Residue depth: a novel parameter for the analysis of protein structure and stability

Suvobrata Chakravarty[1] and Raghavan Varadarajan[1,2]*

Background: Accessible surface area is a parameter that is widely used in analyses of protein structure and stability. Accessible surface area does not, however, distinguish between atoms just below the protein surface and those in the core of the protein. In order to differentiate between such buried residues we describe a computational procedure for calculating the depth of a residue from the protein surface.

Results: Residue depth correlates significantly better than accessibility with effects of mutations on protein stability and on protein–protein interactions. The deepest residues in the native state invariably undergo hydrogen exchange by global unfolding of the protein and are often significantly protected in the corresponding molten-globule states.

Conclusions: Depth is often a more useful gage of residue burial than accessibility. This is probably related to the fact that the protein interior and surrounding solvent differ significantly in polarity and packing density. Hence, the strengths of van der Waals and electrostatic interactions between residues in a protein might be expected to depend on the distance of the residue(s) from the protein surface.

Addresses: [1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India and [2]Chemical Biology Unit, Jawaharlal Nehru Center for Advanced Scientific Research, Jakkur, Bangalore 560004, India.

*Corresponding author.
E-mail: varadar@mbu.iisc.ernet.in

Key words: accessibility, hydrogen exchange, mutations, stability

Received: 23 December 1998
Revisions requested: 15 February 1999
Revisions received: 9 March 1999
Accepted: 22 March 1999

Published: 14 June 1999

Structure July 1999, 7:723–732
http://biomednet.com/elecref/0969212600700723

 ISSN 0969-2126

Introduction

Upon folding of a protein, a large fraction of its residues become inaccessible to solvent. Burial of nonpolar surface area in the protein interior is an energetically favorable process and the hydrophobic driving force is one of the most important determinants of protein structure and stability. Residue burial is conventionally quantitated by a parameter called the solvent-accessible surface area (ASA). The ASA is the area of the surface generated by rolling a spherical probe over the surface of the protein [1]. This area is a function of the radius of the solvent molecule. A typical value for this radius is 1.4 Å, the van der Waals radius of a water molecule. Since its introduction in 1971, ASA has been widely used in the analysis of protein structure, stability and protein–protein interactions [2–7]. ASA, however, does not distinguish between atoms just below the protein surface and those buried deep within the protein interior. The coordinates of protein atoms derived from X-ray crystallography are a good approximation of the mean positions of the atoms in solution. However, in solution atoms can undergo large fluctuations about these positions, and, hence, many atoms that are apparently just below the surface in the static X-ray crystal structure can come transiently into contact with solvent. Such atoms and residues might make a different contribution to protein stability than more deeply buried residues do.

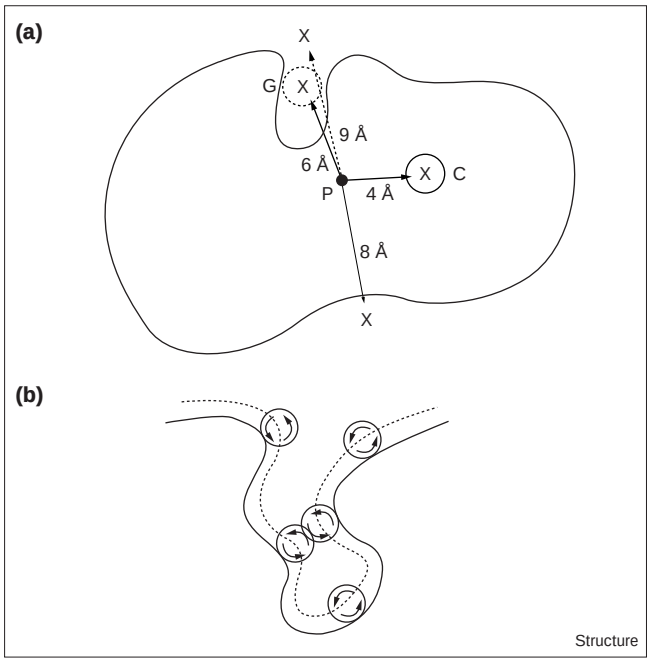
In addition to the hydrophobic effect, van der Waals and electrostatic interactions are also important contributors to protein stability. The packing density and dielectric properties of the protein interior differ significantly from those of surrounding aqueous solvent [8]. Hence, the strengths of van der Waals and electrostatic interactions between residues in a protein might be expected to depend on the distance of the residue(s) from the protein surface. It is therefore desirable to have a parameter that is a measure of the distance of an atom or residue in the molecule from the protein surface. The present work describes a procedure for measuring this parameter (which we call depth) and outlines some useful potential applications. We show that such a parameter correlates significantly better with mutational data on protein stability and protein–protein interactions than accessibility. Depth also correlates better than accessibility with experimental measurements of amide hydrogen exchange rates. We suggest that measurement of protection factors of the deepest residues will be a convenient method for estimating the free energy of folding of a protein in the absence of denaturant.

Results and discussion

Description of the algorithm

We define the depth of an atom in a protein as the distance of the atom from the nearest surface water molecule. Calculation of this distance requires the position of surface water molecules to be known. As surface waters are often disordered, they are not always detected in protein crystal or nuclear magnetic resonance (NMR) structures. Positions of such surface waters are therefore determined as

Figure 1



Schematic representation of depth calculation for an interior protein atom P. (a) Water molecules are represented by X. C and G represent an internal cavity and a surface groove, respectively. Water molecules present in C and G are not considered in depth calculations (see text). The depth of P is 8 Å, the distance to the nearest surface water. (b) A representation of the molecular surface groove G. Although protein atoms lining the groove are all solvent-accessible, two water molecules cannot be simultaneously placed in the groove without a steric clash. The relative sizes of water and protein are not to scale. Water molecules in C and G are not taken into account in depth calculations.

follows: the protein molecule of interest is placed at the center of a pre-equilibrated box of TIP3P (transferable intermolecular potential) water obtained from a Monte Carlo simulation [9]. The protein is then rotated about an axis passing through its center of mass by a random angle and translated along the x axis by a random number n such that $-d \leq n \leq d$ (where $d = 2.8$ Å is the average distance between neighboring water molecules in the box). All water molecules that satisfy either of the following two criteria are removed: A) the water molecule that lies within a distance of 2.6 Å of a protein atom; or B) the water molecule that has less than two neighboring waters within a sphere of radius of 4.2 Å (one and a half hydration shells). The second criterion removes water molecules that occur inside cavities in the protein interior or in surface concavities. It is based on the assumption that cavities found in naturally occurring proteins will not accommodate more than two hypothetical water molecules [10]. In addition, the maximum volumes of cavities determined in a case study of cavity-creating mutants of T4 lysozyme and ribonuclease S (RNase S) [11,12] were of the order of 200 Å^3. This volume corresponds to a sphere of radius of 3.7 Å. Two water molecules separated by 2.8 Å can not be placed inside such a cavity without clashing with surrounding protein atoms. Removal of internal waters is crucial, as internal waters are not taken into consideration in depth calculations. Figure 1a is a cartoon of a protein structure depicting an interior protein atom P. The positions of four hypothetical waters are indicated by the symbol X. The waters in the internal cavity C and in the surface groove G (at 4 Å and 6 Å from P, respectively) are eliminated using criterion B (discussed in more detail below). The closer of the two remaining water molecules located at a distance of 8 Å is the depth defining water for atom P. Figure 1b is a magnified representation of the groove G.

The rotation and translation procedures described above are repeated a number of times in order to approximate the dynamics of a molecule in solution. Each repetition of rotation and translation generates a distinct configuration of waters around the protein. The process of rotation, translation, water removal and depth calculation is carried out for a minimum of ten iterations. The iterations are repeated until the mean depth of each atom averaged over all the iterations approaches a convergent value. Convergence is defined as follows: for every atom in the structure, the mean depth averaged over all the iterations and the standard deviation in depth are calculated. When the coefficient of variance $((SD/mean) \times 100)$ of atom depth is less than 25% for every atom, the calculation is said to have converged. In each of the calculations reported here, 25 iterations were sufficient for convergence. The depth of a residue is the average of the constituent atom depths. For a 100-residue protein, the entire depth calculation takes approximately ten minutes on an IBM RS 6000 340 workstation.

Comparison of residue depths calculated by different methods

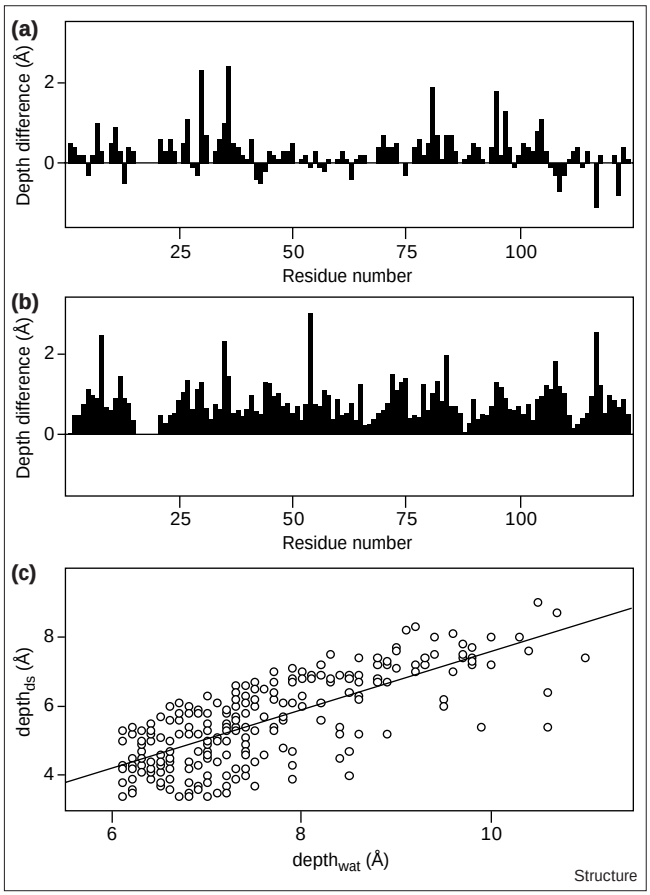
In order to check the accuracy of the depth calculations, we have compared the residue depths calculated using our procedure with structures obtained during the course of a solvated molecular dynamics (MD) simulation. 1000 structures extracted at different time points in the course of a 100 ps simulation [13] were used for the comparison. The simulation was carried out on the 124-residue protein RNase S solvated by about 2400 water molecules. For each protein atom in a structure extracted from the simulation, the distance to the closest water molecule was calculated. This distance averaged over all the structures corresponds to the average depth of the protein atom during the simulation. Depths calculated by our procedure are in close agreement with those obtained from the MD simulation. Very few residues show a difference in depth greater than 2 Å (Figure 2). Residue depths of disordered residues 16–20 are not shown. As a result of the breathing motion of a protein molecule during the course of the MD

simulation, certain regions of the protein become more accessible to water than in the crystal structure. The average difference in residue depths computed by the two methods is 0.4 ± 0.42 Å. Although MD simulations are, in principle, the most accurate method of calculating residue depths for small proteins, such procedures are far more computationally expensive than our procedure.

In principle, the distance of a protein atom from the surface of the protein [1] can also be used as a measure of depth. The molecular surface [8] of the protein has two components. The convex contact surface comprises the part of a van der Waals envelope directly in contact with the probe water molecule. Separated patches of contact surface are connected by concave and saddle-shaped re-entrant surface (inward facing part of the probe sphere when it makes contact with more than two atoms). The molecular surface used in the present work is calculated by the MS (molecular surface) program developed by Connolly [14] using a probe radius of 1.4 Å. The output of the MS program is a set of points evenly distributed on the molecular surface at a dot density of 5.5 dots per Å^2. The distance of the closest surface dot to a given atom was calculated. The probe radius of 1.4 Å was added to this distance for atom-depth definition. The depth of a residue is the average of the constituent atom depths. Residue depths for RNase S computed using our method were compared with MD-simulation depths and with depths calculated using Connolly's dot-surface estimation (Figure 2). Depths calculated using our procedure are in closer agreement with MD-simulation depths than those calculated from the dot-surface procedure. For several residues, depths calculated using our procedure are significantly different from those calculated using the dot-surface method. A regression analysis on residue depths calculated using our method and Connolly's dot-surface method was carried for a set of seven proteins (PDB codes 5icb, 1pcy, 1ruv, 1stn, 2lzm, 1cpn, and 1omp). The correlation coefficient between residue depths is 0.71 for residues deeper than 6 Å (Figure 2c). In several cases there are substantial differences between the two depth estimates.

Although molecular surfaces have proven to be very useful, particularly in studies involving docking and ligand binding [15], the interaction of water with the protein surface is not well described by a rolling sphere. Interactions with water are dictated by highly directional hydrogen bonds. Water molecules cannot pack as tightly against the protein surface as Lennard–Jones spheres, which have no hydrogen-bonding constraint [16]. Figure 1b shows that two water molecules cannot in reality simultaneously occupy the same space in a surface groove. Although there is sufficient space for a single water molecule to be rolled all over the surface of the groove, in solution the solvation of atoms in the groove might differ appreciably from

Figure 2

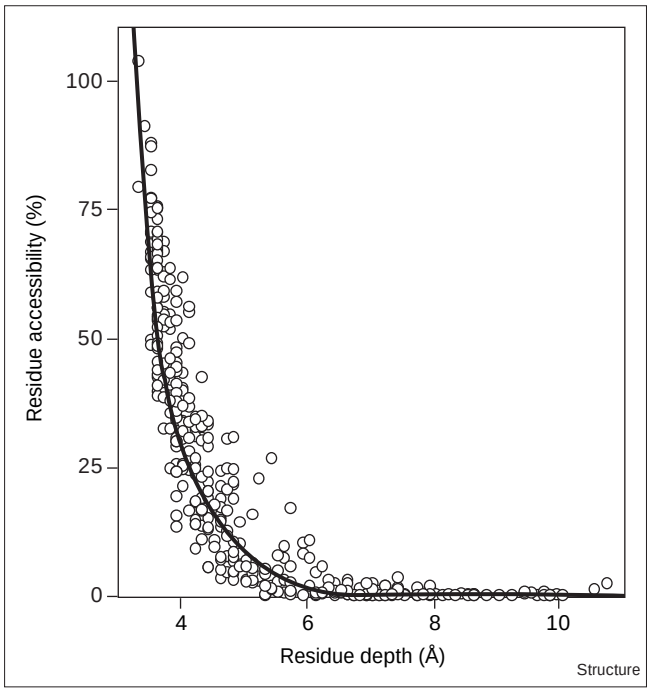


A comparison of residue depths calculated using different procedures. (a) Depths obtained from an MD simulation of RNase S subtracted from depths obtained using our procedure. (b) Depths calculated using a dot surface of RNase S subtracted from depths obtained in our procedure. (c) Correlation between depth$_{wat}$ (depth obtained using surface waters from our procedure) and depth$_{ds}$ (depth calculated using dot-surface procedure).

exposed and convex regions of the protein surface. Use of the molecular dot-surface procedure might lead to an over estimate of the molecular surface actually in contact with solvent. A water molecule X in the groove G at a distance of 6 Å from protein atom P (Figure 1a) might well correspond to a surface dot. In our depth calculations, however, such a water molecule is not considered. Visual inspection using computer graphics shows that most positions in which there are large differences in depths estimated using our procedure and the dot-surface method occur when the depth-defining dot is located in a surface groove. Given that many of these surface grooves are unlikely to be occupied by water in solution, our procedure may provide a more realistic depth estimate than that derived from the dot-surface procedure.

A fourth estimate of depth can be obtained by using water molecules present in high-resolution crystal

Figure 3



Relationship between residue accessibility and depth for maltose-binding protein. The line is a fit of an exponential function to the data.

structures. As an example we have used the 1.3 Å crystal structure of RNase A (PDB code 1ruv). However, even in high resolution crystal structures only a small fraction of surface waters are visible. The average residue depth difference between depths calculated using crystallographic waters and the present procedure (see the Materials and methods section) is 0.3 ± 0.25 Å. Thus, depth calculated using our procedure is in close agreement with values obtained from MD simulations or from high-resolution protein crystal structures.

Correlation of depth with protein size and accessibility

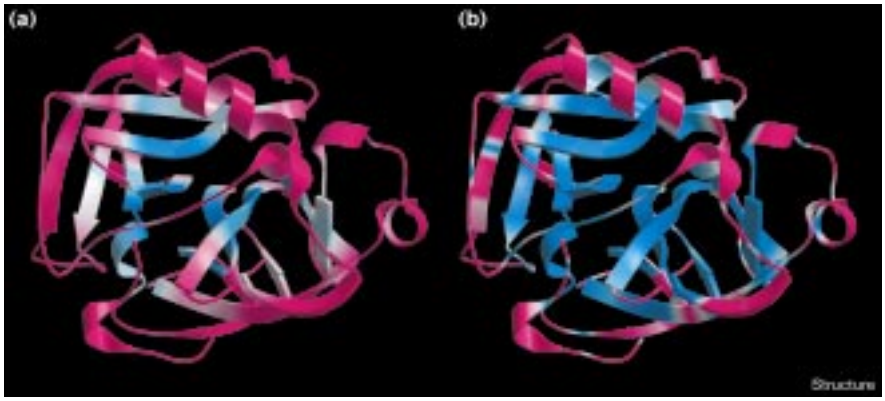
Shown in Figure 3 is a plot of residue accessibility (residue ASA normalized to ASA in peptide Gly–X–Gly) as a function of depth for residues in a 370 amino acid protein, maltose-binding protein. The data fit well to a simple exponential function. Accessibility decreases very steeply with depth up to a depth of about 4 Å. In this depth range, residues with similar depths can differ in accessibility by about 25%. Residues deeper than 6 Å have zero accessibility but differ in depth. Figure 4 shows a ribbon diagram of chymotrypsin colored as a function of either depth (Figure 4a) or accessibility (Figure 4b). The color changes from magenta (surface residues) via white (intermediate) to blue (buried residues). The figure clearly shows that depth can discriminate between buried residues with similar accessibility.

In order to examine how depths of the deepest atoms vary with protein size we performed depth calculations on a set of 65 monomeric and 35 dimeric proteins. Figure 5 shows the average depth of the five deepest atoms in a protein as a function of protein size. For monomeric proteins this number appears to plateau at about 12 Å in the size range of 200–250 residues. This plateau is a result of formation of multiple domains in larger sized proteins, a phenomenon that has been observed previously [17]. In contrast, for dimeric proteins this depth has not reached a plateau even at a size of 900 residues. This may be related to the earlier observation [18] that surface area:volume correlations extend to larger sizes for oligomeric proteins than for monomeric ones. In the following sections we show that buried residues that differ in depth also have differing contributions to protein stability and dynamics.

Analysis of the stability of mutants

To design novel proteins, or rationally alter existing ones, a quantitative understanding of the factors that affect the stability of the native state of protein is required. Protein engineering studies have provided a wealth of information

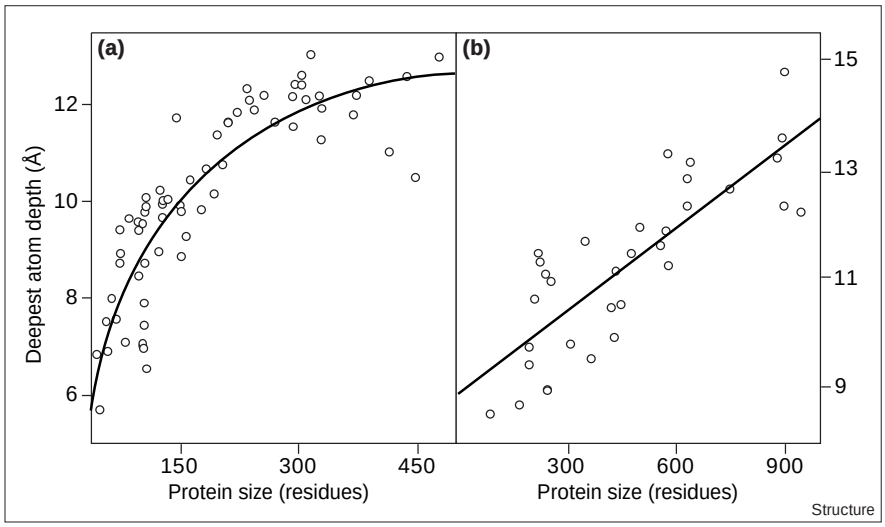
Figure 4



Residue burial in α-chymotrypsin (PDB code 4cha) characterized by (a) depth and (b) accessibility. The color changes from magenta (exposed residues) via white to blue (buried residues). The figure was generated by replacing the B factors in the PDB file with either depths or accessibilities. All residues with accessibility greater than 40% were considered to be fully exposed [42]. The figure was created using MOLSCRIPT [43] and Raster3D [44].

Figure 5

The relationship between the average depths of the deepest atoms and total protein size for (a) monomeric and (b) dimeric proteins. The lines in (a) and (b) represent, respectively, the best fits of hyperbolic and linear functions to the data.



on the relationship between protein structure and stability. Studies on mutations of buried residues [11,19–21] have shown that the packing of nonpolar groups and burial of hydrophobic surface are important factors in the stabilization of proteins. Studies on both fully and partially buried hydrophobic residues in barnase, chymotrypsin inhibitor 2 (CI2), staphylococcal nuclease and FK506-binding protein (FKBP12) have shown a correlation between the change in protein stability on mutation ($\Delta\Delta G_{U-F}$) and the following two parameters: the number of methyl(ene) groups within a certain radius of the nonpolar groups [2–4,20,22]; and the difference between the solvent-accessible surface area that is buried upon folding of the wild-type and the mutant sidechain [2–4].

We have examined the correlation of $\Delta\Delta G_{U-F}$ in a set of cavity-creating mutations in eight proteins with residue depth (Table 1). In this set of proteins we analyze the energetic cost of deleting sidechains of large residues from the wild-type protein that result in the loss of one (Ile→Val, Ala→Gly), two (Val→Ala), three (Leu→Ala, Val→Gly, Met→Ala) and four (Met→Gly, Leu→Gly) methyl(ene) groups from within the core of a protein. The sum of the depths ($depth_{sc}$) of sidechain atoms that are deleted upon mutation are calculated. The number of methyl(ene) groups within 6 Å (nCH_2) of the sidechain atoms deleted upon mutation, as well as $\Delta\Delta$ASA, the difference between the wild-type and mutant sidechain ASA buried upon folding, were calculated. The calculations assumed that, apart from removal of sidechain atoms, the protein structure is not changed upon mutation. All ASA calculations were performed using the algorithm of Lee and Richards [1] with a probe radius of 1.4 Å. Values of $\Delta\Delta G_{U-F}$ show a stronger correlation with $depth_{sc}$ (r = 0.71, slope = –0.14) than with nCH_2 (r = 0.61, slope = –0.05) or $\Delta\Delta$ASA (r = 0.62, slope = –0.04) (Figure 6). This analysis suggests that depth is a useful index for measuring the relative contributions of different buried residues to the thermodynamic stability of the protein. It should be noted, however, that certain contributions to $\Delta\Delta G_{U-F}$ such as secondary-structural propensities and hydrogen-bonding potential will not be correlated with depth.

Analysis of the stability of complexes

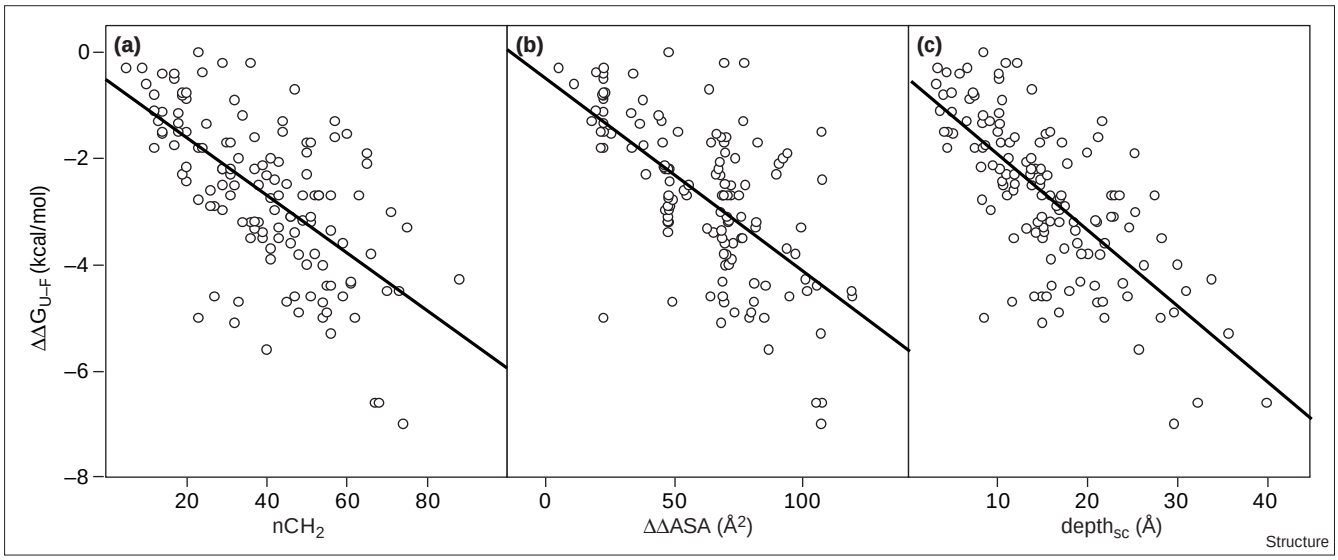
Specific protein–protein interactions are critical events in many biological processes. However, the principles that govern the interaction of two protein surfaces remain elusive [23]. A quantitative understanding of

Table 1

Proteins used for various analyses.

Protein name	PDB code	Reference
Cavity-creating mutations		
Staphylococcal nuclease	1stn	[20,45,46]
T4 Lysozyme	2lzm	[11,47–49]
Barnase	2a2p	[2]
Calbindin D9K	5icb	[50]
FK506-binding protein, FKBP12	1fkd	[22]
Chymotrypsin inhibitor 2	2ci2	[3,4,51]
Human lysozyme	1lz1	[52]
RNase S	1rnu	[53]
Interface mutations		
BPTI–chymotrypsin	1cbw, 5pti, 4cha	[7]
Barnase-barstar	2brs, 2a2p, 1a19	[7]
RNase inhibitor–RNase A	1dfj, 2bnh, 1ruv	[7]
Protein A–IgG1	1fc2, 1bdd	[7]
D1.3–HEW lysozyme	1vfb	[7]
D1.3–E5.2	1dvf	[7]
HyHEL-10–HEW lysozyme	3hfm	[7]
hHG–hGHbp	3hhr, 1hgu	[7]
Tissue factor–factor VIIa	1dan, 2hft	[7]

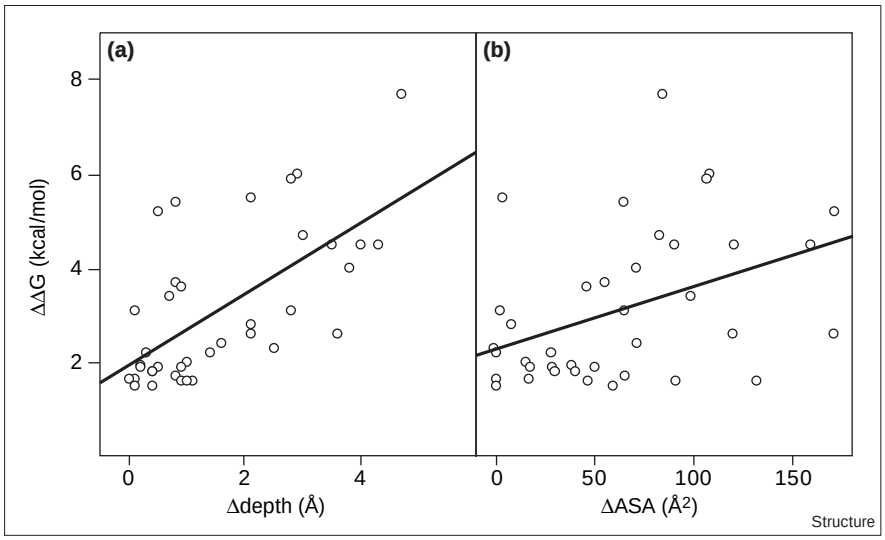
Figure 6



The effect of cavity-creating mutations on the stability of proteins. The observed change in ΔG_{U-F} of folding upon deleting sidechain atoms from the wild-type protein ($\Delta\Delta G_{U-F}$) is plotted versus various structural parameters. **(a)** nCH_2, the total number of methyl(ene) sidechain groups within 6 Å of the sidechain atoms that are deleted upon mutation. **(b)** ΔΔASA, the difference in the solvent-accessible surface area that is buried upon folding between the wild-type and the mutant sidechain. **(c)** $depth_{sc}$, the sum of the depths of sidechain atoms that are deleted upon mutation.

protein–protein interactions requires accurate structural and thermodynamic information. With the recent explosion in co-crystal structures of protein heterodimers and mutational studies of a very large number of protein complexes, there is extensive experimental information about contributions of individual residues to binding. Unfortunately, even in complexes of known structure it has not been possible to deduce the relative contributions of different interface residues to the stability of the complex solely from the structure. Most interfaces are composed of two relatively large protein surfaces with good shape and charge complementarity for one another [6,23,24]. The total ASA buried is typically in the range of 600–1200 Å^2 per subunit [5,6] and it is often assumed that the energy of protein–protein binding is directly related to total buried hydrophobic surface area [5,6]. However, this assumption gives little insight about contributions of individual residues to complex stability. In order to study

Figure 7



Analyses of mutations at the interface of protein–protein complexes. The dependence of free energy of destabilization of the complex (ΔΔG) on **(a)** Δdepth (change in residue depth upon complex formation) and **(b)** ΔASA (change in residue solvent-accessible surface area upon complex formation).

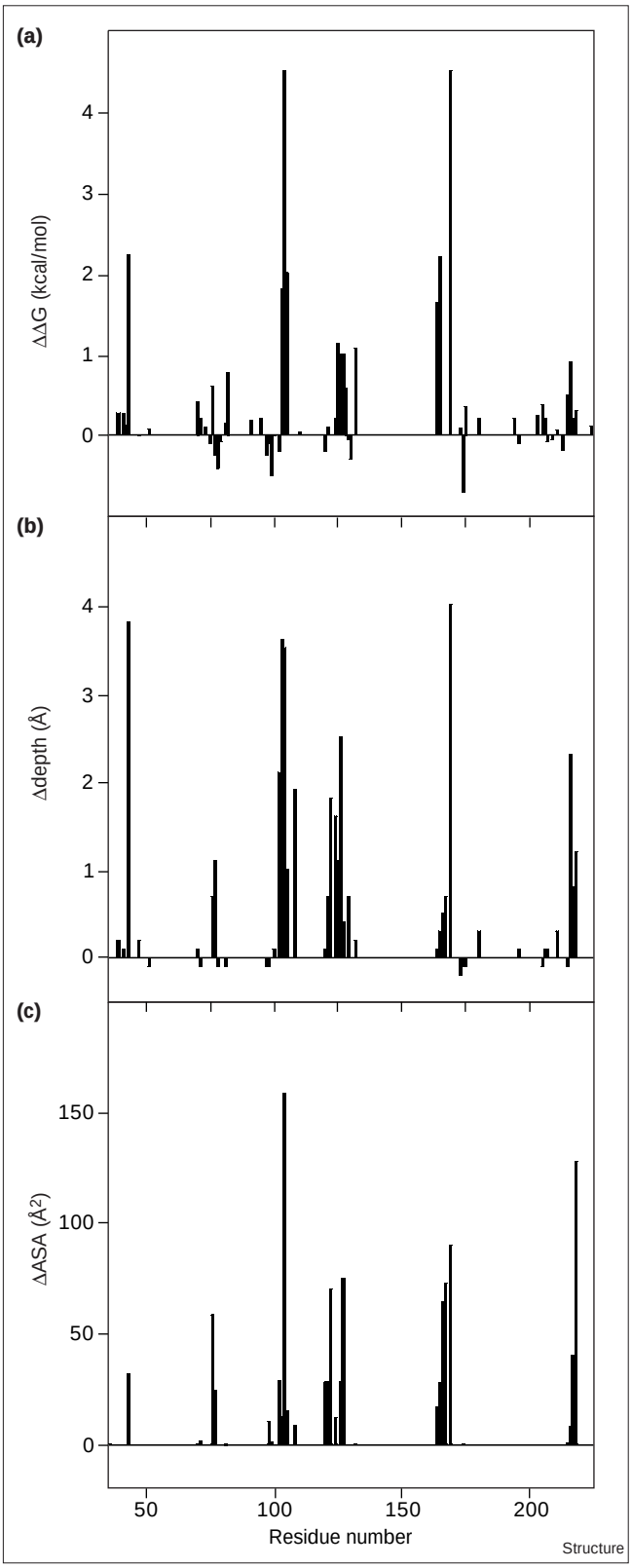


Figure 8

Analyses of human growth hormone binding protein mutations at the human growth hormone–human growth hormone binding protein interface. The bars in (a), (b) and (c) show ΔΔG (energy of destabilization), Δdepth (change in the residue depth upon complex formation) and ΔASA (change in the residue solvent-accessible surface area upon complex formation), respectively, along the sequence of the protein.

protein–protein association at the level of individual residues, we used residue depth to analyze a database of alanine mutants of nine heterodimeric protein–protein complexes for which affinity measurements and structural data are available [7]. The interfaces included in this study are listed in Table 1. We show that residues that contribute significantly to complex formation can be identified using the depth index. We selected only those mutations that alter the free energy of complex formation by more than 1.5 kcal/mol ($\Delta\Delta G < -1.5$ kcal/mol). Depth calculations were performed on a protein complex as well as on each of the individual members independently. The increase in depth of each residue (Δdepth) upon complex formation was calculated. ΔASA, the accessible area of each residue that is lost upon complex formation was also calculated. The change in the free energy of association shows a stronger correlation with Δdepth ($r = 0.68$, slope = 0.75) than with ΔASA ($r = 0.43$, slope = 0.01) (Figure 7). The data indicate that depth changes can be used to quantify the 'O ring model' of Bogan and Thorn [7]. Δdepth is particularly useful for identifying residues that have the largest contributions to ΔΔG. Data for binding of human growth hormone binding protein (hGHbp) to human growth hormone (hGH) illustrate this point (Figure 8). Extensive mutagenesis studies had earlier shown the difficulty in predicting residues contributing to binding even after the structure of the complex had been determined [7]. The two residues of hGHbp that make the greatest contributions to binding hGH are W104 and W169 [25]. Both these residues have large values of Δdepth but only one of them (W104) has a large value of ΔASA. All four of the residues with large depth changes (R43, I103, W104, W169) contribute significantly to binding.

Correlation of amide exchange rates with depth

Hydrogen exchange [26] of backbone amide protons in proteins is a powerful and sensitive method for studying protein stability, dynamics and folding [27]. Hydrogen exchange in proteins is typically believed to take place in two steps [28]. The protein first unfolds by either a local or a global unfolding event to an exchange-competent state. The second step involves the actual exchange in this unfolded state. Under appropriate conditions, measurements of exchange rates can provide information about the free energy difference, ΔG_{ex}, between the folded and exchange-competent states [29–33].

We have analyzed a set of eight proteins (Table 2), for which extensive exchange-rate data are available, to establish a correlation between the backbone amide proton exchange rate of a residue and its depth. The dependence

Table 2

Proteins used for equilibrium proton-exchange analysis.

Protein name	Experimental condition	r_{depth}*	r_{Acc}†	PDB code	Reference
HEW lysozyme	pH 7.5, 30°C	0.60	0.32	193l	[54]
E. coli thioredoxin	pH 6.1, 25°C	0.40	0.27	2trx	[55]
Calbindin D9K	pH 7.5, 30°C	0.72	0.55	1clb	[56]
Pseudomonas aeruginosa cytochrome *c*-551	pH 7.0, 27°C	0.66	0.53	351c	[57]
Ribonuclease T1	pH 7.4, 35°C	0.70	0.30	1bvi	[58]
Ribonuclease A	pH 5,6, 25°C	0.60	0.45	1ruv	[59]
Anti-digoxin VL domain	pH 5.5, 22°C	0.60	0.41	1maj	[60]
Equine lysozyme	pH 4.5, 30°C	0.63	0.36	1eql	[39]

*r_{depth} is the correlation coefficient between the log(amide proton exchange rate constant) and residue amide depth. †r_{Acc} is the correlation coefficient between the log(amide proton exchange rate constant) and residue accessibility.

of exchange rates on amide proton depth varies from protein to protein, probably because of differences in the mechanism of exchange. However, in all the proteins examined the rates of amide exchange correlate much better with depth than with the accessibility of the residue. It is physically reasonable to expect that the deepest amide protons in the protein will exchange by global unfolding. In several proteins, amide protons exchanging by global exchange have previously been identified. Analysis of this data confirmed that the deepest protons exchange via a global unfolding event.

In hen egg white (HEW) lysozyme it was shown that exchange rates of residues undergoing exchange by global unfolding are drastically affected by cross-linking of the protein [34]. Four of the five deepest residues (W28, A31, V29, A32) with amide proton depths ranging from 8 to 9 Å undergo exchange via global unfolding. In three other proteins, residues undergoing exchange by global unfolding were identified from the observation that ΔG_{ex} was equal to values of ΔG_{U-F}, the free energy of unfolding obtained from conventional denaturation studies. All three protons (A74, I88, L89) of the deepest residues of barnase [29,33], all four (L94, I95, L68, L98) of the deepest residues of yeast iso-1-ferricytochrome [35] and three of the four deepest residues (N100, Y91, A90) of staphylococcal nuclease [31] all appear to undergo exchange by global unfolding. No exchange-rate data were reported for one of the deepest residues in either HEW lysozyme (I55) [34] or staphylococcal nuclease (I92) [31]. These observations led us to believe that residues that are within 1 Å of the deepest residue typically exchange by a global-unfolding mechanism. This is an important result as protection factors of the deepest residues measured under native conditions can henceforth be used to obtain a lower estimate of ΔG_{U-F}. This will be especially useful in cases where denaturant or temperature-induced denaturation is irreversible. Such a situation often arises for proteins from hyperthermophilic organisms.

We have also examined the correlation between residue depths in the native state and protection factors in corresponding molten-globule states of a protein. The four proteins analyzed were β-lactoglobulin (PDB code 1beb), guinea pig and human α-lacatalbumins (1hfx, 1alc) and equine lysozyme (1eql) [36–39]. In each of these proteins, hydrophobic core residues were defined using the following criterion: residues within 1 Å of the deepest residue were selected, and additional residues that were in van der Waals contact with these selected residues were clustered to find distinct hydrophobic cores. It was found that 40%–60% of such core residues have amide protons with protection factors greater than or close to 100. This suggests that native-state hydrophobic cores identified by depth criteria in many cases resemble hydrophobic clusters found in corresponding molten globules and implies that molten globules retain some characteristics of native topology.

Biological implications

Burial of hydrophobic surface area is an important driving force for protein folding and macromolecular recognition. Residue burial is typically quantitated in terms of the magnitude of surface area that becomes inaccessible to solvent (ΔASA). Given a protein or protein complex of known three dimensional structure, it would be highly desirable to identify the subset of residues that make the largest contributions to the stability of the molecule. In the present work we show that the depth of a residue from the protein surface correlates significantly better than ΔASA with its energetic contribution to stability. Residues that undergo a large change in depth upon complex formation also contribute significantly to the stability of the complex. The depth parameter correctly identified the two residues making the largest contributions to the stability of a complex of human growth hormone and its binding protein. Depth is probably a better predictor than solvent-accessible surface area because a protein and the aqueous solution surrounding it have significantly different packing densities and dielectric properties. Hence, the strengths of van der Waals and electrostatic interactions between residues in a protein might be expected to depend on the distance of the residue(s) from the protein surface. A

simple procedure for calculating depth is described. In several cases, the deepest residues in the native state of protein are amongst the few residues protected from hydrogen exchange in the corresponding molten-globule intermediate states of the same protein. This suggests that native-like hydrophobic cores involving the deepest residues may also be present in molten-globule folding intermediates. Measurements of the exchange rates of the deepest residues can be used to estimate the stability of the protein under physiological conditions in the absence of denaturant. Depth calculations can serve as a useful guide for experimentalists interested in altering and measuring protein stability and strengths of protein–protein interactions or in analyses of protein folding and dynamics.

Materials and methods

Depth calculations were performed using a nonhomologous dataset consisting of 65 monomeric and 35 dimeric proteins. The remaining proteins listed in Table 1 and 2 were chosen on the basis of previously published studies on protein stability, binding affinity and hydrogen exchange. All the proteins were taken from the Brookhaven Protein Data Bank [40].

The water-bath coordinates were derived from the AMBER 4.1 suite of programs [9]. The original bath consisted of a $37 \times 37 \times 37$ Å^3 box of waters. The size of the actual bath used was determined by the size of proteins used for depth calculations. For monomeric proteins with a maximum size of 500 residues a bath of $111 \times 111 \times 111$ Å^3 was used, whereas for dimers with a maximum size of 900 residues a bath of $185 \times 185 \times 185$ Å^3 was used. These were derived by appropriately translating the original bath. The sidechain solvent-accessible surface area (ASA) was calculated using an implementation of the Lee and Richards [1] algorithm with a probe radius of 1.4 Å and a z section of 0.05 Å. Dot-surface calculations were performed using the MS package [14]. In both ASA and dot-surface calculations a probe of radius of 1.4 Å and atomic van der Waals radii taken from Chothia [41] were used. For depth calculated using the high-resolution crystal structure of ribonuclease A (1ruv), the number of crystallographic water molecules (130) was insufficient to solvate the protein. Therefore, additional water molecules from the water bath were used to fully solvate the protein. Depth comparisons were performed only for residues whose depths were defined by crystallographic water molecules. Residue accessibility was calculated as the ratio of the observed ASA of a residue to its ASA in a Gly–X–Gly peptide of extended conformation. Each of the proton-exchange correlations in Table 2 used the following numbers of amide protons: 36 (ribonuclease T1), 44 (ribonuclease A), 51 (calbindin D9K), 61 (anti-digoxin V_L domain), 63 (HEW lysozyme), 66 (equine lysozyme), 72 (cytochrome c-551) and 99 (thioredoxin).

Thermodynamic data on protein–protein interactions were obtained from the database located at *http://motorhead.ucsf.edu/~thorn/hotspot* [7]. The depth program is written in Fortran 77 and is available from the authors upon request. The program will perform depth calculations either using the Monte Carlo procedure outlined in this work or from the Connolly dot-surface file if this is available.

Acknowledgements

We thank Dr Gautham Nadig for helpful discussions, for initial ideas and implementation of the depth algorithm and for the coordinates of the RNase S MD simulations. We thank Dr N Srinivasan and Dr R Sowdhamini for useful suggestions. We acknowledge the SERC and Bioinformatics and Interactive Graphics Facility at the Indian Institute of Science for use of computational and computer graphics facilities. This work was supported by grants from DBT and DST to RV.

References

1. Lee, B. & Richards, F.M. (1971). The interpretation of protein structure: Estimation of static accessibility. *J. Mol. Biol.* 55, 379-400
2. Serrano, L., Kellis, J.T, Jr, Cann, P., Matouschek, A. & Fersht, A.R. (1992). The folding of an enzyme II. Substructure of Barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* 224, 783-804.
3. Jackson, S.E., Moracci, M., elMasry, N., Johnson, C.M. & Fersht, A.R. (1993). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* 32, 11259-11269
4. Otzen, D.E., Rheinnecker, M. & Fersht, A.R. (1995). Structural factors contributing to the hydrophobic effect: the partly exposed hydrophobic minicore in chymotrypsin inhibitor 2. *Biochemistry* 34, 13051-13058.
5. Janin, J. & Chothia, C. (1990). The structure of protein-protein recognition sites. *J. Biol. Chem.* 265, 16027-16030.
6. Jones, S. & Thornton, J.M. (1996). Principles of protein-protein interactions. *Proc. Natl Acad. Sci., USA* 93, 13-20.
7. Bogan, A.A. & Thorn, K.S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280, 1-9.
8. Richards, F.M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6, 151-176
9. Pearlman, D.A., *et al*., & Kollman, P.A. (1995). AMBER 4.1. University of California, San Francisco, USA.
10. Hubbard, S.J., Gross, K.H. & Argos, P. (1994). Intramolecular cavities in globular proteins. *Prot. Eng.* 7, 613-626
11. Eriksson, A.E., *et al*., & Matthews, W.B. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 225, 178-183.
12. Varadarajan, R. & Richards F.M. (1992). Crystallographic structures of ribonuclease S variants with nonpolar substitution at position 13: packing and cavities. *Biochemistry* 31, 12315-12327.
13. Nadig, G., Ratnaparkhi, G.S., Varadarajan, R. & Vishveshwara, S. (1996). Dynamics of ribonuclease A and ribonuclease S: computational and experimental studies. *Prot. Sci.* 5, 2104-2114.
14. Connolly, M. (1993). The molecular surface package. *J. Mol. Graph.* 11, 139-141.
15. Shoichet, B.K. & Kuntz, I.D. (1991). Protein docking and complementarity. *J. Mol. Biol.* 221, 327-346.
16. Gerstein, M. & Lynden-Bell, R.M. (1993). What is the natural boundary for a protein in solution? *J. Mol. Biol.* 230, 641-650.
17. Rose G.D. (1979). Hierarchic organization of domains in globular proteins. *J. Mol. Biol.* 134, 447-470.
18. Janin, J., Miller, S. & Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* 204, 155-164.
19. Kellis, J.T, Jr, Nyberg, K., Sali, D. & Fersht, A.R. (1988). Contribution of hydrophobic interactions to protein stability. *Nature* 333, 784-786.
20. Shortle, D., Stites, W.E. & Meeker, A.K. (1990). Contribution of the large hydrophobic amino acids to the stability of Staphylococcal Nuclease. *Biochemistry* 29, 8033-8041
21. Lim, W.A., Farruggio, D.C. & Sauer, R.T. (1992). Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* 31, 4324-4333.
22. Main, E.R.G., Fulton, K.F. & Jackson, S.E. (1998). Context-dependent nature of destabilizing mutations on the stability of FKBP12. *Biochemistry* 37, 6145-6153.
23. Janin, J. (1995). Elusive affinities. *Proteins* 21, 30-39.
24. Lawrence, M.C. & Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946-950.
25. Clackson, T. & Wells, J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science* 267, 383-386.
26. Lingderstrøm-Lang, K.U. (1955). Deuterium exchange between peptides and water. *Spec. Publ. Chem. Soc.* 2, 1-20.
27. Udgaonkar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathways of ribonuclease A. *Nature* 335, 694-699
28. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. (1993). Protein stability parameters measured by hydrogen exchange. *Proteins* 17, 75-86.
29. Clarke, J., Hounslow, A.M., Bycroft, M. & Fersht, A.R. (1993). Local breathing and global unfolding in hydrogen exchange of barnase and its relationship to protein folding pathways. *Proc. Natl Acad. Sci. USA* 90, 9837-9841.
30. Kim, K.S., Fuchs, J.A. & Woodward, C.K. (1993). Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry* 32, 9600-9608.

31. Loh, S.N., Prehoda, K.E., Wang, J. & Markley, J.L. (1993). Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry* 32, 11022-11028.
32. Mayo, S.L. & Baldwin, R.L. (1993). Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* 262, 873-876.
33. Perrett, S., Clarke, J., Hounslow, A.M. & Fersht, A.R. (1995). Relationship between equilibrium amide proton exchange behavior and the folding pathway of barnase. *Biochemistry* 34, 9288-9298.
34. Noda, Y., Fukuda, Y. & Segawa, S. (1997). A two-dimensional NMR study of exchange behavior of amide hydrogens in a lysozyme derivative with an extra cross-link between Glu35 and Trp108 M – quenching of cooperative fluctuations and effects on the protein stability. *Biopolymers* 41, 131-143.
35. Bentz, S.F., Marmorino, J.L., Saunders, A.J., Doyle, D.F., Young, G.B. & Peilak, G.J. (1996). Unusual effects of an engineered disulfide on global and local protein stability. *Biochemistry* 35, 7422-7428
36. Ragona, L., Pusterla, F., Zetta, L., Monaco, H.L. & Molinari, H. (1997). Identification of a conserved hydrophobic cluster in partially folded bovine beta-lactoglobulin at pH 2. *Fold. Des*. 2, 281-290.
37. Chyan, C., Wormald, C., Dobson, C.M., Evans, P. & Baum, J. (1993). Structure and stability of the molten globule state of guinea-pig α-lactalbumin: A hydrogen exchange study. *Biochemistry* 32, 5681-5691.
38. Schulman, B.A., Redfield, C., Peng, Z., Dobson, C.M. & Kim, P.S. (1995). Different subdomains are most protected from hydrogen exchange in molten globule and native states of human α-Lactalbumin. *J. Mol. Biol.* 253, 651-657.
39. Morozova-Roche, L.A., Aciro-Muendel, C.C., Haynie, D.T., Emelyanenko, V.I., Dael, H.V. & Dobson, C.M. (1997). Structural characterization and composition of the native and A-state of equine Lysozyme. *J. Mol. Biol.* 268, 903-921.
40. Berstein, F.C., *et al*., & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 122, 535-542.
41. Chothia, C. (1976). The nature of the accessible and buried surface in proteins. *J. Mol. Biol.* 105, 1-14.
42. Holbrook, S.R., Muskal, S.M., & Kim, S.H. (1990). Predicting surface exposure of amino acids from protein sequence. *Prot. Eng.* 3, 659-665.
43. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.
44. Merrit, E.A. & Murphy, M.E.P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* 50, 869-873.
45. Green, S.M., Meeker, A.K. & Shortle, D. (1992).Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state. *Biochemistry* 31, 5717-5728.
46. Green, S.M., & Shortle, D. (1993). Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease. *Biochemistry* 32,10131-10139.
47. Matsumura, M., Becktel, W. J. & Matthews, B.W. (1988). Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* 334, 406-410.
48. Eriksson, A.E., Baase, W.A. & Matthews, B.W. (1993). Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* 229, 747-769.
49. Xu, J., Basse, W.A., Baldwin, E. & Matthews, B.W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Prot. Sci.* 4, 158-177.
50. Julenius K., Thulin, E., Linse, S. & Finn, B.E. (1998). Hydrophobic core substitution in Calbindin D9k: effects on stability and structure. *Biochemistry* 37, 8915-8925
51. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. (1995). 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, 260-288
52. Takano, K., *et al*., & Yutani, K. (1995). Contribution of hydrophobic residues to the stability of human lysozyme: calorimetric studies and X-ray structural analysis of the five isoleucine to valine mutants. *J. Mol. Biol.* 254, 62-76.
53. Connolly, P.R., Varadarajan, R., Sturtevant, J.M. & Richards, F.M. (1990). Thermodynamics of protein-peptide interactions in the ribonuclease S system studies by titration calorimetry. *Biochemistry* 29, 6108-6114.
54. Radford, S.E., Buck, M., Topping, K.D., Dobson, C.M. & Evans, P.A. (1992). Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins* 14, 237-248.
55. Jeng, M. & Dyson, H.J. (1995). Comparison of the hydrogen-exchange behavior of reduced and oxidized *Escherichia coli* thioredoxin. *Biochemistry* 34, 611-619.
56. Skelton, N.J., Kordel, J., Akke, M. & Chazin, W.J. (1992). Nuclear magnetic studies of the internal dynamics in apo, $(Cd^{2+})_1$ and $(Ca^{2+})_2$ calbindin D_{9k}. *J. Mol. Biol.* 227, 1100-1117.
57. Timkovich, R., Walker II, L.A. & Cai, M. (1992). Hydrogen exchange in *Pseudomonas* cytochrome *c*-551. *Biochim. Biophys. Acta* 1121, 8-15.
58. Mullins, L.S., Pace, C.N. & Raushel, F.M. (1997). Conformational stability of ribonuclease T1 determined by hydrogen-exchange. *Prot. Sci.* 6, 1387-1395.
59. Wang, A., Robertson, A.D. & Bolen, D.W. (1995). Effects of a naturally occurring compatible osmolyte on the internal dynamics of ribonuclease A. *Biochemistry* 34, 15096-15104.
60. Constantine, K.L., Friedrichs, M.S., Goldfarb, V., Jeffrey, P.D., Sheriff, S. & Mueller, L. (1993). Characterization of the backbone dynamics of an anti-digoxin antibody V_L domain by inverse detected ^1H-^{15}N NMR: comparison with X-ray data for Fab. *Proteins* 15, 290-311.

Because *Structure with Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed (accessed from http://biomednet.com/cbiology/str). For further information, see the explanation on the contents page.