The amino acid side chains with polar groups are **amphiphilic**, having both polar and nonpolar segments. Other amphiphiles include lipids and detergents. Amphiphilic molecules tend to interact in aqueous solution in such a way that their nonpolar segments interact with other nonpolar groups and their polar groups are in contact with water. This is the basic principle of the formation of membranes, lipid bilayers, and micelles. A similar phenomenon can be seen to produce the folded conformations of proteins (see Chap. 6).

A useful concept in considering amphiphilic molecules is the **hydrophobic moment**, which is exactly analogous to the dipole moment of electrical charge (Sec. 4.1.2.b) but represents a vector from the hydrophilic to the hydrophobic parts of a molecule. The hydrophobic moment of a polypeptide chain in a particular three-dimensional structure is calculated from the vector sum of the contributions of each of the amino acid residues. This contribution is given by a vector that points from the $C^\alpha$ atom to the center of the side chain and whose length is proportional to the hydrophobicity of the side chain. This parameter can often account for the architecture and interactions of large molecules such as proteins.

**References**


**4.4 Intramolecular Interactions**

The analysis presented in this chapter of the electrostatic, van der Waals, and hydrogen-bond interactions among the atoms of proteins indicates that all such interactions are weak in the presence of water. Ionized or polar groups interact with water almost as favorably as they interact with other suitable ionized or polar groups (Table 4.9), and it is energetically unfavorable to remove them from aqueous solution (Fig. 4.8). Nonpolar groups prefer to interact with each other rather than with water (Table 4.8), but even the resulting hydrophobic interaction is not very strong (Table 4.9). Yet such interactions will be shown in Chapter 6 to produce stable folded conformations of proteins. How can they do this?

For molecules to interact, they must lose entropy, which is energetically unfavorable. Were it not for entropy, all matter would be solid. This contribution to the free energy, which arises from molecules having freedom, makes the liquid and gaseous states possible. How much entropy is lost in an interaction depends on the number of degrees of freedom that must be fixed. For example, van der Waals interactions require the least entropy loss because only the distance between two atoms is fixed (Fig. 4.2), whereas hydrogen bonding requires that both proximity and orientation be fixed to some extent (Figs. 4.3 and 4.4).

Entropic considerations are especially important when two or more interactions are possible simultaneously in a single molecule because, in favorable cases, much less entropy need be lost in the second and subsequent interactions than in the first. Two interactions that can occur simultaneously can be much more favorable energetically than might be expected from their individual strengths.
4.4.1 Effective Concentrations

The magnitude of entropic cooperativity can be illustrated by intramolecular interactions. Two parts of the same molecule can interact without losing as much entropy as must be lost to bring two independent molecules together. Because the two parts of a molecule are already fixed to some degree in proximity and orientation, only some fraction (which depends on the molecule and the interaction) of the internal flexibility of the molecule has to be lost in the intramolecular interaction.

Intramolecular and bimolecular examples of the same interaction can be compared by means of the ratio of their equilibrium constants, which for the intramolecular interaction is dimensionless and for the intermolecular interaction has dimensions of $(\text{concentration})^{-1}$. Therefore, the ratio of the two has the dimensions of concentration, which can be thought of as the effective concentration of the two groups when they are part of the same molecule in the intramolecular interaction:

$$
A + B \xleftrightarrow{K_{\text{inter}}} A \cdot B
\quad (4.20)
$$

$$
A - B \xleftrightarrow{K_{\text{intr}}} A \cdot B
\quad (4.21)
$$

$$
\frac{K_{\text{intr}}}{K_{\text{inter}}} = \text{effective concentration of } A - B \quad (4.22)
$$

It was thought for a long time that the maximum effective concentration of two groups in aqueous solution was about 55 M, the concentration of pure water, when one group could be considered to be immersed in a liquid environment of the second component. Consequently, one often finds instances of the magnitudes of intramolecular and intermolecular interactions being interconverted by using the factor of 55 M.

Many experimental measurements have been made for various chemical reactions, however, and much greater values of effective concentrations are generally found: representative examples are given in Table 4.11. These examples represent chemical reactions involving reversible covalent bond formation that can be considered analogous to noncovalent interactions. The covalent nature of these interactions, however, probably increases the magnitude of the entropic effect, due to the more stringent geometrical requirements of covalent bond formation. In any case, the first three examples in the table (A–C) involve flexible molecules with relatively free rotations about three single bonds, which must be restricted to form the product. In spite of this considerable entropic loss, the effective concentrations measured are in the range of $10^2$ to $10^3$ M. Therefore, merely keeping two groups in reasonable proximity by linking them covalently through several bonds causes their concentration relative to each other to be much higher than would be possible if the groups were on separate molecules, even in the most concentrated liquid state. The last example of Table 4.11 (D) has an enormous effective concentration of $5 \times 10^6$ M, which is undoubtedly due primarily to the small entropy difference between the molecule with and without the anhydride interaction. In this case, the planar aromatic structure of the molecule keeps the carboxyl groups in close proximity whether or not the anhydride is present. The very small increase in flexibility and entropy that occurs when the anhydride interaction is broken results in an enormous effective concentration that is close to the maximum considered possible theoretically (approximately $10^6$ M). Of course, other factors can cause large apparent effective concentrations, such as strain in the molecule that is relieved upon forming the interaction, but numerous examples of large values are known that illustrate the entropic contribution.

When there is no entropic difference between the molecules with and without the interaction, the effective concentration is at its maximum value. This value depends on the type of interaction. Those in which the proximity and orientation of the interacting groups are very important, as in a hydrogen bond and especially when a covalent bond is formed, have very high maximum effective concentrations, up to $10^6$ M. When these factors are not so important, as in van der Waals interactions, the interacting groups have significant degrees of freedom and have less entropy to gain upon dissociating, so lower values of maximum effective concentrations apply. Even in this last case, though, the maximum values are substantially greater than 55 M. The reason for this high maximum value is that the molecules of a liquid have a high degree of translational and rotational freedom, so they are far from being in the optimal situation for interacting.

Unfortunately, the magnitudes of the effective concentrations expected for interactions of the type observed in proteins are not known. Only in the case of the disulfide interaction between thiol groups have values been measured in proteins (Sec. 7.5.4). The maximum value observed is somewhat greater than $10^3$ M, but the disulfide bond is a covalent interaction, which tends to enhance the effective concentration. Hydrogen bonds are moderately sensitive to orientation and probably have a partial covalent character, so substantial maximum values would be expected, but probably much less than $10^4$ M and less than those involving disulfide bonds. Ionic and hydrophobic interactions are not stereochemically very stringent, so maximum values of $10^2$–$10^3$ M may apply in these instances.
<table>
<thead>
<tr>
<th>Example</th>
<th>Equilibrium reaction</th>
<th>Effective concentration (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>( \text{CO}_2\text{H} + \text{CH}_2 + \text{SH} \rightleftharpoons \text{H}_2\text{C} - \text{S} - \text{C} + \text{H}_2\text{O} )</td>
<td>( 3.7 \times 10^3 ) ( ^a )</td>
</tr>
<tr>
<td>B</td>
<td>( \text{HOCH} + \text{GSSG} \rightleftharpoons \text{HOCH} + 2\text{SH} )</td>
<td>( 2 \times 10^2 ) ( ^b )</td>
</tr>
<tr>
<td>C</td>
<td>( \text{CO}_2\text{H} + \text{CH}_2 + \text{SH} \rightleftharpoons \text{H}_2\text{C} - \text{O} + \text{H}_2\text{O} )</td>
<td>( 1.9 \times 10^5 ) ( ^a )</td>
</tr>
<tr>
<td>D</td>
<td>( \text{CH}_3\text{CO}_2\text{H} \rightleftharpoons \text{CH}_3\text{CO}_2\text{H} + \text{H}_2\text{O} )</td>
<td>( 5.4 \times 10^9 ) ( ^a )</td>
</tr>
</tbody>
</table>


In contrast to the very high effective concentrations that are possible when interacting groups are held in the appropriate proximity and orientation, constituent groups that are kept apart by the structure of their molecules have very low, or zero, effective concentrations. Intramolecular interactions are much more sensitive to their environment than are interactions between independent molecules in the liquid state.

Detailed explanations of most values of the effective concentrations measured are complicated by the presence of unfavorable steric or physical interactions in the molecules with or without the interaction. Consequently, there is no ideal example with which to illustrate the solely entropic contribution to the effective concentration, but the many experimental examples available indicate that the effect is very substantial.

**References**

4.4.2 Multiple Interactions

Multiple groups on a molecule can behave very differently from the same groups in isolation. For example, individual ions do not associate very strongly in aqueous solution (Table 4.9) because of their favorable entropies and strong interactions with water; but a polyelectrolyte molecule that has a number of such charged groups binds ions of the opposite charge very tightly due to interactions among the charged groups. Being part of the same molecule, the charged groups are constrained to be close to each other by the covalent bonds. The electrostatic repulsion between groups with the same charge is compensated by the counterions that they attract from the solution and bind very tightly; just how tightly depends on the charge density of the polyelectrolyte and the valence of the counterions. The degree of binding is almost independent of the concentration of counterions in the bulk solvent because the diminution of the electrostatic repulsions in the polyelectrolyte is energetically much more important than is the equilibration of the ions with the bulk solvent. The counterions are not necessarily bound at specific sites, and they can retain their water of hydration and move in an unrestricted and random manner along the polyelectrolyte chain. This phenomenon can be very important for binding other ligands that have the same charge as the counterions; displacement of the counterions into a very dilute bulk solution by the binding of the ligand can provide a strong driving force for its binding.

Another means of compensating unfavorable electrostatic repulsions in a polyelectrolyte is suppression of the ionization of a fraction of its groups. Consequently, groups on a polyelectrolyte can have pKₐ values that are very different from those found when they are isolated. These electrostatic effects among multiple groups on a polyelectrolyte are important for the function of proteins, but especially for the function of nucleic acids, with their numerous phosphate groups, and for the interactions of nucleic acids with proteins (Sec. 8.3.2).

4.4.3 Cooperativity of Multiple Interactions

The simultaneous presence of multiple interactions in a single molecule produces cooperativity between them, and together they can be much stronger than might be expected from the sum of their individual strengths. Cooperativity is essential for proteins, in which the noncovalent interactions are intrinsically very weak (Table 4.9). Only when such interactions cooperate is a stable single conformation produced.

Consider an unfolded polypeptide chain in which two groups A and B are capable of interacting favorably as in a hydrogen bond, a salt bridge, or a nonpolar hydrophobic interaction:

\[
\begin{array}{c}
\text{A} \\
\text{K}_{AB}[A/B]_U \\
\text{B}
\end{array}
\]

The observed equilibrium constant for interaction of the two groups, \(K_{\text{obs,U}}\), can be expressed as

\[
K_{\text{obs,U}} = K_{AB}[A/B]_U
\]

where \(K_{AB}\) is the association constant measured with groups A and B on individual molecules (Table 4.9) and \([A/B]_U\) is the effective concentration of the two groups relative to each other on the unfolded polypeptide U. Groups attached to moderate sized random polypeptides have effective concentrations in the range \(10^{-2}\) to \(10^{-5}\) M, depending on their relative positions in the polypeptide chain (see Sec. 5.2). With typical values of \(K_{AB}\) (Table 4.9), values for the observed equilibrium constant \(K_{\text{obs,U}}\) of between \(4 \times 10^{-3}\) and \(10^{-1}\) are expected for individual hydrogen bonds, salt bridges, and so on. Consequently, a single interaction between two groups on a polypeptide chain is not expected to be stable unless the groups are close in the covalent structure in such a way that they have an especially high effective concentration.

Multiple interactions among two or more pairs of groups on the same molecule are often not independent but assist or interfere with each other. The following

References


Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening and ion
The equilibria are possible with two pairs of groups on a polypeptide:

\[ \frac{[A/B]_I}{[A/B]_U} = \frac{[C/D]_I}{[C/D]_U} = \text{Coop} \] (4.26)

The factor \( \text{Coop} \) is the degree of cooperativity between the two interactions. Consequently, each interaction can be more stable in the presence of the other interaction than when it takes place alone.

If additional groups that may also interact simultaneously are present on the polypeptide chain, these equilibria are extended in a similar way. The overall equilibrium constant between the final state (with all the interactions present) and the unfolded state (with none) is the product of the individual equilibrium constants along any of the conceivable reaction paths; for example,

\[ K_{\text{net}} = (K_{AB}[A/B]_U)(K_{CD}[C/D]_U)(K_{EF}[E/F]_U) \cdots \] (4.27)

The value of \( K_{\text{net}} \) is independent of the reaction path, so we need not know or propose a specific "folding pathway."

The final conformation is stable—that is, populated by most of the molecules—only if the value of \( K_{\text{net}} \) is greater than unity. Consider a series of weak interactions. The first will be very weak, with an equilibrium constant of \( 10^{-3} \) to \( 10^{-7} \). But the presence of the first interaction can increase the effective concentration of the second pair of groups, so the equilibrium constant for the second interaction may be somewhat larger than that of the first, by the factor \( \text{Coop} \). If the second equilibrium constant is also less than unity, however, the product of the two equilibrium constants is even smaller than the first (Fig. 4.14). Similarly, the net stabilities of conformations with additional weak interactions are even lower than that of the conformation with a single interaction. This process continues until the effective concentrations of additional interacting groups are suf-

---

**FIGURE 4.14**

Hypothetical illustration of the cooperativity produced by multiple weak interactions. Up to 10 interactions are possible simultaneously, and the contribution of the \( i \)th interaction is given at the top. The initial interaction has an equilibrium constant of \( 10^{-4} \), and each additional interaction is 10 times stronger than the previous one as a result of an increase in the effective concentration. The overall equilibrium constant \( K_{\text{net}} \) (bottom) is the product of the contributions of the \( i \) interactions present (see Eq. 4.27). Only with 10 such interactions is \( K_{\text{net}} > 1 \), implying stability of the folded structure. The free energy of each state relative to \( U \) is given by \( \Delta G^*_{\text{net}} = -RT \ln K_{\text{net}} \), with the scale on the right pertaining to 25°C.
FIGURE 4.15
Simple schematic diagram of cooperativity among three simultaneous interactions occurring between groups A and B, C and D, and E and F. The strength of each interaction is determined by the effective concentration of the two groups when they are not interacting, as in conformations I, II, and III. Assuming there are no other considerations, the value of the effective concentration will be inversely proportional to the degree of flexibility permitted. Therefore, the most stable interactions should be those between groups that are held most rigidly by the other interactions, in this case C and D, and the stability of each interaction should depend on the stabilities of all the others.

sufficiently increased to make the equilibrium constant for each additional interaction greater than unity. The value of $K_{net}$ then increases in magnitude with each additional interaction. A sufficient number of simultaneous weak interactions can make the value of $K_{net}$ greater than unity and the folded conformation stable.

An example with Coop = 10 for each additional interaction is given in Figure 4.14. Partially folded structures, those with incomplete stabilizing interactions, are unstable relative to the initial and final states, which means that the transition is cooperative. The degree of cooperativity will be even greater if the intermediate structures have nonbonded groups in unfavorable environments, such as polar groups present in nonpolar environments without being hydrogen bonded. Such situations have been ignored here.

In summary, weak interactions are expected to stabilize a particular folded conformation only when they cooperate so that the interacting groups have very high effective concentrations in that structure. The effective concentration of two groups in a folded structure depends on the extent to which the groups are held in proximity when not interacting (Fig. 4.15), which in turn depends on the stability of all the surrounding interactions. All parts of such a structure, therefore, are expected to be mutually dependent to varying degrees.

As just described, the contribution of each interaction to net stability of the folded structure should depend on the effective concentration of the interacting groups in that folded structure. If the groups are on the surface or in a flexible part of the folded structure, their effective concentration will be low and the interaction will provide little, if any, net stability. Breaking that interaction will have little effect on the folded state. On the other hand, groups in relatively rigid parts of the folded structure will have high effective concentrations, and their interaction will provide a substantial contribution to the net stability of the conformation; removing or altering such an interaction would have a large effect on the stability of the folded conformation.

References


Exercises

1. Ion pairs in proteins involving Arg residues have been observed to be energetically stronger than those involving Lys residues. If this were the case, in what ways might
From these general principles, it is often possible to guess correctly the structure of a protein–ligand complex if the structures of the two components are known. Nevertheless, do not underestimate the difficulty of fitting together two molecules of known structure to make the most stable complex. There are an enormous number of ways that two molecules can associate, especially if at least one is the size of a protein, and only one of the ways is likely to be the correct one.

### References


### 8.2 Energetics and Dynamics of Binding

A fundamental aspect of the interaction of a protein with a ligand is the affinity of the two for each other, which is a measure of the overall free energy of the interaction. The magnitude of the affinity determines whether a particular interaction is relevant under a given set of conditions. The observed affinities of proteins for ligands vary enormously, ranging from very high values, for which dissociation is immeasurably small, to very low values, for which the concentration of free ligand required for a significant degree of binding is so great as to cast doubt on its relevance. Whether or not any particular affinity of a protein for a ligand is significant depends on the concentration of the ligand that the protein is likely to encounter; no other generalizations are possible.

If the affinity is very high, the protein is likely to be found and isolated as the complex; if such a ligand is relatively small, it is designated a prosthetic group. Examples are the heme groups of the globins and cy-
tochromes, some coenzymes that bind tightly to enzymes, and metal ions that are integral parts of the protein structure. With lower affinities, ligands that are
originally bound to a protein are likely to be lost during purification, unless they are added to the protein solution.

8.2.1 Binding Affinities

The affinity between a protein P and a ligand A is measured by the association constant $K_a$ for the binding reaction at equilibrium:

$$ P + A \underset{K_a}{\overset{}{\rightleftharpoons}} P \cdot A $$

(8.2)

$$ K_a = \frac{[P \cdot A]}{[P][A]} $$

(8.3)

All species are presumed to be present at sufficiently low concentrations for thermodynamic ideality to apply; if not, activities rather than concentrations must be measured. $K_a$ is a constant under a given set of conditions and is measured experimentally by the dependence of binding on the free ligand concentration. Several common used graphic methods of analyzing binding data are illustrated in Figure 8.6.

The ratio of bound to free protein should be, according to Equation (8.3), directly proportional to the free-ligand concentration:

$$ \frac{[P \cdot A]}{[P]} = K_a[A] $$

(8.4)

An experimentally more useful measure of binding is the fraction $y$ of protein molecules with bound ligand:

$$ y = \frac{[P \cdot A]}{[P] + [P \cdot A]} = \frac{K_a[A]}{1 + K_a[A]} $$

(8.5)

The greater the value of $K_a$, the greater the affinity. The value of $K_a$ has units of concentration, however, and it is often intuitively easier to consider the dissociation constant $K_d$, which is simply the reciprocal of $K_a$ and has units of concentration. With concentrations of free ligand below $K_d$, little binding to the protein occurs. With a concentration equal to $K_d$, half the protein molecules have bound ligand. An occupancy of 90% requires a nine times greater concentration of free ligand, whereas 99% occupancy requires that the concentration be 99 times $K_d$. Binding equilibria are simplest when the ligand is present at a concentration much greater than that of the protein binding sites. Uptake of the ligand by the protein does not then significantly alter the concentration of free ligand.

Specific binding by a protein of one ligand, and not another, depends on their relative affinities, their concentrations, and whether they bind at the same site. If two ligands are present at a concentration of $10^{-5} M$ but have different values of $K_a$—say, $10^{-3} M$ and $10^{-6} M$—only the ligand with the lower $K_a$ is bound significantly. If both are present at much higher concentrations—say, $10^{-2} M$—both are bound to the protein to the maximum extent if they bind at separate sites. In this case, the higher affinity of one ligand is almost immaterial. If the two ligands compete for the same site, however,

$$ P \cdot A \underset{K_a}{\overset{K_d}{\rightleftharpoons}} P \cdot B $$

(8.6)

$$ [P] = \frac{[P \cdot A]}{[A]} \frac{K_d}{K_a} \frac{[P \cdot B]}{[B]} $$

(8.7)

$$ \frac{[P \cdot A]}{[P \cdot B]} = \frac{K_d[A]}{K_a[B]} $$

(8.8)

the ligand with the higher affinity is bound to a correspondingly greater extent when the ligands are present at the same concentration. Weaker affinity can always be overcome by a higher concentration of that ligand, however, so binding affinities should always be considered relative to the concentration of the ligand.

The energetics of binding are often expressed by the Gibbs free energy of binding, $\Delta G_{bind}$:

$$ \Delta G_{bind} = -RT \ln K_a = RT \ln K_d $$

(8.9)

It must be kept in mind, however, that $K_a$ and $K_d$ have units of concentration and that the value of $\Delta G_{bind}$ depends on which units are used (i.e., the standard state). If the units are moles per liter, the standard state is 1 M, and the calculated value of $\Delta G_{bind}$ applies only under the rather arbitrary situation when the concentration of free ligand is 1 M. In many instances a "unitary" free energy of binding is used as a measure of the intrinsic affinity; this is the free energy of binding that would occur with ligand at a hypothetical concentration of 55 M, the normal concentration of water. This parameter is not of any special significance, however, except when the ligand is water, and it does not represent the free energy of interaction that would occur in a unimolecular interaction (see Sec. 8.2.2).

The energetics of binding are defined more explicitly as the difference in free energies of the free and liganded protein, $\Delta G_b$:

$$ \Delta G_b = -RT \ln (K_a[A]) = -RT \ln \left( \frac{[A]}{K_d} \right) $$

(8.10)

In this case, the concentration of free ligand must be specified. In a similar way, the enthalpy and entropy of
8.2 Energetics and Dynamics of Binding

FIGURE 8.6
Some common methods of plotting binding data, using theoretical curves for the simple binding reaction $P + A \rightleftharpoons P \cdot A$. The concentration of free ligand is expressed relative to its dissociation constant, which is that concentration of free ligand that gives half-maximal binding. The concentrations of free and liganded forms of the proteins are given relative to the concentration of total protein. A: The normal hyperbolic relationship between binding and free-ligand concentration, demonstrating that a free-ligand concentration 9-fold greater than its dissociation constant produces only 90% of maximal binding (indicated by the dashed line); a 99-fold greater concentration is required for 99% saturation. B: A logarithmic scale emphasizes the wide range of free-ligand concentrations required for a complete binding curve. C: Scatchard plot. The negative slope gives the value of the association constant (the reciprocal of the dissociation constant). The horizontal intercept gives the extrapolated extent of the maximal binding. D: Hill plot. An accurate value for the maximum binding is required for this plot because both the liganded and the free protein concentrations are required. The value of the dissociation constant is given by the value of the free-ligand concentration where the vertical axis is zero (i.e., at half-maximal binding). This plot is used primarily for analyzing cooperative binding (see Fig. 8.24C).

binding are defined by the temperature dependence of the binding affinity.

A general consequence of ligand binding is that the protein is stabilized against unfolding and is less flexible. Neither of these observations need imply that the ligand has altered the structure of the protein. Instead, they are simply a consequence of the ligand binding more tightly to the fully folded conformation (N) than to the fully unfolded state (U) and any distorted or partially unfolded forms that result from flexibility of the structure. This can be illustrated very simply for the case of unfolding of the protein when the ligand L binds solely to the folded state N:

$$\frac{L}{K_a} \rightleftharpoons \frac{N}{Y} \rightleftharpoons \frac{N}{L} \rightleftharpoons \frac{U}{L} \rightleftharpoons \frac{U}{L} \rightleftharpoons \frac{U}{L} \rightleftharpoons \frac{U}{L} \rightleftharpoons \frac{U}{L}$$

(8.11)

$$K_{app} = \frac{[N \cdot L] + [N]}{[U]} = K_u \left(1 + \frac{[L]}{K_a^2}\right)$$

(8.12)

The protein is stabilized against unfolding by the presence of the ligand. Even at very high ligand concentra-
tions, above those at which the folded protein is fully saturated, the apparent stability of the protein is increased in proportion to the concentration of free ligand.

Ligand binding is simple in dilute solutions, but proteins often function in extremely concentrated aqueous solutions, as in the cytosol. For example, the interior of the red blood cell is about 35% hemoglobin by weight. Such solutions are very nonideal. The pertinent equilibria must be expressed in terms of the thermodynamic activities of the protein and of the ligand, which can be very different from their concentrations. Even though a particular protein might not be present in high concentration, the presence of molecules other than water in the environment can lead to substantial excluded volume effects (see Fig. 7.2). Added molecules favor any conformational or binding reaction that leads to a more spherical shape of a protein molecule, with less surface area exposed to solvent (Sec. 7.1.1). Consequently, binding of a ligand to a protein is often considerably greater in a concentrated solution than might be expected. It is possible that most proteins in the cytosol usually exist bound to each other, to membranes, to cytoskeleton, or to some other organized structure.

References


8.2.2 Accounting for Relative Affinities

Can we account for the particular affinity of a protein for a given ligand? Do the structures of the two give any insight into why the affinity is high or low? A qualified yes is permissible in answer to these questions in certain cases. Generally, however, only qualitative conclusions are feasible. Yet it should be possible eventually to account for, or even to predict, binding affinities for ligands and to design ligands or proteins with useful binding properties. The implications are immense for chemotherapy and for drug design.

It is currently impossible to rationalize the values of $K_a$ or $K_d$ of any protein for any ligand, even when the structures of the complex and of the free components are known. The practical difficulties in rationalizing ligand affinities arise from our poor understanding of the energetics of protein structure (Chap. 7) and of the strengths of the basic interactions in aqueous solvent (Chap. 4). The observed affinity depends on the relative free energies of the complex and of the components. This includes not only the interactions between the two in the complex but also any changes in their average conformations and their flexibilities produced by complex formation, any differences in their various interactions with solvent, the loss of translational and rotational freedom of each component, plus the displacement of solvent and any other ligands present in the binding sites before formation of the complex. Many of these factors compensate each other, and the net observed effect is a small difference between several terms of large and uncertain magnitude. It is not yet possible to calculate these quantities accurately enough to predict the values of $K_a$ and $K_d$.

The relative affinities of two related ligands for the same protein (e.g., A and B), or of two closely related proteins for the same ligand, are more easily analyzed because many of the factors are the same in the two cases; differences in affinities can often be related to just one or two factors. Also, the ratio of their affinities is dimensionless:

$$\Delta (\Delta G^\circ)_{A-B} = -RT \ln \frac{K_A}{K_B} = +RT \ln \frac{K_B}{K_A}$$

There is then no complication in defining standard states. The most successful method for analyzing differences in binding interactions between closely related ligands or proteins is the free-energy perturbation method (Sec. 7.4.4.b) for simulating the effects of differences in chemical structure of the ligand or the protein, where the free energy is calculated in both the complex and the free molecule as the group that differs is gradually "mutated" during the calculation.

Interactions between a protein and a ligand always involve a substantial number of groups. The general approach to understanding the observed affinity has been to dissect it into the contributions of each group by measuring the effect on the affinity of removing then
individually. Varying the ligand has traditionally been easiest, but even closely related ligands are occasionally observed to bind to the same protein in very different ways; in this case, a comparison of their measured binding affinities is largely meaningless. It is now more acceptable to use the same ligand but to vary the protein, using site-directed mutagenesis (Sec. 2.2), because the structures of the variant proteins tend to remain more constant. It is still advisable, however, to determine that the ligand binds in the same way to the variant proteins.

Given a series of binding affinities of related ligands for the same protein or of variant proteins for the same ligand, how is the binding energy dissected? It might be thought that the total binding energy (Eq. 8.9) is simply the sum of the contributions of each group, but it is not that straightforward. This can be illustrated in a manner first presented by Jencks.

Consider a ligand composed of two parts A and B; A might be capable of hydrogen bonding, and B might be hydrophobic. The affinity of ligand AB is compared with the affinities of A and B separately:

\[
\text{AB} \quad + \quad K^{AB} \\
\text{A} \quad + \quad K^{A} \\
\text{B} \quad + \quad K^{B}
\]  

where the equilibrium constants \(K^{AB}, K^{A}, \) and \(K^{B}\) are for either association or dissociation. In general, there is no simple relationship among these constants, and the classical binding energies calculated from them using Equation (8.9) are generally not additive:

\[
-RT \ln K^{AB} \neq -RT \ln K^{A} - RT \ln K^{B}
\]  

even if the standard state is taken as 55 M so that unitary binding energies are calculated (Sec. 8.2.1). This nonadditivity is illustrated by the binding of biotin and some derivatives to the protein avidin (Table 8.2).

**Table 8.2 Binding of Biotin Derivatives to Avidin**

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Dissociation constant (M)</th>
<th>Free-energy contribution to binding (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotin</td>
<td>1.5 \times 10^{-15}</td>
<td></td>
</tr>
<tr>
<td>Desthiobiotin</td>
<td>5 \times 10^{-15}</td>
<td></td>
</tr>
<tr>
<td>CH_3-(CH_2)_4-CO_2^-</td>
<td>3 \times 10^{-3}</td>
<td>-10.7</td>
</tr>
</tbody>
</table>


The reason for the nonadditive nature of specific binding affinities becomes clear if the binding of AB is dissected into steps:

\[
\text{AB} \quad + \quad K^{AB} \\
\text{A} \quad + \quad K^{A} \\
\text{B} \quad + \quad K^{B}
\]  

The first step in the binding of the entire ligand should be analogous to the binding of each part when present
alone (Eqs. 8.15 and 8.16), so the first two steps are assigned the bimolecular binding constants of each part. The second step, binding of the second part of the ligand, however, is now a unimolecular step, rather than bimolecular. Consequently, the bimolecular binding constants $K^A$ and $K^B$ do not apply to the second step; instead, these second steps are assigned the unimolecular equilibrium constants $K^a$ and $K^b$. The values of these two constants are not independent because they are linked functions:

$$K^a = K^a K^b - K^a K^b (8.19)$$

From Equation (8.19), it is apparent that the constants $K^a$ and $K^b$ are the ratios of the binding constants of ligands with and without each of the respective moieties:

$$K^a = \frac{K^a}{K^b}, \quad K^b = \frac{K^b}{K^a} (8.20)$$

The contribution to the free energy of binding of each moiety can then be calculated:

$$\Delta G^a = -RT \ln K^a = -RT \ln \frac{K^a}{K^b} (8.21)$$

$$\Delta G^b = -RT \ln K^b = -RT \ln \frac{K^b}{K^a} (8.22)$$

For example, the free-energy contribution to binding of the sulfur atom of biotin can be estimated from the relative affinities of the first two compounds of Table 8.2 to be $-3.5$ kcal/mol:

$$\Delta G^f = -RT \ln \frac{5 \times 10^{-13} M}{1.3 \times 10^{-15} M} = -3.5 \text{ kcal/mol} (8.23)$$

Similarly, the contributions of the remaining five-membered ring and of the acidic hydrocarbon group can be estimated to be $-13.3$ and $-10.7$ kcal/mol, respectively (Table 8.2).

The incremental binding energy contributions calculated in this way give a measure of the increased affinity caused by the presence of each group of the ligand. Their values depend critically on the relationship between the two parts of the ligand during binding, that is, on the effective concentration of the second part of the ligand when the first is bound, which can be designated as $[A/B]$.

$$K^a = K^a [A/B] = \frac{[A/B]}{K^a} (8.24)$$

$$K^b = K^b [A/B] = \frac{[A/B]}{K^b} (8.25)$$

The same effective concentration applies to both parts because these are linked functions (see Eq. 8.19).

The effective concentration of either part in an intermediate complex can conceivably be zero, when it is kept away from the binding site and so provides no contribution to binding. Or, it might have a value of up to $10^{10} M$ (see Table 4.11), when parts A and B of the ligand are always in optimal orientation for simultaneous binding to a perfectly complementary binding site. The large values of effective concentrations in intramolecular reactions result from the entropic effect of the covalent linkage of the two parts. A ligand must lose a substantial amount of translational and rotational entropy upon binding; this is one of the factors determining the values of both $K^a$ and $K^b$. In the case of ligand AB, however, at least some of this entropy is lost when the first part is bound; the second part of the ligand is then fixed to some extent and need not lose as much entropy upon completion of the binding as if that part were binding by itself. Consequently, the greater the rigidity between the two parts of a ligand, the greater the entropic contribution to the effective concentration is likely to be. This entropic contribution is the primary reason that the binding contributions of individual parts of a ligand do not add to give the observed affinity. For example, the contributions to the binding of biotin by the sulfur atom, the five-membered ring, and the acidic hydrocarbon (Table 8.2) total $-27.5$ kcal/mol. In contrast, the free energy of binding that would be calculated from Equation (8.9) is only $-20.2$ kcal/mol. The difference between these values reflects primarily the greater entropy that must be lost when parts of a ligand bind as separate molecules relative to the entropy that must be lost when they bind as parts of the same molecule.

If neither the ligand nor the protein is strained by binding, very high effective concentrations and free-energy contributions to binding may be observed. For example, the data of Table 8.2 for the two halves of desulfobiotin imply that their effective concentrations in the hypothetical intermediate complex are $2 \times 10^5 M$ because

$$[A/B] = \frac{K^a K^b}{K^a K^b} = \frac{K^a}{K^a} (8.26)$$

With such high effective concentrations, ionic and hydrogen-bond interactions between ligand and protein may contribute substantially to binding, even in aqueous solution, where they must compete with intramolecular interactions between the solvent and the free protein and free ligand.

Because effective concentrations of the different parts of ligands are likely to vary substantially in different ligands and different binding situations, it is unrealistic to expect a constant contribution to binding of a hydrogen bond, a van der Waals interaction, and so on, in all ligand-binding interactions.
Table 8.5 Large Contributions to Ligand Affinities for Proteins

<table>
<thead>
<tr>
<th>Group of ligand</th>
<th>Free-energy contribution to binding to proteina (kcal/mol)</th>
<th>Free energy of transfer from water to nonpolar liquidb (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>—CH₃</td>
<td>-2.0 to -3.9</td>
<td>-0.5</td>
</tr>
<tr>
<td>—CH₂CH₃</td>
<td>-6.5</td>
<td>-1.0</td>
</tr>
<tr>
<td>—CH —(CH₃)₂</td>
<td>-9.6</td>
<td>-1.5</td>
</tr>
<tr>
<td>—CH₂—CH₂—CH₂—CH₃</td>
<td>-7 to -8</td>
<td>-2.6</td>
</tr>
<tr>
<td>—SCH₃</td>
<td>-4.9</td>
<td></td>
</tr>
<tr>
<td>—CH₂—CH₂—S—CH₃</td>
<td>-10 to -11</td>
<td>-1.3</td>
</tr>
<tr>
<td>—SH</td>
<td>-5.4 to -9.1</td>
<td></td>
</tr>
<tr>
<td>—OH</td>
<td>-8</td>
<td></td>
</tr>
<tr>
<td>—NH₂</td>
<td>-4.5</td>
<td></td>
</tr>
<tr>
<td>—NH₃⁺</td>
<td>-6.7</td>
<td></td>
</tr>
<tr>
<td>—CO₂⁻</td>
<td>-4.3</td>
<td></td>
</tr>
</tbody>
</table>


Some examples of large incremental contributions to binding by various groups, measured by the relative affinities of ligands that differ only in that group, are tabulated in Table 8.3. The values for the nonpolar groups are considerably greater than their free energies of transfer from water to nonpolar liquids, which often is considered an analogous process. This discrepancy is further evidence that a protein—at least, its binding site—is not equivalent to an organic liquid. Instead, the folded protein has a higher concentration of atoms, and a binding site for nonpolar groups probably presents a more rigidly defined cavity with greater van der Waals interactions than is possible for a liquid. If part of the ligand or the protein normally involved in binding is missing, there might be a void at the interface between protein and ligand. Such a void could be filled by an isolated solvent molecule, or the protein and ligand could adapt to attain complementarity; but both are energetically costly. If a polar group normally involved in hydrogen bonding is deleted, its partner can be left in an energetically unfavorable situation without an alternative group to hydrogen-bond to. Consequently, interpreting such binding data in terms of individual interactions is not straightforward.

The data of Table 8.3 demonstrate that a protein can discriminate very effectively between its proper ligand and a ligand that lacks just one small part. Discrimination of ligands containing extra groups can be even more powerful because additional groups can interfere sterically with the complementarity between ligand and binding site.

Nevertheless, there are limits to the specificity of binding that is possible, set by the energetics of the interactions between groups. These limits are exceeded in some instances for which extreme specificity is necessary: for example, in the replication, transcription, and translation of genetic information. DNA replication occurs with an error frequency of only $10^{-10}$ even though the tautomeration of the nucleic acid bases, which will cause incorrect base-pairing, occurs with a frequency of $10^{-3}$. Amino acids are also incorporated into proteins with considerably greater fidelity than expected, even from the data of Table 8.3. Much of that data comes from binding of amino acids to tRNA synthetases, which carry out the most crucial step of attaching the correct amino acid to the correct tRNA molecule. For example, how does a tRNA synthetase discriminate effectively against Gly when adding Ala to its tRNA? Val in the case of Ile, and Ser in the case of Thr? These pairs differ only by one —CH₂—group and might be expected (Table 8.3) to differ in affinity by only a factor of 10². The answer in this case is that the enzyme seems to check the amino acid twice, discriminating at the first binding step and then subjecting the selected
amino acid to a second check designed to detect the most likely fraudulent amino acid. Any caught by the second step are hydrolyzed from the tRNA and expelled (see Sec. 9.3.1.b). In such a double-sieve editing mechanism, the probability that an incorrect amino acid will be missed by both steps is the product of the two separate probabilities (e.g., $10^{-5} \times 10^{-5} = 10^{-10}$). In this way, biological specificities can be greatly enhanced over those possible with simple physical principles. Comparable multiple checks on specificity appear to be used in DNA replication. Such methods are used only when absolutely necessary, however, because there is a cost involved, in that a certain fraction of correct molecules are also removed at the subsequent recognition steps due to the intrinsic limitations on binding specificity.

References


8.2.3 Rates of Binding and Dissociation

The rates of binding and of dissociation of a ligand from a protein are determined by the respective rate constants $k_a$ and $k_d$:

$$ P + A \xrightleftharpoons[k_d]{k_a} P \cdot A $$  

(8.27)

Their ratio gives the association constant:

$$ K_a = \frac{k_a}{k_d} $$  

(8.28)

The rate constants for binding ligands to proteins vary considerably, depending on the sizes of both and on any conformational changes that must take place in each upon binding. Many small ligands are found to bind very rapidly, at rates approaching those expected for diffusion control, $k_D$. This expected rate can be estimated from the diffusion coefficients of the protein and ligand, $D_p$ and $D_A$, respectively, treating them as small spherical molecules that must approach within a distance $r_{PA}$ for binding to occur:

$$ k_D = 4\pi N_A (D_p + D_A)r_{PA} $$  

(8.29)

where $N_A$ is Avogadro’s number. For molecules with typical diffusion coefficients under normal circumstances (see Table 7.2), values of $k_D$ in the region of $10^9$ M$^{-1}$ s$^{-1}$ are expected. Larger molecules have smaller diffusion coefficients, but the value of $k_D$ does not decrease accordingly because the value of $r_{PA}$ is correspondingly larger. If the two molecules attract or repel each other at a distance, the term $r_{PA}$ in Equation (8.29) should be replaced by a term containing the energy of interaction as a function of distance. For example, electrostatic interactions are significant over substantial distances and, when favorable, can increase rate constants for association to $10^{13}$ M$^{-1}$ s$^{-1}$.

Rates of binding that are observed to be lower than $k_D$ imply either that the two molecules must be in defined orientations for productive binding to occur or that changes occur during binding to produce a multistep association reaction. Both phenomena are undoubtedly important with proteins. The binding sites on proteins usually bind ligands only in defined orientations, and they generally comprise only small fractions of the protein surface. Consequently, most encounters between ligand and protein would be expected to be unproductive, and association to be relatively slow, but there are exceptions. For example, cytochromes c are thought to transfer electrons to and from other proteins through only 0.6% of their surfaces, where the heme group is accessible (see Fig. 6.31), and only when the two proteins interact in very specific orientations. On this basis, the rate of their interaction would be expected to be lower than that for diffusion-controlled encounters by a factor of at least 1000, but it is not. The reason is thought to be that asymmetric distributions of charges on the proteins orient them so that they tend to approach each other rapidly in a productive manner. The charge distribution of horse cytochrome c indicates a large dipole moment of just over 300 Debye units, and the dipole axis passes through the presumed binding site. Electrostatic interactions have also been shown to guide charged ligands to their binding sites on other proteins.