Aspects of weak interactions

Dudley H. Williams* and Martin S. Westwell†

Cambridge Centre for Molecular Recognition, University Chemical Laboratory, Lensfield Road, Cambridge, UK CB2 1EW

Weak interactions, such as those non-covalent interactions that occur in biological systems, are less well characterised than their strong, covalent counterparts. Here, we discuss associations between two or more molecules and consider the effect of interactions with solvent molecules (particularly water) and changes in the internal structure of the associating molecules on binding. We go on to discuss some of the progress that has been made in the estimation of binding constants.

1 Introduction

In this article, we use the definition that 'weak interactions' are those involving bonds which are comparable to thermal energies. In such systems, equilibrium constants between two possible states can frequently be varied (often drastically) in the temperature range 0–100 °C. The study of weak interactions is a topic of great current interest because it is these non-covalent interactions which determine the stability, for example, of DNA duplexes, of the folded states of proteins, of enzyme–substrate complexes, and of ligand–receptor interactions. A deeper understanding of weak interactions is highly desirable not simply because of the intellectual drive to understand the above systems, but also because of the wish of the pharmaceutical industry to 'rationally' design new drugs. We largely discuss an approach developed in our laboratory, which builds on the earlier work of others (particularly Jencks).[1,2] We first discuss associations where in the ideal case the associating entities do not interact strongly with solvent, nor do they change their structures significantly upon association. Second, we go on to consider associations in polar media, particularly water, and where the associating molecules may modify their internal structures upon binding. Lastly, we discuss some of the progress that has been made in the estimation of binding constants.

† *Current Address*: Dyson Perrins Laboratory, University of Oxford, South Parks Road, Oxford, UK OX1 3QR.

2 Model 1: Where A and B do not interact strongly with solvent, nor do they extensively modify their structures upon association

Perhaps the most commonly used expression for the representation of a reversible association is eqn. (1).

$$\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{A}\cdot\mathbf{B} \qquad (1)$$

This kind of formalism has arisen because it is very useful in considering the association of, for example, two water molecules to give a hydrogen bonded dimer. To a good approximation, each water molecule (one designated **A** and the other **B**) retains the same internal structure in the dimer as in the dissociated state. Thus, as a useful approximation, the bonding between **A** and **B** in **A·B** can be represented as a property of the interface between **A** and **B**. The approximation works well because the bonds within **A** and **B** are strong compared to the bonding between **A** and **B**. In such circumstances, we have previously argued[3,4] that the restriction of motion which occurs when the association of eqn. (1) takes place (measured in terms $T\Delta S$, where T is the temperature, 298 K, and ΔS is the overall loss of entropy of translation and rotation) will be related to the exothermicity of the association (ΔH) by a curve of the general form shown in Fig. 1.

The direction of curvature arises because motional restriction reaches a limit of $T\Delta S \approx 50$ to 60 kJ mol^{-1} for the immobilisation of a molecule of mass of *ca.* 100–300 at room temperature, and this limiting entropy loss is approached much before covalent bond strengths are reached. The general shape of this curve has subsequently received theoretical support,[5] and we reproduce here experimental data for associations occurring in methylene chloride solution (Fig. 2).[4] As a good approximation, the species involved in these associations do not interact strongly with the solvent, nor do they modify their structures on association. Since $\Delta G = -RT\ln K = \Delta H - T\Delta S$, the direction of curvature satisfies the requirement that in these simple systems the equilibrium constant for association will increase with increasing exothermicity of association. More importantly, the curve qualitatively emphasises how the adverse entropy of

Dudley Williams

Dudley Williams has bachelor and PhD degrees, in chemistry and organic chemistry, respectively, from the University of Leeds in 1958 and 1961. He subsequently studied as post-doctoral fellow at Stanford University from 1961–64. Since 1964, he has worked at the University of Cambridge, where he is a Fellow of Churchill College and Professor of Biological Chemistry. He was elected a Fellow of the Royal Society in 1983.

Martin Westwell

Martin Westwell was born in 1971 in Lancashire. He obtained his undergraduate and doctoral degrees from Churchill College, Cambridge University. After a further year working in the Chemistry Department with Professor Williams, Dr Westwell has moved to Oxford University where he is the Glaxo Junior Research Fellow in Biological or Medical Sciences at Lincoln College.

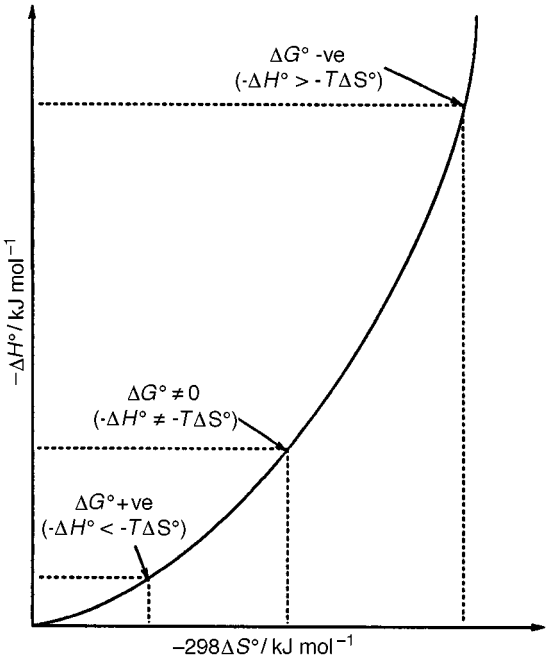


Fig. 1 The general form of the extent of the exothermicity of association ($\Delta H°$) $\mathbf{A} + \mathbf{B} \rightarrow \mathbf{A{\cdot}B}$ as a function of the entropic cost ($\Delta S°$) at a temperature of 298 K. There is a limit in the price in entropy to be paid (due to loss of translational and rotational freedom) and this limit is approached before covalent bond strengths are approached.

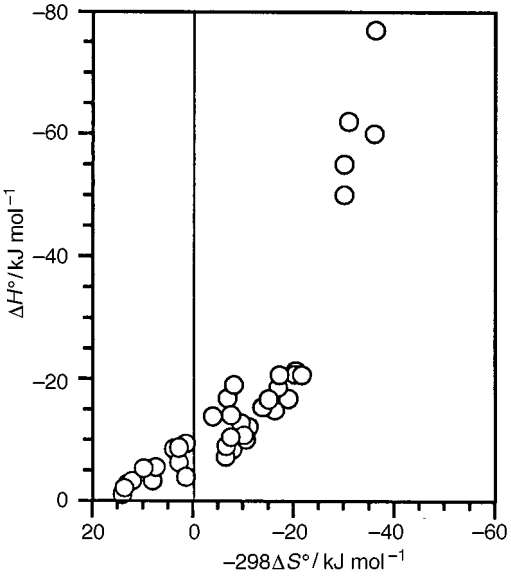


Fig. 2 Enthalpy ($\Delta H°$) *vs.* entropy ($298\Delta S°$) for the association of macrocycles with neutral molecules in dichloromethane. Data collected by Izatt *et al.*[39] Very weak exothermic associations have favourable entropies due to desolvation [an effect which is absent for an idealised non-polar solvent (or gas phase) (Fig. 1)].

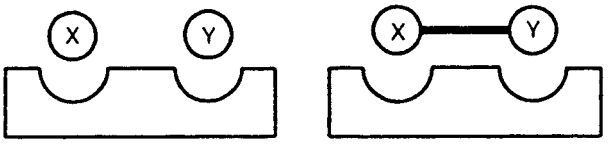
association increases gradually as a function of increased bonding of the associating entities, *i.e.* how increased bonding gradually restricts the dynamic motion in **A·B**.

3 Problems in trying to obtain the free energy of binding of specific groups in networks of weak interactions

Much effort has gone into attempts to estimate free energies of binding for common types of weak interactions [*e.g.* the hydrophobic effect ($Å^{-2}$), amide–amide hydrogen bonds, and salt bridges] as they occur as parts of a network of weak interactions. To understand the problems inherent in such an approach, let us first consider binding in a non-polar solvent where there are no internal rotations to be restricted on binding. A favourable free energy contribution to binding (ΔG_p) can occur for any pair of electrostatic interactions (*e.g.* amide–amide hydrogen bond formation) found in the binding site. It is possible to consider the sum of all such pairs of interactions ($\Sigma\Delta G_p$). These binding terms are opposed by the entropic cost of reducing the overall motion of the ligand when it binds to its receptor. This is described as ΔG_{t+r}, the adverse free energy change due to loss of entropy of translation and overall rotation. Thus, if the observed free energy of binding is ΔG_{obs}, an attempted partitioning can consider the approximation eqn. (2).

$$\Delta G_{obs} = \Delta G_{t+r} + \Sigma\Delta G_p \quad (2)$$

The reason for believing that eqn. (2) might serve as a useful approximation derives from a limiting case considered in 1981 by Jencks.[2] Consider that two species **X** and **Y** can associate, separately as at left, or connected together by a strain free connection as at right, into two distinct binding sites of a receptor (Scheme 1). In the ideal and limiting case, **X**, **Y**, and **X–Y** all lose all their translational and (overall rather than internal) rotational entropy on binding to the receptor; this corresponds to the value of $T\Delta S \approx 50$ to 60 kJ mol^{-1} for immobilisation referred to above. Thus, in applying eqn. (2) to all three possible binding events (of **X**, **Y**, and **X–Y** to the receptor) the adverse value of ΔG_{t+r} would be the same 50 to 60 kJ mol^{-1} in all three cases.



Scheme 1

Let us hypothetically (and unreasonably—see later) select the binding constants of **X** and **Y** as $K_X = K_Y = 10^3$ dm^3 mol^{-1}, and take the loss of entropy as the limit, $\Delta G_{t+r} \approx 57$ kJ mol^{-1} (because each 5.7 kJ mol^{-1} opposes binding by a factor of 10 at room temperature). The price in entropy only has to be paid once in the binding of **X**, **Y**, or **X–Y**. Applying eqn. (2), the intrinsic binding affinity (the limiting ΔG_p value, defined by Jencks[2] as the binding affinity expressed when association occurs without adverse entropy) of both **X** and **Y** can be derived. When **X** (or **Y**) binds alone to the receptor, we know that $\Delta G_{obs} \approx -17.1$ kJ mol^{-1} ($K = 10^3$ dm^3 mol^{-1}) and $\Delta G_{t+r} \approx +57.0$ kJ mol^{-1}, such that eqn. (3) holds.

$$\Delta G_p(\mathbf{X}) = \Delta G_{obs} - \Delta G_{t+r} = -17.1 - 57 = -74.1 \text{ kJ mol}^{-1} \quad (3)$$

Now, if we consider the binding of **X–Y**, the lost entropy is again in the limit $\Delta G_{t+r} \approx +57.0$ kJ mol^{-1} and the hypothetical ΔG_{obs} from eqn. (2) is given by eqn. (4).

$$\Delta G_{obs}(\mathbf{X–Y}) = \Delta G_{t+r} + \Sigma\Delta G_p = \Delta G_{t+r} = +\Delta G_p(\mathbf{X}) + \Delta G_p(\mathbf{Y})$$
$$= +57.0 - 74.1 - 74.1 = -91.2 \text{ kJ mol}^{-1} \quad (4)$$

Thus, if $K_X = K_Y = 10^3$ dm^3 mol^{-1}, then from the above analysis, $K_{X-Y} \approx 10^{16}$ dm^3 mol^{-1}.

Although this example illustrates the principle of an extreme case, it is also clear that it involves an assumption which is physically completely unrealistic. Any association with $K = 10^3$ dm^3 mol^{-1} would not occur with complete loss of translational and rotational entropy, but rather with only partial loss of this entropy. The contribution of Fig. 1 towards an understanding of the problem is that a semi-quantitated form would allow a crude guideline to the entropic cost (due to loss of translational and rotational entropy) as a function of the strength of the electrostatic interaction formed. That is, for a gas phase association the horizontal axis of Fig. 1 is the ΔG_{t+r} term

and the vertical axis is the ΔG_p (or $\Sigma\Delta G_p$) term (although not the limiting ΔG_p term as defined by Jencks). We see how, in qualitative terms, the term opposing association (ΔG_{t+r}) by restricting motion is played off against the term promoting association [ΔG_p (or $\Sigma\Delta G_p$)] by favourable bonding interactions. The gas phase description is carried over to the case of association in non-polar solvents as a useful approximation. Since even a non-polar solvent will always interact finitely with the associating species that are dissolved in it, the nature of the approximation is exposed by the fact that the plot does not pass through the origin in Fig. 2. Where **X** and **Y** associate with very low exothermicity, the adverse ΔG_{t+r} term for the association is very small, and is more than offset by the favourable entropy of release of solvent. The curve therefore tails into the $T\Delta S > 0$ of Fig. 2, very weakly exothermic associations occurring with a net favourable entropy.

Let us now consider some physically plausible cases in terms of the entropy–enthalpy compensation curve. In the situation we consider first (solid lines in Fig. 3), the simplifying assumption is made that the exothermicity of the association of **X**–**Y** with the receptor is simply the sum of the exothermicities when **X** and **Y** bind separately ($\Delta H_{X-Y} = \Delta H_X + \Delta H_Y$). Suppose **X** alone binds with an exothermicity of 20 kJ mol^{-1}, and **Y** with an exothermicity of 50 kJ mol^{-1}. On the basis of Fig. 3, the respective adverse $T\Delta S_{t+r}$ terms would be approximately 20 and 38 kJ mol^{-1}, giving rise to ΔG_X and ΔG_Y of 0 and 12 kJ mol^{-1} ($K_X = 1$ and $K_Y = 1.3 \times 10^2$ dm^3 mol^{-1} at 298 K). Taking the exothermicity of binding of **X**–**Y** as 70 kJ mol^{-1}, from Fig. 3 the cost in $T\Delta S_{t+r}$ is 46 kJ mol^{-1}, giving ΔG_{X-Y} as 24 kJ mol^{-1} ($K_{X-Y} = 1.6 \times 10^4$ dm^3 mol^{-1}). Thus, in this case of **X** binding with a relatively small exothermicity, and **Y** binding with a moderate exothermicity (case 1, Table 1), the enhancement of binding constant of **X**–**Y** relative to its separate components (expressed as K_{X-Y}/K_XK_Y) is *ca.* 10^2. In contrast, if the exothermicities of binding of **X** and **Y** are respectively 50 and 70 kJ mol^{-1}, then the respective $T\Delta S_{t+r}$ values for **X** and **Y** are estimated as 38 and 46 kJ mol^{-1}, giving rise to ΔG_X and ΔG_Y of 12 and 24 kJ mol^{-1} ($K_X = 1.3 \times 10^2$ and $K_Y = 1.6 \times 10^4$ dm^3 mol^{-1} at 298 K); **X**–**Y** binds with an exothermicity of 120 kJ mol^{-1}, and an adverse $T\Delta S_{t+r}$ value of 50 kJ mol^{-1} ($T\Delta S_{t+r}$ is approaching its limit), giving ΔG_{X-Y} as 70 kJ mol^{-1} ($K_{X-Y} = 1.9 \times 10^{12}$ dm^3 mol^{-1}). In this case of **X** binding with a moderate exothermicity, and **Y** binding with a larger exothermicity (case 3), the enhancement of binding constant of **X**–**Y** relative to its separate components (expressed as K_{X-Y}/K_XK_Y) is *ca.* 9×10^5 dm^3 mol^{-1}. The most important of these data, along with the analysis of one where the **X** and **Y** exothermicities are respectively small and large (case 2), are summarised in Table 1. Whatever a more precise form of Fig. 3 may be, given only the generality that more exothermic interactions approach a limiting cost in entropy, it is seen that this expression of cooperativity (the classical chelate effect) is greatest where each of the associations of **X** and **Y** are quite strongly exothermic. (It is for this reason that the hypothesis $K_X = K_Y = 10^3$ dm^3 mol^{-1} made at the beginning of this section, for the purposes of illustration of an intrinsic binding affinity, is an unreasonable one: if **X** and **Y** lose essentially all their entropy in binding to a receptor, the binding into the receptor sites would have to be very strong, and K_X and K_Y would have to much greater than 10^3 dm^3 mol^{-1}).

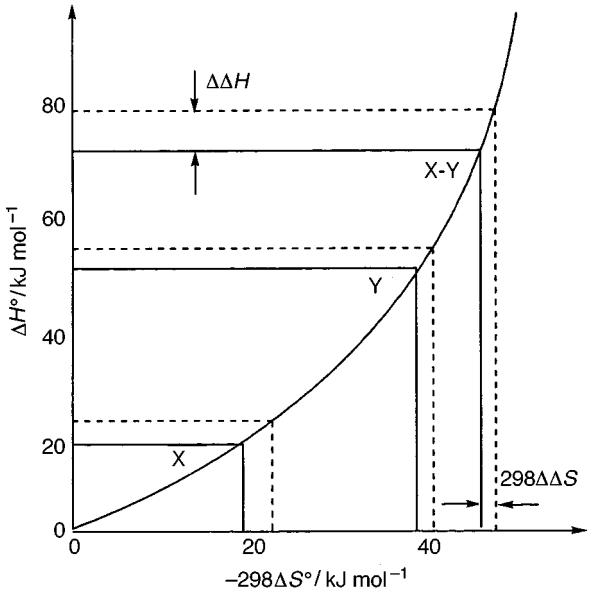


Fig. 3 Entropy–enthalpy compensation curve illustrating the enthalpy and related entropy values for the association of **X** and **Y** with a substrate where $\Delta H_X = 20$ kJ mol^{-1}, $\Delta H_Y = 50$ kJ mol^{-1}, $\Delta H_{X-Y} = 70$ kJ mol^{-1} (solid lines). When **X** and **Y** are tethered together, the binding of **X** will enhance the binding of **Y** and *vice versa*. This is represented by the dashed lines.

Table 1 First approximation for estimating some hypothetical relative magnitudes of chelate effects as a function of exothermicities of association

Case	ΔH_X[a]	ΔH_Y[a]	K_X[b]	K_Y[b]	K_{X-Y}[b]	K_{X-Y}/K_XK_Y[c]
1	−20	−50	1	1.3×10^2	1.6×10^4	1.2×10^2
2	−20	−70	1	1.6×10^4	2.3×10^7	1.4×10^3
3	−50	−70	1.3×10^2	1.6×10^4	1.9×10^{12}	9.0×10^5

[a] kJ mol^{-1}. [b] dm^3 mol^{-1}. [c] mol dm^{-3}.

We can in fact make a simple refinement of the situation considered thus far. The consequences of this refinement will be considered for case 1 in Table 1. For each of the **X** and **Y** interactions, only a part of the total theoretical maximum translational and rotational entropy which could be lost is lost (*cf.* Fig. 3). This is because, in a weak interaction, the theoretical maximum bonding (which could be expressed at 0 K) is not expressed at room temperature due to the opposing entropic advantage of residual motion, resulting in an average position in the enthalpic well corresponding to an exothermicity of 20 kJ mol^{-1} when **X** binds alone and 50 kJ mol^{-1} when **Y** binds alone. However, if the restriction of motion is aided by a neighbouring exothermic interaction, then the average position in the enthalpic well will correspond to some larger exothermicity. This is because the binding of the **X** part of **X**–**Y** is aided by the binding of the **Y** part of **X**–**Y**, and *vice versa*. Thus in the general, and real, case it is to be expected that the exothermicity of binding of **X**–**Y** will be greater than the sum of the exothermicities with which **X** and **Y** separately bind. The effect of this second cooperative factor, which is quite distinct from that described to give the numbers in Table 1, is illustrated by a hypothetical example in Fig. 3. The solid horizontal lines indicate the enthalpies of association of **X**, **Y**, and **X**–**Y** (20, 50 and 70 kJ mol^{-1}, respectively) for case 1 (Table 1), analysed with the previous simple assumption that the exothermicity of association of **X**–**Y** is equal to the sum of **X** and **Y**. The dotted lines indicate the corresponding analysis allowing for the effect now being considered, whereby **X**–**Y** binds more exothermically than the sum of its parts, since in **X**–**Y**, **X** helps to anchor **Y**, and **Y** helps to anchor **X**. The increase in binding energy of **X**–**Y** due to this effect is $\Delta\Delta H - 298\Delta\Delta S$ (Fig. 3). Fig. 3 has been drawn to illustrate the physically likely consequence that a weakly exothermic interaction (of **X**) is likely to be strengthened more by the assistance of a stronger adjacent interaction (of **Y**) than *vice versa*. The importance of this last analysis is in a case where **X** and **Y** bind in a strain free manner as **X**–**Y** through a direct connection of **X** and **Y**. The enhancement of binding should occur not simply through a classical entropy driven chelate effect, but also as a consequence of an improved enthalpy of binding. It is this latter consequence that precludes the derivation of free energies of binding which are character-

istic of common functional groups, even in relatively simple systems. The binding energy obtainable from any specified weak interaction will always be context dependent. In sections 5 and 6 we present experimental data to support this view.

4 Model 2: Where associations occur in water

Having considered weak associations occurring where the interaction with solvent is relatively weak, we now turn to the more complex situation of associations in water. Much effort has gone into attempts to estimate free energies of binding for common types of weak interactions [*e.g.* the hydrophobic effect (Å^{-2}),[6–9] amide–amide hydrogen bonds[10–16] and salt bridges[10,17–19]] as they occur as parts of a network of weak interactions in water as solvent. For convenience, we may consider two kinds of models in this area. First, those systems where it might be a useful approximation to consider that the associating entities **A** and **B** retain their structures in the associated state, except insofar as there might be restrictions of internal rotations of **A** and/or **B**. Second, where **A** and **B** do not retain their internal structures in the same form; we reserve considerations of these cases until section 6. In the former case, eqn. (5) would form a simple extension of eqn. (2) for strain-free systems: eqn. (5).

$$\Delta G_{obs} = \Delta G_{t+r} + \Delta G_r + \Delta G_h + \Sigma\Delta G_p \quad (5)$$

Here, the contributions which promote binding are the familiar term $\Sigma\Delta G_p$, and the new term: ΔG_h, the contribution from the hydrophobic effect, which corresponds to the favourable free energy of binding arising from the removal of hydrocarbon surface area from water upon association. It is a term which is conveniently separated from other binding terms for two reasons. First, at room temperature it is essentially purely an entropy term—favourable because when hydrocarbon surface area is removed from exposure to water, the water structure becomes more disordered.[20] Second, its magnitude is proportional to the surface area of hydrocarbon removed from exposure to water.[6–9] This surface area is frequently conveniently measured from modelling studies. The terms which oppose binding are the familiar ΔG_{t+r} term, and the new term ΔG_r due to loss of entropy associated with the restriction of any internal rotations which may occur on binding. If this loss of entropy is ΔS_r, then at 298 K, $\Delta G_r = -298\Delta S_r$.

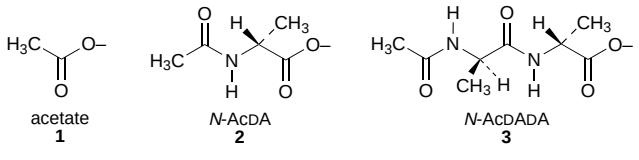
A problem in applying eqn. (5) will presumably derive from the same source as in applying eqn. (2). For example, if we add a polar group into a binding site, and then measure the change in ΔG_{obs}, we will in the general case not obtain a true ΔG_p value because the addition of the polar group increases the adjacent interactions in a manner analogous to the arguments already presented for the case in non-polar solvents. It is true that the addition of the polar group will also increase the adverse ΔG_{t+r} term, but the crux of Fig. 1 and 3 is that the benefit to binding of the extra bonding will outweigh the extra cost in entropy. The key conclusion is that deletions (including mutation studies on proteins) of groups which contribute to binding in strain-free systems will normally be expected to give ΔG_p values which are too large. In the following section we present data for binding in water to support this conclusion.

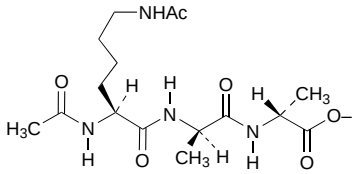
5 Attempts to obtain individual group contributions to binding and cooperativity at an interface

In the early 1990s, we selected glycopeptide antibiotics of the vancomycin group, in their binding of bacterial cell-wall peptide analogues as a vehicle to test the application of the approximation represented by eqn. (5). Since we now know in what way this approximation is likely to err, what have we been able to learn from the studies?

Since the study involved restriction of internal rotations within the bacterial cell-wall peptide analogues upon binding to the antibiotics, we required guides to ΔG_r. This parameter was taken to lie in the range 2–5 kJ mol^{-1}; the lower end of this range came from the restriction of internal rotations for the formation of crystals from the liquid state,[6,21] and the upper end from the restriction of internal rotations in the formation of small rings from linear hydrocarbons.[22] The smaller values (2–3 kJ mol^{-1}) presumably reflect the larger residual motions in crystals compared to tight ring structures, and therefore are probably more appropriate to the associations commonly found in biology.

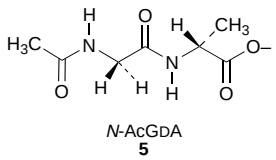
We were able to delete methyl groups, and amide–amide hydrogen bonds in the antibiotic binding sites, and so evaluate the apparent binding energies associated with these entities. Our first attempts carried out the partitioning incorrectly, and gave amide–amide hydrogen bond strengths which were far too high (*ca.* 20 kJ mol^{-1} in water).[23,24] A more appropriate partitioning gave these bond strengths in the range 0–7 kJ mol^{-1}, and a hydrophobic effect of 0.20 kJ mol^{-1} Å^{-2} (of hydrocarbon buried from water on binding).[6] Interestingly, the values are in good accord with the apparent binding energies obtained by protein engineering experiments (1–8 kJ mol^{-1}, and a hydrophobic effect of 0.23 kJ mol^{-1} Å^{-2}).[7] Despite this agreement, the arguments presented in sections 2 and 3 suggest that these values should be larger than the true local binding energies. We now present two of our most recent studies which support this conclusion.

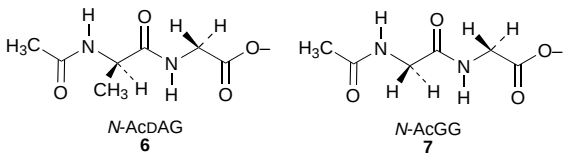


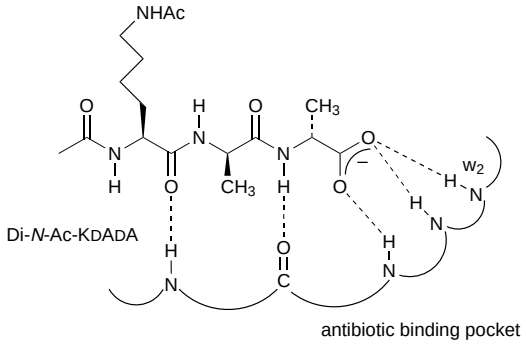


(i) We have examined binding of the ligands **1** to **4** into the binding site of vancomycin group antibiotics. As the network of interactions which increases along the series **1** to **4** is extended (to the left as displayed in Scheme 2), the strength of the hydrogen bond between the carboxylate oxygen atom of the ligands and the antibiotic NH (designated w_2) gradually increases.[25] Thus, the carboxylate anion is bound more strongly into the pocket which receives it as it is aided in this binding by adjacent interactions which help to restrict the ligand motion. This result indicates that if amide–amide hydrogen bond free energies are inferred from free energies of binding in the series **1** → **2** → **3**, then the derived hydrogen bond strengths will be too large, for the addition of these hydrogen bonds to the network increases the strengths of adjacent interactions.

(ii) We have also examined the ligand series **3**, **5**, **6** and **7**. In every case, it is observed that a change in the ligand of a Gly







Scheme 2

to an Ala increases the strength of the hydrogen bond from carboxylate oxygen to w_2, the antibiotic NH.[26] Thus the apparent increased binding energies from the Gly → Ala 'mutations' must reflect in part a contribution from the increase in strength of this adjacent hydrogen bond. This cooperative strengthening of adjacent interactions can account for the fact that the values cited earlier in this section for the hydrophobic effect (0.20 and 0.23 kJ mol^{-1} Å^{-2})[7,21] are larger than the solvent transfer value[9] (0.125 kJ mol^{-1} Å^{-2}, which cannot benefit from the type of cooperativity described). While it is of course also possible that the conformational bias of the Ala-containing peptides (relative to Gly-containing) might improve binding site affinity and so contribute to the strengthening of the adjacent hydrogen bond, the larger apparent value of the hydrophobic effect from the 'binding site' *vs.* 'solvent transfer' experiments is understandable from a common viewpoint: the addition of binding affinity, or the restriction of motion, at one point in the binding site can improve the binding affinity in an adjacent site.

In summary, real systems will always have residual motion. The addition of an extra interaction will typically (in a strain-free system) reduce adjacent motions and improve the free energy of binding of adjacent groups. Therefore, attempts to estimate the binding affinities of specified groups will, in strain free systems, tend to give values that are too large. This consequence has been exemplified for the hydrophobic effect above, and also implies that the apparent amide–amide bond strengths in water of 0–8 kJ mol^{-1} are probably benefiting from other cooperative interactions, and are therefore likely to be over-estimates. This conclusion is consistent with recent calculations, which suggest that the enthalpy of formation of peptide hydrogen bonds is close to zero.[27]

6 Model 3: Binding where A and B can adjust their structures on association

In the light of the formalism of eqn. (1), it has perhaps been inevitable that many studies of weak interactions have sought the origins of experimental binding energies by an examination of the interactions of **A** and **B** with solvent and of the interface between these two entities. This approach will be invalid if **A** and **B** change their internal structures upon association. Therefore, if one or more of the associating components is a folded polypeptide (essentially all biological receptors) or a polymer of DNA or RNA which is involved in duplex or folded structures, then the binding energy of the two components cannot reliably be sought at the binding interface, even after consideration of the energetics of desolvation of this interface. If we temporarily ignore the interactions with solvent, then in such cases a more appropriate form of the equilibrium constant would be given by eqn. (6).

$$\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{A'}\cdot\mathbf{B'} \qquad (6)$$

Eqn. (6) recognises that once **A** and **B** have associated, then typically, they no longer exist. Rather, they have been replaced by modified entities **A′** and **B′**. Crucially, the formalism of eqn. (6) emphasises that the binding energy between the two entities that come together is not simply a property of the interface between them, but also is dependent upon the modifications of the internal structures of **A** and **B** (**A** → **A′** and **B** → **B′**). Although it might be argued that so much is self-evident, reference to the literature indicates that this is so for many authors, but equally not so for many others. It is a common practice to rationalise the observed binding energy between **A** and **B** by examination of the interface between them, and to ignore the consequences of the changes **A** → **A′** and/or **B** → **B′**. These changes may at one extreme take the form of obvious structural modifications, but at the other extreme may in principle involve essentially no structural reorganisation but simply a 'tightening' (or a 'loosening') of the internal structure of **A** when it is modified to **A′** (or of **B** when it is modified to **B′**). Where these consequences can be considered, it may be possible to make semi-quantitative adjustments for the re-organisation **B** → **B′**, where **B** is a small substrate but, so far as we are aware, never for **A** → **A′** where **A** is a large receptor. Calculations might attempt to account for the free energy change and even to include the effect of solvent, but the forcefields currently in use are not sufficiently accurate to give reliable free energy changes for systems involving large receptors.‡ So, the binding affinities of greatest interest cannot be readily understood in molecular terms. We present below a relatively simple example of this complication.

7 Cooperativity beyond the interface

While some glycopeptide antibiotics show no measurable propensity to dimerise (*e.g.* teicoplanin), some do strongly (*e.g.* eremomycin).[29] In general, the antibiotics dimerise more strongly in the presence of bacterial cell-wall mucopeptide precursor analogues than in their absence.[29] For example, the dimer of the glycopeptide antibiotic eremomycin has $K_{dim} = 3 \times 10^6$ dm^3 mol^{-1} in the absence of di-*N*-acetyl-Lys-D-Ala-D-Ala, but $K_{dim} = 3 \times 10^8$ dm^3 mol^{-1} in its presence. It follows from a thermodynamic cycle that di-*N*-acetyl-Lys-D-Ala-D-Ala is bound by a factor of 10 more strongly by the dimer than by the monomer (each site of the dimer binds cell-wall analogue with the same affinity.[30]) The basis for at least part of this cooperativity can be seen from the structure of the ligand-bound dimer [Fig. 4(*a*)]. The antibiotics showing the largest dimerisation constants have, in addition to the four hydrogen bonds at the dimer interface [heavy dashed lines in Fig. 4(*a*)], two additional hydrogen bonds. These two bonds are from the alkylammonium ions of the amino sugars (which are unique to the strongly dimerising antibiotics) to amide carbonyl groups in the other half of the dimer [Fig. 4(*b*)]. The alkyl ammonium ions, one at each end of the head-to-tail dimer, are brought into the proximity of the cell-wall analogue carboxylate anion [one in each half of the dimer; Fig. 4(*b*)]. The resulting Coulombic attraction can tighten binding at both the ligand–antibiotic and dimer interfaces (Fig. 4). One way of looking at this cooperativity is that strong dimer formation also makes for the formation of a salt bridge which is mediated through an intervening amide bond [arrowed in Fig. 4(*b*)]. The evolution of this sophisticated interaction suggests that dimers may work more efficiently than monomers in antibacterial action, and indeed this has been shown to be the case.[31] However, in the present context, the relevant questions are:

(i) in the promotion of the dimerisation constant of eremomycin from $K_{dim} = 3 \times 10^6$ to $K_{dim} = 3 \times 10^8$ dm^3 mol^{-1} by ligand, can we infer anything useful about the origin of the extra binding energy?

‡ The difference in binding affinities ($\Delta\Delta G$) of two closely related ligands to a common receptor of moderate size had been calculated with impressive accuracy,[28] but this accuracy is only possible because relatively large systematic errors in the calculation are removed by difference.

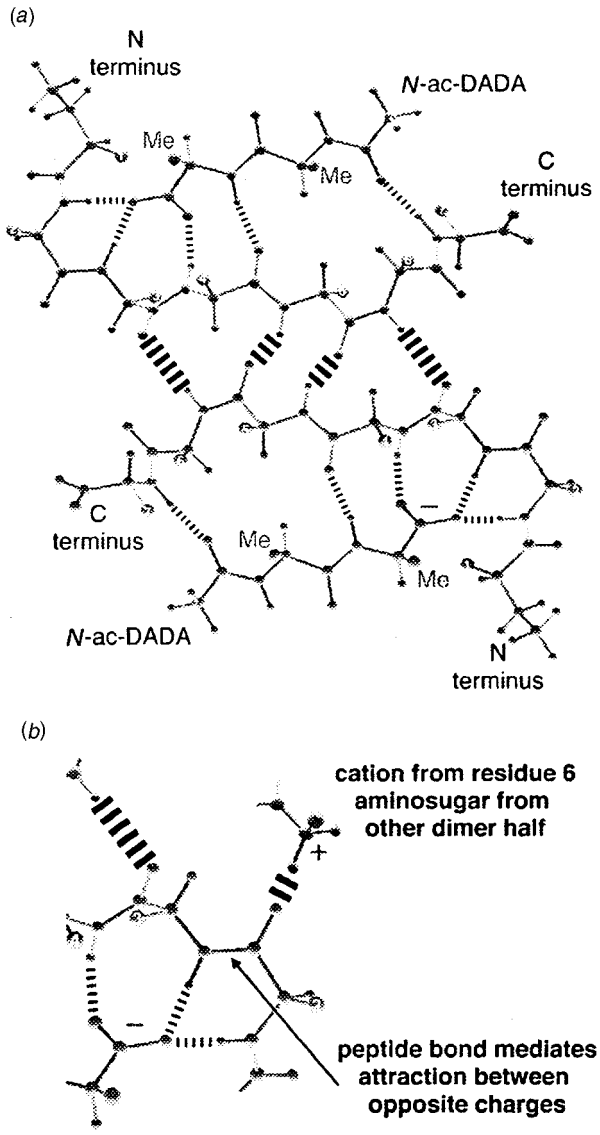
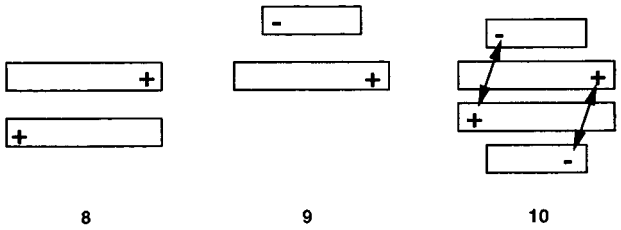


Fig. 4 (*a*) The structure of the antibiotic dimer (backbone only) bound to ligand (*N*-acetyl-D-Ala-D-Ala). The hydrogen bonds at the dimer interface are represented as broad dashed lines and those in the ligand binding pocket as narrower dashed lines. (*b*) The strongly dimerising antibiotics have an amino sugar on residue 6 which has a Coulombic interaction with the carboxylate of the ligand. This is essentially a salt bridge mediated by the peptide bond between residues 2 and 3.

(ii) similarly, in the promotion of the binding of di-*N*-acetyl-Lys-D-Ala-D-Ala to the dimer over the monomer by a factor of 10, can we infer anything useful about the origin of the extra binding energy?

These questions have not yet been addressed by experiment. Yet what seems to be a physically reasonable model, representing dimer by **8**, ligand-bound monomer by **9**, and ligand-bound dimer by **10**, gives plausible insights (in these diagrams, the + signs represent the ammonium ions of the amino sugar, and the − signs represent the negative charge of the carboxylate ion of the cell-wall analogue ligand). Since there is little doubt that at least part of the cooperativity is due to the Coulombic attraction between these two opposite charges, the models **8** to **10** suggest the following.

In the case of question (i), part of the extra stability of the dimer structure **10** over the dimer structure **8** lies in the stronger binding of the ligand in **10** relative to its binding in **9**. That is, when two of the entities shown in **9** come together to give one of **10** then, compared to the formation of **8** from two antibiotic monomers, part of the increased binding energy will come from the fact that ligand is bound more tightly in **10** than **9**. In other words, in the increase of the dimerisation constant by a factor of 100, some of the favourable free energy should come from the 'tightening' of the ligand–antibiotic interface upon dimerisation. The Coulombic attractions which are unique to **10** (and indicated by double-headed arrows in this structure) can be expected to lead to strengthening of the weak interactions at all three interfaces present in **10**. It is for this reason that we have schematically inferred bond-shortening at all three interfaces in **10** relative to **8** and **9**.



In the case of question (ii), the analysis is of course equally relevant in illustrating how the increased affinity of the ligand for the dimeric receptor (**10**) over the monomeric receptor (**8**) cannot be simply ascribed to the strengthening of the weak interactions at the ligand–receptor interface in **10** relative to **9**. It will in part be also due to the strengthening of the weak interactions at the dimer interface.

These models illustrate in a simple way the potential origins of binding affinity which is remote from the binding interface. In large systems, this makes the prediction of binding constants more difficult because interactions remote from the binding site may significantly contribute to the overall binding constant.

8 Can binding affinities be reliably predicted?

The extensive and subtle changes which can affect binding affinities, as outlined in the preceding sections, suggest that relatively accurate *de novo* predictions of equilibrium constants for associations of extended networks are not likely to be achieved in the general case in the near future.

Despite the problems in the search for a solution that is universally valid, the use of eqn. (5)[21,32] has recently been extended by Bohm, and with a surprisingly good outcome for a limited data set.[33] In this work, the four terms (ΔG_{t+r} + ΔG_r + ΔG_h + $\Sigma\Delta G_p$) of eqn. (5) are used, but the last of these (for binding electrostatic attractions) is sub-divided into two terms—one for formally uncharged hydrogen bonds (ΔG_{hb}) and a second for a polar interaction involving a charged entity (ΔG_{ionic}). A set of 45 interactions of experimentally known binding constants was then considered, where ligands of relatively small molecular mass (66 to 1047) interact with proteins. Since the modified form of eqn. (5) has only five unknowns, and the 45 interactions involve different combinations of these unknowns, average values for the five parameters can be obtained. From this training set, the average contribution from a neutral hydrogen bond of ideal geometry was −4.7 kJ mol^{-1}, from the corresponding ionic interaction was −8.3 kJ mol^{-1}, from the hydrophobic effect was −0.17 kJ mol^{-1} Å^{-2}, and from the restriction of a rotatable bond was +1.4 kJ mol^{-1}. The average opposition to binding from ΔG_{t+r} was +5.4 kJ mol^{-1}. This energy function reproduced the binding constants of the training set (which experimentally range from 40 to 2.5 × 10^{13} dm^3 mol^{-1}) with a standard deviation of 7.9 kJ mol^{-1}, corresponding to 1.4 orders of magnitude of binding affinity.

These results are approximations which may be very useful, and merit some comment. First, the values obtained for the polar interactions are in good agreement with those obtained by other approaches,[11–13,16] although for the reasons presented earlier these should perhaps best be regarded as apparent binding energies[12,34,35] rather than as localised bond energies. Second,

the value for the hydrophobic effect is in reasonable accord with that from other work, being intermediate between that from solvent transfer experiments (-0.12 kJ mol^{-1} Å^{-2})[9] and from the deletion of methyl groups in binding sites (-0.20 to -0.23 kJ mol^{-1} Å^{-2}).[7,21] Third, the cost of restricting a rotatable bond is somewhat less than that found for the melting of crystals (2 to 3 kJ mol^{-1}),[6] and may reflect the fact that rotations are somewhat less restricted in these binding sites than they are in crystals. Fourth, the average value of ΔG_{t+r} of +5.4 kJ mol^{-1} is remarkably small. It represents only about one tenth of the maximum theoretical entropy loss (corresponding to complete immobilisation of the ligand). If it is assumed that this estimate is realistic (Bohm notes that this parameter is the most uncertain of all those derived), then it may reflect small average exothermicities of association of these ligands to their protein receptors in water—a suggestion which seems physically plausible. Last, it is noteworthy that one of the largest errors in the estimated binding constants is found for the streptavidin–biotin interaction ($K_{calc} = 10^{10.75}$ *vs.* $K_{exp} = 10^{13.4}$ dm^3 mol^{-1}). In a sense, this is remarkably good agreement for such high affinity binding, when computed by a simple method. But it should be noted that this measure of agreement was only obtained by regarding the ureido group of biotin as a charged entity (following the suggestion of Weber *et al.*),[36] despite the fact that it is formally uncharged. This is perhaps an attempt to rationalise the origins of binding energy at an interface, where in fact the origins are probably much more complex in this case.[37] Despite such problems, the application of eqn. (5) in the approach of Bohm gives impressively accurate predictions overall, and as a pragmatic and simple approach, it has much to commend it. It will be of great interest to see how it performs with a wider set of associations, or possibly also after further refinement.

Another promising approach to predict the binding affinity of novel ligands for receptors of known three-dimensional structure is currently being explored.[38] This method, known as VALIDATE, gives an absolute average error of only 1.45 log units between experimental and computed binding constants for 11 thermolysin inhibitors which were not part of the training set (though other thermolysin–ligand interactions were in the training set). Additionally, for a set of 14 inhibitors where neither the ligands nor the specific receptors were included in the training set, the technique gave estimated binding constants that had an absolute average error of only 0.68 log units relative to the experimental values.

Both of the above semi-empirical methods may lack an accurate physical description of binding processes, but their importance lies in the reasonable success of their predictions.

Conclusions

In studying weak interactions, we have considered the play-off or compensation between electrostatic bonding and the restriction in motion. These two terms form the basis of our attempt, for gas phase interactions and those occurring in non-polar solvents, to factorise the observed free energy of binding into its component parts. For associations in water, we have developed an approach where the hydrophobic effect, and internal rotations in the associating molecules are taken into account such that some prediction of binding constants and their component parts can be made.

Further, we note that in biological systems, cooperativity both at the binding interface and remote from it, can contribute significantly to the observed free energy; see also ref. 37. That is, the sum of the isolated component parts is less than the whole observed binding energy. Because the entropic cost of an association is variable, this complication cannot be removed by simply trying to factor out the entropic term. Motion and bonding are intrinsically linked. Despite these complications, current semi-empirical approaches for the estimation of binding affinities are giving promising results. Weak interactions may still be ill understood, but they are giving up their secrets.

10 Acknowledgements

We thank the EPSRC, BBSRC, and Glaxo Research and Development for financial support. Lilly Laboratories, Indianapolis are thanked for the exchange of information and useful discussions regarding the vancomycin group antibiotics.

11 References

1 W. P. Jencks, *Adv. Enzymol.*, 1975, **43**, 219.
2 W. P. Jencks, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 4046.
3 M. S. Searle, M. S. Westwell and D. H. Williams, *J. Chem. Soc., Perkin Trans. 2*, 1995, 141.
4 M. S. Westwell, J. Klein and D. H. Williams, *J. Phys. Chem.*, 1996, **100**, 16000.
5 J. D. Dunitz, *Chem. Biol.*, 1995, **2**, 709.
6 M. S. Searle and D. H. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 10690.
7 L. Serrano, J.-L. Neira, J. Sancho and A. R. Fersht, *Nature*, 1992, **356**, 453.
8 K. A. Sharp, A. Nicholls, R. Friedman and B. Honig, *Biochemistry*, 1991, **30**, 9686.
9 R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1986, **82**, 8069.
10 A. Fersht, *Enzyme Structure and Mechanism*, W H Freeman and Company, New York, 2nd edn. 1985.
11 A. R. Fersht, *Trends Biochem. Sci.*, 1987, **12**, 301.
12 A. R. Fersht, J.-P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye and G. Winter, *Nature*, 1985, **314**, 235.
13 B. A. Shirley, P. Stanssens, U. Hahn and C. N. Pace, *Biochemistry*, 1992, **31**, 725.
14 R. L. Baldwin, *Trends Biochem. Sci.*, 1989, **14**, 291.
15 K. A. Dill, *Biochemistry*, 1990, **29**, 7133.
16 D. H. Williams, M. S. Searle, M. S. Westwell, U. Gerhard and S. E. Holroyd, *Phil. Trans. R. Soc. London, A*, 1993, **345**, 11.
17 A. R. Fersht, *J. Mol. Biol.*, 1972, **64**, 497.
18 M. F. Perutz, *Nature*, 1970, **228**, 726.
19 A. Bierzynski, P. S. Kim and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 2470.
20 N. Muller, *Acc. Chem. Res.*, 1990, **23**, 23.
21 M. S. Searle, D. H. Williams and U. Gerhard, *J. Am. Chem. Soc.*, 1992, **114**, 10697.
22 M. I. Page and W. P. Jencks, *Proc. Natl. Acad. Sci. USA*, 1971, **68**, 1678.
23 D. H. Williams, *Aldrichim. Acta*, 1991, **24**, 71.
24 A. J. Doig and D. H. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 338.
25 P. Groves, M. S. Searle, M. S. Westwell and D. H. Williams, *J. Chem. Soc., Chem. Commun.*, 1994, 1519.
26 G. J. Sharman, M. S. Searle, B. Benhamu, P. Groves and D. H. Williams, *Angew. Chem.*, 1995, **34**, 1483.
27 M. J. Sippl, *J. Mol. Biol.*, 1996, **260**, 664.
28 W. C. Still and R. Liu, *Phil. Trans. R. Soc. London, A*, 1993, **345**, 97.
29 J. P. Mackay, U. Gerhard, D. A. Beauregard, M. S. Westwell, M. S. Searle and D. H. Williams, *J. Am. Chem. Soc.*, 1994, **116**, 4581.
30 G. Batta, M. F. Cristofaro, G. J. Sharman and D. H. Williams, *Chem. Commun.*, 1996, 101.
31 D. A. Beauregard, D. H. Williams, M. N. Gwynn and D. J. Knowles, *Antimicrob. Agents Chemother.*, 1995, **39**, 781.
32 P. R. Andrews, D. J. Craik and J. L. Martin, *J. Med. Chem.*, 1984, **27**, 1648.
33 H.-J. Bohm, *J. Comp.-Aided Mol. Design*, 1994, **8**, 243.
34 A. E. Mark and W. F. van Gunsteren, *J. Mol. Biol.*, 1994, **240**, 167.
35 D. H. Williams, M. S. Searle, M. S. Westwell, J. P. Mackay, P. Groves and D. A. Beauregard, *Chemtracts*, 1994, **7**, 133.
36 P. C. Weber, J. J. Wendoloski, M. W. Pantoliano and F. R. Salemme, *J. Am. Chem. Soc.*, 1992, **114**, 3197.
37 D. H. Williams and M. S. Westwell, *Chem. Biol.*, 1996, **3**, 695.
38 R. D. Head, M. L. Smythe, T. I. Oprea, C. L. Waller, S. M. Green and G. R. Marshall, *J. Am. Chem. Soc.*, 1996, **118**, 3959.
39 R. M. Izatt, J. S. Bradshaw, K. Pawlak, R. L. Bruening and B. J. Tarbet, *Chem. Rev.*, 1992, **92**, 1261.

Received, 10th July 1997
Accepted, 15th September 1997