SOLVENT EFFECTS AND POLAR INTERACTIONS IN THE STRUCTURAL STABILITY AND DYNAMICS OF GLOBULAR PROTEINS

J. L. Finney, B. J. Gellatly, I. C. Golton, and J. Goodfellow, *Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England*

ABSTRACT Using detailed hydrogen bonding, surface exposure, internal environment, and solvent interaction calculations on several proteins, in conjunction with data from quantum mechanical hydrogen-bonding studies, various contributions to the free energy of folding are estimated and their likely relative significance discussed. A picture emerges of globular proteins as extremely well-fitting jigsaw-puzzles, in which no single driving force dominates the marginal stability of the native conformation. Rather, the folded structure is seen as the result of a complex global maximization of several strongly-interacting driving forces. In particular, the necessity to maintain very efficient internal hydrogen-bonding, and the role of the solvent as a hydrogen-bond sink, are stressed as strong constraints on the (incomplete) maximization of hydrophobic effects. The possible significance of internal dipole-induced dipole interactions is discussed tentatively.

Although quantitative estimates of the various contributions remain uncertain, consideration of effective force constants suggests that polar, including solvent, interactions may largely determine the overall curvatures of the native conformation's potential well, and be important in controlling the flexibility of local regions which are important for the exact positioning of groups during enzyme catalysis, as well as the molecule's overall dynamics. In contrast, hydrophobic interactions change less for small geometrical perturbations, and seem more relevant to directing the folding protein. along a path to a region in configurational space where the polar interactions can switch on for the final "docking."

INTRODUCTION

Solvent effects are relevant to the folding, stability, and dynamics of globular protein molecules. Although the significance of so-called hydrophobic interactions has been recognized for over two decades (1), their operation at the molecular level is not understood. Moreover, the relative importance of direct polar interactions within the molecule, between molecule and solvent, and within the solvent itself, is unclear.

As the structural consequences of the operation of these and other relevant driving forces are found in the native protein's tertiary structure, we might examine such structures for clues concerning the significance of the various effects. For example, the earliest protein structures solved by x-ray crystallography appeared to show a large degree of burial of apolar groups within the molecule, a result consistent with the dominance of a hydrophobic driving force. More detailed quantitative analysis of solvent accessible surface by Lee and Richards (2) and others (3, 4, 5, 6) showed this to be an over-simplification, with up to half the exposed surface of a typical protein being apolar.

In this paper, we attempt to reassess the relative contributions of the various "hydrophobic," polar, entropic, nonbonding, and solvent effects to the folding, stability, and dynamics of globular proteins. We make use of data on internal interactions, packing, and the nature of the

exposed surface, together with the results of *ab initio* calculations on polar (hydrogen-bonding) interactions in an attempt to quantify as far as possible free energy differences between the folded and fully extended chain conformations. Despite large quantitative uncertainties in the estimates made, we conclude that no one effect dominates the folding process; the biologically-important delicate stability of the native structure is rather the result of a complex joint maximization of hydrophobic, hydrogen bonding, and related entropic effects which are strongly connected through the topology of the molecule. It is argued that direct polar interactions, made both internally and to solvent, appear to determine the overall shape of the native structure's potential well, with the nonbonded internal interactions controlling the fine structure of the molecule's flexibility and dynamics

METHODS

Using atomic coordinates obtained from x-ray crystallographic studies, the internal and surface structures of ribonuclease-S at 2.0 Å (9), lysozyme (2.0 Å) (10), and pancreatic trypsin inhibitor (1.5 Å)(11) were examined in detail. A previously described procedure (5, 7, 8) partitioned each protein into a unique set of space-filling polyhedra, one for each atom or atom group. To facilitate polyhedron closure at the molecular surface, hypothetical solvent molecules were placed in a uniform density on the surfaces of spheres centered on the surface atoms. Up to 24 surface points could be added at relative positions corresponding to the snub cube arrangement (53); additional "tetrahedral" surface points were placed at the sphere intersections as described in reference 7. The sphere radii were chosen with reference to equivalent group-solvent distances observed in small molecule hydrate crystals (5); no van der Waals radii thus need be assigned. Fig. 1 illustrates the resulting quantification of internal volume occupation, internal interactions, and molecular (surface-exposed) area (related to solvent accessible area [5]). The

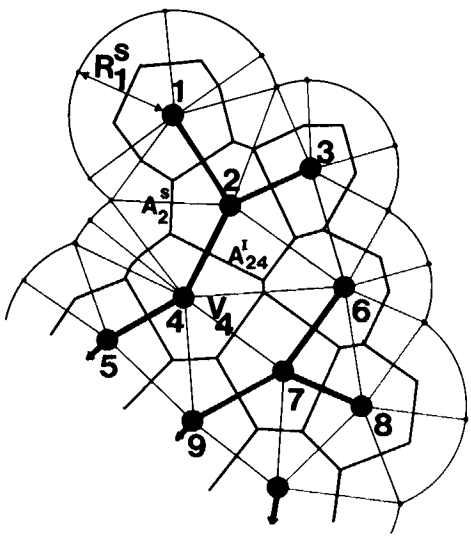

Figure 1 A two dimensional analogue of the volume and surface partitioning scheme used. Each atom or atom groups (●) (eg, CH_2) is surrounded by the smallest polyhedron constructed from planes perpendicularly bisecting atom-atom (or atom-surface point) vectors. Hypothetical surface points (•) are placed at a given density (up to 24 points) on spheres drawn at radii R_i^s around each atom i; in addition, points are placed at the intersections of three such spheres (two circles in two dimensions). Values of R_i^s ("Solvent interaction distances") are chosen from a set derived from actual group-to-water distances found in small molecule hydrates. The volume V_i occupied by an atom i is taken to be the volume of the associated polyhedron; here only atom 7 is fully internal. The interaction between two atoms (ij) is taken to be proportional to the "interaction area" A_{ij}^I of the common polyhedron plane. Molecular surface area A_i^s is the sum of all area segments for a given atom which are shared with hypothetical solvent points. The curved area (locus of hypothetical solvent points) is the "surface accessible area."

methods are similar to those developed by Richards (2, 4, 12), and Shrake and Rupley (6) and used extensively by Chothia (3); results of these investigations are also drawn upon in the following discussion.

In addition, the geometry of possible internal polar-polar (hydrogen-bonding) interactions was examined in detail, and internal contact distances in insulin (1.5 Å) (G. Dodson, personal communication) and rubredoxin (1.5 Å) (13) were calculated.

RESULTS

Internal Hydrogen Bonding

Assuming that NH, NH_2, CO, and OH groups can ideally participate in 1,2,2, and 2–3 hydrogen bonds respectively,[1] we find 586–627 and 571–600 hydrogen bonding sites respectively in RNase-S and lysozyme. In both the folded structures, ~300, or 48–52% of the polar sites, make internal hydrogen bonds.

For those 250–300 polar groups with no internal hydrogen bond interaction, a strong relation is found with surface exposed area (see below): in general, each polar group is exposed sufficiently to the surface to be geometrically capable of making up its full hydrogen-bond complement with solvent. For RNase-S ~7% of main chain NH groups and ~15% of the carbonyl oxygens appear unable to fully saturate their hydrogen bonding capabilities internally or to solvent. For lysozyme, examination of a three dimensional model in conjunction with the estimated errors in the atomic coordinates, reduced these values to only 3% and 3%, respectively. Possible partially-compensating local interactions in the neighborhood of the remaining unsatisfied groups (largely within the β-sheet structure) are discussed elsewhere (5).

Thus, despite the molecule's complexity, the folded structures of RNase and lysozyme allow almost all polar groups to satisfy their hydrogen bonding capabilities either internally or to solvent. The accessibility data of Lee and Richards (2) implies the fully extended chain can also saturate its hydrogen bonding but through interactions with solvent.

Surface Exposure

(*a*) 40–50% of the surface accessible area of a typical globular protein is made up of apolar groups (2, 5, 6). Thus, the classical hydrophobic burial, or "grease in" principle, is incompletely observed. Moreover, in folding from a fully extended chain, about the same fraction (30–35%) of polar and apolar atom groups become internally buried. (*b*) Charged groups are in general fully exposed to solvent. Where this is not the case, internal salt bridges are usually formed. There are a few cases of unbridged internal charged groups, e.g., in the serine proteases (15–17); here, the local neighborhood of the buried charge contains several water molecules, which may serve to reduce the destabilizing effect of the unbalanced internal charge (14, 15, 18) by what might be termed local dielectric effects. (*c*) As discussed above, the exposure of polar groups is strongly related to the number of potential hydrogen bonds made internally. For lysozyme, the polar group exposure is enough to support geometrically polar group-water interactions with ~260 solvent water molecules. This amounts to 0.32 g water/g protein, and is numerically the same as the nonfreezing water fraction (0.32 ± 0.02 g/g) obtained from both infrared and differential scanning calorimetry measurements on lysozyme glasses (19).

---

[1]The validity of these "classical" numbers is discussed elsewhere (5, 14). In particular, the variability assigned to the hydroxyl group derives from an incomplete lone pair separation. All groups are here assumed unchanged; removing charged groups from consideration makes little difference to the discussion.

Internal Packing

Despite its complexity, the internal packing of groups within a protein is generally as efficient as that in small molecule crystals (4, 7, 12). However, there are significant fluctuations in local density: for a particular atom group, the standard deviation of the volume occupied is typically 10–15% of the mean (7), which compares poorly with the corresponding figure of 5% for close-packed model monatomic glasses (20). Thus the local maximization of the van der Waals interaction is far from optimal, their relative weakness allowing significant geometrical movements with only small energy costs: for a single $CH_2...CH_2$ interaction we can estimate that a local volume increase of 10% carries only a small energy penalty of ~7 cal mol^{-1}. This uses the van der Waals well depth given in reference 40, and may be an overestimate. Other literature values would give corresponding energies down to 1 cal mol^{-1}. This problem is referred to later in the text.

Internal Environment

Table I shows the nature of the internal interactions, quantified (5, 54) as the area of the dividing plane constructed as in Fig. 1 and classified in terms of the polar or apolar nature of the neighboring groups. Carbon and sulphur are assumed "apolar," nitrogen and oxygen "polar." Previous work shows that a reasonable alternative scheme makes no significant difference to the overall statistics (5). The major point of interest is the high proportion of nominally mixed polar-apolar interactions. Moreover, in Ooi plots drawn for each of these types of contact in RNase-S (5), the secondary structure is more visible in the mixed polar-apolar plot than in the apolar-apolar one.

Fig. 2 plots the distribution of separation distances of subsets of these polar-apolar and apolar-apolar proximities for insulin—one of the best characterized protein structures available, and refined without strong constraints on nonbonded contacts. The apolar-apolar plot looks normal, and indicates the order of magnitude of the errors in the coordinate set. In contrast, the polar-apolar plot shows a significant bias to distances shorter than the van der Waals sum. Comparable plots for rubredoxin shows the same behavior. A similar effect noted for RNase-S (5) was tentatively rationalized in terms of dipole-induced dipole attractions; our preliminary calculations suggest such an effect might result in shortening of a classic van der Waals contact distance by 0.2–0.4 Å, depending on the choice of van der Waals parameters.

DISCUSSION

Stability of the Native Conformation

In comparing the native conformation (F) with the fully extended chain structure (E), we require measurements of the Gibbs free energy change $\Delta G_{E \to F}$ relating the two structures. For

TABLE I
TOTAL OF INTERACTION AREA ASSIGNED TO THE VARIOUS CATEGORIES

Interaction type	RNase-S*	Lysozyme	BPTI
	(%)	(%)	(%)
Polar-polar	26	35	28
Apolar-apolar	30	26	29
Polar-apolar	44	39	43

*From reference 5.

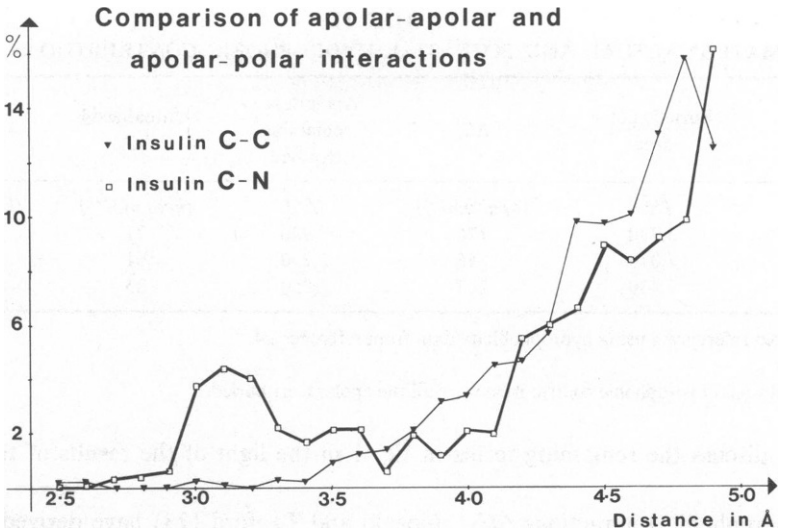

Figure 2 A subset of polar-polar and polar-apolar distances in insulin. The C-C distribution is for tetrahedral carbon atoms, and shows a smooth curve beginning to rise significantly around a "van der Waals radius" of 3.7–4.0 Å. The tetrahedral carbon-main chain nitrogen distribution, in contrast, shows a strong peak around 3.0–3.2 Å. This distance is well below the normally assigned van der Waals radius sum for the two atom groups (3.4–3.7 Å). The ordinate is expressed as a percentage of all relevant separation distances below 5.0 Å.

denaturation in a strongly solvent-perturbing medium such as guanadinium hydrochloride, which arguably results in an unfolded reference state of the protein, values of $\Delta G_{F\to E}$ of 10–20 kcal mol^{-1} are given in the literature for proteins such as hen egg-white lysozyme and ribonuclease A (21) near neutral pH.

Those terms thought to be significant to $\Delta G_{E\to F}$ are given in the equation:

$$\Delta G_{E\to F} = \Delta G_{\text{titr}} + \Delta G_{\text{charge}} + \Delta G_{\phi} + \Delta H_{\psi} - n_c T\Delta s_{\text{config}} - n_w T\Delta s_{\text{release}}, \quad (1)$$

where ΔG_{titr} is the result of ionizable groups undergoing a change in pK_a on folding, ΔG_{charge} refers to changes in charge-charge interactions between ionized groups on the surface, ΔG_{ϕ} is the hydrophobic free energy change, ΔH_{ψ} refers to a variety of hydrogen bonding effects, Δs_{config} is the average entropy change for each of n_c residues on folding, and $\Delta s_{\text{release}}$ is the average entropy change for each of n_w water molecules freed from interactions with surface polar groups on folding.

Of these, ΔG_{charge} is usually assumed to be zero near the isoelectric point of a protein (21). Complete burial of charges as salt bridges, or partial burials in surface cavities, are likely to contribute significantly through increased effective charge-charge interactions in regions of significantly lowered effective dielectric constant. At present there seems no way of estimating reliably this contribution; a value of ~-5kcal mol^{-1} has been given for ΔG_{titr} for ribonuclease A (21). Although this is a very small value in relation to most of the other terms in Eq. (1) (Table IV), it still amounts to 25–50% of the total $\Delta G_{E\to F}$.

The change in chain configurational entropy Δs_{config} is in contrast large. Again it is difficult to estimate, depending upon the configurational freedom of the extended chain. Using estimates of between 2 and 5 cal K^{-1} (residue mol)$^{-1}$(22) suggests an unfavorable free energy increase on folding of lysozyme of 70–200 kcal mol^{-1} (Table II).

TABLE II

ESTIMATED* ACTUAL AND POTENTIAL HYDROPHOBIC CONTRIBUTIONS ΔG_ϕ

Protein	Buried apolar area	ΔG_ϕ	Apolar area remaining exposed	Unrealized§ ΔG_ϕ	$-T\Delta S_{config}$‡
	(Å²)	(*kcal mol⁻¹*)	(Å²)	(*kcal mol⁻¹*)	(*kcal mol⁻¹*)
RNase-S	5,730	126	3,220	71	70–190
Lysozyme	7,070	156	2,750	61	70–200
Myoglobin	9,430	207	3,850	85	85–240

*Calculated from reference 2 using hydrophobicity data from reference 24.

‡See text.

§Estimated additional hydrophobic contribution were all the apolar area buried.

We now discuss the remaining terms in Eq. 1 in the light of the results of the previous section.

Hydrophobic Interactions ΔG_ϕ Nozaki and Tanford (23) have derived values for the free energy of transfer ("hydrophobicity") of amino acid side chains from organic solvent (mainly ethanol) to water. By correlating this data with Lee and Richards' (2) accessible surface area data, Chothia (24) proposed that apolar groups in proteins could be allocated a hydrophobicity of 22 cal (Å² of accessible area)$^{-1}$. Although the assumptions involved are questionable, we use this figure together with Lee and Richards' data on changes in accessibility on folding, to estimate the values of ΔG_ϕ given in Table II. Two points can be made from the table: (*a*) ΔG_ϕ estimates fall in the mid to upper end of the estimated $T\Delta S_{config}$ range. Although the entropy estimate is insufficiently precise for definite conclusions to be drawn, the hydrophobic driving force may be inadequate to overcome the entropic loss from chain folding. (*b*) Were all apolar groups buried, ΔG_ϕ would increase further by some 40–60%. ΔG_ϕ is presumably prevented from being maximized by consequent energetically-unfavorable effects which would more than offset the increased hydrophobic gain. Such constraining effects could be of several kinds: for example, the connectedness of the protein may make such additional burial topologically impossible, or the internal and surface hydrogen bonding pattern might be seriously disrupted thereby. The magnitude of the hydrophobic contribution is thus conditional on satisfying other constraints.

Internal and Surface Polar Interactions ΔH_ψ We noted earlier that all polar groups in the extended chain are exposed sufficiently to solvent to permit (geometrically) full hydrogen bonding to solvent. Assuming each such interaction is of ideal geometry, we estimate each contributes to the energy of the unfolded conformation an energy of ~4 kcal mol^{-1}, this being an average of several *ab initio* calculations (25–27) which is consistent with spectroscopy (22, 28). (For RNase and lysozyme, this scales up to some 2,200–2,400 kcal mol^{-1}.) This may well be an underestimate by perhaps 1–2 kcal mol^{-1}, but it provides a reasonable base from which changes can be discussed. A larger (numerical) value will increase the energetic significance of hydrogen bonding interactions. Moreover, if the extended chain polar group-water hydrogen bonding is incomplete, as suggested by Biltonen's calorimetric measurements (29), then any extensive hydrogen bonding in the native structure made either internally and/or to solvent will contribute significantly to the stability of the folded structure.

In the folded conformation all polar groups were optimally hydrogen bonded either internally or to solvent, and made hydrogen bonds of equal strength, the hydrogen bond contribution to the folded energy would be identical to that of the unfolded state, and ΔH_ψ

would be zero. However, at least four factors need to be considered, each of which could be the source of energy changes of several times the total experimental $\Delta G_{E\rightarrow F}$, and commensurate with the entropic and hydrophobic effects already discussed.

Relative strengths of hydrogen bonds. Polar group-polar group (PP) interactions may be significantly stronger than those assumed made to solvent (PW) in the extended chain. Although estimated errors are large (up to 2–3 kcal mol^{-1} for an interaction of 4–7 kcal mol^{-1}), there is quantum mechanical evidence for peptide hydrogen bonds being up to 20% stronger than the average interaction with water (30). Assuming additionally no difference in strength between ideal PW and WW interactions (PW $\simeq$ WW), a conservative figure for an additional strength of 0.5 kcal mol^{-1} (~12%) for PP over PW interactions yields for the 150 internal hydrogen bonds in RNase-S a contribution to the stability of the folded conformation of ~75 kcal mol^{-1}. This is over half the estimated hydrophobic ΔG_ϕ (Table II). A slightly hydrogen bond, would increase this contribution to stability, possibly even to dominance over ΔG_ϕ. The estimated ΔG_ϕ of 126 kcal mol^{-1} (Table II) is equivalent to ~0.85 kcal mol^{-1} per internal PP interaction; this represents an average PP hydrogen bond only 21% stronger than a PW interaction.

Internal hydrogen bond distortion. Any departure from idealized hydrogen bond geometry will tend to offset the above effect. Table III shows the RMS deviations from ideality of distance and angles of the peptide hydrogen bonds in PTI, together with the corresponding estimated energy penalties. Even making the maximum justifiable allowance for coordinate errors (figures in brackets), the energy penalty of distortion is far from negilgible. For the 150 internal hydrogen bonds in RNase-S, the total stability loss from this source could amount to 100–200 kcal mol^{-1} (see footnote Table III).

Polar group unsaturation. Only a minority of polar groups in the folded structures of lysozyme and RNase-S either fail to hydrogen bond internally, or are insufficiently exposed to solvent to make up for unmade internal interactions (see Results). Assuming as before full hydrogen bond saturation in the extended chain, the energetic consequences of unsatisfied polar groups in the folded structure are serious. Each unmade hydrogen bond carries a penalty of ~0.5 $\times$ 4 kcal mol^{-1}; the total penalties for RNase-S and lysozyme (using data quoted in Results) may amount thus to some 70 kcal mol^{-1} and 24 kcal mol^{-1}, respectively, from this source. Although compensating mechanisms mentioned previously (see Results and reference 5) may reduce this penalty, they cannot match energetically the lost hydrogen bond.

A highly saturated, minimally distorted hydrogen-bonding network thus appears to be a severe constraint on possible folded structures. Valuing each lost "half hydrogen bond" at ~2

TABLE III
DEVIATIONS FROM IDEALITY OF PEPTIDE HYDROGEN BONDS IN PTI

Variable (ideal)*		RMS deviation‡ from ideal		Estimated energy penalty§	
				(kcal mol⁻)	
$R_{O\cdot N}$	(2.85 Å)	0.17 Å	(0.0 Å)	0.4	(0.0)
NÔC	(120°)	27°	(21°)	0.6	(0.3)
NĤO	(180°)	26°	(20°)	0.8	(0.5)

*From reference 31.

‡Bracketed figures are estimated values after maximum correction for coordinate errors.

§Estimated using data taken from quantum mechanical calculations on the OH—O hydrogen bond (50–52; P. Barnes, J. L. Finney, and J. E. Quinn, manuscript in prepartion). The quoted penalties are thought likely to be an underestimate for NH—OC interactions.

kcal mol^{-1}, we need to lose only ~60 (or one polar group in ten) to wipe out completely the hydrophobic free energy change ΔG_ϕ (Table II). Only 5–10 unsaturated polar groups represent the same energy as the free energy (10–20 kcal mol^{-1}) that stabilizes the whole tertiary structure. We might note in passing the possible significance for amino acid substitutions and protein evolution in general.

In this context, the strong correlation observed between exposed area of polar groups and unsatisfied hydrogen bonding capacity (see Results and references 5 and 14) seems not only reasonable, but necessary. Although uncertainty remains concerning the relative strengths of the various internal and surface hydrogen bonds, the solvent's "hydrogen-bond sink" role of providing needed interactions seems significant. The fluidity, structural versatility, and double donor-double acceptor capacity of water enables it to fulfil this role very efficiently.

Entropy loss from release of extended-chain "bound" water, $T\Delta s_{\text{release}}$. For every polar group-water interaction transformed into an internal hydrogen bond, there is release of solvent to the surrounding fluid medium, with a consequent favorable entropic gain to the whole system. Quantitative estimates of the magnitude of this effect are extremely difficult, depending upon the restrictions on rotational and translational motion experienced by the (rapidly exchanging) water molecules concerned in both folded and unfolded conformations. A maximum limit can be deduced from the melting entropy of ice Ih; this value of 5.26 cal $K^{-1}mol^{-1}$ (32) corresponds to a $T\Delta s_{\text{release}}$ of ~1.6 kcal mol^{-1} at room temperature. This seems very large—almost comparable to the energy loss from an umade internal hydrogen bond (see above). Although nuclear magnetic resonance evidence (33) suggests a few waters in the folded conformation may well be sufficiently strongly bound to contribute to this extent, most will be less restricted. An average $T\Delta s_{\text{release}}$ of about a quarter of the ice value scales up to ~230 kcal mol^{-1} for the ~300 water molecules released on folding by ribonuclease and lysozyme. Even though this value probably underestimates the effect, it is again comparable with our hydrophobic ΔG_ϕ estimate.

Relative Magnitudes of Contributions to $\Delta G_{E\to F}$ Table IV collects together the various estimates made above, and lists the losses and gains likely for RNase-S from the several effects discussed.

It should be stressed that the errors on the quoted estimates are large—in almost all cases greater than the total $\Delta G_{E\to F}$. This underlies the necessary lack of certainty in any assertion of the dominance of one particular effect. Rather, apart from the charge-related contributions (which can still be of the order of $\Delta G_{E\to F}$ for cases of significant buried charge or for proteins away from their isoelectric points), all estimated contributions are of the same order of magnitude (positive or negative), and much greater than the total free energy of folding $\Delta G_{E\to F}$. We note also that neither the hydrophobic nor hydrogen-bonding contributions are independently maximized; the folded structure appears to be a compromise in which further augmenting one contribution would have serious consequences for the other which would more than offset the expected marginal gain.

Thus, we cannot assert that a protein owes the stability of its tertiary structure to a single effect; its free energy of stability is a small difference between half a dozen or so effects whose individual contributions to increasing or decreasing stability are much larger than the final difference. The stability is an extremely delicate one to which hydrophobic effects, hydrogen bond distortion and unsaturation, chain configurational entropy, and entropy of release of solvent all contribute. None can be said to dominate.

The near perfection of the three-dimensional jigsaw puzzle that is the folded protein is impressive, particularly with respect to the efficiency of internal hydrogen-bonding. This

TABLE IV

ESTIMATED CONTRIBUTIONS TO THE FREE ENERGY OF FOLDING OF RNASE-S: A TENTATIVE BALANCE SHEET*

Effects likely to stabilize folded conformation (gains)	Quantity estimated (Eq. 1)	Estimated magnitude (kcal mol^{-1}) (rounded to nearest 5 kcal mol^{-1})
Hydrophobic interactions	ΔG_ϕ	−130
Differences between polar-polar and polar-water interaction strengths	$\Delta H_{\phi 1}$‡	−75
Entropy of release of water	$-n_w T\Delta s_{\text{release}}$	−120
Charged groups	ΔG_{titr}	0
	ΔG_{charge}	−5
Effects likely to destabilise folded conformation (losses)		
Configurational entropy loss of chain	$-n_c T\Delta s_{\text{config}}$	+70 to +190
Hydrogen bond distortion	$\Delta H_{\phi 2}$‡	+100 to +200
Unsaturated polar groups	$\Delta H_{\phi 3}$	+70§
Total balance (Exp.)	$\Delta G_{E\to F}$	−10 to −20

*Probably minor contributions such as dielectric effects on polar group-water interactions in surface cavities, and possible changes in the strength of water-water interactions on moving from the extended-chain protein-water interface region to the bulk, have been ignored.

‡The ΔH_ϕ term in Eq. 1 is here split into three separate terms. The $P\Delta V$ term is negligible.

§Possible compensating effects discussed in the text are not allowed for.

suggests that maintaining efficient hydrogen bonding is a severe constraint on maximizing other contributions (in particular the hydrophobic driving force). That each single unmade hydrogen bond may well carry a penalty of 10–20% of the total stability of the molecule underlines this. In maintaining the hydrogen bonding complement, the solvent appears to play an important role as a hydrogen bond sink, facilitating hydrogen bonds at the surface (and internally in certain cases [14, 18]) where good internal interactions cannot otherwise be made. In addition, the double donor-double acceptor nature of the water molecule enables it to soak up the hydrogen donor-acceptor imbalance (the ratio being 1.4:1.0 in ribonuclease and lysozyme) among the polar groups of the molecule itself.

One possible effect not listed explicitly in Table IV concerns the strength of internal dipole-induced dipole interactions between neighboring polar and apolar groups. This effect is implied by the statistics of internal contacts shown in Table I and Fig. 2. Although much information exists in the literature on so-called "CH-hydrogen bonds" (34, 35), the possibilities of such interactions within proteins is normally ignored. It has been further argued that an internal peptide hydrogen bond presents an effectively hydrophobic surface to its surroundings (24).

The enhanced strength of certain polar-apolar group interactions implied in Fig. 2 is, however, supported both by other experimental observations and theoretical calculations. For example, although "bifurcated H-bonds," in which the NH group acts as donor to two acceptors have been described in x-ray studies, accurate location of the hydrogen (or deuterium) by neutron diffraction supports an interpretation in terms of a second non-H-bonded interaction that is weaker than a hydrogen bond, but significantly stronger than a

classical nonbonded interaction (31, 36). This is inconsistent with the peptide hydrogen bond presenting an apolar surface to its surroundings.

Estimating the quantitative significance of these possible enhanced polar-apolar interactions is less straightforward. Using a classical polarizability calculation for a model polar-apolar system (here the methane-water dimer), our preliminary calculations (P. Barnes, J. L. Finney, and J. E. Quinn, unpublished observations), show an additional dipole-induced dipole contribution of 0.12 (at 4.0 Å separation) to 0.32 kcal mol^{-1} (at 3.4 Å separation). A comparison between the magnitude of this energy and the straight van der Waals interaction is made difficult by the wide range in published van der Waals parameters (37–40): quoted well-depth for $CH...CH_2$ interactions range between 0.04 and 0.28 kcal mol^{-1}. Corresponding reductions in contact distances (Fig. 2) between CH_n and NH groups estimated from dipole-induced dipole interactions would vary between 0.1 and 0.7 Å less than the expected van der Waals radius sum. Although each individual polar-apolar interaction is thus relatively weak (this estimate of 3–8% of the average energy of a hydrogen bond certainly precludes us from calling it a hydrogen bond), the large number of such interactions within the protein (Table I) suggests the total effect may not be insignificant. For example, they could help buffer the protein against disadvantageous local configurations such as unsatisfied polar groups: rather than pay a full penalty of 2 kcal mol^{-1}, we might save ~0.5 kcal mol^{-1}(see above).

How much of this effect is included already in the free energy of transfer data used to estimate hydrophobicities of side-chains (23) is not clear: the solvent serving as a model for the protein interior (ethanol) contains one polar group, and thus will at least partly allow for the energetic consequences of such polar-apolar contacts. The relevance of this discussion to direct apolar group-water interactions on the protein surface is also not clear. More extensive quantitative work, e.g., very accurate quantum mechanical calculations, is needed to evaluate the significance or otherwise of this possible effect.

Dynamics, Denaturation, and Structural Metastibility

For small perturbations around the native conformation we need to discuss energy changes consequent upon small distortions. Of those effects listed in Table IV, only hydrophobic interactions and hydrogen bond distortion are likely to be significant; in addition, weaker local interactions such as van der Waals, or the dipole-induced dipole effects discussed above, need be considered.

Hydrogen Bond *Ab initio* molecular orbital calculations on effective force constants of the various hydrogen bonds involved in proteins are sparse; what information we do have suggests they are of the same order of magnitude of those of the water-water hydrogen bond. Using the data source quoted in Table III, an energy of kT(0.5 kcal mol^{-1} at 300 K) corresponds to distortions of 0.2 Å in bond length, and to 20° and 25° in hydrogen acceptor and donor angles, respectively.

Hydrophobic Interaction For a molecule of molecular weight ~14,000 (e.g., lysozyme, RNase) we expect an RMS volume fluctuation at room temperature of ~38 $Å^3$ $molecule^{-1}$ (41). Assuming this to be an isotropic breathing motion, this is equivalent to an RMS surface area fluctuation of ~7 $Å^2$ $molecule^{-1}$, or ~0.1% of the total surface area. Assuming all this additional exposure is of apolar surface, the corresponding free energy fluctuation would be ~0.15 kcal mol^{-1}. Alternatively, localizing the RMS volume fluctuation within a channel which is just accessible to solvent gives a maximum RMS area fluctuation of 50 $Å^2$. This corresponds to a maximum RMS hydrophobic free energy over the whole molecule of 1.1 kcal mol^{-1}.

The relative magnitudes of these two "effective force constants" is interesting; a localized thermal fluctuation of kT might expose ~25 Å^2 of additional hydrophobic surface (0.4% of the total exposed surface area), or stretch or bend a single hydrogen bond through only 0.2 Å or 20°–25°. Thus it appears that the overall shape, or basic curvatures, of the potential well in which the native structure is moving is determined largely by the stiffer hydrogen bonding interactions. These may act as local stiff but elastic "sheet anchors" restricting the flexibility of local regions within the protein. This is consistent with earlier suggestions that certain hydrogen bonds are necessary to ensure critical groups in a protein are optimally positioned for catalytic activity. Such hydrogen bonds involve in some cases internal water molecules, an internal "hydrogen bond sink." (14, 15, 28, 42, 43). Within this framework, the even finer-scale local topography of the potential well on the scale of kT or less will in turn depend on local even weaker, short-range van der Waals and perhaps the somewhat stronger dipole-induced dipole effects discussed earlier. This picture is consistent with that recently put forward by Privalov and Tsalkova (44) to rationalize the "microstability" of globular proteins.

Efficient denaturing of the molecule thus would appear to require disruption of the hydrogen bonded structure. Refolding from the extended chain would appear to be directed by the effectively "longer-range" hydrophobic effects, which drive the protein through a path in phase space to a region in which enough short-range polar interactions can "switch on" to stabilize the structure. The hydrophobic driving forces may control the folding path in its early stages, but do not control the final conformation.

Thus, once denatured, the structure may be expected to be stable (perhaps in some cases metastable) to some changes in the surrounding solvent, as long as the hydrogen-bonding structure, including interactions to solvent, is not thereby disrupted. This is consistent with experiments in which changing of the solvent subject to certain constraints of dielectric constant and ionic strength does not lead to denaturation (45–48). It is also possible to envisage removing a large part of the solvent without denaturing the protein. The internal hydrogen bonds would then have no competition from water which might lead to unfolding. There is indeed some experimental evidence in this regard (49).

SUMMARY

On the basis of detailed structural studies of several globular proteins, together with currently available estimates of the magnitudes of the thermodynamic quantities involved, the stability of the native structure cannot be said to be due largely to a single effect. A picture is developed of a very well-fitting but delicately balanced three dimensional jigsaw puzzle in which the internal hydrogen bonding is almost fully maximized, with potential solvent interactions making the native conformation potentially almost as fully hydrogen bonded as the extended chain. The estimated energy penalty associated with unsatisfied polar groups is so large with regard to the total stability of the folded structure (10–20%) that the maintenance of highly efficient hydrogen bonding is a strong constraint on maximizing the hydrophobic free energy change by apolar group burial. The relative magnitudes of these and other possible effects are considered; although extensive uncertainty remains in these estimates, the global stability is seen as due to the joint action of several different energetic and entropic effects. Mathematically, the problem is a complex one of global magnetization of several strongly-interacting variables of which the hydrophobic contribution is one which is of only average potential magnitude. The significance of the solvent is discussed in terms of a concept of a "hydrogen-bond sink," which facilitates otherwise impossible polar interactions both internally and on

the surface. The quantitative significance of internal dipole-induced dipole interactions is suggested.

Consideration of effective force constants of polar and hydrophobic interactions suggests that the overall curvatures of the potential well which controls the native protein's dynamics are determined largely by the relatively stiff polar interactions, again including solvent. The detailed dynamics within these constraints, which limit local movements, is probably dependent on the details of the local van der Waals interactions and probably also dipole-induced dipole forces which are probably somewhat stronger. The folding path would seem to be hydrophobically controlled, with hydrogen bonds (including solvent where necessary) switching on for multiple local docking procedures as relevant regions of the molecule come together. Considering the relative magnitudes of kT and the hydrogen bond energy, this must take place progressively rather than explosively: folding would seem to be a reconstructive transition of bond (polar group-solvent) breaking and reforming, there being insufficient thermal energy to break more than a few such interactions simultaneously. Similarly, thermal denaturation appears as a progressive, cooperative transition. The ability to change the solvent within limits without disrupting the structure is consistent with the ideas developed; the folded structure may be only metastable in such changed environments.

Received for publicaton 27 November 1979.

REFERENCES

1. Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* **14**:1–68.
2. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**: 379–440.
3. Chothia, C. 1975. Structural invariants in protein folding. *Nature (Lond.).* **254**:304–308.
4. Richards, F. M. 1978. Areas, volumes, packing and protein structure. *Ann. Rev. Biophys. Bioeng.* **6**:151–176.
5. Finney, J. L. 1978. Volume occupation, environment and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* **119**:415–441.
6. Shrake, A., and J. A. Rupley. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**:351–372.
7. Finney, J. L. 1975. Volume occupation environment and accessibility in proteins. The problem of the protein surface. *J. Mol. Biol.* **96**:721–732.
8. Finney, J. L. 1979. A procedure for the construction of Voronoi polyhedra. *J. Comp. Phys.* **32**:137–143.
9. Wyckoff, H. W., P. Tsernoglou, A. W. Hanson, J. R. Knox, B. Lee, and F. M. Richards. 1970. The three-dimensional structure of ribonuclease-S. Interpretation of an electron-density map at a nominal resolution of 2 Å. *J. Biol. Chem.* **245**:305–328.
10. Blake, C. C. F., G . A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. 1967. On the conformation of the hen egg-white lysozyme molecule. *Proc. R. Soc. London B. Biol. Sci.* **167**:365–377.
11. Deisenhofer, J., and W. Steigemann. 1975. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Cryst.* **B31**:238–250.
12. Richards, F. M. 1974. The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**:1–14.
13. Watenpaugh, K. D., L. C. Sieker, J. R. Herriott, and L. H. Jensen. 1973. Refinement of the model of a protein: rubredoxin at 1.5 Å resolution. *Acta. Cryst.* **B29**:943–956.
14. Finney, J. L. 1979. The organization and function of water in protein crystals. *In* Water: A Comprehensive Treatise. F. Franks, editor. Plenum Press, Inc., New York. **6**:47–122.
15. Birktoft, J. J., and D. M. Blow. 1972. Structure of crystalline α-chymotrypsin. *J. Mol. Biol.* **68**:187–240.
16. Birktoft, J. J., D. M. Blow, R. Henderson, and T. A. Steitz. 1970. The structure of α-chymotrypsin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **257**:67–76.
17. Bode, W., and P. Schwager. 1975. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. *J. Mol. Biol.* **98**:693–717.
18. Finney, J. L. 1977. The organization and function of water in protein crystals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **278**:3–32.
19. Golton, I. C. 1980. An experimental and theoretical study of the interaction between water and a globular protein. Ph. D. Thesis, University of London.

20. Finney, J. L. 1975. The structure of laboratory and computer-built random packings. *J. Phys. Colloque.* C2, suppl. No. 4. 36: C2-1-C2-11.
21. Pain, R . H. 1978. The conformation and stability of folded globular proteins. *In* Characterization of Protein Conformation and Function. F. Franks, editor. Symposium Pres, London. 19–36.
22. Brandts, J. F. 1969. Conformational transitions of proteins in water and aqueous mixtures. *In* Structures and Stability of Biological Macromolecules. S. N. Timasheff and G. D. Fasman, editors. Marcel Dekker Inc., New York. 213–290.
23. Nozaki, Y., and C. Tanford. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* **246**:2211–2217.
24. Chothia, C. 1974. Hydrophobic bonding and accessible surface area in proteins. *Nature (Lond.).* **248**:338–339.
25. Diercksen, G. H. F., W. P. Kraemer, and W. van Niessen. 1972. SCF-MO-LCGO studies on hydrogen bonding. Ammonia-water system. *Theor. Chim. Acta.28:67–74.*
26. Dill, J. D., L. C. Allen, W. C. Topp, and J. A. Pople. 1975. A systematic study of the nine hydrogen-bonded dimers involving NH_3, OH_2 and HF. *J. Am. Chem. Soc.* 97:7220–7226.
27. Johansson, A., and P. A. Kollman. 1972. Amide-water hydrogen bonding. *J. Am. Chem. Soc.* **94**:6196–6198.
28. Eagland, D. 1975. Nucleic acids, peptides and proteins. *In* Water: A Comprehensive Treatise. F. Franks, Editor. Plenum Press, Inc., New York. 4:305–518.
29. Biltonen, R. L. 1976. The influence of water on interacting biological systems. *In* Colloques Int. CNRS No. 246, "L'eau et les systemes biologiques," CNRS, Paris. 13–18.
30. Schuster, P. 1976. Energy surfaces for hydrogen bonded systems. *In* The Hydrogen Bond. P. Schuster, G. Zundel, and C. Sandorfy, editors. Elsevier North-Holland, Amsterdam. 1:25–163.
31. Olovsson, I., and P.-G. Jönsson. 1976. X-ray and neutron diffraction studies of hydrogen bonded systems. *In* The Hydrogen Bond. P. Schuster, G. Zundel, and C. Sandorfy, editors. Elsevier North-Holland, Amsterdam. 2:393–456.
32. Rossini, F. D., D. D. Wagman, W. H. Evans, S. Levine, and I. Joffe. 1952. Selected values of chemical thermodynamic properties. Circular of the National Bureau of Standards 500.
33. Cooke, R., and I. D. Kuntz. 1974. The properties of water in biological systems. *Ann. Rev. Biophys. Bioeng.* 3:95–126.
34. Sutor, J. D. 1963. Evidence for the existence of C-H---O hydrogen bonds. *J. Chem. Soc.* **1963**:1105–1110.
35. Green, R. D. 1974. Hydrogen Bonding by CH Groups. Macmillan, London.
36. Jönsson, P.-G. and Å. Kvick. 1972. Structure of protein and nucleic acid components. III. *Acta Cryst.* **B28**:1827–1833.
37. Scott, R. A., and H. A. Scheraga. 1966. Conformational analysis of macromolecules. III. Helical structures of polyglycine and poly-L-alanine. *J. Chem. Phys.* **45**:2091–2101.
38. Momany, F. A., L. M. Carruthers, R. F. McGuire, and H. A. Scheraga. 1974. Intermolecular potentials from crystal data. III. *J. Phys. Chem.* **78**:1595–1620.
39. Levitt, M. 1974. Energy refinement of hen egg-white lysozyme. *J. Mol. Biol.* **82**:393–420.
40. Hirschfelder, J. O., C. F. Curtiss, and R. B. Bird. 1954. Molecular theory of gases and liquids. John Wiley & Sons, New York.
41. Cooper, A. 1976. Thermodynamic fluctuations in protein molecules. *Proc. Natl. Acad. Sci. U.S.A.* **73**:2740–2741.
42. Fehlhammer, H., W. Bode, and R. Huber. 1977. Crystal structure of bovine trypsinogen at 1.8 Å resolution. *J. Mol. Biol.* **111**:415–438.
43. Wright, H. J. 1973. Comparison of the crystal structures of chymotrypsinogen-A and α-chymotrypsin. *J. Mol. Biol.* **79**:1–11.
44. Privalov, P. L., and T. N. Tsalkova. 1979. Micro- and macro-stabilities of globular proteins. *Nature (Lond.).* **280**:693–696.
45. Alber, J., G. A. Petsko, and D. Tsernoglou. 1976. Crystal structure of elastase-substrate complex at -55°C. *Nature (Lond.).* **263**:297–300.
46. Douzou, P., G. H. B. Hoa, and G. A. Petsko. 1975. Protein crystallography at sub-zero temperatures: lysozyme-substrate complexes in cooled mixed solvents. *J. Mol. Biol.* **96**:367–380.
47. Fink, A. L. 1974. Effect of dimethyl sulfoxide on the interaction of proflavine with α-chymotrypsin. *Biochemistry.* **13**:277–280.
48. Fink, A. L, and I. A. Ahmed. 1976. Formation of stable crystalline enzyme-substrate intermediates at sub-zero temperatures. *Nature (Lond.)* **263**:294–297.
49. Yang, P.-H., and J. A. Rupley. 1979. Protein-water interactions, heat capacity of the lysozyme-water system. *Biochemistry.* **18**:2654– 2661.
50. Popkie, H., H. Kistenmacher, and E. Clementi. 1973. Study of the structure of molecular complexes IV. *J. Chem. Phys.* **59**:1325–1336.
51. Diercksen, G. H. F., W. P. Kraemer, and B. Roos. 1975. SCF-CI studies of correlation effects on hydrogen bonding and ion hydration. *Theor. Chim. Acta.* 249–274.

52. Hankins, D., J. W. Moskowitz, and F. H. Stillinger. 1970. Water molecule interactions. *J. Chem. Phys.* **53**:4544–4554.
53. Mackay, A. L., J. L. Finney, and K. Gotoh. 1977. The closest packing of equal spheres on a spherical surface. *Acta Cryst.* **A33**:98–100.
54. Finney, J. L. 1968. Random packings and the structure of the liquid state. Ph. D. Thesis, University of London.
55. Barnes, P., J. L. Finney, J. D. Nicholas, and J. E. Quinn. 1979. Cooperative effects in simulated water. *Nature (Lond.).* **282**:459–464.
56. Barnes, P., D. V. Bliss, J. L. Finney, and J. E. Quinn. 1980. Cooperative and quadrupole effects in order-disorder transitions in crystalline ices. *Farad. Disc. Chem. Soc.* **69**. In press.

DISCUSSION

Session Chairman: V. Adrian Parsegian *Scribe*: Carolyn Ritz-Gold

NÉMETHY: The discussion and the data in Table IV certainly illustrate the concept that various interactions contribute to the free energy of folding of a protein, and that several of them may be of significant magnitude. In view of the large uncertainties in the numbers cited, do you see a way to compare the effects quantitatively? The errors are so large that combination of the numbers in the table could justify any conclusion about the overall stability of the protein; even the relative importance of the various contributions could be assessed in widely different ways.

FINNEY: There are certainly limitations to this type of analysis since we don't have the quantitative information that we need. However, we can certainly improve on the theoretical basis of hydrogen bonding interactions. We have learned a lot about the nature of the water-water hydrogen bond, both experimentally and theoretically, and quantum mechanical techniques are now sufficiently well developed to formulate good models of simple hydrogen bonding interactions. The force constant data that I showed can be extended to polar groups in general. Adding such information to the crystallographic data may tell us whether the tentative ideas proposed here on hydrogen bond distortions, and strength differences between polar-polar and polar-water by interactions, are realistic. Moreover, by using such derived potential functions in simulation calculations of water-protein systems and protein dynamics we might learn something about what may be occurring in almost real systems.

NÉMETHY: The possible differences between various hydrogen bonds (PP, PW, WW) are at least as large as the other effects cited, such as distortions. You distinguish between the polar-polar, polar-water, and water-water hydrogen bond strengths and assign definite orders of differences between them. What was this based on? What is the rationale for assuming the strengths of polar-water and water-water interactions are equal?

FINNEY: This was based on information whose sources are indicated as footnotes to the text. The polar-polar and polar-water interaction data are averages of information from a variety of quantum mechanical calculations. They show wide variations but the emerging pattern suggests that my estimate of their relative strengths is reasonable. As far as the polar-water and water-water interactions are concerned, a similar assessment suggests that there is inadequate justification to assert a significant difference in strength.

NÉMETHY: The differences are consistent with earlier estimates, for example, that polar-polar interactions are stronger than polar-water interactions and thus help to stabilize the folded molecule. Can quantum mechanical calculations definitely show sufficient differentiation?

FINNEY: I think so, provided great care is taken that the calculations are taken beyond the Hartree-Fock level.

NÉMETHY: In your attempt to assess the relative effects of hydrogen bonds and hydrophobic interactions on dynamics and on the overall shape of the potential surface (and consequently to estimate the stage of folding at which each interaction is dominant), either the free energy of individual interactions was compared, or the total estimated free energy was given, averaged over the whole molecule. In neither form does the estimation take into account the fact that these interactions are not distributed uniformly within the protein molecule, but that there is some local clustering of nonpolar and polar groups. Such clustering has been described earlier by Kuntz. It also was discussed in a recent model derived by Wüthrich from his NMR studies. Consequently, certain steps of conformational changes may involve primarily one or the other of these types of interactions, at early or at late stages of folding. Either the structure or the folding dynamics of local regions might be dominated by one or the other type of interaction.

FINNEY: What you suggest may turn out to be the case, but the state of the art of correlating structural features and thermodynamic measurements is not yet up to making such assertion, as I imply in the discussion text. Considering the mine field I am walking in with the numbers I have shown, I would certainly not attempt to refine the statements I made at the level you mention.

KUNTZ: You base the hydrogen bond distortion picture on trypsin inhibitor and that structure is probably not at the level of resolution of, say, rubredoxin for example.

FINNEY: Pancreatic trypsin inhibitor is one of the few structures refined to a resolution high enough to justify such calculations as hydrogen bond distortion. We have done similar calculations on lysozyme and ribonuclease S, but I would accept your point that one would not want to put too much emphasis on these numbers considering the less good resolution of this latter data. I have included in the numbers quoted an estimate of what the likely errors are for a structure of 1.5 Å resolution. Certainly one should look at more precise data when it is available. Rubredoxin is an obvious case in point. In fact, we are processing it for other reasons and similar hydrogen-bond data will come out of it.

KUNTZ: You speak of a mine field in terms of the uncertainty of some of your numbers. I think we all agree that that's a real problem, but beyond that there is the free floating factor of temperature; the heat capacity is a large number in all of these protein water systems. An analysis at one temperature is fraught with real interpretative problems if one changes the temperature even 10–20 degrees. I'm sure you have thought about this problem.

FINNEY: Yes. What I have given here is a framework which holds as far as it goes. I wouldn't want to push it any further without much more reliable data. Concerning heat capacities, it can be argued quite reasonably that the heat capacity changes during folding are largely due to the so-called hydrophobic effects plus released water molecules from polar groups, but we cannot quantify the values in molecular terms. There is certainly more information which would be obtained by looking at the temperature variation. One could also look at temperature variations on the dynamics of a simple protein and see what difference this makes in the contributions from some of these internal interactions. But this requires more realistic potential functions than are currently available, and is not worth doing until we have more reliable potentials.

KUNTZ: The real difficulty here is that as you go towards zero degrees, both ΔH and ΔS rapidly approach zero.

FINNEY: You would see a change in the balance of contributions. There is information there to extract.

B. K. LEE: I agree with your conclusion that the relatively small ΔG between native and denatured protein is the result of addition and subtraction of many large terms. The situation is similar to the one in quantum mechanics where one tries to calculate the bond energy by balancing out many large terms. I'm struck by your comment that you hope to obtain more accurate information from quantum mechanical studies. One has inaccuracies in the results of quantum mechanics, and if one uses this to solve our problem, errors will mount.

My first question has to do with this. As in quantum mechanics, there is a partition problem here when one tries to divide the total ΔG into a number of factors. One has to be careful in this process neither to leave out nor duplicate a term arising from one physical effect. For example, shouldn't one include the reduction in surface area due to polar groups, as well as that due to apolar groups, in the hydrophobic term? A second example is in comparing the hydrogen bond strengths between polar groups and water and polar-polar groups inside the protein. When hydrogen bonds between polar groups and solvent water are broken, there will be an increase in entropy because water is released. Should this term be included in your hydrogen bond strength estimation when you change the polar-water interaction into a polar-polar internal hydrogen bond?

FINNEY: I agree that there is a problem taking small differences from two large numbers, but I think that we are now able to obtain answers to 20% and that this is a significant improvement over our current knowledge. We are now setting up on the Cray 1 at Daresbury a large configuration interaction calculation looking at cooperative effects in water trimers. Although this is even more problematical than the hydrogen-bonded pair, we think we will obtain useful estimates. So I don't think the difficulties are as great as you suggest though I will agree the method must be treated with great respect.

B. K. LEE: Do you have an estimate of absolute error in your calculation of the quantum mechanics? How does one know what the true value is?

FINNEY: You don't; you have comparisons. One can make use of the accumulated knowledge of *ab initio* calculations beyond the Hartree-Fock level to get a feeling for what good basis sets are and how best to account for configuration interactions. I agree that it is not a simple problem, but I think we do have adequate pointers to tell us that we can in fact do something about it.

B. K. LEE: My comment has to do with the assignment of a certain type of force that determines the overall curvature of the potential curve. Here one has to be careful about the coordinate system used. What curvature are you talking about? The potential energy surface is not a one-dimensional surface, but is multidimensional, and the

curvature depends on the coordinate one chooses. If, for instance, you chose as your coordinate the radius of the whole molecule, then the hydrophobic force may still be important in determining the curvature of the potential energy with respect to that coordinate, which is simply the size of the molecule. One must choose the coordinate according to the kind of physical phenomenon one is interested in. The nature of the force that determines the curvature of the potential curve along a particular direction will generally vary.

FINNEY: I prefer to treat the interaction of polar groups with water differently from the interaction of apolar groups with water. The polar interactions are discussed in terms of the direct bonding which occurs, together with a solvent-release term, rather than to lump together a number of entropic and enthalpic factors as is normally (probably inadequately) done in the apolar group case. With respect to loss of solvent, an estimate of this term is given in the paper; it is potentially a very large contribution. We need experimental data on model systems to obtain realistic estimates of the contribution. With respect to the curvature of the potential energy surfaces, I mean N-dimensional surfaces defined by, e.g., the Cartesian ccordinate system of the components. So it's a very general statement which can be made more specific if you wish.

B. K. LEE: Did you include in your term the change in van der Waals forces i.e., induced dipole-induced dipole interactions? There are going to be changes in that, and I wonder if you have an estimate.

FINNEY: The van der Waals interactions I have skated around, because there's a lot I do not understand. Although a typical van der Waals interaction inside the protein may be relatively weak (in the order of 0.1–0.3 kcal/mol), there are a lot of them—up to a dozen around a given atom. On the other hand, a polar group generally makes no more than 4 classical hydrogen bonds, and normally less. We might consider also van der Waals interactions involving the polar groups themselves; these are probably stronger than induced dipole-induced dipole interactions.

I would appreciate comments on Fig. 2, the ideas of which arose from work we did earlier with Fred Richards' RNAse S data. We found many "abnormally" close contacts. Similar effects were found with the higher resolution insulin data, which has been refined very carefully without making restrictive chemical assumptions. The point I want to show is the distribution of contact distances between tetrahedral carbons and nitrogens. There is a peak at ~3.1 Å. The van der Waals sum for those pairs is at 3.4-3.8 Å, depending on your preferred set of constants. So the indication is that the interaction in those groups is somewhat stronger than the classic van der Waals interaction. Preliminary calculations on the kind of energy you might expect for such dipole-induced dipole interactions, suggest a fairly significant dipole can be induced on, say, a methane molecule, by a neighboring permanent dipole. Energies of between 0.5 and 1 kcal/mol might well be realistic, depending on the parmeters assumed. These possible additional contributions are, individually, small. However, when scaled up by the total polar-apolar interaction area within the molecule (Table I) (larger than that for the other types of interaction) the total contribution may be significant.

B. K. LEE: The reduction in surface area due to polar groups should also be included. If you consider the ΔG for the process of conversion of a denatured protein in an aqueous solution to its native state in that solution, this process can be broken down as follows: (*a*) lifting the denatured protein out of the solvent (without changing anything else) and creating a hole in the solvent plus the denatured "gas" molecule ($-X_D$); (*b*) converting the denatured "gas" molecule to the native "gas" molecule (ΔI); (*c*) converting the hole left in the solvent by the denatured molecule into a hole that would fit the native one ($K \Delta A$); and (*d*) dumping the native "gas" molecule into this hole (X_N) to get the native protein in aqueous solution. The ΔG associated with process (*c*) can then be calculated. It's difficult to break ΔG down into ΔH and ΔS using certain kinds of statistical mechanical procedures, e.g., scaled particle theory, and then to say these quantities are roughly proportional to the area. This ΔG has many of the properties commonly associated with hydrophobic effects. This term depends little on the nature of the groups that lined the inside-wall of the cavity initially—it depends primarily on the amount of reduction in the surface area of the cavity. This, then, is my argument for including a term that reflects the reduction in the surface area due to polar groups.

FINNEY: I suspect the ΔG for general cavity formation will be different from that used solely to try to quantify hydrophobic effects. The partitioning of the various contributions will also be different.

D. EISENBERG: The carbon-carbon and carbon-nitrogen distances plotted in Fig. 2 for insulin show a maximum for carbon-nitrogen and suggest that this is due to dipole-induced dipole interactions. What about carbon-oxygen? Does that show a similar maximum?

FINNEY: It shows a similar but weaker effect, presumably because there is no "spare" lone pair on the oxygen as there is on the nitrogen. The rubredoxin data show the same kind of statistics and in that case specific effects of secondary structure will be absent.

P. ROSS: Dr. Subramanian and I have looked at what is known about the thermodynamics of various protein association processes. Examining those processes in terms of the entropic and enthalpic contributions to the free

energy, the thermodynamic data seem to show that hydrogen bonding and what we loosely call van der Waals' interactions occurring in a low dielectric environment are mainly responsible for the observed thermodynamic parameters. We conclude that the exact location of these interactions will be important in understanding the thermodynamic data in terms of structure. Finally, the temperature effect brought up by Kuntz appears to us to be primarily a property of the aqueous solvent.

FINNEY: Concerning your first point, I tried to ignore that because of the problems involved with the concept of dielectric constant at the molecular level. Rufus Lumry calls such an operation "bootstrapping," and argues that is illegal. If the effect is significant, it further de-emphasizes hydrophobic interactions with respect to hydrogen-bonding effects.

S. SUBRAMANIAN: In Table IV, the values given in the first part of the table do not show any uncertainty associated with them (rounded to nearest 5 kcal/mol), while those in the second part cover a wide range resulting in a large uncertainty when the total balance is estimated. This is another limitation of these calculations. Secondly, for the hydrophobic contribution, you make a statement that it is not clear how much of the van der Waals interactions is included in the free energy of transfer. Could other transfer experiments be devised, e.g. transfer of a methyl or other aliphatic group from a nonpolar solvent to water, and compare the transfer data with the ethanol-water transfer data of Tanford to get an estimate of the van der Waals interactions?

FINNEY: Your last point is an interesting possibility which could be pursued experimentally. We now know what the interior of a protein is like,. and it should be possible to model that by some mixed solvent and do such measurements. On your first point, you're right: I should have specified ranges for the values given in the first part of the Table. For the hydrophobic contribution, I used what is generally used in the literature, though I have severe reservations about the use of transfer free energies in this context. The other values must not be treated as more than order of magnitude estimates.