Thermodynamic aspects of hydrophobicity and biological QSAR

Ki H. Kim

Department of Structural Biology, Division of Advanced Technology, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60061-6100, USA

Received 2 August 2000; accepted 23 October 2000

Key words: thermodynamic QSAR, enthalp-entropy compensation, drug-receptor interaction, 2D- and 3D-QSAR, ligand binding site, QSAR in organic solvent

Summary

A protein contains a large amount of water molecules, and the nature of the interactions of the water molecules with a protein play an important role in the thermodynamics of the ligand binding process. In this paper, thermodynamic aspects of drug-receptor interactions, enthalpy-entropy compensation or reinforcement, hydrophobicity, and biological 2D- and 3D-QSAR are discussed. Comparisons of the thermodynamic QSAR of phenyl esters of N-benzoyl L-alanine in phosphate buffer and pentanol provide useful insight for the ligand-enzyme interactions.

Introduction

In biological 3D-QSAR, it is often assumed that steric and electrostatic forces are enough to account precisely for the observed biological activities. Although steric and electrostatic complementarity between protein and ligand is considered to be the most important features for protein-ligand interactions, these alone are not sufficient in many cases. Hydrophobicity has long been recognized as an important property in drug-receptor interactions and in classical quantitative structure-activity relationship (QSAR) studies. And various forms of hydrophobic properties of drug molecules are shown to be useful in 3D-QSAR

Furthermore, it is also assumed that entropic contributions for structurally related compounds are similar. However, this is not always the case. A protein contains a large amount of water molecules, and the degrees in tightness of these waters and the nature of the interactions of the water molecules with a protein play an important role in the thermodynamics of the ligand binding process. Therefore, thermodynamic analyses in QSAR can provide useful information about the binding of drug molecule into a protein.

This paper describes some of the thermodynamic aspects of hydrophobicity and biological QSAR.

Ordered water positions in protein

The protein surface is covered with water molecules, and they contribute to the overall stability and characteristics of the protein. The structural and functional importance of water molecules is well recognized. X-ray or neutron crystallography and NMR as well as theoretical methods are useful to reveal the favored average positions occupied by water molecules in a protein and to understand the ordered water molecules [1]. These water molecules show varying degrees in tightness, but most of them are freely exchangeable with the solvent water.

The ordered water sites range from highly localized ones with high occupancy (low B-factor) to sites which show a weak preference. The water molecules that interact with apolar surfaces often occupy a different position each time showing a fairly evenly distributed pattern of hydration. However, the water molecules that interact with polar surfaces tend to move to favored positions for hydrogen-bonding interactions. Such hydrogen-bonding interactions give these water molecules less freedom in motion and thus diffuse more slowly. Fully buried water molecules with multiple hydrogen-bonding interactions within the protein have still less freedom in motion and ex-

pect to give the greatest loss of entropic energy to displace.

The water molecules observed in X-ray structures can be classified into three categories as suggested by Ringe and her co-workers [2]. The first is the water molecules observed at the same site, making the same interactions with the protein surface in every independently determined structure. These are category one water molecules. When comparing two structures, approximately one-half of the water molecules observed usually belong to this category. When multiple structures are compared, the number of category one water molecules drops. The second is the water molecules observed at a particular site in only one structure but not conserved in others. These are category two water molecules. These waters are an integral part of the protein structure. The third is the disordered water molecules that must be present at the surface of the protein, but that are not observed crystallographically. These are category three waters. These water molecules leave patches on the protein surface that look as if none are bound.

It was suggested that the slightly disordered water molecules that belong to the second category are more conducive to allowing a ligand to bind than are strictly ordered water molecules. When a ligand displaces a category two water molecule, there is a gain in entropy as the partially ordered solvent molecules are freed from the surface.

It is interesting to note that even in pure acetonitrile, a shell of ordered water molecules was associated with the surface of a protein (subtilisin) [3]. It appears that most of the water molecules found associated with the protein surface cannot easily be displaced by another solvent (even a water miscible one). However, some of the water molecules in the binding site can be displaced in order for a ligand molecule to interact with the binding site protein residues.

The most accurate water positions come from high-resolution neutron diffraction studies. However, even at a relatively high resolution (<2 Å) and well-refined (*R* factor <20%) structures, the water sites can be unreliable and incomplete near the noise level. For example, a thorough analysis of solvent structure in a trypsin crystal with 1.35 Å resolution X-ray data and 2.1 Å resolution neutron data revealed 184 and 291 water sites, respectively. Comparing the two results, 34 new waters were added to the final X-ray model and six weak water sites were removed. Twenty sites including five of the 58 strongest sites clearly visible in the X-ray structure were not visible in the neutron

analysis. Karplus and Faerman [1] also showed surprising variability of water structure in a comparison study of three independently determined interleukin-1 structures. The three structures (1i1b, 2i1b, 4i1b) refined at 2.0 Å resolution and R-factors of <20% contain 83, 91, and 168 water molecules, respectively. Therefore, the reliability and completeness of water positions depend not only on the resolution, but also the number of molecules per asymmetric unit (density averaging can decrease the noise) and the criteria used for placing water molecules [1]. Solvent structure may also be sensitive to subtle changes in experimental conditions, such as pH, buffer, ionic strength, data collection temperature, etc., but the effects of these parameters on water structure have not yet been well documented.

Ligand binding sites on protein

The three-dimensional structure of a protein is usually described by the sequence of the peptides and the folding of the peptides as well as the packing of the individual side chains that produces the overall tertiary structure of the protein. In addition, the water molecules that are associated with the protein surface contribute to the overall stability and the characteristics of the protein.

As described above, the surface of any protein is covered with water molecules with varying degrees in tightness. When a ligand binds to a site on a protein surface, it must displace one or more of these water molecules. Therefore, the nature of the interactions of the water molecules with the protein is important in the thermodynamics of the ligand binding process.

Different approaches have been used to study ligand-binding sites on protein. One of them utilizes organic solvents. Organic solvents can influence the enzymatic activity as well as the stability [4, 5]. Thus, by exposing the protein crystals to an organic solvent, the integrity of the protein can be altered. However, it is also known that certain proteins tolerate in neat organic solvents and still active.

Ringe et al.[3] determined the crystal structure of subtilisin in pure acetonitrile at 2.3 Å and compared to that obtained from an aqueous solvent. Comparison of these two structures showed that the structure was not affected by the organic solvent, and a similar number of hydration sites were occupied by water [3]. RMS deviations for the backbone atoms were less than 0.3 Å and for all atoms were less than 0.5 Å. These are

within the range of deviations found for two structures solved independently using crystals treated in the same way.

In order to find for the ligand binding sites on a protein using X-ray crystallography, Ringe and her coworkers [6] utilized organic solvents as search probes. For example, crystals of a protein such as elastase from aqueous mother liquor were transferred into an organic solvent (either neat or as an aqueous mixture) to allow the new mother liquor to diffuse through the crystal. By exchanging the mother liquor several times the solvent inside and outside the crystal was equilibrated allowing the organic solvent molecules to interact with the surface and favorable binding sites. Using such an experimental approach, they were able to map the complete binding surface of the protein and found the active site cleft.

A binding site is generally considered to be a region that is usually depression in the protein surface and has a greater than average degree of exposure of hydrophobic groups [2]. It contains bound water molecules (category two waters) that are not tightly held but which make specific interactions with the polar groups of the protein. These water molecules are easily displaced, allowing the polar functional groups on the ligand to make the similar interactions with the protein.

Ligand binding on protein

Most important features for protein-ligand interactions appear to be good steric and electrostatic complementarity between them [7, 8]. However, in many cases these alone are not sufficient enough to describe tight binding of a ligand especially without including desolvation effects. Strong ligand binding to protein has sometimes been observed with no hydrogen bonds or very few hydrogen bonds. It is generally considered that tight binding can be achieved if the total number of hydrogen bonds increases upon ligand binding, and that water molecules can play important roles in hydrogen bonding. If a ligand simply replaces tightly bound water molecules upon binding to a protein and does not yield additional interactions, its binding affinity will be small. On the other hand, a ligand replacing loosely bound water molecules will increase the binding affinity.

Hydrophobic interactions are considered to be mainly due to the replacement and release of ordered water molecules and are entropy-driven [9]. It is often

assumed that the entropic contributions to the binding are similar for different ligands [10]. However, this is not usually the case. Therefore, the interaction energies calculated without the entropic contributions are not always expected to correlate well with the measured free energies of binding.

With respect to the role of enthalpy and entropy contribution to the ligand binding on protein, the phenomenon of enthalpy-entropy compensation has important consequences for the theoretical derivation of structure-activity relationships. The enthalpy-entropy compensation phenomenon may be strong indication of a common mechanism which is able to control the binding process, irrespective of the nature of ligands and of their macromolecular targets [11]. Before this important phenomenon is discussed, however, we will first consider the thermodynamic aspects of drug-receptor interactions.

Thermodynamic aspects of drug-receptor interactions

Thermodynamic analyses are sometimes used to describe the binding of drug molecules. Two types of information are generally obtained experimentally [12]. First, equilibrium thermodynamic parameters ΔG^0 , ΔH^0 , and ΔS^0 can be used to evaluate the overall forces which drive the binding reaction. The standard free energy (ΔG^0) is calculated from the relationship, $\Delta G^0 = -RT \ln K_a$, where R is the gas constant, T is the absolute temperature, and K_a is the equilibrium association constant. K_a and the standard enthalpy (ΔH^0) are obtained from experiments, and the standard entropy (ΔS^0) is obtained from the relationship, $\Delta S^0 = (\Delta G^0 - \Delta H^0)/T$.

Determination of whether a specific ligandreceptor interaction at equilibrium is enthalpy- and/or entropy-driven can be achieved by thermodynamic analysis. An enthalpy-driven process is usually associated with the formation of new bond such as hydrogenbonds and van der Waals interactions, whereas an entropy-driven process is usually characterized by the displacement of ordered water molecules coupled with the formation of new hydrophobic interactions [13].

The initial process in the receptor association is generally entropy-driven. This initial process is followed by one or more steps that are enthalpy-driven processes. There are a wide diversity in the thermodynamics of a ligand-receptor interactions. Some show entropy-driven, and others show enthalpy-driven [13].

The free energy of the binding (ΔG) shows in general no correlation with the enthalpy (ΔH) , and the binding of different ligands for a given protein can be driven by enthalpy and/or entropy [14]. However, if ΔH is plotted against ΔS for a set of related ligands, a straight line is frequently obtained. This phenomenon is called enthalpy-entropy compensation or reinforcement depending on the sign of the slope. The phenomenon is important in understanding the binding of drug molecules as well as the thermodynamic aspects of hydrophobicity and biological QSAR.

Enthalpy-entropy compensation or reinforcement

The phenomenon of enthalpy-entropy compensation refers to an empirical observation that the change in enthalpy is partially compensated by a corresponding change in entropy, resulting in a smaller net free energy change. Relative free energies measure the change in free energy caused by replacing one molecule with another during thermodynamic cycles involving either solvation or ligand binding. The enthalpy changes measure a change in the strength of the interactions between molecules while the entropy changes measure a change in the order of the solvent. Stronger interactions between molecules will result in a reduction of the configurational freedom of the solvent and thus a reduction of the entropy. On the other hand, weaker molecular interactions will result in a looser molecular association and an increase of the entropy.

The enthalpy-entropy compensation is commonly observed in the process involving biological molecules in solution and characteristic of all weak interactions, such as hydrogen-bonds and van der Waals interactions which are found between ligands and proteins [15]. Gilli et al. [11] suggested that the observed enthalpy-entropy compensation phenomenon is strong indication of a common mechanism which is able to control the binding process, irrespective of the nature of ligands and of their macromolecular targets. They also suggested that the enthalpy-entropy compensation arises from an intrinsic property of the hydrogen bond, which is the main force determining the association of the participants (water, ligand, binding site) in the ligand-receptor binding equilibrium. Lemieux [16] suggested that enthalpy-entropy compensation arises from changes in the dynamics of the water structure.

Figure 1 shows two examples of such enthalpyentropy compensations: the transfer of substituted phenols in the octanol-water system [17] and for the ligands binding to OppA [14]. As a consequence of the enthalpy-entropy compensations indicated by the positive slopes in these examples, the corresponding free energy changes in these transformations became small

Recently, Gallicchio et al. [18] addressed the role of enthalpy-entropy compensation on the relative free energies of solvation and relative free energies of binding. The process of designing molecules that optimize free energy of ligand binding is often impeded by the fact that changes directed to strengthen the association of the molecules with the solvent or the ligand to the protein are accompanied by a compensating reduction of the entropy. This results in a small change in free energy and sometimes is not even in the expected direction.

However, there are many exceptions to such an enthalpy-entropy compensation. Two examples of noncompensating behavior are shown in Figure 2: for the hydration of a series of miscellaneous solute pairs and for the binding of related peptide pairs to a host [18]. In these examples, the enthalpy and entropy changes are similar in magnitude but opposite in sign resulting in enthalpy-entropy reinforcement. Thus, they add constructively to produce a free energy change approximately twice as large in magnitude. Gallicchio et al. [18] reported that only about 10% of the databases of relative thermodynamic quantities they constructed showed clear noncompensating behavior. They suggested that it is more likely to observe compensating processes, but that the phenomenon of enthalpy-entropy compensation is neither universal nor thermodynamically necessary. Krug et al. [19-21] suggested that detectable extrathermodynamic enthalpy-entropy effects are rare, and that linear plots of enthalpy-entropy data do not necessarily imply that a chemical phenomenon is responsible for the observed correlation. They claimed that most reported examples of linear enthalpy-entropy compensations in the literature appear to be merely the statistical compensation pattern that is independent of the chemistry of the observed reactions and equilibria [21].

Thermodynamic aspects of partitioning properties and biological QSAR

Hydrophobicity has long been recognized to play an important role in drug-receptor interaction, and hy-

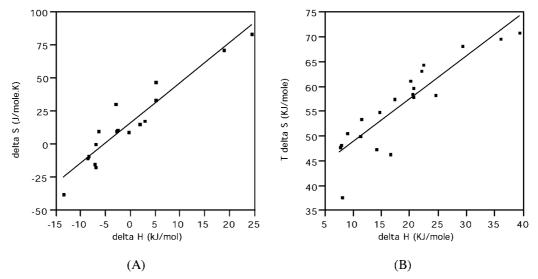


Figure 1. Enthapy-entropy compensation plot for the transfer of substituted phenols in the octanol-water system [17] (A) n = 17, $R^2 = 0.90$, slope = 3.02) and for the ligands binding to OppA [14] (B) n = 20, $R^2 = 0.79$, slope = 0.87).

drophobic parameter represented by the logarithm of octanol-water partition coefficient has been the most popular parameter in classical quantitative structure-activity relationship (QSAR) studies. Therefore, understanding the thermodynamic aspects of partition coefficient is useful to understand the binding of drug molecules.

Fujiwara and co-workers [22, 23] studied the thermodynamic aspects of partition properties by experimentally determining the equilibrium thermodynamic parameters ΔG^0 , ΔH^0 , and ΔS^0 for benzoic acids in the octanol-water system and their role in QSAR. The standard free energy change in partition (ΔG_p^0) was calculated from the relationship, $\delta G_p^0 = -RT \ln P$, and the free energy change can be divided into the enthalpy change (ΔH_p^0) and the entropy change (ΔS_p^0) as described in Equation 1.

$$\log P = -\Delta G_p^0 / (2.303RT)$$

$$= -\Delta H_p^0 / (2.303RT) + \Delta S_p^0 / (2.303R) \quad (1)$$

$$= P_H + P_S$$

They defined the hydrophobic enthalpy parameter P_H and the hydrophobic entropy parameter P_S from this relationship. Therefore, P_H can be related to the heat evolved when a solute is transferred from water to the octanol phase, and P_S to the heat reflecting the change of randomness of mobility induced in solution when a solute is transferred from water to octanol phase. Correlation between π and π_H or π and π_S of 50 substituted benzoic acids is low (r = 0.69 and 0.20,

respectively) and shows that the π constants are determined exclusively by neither the enthalpy nor the entropy term, but that the two terms contribute cooperatively to π . Here, π , π_H or π_S is the hydrophobic parameter derived from octanol-water partition coefficients (log P), P_H or P_S referenced to unsubstituted benzene or benzoic acid, respectively.

An application of hydrophobic enthalpy parameter P_H and hydrophobic entropy parameter P_S (thermodynamic parameters obtained for substituted alcohols), to the inhibitory potency (pID₅₀) of alcohol toxicity not only provided a better correlation than log P, but also revealed the relative contributions of the enthalpy and entropy contribution (Compare Equation 2 and 3) [22]. From the coefficients of Equation 3, it can be seen that the entropy term P_S contributes about one-third of the enthalpy term P_H to the inhibitory potency. (The enthalpy and entropy terms are normalized to contribute equally to the standard free energy change. Therefore, the ratio of the coefficient of the entropy term to the coefficient of the enthalpy term expresses the relative contribution of these two terms.)

$$pID_{50} = 0.678(\pm 0.615) \log P + 1.475(\pm 0.466)$$

 $n = 11, r = 0.639, s = 0.570$ (2)

$$pID_{50} = 1.006(\pm 0.289)P_H + 0.335(\pm 0.291)P_S$$

+2.680(\pm 0.472) (3)
$$n = 11, r = 0.952, s = 0.241$$

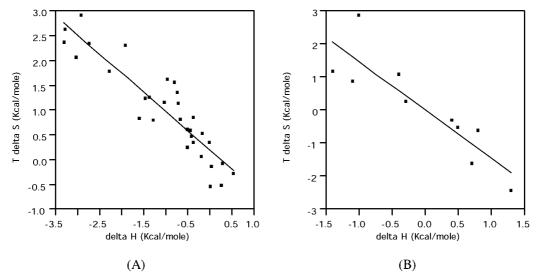


Figure 2. Relative entropy (T delta S) versus enthalpy (delta H) of hydration for a series of enthalpy-entropy noncompensating solute pair (A) n = 31, $R^2 = 0.83$, slope = -0.77) and ligand pairs (B) n = 10, $R^2 = 0.78$, slope = -1.46) [18].

The contributions of the enthalpy term and entropy term vary depending on the biological system [13]. In some cases, the enthalpy term is very important, and in other cases both terms are important. For example, in the correlation for the 50% inhibition of thiopurine methyltransferase's activity by substituted benzoic acids, a remarkable improvement was achieved when the traditional π was just replaced by π_H . The results indicated an important role of the enthalpy term for the inhibitory activity (Compare Equations 4 and 5) [22].

$$pI_{50} = 0.214(\pm 0.347)\pi + 3.373(\pm 0.280)$$

$$n = 13, r = 0.379, s = 0.458$$
(4)

$$pI_{50} = 0.867(\pm 0.172)\pi_H + 3.431(\pm 0.087)$$

$$n = 13, r = 0.958, s = 0.142$$
(5)

Likewise, the contributions of hydrophobic enthalpy and entropy terms depend on the partitioning system. The partitioning of phenol in the cyclohexanewater system is entropy-driven, whereas the partitioning in the octanol-water system is mainly enthalpy-driven [17]. It appears that solute-solvent interactions are important in octanol due to the hydrogen-bonding interactions. Therefore, the observed small gain in the entropy of phenol in the octanol-water system could be explained as the net effect when the usually large positive entropy gain resulting from transfer of hydrocarbon out of water was balanced by an almost equal loss of entropy due to restriction in movement of phenol as a result of their interaction with octanol.

These examples presented above and others [22] show that use of the hydrophobic enthalpy and entropy terms in QSAR can be useful not only for the improved correlations, but also for deciphering the relative contributions of the enthalpic and entropic portions toward the biological activity. Da et al. [22] suggested that such an improvement could be made when relative contributions of enthalpic and entropic portion of the partition properties differ between the biophase and the model system.

Thermodynamic aspects of biological QSAR in organic solvents

Thermodynamic analyses in QSAR can also provide useful information about the binding of drug molecules. The QSAR study of phenyl esters of N-benzoyl L-alanine in phosphate buffer and pentanol is an example. From the binding constants (K_m) of 18 substituted phenyl esters of N-benzoyl L-alanine analogs (PhCONHCH(Me)CO₂C₆H₄-p-X, I) to α -chymotrypsin in phosphate buffer (pH 7.4) and in pentanol, Selassie et al. [24] developed QSAR equations described in Equations 6 and 7.

For the binding of I in phosphate buffer:

$$\log 1/K_m = 0.28(\pm 0.11)\pi + 0.51(\pm 0.24)\sigma^-$$
$$+0.38(\pm 0.23)MR + 3.70(\pm 0.24)$$
(6)
$$n = 16, r = 0.913, s = 0.198$$

For the binding of I in pentanol:

$$\log 1/K_m = 0.25(\pm 0.09)\pi + 0.24(\pm 0.18)\sigma^-$$

$$+4.10(\pm 0.09)$$

$$n = 17, r = 0.873, s = 0.156$$
(7)

The binding constants ($\log 1/K_m$) obtained from phosphate buffer correlate well with π , σ^- , and M R, whereas those from pentanol show a dependence only on π and σ^- . In these equations, π is the Hansch hydrophobic parameter derived from octanol-water partition coefficients, σ^- is the Hammett electronic parameter, and M R is the molar refractivity. Because of poor fit, two compounds (4-t-Bu and 4-OH) were not included in deriving Equation 6, and one compound (4-CONH₂) was not included in deriving Equation 7.

Equation 6 (from phosphate buffer) shows that the binding of the substrates to the enzyme in the phosphate buffer is influenced by the hydrophobic, electronic, and steric effects of the para-substituents X. The positive coefficient of π and σ^- indicates that hydrophobic electron-withdrawing para-substituents increase the binding affinity of Compound I to the α-chymotrypsin binding site. The positive coefficient of M R suggests that large size substituents enhance the binding. The small coefficient of π indicates that the binding involves partial desolvation on the enzyme surface. The electron-withdrawing substituents help polarize the carbonyl bond and facilitate the nucleophilic attack by the serine hydroxyl group of the enzyme and thus increase the binding by stabilizing the oxyanion tetrahedral intermediate.

Equation 7 (from pentanol) indicates that the binding of the substrates to the enzyme in the phosphate buffer is only influenced by the hydrophobic and electronic effects of the para-substituents X. The effects described by MR in Equation 6 for the binding of Compound I in phosphate buffer is not needed in Equation 7 for the binding in pentanol.

The coefficients of π and σ^- in Equations 6 and 7 are essentially identical considering their error of esti-

mates indicating that the partial desolvation described by π and electronic influence of the para-substituents remain unchanged both in phosphate buffer and in pentanol. It is worth to note that the binding in pentanol was enhanced by 3 fold (the intercept of 3.70 in phosphate buffer versus 4.10 in pentanol), in addition to the fact that MR term was not needed to describe the binding in the organic solvent. Based on Equations 6 and 7, Selassie et al. [24] suggested that the pentanol competes with the enzyme for the buffer water resulting in desolvation of the enzyme surface and enhancing easier access of the ligand to the binding site.

In view of the works reported by Ringe and coworkers [2, 3, 6], describing that the protein structure was not affected by the organic solvent, and the QSAR equations (Equations 6 and 7) reported by Selassie et al. [24], from the binding constants (K_m) of phenyl esters of N-benzoyl L-alanine analogs to αchymotrypsin in phosphate buffer and in pentanol, it is possible that the integrity of α -chymotrypsin was not affected by the organic solvent. It is also likely that pentanol solvent molecules displaced one or more category two water molecules when the protein was in the organic solvent. This process of displacing the ordered water molecules to form new hydrophobic interactions by the pentanol would be an entropy-driven process. When a substrate was added to the α -chymotrypsin in pentanol, the substrate must replace one or more pentanol molecules bound to the active site in order for the hydrolysis by the enzyme to occur. The binding of substrate replacing the pentanol occupied in the active site would be associated with the formation of new bond such as hydrogen-bonds and van der Waals interactions. On the other hand, the binding of Compound I to α-chymotrypsin in phosphate buffer involves directly displacing the ordered water molecules by the substrate.

Equations 8 and 9 were developed [25] from the same data used in Equations 6 and 7 using the hydrophobic enthalpy parameter π_H and the hydrophobic entropy parameter π_S [22] instead of the hydrophobic constant π . One compound (4-t-Bu) was excluded in deriving Equations 6 was included in deriving Equations 8 and 9. However, two different compounds (4-CONH₂ and 4-SMe) were not included in Equations 8 and 9 because of missing π_H and π_S parameter values. Since the hydrophobic constant π values are exclusively determined by neither the enthalpy nor the entropy term, but by the two terms

cooperatively, it is not surprising to see that both π_H and π_S terms are present in Equations 8 and 9.

Equation 8 shows that the enthalpy term (π_H) plays more important role than the entropy term (π_S) in the binding process of Compound I in phosphate buffer. However, their roles are reversed for the binding process in the organic solvent (Equation 9). Although it appears that the binding in aqueous phosphate buffer is largely controlled by an enthalpy-driven process and that the binding in pentanol is largely controlled by an entropy-driven process, it needs further confirmation by additional data because of the large confidence intervals associated with the coefficients at this time.

For the binding of I in aqueous phosphate buffer:

$$\log 1/K_m = 0.38(\pm 0.11)\pi_H + 0.19(\pm 0.07)\pi_S$$
$$+0.53(\pm 0.11)\sigma^-$$
$$+0.26(\pm 0.10)M R + 3.77(\pm 0.11)$$
$$n = 15, r = 0.898, s = 0.200 \quad (8)$$

For the binding of I in pentanol:

$$\log 1/K_m = 0.21(\pm 0.08)\pi_H + 0.31(\pm 0.05)\pi_S$$
$$+0.20(\pm 0.08)\sigma^- + 4.16(\pm 0.04)$$
$$n = 15, r = 0.887, s = 0.160 \quad (9)$$

The coefficients of σ^- and the intercept in Equations 8 and 9 remain essentially the same as those in Equations 6 and 7. It is worth noting that the MR term is not significant for the binding of Compound I in pentanol. MR term is usually considered to describe the nonspecific dispersion interaction in the polar space [26]. In view of the above aspects, it may be possible that MR term in Equations 6 and 8 may describe the process of displacing the ordered category two water molecules (which is polar in nature) by the substrate. When a substrate was added to the α -chymotrypsin in pentanol, the substrate must now replace one or more pentanol molecules (which is nonpolar in nature) rather than the water molecules, and thus MR term may be no longer needed.

Figure 3 shows the binding pocket of eight α -chymotrypsin (1acb, 1cho, 1cgi, 1cgj, 2cha, 4cha, 5cha, 6cha) structures superimposed over the backbone. Figure 3A is shown without a ligand, and Figure 3B is shown with a ligand (3,5-dimethylphenyl ester of N-benzoyl L-alanine analog; PhCONHCH(Me)CO₂C₆H₄-3,5-Me₂) docked into the binding site. Water molecules are shown in balls in

these figures, and a comparison of the two figures shows the molecules that should be replaced by the bound ligand. Water molecules in different colors indicate that they are from different α -chymotrypsin structures.

Similar results were obtained by Compadre et al. [27] for the study of hydrolysis of N-substituted glycine analogs (PhCONHCH₂CO₂C₆H₄-p-X; II) by cathepsin b in an aqueous buffered solvent and in acetonitrile (Equations 6 and 7).

For the hydrolysis of II in aqueous buffer:

$$\log(k_{\text{cat}}/K_m) = 0.45(\pm 0.20)\sigma^{-}$$

$$+1.07(\pm 0.46) \text{M R} + 2.84(\pm 0.33)$$

$$n = 8, r = 0.970, s = 0.145$$
(10)

For the hydrolysis of II in acetonitrile:

$$\log(k_{\text{cat}}/K_m) = 0.64(\pm 0.24)\sigma^- + 4.35(\pm 0.16)$$

$$n = 8, r = 0.937, s = 0.163$$
(11)

Equations 10 and 11 show that the correlation for the hydrolysis of Compound II in aqueous buffer is influenced by σ^- and MR (Equation 10), whereas the hydrolysis in acetonitrile was influence only by σ^- (Equation 11).

Since no binding data were reported for this system, further analysis and comparison could not be made between the N-substituted L-alanine analogs (I) and the N-substituted glycine analogs (II).

Calculation of hydrophobic descriptors and its thermodynamic implication

As seen in the previous examples, partition coefficients or hydrophobic parameter are often used in biological QSAR. Hydrophobic parameter is usually expressed as the logarithm of n-octanol-water partition coefficient (log P) along with the Hansch hydrophobic constant π . Although experimentally determined values, such as those from the shake-flask method that has conventionally been used, are always preferred, it is often convenient to use calculated values. Therefore, various methods of calculating partition coefficients have been proposed [28].

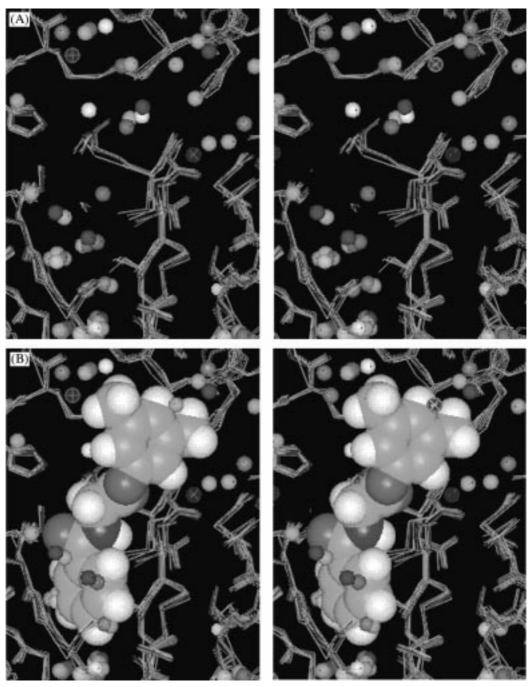


Figure 3. α-chymotrypsin and 3,5-dimethylphenyl ester of N-benzoyl L-alanine analog, PhCONHCH(Me)CO₂C₆H₄-3,5-Me₂ docked into the binding site. (A) and (B) show the binding pocket of eight α-chymotrypsin (1acb, 1cho, 1cgi, 1cgj, 2cha, 4cha, 5cha, 6cha) structures superimposed over the backbone. (A) is shown without a ligand, and (B) is shown with a ligand (3,5-dimethylphenyl ester of N-benzoyl L-alanine analog; PhCONHCH(Me)CO₂C₆H₄-3,5-Me₂) docked into the binding site. Water molecules are shown in balls and shows that should be replaced by the bound ligand. Water molecules in different colors indicate that they are from different α-chymotrypsin structures.

Since the introduction of three-dimensional quantitative structure-activity relationship(3D-QSAR) method [29] such as CoMFA [7, 8, 30], methods of calculating hydrophobic parameters from 3D structure of molecules [31] became increasingly important. In the original CoMFA paper, it was assumed that steric and electrostatic forces are enough to account precisely for a great variety of observed biological activities [10]. However, it was soon realized that some form of hydrophobic field is also quite important in CoMFA [32, 33].

We proposed a novel method of describing hydrophobic effects directly from 3D structures in 3D-QSAR studies using GRID-CoMFA approach [32]. In this approach, the 'hydrophobic' interaction energies were calculated with a water probe using the GRID program [34]. In general, molecular surface area or volume and hydrogen-bond donating and accepting properties have been considered to be the two most important properties contributing toward the log Pvalues of organic compounds. The water probe in GRID is an electrically neutral probe and can donate up to two hydrogen-bonds and also accept up to two hydrogen bonds. As a result, the energies calculated with the water probe were from steric and hydrogenbonding interactions in nature. Thus, it was suggested that the hydrogen-bond donating and accepting components in combination with the steric component in the 'hydrophobic' interaction energies with the water probe might mimic $\log P$ [32, 35].

The water probe used in the GRID-CoMFA approach was originally developed in GRID to locate water binding sites in a protein. Then, one may ask why GRID-CoMFA descriptors, from the interaction energies calculated using the water probe, give such good results as log *P* does. In the thermodynamic point of view, it is possible that the GRID-CoMFA approach may very well describe the desolvation process in the partitioning of molecules.

The hydrophobic parameters of a variety of compounds were calculated using this approach [7]. They include the octanol-water partition coefficients of pyrazines, pyridines, triazines, furans, and benzyl N,N-dimethylcarbamates, as well as the capacity factors ($\log k'$) of a set of furan, benzene, pyrrole, 1-methylpyrrol, benzofuran, indole, 1-methylindole, and orthopramides. Table 1 shows some of the sets we investigated. This approach was applied not only to congeneric series, but also to a mixed set of noncongeneric series, distribution coefficients of diazine analogs of ridogrel and amino acids, hydrophobicity

of cytosine nucleosides, and to the partition coefficients and solubilities of amino acid derivatives [7]. Waller [36] used a similar approach to calculate partition coefficients of structural isomers. Various other applications were discussed in a recent review [7].

The GRID-CoMFA approach of calculating hydrophobicity has been applied not only in calculating parameter values, but also in correlating biological activities [37]. These results compare favorably with those of the classical QSAR [38]. A few examples are shown below.

Description of hydrophobic properties in 3D-QSAR

This GRID-CoMFA approach of calculating hydrophobicity was extended in correlating biological activities. For example, a significant correlation was obtained for the ability of procaine analogs (III) to inhibit valinomycin-induced uptake of potassium (log $1/K_i$) (Equation 13). Equation 12 shows the corresponding traditional QSAR derived from the same data set [35].

$$\log 1/K_i = 0.74(\pm 0.20)\pi + 2.82(\pm 0.17)$$

$$n = 11, r = 0.943, s = 0.218$$
(12)

$$\log 1/K_i = 0.09(\pm 0.01)Z1 + 0.03(\pm 0.01)Z2$$

$$+3.13(\pm 0.01)$$

$$n = 11, r = 0.974, s = 0.156, s(cv) = 0.274$$
(13)

In Equations 13 (and 15 shown below), Z1 and Z2 are the first and the second latent variables obtained from the partial least square analysis in CoMFA.

The correlation for the ability to form the enzymesubstrate complex of papain-catalyzed hydrolysis of substituted phenyl hippurates (IV) provides another example (Equations 14 and 15) [35].

Table 1. Summary of the correlation between log P and log k' for different series of compounds studied. a

Set	L	N	R^2	RMSE	R^2 (cv)	RMSE (cv)	Ref.
$\log P$							
Carbamate	4	19	0.941	0.201	0.384	0.590	40
Furan	5	17	0.982	0.116	0.548	0.500	40
Pyrazine	4	20	0.935	0.186	0.668	0.385	40
Pyridine	5	20	0.974	0.124	0.631	0.406	40
$\log k'$							
Carbamate	4	19	0.939	0.125	0.355	0.368	41
Furan	5	17	0.978	0.093	0.541	0.434	41
Pyrazine	4	19	0.910	0.154	0.518	0.322	41
Pyridine	5	19	0.978	0.072	0.642	0.250	41
Triazine	5	54	0.958	0.100	0.865	0.171	41
Miscellaneous	8	59	0.928	0.158	0.775	0.261	41

 $[^]aL$ = number of latent variable, N = number of compounds used in the correlation; R^2 = squared correlation coefficient, RMSE = root mean square error, R^2 (cv) = squared correlation coefficient for the cross-validation, RMSE (cv) = root mean square error for the cross-validation.

$$\log 1/K_m = 0.44(\pm 0.19)\pi + 4.08(\pm 0.15)$$

$$n = 8, r = 0.919, s = 0.149$$
(14)

$$\log 1/K_m = 0.07(\pm 0.01)Z1 + 4.26(\pm 0.05)$$

$$n = 11, r = 0.919, s = 0.149, s(cv) = 0.201$$
(15)

Additional examples were summarized in a recent review paper [37].

In the thermodynamic point of view, it is possible that the correlations obtained by the GRID-CoMFA approach may very well describe the desolvation process in the partitioning of molecules or the process of a ligand replacing loosely bound water molecules in the ligand binding site of a protein.

Enthalpy or entropy, or both in 3D-QSAR?

In 3D-QSAR such as CoMFA, it is usually assumed that the entropic contributions for structurally related compounds are similar. However, this is not always the case as discussed above. Nonetheless, hundreds of

excellent correlation from CoMFA have been reported in the literature, and many of them are for binding data [7]. How was it possible? What made it possible? In this respect, the work reported by Nakamura et al. [23] is interesting for a close look.

In a study to investigate the factors contributing to the enthalpy and entropy changes of partitioning, Nakamura et al. [23] found that the enthalpy and entropy changes of partition for monosubstituted phenols can be expressed by four parameters, ΔV_w , σ^- , α , and β , where ΔV_w is the van der Waals volume referenced to unsubstituted phenol, σ^- is the Hammett electronic constant for phenols, α and β , are hydrogenbonding related parameters derived from proton donor and proton acceptor scales. Equations 16 and 17 show the correlations.

$$\Delta H_p^0 = -15.63(\pm 2.41) \Delta V_w - 2.08(\pm 0.56) \sigma^- +1.05(\pm 0.67) \alpha - 0.86(\pm 0.38) \beta -7.00(\pm 0.49)$$

$$n = 28, r = 0.92, s = 0.94$$

$$\Delta S_p^0 = 11.74(\pm 7.41) \Delta V_w - 3.32(\pm 1.74) \sigma^-$$

$$-2.29(\pm 2.05)\alpha - 11.19(\pm 1.19)\beta$$

$$-6.30(\pm 1.50)$$

$$n = 28, r = 0.96, s = 8.39$$
(17)

Equation 16 revealed that the molecular size represented by the van der Waals volume (ΔV_w) contributes almost half of the substituent dependence, and the other hydrogen-bonding related terms (σ^- , α , and β) account for the remaining portion. After normalization of all parameters, the following coefficients were obtained for ΔH_p^0 : $-0.71\Delta V_w - 0.35\sigma^- + 0.21\alpha - 0.30\beta$. Nakamura et al. suggested that the term described by the molecular size may reflect the cavity in which ligand molecule is incorporated into the solvent, express the energy of cavity formation, and include the volumes of ligand and solvent molecules, surface tension, and others. The other molecular interaction terms express the energy of ligand-solvent interactions.

Equation 17 shows that the entropy change (ΔS_n^0) can be correlated with the same four parameters with almost equal statistical quality. After normalization, Equation 17 yielded the following coefficients for ΔS_p^0 : $0.13\Delta V_w - 0.13\sigma^- - 0.11\alpha - 0.92\beta$. The hydrogen-bonding ability term β accounts for more than 70% of the variance in ΔS_p^0 , and the other three terms explain the remaining portion. The negative coefficient of β variable means that when β is large and hydrogen-bond acceptor ability is increased, ΔS_n^0 is decreased. It was suggested that when a phenol molecule is transferred from water to micelle, a definite number of water molecules will be released from the solvation sphere to the phenol molecule. But if the ligand-solvent hydrogen-bonding is strengthened, fewer water molecules will be released. That is, a smaller number of water molecules will be released for phenols with higher β value (decreased ΔS_p^0 for phenols with high β value). It is considered that ΔS_n^0 reflects the change in the degree of randomness resulting from the change in the degree of hydrogen-bonding.

Similar analyses with π_H , π_S and π yielded Equations 18, 19 and 20. Equations 18 and 19 correspond to Equations 16 and 17.

$$\pi_{H} = 2.18(\pm 0.34) \Delta V_{w} + 0.40(\pm 0.08) \sigma^{-}$$

$$-0.07(\pm 0.09) \alpha + 0.12(\pm 0.06) \beta$$

$$-0.06(\pm 0.07)$$

$$n = 28, r = 0.93, s = 0.13$$

$$\pi_{S} = 0.71(\pm 0.42) \Delta V_{w} - 0.20(\pm 0.10) \sigma^{-}$$

$$-0.17(\pm 0.12) \alpha - 0.56(\pm 0.07) \beta$$

$$-0.07(\pm 0.08)$$

$$n = 28, r = 0.94, s = 0.13$$
(18)

$$\pi = 2.90(\pm 0.32) \Delta V_w + 0.20(\pm 0.08) \sigma^{-}$$

$$-0.24(\pm 0.09) \alpha - 0.44(\pm 0.05) \beta$$

$$-0.01(\pm 0.06)$$

$$n = 28, r = 0.94, s = 0.13$$
(20)

Equations 18–20 show that ΔV_w and σ^- are more important in π_H whereas ΔV_w and β are more important in π_S . The Hammett σ^- constant shows a compensating effect on π_H and π_S resulting in only slight influence on π , which is equal to $\pi_H + \pi_S$ by definition. ΔV_w , σ^- and β are all important in π .

Considering Equations 16-20 together [showing that both the enthalpic and entropic terms (ΔH_p^0 and ΔS_p^0 or π_H and π_S) can be described by the same four parameters, ΔV_w , σ^- , α , and β], along with other studies [showing that CoMFA can describe the size, electronic, and hydrogen-bonding effects of substituents [39-44], these correlations could provide an explanation why 3D-QSAR such as CoMFA yielded hundreds of excellent correlation even though such an approach only take into account the enthalpic contribution without any account for the entropic contribution. Nonetheless, it does not make sense in terms of physical chemistry to describe entropic contribution by an approach like CoMFA. Besides, quite a high correlation exists between ΔS_p^0 and β (r = 0.948), and between ΔH_p^0 and ΔV_w (r=0.784). (The correlation coefficients are relatively low (r < 0.69) for any pair of ΔV_w , σ^- , α , and β .)

In view of the Nakamura et al.'s [23] study described above, it is interesting to remember the CoMFA study reported by Klebe and Abraham [45]. Based on their CoMFA study of binding properties (expressed as ΔG , ΔH , and ΔS) of renin inhibitors, Klebe and Abraham concluded that only enthalpies (ΔH) can be properly predicted by CoMFA, and not the corresponding ΔG , ΔS , or the binding constants. They suggested that a prediction of ΔG (or $\ln K$) would only be possible if the entropic contribution to ΔG is negligible or constant over the entire data set. In the meantime, Steinmetz [46] reported excellent correlations with ΔG of a number of phosphorus ligand sets.

Conclusion

Several examples presented in this paper clearly demonstrate that thermodynamic analyses in QSAR

can provide useful information about the binding of drug molecule into a protein.

The necessity to explicitly take into account the entropic effects into 3D QSAR studies is well recognized [37, 39]. However, the ability of 3D-QSAR methodology such as CoMFA to describe the entropic contribution to the free energy of binding has not yet been well sorted out.

Nevertheless, new insights could be obtained considering the waters surrounding macromolecules along with the traditional QSAR results. At present, only limited attention has been paid in this field, and the appropriate amount of data are scarce in the literature. It is expected that rich information can be extracted from both thermodynamic analyses and protein structures including water molecules. Molecular modeling can play an important role in this respect.

References

- Karplus, P.A. and Faerman, C., Curr. Opinion Struct. Biol., 4 (1994) 770.
- 2. Ringe, D., Curr. Opinion Struct. Biol., 5 (1995) 825.
- Fitzpatrick, P.A., Steinmetz, A.C.U., Ringe, D. and Klibanov, A.M., Proc. Natl. Acad. Sci. USA, 90 (1993) 8653.
- 4. Reslow, M., Adlercreutz, P. and Mattiasson, B., Appl. Microbiol. Biotechnol., 26 (1987) 1.
- Stevenson, D.E. and Storer, A.C., Biotechnol. Bioeng., 37 (1991) 519.
- Allen, K.N., Bellamacina, C.R., Ding, X., Jeffery, C.J., Mattos, C., Petsko, G.A. and Ringe, D., J. Phys. Chem., 100 (1996) 2605.
- Kim, K.H., Greco, G. and Novellino, E., in Kubinyi, H., Golkers, G. and Martin, Y.C. (Eds) A Critical Review on Recent CoMFA Applications. 3D QSAR in Drug Design. Recent Advances, Vol. 3. Kluwer Academic Publishers, Dordrecht, 1998, pp. 257–315.
- Martin, Y.C., Kim, K.H. and Lin, C.T., in Charton, M. (ed.), Comparative Molecular Field Analysis: CoMFA. Advances in Quantitative Structure-Property Relationships, Vol. 1. JAI Press Inc., Greenwich, Connecticut, 1996, pp. 1–52.
- Bohm, H.-J. and Klebe, G., Angew. Chem. Int. Ed. Engl., 35 (1996) 2588.
- Cramer, R.D. III, Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
- Gilli, P., Ferretti, V., Gilli, G. and Borea, P.A., J. Phys. Chem., 98 (1994) 1515.
- 12. Hitzemann, R., TIPS, 9 (1988) 408.
- Leffler, J.E. and Grunwald, E., Rates and Equilibria of Organic Reactions. Wiley, New York, NY, 1963.
- Sleigh, S.H., Seavers, P.R., Wilkinson, A.J., Ladbury, J.E. and Tame, J.R.H., J. Mol. Biol., 291 (1999) 393.
- 15. Dunitz, J.D., Chem. Biol., 2 (1995) 709.
- 16. Lemieux, R.U., Acc. Chem. Res., 29 (1996) 373.
- 17. Rogers, J.A. and Wong, A., J. Pharm., 6 (1980) 339.
- Gallicchio, E., Kubo, M.M. and Levy, R.M., J. Am. Chem. Soc., 120 (1998) 4526.

- Krug, R.R., Hunter, W.G. and Grieger, R.A., Nature, 261 (1976) 566.
- Krug, R.R., Hunter, W.G. and Grieger, R.A., J. Phys. Chem., 80 (1976) 2335.
- Krug, R.R., Hunter, W.G. and Grieger, R.A., J. Phys. Chem., 80 (1976) 2341.
- Da, Y.-Z., Ito, K. and Fujiwara, H., J. Med. Chem., 35 (1992)
 3382
- Nakamura, K., Hayashi, K., Ueda, I. and Fujiwara, H., Chem. Pharm. Bull., 43 (1995) 369.
- 24. Selassie, C.D., Gan, W.X., Fung, M. and Shortle, R., in Sanz, J.G.F. and Manaut, F. (eds), Chymotrypsin-ligand interactions in non-aqueous solvents. QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications, Proceedings of the 10th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling, Barcelona, September 4–9, 1994. J. R. Pros Science Publishers, Barcelona, 1994, pp. 128–130.
- 25. Kim, K.H., Bioorg. Med. Chem., 9 (2001) in press.
- Blaney, J.M. and Hansch, C., in Ramsden, C.A. (ed.), Application of Molecular Graphics to the Analysis of Macromolecular Structures. Quantitative Drug Design. Pergamon, Oxford, 1990, pp. 459–496.
- Compadre, C.M., Sanzhez, R.I., Bhuvaneswaran, C., Compadre, R.L., Plunkett, D. and Novick, S.G., in Wermuth, C.G. (ed.), Analysis of enzyme-ligand interactions in organic solvents: A QSAR approach. Trends in QSAR and Molecular Modelling 92, Proceedings of the 9th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modeling. September 7–11, 1992, Strasbourg, France. ESCOM, Leiden, 1993, pp. 112–115.
- 28. Leo, A.J., Chem. Rev., 93 (1993) 1281.
- Greco, G., Novellino, E. and Martin, Y.C., in Martin, Y.C. and Willet, P. (eds), Approaches to Three-Dimensional Quantitative Structure-Activity Relationships. Designing Bioactive Molecules: Three-Dimensional Techniques and Applications. American Chemical Society, Washington, DC, 1997, pp. 219–252.
- Kim, K.H., in Dean, P.M. (ed.), Comparative Molecular Field Analysis (CoMFA). Molecular Similarity in Drug Design. Blackie Academic & Professional, London, 1995, pp. 291–331.
- Abraham, D.J. and Kellogg, G.E., in Kubinyi, H. (ed.), Hydrophobic Fields. 3D QSAR in Drug Design. Theory Methods and Applications. ESCOM, Leiden, 1993, pp. 506–522.
- 32. Kim, K.H., Med. Chem. Res., 1 (1991) 259.
- Cramer III, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., in Kubinyi, H. (ed.), The Developing Practice of Comparative Molecular Field Analysis. 3D QSAR in Drug Design. ESCOM, Leiden, 1993, pp. 443–485.
- GRID program. Molecular Discovery Ltd., West Way House, Elms Parade, Oxford, U.K.
- 35. Kim, K.H., Quant. Struct.-Act. Relat., 12 (1993) 232.
- 36. Waller, C.L., Quant. Struct.-Act. Relat., 13 (1994) 172.
- Kim, K.H., in Kubinyi, H. (ed.), Comparison of Classical and 3D QSAR. 3D QSAR in Drug Design. Theory Methods and Applications. ESCOM, Leiden, 1993, pp. 619.
- 38. Klebe, G. and Abraham, U., J. Med. Chem., 36 (1993) 70.
- 39. Steinmetz, W.E., Quant. Struct.-Act. Relat., 15 (1996) 1.
- Wade, R.C., in Kubinyi, H. (ed.), Molecular Interaction Fields.
 QSAR in Drug Design. ESCOM, Leiden, 1993, pp. 486–506.
- 41. Kim, K.H., Quant. Struct.-Act. Relat., 11 (1992) 127.

- 42. Kim, K.H., Greco, G., Novellino, E., Silipo, C., Vittoria, A., J. Comput. Aided Mol. Des., 7 (1993) 263.
- 43. Kim, K.H., in Wermuth, C.G. (ed.), Use of the hydrogen-bond potential function on comparative molecular field analysis (CoMFA): An extension of CoMFA. Trends in QSAR and Molecular Modelling 92, Proceedings of the 9th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling. September 7–11, 1992, Strasbourg, France. ESCOM, Leiden, 1993, pp. 245–251.
- 44. Klebe, G. and Abraham, U., J. Med. Chem., 36 (1993) 70.
- 45. Steinmetz, W.E., Quant. Struct. Act. Relat., 15 (1996) 1.
- Wade, R.C., in Kubinyi, H. (ed.), Molecular interaction fields.
 3D QSAR in Drug Design. Theory Methods and Applications.
 ESCOM, Leiden, 1993, pp. 486–506.
- 47. Kim, K.H., J. Comput.-Aid. Mol. Design, 9 (1995) 308.
- 48. Kim, K.H. and Kim, D.H., Bioorg. Med. Chem., 3 (1995)