Eur Biophys J (2010) 39:911–919
DOI 10.1007/s00249-009-0441-1

ORIGINAL PAPER

Identification of the folding inhibitors of hen-egg lysozyme: gathering the right tools

Martina Caldarini · Ludovico Sutto · Carlo Camilloni · Francesca Vasile · Ricardo A. Broglia · Guido Tiana

Received: 15 January 2009 / Revised: 3 March 2009 / Accepted: 9 March 2009 / Published online: 27 March 2009

Abstract The unfolded state of proteins displays a surprisingly rich amount of local native structure, which appears to be critical for driving the protein to its native state. Peptides with the same sequence of the corresponding structured segments can be used to interfere with the correct folding of the protein. Using model simulations, we investigate the folding of hen-egg lysozyme, identifying its key segments. Activity assays, NMR and circular dichroism experiments are used to screen the peptides which are able to inhibit the folding of lysozyme. Few peptides, corresponding to the segments of the protein which are structured in the unfolded state, are identified to have significant inhibitory effects.

Keywords Folding inhibition

Introduction

Until few years ago, it was widely accepted that the denatured state of proteins is a random coil. This idea was based on measurements carried out by Tanford in the 1960s [1], who showed that the size of a series of proteins denatured by guanidinium chloride is the size of a random coil. But, as pointed out only recently [2], something which has the size of a random coil is not necessarily a random coil. Moreover, the denatured state in guanidinium chloride can be very different from the denatured state in biological conditions, which is the one controlling folding [3].

As a matter of fact, NMR experiments on several proteins point out the presence of weak but detectable structure in the denatured state [4–7], even under strongly denaturant conditions [8]; often such structure is native and involves the same residues which build out the folding nucleus (FN) [4, 5], that is the set of native contacts whose formation allows the protein to overcome the main free-energy barrier towards the native state. Simulations with simplified models confirm that a number of small, single-domain proteins fold by structuring in the denatured states in some segments, which dock together to form the FN [9]. We call these segments local elementary structures (LES).

These findings suggested the possibility of inhibiting the biological activity of proteins by blocking their folding, using for the purpose peptides (called p-LES) with the same sequence of a LES of the target protein [10]. As a p-LES has the same molecular structure as a LES, its specificity with respect to the complementary LES, having been tooled by evolution, is high and its affinity strong. In any case, as strong as the affinity which stabilizes the FN. Adding to the protein solution a concentration of p-LES large enough, the entropic cost of binding is small and one stabilizes a state which is partially denatured (and thus inactive) with the p-LES bound in the interior. This mechanism has been proven to be particularly efficient to inhibit the folding of HIV-1 protease [11]. In this case, as in all the cases of viral proteins, p-LES inhibitors have a

Proceedings of the XIX congress of the Italian society of pure and applied biophysics (SIBPA), Rome, September 2008.

M. Caldarini · L. Sutto · C. Camilloni · R. A. Broglia · G. Tiana (✉)
Dipartimento di Fisica, Universitá di Milano,
via Celoria 16, 20133 Milan, Italy
e-mail: tiana@mi.infn.it

F. Vasile
CISI, Universitá di Milano, via Fantoli 16/15, 20138 Milan, Italy

R. A. Broglia
Niels Bohr Institute, University of Copenhagen,
Bledgamsvej 17, 2100 Copenhagen, Denmark

great advantage with respect to traditional drugs. In fact, as they target a segment of the protein which is critical for the folding of the enzyme, it is unlikely that the virus can mutate it in order to decrease its affinity for the inhibitor, producing pharmacological resistance.

The main problem in the application of the p-LES strategy is to locate the LES of the protein one wishes to inhibit. This can be done in different ways, for example studying the early events of protein folding by model simulations [13, 14]; studying the amino acid conservation pattern and making the assumption that, because the LES are so important for the folding of the protein, they should be highly conserved [14]; testing the effect on folding/activity of a set of peptides whose sequence overlap with the full-length sequence of the protein [15, 16].

In the following sections we shall discuss the inhibition of the folding of hen-egg lysozyme. Although this protein is of no pharmaceutical interest, its folding has been widely characterized [12, 17], it is easy to obtain and handle and, consequently, is a good model to study and describe the p-LES strategy.

Lysozyme is composed of 129 residues, its native conformation displays five helices and a three-stranded β-sheet, and is stabilized by four disulphide bridges. Its native conformation is remarkably stable: at neutral pH, its folding temperature is around 70^{o}C [18] and it does not unfold in 8M urea unless upon consistent addition of salts [19]. At neutral pH only two states are populated at equilibrium, while an intermediate appears at acidic conditions [20]. Its folding kinetics is anyway complex: it displays at least three time scales and thus suggests that transient intermediates exist also at neutral pH, even if not populated at equilibrium [17]. Its role is that of catalyzing the hydrolysis of a polysaccharide which builds out the cellular wall of several bacteria. Also the enzymatic reaction has been widely studied [21] and involves Glu-35 and Asp-52.

First, we will study the folding of lysozyme with a simplified model, identifying residual structure in the unfolded state both in native- and strongly denaturant conditions. Next, different experimental techniques are used to screen the inhibition properties of a set of peptides whose sequence overlaps exhaustively the sequence of lysozyme. Few of these peptides, corresponding to regions which are structured in the unfolded state, display significant inhibitory effects, preventing lysozyme from folding.

Materials and methods

Model simulations

Lysozyme (PDB code: 2LYZ) is simulated using a coarse grained geometry representing both the backbone and the side-chains. In particular, the backbone is represented by a connected chain of C_α atoms separated by 3.8 Å. The ith side-chain is geometrically represented by a second bead, $C_\beta(i)$, linked at fixed distance $l = 3.0$ Å to the $C_\alpha(i)$ atom and whose position is uniquely determined by the positions of $C_\alpha(i-1)$, $C_\alpha(i)$, $_\alpha(i+1)$ as described in the work of Park and Levitt [22]. It follows that for the first and last atoms of the chain, the C_β position is undetermined and therefore replaced with the C_α atom itself.

The C_α atoms interact with any other atom through a generic excluded volume term while the interactions between C_β atoms are given by a heterogeneous Gō-like potential $U_{\beta\beta}$ written as follows

$$U_{\beta\beta}\{r_{ij}\} = \sum_{i+2<j}\Big[U_{ij}\cdot\theta(R_C - r_{ij}^N) + \epsilon_R\cdot\theta(r_{ij}^N - R_C)\Big] \cdot\theta(R_C - r_{ij}) + \epsilon_H\cdot\theta(R_H - r_{ij}), \quad (1)$$

where $\{r_{ij}\}$ are the positions of the C_β atoms in a given conformation and $\{r_{ij}^N\}$ are the C_β positions in the native structure. $R_C = 5.5$ Å and $R_H = 1.0$ Å are the interaction and hardcore range, respectively, and $\theta(x) = 1$ if $x > 0$ and 0 otherwise. The energies constants $\varepsilon_H = 1{,}000$ and $\varepsilon_R = 5.0$ are respectively the hardcore barrier and the repulsive energy for non-native pairs, while U_{ij} is the attractive energy between native pairs. These matrix elements are the average Van der Waals interaction between all the atoms of residues i and j averaged over a 10 ns all atom explicit-solvent simulation of the native conformation at $T = 300$ K using the GROMOS force-field and the GROMACS package. The four disulfide bonds between residues 6–127, 30–115, 64–80 and 76–94 are implemented setting $U_{ij} = -1{,}000$ resulting in unbreakable bonds at the simulated temperatures. All the energies and temperatures are expressed in kJ/mol unless otherwise specified.

The specific heat shown in the inset of Fig. 1 is calculated by applying the multiple histogram method to 10 Monte Carlo simulation runs of 5×10^9 steps at a constant temperature ranging from $T = 1.5$ to $T = 4.0$. From its peak the folding temperature is found to be $T_f = 1.97$. The protein is then simulated with 100 Monte Carlo runs of 1×10^9 steps each at two different temperatures $T = 1.5 < T_f$ and $T = 2.5 > T_f$ leading to 10^6 conformations for analysis. All the starting conformations were randomly generated keeping the four disulfide bonds formed.

We used the C_α dRMSD$= \left(\frac{1}{\mathcal{N}}\sum_{i<j}(r_{ij} - r_{ij}^N)^2\right)^{1/2}$ as order parameter to measure the structural distance between a conformation $\{r_{ij}\}$ and the native one $\{r_{ij}^N\}$, where $\mathcal{N}$ is the normalization constant. Observing that the unfolded state is separated from the native one by a free-energy barrier located at dRMSD ≈ 6 Å (see Fig. 1), we

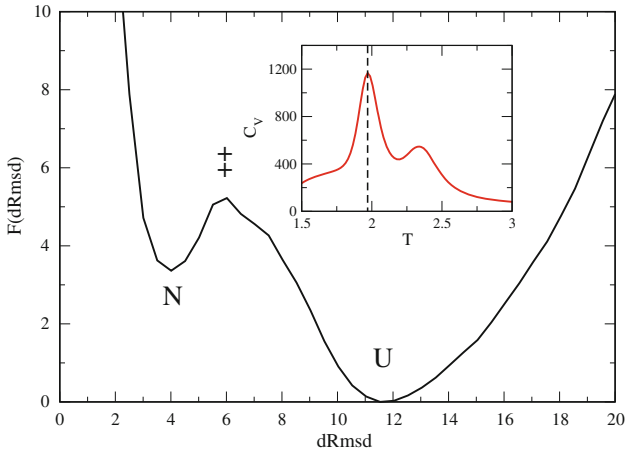

Fig. 1 Lysozyme free energy as a function of dRMSD obtained from Monte Carlo simulations at $T = 2.0$ ($\approx T_f$). The native (N) and unfolded (U) states are separated by a free energy barrier whose maximum defines the transition (‡) state. The specific heat as a function of temperature is shown in the inset. The peak identifies the folding temperature $T_f = 1.97$ (in the energy units of the system)

considered all the conformations with dRMSD $\geq$ 10 Å as belonging to the unfolded state.

The pair stability shown above the diagonal in the unfolded contact maps of Fig. 2 is given by the probability p_{ij} atom $C_\beta(i)$,$C_\beta(j)$ are within the interaction range R_C. This is given by the number of times the contact is formed over the number of conformations considered. The native contact map Δ_{ij} is represented below the diagonal in both the contact maps. The average stability of a fragment $P_{[n;m]}$ starting from residue n and ending with residue m is given by

$$P_{[n;m]} = \sum_{n \leq i,j \leq m} p_{ij} \cdot \Delta_{ij} \Big/ \sum_{n \leq i,j \leq m} \Delta_{ij} \qquad (2)$$

Analogously, the average mutual stability between two fragments $[n;m]$ and $[n';m']$ is given by

$$P_{[n;m]-[n';m']} = \sum_{\substack{n \leq i \leq m \\ n' \leq j \leq m'}} p_{ij} \cdot \Delta_{ij} \Big/ \sum_{\substack{n \leq i \leq m \\ n' \leq j \leq m'}} \Delta_{ij} \qquad (3)$$

We will call "good candidates" those fragments characterized by $P_{[n;m]-[n';m']} < 0.0001$. This enforces the presence of at least one native contact between the two fragments (i.e. non null denominator) and a low stability of this same contact(s) (low numerator). Using these two parameters, we say that a LES must display a high $P_{[n;m]}$ in the unfolded state while keeping $P_{[n;m]-[n';m']}$ low, reflecting both local stability and poor non-local interactions in the unfolded state.

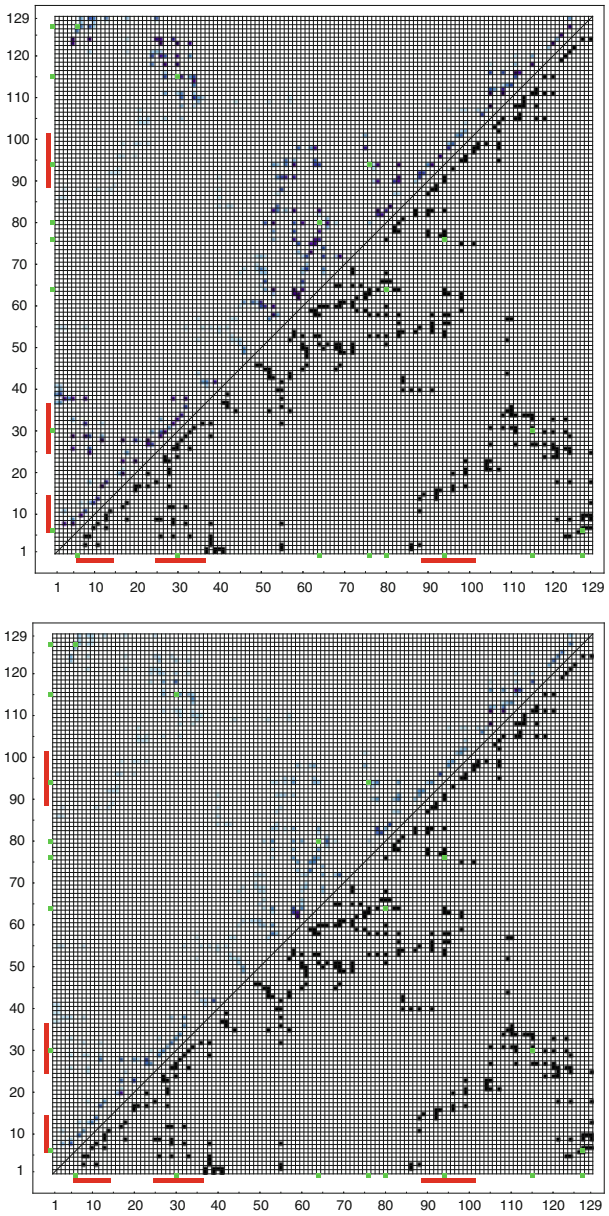

Fig. 2 Contact map for lysozyme at $T = 1.5$ (*upper panel*) and $T = 2.5$ (*lower panel*). The tone of the *squares* in the higher half of each map reflects the stabilities p_{ij} of the contacts (see "Materials and methods") while the native contact map is shown below the diagonal. The *green squares* identify the cysteines and the disulfide bonds, the *red bars* the three largest helices

Experimental characterization

Hen egg-white lysozyme (thrice crystallized, dialyzed and lyophilized), dried *Micrococcus lysodeikticus* cells, DMSO (at least 99.5% pure) were purchased from Sigma. 25 peptides from the sequence of the hen-egg lysozyme (of length of 10 residues overlapped by 5 amminoacids, corresponding to the sequence 1–10, 6–15, 11–20, etc.) were produced by solid phase synthesis with acetyl and amide as

terminal protection groups and were estimated to be >70% pure by analytical HPLC after purification.

Enzymatic activity of lysozyme was determined by measuring the clearing of turbidity of *Micrococcus lysodeicticus* suspension (substrate solution) at fixed wavelength. The suspension of *Micrococcus lysodeicticus* was prepared by mixing *Micrococcus lysodeicticus* (5 mg) with a 66 mM potassium phosphate buffer at pH 7.0 (50 ml). The mixture was suspended by gentle swirling and incubated at 25°C for 1 h. The samples of lysozyme and of different complexes with peptides (obtained as described below) were mixed with 1 ml of substrate suspension. The intensity decrease in light scattering was monitored by a Jasco UV–vis spectrophotometer (Jasco V-550) at 450 nm for 180 s. The value of the initial velocity of the reaction was determined by a linear extrapolation of the data.

In order to measure the enzymatic activity of lysozyme incubated with 25 different peptides in DMSO 80%, we use the following procedure. A 4.2 M lysozyme solution was prepared in a 66 mM potassium phosphate buffer at pH 7.0. Its concentration was determined spectrophotometrically at 280 nm, using an extinction coefficients of 2.63 mg ml^{-1} cm^{-1}. 4 mg of each peptide was dissolved in 1 ml of DMSO and 20 µl of this solution was diluited with 980 µl of DMSO. The concentration of the obtained peptide solution was 80 µM. In order to incubate the protein with peptide, 25 µl of lysozyme solution was mixed with 100 µl of peptide solution. The lysozyme reference solution in absence of peptide was prepared in the same way mixing 25 µl of lysozyme solution with 100 µl of DMSO. The samples were incubated at 25°C for 45 min. In the incubated samples the enzyme and the peptide had a nominal concentration of 0.9 and 64 µM, respectively. In the case of peptides 6–15, 71–80, 91–100, 106–115 and 121–129, the same experiments were performed at different concentrations (2, 10, 20 and 40 µM).

Activity assays for the determination of inhibition constant within the Michaelis–Menten theory were performed on a range of substrate concentrations from 0.3 to 0.02 mg/ml. 1 ml of *Micrococcus lysodeicticus* suspension was mixed with 12 µl of the incubated solutions prepared as above with peptide 91–100 and peptide 106–115 and the value of the initial velocity of the reaction was determined.

In the case of incubation of lysozyme with peptides 6–15, 91–100, 106–115 and 16–25 in the presence of urea, we use the following procedure. 2 µl of the solution of peptide dissolved in DMSO were diluted with 98 µl of buffer (pH 7.00) with urea 9 M. The concentration of the obtained peptide solution was 80 µM. 25 µl of lysozyme solution 4.2 µM was mixed with 100 µl of peptide solution. The sample was then heated at 85°C for 90 s. After 45 min at 25°C, the activity was measured. The lysozyme reference solution in the absence of the peptide was prepared in the same way mixing 25 µl of lysozyme solution with 98 µl of buffer with urea 9 M and 2 µl of DMSO. In the incubated samples, the enzyme and the peptide had a nominal concentration of 0.9 µM and 64 µM, respectively and the urea was 7.1 M.

For nuclear magnetic resonance spectroscopy, the lysozyme solution is prepared by mixing 4.6 mg of protein in 500 µl of phosphate buffer at pH 6 together with 9 M deutered urea. The solution of lysozyme and peptide 91–100 is prepared by mixing 4.6 mg of protein and 1 mg of peptide in 500 µl of phosphate buffer together with 9 M deutered urea. The samples are incubated at 85°C for 90 s in order to unfold it, and at 25°C for 45 min. Homonuclear 1 H NMR experiments were performed at 600 MHz on a Bruker Avance spectrometer. The probe temperature was maintained at 298 K and water suppression was carried out using the presaturation or the excitation sculpting with gradients sequence. In TOCSY experiments, a mixing time of 80 ms was applied to obtain remote scalar connectivities. NOESY spectra were recorded with mixing time of 300 ms. The spectral width was about 16 ppm.

Results

Lysozyme is, under biological conditions, a two-state folder, displaying a phase-transition between a native (N) and an unfolded (U) state [12]. While the native state has been widely characterized in the literature, little is known about the unfolded state due to the difficulty to study it both experimentally and resorting to simulations.

It is important to notice that in all the simulations and the experiments described below, the four disulphide bridges which stabilize the native conformation are kept intact.

The basis of folding inhibition: model simulations of the unfolded state

All atom models in explicit solvent can appropriately describe the interactions within proteins, but are computationally too demanding to sample the unfolded state. For this purpose we make use of a coarse-grained protein model (see "Materials and methods"), which is able to reproduce some experimental data [13] and can sample exhaustively the conformational space of small proteins.

The free energy of lysozyme as a function of the dRMSD to its crystallographic native state is shown in Fig. 1 at a temperature slightly higher than its folding temperature T_f, showing the expected two-state behavior. The transition state, corresponding to the top of the free-energy barrier which separates the two states, is located at dRMSD $\approx$ 6 Å. The folding temperature at which the

specific heat shows a peak is $T_f = 1.97$ (cf. the inset of Fig. 1).

The unfolded state is studied extracting from the simulations a representative set of conformations with high dRMSD (see "Materials and methods"). The native contact maps associated with the unfolded states obtained at $T = 1.5$, to describe the unfolded state under native conditions, and $T = 2.5$, to describe the unfolded state under strongly denaturing conditions, are displayed in the higher halves of Fig. 2 (upper and lower panel, respectively). The tone of each square reflects the statistical weight of the associated contact.

Looking at the high-temperature contact map, the most stable contacts are mostly local (i.e., close to the diagonal) and involve regions of the first two helices (residues 6–36) and the regions 58–65, 78–84 and 105–116. The only non-local contacts are concentrated in regions close to the four disulfide bonds between the eight cysteine residues. On the other hand, the low-temperature contact map, aside from an expected overall higher stability, displays several non-local stable contacts, like those between region 1–12 and region 25–41, that is the first alpha helix with the second one.

We are interested in locating the LES of the protein. Since they are expected to drive the protein from the unfolded to the native state, crossing the transition state, we look for them among the segments of the protein which build in the unfolded state stable local contacts (resulting in a high $P_{[n;m]}$), but do not stabilize non-local contacts (low $P_{[n;m]-[n';m']}$). In fact, if two segments of the protein build non-local contacts already in the unfolded state, their docking cannot be associated with the crossing of the transition state and hence cannot qualify as LES.

Operatively, we calculate the average stability of all the possible segments of lysozyme of length between 8 and 15, allowing for a maximum overlap of 4 residues. Within this set, we keep only the "good candidates", those segments which interact with at least another segment in the native state, but whose mutual interaction in the unfolded state is small (for details, see "Materials and methods"). The resulting data is shown in Fig. 3 for both cases of high and low temperatures. At high temperature there are three main regions where such segments are concentrated, namely 1–25, 42–69 and 73–105 (named I, II and III in Fig. 3). The same regions are also identified in the low-temperature data, together with a fourth region corresponding to residues 29–41. The segment 105–116, although resulting stable in the unfolded state, is already docked to its native counterpart and thus does not appear in Fig. 3.

From the contact map and the native tridimensional conformation we observe that both regions 1–25 and 42–69 are in contact with region 73–105. Consequently, we expect them to contain the LES of the protein and their docking to nucleate its folding.

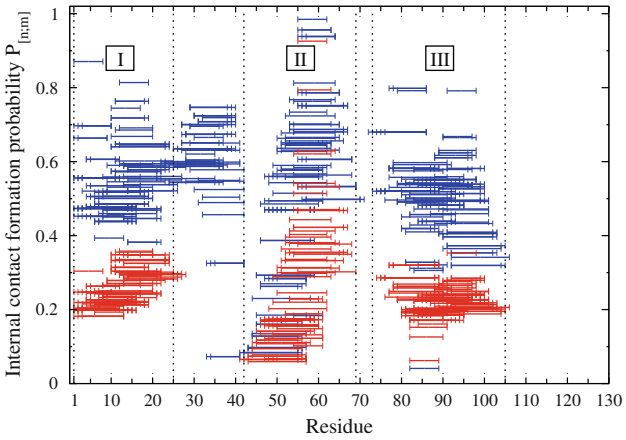

Fig. 3 The stability $P_{[n;m]}$ in the unfolded state of the "good candidates" segments at $T_L = 1.5$ (*blue*) and $T_H = 2.5$ (*red*). To qualify as "good candidate" a segment $[n;m]$ must have at least one native contact with another segment and the set of native contacts between them must be poorly stable in the unfolded state (see "Materials and methods"). The three regions I: [1–25], II: [42–69] and III: [73–105] are separated by *dotted lines*

Folding inhibition: experimental results

It was shown in [10, 11, 15] that peptides (p–LES) with the same sequence than that of a LES stabilize the unfolded state. This effect is due to the fact that the p–LES binds to its complementary LES in the protein, preventing the formation of the correct FN and thus the crossing of the transition state. The result is an inactive protein.

To identify experimentally the p–LES of lysozyme we have measured the activity of hen egg lysozyme in the presence of 25 peptides. Each of them is built out of ten amino acids (except for the one corresponding to the C-terminal, which is of length 9), and with sequence identical to that of the segment of the enzyme delimited by amino acids 1–10, 6–15,..., 111–120, 116–125 and 121–129. The initial velocity of enzymatic reaction of lysozyme by itself (v_0) and of lysozyme incubated with each of the 25 peptides (v_p) has been measured and the corresponding ratios v_0/v_p are displayed in Fig. 4. Peptides 6–15, 91–100 and 106–115 show a $v_0/v_p < 0.5$ and thus are good inhibitors of folding. Looking at the structure of the protein they correspond to the first helix A, to the last three turn of helix C and to the helix D.

Unlike HIV-1 protease, which is marginally stable and allows the corresponding p–LES to penetrate its native conformation under biological conditions [11], lysozyme is remarkably stable. Consequently, the incubation was carried out placing the protein and the peptide in 80% DMSO, to ensure enzyme denaturation, and then return to folding

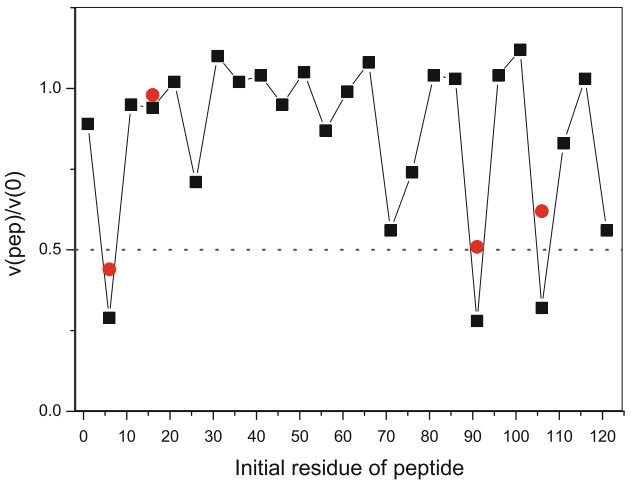

Fig. 4 The ratio v_0/v_{pep} of the initial velocity v_{pep} of the enzymatic reaction of lysozyme in presence of each of the 25 peptides and the initial velocity v_0 of the enzyme by itself. Each peptide contains ten aminoacids, except for one of length 9, with sequence identical to that of the segment of the enzyme delimited by amino acids 1–10, 6–15,...,111–120, 116–125 and 121–129. In this way they exhaustively overlap the protein sequence. On the *x*-axis the initial residue along the lysozyme sequence corresponding to the first amino acid of the 10-mer peptides is reported. *Solid black squares* refer to incubation in 80% DMSO and peptidic concentration of 64 μM (see "Materials and methods"). The four *red squares* corresponding to peptides 6–15, 16–25 91–100 and 106–115 are results of enzymatic activity measurements after incubation of 64 μM of peptide in 7.1 μM urea, and denaturation through temperature

conditions (8% DMSO). During this process the peptide can penetrate the unfolded enzyme and compete with the corresponding segment of the protein in forming its native contacts.

The effect of the most efficient inhibitory peptides 91–100 and 106–115 on the protein has been characterized in detail. The enzymatic kinetics of the protein is formally consistent with a noncompetitive inhibitions [15], due to the presence of large kinetic barriers that the peptide meets for binding the protein under native conditions. During the incubation of the peptide with lysozyme in denaturating conditions, the kinetic barriers are small and there is an equilibrium between the two molecules. As the system is brought back to native conditions, the kinetic barriers return to be large and partitioning takes place between the native enzyme and the inactive enzyme complexed with peptide. From standard analysis we obtained the two constants of inhibition $K_I = 19$ μM for peptide 91–100 and $K_I = 60$ μM for peptide 106–115 [15]. These can be compared with the values of K_I obtained from the ratios v_0/v_{pep} through a two-state model, which gave [15] $K_I = 21$ μM for peptide 91–100, $K_I = 57$ μM for 106–115. From the velocity ratios displayed in Fig. 4 one can also obtain $K_I = 43$ μM (peptide 6–15), $K_I = 83$ μM (peptide 71–80) and $K_I = 59$ μM (peptide 121–125).

If the mechanism we propose is correct, the inhibitory efficiency of peptides is not expected to be dramatically dependent on the denaturing agent used during incubation.The activity of lysozyme incubated at 85°C for 90 s in solution with peptides 6–15, 91–100, 106–115 and 16–25 (as negative control) has been measured in 9 M urea, to avoid aggregation (lysozyme in 9 M urea at neutral pH is folded and active [19], cf. Fig. 5). Their velocity ratios v_0/v_{pep} are 0.44, 0.51, 0.62, respectively, for the three inhibitory peptides (see Fig. 4), confirming their ability to inactivate lysozyme also upon heat denaturation, while it is 0.98 for the control peptide 16–25.

To prove that inhibition is due to unfolding of lysozyme, NMR spectra have been recorded for the complex of lysozyme with the peptide 91–100. Incubation was carried out in the presence of 9 M urea solution as in the activity assays in order to prevent aggregation and incubation has

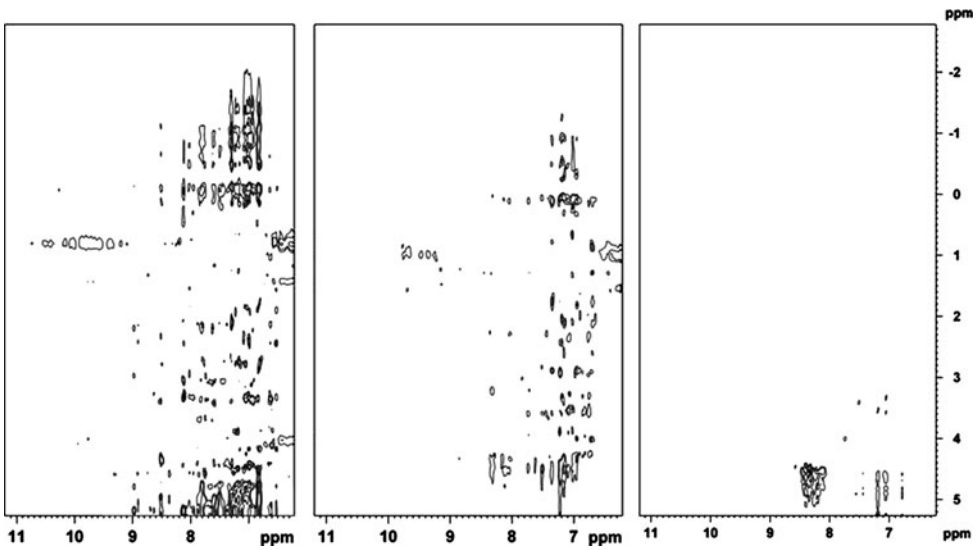

Fig. 5 The NOESY spectra of lysozyme in water, in 9 M urea and in solution with peptide 91–100, again in 9 M urea

been carried out at 85°C. In Fig. 5, the 2D NOESY spectra of lysozyme in water is shown (left panel), in the presence of 9 M urea (central panel) and incubated with peptide 91–100 (right panel). The shapes of the first two spectra are similar and display a good dispersion of crosspeaks between amide and sidechains signals, consistently with the native character of the protein. In the spectrum of lysozyme incubated with peptide 91–100 the peaks collapse towards the chemical shifts associated with random conformations, testifying to the fact that the protein is no longer in its native conformation. In particular, all the assigned peaks are lost and no more recognizable.

In order to quantify the ability of peptide 91–100 to inhibit the folding of lysozyme, circular dichroism experiments have been performed at different peptide concentrations (cf. ref. [15]). The resulting inhibition constant extracted from the values of ellipticity at a fixed wavelength ($\lambda = 233$ nm) through a two-state model (see ref. [15]) is $K_I = 17$ μM and it is not inconsistent with those obtained from activity assays (i.e., $K_I = 21$ μM and $K_I = 19$ μM, see above).

Can we predict the inhibitors from easily accessible data?

So far, we have identified the LES (and thus the p–LES) of lysozyme through extensive folding simulations and through the direct measurement of the effect of peptides on the enzyme. It is interesting to check if there are some features of the protein which can be obtained more readily and which allow to locate the LES or, at least, provide some information about them.

A first consideration which can be useful is that the segments of the protein which are critical for its folding are likely to be evolutionary conserved. Given a family of proteins displaying the same native conformation, a measure of the degree of conservation of each residue is the entropy per site $S(i) = -\sum p_i(\sigma) ln(p_i(\sigma))$, where $p_i(\sigma)$ is the frequency of appearance of residue of type σ at site i. Regions where sequence variation is high-yield low conservation and consequently have a high value of entropy while regions where the sequence variation is low are expected to display a low value of entropy.

Making use of a family of proteins structurally similar to lysozyme but with sequence identity lower than 25% we calculate the entropy per residue as shown in Fig. 6. In particular residues 8, 12, 16, 30, 42, 54, 57, 84, 86, 87, 88, 89, 95, 96, 97 display a value of entropy lower than the mean value minus standard deviations ($<S> = 1.80$, $\sigma = 0.28$). The most conserved regions are then 1–20, 52–64 and 76–99. These segments are subsets of the three LES identified by the folding simulations (1–25, 42–69 and 73–105) and two of them overlap with the p–LES found experimentally (i.e., 6–15 and 91–100).

A second element playing a fundamental role during the folding process is the interaction energy between amino acids. In order to become stable at early stages of folding, the segment of the protein must carry a significant fraction of the stabilization energy of the protein. Following the method described in [25], we have calculated this quantity making use of short all atom molecular dynamics simulations in explicit water using GROMACS, extracting the matrix U_{ij} of the average interaction energy between the pairs of amino acids. If the lowest eigenvalue displays a large gap to the others, then the protein is stabilized cooperatively by the residues identified by the eigenvector corresponding to the lowest eigenvalue. The lowest eigenvalue of lysozyme ($\lambda_1 = -80.4$ kJ/mol) displays a large energy gap, that is 11.3 kJ/mol (corresponding to $4k_BT$), with respect to the next one. The associated eigenvector is displayed in Fig. 7. Two regions display large

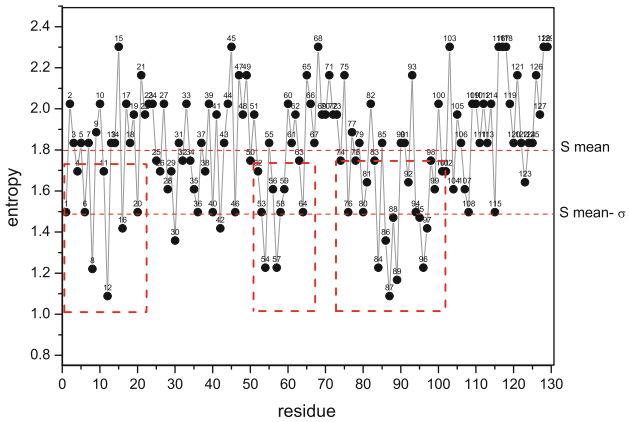

Fig. 6 The entropy per site (see text) of lysozyme. It is calculated over those proteins displaying sequence similarity <25% with lysozyme. Also given are the average value and the associated standard deviation $<S> = 1.80$ and $\sigma = 0.28$, respectively

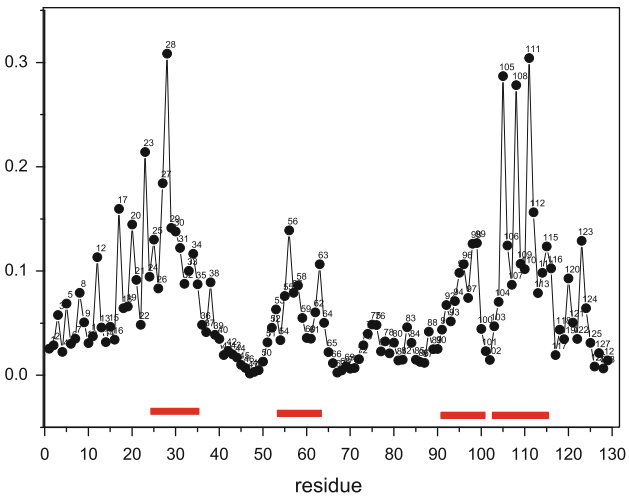

Fig. 7 The normalized components of the eigenvector $|c_i|$ associated with the lowest eigenvalue of the 129 × 129 interaction matrix U_{ij}

amplitudes (23–35 and 105–116) and other two regions with a lower but significant amplitude (50–65 and 90–101). Of these regions, 90–101 matches both the simulated LES and the p–LES determined experimentally, 50–65 only the former and 105–116 only the latter.

To be noted that a more naive analysis of the distribution of stabilization energies across the protein is not easy to interpret. In fact, from the average energy per site $E_i = \sum_{j\neq i}^{129} U_{ij}$ in the native conformation shown in Fig. 8 one cannot clearly identify any LES of the protein.

The same is true for the distribution of hydrophobic residues along the chain. In Fig. 9, the degree of hydrophobicity according to the scale of Hopp and Woods [26] is shown. Also in this case no clear pattern can be found.

From this analysis, we conclude that evolutionary data of conserved residues in analogous proteins together with the decomposition of the interaction matrix in its principal eigenvector can help spotting the LES. On the other hand, in more trivial analysis such as the average energy per site or the residue hydophobicity, does not give any sensible information. Anyhow, those methods must be complemented with simulations or experimental assays to have a clearer picture of the early structuring fragments and are not sufficient by themselves to identify the LES.

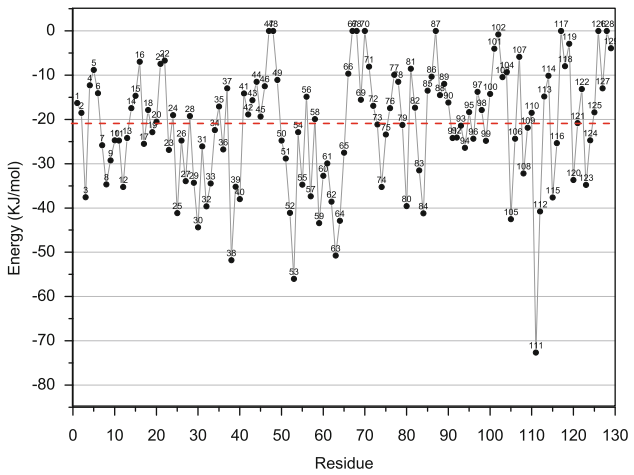

Fig. 8 The total interaction energy $E_i = \sum_{j\neq i}^{129} U_{ij}$ of the amino acid which occupies site i associated with all the other 128 amino acids j of lysozyme in its native conformation. The line in red is the associated average $< E > = -21.63$ kJ/mol, while the standard deviation is $\sigma = 13.40$ kJ/mol

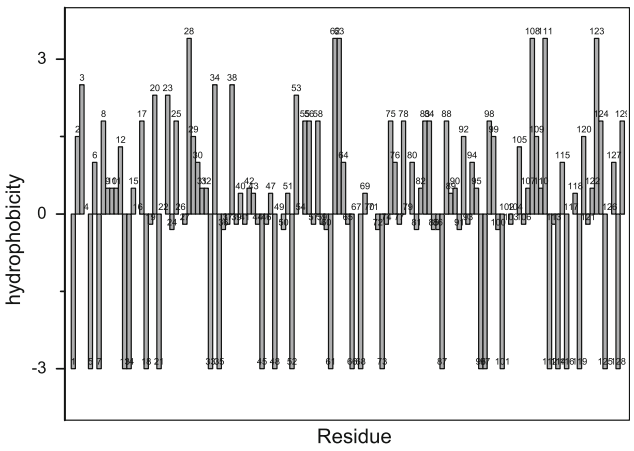

Fig. 9 The hydrophobicity of the residues of lysozyme

Discussion and conclusions

The characterization of the unfolded state is of great interest to understand protein folding, because not only the folding path is determined there, but it is also very difficult to carry out. Standard all-atom simulations in explicit solvents are computationally very demanding. Experimentally, the unfolded state can be stabilized and studied by means of chemical denaturants; but, aside from the tendency of the protein to aggregate, the denaturant alters the interactions within the protein, and consequently one studies a different system.

Simplified models can sample efficiently the conformational space of small proteins and show that some segments of lysozyme (1–25, 42–69 and 73–105) build native contacts within each segment in the unfolded state, but dock together only when they cross the transition state, qualifying as LES. In fact, NMR experiments of reduced lysozyme in water, in which case the disulphide bonds which stabilize the native state are broken, indicate that regions 19–32, 58–64, 106–116 and 119–124 are partially structured in the unfolded state [23]. Circular dichroism measurements of four peptides spanning the sequence of lysozyme at pH 2 (in water) show that, while fragments 1–40, 41–60 and 61–82 appear to be unstructured, fragment 84–129 has considerable helical structure [24].

The importance of the experiments done with lysozyme interacting with peptides displaying the same sequence as segments of the protein itself is twofold. First, they are a tool to investigate the unfolded state of a protein under biological (i.e., non–denaturing) conditions. By means of these peptides one can probe the assembly of the different parts of the protein, locating the LES which are critical to determine the folding pathway. Moreover, one can use the p-LES, that is the peptides with the same sequence as a LES, as an inhibitor or as a lead to design a peptidomimetic. This has been done with success, for example, in the case of HIV-1 protease.

An interesting question is whether one can use any easily available data to predict the LES of a protein without the need of carrying out lengthy simulations or exhaustive peptide screenings. Our results indicate that evolutionary data and a proper analysis of how the stabilization energy

of the protein is distributed among the amino acids give valuable information to locate the LES.

References

1. Tanford C, Kawahara K, Lapanje S (1966) Proteins in 6 m guanidine hydrochloride. J Biol Chem 241:1921–1923
2. Fitzkee N, Rose G (2004) Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci USA 101:12497–12502
3. Mok YK, Kay CM, Kay LE, Forman-Kay J (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. J Mol Biol 289:619–638
4. Dill KA, Shortle D (1991) Denatured states of proteins. Annu Rev Biochem 60:795–825
5. Yi Q, Scalley-Kim ML, Alm EJ, Baker D (2000) NMR characterization of residual structure in the denatured state of protein L. J Mol Biol 299:1341–1351
6. Fieber W, Kristjansdottir S, Poulsen FM (2004) Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. J Mol Biol 339:1191–1199
7. Blanco FJ, Serrano L, Forman-Kay J (1998) High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. J Mol Biol 284:1153–1164
8. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293:487–489
9. Broglia RA, Tiana G, Provasi D (2004) Simple models of protein folding and of non-conventional drug design. J Phys Cond Mat 16:R111–R144
10. Broglia RA, Tiana G, Berera R (2003) Resistance-proof, folding-inhibitor drugs. J Chem Phys 118:4754–4758
11. Broglia RA, Provasi D, Vasile F, Ottolina G, Longhi R, Tiana G (2006) A folding inhibitor of the HIV-1 protease. Proteins 62:928–933
12. Dobson CM, Evans PA, Radford SE (1994) Understanding how proteins fold: the lysozyme story so far. Trends Biochem Sci 19:31–37
13. Sutto L, Tiana G, Broglia RA (2006) Sequence of events in folding mechanism: beyond the Go model. Protein Sci 15:1638–1652
14. Broglia RA, Tiana G, Sutto L, Provasi D, Simona F (2005) Design of HIV-1-PR inhibitors that do not create resistance: blocking the folding of single monomers. Protein Sci 14:2668–2681
15. Caldarini M, Vasile F, Provasi D, Longhi R, Tiana G, Broglia RA (2008) Identification and characterization of folding inhibitors of hen egg lysozyme: an example of a new paradigm for drug design. Proteins 74:390–399
16. Broglia RA, Tiana G, Sutto L, Provasi D, Perelli V (2007) Low-throughput model design of protein folding inhibitors. Proteins 67:469
17. Matagne A, Dobson CM (1998) The folding process of hen lysozyme: a perspective from the'new view'. Cell Mol Life Sci 54:363–371
18. Arai S, Hirai M (1999) Reversibility and hierarchy of thermal transition of hen egg-white lysozyme studied by small-angle X-ray scattering. Biophys J 76:2192–2197
19. Matsubara M, Nohara D, Sakai T (1992) Difference between guanidinium chloride and urea as denaturants of globular proteins : the possibility of application to improved refolding processes. Chem Pharm Bull 40:550–552
20. Sasahara K, Demura M, Nitta K (2002) Equilibrium and kinetic folding of hen egg-white lysozyme under acidic conditions. Proteins 49:472–482
21. Fersht AR (1999) Structure and function in protein science. W. H. Freeman, San Francisco
22. Park B, Levitt M (1996) Energy functions that discrimante X-ray and near-native folds from well-constructed decoys. J Mol Biol 258:367–392
23. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchard E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. Science 295:1719–1722
24. Yang JJ, Buck M, Pitkeathly M, Kotik M, Haynie DT, Dobson CM, Radford SE SE (1995) Conformational properties of four peptides spanning the sequence of hen lysozyme. J Mol Biol 252:483–491
25. Tiana G, Simona F, De Mori GMS, Broglia RA, Colombo G (2004) Understanding the determinants of stability and folding of small globular proteins from their energetics. Protein Sci 13:113–124
26. Hoop TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78:3824