

1 Determinants of protein-protein interfaces

At the simplest level, one would expect the sort of bonds that help proteins form their basic structure would also be involved in joining two different proteins together.

Both hydrogen bonds and salt bridges play a significant role at protein interfaces [29].

The density of hydrogen bonds between two different proteins at an interface is about one per two square nanometers.

If you think of a checkerboard with nanometer sized squares, then it is like having one hydrogen bond on each of the red squares.

The average number of hydrogen bonds per interface is about ten.

On the other hand, the average number of salt bridges per interface is only two.

2 Protein-protein interface story

It might be that the story of protein-protein interactions ends here, with the intermolecular hydrogen bonds and salt bridges being the whole story.

However, three of the 54 high-resolution structures studied in [29] have no hydrogen bonds or salt bridges, and another dozen have no salt bridges and five or fewer hydrogen bonds.

Thus, we will delve deeper to see what other factors can determine protein-protein interactions.

One factor that complicates the picture of protein-protein interactions is the appearance of water molecules which appear to play a structural role, as opposed to simply mediating interactions via dielectric effects.

In the protein interfaces studied in [29], polar atom pairs bridged by water across the interface with hydrogen bonds were more numerous than direct hydrogen bond pairs, with each water molecule connecting 3.8 cross-chain atom pairs on average.

3 Amino acids at protein-protein interfaces

We begin with a simple use of datamining applied to the understanding of amino acid tendencies at interfaces.

There are different questions that one can ask, and of course it is natural that amino acids get ranked in different orders accordingly.

For simplicity, we contrast just two, but we also review others in Section 8.

The data here is drawn primarily from [12, 10, 4].

The site specificity of protein-protein interactions has been widely studied due to its central biological significance [15, 17, 16, 8, 14, 12].

One finds that hydrophobic residues such as Leu and Val are more abundant at protein-ligand interfaces, but it is also true that such residues are abundant over-all (see Table 3).

As a result, the removal of water surrounding hydrophobic residues on the protein surface has been assumed to be a driving force for association [27, 9].

3.1 Amino acid composition of protein–protein interfaces

The first question [10] we consider is about the amino acid composition of protein–protein interfaces.

This can be done by simply counting, once an identification has been made regarding which amino acids are at an interface.

But simple frequencies are misleading: Leu is the most common residue at interfaces, but it is also overwhelmingly the most common residue in most proteins.

Thus one has to normalize by the natural frequencies of amino acids in proteins [4].

3.2 Composition of amino acid pairs at interfaces

The second question [12] is about the amino acid composition for pairs of amino acids at interfaces *that are interacting*.

There are many ways to define interaction, but proximity [12] is a natural metric.

That is, two residues are defined [12] as interacting if their C_β coordinates differ by at most 6Å (with a similar scheme to include Gly).

This notion is simplistic in that the C_β atom is only the first in the sequence, but it is notable that the same sort of simple measure based on the initial segment is successful in other contexts [19].

3.3 Compare and contrast conflicting coompositions

Let us compare and contrast the two questions.

The first question seeks to determine clues for protein-protein association by investigating all residues, suitably normalized.

The second question assumes that proximity of sidechain pairs is a significant factor in protein-protein association, and thus looks for consequences of restricting to such pairs.

Not surprisingly, each question returns different answers regarding the relative significance of different residues.

3.4 Some caveats

The distribution of amino acid composition in proteins displays evolutionary trends [4], and this can require extra care to reveal subtle relationships.

Here we limit our investigations to fairly strong trends for simplicity.

However, the precise numerical data presented would differ if different databases were chosen for the primary data being used.

4 Interface propensity

The common belief is that hydrophobic residues on the surface of proteins are likely candidates to support interfaces in protein-protein association.

In Section 6, we present evidence that supports this case with suitable clarifications.

However, [10] presents data with a distinctively different conclusion, by normalizing the relative abundance of residues at the interface by their over-all abundances.

The residues with the highest relative propensity [10] to be at interfaces are, in decreasing order of frequency, Asn, Thr, Gly, Ser, Asp, Ala, and Cys, the group depicted in Figure 1.

None of these residues is distinctively hydrophobic.

This is quite a surprising result, and it demands an explanation.

4.1 Sidechains involved in interfaces

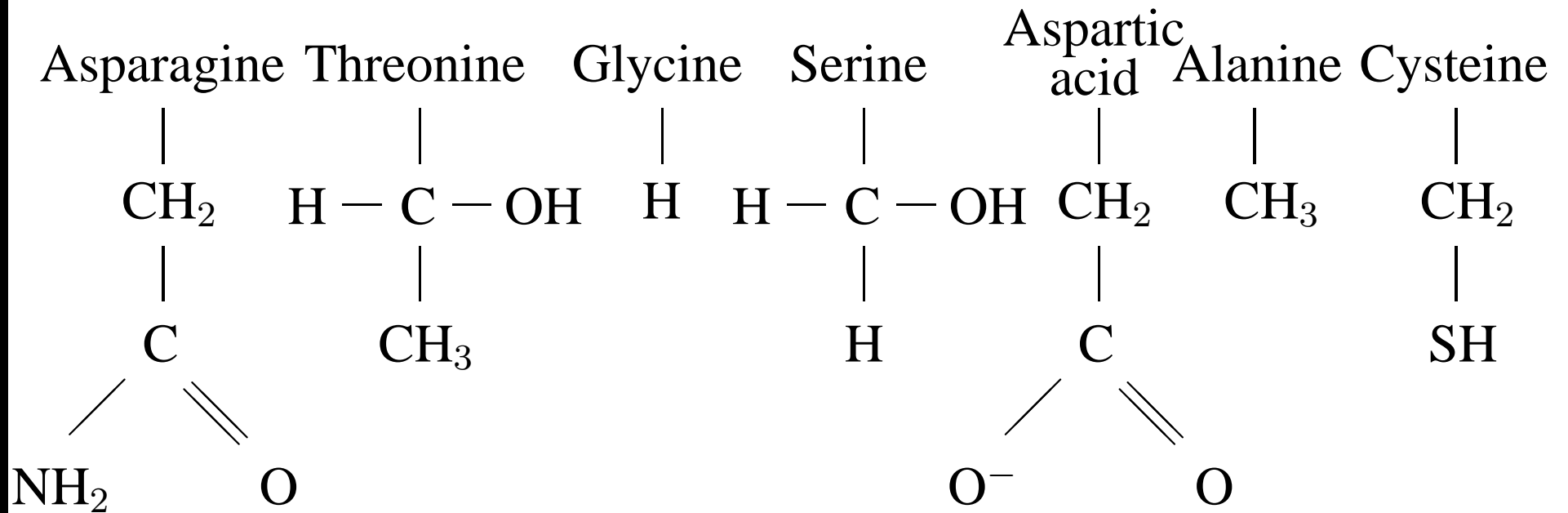


Figure 1: Periodic table of amino acid sidechains. The most likely to be involved in interactions, ordered from the left (asparagine).

4.2 Basic statistical definitions

To begin with, let us clarify the basic notions.

If we have a dataset with N different types of characteristics (e.g., $N = 20$ and the characteristics are the different amino acids), then the **frequency** f_i of the i -th characteristic is defined by

$$f_i = \frac{o_i}{\sum_{j=1}^N o_j} \quad (4.1)$$

where o_j is the number of occurrences of the j -th characteristic in the dataset.

In some cases, frequencies are represented as percentages, in which case we simply multiply by 100 in (4.1).

4.3 Basic statistical definitions continued

If we have two datasets with the same characteristics, with frequencies f_i and g_i , respectively, then one can define a **relative frequency**

$$r_i = f_i/g_i \quad (4.2)$$

of the characteristics between the two datasets.

There are some problems with this measure of occurrence.

First of all, it might happen that $g_k = 0$ for some k , making the interpretation difficult.

Related to this is the need for normalization in order to be able to compare two different comparisons.

In [10], the following approach was taken.

Define a normalized **relative propensity** via

$$R_i = \frac{r_i}{\sum_{j=1}^N r_j}. \quad (4.3)$$

These relative propensities sum to one, so we can think of them like ordinary frequencies.

Similarly, we multiply by 100 in (4.1) to convert to percentages as the unit of “frequency.”

If we apply this approach to datasets of proteins, and the characteristics are the different amino acid constituents, then we obtain the scheme used in [10].

In this case, the sum of the relative propensities (in percentage units) is one hundred, so the mean is five.

In Table 1, data is presented in terms of the deviation of these relative propensities from the mean of five.

That is, the data represent $R_i - 5$.

3-letter code	1-letter code	Carbon groups	Interface Propensity	Dehydron Propensity	Hydropathy Kyte et al.[18]
Asn	N	1	+1.28	+1.63	-3.5
Thr	T	1	+1.10	+1.41	-0.7
Gly	G	0	+0.99	+1.42	-0.4
Ser	S	0	+0.60	+0.80	-0.8
Asp	D	1	+0.34	+0.76	-3.5
Ala	A	1	+0.29	+0.6	1.8
Cys	C	1	+0.25	+0.24	2.5
Val	V	3	+0.20	-0.31	4.2
Met	M	3	+0.10	+0.10	1.9
Tyr	Y	7	+0.10	+0.10	-1.3

Table 1: Amino acids ranked according to their likelihood of being found at protein-protein interfaces.

3-letter code	1-letter code	Carbon groups	Interface Propensity	Dehydron Propensity	Hydropathy Kyte et al.[18]
His	H	2	-0.25	-0.25	-3.2
Pro	P	3	-0.25	-0.25	-1.6
Trp	W	9	-0.33	-0.4	-0.9
Arg	R	3	-0.35	-0.4	-4.5
Leu	L	4	-0.35	-1.10	3.8
Phe	F	7	-0.40	-0.40	2.8
Lys	K	4	-0.42	-0.38	-3.9
Glu	E	2	-0.50	-0.11	-3.5
Gln	Q	2	-0.62	-0.6	-3.5
Ile	I	4	-0.70	-0.92	4.5

Table 2: Amino acids ranked according to their likelihood of being found at protein-protein interfaces.

4.4 Explanation of ranking

The unusual ranking of residues in Table 1 was explained in [10] by noting that it correlates closely with the propensity to be engaged in under-wrapped backbone hydrogen bonds, among amino acids acting as either proton donors or acceptors for main-chain hydrogen bonds.

We will review the concept of wrapping of hydrogen bonds shortly.

The data on wrapping are presented in the fifth column in Table 1, and the correlation is striking.

Such under-wrapped backbone hydrogen bonds, in turn, are determinants of protein-protein associations, as discussed subsequently.

As noted in [10], the seven residues with the highest propensity for being engaged in under-desolvated hydrogen bonds also have at most one torsional degree of freedom in their side chain.

Thus, the entropic loss resulting from the conformational hindrance of the sidechains upon protein association is minimal with these sidechains, so that the energetic benefit of intermolecular protection of pre-formed hydrogen bonds is most beneficial.

The only purely hydrophobic residue that has an appreciable propensity to be in an interface is Val, with only one sidechain rotameric state.

Therefore, its conformational hindrance upon binding entails the lowest loss in conformational entropy.

Considering the residues ranked at the bottom of Table 1 demonstrates that hydrophobic residues on the protein surface are infrequent relative to their over-all abundance.

This implies that are negatively selected to be part of binding regions, and thus they must play a secondary role in terms of binding.

Note that the polar residues (Asn, Asp, Ser, Cys and Thr) with a minimal distance from their polar groups to the backbone are likely to be engaged in dehydrons, according to Table 1.

It is presumed [10] that this arises not only because they have minimal nonpolar carbonaceous groups, but also because the relative proximity of their polar groups to a backbone hydrogen bond may limit further clustering of hydrophobic groups around the bond.

Gly is itself the greatest under-wrapper and can even be thought of as polar due to the fact that the polar environment of the peptide bond is exposed;

Ala is the penultimate under-wrapper and may also exhibit some of the polar qualities of Gly.

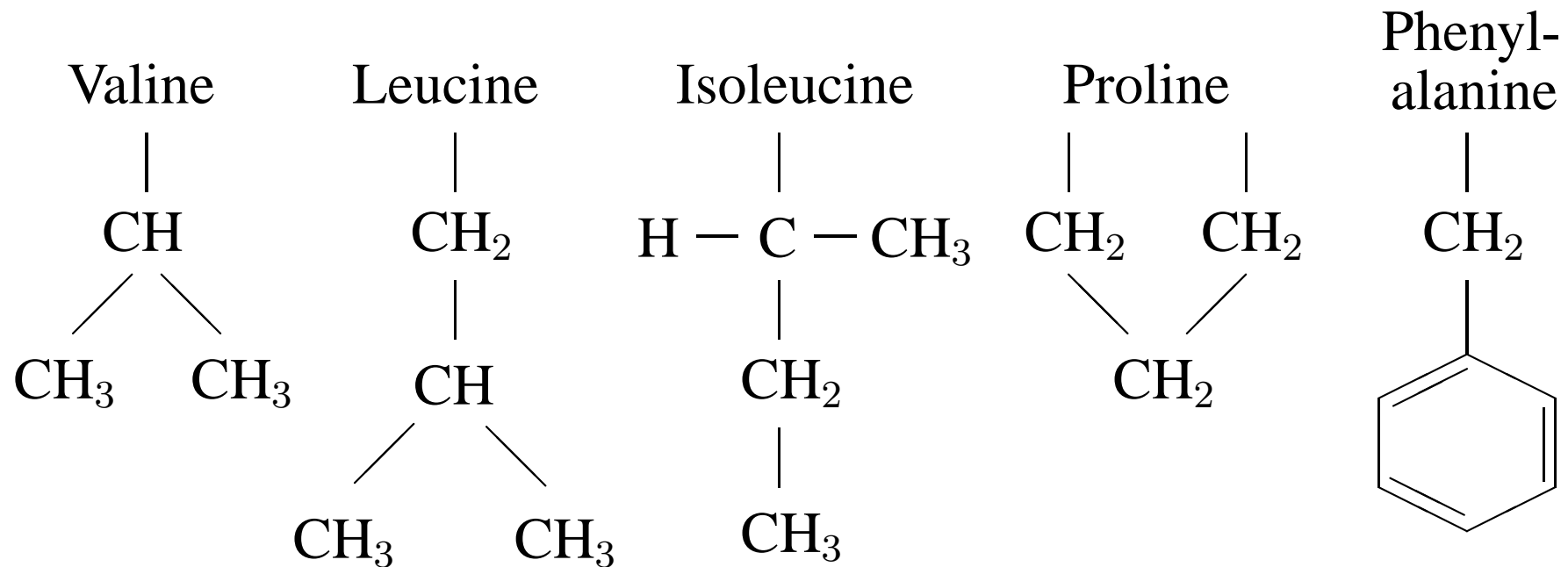
5 Tutorial on hydrophobic wrapping

Effect of modulation of dielectric by hydrophobic groups.

- Amino acid side chains have different properties
- Tutorial on hydrophobicity: carbonaceous groups
- Tutorial on dielectrics: more on this later
- Extent of wrapping changes nature of hydrogen bond
- Dehydrons: Under-wrapped hydrogen bonds
- A model for protein-protein interaction
- Extreme interaction: amyloid formation
- Stickiness of dehydrons

5.1 Amino acid side chains have different properties

Carbonaceous groups on certain side chains are hydrophobic:



Amino acids (side chains only shown) with carbonaceous groups.

5.2 Tutorial on hydrophobicity

Carbonaceous groups (CH, CH₂, CH₃) are hydrophobic because

- they are non-polar and thus do not attract water strongly
- they are polarizable and thus damp nearby water fluctuations

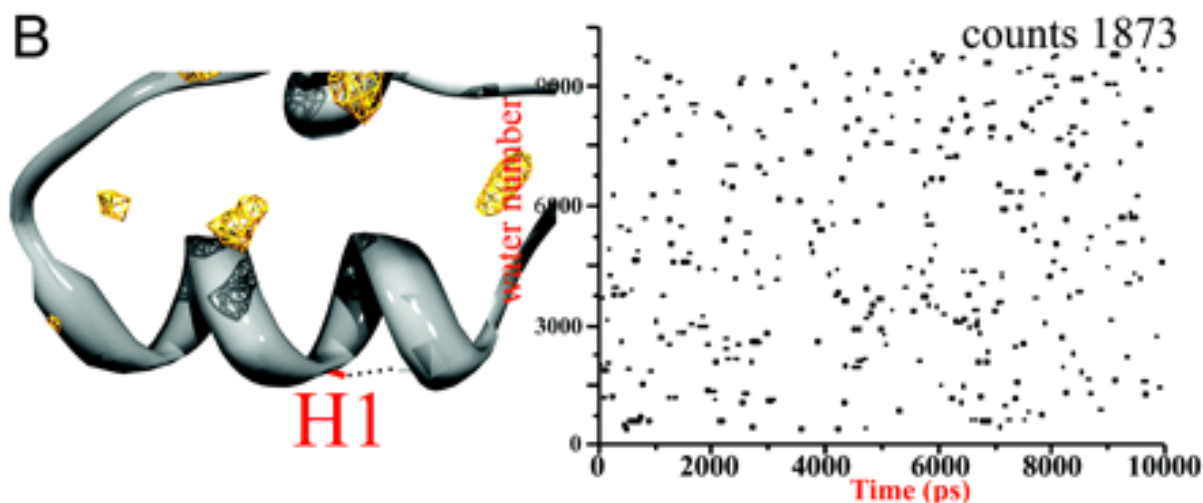
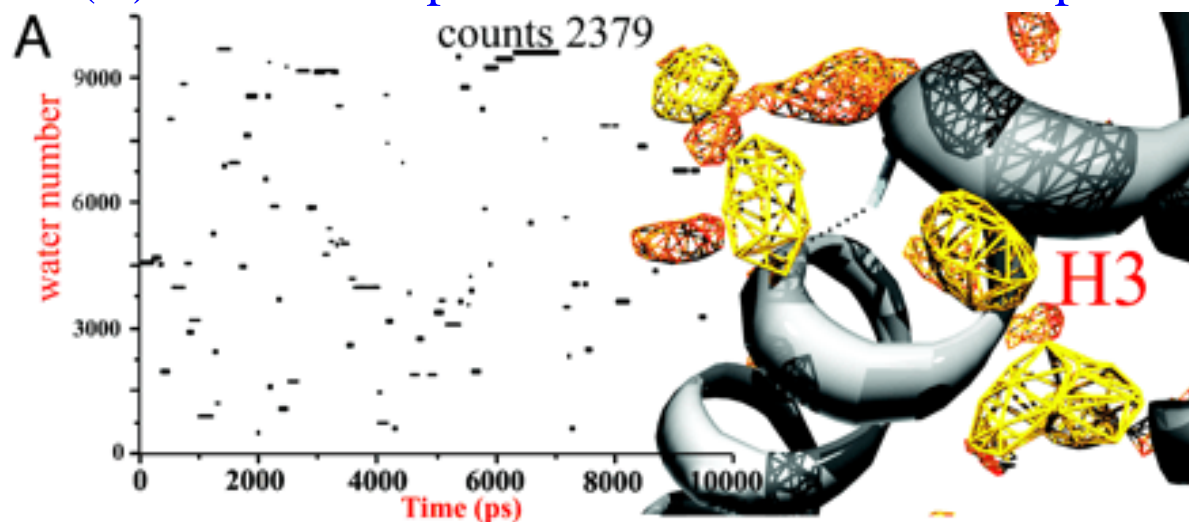
5.3 Tutorial on dielectrics

Water removal reduces the dielectric effect and makes electronic bonds stronger.

Number of carbonaceous groups in a region determine extent of water removal and strength of electronic bonds.

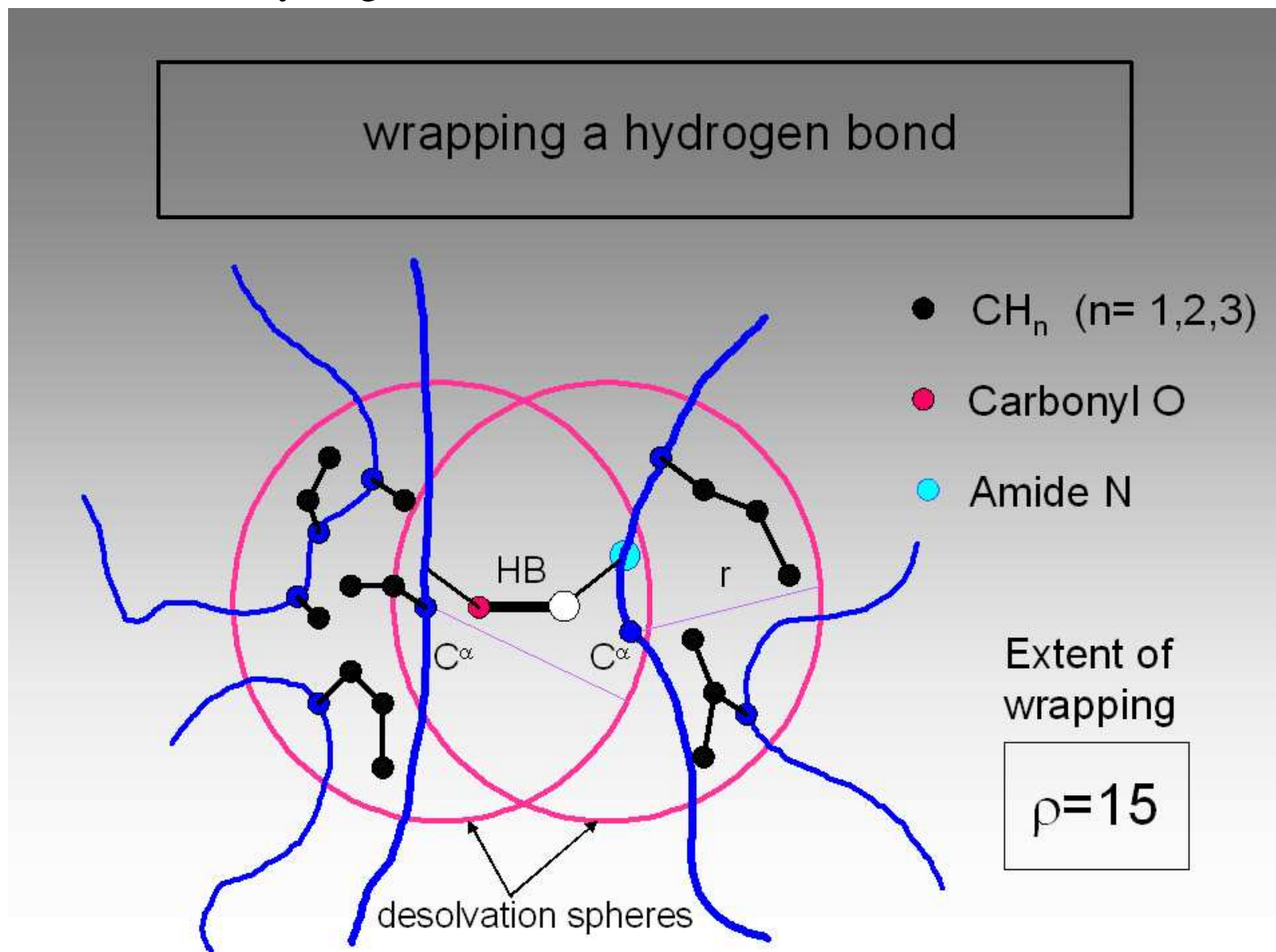
5.4 Extent of wrapping changes nature of hydrogen bond

Hydrogen bonds (B) that are not protected from water do not persist.



From De Simone, et al., PNAS 102 no 21 7535-7540 (2005)

Wrapping made quantitative by counting carbonaceous groups in the neighborhood of a hydrogen bond.



5.5 Under-wrapped hydrogen bonds

Hydrogen bonds with insufficient wrapping in one context can become well wrapped by a partner.

The hydrogen bond is much stronger when wrapped.

The change in energy makes these hydrogen bonds sticky.

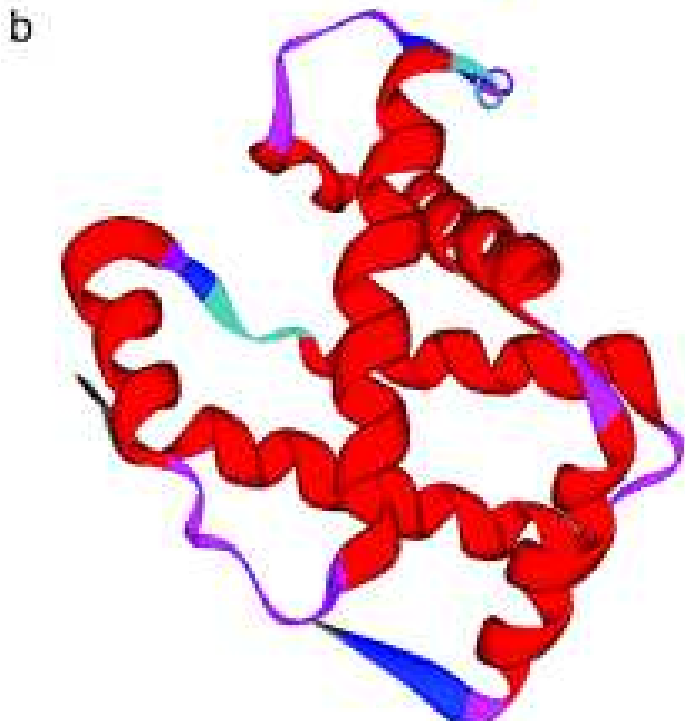
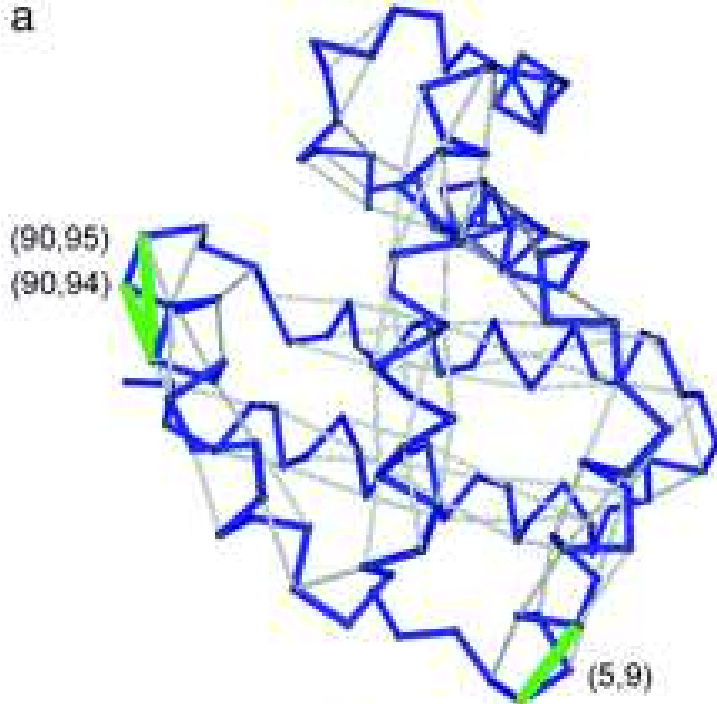
We call such under-wrapped hydrogen bonds

dehydrons

because they can benefit from becoming dehydrated.

The force associated with dehydrons is not huge, but they can act as a guide in protein-protein association.

In our pictures, **our new lens colors dehydrons GREEN** to distinguish from ordinary hydrogen bonds.



Dehydrons

in human hemoglobin, From PNAS

100: 6446-6451 (2003) Ariel Fernandez,
Jozsef Kardos, L. Ridgway Scott, Yuji Goto,
and R. Stephen Berry. Structural defects and
the diagnosis of amyloidogenic propensity.

Well-wrapped

hydrogen bonds are

grey, and dehydrons are green.

The standard ribbon model
of “structure” lacks indicators
of electronic environment.

6 Amino acid pairs at interfaces

We now return to the second question raised at the beginning of the chapter regarding the amino acid composition for interacting pairs of amino acids at interfaces.

We review the results in [12] which use proximity as an interaction metric in which two residues are defined as interacting if their C_β coordinates differ by at most 6Å.

In this setting, some dominant residues are indeed found to be hydrophobic.

We present in Table 3 the residues and their relative propensities, as defined in (4.3), in decreasing order.

Res. Code	Pairing Rel. Prop.	Pairing Rel. Freq.	Pairing Freq. [12]	Total Abundance [4]	Interface Propensity	Rim/Core freq. [6]
Cys	5.4	2.40	1.87	0.78	+0.25	0.45
Trp	1.9	1.60	1.63	1.02	-0.33	0.32
Pro	1.7	1.55	6.74	4.35	-0.25	1.24
Ser	1.5	1.50	7.01	4.66	+0.6	1.04
Asn	1.3	1.46	4.90	3.36	+1.28	1.19
Thr	1.1	1.41	6.87	4.87	+1.1	1.19
His	0.76	1.33	2.56	1.92	-0.25	0.52
Tyr	0.32	1.23	3.70	3.00	+0.1	0.67
Gly	0.11	1.18	8.59	7.30	+0.99	1.16
Ala	0.11	1.18	9.18	7.77	+0.29	0.95

Table 3: Amino acids which occur in pairs at interfaces and their relative abundances. Primary data is taken from the indicated references. Relative Propensity is defined in (4.3) and Relative Frequency is defined in (4.2).

Res. Code	Pairing Rel. Prop.	Pairing Rel. Freq.	Pairing Freq. [12]	Total Abundance [4]	Interface Propensity	Rim/Core freq. [6]
Phe	-0.15	1.12	4.02	3.61	-0.4	0.33
Gln	-0.33	1.08	3.41	3.15	-0.62	1.03
Met	-0.72	0.99	2.38	2.41	+0.1	0.54
Asp	-0.98	0.93	5.06	5.42	+0.34	1.48
Val	-1.2	0.87	7.12	8.17	+0.2	1.09
Leu	-1.6	0.79	7.05	8.91	-0.35	0.82
Ile	-1.8	0.75	5.00	6.66	-0.7	0.76
Arg	-1.9	0.71	4.46	6.27	-0.35	1.19
Glu	-2.6	0.55	4.71	8.59	-0.5	1.87
Lys	-2.9	0.48	3.73	7.76	-0.42	2.16

Table 4: Amino acids which occur in pairs at interfaces and their relative abundances. Primary data is taken from the indicated references. Relative Propensity is defined in (4.3) and Relative Frequency is defined in (4.2).

6.1 Explaining the differences

Two of the residues in Table 3 with greatest relative propensity, namely Trp and Pro, are distinctively hydrophobic, as we might expect.

However, these are also two of the most unique residues.

Moreover, other high-ranking residues are as found in Table 1.

The differences between this table and Table 1 reflect the fact that we are now asking about residues which are in proximity and thus are interacting in some direct way.

It is natural to expect that hydrophobicity would be a prominent type of interaction.

Since Table 3 does not provide relative abundances directly, we need to say how these have been derived.

The fundamental data in Table 3 is Table II on page 93 in [12], which lists the “contact” matrix C_{ij} .

This is a matrix that counts the number of times that residue i contacts (is within the proximity radius of) residue j .

Summing a column (or row) of C_{ij} and normalizing appropriately gives the total frequency F_i of the i -th amino acid involved in such pairings.

More precisely, to report frequencies as a percentage, define

$$F_i = 100 \frac{\sum_{j=1}^{20} C_{ij}}{\sum_{i,j=1}^{20} C_{ij}} \quad (6.4)$$

to be the amino acid pairing frequency, shown in the column entitled Pairing Freq. [12] in Table 3.

6.2 Normalized abundance

The abundance of each amino acid in such pairings needs to be normalized by an appropriate measure.

Here we have taken for simplicity the abundances published in [4] which are reproduced in the column entitled Total Abundance [4] in Table 3.

We do not claim that this provides the optimal reference to measure relative abundance in this setting, but it certainly is a plausible data set to use.

The data shown in the column entitled Pairing Rel. Freq. in Table 3 represents the ratio of F_i , defined in (6.4), to the abundances reported in [4].

The fact that Cys appears to have the highest relative abundance in pairs at interfaces reflects the simple fact that when Cys appears paired with another residue, it is unusually frequently paired with another Cys to form a disulfide bond, as confirmed in [12].

7 Pair frequencies

In addition to looking at the frequencies of individual residues, one can also look at the frequencies of pairings.

A standard tool for doing this is the odds ratio.

Suppose that f_i is the frequency of the i -th amino acid in some dataset, and suppose that $C_{i,j}$ is the frequency of the pairing of the i -th amino acid with the j -th amino acid. Then the odds ratio O_{ij} is defined as

$$O_{ij} = \frac{C_{ij}}{f_i f_j} \quad (7.5)$$

and has the following simple interpretation. If the pairing of the i -th amino acid with the j -th amino acid were random, then we would have $C_{ij} = f_i f_j$, and thus $O_{ij} = 1$. Therefore an odds ratio bigger than one implies that the pairing is more common than would be expected for a random pairing, and conversely if it is less than one.

The **log odds ratio** is often defined by simply taking the logarithm of the odds ratio. This has the benefit of making the more likely pairings positive and the less likely pairings negative. In [12], a quantity G_{ij} is defined by multiplying the log odds ratio by a numerical factor of ten.

It is noteworthy that the odds ratios indicated in Table III of [4] are all between a half and two. That is, there are no pairs which occur even as much as twice as frequently as would be expected randomly (or half as frequently).

The pair with the highest odds ratio (1.87) is Cys-Cys, a disulfide bridge.

Although Cys is uncommon, when it does appear we can expect it to be involved in a disulfide bridge.

The next highest odds ratio pair is Trp-Pro (1.42), which pairs two of the most unique sidechains.

The lack of rotational freedom in proline may be significant since there is no entropic loss in the pairing, but the story is likely much more complex, e.g., Trp-Pro can be involved in a sandwich [24].

In Table III of [4], The following four pairs with the next highest odds ratios involve charged residues:

Asp-His (1.25),

Arg-Trp (1.23),

Asp-Ser (1.22) and

Asp-Thr (1.21).

The first of these is a salt-bridge, and the second is a charge-polar interaction known as a cation- π interaction [11, 30, 7] based on the polarity of aromatic residues.

The latter two pairs are charged and polar residues as well.

The next four pairs in ranking of odds ratio are Cys-Ser (1.20), Asp-Arg (1.19), Met-Met (1.16) and Cys-His (1.15).

These show a similar mix of polar interactions.

So how does hydrophobicity play a role?

7.1 Comparing odds ratios

There is no absolute scale on which to measure odds ratios, and the significance of any deviation from one is context dependent.

But it is notable that the pair frequencies reported in [12] are much smaller than found for alpha helices or beta sheets [19].

The top thirty values for the odds ratios for amino acid pairs with $\theta < 50$ are all greater than two, with the highest being 3.75 [19].

Moreover, the top fifteen values for the odds ratios for amino acid pairs with $\theta > 155$, that is pairs in β sheets, are all greater than two [19].

We interpret that to mean that the hydrophobic pairs involved in interfaces are more nearly random, none of which occur with very high odds ratios.

7.2 Core versus rim pairs

When we add the further analysis in [6] which differentiated the prevalence of core versus rim residues in protein interfaces, the picture is amplified.

In [6], interface topology was characterized in detail, and it was found that interfaces could typically be described in terms of discrete patches of about 1600 Å² in area.

For each patch, the boundary (rim) residues were identified versus the interior (core) residues.

The statistics for amino acid preferences for the rim versus the core are reproduced in Table 3.

There is a strong correlation between being charged or polar and preferring the rim, as indicated in Table 5.

Res. Code	Rel. Prop.	Rel. Freq.	Pair Freq.[12]	Total Abundance [4]	Rim/Core freq. [6]	Homodimer Rim/Core [1]
Lys	-2.9	0.48	3.73	7.76	2.16	2.19
Glu	-2.6	0.55	4.71	8.59	1.87	1.48
Asp	-0.98	0.93	5.06	5.42	1.48	1.61
Pro	1.7	1.55	6.74	4.35	1.24	1.51
Asn	1.3	1.46	4.90	3.36	1.19	1.49
Thr	1.1	1.41	6.87	4.87	1.19	1.16
Gly	0.11	1.18	8.59	7.30	1.16	1.38
Arg	-1.9	0.71	4.46	6.27	1.19	0.85
Val	-1.2	0.87	7.12	8.17	1.09	0.83
Ser	1.5	1.50	7.01	4.66	1.04	1.15

Table 5: Amino acids which occur in pairs at interfaces and their relative abundances. Primary data is taken from the indicated references.

Res. Code	Rel. Prop.	Rel. Freq.	Pair Freq.[12]	Total Abundance [4]	Rim/Core freq. [6]	Homodimer Rim/Core [1]
Gln	-0.33	1.08	3.41	3.15	1.03	1.22
Ala	0.11	1.18	9.18	7.77	0.95	0.93
Leu	-1.6	0.79	7.05	8.91	0.82	0.61
Ile	-1.8	0.75	5.00	6.66	0.76	0.55
Tyr	0.32	1.23	3.70	3.00	0.67	0.58
Met	-0.72	0.99	2.38	2.41	0.54	0.68
His	0.76	1.33	2.56	1.92	0.52	0.85
Cys	5.4	2.40	1.87	0.78	0.45	0.81
Phe	-0.15	1.12	4.02	3.61	0.33	0.40
Trp	1.9	1.60	1.63	1.02	0.32	0.60

Table 6: Amino acids which occur in pairs at interfaces and their relative abundances. Primary data is taken from the indicated references.

7.3 Special interfaces

Protein-ligand interfaces differ in function, and interfaces with different function can have different composition.

In [15], basic differences between protein-antibody and enzyme-inhibitor pairs, as well as others, are explored.

Using more extensive datasets available more recently, this approach has been refined to allow classification of interface type based on aminoacid composition [23].

8 Comparisons and caveats

We have made several observations based on analyzing existing data sets. These conclusions should be viewed as preliminary since these data sets must be viewed as incomplete. Our primary intent was to introduce a methodology for exploring such data sets and to indicate the type of results that can be obtained.

Our basic analysis of pairwise interaction data was taken from [12]. However, the methodology is quite similar to that of the earlier paper [28], although there are differences in the way the interior (and non-interior) sidechains in the interaction zone are defined.

That is, the classification of rim and core residues in the interface [12] is different in definition from exposed and interior residues in the interface in [28], although similar in spirit.

Figure 3B of [28] shows how the residues that are interacting (proximate) in an interface are very similar in composition to ones in the interior of proteins.

8.1 Comparing interfaces

To illustrate the sensitivity of results depending on the database chosen, we review the results in [1] which is very similar in spirit to [6], the difference being the use of homodimers for the study of interfaces.

In Table 5, we present this data, with the residues reordered to give the rim/core preferences in order for the data in [6] to facilitate comparison with the data in [1].

What we see is the same general trend, namely that charged and polar residues prefer the rim, but with changes in the particular rankings among the different groups.

However, there is a significant reversal in the roles of arginine and valine [1].

8.2 Nonspecific interactions

The dissection trilogy is completed in [2] in which an attempt is made to determine aminoacid distributions for “nonspecific” interactions.

This is intended to be a proxy for any surfaces which might bind however briefly to other protein surfaces. The dataset is determined by looking at crystal contact surfaces in the PDB.

We leave as an exercise to compare the data for these surfaces with the other data presented here.

See [2] for a comparison with the data in [6] and [1].

8.3 Hot spots and alanine scanning

In [3], an attempt is made to identify so-called “hot spots” on protein surfaces.

They report on the results of an experimental technique called **alanine scanning** in which residues are replaced by alanine and compared with the original protein by some activity assay.

What they discover is that the most common sidechains at hot spots are the ones that are bulkiest, Trp, Tyr and Arg. This is not surprising since the replacement by Ala has the greatest change in geometry for these residues.

However, such substitutions might be extremely rare.

What might be a better test of importance would be other mutations, e.g., ones which do not change the volume or geometry of the side chain. Systematic replacement of all amino acids by all other amino acids is clearly an order of magnitude more work than just replacing by a fixed side chain.

Having a better model of what governs protein-protein interactions could lead to a more directed study of sidechain mutation effects.

8.4 Cation- π interactions

The aromatic sidechains do play a special role in protein interfaces through what is called a cation- π interaction [11].

The special polar nature of the aromatic residues provides the opportunity for interaction with positively charged (cation) residues (Lys, Arg, His).

The cation- π motifs play a special role in protein interfaces [30, 7].

The cation- π interaction also has a significant role in α -helix stabilization [26].

A study of the role of evolution on protein interface composition can be found in [5]. In [13, 20], interacting amino acids across interfaces are studied and compared with regard to conservation and hot spots.

Protein-protein interactions can be classified in different ways, e.g., by how transient they are, and studies have been done to examine differences in size of interaction zones and sidechain propensities [22, 21].

Identification of individual sidechains that may play the role of “anchors” in protein-ligand recognition is studied in [25] via molecular dynamics simulations. Individual residues are identified that appear to fit into geometric features on paired protein surfaces both in crystal structures and in the dynamic simulations.

9 Conclusions

Two main conclusions were obtained. The first is that hydrophobicity is not the primary variable that determines proximity to interaction sites. Instead, there is a different “interactivity” order that governs the likelihood of an amino acid residue being in an active zone. Electrical forces have a more prominent role than simple steric effects.

On the other hand, pairwise interactions of hydrophobic residues do play a secondary role in protein-protein interactions, especially in the interior, or core, regions of interaction domains. Moreover, their interactions tend to be less specific than might be the case in other pairings, such as in alpha helices and beta sheets. Nonspecific interactions can be problematic [1].

References

- [1] Ranjit Prasad Bahadur, Pinak Chakrabarti, Francis Rodier, and Joël Janin. Dissecting subunit interfaces in homodimeric proteins. *Proteins: Str. Func. Gen.*, 53:708–719, 2003.
- [2] Ranjit Prasad Bahadur, Pinak Chakrabarti, Francis Rodier, and Joël Janin. A dissection of specific and non-specific protein-protein interfaces. *Journal of Molecular Biology*, 336:943–955, 2004.
- [3] A. A. Bogan and K. S. Thorn. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280:1–9, 1998.
- [4] Dawn J. Brooks, Jacques R. Fresco, Arthur M. Lesk, and Mona Singh. Evolution of amino acid frequencies in proteins over deep time: Inferred order of introduction of amino acids into the genetic code. *Molecular Biology and Evolution*, 19:1645–1655, 2002.
- [5] Daniel R. Caffrey, Shyamal Somaroo, Jason D. Hughes, Julian Mintseris, and Enoch S. Huang. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci*, 13(1):190–202,

2004.

- [6] Pinak Chakrabarti and Joël Janin. Dissecting protein-protein recognition sites. *Proteins: Str. Func. Gen.*, 47:334–343, 2002.
- [7] Peter B. Crowley and Adel Golovin. Cation- π interactions in protein-protein interfaces. *PROTEINS: Structure, Function, and Bioinformatics*, 59:231–239, 2005.
- [8] W. L. DeLano, M. H. Ultsch, A. M. de Vos, and J. A. Wells. Convergent solutions to binding at a protein-protein interface. *Science*, 287:1279–1283, 2000.
- [9] Ariel Fernández and Harold A. Scheraga. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc Natl Acad Sci USA*, 100(1):113–118, Jan 2003.
- [10] Ariel Fernández, L. Ridgway Scott, and Harold A. Scheraga. Amino-acid residues at protein-protein interfaces: Why is propensity so different from relative abundance? *J. Phys. Chem. B*, 107(36):9929–9932, 2003.
- [11] Justin P. Gallivan and Dennis A. Dougherty. Cation- π interactions in

structural biology. *PNAS*, 96(17):9459–9464, 1999.

- [12] Fabian Glaser, David M. Steinberg, Ilya A. Vakser, and Nir Ben-Tal. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins: Str. Func. Gen.*, 43:89–102, 2001.
- [13] Inbal Halperin, Haim Wolfson, and Ruth Nussinov. Protein-protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. implications for docking. *Structure*, 12:1027–1038, 2004.
- [14] Zengjian Hu, Buyong Ma, Haim Wolfson, and Ruth Nussinov. Conservation of polar residues as hot spots at protein interfaces. *Proteins: Str. Func. Gen.*, 39:331–342, 2000.
- [15] Susan Jones and Janet M. Thornton. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA*, 93:13–20, 1996.
- [16] Susan Jones and Janet M. Thornton. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.*, 272:121–132, 1997.
- [17] Susan Jones and Janet M. Thornton. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.*, 272:133–143, 1997.

- [18] J. Kyte and R. F. Doolittle. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157:105–132, 1982.
- [19] Jing Liu, L. Ridgway Scott, and Ariel Fernández. Interactions of aligned nearest neighbor protein side chains. *Journal of Bioinformatics and Computational Biology*, 7:submitted–, 2006.
- [20] Buyong Ma, Tal Elkayam, Haim Wolfson, and Ruth Nussinov. Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci USA*, 100:5772–5777, 2003.
- [21] Irene M. A. Nooren and Janet M. Thornton. Diversity of proteinprotein interactions. *EMBO Journal*, 22:3486–3492, 2003.
- [22] Irene M. A. Nooren and Janet M. Thornton. Structural characterisation and functional significance of transient protein-protein interactions. *Journal of Molecular Biology*, 325(5):991–1018, 2003.
- [23] Yanay Ofren and Burkhard Rost. Analysing six types of protein-protein interfaces. *Journal of Molecular Biology*, 335:377–387, 2003.

- [24] Sergei Radaev and Peter Sun. Recognition of immunoglobulins by Fc γ receptors. *Molecular Immunology*, 38(14):1073–1083, 2002.
- [25] Deepa Rajamani, Spencer Thiel, Sandor Vajda, and Carlos J. Camacho. Anchor residues in protein-protein interactions. *Proc Natl Acad Sci USA*, 101:11287–11292, 2004.
- [26] Z. Shi, C.A. Olson, and N.R. Kallenbach. Cation- π interaction in model α -helical peptides. *Journal of the American Chemical Society*, 124(13):3284–3291, 2002.
- [27] G. R. Smith and M. J. Sternberg. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol*, 12:28–35, 2002.
- [28] Chung-Jung Tsai, S. L. Lin, Haim J. Wolfson, and Ruth Nussinov. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Science*, 6:53–64, 1997.
- [29] Dong Xu, Chung-Jung Tsai, and Ruth Nussinov. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering*, 10(9):999–1012, 1997.

[30] Wai-Ming Yau, William C. Wimley, Klaus Gawrisc, and Stephen H. White. The preference of tryptophan for membrane interfaces. *Biochemistry*, 37:14713–14718, 1998.