ELSEVIER

Biophysical Chemistry 101–102 (2002) 309–319

Biophysical Chemistry

www.elsevier.com/locate/bpc

Molecular recognition—viewed through the eyes of the solvent

A. Ben-Naim*

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

Abstract

The binding of a ligand **L** to a specific site on an adsorbent molecule **P** is usually studied within the lock-and-key model. In this article we argue that the presence of an aqueous solvent may change dramatically the means by which a ligand recognizes its binding site. The main factor, originating from the specific property of water, is the ability of a water molecule to form a hydrogen-bond-bridge between a functional group on **L** and a functional group on **P**. This factor will change the criterion for selecting the best binding site.

Keywords: Functional groups; Gibbs energy; Binding site; Molecular recognition; Hydrophilic interactions

1. Introduction

Binding processes are ubiquitous in biological systems [1–4]. These range from binding of small molecules, like drugs to plasma proteins to binding of proteins to DNA. Most of these binding processes are highly specific with respect to the selection of the binding site.

How does a given ligand select the specific site, on a relatively large surface of the adsorbent molecule, on which it binds? This is the problem of molecular recognition.

Almost all the studies of the means by which a ligand molecule 'recognizes' the selected binding site have focussed on the *geometry*, or the *form*, of the binding partner. The geometrical fit between the ligand and the binding site lies at the heart of the well known lock-and-key model (Fig. 1a), proposed more than 100 years ago by Emil Fischer (1897) [1–3]. Modern drug-design methodologies are also based on the search for the best geometrical fits between ligands and potential binding sites [5–9].

Indeed, if one assumes that the major driving forces for binding are weak van der Waals forces [10], then the better the geometrical fit between the ligand and the site, the larger the binding constant. Hence, recognition of the best binding site is equivalent to finding the site at which the ligand–site interaction energy is the strongest (i.e. lowest energy of the system of the two binding partners).

This, very simple recognition model has evolved through several more complicated recognition models; one is by means of matching functional groups (FG), such as the one depicted in Fig. 1b. The second is the so-called induced-fit model [10]. The latter may be viewed as a generalization of the lock-and-key model where the adsorbent mol-

*Tel.: +972 2651 3742; fax: +972 265 85733.
E-mail address: arieh@batata.fh.huji.ac.il (A. Ben-Naim).

0301-4622/02/$ - see front matter
PII: S0301-4622(02)00165-5

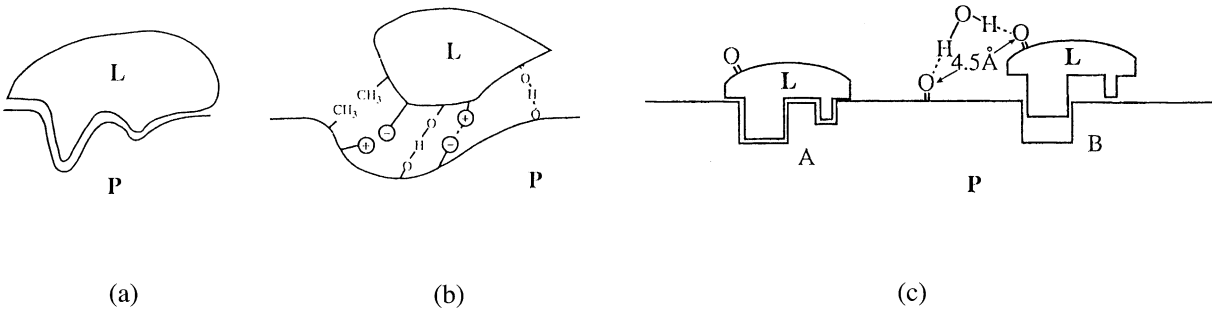

Fig. 1. (a) Lock-and-key model. Here the geometrical fit maximizes the van der Waals interaction between **L** and **P**; (b) Recognition by direct interaction through matching FG; (c) Possible switch of preference for binding sites. In vacuum, site A is preferred (lock-and-key model). In water, site B becomes preferred, due to the formation of hydrogen-bond-bridge between **L** and **P**.

ecule **P** is assumed to be an equilibrated mixture of conformers, each of which offering different binding sites to the ligand **L**. The process of binding induces a shift in the equilibrium distribution of conformers-favoring the conformers with the largest binding constant [4].

In all of these molecules, the driving forces for binding originates from *direct* interaction between the ligand **L** and the site on **P**. The criterion for selection of the best-binding-site is based on the interaction energy between the ligand and the *i*th site $\Delta U(i)$, namely

$$\min_i \Delta U(i) \tag{1.1}$$

$\Delta U(i)$ is the change in the potential energy of interaction when the ligand **L** is brought from infinite separation to the specific site i on **P**. (In more general cases both **L** and **P** can offer different binding sites). This criterion is used either explicitly or implicitly in most studies of molecular recognition in model compounds [11–13]. However, most binding processes in living systems occur in a solvent, the major component of which is water. In this case the criterion for selecting the best-binding-site, [Eq. (1.1)] should be replaced by

$$\min_i \Delta G(i) \tag{1.2}$$

where $\Delta G(i)$ is the Gibbs energy change for the same process of binding as described above.

While the ingredients that contribute to $\Delta U(i)$ are relatively well recognized and understood, the same cannot be said for $\Delta G(i)$. Only very recently an analysis of the ingredients that make up $\Delta G(i)$ has been undertaken [14,15]. Yet, still very little is known on the relative magnitude of the various ingredients that contribute to $\Delta G(i)$.

The solvent-induced effect on $\Delta G(i)$ is defined by the difference [14,15]

$$\delta G(i) = \Delta G(i) - \Delta U(i) \tag{1.3}$$

In this article we explore some of the solvent induced effects on molecular recognition. We shall see that switching from the energy criterion Eq. (1.1) to the free energy criterion Eq. (1.2), can lead to switching the preference for the best binding sites. Fig. 1c shows schematically this switch of preferences. In vacuum site A is preferred (by means of the lock-and-key model). In the presence of the solvent site B becomes preferred due to solvent induced effects, discussed in Section 2.

2. Recognition by indirect, solvent-induced, interactions

The general statistical mechanical expression for the solvent-induced contribution to $\Delta G(i)$ is [15]

$$\delta G(i) = -k_B T \ln\left[\frac{\langle \exp[-\beta B(\mathbf{LP}(i))]\rangle_o}{\langle \exp[-\beta B(\mathbf{L})]\rangle_o \langle \exp[-\beta B(\mathbf{P})]\rangle_o}\right] \tag{2.1}$$

Here, k_B is the Boltzmann constant, T, the absolute temperature and $\beta = (k_B T)^{-1}$. The quantity $B(\alpha)$

is the total interaction energy of a solute α with all solvent molecules (here, water molecules only) being at some specific configuration $\mathbf{X}^N = \mathbf{X}_1 \ldots \mathbf{X}_N$. This is defined as

$$B(\alpha) = \sum_{i=1}^{N} U(\mathbf{X}_\alpha, \mathbf{X}_i) \tag{2.2}$$

where $U(\mathbf{X}_\alpha, \mathbf{X}_i)$ is the pair potential between the solute α and the *i*th water molecule, being at some specific configuration $\mathbf{X}_i$ (this comprises both location and orientation coordinates specifying the configuration of the *i*th molecule).

In the denominator of Eq. (2.1) we have the binding energies $B(\mathbf{L})$ and $B(\mathbf{P})$ of the two solutes **L** and **P** before they bind. In the numerator we have the binding energy $B(\mathbf{LP}(i))$ of the bound pair **LP** at the specific site *i*. In all our discussions we assume that both **L** and **P** have a fixed conformation [16]. The average, denoted by $\langle\ \rangle_o$ in Eq. (2.1) is over all the configurations of the solvent molecules (here in the *T*, *V*, *N* ensemble, but in actual application the more appropriate *T*, *P*, *N* ensemble should be used). This is defined as

$$\langle \phi \rangle_o = \int \phi(\mathbf{X}_1 \ldots \mathbf{X}_N) P_0(\mathbf{X}_1 \ldots \mathbf{X}_N) \mathrm{d}\mathbf{X}_1 \ldots \mathrm{d}\mathbf{X}_N \tag{2.3}$$

where $P_0(\mathbf{X}_1 \ldots \mathbf{X}_N)$ is the densitry distribution for the configurations of all *N* solvent molecules, *before* the solute α has been inserted into the solvent, i.e.

$$P_0(\mathbf{X}_1 \ldots \mathbf{X}_N) = \frac{\exp\left[-\beta U_N(\mathbf{X}_1 \ldots \mathbf{X}_N)\right]}{\int \ldots \int \exp\left[-\beta U_N(\mathbf{X}_1 \ldots \mathbf{X}_N)\right] \mathrm{d}\mathbf{X}^N} \tag{2.4}$$

The entire expression in Eq. (2.1) is the *indirect* Gibbs energy change (i.e. $\Delta G(i) - \Delta U(i)$) for the process of binding at site *i*. Note that only the numerator in Eq. (2.1) depends on the site *i*, the two average quantities in the denominator depend only on the solvation of **L** and **P** in isolation, i.e. before binding.

In terms of solvation Gibbs energies, $\delta G(i)$ may be expressed as [15,17]

$$\delta G(i) = \Delta G^*(\mathbf{LP}(i)) - \Delta G^*(\mathbf{P}) - \Delta G^*(\mathbf{L}) \tag{2.5}$$

where $\Delta G^*(\alpha)$ is the solvation Gibbs energy of the solute α [17].

Clearly, since the average quantities in Eq. (2.1) involve integrals over *N* particles, (*N* being on the order of Avogadro number), it is difficult, if not impossible, to analyze the types and magnitudes of the different solvent-induced effects hidden within these integrals. Therefore, for the remainder of this section we shall focus onto a very simple 'solvent'; a single water molecule. This is clearly an extremely simple solvent, for which a detailed analysis of the integrals in Eq. (2.1) can be achieved. However, this is not a trivial 'solvent'. As we shall see below, the most important types of solvent-induced effects of real liquid water can be studied even for such a simple solvent. We shall comment on some missing contributions and their probable magnitudes—in real liquid water—at the end of this section. (Similar arguments were employed recently in examining the solvent-induced forces in the process of protein folding [18].)

Each of the average quantities in Eq. (2.1), when applied to the one-water-mole-solvent, reduces to

$$\langle \exp[-\beta B(\alpha)] \rangle_o = \int \exp[-\beta B(\alpha)] P_0(\mathbf{X}_w) \mathrm{d}\mathbf{X}_w \tag{2.6}$$

Since there is only *one* solvent molecule, the binding energy Eq. (2.2) reduces to

$$B(\alpha) = U(\mathbf{X}_\alpha, \mathbf{X}_w) \tag{2.7}$$

i.e. this is simply the solute–solvent pair interaction and the density distribution for solvent configurations Eq. (2.4) reduces to

$$P_0(\mathbf{X}_w) = \frac{\exp[-\beta U_1]}{\int \exp[-\beta U_1] \mathrm{d}\mathbf{X}_1} = \frac{1}{8\pi^2 V} \tag{2.8}$$

where *V* is the volume of the system (in the *T*, *V*, *N* ensemble) and $8\pi^2$ arises from the integration over all possible orientations of a single water molecule. The quantity $\mathrm{d}\mathbf{X}_w / 8\pi^2 V$ is the probability of finding the solvent molecule w at some configuration within the infinitesimal element $\mathrm{d}\mathbf{X}_w$ [15].

Now, each of the pair interactions in Eq. (2.7) for $\alpha = \mathbf{L}$, $\mathbf{P}$ or $\mathbf{LP}(i)$ can be written as consisting of three components

$$U(\mathbf{X}_\alpha,\mathbf{X}_w) = U^H + U^S + U^{HB} \tag{2.9}$$

where U^H is the strong repulsive part of the interaction energy between α at $\mathbf{X}_\alpha$ and the water molecule w, at $\mathbf{X}_w$. U^S is the 'soft', or the van der Waals, interaction between a water molecule and α, and U^{HB} is the hydrogen-bond (HB), part of the interaction, this part operates between a water molecule and FG on α, that can form HB's, such as carbonyl, carboyxl, amine, etc. We shall see below that this part of the pair potential provides the most important contribution to $\delta G(i)$ in liquid water.

Before analyzing the relative contributions of these three parts of the interaction, we shall rewrite any of the solvation quantities on the rhs of Eq. (2.5) in the form

$$\Delta G^*(\alpha) = \Delta G^{*H}(\alpha) + \Delta G^{*S/H}(\alpha) + \Delta G^{*HB/H,S}(\alpha) \tag{2.10}$$

Some details of the derivation of this identity are given in Appendix A, and more can be found in Refs. [15,17]. Here, we shall only give a qualitative reasoning underlying this identity for any solute α in a solvent (any solvents, not necessarily the simple one discussed in this section) for which the pair potential can be expressed as a sum of the form Eq. (2.9). The solvation Gibbs (or Helmholtz) energy, written in the form Eq. (2.10), can be rationalized as follows: suppose we rewrite Eq. (2.9) as

$$U(\mathbf{X}_\alpha,\mathbf{X}_w) = U^H + \varepsilon_1 U^S + \varepsilon_2 U^{HB} \tag{2.11}$$

We begin by 'switching off' the soft and the HB parts, i.e. by letting $\varepsilon_1 = \varepsilon_2 = 0$. In which case only the hard part of the potential is active. The solvation Gibbs energy of the hard part only is $\Delta G^{*H}(\alpha)$. This is the first term on the rhs of Eq. (2.10). Next, we 'turn on' the soft part U^S, by letting ε_1 increase from zero to unity. This turning-on of U^S would result in a change in the solvation Gibbs energy which we denote by $\Delta G^{*S/H}$. It can be shown (Appendix A) that this part is the same type of an average quantity as in Eq. (2.3) but instead of P_0 one must use the *conditional* density ($P(\mathbf{X}^N/H)$). In other words, $\Delta G^{*S/H}$ is the *conditional* Gibbs energy of solvation of the soft interaction, given that the hard part of the interaction has already been turned on. Similarly, we next 'turn on' the HB part of the pair potential, U^{HB}, by letting ε_2 change from zero to unity. The additional change in the Gibbs energy of solvation is denoted by $\Delta G^{*HB/H,S}(\alpha)$. This is the *conditional* solvation Gibbs energy of the HB part, *given* that the soft and the hard parts of the potential have already been turned on.

Substituting the solvation quantities $\Delta G^*(\alpha)$ in Eq. (2.10) for $\alpha = \mathbf{L}$, $\mathbf{P}$ and $\mathbf{LP}(i)$ in Eq. (2.5) we obtain the corresponding split of $\delta G(i)$ into the following three terms

$$\delta G(i) = \delta G^H(i) + \delta G^{S/H}(i) + \delta G^{HB/S,H}(i) \tag{2.12}$$

Now, we can analyze the various contributions to $\delta G(i)$, according to Eq. (2.12). We shall discuss the first two terms briefly, since these are the less important quantities. We shall devote more space to discuss the third term on the rhs of Eq. (2.12).

2.1. The hard part of $\delta G(i)$

Each of the solvation quantities of the hard part, for the special one-water-solvent, has the form Eqs. (2.6) and (2.8)

$$\exp[-\beta\Delta G_\alpha^{*H}] = \int \exp[-\beta U^H(\alpha)] P_0(\mathbf{X}_w)\mathrm{d}\mathbf{X}_w = \frac{8\pi^2(V - V_\alpha^{EX})}{8\pi^2 V} = 1 - \frac{V_\alpha^{EX}}{V} \tag{2.13}$$

where V_α^{EX} is the excluded volume for the water molecule produced by the solute α [15].

Combining the three expressions for $\alpha = \mathbf{L}$, $\mathbf{P}$, $\mathbf{LP}(i)$, we obtain

$$\delta G^H(i) = \frac{k_B T}{V}(V_{\mathbf{LP}(i)}^{EX} - V_{\mathbf{P}}^{EX} - V_{\mathbf{L}}^{EX}) = p\Delta V^{EX}(i) \tag{2.14}$$

where p is the pressure (of the ideal gas, $k_B T/V$) and $\Delta V^{EX}(i)$ is the change in the excluded volume upon binding at the site i. This quantity will always

be negative, i.e. the excluded volume of the pair **LP** at site i is always smaller than the sum of the excluded volumes of **L** and **P**, separately. Clearly, this term depends only on the sizes and forms of the particles involved and not on any specific property of water. When searching for the minimum of $\delta G^{H}(i)$ for different binding sites i, we do not expect that variation of $\delta G^{H}(i)$ would be large. Some more details on this quantity can be found in Ref. [15], p. 600.

2.2. The soft part of $\delta G(i)$

The second term on the rhs of Eq. (2.12) is due to turning-on the soft part of the solute–solvent interaction, given that the hard part of the interaction has already been turned on. One can show, using geometrical arguments as in the previous subsection, that this term depends on the surface aras of the various solutes, i.e.

$$\Delta G_{\alpha}^{*S/H} \sim -\varepsilon A_{\alpha} \tag{2.15}$$

where A_{α} is the surface area of the solute α and ε is some energy parameter, on the order of magnitude of the Lennard–Jones parameter between two small non polar solutes, say neon–methane, or methane–methane.The corresponding contribution to $\delta G(i)$ is

$$\delta G^{S/H}(i) = -\varepsilon[A_{\mathbf{LP}(i)} - A_{\mathbf{P}} - A_{\mathbf{L}}] \tag{2.16}$$

We expect that this term will be positive, and its variation with the site i will be small. More details on this term may be found in Ref. [15], p. 601.

2.3. The hydrogen-bond part of $\delta G(i)$

The third contribution to $\delta G(i)$ in Eq. (2.12) is due to turning on the HB contribution to the interaction energies between a water molecule and the various solutes involved in the binding process. This contribution is likely to be not only the largest in magnitude, but also the one that is specific to water as a solvent. As we shall argue below, this term will be decisive in the selection of the best binding site (through the eyes of the solvent). Fig. 2 shows schematically the interaction between a water molecule and four hydroxyl groups distributed over the surface of the solute **L**.

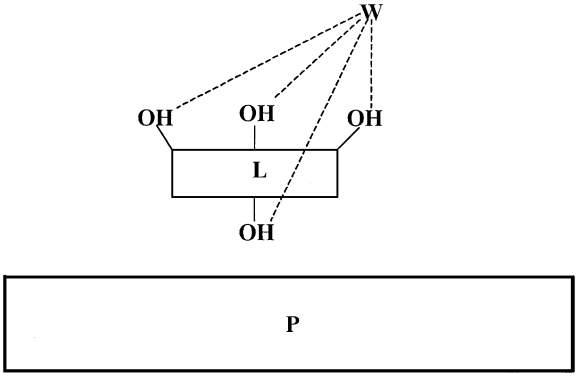

Fig. 2. Schematic representation of the interactions of a water molecule with the hydroxyl groups on **L**.

The corresponding contribution to the binding energy due to HB is written as

$$U_{\alpha}^{HB} = \sum_{k_{oH}} U^{HB}(k_{oH}, \mathbf{X}_{w}) \tag{2.17}$$

where $U^{HB}(k_{oH}, \mathbf{X}_{w})$ is the HB-interaction between the kth hydroxyl group and the water molecule at $\mathbf{X}_{w}$. Each of the terms on the rhs of Eq. (2.17) has a very short range and strongly depends on the distance and relative orientation between the water molecule and the hydroxyl group. This is exactly the reason for claiming that this part of the potential energy will be the most crucial for the selection of the specific binding site (by the water molecule).

To further analyze the general case of $\delta G^{HB}(i)$ would depend on the precise distribution of the FG on the surface of **L**, **P** and **LP**(i). We have recently made an approximate analysis of this kind to elucidate the factors that determine the relatively high solubility of globular proteins [19]. Here, we shall limit ourselves to one representative example shown schematically in Fig. 2.

The most important step in the analysis of δG^{HB} is to recognize three different types of FG's on the surface of **L** and **P**.

In Fig. 3, **L** is the same solute as in Fig. 2, but here we enumerate the hydroxyl groups on the surface of **L** by the numbers 1–4. In additon, we added some hydroxyl groups on the surface of **P**,

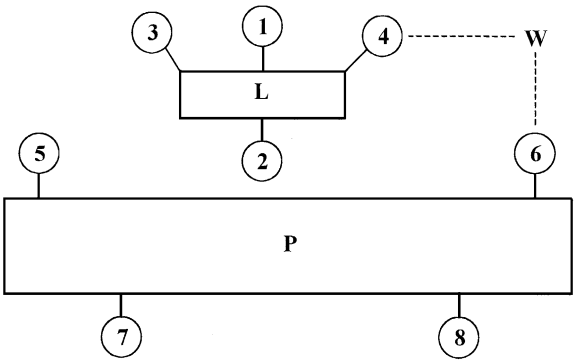

Fig. 3. Same as Fig. 2, but the hydroxyl groups on **L** are numbered 1–4. Some more hydroxyl groups are shown (numbered 5–8) on **P**.

numbered 5–8. We now distinguish between the following types of FGs on the surface of the solutes **L** and **P**.

2.3.1. The external FGs

Representatives of the external FG's are numbered 1, 7 and 8. The characteristic properties of these FG's is that their solvation does not change upon the formation of the pair **LP**. The qualitative reason is simple; a FG, such (1) in Fig. 3, is solvated by water (either real liquid water or the one-water-molecule-solvent). When **L** binds to **P** the conditional solvation Gibbs energy of this FG will not change upon binding, hence, the contribution of this FG to $\delta G(i)$ in Eq. (2.1) will cancel out. A similar cancellation will occur for any FG which is far away from the binding region between **L** and **P**. Hence the Gibbs energy of the binding process is not affected by these groups.

2.3.2. The internal FGs

We refer to the hydroxyl group numbered 2 in Fig. 3 as an *internal* FG. By internal we mean a FG that is fully solvated when **L** and **P** are isolated. Once **L** is bound to **P**, this FG becomes buried within the complex formed by **L** and **P**. Hence, after the binding process, this FG is not exposed to the solvent molecules. It is clear that upon binding, the conditional solvation Gibbs energy of any internal FG will be lost. This will have a large positive contribution to the quantity $\delta G(i)$. Formally, one can show that such groups contribute to the average quantities in the denominator of Eq. (2.1), but not the the numerator of Eq. (2.1) (for details, see Ref. [15], p. 596). Earlier estimates indicate that the loss of solvation of each FG, such as OH or C=O can contribute between -5 and -6 kcal/mol to $\delta G(i)$. Normally such groups will form direct HB with a FG on **P**, the energy of this HB is part of $\Delta U(i)$, not of $\delta G(i)$.

2.3.3. The joint FGs

Once we have defined and characterized FGs as external or internal, we can define all other FGs as belonging to the joint group. Examples of such FGs are the hydroxyl groups numbered 3, 4, 5 and 6. These FGs are solvated both *before* and *after* the binding process. However, the extent of solvation changes upon binding. (Note that the solvation of FG in the external region *does not* change upon binding. The solvation of FG in the internal region is *totally lost* upon binding.) This change could lead to either positive or negative contributions to $\delta G(i)$. A positive contribution will occur when the solvation of a FG say on **L** is partially hindered by a FG on **P**. It is negative when the solvation of a FG on **L** is enhanced by a FG on **P**, We have recently shown that such an enhancement of solvation will occur when a water molecule (even our one-water-molecule-solvent) forms a HB bridge between the two FGs. One water-bridge between FGs 4 and 6 is shown in Fig. 3. Numerical estimates indicate that under ideal conditions such a HB bridge could contribute approximately -2.5 to -3 kcal/mol [15]. This type of contribution has been referred to as hydrophilic interaction between the two hydrophilic (OH) FG. We believe that this particular solvent-induced effect is one of the most important factors that determine the strength of the binding Gibbs energy between **L** and **P**.

Of course, in the case of a one-water-molecule-solvent the possibility of finding a solvent molecule at the right location and orientation to form such a HB is very small. It is clearly much larger in a real solvent where the local density of water molecules around the FGs is much larger [15].

The two FG, say 3 and 5 in Fig. 3, that were independently solvated before the binding, become

correlated after the binding process. Clearly, there could be also a correlation between three and even four FGs due to the solvation by water molecules. (These are discussed in Ref. [15], p. 628 and 666.)

In the above discussion we have listed the main contributions of the solvation to the Gibbs energy of binding. We have used a very simplified solvent constituting a single water molecule. However, the main picture does not change much when dealing with a real solvent, at normal liquid densities. We have pointed out that the major difference between our simplified solvent and a real aqueous solvent is the considerably enhanced probability of finding a water molecule suitably located and oriented to form a HB bridge.

There is one important new effect in a real solvent that does not feature in our simple solvent. This is the possibility of the formation of two, or three, water molecule HB-bridges. Clearly, in our simplified solvent, HB-bridges can be formed only by a single water molecule. It has been shown, however, that HB-bridges involving two or more water molecules are not likely to contribute significantly to $\delta G(i)$ [15].

3. A simple two-dimensional example

Consider a ligand travelling along the 'surface' of an adsorbent molecule **P** in a two-dimensional system depicted in Fig. 4a. If only weak van der Waals forces are operating between **L** and **P**, then the interaction energy is determined by the size of the contact 'area' between **L** and **P**. Clearly the site D is the preferable binding site-this is also the site where the best geometrical fit between **L** and **P** is achieved.

Now, suppose that the surface and the ligand are featureless, as far as their geometry is concerned. This means that when the ligand **L** travels along the surface, the interaction energy is constant, and there is no preferential binding site. Now, we place some FG such as hydroxyl (OH) or carbonyl (C=O) along the surface, Fig. 4b. We distinguish between a FG that can be a donor (d) for HB, such as OH or NH, and an acceptor (a) for HB, such as C=O. The ligand **L** travelling along the surface will bind most strongly to the site at which the maximum number of HBs can be formed. Here, the best binding site is the site at which two *direct* HBs can be formed between the ligand **L** and the surface (this site is denoted A in Fig. 4b). If each HB contributes approximately 6 kcal/mol to the interaction energy, then the ratio of the probability of binding to the site where *two* HBs are formed relatively to the probability of binding to a site where only *one* HB is formed is

$$\frac{P(2\mathrm{HB})}{P(1\mathrm{HB})}=\exp\left[\frac{6}{0.6}\right]\approx 2\times 10^4 \tag{3.1}$$

Thus, for all practical purposes we can view the two-HB-site as the only binding site for this system. Note that in this example we have considered only the *pattern* of FGs on the surface, not the geometry of the surface. The recognition has been achieved by matching pairs of FGs on the interface between **L** and **P**.

Now we take the same system of **L** and **P**, with the same distribution of FGs as in Fig. 4b. We immerse the system in water, and consider again the potential sites for binding. There are two major modifications that result from switching from the energy criterion Eq. (1.1) to the free energy criterion Eq. (1.2). In the latter case, the best binding site is the one for which the interaction free energy is the lowest. The first modification applies to the direct HBs we have discussed above. Instead of the HB *energy*, estimated to be approximately -6 kcal/mol, we now have the HB *free energy*, estimated to be approximately -1.5 kcal/mol, clearly providing a far weaker interaction between the ligand and the surface [15]. The qualitative reason for this reduction in the interaction is simple. Before the binding occurs, each FG (either d or a) is fully solvated by the solvent. Upon binding, this solvation free energy is lost (partially or totally, depending on the extent of exposure to the solvent of the pair of FGs). Thus the gain of HB energy is partially compensated by the loss of the solvation free energy of the two FGs. A rough estimate indicates that the net free energy change per one direct HB is approximately -1.5 kcal/mol (more details on this can be found in Refs. [14,15]).

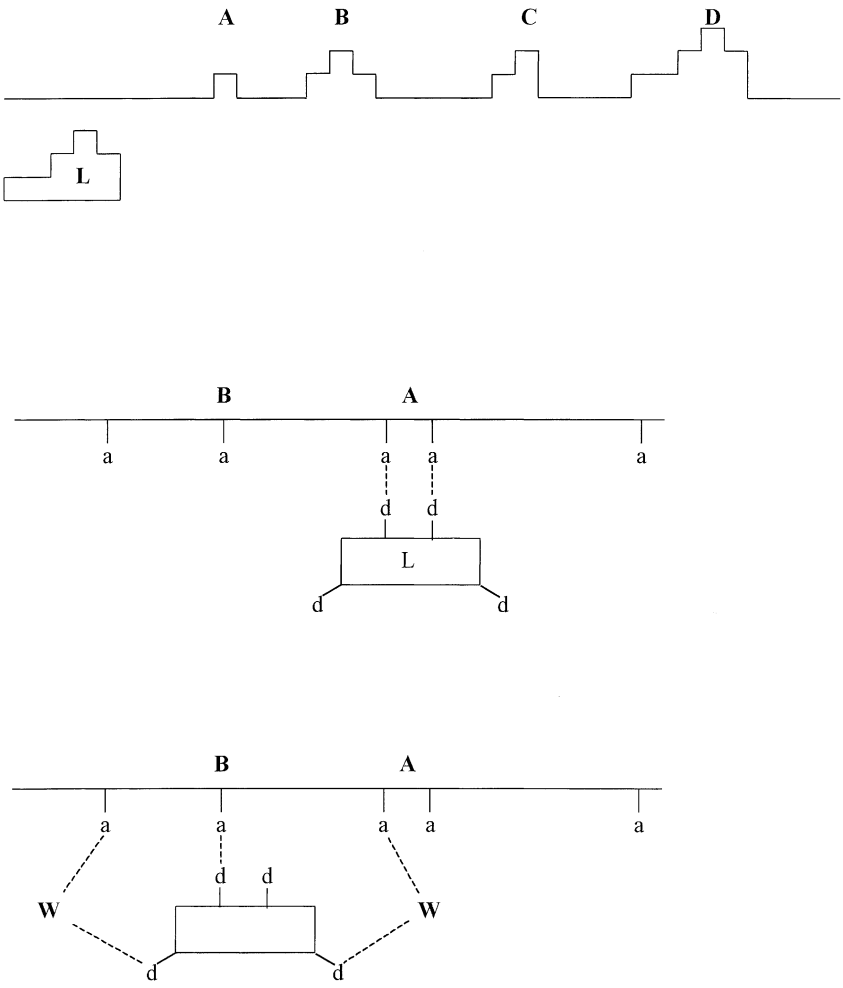

Fig. 4. (a) Selection of site D as a binding site by means of the lock-and-key model; (b) Selection of site A for binding by means of direct hydrogen-bonds between **L** and **P**; (c) Selection of site B for binding by both direct and indirect interactions.

The second modification is due to indirect interaction between FGs that are not in the interface between **L** and **P**. These are pairs of FGs that can interact by means of HB-bridge formed by a water molecule. Two such pairs of FGs are indicated in Fig. 4c. It should be noted that these pairs of FGs do not contribute to the *direct* interaction (since they are relatively far apart) and enter into play only in the presence of the solvent. The details of such an indirect interaction mediated by the solvent has been discussed in great in Refs. [14,15]. Here, we note only that each HB-bridge connecting two such FGs, under ideal configuration (distance and orientation of the FGs) can contribute between -2.5 and -3 kcal/mol to $\delta G(i)$, thus for the specific example of Fig. 4c we can estimate that in the absence of the solvent the best binding site (denoted A in Fig. 4b) is due to the formation of two direct HB, contributing approximately -12 kcal/mol to $\Delta U(i)$. In a solvent, the same site will have a binding free energy of interaction of approximately -3 kcal/mol. On the other hand, the site indicated B in Fig. 4c will contribute approximately -1.5 kcal/mol for the HB between the two FGs at the interface between **L** and **P**, and in addition, two HB-bridges will contribute

approximately −5 kcal/mol. Altogether the preference for binding to the new site (denoted B in Fig. 4c) relative to the previous one is approximately

$$\frac{P(2\text{ HB bridges}+1\text{ direct HB})}{P(2\text{ direct HB})} \approx \exp\left[\frac{+5+1.5-3}{0.6}\right]\sim 340 \tag{3.2}$$

Clearly, in the absence of solvent, site A in Fig. 4b in the preferred site for binding. In the presence of the solvent, there is a switch to the preference of site B relative to A.

It should be noted that in both cases the selection of the preferred site is determined by the distribution of FGs on the surface of **P** and **L**, the criterion for the selection of the best binding site is different when the direct interaction mode is operating and when indirect interactions are involved, i.e. recognition viewed via the eyes of the solvent. In the absence of the solvent we must count only the the number of matching FGs in the interface region between **L** and **P**. Viewed by the eyes of the solvent, we must count both matching groups in the interface, as well as HB-bridges formed by FG that are far from the interaction between **L** and **P**, i.e. FGs belonging to the joint region (Section 2).

4. Conclusion

Most of the theoretical work done on the factors involved in binding processes have focused on the direct interaction between the ligand and the binding sites. These interactions are at the heart of the lock-and-key model, either as in its original formulation or in one of the new variants.

The indirect, or the solvent-induced effect, has largely been neglected. The reason is that these effects are very difficult to analyze and it is almost impossible to estimate their magnitude. We have argued that solvent-induced effects could be large, and perhaps even decisive in selecting the best binding-site. We have discussed some of the solvent-induced effects for a highly simplified solvent consisting of a *single* water molecule. We believe that even in this simplified solvent we have captured most of the important solvent-induced effects. Of course, in a real binding process both the direct and the indirect interactions operate simultaneously in recognizing the best binding site.

For any specific binding system one should take into account both the geometry of the surface as well as the distribution of FG on both the ligand **L** and the adsorbent molecule **P**. This principle applies also for more complicated processes such as self assembly of subunits to form macro biochemical molecules [20].

Acknowledgments

This work was partially supported by a grant from INTAS number 1899, for which the author is most grateful.

Appendix A: Solvation and conditional solvation of Gibbs energies

In Section 2 we have written the solvation Gibbs energy of a solute α as consisting of three terms, corresponding to the three terms of the pair potential $U(\mathbf{X}_\alpha, \mathbf{X}_w)$ given in Eq. (2.9). We derive here the expression for the solvation Helmholtz energy of a solute α in any solvent having N solvent molecules. We assume that the intermolecular potential energy may be written as a sum of two contributions

$$U(\mathbf{X}_\alpha,\mathbf{X}_i)=U^{\mathrm{H}}(\mathbf{X}_\alpha,\mathbf{X}_i)+U^{\mathrm{S}}(\mathbf{X}_\alpha,\mathbf{X}_i) \tag{A.1}$$

The solvation Helmholtz energy of α is

$$\begin{aligned}\Delta G^*_\alpha &= -k_{\mathrm{B}}T\ln\left\langle \exp\left[-\beta\sum_{i=1}^{N}U(\mathbf{X}_\alpha,\mathbf{X}_i)\right]\right\rangle_{\mathrm{o}}\\ &= -k_{\mathrm{B}}T\ln\left\langle \exp\left[-\beta\sum_{i}U^{\mathrm{H}}(\mathbf{X}_\alpha,\mathbf{X}_i)\right]\exp\left[-\beta\sum_{i}U^{\mathrm{S}}(\mathbf{X}_\alpha,\mathbf{X}_i)\right]\right\rangle_{\mathrm{o}}\end{aligned} \tag{A.2}$$

The two factors under the average sign are not independent. Therefore, the average of the product

cannot be factored into products of two average quantities. Instead, we proceed by rewriting the average quantity in Eq. (A.2) as follows

$$\int \exp[-\beta(B^{\mathrm{H}}+B^{\mathrm{S}})]P_0(\mathbf{X}^N)\mathrm{d}\mathbf{X}^N = \frac{\int \exp[-\beta B^{\mathrm{H}}]\exp[-\beta U_N]\mathrm{d}\mathbf{X}^N}{\int \exp[-\beta U_N]\mathrm{d}\mathbf{X}^N} \frac{\int \exp[-\beta B^{\mathrm{S}}]\exp[-\beta B^{\mathrm{H}}-\beta U_N]\mathrm{d}\mathbf{X}^N}{\int \exp[-\beta B^{\mathrm{H}}]\exp[-\beta U_N]\mathrm{d}\mathbf{X}^N} = =\langle \exp[-\beta B^{\mathrm{H}}]\rangle_{\mathrm{o}}\langle \exp[-\beta B^{\mathrm{S}}]\rangle_{\mathrm{H}} \tag{A.3}$$

In Eq. (A.3) we used a shorthand notation for B^{H} and B^{S}, as the total binding energies due to the hard and the soft parts of the interaction energies, and U_N for the total interaction energies among all N solvent molecules, being at some specific configuration $\mathbf{X}_1 \ldots \mathbf{X}_N$. The first average on the rhs of Eq. (A.3) is the average over the hard part of the interaction energy, with the density distribution of the pure solvent, $P_0(\mathbf{X}^N)$, hence the subscript o. The second average is a conditional average, i.e. it is an average of the quantity $\exp[-\beta B^{\mathrm{S}}]$ with the conditional density $P(\mathbf{X}_1 \ldots \mathbf{X}_N/\mathbf{X}_\alpha^{\mathrm{H}})$ of all the configurations solvent molecules, *given* that the hard part of the solute–solvent interaction has already been turned on, hence, the subscript H under the average sign.

The conditional density used in Eq. (A.3) is obtained as follows: in a system of N solvent molecules and a single hard particle, the general density distribution is

$$P(\mathbf{X}^N,\mathbf{X}_{\mathrm{H}}) = \frac{\exp[-\beta U_N - \beta B^{\mathrm{H}}]}{\int \exp[-\beta U_N - \beta B^{\mathrm{H}}]\mathrm{d}\mathbf{X}^N \mathrm{d}\mathbf{X}_{\mathrm{H}}} \tag{A.4}$$

B^{H} is the binding energy of the hard particle to the rest of the system. The density distribution for finding the configuration $\mathbf{X}_{\mathrm{H}}$ of the hard particle is

$$P(\mathbf{X}_{\mathrm{H}}) = \frac{\int \exp[-\beta U_N - \beta B^{\mathrm{H}}]\mathrm{d}\mathbf{X}^N}{\int \exp[-\beta U_N - \beta B^{\mathrm{H}}]\mathrm{d}\mathbf{X}^N \mathrm{d}\mathbf{X}_{\mathrm{H}}} \tag{A.5}$$

Hence, the conditional density distribution is obtained from the ratio of Eqs. (A.4) and (A.5), i.e.

$$P(\mathbf{X}^N/\mathbf{X}_{\mathrm{H}}) = \frac{P(\mathbf{X}^N,\mathbf{X}_{\mathrm{H}})}{P(\mathbf{X}_{\mathrm{H}})} = \frac{\exp[-\beta U_N - \beta B^{\mathrm{H}}]}{\int \exp[-\beta U_N - \beta B^{\mathrm{H}}]\mathrm{d}\mathbf{X}^N} \tag{A.6}$$

This is the density distribution used in the second average quantity on the rhs of Eq. (A.3).

Translating Eq. (A.3) into solvation Helmholtz energy, leads to

$$\Delta G_\alpha^* = \Delta G_\alpha^{*\mathrm{H}} + \Delta G^{*\mathrm{S/H}} \tag{A.7}$$

In words, instead of solvating α in one step, we first solvate the hard part of the solute–solvent interaction. Then, we solvate the soft part—but now this is a conditional solvation, since the hard part of the interaction has already been turned on.

Clearly, one can extend the same argument for any number of terms in the pair potential such as three terms in Eq. (2.9), and for a more general case, as discussed in Section 2.

References

[1] P. Ritter, Biochemistry, A Foundation, Brooks/Cole Publ Comp, New York, 1996, p. 180.
[2] B.E. Tropp, Biochemistry Concepts and Applications, Brooks/Cole Publ Comp, New York, 1997, p. 179.
[3] L. Stryer, Biochemistry, W.H. Freeman and Comp, San Francisco, 1975, p. 122.
[4] A. Ben-Naim, Cooperativity and Regulation in Biochemical Processes, Kluwer Academic/Plenum Publishers, New York, 2001.
[5] I.D. Kuntz, Structure-based strategies for drug design and discovery, Science 257 (1992) 1078.
[6] T.J. Perun, C.L. Propst, Computer Aided Drug Design, Marcel Dekker, New York, 1989.
[7] C.L. Propst, T.J. Perun, Nucleic Acid Targeted Drug Design, Marcel Dekker, New York, 1992.
[8] H. Wang, A. Ben-Naim, A possible involvement of solvent induced interactions in drug design, J. Med. Chem. 39 (1996) 1531.
[9] C.E. Bugg, W.M. Carson, J.A. Montgomery, Drug design, Scientific Am. December (1993).
[10] By driving *force* one does not refer to the actual force operating between the binding partners. In most general cases it refers to the interaction free energy (rather than its gradient with respect to the distance) between the ligand and the site.
[11] J. Rebek, Molecular recognition: model studies with convergent functional groups, J. Mol. Recognition 1 (1988) 1.

[12] J. Rebek, Model studies in molecular recognition, Science 235 (1987) 1478.
[13] F.H. Stillinger, Z. Wasserman, Molecular recognition and self organization in fluorinated hydrocarbons, J. Phys. Chem. 82 (1978) 929.
[14] A. Ben-Naim, Solvent effects on protein association and protein folding, Biopolymers 29 (1990) 567.
[15] A. Ben-Naim, Statistical Thermodynamics for Chemists and Biochemists, Plenum Press, New York, 1992.
[16] Otherwise one needs to take appropriate averages over all conformations of the solute α. For more details, see Ref. [15].
[17] A. Ben-Naim, Solvation Thermodynamics, Plenum Press, New York, 1987.
[18] A. Ben-Naim, Hydrophobic–hydrophilic forces in protein folding, in: C.J. Cramer, D.G. Truhlar (Eds.), Structure and Reactivity in Aqueous Solution, ACS Symposium Series 568, 1994, p. 371, Chapter 25.
[19] H. Wang, A. Ben-Naim, Solvation and solubility of globular proteins, J. Phys. Chem. B101 (1997) 1077.
[20] A. Ben-Naim, On the role of water in molecular recognition and self assembly, Proc. Indian Acad. Sci. 98 (1987) 357.