2. Black empty circles refer to the homopolymer, black full circles to the good folder and grey circles to the bad folder.

Fig. 1 shows that the three analyzed sequences do not exhibit any significant difference in the shape of the distribution of the energy of the minima and of their angular distance from the NC.[1] Nonetheless the two quantities show to be significantly correlated as one can see in the top part of Fig. 2, where, after dividing the range of possible angular distances from the NC in bins of amplitude 0.2, I plot the

---

[1] Actually the distribution of the angular distance from the NC for the homopolymer shows a periodic pattern also detectable, for low distances, in the case of the bad folder. The visual analysis of the conformation of several minima shows that the bond angles tend to be distributed near to the values $\pi/3$, $2\pi/3$ or π thus indicating that, in spite of the harmonic and angular contributions, the lower and intermediate energy minima are still reminiscent of the regular structure of the minima of a Lennard–Jones cluster.

average energy of the minima within each bin versus the angular distance. The graphs reported do not depend on the bin size, thus showing that the statistics is sufficiently high. The steepness of the curves reported in the top part of Fig. 2 is a good measure of the average landscape steepness. According to Eq. (4), the transition between the basins of attraction of two different minima is more probable to occur from the higher minimum to the lower one, than in the opposite direction. As a consequence, a "steep" dependence of δ on the energy implies that the dynamics is strongly biased toward the NC. Fig. 2 highlights that the steepness of the landscape is higher for the good folder than for the bad folder: this confirms that the former exhibit a more pronounced trend to move towards the native configuration than the latter which is in qualitative agreement with the standard energy landscape picture. The case of the homopolymer is a bit richer: the landscape is very steep at high distances from the NC, but it is almost flat in a large neighborhood of the NC. Jumps between metastable minima will then be strongly biased toward the NC at high configurational distances from the NC but will show no preferential direction at intermediate and small distances from the NC thus letting the protein wander around in a large neighborhood of the NC.

The above analysis is based just on the energetic term of Eq. (4) and must be complemented by analogous considerations about the contribution of the entropic term. In the bottom part of Fig. 2 the average curvature of the minima is reported in a similar fashion to what was done for the average energy, i.e. by adopting the same binning procedure in δ. The curvature of every minimum has been estimated by the geometric mean of the nonzero eigenfrequencies ω_i of the potential Hessian estimated at the considered minimum (37 in this case, the number of degrees of freedom minus 1 for rotation and 2 for translations):

$$\omega_{\mathrm{med}} = \left(\prod_{i=1,37} \omega_i \right)^{\frac{1}{37}} = (\Omega_i)^{\frac{1}{37}} \tag{6}$$

Coherently with Eq. (4) negative the dependence of this number from δ implies a directionality of the jump dynamics toward the unfolding direction. As far as the homopolymer and the bad folder are concerned the scenario arising form the analysis of the landscape steepness is substantially confirmed. Actually the curvature of the minima of the bad folder is essentially flat thus not favoring jumps in any preferential direction. The same holds for the homopolymer in a large neighborhood of the NC, though at larger distances the curvature decreases contrasting the initial collapse encouraged by the landscape steepness. The situation gets more complicated when considering the good folder. In this case the curvature always decreases with the distance from the NC. The folding propensity of the good folder is then determined by an interplay between an entropic factor favoring unfolding and an energetic bias toward folding. According to Eq. (4) the balance of this two terms depends on temperature. Decreasing the temperature the importance of the energetic term increases while the entropic term remains unchanged. In the case of the good folder this means that at low temperatures it will tend to fold following the landscape steepness while at higher temperatures it will tend to unfold following the curvature gradient. The temperature at which this two effects balance in average on all the landscape is then a good candidate for the folding transition.

In order to determine the value of the folding transition temperature one may follow a thermodynamical approach. This will also test the reliability of the considered dataset and of the statistical analysis performed. Actually the qualitative insight gained so far on the folding dynamics depends on the statistical reliability of the minima database and on the approximate dependencies of the energy of the minima and their amplitude form their distance form the native configuration. In order to check the approximations introduced one can use them to compute the folding transition temperature of the system and compare it with the same quantity obtained by direct numerical simulation. From the energy of the minima and their amplitude one can compute a first-order approximation of the partition function of the system [17]. Thus, knowing the dependence of the number of minima and of their average energy and amplitude from the distance from the NC δ, one could further approximate the partition function introducing in it a dependence on δ:

$$Z(T,\delta) = C(T)\frac{N(\delta)\exp^{\frac{-E(\delta)}{K_b T}}}{\omega_{\mathrm{med}}(\delta)^{36}} \tag{7}$$

where $N(\delta)$ is the number of minima, $E(\delta)$ their average energy and $\omega_{\mathrm{med}}(\delta)$ their average curvature as a function of the angular distance from the native configuration and $C(T)$ is a temperature dependent normalization factor. In order to estimate $Z(T;\delta)$ for the three considered sequences I assume $N(\delta)$ to be a Gaussian of average 15 and standard deviation 8 and I approximate as linear in the interval $0<\delta<20$ both $E(\delta)$ and $\omega_{\mathrm{med}}(\delta)$. The approximate partition function corresponding to the good folder is plotted in Fig. 3 at three different temperatures. For low temperatures the partition function has a maximum in $\delta=0$ thus showing that the most probable conformations are the NC and it's nearest neighboring minima. The appearance of a different maximum of the partition function at higher temperatures clearly indicates the occurrence of a folding transition. In the inset of Fig. 3 the position of the maximum $\delta_{\max}$ of the partition function is shown as a function of temperature for the three considered sequences. The intercept of one of these lines with the T axis corresponds to the folding transition temperature T_{f} for the considered sequence. The actual numerical values of T_{f} estimated in this way are systematically lower than those computed numerically in a previous work ($T_{\mathrm{f}}=0.061$ for the good folder and 0.044 both for the

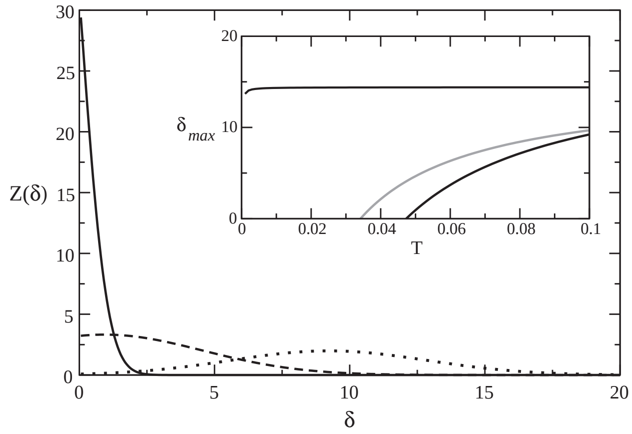

Fig. 3. Approximate partition function for the good folder as a function of δ at T=0.04 (continuous line), T=0.05 (dashed line) and T=0.1 (grey line). In the inbox I draw the maximum of the partition function as a function of temperature for the homopolymer (black thin line), the good folder (black bold line) and the bad folder (grey).

homopolymer and the bad folder), especially in the case of the homopolymer, but the order of magnitude and the ordering of the sequences is the correct one, suggesting that the procedure I described is able to capture the gross features of the folding process.

5. Saddles

A major concern in this analysis is the finding of a good indicator for the connectivity between two minima. In practice, I aim at identifying a proximity criterion, able to predict when two minima can be considered sufficiently close to each other to be connected by a single first-order saddle. This information would be very useful in order to speed up the algorithm for finding saddles. It is actually useless to look for a saddle between two minima, which are very far apart. Moreover, I am going to show that this analysis provides useful insights onto the mechanism of saddles jumping and onto the underlying landscape structure.

The distance between two minima may be defined in several ways. The angular distance, which I have previously introduced, is actually just one of the many possible ways to define a distance that measures the difference in the angular conformation of two chain configurations M and Q. A more general formulation is:

$$\delta_{\mathrm{A}}^{(n)}(M,Q) = \left(\sum_{i=2}^{L-1} \mid \theta_i^M - \theta_i^{Qn} \mid \right)^{\frac{1}{n}} \tag{8}$$

where θ_i^Q is the ith bond angle of configuration the configuration Q. In this notation, the previously introduced angular distance δ corresponds to $\delta_{\mathrm{A}}^{(1)}$, while I define $\delta_{\mathrm{A}}^{(\infty)}$=max($|\theta_1^M - \theta_1^Q|$, ..., $|\theta_{L-1}^M - \theta_{L-1}^Q|$).

The angular distances just defined convey an information about differences in the bond angles which are local quantities.

Now, I wish to investigate if the connectivity between different minima depends on more global differences in conformations. For this aim I introduce a second family of distances

$$\delta_{\mathrm{G}}^{(n)}(M,Q) = \left(\sum_{i<j+1}^{L} \mid r_{i,j}^M - r_{i,j}^Q \mid^n \right)^{\frac{1}{n}} \tag{9}$$

which depend on all the intra-bead distances $r_{i,j}$. In order to test which one among these configurational distances better captures the configurational proximity between minima, I have performed the statistical analysis hereafter described.

For each one of the considered metrics I compute the distribution of distances among all minima connected via a single first-order saddle. I determine the distance Δ, which includes 90% of all connected pairs. I compute also the statistical distribution of the distances between all possible pairs of minima and determine the percentage of generic (connected or unconnected) pairs characterized by a distance lower than Δ. The smaller is this quantity the better is the predictive capability of the metric. For each family of metrics the analysis has been performed only for $\delta^{(\infty)}$ and $\delta^{(1)}$, because any other member of the family satisfies the inequality $\delta^{(\infty)}(M, Q) \leq \delta^{(n)}(M, Q) \leq \delta^{(1)}(M, Q)$ for any couple of chain configurations M and Q.

The data are reported in Table 2 showing that for both metric families $\delta^{(1)}$ performs better than $\delta^{(\infty)}$ and that the angular distance performs significantly better than the global distance in the case of the heteropolymers, while the two behave similarly in the homopolymer case. I can conclude that the best indicator for distinguishing between connected and unconnected minima is $\delta_{\mathrm{A}}^{(1)}$, which is a distance that averages along the chain a local configurational difference. This means that, as far as the two heteropolymers are concerned, the attraction basins of two minima are likely to be connected by a first-order saddle if their configurations are locally similar along all the chain. Accordingly, the jumps between the basins of attraction of connected minima typically involves small local changes of the chain conformation, which are distributed along the entire chain. Eventually, a huge global change may appear as the combined effect of a long series of jumps.

Table 2
Performances of different distances in predicting the connection between minima

	S0	S4	S1
$\delta_{\mathrm{A}}^{(\infty)}$	0.3	10^{-2}	5×10^{-2}
$\delta_{\mathrm{A}}^{(1)}$	10^{-3}	4×10^{-4}	2.7×10^{-4}
$\delta_{\mathrm{G}}^{(\infty)}$	2×10^{-2}	1.8×10^{-2}	4×10^{-3}
$\delta_{\mathrm{G}}^{(1)}$	10^{-3}	3×10^{-3}	8×10^{-4}

The reported number is the percentage of all the minima in the database having distance lower than that characterizing the 90% of connected pairs.

6. Conclusions

In summary, I have shown that some simple features of the energy landscape of a model protein can be related to its folding capabilities. In particular, the statistical analysis of the minima indicates that the folding propensity of a sequence is determined by the interplay between the steepness of the energy funnel and the gradient of the curvature of the minima along the reaction coordinate.

Moreover, the identification of the first-order saddles of the potential yields a quantitative analysis of the connections between the basins of attraction of different minima. The use of suitable configurational distances indicate that the conformational changes of the chain typically result as the effect of a sequence of small local rearrangements, distributed all over the chain length.

Acknowledgements

I warmly acknowledge R. Livi, A. Politi and A. Torcini for their valuable help and advise. Moreover I wish to thank CINECA and INFM for providing me access to the CINECA Linux-cluster under the grant “Iniziativa Calcolo Parallelo”. This work has been partially supported by a grant of the Ente Cassa di Risparmio di Firenze, Italy.

Appendix A

In this appendix sketch the procedure for the identification of the metastable minima of the potential energy and of the first-order saddles connecting them.

The method for obtaining the “inherent” minima [2] has been explained in details for the present model in a previous paper [9]. It essentially consists in performing a Langevin dynamics at a sufficiently high temperature T (typically higher than the folding temperature). Accordingly, the system visits a wide portion of the accessible phase space. A steepest descent minimization is used for identifying the minima underlying the dynamics, at fixed sampling time of the trajectory. A refinement of the location of the obtained minima is achieved by employing a standard Newton algorithm. The database of minima is used for finding out the first-order saddles connecting directly couples of them. This goal is accomplished by the following algorithm:

(i) for every pair of minima A and B an intermediate configuration C is constructed with bond angles obtained as the average of the corresponding angles of A and B;

(ii) a steepest descent algorithm is applied to C until a minimum is reached;

(iii) (a) if the new minimum is A (B) I construct a new configuration C′ intermediate between C and A (B) and I go back to item (ii);

  (b) if the new minimum is neither A nor B, I assume that the basins of attraction of the two minima are not directly connected, if the new minimum is not included in the minima database I update the database with this new configuration, I go back to item (i) and I analyze a new couple of minima;

  (c) if the Euclidean distance between C and C′ is lower than a threshold δ, then C and C′ are on the ridge, (i.e. the stable manifold that divides the basins of attraction of A and B) and I can pass to item (iv);

(iv) I start two steepest descent relaxations from C and C′, while monitoring their relative distance when they move along the ridge; if their Euclidean distance overcomes the threshold δ I go back to item (i) by using updated C and C′ as new initial configurations;

(v) if during the steepest descent the gradient of the potential becomes lower than a given threshold ϵ, I assume to be on the saddle: the corresponding configuration is then refined by a Newton algorithm and the eigenvalues of the Hessian computed in order to verify the order of the fixed point found.

As already explained the goal of items (i), (ii) and (iii) is finding the ridge separating the basins of attraction of the couple of minima, while the aim of items (iv) and (v) is moving along the ridge in the direction of the first-order saddle, i.e. towards the point with lowest energy on the ridge. Though the algorithm is not particularly sensitive to the threshold ϵ (set to 10^{-7} during the database analysis), some care must be devoted to the choice of the threshold δ. Too much a small value for this parameter will force the algorithm to perform too many steepest descents (the lengthiest part of the computation) and prevent it to profit of the acceleration provided by the Newton refinement.

On the other hand too much a high δ will prevent convergence in the most pathological cases causing the algorithm to continuously jump between steps (iv) and (i). The choice δ=10^{-3} has shown sufficient to eliminate the problem.

References

[1] D.J. Wales, Energy Landscapes, Cambridge University Press, Cambridge, 2003.

[2] F.H. Stillinger, T.A. Weber, Packing structures and transitions in liquids and solids, Science 225 (1984) 983–986.

[3] S. Sastry, P.G. Debenedetti, F.H. Stillinger, Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid, Nature 393 (1998) 554–557.

[4] L. Angelani, R. Di Leonardo, G. Ruocco, A. Scala, F. Sciortino, Saddles in the energy landscape probed by supercooled liquids, Phys. Rev. Lett. 85 (2000) 5356–5359.

[5] S.V. Krivov, M. Karplus, Free energy disconnectivity graphs: application to peptide models, J. Chem. Phys. 117 (2002) 10894–10903.

[6] D.A. Evans, D.J. Wales, Free energy landscapes of model peptides and proteins, J. Chem. Phys. 118 (2003) 3891–3897.

[7] A. Baumketner, J.-E. Shea, Y. Hiwatari, Glass transition in an off-lattice protein model studied by molecular dynamics simulations, Phys. Rev., E 67 (2003) 011912–011918.

[8] H. Nymeyer, A.E. García, J.N. Onuchic, Folding funnels and frustration in off-lattice minimalist protein landscapes, Proc. Natl. Acad. Sci. 95 (1998) 5921–5928.

[9] L. Bongini, R. Livi, A. Politi, A. Torcini, Thermally activated processes in polymer dynamics, Phys. Rev., E 68 (2003) 061111–061127.

[10] A. Torcini, R. Livi, A. Politi, A dynamical approach to protein folding, J. Biol. Phys. 27 (2001) 181–203.

[11] J.D. Bryngelson, P.G. Wolynes, Intermediates and barrier crossing in a random model (with applications to protein folding), J. Phys. Chem. 93 (1989) 6902–6915.

[12] N.D. Socci, J.N. Onuchic, P.G. Wolynes, Diffusive dynamics of the reaction coordinate for protein folding funnels, J. Chem. Phys. 104 (1996) 5860–5868.

[13] H. Frauenfelder, F. Parak, R.D. Young, Conformational substates in proteins, Annu. Rev. Biophys. Biophys. Chem. 17 (1988) 451–479.

[14] R.S. Berry, N. Elmaci, J.P. Rose, B. Vekhter, Linking topography of its potential surface with the dynamics of folding of a protein model, Proc. Natl. Acad. Sci. 94 (1997) 9520–9524.

[15] A. Miller, J. Wales, Energy landscape of a model protein, J. Chem. Phys. 111 (1999) 6610–6616.

[16] F.H. Stillinger, T.H. Gordon, C.L. Hirshfeld, Toy model for protein folding, Phys. Rev., E 48 (1993) 1469–1477.

[17] E. Marinari, G. Parisi, Simulated tempering: a new Monte Carlo scheme, Europhys. Lett. 19 (1992) 451–458.

[18] A. Irbäck, F. Potthast, Studies of an off-lattice model for protein folding: sequence dependence and improved sampling at finite temperature, J. Chem. Phys. 103 (1995) 10298–10305.

[19] A. Irbäck, C. Peterson, F. Potthast, Identification of amino acid sequences with good folding properties in an off-lattice model, Phys. Rev., E 55 (1997) 860–867.

[20] P.G. De Gennes, Scaling Concepts in Polymer Physics, Cornell University Press, New York, 1979.

[21] R.S. Berry, H.L. Davis, T.L. Beck, Finding saddles on multidimensional potential surfaces, Chem. Phys. Lett. 147 (1988) 13–17.

[22] F.H. Stillinger, Exponential multiplicity of inherent structures, Phys. Rev., E 59 (1999) 48–51.

[23] P. Hänggi, P. Talkner, M. Borkovec, Reaction rate theory: fifty years after Kramers, Rev. Mod. Phys. 62 (1990) 251–342.