Extent of Hydrogen-Bond Protection in Folded Proteins: A Constraint on Packing Architectures

Ariel Fernández*[†] and R. Stephen Berry[‡]

*Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637 USA, [†]Instituto de Matemática, Universidad Nacional del Sur, Consejo Nacional de Investigaciones Científicas y Técnicas, Bahía Blanca 8000, Argentina, and [‡]James Franck Institute and Department of Chemistry, The University of Chicago, Chicago, Illinois 60637 USA

ABSTRACT   Progressive structuring and ultimately exclusion of water by hydrophobes surrounding backbone hydrogen bonds turn the latter into guiding factors of protein folding. Here we demonstrate that an arrangement of five hydrophobes yields an optimal hydrogen-bond stabilization. This motif is shown to be nearly ubiquitous in native folds.

INTRODUCTION

The progressive structuring and ultimate removal of water surrounding amide–carbonyl hydrogen bonds turn the latter into major factors guiding the protein folding process (Makhatadze and Privalov, 1995; Krantz et al., 2000, 2002; Fernández, 2001, 2002; Baldwin, 2002; Fernández et al., 2002). This is so regardless of whether hydrophobic collapse triggers or is concurrent with secondary structure formation (Krantz et al. 2000; Baldwin, 2002). Increasing inaccessibility of hydrogen bonds to solvent takes place as the protein strategically places hydrophobes around its backbone polar groups (Vila et al., 2000; Garcia and Sanbonmatsu, 2002). This process induces hydrogen-bond formation as a means to compensate for the otherwise unfavorable hydrophobic-polar mismatches in the burial of the amides and carbonyls (Makhatadze and Privalov, 1995; Krantz et al. 2000). In this regard, natural questions arise and are addressed in this paper: What is the most effective way for a protein to protect its backbone hydrogen bonds? Is this efficiency associated with an optimal arrangement of hydrophobes? and Is this protective motif found in natural folds?

Here we approach these problems by establishing a relationship between the influence of shielding on electrostatics and the packing of hydrogen bonds. Side-chain packing regularities were first noted by others, notably by Jernigan and collaborators (Bahar and Jernigan, 1997), who have introduced effective context-dependent contact potentials. However, the regularity of the hydrophobic pattern has not been explicitly described in relation to the network of native hydrogen bonds. To the best of our knowledge, no regularity in the extent of hydrogen-bond desolvation across native folds has been reported so far.

We find that the maximum stabilization of a hydrogen bond is reached by surrounding it with five hydrophobes.

An examination of the protein data bank (PDB) revealed that this extent of protection is ubiquitous in native folds. The examination of the PDB was selective only in the sense that the protein sequences were obtained from the Orthopedic Web Links (OWL) database (Bleasby et al., 1994). This database emphasizes non-redundancy, an important factor in assessing the frequency of a structural pattern. Structural redundancies, in contrast, were avoided by intersecting our original data base with that containing only representative proteins and used for protein structure alignment by incremental combinatorial extension of the optimal path (Shindyalov and Bourne, 1998).

Our approach is rooted in the continuous dielectric model inspired by several seminal references (Pettitt and Karplus, 1988; Beglov and Roux, 1996; Bryant, 1996; Warshel and Papazyan, 1998; Petrey and Honig, 2000; and others), with the caveat that the characteristic length for solvent structuring, the size of the objects involved, and the ranges of their interactions are within the same order of magnitude, suggesting that future refinements should be based on a discrete counterpart tailored to the molecular scale.

Conformation-dependent environments can dramatically stabilize intramolecular dielectric-dependent interactions (Bryant, 1996; Warshel and Papazyan, 1998; Fernández, 2001; Fernández et al., 2002). Thus, such interactions may be protected from water attack when hydrophobic groups are in their vicinity. For example, the adjacency of a hydrophobic residue to a pair of hydrogen-bonded residues of the backbone creates a structured cavity in the local solvent environment, raising the free energy barrier for water to solvate the peptide backbone (Vila et al., 2000; Fernández, 2002; Fernández et al., 2002; Garcia and Sanbonmatsu, 2002). This translates, at the continuum level of description, as a decrease in solvent polarizability with a concurrent lowering of the dielectric function, thereby stabilizing the pairwise interaction.

To illustrate these ideas, the first part of this paper reports factual information on hydrophobic clustering around native backbone hydrogen bonds, whereas the second part describes a theoretical treatment to explain and justify the relative abundance of hydrophobic clusters of different sizes in relation to the optimization of their protective roles.

RESULTS

To report our results, we first define a desolvation sphere radius for a backbone hydrogen bond, fixing it at $R = 7.2$ Å. This value may be obtained relating R to the characteristic length Λ of the solvent-structuring effect due to the presence of a hydrophobe (see Theoretical Method). Thus, fixing Λ at 1.8 Å, the effective thickness of a single-layer water cavity (see below), and assuming an exponential decay of the structuring influence, we adopt $R = 4\Lambda = 7.2$ Å, a cutoff distance at which the structuring influence is 1% of its maximum value.

We now determine the average extent, ρ, of hydrogen-bond protection in a native fold. We define ρ for a single protein as the average number of hydrophobes that surround a backbone hydrogen bond. To obtain this value, we define the number $C_3 = C_3(R)$ of three-body correlations (Fernández, 2001; Fernández et al., 2002) in a native structure as the number of hydrophobic residues whose α carbon is contained in a desolvation sphere of radius R centered at the α carbon of one of the residues paired by a hydrogen bond. The results are qualitatively invariant if we adopt a β carbon representation. The counting includes the participants of the hydrogen bond itself if their side chains happen to be hydrophobic. Thus, if Q is the number of native backbone hydrogen bonds, we get: $\rho = \rho(R) = C_3(R)/Q$.

Fixing the Λ value at 1.8 Å, we find (see Table 1 and Fig. 1) a ρ value in the range 5.00 ± 0.38 for 95.66% of the 3358 PDB soluble proteins examined. Within this ensemble, 70.02% of the proteins lie in the range 5.00 ± 0.16. The dispersion σ in the extent of protection within a protein remained invariably below 18% of the average value. We also note that every hydrophobe of an autonomous folding protein is a hydrogen-bond protector and that no backbone hydrogen bond has less than two protectors or more than eight (in the latter case, most of them are alanines).

In the present calculations, the criterion for counting a backbone hydrogen bond as such was an interaction energy less than $-kT/2$. This definition results in the number of hydrogen bonds being generally less by about 20% (19% on average) than the number obtained using a definition based on a geometric criterion allowing for a latitude of 45° in the angle between the amide and carbonyl vectors, and nitrogen–oxygen distance less than 4 Å. The use of the geometric definition, along with the desolvator centered at the β carbon, rather than α carbon, has a low quantitative impact on the results shown. For example, the average ratio of the number of three-body correlations versus the number of hydrogen bonds remained near five (5.31) across our PDB sample. An exception was the dispersion in the average number of three-body correlations per hydrogen bond, which was about two-fold larger than that reported in Table 1.

A more precise but less direct measure consists of counting the average number of carbonaceous groups (CH_i, $i = 1$, 2, 3) within the desolvation spheres of the hydrogen bonds.

TABLE 1 Number of three-body correlations (C_3), amide–carbonyl hydrogen bonds (Q), average extent of hydrogen bond protection (ρ) and dispersion in the extent of protection (σ) for native structures of autonomous folders identified by their PDB accesion codes

PDB Code	C_3	Q	ρ	σ (%)
1aa2	257	52	5.04	10.18
1lou	242	47	5.15	13.05
1ris	230	45	5.11	12.87
1alf	20	4	5.00	10.02
1aue	250	49	5.10	11.80
256b	394	75	5.25	16.05
1aac	170	33	5.15	11.28
1abq	80	15	5.33	14.06
1aoj	39	8	4.87	15.47
1bo3	5	1	5.00	0.00
1ubi	155	31	5.00	10.06
1gb4	80	16	5.00	10.14
1lmb	260	51	5.09	8.67
1srl	40	8	5.00	12.83
2ptl	74	16	4.62	16.33
1crc	136	28	4.85	9.60
1cw6	32	6	5.33	14.02
1vii	30	6	5.00	12.55
1hhh	446	86	5.18	12.68
1mim	318	64	4.96	17.62
1ifb	215	43	5.00	8.83
1hhg	468	95	4.92	11.09
1e4j	225	45	5.00	12.11
1e4k	233	46	5.07	11.15
1gff-1	612	124	4.93	11.58
1csk-A	111	22	5.04	12.01
1c3t	105	21	5.00	10.78
1a6v	172	33	5.21	17.91
1beb	345	57	6.05	14.06
1cqx	1142	188	6.07	12.89

The radius of the desolvation sphere for a hydrogen bond was fixed at 7.2 Å. Two definite outliers to $\rho = 5$ are indicated in the last two separated rows.

This number is 15.08 ± 1.20 for the same percentage of proteins, with a dispersion invariably lower than 22.25%. In contrast, the average number of polar moieties within the desolvation spheres varies widely, from zero to eight across the PDB, and no straightforward statistical inference can be made, not even by sampling hydrogen bonds within similar regions (surface or interior).

Regardless of the measure of hydrogen-bond desolvation adopted, the nearly constant ρ value and its relatively low dispersion over ensembles of soluble proteins determines an architectural constraint on the packing of protein structure. Furthermore, the nearly constant ρ value is a reflection of the generic composition of natural protein chains. In contrast, the dispersion is essentially due to the wide range of side-chain sizes, which implies that proper desolvation of a hydrogen bond may be achieved with less than five large hydrophobes (Trp, Phe), or alternatively might require more than five small hydrophobes (Ala). With the sole exception of cellular prion proteins (Prusiner, 1998), very few hydro-
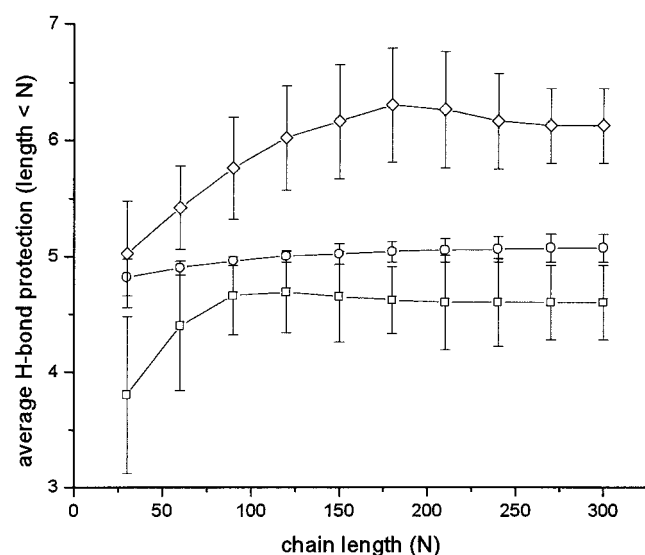
FIGURE 1 Extent of protection of amide–carbonyl hydrogen bonds in the native fold averaged over ensembles of 1000 proteins of length < N. The extent of protection is defined in relation to a desolvation sphere of radius 7.2 Å (*circles*), 6 Å (*squares*), and 8 Å (*diamonds*).

FIGURE 2 Illustration of the most effective protective motif ($n = 5$) for an amide–carbonyl hydrogen bond (*thick segment*). The hydrophobic residues (*spheres*) are arranged in a trigonal bipyramid surrounding the hydrogen bond with three residues equidistant from the carbonyl oxygen (O) and amide hydrogen (H).

gen bonds are strictly under-desolvated in the native folds of soluble proteins, as described below.

The mean and dispersion of ρ averaged over ensembles of 600 soluble proteins of length $<N$ is given in Fig. 1. The circle, diamond, and square plots correspond, respectively, to desolvation radii $4\Lambda = 7.2$, 8, and 6 Å. For each N ensemble, every chain length is equally represented. These statistics reveal large fluctuations in the average extent of protection as we depart from the characteristic length $\Lambda = 1.8$ Å, suggesting a fine tuning of ρ to the spatial scale at which native structure is examined and important irregularities in the packing of hydrogen bonds at other scales.

As Tables 1 and 2 reveal, $\rho = 5$ imposes a constraint on the packing of stable soluble proteins. Furthermore, this ubiquitous average extent of backbone hydrogen-bond protection suggests that an arrangement of five hydrophobes
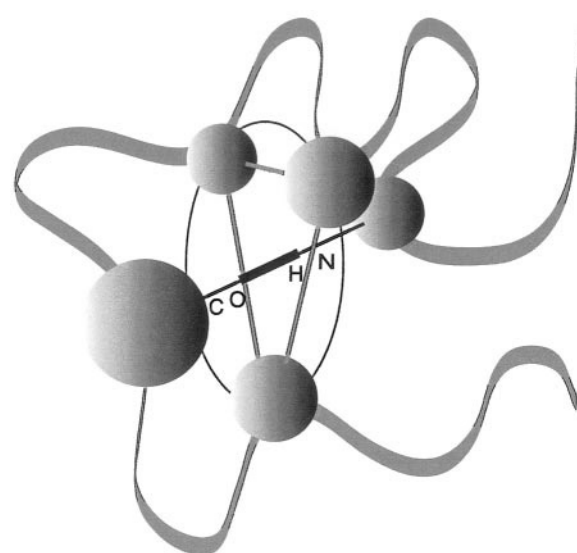
TABLE 2 **Number of protective clusters of different size for proteins identified by their pdb accession codes**

PDB Code	Chain Length	Protective Cluster Size				
		$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
1gb4	57	2	2	9	1	2
1ubi	76	4	2	18	3	4
1lmb	92	1	6	36	4	4
1aa2	108	3	2	39	4	3
1a6v	110	3	1	20	4	5
1e4j	176	4	2	33	2	4
1mim	215	6	4	42	3	8
1hhg	275	9	2	72	5	6
1hhh	275	2	3	68	5	8
1gff-1	426	14	5	91	5	9

represents the best compromise between proximity to the hydrogen bond and number of units that may be placed within its desolvation sphere (Fig. 2), a result corroborated by our theoretical results.

Of the two five-hydrophobe arrangements, the square pyramid and the trigonal bipyramid (Fig. 2), the latter, albeit sometimes distorted, proved to be the one shaping the interior for some of the shortest ($\sim$1.77 Å) and almost collinear amide–carbonyl hydrogen bonds. Its relative abundance with respect to the square pyramid is 3:1.

Figure 3, *A–E*, displays the optimal protective pattern on selected native hydrogen bonds. Filled circles represent the α and β carbon of hydrophobic residues, thick two-color segments joining α carbons denote backbone hydrogen bonds (blue = amide, red = carbonyl), thick light gray segments indicate a double hydrogen bond where each residue contributes both its amide and carbonyl group, and thin blue lines to the center of hydrogen bonds represent 3-body correlations indicating hydrogen-bond protection by hydrophobes. The Val-73–Phe-76 hydrogen bond in *lambda repressor* (pdb.1lmb) is displayed in Fig. 3 *A*. It is protected by Leu-18, Ala-66, Ala-81, Val-73, and Phe-76. Figure 3 *B* displays the Ile-8–Gly-15 hydrogen bond for *hyperthermophile variant protein G* (pdb.1gb4). In this case we have an $n = 6$ cluster of protectors: Leu-6, Ile-7, Ile-8, Leu-13, Ile-17, and Val-55 containing a distorted trigonal bipyramid arrangement with the sixth residue, the distant Val-55 (7.22 Å), being almost collinear with Ile-8. Figure 3 *C* displays the $n = 5$ protective cluster for the strong Met-1–Val-17 hydrogen bond of *ubiquitin* (pdb.1ubi). Again we find that the protective units Met-1, Ile-3, Leu-15, Val-17 and Pro-19
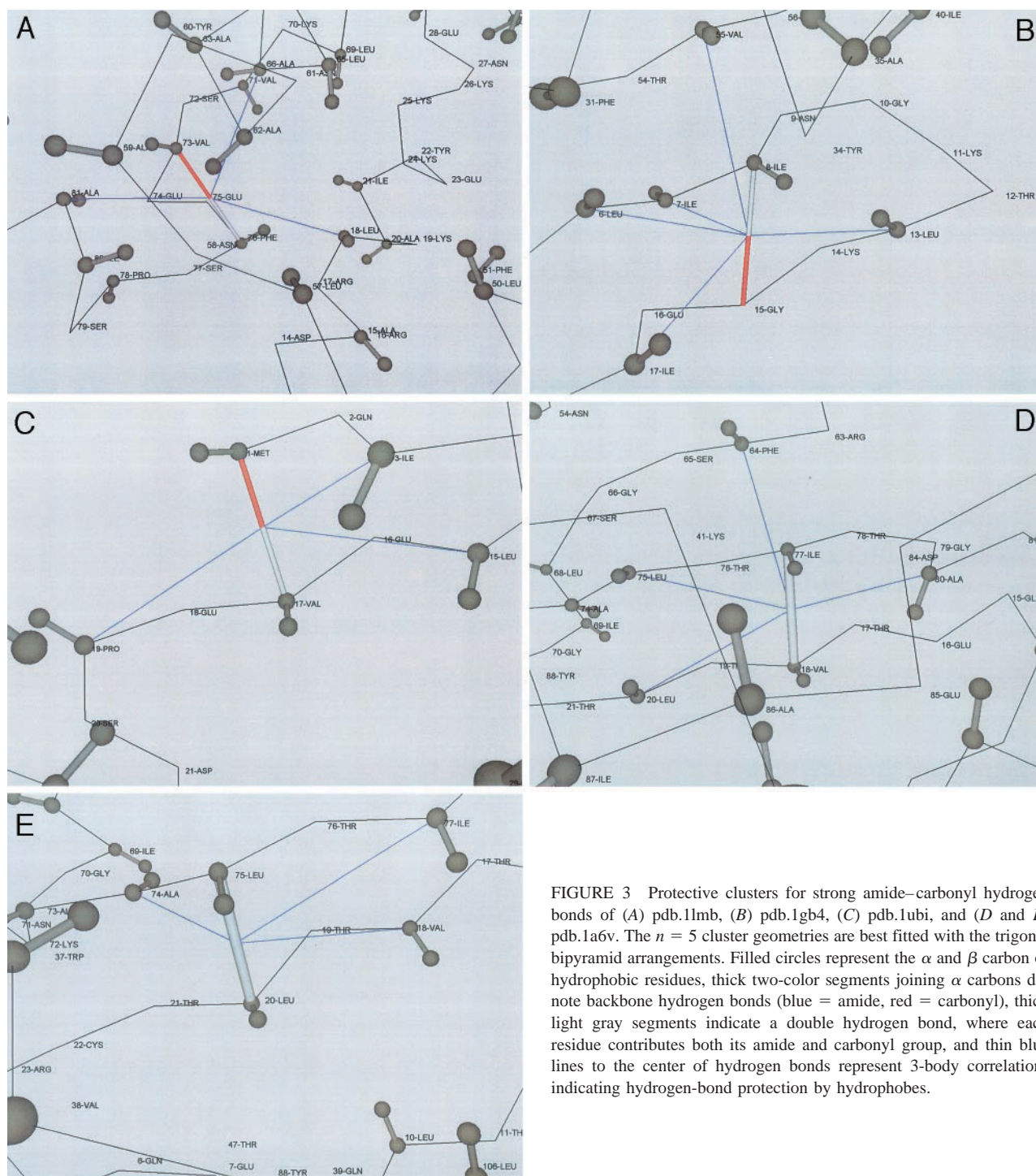
FIGURE 3 Protective clusters for strong amide–carbonyl hydrogen bonds of (A) pdb.1lmb, (B) pdb.1gb4, (C) pdb.1ubi, and (D and E) pdb.1a6v. The $n = 5$ cluster geometries are best fitted with the trigonal bipyramid arrangements. Filled circles represent the α and β carbon of hydrophobic residues, thick two-color segments joining α carbons denote backbone hydrogen bonds (blue = amide, red = carbonyl), thick light gray segments indicate a double hydrogen bond, where each residue contributes both its amide and carbonyl group, and thin blue lines to the center of hydrogen bonds represent 3-body correlations indicating hydrogen-bond protection by hydrophobes.

are assembled in an approximate trigonal bipyramid cluster. Finally, two adjacent double hydrogen bonds with $n = 5$ protective cluster are displayed for immunoglobulin (pdb.1a6v) in Fig. 3, D and E. They are, respectively, Val-18–Ile-77 and Leu-20–Leu-75. Their protective clusters are, respectively, Val-18, Ile-77, Ala-80, Leu-20, Leu-75, and Leu-20, Leu-75, Val-18, Ala-74, Ile-77. The rela-

tively close (~7.30 Å), Phe-64 is again virtually collinear with Ile-77. Thus, although we have identified an $n = 5$ cluster protecting the Val-18–Ile-77 double hydrogen bond, strictly speaking, its extent of protection should be taken to be either $n = 5$ or $n = 6$.

If we allow for more latitude in the desolvation radius, we find that the trigonal bipyramid arrangement is sometimes

(~⅓ of the cases) part of a larger $n = 6$ protective cluster, with the sixth protector being almost as close as the other five.

Severely under-desolvated hydrogen bonds, i.e., those with less than three hydrophobes in their desolvation shells, are rare and constitute, on average, less than 2% of the total number of hydrogen bonds in native folds. The β subunit for hemoglobin has two, as indicated in Fig. 4 A. Strikingly, these bonds are located next to Glu-6, the residue whose mutation is responsible for sickle cell anemia (Branden and Tooze, 1991). In contrast, cellular prion proteins (Prusiner, 1998) have ~40% of their hydrogen bonds severely under-desolvated (Fig. 4 B). Significantly, with $\rho = 3.4$–3.7, prions are the definite outliers to the $\rho = 5$ constraint.

THEORETICAL METHOD

The purpose of this section is to demonstrate that the commonly found average extent of hydrogen-bond protection $\rho = 5$ may be understood by identifying the geometry of the hydrophobic cluster that best exerts the protection. Let us place the carbonyl oxygen effective charge q at the center of coordinates, define the x axis as that along the carbonyl–amide hydrogen bond, and place the amide hydrogen at position $\mathbf{r}$, 1.4 to 2.1 Å away along the positive x axis. We assume the hydrogen bond to be surrounded by a discrete number of identical spherical hydrophobic units of radius $d/2$ (the parameter d will be defined below) centered at fixed positions $\mathbf{r}_j$, $j = 1, 2, \ldots, n$. This is clearly an idealized case but one that can be dealt with analytically.

Previous extrapolations of macroscopic treatments used to evaluate an electric field $E(\mathbf{r})$ at position $\mathbf{r}$ are based on a local averaging of the solvent environment and thereby, on a mean-field position-dependent dielectric (Beglov and Roux, 1996; Bryant, 1996; Warshel and Papazyan, 1998; Petrey and Honig, 2000). To deal with the context of interest here, this approach would require that we take into account the solvent structuring in the proximity of hydrophobes, estimate the spatial propagation of such effects and determine their influence on the coulomb screening. In accord with current research, the adaptation of a macroscopic electrostatic approach to deal with such a microscopic context would impose a breakdown in the isotropy of the local dielectric and demand a careful identification of the spatial boundary that defines the region where a bulk-like macroscopic treatment may still hold valid.

An ultimately more convenient approach applicable in our context of interest is based on three operational tenets: determination of the perturbation in the diffraction structure of bulk water given in frequency space as hydrophobes are incorporated; recovery of their solvent-structuring effect by inverse Fourier transforming the previous result; and propagation in space of the solvent-structuring effect around hydrophobes by assuming that the field at position $\mathbf{r}$ is correlated with the field at any neighboring position $\mathbf{r}'$.

Thus, to effectively propagate the solvent-structuring effect due to the presence of hydrophobes, we start by replacing the position-dependent dielectric by a convolution of the electric field with a kernel representing the correlations. This leads to

$$\mathrm{div}\left[\int K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})E(\mathbf{r}')\, d\mathbf{r}'\right] = 4\pi q \delta(\mathbf{r}), \qquad (1)$$

where the integral kernel $K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})$ is parametrically dependent on the fixed hydrophobe positions. In the absence of vicinal hydrophobic units, the correlations decay as $\exp(-\|\mathbf{r} - \mathbf{r}'\|/\xi)$ (ξ = characteristic correlation length defined below). In the limit $\xi \to 0$, we get: $K(\mathbf{r}, \mathbf{r}') \sim \delta(\mathbf{r}' - \mathbf{r})$, and thus Eq. 1 becomes the standard Poisson equation.
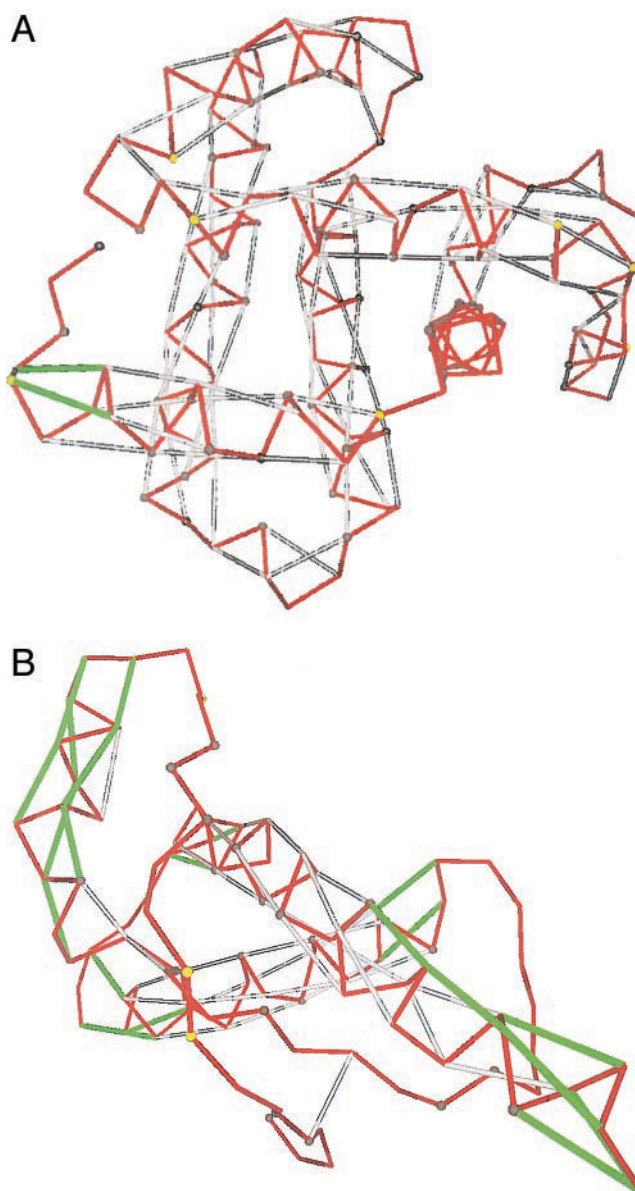
FIGURE 4 Identification of severely under-desolvated hydrogen bonds in (A) hemoglobin β subunit (pdb.1bz0-chain B) and (B) human cellular prion protein (pdb.1qm0). A simplified description has been adopted for clarity. The backbone is represented as a red line made of α-carbon virtual bonds. Desolvated hydrogen bonds are simply represented as gray segments joining α carbons (no distinction is made between single and double bonds), whereas severely underdesolvated hydrogen bonds are shown as green segments. Hydrophobic residues are shown as gray disks on α carbons, whereas over-exposed hydrophobes (less than 33% buried) are indicated as yellow disks.

The correlation kernel may be identified taking into account the relationship between diffraction and dielectric. For bulk water, we get $K(\mathbf{r}, \mathbf{r}') = K(\mathbf{r} - \mathbf{r}')$ by inverse transforming its frequency $\mathbf{k}$ vector representation

$$K(\mathbf{r} - \mathbf{r}') = \int \exp[i\mathbf{k}(\mathbf{r} - \mathbf{r}')]L(\mathbf{k})\, d\mathbf{k}. \qquad (2)$$

Where the distribution $L(\mathbf{k}) = [(\epsilon_w - \epsilon_0)/(1 + \epsilon_w\|\mathbf{k}\|^2\xi^2/\epsilon_0) + \epsilon_0]$ reflects the fact that, due to dipole reorientation inertia, water becomes significantly polarized only when interacting with low-frequency radiation. In Eq. 2, ξ denotes the characteristic length, here fixed at $\xi \approx 5$ Å, and ϵ_w, ϵ_0 are the permittivities of water and vacuum, respectively.

To obtain the correlation kernel with n hydrophobic units at fixed positions, we modify Eq. 2 to incorporate their solvent-structuring effect,

$$K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\}) = \left[\int \exp[i\mathbf{k}(\mathbf{r} - \mathbf{r}')]L(\mathbf{k})\,d\mathbf{k}\right]$$

$$\times \left[1 + \sum_{j=1,\ldots,n}\Gamma_j(\mathbf{r}, \mathbf{r}')\right], \quad (3)$$

where $\Gamma_j(\mathbf{r}, \mathbf{r}') \sim \exp[-(\|\mathbf{r} - \mathbf{r}_j\| + \|\mathbf{r}' - \mathbf{r}_j\|)/\Lambda]$, for $\|\mathbf{r} - \mathbf{r}_j\|$ and $\|\mathbf{r}' - \mathbf{r}_j\|$ d/2 (Fig. 3 C). For Eq. 3, we assume a characteristic length Λ for water structuring. This parameter will be fixed at 1.8 Å, the effective thickness of a single-layer water cavity. Our conclusions are qualitatively invariant for different ξ values provided $\xi' > \Lambda$ and remain robust under changes in Λ, as shown below.

We now solve Eq. 1 using Eq. 3 and performing a Fourier transformation. In Fourier conjugate $\mathbf{k}$-vector representation, the kernel has a perturbation term involving the convolution of the normal distribution of frequencies for bulk water with a structure factor defined by the set of frequency vectors $\{\mathbf{k}_j\}$, conjugate to the set of fixed hydrophobe positions. Thus, we get

$$\int \exp(-i\mathbf{k}\mathbf{r})\int \exp(-i\mathbf{k}\mathbf{r}')K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})\,d\mathbf{r}'\,d\mathbf{r}$$

$$= K(\mathbf{k}, \{\mathbf{k}_j\})$$

$$= L(\mathbf{k}) + L(\mathbf{k})$$

$$\otimes \left\{\sum_{j=1,\ldots,n}\exp(-i\mathbf{k}\mathbf{r}_j)[1/(1 + \|\mathbf{k} - \mathbf{k}_j\|^2\Lambda^2)]\right\}, \quad (4)$$

where the symbol 196 stands for convolution, and the factor

$$\left\{\sum_{j=1,\ldots,n}\exp(-i\mathbf{k}\mathbf{r}_j)[1/(1 + \|\mathbf{k} - \mathbf{k}_j\|^2\Lambda^2)]\right\}$$

gives the structure determined by the spatial distribution of the hydrophobes.

Now we may get the electric field form by inverse Fourier transformation of the solution to Eq. 1 given in $\mathbf{k}$ representation,

$$\int E(\mathbf{r})\,d\mathbf{r} = -(4\pi q)\int \exp(i\mathbf{k}\mathbf{r})\|\mathbf{k}\|^{-2}[K(\mathbf{k}, \{\mathbf{k}_j\})]^{-1}\,d\mathbf{k}.$$

$$(5)$$

Direct residue evaluation at the first-order poles $k = \pm i(\epsilon_0/\epsilon_\omega)^{1/2}\xi^{-1}$ ($k = \|\mathbf{k}\|$) and $\mathbf{k} = \mathbf{k}_j \pm i\Lambda^{-1}$, yields the electric field $E(\mathbf{r})$ by retaining only the real part of the residue calculation,

$$E(\mathbf{r}) = (q/\mathbf{r}^2)[(\varepsilon_0^{-1} - \varepsilon_w^{-1})$$

$$\times \Omega(\{\mathbf{r}_j\})(1 + \mathbf{r}/\xi)\exp(-\mathbf{r}/\xi) + \varepsilon_w^{-1}], \quad (6)$$

where

$$\Omega(\{\mathbf{r}_j\}) = \prod_{j=1,\ldots,n}[1 + \exp(-\|\mathbf{r}_j\|/\Lambda)]$$

$$\times [(1 + \exp(-\|\mathbf{r} - \mathbf{r}_j\|/\Lambda)], \quad (7)$$

and $r = \|\mathbf{r}\|$. Eqs. 6 and 7 reveal that the net effect of the hydrophobic arrangement on the electric field can be captured by replacing the reciprocal permittivity constant ϵ_w^{-1} for bulk water by the quantity

$$[(\varepsilon_0^{-1} - \varepsilon_w^{-1})\Omega(\{\mathbf{r}_j\})(1 + \mathbf{r}/\xi)\exp(-\mathbf{r}/\xi) + \varepsilon_w^{-1}],$$

which is dependent on $\mathbf{r}$ and parametrically dependent on the hydrophobe positions. As expected, this quantity tends to the bulk-limit value ϵ_w^{-1} as $r/\xi \to \infty$. Thus, we may define an effective permittivity $\epsilon = \epsilon(\mathbf{r}, \{\mathbf{r}_j\})$ as

$$\varepsilon = [(\varepsilon_0^{-1} - \varepsilon_w^{-1})\Omega(\{\mathbf{r}_j\})(1 + \mathbf{r}/\xi)\exp(-\mathbf{r}/\xi) + \varepsilon_w^{-1}]^{-1}.$$

$$(8)$$

Because $\epsilon_0^{-1} \gg \epsilon_w^{-1}$, Eq. 8 implies that, for a fixed $\mathbf{r}$, finding the surrounding hydrophobic cluster with the lowest dielectric in its interior is tantamount to finding the arrangement $\{\mathbf{r}_j\}$ that maximizes the function $\Omega(\{\mathbf{r}_j\})$.

Eqs. 6–8 reveal that the spatial distribution of the hydrophobic moieties around a charged group is responsible for an enhancement of the electric field when their positions lie at distances comparable to the characteristic length ξ. The neighboring hydrophobes introduce a local inhibition of solvent polarizability responsible for a drastic decrease in the permittivity constant.

We now find the optimal arrangement $\{\mathbf{r}_j\}$ of hydrophobes that yields the maximum value Ω^* for $\Omega(\{\mathbf{r}_j\})$. First, we compute the maximum $\Omega^*(n)$ of $\Omega(\{\mathbf{r}_j\})$ for each fixed n using the method of Lagrange multipliers, and incorporating as constraints, a fixed minimum distance d between any two hydrophobes and the fact that all hydrophobes are placed within a desolvation sphere for the hydrogen bond. The distance d will be taken to be 5 or 6 Å in accord with typical minimal distances between α carbons of nonadjacent (noncovalently linked) residues in standard secondary or tertiary structure motifs (Branden and Tooze, 1999). Our results are qualitatively invariant down to the lowest boundary value $d = 4.5$ Å. Two cases must be distinguished: $1 \leq n < 4$ and $n \geq 4$.

For $n = 1, 2, 3$, the lowest dielectric is achieved by distributing the hydrophobes equidistantly from the O and H atoms, that is, in circles centered in the middle of the hydrogen bond and orthogonal to the x axis. However, because it follows from direct inspection of the PDB, no native amide–carbonyl hydrogen bond has fewer than two hydrophobic protectors in a native structure, even if we adopt the stringent definition of protector as a hydrophobe lying within 6 Å from the center of the hydrogen bond. So, we shall only treat the case $n = 2, 3$ and, separately, the case $n \geq 4$.

For $n = 2, 3$, the arrangements yielding the lowest dielectric are obtained, respectively, by placing the hydrophobes as antipodes on a circle of radius d or in an equilateral triangle with side equal to d, and with the hydrophobes equidistant from both the O and H atoms. Thus, we get

$$\Omega^*(2) = 2.771 \quad d = 5 \text{ Å};$$

$$\Omega^*(3) = 2.090 \quad d = 6 \text{ Å},$$

$$\Omega^*(3) = 2.869 \quad d = 5 \text{ Å};$$

$$\Omega^*(3) = 2.199 \quad d = 6 \text{ Å}.$$

$$(9)$$

For $n \geq 4$, the optimal arrangement is invariably obtained by fixing $n - 2$ hydrophobes at distance d from each other and equidistantly from the O and H atoms, and placing the remaining two hydrophobes along the x axis

at distance $(\Lambda + \eta)(1 - n^{-2})$ (to first approximation) away from the C and N atoms, with $\eta = $ C–O distance in the carbonyl group. An illustration of one such arrangement for $n = 5$ is given in Fig. 2. To fix notation, we shall denote by R(n) the distance to the O or H atom from any of the $n - 2$ hydrophobes equidistant to those atoms.

This gives, for $n = 4$ (tetrahedron), R(4) = 2.596 Å ($d = 5$ Å), R(4) = 3.08 Å ($d = 6$ Å):

$$\Omega^*(4) = 3.419 \quad d = 5 \text{ Å};$$
$$\Omega^*(4) = 2.843 \quad d = 6 \text{ Å}. \tag{10}$$

For $n = 5$ (trigonal bipyramid, Fig. 2), R(5) = 2.970 Å ($d = 5$ Å), R(5) = 3.534 Å ($d = 6$ Å), we get

$$\Omega^*(5) = 4.144 \quad d = 5 \text{ Å};$$
$$\Omega^*(5) = 3.177 \quad d = 6 \text{ Å}. \tag{11}$$

The alternative square pyramid arrangement for $n = 5$ does not obey the constraints to be satisfied by an optimal solution given above. Its values are

$$\Omega^*(5) = 3.955 \quad d = 5 \text{ Å};$$
$$\Omega^*(5) = 2.900 \quad d = 6 \text{ Å}. \tag{12}$$

For $n = 6$ (square bipyramid), R(6) = 3.604 Å ($d = 5$ Å), R(6) = 4.297 Å ($d = 6$ Å), we get

$$\Omega^*(6) = 3.952 \quad d = 5 \text{ Å};$$
$$\Omega^*(6) = 2.898 \quad d = 6 \text{ Å}. \tag{13}$$

For $n = 7$ (pentagonal bipyramid), R(7) = 4.310 Å ($d = 5$ Å), R(7) = 5.120 Å ($d = 6$ Å), we get

$$\Omega^*(7) = 3.421 \quad d = 5 \text{ Å};$$
$$\Omega^*(7) = 2.515 \quad d = 6 \text{ Å}. \tag{14}$$

Similar calculations for the entire range $0 < \Lambda < 2.1$ Å adopting any of the two minimum hydrophobe distances invariably yield the order

$$\Omega^*(3) < \Omega^*(4) < \Omega^*(5) \quad \text{and} \quad \Omega^*(n) > \Omega^*(n + m)$$
$$\text{for} \quad (n \geq 5, m > 0). \tag{15}$$

Thus, within the ranges of parametrization given above, we have proven the following result: *A hydrogen bond is embedded in the lowest dielectric when surrounded by five hydrophobes and, given the constraints on optimal solutions resulting from the Lagrange multipliers method, the trigonal bipyramid arrangement (Fig. 2) is the most effective protecting motif.* In practice, this motif is realized only approximately due to the diversity of shapes and sizes of the hydrophobic side chains (cf. Fig. 3, *A–E*). In contrast, the protection number $n = 5$ and the average extent of protection $\rho = 5$ appear to be by far the most common (Tables 1 and 2).

CONCLUSION

Our results reveal that the hydrophobic surface burial preceding or concurrent with hydrogen bond formation is needed to modulate the electrostatics that warrants the ultimate survival of hydrogen bonds. Furthermore, we have shown that the optimal extent of hydrogen bond protection is achieved by $n = 5$ hydrophobic clusters, a nearly ubiquitous arrangement in native folds.

REFERENCES

Bahar, I., and R. L. Jernigan. 1997. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophobic interactions at close separation. *J. Mol. Biol.* 266:195–214.

Baldwin, R. L. 2002. Making a network of hydrophobic clusters. *Science.* 295:1657–1658.

Beglov, D., and B. Roux. 1996. Solvation of complex molecules in a polar liquid. An integral equation theory. *J. Chem. Phys.* 104:8678–8689.

Bleasby, A. J., D. Akrigg, and T. K. Attwood. 1994. OWL: a nonredundant composite protein sequence database. *Nucl. Acids Res.* 22:3574–3577.

Branden, C., and J. Tooze. 1991. Introduction to Protein Structure. Garland Publishing, Inc., New York.

Bryant, R. G. 1996. The dynamics of water–protein interactions. *Annu. Rev. Biophys. Biomol. Struct.* 25:29–53.

Fernández, A. 2001. Conformation-dependent environments in folding proteins. *J. Chem. Phys.* 114:2489–2502.

Fernández, A. 2002. Time-resolved backbone desolvation and mutational hot spots in folding proteins. *Proteins.* 47:447–457.

Fernández, A., A. Colubri, and R. S. Berry. 2002. Three-body correlations in protein folding: the origin of cooperativity. *Physica A.* 307:235–259.

Garcia, A. E., and K. Y. Sanbonmatsu. 2002. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* 99:2782–2787.

Krantz, B. A., L. B. Moran, A. Kentsis, and T. R. Sosnick. 2000. D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nature Struct. Biol.* 7:62–71.

Krantz, B. A., A. K. Srivastava, S. Navli, D. Baker, R. T. Sauer, and T. R. Sosnick. 2002. Understanding protein hydrogen bond network formation with kinetic D/H amide isotope effects. *Nature Struct. Biol.* 9:458–463.

Makhatadze, G. I., and P. L. Privalov. 1995. Energetics of protein structure. *Adv. Protein Chem.* 47:307–425.

Petrey, D., and B. Honig. 2000. Free energy determinants of tertiary structure and the evaluation of protein models. *Protein Sci.* 9:2181–2191.

Pettitt, B. M., and M. Karplus. 1988. Conformational free energy of hydration for an alanine dipeptide: thermodynamic analysis. *J. Phys. Chem.* 92:3994–3997.

Prusiner, S. B. 1998. Prions. *Proc. Natl. Acad. Sci. U.S.A.* 95: 13363–13383.

Shindyalov, I. N., and P. E. Bourne. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 11:739–747.

Vila, J. A., D. R. Ripoll, and H. A. Scheraga. 2000. Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proc. Natl. Acad. Sci. U.S.A.* 97: 13075–13079.

Warshel, A., and A. Papazyan. 1998. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* 8:211–217.