On the relation between native geometry and conformational plasticity

Patrícia F.N. Faísca [a,*], Cláudio M. Gomes [b]

[a] Centro de Física Teórica e Computacional, Universidade de Lisboa, Av. Prof. Gama Pinto 2, 1649-003 Lisboa, Portugal
[b] Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da Republica EAN, 2785-572 Oeiras, Portugal

ARTICLE INFO

ABSTRACT

In protein folding the term plasticity refers to the number of alternative folding pathways encountered in response to free energy perturbations such as those induced by mutation. Here we explore the relation between folding plasticity and a gross, generic feature of the native geometry, namely, the relative number of local and non-local native contacts. The results from our study, which is based on Monte Carlo simulations of simple lattice proteins, show that folding to a structure that is rich in local contacts is considerably more plastic than folding to a native geometry characterized by having a very large number of long-range contacts (i.e., contacts between amino acids that are separated by more than 12 units of backbone distance). The smaller folding plasticity of native geometries is probably a direct consequence of their higher folding cooperativity that renders the folding reaction more robust against single- and multiple-point mutations.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

During the last 15 years significant progress has been achieved towards a complete understanding of the kinetics and mechanisms of protein folding. The synergistic link between computer simulations and *in vitro* experiments has proven particularly fruitful in this endeavour [1,2]. Much of our current knowledge on protein folding has been gathered by studying small (i.e., with less than 100 amino acids) monomeric proteins, epitomized by the likes of the 64-residue protein Chymotrypsin Inhibitor 2 (CI2). Indeed, their two-state folding kinetics [3] renders them particularly suitable models to investigate, both *in vitro* and *in silico*, the rather complex phenomenon that is the folding 'reaction'.

A major and challenging task in studying two-state proteins is the structural characterization of the folding transition state (TS), located on the top of the free energy barrier that separates the native fold from the ensemble of unfolded conformations. Indeed, due to its fleeting nature, the commonly available biophysical tools have revealed inappropriate to probe the TS' structure. Thus, experimental studies of the TS have remained predominantly rooted in the use of a particular class of protein engineering methods, the so-called ϕ-value analysis, developed by Fersht [8] back in the late 1980s. In the ϕ-value analysis a non-disruptive mutation (i.e., a mutation that does not change the structure of the native state, and does not alter the folding

pathway either) is made at some position in the protein sequence [12]. The change in the activation energy of folding of the mutant with respect to that of the wild-type (WT) protein, denoted by $\Delta\Delta G^{TS-D}$, is measured together with the change in the free energy of folding, $\Delta\Delta G^{N-D}$, caused by the mutation. The corresponding ϕ value is then defined as the ratio between these quantities, namely, $\phi = \Delta\Delta G^{TS-D}/\Delta\Delta G^{N-D}$. A ϕ value of unity means that the energy of the TS is perturbed upon mutation exactly as much as that of the native state, which has traditionally been interpreted as if the protein structure is folded at the site of mutation in the TS. Conversely, residues that are unfolded in the TS exhibit ϕ-values of zero.

The traditional interpretation of fractional ϕ values is, however, not straightforward as they might indicate the existence of multiple folding pathways or a unique TS with genuinely weakened interactions [8]. Moreover, the so-called nonclassical ϕ-values ($\phi > 1$ or $\phi < 0$) are difficult to interpret in the traditional ϕ-value model. Recently, Weikl et al. [9,10] have proposed a new model that is able to capture and interpret nonclassical ϕ-values. The model assumes that cooperative structural elements (e.g., and α-helix or a β-hairpin) are fully formed in the TS, and that a mutation on a single residue affects the whole structural element where the native contacts established by the mutated residue are located. Likewise, the activation energy of folding in the definition of the ϕ-values contain explicit free energy contributions from different substructural elements of the protein. This model was recently applied to study two small β-sheet domains, namely the FBP and PIN WW domains, which were assumed to fold

via two distinct folding pathways (and transition states), and it was able to reproduce ϕ-values in good agreement with those reported experimentally [11].

The concept of conformational/folding plasticity [13], or folding malleability [14], refers to the number of alternative folding pathways encountered in response to free energy perturbations such as those induced by solvent changes or mutation; the larger that number, the broader the TS, and the larger the plasticity. Although it is experimentally very difficult to investigate in a direct manner the number of folding pathways leading to the native state, evidence for the existence of conformational plasticity was found in a few cases by means of protein engineering experiments. For example, single point mutations in the N-terminal domain of the monomeric λ_{6-85}-repressor revealed a broad TS [24]. The use of circular permutation methods – a less milder protein engineering technique where the protein's backbone is cut-up at some point, and its original N- and C-terminal parts are linked together in a process that leads to minimal alteration of the native fold – revealed considerable structural changes in the TSs of both the α-spectrin SH3 domain [15] and src SH3 domain [16], and even more dramatic modifications in the TS of the ribosomal protein S6 [17], a finding that was corroborated by computer simulation studies [18–20]. The plasticity exhibited by these proteins has been taken has evidence for the existence of multiple folding pathways [13].

In a recent study, Klimov and Thirumalai have studied the influence of the distribution of secondary structural elements along the protein sequence using an off-lattice coarse grained model and Langevin Molecular Dynamics simulations. The major result that came out of their study is that a symmetric distribution of α helices and β sheets with respect to sequence midpoint favours multiple folding pathways [25].

A study of the validity of the ϕ-value analysis as a tool for identifying critical (i.e., nucleating) residues in protein folding and providing a structural characterization of the TS was recently presented in [41]. This study involved measuring the folding time (i.e., the inverse of the folding rate) of all possible single point mutations and many double point mutations in two model proteins with different native geometries. Here we propose to explore the link between conformational plasticity and a gross geometric trait of the native structure, the relative number of local and nonlocal native contacts, by using as starting point the mutational data reported in [41]. For each native geometry, we selected a number of mutations with different kinetic effects. We investigate the relation between folding plasticity and native geometry by studying the impact of these mutations on the folding 'reaction' from the point of view of the conformational changes it encompasses. This is done by monitoring for each mutant the degree of nativeness of each residue along the folding process and comparing it with the wild type protein.
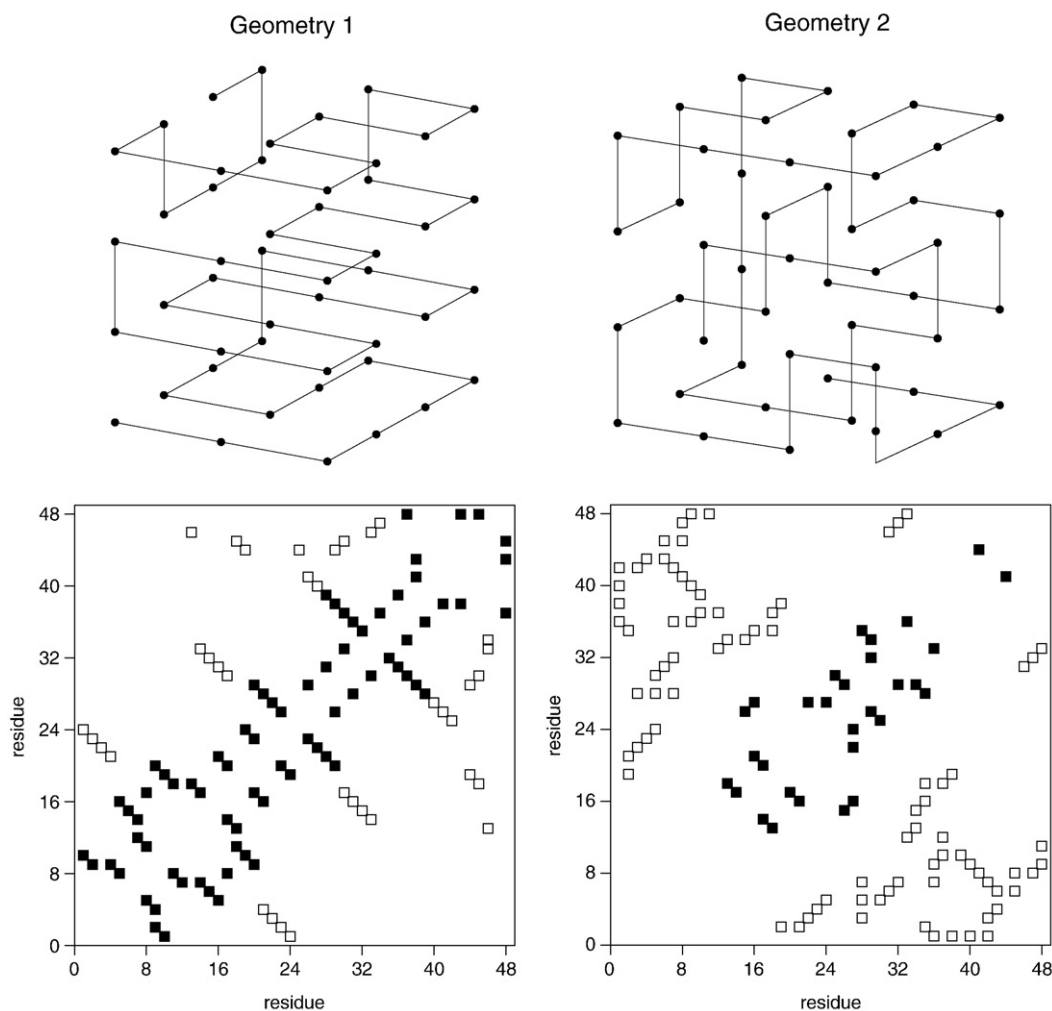

Fig. 1. Three dimensional representation of geometry 1 and 2 (top row) and corresponding contact maps (bottom row). Each square in the contact map represents a native contact. For structures that like ours are maximally compact cuboids with $N=48$ residues there are 57 native contacts. A nonlocal contact between two residues i and j is defined as LR if their sequence separation is at least 12 units, i.e., $|i-j| \geq 12$ [5]. Accordingly, the number of LR (white squares) contacts in geometry 1 is 19 and in geometry 2 is 42.

Our analysis shows that folding to the native geometry that is rich in nonlocal long-range (LR) contacts is considerably less plastic (i.e., more conformationally robust) than folding to the model protein that has predominantly local contacts, and that this is possibly a direct consequence of the more cooperative folding transition exhibited by the nonlocal geometry. This reinforces the results of [41], according to which the picture of the TS emerging from the ϕ-value analysis is more reliable when applied to target proteins having a distinctively large number of nonlocal native contacts. Therefore, if ϕ-value mutational data is to be interpreted in the traditional way, which is so far the most commonly adopted view, it is important to select native folds where the number of LR native contacts is sharply dominant; they show a higher robustness against mutation and, at the coarse level of contact cluster, they tend to fold in a Levinthal-like manner, i.e., as single route folding proteins.

This article is organized in the following manner. After the introductory section, the protein models and simulation methodologies are described. Then, we present and discuss the results from simulations. Finally, in the last section we draw some concluding remarks.
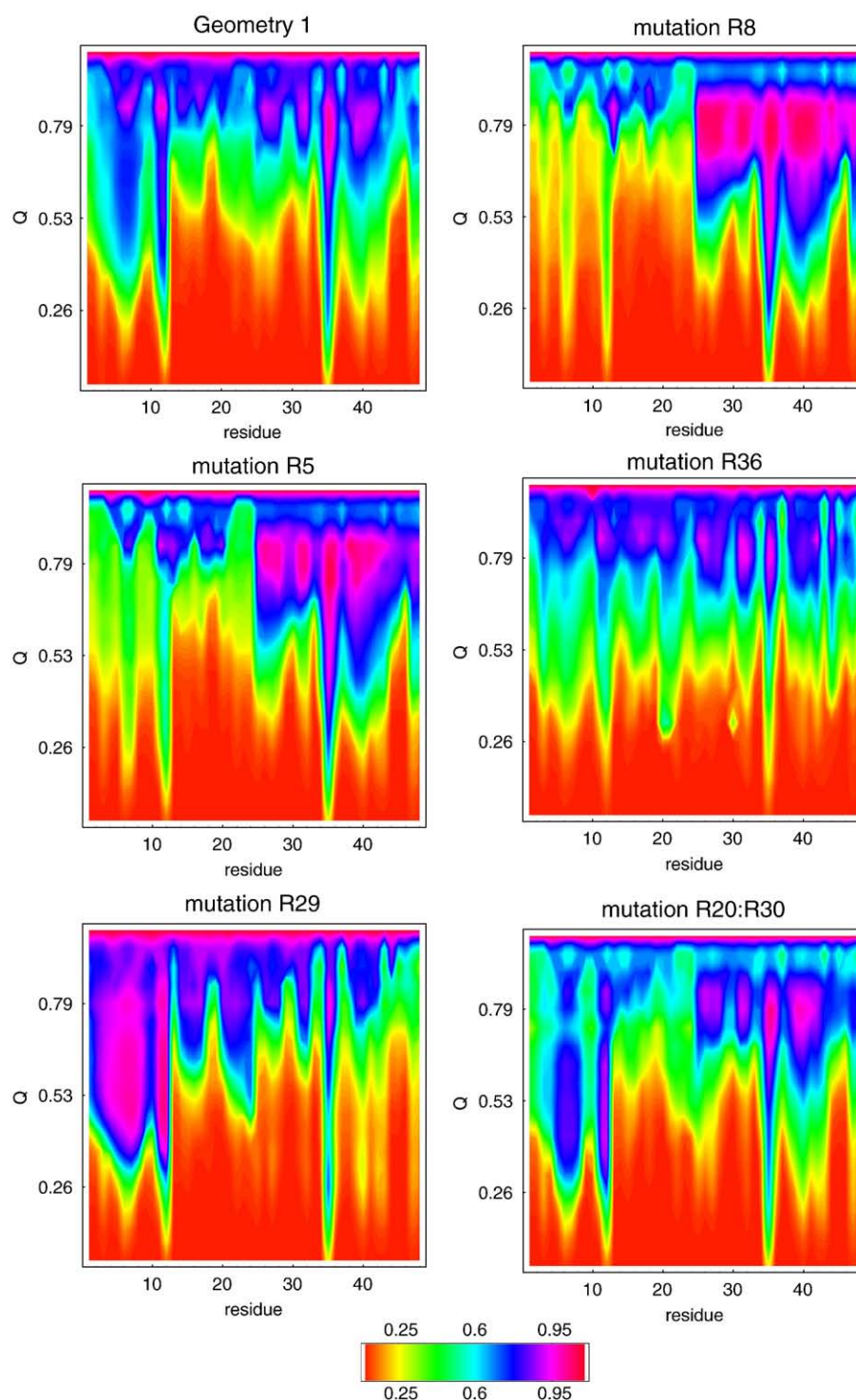
Fig. 2. Folding patterns of geometry 1: probability that each residue is in its native environment along the reaction coordinate fraction of native contacts, Q, for the wild type (top left) and mutated sequences in geometry 1.

2. Models and methods

2.1. The Gō model and simulation details

We consider a simple three-dimensional lattice model of a protein molecule with chain length $N=48$. In such a minimalist model amino acids, represented by beads of uniform size, occupy the lattice vertices and the peptide bond, which covalently connects amino acids along the polypeptide chain, is represented by sticks with uniform (unit) length corresponding to the lattice spacing.

To mimic protein energetics we use the Gō model [26]. In the Gō model the energy of a conformation, defined by the set of bead coordinates $\{\vec{r_i}\}$, is given by the contact Hamiltonian

$$H(\{\vec{r_i}\}) = \sum_{i>j}^{N} \epsilon \Delta(\vec{r_i} - \vec{r_j}),$$ (1)

where the contact function $\Delta(\vec{r_i} - \vec{r_j})$, is unity only if beads i and j form a non-covalent native contact, i.e., a contact between a pair of beads that is present in the native structure, and is zero otherwise. The
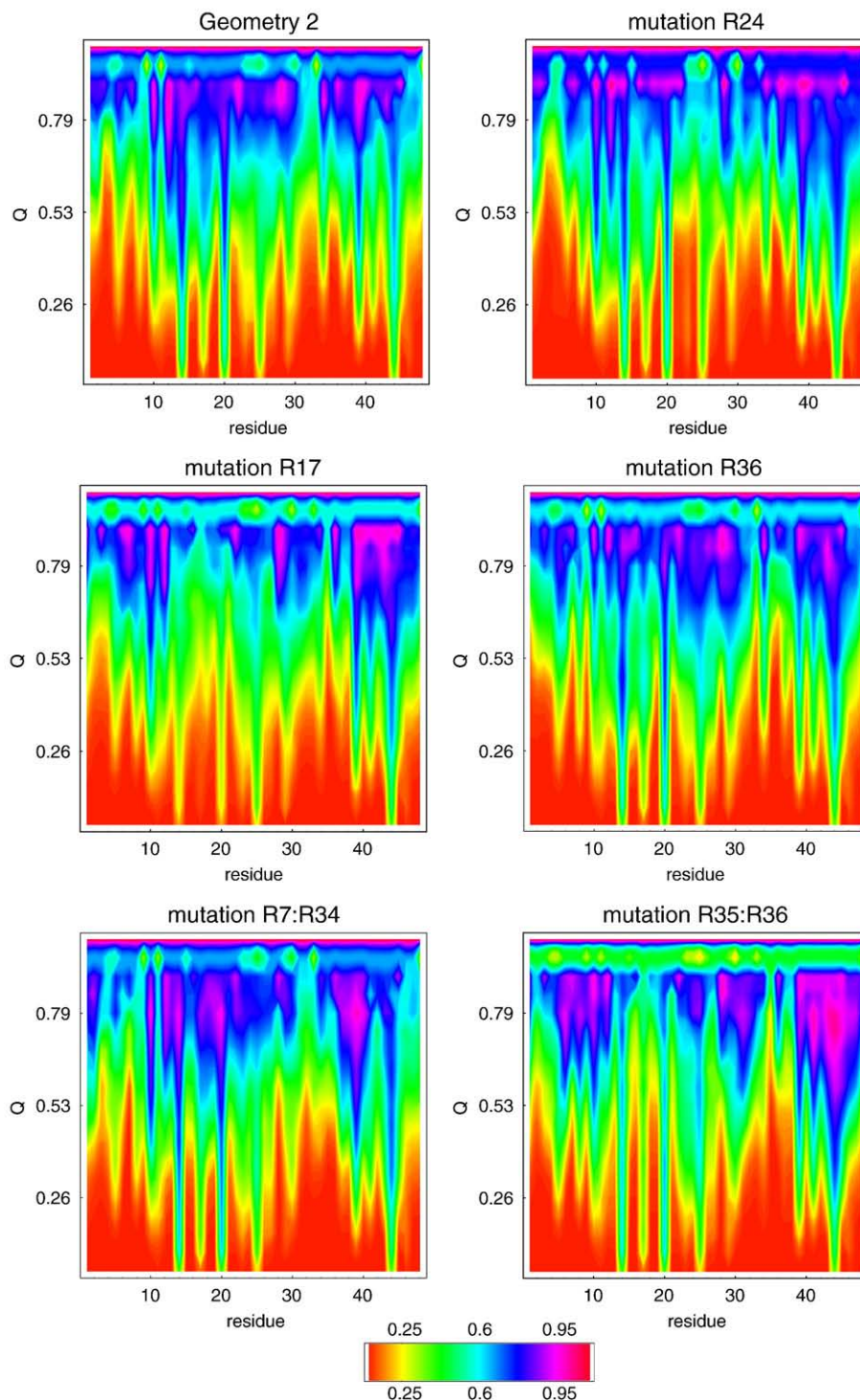
Fig. 3. Folding patterns of geometry 2: probability that each residue is in its native environment along the reaction coordinate fraction of native contacts, Q, for the wild type (top left) and mutated sequences in geometry 2.

Gō potential is based on the idea that the native fold is very well optimized energetically. Accordingly, it ascribes equal stabilizing energies (e.g., $\epsilon = -1.0$) to all the native contacts and neutral energies ($\epsilon = 0$) to all non-native contacts. The motivation to use the Gō model is based on the well accepted finding that for small, single domain two-state proteins, the geometry of the native fold is the major determinant of folding kinetics [4–6].

In order to mimic the protein's relaxation towards the native state we use a standard Monte Carlo (MC) algorithm together with the kink-jump move set [27]. Accordingly, local random displacements of one or two beads (at the same time) are repeatedly accepted or rejected in accordance with the standard Metropolis MC rule [28]. A MC simulation starts from a randomly generated unfolded conformation and the folding dynamics is monitored by following the evolution of the fraction of native contacts, $Q = q/L$, where L is number of contacts in the native fold and q is the number of native contacts formed at each MC step. The number of MC steps required to fold to the native state (i.e., to achieve $Q = 1.0$) is the first passage time (FPT) and the folding time, t, is computed as the mean FPT of 100 simulations. Except otherwise stated folding is studied at the so-called optimal folding temperature, the temperature that minimizes the folding time [29–32]. The folding transition temperature, T_f, is defined as the temperature at which denatured states and the native state are equally populated at equilibrium. In the context of a lattice model it can be defined as the temperature at which the average value $<Q>$ of the fraction of native contacts is equal to 0.5 [33]. In order to determine T_f we averaged Q, after collapse to the native state, over MC simulations lasting $\sim 10^9$ MCS.

2.2. Target geometries

Two native folds, which are amongst the 'simplest' (geometry 1) and the most 'complex' (geometry 2) cuboid geometries found through lattice simulations of homopolymer relaxation, were considered in this study (Fig. 1). A nonlocal contact is considered long-range (LR) if the two beads participating in the contact are separated by more than 12 units of backbone distance [5]. Accordingly, 33% of the native contacts in geometry 1 are LR, while geometry 2 has 74% LR contacts.

3. Results

3.1. A picture of the folding reaction

The mechanistic equivalent of the experimental ϕ-value of residue i at time t, $\phi_i^{\mathrm{mec}}(t)$, is given by the number of native contacts $q_i^{\Gamma}(t)$, the residue establishes in conformation Γ, normalised to the number of contacts it establishes in the native fold, q^{native}, i.e., $\phi_i^{\mathrm{mec}} = q_i^{\Gamma}(t)/q^{\mathrm{native}}$ [34–36]; a residue is fully native if $\phi^{\mathrm{mec}} = 1$. Thus, ϕ^{mec} provides a *direct* measure of structure formation.

In what follows we use ϕ^{mec} to obtain a picture of the folding reactions leading to geometry 1 and geometry 2, and we use the fraction of native contacts as a progress variable. In particular, for fraction of native contacts, Q, we compute the number of times each residue is fully native (i.e., $\phi^{\mathrm{mec}} = 1$) and normalize it to the total number of times Q is counted in the course of a folding event. The probability thus computed is then averaged over an ensemble of 100 independent folding simulations. We remark that although the probability to fold, P_{fold}, is the most accurate progress variable in lattice simulations of protein folding [37], its use for the purposes of the present goal is computationally prohibitive as it would have to be evaluated for *every* conformation sampled in a folding event. This is why we use instead the fraction of native contacts, which was recently shown to measure correctly the degree of closeness to the native fold [39] for proteins with smooth energy landscapes (i.e., single exponential kinetics) [38,40], like those considered here.

Not surprisingly, the folding pattern of geometry 1 (Fig. 2, top left), i.e., the conformational changes leading to the native fold, is readily distinct from that exhibited by geometry 2 (Fig. 3, top left). For example, from early to mid folding (i.e., $0.26 < Q < 0.53$), there are more residues in geometry 2 with a higher probability of being in their native environment. These probabilities decrease sharply immediately prior to collapse into the native fold (i.e., $0.79 < Q < 1.0$). On the other hand, for the vast majority of the residues belonging to geometry 1, the probability of being fully native increases in a rather continuous way from early to late folding.

3.2. Native geometry and conformational plasticity

Within the context of the Gō model a single-point mutation is equivalent to replace the set of native contacts established by the mutated residue with neutral contacts, i.e., contacts to which zero energy is ascribed. In a recent study we have mutated every single residue in each protein model and measured the change in the folding time of the mutant relative to the wild-type (WT) sequence (Fig. 4); an extensive number of double point mutations was also performed [41]. We now ask how does the folding pattern of both protein models previously described alters upon mutation? To answer this question we have selected single point mutations that produce different kinetic effects. For both geometries we have investigated the effect on conformational plasticity of the most deleterious and of the milder forms of single point mutations. We have also considered the impact on the conformational changes occurring during folding of the most deleterious double point mutations. Thus, for geometry 1 we have looked into the folding patterns obtained upon mutating residue 29 (which leads to the largest folding time increase of 200%), residues 5 and 36 (neutral mutations located below and above the sequence midpoint respectively, and also residue 8 (which decreases the WT protein folding time by 16.2%) (Fig. 4). A simultaneous mutation on the residues 20 and 30, which is the most deleterious double point mutation for geometry 1, increases the folding time of the WT protein by almost two orders of magnitude [41]. We have also investigated its impact on conformational plasticity. The pattern of the folding reaction associated with any one of these mutations is sharply different from that exhibited by the WT protein (Fig. 2), which is indicative that the less complex geometry has access to several folding pathways. For the more complex geometry 2, one has selected the mutation on residue 36 (which produces the largest increase in folding
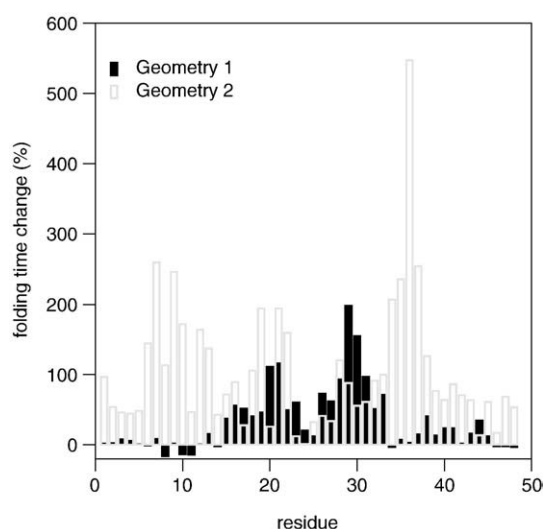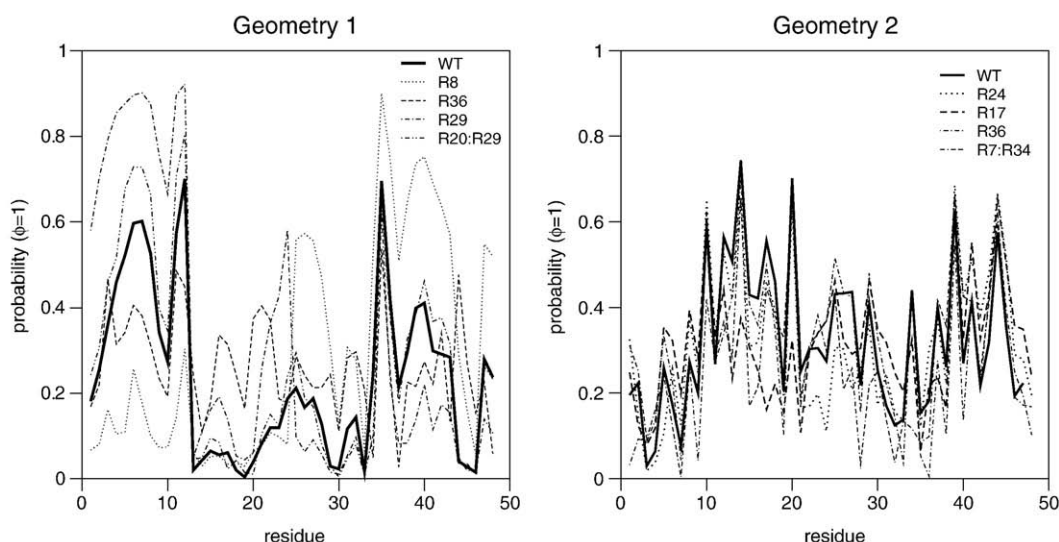

Fig. 4. Change in folding time relative to the wild-type protein resulting from single point mutations [42].

Fig. 5. Probability that a residue is in its native environment in the wild-type transition state (TS) of geometry 1 (left) and geometry 2 (right). Also shown is the TS change induced by single and double-point mutations.

time of 600%), on residue 24 (which is the only mutation for which a vanishingly small change of 1.2% in the folding time is observed), and also on residue 17 (which leads to a mild increase of 27% the WT protein's folding time). We have also investigated the effects of two double point mutations, namely on residues 7 and 34, and on residues 35 and 36. These mutations increase the WT protein's folding time by more than two orders of magnitude [41]. A scenario, considerably different from that reported for geometry 1, is observed for geometry 2 (Fig. 3). Indeed, in this case, the folding pattern is considerably more robust, which suggests that the search for the native fold is constrained to follow a fixed sequence of conformational changes, i.e., that this geometry has a smaller conformational plasticity.

3.3. Long-range contacts, the structure of the transition state and the cooperativity of the folding transition

As discussed below, the large number of LR contacts in geometry 2 protects folding from perturbation, which decreases the conformational plasticity, in two ways: i) by increasing the robustness of the

transition state' structure against mutation and ii) by increasing the cooperativity of the folding transition.

Fig. 5 (straight line) reports the probability that each protein residue is fully native in the TS ($Q \approx 0.5$) of geometry 1 (left) and geometry 2 (right), and how this probability changes upon mutation. The structure of the TS of geometry 2 is clearly more robust against mutation than that of geometry 1. Indeed, most mutations in geometry 1 lead to a large structural consolidation between residues 12 and 33. In geometry 2, the most deleterious double-point mutations disrupt networks of nonlocal LR contacts that form the core structure of the TS [41]. Since the formation of LR contacts greatly restricts the number of conformations available to the folding chain, the establishment of LR contacts is generally entropically costly. Because the LR contacts are clearly dominant in the native fold of geometry 2 it is highly unlikely that upon the mutation of one or more of its residues, folding can occur through an energetically competitive TS, as this would imply changing the TS structure by increasing substantially its content in local (i.e., less entropically costly) native contacts [43,44]. Thus, for the nonlocal geometry 2, the robustness of its TS is a direct consequence of the fold's high content in LR contacts.
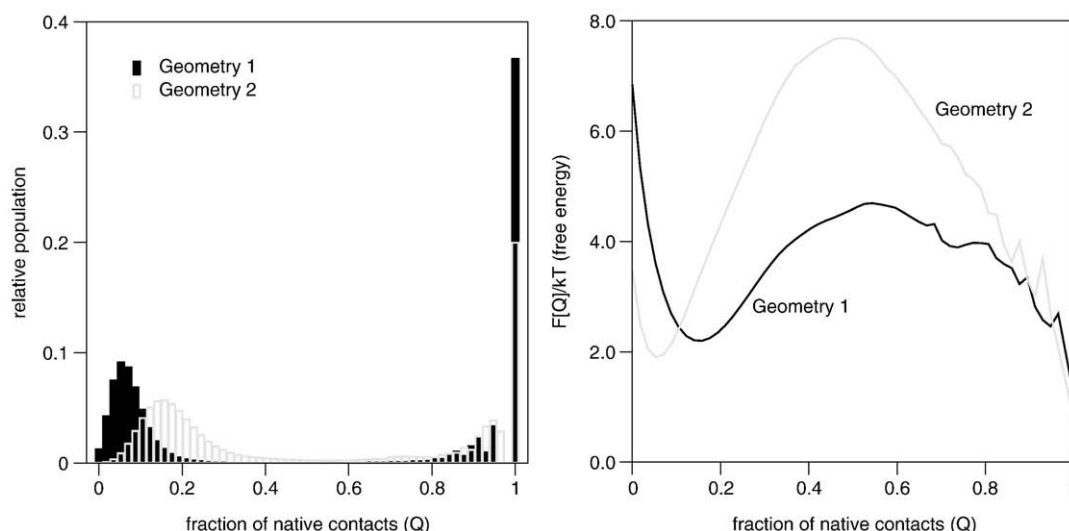
Fig. 6. Cooperativity of the folding transition. Population histogram for the frequency of occurrence of conformations with fraction of native contacts Q, $P(Q)$ (left), and the free energy computed as $F(Q) = -kT \ln P(Q)$ (right). To compute $P(Q)$ data were averaged after collapse to the native state in long MC simulations lasting over $\sim 10^9$ MCS.

In protein folding the formation (and breaking) of native contacts in a non-independent manner (i.e., cooperatively) results into the depletion of partially folded conformational states, which translates into a kinetic behaviour that fits remarkably well a two-state model. Indeed, at the transition midpoint of a two-state folding 'reaction', half of the protein molecules in the test tube are folded and half of them are coil. Thus, microscopically, a cooperative folding transition is characterized by a bimodal distribution of protein molecules over energy, fraction of native contacts, Q, or any other observable parameter [21]. Data reported in Fig. 6 (top) shows that at the transition temperature, T_m, the population distribution of the fraction of native contacts for geometry 2, $P(Q)$, is more strongly bimodal than that of geometry 1, showing that for geometry 2 the folding transition is clearly more cooperative [34]. The stronger cooperative stabilization of the native state of geometry 2 relative to that of geometry 1 is more evident from the corresponding free energy curve, $F(Q) = -kT \ln P(Q)$, where a very high free energy maximum separates the native conformation from the of ensemble of unfolded conformers (Fig. 6, bottom) [27]. Mechanistically, the strong cooperative folding of geometry 2 restricts the number of allowed conformational changes, and as a consequence, the number of alternative folding trajectories (i.e., the conformational plasticity) is smaller for this more complex geometry.

4. Conclusion

According to the statistical or landscape view of protein folding the native state can be reached from a myriad of microscopic parallel pathways. This 'new' view contrasts the so-called traditional view that envisages folding as a Levinthal-like search, where the elements of the native structure assemble in a well-defined order. Thus, identifying the number of folding pathways accessible to a protein chain is a crucial question in molecular biology. Recent molecular dynamics simulations of the unfolding of CI2 at high temperature showed that the observation of sequential or multiple unfolding pathways depends on the 'resolution', i.e. level of structural detail, at which one observes the unfolding process. Indeed, at the level of individual contact formation the unfolding of CI2 happens through highly parallel folding pathways while at the coarser level of contact clusters sequential folding events emerge [22]. Experimentally, however, the observation of multiple folding pathways has proven a much more challenging task, and indirect evidence for alternative routes to the native state has been reported for engineered proteins only and the term folding or conformational plasticity was coined to denote the multiplicity of folding pathways identified for the perturbed (i.e., mutated) system. Folding plasticity for small perturbations (i.e., mild mutations) is taken as an indication that multiple folding routes exist for the wild-type protein. Here we have investigated the effect of single- and double-point mutations on the robustness of the folding reaction (and TS structure) leading to two distinct protein geometries. More specifically, we have investigated how the probability that each residue is fully native (i.e., that is has all its native contacts formed) evolves during the folding process, and how this evolution responds to mutation. Our findings suggest that folding to native geometries which are distinctively rich in long-range native contacts is more robust against mutation than folding to native geometries where the number of local native contacts is dominant. Indeed, 'local' geometries seem to be able to find alternative folding routes in response to mutation while nonlocal geometries appear to behave in a Levinthal-like manner, their folding reaction being constrained to follow a fixed sequence of conformational changes, when monitored at the coarse level of contact cluster (i.e., the set of native contacts established by each residue). In other words, the conformational plasticity is expected to be small for target proteins where the number of nonlocal native contacts is sharply larger than the number of local contacts. These findings are supported by experimental results reported for the λ_{6-85}-repressor (an α-protein that is rich local contacts) and for protein CI2, and src SH3 domain (α/β proteins rich in nonlocal contacts); while the TSs of the latter were found to be largely invariant against single and double point mutations [7,23], substantial changes were found for the former. Interestingly, the more complex native fold, i.e., the one that is rich is native contacts is clearly less symmetric than the more local geometry, suggesting that a relation similar to that found by Klimov and Thirumalai between the symmetry of the tertiary structure and the folding plasticity [25] may also hold for lattice proteins.

Mechanistically, the lower folding plasticity of the more complex native geometries is most possibly a direct consequence of their more cooperative folding transition, which results from a large content in long-range native contacts. In agreement with this explanation, experimental evidence for high conformational plasticity, resulting from a weak cooperative behaviour, was recently reported. Indeed, a change in the folding mechanism was found for two members of the *cks* family as a result of weakening the cooperativity of the core of the protein [42].

In our previous work [41] we have concluded that native folds having a distinctively large number of non-local native contacts are more suitable targets for TS studies based on the use of ϕ-value analysis. The results reported here strengthen this conclusion; indeed, due to their smaller folding plasticity, model proteins with a distinctively large number of LR contacts are preferable targets for TS studies based on the traditional interpretation of ϕ-values.

Acknowledgment

References

[1] C.M. Dobson, M. Karplus, The fundamentals of protein folding: bringing together theory and experiment, Curr. Opinion Struct. Biol. 9 (1999) 92–101.
[2] G. Weihua, J.E. Shea, S. Berry, The physics of the interactions governing folding and association of proteins, Ann. N. Y. Acad. Sci. 1066 (2005) 34–53.
[3] S.E. Jackson, How do small single-domain proteins fold? Fold Des. 3 (1998) R81–91.
[4] K.W. Plaxco, K.T. Simmons, I. Ruczinski, D. Baker, Topology, stability, sequence and length: defining the determinants of two-state protein folding kinetics, Biochemistry 39 (2000) 11177–11183.
[5] M.M. Gromiha, S. Selvaraj, Comparison between long-range interactions and contact order in determining the folding rate of two-state folders: application of long-range order to folding rate prediction, J. Mol. Biol. 310 (2001) 27–32.
[6] H. Zhou, Y. Zhou, Folding rate prediction using total contact distance, Biophys. J. 82 (2002) 458–463.
[7] L.S. Itzhaki, D.E. Otzen, A.R. Fersht, The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein-folding, J. Mol. Biol. 254 (1995) 260–288.
[8] A. Fersht, in: W.H. Freeman (Ed.), Structure and Mechanism in Protein Science: A guide to enzyme catalysis and protein folding, 3rd Edition, 1998.
[9] C. Merlo, K.A. Dill, T.R. Weikl, ϕ-values in protein-folding kinetics have energetic and structural components, Proc. Natl. Acad. Sci. U. S. A 102 (2005) 10171–10175.
[10] T.R. Weikl, K.A. Dill, Transition states in protein folding kinetics: the structural interpretation of ϕ-values, J. Mol. Biol. 365 (2007) 1578–1586.
[11] T.R. Weikl, Transition states in protein folding kinetics: modeling ϕ-values of small β-sheet proteins, Biophys. J. 94 (2008) 929–937.
[12] A.R. Fersht, S. Sato, ϕ-value analysis and the nature of protein-folding transition states, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7976–7981.
[13] V. Munoz, Folding plasticity, Nature Struct. Biol. 9 (2002) 792–794.
[14] M.O. Lindberg, M. Oliveberg, Malleability of protein folding pathways: a simple reason for complex behaviour, Curr. Opin. Struct. Biol. 17 (2007) 21–29.
[15] A.R. Viguera, L. Serrano, M. Wilmanns, Different folding transition states may result in the same native structure, Nature Struct. Biol. 3 (1996) 874–880.
[16] V.P. Grantcharova, D. Baker, Circularization changes the folding transition state of the src SH3 domain, J. Mol. Biol. 306 (2001) 555–563.
[17] M. Lindberg, J. Tangrot, M. Oliveberg, Complete change of the protein folding transition state upon circular permutation, Nature Struct. Biol. 9 (2002) 818–822.
[18] I.A. Hubner, M. Lindberg, E. Haglund, M. Oliveberg, E.I. Shakhnovich, Common motifs and topological effects in the protein folding transition state, J. Mol. Biol. 359 (2006) 1075–1085.

[19] I.A. Hubner, M. Oliveberg, E.I. Shakhnovich, Simulation, experiment, and evolution: Understanding nucleation in protein S6 folding, Proc. Natl. Acad. Sci. U. S. A. 101 (2006) 8354–8359.

[20] S. Matysiak, C. Clementi, Optimal combination of theory and experiment for the characterization of the protein folding landscape of S6: how far can a minimalist model go? J. Mol. Biol. 343 (2004) 235–248.

[21] G.I. Makhatadze, P.L. Privalov, Energetics of protein structure, Adv. Protein Chem. 47 (1995) 307–309.

[22] L. Reich, T.R. Weikl, Substructural cooperativity and parallel versus sequential events during protein unfolding, Proteins 63 (2006) 1052–1058.

[23] V.P. Grantcharova, D.S. Riddle, D. Baker, Long-range order in the src SH3 folding transition state, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 7084–7089.

[24] R.E. Burton, G.S. Huang, M.A. Daugherty, T.L. Calderone, T.G. Oas, The energy landscape of a fast-folding protein mapped by Ala-Gly substitutions, Nat. Struct. Biol. 4 (1997) 305–310.

[25] D.K. Klimov, D. Thirumalai, Symmetric connectivity of secondary structure elements enhances the diversity of folding pathways, J. Mol. Biol. 353 (2005) 1171–1186.

[26] N. Go, H. Taketomi, Respective roles of short- and long-range interactions in protein folding, Proc Natl. Acad. Sci. U. S. A. 75 (1978) 559–563.

[27] D.P. Landau, K. Binder, A guide to Monte Carlo Simulations in statistical physics, Cambridge University Press, 2000.

[28] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.

[29] M. Oliveberg, Y. Tan, A.R. Fersht, Negative activation enthalpies in the kinetics of protein folding, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 8926–8929.

[30] A. Gutin, A. Sali, V. Abkevich, M. Karplus, E.I. Shakhnovich, Temperature dependence of the folding rate in a simple protein model: search for a "glass" transition, J. Chem. Phys. 108 (1998) 6466–6483.

[31] M. Cieplak, T.X. Hoang, M.S. Li, Scaling of folding properties in simple models of proteins, Phys. Rev. Lett. 83 (1999) 1684–1687.

[32] P.F.N. Faisca, R.C. Ball, Thermodynamic control and dynamical regimes in protein folding, J. Chem. Phys. 116 (2002) 7231–7238.

[33] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Impact of local and non-local interactions on thermodynamics and kinetics of protein folding, J. Mol. Biol. 252 (1995) 460–471.

[34] M. Vendruscolo, E. Paci, C.M. Dobson, M. Karplus, Three key residues form a critical contact network in a protein folding transition state, Nature 409 (2001) 641–645.

[35] I.A. Hubner, J. Shimada, E.I. Shakhnovich, Commitment and nucleation in the protein G transition state, J. Mol. Biol. 336 (2004) 745–761.

[36] L.R. Allen, E. Paci, Transition states for protein folding using molecular dynamics and experimental restraints, J. Phys.: Condens. Matter 19 (2007) 285211.

[37] R. Du, V.S. Pande, A.Y. Grosberg, T. Tanaka, E.S. Shakhnovich, On the transition coordinate for protein folding, J. Chem. Phys. 108 (1998) 334–350.

[38] B. Gillespie, K.W. Plaxco, Non-glassy kinetics in the folding of a simple, single domain protein, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 12014–12019.

[39] S. Cho, Y. Levy, P.G. Wolynes, Q. P versus, Structure reaction coordinates capture protein folding on smooth landscapes, Proc. Natl. Acad. U. S. A. 103 (2006) 586–591.

[40] C. Clementi, P.A. Jennings, J.N. Onuchic, Prediction of folding mechanism for circular-permuted proteins, J. Mol. Biol. 311 (2001) 879–890.

[41] P.F.N. Faisca, R. Travasso, R.C. Ball, E.I. Shakhnovich, Identifying critical residues in protein folding: Insights from ϕ-values and Pf old analysis, J. Chem Phys. 129 (2008) 095108.

[42] M.A. Seeliger, S.E. Breward, L.S. Itzhaki, Weak cooperativity in the core causes a switch in folding mechanism between two proteins of the cks family, J. Mol. Biol. 325 (2003) 189–199.

[43] D. Baker, A surprising simplicity to protein folding, Nature 405 (2000) 39–42.

[44] N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, Identifying the protein folding nucleus using molecular dynamics, J. Mol. Biol. 296 (2000) 1183–1188.