Solvation and Solubility of Globular Proteins

Hongwu Wang[†] and Arieh Ben-Naim*

Department of Physical Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received: May 30, 1996; In Final Form: August 22, 1996[⊗]

The solvation Gibbs energy of a globular protein is expressed in terms of its various ingredients. Estimating each of these ingredients leads to an estimate of the overall solvation Gibbs energy of globular protein. As expected, the contribution of the sum of all the hydrophilic groups on the surface of the protein makes the solvation Gibbs energy of the protein less positive. However, this may not be enough to make the solvation Gibbs energy negative. We found, quite unexpectedly, that correlations between pairs and triplets of hydrophilic groups substantially decrease the solvation Gibbs energy of the protein. Therefore we conclude that pair and higher order correlations between hydrophilic groups on the surface of the protein has a significant contribution to the high solubility of globular proteins.

1. Introduction

Knowledge of the solvation of proteins is an essential ingredient in the study of the driving forces for any biochemical process involving proteins, such as protein folding, protein−protein association, and binding of protein to nucleic acids. The solvation Gibbs energy of a protein is also a major factor in the determination of the solubility of the protein.

To function in an aqueous medium, a protein must be soluble. But what makes such a huge organic molecule soluble in water?

Consider a typical protein of 150−200 amino acid residues. There are some $20^{150}-20^{200}$ possible sequences of polypeptides of such length. We know that only a tiny fraction of this immense number of possibilities do function in a living cell. These are the ones that had survived evolution: they were probably selected not from the total number of possible sequences, but from the highly limited number of polypeptides that were soluble. We also know that globular proteins, when undergoing denaturation, i.e., breakdown of their 3-D structure, lose their solubility. Therefore, it is clear that the solubility of the protein is intimately related to its 3-D structure. Specifically, it is the distribution of hydrophobic (HΦO) and hydrophilic (HΦI) groups on the surface of the protein that determines the high solubility of the protein. We explore this aspect of protein structure in the next sections. The general relationship between solvation Gibbs energy and solubility of any solute is presented in section 2, along with a specific implementation of this relationship for proteins as solutes. The latter is essentially an expansion of the solvation Gibbs energy (see (2.16)) in terms of its various ingredients. Section 3 is devoted to details of the computational procedure and the various assumptions made in the calculation of the solvation Gibbs energies of proteins. Results and some conclusions are presented in section 4. The main conclusion of this paper can be summarized as follows: Suppose that we "turn off" all functional groups that are exposed to the solvent—we will be left with a large globular nonpolar molecule—the solvation Gibbs energy of which can be estimated to be very large and positive. Next, we "turn on" the sum of all hydrophilic groups and estimate their contribution to the total solvation Gibbs energy. As expected, this contribution is negative. However, a less expected finding is that, when adding the contributions from pair and triplet correlations between hydrophilic groups, the solvation Gibbs energy of the protein becomes significantly more negative. The precise amount of the decrease in the solvation Gibbs energy is dependent on the many approximations made in the course of these calculations. However, we believe that correlations between hydrophilic groups are significant in making the large globular proteins soluble in water.

2. Solvation Gibbs Energy and Solubility

In defining the solubility of a substance α, in a liquid phase l, one needs to specify, in addition to the temperature T, pressure P, and composition $x = (x_1...x_c)$, also the second phase with which α is in equilibrium. Normally, for proteins, the second phase is a solid phase s. The equilibrium condition would be

$$\mu_\alpha^s(P,T) = \mu_\alpha^l(T,P,x,\rho_\alpha^l) \tag{2.1}$$

where μ_α^s is the chemical potential (CP) of the protein α in the phase s (presuming that the solid is pure α), μ_α^l is the CP of α in l, x_i is the mole fraction of the ith component in the *solvent* l, and ρ_α^l is the number density of α in l (α is not counted as one of the components of l).

It is convenient to write the CP of α in l as[1,2]

$$\mu_\alpha^l = \mu_\alpha^{*l} + kT\ln(\rho_\alpha\Lambda^3{}_\alpha) \tag{2.2}$$

where $\Lambda^3{}_\alpha$ is the momentum partition function of α (or the de Broglie wavelength) and μ_α^{*l} is the so-called pseudo CP of α in l. In general, μ_α^{*l} can depend on ρ_α^l, but at sufficiently low density $\rho_\alpha^l \rightarrow 0$, μ_α^{*l} becomes independent of ρ_α^l and the *solubility* is the value of ρ_α^l that fulfills (2.1) which we denote by $(\rho_\alpha^l)_{eq}$ and is given by

$$(\rho_\alpha^l)_{eq} = \Lambda_\alpha^{-3}\exp[-\beta(\mu_\alpha^{*l} - \mu_\alpha^s)] \tag{2.3}$$

where $\beta = (kT)^{-1}$.

The solvation Gibbs energy of α is defined by[1,2]

$$\Delta G_\alpha^{*l} = \mu_\alpha^{*l} - \mu_\alpha^{*g} \tag{2.4}$$

where μ_α^{*g} is the pseudo CP of α in an ideal gas phase (i.e.,

* Author to whom correspondence should be addressed. E-mail: arieh@batata.fh.huji.ac.il.

[†] Current address: Research Institute, FF1, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195. E-mail: hwang@iris1.hh.ri.ccf.org.

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1997.

S1089-5647(96)01591-X CCC: $14.00 © 1997 American Chemical Society

where α does not interact with any other molecules which might be present in that phase).

For simple solutes α, such that have measurable vapor pressure, the solvation Gibbs energy can be easily determined from the equilibrium condition

$$\mu_\alpha^l = \mu_\alpha^{*l} + kT \ln \rho_\alpha^l \Lambda_\alpha^3 \tag{2.5}$$

or equivalently

$$(\rho_\alpha^l/\rho_\alpha^l)_{eq} = \exp[-\beta(\mu_\alpha^{*l} - \mu_\alpha^{*g})] \tag{2.6}$$

Hence from 2.3 and 2.5 we obtain

$$(\rho_\alpha^l)_{eq} = \Lambda_\alpha^{-3} \exp[-\beta(\mu_\alpha^{*g} - \mu_\alpha^s) - \beta(\mu_\alpha^{*l} - \mu_\alpha^{*g})] = (\rho_\alpha^g)_{eq} \exp[-\beta\Delta G_\alpha^*] \tag{2.7}$$

Here $(\rho_\alpha^g)_{eq}$ is seen to be the density of α in the ideal gas phase at equilibrium with the pure solid phase. Since proteins have no measurable vapor pressure, $(\rho_\alpha^g)_{eq}$ cannot be determined experimentally. The first factor in (2.7) determines the equilibrium density $(\rho_\alpha^g)_{eq}$ in the gaseous phase. The second factor, depending on the solvation Gibbs energy, transforms $(\rho_\alpha^g)_{eq}$ into the equilibrium density of α in the liquid phase. We shall focus on the latter factor only.

It is also convenient to assume that the internal degrees of freedom of the solute α are unaffected by the presence of the solvent, in which case the pseudo CP in the liquid phase may be written as

$$\mu_\alpha^{*l} = W(\alpha|l) - kT \ln q_\alpha \tag{2.8}$$

and

$$\Delta G_\alpha^* = W(\alpha|l) \tag{2.9}$$

where q_α is the internal partition function of α and W(α|l) is the coupling work, i.e., the work of inserting α at some fixed position in the liquid. (Note that whenever α can be in several conformations, ΔG_α^* is understood to be an average coupling work over all possible conformations Elaboration on the way this average is taken is discussed in details in refs 1 and 2).

With the above assumption we can rewrite (2.6) as

$$(\rho_\alpha^l/\rho_\alpha^g)_{eq} = \exp[\beta W(\alpha|l)] \tag{2.10}$$

Usually the last relation is used to determine the solvation Gibbs energy from the density ratio at equilibrium. In this article we shall try to estimate the various contributions to the solvation Gibbs energy and thereby estimate their relative contributions to the solubility of α in l.

The fundamental connection between W(α|l) and molecular interactions is given by

$$\Delta G_\alpha^* = W(\alpha|l) = -kT \ln\langle \exp(-\beta B_\alpha)\rangle \tag{2.11}$$

where B_α is the total interaction energy between the solute α and all the surrounding molecules at some fixed conformation $X^N = X_1 ... X_N$, X_i being the configuration (location and orientation) of the ith molecules.

$$B_\alpha = \sum_{i=1}^{N} U(X_\alpha, X_i) \tag{2.12}$$

The sum in (2.12) runs over all the molecules in l except the one, α-molecule, under consideration. The average in (2.11) is over all the configurations of the solvent molecules in the appropriate ensemble.[1,2]

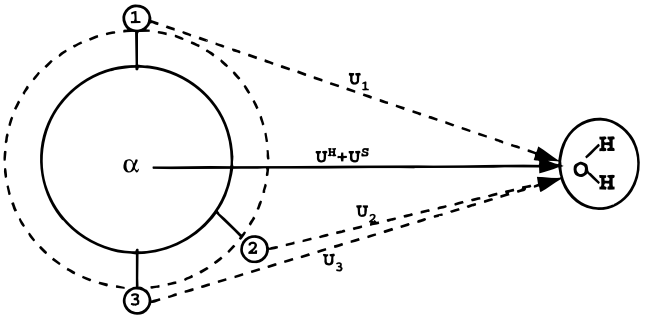

Figure 1. Schematic illustration of the split of the total solute (α)-solvent (water) interaction. $U^H + U^S$ is the hard and soft interaction of the globular protein. U_k is the interaction between the kth functional group and a water molecule.

For simple solute α, such as argon, one can separate each of the solute−solvent pair potentials in (2.12) into two contributions.[1,2]

$$U(X_\alpha, X_i) = U^H + U^S \tag{2.13}$$

where U^H is the "hard", or the repulsive part of the interaction potential function, and U^S is the "soft", or the van der Waals part of the interaction. For such a solute, the solvation Gibbs energy may be written as a sum of two terms

$$\Delta G_\alpha^* = \Delta G_\alpha^{*H} + \Delta G_\alpha^{*S/H} \tag{2.14}$$

The first term corresponds to the solvation of the hard core of the solute. This is equivalent to the work required to create a cavity of a suitable size and shape to accommodate the solute. The second term is the *conditional* solvation Gibbs energy of the soft part, given that the hard part has already been turned on.

For proteins, the simple split of the solute−solvent pair potential as in (2.13) is not appropriate; one needs a more elaborate description of the ingredients of this pair potential. There are several ways of performing such a split into a sum of contributions. One way is to recognize that, in addition to the hard and the soft parts of the potential, there are also specific functional groups such as charged or polar groups that interact with water in a way different from a simple nonpolar group.[1−4]

To account for these specific interactions, we write the solute−solvent pair interaction for a protein α, in generalization of (2.13), as

$$U(X_\alpha, X_i) = U^H + U^S + \sum_k U_k \tag{2.15}$$

where U_k is the contribution of the kth functional group to the solute−solvent interaction. Note that U_k includes both hard and soft, as well as any specific, interaction with the solvent. Figure 1 shows a schematic example of such a split of the total solute−solvent interaction. When the solvation Gibbs energy is computed through (2.11), using (2.15), we obtain[2]

$$\Delta G_\alpha^* = \Delta G_\alpha^{*H} + \Delta G_\alpha^{*S/H} + \sum_k \Delta G_\alpha^{*k/H,S} + \sum_{k,j} \Delta G_\alpha^{*k,j/H,S} + \sum_{i,j,k} \Delta G_\alpha^{*i,j,k/H,S} + ... \tag{2.16}$$

Thus, even when the additivity assumption in (2.15) is a good approximation, or even an exact, the solvation Gibbs energy of α is, in general, not additive with respect to the contributions of all the functional groups. The reason is that the solvation of

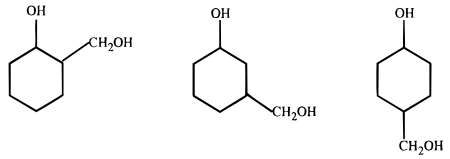

Figure 2. Three isomers of hydroxybenzyl alcohol. These three isomers would have the same solubility if additivity of the form (2.17) is assumed.

a group of two or more functional groups (FG) might be correlated and, therefore, depending on the extent of correlation, one must account for independent FGs, pair correlated FGs, triplet correlated FGs etc. Thus, the first sum on the rhs of (2.16) is over all the independently solvated FGs, the second term is over all pair correlated FGs, and so on.

The expansion of the solvation Gibbs energy as done in (2.16) is useful in the study of the various contributions to the solubility of a globular protein. Of course the specific implementation of such an expansion depends upon many, some arbitrary, decisions on how to split the potential function into its ingredients in (2.15). For instance, one must decide where to make the cutoff between the hard and soft part of the interaction (a problem which exists even for the simplest real solute such as argon), how to select the criterion that distinguishes between FGs that are exposed or unexposed to the solvent, etc. These details will be described in the next section.

We end this section by a simple demonstration of the effect of pair correlation on the solubility of a small model compound.

Consider a small molecule having two FGs that can form hydrogen bonds to a water molecule. Figure 2 shows the three isomers of hydroxylbenzyl alcohol. Had we assumed that the FGs were independently solvated, we would have written the solvation Gibbs energy as[5]

$$\Delta G^* = \Delta G^{*\mathrm{H,S}} + \Delta G^{*\mathrm{OH/H,S}} + \Delta G^{*\mathrm{CH_2OH/H,S}} \tag{2.17}$$

where in the first term we have combined the hard and soft parts of the potential function for the benzyl−water interaction. The last two terms are the conditional solvation Gibbs energies of the FGs OH and CH_2OH. Had we assumed total independence of the FGs, which is equivalent to group additivity of the solvation Gibbs energy, we would have predicted that the three isomers in Figure 2 would have the same solvation Gibbs energy and hence the same solubility, in contrast to experiment. The reason for this discrepancy is that we have not taken into account possible correlations, i.e., dependence between the solvation of the FGs. We know that the solubilities of the three isomers (in the sense of (2.6) widely differ and this variation is due to the effect of pair correlation. To account for this variation, we consider the difference in the solvation Gibbs energy between the two isomers, say 1,4 and 1,2. We assume that in the 1,4 isomer the two FGs are far apart and therefore they are independently solvated. Also, we assume that the quantity $\Delta G^{*\mathrm{H,S}}$ includes only factors that depend on the benzyl ring and not on the FGs. We take the difference in the solvation Gibbs energies of the two isomers

$$\delta G(1,4\rightarrow 1,2) = \Delta G^*_{1,2} - \Delta G^*_{1,4} = \Delta G^{*\mathrm{OH,CH_2OH/H,S}} - \Delta G^{*\mathrm{OH/H,S}} - \Delta G^{*\mathrm{CH_2OH/H,S}} \tag{2.18}$$

we see that the quantity δG is exactly the indirect work of transferring one functional group from position 4 to position 2. Hence, this is a measure of the solvent-induced part of the correlation between the two FGs. (There is also a direct correlation due to the direct interaction between the two FGs. Note also that this correlation is *conditional*, in the sense that the hard and soft parts (H and S) of the benzyl ring are present.)[1,5]

Had we known $\Delta G^*_{1,2} - \Delta G^*_{1,4}$, we could have used (2.18) to estimate the (indirect) pair correlation between the two FGs. Unfortunately this information is not available. Instead we have only the *difference* in the solvation Gibbs energies between water and hexane (or cyclohexane). Hence, we can only estimate the difference in correlation between the two FGs in the two liquids. The experimental results are (in kcal/mol)[5]

	water−hexane	water−cyclohexane
$\delta G(1,4\rightarrow 1,2)$	0.998	0.918
$\delta G(1,4\rightarrow 1,3)$	−1.278	−1.613

We have interpreted these results as follows. Assuming that the major part of the correlations is due to the water and not to the organic solvent, the first transfer 1,4 → 1,2 brings the two FGs too close to each other in such a way that solvation of one FG interferes with the solvation of the second FG; hence $\delta G > 0$. On the other hand, transferring from position 4 to position 3 makes δG negative, indicating an enhancement of the solvation of the pair of FGs relative to the independently solvated FGs. Translating these findings into solubilities, or more precisely into solubility ratios, we obtain

$$\left[\frac{\rho_{1,2}(w)}{\rho_{1,4}(w)}\frac{\rho_{1,4}(H)}{\rho_{1,2}(H)}\right]_{\mathrm{eq}} = \exp(-\beta\delta G) \tag{2.19}$$

Again, assuming that the density ratio in hexane is of the order of unity, we conclude that the density ratio in water can be of the order of 10^{-1} in the first case (1,4 → 1,2) and 10 in the second case (1,4 → 1,3). We have recently estimated δG between two FGs in the most favorable configuration to be of the order of −2.5 to −3.0 kcal/mol,[2,6] which translates into a ratio of densities of about 100.

3. Computational Procedure in the Estimation of the Solvation Gibbs Energy

In this section we present the details of our assumptions and approximations in the estimation of the solvation Gibbs energy of proteins. The total solvation Gibbs energy of a protein is calculated by estimating each of the terms in (2.16).

A. Hard Part. The first term in (2.16) $\Delta G^{*\mathrm{H}}$ corresponds to turning on the hard part of the protein−solvent interaction potential. This is the same as the free energy of creating a cavity of suitable size at some fixed position in the solvent.[2,7,8] We assume that the globular protein is spherical with an effective diameter d_p, so we can calculate $\Delta G^{*\mathrm{H}}$ using scaled-particle theory (SPT).[7,8] For a specific protein, its effective diameter d_p is estimated from its solvent-accessible surface area (SASA). The SASA of a protein is calculated by the program ACCESS.[9,10] Then

$$d_p = \sqrt{\mathrm{SASA}/\pi} - d_w \tag{3.1}$$

where SASA is the solvent-accessible surface area of the protein and $d_w = 2.8$ Å is the diameter of the water molecule rolling on the surface of the protein. Thus, the cavity suitable to accommodate the protein has a radius of

$$R_{\mathrm{cav}} = (d_p + d_w)/2 \tag{3.2}$$

According to SPT, the Gibbs energy of cavity formation is calculated by[2,8]

$$\Delta G^{*\mathrm{H}} = K_0 + K_1 R_{\mathrm{cav}} + K_2 R_{\mathrm{cav}}^2 + K_3 R_{\mathrm{cav}}^3 \tag{3.3}$$

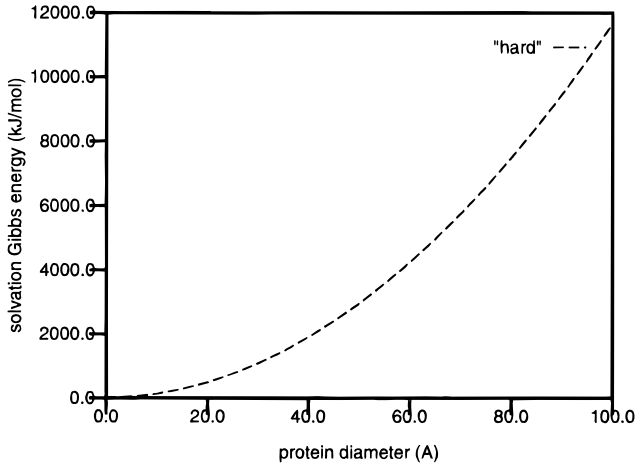

Figure 3. Solvation Gibbs energy of the hard part (cavity formation) of the interaction between protein and water as a function of protein diameter at $T = 298.15$ K.

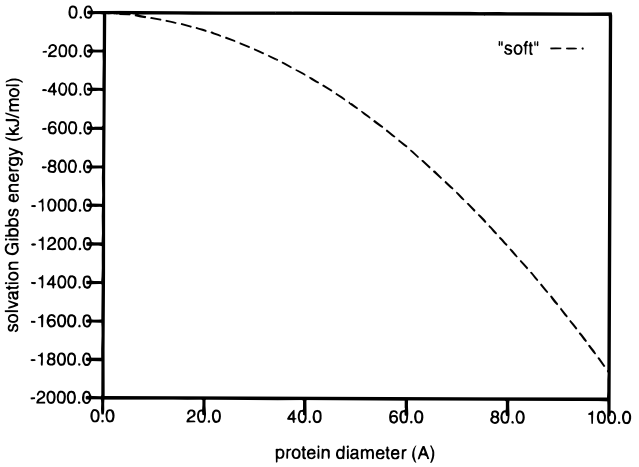

Figure 4. Solvation Gibbs energy of the soft part of the interaction between protein and water as a function of protein diameter.

where the coefficients K_i are

$$K_0 = kT(-\ln(1-y) + 4.5z^2) - \pi p d_w^3/6$$

$$K_1 = -kT(6z + 18z^2)/d_w + \pi p d_w^2$$

$$K_2 = kT(12z + 18z^2)/d_w^2 - 2\pi p d_w \qquad (3.4)$$

$$K_3 = 4\pi p/3$$

$$y = \pi\rho w d_w^3/6, \quad z = y/(1-y) \qquad (3.5)$$

where ρ_w is the solvent density and p is the pressure. Taking the density of water at 298.15 K, $\rho_w = 3.344 \times 10^{22}$ molecules/cm^3 and p as 1 atmosphere, we calculated the solvation Gibbs energy of the hard part as a function of protein diameter d_p. The result is given in Figure 3. Note that the steep increase in the solvent Gibbs energy corresponds to a very steep decrease of the solubility, caused only by the creation of a cavity in the solvent.

B. Soft Part. The second term in (2.16) corresponds to the soft part of the protein−water interaction potential. This is given by

$$\Delta G^{*S/H} = -kT \ln\langle \exp[-\beta B_p^S] \rangle_h \qquad (3.6)$$

where B^S is the total van der Waals interaction of the protein with the surrounding water molecules. We have shown that $\Delta G^{*S/H}$ can be estimated approximately by[3,4]

$$\Delta G^{*S/H} = -\epsilon(\mathrm{Ne,CH_4})\rho_w \frac{4\pi}{3}[(d_p/2 + d_w)^3 - (d_p/2)^3] \qquad (3.7)$$

where

$$\epsilon(\mathrm{Ne,CH_4}) = \sqrt{\epsilon(\mathrm{Ne,Ne})\epsilon(\mathrm{CH_4,CH_4})} \qquad (3.8)$$

in which ϵ(Ne,Ne), ϵ(CH$_4$,CH$_4$) are the Lennard-Jones parameters of neon and methane respectively, ρ_w is the water density at 298.15 K, d_p is the effective diameter of protein, and d_w is the diameter of water molecule. Figure 4 shows $\Delta G^{*S/H}$ as a function of protein diameter. As expected, the soft interaction adds a negative contribution to ΔG^*, but this is relatively small compared with the corresponding values of ΔG^{*H}.

C. Contribution of Hydrogen Bond Interaction between the FGs and the Solvent Molecules to the Solvation Gibbs Energy of the Protein: the FGs Are Treated as Independent of Each Other. The third term in (2.16) consists of the sum of all the independently solvated FGs whose interactions with the solvent were not included in the previous two terms. In principle, we should have taken the sum over all possible FGs, both HΦO and HΦI groups. However, in the procedure of calculating the diameter of the protein molecule, we have actually taken an upper limit for the diameter. This is equivalent to including all van der Waals interactions between the protein and all the water molecules in the system. What remains is only that part of the protein−water interaction that has not been accounted for in the first two terms in (2.16). In the following, we focus only on the contributions of those FGs that are capable of forming hydrogen bonds with water molecules. In a more complete treatment, one must include also ionized groups on the surface of the protein. We shall not include these effects, since we do not have the relevant data. We first assume that all the exposed HΦI FGs of the protein are independently solvated. We will consider the correlations (both negative and positive) among these groups later. Thus the sum over k in (2.16) runs over all HΦI FGs of the protein that are exposed to the solvent.

We have previously estimated that a FG capable of forming one hydrogen bond with the solvent, (i.e., one hydrogen bond "arm" is solvated), can contribute approximately −9.5 kJ/mol to the solvation Gibbs energy.[2,11] Thus, to estimate the hydrogen bond contribution to the solvation Gibbs energy of a protein, we need only to count the total number of solvated arms of all FGs on the surface of protein.

Table 1 listed the maximum number of solvated arms of each kind of HΦI atom in the protein. Note that the values given in Table 1 are the upper limits of the number of possible solvated arms of each kind of a HΦI atom.[12] The actual number of solvated arms also depends on the extent the HΦI atoms are exposed to the solvent. Finney et al. pointed out that a molecular area greater than 5 Å^2 is large enough to accommodate a single protein−solvent hydrogen bond interaction,[13,14] which is roughly 5% of the total SASA of an oxygen atom (the SASA of a separate oxygen atom). Correspondingly, the areas for two and three such interaction are 15−25 Å^2 and >25 Å^2, respectively, in their estimation, which are roughly 15%−25% and >25% the total SASA of an oxygen atom. Other authors use smaller SASA criteria in the assignment of the protein-solvent hydrogen bond interaction.[3,4,15] We use both of these criteria in our estimation of the hydrogen bond contribution to solvation Gibbs energy of proteins. These are (i) If the SASA of a HΦI atom is less than 5% of the atom's total SASA, there is no hydrogen bond interaction between this HΦI atom and the water molecules. Note that there are still hard and soft interactions

TABLE 1: Maximum Solvated Arms of Each Kind of Hydrophilic Atoms in Protein[a]

residue	group	no. of lone pairs	no. of hydrogens	max no. of solvated arms
Asn, Gln	NH_2	0	2	2
	C═O	2	0	2
Asp, Glu	CO_2-	4	0	4
Ser, Thr	OH	2	1	3
Tyr	OH	1	1	2
Arg	NH	0	1	1
	$2NH_2$	0	4	4
Lys	NH_3	0	3	3
Trp	NH	0	1	1
His	NH	0	1	1
	NH^+	0	1	1
main chain	NH	0	1	1
	C═O	2	0	2

[a] The listed values are from Thanki et al.[12] The number of lone pairs and hydrogen atoms are those available in each group that are able to participate in hydrogen bonding with solvent molecules.

between this HΦI atom and the solvent, which are included in the first two term of (2.16). If the SASA of a HΦI atom is in the range of 5%−15% of its total SASA, one hydrogen bond arm is solvated. If the SASA of a HΦI atom is in the range of 15%−25% of its total SASA, then at most two hydrogen bond arms are solvated. (By "at most", we mean that, for example, for a main chain nitrogen even when its SASA is larger than 15% of its total SASA, it still has only one arm solvated). If the SASA of a HΦI atom is greater than 25% of its total SASA, then at most three hydrogen bond arms are solvated. (ii) If the SASA of a HΦI atom is less than 1% of the atom's total SASA, there is no hydrogen bond interaction between this HΦI atom and the water molecules. If the SASA of a HΦI atom is in the range of 1%−10% of its total SASA, one hydrogen bond arm is solvated. If the SASA of a HΦI atom is in the range of 10%−20% of its total SASA, then at most two hydrogen bond arms are solvated. If the SASA of a HΦI atom is greater than 20% of its total SASA, then at most three hydrogen bond arms are solvated.

By breaking the FGs, such as COOH in Asp, into HΦI atoms O and OH, in counting the solvated arms we may overestimate the contribution of hydrogen bond interaction to the solvation Gibbs energy of protein from side chains such as Asp, Glu, Asn, Gln, and Arg. This could be exemplified by the experimental solvation Gibbs energy of simple compounds listed in Table 2.[16,17]

We use carboxylic acids to represent the carboxyl group in the side chain of Arg and Glu. By subtracting the solvation Gibbs energy of the corresponding alkene, which represent the hard and soft parts of solvation Gibbs energy, we get the hydrogen bond interaction contribution to the solvation Gibbs energy of carboxyl group. The average value of this contribution from the available experimental data is −32.74 kJ/mol; see Table 2.

In our counting of the solvated arms of carboxylic group, we count the solvated arms of O and OH separately, that is, we treat the carboxyl O as if it were a carbonyl oxygen and the carboxyl OH as an independent hydroxyl group. From Table 2, we see that the average hydrogen bond contribution to solvation Gibbs energy of a carbonyl oxygen is −20.87 kJ/mol and the average hydrogen bond contribution to solvation Gibbs energy of a hydroxyl is −29.22 kJ/mol. The summation of these hydrogen bond contributions to solvation Gibbs energy is −50.09 kJ/mol, which is 17.35 kJ/mol more negative than that of a carboxylic group. Thus, due to the negative interference (see below) between O and OH in the carboxylic group, they lose approximately 1.8 solvated arms compared with independent O (2 arms in carbonyl) and OH (3 arms in hydroxyl). We assign this negative interference evenly among all five possible solvated arms, that is, each solvated arm of a carboxylic group will lose 0.37 arms, or the equivalent of 0.37 × 9.5 kJ/mol in solvation Gibbs energy.

In analogy with this, we assign negative interference for amide group in Asn and Gln as 0.06 arm per solvated arm and for guanidinyl group in Asp as 0.21 arm per solvated arm.

D. Correlation among HΦI Atoms. Both theoretical and simulation studies[18−20] point out that if two HΦI atoms are too close to each other, e.g., at a distance less than 5 Å, there will be a strong correlation due to solvation effects. If the distance is larger than 5 Å, the correlation is much smaller and hence the two HΦI atoms may be treated as if they were independent.

(1) Negative Correlation. When the distance between two HΦI atoms is less than 4 Å, the correlation between them will

TABLE 2: Experimental Solvation Gibbs Energy (kJ/mol) of Some Model Compounds[a]

compd	Δ*G**	compd	Δ*G**	ΔΔ*G**
acetic acid	−28.05	2-methyl-1-propene	4.87	−32.92
propanoic acid	−27.09	2-methyl-1-butene	5.48	−32.57
butanoic acid	−26.59	2-methyl-1-pentene	6.15	−32.74
functional group −COOH		average HB[c] contribution to Δ*G**		−32.74
2-propanone	−16.12	2-methyl-1-propene	4.87	−20.99
2-butanone	−15.22	2-methyl-1-butene	5.48	−20.70
2-pentanone	−14.76	2-methyl-1-pentene	6.15	−20.91
functional group −C═O		average HB contribution to Δ*G**		−20.87
2-propanol	−19.90	2-methylpropane	9.70	−29.60
2-butanol	−19.15	2-methylbutane	9.97	−29.12
2-pentanol	−18.38	2-methylpentane	10.56	−28.94
functional group −OH		average HB contribution to Δ*G**		−29.22
acetamide	−40.63	2-methyl-1-propene	4.87	−45.50
propionamide	−39.39[b]	2-methyl-1-butane	5.48	−44.87
functional group $-C(O)NH_2$		average HB contribution to Δ*G**		−45.19
ethanamide	−18.48	propane	8.18	−27.02
1-propanamide	−18.37	butane	8.70	−27.07
1-butanamide	−17.97	pentane	9.76	−27.73
1-pentanamide	−17.14	hexane	10.40	−27.54
functional group $-NH_2$		average HB contribution to Δ*G**		−27.34
methylguanidine	−46.81[b]	2-methyl-2-butene	5.48	−52.29
functional group $-NC(NH_2)_2$		average HB contribution to Δ*G**		−52.29

[a] All the experimental data from Cabani et al.[16] except those specifically stated. [b] Data from Wolfenden et al.[17] [c] HB = hydrogen bond.

TABLE 3: Proteins Used in Our Calculation of Solvation Gibbs Energy

PDB file	description	SASA (Å^2)	diameter (Å)	no. of exposed hydrophilic atoms 5%[a]	no. of exposed hydrophilic atoms 1%[b]
1crn	crambin	3013.41	28.17	42	60
4pti	trypsin inhibitor	4028.36	33.01	64	90
1ubi	ubiquitin	4821.25	36.38	91	119
2pcy	plastocyanin	4999.18	37.09	96	139
4cpv	calcium-binding parvalbumin b	5644.19	39.59	86	122
2lyz	lysozyme	6660.54	43.25	130	178
2myc	myoglobin	8212.87	48.33	131	184
1ast	astacin	9051.21	50.88	181	257
3sgb	ovomucoid third domain	9641.85	52.60	198	282
1cim	carbonic anhydrase II	11671.41	58.15	206	310
1tec	eglin c	12083.61	59.22	240	338
4pep	pepsin	13201.34	62.02	276	392
3enl	enolase	16082.13	68.75	319	455
1coy	cholesterol oxidase	17584.45	72.02	354	507
1gal	glucose oxidase	19500.63	75.99	399	583
1afg	fibroblast growth factor (acidic)	23769.14	84.18	478	627
1nis	aconitase	24732.55	85.93	488	779
2sod	superoxide dismutase	26739.49	89.46	489	706
1gpb	glycogen phosphorylase b	31911.62	97.99	627	913
1hpl	lipase	34000.11	101.23	712	1035

[a] An atom is regarded as exposed to the solvent when its SASA is greater than 5% of its total SASA. [b] An atom is regarded as exposed to the solvent when its SASA is greater than 1% of its total SASA.

be negative. We call this *negative correlation* because at these distances the solvation of one HΦI group interferes with the solvation of the second HΦI group. The reason for such an interference is purely geometrical, i.e., at very short distance there is not enough space for water molecules to solvate the two HΦI groups as efficiently as when they are widely separated. This makes the solvation Gibbs energy of the pair more positive than the sum of the solvation Gibbs energies when they are separated. In our calculation, we found that most of the HΦI atoms on the surface of the protein have one or two neighbor HΦI atoms at the distance less than 4 Å; some of them have more than four such neighbors.

To the best of our knowledge, there are neither theoretical nor simulated results providing quantitative estimates for this negative correlation. We therefore use a rule-of-thumb method in estimating its contribution to the solvation Gibbs energy of a protein. First, we refer to the experimental results for carboxylic group shown in Table 2. Assuming that the loss of solvation Gibbs energy compared with independently solvated O and OH results from the negative correlation between them, and that the negative correlation is evenly distributed among the five possible solvated arms, we came to the conclusion that each solvated arm of carboxylic group lost 3.5 kJ/mol of solvation Gibbs energy due to the short distance between the two oxygen atoms. The distance between the two oxygen atoms is 2.25 Å. On the other hand, simulation results indicated that the correlation between two HΦI atoms turns to positive around 4 Å.[19,20] Furthermore, assuming that the negative correlation is linear in the distance between the two HΦI atoms, we take

$$y = 8.0 - 2.0d \qquad (3.9)$$

where d is the distance between two HΦI atoms (in Å) and y is the contribution of the negative correlation to the solvation Gibbs energy (in kJ/mol).

Since those regions from which water molecules can form hydrogen bond with a HΦI atom are very small, a HΦI atom can still be solvated by hydrogen bonds even if there are other HΦI atoms at very short distances. In order to deal with this situation, we adopted the following rules. (i) The HΦI atom has only one solvated arm. If the number of neighbors is less than 2 (by "neighbor" we mean that the distance between them is less than 4 Å), then the contribution of negative correlation to solvation Gibbs energy is 0, i.e., $y = 0$; If the number of neighbors is 2 or 3, then

$$y = 8.0 - 2.0d$$

where d is the average distance between the HΦI atom and all its neighbors. If the number of neighbors is equal to or larger than 4, then this HΦI atom loses its hydrogen bond part of the solvation completely, i.e., $y = 9.5$ kJ/mol.

(ii) The HΦI atom has two solvated arms. The contribution due to negative correlation can be calculated by

number of neighbor atoms	contribution to solvation Gibbs energy
≈1 to 2	$y = 8.0 - 2.0d$
3	$y = 9.5$
4	$y = 9.5 + (8.0 - 2.0d)$
≥5	$y = 9.5 \times 2$

(iii) The HΦI atom has three solvated arms. The contribution due to negative correlation can be calculated by

number of neighbor atoms	contribution to solvation Gibbs energy
1	$y = 8.0-2.0d$
2	$y = 9.5$
3	$y = 9.5 + (8.0 - 2.0d)$
4	$y = 9.5 \times 2$
≥5	$y = 9.5 \times 3$

Using these rules, we studied 20 proteins (Tables 4 and 5). We found that the negative correlation contribution to solvation Gibbs energy is relatively small, less than 10% of the value of $\Delta G^{*\text{indep/S,H}}$. This contribution is also much smaller compared with the corresponding contribution due to the positive correlations, discussed below. The relative insignificance of the negative correlation assures us that the precise choice of the rules of estimating of this correlation will not greatly affect the final conclusion.

(2) Positive Correlation. In a recent examination of the solvent induced effects on protein folding and association,[6] we found that, at a distance of about 4.5 Å, there should be strong positive correlation between the solvation of the two HΦI groups. This is a result of an enhancement of the solvation of the pair of HΦI groups due to the possibility that a water

TABLE 4: Contribution of Each Kind of Interaction to Solvation Gibbs Energy (kJ/mol)

PDB file	ΔG^{*H}	$\Delta G^{*S/H}$	$\Delta G^{*H} + \Delta G^{*S/H}$	$\Delta G^{*k/S,H}$	$\Delta G^{*NC/S,H}$	no. of triplets	no. of pairs	$\Delta G^{*PC/S,H}$	ΔG^{*}
1crn	958.69	−169.18	789.51	−586.72	3.03	0	13	−162.5	43.3
4pti	1306.73	−226.03	1080.70	−945.47	22.06	0	13	−162.5	−5.2
1ubi	1580.89	−270.49	1310.40	−1302.74	32.15	1	31	−415.4	−375.6
2pcy	1641.98	−280.36	1361.62	−1399.73	37.82	2	35	−493.3	−493.6
4pcv	1866.32	−316.54	1549.78	−1315.28	11.42	1	31	−415.4	−169.5
2lyz	2220.70	−373.48	1847.22	−1840.06	33.83	3	38	−558.7	−517.7
2myc	2763.68	−460.32	2303.36	−1842.53	40.64	2	41	−568.3	−66.8
1ast	3058.64	−507.33	2551.31	−2475.79	84.54	4	59	−849.5	−689.4
3sgb	3266.05	−540.34	2725.71	−2785.59	70.76	3	67	−921.2	−910.3
1cim	3981.80	−653.93	3327.87	−2906.05	66.82	3	65	−896.2	−407.6
1tec	4127.95	−677.07	3450.88	−3513.39	86.07	3	80	−1083.7	−1060.1
4pep	4522.89	−739.54	3783.35	−3805.89	80.91	5	89	−1252.0	−1193.6
3enl	5546.11	−900.98	4645.13	−4304.45	64.99	3	105	−1396.2	−990.5
1coy	6080.98	−985.18	5095.81	−4798.92	114.74	7	105	−1507.8	−1096.2
1gal	6763.52	−1092.45	5671.08	−5288.46	46.76	6	127	−1754.9	−1325.5
1afg	8286.53	−1331.26	6955.27	−6598.23	121.26	5	155	−2077.0	−1598.7
1nis	8632.05	−1385.35	7246.70	−5890.38	108.99	7	137	−1907.8	−442.5
2sod	9350.53	−1497.74	7852.80	−6790.32	147.78	3	157	−2046.2	−831.9
1gpb	11205.57	−1787.40	9418.17	−7917.21	95.21	5	175	−2327.0	−730.8
1hpl	11954.27	−1904.13	10050.13	−9191.44	110.70	13	217	−3075.2	−2105.8

TABLE 5: Contribution of Hydrophilic Groups to the Solvation Gibbs Energy (kJ/mol) when 1% Is Used as the Criteria To Pick the Exposed Atoms

PDB file	$\Delta G^{*k/S,H}$	$\Delta G^{*NC/S,H}$	no. of triplets	no. of pairs	$\Delta G^{*PC/S,H}$	ΔG^{*}
1crn	−817.66	67.53	0	26	−325.0	−312.6
4pti	−1271.95	132.63	2	27	−393.3	−451.9
1ubi	−1668.20	92.16	1	52	−677.9	−943.5
2pcy	−2000.99	236.71	5	55	−827.0	−1229.7
4cpv	−1747.15	58.77	2	52	−705.8	−844.4
2lyz	−2455.65	180.28	5	60	−889.5	−1317.7
2myc	−2496.60	158.14	4	73	−1024.1	−1059.2
1ast	−3464.08	349.01	13	94	−1537.7	−2101.5
3sgb	−3990.85	388.78	7	115	−1632.8	−2509.2
1cim	−4150.17	376.14	7	116	−1645.3	−2091.5
1tec	−4719.98	363.74	7	130	−1820.3	−2725.7
4pep	−5297.11	550.82	16	150	−2321.4	−3284.3
3enl	−6115.72	408.31	10	166	−2354.0	−3416.3
1coy	−6723.72	490.65	15	174	−2593.5	−3730.8
1gal	−7541.96	401.22	14	217	−3103.1	−4572.8
1afg	−8668.18	575.45	12	243	−3372.3	−4509.8
1nis	−9516.91	873.51	29	265	−4121.6	−5518.3
2sod	−9554.72	722.96	10	290	−3904.0	−4883.0
1gpb	−11376.44	665.07	23	317	−4604.2	−5897.4
1hpl	−13516.13	1035.21	30	388	−5687.0	−8117.8

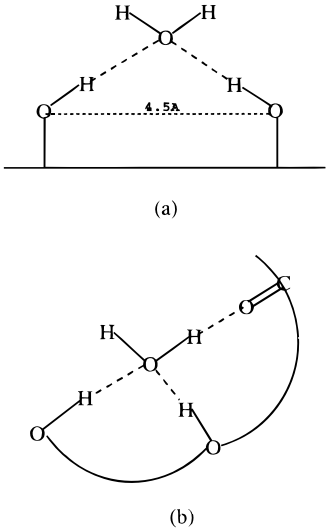

Figure 5. Schematical representation of pair and triplet correlations among functional groups on the surface of protein. (a) Two functional groups bridged by a single water molecule. (b) Three functional groups bridged by a single water molecule.

molecule will simultaneously form hydrogen bonds with these two groups. We have estimated that the correlation between two such HΦI groups at a configuration similar to the second-nearest neighbors in ordinary ice (as in Figure 5a, the distance between the two oxygen is 4.5 Å), contributes about 12.5 kJ/mol to the solvation Gibbs energy. Some simulation results further support this conclusion.[20] We have also estimated the solvent-induced interaction among three HΦI groups, when they form an equilateral triangle with an edge of about 4.5 Å and oriented so that they can simultaneously form three hydrogen bonds with one water molecule (as in Figure 5b, triplet correlation), to be around 27.9 kJ/mol [2].

To estimate the contribution of these positive correlation to the solvation Gibbs energy of protein, we first calculated the distance distribution of all HΦI atoms on the surface of protein. Then, we identified all the possible triplet, i.e., three HΦI atoms on the surface of protein forming a triangle with an edge of 4.0−5.0 Å. To form three simultaneously hydrogen bonds with one water molecule, there should be at least one hydrogen bond donor and one hydrogen bond acceptor among the three HΦI atoms. Thus, neither three main chain N atoms nor three main chain O atoms can form this kind of triplet correlation even though they possess the suitable geometry. There are also other cases that one atom participating in two geometrically possible triplets, but the atom has only one solvated arm. In this case, we assume that only one triplet correlation can be formed.

Finally, we assign and count the number of pair correlations. To be correlated with each other through a water bridge, two HΦI atoms should be at a distance ranging between 4.0 and 5.0 Å, and both of them must have the sufficient number of solvated arms. For example, suppose atom a_1 has one arm, and the distances between a_1 and a_2 and between a_1 and a_3 are both in the range of 4.0−5.0 Å, then a_1 has two choices of being positively pair-correlated according to this geometry. If we assign a positive pair correlation to a_1-a_2, then a_1-a_3 cannot form such a positive correlation due to the lack of solvated arm of atom a_1. Accordingly, we adopted the following set of rules to assign the pair correlation. (i) If a three-arm atom has only one choice of forming positive pair correlation, then we assign a pair between this and the second atom. (We do this because a three-arm atom has a higher possibility of having the right orientation). If the second atom has only one arm, then it can no longer form positive pair correlation with other atoms even when they also possess the suitable geometry. (ii) If a two-arm atom has only one choice of forming positive pair correlation, then we assign a pair between this and the second

atom. (iii) If an atom has n arms, it can participate at most in n such pairs. (iv) Considering the restriction of orientation, we always avoid assigning a three-arm atom to participate in 3 pairs and also avoid assigning a two-arm atom to participate in two pairs if possible. (v) Whenever there are two more choices still exist for an atom after applying the previous procedure, we assign the pair to the HΦI atom with a distance closer to 4.5 Å between them.

The above estimates of the contribution due to positive correlation probably provides a lower limit for the following reason. When we considered a linear sequence of HΦI group at the appropriate distance of about 4.5 Å, with a single-arm group at the center, we have assigned the full positive correlation only to one pair. The reason was that a single-arm HΦI group forming a hydrogen-bonded bridge with one partner cannot at the same time participate in a second bridge to another partner. This is true when the bridge consists of permanent hydrogen bonds. Because the positive correlation is a statistical phenomenon, therefore even a single-arm HΦI group can participate simultaneously in two hydrogen-bonded bridges with two different partners. Although each bridge will probably contribute less than −12.5 kJ/mol to the Gibbs energy, adding these possibilities will make the total solvation Gibbs energy more negative.

4. Results and Discussion

To get a general feeling for the relative contribution of each kind of interactions to the solvation Gibbs energy of the proteins and hence their solubilities, we studied 20 globular proteins with different sizes. These proteins are listed in Table 3.

The data for the coordinates of these proteins were taken from PDB.[21] The SASAs of these protein were calculated by ACCESS and their diameters estimated by (3.1). The number of exposed HΦI atoms on the surface of the proteins was calculated using two criteria. In the first, a HΦI atom is regarded as exposed when its SASA is larger than 1% of its total SASA. In the second, a HΦI atom is regarded as exposed when its SASA is larger than 5% of its total SASA.

Table 4 gives the contribution of each kind of interaction to the solvation Gibbs energy with each term calculated by the method described in the previous section. The contribution of the hydrogen bond part and the correlation of HΦI atoms are calculated using 5% as the criteria in determining whether the atoms are exposed to the solvent.

The contribution of hydrogen bonding to the solvation Gibbs energy of proteins was calculated in three stages. First, we treated all the HΦI atoms as if they were independent. The corresponding values of $\Delta G^{*\mathrm{indep/S,H}}$ for all the proteins were fitted to a square polynomial of the form

$$\Delta G^{*\mathrm{indep/S,H}} = 743.87 - 31.72d_\mathrm{p} - 0.61d_\mathrm{p}^2 \qquad (4.1)$$

where $\Delta G^{*\mathrm{indep/S,H}}$ is the solvation Gibbs energy in kJ/mol and d_p is the diameter of protein. Figure 6(a) shows the resulting fitted curve. Second, the contribution of the negative correlation (NC) is rather small in magnitude, we therefore fitted a linear relation of the form

$$\Delta G^{*\mathrm{NC/S,H}} = -23.60 + 1.51d_\mathrm{p} \qquad (4.2)$$

shown in Figure 6(b). Finally, the contribution of the positive correlation (PC) is shown in Figure 6(c); the fitted curve is

$$\Delta G^{*\mathrm{PC/S,H}} = 378.41 - 14.72d_\mathrm{p} - 0.16d_\mathrm{p}^2 \qquad (4.3)$$

On the basis of these results, we calculated the total solvation Gibbs energy of an "average" protein and the relative contribution of each kind of interactions to the solvation Gibbs energy as a function of the protein diameter. The result is shown in Figure 7. Curve 1 in Figure 7 is the solvation Gibbs energy due to the repulsive interaction between protein and the solvent molecules, $\Delta G^{*\mathrm{H}}$, calculated by the scaled-particle theory. In curve 2, we turned on the van der Waals interaction between protein and the solvent molecules, $\Delta G^{*\mathrm{S/H}}$. We see a drop in solvation Gibbs energy, which is not sufficient to overcome the

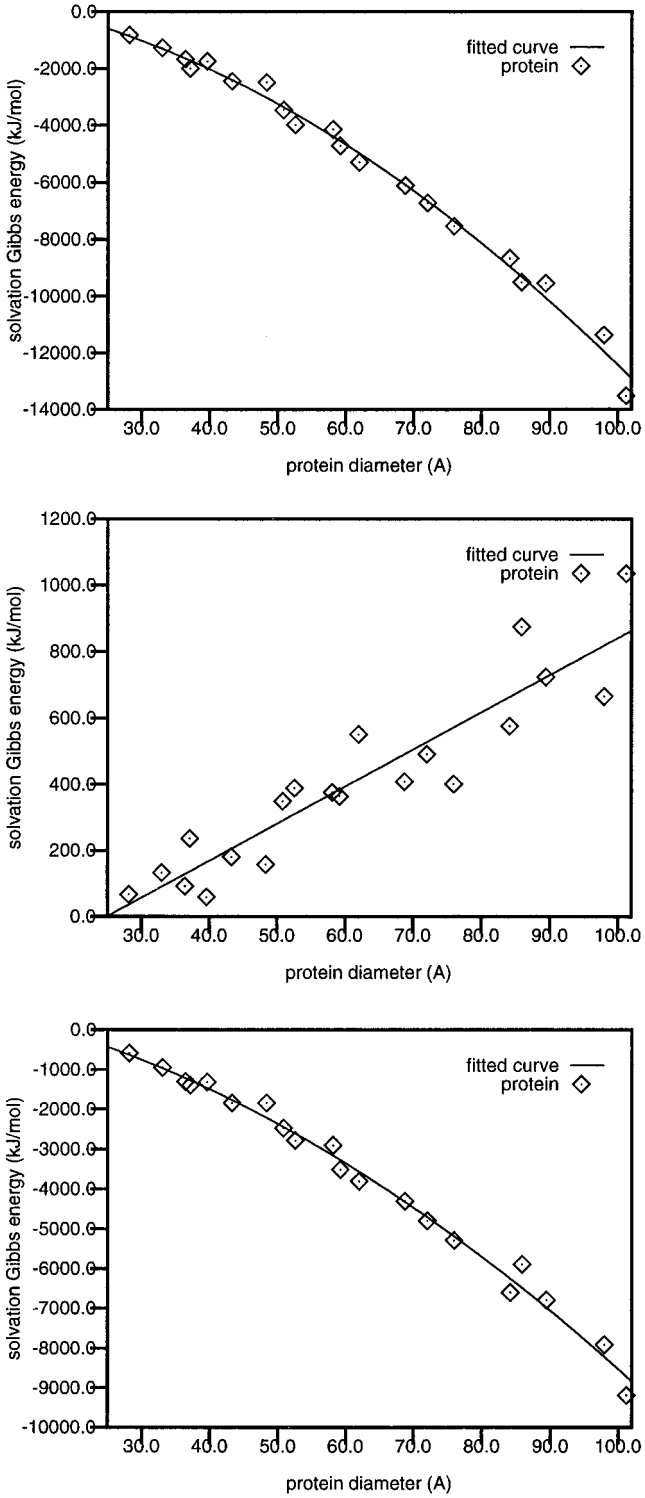

Figure 6. Solvation Gibbs energy contribution from the hydrogen bond interaction of the HΦI atoms on the surface of proteins. An atom is regarded as exposed to the solvent when its SASA is larger than 5% of its total SASA. (a, top) Solvation Gibbs energy of HΦI atoms when treated as if they were independent of each other. (b, middle) Solvation Gibbs energy contributed from the negative correlation among HΦI atoms. (c, bottom) Solvation Gibbs energy contributed from the positive correlation among HΦI atoms.

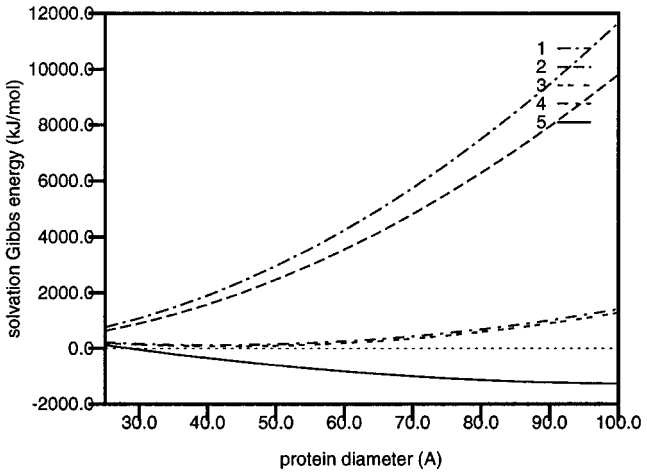

Figure 7. Calculated curves of solvation Gibbs energy as a function of the protein diameter, using 5% as the criteria in determining whether a HΦI atom is exposed to the solvent. Curve 1 is the solvation Gibbs energy of cavity formation. In curve 2, the soft part of interaction between protein and the solvent is turned on. In curve 3, we further turn on the hydrogen bond part of interaction by treating the HΦI atoms as if they were independent of each other. In curve 4, we add the negative correlation among HΦI atoms. Curve 5 is the total solvation Gibbs energy of protein, which results from the addition of positive correlation among HΦI atoms to curve 4.

large positive solvation Gibbs energy of the cavity formation. We next add $\Delta G^{*\text{indep/S,H}}$, which is the solvation Gibbs energy contributed by the hydrogen bond interaction of the surface HΦI atoms when they are treated as if they were independently solvated, and get curve 3. The addition of $\Delta G^{*\text{indep/S,H}}$ brings a large drop in solvation Gibbs energy. The addition of negative correlation among HΦI atoms only results in a slight increase in solvation Gibbs energy, which is shown as curve 4 in Figure 7. Finally, we added the positive correlation, and obtain the total solvation Gibbs energy of protein shown as curve 5. Clearly the positive correlation is quite important in the sense that it brings about a large drop in solvation Gibbs energy. For proteins of large size, the solvation Gibbs energy turns from increasingly positive to increasingly negative in relation to protein diameter.

Since the solvation Gibbs energy contributed by the hydrogen bond interaction of HΦI atoms is affected by the criteria one uses to determine whether a HΦI atom is exposed to the solvent or not, we carried out our calculation using 1% as the criteria. The results are listed in Table 5 and shown in Figure 8. In this case, each term of solvation Gibbs energy can be fitted as follows

$$\Delta G^{*\text{indep/S,H}} = 758.14 - 28.57d_p - 1.03d_p^2 \qquad (4.4)$$

$$\Delta G^{*\text{NC/S,H}} = -278.63 + 11.19d_p \qquad (4.5)$$

$$\Delta G^{*\text{PC/S,H}} = 241.05 - 7.72d_p - 0.46d_p^2 \qquad (4.6)$$

Figure 9 shows the solvation Gibbs energy of protein as a function of protein diameter when we used the 1% criteria. In this more "liberal" criteria for exposure to the solvent, we find that addition of the "independent" solvation Gibbs energies of the hydrophobic groups already makes the solvation Gibbs energy of the protein negative. A further significant reduction in the solvation Gibbs energy is obtained by adding the correlations between the hydrophilic groups.

The conclusion reached above is in sharp contrast to a conclusion reached a few years ago.[3,4] At that time, we have

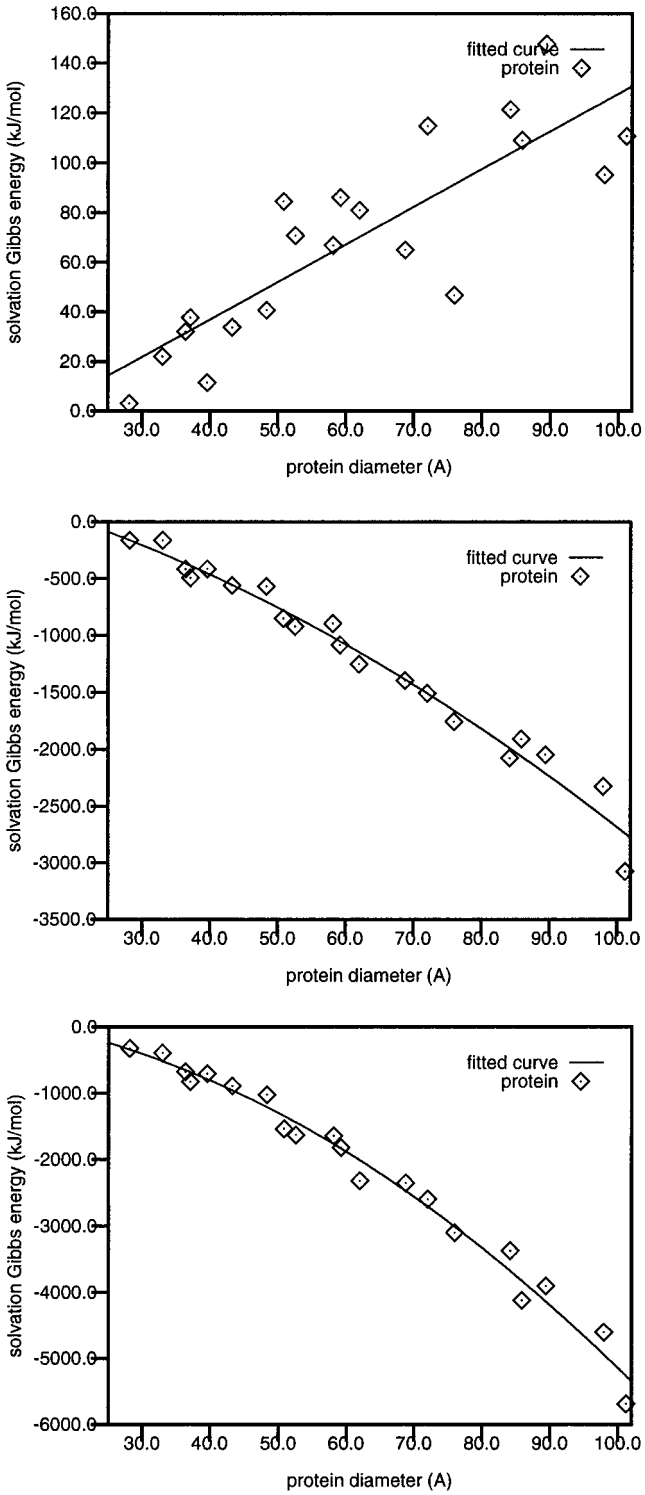

Figure 8. Solvation Gibbs energy contribution from the hydrogen bond interaction of the HΦI atoms on the surface of proteins. An atom is regarded as exposed to the solvent when its SASA is larger than 1% of its total SASA. (a, top) Solvation Gibbs energy of HΦI atoms when treated as if they were independent of each other. (b, middle) Solvation Gibbs energy contributed from the negative correlation among HΦI atoms. (c, bottom) Solvation Gibbs energy contributed from the positive correlation among HΦI atoms.

looked at the distance distribution of all HΦI groups exposed to the solvent. We found that there are roughly equal numbers of negatively correlated and positively correlated pairs; hence, assuming that magnitude of each correlation is the same, we have concluded that there is a net cancellation, i.e., the HΦI groups can be treated as if they were independently solvated. In the present work we found that the magnitude of negative correlations is relatively small compare with the positive

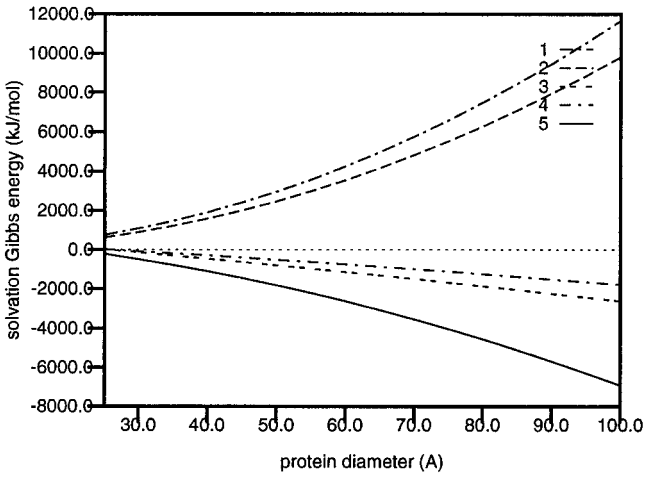

Figure 9. Calculated curves of solvation Gibbs energy as a function of the protein diameter, using 1% as the criteria in determining whether a HΦI atom is exposed to the solvent. Curve 1 is the solvation Gibbs energy of cavity formation. In curve 2, the soft part of interaction between protein and the solvent is turned on. In curve 3, we further turn on the hydrogen bond part of interaction by treating the HΦI atoms as if they were independent of each other. In curve 4, we add the negative correlation among HΦI atoms. Curve 5 is the total solvation Gibbs energy of protein, which results from the addition of positive correlation among HΦI atoms to curve 4.

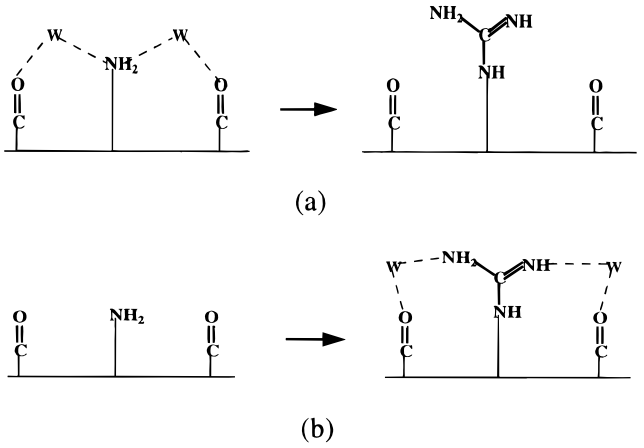

Figure 10. (a) An NH_2 group, forming two water bridges with carbonyl groups, is replaced by guanidino group, that cannot form such a bridge. The effect of such a replacement would cause a decrease in solubility. (b) An NH_2 group, which initially cannot form water bridges, can, after replacement form two water bridges, and hence increase the solubility of the protein.

correlations; hence, we have found that the latter are probably crucial in making the globular proteins soluble.

One more piece of experimental evidence for the importance of correlation is the following: Hughes et al.[22] found long ago that replacing an amino group on the surface of a protein by a guanidino group causes a marked decrease in the solubility of the protein. This finding is quite surprising. Assuming that the 3-D structure of the protein does not change upon this substitution (since it is done on amino groups on its surface), and since the guanidino group is more polar than the amino group, in the sense that it can form more hydrogen bonds with solvent molecules, we should have expected that such a replacement would bring about an *increase* in solubility rather than a *decrease*.

We are tempted to interpret these findings by invoking the correlations between HΦI groups. Thus, replacing an amino group by guanidino can bring about either the elimination of positive correlation (Figure 10(a)) or making new negative correlations, or both. If this is the case, then such a replacement would cause a net decrease in solubility. It is also conceivable that such a replacement would lead to an increase in solubility; an example of such a situation is depicted in Figure 10(b). We have recently estimated that the increase in solubility, as a result of the formation of one hydrogen-bonded bridge, could be of the order of $\exp[^3/_{0.6}] \approx 150$ (in an ideal geometrical situation). It is easy to imagine that replacements of a few amine groups on the surface of a protein could result in a very large change (increase or decrease) in the solubility of that protein.

It should be noted that any factor that makes the solvation Gibbs energy of a protein more negative, is also contributing to the relative stability of the protein. For instance, in the process of folding of a protein from the unfolded (U) to the folded (F) form

$$U \rightarrow F$$

the solvent induced driving force is

$$\delta G(U \rightarrow F) = \Delta G_F^* - \Delta G_U^* \qquad (4.7)$$

Hence, any factor that makes ΔG_F^* more negative will also increase the strength of the driving force for this process.

Acknowledgment. This work was partially supported by the Basic Research Foundation, administered by the Israel Academy of Sciences and Humanities.

References and Notes

(1) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987.
(2) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum Press: New York, 1992.
(3) Ben-Naim, A.; Ting, K.-L.; Jernigan, R. L. *Biopolymers* **1989**, *28*, 1309.
(4) Ben-Naim, A.; Ting, K.-K.; Jernigan, R. L. *Biopolymers* **1989**, *28*, 1327.
(5) Wilf, J.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 8594.
(6) Ben-Naim, A. *Biopolymers* **1990**, *29*, 567.
(7) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. *J. Chem. Phys.* **1959**, *31*, 369.
(8) Reiss, H. *Adv. Chem. Phys.* **1966**, *9*, 1.
(9) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(10) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(11) Ben-Naim, A. *J. Phys. Chem.* **1991**, *95*, 1437.
(12) Thanki, N.; Thornton, J. M.; Goodfellow, J. M. *J. Mol. Biol.* **1988**, *202*, 637.
(13) Finney, J. L. *J. Mol. Biol.* **1978**, *119*, 415.
(14) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(15) Savage, H. J.; Elliott, C. J.; Freeman, C. M.; Finney, J. L. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 2609.
(16) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(17) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.
(18) Ben-Naim, A. *J. Chem. Phys.* **1990**, *90*, 7412.
(19) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198.
(20) Mezei, M.; Ben-Naim, A. *J. Chem. Phys.* **1990**, *92*, 1359.
(21) Bernstein, F. C.; Koetzle, T. F.; William, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Eur. J. Biochem.* **1977**, *80*, 319.
(22) Hughes, W. L., Jr.; Saroff, H. A.; Carney, A. L. *J. Am. Chem. Soc.* **1949**, *71*, 2476.