
SHORT COMMUNICATION

The “First In–Last Out” Hypothesis on Protein Folding Revisited

Alessandro Pintar* and Sándor Pongor

Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), AREA Science Park, Trieste, Italy

ABSTRACT We calculated profiles for mean residue depth, contact order, and number of contacts in the native structure of a series of proteins for which folding has been studied extensively, the chymotrypsin inhibitor 2, the SH3 module from the src tyrosine kinase, the small ribonuclease barnase, the bacterial immunity protein Im7, and apomyoglobin. We compared these profiles with experimental data from equilibrium or pulse labeling hydrogen-deuterium exchange obtained from NMR and ϕ values obtained from the protein engineering approach. We find a good qualitative agreement between the hierarchy of formation of topological elements during the folding process and the ranking of secondary structure elements in terms of residue depth. Residues that are most deeply buried in the core of the native protein usually belong to stretches of secondary structure elements that are formed early in the folding pathway. Residue depth can thus provide a useful and simple tool for the design of folding experiments. *Proteins* 2005; 60:584–590. © 2005 Wiley-Liss, Inc.

Key words: depth; contact order; H/D exchange; topology; protein engineering

INTRODUCTION

A simple linear inverse correlation exists between the average sequence separation among residues that make contacts in the native 3D structure of a protein (the contact order, CO) and its global folding rate.¹ Protein topology can thus provide a simple key to predict the overall folding rate of a polypeptide chain,² proteins with a simple topology (low CO, for example all α proteins) folding faster than those having a complex one (high CO, for example all β proteins). However, the contact order contains little information about specific interactions occurring at the residue level in the transition state.³ To get insight into the atomic details of the folding pathway, different experimental approaches have been devised. On the one hand, backbone amide hydrogen-deuterium (H/D) exchange monitored either by NMR^{4–6} or mass spectrometry⁷ provided a well of information about the folding of several proteins in the msec time scale and longer. On the other, the protein

engineering approach⁸ provided a residue-specific, though indirect method to evaluate the contribution of a residue side chain to the transition state, as measured by the residue's ϕ value. Given the native protein and a mutant, ϕ is defined as $\Delta \ln k_f / \Delta \ln K$, where k_f is the folding rate constant, K is the equilibrium constant (k_f/k_u) and Δ represents the mutation-induced difference in the corresponding values. Residues having ϕ values close or equal to 1 define the folding nucleus. It has been suggested that selected residues displaying a high degree of conservation within structurally related proteins belong to the folding nucleus and can act as folding accelerators.⁹ It has also been observed that there is a correlation between residue conservation and residue depth, or, in other words, that the most conserved residues within the same family of structurally aligned proteins are also the most deeply buried in the native structure.¹⁰ Moreover, it is known that residue depth correlate well with H/D exchange rates,¹¹ and it has been proposed that the last backbone NH hydrogens to exchange with solvent in the native protein might identify the first parts of the protein to fold.^{12,13} This hypothesis, abridged as “first in–last out,” has been a subject of debate.^{13–15} Here, we examine the link between protein topology, residue depth,^{10,11} and experimental results available from folding experiments in vitro of a set of well studied proteins, as obtained from NMR data and from the protein engineering method.

METHODS

Atom depths (Å) were calculated using DPX¹⁶ (see also <http://hydra.icgeb.trieste.it/dpx/>). With this simple algorithm, the atomic solvent accessible surface is calculated first from the 3D structure of the native protein, and depth of any nonhydrogen atom i is defined as the shortest distance between i and its closest solvent accessible neigh-

*Correspondence to: Alessandro Pintar, Protein Structure and Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), AREA Science Park, Padriciano 99, 34012 Trieste, Italy. E-mail: pintar@icgeb.org

Received 22 October 2004; Revised 25 January 2005; Accepted 1 March 2005

Published online 14 July 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20529

bor. Mean residue depth ($\langle \text{dpxr} \rangle$, Å) is calculated as the average of atom depths for a certain residue. For chymotrypsin inhibitor 2 (CI2), dpxr values were calculated for each of the 20 models of the NMR structure ensemble¹⁷ (PDB: 3ci2) and averaged. Coordinates of the src-SH3 module were extracted from the crystal structure comprising the SH3, SH2, and kinase domains of src¹⁸ (PDB: 1fmk) and residues renumbered according to Riddle et al.¹⁹ For the wild-type barnase²⁰ (PDB: 1a2p) the coordinates of chain A only were considered. Coordinates for Im7 refer to the PDB entry 1ayi. For apomyoglobin, dpxr values were calculated from the high resolution crystal structure of sperm whale carbonmonoxymyoglobin (PDB: 1a6g) from which heme atoms were removed. The ϕ values of refolding in water were used for CI2,²¹ src-SH3,¹⁹ barnase,²² and Im7.²³ When ϕ values for several mutants at the same position were reported, the $\rightarrow A$ mutation was chosen as representative. Contact order for residue k is defined as $\text{CO}_k = 1/N \cdot \sum \Delta S_{kj}$, where ΔS_{kj} is the sequence distance separation between residue k and the contacting residues j , and N is the total number of contacts for residue k . Contact order (CO) profiles and total number of contacts (NC) were calculated within a 8.5 Å distance from the side chain center using the Protein Structure Workbench web server (<http://mmlsun4.pha.unc.edu/psw/3dworkbench.html>).

RESULTS

Exhaustive information on folding kinetics and hydrogen exchange is available only for a limited number of proteins,²⁴ mainly due to the difficulties of following the folding process in real time.^{25,26} We therefore focused our analysis on a set of small, globular, monomeric, well studied proteins, the chymotrypsin inhibitor 2 (CI2) from barley (CATH 3.30.10.10, $\alpha+\beta$ two-layer sandwich), the SH3 module from the src tyrosine kinase (CATH 2.30.30.40, β -roll), barnase (CATH 2.20.25.50, mainly β , single sheet) a small ribonuclease from *Bacillus amyloliquefaciens*, the bacterial immunity protein Im7 (CATH 1.10.1200.20, mainly α , orthogonal bundle), and sperm whale apomyoglobin (CATH 1.10.490.10, mainly α , orthogonal bundle).

From extensive mutational studies²¹ it has been proposed that the folding nucleus of CI2 is formed by the N-terminal region of the α -helix ($\alpha 1$) that is being stabilized by interactions of A16 ($\alpha 1$) with L49 ($\beta 4$) and I57 ($\beta 5$). Together with $\beta 3$, $\alpha 1$ and $\beta 4$ also belong to the slow exchange core of CI2.

A plot of mean residue depth ($\langle \text{dpxr} \rangle$) versus residue number (Fig. 1) for CI2 shows that A16, L49, I57 also belong to the set of residues that are most deeply buried in the native structure. Of the secondary structure elements, the central part of the helix (residues 16–20) and the central β -strand ($\beta 4$, 45–52) also show the highest dpxr values ($\langle \text{dpxr} \rangle_{16-20} = 1.75$ and $\langle \text{dpxr} \rangle_{45-52} = 1.65$, respectively). I20 and V47 show high dpxr values but only intermediate ϕ values, although belonging to the helix and the $\beta 4$ strand. For W5, no information is available from mutational studies as this residue is routinely used as

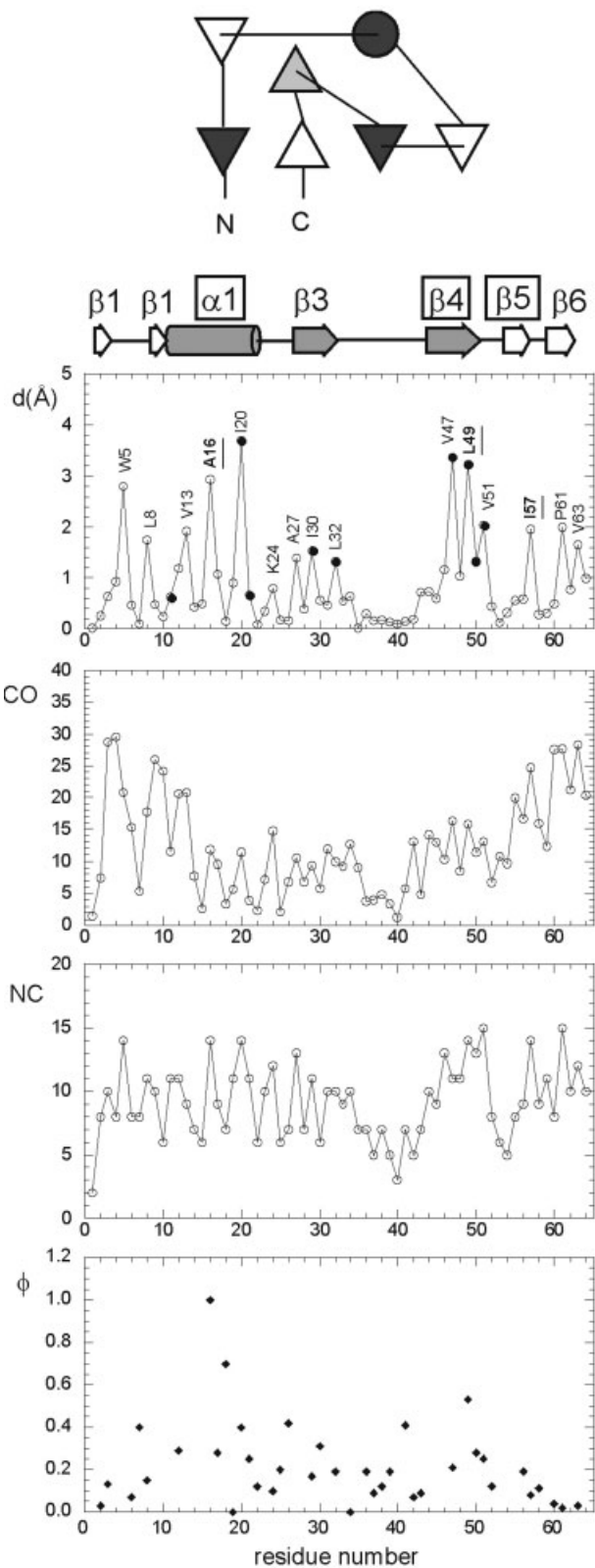
reporter in the fluorescence quenching experiments used to measure folding kinetics. Of the nine residues that display a particularly slow exchange rate at equilibrium, six correspond to peaks in the dpxr profile. The quantitative correlation between dpxr and ϕ values is low ($R = 0.32$) but it improves ($R^* = 0.53$) when considering only buried residues (residue solvent accessibility $< 20\%$). The number of buried residues for which ϕ values were measured unfortunately is limited (12) despite the relatively high sequence coverage (59%) but the correlation probability ($P = 0.07$) is not negligible.

For the SH3 domain, predictions of the folding nucleus based on the search for saddle points along simulated folding pathways turned out to be particularly difficult.²⁷ From both the protein engineering approach, which implied the mutation of 52 of the 57 residues in the protein^{28,19} and NMR studies of protein fragments,²⁹ it has been proposed that the β -hairpin formed by strands $\beta 3$ and $\beta 4$ is the most ordered structure in the transition state, with the diverging turn (residues 28–30) and strand $\beta 2$ partially ordered, while the N- and C-terminal regions are largely unstructured. A clearly defined slow exchange core does not emerge from H/D exchange NMR studies,³⁰ with only three residues, Y14, A45, I56 as the slowest exchanging ones.³¹

From the dpxr profile, L32, S47, and I56 (respectively in $\beta 2$, $\beta 3$, and $\beta 4$) display the highest values (Fig. 2). Other residues, like F10/A12 ($\beta 1$), V61 ($\beta 5$) and F26 show somewhat lower values, although depth calculated for side chain atoms separately points to F26 as an important residue in the hydrophobic cluster forming the core of the structure. Dpxr values averaged over secondary structure elements do not show a clear hierarchy: $\beta 1$, $\langle \text{dpxr} \rangle_{84-88} = 1.32$; $\beta 2$, $\langle \text{dpxr} \rangle_{107-110} = 1.42$; $\beta 3$, $\langle \text{dpxr} \rangle_{118-123} = 0.97$; $\beta 4$, $\langle \text{dpxr} \rangle_{129-133} = 1.54$; $\beta 5$, $\langle \text{dpxr} \rangle_{137-139} = 1.24$. A higher polarization of the transition state is expected from ϕ rather than from dpxr values, some discrepancy occurring in the N- and C-terminal regions ($\beta 1$ and $\beta 5$). Nevertheless, many of the residues proposed to constitute the folding nucleus also in this case correspond to peaks in residue depth in the native structure (Fig. 2). The sequence coverage of ϕ values is 64%. The correlation between dpxr and ϕ values is very low ($R = 0.08$), but it drastically improves when considering only buried residues ($R^* = 0.70$, 11 residues, $P = 0.02$).

In barnase, extensive mutational²² and NMR studies³² suggest that the docking of the β -sheet and α -helix 1 ($\alpha 1$) occurs early in the folding process, with the central strand ($\beta 3$) of the sheet forming first, followed by the two flanking ones ($\beta 2$ and $\beta 4$). In more detail, the second and third turn of $\alpha 1$, but not the first, form early in folding. This was confirmed by NMR studies of protein fragments and NH protection from solvent exchange. It has also been proposed that the central parts of $\alpha 2$ and $\alpha 3$ form early, but dock one onto the other and on the rest of the protein only at a later stage. The slow exchange core is formed by the same secondary structure elements, the helix $\alpha 1$ and the central part of the β -sheet ($\beta 3$ with the flanking $\beta 2$ and $\beta 4$).

In barnase, residues I88 and L89 in $\beta 3$ show the highest dpxr values (Fig. 3), and strand $\beta 3$ the highest average dpxr of all the secondary structure elements ($\langle \text{dpxr} \rangle_{85-91} =$



2.66). The two flanking β -strands $\beta 4$ and $\beta 2$ follow in the dpxr ranking ($\langle \text{dpxr} \rangle_{94-99} = 1.79$; $\langle \text{dpxr} \rangle_{70-76} = 1.38$, respectively) contributing residues W71 and I76 in $\beta 2$, I96 and Y97 in $\beta 4$. The two external strands, $\beta 1$ ($\langle \text{dpxr} \rangle_{50-55} = 0.88$) and $\beta 5$ ($\langle \text{dpxr} \rangle_{106-108} = 0.43$) contribute marginally to dpxr. Analyzing in detail the dpxr profiles for α -helices, we find that in $\alpha 1$ ($\langle \text{dpxr} \rangle_{6-18} = 1.14$) most of the contribution comes from the second turn (V10, A11; $\langle \text{dpxr} \rangle_{10-14} = 2.05$) and to a lesser extent from the third (L14, $\langle \text{dpxr} \rangle_{10-18} = 1.33$) while $\alpha 2$ and $\alpha 3$ contribute only marginally ($\langle \text{dpxr} \rangle_{26-34} = 0.71$; $\langle \text{dpxr} \rangle_{41-46} = 0.47$, respectively). Only A30, in the middle of $\alpha 2$, is deeply buried in the native structure (dpxr = 2.93; dpxr = 3.81 for the side chain only) and makes an exception. Overall, the subdomain formed by $\alpha 2$, $\alpha 3$, and $\beta 1$ contribute much less to residue depth than the core domain ($\alpha 1$, $\beta 2$, $\beta 3$, $\beta 4$), which is thought to form at an early stage in the folding pathway. Also in this case the correlation between dpxr and ϕ values (sequence coverage 30%) is low ($R = 0.13$) and improves when considering only buried residues ($R^* = 0.62$, 10 residues, $P = 0.05$).

We also considered two α -helical proteins, Im7 and apomyoglobin. In Im7, both equilibrium hydrogen exchange data³³ and mutational studies²³ pointed towards a rapid formation of helices h1, h2, and h4, which are already present in the folding intermediate, while h3 is formed later. Depth profiles (Fig. 4) are clearly consistent with this view, with h3 displaying a distinctly lower $\langle \text{dpxr} \rangle$ value (0.33) in respect to the other helices (0.89, 0.99, 0.72 for h1, h2, h4, respectively). The sequence coverage of ϕ values in Im7 is 28% and the correlation between dpxr and ϕ is $R = 0.29$. R^* is 0.48, and increases to $R^* = 0.87$ when discarding two outliers (I72, for which $\phi \gg 1$, and F15; 12 residues, $P < 10^{-3}$).

In apomyoglobin, hydrogen exchange pulse labeling³⁴ showed that a stable intermediate is formed within the 5 msec of the burst phase, and comprises helices A, G, H, and part of helix B. Depth data (Fig. 5) are once again consistent with experimental results. Ranking residues in order of decreasing $\langle \text{dpxr} \rangle$, we find that all the top 10 residues belong either to helix A, G, H or to the C-terminal part of helix B. Of the top 20 residues, 16 belong to the same helices. Moreover, the top ranking residues belonging to the same helices often follow the helix periodicity (V10/V13/V17 in helix A; H24/I28,

Fig. 1. Topology, secondary structure, mean residue depth (d , Å), residue contact order (CO), total number of contacts per residue (NC) and ϕ values for Cl2. In the topology cartoon (TOPS)⁴¹ helices and β -strands are drawn as circles and triangles, respectively, and colored according to mean residue depth values averaged over the secondary structure element. The gray scale is from white (low depth values) to black (high depth values). In the secondary structure diagram helices are shown as cylinders and β -strands as arrows. Stretches of secondary structure that are forming at an early stage of folding according to the protein engineering approach are boxed. Secondary structure elements belonging to the slow exchange core are filled in gray. Residues reported to belong to the folding nucleus in the original references are in bold and underlined. Residues reported to have particularly slow exchanging backbone NH protons in the native structure are shown as filled circles. Residue numbers and assignment of secondary structure can differ from one reference to another.

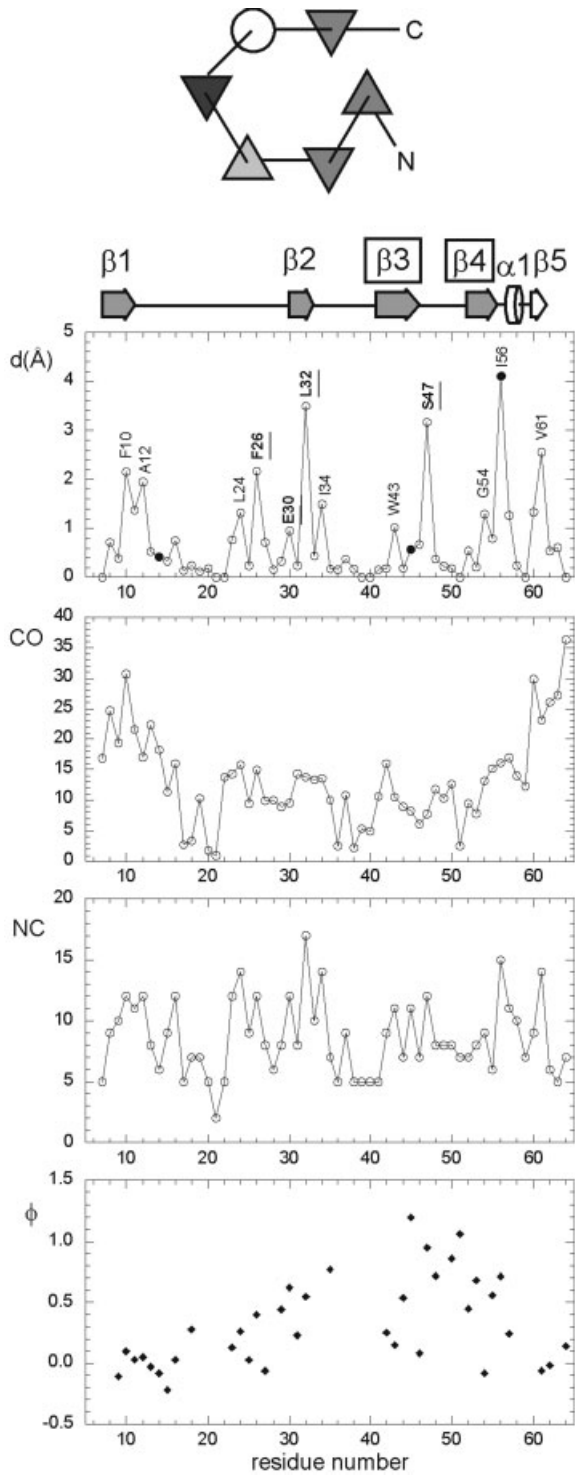


Fig. 2. Topology, secondary structure, mean residue depth (d , Å), residue contact order (CO), total number of contacts per residue (NC) and ϕ values for src-SH3.

L29/F33 in helix B; I111/L115 in helix G; M131/A134/L135, I142/Y146 in helix H).

The comparison between the sets of experimental data and depth profiles shows that there is a connection between residue depth in the native structure and key

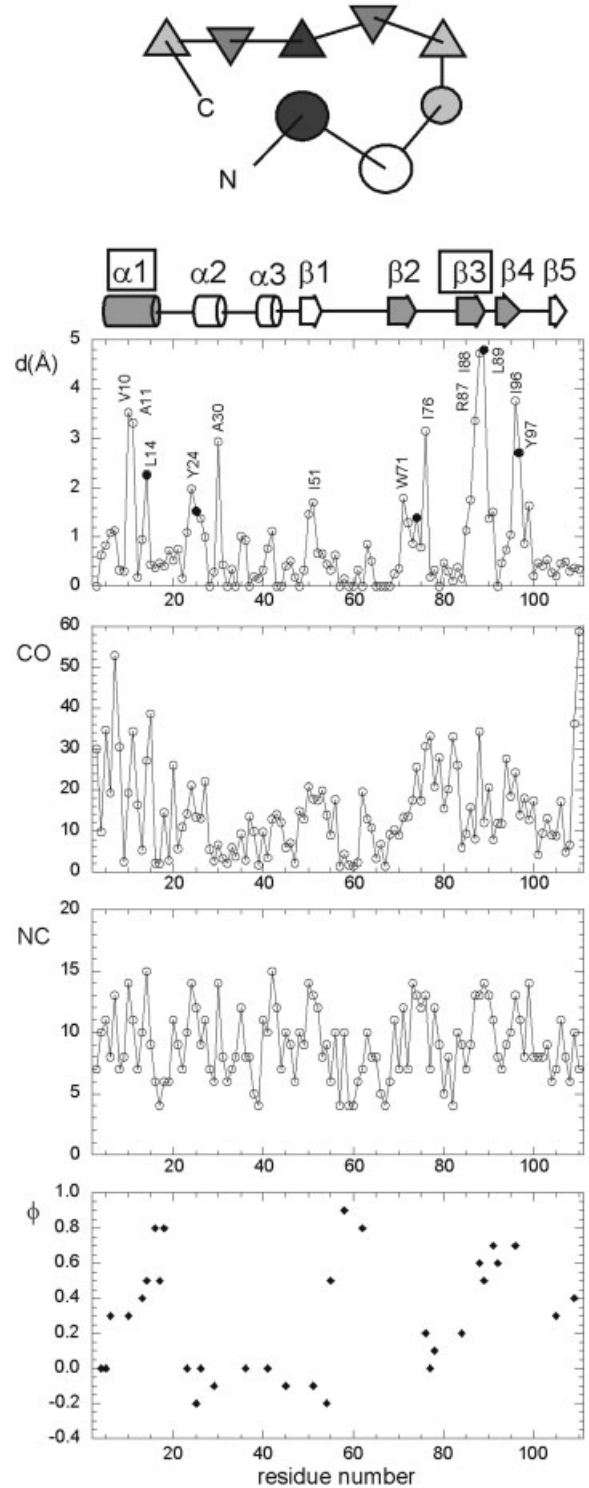


Fig. 3. Topology, secondary structure, mean residue depth (d , Å), residue contact order (CO), total number of contacts per residue (NC) and ϕ values for barnase.

elements in the folding pathway. Residues with high d_{pxr} values participate directly to the folding nucleus or belong to stretches of the polypeptide chain that play an important role in the folding process, as determined by ϕ values

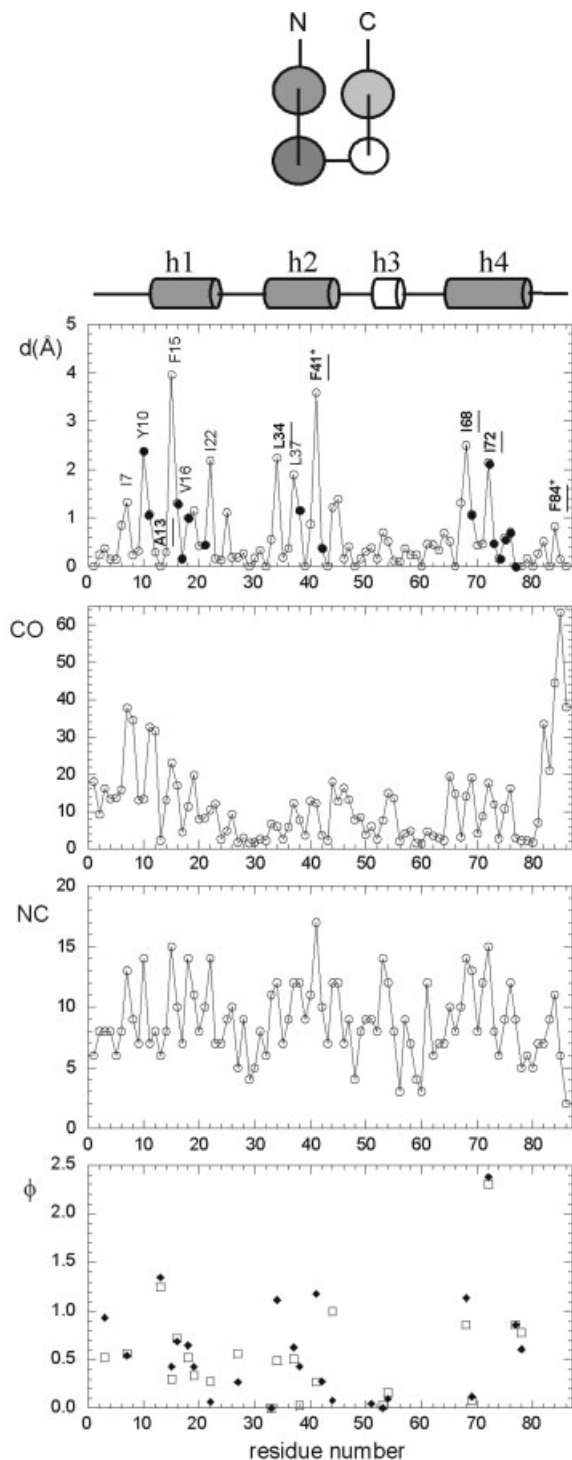


Fig. 4. Topology, secondary structure, mean residue depth (d , Å), residue contact order (CO), total number of contacts per residue (NC) and ϕ values for Im7. Mutations involving residues F41 and F84 (labeled with an asterisk) are leading to an unfolded protein. ϕ values for both the transition state (open box) and the intermediate (filled diamond) are reported.

and H/D exchange data. There is not a unique correspondence, but dpxr profiles are more effective than residue CO and contact profiles in pointing to key residues or regions

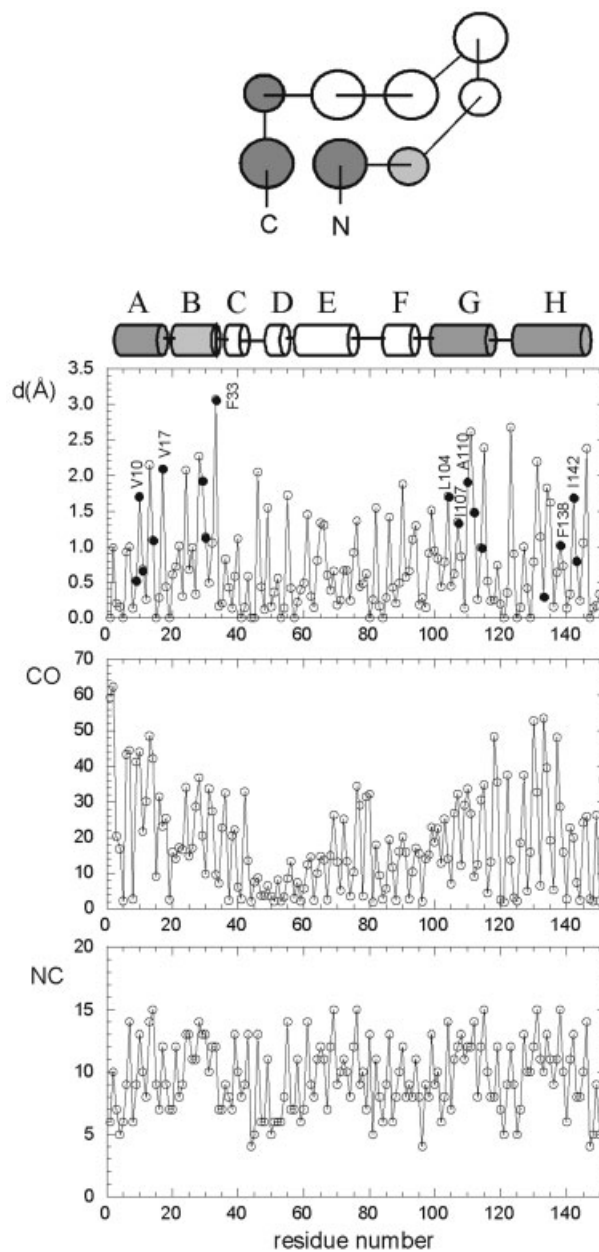


Fig. 5. Topology, secondary structure, mean residue depth (d , Å), residue contact order (CO), total number of contacts per residue (NC) for apomyoglobin.

(Figs. 1–5). Although the overall quantitative correlation between dpxr and ϕ values is usually low, it drastically improves when considering only buried residues (residue solvent accessibility < 20%) and the R values are comparable to those obtained from free energy calculations. For CI2, barnase, Im7, and apomyoglobin there is also a clear hierarchy in the topology of secondary structure elements, as measured by residue depth, and this hierarchy is maintained in folding (Figs. 1–5).

DISCUSSION

The tight packing of hydrophobic residues in the protein interior has been early recognized as one of the main

effects of folding.^{35–37} Accordingly, the hydrophobic surface burial of selected residues has been proposed as one of the driving events for folding,³⁸ residues having a large buried hydrophobic surface and a high number of contacts in the native structure being involved in the formation of the folding nucleus.³⁹ The relationship between several physico-chemical properties of amino acid types and ϕ values for selected proteins has been analyzed.⁴⁰ Depending on the degree of burial of a residue, ϕ values were found to correlate well with general properties like volume or long-range interactions.⁴⁰ Atom depth, defined as the distance of an atom in a protein from its closest solvent accessible neighbor is a geometrical parameter that can be easily calculated from the 3D structure of a protein.¹⁶ Mean residue depth (dpxr) profiles then provide a sensitive and accurate description of the protein interior.¹⁰

Taking the cue from the hypothesis^{12–15} that the slow exchange core is the folding core, we suggest that this could be extended and reformulated as “the most deeply buried residues in the native protein might identify the first parts of the polypeptide chain to fold”. Once the polypeptide chain has collapsed to its native state, major rearrangements of deeply buried residues would be energetically disfavored because they would require extensive unfolding, and the assumption that deeply buried residues are expected to be set in place early should be legitimate.

Different lines of evidence give support to this hypothesis: (1) specific residues display a high degree of conservation within structurally related proteins; there is a good correspondence between these residues and the folding nucleus as determined by ϕ values; (2) the most conserved residues within the same family of structurally aligned proteins are also the most deeply buried in the native structure;¹⁰ (3) residues or pieces of secondary structure elements supposed to belong to the folding nucleus often, albeit not always display high mean residue depth values; (4) slowly H/D exchanging residues in the native structure, or slow exchange core regions often display high mean residue depth values; (5) backbone NHs that are quickly protected in pulse labeling H/D exchange experiments monitored by NMR point to regions that are deeply buried in the native structure; (6) topologically “internal” secondary structure elements display high mean residue depth values, and often high ϕ values and slow exchange rates.

Because of the complexity of the folding process, it is not likely that a single geometrical or physical parameter can describe it effectively. Residues that are deeply buried in the native structure are by no means the only ones that can have a role at the early stages of folding. The local formation of transient or stable hydrogen bonds associated with pieces of secondary structure (turns or helices) and the formation of micro-clusters of hydrophobic residues through local or long range interactions are all events that are likely to occur during the folding process. The contribution of these events to folding will be different along the different folding trajectories and can occur at different times. They will critically depend on the residual structure of the protein in its denatured state and eventually on the amino acid sequence. At least part of these events will not

be rate-limiting steps in the folding process, and will thus escape experimental detection. Moreover, the formation of nonnative structures not retained in the folded protein have been experimentally observed during folding. Depth calculations will then miss the contribution of non-native interactions and under-estimate the role of solvent exposed residues but will spot residues that form the core of the native structure.

When comparing different structural parameters (CO, number of contacts, depth and others) with experimental results, it should be born in mind that even experimental techniques can give only a partial overview of the folding process. The protein engineering approach places its theoretical grounds on a number of assumptions,⁸ and the residue-specific effects are deduced from global folding kinetics, as measured from tryptophan fluorescence quenching of a reporter residue. The experimentally determined ϕ values usually cover only a fraction of the amino acid sequence, they often display a high variability depending on the type of mutation, and sometimes are negative or > 1 , which creates some difficulties in the interpretation. On the other hand, H/D exchange data as determined by NMR provide direct, site-specific information, which is however restricted to backbone NH protons. While there is a correlation between depth and H/D exchange rates, these depend more on hydrogen bond strength and local dynamics, which explains why slowly exchanging NHs are not necessarily deeply buried. As it has been pointed out,^{5,13} the protein engineering and H/D exchange methods are in many aspects complementary. The first is focused on side chains, the second on backbone NHs, and they mainly monitor hydrophobic interactions and hydrogen bonding, respectively. Moreover, the first can be applied in the sub-msec time scale, the latter in the msec time scale and at equilibrium. To date, a residue-specific, experimental measure of depth that could be applied in the msec time scale, to our knowledge is still lacking. At this stage, we propose that residue depth can be used as a simple and efficient tool in the design of folding experiments *in vitro*, and possibly as a geometrical constraint in computer simulations of protein folding.

REFERENCES

1. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998;277:985–994.
2. Baker D. A surprising simplicity to protein folding. *Nature* 2000;405:39–42.
3. Mirny L, Shakhnovich E. Protein folding theory: from lattice to all-atom models. *Annu Rev Biophys Biomol Struct* 2001;30:361–396.
4. Krishna MM, Hoang L, Lin Y, Englander SW. Hydrogen exchange methods to study protein folding. *Methods* 2004;34:51–64.
5. Li R, Woodward C. The hydrogen exchange core and protein folding. *Protein Sci* 1999; 8:1571–1590.
6. Woodward C, Carulla N, Barany G. Native state hydrogen-exchange analysis of protein folding and protein motional domains. *Methods Enzymol* 2004;380:379–400.
7. Konermann L, Simmons DA. Protein-folding kinetics and mechanisms studied by pulse-labeling and mass spectrometry. *Mass Spectrom Rev* 2003;22:1–26.
8. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 1992;224:771–782.

9. Mirny LA, Shakhnovich EI. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol* 1999;291:177–196.
10. Pintar A, Carugo O, Pongor S. Atom depth as a descriptor of the protein interior. *Biophys J* 2003;84:2553–2561.
11. Chakravarty S, Varadarajan R. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure Fold Des* 1999;7:723–732.
12. Woodward C. Is the slow exchange core the protein folding core? *Trends Biochem. Sci* 1993;18:359–360.
13. Woodward C, Li R. The slow-exchange core and protein folding. *Trends Biochem Sci* 1998;23:379.
14. Englander SW. Native-state HX. *Trends Biochem Sci* 1998;23:378.
15. Clarke J, Itzhaki LS, Fersht AR. A reply to Englander and Woodward. *Trends Biochem Sci* 1998;23:379–381.
16. Pintar A, Carugo O, Pongor S. DPX: for the analysis of the protein core. *Bioinformatics* 2003;19:313–314.
17. Ludvigsen S, Shen HY, Kjaer M, Madsen JC, Poulsen FM. Refinement of the three-dimensional solution structure of barley serine proteinase inhibitor 2 and comparison with the structures in crystals. *J Mol Biol* 1991;222:621–635.
18. Xu W, Harrison SC, Eck MJ. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 1997;385:595–602.
19. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 1999;6:1016–1024.
20. Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C, Jack A. Molecular structure of a new family of ribonucleases. *Nature* 1982;297:162–164.
21. Itzhaki LS, Otzen DE, Fersht AR. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 1995;254:260–288.
22. Matouschek A, Serrano L, Fersht AR. The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 1992;224:819–835.
23. Capaldi AP, Kleanthous C, Radford SE. Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* 2002;9:209–216.
24. Grantcharova V, Alm EJ, Baker D, Horwich AL. Mechanisms of protein folding. *Curr Opin Struct Biol* 2001;11:70–82.
25. Brockwell DJ, Smith DA, Radford SE. Protein folding mechanisms: new methods and emerging ideas. *Curr Opin Struct Biol* 2000;10:16–25.
26. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J. Fast kinetics and mechanisms in protein folding. *Annu Rev Biophys Biomol Struct* 2000;29:327–359.
27. Galzitskaya OV, Finkelstein AV. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc Natl Acad Sci USA* 1999;96:11299–11304.
28. Grantcharova VP, Riddle DS, Santiago JV, Baker D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 1998;5:714–720.
29. Yi Q, Bystroff C, Rajagopal P, Klevit RE, Baker D. Prediction and structural characterization of an independently folding substructure in the src SH3 domain. *J Mol Biol* 1998;283:293–300.
30. Grantcharova VP, Baker D. Folding dynamics of the src SH3 domain. *Biochemistry* 1997;36:15685–15692.
31. Yu H, Rosen MK, Schreiber SL. 1H and 15N assignments and secondary structure of the Src SH3 domain. *FEBS Lett* 1993;324:87–92.
32. Matouschek A, Serrano L, Meiering EM, Bycroft M, Fersht AR. The folding of an enzyme. V. H/2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J Mol Biol* 1992;224:837–845.
33. Gorski SA, Le Duff CS, Capaldi AP, Kalverda AP, Beddard GS, Moore GR, Radford SE. Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J Mol Biol* 2004;337:183–193.
34. Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 1993;262:892–896.
35. Richards FM, Lim WA. An analysis of packing in the protein folding problem. *Q Rev Biophys* 1993;26:423–498.
36. Rose GD, Wolfenden R. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu Rev Biophys Biomol Struct* 1993;22:381–415.
37. Levitt M, Gerstein M, Huang E, Subbiah S, Tsai J. Protein folding: the endgame. *Annu Rev Biochem* 1997;66:549–579.
38. Moulton J, Unger R. An analysis of protein folding pathways. *Biochemistry* 1991;30:3816–3824.
39. Serrano L, Matouschek A, Fersht AR. The folding of an enzyme. VI. The folding pathway of barnase: comparison with theoretical models. *J Mol Biol* 1992;224:847–859.
40. Gromiha MM, Selvaraj S. Important amino acid properties for determining the transition state structures of two-state protein mutants. *FEBS Lett* 2002;526:129–134.
41. Flores TP, Moss DS, Thornton JM. An algorithm for automatically generating protein topology cartoons. *Protein Eng* 1994;7:31–37.