Volume changes on protein folding*

Yehouda Harpaz[1], Mark Gerstein[2,3] and Cyrus Chothia[1,2]

[1]Cambridge Centre for Protein Engineering, Hills Road, Cambridge, CB2 2QH, UK, [2]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK and [3]Beckman Laboratories for Structural Biology, Department of Cell Biology, Stanford Medical School, Stanford, CA 94305, USA

Background: Protein volumes change very little on folding at low pressure, but at high pressure the unfolded state is more compact. So far, the molecular origins of this behaviour have not been explained: it is the opposite of that expected from the model of the hydrophobic effect based on the transfer of non-polar solutes from water to organic solvent.

Results: We redetermined the mean volumes occupied by residues in the interior of proteins. The new residue volumes are smaller than those given by previous calculations which were based on much more limited data. They show that the packing density in protein interiors is exceptionally high. Comparison of the volumes that residues occupy in proteins with those they occupy in solution shows that aliphatic groups have smaller volumes in protein interiors than in solution, while peptide and charged groups have larger volumes. The cancellation of these volume changes is the reason that the net change on folding is very small.

Conclusions: The exceptionally high density of the protein interior shown here implies that packing forces play a more important role in protein stability than has been believed hitherto.

Structure 15 July 1994, **2**:641–649

Key words: models for protein stability, packing density in proteins

Introduction

The volume changes that occur on protein folding are directly related to the forces responsible for the stability of proteins. They are the result of the differences in the molecular interactions that occur within folded proteins and those that occur between unfolded proteins and water. More than 20 years ago, several groups determined experimentally the extent of volume changes that occur on folding [1–3]. These experiments were carried out to test the hydrophobic model for protein stability [4]: the results were not those expected.

It was found that on protein folding the volume changes are very small (<0.5%) at normal pressures but large and positive at high pressures — i.e. at low pressures the folded and unfolded states occupy the same volume but at high pressure the unfolded state is more compact. Models of the hydrophobic effect, based on the transfer of non-polar solutes from water to organic solvent, predict that the volume changes on folding would be positive at low pressure and negative at high pressure — i.e. the unfolded state should be more compact at low pressure and less compact at high pressure [4].

A consequence of the failure of the solution transfer models to account for the observed volume changes was that the molecular mechanisms that underlay the changes were not understood [1–3]. However, as Kauzmann has argued [5], for any model of protein stability to be successful it must account for these observations.

Soon after these experiments, Richards introduced a procedure for determining directly the volumes of residues buried in the interior of protein structures [6]. Its application to the early protein crystal structures showed clearly that their interiors are close-packed as in crystals of amino acids [6,7]. Though this discovery has been important for subsequent calculations and experiments on protein stability it raised a new problem: the total protein volumes given by these calculations were significantly larger than the volumes determined by solution experiments [6].

Here we provide an explanation for the volume changes on protein folding. This explanation is based upon a redetermination of the volumes residues occupy in folded proteins (which demonstrates that they are smaller than previously thought) and a comparison of these new volumes with the volumes residues occupy in solution.

Results and discussion

Volumes of residues in folded proteins

Redetermination of the mean volumes of residues in protein interiors

The current values for the mean volumes of residues buried in protein interiors were determined from the atomic coordinates of 15 proteins (the set of structures that was available at the time) [7]. Much more accurate and extensive crystallographic data are now available and we used these to redetermine the volumes of buried residues.

We initially took 119 different proteins from the protein structure data bank [8]. All had different sequences, were determined at high resolution (between 1.0 Å and

*This paper is dedicated to John Edsall on the sixtieth anniversary of his paper with Cohn, McMeekin and Blanchard [14] on the volumes of residues in solution.

1.9 Å), had been refined to R-factors of 20 % or less, and had good sterochemistry [9]. Four proteins were found to have no completely buried residues and seven had sequence identities of between 67 % and 90 % with other members of the set. These 11 proteins were removed from our set and calculations carried out on the remaining 108. Of these, 74 had no significant sequence identity with each other, 25 had identities of between 17 % and 45 % with one of the 74 and 9 had identities of between 45 % and 63 %. (A list of these structures is available on request from the authors.)

The volumes of atoms buried in the interior of the protein structures were calculated by the Voronoi polyhedra procedure [6,7,10] (Fig. 1) using a computer program based on the one originally written by Richards [6]. In addition, the accessible surface areas of the residues in each protein were calculated using the procedure described by Lee and Richards [11]. From these data, we determined the volumes of all residues that are completely buried, i.e. those that have no surface area accessible to solvent. We give in Table 1, for each residue type, the number of residues found buried in our set of proteins (N), the mean volume (in $Å^3$), the standard error of the mean volume ($\sigma/\sqrt{N}$) and the standard deviation of the observed volumes (σ) in cubic Å and as a percentage of the mean.

The new mean residue volumes are smaller than those determined in the previous work [7]: values for individual residue types are smaller by between 1 % and 10 %, and by 3.7 % on average. The aliphatic and sulphur-containing residues, alanine, valine, leucine, isoleucine, methionine and cysteine are smaller by 1–2.5 %; the aromatic residues, histidine, phenylalanine, tyrosine and tryptophan, and the small residues, glycine, alanine, serine and threonine by 2–5 % and the charged and amide residues, aspartate, asparagine, glutamate, glutamine and arginine by 4–10 %.

The new and old calculations differ in the accuracy and number of the protein structures used in the calculations. They also differ in the definition of buried residues: here, buried residues are those with no surface accessible to water; previously, they had 0–5 % of their surface exposed to water. When our present calculations are repeated using the old definition for buried residues, the mean volumes are 2.5 % larger. This suggests that, of the 3.7 % difference between the old and new values, ~1 % comes from using more accurate structures and ~2.5 % from using a strict definition for buried residues. These results explain why sets of residues have characteristic values for the differences between the old and new volumes (see above). Within the 0–5 % accessibility range used in the

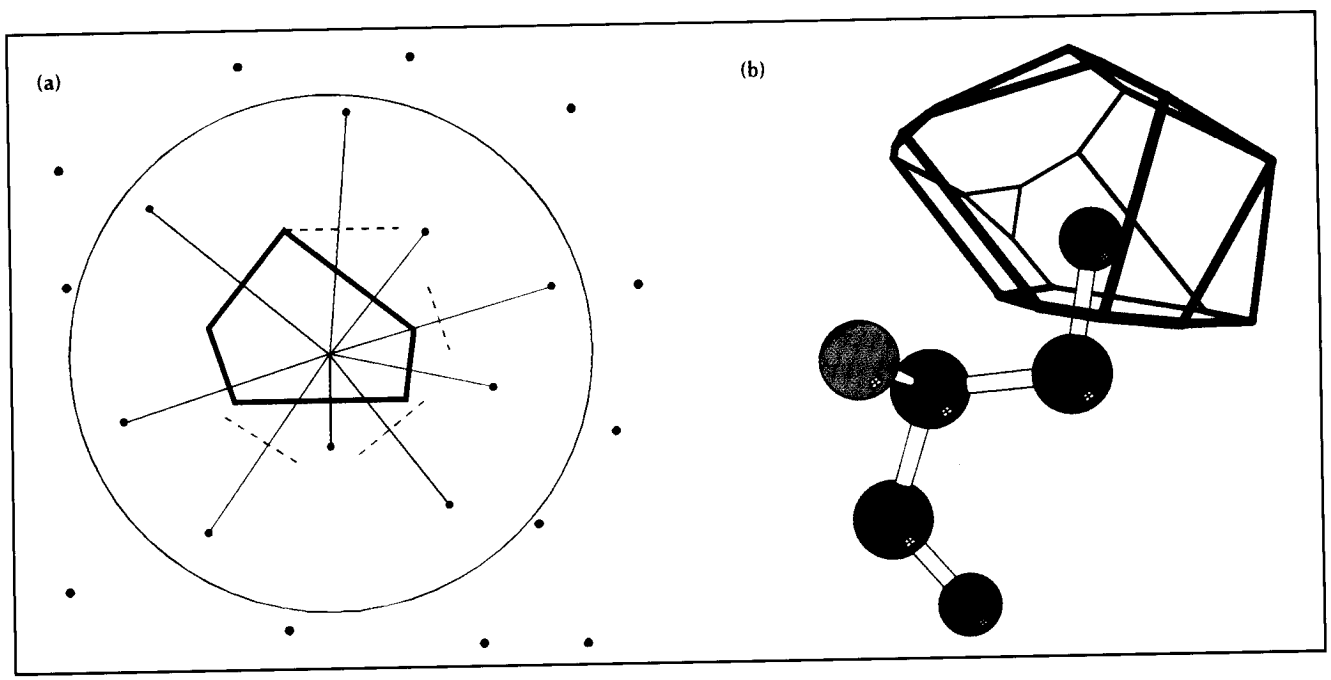

Fig. 1. The determination of the volumes occupied by residues in protein interiors. **(a)** A two-dimensional graphical illustration of the Voronoi procedure [6,10] for determining the volumes occupied by atoms buried in an assembly of atoms. The points represent the position of atoms. We show how the space effectively occupied by one atom is determined. Vectors are drawn to all neighbours within a radius of 8 Å. Planes perpendicular to each of these vectors are then constructed at a point related to the van der Waals radii of the two atoms forming the vector. The smallest polygon (polyhedron in the three-dimensional case) so constructed is the Voronoi polygon and defines the space effectively occupied by the atom. If the procedure is carried out for all buried atoms the whole of the interior space is allocated to them. The packing density of the interior of the assembly is proportional to the mean area of the polygons (or, in three dimensions, the mean volumes of the polyhedra). **(b)** The Voronoi polyhedron of the Oγ group in a buried serine residue. The atoms in the residue are drawn as spheres with a radius one-quarter of the van der Waals radii. For reasons of clarity, we do not show the atoms in neighbouring residues which bury the Oγ group and which are used in the construction of the polyhedron [see (a)]. Each of the other atoms in the buried residue is surrounded by similar polyhedra; the sum of their volumes gives the volume occupied by the residue in the protein interior.

Table 1. Mean volumes of residues buried in protein interiors.

Residue	Number (N) of buried residues	Mean volume (Å^3)	Standard error of the mean volume (Å^3)	Standard deviation (σ) of the observed volumes (Å^3)	Standard deviation (σ) of the observed volumes (%)
Gly	323	63.8	0.2	2.9	4.6
Ala	387	90.1	0.2	4.2	4.6
Pro	64	123.1	0.7	5.9	4.8
Val	353	139.1	0.2	4.7	3.4
Leu	276	164.6	0.4	5.9	3.6
Ile	234	164.9	0.4	6.2	3.7
Met	72	167.7	0.8	6.7	4.0
His	23	159.3	1.0	4.9	3.1
Phe	115	193.5	0.6	5.9	3.1
Tyr	41	197.1	1.0	6.5	3.3
Trp	26	231.7	1.1	5.6	2.4
Ser	137	94.2	0.3	3.7	3.9
Thr	102	120.0	0.5	4.8	4.0
Cys	43	103.5	0.8	5.0	4.8
Cyh	30	113.2	0.7	3.8	3.3
Asp	36	117.1	0.7	4.0	3.4
Asn	41	127.5	0.6	4.2	3.3
Glu	7	140.8	2.0	5.3	3.7
Gln	17	149.4	1.2	4.9	3.3
Lys	6	170.0	2.1	5.1	3.0
Arg	13	192.8	1.8	6.6	3.4
Peptide	1866	39.2	0.1	5.4	2.1

Volumes are given as Å^3 residue^{-1}; to convert these values to cm^3 mol^{-1}, they should be multiplied by 0.6023 (see text). Cys is involved in a disulfide linkage while Cyh is not.

old calculations, the aliphatic and sulphur-containing residues tend to be more buried than other residue types and the differences in their volumes are relatively small (1–2.5 %) because the change in the definition for buried residues has only a small effect. The charged and amide residues, on the other hand, tend to be more accessible than other residue types and the strict definition for buried residues will make an important contribution to their larger differences (4–10 %).

In the new calculations, the standard deviations of the observed volumes around the mean values are also smaller than those in the previous work: the average deviation is 3.7 % in place of 6 % [7] (Table 1). This reflects again the accuracy of the structures used in the present calculations.

In the following sections, we discuss the differences between the mean volumes residues occupy in the protein interior relative to the volumes residues occupy in solution. The significance of these differences will depend upon the standard error of the mean residue volumes. This is not the same as the standard deviations of the observed volumes, σ, but is related to it by the expression $\sigma/\sqrt{N}$, where N is the number of observations used to calculate the mean. The standard error of each mean residue volume is given in Table 1. For 18 residues, the values are between 0.2 Å^3 and 1.2 Å^3. Three residues, arginine, glutamate and lysine, have larger standard error values (between 1.8 Å^3 and 2.1 Å^3) which arise from the small number (N) of these residues that are buried.

Calculation of total protein volumes using the volumes of buried residues

The values for the residue volumes in Table 1 are derived from the relatively small proportion of residues that are completely buried within the proteins: those that form its 'deep' structure. To determine whether these values have a wider application, they were used to calculate the volumes of whole protein molecules in solution.

A major difference between residues in protein interiors and those on the surface is the electrostriction that occurs around charged groups in contact with water. The effect of electrostriction on the volumes of atomic groups was discussed in some detail by Cohn and Edsall [12] who concluded that it reduces the volumes of carboxyl oxygens by 10 Å^3 and amino and guanadino groups by 18 Å^3 — values that agree with more recent determinations [13].

The experimental solution volume of a protein is usually expressed as a partial specific volume (PSV), its volume (V) divided by its molecular weight (W). PSVs were calculated for 13 proteins (Table 2). The volume, V, was computed by summing the volumes of each residue in the protein using the mean values given in Table 1; adding the atomic volume of a carboxyl oxygen for the carboxyl terminus and, when necessary, an acetyl group for the amino terminus; and correcting the resulting volume for electrostriction by subtracting 10 Å^3 for each carboxyl oxygen and 18 Å^3 for each amino and guanadino group in the protein (inclusive of the termini). The molecular weight, M, was computed from the amino acid sequences. Now, by convention, the units of PSV are ml g^{-1} and the atomic units are Å^3 and Daltons. To calculate the PSVs in conventional units we used the formula:

$$6.023 \times 10^{23} \times (10^{-8})^3 \times (V/W) = 0.6023(V/W)$$

where 6.023×10^{23} is Avogadro's number and $(10^{-8})^3$ converts cubic Å to millilitres.

The calculated and experimental PSVs for the 13 proteins are given in Table 2. They differ by between 0.0 % and 2.2 %, and by 0.65 % on average. This close agreement between the calculated and observed values shows that, if allowance is made for electrostriction around the charged groups on the protein surface, the volumes of folded proteins can be accurately determined from the mean volumes of buried residues given in Table 1.

Volumes of residues in solution

Volumes of amino acids in solutions

There are no data for the volumes residues occupy in solution *per se*. However, since the work of Cohn *et al.* [14], residue volumes have been estimated from the volumes of amino acids and peptides in solution. Data

Table 2. Observed and calculated partial specific volumes for proteins.

Protein	Partial specific volumes ($ml\,g^{-1}$)		
	Observed[a]	**Calculated[b] from volumes of: (i) residues in folded proteins**	**(ii) amino acids in solution**
Ribonuclease A (bovine)	0.703	0.707	0.704
Lysozyme (chicken)	0.712	0.714	0.714
Adenylate kinase (porcine)	0.740	0.724	0.727
Chymotrypsinogen (bovine)	0.732	0.731	0.728
Elastase (porcine)	0.730	0.730	0.725
Subtilisin (*B. amyloliquefaciens*)	0.731	0.734	0.727
Carbonic anhydrase B (human)	0.729	0.725	0.721
Superoxide dismutase (bovine)	0.729	0.721	0.717
Carboxypeptidase A (bovine)	0.733	0.731	0.726
Concanavalin A (jack bean)	0.732	0.727	0.722
Malate dehydrogenase (porcine)	0.742	0.741	0.741
Alcohol dehydrogenase (equine)	0.750	0.736	0.735
Lactate dehydrogenase (bovine)	0.740	0.740	0.738

[a]Observed values taken from the compilation of Squire and Himmel [38] except that for lysozyme, which is taken from Gekko and Noguchi [21]. [b]The calculation of the partial specific volumes is described in the text.

are available for the volumes in solution of all 20 amino acids [12–16] and are given in Table 3. Data are also available for certain tripeptides of the form Gly-X-Gly (see below for discussion of these data).

Calculation of protein volumes from the volumes of amino acids in solution

It was shown more than 50 years ago that the volumes of proteins in solution can be accurately calculated from the volumes of amino acids in solution [12], and we used the amino acid volumes in Table 3 to calculate the PSVs of the 13 proteins in Table 2. The protein volumes were calculated by summing the amino acid volumes (taken from Table 3) and correcting for the formation of the peptide links by subtracting 10.4 Å^3 for each such link in the protein.

The PSVs calculated from the amino acid solution volumes differ from the experimental values by between 0.0% and 2.0%, and by 0.8% on average. Note that these calculated values are as accurate as those calculated using the mean volumes of residues in folded proteins (Table 2).

Volume changes on protein folding

The calculations described in the previous sections show that protein volumes can be accurately determined from both the volumes amino acids occupy in solution and from the volumes residues occupy in protein interiors. This conclusion might have been expected from the experimental observations that volume changes on folding are small (<0.5%; see above). However, it does not arise because there are no changes in residue volumes on folding. Changes in residue volume do occur. The size of these changes can be calculated from the comparison of the volumes of residues in the protein interior with those of amino acids in solution.

Table 3. The volumes[a] side chains occupy in protein interiors and in amino acids in solutions.

	Residue[b] and side chain[d] in protein interior		Amino acid[c] and side chain[d] in water		Changes in side chain[d] volume on folding
	V_p	SC_p	V_s	SC_s	$SC_p - SC_s$
Gly	63.8		71.7		
Aliphatic residues					
Ala	90.1	26.3	100.3	28.9	−2.6
Val	139.1	75.3	150.6	78.9	−3.6
Leu	164.6	100.8	178.7	107.0	−6.2
Ile	164.9	101.1	175.4	103.7	−2.6
Pro	123.1	59.3	137.2	65.5	−6.2
Aromatic, hydroxyl, or sulphur-containing residues					
Met	167.7	103.9	174.9	103.2	+0.7
Cyh	113.2	49.4	122.1	50.4	−1.0
Cys	103.5	39.7			
Phe	193.5	129.7	202.3	130.6	−0.9
Tyr	197.1	133.3	205.3	133.6	−0.3
Trp	231.7	167.9	239.0	167.3	+0.6
Ser	94.2	30.4	100.7	29.0	+1.4
Thr	120.0	56.2	127.6	55.9	+0.3
Charged and amide residues					
His	159.3	95.5	163.9	92.2	+3.3
Asn	127.5	63.7	128.4	56.7	+7.0
Asp	117.1	53.3	113.1	41.4	+11.9
Gln	149.4	85.6	156.0	84.3	+1.3
Glu	140.8	77.0	140.2	68.5	+8.5
Arg	192.8	129.0	192.8	121.1	+7.9
Lys	170.0	106.2	170.3	98.6	+7.6

[a]All volumes are given here as Å^3 residue^{-1} etc.; to convert these values to cm mol^{-1}, they should be multiplied by 0.6023 (see text). [b]Residue volumes in the protein interior, V_p, are from Table 1. [c]Amino acid volumes in water, V_s, are from Mishra and Ahluwalia [13], Rao *et al.* [15] and Jolicoeur *et al.* [16]. Many of these values are close to the original measurements of Cohn and Edsall [12] and Cohn *et al.* [14]. [d]The volumes occupied by side chains in proteins and amino acids were calculated by simply subtracting the corresponding volume of the glycyl residue or glycine from the rest of the residue or amino acid volumes: e.g., for an alanine in the protein interior: $SC_p(Ala) = V_p(Ala) - V_p(Gly)$.

The comparison of residue volumes with those of amino acids is not entirely straightforward because of the additional atoms in amino acids. We overcome this problem by treating the side chains and the peptide group separately. The volume of a side chain in solution is taken to be the volume of its amino acid less that of glycine; the volume of a side chain in the folded protein is taken to be that of its residue less that of the glycyl residue. The change in the volume of the side chain on folding is the difference between these two values.

As shown in Table 3, the side chain volumes in folded and unfolded proteins have differences that are systematic enough to place them in three distinct groups. The aliphatic side chains occupy volumes that are 2.6 Å^3 to 6.2 Å^3 smaller in proteins than in solution. Aromatic side chains, and those with sulphur or hydroxyl groups, show only small volume changes on going between the two environments. Side chains that are charged or have amide groups have volumes that are 3.3 Å to 11.9 Å^3

larger in proteins than in solution. The volume changes for side chains with polar atoms are the net result of the increase in volume of the polar atoms and the decrease in volume of the non-polar atoms. Note that the changes for the aliphatic, charged and amide side chains are much larger than the standard errors of the mean volumes (Table 1) and are very significant.

Experimental measurements of the volume of a peptide group in solution [12,13] give a value of $\sim 33\,Å^3$. For peptides buried in protein interiors, our calculations give a mean volume of $39.2\,Å^3$ (Table 1). An increase in volume of $6\,Å^3$ may seem large, but a peptide contains both an NH and an O group and the increase is smaller than that which occurs on loss of electrostriction by charged polar groups: $10\,Å^3$ for a carboxylate group and $18\,Å^3$ for amino and guanadino groups [12,13].

The positive and negative changes in volume are not due to any special interactions within folded proteins. The gain in volume of the polar side chains arises from efficient internal packing not fully compensating for the loss of electrostriction and hydrogen-bonded hydration. The reduction in the volume of aliphatic side chains results from the high efficiency of the protein packing in contrast to the low density of the 'iceberg' structure formed by hydrophobic hydration (see also [17–20]).

The systematic differences in the volumes of folded and unfolded side chains and peptides provide a clear explanation of why the volume changes on folding are small at low pressures and large and positive at high pressures. The change in volume at low pressure (~ 1 atm) and room temperature is less than 0.5 % ([1–3] and references therein). Such a small change requires either that residues have the same volumes in the folded and unfolded states or that positive and negative changes cancel on folding. Here we have shown that, upon folding, hydrophobic groups decrease in volume while hydrophilic groups increase in volume. The cancellation of these effects is demonstrated by the accurate calculation of PSVs from the volumes of both residues in proteins and amino acids in solution.

Furthermore, the cancellation of residue volume changes implies that the total volume change at high pressure depends primarily on the compressibility of the solution around the protein. The adiabatic compressibility of a protein is typically $\sim 5 \times 10^{-11}\,m^2\,N^{-1}$ [21,22], a value between that of organic solids and of soft metals. Unfolded proteins have compressibilities $\sim 25\,\%$ greater than folded proteins [1,2] and so are more compact at high pressures. Most of this increase in compressibility must come from the many non-polar groups exposed to solvent upon denaturation since the compressibility of water around non-polar groups ($\sim 37 \times 10^{-11}\,m^2\,N^{-1}$) is greater than that of water around polar and charged groups [23]. Thus, the decrease in volume that accompanies protein unfolding at high pressure results from the great compressibility of the water around non-polar groups that become exposed.

Solution side chain volumes from tripeptides

Experiments on amino acids provide the most complete data for side chain volumes in solution. These side chain volumes, however, are not exactly comparable with those in proteins because, in amino acids, the electrostriction around charged amino and carboxyl groups affects the solvation of side chains so as to reduce their volumes [17,24]. This effect can be mitigated by separating the side chains from these charged groups. Thus, more accurate but less extensive data are given by the work of Reading and Hedwig [18,19], who determined the volumes in solution of tripeptides of the form Gly-X-Gly, where X is glycine, alanine, valine, leucine, serine or asparagine. The volumes of these peptides and their side chain components are listed in Table 4. The solution side chain volumes calculated from the tripeptide data are 0.6–2.9 $Å^3$ larger than those calculated from amino acids.

Table 4. The volumes[a] side chains occupy in protein interiors and in Gly-X-Gly tripeptides in solution.

	Volume of residue[b] and side chain[d] in protein interior		Volume of Gly-X-Gly tripeptide[c] and side chain[d] in solution		Differences in side chain volumes
X	V_p	SC_p	V_t	SC_t	$SC_p - SC_t$
Gly	63.8		185.8		
Ala	90.1	26.3	215.3	29.5	−3.2
Val	139.1	75.3	266.3	80.5	−5.2
Leu	164.6	100.8	294.5	108.6	−7.8
Ser	94.2	30.4	217.7	31.9	−1.5
Asn	127.5	63.7	244.7	58.8	+4.9

[a]All volumes are given here as $Å^3$ residue^{-1} etc.; to convert these values to $cm^3\,mol^{-1}$, they should be multiplied by 0.6023 (see text). [b]Residue volumes in the protein interior, V_p, are from Table 1. [c]Volumes of the Gly-X-Gly tripeptides in solution, V_s, are from Reading and Hedwig [18,19]. [d]The volumes occupied by side chains in proteins and in tripeptides were calculated by simply subtracting the corresponding volume of the glycyl residue, or glycine tripeptide, from the residue, or tripeptide, volumes: e.g., for an alanine in the protein interior: $SC_p(Ala) = V_p(Ala) - V_p(Gly)$.

This means that on folding, the negative volume changes that occur for aliphatic residues are somewhat larger than those calculated from the amino acid data and the positve changes for polar side chains are somewhat smaller (see Tables 3 and 4). Note that these corrections do not affect the general conclusions obtained from the amino acid data on the nature of the volume changes that occur on protein folding.

High packing density of protein interiors

The earlier calculations of the volumes residues occupy in protein interiors indicated that the packing density is the same as that in crystals of amino acids [6,7]. The new smaller values show that the density is, in fact,

somewhat higher. To determine the extent to which this is the case, we compared the mean volumes of side chains in protein interiors with the volumes of the side chains in crystals of amino acids (Table 5).

Table 5. The volumes[a] side chains occupy in protein interiors and in amino acids in crystals.

	Volume of residue[b] and side chain[d] in protein interior		Volume of amino acid[c] and side chain[d] in crystals		Differences in side chain volumes
	V_p	SC_p	V_x	SC_x	$SC_p - SC_x$
Gly	63.8		77.4		
Aliphatic residues					
Ala	90.1	26.3	107.7	30.3	−4.0
Val	139.1	75.3	154.6	77.2	−1.9
Leu	164.6	100.8	187.4	110.0	−9.2
Ile	164.9	101.1	182.2	104.8	−3.7
Pro	123.1	59.3	135.4	58.0	+1.3
Aromatic, hydroxyl, or sulphur-containing residues					
Met	167.7	103.9	187.1	109.7	−5.8
Cyh	113.2	49.4	134.2	56.8	−7.4
Cys	103.5	39.7	119.4	42.0	−2.3
Phe	193.5	129.7	208.5	131.1	−1.4
Tyr	197.1	133.3	212.9	135.5	−2.2
Trp	231.7	167.9	251.3	173.9	−6.0
Ser	94.2	30.4	113.5	36.1	−5.7
Thr	120.0	56.2	135.4	58.0	−1.8
Charged and amide residues					
His	159.3	95.5	177.4	100.0	−4.5
Asp	117.1	53.3	134.7	57.3	−4.0
Gln	149.4	85.6	159.1	81.7	+3.9
Glu	140.8	77.0	155.0	77.6	−0.6

[a]All volumes are given here as $Å^3$ residue^{-1} etc.; to convert these values to $cm^3\,mol^{-1}$, they should be multiplied by 0.6023. [b]Residue volumes in the protein interior, V_p, are from Table 1. [c]Amino acid volumes in crystals, V_x, are from [7]. There are no data for crystals of asparagine, lysine or arginine. [d]The volumes occupied by side chains in proteins and amino acids were calculated by simply subtracting the corresponding volume of the glycyl residue or glycine from the rest of the residue or amino acid volumes: e.g., for an alanine in the protein interior: $SC_p(Ala) = V_p(Ala) - V_p(Gly)$.

The volumes of both polar and non-polar side chains in protein interiors are generally smaller than in crystals: of the 17 side chains for which we have data, 15 have smaller volumes in proteins and 2 have smaller volumes in amino acid crystals (Table 5). The extent of the differences varies but, on average, side chains in protein interiors occupy 4% less volume. This value can be compared with the 15% volume reduction that occurs typically when organic molecules go from the liquid to the solid state.

These results mean that the packing in protein interiors is more efficient than that in organic crystals. This tight packing arises from two factors. First, proteins, unlike amino acids, are formed from heterogeneous side chains and, in general, objects of different sizes and shapes can pack more tightly than objects of the same size and shape. Second, and perhaps more importantly, the efficiency of packing within the unit cells of crystals is limited by the necessity of forming a three-dimensional lattice. Protein structures have no such requirement: they are free to shift their component parts relative to each other to produce an exceptionally high packing density (see also [25–27]).

Constant chemical character of the surfaces buried in proteins

In the previous sections we have argued that the very small net change in volume that occurs on folding arises from the positive changes that take place when polar groups are buried, cancelling the negative changes produced by aliphatic groups. An implication of this argument is that to maintain this balance, the ratio of different types of groups buried in proteins should be essentially constant.

Previous work has shown that the amount of surface buried in a protein is a function of molecular weight and that, for monomeric proteins in the size range 15 000–35 000 kDa, the chemical composition of the buried surface is essentially constant [7,28]. (Most smaller proteins have the same composition for their buried surfaces but there are a few exceptions involving structures in which several buried disulphide bridges form a significant fraction of the interior [28].)

We extended this work by calculating, as described previously [28], the chemical composition of the buried surface in well-determined proteins outside the range of the previous work: for monomers with molecular weights between 42 500 kDa and 82 500 kDa, and for oligomeric proteins with molecular weights between 15 800 kDa and 229 300 kDa. The results of the new calculations are given in Table 6. They show that the chemical character of surface buried in both sets of proteins is also constant: non-polar groups (those based on carbon) contribute 62 ± 1%, polar groups 31 ± 2%, charged groups 7 ± 2%, and polar and charged groups together 38 ± 1%.

The groups that the make the largest contribution to the volume changes are those totally buried in the protein interiors. We determined the composition of these groups in the proteins used to calculate the residue volumes (see Table 7). Table 7 shows that the major changes on folding come from the volume reductions produced by burying aliphatic groups and the volume gains produced by burying peptide groups.

Implications of the volume changes for models of protein stability

The explanation of the volume changes on protein folding given here has implications for experimental models of protein stability. The model most widely used is the transfer of small molecules from water to organic solvent [4,29–31]. At normal pressures, the transfer of hydrophobic molecules from water to organic solvent is accompanied by a volume increase. For example, methylene groups, which have a volume in water of 26.5 $Å^3$, increase in volume by 2–3 $Å^3$ to about 29 $Å^3$ [23,32]. As discussed above, such volume

Table 6. Chemical character of buried and accessible surface in proteins.

Protein[a]	Number of residues	Molecular weight	Z[b]	Surface of unfolded protein (Å²)	Buried surface				Accessible surface			
					Total (Å²)	Non-polar (%)	Polar (%)	Charged (%)	Total (Å²)	Non-polar (%)	Polar (%)	Charged (%)
Uteroglobin	140	15 760	2	23 690	16 260	62	33	5	7 430	56	23	21
trp repressor	202	23 060	2	34 660	23 480	63	29	9	11 180	53	25	21
SOD	304	31 070	2	46 380	32 630	61	33	6	13 750	57	25	18
Ovalbumin	385	42 570	1	63 790	47 820	63	31	6	15 920	55	27	18
PRAI-1GPS	452	49 370	1	74 030	55 440	63	30	7	18 590	55	30	15
Taka α-amylase	476	52 380	1	77 270	60 830	62	33	5	16 440	57	31	13
Aconitase	754	82 550	1	123 540	97 250	61	31	8	26 280	60	23	17
AAT	802	89 580	2	134 160	103 430	62	31	7	30 730	57	23	19
G-3-PDH	1332	145 210	4	214 870	172 570	63	30	7	42 300	53	24	23
Glycogen phosphorylase	1646	189 480	2	283 270	222 660	62	29	9	60 610	55	23	22
Catalase	1992	229 290	4	336 350	276 420	61	30	8	59 930	57	25	18
Composition range						62 ± 1	31 ± 2	7 ± 2		56.5 ± 3.5	27 ± 4	18 ± 5

[a]Protein abbreviations are as follows: SOD, superoxide dismutase; PRAI-1GPS, phosphoribosylanthranilate isomerase indoleglycerolphosphate synthase; AAT, aspartate aminotransferase; G-3-PDH, glyceraldehyde-3-phosphate dehydrogenase. The accessible surface areas of the unfolded protein and of the surface and interior of the folded proteins were calculated from the protein atomic coordinates using the procedures described previously [27]. The PDB coordinate files and references for the structure determinations of the proteins are: 1UTG [39]; 3WRP [40]; 2SOD [41]; 1OVA [42]; 1PII [43]; 6TAA [44]; 5ACN [45]; 7AAT [46]; 1GD1 [47]; 1GPB [48] and 8CAT [49]. [b]The number of subunits per molecule.

Table 7. The composition of the atomic groups buried in the protein interiors.

Atomic groups	Constituents	Constituent totals (%)	Group totals (%)
Tetrahedral carbons	Main chain Cα	14	
	Side chains	24	38
Peptides	N	15	
	C	16	
	O	13	44
Trigonal carbons	Aromatic side chains	12	
	Charged groups	1.5	13.5
Polar side chains	Neutral N, S and O	3.5	
	Charged N and O	1	4.5

increases can not be reconciled with the observations that the volume changes on protein unfolding are small and that protein interiors are well packed [1,3,5,33]. Here we have shown that this difficulty with the solution transfer model is even more serious since we have found that the close packing that occurs on folding actually involves reductions in the volumes of hydrophobic side chains. So, for example, methylene groups in proteins have a mean volume of 23.5 Å^3, about one-fifth less than their volume in organic solvents [23,32].

As the chemical environments of non-polar side chains in organic solvents and in proteins are similar, van der Waals forces will have a fairly direct relation to packing density. Thus, the large differences in packing density in the two environments means that the role van der Waals forces play in protein folding are not properly represented by solution transfer models.

An alternative model for protein stability is based on the thermodynamics of dissolving crystalline cyclic dipeptides in water [34,35]. Murphy and Gill [35] argue that crystalline dipeptides more accurately represent the protein interior because of the peptide hydrogen bonding that occurs in both. Crystalline dipeptides should also more accurately represent the packing of the side chains in protein interiors. We compared the volumes of side chains in the few known cyclic dipeptide crystal structures with those of side chains in the protein interior. As is found for amino acids, the volumes of both polar and non-polar side chains in the cyclic dipeptide crystals tend to be slightly larger (4% on average) than those in proteins (unpublished data). Thus, the solid model compounds are a good but not exact representation of the close packing in protein interiors.

Biological implications

On protein folding, residues are transferred from solution to the close-packed interior of the molecule. This change produces different effects on different groups: aliphatic groups occupy smaller volumes in proteins than they do in solution; polar groups occupy larger volumes. Examination of the surfaces buried in a wide range of proteins shows that the proportions of these groups that become buried is essentially constant and are such that the positive and negative

volume changes cancel each other out. This accounts for the experimental observations [1–3] that the net volume changes on folding are very small.

We have demonstrated here that the overall packing density of both polar and non-polar residues in protein interiors is exceptionally high. This high density produces the well-ordered structures that give proteins their high specificity in recognition and catalysis and which are essential for their biological functions.

What occurs in natural proteins can be contrasted with what is found in the proteins that have been created by *de novo* design. Designed proteins are compact with stable secondary structures and high stability, but are like molten globules in having no fixed tertiary structure and interiors with mobile side chains [36]. The stability of these structures, which may be greater than that of natural proteins, is produced by hydrophobic forces: the design experiments demonstrate the essential role of close packing in producing structures that are specific [36].

In recent discussions of protein stability, the emphasis has been placed on the contributions made by hydrophobic forces and hydrogen bonds [29–31,34,35,37]. It is difficult to determine from our results the extent of the contribution that packing forces make to the balance between the folded and unfolded states: the chemical environment of residues in solution and in protein interiors is very different, so the volumes they occupy in the two states cannot simply be related to energies. However, the high packing density observed in proteins must play a significant role because any diminution of density would reduce the contribution packing forces make to the stability of the folded state. Thus, an essential condition for an amino acid sequence to form a structure that is both stable and specific is the ability of the buried residues to pack not only in low-energy conformations, but also with high density.

Acknowledgements: We thank Drs JT Edsall, FM Richards, RL Baldwin and M Han for discussions of our results and J Goodfellow and TH Lilley for information. MG is supported by an Damon-Runyon Walter-Winchell Fellowship.

References

1. Brandts, J.F., Oliveira, R.J. & Westort, C. (1970). Thermodynamics of protein denaturation. Effect of pressure on the denaturation of ribonuclease A. *Biochemistry* 9, 1038–1047.
2. Hawley, S.A. (1971). Reversible pressure–temperature denaturation of chymotrypsinogen. *Biochemistry* 10, 2436–2442.
3. Zipp, A. & Kauzmann,W. (1973). Pressure denaturation of metmyoglobin. *Biochemistry* 12, 4217–4228.
4. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14, 1–63.
5. Kauzmann, W. (1987). Thermodynamics of unfolding. *Nature* 325, 763–764.
6. Richards, F.M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* 82, 1–14.
7. Chothia, C. (1975). Structural invariants in protein folding. *Nature* 254, 304–308.
8. Bernstein, F., *et al.*, & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542.
9. Morris, A.L., MacArthur, M.W., Huchinson E.G. & Thorton, J. (1992). Stereochemical quality of protein structure coordinates. *Proteins* 12, 345–364.
10. Bernal, J.D. & Finney, J.L. (1967). Random-close-packed hard-sphere model II. Geometry of random packing of hard spheres. *Discussions Faraday Soc.* 43, 62–69.
11. Lee, B. & Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400.
12. Cohn, E.J. & Edsall, J.T. (1943). In *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions.* pp. 370–381, Reinhold, New York.
13. Mishra, A.K. & Ahluwalia, J.C. (1984). Apparent molal volumes of amino acids, *N*-acetylamino acids and peptides in aqueous solutions. *J. Phys. Chem.* 88, 86–92.
14. Cohn, E.J., McMeekin, T.L., Edsall, J.T. & Blanchard, M.H. (1934). Studies in the physical chemistry of amino acids, peptides and related substances. I. The apparent molal volume and the electrostriction of the solvent. *J. Am. Chem. Soc.* 56, 784–794.
15. Rao, M.V.R., Atreyi M. & Rajeswari, M.R. (1984). Partial molar volumes of α-amino acids with ionogenic side chains in water. *J. Phys. Chem.* 88, 3129–3131.
16. Jolicoeur, C., Riedl, B., Desrochers, D., Lemelin, L.L., Zamojska, R. & Enea, O. (1986). Solvation of amino acid residues in water and urea–water mixtures: volumes and heat capacities of 20 amino acids in water and 8 molar urea at 25°C. *J. Solution Chem.* 15, 109–128.
17. Kauzmann, W., Bodanzski, A. & Rasper, J. (1962). Volume changes in protein reactions. II Comparison of ionization reactions in proteins and small molecules. *J. Am. Chem. Soc.* 84, 1771–1777.
18. Reading, J.F. & Hedwig, G.R (1990). Thermodynamic properties of peptide solutions. *J. Chem. Soc. Faraday Trans.* 86, 3117–3123.
19. Hedwig, G.R. (1992). Partial molar volumes of the amino acid side-chains of proteins in aqueous solution: some comments on their estimation using partial molar volumes of amino acids and small peptides. *Biopolymers* 32, 537–540.
20. Leslie, T.E. & Lilley, T.H. (1985). Aqueous solutions containing amino acids and peptides XX: volumetric behaviour of some terminally substituted amino-acids and peptides at 298.15 K. *Biopolymers* 24, 695–710.
21. Gekko, K. & Noguchi, H. (1979). Compressibility of globular proteins in water at 25°C. *J. Phys. Chem.* 83, 2706–2714.
22. Gavish, B., Gratton E. & Hardy, C.J. (1983). Adiabatic compressibility of globular proteins. *Proc. Natl. Acad. Sci. USA* 80, 750–754.
23. Gekko, K. & Noguchi, H. (1974). Hydration behaviour of ionic dextran derivatives. *Macromolecules* 7, 224–229.
24. Zana, R. (1977). On the volume changes upon the protonation of *N*-alkylcarboxylate ions and *N*-alkylamines in aqueous solution. *J. Phys. Chem.* 81, 1817–1822.
25. Lesk, A.M. & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* 136, 225–270.
26. Eriksson, A.E., *et al.*, & Matthews, B.W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255, 178–183.
27. Buckle, A.M., Henrick, K. & Fersht, A.R. (1993). Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J. Mol. Biol.* 234, 847–860.

28. Miller, S., Janin, J., Lesk, A.M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656.
29. Baldwin, R.L. & Muller, N. (1992). Relation between the convergence temperatures T_h^* and T_s^* in protein unfolding. *Proc. Natl. Acad. Sci. USA* **89**, 7110–7113.
30. Spolar, R.S., Livingstone, J.R. & Record, M.T. (1992). Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surfaces from water. *Biochemistry* **31**, 3947–3955.
31. Lee, B. (1991). Isoenthalpic and isoentropic temperatures and the thermodynamics of protein denaturation. *Proc. Natl. Acad. Sci. USA* **88**, 5154–5158.
32. Friedman, M. & Scheraga, H.A. (1965). Volume changes in hydrocarbon–water systems. Partial molal volumes of alcohol–water solutions. *J. Phys. Chem.* **69**, 3795–3800.
33. Bello, J. (1978). Tight packing of protein cores and interfaces. *Int. J. Pept. Protein Res.* **12**, 38–41.
34. Murphy, K.P., Privalov, P.L & Gill, S.J. (1990). Common features of protein unfolding and dissolution of hydrophobic compounds. *Science* **247**, 559–561.
35. Murphy, K.P. & Gill, S.J. (1991). Solid model compounds and the thermodynamics of protein unfolding. *J. Mol. Biol.* **222**, 699–709.
36. Betz, S.F., Raleigh, D.P. and DeGrado, W.F. (1993). *De novo* protein design: from molten globules to native-like states. *Curr. Opin. Struct. Biol.* **3**, 601–610.
37. Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155.
38. Squire, P.G. & Himmel, M.E. (1979). Hydrodynamics and protein hydration. *Arch. Biochem. Biophys.* **196**, 165–177.
39. Morize, I., *et al.*, & Moron, J.P. (1987). Refinement of the $C222_1$ crystal form of oxidized uteroglobin at 1.34 Å resolution. *J. Mol. Biol.* **194**, 725–739.
40. Zhang, R.-G., Joachimiak, A., Lawson, L., Schevitz, R.W., Otwinowski, Z. & Sigler, P.B. (1987). The crystal structure of *trp* aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature* **327**, 591–597.
41. Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S. & Richardson, D.C. (1982). Structure of copper, zinc superoxide dismutase. *J. Mol. Biol.* **160**, 181–217.
42. Stein, P.E., Leslie, A.G.W., Finch, J.T. & Carrell, R.W. (1991). Crystal structure of uncleaved ovalbumin at 1.95 Å resolution. *J. Mol. Biol.* **221**, 941–959.
43. Wilmanns, M., Priestle, J.P. & Jansonius, J. (1992). Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *J. Mol. Biol.* **223**, 477–507.
44. Swift, H.J., *et al.*, & Wilkinson, A.J. (1991). Structure and molecular model refinement of *Aspergillus oryzae* (taka) α-amylase: an application of the stimulated-annealing method. *Acta Crystallogr. B* **47**, 535–544.
45. Robbins, A.H. & Stout, C.D. (1989). The structure of aconitase. *Proteins* **5**, 289–312.
46. Mcphalen, C.A., Vincent, M.G. & Jansonius, J. (1992). The x-ray structure refinement and comparison of three forms of mitochondrial aspartate aminotransferase. *J. Mol. Biol.* **225**, 495–517.
47. Skarzynski, T., Moody, C.E. & Wonacott, A.J. (1987). The structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.* **193**, 171–187.
48. Barford, D., Hu, S.-H. & Johnson, L.N. (1991). Structural mechanism for glycogen phosphorylase: control by phosphorylation and AMP. *J. Mol. Biol.* **218**, 233–260.
49. Fita, I., Silva, A.A., Murthy, R.N. & Rossmann, M.G. (1986). The refined structure of beef liver catalase at 2.5 Å resolution. *Acta Crystsllogr. B* **42**, 497–515.

Received: **5 Apr 1994**; revisions requested: **28 Apr 1994**; revisions received: **11 May 1994**. Accepted: **17 May 1994**.