

## On the role of water in molecular recognition and self-assembly

A BEN-NAIM

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

**Abstract.** The conventional concept of hydrophobic interaction is generalized to include any kind of solvent-induced effects on the binding of two or more solutes in aqueous solutions. Specifically, we focus on the role of hydrogen-bonding between the solutes and solvent molecules. A qualitative examination of the solute-solvent hydrogen-bonding effect on molecular recognition, self-assembly, and stabilization of biopolymers shows that these effects might be quite large and possibly more important than direct interactions between solute particles.

**Keywords.** Molecular recognition; self-assembly; role of water; solvent-induced effects; hydrophobic interaction.

### 1. Introduction

The importance of the presence of water in biological processes has long been recognized. However, very little quantitative work has been done to elucidate the role of water in these processes. One concept that has emerged from these studies is the concept of hydrophobic interaction (HI) (Kauzmann 1959; Tanford 1973; Ben-Naim 1980). In its most primitive sense, this concept was coined to describe the tendency of nonpolar molecules, or nonpolar side-chain groups on biomolecules, to adhere to each other, thereby minimizing their exposure to the solvent. The concept itself, although far from being understood on a molecular level, has been applied to many complex processes, such as binding of substrate to enzymes, association of subunits to form multisubunit enzymes, association of proteins to DNA, self-assembly of macromolecules, and even to interaction between entire cells. Clearly, these processes involve many factors in addition to the so-called HI, such as hydrogen-bonds (HB) and electrostatic interactions. Some of these originate from the interacting particles themselves, and others have their origin in the solvent. We shall, therefore, start in §2 with an exact separation of the two sets of factors. Next, we shall focus on the solvent-induced factors; these will also be separated into what we may call the conventional and nonconventional HI. This is described in §3. In §4 we further classify the various contributions to the HI according to the various functional groups (FG) and according to their location on the surface of the biomolecules.

In §5 we define the concept of molecular recognition, and in §6 we make one numerical estimate of the solvent-induced contribution to this process. Two possible extensions of the study of solvent effect, the self-assembly and protein folding, are briefly discussed in §§7 and 8.

## 2. Hydrophobic interactions and binding thermodynamics

The concept of HI has been used in the literature to describe various phenomena. Kauzmann coined the term hydrophobic bond to describe the tendency of nonpolar solutes (or of side-chain groups on a protein) to adhere to each other in aqueous media (Kauzmann 1959; Ben-Naim 1980). This term has evolved into HI and is currently used to describe the binding tendency of two or more nonpolar solutes in water.

A fundamentally different process, which is also included under the concept of HI or the hydrophobic effect (Tanford 1973), is the transfer of a nonpolar solute from a nonaqueous into aqueous solvent. The latter is more recently referred to as hydrophobic solvation (Ben-Naim 1980). Némethy and Scheraga (1962) argued that the formation of dimers by nonpolar solutes can be viewed as a partial reversal of the process of hydrophobic solvation and, therefore, the thermodynamics of the two concepts; HI and hydrophobic solvation should be simply related to each other. A close analysis of the thermodynamics of these two processes reveals that such a relationship does exist only in very special cases (Ben-Naim 1980) (e.g., the formation of a "dimer" at zero separation or aggregation of a very large number of solute particles to form a compact droplet in aqueous solutions).

Perhaps the simplest experimental model that has been used to study HI is the dimerization of a series of carboxylic acids in various solvents.

Schrier *et al* (1964) proposed a method of extracting information on HI from the standard free energy of dimerization of carboxylic acids, which we denote by  $\Delta G_B^0(R_n \text{ COOH})$ , where  $R_n$  is the alkyl chain with  $n$  carbon atoms.

Assuming that for each carboxylic acid one can write:

$$\Delta G_B^0 = \Delta G_{\text{HI}}^0 + \Delta G_{\text{HB}}^0, \quad (1)$$

where  $\Delta G_{\text{HI}}^0$  is the contribution due to HI, and  $\Delta G_{\text{HB}}^0$  is the contribution due to hydrogen-bonding. Furthermore, assuming that  $\Delta G_{\text{HB}}^0$  is common to all the carboxylic acids, and in particular, in the case of formic acid where we have only HB, i.e.,

$$\Delta G_B^0(\text{H COOH}) = \Delta G_{\text{HB}}^0, \quad (2)$$

One can extract information on the HI from  $\Delta G_B^0$  by subtracting the HB contribution, i.e.:

$$\Delta G_{\text{HI}}^0(R_n \text{ COOH}) = \Delta G_B^0(R_n \text{ COOH}) - \Delta G_B^0(\text{H COOH}). \quad (3)$$

However, a close examination of the content of the free energy of dimerization shows that theory does not support the above procedure (Ben-Naim 1980).

The standard free energy of dimerization (based on the molar concentration scale) may be written as

$$\Delta G_B^0 = \Delta U_B + \delta G + kT \ln (q_M^2/q_D), \quad (4)$$

where  $\Delta U_B$  is the direct binding energy between the two monomers,  $\delta G$  is the indirect or solvent-induced contribution to the binding free energy, and the last term on the right-hand side (RHS) of (4) contains the momentum and the rotational-vibrational partition functions of the monomers (M) and the dimer (D).

Clearly, even without any further analysis of the quantity  $\delta G$  (which we shall carry out in the next section), the form of  $\Delta G_B^0$  in (4) does not suggest a split of the form (1) into two contributions.

In 1971, a new definition of HI was suggested (Ben-Naim 1971, 1980). Recognizing the almost universal application of this concept to nonpolar solutes in aqueous solutions, it was argued that one should focus on  $\delta G$  as being a measure of the HI. The two other terms in (4) are presumed to be (approximately) invariant to changes of the solvent. Taking the extreme case of an ideal gas (i.g.) in which the same dimerization process is carried out, we find that:

$$\Delta G_B^{0,\text{i.g.}} = \Delta U_B + kT \ln (q_M^2/q_D), \quad (5)$$

i.e., in the ideal gas phase,  $\delta G = 0$ . Hence, from (4) and (5), one can extract the solvent-induced contribution, namely:

$$\delta G = \Delta G_B^0 - \Delta G_B^{0,\text{i.g.}}. \quad (6)$$

The general split of  $\Delta G_B^0$  in (4) is useful, since it does separate the driving force for binding into three factors. One depends on the direct interaction between the two monomers. This could include any type of direct interaction, such as van der Waal, HB, or electrostatic interaction. The third term on the RHS of (4) includes properties of the monomers and dimers, such as the mass, moment of inertia, vibrational frequencies etc. Both of these terms depend *only* on properties of the monomers and the dimers. In contrast, the quantity  $\delta G$  is a statistical mechanical quantity and involves averages over all configurations of the solvent molecules. This is essentially a many-body-interaction quantity, depending on the properties of the solvent and on the solute-solvent interaction. This feature of  $\delta G$  makes it more difficult to study by theoretical means.

For two simple solutes such as methane in water, an approximate measure of  $\delta G$  in terms of experimental quantities has been suggested (Ben-Naim 1971, 1980), namely,

$$\delta G = \Delta G_{\text{ethane}}^* - 2\Delta G_{\text{methane}}^*, \quad (7)$$

where on the RHS of (7) we have the solvation free energies of ethane and methane (Ben-Naim 1987). This quantity has been found useful and convenient for comparing various liquids as solvents.

Originally, the quantity  $\delta G$  has been suggested as a measure of HI in the conventional sense, i.e., for the indirect or solvent-induced part of the interaction between two *nonpolar* solutes in aqueous solutions. However, the significance of  $\delta G$ , as a solvent-induced contribution to  $\Delta G_B^0$ , is retained even when the binding solutes are not simple nonpolar particles, e.g., carboxylic acids, as discussed above, or proteins, as we shall discuss in subsequent sections.

Clearly, in the case of carboxylic acids, the conventional HI may be applied to describe the interaction between the nonpolar alkyl groups, as was suggested by Schrier *et al* (1964). However, since the entire molecules are certainly not nonpolar, we could not have used the term HI, in its conventional sense, to apply to the entire solvent-induced quantity  $\delta G$  as defined in (6).

Fortunately, the split of  $\Delta G_B^0$  in (4) holds true for any binding process (assuming that classical statistical mechanics can be applied to our systems). Therefore, we shall generalize the concept of HI to include any type of solvent-induced

contribution to  $\Delta G_B^0$ . We shall thus use the concept of HI in a nonconventional sense when it is applied to complex solutes. It will coincide with the conventional sense of HI when our solutes are simple nonpolar. The exact formalism that makes this generalization is discussed in the next section.

### 3. Formal split of $\delta G$ into conventional and nonconventional HI

Consider two solutes, say a protein P and a ligand L, at infinite dilution in water. In §2 we wrote the standard free energy of binding  $\Delta G_B^0$  as comprising three terms [see (4)]. We shall now focus on the solvent-induced quantity  $\delta G$ , which may be written as (Ben-Naim 1980):

$$\delta G = -kT \ln \left\{ \frac{\langle \exp[-\beta B_{PL}] \rangle_0}{\langle \exp[-\beta B_P] \rangle_0 \langle \exp[-\beta B_L] \rangle_0} \right\}. \quad (8)$$

Here,  $\beta = (kT)^{-1}$  with  $k$  the Boltzmann constant and  $T$  the absolute temperature.  $B_\alpha$  is referred to as the total binding energy of  $\alpha$  to the solvent, i.e.,

$$B_\alpha = \sum_{i=1}^N U(\mathbf{X}_\alpha, \mathbf{X}_i), \quad (9)$$

where  $U(\mathbf{X}_\alpha, \mathbf{X}_i)$  is the pair potential function for the interaction between the solute  $\alpha$  and the  $i$ th water molecule, the pair being at a specific configuration, denoted by  $(\mathbf{X}_\alpha, \mathbf{X}_i)$ . The symbol  $\langle \rangle_0$  denotes an average over all configurations of the solvent molecules in the TPN ensemble. In what follows, we shall use, for notational simplicity, the T, V, N ensemble. One can show that the difference between the Helmholtz and the Gibbs free energies is quite negligible for the processes discussed in this article (Ben-Naim 1987).

We also assume for simplicity that the solutes P, L and the complex PL are rigid molecules. Normally one should allow for some conformational flexibility of the molecule, in which case an additional average over all possible conformations must be carried out in (8) (Ben-Naim 1980, 1987).

The quantity  $\delta G$  may be interpreted as the indirect (or solvent-induced) part of the free energy change for the process of bringing P and L from fixed configuration at infinite separation to a fixed configuration at the final state of the complex, which we denote by PL.

Each of the average quantities in (8) has the form,

$$\langle \exp[-\beta B_\alpha] \rangle_0 = \int d\mathbf{X}^N \exp[-\beta B_\alpha] P_0(\mathbf{X}^N), \quad (10)$$

where  $P_0(\mathbf{X}^N)$  is the probability distribution function for finding a configuration  $\mathbf{X}^N = \mathbf{X}_1 \dots \mathbf{X}_N$  of all the solvent molecules in the *absence* of the solutes (i.e., in the pure solvent, hence the subscript 0 used in the notation  $\langle \rangle_0$ ). More explicitly,

$$P_0(\mathbf{X}^N) = \{\exp[-\beta U_N(\mathbf{X}^N)]\} / \{\int d\mathbf{X} \exp[-\beta U_N(\mathbf{X}^N)]\}. \quad (11)$$

In what follows, we assume that for each of the solute-solvent pair potentials we can write

$$U(\mathbf{X}_\alpha, \mathbf{X}_i) = U^H(\mathbf{X}_\alpha, \mathbf{X}_i) + U^F(\mathbf{X}_\alpha, \mathbf{X}_i), \quad (12)$$

where  $U^H$  is the hard-core interaction, i.e., the steep repulsive interaction between  $\alpha$  and the  $i$ th water molecule at a very short distance. The second term  $U^F$  includes all other interactions between  $\alpha$  and a water molecule, such as van der Waals, hydrogen-bond, and electrostatic. These arise from functional groups (FG) on the surface of  $\alpha$  (by surface, we mean any region of  $\alpha$  which is exposed to the solvent).

For some specific applications, we may further split  $U^F$  into a van der Waals term and a specific term due to, say, HB or electrostatic interactions, i.e.,

$$U^F(\mathbf{X}_\alpha, \mathbf{X}_i) = U^S(\mathbf{X}_\alpha, \mathbf{X}_i) + U^{\text{HB}}(\mathbf{X}_\alpha, \mathbf{X}_i) + \dots \quad (13)$$

Corresponding to this split of the pair potential, we have for the total binding energy of each solute  $\alpha$ ,

$$B_\alpha = B_\alpha^H + B_\alpha^S + B_\alpha^{\text{HB}} + \dots \quad (14)$$

From now on, we shall assume that the surface of the solute contains either nonpolar FG (such as methyl, ethyl etc.) or FG that form HB (such as hydroxyl, amine etc). We shall not discuss the case of charged FG (such as ionized-carboxyl or ammonium groups). Also, to simplify the notation, we combine  $B_\alpha^H$  and  $B_\alpha^S$  into one term  $B_\alpha^{\text{HS}}$  and rewrite  $B_\alpha$  simply,

$$B_\alpha = B_\alpha^{\text{HS}} + B_\alpha^{\text{HB}}. \quad (15)$$

Using (15) in each of the average quantities in (8) induces a factorization of each average term as follows,

$$\begin{aligned} \langle \exp[-\beta B_\alpha] \rangle_0 &= \int d\mathbf{X}^N \exp[-\beta B_\alpha^{\text{HS}} - \beta B_\alpha^{\text{HB}}] P_0(\mathbf{X}^N) \\ &= \frac{\int d\mathbf{X}^N \exp[-\beta U_N - \beta B_\alpha^{\text{HS}} - \beta B_\alpha^{\text{HB}}] \int d\mathbf{X}^N \exp[-\beta U_N - \beta B_\alpha^{\text{HS}}]}{\int d\mathbf{X}^N \exp[-\beta U_N - \beta B_\alpha^{\text{HS}}] \int d\mathbf{X}^N \exp[-\beta U_N]} \\ &= \langle \exp[-\beta B_\alpha^{\text{HB}}] \rangle_{\text{HS}} \langle \exp[-\beta B_\alpha^{\text{HS}}] \rangle_0, \end{aligned} \quad (16)$$

where the symbol  $\langle \rangle_{\text{HS}}$  signifies a *conditional* average, i.e., this is an average over all configurations of the solvent molecules, given that the hard (H) and soft (S) parts of the potential have been "turned on." This average employs the conditional distribution function,

$$P_{\text{HS}}(\mathbf{X}^N/\mathbf{X}_\alpha) = \{\exp[-\beta U_N - \beta B_\alpha^{\text{HS}}]\} / \{\int d\mathbf{X}^N \exp[-\beta U_N - \beta B_\alpha^{\text{HS}}]\}, \quad (17)$$

which is the probability density of finding a configuration  $\mathbf{X}^N$ , given that the hard and soft part of  $B_\alpha$  have been "turned on" at some specific configuration  $\mathbf{X}_\alpha$ .

The factorization (16) has a simple interpretation in terms of the solvation free energy of the solute  $\alpha$  (Ben-Naim 1967), i.e.,

$$\begin{aligned} \Delta G_\alpha^* &= -kT \ln \langle \exp[-\beta B_\alpha] \rangle_0 = -kT \ln \langle \exp[-\beta B_\alpha^{\text{HS}}] \rangle_0 \\ &\quad - kT \ln \langle \exp[-\beta B_\alpha^{\text{HB}}] \rangle_{\text{HS}} = \Delta G_\alpha^{*\text{HS}} + \Delta G_\alpha^{\text{HB/HS}}, \end{aligned} \quad (18)$$

where  $\Delta G_\alpha^{*\text{HS}}$  is the solvation free energy of the hard and soft parts of the

solute-solvent interaction, and  $\Delta G_{\alpha}^{*HB/HS}$  is the *conditional* solvation free energy of the HB part, given that the hard and soft parts have already been solvated.

Using the factorization (16) for each of the average quantities in (8), one can write  $\delta G$  as,

$$\delta G = \delta G^{HS} + \delta G^{HB/HS}, \quad (19)$$

where

$$\delta G^{HS} = -kT \ln \left\{ \frac{\langle \exp[-\beta B_{PL}^{HS}] \rangle_0}{\langle \exp[-\beta B_P^{HS}] \rangle_0 \langle \exp[-\beta B_L^{HS}] \rangle_0} \right\}, \quad (20)$$

and 
$$\delta G^{HB/HS} = -kT \ln \left\{ \frac{\langle \exp[-\beta B_{PL}^{HB}] \rangle_{HS}}{\langle \exp[-\beta B_P^{HB}] \rangle_{HS} \langle \exp[-\beta B_L^{HB}] \rangle_{HS}} \right\}, \quad (21)$$

Thus,  $\delta G^{HS}$  is the solvent-induced contribution to  $\delta G$  that arises only from the hard and soft (HS) parts of the solute-solvent interactions. The HB (or also charged groups, if they exist) are presumed to be "turned off." This part may be referred to as the *conventional* HI, since it is applied to two nonpolar solutes in water. The second term  $\delta G^{HB/HS}$  is the contribution due to turning on the HB part of the interaction. This is still a solvent-induced contribution to  $\delta G$  and may be referred to as a *nonconventional* HI. It is nonconventional only in the traditional usage of the term HI, but as we stressed above, it is well within the general definition of HI in terms of  $\delta G$  (Ben-Naim 1971, 1980).

Perhaps we should point out that the term HI, in the traditional sense, has been applied to hydrophobic solutes, i.e., solutes with relatively low solubility in water. The generalized concept of HI, as defined above in terms of  $\delta G$ , include also the case of very hydrophilic solutes, such as polyalcohols or hydrophilic proteins. Although we shall find in the next sections that these solutes could be effectively attracted to each other in aqueous solutions, such an attraction cannot conceptually be attributed to "phobia" toward water.

Clearly, for some specific applications, one may choose to split  $B_{\alpha}$  in different ways, e.g., to separate the hard and soft part or to add an electrostatic part to the solute-solvent interaction. The formalism developed above can be adapted with minor notational modification only.

#### 4. Classification of regions on the surface of the solutes

In the previous section, we have examined the formal split of  $\delta G$  into a conventional and nonconventional HI. In order to proceed with the examination of the different contributions to  $\delta G$ , it is convenient to distinguish between three different regions on the polymers.

Letting  $\alpha$  be either P or L, we assume that  $\alpha$  is a large protein or nucleic acid. We also assume that the solute-solvent pair potential may be written as,

$$U(\mathbf{X}_{\alpha}, \mathbf{X}_i) = U^H(\mathbf{X}_{\alpha}, \mathbf{X}_i) + \sum_{k=1}^M U^F(\mathbf{X}_k, \mathbf{X}_i), \quad (22)$$

where  $U^H(\mathbf{X}_{\alpha}, \mathbf{X}_i)$  is the hard-core repulsive interaction between the solute  $\alpha$  and the  $i$ th water molecule.  $U^F(\mathbf{X}_k, \mathbf{X}_i)$  is the interaction between the  $k$ th functional

group on the surface of  $\alpha$  and the  $i$ th water molecule. This function may include van der Waals, charge-charge, HB etc., interactions. The main distinction between the two terms of the RHS of (22) is that, while  $U^H$  depends on the entire solute  $\alpha$ , the second part  $U^F$  depends only on groups that are in contact with the water, and by definition, these groups belong to the surface of  $\alpha$ .

The binding energy of  $\alpha$  to the solvent can be written as,

$$B_\alpha = \sum_{i=1}^N U(\mathbf{X}_\alpha, \mathbf{X}_i) = B_\alpha^H + \sum_{k=1}^M B_{\alpha,k}^F, \quad (23)$$

where  $B_{\alpha,k}^F$  is the total binding energy of the  $k$ th FG on the surface of  $\alpha$  (excluding the hard-core repulsive interaction).

Using a similar procedure as in §3, we can write the solvation free energy of  $\alpha$  as,

$$\begin{aligned} \Delta G_\alpha^* &= -kT \ln \langle \exp[-\beta B_\alpha] \rangle_0 \\ &= -kT \ln \left\{ \langle \exp[-\beta B_\alpha^H] \rangle_0 \left\langle \exp \left[ -\beta \sum_{k=1}^M B_{\alpha,k}^F \right] \right\rangle_H \right\} \\ &= \Delta G_\alpha^{*H} + \Delta G_\alpha^{*F/H}, \end{aligned} \quad (24)$$

where  $\Delta G_\alpha^{*H}$  is the solvation free energy of the hard-core part of the interaction (essentially the work required to create a cavity in the solvent), and  $\Delta G_\alpha^{*F/H}$  is the conditional solvation free energy of all the FG on the surface of  $\alpha$ , given that the hard part of the interaction is already “turned on.” This quantity depends on the entire volume of  $\alpha$  through the condition “H” which essentially excludes the water molecules from the excluded volume of  $\alpha$  with respect to the solvent. On the other hand, the quantity  $B_{\alpha,k}^F$  depends only on the constituency of the surface of  $\alpha$ , i.e. on the FG that are in direct contact with the solvent. This separation is useful for the classification of the various sources of contributions to  $\delta G$  as discussed below.

We now classify all the FG on the surface of either P or L into three groups (see figure 1):

- (1) The region E (for external) includes all FG which are not affected by the binding of P and L to form the complex PL. Intuitively, any FG that is far away from the binding region between P and L will not “notice” the occurrence of the binding (in a sense discussed below).
- (2) The region I (for inner) includes all FG on either P or L which contribute to  $\Delta G_P^{*F/H}$  and to  $\Delta G_L^{*F/H}$  but not to  $\Delta G_{PL}^{*F/H}$ . Intuitively, these FG are exposed to the solvent when P and L are separated but are not exposed to the solvent in the complex PL.
- (3) The third region J (for joint) includes all FG that are neither in E nor in I. These are the FGs that *are* exposed to the solvent before and after the binding, but the extent of solvation might be different in the two states of the solutes P and L.

We now make the requirements for the classification of the FG more precise. First, for each of the solutes P and L, we write,

$$\begin{aligned} \langle \exp[-\beta B_{P,E}^F - \beta B_{P,I}^F - \beta B_{P,J}^F] \rangle_H &= \langle \exp[-\beta B_{P,J}^F] \rangle_H \times \\ &\quad \langle \exp[-\beta B_{P,E}^F - \beta B_{P,I}^F] \rangle_{H,P}, \end{aligned}$$

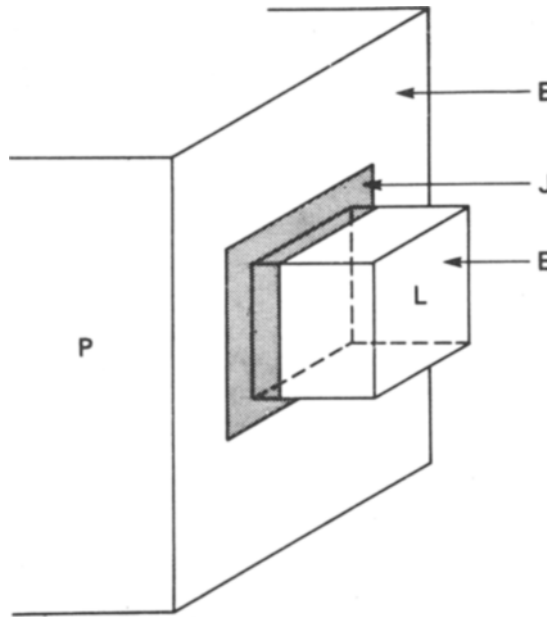


Figure 1. The three regions, E, I, and J, on the two solutes P and L.

$$= \langle \exp[-\beta B_{P,J}^F] \rangle_H \langle \exp[-\beta B_{P,E}^F] \rangle_{H,P} \langle \exp[-\beta B_{P,I}^F] \rangle_{H,P}, \tag{25}$$

and similarly

$$\begin{aligned} \langle \exp[-\beta B_{L,E}^F - \beta B_{L,I}^F - \beta B_{L,J}^F] \rangle_H &= \langle \exp[-\beta B_{L,J}^F] \rangle_H \times \\ &\quad \langle \exp[-\beta B_{L,E}^F - \beta B_{L,I}^F] \rangle_{H,L} \\ &= \langle \exp[-\beta B_{L,J}^F] \rangle_H \langle \exp[-\beta B_{L,E}^F] \rangle_{H,L} \langle \exp[-\beta B_{L,I}^F] \rangle_{H,L}. \end{aligned} \tag{26}$$

In (25) and (26) we first write  $B_\alpha^F$  as a sum of three contributions due to the FG in regions E, I, and J. Next, we transform it into a product of two conditional averages, as we have done in §3. Note that the conditions  $P_J$  or  $L_J$  means that the interaction due to FG in the J region (of P or L) are turned on. Finally, we assume that the regions E and I are independent, which leads to a factorization of the average over the product of  $\exp[-\beta B_{\alpha,E}^F] \exp[-\beta B_{\alpha,I}^F]$  into a product of two averages.

Similarly, for the PL complex, we write,

$$\begin{aligned} \langle \exp[-\beta B_{PL,E}^F - \beta B_{PL,J}^F] \rangle_H &= \langle \exp[-\beta B_{PL,J}^F] \rangle_H \times \\ &\quad \langle \exp[-\beta B_{P,E}^F - \beta B_{L,E}^F] \rangle_{H,P,L} \\ &= \langle \exp[-\beta B_{PL,J}^F] \rangle_H \langle \exp[-\beta B_{P,E}^F] \rangle_{H,P,L} \times \\ &\quad \langle \exp[-\beta B_{L,E}^F] \rangle_{H,P,L}. \end{aligned} \tag{27}$$



In the PL complex, by definition, there is no contribution from the region I to  $B_{PL}^F$ . We again transformed into conditional averages and assume that the E region on P and the E region on L are independent.

We can now combine (25), (26), and (27) and rewrite (8) as follows [note (24)],

$$\begin{aligned} \delta G &= \delta G^H + \delta G^{F/H} = [\Delta G_{PL}^{*H} - \Delta G_P^{*H} - \Delta G_L^{*H}] \\ &\quad - kT \ln \langle \exp[-\beta B_{PL,J}^F] \rangle_H \\ &\quad + kT \ln \langle \exp[-\beta B_{P,I}^F - \beta B_{P,J}^F] \rangle_H \langle \exp[-\beta B_{L,I}^F - \beta B_{L,J}^F] \rangle_H, \end{aligned} \quad (28)$$

where all the contributions due to region E are cancelled out. [Note that the conditions  $P_J$  in (25) and  $L_J$  in (26) are slightly different from the condition  $P_J L_J$  in (27), but we ignore this difference in cancelling the terms that include contributions from region E].

In (28) we have three distinctly different contributions to the solvent-induced quantity  $\delta G$ . The first is due to the change in the “cavities” of P and L upon the formation of the complex PL. This term will always be positive, since the excluded volume of PL is always smaller than the excluded volume of P and L separately. The second term on the RHS of (28) is the “gain” of solvation in the J region due to the binding. The third term is due to the “loss” of solvation of regions I and J of the separate solutes P and L.

If we can assume that the regions I and J are independent (or weakly dependent), we can rewrite (28) in a somewhat more convenient form, i.e.,

$$\begin{aligned} \delta G &= \delta G^H - kT \ln \left\{ \frac{\langle \exp[-\beta B_{PL,J}^F] \rangle_H}{\langle \exp[-\beta B_{P,J}^F] \rangle_H \langle \exp[-\beta B_{L,J}^F] \rangle_H} \right\} \\ &\quad + kT \ln \langle \exp[-\beta B_{P,I}^F] \rangle_H \langle \exp[-\beta B_{L,I}^F] \rangle_H. \end{aligned} \quad (29)$$

In this form, the last term contains only the “loss” of solvation of the region I, whereas the second term has the form of a correlation function. It measures the extent of dependence between FG in the region J on P and L in the state PL relative to the state when P and L are separated (and hence independent). We can rewrite (29) in a shorthand notation as follows,

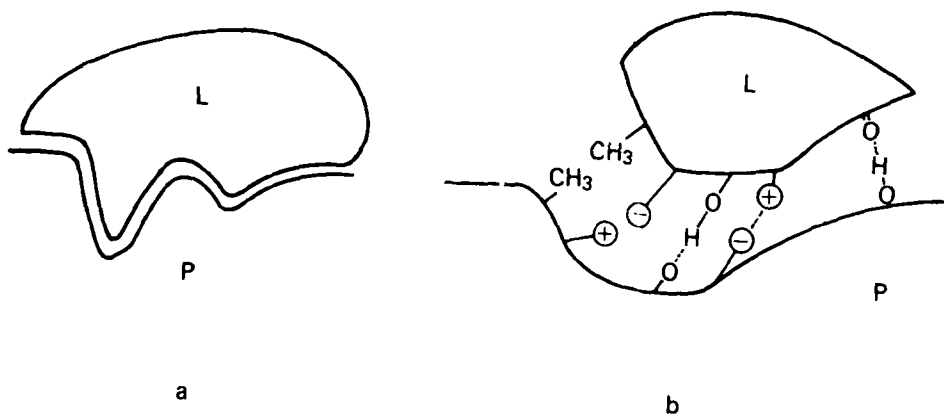
$$\delta G = \delta G^H + \delta G_J^{F/H} + \delta G_I^{F/H}, \quad (30)$$

where each term on the RHS of (30) denotes the corresponding terms on the RHS of (29).

## 5. The various contributions to molecular recognition

The concept of molecular recognition is usually discussed in terms of the so-called “key and lock” model (figure 2). This is essentially equivalent to the statement that the direct binding energy  $\Delta U_B$  to a particular site A (which is the one to be recognized) is minimal with respect to all other possible binding sites  $i$ , i.e.

$$\Delta U_B(A) = \min \Delta U_B(i). \quad (31)$$



**Figure 2.** (a) The classical "lock and key" model. Here, the geometrical fit maximizes the van der Waals interaction between P and L. (b) Various possible contributions to the direct interaction between P and L, such as van der Waals, hydrogen bonds, and electrostatic interactions.

Clearly, this statement is valid as a criterion for recognition if the binding process takes place in a vacuum. An interesting model for molecular recognition based on direct interactions has been presented by Stillinger and Wasserman (1978).

However, in a solvent the recognition criterion (31) should be replaced by

$$\Delta G_B(A) = \min_i \Delta G_B(i) = \min_i [\Delta U_B(i) + \delta G(i)], \quad (32)$$

where  $\Delta G_B(i)$  is the sum of the direct and indirect parts of the binding free energy at the site  $i^*$ .

The indirect part has been split into three terms [(30)]:

$$\delta G = \delta G^H + \delta G_j^{F/H} + \delta G_l^{F/H} \quad (33)$$

Each of the terms in (32) and (33) can contribute to the specificity of the molecular recognition. The simplest to visualize is clearly the role of  $\Delta U_B$ . The strongest interaction between P and L at a particular site would lead to recognition. This can be achieved either through the "key and lock" mechanism or by a specific pattern of hydrogen-bonding FG on P and L. Both of these effects can operate simultaneously to minimize  $\Delta U_B$ . The various contributions to  $\delta G$  are more subtle.  $\delta G^H$  depends essentially on the change of the excluded volumes of P and L upon the formation of the complex. Since binding will always reduce the excluded volume of P and L,  $\delta G^H$  is always negative and tends to enhance the association. This effect is similar to the "key and lock" mechanism. The tighter the binding between L and P, the larger is the reduction in the excluded volume and the larger the contribution to  $\delta G^H$ . In contrast to  $\Delta U_B$ , the term  $\delta G^H$  depends only on the overlapping of the excluded volumes of P and L but is independent of the type of the FG in the I regions of P and L.

\* In this article we assume that the concept of recognition is an equilibrium concept, i.e., the binding partners find each other under equilibrium conditions. It is, in principle, possible that kinetic effects other than, and independent of,  $\Delta G_B(i)$  will contribute to the recognition under nonequilibrium conditions (Yi-der Chen and Ben-Naim 1987).

The quantity  $\delta G_1^{F/H}$  is due to the loss of the solvation free energy of the FG in the region I of P and L. It is clear from (29) that the larger the binding energy of these FG to the solvent, the larger will be the tendency for dissociation. From the point of view of recognition, the site which will be most favoured for binding is the site that provides the weakest binding energy to the solvent. Qualitatively, this effect is in conformity with the experimental observation that binding sites are often characterized by being "hydrophobic," meaning that they are dominated by nonpolar FG, hence providing relatively weak interaction with the solvent.

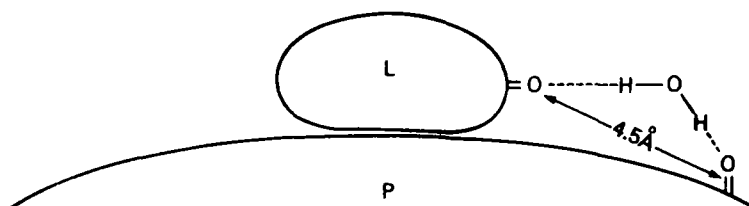
Perhaps the most subtle and potentially important contribution to  $\delta G$  is contained in  $\delta G_J^{F/H}$ . This term depends on the correlation between FG in the J region when P and L form the complex PL. One of the characteristic features of this quantity is that, if we have two FG – one on P and the other on L – which can be connected by a chain of water molecules through HB, then we should obtain a large negative contribution to  $\delta G$ . This is quite unique to liquid water because of the ability of a water molecule to form simultaneous HB with two FG on P and L. This contribution is highly specific in the sense that its value will strongly depend on the locations and orientations of these FG. In the next section we shall discuss some quantitative aspects of this particular effect on  $\delta G$ . In another paper, we have also suggested an experimental way of studying this effect by using small model-compound molecules (Ben-Naim *et al* 1987).

## 6. Estimates of the contribution of hydrogen bonding in the J region to molecular recognition

In §§4 and 5 we identified three different contributions to  $\delta G$ , each of which can be specific in the sense of contributing to the recognition of a particular binding site. In this and in the following sections, we shall focus on one of these contributions, namely  $\delta G_J^{F/H}$ , which we believe is the most sensitive to subtle changes in the locational and orientational distribution of FG. Furthermore, we also restrict ourselves to FG that can form HB with solvent molecules.

With these restrictions in mind, we rewrite the term  $\delta G_J^{F/H}$  from (30) in the modified form,

$$\delta G_J^{\text{HB/H}} = -kT \ln \left\{ \frac{\langle \exp[-\beta B_{\text{PL}}^{\text{HB}}] \rangle_{\text{H}}}{\langle \exp[-\beta B_{\text{P}}^{\text{HB}}] \rangle_{\text{H}} \langle \exp[-\beta B_{\text{L}}^{\text{HB}}] \rangle_{\text{H}}} \right\}, \quad (34)$$



**Figure 3.** Two carbonyl groups, one in the J region of P and the second in the J region of L, at a distance of 4.5 Å. The orientations of these two groups is such that they can form a bridge by means of HB to one water molecule.

where  $F$  is replaced by the superscript HB. For notational simplicity we also omit the subscript  $J$  on the RHS of (34).

We further assume that in the bound state PL, the  $J$  region contains pairs of FG, say hydroxyl or carbonyl groups, one on P and one on L, and that they are at a distance close to  $4.5 \text{ \AA}$  and oriented in such a way that they can form simultaneous HB with a single water molecule (figure 3).

Furthermore, if there are  $N_{PL}^{HB}$  such pairs of FG in this region, they are presumed to be far apart so that they can be treated as independent (figure 4).

We now write the total binding energy  $B_{PL}^{HB}$  as:

$$B_{PL}^{HB} = \sum_{i=1}^{N_{PL}^{HB}} B_i^{HB}, \quad (35)$$

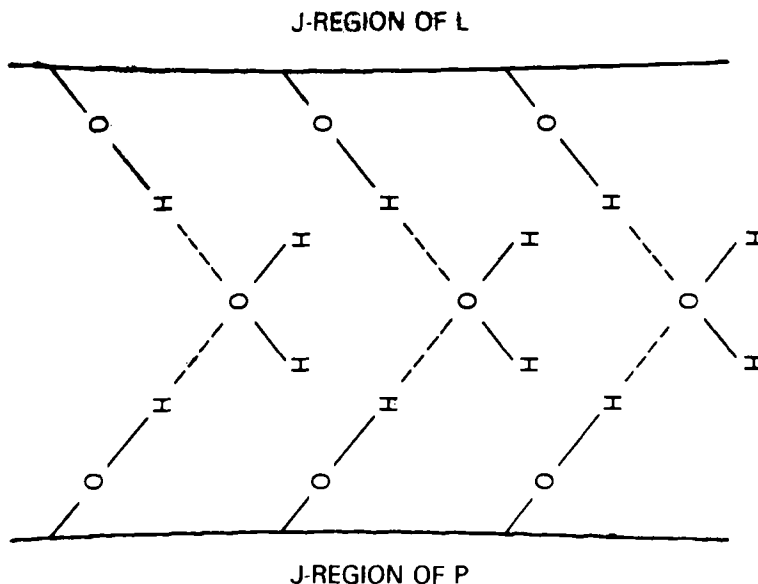
where

$$B_i^{HB} = \sum_{j=1}^N U_j^{HB}(\mathbf{X}_{PL}, \mathbf{X}_j). \quad (36)$$

Here,  $B_i^{HB}$  is the binding energy through HB of the  $i$ th pair of FG (in the  $J$  region), and  $U_j^{HB}$  is the HB part of the interaction between the  $i$ th pair and the  $j$ th water molecule at the specified configuration  $(\mathbf{X}_{PL}, \mathbf{X}_j)$ .

Following the assumption of independence of the pairs of FGs, each of the average quantities in (34) may be factored into  $N_{PL}^{HB}$  factors; thus, we obtain:

$$\delta G_J^{HB/H} = -kTN_{PL}^{HB} \ln \left\{ \frac{\langle \exp[-\beta B_{PL,i}^{HB}] \rangle_H}{(\langle \exp[-\beta B_{P,i}^{HB}] \rangle_H \langle \exp[-\beta B_{L,i}^{HB}] \rangle_H)} \right\}, \quad (37)$$



**Figure 4.** Pairs of FG along the  $J$  regions of P and L each of which can form a hydrogen-bonded bridge with one water molecule. The pairs are presumed to be far apart so that they act independently.

where we assume, for simplicity, that all FG have the same binding energy to the solvent; therefore, instead of summing over all kinds of FG, we simply take  $N_{PL}^{HB}$  identical terms in (37).

We now wish to estimate the contribution to  $\delta G_J^{HB/H}$  from each of the pairs of FG in the J region. Since the pair of FG is oriented as described in figure 3, only one water molecule can interact with this pair, either by forming one or two HB with this pair. Therefore, the entire range of configurations of the solvent molecule may be split as follows:

$$\int \exp[-\beta B_{PL,l}^{HB}] P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^N =$$

$$\sum_{i=1}^N \int \exp[-\beta U_i^{HB}(\mathbf{X}_{PL}, \mathbf{X}_i)] P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^N$$

ith water molecule  
hydrogen-bonded to l

$$+ \int P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^N$$

no hydrogen bonds to l

(38)

The first term on the RHS of (38) includes all configurations of the solvent molecules, for which one water molecule is hydrogen bonded to the lth pair of FG. Since only one water molecule can form such HB with the lth pair, one can simply rewrite this as N times an integral for which one specific water molecule, say i = 1, is hydrogen bonded to this pair. This integral can be further split into three integrals as follows:

$$\int = \int + \int + \int$$

water 1	water 1	water 1	water 1
hydrogen	hydrogen	hydrogen	hydrogen
bonded	bonded	bonded	bonded to
to l	to P only	to L only	both P and L

(39)

where the three terms on the RHS of (39) correspond to the three cases for which water number 1 is hydrogen bonded to either P, L, or to both P and L, respectively.

The second term on the RHS of (38) includes all configurations for which no water molecule is hydrogen bonded to the lth pair of FG. This may be written as\*:

$$\int P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^N = 1 - N \int \dots - N \int \dots - N \int \dots$$

no HB to l	water 1	water 1	water 1
	hydrogen	hydrogen	hydrogen
	bonded	bonded	bonded to
	to P only	to L only	both P and L

(40)

Combining (38) and (40), we have:

\* "... " means the same integrand as in the preceding integral.

$$\int \exp[-\beta B_{PL,l}^{HB}] P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^N = 1 + N \int \left\{ \exp[-\beta U_l^{HB}] - 1 \right\} P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^N$$

water 1  
hydrogen  
bonded to P only

$$+ N \int \dots \qquad + N \int \dots \qquad (41)$$

water 1  
hydrogen  
bonded to L only

water 1  
hydrogen  
bonded to  
both P and L

Since the range of configurations of a water molecule allowable for the formation of HB is very restricted, we can approximate each of the integrals on the RHS of (41) as follows:

$$N \int \left\{ \exp[-\beta U_l^{HB}] - 1 \right\} d\mathbf{X}_1 \int P(\mathbf{X}^N/\mathbf{X}_{PL}^H) d\mathbf{X}^{N-1}$$

water 1  
hydrogen  
bonded to P only

all  
configurations

$$= \int \left\{ \exp[-\beta U_l^{HB}] - 1 \right\} \rho(\mathbf{X}_1/\mathbf{X}_{PL}^H) d\mathbf{X}_1 \approx [\exp(10) - 1] \rho_w (\eta - \eta')$$

water 1  
hydrogen  
bonded to P only

(42)

On the last step on the RHS of (42), we have made the following approximations. Since the region of integration for which the integrand is nonzero is presumed to be small, we replaced the integral by a product of the value of the integrand at some point within the region of integration times the volume of this region. In this region  $-\beta U_l^{HB} \approx 10$  (for hydrogen bond energy at room temperature). The conditional density  $\rho(\mathbf{X}_1/\mathbf{X}_{PL}^H)$  is replaced by the density of water at room temperature  $\rho_w$ .  $\eta$  is the volume of the region for which a water molecule can form a HB with one FG. From this we subtract  $\eta'$ , the volume of the region for which a water molecule can form two HB with the two FG. See appendix A for estimates of  $\eta$  and  $\eta'$ . Applying similar approximations to each of the integrals in (41), we can write the entire average quantity in the numerator of (37) as:

$$\langle \exp[-\beta B_{PL,l}^{HB}] \rangle_H = 1 + 2\rho_w(\eta - \eta') [\exp(10) - 1] + \rho_w \eta' [\exp(20) - 1]. \qquad (43)$$

Similarly, for each of the average quantities in the denominator of (37), we write the approximation:

$$\langle \exp[-\beta B_{P,l}^{HB}] \rangle_H \approx \langle \exp[-\beta B_{L,l}^{HB}] \rangle_H \approx 1 + \rho_w \eta [\exp(10) - 1]. \qquad (44)$$

Note that in (44) we have the full range of configurations for which a water

molecule can form a HB with *one* FG. This is denoted by  $\eta$ . On the other hand, in (42) we put  $\eta - \eta'$  to take into account the reduction of this range of configurations when a second FG is present. For more details, see appendix A.

Thus, for the entire quantity in (37), we have the approximation:

$$\delta G_J^{\text{HB/H}} = -kTN_{\text{PL}}^{\text{HB}} \ln \left\{ \frac{1 + 2\rho_w(\eta - \eta') \exp(10) + \rho_w\eta' \exp(10)}{1 + 2\rho_w\eta \exp(10) + (\rho_w\eta)^2 \exp(20)} \right\}. \quad (45)$$

Taking the density of water at room temperature as  $\rho_w = 5.5 \times 10^{-2} \text{ mol cm}^{-3}$  and using the estimates of  $\eta = 5.74 \times 10^{-4} \text{ cm}^3 \text{ mol}^{-1}$  and  $\eta' = 20/360 \eta$  from appendix A, we may estimate the entire quantity  $\delta G_J^{\text{HB/H}}$  per pair of FG as:

$$\delta G_J^{\text{HB/H}}/N_{\text{PL}}^{\text{HB}} = -3.37 \text{ kcal/mol}. \quad (46)$$

Although this is a very crude estimate, it does indicate that the solvent-induced effect through hydrogen bonding might be of an order of magnitude similar to the direct hydrogen bonding between P and L. Clearly, for binding between two large proteins or between proteins and nucleic acids, we can expect many pairs of FG, each of which contributes about 3 kcal/mol to the total binding free energy.

Having established that the solvent-induced effect can provide a considerable contribution to the binding free energy, we may claim that in any real example this effect might operate either separately or in combination with the "lock and key" effect to produce a selective recognition of a binding site.

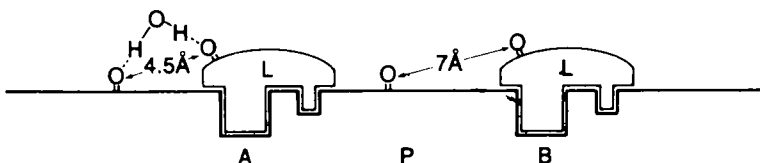
Figure 5 shows two identical sites on P, which from the point of view of the direct interaction (as well as  $\delta G^{\text{H}}$  and  $\delta G_i^{\text{1/H}}$ ) will have the same probability of binding a given ligand. However, if there is at least one pair of FG in the J region with the correct orientation, one of these sites, say A, may become more favourable for binding the ligand. Using the estimate we had for  $\delta G_J^{\text{HB/H}}$ , the ratio of the probabilities of binding to these two sites is:

$$\frac{y_A}{y_B} = \exp[-\beta\delta G(A) + \beta\delta G(B)] \approx \exp \left[ \frac{3.37}{0.592} \right] = 297. \quad (47)$$

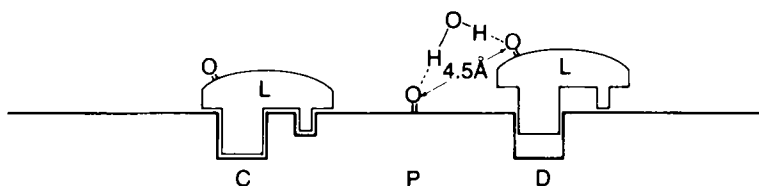
Thus, the site A is almost 300 times more likely to bind the ligand as site B. In this case we might say that L recognizes the site A, although this recognition is not obvious from the "key and lock" model.

Figure 6 shows a slightly different situation. Here, from the point of view of the direct interaction, site C will be considered more favourable (more contacts between P and L). However, one pair of FG in the J region may provide sufficient solvent-induced driving force that will make site D more likely to bind the ligand.

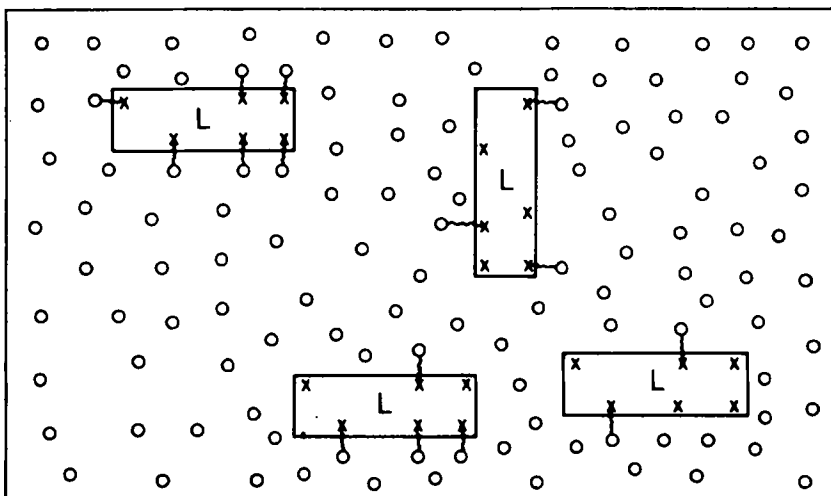
Another example of solvent-induced effect is shown in figure 7. Suppose that the



**Figure 5.** Two identical sites A and B on P. From the point of view of direct interaction, L will have the same binding energy with A and B. The indirect solvent-induced effect will favor binding of L to A.



**Figure 6.** Two different binding sites C and D. From the point of view of direct interaction, L will favor binding to C (better fit). However, the solvent-induced effect might give preference to binding to D.



**Figure 7.** FG on the surface of P are distributed randomly on the surface of P so that there is no specific pattern that can be recognized by L through direct interactions. However, indirect solvent-induced effects can give considerable preferences to certain sites (see §6 for estimates of relative probabilities to bind on different sites). Here and in figures 8, 9, and 10, an open circle represents a HB acceptor and  $\times$  represents a HB donor. A line connecting  $\times$  and  $\circ$  represents a bridge through one water molecule.

surface of the polymer P contains FG which are more or less uniformly distributed on its surface. There are no “locks” nor any specific patterns of FG that could be recognized by the direct interaction between L and P. Thus, from the point of view of the direct interaction, all possible sites on P are nearly equally probable for binding. However, if some of these FG can form bridges through water molecules, then some specific locations on the surface of P may become preferable. The probabilities of binding to the various sites indicated in figure 7, relative to binding to an arbitrary point on the surface of P, are:

$$P(2)/P(0) = \exp \left[ 2 \times \frac{3 \cdot 37}{0.592} \right] = 8 \cdot 8 \times 10^4,$$

$$P(3)/P(0) = \exp \left[ 3 \times \frac{3 \cdot 37}{0.592} \right] = 2 \cdot 6 \times 10^7,$$

$$P(4)/P(0) = \exp \left[ 4 \times \frac{3 \cdot 37}{0.592} \right] = 7 \cdot 7 \times 10^9,$$



$$P(5)/P(0) = \exp \left[ 5 \times \frac{3 \cdot 37}{0 \cdot 592} \right] = 2 \cdot 3 \times 10^{12}. \quad (48)$$

It is quite clear that even a few pairs of favorably oriented FG may lead to a very large solvent-induced preference to binding.

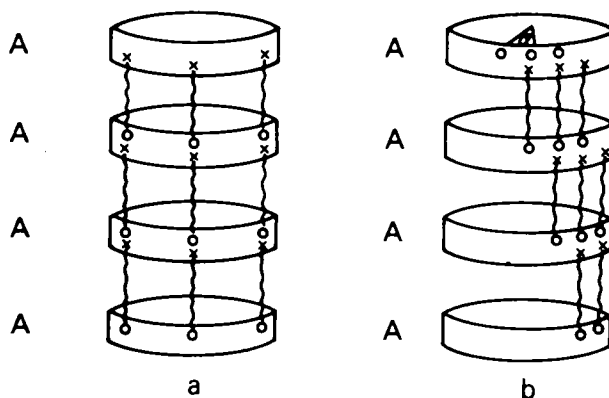
## 7. Solvent effect on self assembly

As noted in §5, there are a few distinctly different effects of the solvent on the binding thermodynamics. In §6 we chose one of these effects, which we believe can be very significant in the determination of the specificity of the binding. We now apply the same effect to describe other processes of binding and aggregation.

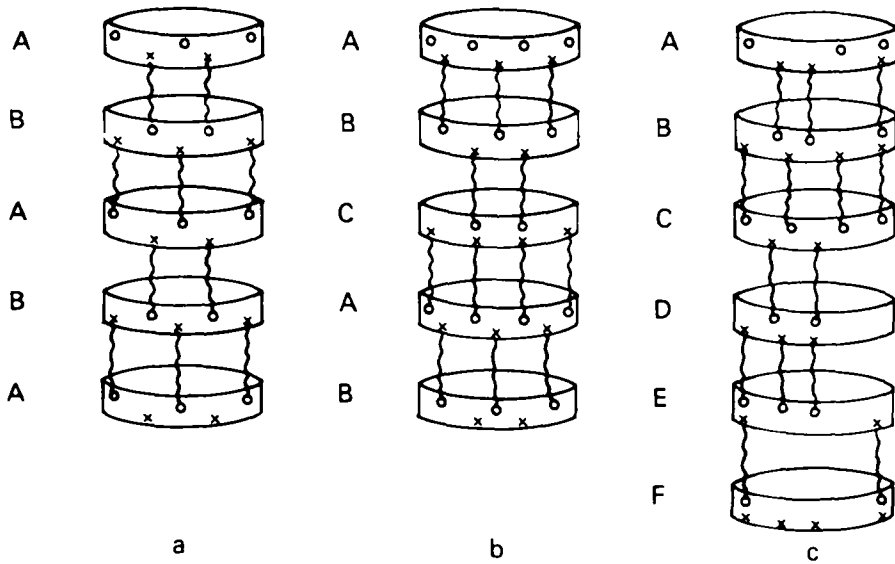
For simplicity, we assume that our building blocks are featureless, except for having FG in the J region that can form HB with the solvent. Two kinds of FG are assumed to occur, HB donors (denoted by  $\times$  in figures 7, 8, 9, and 10) and HB acceptors (denoted by  $\circ$  in the figures). These FG can form a bridge through one water molecule by means of two HB, presented in the figures by lines connecting  $\times$  to  $\circ$ . We also assume that pairs of FG,  $\times$  and  $\circ$ , are far apart and, therefore, operate independently.

Two examples of solvent-induced self-assembly are shown in figure 8. In (a), we have a homopolymer built up of a single subunit A, which has a given pattern of donor and acceptor FG that matches so that the solvent effect would enhance stacking of these monomers one on top of the other. In (b), we have a homopolymer of subunits A, but each monomer is rotated relative to its neighbour by an angle  $\alpha$ . This mode of packing may or may not be periodic depending on whether there is an integer  $n$  such that  $\alpha \times n = 2\pi$ .

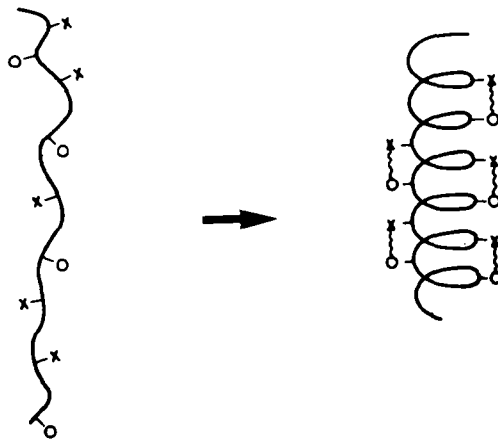
In figure 9 we demonstrate a few examples of heteropolymers. These may or may not have periodicity in the sequence of monomers (e.g., A recognizes B, B recognizes C, C recognizes A, and so on).



**Figure 8.** Self assembly of identical units to form homopolymers. (a) Stacking of identical monomers A in a regular manner to form a homopolymer. (b) Stacking of identical monomers. Each monomer is rotated by an angle  $\alpha$  relative to its adjacent monomer.



**Figure 9.** Various possible heteropolymers. (a) Here, monomer A recognizes B, B recognizes A, and so on. (b) Here, monomer A recognizes B, B recognizes C, C recognizes A, and so on. (c) Here, A recognizes B, B recognizes C, etc. No periodicity in the pattern of monomers.



**Figure 10.** Possible contribution to the stabilization of a certain conformation of a biopolymer by means of hydrogen-bonded bridges through water molecules.

Clearly, association can occur to form structures other than simple stacking as shown in figures 8 and 9. The important point that should be stressed here is that, even in cases when there is no apparent *direct* means of recognition, a pattern of FG on biopolymers can provide means for indirect recognition. We can say that such a pattern of FG store information which leads to a highly specific mode of packing. Furthermore, the “expression” of this information is achieved *only* in aqueous solutions. We note also that, in contrast to the direct recognition which depends *only* on the properties of the monomers, the indirect recognition can be controlled also by changes in the composition of the solvent. This gives us a finer

mechanism to regulate the extent of binding or self-assembly, a mechanism which certainly confers considerable advantages over the direct recognition and which is likely to be exploited by biological systems.

Although we have endeavored to show that recognition can be achieved by purely solvent-induced effects, it is likely that in real cases both direct and indirect effects will be operating simultaneously.

### **8. Solvent effect on protein folding and stabilization**

Much has been studied about the mechanism of protein folding by means of the direct intramolecular interactions. Also, the solvent effect has been introduced to explain the tendency of nonpolar groups to be either completely or partially removed from being exposed to the solvent by approaching each other. This effect falls within the conventional concept of HI. Here, we present also the possibility of solvent effect on the stabilization of a particular conformation due to bridges by means of water molecules. Figure 10 depicts one such example where acceptor and donor FG (denoted by  $\circ$  and  $\times$ , respectively) are brought to a distance of about 4.5 Å where they can form two HB to a water molecule. This effect (operating in combination with others) can contribute significantly to the stabilization of a particular conformation, an effect which will be highly specific to liquid water and, as in the binding process, could be controlled by changes in the composition of the solvent.

### **9. Conclusion**

The traditional concept of HI has been generalized to include any kind of solvent-induced effect on binding thermodynamics. We have seen that there are at least three distinctly different ways that the solvent can contribute to the binding free energy. Focusing on one of these, we have demonstrated that the solvent effect could be large and highly specific in molecular recognition processes.

Although direct recognition probably plays an important role in biochemical processes, it is inconceivable that nature would not have utilized the advantages of the solvent-induced effect. The latter, being highly specific to the composition of the solvent offers a fine way of controlling and regulating these processes. It can also be used to "switch preferences." For example, in the case of figure 6, changes in the solvent composition can change the ability of water molecules to form HB with the solute. Adding a solute which increases the ability of hydrogen bonding will give preference to site D; an opposite effect can be achieved by adding a solute that decreases the ability of hydrogen bonding, leading to preference of site C. Such a mechanism clearly has an evolutionary advantage over the conventional "lock and key" mechanism based on direct interactions.

Of course, much more effort should be expended to clarify the various contributions to the solvent-induced effect and to obtain more reliable estimates of their magnitudes. We hope that this paper will encourage further theoretical study of this topic.

**Appendix A**

*Estimates of the volumes of the configurational space of a water molecule forming one and two HB with a given FG*

In §6 we introduced the quantities  $\eta$  and  $\eta'$  to approximate some integrals over all the configurations of a water molecule that forms one or two HB with a given pair of FG. In a previous paper (Ben-Naim *et al* 1987) these quantities were estimated based on the experimental values of the second virial coefficient for water, which at room temperature is approximately:

$$B_2(T \approx 298 \text{ K}) = \frac{1}{16\pi^2} \int \{\exp[-\beta U_{ww}(\mathbf{X}_1, \mathbf{X}_2)] - 1\} d\mathbf{X}_1 \\ \cong -110 \text{ cm}^3 \text{ mol}^{-1}, \quad (\text{A1})$$

where  $U_{ww}$  is the water-water pair potential, and the integration is over all possible configurations of one water molecule (the second being at a fixed configuration denoted by  $\mathbf{X}_2$ ). Thus:

$$\int d\mathbf{X}_1 = \int dx_1 \int dy_1 \int dz_1 \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi \int_0^{2\pi} d\psi. \quad (\text{A2})$$

We now split the total range of integration in (A1) into two parts. The first is when  $R_{12} \leq \sigma_{ww}$  where  $\sigma_{ww} \sim 2.8 \text{ \AA}$  is the hard-core diameter of a water molecule. The second is over all configurations for which  $R_{12} > \sigma_{ww}$ . In this part we assume that the main contribution to the integral comes from those configurations for which the two molecules are hydrogen-bonded (and possibly additional van der Waals interaction).

Thus, we write (A1) as:

$$B_2 = -\frac{1}{16\pi^2} \int_{R_{12} \leq \sigma_{ww}} d\mathbf{X}_1 + \int_{R_{12} > \sigma_{ww}} \{\exp[-\beta U_{ww}(\mathbf{X}_1, \mathbf{X}_2)] - 1\} d\mathbf{X}_1 \\ = \frac{1}{2} \frac{4\pi\sigma_{ww}^3}{3} - \frac{8}{2} [\exp(11) - 1] \eta = -110 \text{ cm}^3 \text{ mol}^{-1} \quad (\text{A3})$$

The first term on the RHS of (A3) is essentially half the excluded volume between two water molecules. This is estimated to be:

$$\frac{1}{2} [4\pi(2.8)^3/3] = 45.99 \text{ \AA}^3 = 27.69 \text{ cm}^3 \text{ mol}^{-1}. \quad (\text{A4})$$

The factor 8 in the second term on the RHS of (A3) accounts for the fact that there are eight identical regions from which one molecule can form a HB with a second molecule. The maximum value of the integrand is taken to be  $\exp(11) - 1$  to account for both the HB energy ( $\beta U^{\text{HB}} \approx 10$ ) and van der Waals interaction ( $\beta U^{\text{dW}} \approx 1$ ). Hence, we have from (A3) and (A4):

$$\eta = \frac{110 + 27.69}{4 \exp(11)} = 5.74 \times 10^{-4} \text{ cm}^3 \text{ mol}^{-1} \quad (\text{A5})$$

The second quantity we want to estimate is the configurational volume of a water molecule forming two HB with the two FG as in figure 3. If the two FG are oriented ideally to form two HB, then the range of configurations for the coordinates  $x_1, y_1,$

and  $z_1$  allowable for the water molecule to form one HB or two HB will be the same. Regarding the rotational configurations, we choose the following three axes of rotation – the O-O axis between the water and one FG, the O-O axis between the water and the second FG, and the axis through the oxygen of the water and perpendicular to the plane containing the three oxygens. Clearly, rotation about one of the O-O axes by  $2\pi$  will not affect the HB energy for the HB along that particular axis. However, when we have a second FG and we require that the water will form a second HB with this FG, then this rotational range should be reduced from  $2\pi$  to some  $\Delta\alpha$ . Hence, the relation between  $\eta'$  and  $\eta$  is approximately:

$$\eta' = \eta \frac{\Delta\alpha}{2\pi} \approx \eta \frac{20}{360},$$

where the choice of  $\Delta\alpha \sim 20^\circ$  was taken as an average between two ranges (the so-called strong and weak HB) suggested by Dahl and Andersen (1983).

## References

- Ben-Naim A 1971 *J. Chem. Phys.* **54** 1387  
 Ben-Naim A 1980 *Hydrophobic interactions* (New York: Plenum Press)  
 Ben-Naim A 1987 *Solvation thermodynamics* (New York: Plenum Press) (in press)  
 Ben-Naim A, Ting K L and Jernigan R L 1987 (to be published)  
 Dahl L W and Andersen H C 1983 *J. Chem. Phys.* **78** 1980  
 Kauzmann W 1959 *Adv. Protein Chem.* **14** 1  
 Nemethy G and Scheraga H A 1962a *J. Chem. Phys.* **36** 3382  
 Nemethy G and Scheraga H A 1962b *J. Chem. Phys.* **36** 3401  
 Nemethy G and Scheraga H A 1962c *J. Phys. Chem.* **66** 1773  
 Stillinger F H and Waserman Z 1978 *J. Phys. Chem.* **82** 929  
 Schrier E E, Pottle M and Scheraga H A 1964 *J. Am. Chem. Soc.* **86** 3444  
 Tanford C 1973 *The hydrophobic effect: Formation of micelles and biological membranes* (New York: Wiley Interscience)  
 Yi-der Chen and Ben-Naim A 1987 (to be published)