Chapter 7

Wrapping electrostatic bonds

For a protein structure to persist in water, its electrostatic bonds must be shielded from water attack [71, 79, 183, 224]. This can be achieved through wrapping by nonpolar groups (such as CH_n, $n = 1, 2, 3$) in the vicinity of electrostatic bonds to exclude surrounding water [71]. Such desolvation enhances the electrostatic contribution and stabilizes backbone hydrogen bonds [17]. In a nonbonded state, exposed polar amide and carbonyl groups which are well wrapped are hindered from being hydrated and more easily return to the bonded state [45], as depicted in Figure 3.7.

The thermodynamic benefit associated with water removal from pre-formed structure makes under-wrapped proteins adhesive [65, 72, 74]. As shown in [71], under-wrapped hydrogen bonds (UWHB's) are determinants of protein associations. In Section 8.1, we describe the average adhesive force exerted by an under-wrapped hydrogen bond on a test hydrophobe.

The dielectric environment of a chemical bond can be enhanced in different ways, but wrapping is a common factor. There are different ways to quantify wrapping. Here we explore two that involve simple counting. One way of assessing a local environment around a hydrogen bond involves just counting the number of 'hydrophobic' residues in the vicinity of a hydrogen bond. This approach is limited for two reasons.

The first difficulty of this approach relates to the taxonomy of residues being used. The concept of 'hydrophobic residue' appears to be ambiguous for several residues. In some taxonomies, Arg, Lys, Gln, and Glu are listed as hydrophilic. However, we will see that they contribute substantially to a hydrophobic environment. On the other hand, Gly, Ala, Ser, Thr, Cys and others are often listed variously as hydrophobic or hydrophilic or amphiphilic. We have identified these five residues in Chapter 4 as among the most likely to be neighbors of underwrapped hydrogen bonds, as will be discussed at more length in Chapter 6. As noted in Section 4.5.1, glycine, and to a lesser extent alanine, can be viewed as polar, and hence hydrophilic, but alanine has only a nonpolar group in its sidechain representation and thus would often be viewed as hydrophobic.

A second weakness of the residue-counting method is that it is based solely on the residue level and does not account for more subtle, 'sub-residue' features. We will see that these limitations can be overcome to a certain extent with the right taxonomy of residues. However, we will also consider (Section 7.3) a measure of wrapping that looks into the sub-residue structure by counting all neighboring non-polar groups. The residue-counting method is included both for historical and

atomic symbol	H	C	N	O	F	Na	Mg	P	S
electronegativity	2.59	2.75	3.19	3.66	4.0	0.56	1.32	2.52	2.96
nuclear charge	1	6	7	8	9	11	12	15	16
outer electrons	1	4	5	6	7	1	2	5	6

Table 7.1: Electronegativity scale [180, 197] of principal atoms in biology. The 'outer electrons' row lists the number of electrons needed to complete the outer shell.

pedagogical reasons, although we would not recommend using it in general.

In our first measure of wrapping, we define precisely two classes of residues relevant to wrapping. This avoids potential confusion caused by using taxonomies of residues based on standard concepts. In Section 7.2.2, we show that this definition is sufficient to give some insight into protein aggregation and make predictions about protein behaviors.

However, it is also possible to provide a more refined measure that looks below the level of the residue abstraction and instead counts all non-polar groups, independent of what type of sidechain they inhabit. We present this more detailed approach in Section 7.3. We will show in Section 8.1 that there is a measurable force associated with an UWHB that can be identified by the second definition. Later we will define this force rigorously and use that as part of the definition of dehydron in Section 7.5. In Section 7.5, we will review a more sophisticated technique that incorporates the geometry of nonpolar groups as well as their number to assess the extent of protection via dielectric modulation.

7.1 Assessing polarity

The key to understanding hydrophobicity is polarity. Nonpolar groups repel water molecules (or at least do not attract them strongly) and polar groups attract them. We have already discussed the concept of polarity, e.g., in the case of dipoles (Section 3.2). Similarly, we have noted that certain sidechains, such as glutamine, are polar, even though there is no apparent charge difference in relevant molecules. Here we explain how such polarity can arise due to more subtle differences in charge distribution.

7.1.1 Electronegativity scale

The key to understanding the polarity of certain molecules is the **electronegativity scale** [180, 197], part of which is reproduced in Table 7.1. Atoms with similar electronegativity tend to form nonpolar groups, such as CH_n and $C-S$. Atomic pairs with differences in electronegativity tend to form polar groups, such as $C-O$ and $N-H$. The scaling of the electronegativity values is arbitrary, and the value for fluorine has been taken to be exactly four.

Let us show how the electronegativity scale can be used to predict polarity. In a C-O group, the O is more electronegative, so it will pull charge from C, yielding a pair with a negative charge associated with the O side of the group, and a positive charge associated with the C side of the pair.

Similarly, in an N-H group, the N is more electronegative, so it pulls charge from the H, leaving a net negative charge near the N and a net positive charge near the H. In Section 7.1.2, we will see that molecular dynamics codes assign such partial charges. The electronegativity difference for C-O is 0.91, and for N-H it is 0.6. Thus, it would be expected to find larger partial charges for C-O than for N-H, as we will see. Of course, the net charge for both C-O and N-H must be zero.

Only the differences in electronegativity have any chemical significance. But these differences can be used to predict the polarity of atomic groups, as we now illustrate for the carbonyl and amide groups. For any atom X, let $\mathcal{E}(X)$ denote the electronegativity of X. Since $\mathcal{E}(O) > \mathcal{E}(C)$, we conclude that the dipole of the carbonyl group $C - O$ can be represented by a positive charge on the carbon and a negative charge on the oxygen. Similarly, because $\mathcal{E}(N) > \mathcal{E}(H)$, the dipole of the amide group $N - H$ can be represented by a positive charge on the hydrogen and a negative charge on the nitrogen. A more detailed comparison of the electronegativities of C, O, N, and H gives

$$\mathcal{E}(O) - \mathcal{E}(C) = 3.66 - 2.75 = 0.91 > 0.60 = 3.19 - 2.59 = \mathcal{E}(N) - \mathcal{E}(H). \tag{7.1}$$

Thus we conclude that the charge difference in the dipole representation of the carbonyl group $(C - O)$ is larger than the charge difference in the dipole representation of the amide $(N - H)$ group.

It is beyond our scope to explain electronegativity here, but there is a simple way to comprehend the data. Electronegativity represents the power of an atom to attract electrons in a covalent bond [180]. Thus a stronger positive charge in the nucleus would lead to a stronger attraction of electrons, which is reflected in the correlation between nuclear charge and electronegativity shown in Table 7.1. More precisely, there is a nearly linear relationship between the electronegativity scale and the number of electrons in the outer shell. The value for hydrogen can be explained by realizing that the outer shell is half full, as it is for carbon.

The atoms with a complete outer shell (helium, neon, argon, etc.) are not part of the electronegativity scale, since they have no room to put electrons that might be attracted to them. Similarly, atoms with just a few electrons in the outer shell seem to be more likely to donate electrons than acquire them, so their electronegativity is quite small, such as sodium and magnesium. Hydrogen and carbon are in the middle of the scale, not surprisingly, since they are halfway from being full and empty of electrons.

7.1.2 Polarity of groups

Using the electronegativity scale, we can now estimate the polarity of groups of atoms. For example, the near match of electronegativity of carbon and hydrogen leads to the correct conclusion that the carbonaceous groups CH_n, $n = 1, 2, 3$ are not polar, at least in appropriate contexts. The typically symmetric arrangement of hydrogens also decreases the polarity of a carbonaceous group, at least when the remaining $4-n$ atoms bonded to it are other carbons or atoms of similar electronegativity.

If a carbon is not covalently attached exclusively to carbon or hydrogen then it is likely polarized and carries a partial charge. Thus, C_α carbons in the peptide bonds of all residues are polar. Sidechain carbons are polar if they are covalently attached to heteroatoms such as N or O. Sulfur (S) is a closer electronegative match with carbon and polarizes carbon to a lesser extent.

Full name of amino acid	three letter	single letter	The various PDB codes for the nonpolar carbonaceous groups
Alanine	Ala	A	CB
Arginine	Arg	R	CB, CG
Asparagine	Asn	N	CB
Aspartate	Asp	D	CB
Cysteine	Cys	C	CB
Glutamine	Gln	Q	CB, CG
Glutamate	Glu	E	CB, CG
Glycine	Gly	G	NA
Histidine	His	H	CB
Isoleucine	Ile	I	CB1, CB2, CG, CD1
Leucine	Leu	L	CB, CG, CD1, CD2
Lysine	Lys	K	CB, CG, CD
Methionine	Met	M	CB (CG, CE)
Phenylalanine	Phe	F	CB, CG, CD1, CD2, CE1, CE2, CZ
Proline	Pro	P	CB, CG
Serine	Ser	S	NA
Threonine	Thr	T	CG2
Tryptophan	Trp	W	CB, CG, CD2, CE1, CE2, CZ3, CH2
Tyrosine	Tyr	Y	CB, CG, CD1, CD2, CE1, CE2
Valine	Val	V	CB, CG1, CG2

Table 7.2: PDB codes for nonpolar carbonaceous groups. The carbonaceous groups surrounding the sulfer in Met may be considered polar.

Full name of compound	PDB code	The various PDB codes for the nonpolar carbonaceous groups
pyroglutamic acid	PCA	CB, CG
phosphorylated tyrosine	PTR	CB, CG, CD1, CD2, CE1, CE2
staurosporine	STU	Ci, $i = 1, \ldots, 7$; $i = 11, \ldots, 16$; C24, C26

Table 7.3: Sample PDB codes and nonpolar carbonaceous groups for some nonstandard amino acids and other compounds.

Residues	atom type	PDB codes	charge
ASP (GLU)	C	CG (CD)	0.27
	OM	ODi (OEi) $i = 1, 2$	-0.635
ASN (GLN)	NT	ND2 (NE2)	-0.83
	H	HD2i (HE2i), $i = 1, 2$	0.415
	C	CG (CD)	0.38
	O	OD1 (OE1)	-0.38
CYS	S	SG	-0.064
	H	HG	0.064
THR	CH1	CB	0.15
	OA	OG1	-0.548
	H	HG1	0.398
SER	CH2	CB	0.15
	OA	OG	-0.548
	H	HG	0.398

Table 7.4: Partial charges from the Gromos force field for polar and negatively charged amino acids.

The case CH_n with $n = 0$ is not encountered in biology unless the carbon is attached to at least one heteroatom.

To illustrate the polarity of the atoms not listed in Table 7.2, we present the partial charges of the remaining atoms as utilized in the Gromos code in Table 7.4 and Table 7.5. In Table 13.1, partial charges for aromatic sidechains are listed.

In addition to the the charges shown for the individual sidechain atoms, the backbone is assigned partial charges as follows: the charges of the amide group are ± 0.28 and the carbonyl group are ± 0.38. That is, in the amide $(N - H)$ group, the N is given a partial charge of -0.28 and the H is given a partial charge of $+0.28$. Similarly, in the carbonyl $(C - O)$ group, the O is given a partial charge of -0.38 and the C is given a partial charge of $+0.38$. Note that the partial charges for $C - O$ are larger than the partial charges for $N - H$, in accord with our prediction using the electronegativity scale in (7.1).

The N-terminal and C-terminal groups also have appropriate modifications. The C-terminal oxygens have a charge of -0.635, and the attached carbon has a charge of 0.27. The N-terminal nitrogen has a charge of 0.129, and the attached three hydrogens have a charge of 0.248. All of the groups listed in Table 7.2 have zero partial charge.

7.2 Counting residues

In [69], the definition of 'well-wrapped' was based on the proximity of certain residues and defined in relation to the observed distribution of rapping among a large sample set of proteins. The extent of hydrogen-bond desolvation was defined by the number of residues ρ_R with at least two *nonpolar* carbonaceous groups (CH_n, $n = 1, 2, 3$) whose β-carbon is contained in a specific desolvation domain,

Residue	atom type	PDB codes	charge
ARG	CH2	CD	0.09
	NE	NE	-0.11
	C	CZ	0.34
	NZ	NHi, $i = 1, 2$	-0.26
	H	HE, HHij, $i, j = 1, 2$	0.24
LYS	CH2	CE	0.127
	NL	NZ	0.129
	H	HZi, $i = 1, 3$	0.248
HIS (A/B)	C	CD2/CG	0.13
	NR	NE2/ND1	-0.58
	CR1	CE1	0.26
	H	HD1/HE2	0.19

Table 7.5: Partial charges from the Gromos force field for positively charged amino acids. The partial charges for His represent two possible ionized states which carry neutral charge.

as depicted in Figure 7.1. In Section 7.1.2, we explained how to determine the polarity of groups using the electronegativity scale. The nonpolar carbonaceous groups are listed in Table 7.2.

The C_α carbons in all residues are covalently bonded to a nitrogen atom. The mismatch in electronegativity between carbon and nitrogen (Table 7.1) implies that the C_α carbons are polar and thus do not contribute to repelling water. Sidechain carbons are counted only if they are not covalently attached to heteroatoms such as N or O. The CH groups in serine and threonine are attached to an oxygen, which renders them polar. However, the CH groups in methionine attached to a sulfur are not polar. Similarly, a lone carbon that is attached to oxygens is also polar. Thus the seven residues listed in Figure 4.4 are eliminated from the group of wrappers.

7.2.1 Desolvation domain

The desolvation domain was chosen in [69] to be the union of two (intersecting) 7Å-radius spheres centered at the C_α-carbons of the residues paired by the hydrogen bond, as shown in Figure 7.1. The choice of the C_α carbons as the centers of the desolvation spheres is justified in Figure 7.2. These figures show that the center of the line joining the centers of the desolvation spheres is often the center of the hydrogen bonds in typical secondary structures. In the case of a parallel β-sheet, the desolvation domain is the same for two parallel hydrogen bonds. The radius represents a typical cutoff distance to evaluate interactions between nearby residues. C_α-carbons which are neighboring in protein sequence are about 3.8Å apart (cf. Exercise 2.1). The distance between other C_α-carbons is easily determined by datamining in the PDB (cf. Exercise 2.2).

An amide-carbonyl hydrogen bond was defined in [69] by an N-O (heavy-atom) distance within the range 2.6–3.4Å (typical extreme bond lengths) and a 60-degree latitude in the N-H-O angle (cf. Section 5.4). As a scale of reference, at maximum density, water occupies a volume that

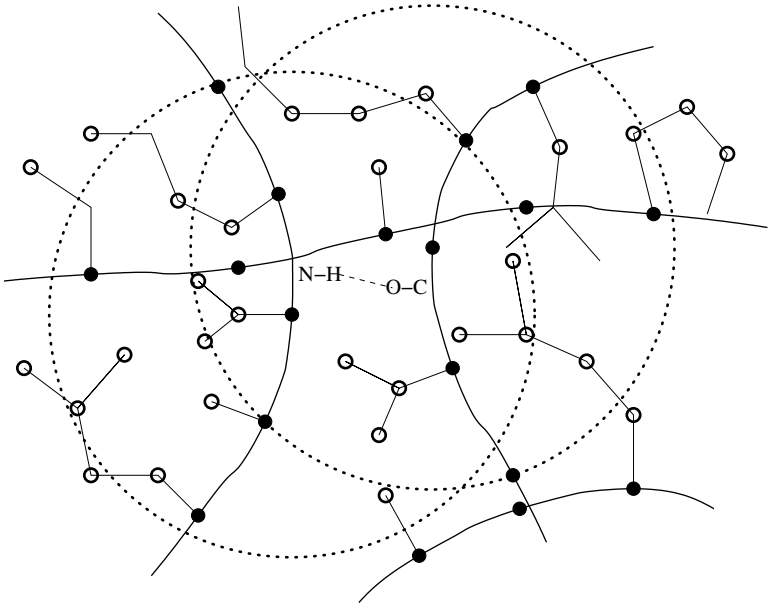

Figure 7.1: Caricature showing desolvation spheres with various side chains. The open circles denote the nonpolar carbonaceous groups, and the solid circles represent the C_α carbons. The hydrogen bond between the amide (N-H) and carbonyl (O-C) groups is shown with a dashed line. Glycines appear without anything attached to the C_α carbon. There are 22 nonpolar carbonaceous groups in the union of the desolvation spheres and six sidechains with two or more carbonaceous groups whose C_β carbon lie in the spheres.

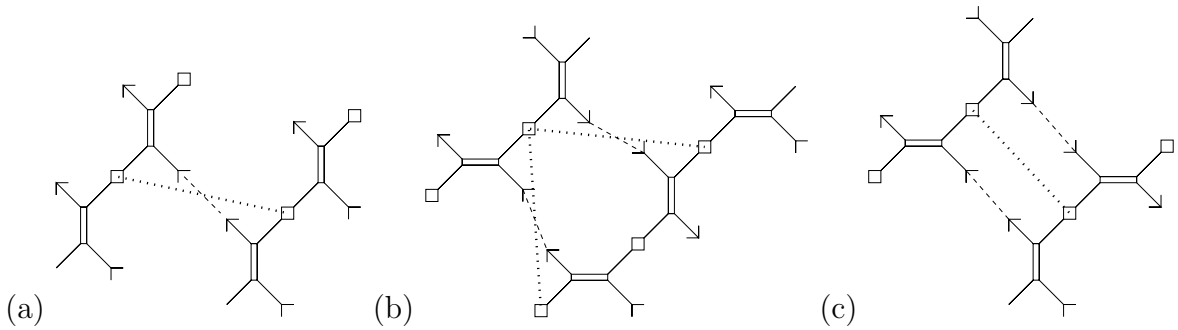

Figure 7.2: The hydrogen bond (dashed line) configuration in (a) α-helix, (b) antiparallel β-sheet, and (c) parallel β-sheet. A dotted line connects the C_α carbons (squares) that provide the centers of the spheres forming the desolvation domains in Figure 7.1. The amide (N-H) groups are depicted by arrow heads and the carbonyl (O-C) groups are depicted by arrow tails.

corresponds to a cube of dimension just over 3.1Å on a side (cf. Section 10.7).

The average extent of desolvation, ρ_R, over all backbone hydrogen bonds of a monomeric structure can be computed from any set of structures. In [69], a nonredundant sample of 2811 PDB-structures was examined. The average ρ_R over the entire sample set was found to be 6.6 [69]. For any given structure, the dispersion (standard deviation) σ from the mean value of ρ_R for that structure can be computed. The dispersion averaged over all sampled structures was found to be $\sigma = 1.46$ [69]. These statistics suggested a way to identify the extreme of the wrapping distribution as containing three or fewer wrapping residues in their desolvation domains. This can be interpreted as defining underwrapped as ρ_R values that are more than two standard deviations from the mean.

The distribution of the selected proteins as a function of their average wrapping as measured by ρ_R is shown in Fig. 5 in [69]. The probability distribution has a distinct inflection point at $\rho = 6.2$. Over 90% of the proteins studied have $\rho_R > 6.2$, and none of these are yet known to yield amyloid aggregation under physiological conditions. In addition, individual sites with low wrapping on selected proteins were examined and found to correlate with known binding sites.

In Section 7.2.2, we will see that the known disease-related amyloidogenic proteins are found in the relatively under-populated $3.5 < \rho_R < 6.2$ range of the distribution, with the cellular prion proteins located at the extreme of the spectrum ($3.5 < \rho_R < 3.75$). We discuss there the implications regarding a propensity for organized aggregation. Approximately 60% of the proteins in the critical region $3.5 < \rho_R < 6.2$ which are not known to be amyloidogenic are toxins whose structures are stabilized mostly by disulfide bonds.

To further assess the virtues of the residue-based assessment of wrapping, we review additional results and predictions of [69].

7.2.2 Predicting aggregation

Prediction of protein aggregation can be based on locating regions of the protein surface with high density of defects which may act as aggregation sites [104, 129, 156]. Figure 3a of [69] depicts the (many) UWHB's for the human cellular prion protein (PDB file 1QM0) [188, 192, 246]. Over half of the hydrogen bonds are UWHB's, indicating that many parts of the structure must be open to water attack. For example, α-helix 1 has the highest concentration of UWHB's, and therefore may be prone to structural rearrangement.

In helix 1 (residues 143 to 156), all of the hydrogen bonds are UWHB's, and this helix has been identified as undergoing an α-helix to β-strand transition [188, 192, 246]. Furthermore, helix 3 (residues 199 to 228) contains a significant concentration of UWHB's at the C-terminus, a region assumed to define the epitope for protein-X binding [188]. The remaining UWHB's occur at the helix-loop junctures and may contribute to flexibility required for rearrangement.

The average underwrapping of hydrogen bonds in an isolated protein may be a significant indicator of aggregation, but it is not likely to be sufficient to determine amyloidogenic propensity. For instance, protein L (PDB file 2PTL) is not known to aggregate even though its $\rho_R = 5.06$ value is outside the standard range of sufficient wrapping. Similarly, trp-repressor (PDB file 2WRP) has $\rho_R = 5.29$, and the factor for inversion stimulation (PDB file 3FIS) has $\rho_R = 4.96$. Many neurotoxins (e.g., PDB file 1CXO with $\rho_R = 3.96$) are in this range as well.

The existence of short fragments endowed with fibrillogenic potential [13, 48, 57, 93, 104, 156, 129] suggests a localization or concentration of amyloid-related structural defects. In view of this, a local wrapping parameter, the maximum density $\delta_{\max}$ of UWHB's on the protein surface was introduced [69]. A statistical analysis involving $\delta_{\max}$ [69] established that a threshold $\delta_{\max} > 0.38/\text{nm}^2$ distinguishes known disease-related amyloidogenic proteins from other proteins with a low extent of hydrogen bond wrapping. On the basis of a combined assessment, identifying both low average wrapping and high maximum density of underwrapping, it was predicted [69] that six proteins might posses amyloidogenic propensity. Three of them,

- angiogenin (cf. PDB files 1B1E and 2ANG),
- meizothrombin (cf. PDB file 1A0H), and
- plasminogen (cf. PDB file 1B2I),

are involved in some form of blood clotting or wound healing, and not something related to disease.

Not all protein aggregation is related to disease. Angiogenesis refers to the growth of new capillaries from an existing capillary network, and many processes involve this, including wound healing. Angiogenin is only one of many proteins involved in the angiogenesis process, but it appears to have certain unique properties [136]. Meizothrombin is formed during prothrombin activation, and is known to be involved in blood clotting [119] and is able to bind to procoagulant phospholipid membranes [182]. Plasminogen has been identified as being a significant factor in wound healing [195].

7.3 Counting nonpolar groups

A more refined measure of hydrogen-bond protection has been proposed based on the number of vicinal nonpolar groups [65, 71]. The desolvation domain for a backbone hydrogen bond is defined again as the union of two intersecting spheres centered at the α-carbons of the residues paired by the hydrogen bond, as depicted in Figure 7.1. In this case, all of the dark circles are counted, whether or not the base of the sidechain lies within the desolvation domain. The extent of intramolecular desolvation of a hydrogen bond, ρ_{PG}, is defined by the number of sidechain nonpolar groups (CH_n, $n = 1, 2, 3$) in the desolvation domain (see Table 7.2).

The distribution of wrapping for a large sample of non-redundant proteins is given in Figure 12.1 for a radius of 6Å for the definition of the desolvation domain. In [72], an UWHB was defined by the inequality $\rho_{PG} < 12$ for this value of the radius. Statistical inferences involving this definition of ρ_{PG} were found to be robust to variations in the range 6.4 ± 0.6 Å for the choice of desolvation radius [71, 79]. In Figure 7.3 the distribution of wrapping is presented for a particular PDB file.

7.3.1 Distribution of wrapping for an antibody complex

It is instructive to consider wrapping of hydrogen bonds from a more detailed statistical point of view. In Figure 7.3 the distribution of wrapping is presented for the antibody complex whose

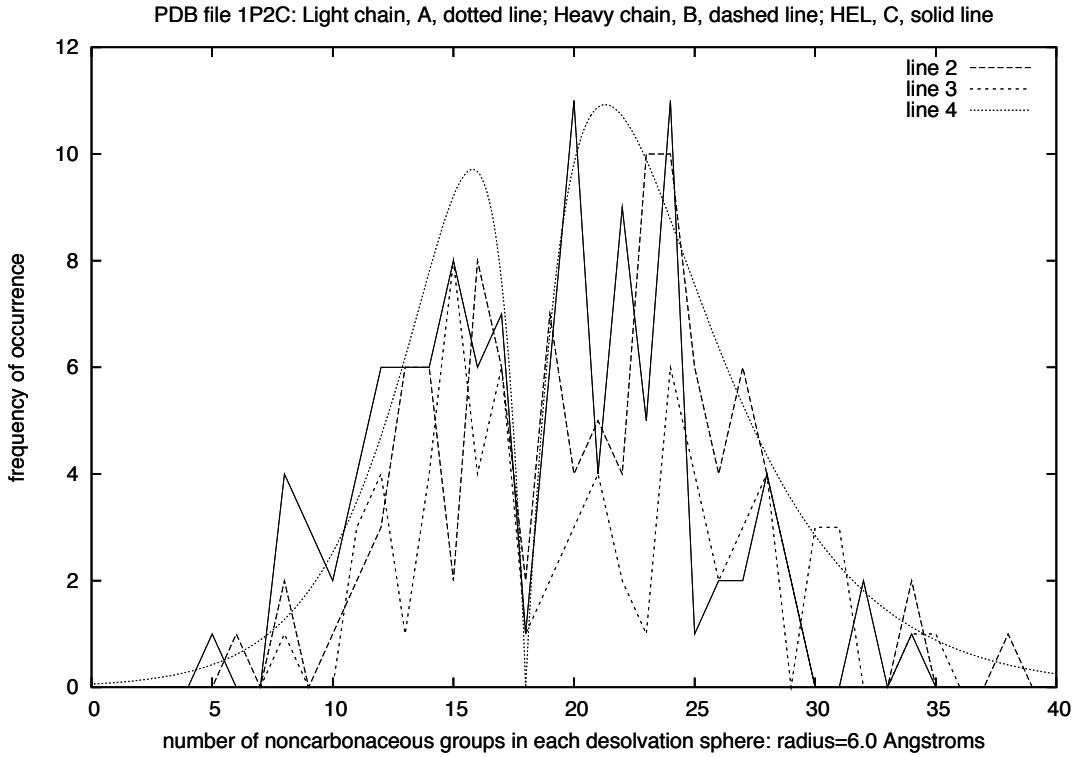

Figure 7.3: Distribution of wrapping for PDB file 1P2C. There are three chains: light, heavy chains of the antibody, and the antigen (HEL) chain. The desolvation radius is 6.0Å. Smooth curves (7.2) are added as a guide to the eye.

structure is recorded PDB file 1P2C. There are three chains, two in the antibody (the light and heavy chains), and one in the antigen, hen egg-white lysozyme (HEL).

What is striking about the distributions is that they are bi-modal, and roughly comparable for all three chains. We have added a smooth curve representing the distributions

$$d_i(r) = a_i|r - r_0|e^{-|r-r_0|/w_i} \tag{7.2}$$

to interpolate the actual distributions. More precisely, d_1 represents the distribution for $r < r_0$, and d_2 represents the distribution for $r > r_0$. The coefficients chosen were $w_1 = 2.2$ and $w_2 = 3.3$. The amplitude coefficients were $a_1 = 12$ and $a_2 = 9$, and the offset $r_0 = 18$ for both distributions. In this example, there seems to be a line of demarcation at $\rho = 18$ between hydrogen bonds that are well wrapped and those that are underwrapped.

The distributions in Figure 7.3 were computed with a desolvation radius of 6.0Å. Larger desolvation radii were also used, and the distributions are qualitatively similar. However the sharp gap at $\rho = 18$ becomes blurred for larger values of the desolvation radius.

7.4 Residues versus polar groups

The two measures considered here for determining UWHB's share some important key features. Both count sidechain indicators which fall inside of desolvation domains that are centered at the C_α backbone carbons. The residue-based method counts the number of residues (of a restricted type) whose C_β carbons fall inside the desolvation domain. The group-based method counts the number of carbonaceous groups that are found inside the desolvation domain.

We observed that the average measure of wrapping based on counting residues was $\rho_R = 6.6$, whereas the average measure of wrapping based on counting non-polar groups is $\rho_{PG} = 15.9$. The residues in the former count represent at least two non-polar groups, so we would expect that $\rho_{PG} > 2\rho_R$. We see that this holds, and that the excess corresponds to the fact that some residues have three or more non-polar groups. Note that these averages were obtained with different desolvation radii, 6.0Å for ρ_{PG} and 7.0Å for ρ_R. Adjusting for this difference would make ρ_{PG} even larger, indicating an even greater discrepancy between the two measures. This implies that ρ_{PG} provides a much finer estimation of local hydrophobicity.

The structural analysis in [69] identified site mutations which might stabilize the part of the cellular prion protein (PDB file 1QM0) believed to nucleate the cellular-to-scrapie transition. The (Met134, Asn159)-hydrogen bond has a residue wrapping factor of only $\rho_R = 3$ and is only protected by Val161 and Arg136 locally, which contribute only a minimal number (five) of non-polar carbonaceous groups. Therefore it is very sensitive to mutations that alter the large-scale context preventing water attack. It was postulated in [69] that a factor that triggers the prion disease is the stabilization of the (Met134, Asn159) β-sheet hydrogen bond by mutations that foster its desolvation beyond wild-type levels.

In the wild type, the only nonadjacent residue in the desolvation domain of hydrogen bond (Met134, Asn159) is Val210, thus conferring marginal stability with $\rho_R = 3$. Two of the three known pathogenic mutations (Val210Ile and Gln217Val) would increase the number of non-polar carbonaceous groups wrapping the hydrogen bond (Met134, Asn159), even though the number of wrapping residues would not change. Thus we see a clearer distinction in the wrapping environment based on counting non-polar carbonaceous groups instead of just residues.

The third known pathogenic mutation, Thr183Ala, may also improve the wrapping of the hydrogen bond (Met134, Asn159) even though our simple counting method will not show this, as both Thr and Ala contribute only one nonpolar carbonaceous group for desolvation. However, Ala is four positions below Thr in Table 6.1 and is less polar than Thr. Table 6.1 reflects a more refined notion of wrapping for different sidechains, but we do not pursue this here.

7.5 Defining dehydrons via geometric requirements

The enhancement of backbone hydrogen-bond strength and stability depends on the partial structuring, immobilization or removal of surrounding water. In this section we review an attempt [73] to quantify this effect using a continuous representation of the local solvent environment surrounding backbone hydrogen bonds [31, 65, 71, 79, 103, 173, 230]. The aim is to estimate the changes in the permittivity (or dielectric coefficient) of such environments and the sensitivity of

the Coulomb energy to local environmental perturbations caused by protein interactions [65, 79]. However, induced-fit distortions of monomeric structures are beyond the scope of these techniques.

The new ingredient is a sensitivity parameter M_k assessing the net decrease in the Coulomb energy contribution of the k-th hydrogen bond which would result from an exogenous immobilization, structuring or removal of water due to the approach by a hydrophobic group. This perturbation causes a net decrease in the permittivity of the surrounding environment which becomes more or less pronounced, depending on the pre-existing configuration of surrounding hydrophobes in the monomeric state of the protein. In general, nearby hydrophobic groups induce a structuring of the solvent needed to create a cavity around them and the net effect of this structuring is a localized reduction in the solvent polarizability with respect to reference bulk levels. This structuring of the solvent environment should be reflected in a decrease of the local dielectric coefficient ϵ. This effect has been quantified in recent work which delineated the role of hydrophobic clustering in the enhancement of dielectric-dependent intramolecular interactions [65, 79].

We now describe an attempt to estimate ϵ as a function of the fixed positions $\{r_j \,:\, j = 1, \ldots, n_k\}$ of surrounding nonpolar hydrophobic groups (CH_n, with $n = 1, 2, 3$, listed in Table 7.2). The simpler estimates of wrapping considered so far could fail to predict an adhesive site when it is produced by an uneven distribution of desolvators around a hydrogen bond, rather than an insufficient number of such desolvators. Based on the fixed atomic framework for the monomeric structure, we now identify Coulomb energy contributions from intramolecular hydrogen bonds that are most sensitive to local environmental perturbations by subsuming the effect of the perturbations as changes in ϵ.

Suppose that the carbonyl oxygen atom is at $\mathbf{r}_O$ and that the partner hydrogen net charge is at $\mathbf{r}_H$. The electrostatic energy contribution $E_{\text{COUL}}(k, \mathbf{r})$ for this hydrogen bond in a dielectric medium with dielectric permittivity $\epsilon(\mathbf{r})$ is approximated (see Chapter 16) by

$$E_{\text{COUL}}(\mathbf{r}) = \frac{-1}{4\pi\epsilon(\mathbf{r})} \frac{qq'}{|\mathbf{r}_O - \mathbf{r}_H|}, \tag{7.3}$$

where q, q' are the net charges involved and where $|\cdot|$ denotes the Euclidean norm.

Now suppose that some agent enters in a way to alter the dielectric field, e.g., a hydrophobe that moves toward the hydrogen bond and disrupts the water that forms the dielectric material. This movement will alter the Coulombic energy as it modifies ϵ, and we can use equation (7.3) to determine an equation for the change in ϵ in terms of the change in E_{COUL}. Such a change in E_{COUL} can be interpreted as a force (cf. Chapter 3). We can compute the resulting effect as a derivative with respect to the position R of the hydrophobe:

$$\nabla_R(1/\epsilon(\mathbf{r})) = \frac{4\pi|\mathbf{r}_O - \mathbf{r}_H|}{qq'} \left(-\nabla_R E_{\text{COUL}}(\mathbf{r})\right) = \frac{4\pi|\mathbf{r}_O - \mathbf{r}_H|}{qq'} F(\mathbf{r}), \tag{7.4}$$

where $F(\mathbf{r}) = -\nabla_R E_{\text{COUL}}(\mathbf{r})$ is a net force exerted on the hydrophobe by the fixed pre-formed hydrogen bond. This force represents a net 3-body effect [65], involving the bond, the dielectric material (water) and the hydrophobe. If E_{COUL} is decreased in this process, the hydrophobe is attracted to the hydrogen bond because in so doing, it decreases the value of $E_{\text{COUL}}(\mathbf{r})$.

To identify the 'opportune spots' for water exclusion on the surface of native structures we need to first cast the problem within the continuous approach, taking into account that $1/\epsilon$ is the factor

in the electrostatic energy that subsumes the influence of the environment. Thus to identify the dehydrons, we need to determine for which Coulombic contributions the exclusion or structuring of surrounding water due to the proximity of a hydrophobic 'test' group produces the most dramatic increase in $1/\epsilon$. The quantity M_k was introduced [73] to quantify the sensitivity of the Coulombic energy for the k-th backbone hydrogen bond to variations in the dielectric. For the k-th backbone hydrogen bond, this sensitivity is quantified as follows.

Define a desolvation domain D_k with border ∂D_k circumscribing the local environment around the k-th backbone hydrogen bond, as depicted in Figure 7.1. In [73], a radius of 7Å was used. The set of vector positions of the n_k hydrophobic groups surrounding the hydrogen bond is extended from $\{\mathbf{r}_j \,:\, j = 1, 2, \ldots, n_k\}$ to $\{\mathbf{r}_j \,:\, j = 1, 2, \ldots, n_k; R\}$ by adding the test hydrophobe at position R. Now compute the gradient $\nabla_R(1/\epsilon)|_{R=R_o}$, taken with respect to a perpendicular approach by the test hydrophobe to the center of the hydrogen bond at the point $R = R_o$ located on the circle consisting of the intersection C of the plane perpendicular to the hydrogen bond with the boundary ∂D_k of the desolvation domain. Finally, determine the number

$$M_k = \max\left\{|\nabla_R(1/\epsilon)|_{R=R_o}| \;:\; R_o \in C\right\}. \tag{7.5}$$

The number M_k quantifies the maximum alteration in the local permittivity due to the approach of the test hydrophobe in the plane perpendicular to the hydrogen bond, centered in the middle of the bond, at the surface of the desolvation domain.

The quantity M_k may be interpreted in physical terms as a measure of the maximum possible attractive force exerted on the test hydrophobic group by the pre-formed hydrogen bond. The only difficulty in estimating M_k is that it requires a suitable model of the dielectric permittivity ϵ as a function of the geometry of surrounding hydrophobic groups. We will consider the behavior of the dielectric permittivity more carefully in Chapter 16, but for now we consider a heuristic model used in [73].

The model in [73] for the dielectric may be written

$$\epsilon^{-1} = (\epsilon_o^{-1} - \epsilon_w^{-1})\Omega(\{\mathbf{r}_j\})\Phi(\mathbf{r}_H - \mathbf{r}_O) + \epsilon_w^{-1}, \tag{7.6}$$

where ϵ_w and ϵ_o are the permittivity coefficients of bulk water and vacuum, respectively, and

$$\Omega(\{\mathbf{r}_j\}) = \prod_{j=1,\ldots,n_k} \left(1 + e^{-|\mathbf{r}_O - \mathbf{r}_j|/\Lambda}\right)\left(1 + e^{-|\mathbf{r}_H - \mathbf{r}_j|/\Lambda}\right) \tag{7.7}$$

provides an estimate of the change in permittivity due to the hydrophobic effects of the carbonaceous groups. In [73], a value of $\Lambda = 1.8$Å was chosen to represent the characteristic length associated with the water-structuring effect induced by the solvent organization around the hydrophobic groups. Further, a cut-off function

$$\Phi(\mathbf{r}) = (1 + |\mathbf{r}|/\xi)\, e^{-|\mathbf{r}|/\xi}, \tag{7.8}$$

where $\xi = 5$Å is a water dipole-dipole correlation length, approximates the effect of hydrogen bond length on its strength [73].

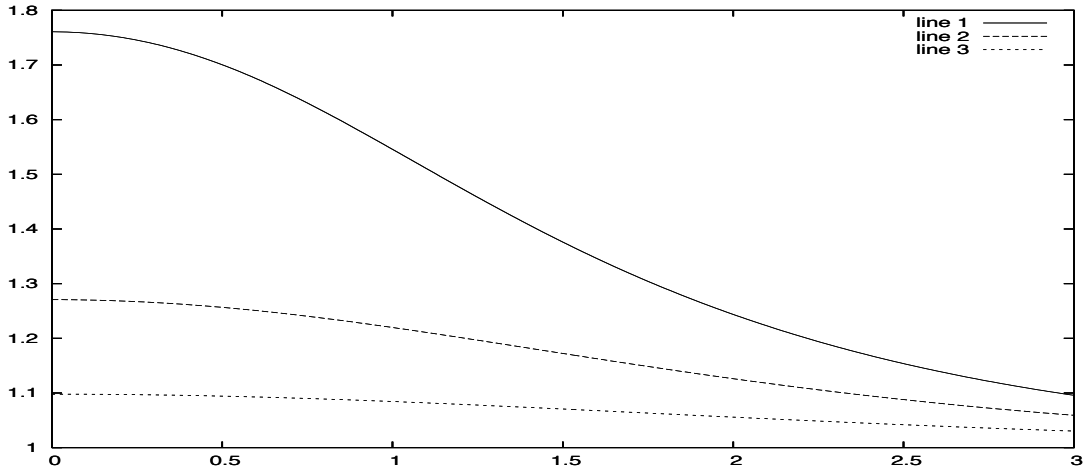

Figure 7.4: The function $\omega(x, y)$ plotted as a function of the distance along the x-axis connecting r_H and r_O, for three different values of the distance y from that axis: $y = 1$ (solid line), $y = 2$ (dashed line), $y = 3$ (dotted line). The coordinates have been scaled by Λ and the value of $|\mathbf{r}_O - \mathbf{r}_H| = 1$ was assumed.

We can write the key expression Ω in (7.7) as

$$\Omega(\{\mathbf{r}_j\}) = \prod_{j=1,\ldots,n_k} \omega(\mathbf{r}_j), \tag{7.9}$$

where the function ω is defined by

$$\omega(\mathbf{r}) = \left(1 + e^{-|\mathbf{r}_O - \mathbf{r}|/\Lambda}\right)\left(1 + e^{-|\mathbf{r}_H - \mathbf{r}|/\Lambda}\right). \tag{7.10}$$

The function ω is never smaller than one, and it is maximal in the plane perpendicular to the line connecting r_H and r_O. Moreover, it is cylindrically symmetric around this axis. The values of ω are plotted in Figure 7.4 as a function of the distance from the perpendicular bissector of the axis connecting $\mathbf{r}_H$ and $\mathbf{r}_O$, for three different values of the distance y from the line connecting $\mathbf{r}_H$ and $\mathbf{r}_O$.

We see that the deviation in ω provides a strong spatial dependence on the dielectric coefficient in this model. Thus hydrophobes close to the plane bissecting the line connecting r_H and r_O are counted more strongly than those away from that plane, for a given distance from the axis, and those closer to the line connecting r_H and r_O are counted more strongly than those further away. When the product $\Phi\Omega = 1$, we get $\epsilon = \epsilon_o$ reflecting the maximal amount of water exclusion possible. Correspondingly, if $\Phi\Omega = 0$, $\epsilon = \epsilon_w$ indicating a dielectric similar to bulk-water. Thus bigger values of Ω correspond to the effect of wrapping.

The definition (7.6) of the dielectric has not been scaled in a way that assures a limiting value of $\epsilon = \epsilon_o$. However, since we are only interested in comparing relative dielectric strength, this scaling is inessential. What matters is that larger values of Ω correspond to a lower dielectric and thus stronger bonds.

The computation of M_k involves computing the gradient of

$$\Omega(\{\mathbf{r}_1, \ldots, \mathbf{r}_{n_k}, R\}) = \omega(R)\Omega(\{\mathbf{r}_1, \ldots, \mathbf{r}_{n_k}\}) \tag{7.11}$$

with respect to R. Due to the cylindrical symmetry of ω, $|\nabla_R \omega|_{R=R_o}$ is a constant depending only on the desolvation radius $|R_o|$ and the hydrogen bond length $|\mathbf{r}_O - \mathbf{r}_H|$ for all $R_o \in C$. Thus, for a fixed desolvation radius $|R_o|$, M_k may be written as a function of $|\mathbf{r}_O - \mathbf{r}_H|$ times $\Omega(\{\mathbf{r}_1, \ldots, \mathbf{r}_{n_k}\})$ when using the model (7.6).

A sensitivity threshold for hydrogen bonds was established in [73] by statistical analysis on a sample of native structures for soluble proteins. Only 8% of backbone hydrogen bonds from a sample of 702 proteins, of moderate sizes ($52 < N < 110$) and free from sequence redundancies [102], were found to be highly sensitive in the sense that

$$M_k > \lambda/10, \tag{7.12}$$

where λ was defined to be

$$\lambda = \frac{\epsilon_o^{-1} - \epsilon_w^{-1}}{2\text{Å}}. \tag{7.13}$$

On the other hand, 91.6% of backbone hydrogen bonds were found to be relatively insensitive to water removal, namely,

$$0 < M_k < \lambda/100 \tag{7.14}$$

This remarkable separation in the (nearly bimodal) distribution of sensitivities led [73] to the definition of a dehydron as a backbone hydrogen bond satisfying (7.12).

7.6 Exercises

Exercise 7.1 *It was predicted [69] that the three proteins*

- *anti-oncogene A (PDB file 1A1U);*
- *RADR zinc finger peptide (PDB file 1A1K) and*
- *rubredoxin (PDB file 1B20).*

might have amyloidogenic tendencies. Investigate these three proteins to see why this might be the case.

Exercise 7.2 *In Section 4.4.3, we noted that sidechains have different conformations. Determine the number of different rotameric states possible for each sidechain (hint: read [146]). Compare the number of rotameric degrees of freedom for the seven residues listed in Figure 4.4 with the remaining group of thirteen sidechains.*

Exercise 7.3 *In Figure 7.2(a), it appears that the dotted line joining the two C_α cabons intersects the dashed line joining the amide and carbonyl groups. By searching the PDB, determine the distribution of distances between the midpoints of these two lines for α-helices.*

Exercise 7.4 *Explain whether you would expect methyl flouride to have a polar carbon, based on the electronegativity scale.*

Chapter 8

Stickiness of dehydrons

We have explained why under-wrapped hydrogen bonds benefit from the removal of water. This makes them susceptible to interaction with molecules that can replace water molecules in the vicinity of the hydrogen bond. Conceptually, this implies that under-wrapped hydrogen bonds attract entities that can dehydrate them. Thus they must be sticky. If so, it must be possible to observe this experimentally. Here we review several papers that substantiate this conclusion. One of them involves a mesoscopic measurement of the force associated with a dehydron [72]. A second presents data on the direct measurement of the dehydronic force using atomic force microscopy [63]. Another paper examines the effect of such a force on a deformable surface [74].

8.1 Surface adherence force

We defined the notion of an under-wrapped hydrogen bond by a simple counting method in Chapter 7 and have asserted that there is a force associated with UHWB's. Here we describe measurements of the adhesion of an under-wrapped hydrogen bond by analyzing the flow-rate dependence of the adsorption uptake of soluble proteins onto a phospholipid bilayer.

8.1.1 Biological surfaces

The principal biological surface of interest is the cell membrane. This is a complex system, but a key component is what is called a **phospholipid bilayer**. The term **lipid** refers to a type of molecule that is a long carbonaceous polymer with a polar (phospho) group at the 'head.' This it is hydrophobic at one end and hydrophilic at the other. These molecules align to form a complex that could be described as a bundle of pencils, with the hydrophilic head group (the eraser) at one side of the surface and the hydrophobic 'tail' on the other side. These bundles can grow to form a surface when enough pencils are added. A second surface can form in the opposite orientation, with the two hydrophobic surfaces in close proximity. This results in a membrane that is hydrophilic on both sides, and thus can persist in an aqueous environment.

One might wonder what holds together a lipid bilayer. We have noted that there is a significant volume change when a hydrophobic molecule gets removed from water contact in Section 4.4.4.

The volume change causes self-assembly of lipids and provides a substantial pressure that holds the surface together. The architecture of a lipid bilayer is extremely adaptive. For example, a curved surface can be formed simply by allocating more lipid to one side than the other. Moreover, it easily allows insertion of other molecules of complex shape but with other composition. Much of a cell membrane is lipid, but there are also proteins with various functions as well as other molecules such as cholesterol. However, a simple lipid bilayer provides a useful model biological surface.

8.1.2 Soluble proteins on a surface

One natural experiment to perform is to release soluble proteins in solution near a lipid bilayer and to see to what extent they attach to the bilayer. Such an experiment [66] indicated a significant correlation between the under-wrapping of hydrogen bonds and bilayer attachment. The results were explained by assuming that the probability of successful landing on the liquid-solid interface is proportional to the ratio of UWHB's to all hydrogen bonds on the protein surface. Here, the number of surface hydrogen bonds is taken simply as a measure of the surface area. Thus the ratio can be thought of as an estimate of the fraction of the surface of the protein that is under-wrapped. The experiments in [66] indicated that more dehydrons lead to more attachments, strongly suggesting that dehydrons are sticky. However, such indications were only qualitative.

A more refined analysis of lipid bilayer experiments was able to quantify a force of attachment [72]. The average magnitude of the attractive force exerted by an UWHB on a surface was assessed based on measuring the dependence of the adsorption uptake on the flow rate of the ambient fluid above the surface. The adhesive force was measured via the decrease in attachment as the flow rate was increased.

Six proteins were investigated in [72], as shown in Table 8.1, together with their numbers of well-wrapped hydrogen bonds as well as dehydrons. The UWHB's for three of these are shown in Fig. 1a-c in [72]. The particular surface was a Langmuir-Blodgett bilayer made of the lipid DLPC (1,2 dilauroyl-sn-glycero-3 phosphatidylcholine) [194]. We now review the model used in [72] to interpret the data.

8.2 A two-zone model

In [72], a two-zone model of surface adhesion was developed. The first zone deals with the experimental geometry and predicts the number of proteins that are likely to reach a fluid boundary layer close to the lipid bylayer. The probability Π of arrival is dependent on the particular experiment, so we only summarize the model results from [72]. The second zone is the fluid boundary layer close to the lipid bilayer, where binding can occur. In this layer, the probability P of binding is determined by the thermal oscillations of the molecules and the solvent as well as the energy of binding.

The number M of adsorbed molecules is given by

$$M = \Pi P(n_{UW}, n_W, T)N \tag{8.1}$$

where Π is the fraction of molecules that reach the (immobile) bottom layer of the fluid, $P(n_{UW}, n_W, T)$ is the conditional probability of a successful attachment at temperature T given that the bottom

layer has been reached, and N is the average number of protein molecules in solution in the cell. The quantities n_{UW} and n_W are the numbers of underwrapped and well-wrapped hydrogen bonds on the surface of the protein, respectively. These will be used to estimate the relative amount of protein surface area related to dehydrons. The fraction Π depends on details of the experimental design, so we focus initially on on the second term P.

8.2.1 Boundary zone model

Suppose that ΔU is the average decrease in Coulombic energy associated with the desolvation of a dehydron upon adhesion. It is the value of ΔU that we are seeking to determine. Let ΔV be the Coulombic energy decrease upon binding at any other site. Let f be the fraction of the surface covered by dehydrons. As a simplified approximation, we assume that

$$f \approx \frac{n_{UW}}{n_{UW} + n_W}. \tag{8.2}$$

Then the probability of attachment at a dehydron is predicted by thermodynamics as

$$P(n_{UW}, n_W, T) = \frac{f e^{\Delta U/k_B T}}{(1-f)e^{\Delta V/k_B T} + f e^{\Delta U/k_B T}} \approx \frac{n_{UW} e^{\Delta U/k_B T}}{n_W e^{\Delta V/k_B T} + n_{UW} e^{\Delta U/k_B T}}, \tag{8.3}$$

with k_B = Boltzmann's constant. In [72], ΔV was assumed to be zero. In this case, (8.3) simplifies to

$$P(n_{UW}, n_W, T) = \frac{f e^{\Delta U/k_B T}}{(1-f) + f e^{\Delta U/k_B T}} \approx \frac{n_{UW} e^{\Delta U/k_B T}}{n_W + n_{UW} e^{\Delta U/k_B T}} \tag{8.4}$$

(cf. equation (2) of [72]). Note that this probability is lower if $\Delta V > 0$.

8.2.2 Diffusion zone model

The probability Π in (8.1) of penetrating the bottom layer of the fluid is estimated in [72] by a model for diffusion via Brownian motion in the plane orthogonal to the flow direction. This depends on the solvent bulk viscosity μ, and the molecular mass, m, and the **hydrodynamic radius** [198] or **Stokes radius** [100] of the protein. This radius R associates with each protein an equivalent sphere that has approximately the same flow characteristics at low Reynolds numbers. This particular instance of a 'spherical cow' approximation [53, 130] is very accurate, since the variation in flow characteristics due to shape variation is quite small [198]. The drag on a sphere of radius R, at low Reynolds numbers, is $F = 6\pi R\mu v$ where v is the velocity. The drag is a force that acts on the sphere through a viscous interaction. The coefficient

$$\xi = 6\pi R\mu/m = F/mv \tag{8.5}$$

where m is the molecular mass, is a temporal frequency (units: inverse time) that characterizes Brownian motion of a protein. The main non-dimensional factor that appears in the model is

$$\alpha = \frac{m\xi^2 L^2}{2k_B T} = \frac{L^2(6\pi R\mu)^2/m}{2k_B T}, \tag{8.6}$$

which has units of energy in numerator and denominator. We have [2]

$$\begin{aligned}\Pi(v,R,m) &= \int_{\Lambda}\int_{\Omega\setminus\Lambda}\int_{[0,\tau]} \frac{\alpha L^{-2}}{\pi\Gamma(t)} e^{-\alpha L^{-2}|\mathbf{r}-\mathbf{r}_0|^2/\Gamma(t)}\, dt d\mathbf{r}_0 d\mathbf{r} \\ &= \int_{\tilde{\Lambda}}\int_{\tilde{\Omega}\setminus\tilde{\Lambda}}\int_{[0,L/v]} \frac{\alpha}{\pi\Gamma(t)} e^{-\alpha|\tilde{\mathbf{r}}-\tilde{\mathbf{r}}_0|^2/\Gamma(t)}\, dt d\tilde{\mathbf{r}}_0 d\tilde{\mathbf{r}}\end{aligned} \tag{8.7}$$

where $\mathbf{r}$ is the two-dimensional position vector representing the cell cross-section Ω, $|\mathbf{r}|$ denotes the Euclidean norm of $\mathbf{r}$, Λ is the 6Å×10^8Å cross-section of the bottom layer, and $\Gamma(t) = 2\xi t - 3 + 4e^{-\xi t} - e^{-2\xi t}$. The domains $\tilde{\Lambda}$ and $\tilde{\Omega}$ represents domains scaled by the length L, and thus the variables $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{r}}_0$ are non-dimensional. In particular, the length of $\tilde{\Lambda}$ and $\tilde{\Omega}$ is one in the horizontal coordinate. Note that $\Gamma(t) = \frac{2}{3}(\xi t)^3 + \mathcal{O}((\xi t)^4)$ for ξt small. Also, since the mass m of a protein tends to grow with the radius cubed, α actually decreases like $1/R$ as the Stokes radius increases.

8.2.3 Model validity

The validity of the model represented by equations (8.1—8.7) was established by data fitting. The only parameter in the model, ΔU, was varied, and a value was found that consistently fits within the confidence band for the adsorption data for the six proteins (see Fig. 3 of [72]) across the entire range of flow velocities v. This value is

$$\Delta U = 3.91 \pm 0.67\,\text{kJ/mole} = \Delta U = 0.934 \pm 0.16\,\text{kcal/mole}. \tag{8.8}$$

This value is within the range of energies associated with typical hydrogen bonds. Thus we can think of a dehydron as a hydrogen bond that gets turned 'on' by the removal of water due to the binding of a ligand.

Using the estimate (8.8) of the binding energy for a dehydron, an estimate was made [72] of the force

$$|F| = 7.78 \pm 1.5\text{pN} \tag{8.9}$$

exerted by the surface on a single protein molecule at a 6Å distance from the dehydron.

8.3 Direct force measurement

The experimental techniques reviewed in the previous section suggest that the density of dehydrons correlates with protein stickiness. However, the techniques are based on measuring the aggregate behavior of a large number of proteins. One might ask for more targeted experiments seeking to isolate the force of a dehydron, or at least a small group of dehydrons. Such experiments were reported in [63] based on atomic force microscopy (AFM).

We will not give the details of the experimental setup, but just describe the main points. The main concept was to attach hydrophobic groups to the tip of an atomic force microscope. These were then lowered onto a surface capable of forming arrays of dehydrons. This surface was formed by

protein name	PDB code	residues	WWHB	dehydrons
apolipoprotein A-I	1AV1	201	121	66
β lactoglobulin	1BEB	150	106	3
hen egg-white lysozyme	133L	130	34	13
human apomyoglobin	2HBC	146	34	3
monomeric human insulin	6INS	50	30	14
human β_2-microglobulin	1I4F	100	17	9

Table 8.1: Six proteins and their hydrogen bond distributions. WWHB=well-wrapped hydrogen bonds.

a self-assembling monolayer of the molecules SH-$(CH_2)_{11}$-OH. The OH "head" groups are capable of making OH-OH hydrogen bonds, but these will be exposed to solvent and not well protected.

The data obtained by lowering a hydrophobic probe on such a monolayer are complex to interpret. However, they become easier when they are compared with a similar monolayer not containing dehydrons. In [63], the molecule SH-$(CH_2)_{11}$-Cl was chosen.

The force-displacement curve provided by the AFM have similarities for both monolayers [63]. For large displacements, there is no force, and for very small displacements the force grows substantially as the tip is driven into the monolayer. However, in between, the characteristics are quite different.

For the OH-headed monolayer, as the displacement is decreased to the point where the hydrophobic group on the tip begins to interact with the monolayer, the force on the tip decreases, indicating a force of attraction. Near the same point of displacement, the force on the tip increases for the chlorine-headed monolayer. Thus we see the action of the dehydronic force in attracting the hydrophobes to the dehydron-rich OH-headed layer. On the other hand, there is a resistance at the similar displacement as the hydrophobic tip begins to dehydrate the chlorine-headed monolayer. Ultimately, the force of resistance reaches a maximum, and then the force actually decreases to a slightly negative (attractive) value as the monolayer becomes fully dehydrated. It is significant that the displacement for the force minimum is approximately the same for both monolayers, indicating that they both correspond to a fully dehydrated state.

The force-displacement curves when the tip is removed from the surface also provide important data on the dehydronic force. The force is negative for rather large displacements, indicating the delay due to the requirements of rehydration. Breaking the hydrophobic bond formed by the hydrophobic groups on the tip and the monolayer requires enough force to be accumulated to completely rehydrate the monlayer. This effect is similar to the force that is required to remove sticky tape, in which one must reintroduce air between the tape and the surface to which it was attached. For the chlorine-headed monolayer, there is little change in force as the displacement is increased by four Ångstroms from the point where the force is minimal. Once the threshold is reached then the force returns abruptly to zero, over a distance of about one Ångstrom. For the OH-headed monolayer, the threshold is delayed by another two Ångstroms, indicating the additional effect of the dehydronic force.

The estimation of the dehydronic force is complicated by the fact that one must estimate the number of dehydrons that will be dehydrated by the hydrophobic groups on the tip. But the geometry of AFM tips is well characterized, and the resulting estimate [63] of

$$|F| = 5.9 \pm 1.2\text{pN} \tag{8.10}$$

at a distance of 5Å is in close agreement with the estimate (8.9) of 7.78 ± 1.5pN at a distance of 6Å in [72]. Part of the discrepancy could be explained by the fact that in [72] no energy of binding was attributed to the attachment to areas of a protein lacking dehydrons. If there were such an energy decrease, due e.g. to the formation of intermolecular interactions, the estimate of the force obtained in [72] would be reduced.

8.4 Membrane morphology

Since dehydrons have an attractive force that causes them to bind to a membrane, then the equal and opposite force must pull on the membrane. Since membranes are flexible, then this will cause the membrane to deform.

The possibility of significant morphological effect of dehydrons on membranes was suggested by the diversity of morphologies [205] of the inner membranes of cellular or subcellular compartments containing soluble proteins [74]. These vary from simple bag-like membranes [56] (e.g., erythrocytes, a.k.a. red blood cells) to highly invaginated membranes [227] (e.g., mitochondrial inner membranes). This raises the question of what might be causing the difference in membrane structure [126, 138, 164, 229].

Some evidence [74] suggests that dehydrons might play a role: hemoglobin subunits (which comprise the bulk of erythrocyte contents) are generally well wrapped, whereas two mitochondrial proteins, cytochrome c and pyruvate dehydrogenase, are less well wrapped. The correlation between the wrapping difference and the morphology difference provided motivation to measure the effect experimentally [74].

8.4.1 Protein adsorption

Morphology induction was tested in fluid phospholipid (DLPC) bilayers (Section 8.1) coating an optical waveguide [74]. The density of bilayer invaginations was measured by a technology called evanescent field spectroscopy which allowed measurement of both the thickness and refractive index of the adlayer [191, 217]. DLPC was added as needed for membrane expansion, with the portion remaining attached to the waveguide serving as a nucleus for further bilayer formation. Stable invaginations in the lipid bilayer formed after 60-hour incubation at T=318K.

8.4.2 Density of invaginations

The density of invaginations correlates with the extent of wrapping, ρ, of the soluble protein structure (Fig. 1, 2a in [74]). Greater surface area increase corresponds with lack of wrapping of backbone

hydrogen bonds. The density of invaginations as a function of concentration (Figure 2b in [74]) shows that protein aggregation is a competing effect in the protection of solvent-exposed hydrogen bonds ([71, 65, 66, 60, 79]): for each protein there appears to be a concentration limit beyond which aggregation becomes more dominant.

8.5 Kinetic model of morphology

The kinetics of morphology development suggest a simple morphological instability similar to the development of moguls on a steep ski run. When proteins attach to the surface, there is a force that binds the protein to the surface. This force pulls upward on the surface (and downward on the protein) and will increase the curvature in proportion to the local density of proteins adsorbed on the surface [66]. The rate of change of curvature $\frac{\partial g}{\partial t}$ is an increasing function of the force f:

$$\frac{\partial g}{\partial t} = \phi(f) \tag{8.11}$$

for some increasing function ϕ. Note that

$$\phi(0) = 0; \tag{8.12}$$

if there is no force, there will be no change. The function ϕ represents a material property of the surface.

The probability p of further attachment increases as a function of the curvature at that point since there is more area for attachment where the curvature is higher. That is, $p(g)$ is also an increasing function.

Of course, attachment also reduces surface area, but we assume this effect is small initially. However, as attachment grows, this neglected term leads to a 'saturation' effect. There is a point at which further reduction of surface area becomes the dominating effect, quenching further growth in curvature. But for the moment, we want to capture the initial growth of curvature in a simple model. We leave as Exercise 8.2 the development of a more complete model.

Assuming equilibrium is attained rapidly, we can assert that the force f is proportional to $p(g)$: $f = cp(g)$ at least up to some saturation limit, which we discuss subsequently. If we wish to be conservative, we can assert only that

$$f = \psi(p(g)) \tag{8.13}$$

with ψ increasing. In any case, we conclude that f may be regarded as an increasing function of the curvature g, say

$$f = F(g) := \phi(\psi(p(g))). \tag{8.14}$$

To normalize forces, we should have no force for a flat surface. That is, we should assume that $p(0) = 0$. This implies, together with the condition $\phi(0) = 0$, that

$$F(0) = 0. \tag{8.15}$$

The greater attachment that occurs locally causes the force to be higher there and thus the curvature to increase even more, creating an exponential runaway (Fig. 4 in [74]). The repeated interactions of these two reinforcing effects causes the curvature to increase in an autocatalytic manner until some other process forces it to stabilize.

The description above can be captured in a semiempirical differential equation for the curvature g at a fixed point on the bilayer. It takes the form

$$\frac{\partial g}{\partial t} = F(g), \tag{8.16}$$

where F is the function in (8.14) that quantifies the relationships between curvature, probability of attachment and local density of protein described in the previous paragraph. Abstractly, we know that F is increasing because it is the composition of increasing functions. Hence F has a positive slope s at $g = 0$. Moreover, it is plausible that $F(0) = 0$ using our assumptions made previously.

Thus the curvature should grow exponentially at first with rate s. In the initial stages of interface development, F may be linearly approximated by virtue of the mean value theorem, yielding the autocatalytic equation:

$$\frac{\partial g}{\partial t} = sg. \tag{8.17}$$

Figure 4 in [74] indicates that the number of invaginations appears to grow exponentially at first, and then saturates.

We have observed that there is a maximum amount of protein that can be utilized to cause morphology (Figure 2b in [74]) beyond which aggregation becomes a significantly competitive process. Thus, a 'crowding problem' at the surface causes the curvature to stop increasing once the number of adsorbed proteins gets too high at a location of high curvature.

8.6 Exercises

Exercise 8.1 *Determine the minimal distance between a hydrophobe and a backbone hydrogen bond in protein structures. That is, determine the number of wrappers as a function of the desolvation radius, and determine when, on average, this tends to zero.*

Exercise 8.2 *Derive a more refined model of morphological instability accounting for the reduction of surface area upon binding. Give properties of a function F as in (8.14) that incorporate the effect of decreasing surface area, and show how it would lead to a model like (8.16) which would saturate (rather than grow exponentially forever), reflecting the crowding effect of the molecules on the lipid surface.*