# Comparative Binding Energy (COMBINE) Analysis of OppA−Peptide Complexes to Relate Structure to Binding Thermodynamics

Ting Wang and Rebecca C. Wade*

*European Media Laboratory, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany, and European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany*

*Received April 5, 2002*

The periplasmic oligopeptide binding component (OppA) of the oligopeptide permease found in Gram-negative bacteria acts as a receptor for peptide transport across the cell membrane and is a potential target for antibacterial drug design. OppA exhibits broad specificity, binding to diverse peptides of 2−5 amino acid residues length. Crystallographic and calorimetric measurements have been carried out by Tame et al. of the binding of 28 peptides of sequence K-X-K to OppA, where X is a natural or nonnatural amino acid. Despite this extensive experimental characterization, a clear relationship between structural and thermodynamic parameters could not be readily identified, with a complicating factor being the observation of varying numbers of water molecules at the binding interface in the different complexes. Consequently, we have applied COMparative BINding Energy (COMBINE) analysis to derive quantitative structure−activity relationships (QSARs) for these 28 OppA−tripeptide complexes. This is the first application of COMBINE analysis to predict binding enthalpies and entropies, and predictive QSAR models were obtained for these quantities as well as for binding free energies. These QSAR models highlight several protein residues and bound water molecules in the binding site, as well as the electrostatic desolvation energies of the protein and the peptides, as responsible for most of the differences in binding thermodynamics between the peptides studied. The QSAR models aid rationalization of the determinants of binding affinity of the OppA:peptide complexes and provide guides for further ligand design. This study also points to the general applicability of COMBINE analysis to estimating thermodynamic parameters for protein−peptide complexes.

## Introduction

The periplasmic oligopeptide binding protein (OppA) of the oligopeptide permease acts as the initial receptor for peptide transport across the cell membrane in Gram-negative bacteria. To supply the cell with a variety of peptide nutrients, OppA has broad specificity, binding to peptides of 2−5 amino acid residues length and diverse sequence.[1−3] A quantitative structure−activity relationship (QSAR) for OppA−peptide binding would aid rational drug design in two ways. First, a QSAR model could guide design of peptide-based antibiotics to bind to OppA so that they can be transported into the cell to take effect. Second, a QSAR model could guide design of OppA inhibitors that block OppA-mediated peptide transport. The latter strategy may, for example, be useful against Lyme disease, the causative agent of which is *Borrelia burgdorferi.* Genome sequence analysis has shown that the *B. burgdorferi* bacterium has no gene for amino acid synthesis.[4] OppA is therefore critical for the uptake of nutrients in this bacterium and consequently a potential target for inhibitor design.

Recently, the binding to the OppA from *Salmonella typhimurium* of a series of 28 tripeptides has been studied by X-ray crystallography and isothermal titration calorimetry (ITC).[5−8] These peptides have the sequence Lys-X-Lys, where X is one of 20 naturally occurring amino acids and eight nonnatural residues (see Figure 1). The nonnatural residues have closely related structures to the natural residues. Orn, Dab, and Dap mimic lysine but have consecutively shorter side chains; Nva and Nle have unbranched alkane chains that are longer than that of alanine by one and two carbon atoms, respectively; Hph has a side chain one carbon atom longer than phenylalanine; Nap is similar to tryptophan but has a naphthalene ring; and Chx is constructed by replacing the phenyl ring in phenylalanine by a cyclohexane ring. The crystal structures of the OppA−tripeptide complexes were solved to good resolution (1.75−2.30 Å). The crystallographic studies demonstrate that OppA undergoes a conformational change upon binding a peptide from an open to a closed form. They show that all of the tripeptides bind to OppA in equivalent extended conformations with the bound peptide entirely buried in the interior of OppA. The main chain and the N and C termini of the peptides form strong parallel and antiparallel $\beta$-sheet hydrogen bond networks with the binding site residues of OppA. The side chains make few direct interactions with OppA (see Figure 2). Instead, ordered water molecules are found around the side chains of the bound peptides and are believed to play a role in mediating binding.[9−11]

* To whom correspondence should be addressed. Tel: +49 6221 533247. Fax: +49 6221 533298. Email: rebecca.wade@eml.villa-bosch.de.

10.1021/jm020900l CCC: $22.00 
Published on Web 09/21/2002

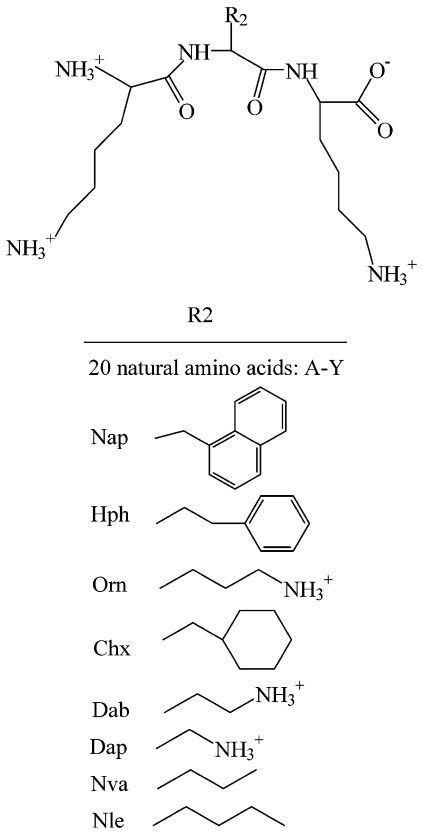
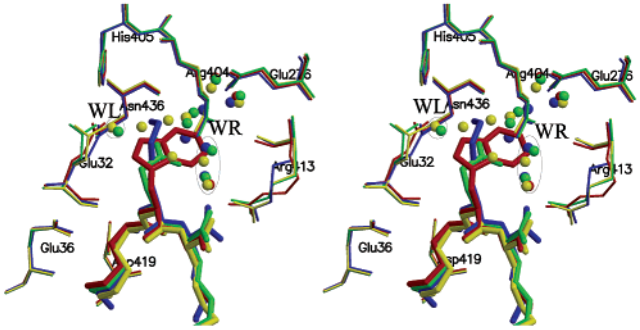



**Figure 1.** Structures of the 28 tripeptides of sequence Lys-X-Lys. Only the central residue R2 is variant.




| | #water | Glu32 |
|---|---|---|
| KGK | 9 | B conformation |
| KEK | 7 | B conformation |
| KKK | 5 | A conformation |
| KWK | 3 | A conformation |

**Figure 2.** Stereoview of the overlay of four OppA−peptide complexes (KGK, KEK, KKK, and KWK). Only the active site residues of OppA are shown and labeled. The peptides are shown in thick sticks. Glu32 is the only residue showing large conformational variability in the different complexes, adopting two alternative conformations (A and B). The number of bound water molecules in the active site varies with the ligand. These water sites were partitioned into three regions for COMBINE analysis: left (WL), right (WR), and central (WC, any water sites not in either WL or WR). See Materials and Methods section for details.

The binding parameters of all 28 OppA−tripeptide complexes have been obtained by ITC.[5,6] The dissociation constant, $K_d$, and the enthalpy change, $\Delta H$, were directly measured from the ITC experiments. The free energy change, $\Delta G$, and the entropy change, $\Delta S$, were calculated from the measured quantities (see Table 1). This extensive data set provides an opportunity to investigate SARs in the OppA−peptide system in detail. However, SARs for OppA−peptide binding are not straightforward to identify. In Table 1, it can be seen that binding of all of the peptides is accompanied by unfavorable enthalpy changes ($\Delta H$) and favorable entropy changes ($\Delta S$); therefore, binding is entropy driven. Both enthalpy and entropy changes span a large range ($\Delta H$: 62 and $T\Delta S$: 57 kJ mol$^{-1}$ at 298 K), but the free energy changes ($\Delta G$) differ relatively little (~12 kJ mol$^{-1}$) among the 20 natural tripeptides, and both the weakest (−25.9 kJ mol$^{-1}$) and the tightest (−46.1 kJ mol$^{-1}$) peptide binders contain nonnatural residues. In addition, the numbers of bound water molecules in the binding site were observed to vary with the ligands, suggesting that water molecules may play an important role in determining protein−peptide binding. Davies et al.[6] tried to correlate binding affinity with structure by using empirical scoring functions, including the LUDI method,[12] but failed to obtain a good correlation.

The lack of strong selectivity between ligands, the flexibility of the binding partners, and the variability in interfacial hydration make the OppA−peptide system a demanding one for structure-based prediction of binding affinities. Global empirical scoring functions, such as LUDI, are designed with the aim of being applicable to all systems. However, such scoring functions display different levels of performance for different systems and are limited in accuracy.[13] In principle, computations based on free energy perturbation theory, exploiting molecular dynamics simulations for sampling of conformations, can account for all of the relevant features of the OppA−peptide complexes and should be applicable to computing binding free energy differences.[14−16] On the other hand, such calculations are computationally demanding and thus difficult to apply in ligand design projects. Here, we report the application of COMparative BINding Energy (COMBINE) analysis[17,18] to the OppA−peptide system. The aim is to derive a system specific QSAR model for estimating binding free energy (or enthalpy or entropy) differences based on energy-minimized structures of ligand−receptor complexes. This is achieved through fitting of a model using experimental binding parameters for a set of receptor−ligand complexes before applying the model to predicting binding parameters of further ligands. COMBINE analysis can be applied to large data sets with modest computational requirements.

The COMBINE method is based upon the assumption that the binding free energy ($\Delta G$) can be correlated with a subset of energy components determined from the structures of receptors and ligands in bound and unbound forms. In this study, the energy terms computed are the electrostatic desolvation energies of OppA and the peptides upon binding, $\Delta G^{R}_{desol}$ and $\Delta G^{L}_{desol}$, respectively, and the Coulombic interactions, $\Delta u_i^{ele}$, and Lennard−Jones interactions, $\Delta u_i^{vdw}$, between each ligand and each protein residue (or interfacial water region) in energy-minimized structures of OppA−

**Table 1.** Experimental and Computed Parameters for the OppA−Tripeptide KXK Complexes[a]

| no. | PDB code | peptide | $\Delta G_{exp}$ (kJ/mol) | $\Delta G_{pred}$ (kJ/mol) | $\Delta H_{exp}$ (kJ/mol) | $\Delta H_{pred}$ (kJ/mol) | $T\Delta S_{exp}$ (kJ/mol) | $T\Delta S_{pred}$ (kJ/mol) | pred $\Delta H - T\Delta S$ (kJ/mol) | Desol_P (kJ/mol) | Desol_L (kJ/mol) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1jet | KAK | −41.1 | −43.4 | 20.1 | 11.9 | 61.1 | 50.3 | −38.43 | 212.81 | 507.57 |
| 2 | 1b05 | KCK | −40.6 | −42.5 | 7.9 | 15.3 | 48.2 | 53.9 | −38.54 | 216.76 | 512.99 |
| 3 | 1b4z | KDK | −29.8 | *−51.0* | 8.1 | *−1.1* | *37.7* | *39.1* | −40.21 | 228.44 | 534.32 |
| 4 | 1jeu | KEK | −38.9 | −47.6 | 11.3 | 12.1 | 50.0 | 48.2 | −36.1 | 245.57 | 547.81 |
| 5 | 1b40 | KFK | −41.5 | −41.4 | 22.0 | 17.7 | 63.2 | 54.8 | −37.12 | 218.27 | 532.10 |
| 6 | 1b3l | KGK | −33.6 | −43.2 | 14.1 | 4.5 | 47.5 | 46.1 | −41.64 | *198.57* | *495.56* |
| 7 | 1b3f | KHK | −39.3 | −40.0 | 20.6 | 18.0 | 59.8 | 60.5 | *−42.54* | 231.71 | 503.16 |
| 8 | 1b3g | KIK | −38.3 | −41.4 | 20.5 | 13.8 | 58.6 | 54.3 | −40.44 | 228.90 | 497.45 |
| 9 | 2olb | KKK | −31.6 | −31.7 | 39.4 | 48.0 | 70.8 | 76.5 | −28.52 | 228.90 | 661.54 |
| 10 | 1b9j | KLK | −33.9 | −40.9 | 24.6 | 17.2 | 58.3 | 56.2 | −39.04 | 224.91 | 515.76 |
| 11 | 1b32 | KMK | −40.5 | −39.7 | 14.6 | 24.1 | 54.9 | 60.7 | −36.65 | 232.13 | 528.78 |
| 12 | 1b5i | KNK | −40.2 | −40.6 | *7.7* | 11.7 | 47.9 | 51.6 | −39.86 | 227.47 | 512.95 |
| 13 | 1b46 | KPK | −30.1 | −42.9 | 16.6 | 12.0 | 46.5 | 51.5 | −39.47 | 214.58 | 511.14 |
| 14 | 1b5j | KQK | −42.4 | −40.2 | 11.4 | 24.1 | 53.4 | 61.9 | −37.87 | 222.60 | 540.71 |
| 15 | 1qka | KRK | −33.8 | −31.6 | 36.0 | 38.5 | 69.7 | 70.8 | −32.32 | 218.61 | 651.59 |
| 16 | 1b51 | KSK | −42.0 | −43.0 | 8.9 | 17.2 | 50.6 | 54.3 | −37.09 | 215.16 | 527.48 |
| 17 | 1b52 | KTK | −40.6 | −42.0 | 17.3 | 15.9 | 57.6 | 53.9 | −38.02 | 217.85 | 515.93 |
| 18 | 1qkb | KVK | −41.9 | −40.2 | 22.4 | 16.9 | 64.3 | 56.1 | −39.28 | 229.57 | 508.33 |
| 19 | 1jev | KWK | −39.2 | −37.6 | 29.3 | 30.8 | 68.2 | 68.7 | −37.91 | *247.72* | 528.36 |
| 20 | 1b58 | KYK | −37.5 | −37.3 | 20.7 | 24.9 | 57.9 | 62.7 | −37.87 | 242.09 | 527.85 |
| 21 | 1b0h | K-Nap-K | −38.2 | −39.9 | 20.8 | 18.9 | 59.0 | 59.3 | −40.41 | 238.81 | 501.61 |
| 22 | 1b1h | K-Hph-K | −40.3 | −39.9 | 7.20 | 18.6 | 47.4 | 57.8 | −39.3 | 226.13 | 512.48 |
| 23 | 1b2h | K-Orn-K | *−25.9* | *−30.1* | *69.4* | 43.1 | *95.3* | *82.6* | −39.46 | 241.04 | 653.39 |
| 24 | 1b3h | K-Chx-K | −35.4 | −39.6 | 20.4 | 20.2 | 56.6 | 60.3 | −40.17 | 234.23 | 508.24 |
| 25 | 1b4h | K-Dab-K | −31.1 | −30.5 | 44.4 | 45.0 | 75.4 | 80.2 | −35.29 | 231.34 | 695.69 |
| 26 | 1b5h | K-Dap-K | −34.3 | −32.6 | 44.7 | *48.1* | 79.0 | 74.8 | *−26.71* | 224.91 | *697.28* |
| 27 | 1b6h | K-Nva-K | −44.7 | −40.3 | 7.93 | 18.1 | 52.4 | 57.2 | −39.07 | 229.82 | 505.60 |
| 28 | 1b7h | K-Nle-K | *−46.1* | −41.0 | 19.6 | 16.9 | 65.5 | 54.7 | −37.79 | 221.42 | 515.26 |

[a] All experimental values are from refs 5 and 6. All thermodynamic measurements were carried out in 50 mM sodium phosphate buffer at pH 7.0 and 298 K. Desol_P and Desol_L are the computed electrostatic desolvation energies of the protein and ligand (peptide), respectively. Highest and lowest values of each parameter are shown in italic.

peptide complexes (see Materials and Methods section for details). The binding free energy, $\Delta G$, was estimated as a weighted linear sum of these energy terms as given in eq 1.

$$\Delta G = w^{R}_{desol}\Delta G^{R}_{desol} + w^{L}_{desol}\Delta G^{L}_{desol} + \sum_i w_i^{vdw}\Delta u_i^{vdw} + \sum_i w_i^{ele}\Delta u_i^{ele} + C \quad (1)$$

The contribution of each interaction energy term is represented by its weight, namely, the parameter $w_i^{vdw}$, $w_i^{ele}$, $w^{R}_{desol}$, or $w^{L}_{desol}$ in eq 1, which is obtained by PLS (partial least squares) analysis.

The COMBINE method has been proved successful for deriving high quality QSAR models for a variety of protein−ligand complexes including enzyme−inhibitor,[19−22] enzyme−substrate,[23,24] and nuclear receptor−DNA complexes.[25] The OppA−peptide system is a novel test for the COMBINE method. It is challenging because of its broad ligand specificity and the variation in numbers of bound water molecules in the binding site. On the other hand, the complete experimental data set for 28 peptides of free energy ($\Delta G$), enthalpy ($\Delta H$), and entropy ($T\Delta S$) of binding provide a unique opportunity to derive QSAR models for all three thermodynamic parameters by COMBINE analysis and to investigate the relationships between these models. Indeed, this is the first time that COMBINE analysis has been used to derive models for $\Delta H$ and $\Delta S$ and this data set provides an opportunity to assess the applicability of the COMBINE analysis procedure to these thermodynamic quantities. In addition to deriving a QSAR model for $\Delta G$ using eq 1, we derived predictive models for $\Delta H$ and $T\Delta S$ by substituting $\Delta H$ or $T\Delta S$ values, respectively, for $\Delta G$ values in eq 1. These COMBINE models serve to resolve some of the open questions about the determinants of peptide binding to OppA.

## Materials and Methods

**Molecular Mechanics Modeling.** The 28 crystal structures of OppA binding to tripeptides of sequence Lys-X-Lys were retrieved from the Brookhaven Protein Data Bank (PDB). The uranium ions and acetate ions ($CH_3-COO^-$) found in the crystal structures of some complexes were assumed unimportant for this study and were removed. Thus, for each complex, the coordinates of the OppA protein of 517 residues, the tripeptide ligand, and a number of ordered water molecules were used in this study.

The two cysteine residues in OppA were defined to make a disulfide bond. The polar hydrogen atoms of OppA, the ligand, and the water molecules were assigned by using the WHATIF program.[26−27] The protonation states of the 11 histidines in OppA and the histidine in the KHK ligand were determined as the following: six histidines (29, 91, 142, 161, 405, and 440) in OppA were protonated on H$\epsilon$ and the other five (55, 75, 117, 371, and 517) on H$\delta$. The histidine in the KHK ligand was protonated on H$\epsilon$. The amino acid side chains of arginine, lysine, aspartate, and glutamate residues of both OppA and peptide, as well as the N and C termini of the peptides, were treated as ionized. The amino groups on the nonnatural residues (Dab, Dap, and Orn) were protonated ($-NH_3^+$).

Because of poor resolution, Asp356 and Lys357 are missing in the crystal structures of the complexes 1b52/KTK and 1b05/KCK. These residues were modeled by using the complex 2olb/KKK as a reference structure.

The terminal nitrogen atom of the side chain of Dab has two conformations in the complex 1b4h/K-Dab-K, and conformation A was arbitrarily selected for modeling. In addition, most complexes contain partially occupied closely spaced pairs of water sites in the binding pocket. For these pairs, only the one with the least steric clash was treated as occupied.

The all atom AMBER 95 force field[28] was used to obtain all of the parameters for the protein, the ligands, and the water

molecules except the atomic partial charges of the side chains of the nonnatural residues (Nap, Hph, Orn, Chx, Dab, Dap, Nva, and Nle). For most nonnatural residues, the partial charges were assigned by analogy to their natural counterparts: Dab, Dap, and Orn referred to lysine; Hph referred to phenylalanine; and Nva and Nle referred to alanine. For Nap and Chx, the atomic partial charges were assigned by using the AMBER potential in the InsightII (version 2000) package.[29]

The xLEaP module of AMBER6.0[30] was used to obtain the topology and coordinate files of each complex. Then, energy minimization was performed by following the same protocol as described in ref 22.

**Binding Energy Decomposition for Chemometric Analysis and Treatment of Water.** In the COMBINE analysis, the interaction energy (arising from Coulombic and Lennard−Jones interactions) between the receptor and the ligand was decomposed on a per residue basis by using the program COMBINE1.0 (provided by A. R. Ortiz). The AMBER topology and coordinate files of the minimized complex were used as input and a distance-dependent dielectric constant, $\epsilon = r$, was used for calculating the Coulombic interaction energies.

The interfacial water molecules were treated as extra protein "residues" while the other ordered water molecules were discarded from the minimized structures. The COMBINE analysis procedure requires the same number of energy descriptors for all of the complexes. Consequently, all of the complexes must have the same number of water residues. This meant that due to the variation in the number and position of water molecules in different OppA−peptide complexes, each water molecule could not be assigned to its own individual residue. Instead, the water molecules in each complex were partitioned into three spatially defined regions, each of which was treated as a residue. These three regions are, with respect to the central residue of the peptide, the left (WL), central (WC), and right (WR) parts of the interfacial water (see Figure 2).

The left region, WL, contains one or no water molecule, depending on the conformation of Glu32 of OppA. Glu32 is the only residue showing large conformational variability in the different complexes and adopts two alternative conformations. In five complexes, it adopts conformation A (1b2h/K-Orn-K, 1b3f/KHK, 1qka/KRK, 1jev/KWK, and 2olb/KKK) and in the other 23 complexes, it adopts conformation B. The water WL forms a hydrogen bond with Glu32 in conformation B but is released from the binding site by Glu32 in conformation A. The right region, WR, contains two water molecules that make a salt link to Arg404, but one water molecule is displaced by the big aromatic side chains of Trp and Nap in complexes 1jev and 1b0h. The central region WC contains the water molecules close to the central residue of the tripeptide, which vary in number from two to six. The partition of water molecules and the alternative conformations of Glu32 are shown in Figure 2, using complexes 1b3l/KGK, 1jeu/KEK, 2olb/KKK, and 1jev/KWK as examples. WL, WC, and WR were considered as three extra residues of the protein in COMBINE analysis.

**Electrostatic Desolvation Energy Calculation.** The crystallographic studies demonstrate that OppA undergoes a conformational change from the open form before binding to the closed form after binding and that OppA completely encloses the ligand. The desolvation effect is believed to play an important role in OppA−peptide binding. The electrostatic contribution to the desolvation energy of the protein, $\Delta G^{R}_{desol}$ (or the ligand, $\Delta G^{L}_{desol}$), was defined as the loss of the electrostatic interaction energy between the solvent and the protein (or the ligand) upon binding, as calculated by the two step procedure proposed by Perez et al.:[20] (i) calculate the electrostatic interactions between the protein (or the ligand) and the surrounding solvent in the absence of the ligand (or the protein), (ii) calculate the electrostatic interactions between the protein (or the ligand) and the surrounding solvent with ligand (or protein) bound but in the absence of the partial charges of the ligand (or the protein). The electrostatic desolvation energy ($\Delta G^{R}_{desol}$ or $\Delta G^{L}_{desol}$) is the difference between the electrostatic energies computed from the two steps.

The program UHBD6.1[31] was used to implement the continuum electrostatic calculations by solving the Poisson−Boltzmann equation using a finite difference method. The interior dielectric constant of both protein and ligand was set to 2, and the solvent dielectric constant was set to 78 with an ionic strength of 150 mM and ionic radius of 1.5 Å. A probe of radius 1.4 Å and a surface with 400 points per Å$^2$ were used to calculate the solvent accessible surface and define the dielectric boundaries. The coarse grid spacing was set to 0.80 Å, and the fine grid spacing was set to 0.225 Å. Both the coarse grid and the fine grid were dimensioned to 110 × 110 × 110.

Before doing the UHBD calculations, the minimized structures of all of the complexes were superposed with the minimized structure of the unbound OppA (PDB entry code 1rkm) to ensure the same reference coordinates. Then, a separate source code was written to convert the superposed structures of the complexes to qcd format files for input to UHBD6.1, with all water molecules removed. Both steps one and two above used the bound forms of proteins and peptides in complexes. The calculated $\Delta G^{R}_{desol}$ and $\Delta G^{L}_{desol}$ values are listed in Table 1.

**Chemometric Analysis.** In the 28 OppA−tripeptide complexes, the ligands are closely superimposable and structural variance of the ligands takes place only in the second residue while the other two residues make identical interactions with OppA. Therefore, only the interactions between the second residue of the ligand and each of the 517 protein residues and each of the three extra water residues WL, WC, and WR were taken into account. (Similar results were obtained when the interactions of the whole ligand, or of all three residues individually, were computed.) Each complex was described by 520 Coulombic energy variables, 520 Lennard−Jones energy variables, and two desolvation energies, totaling 1042 *x*-variables. The *y*-variable was assigned as enthalpy change, $\Delta H$, entropy change, $T\Delta S$, or free energy change, $\Delta G$. The GOLPE4.5.1 program[32] was used to carry out the chemometric analysis.

First, a principal component analysis (PCA) was performed to investigate the distribution of the 28 complexes in the energy space defined by the 1042 *x*-variables. The distances between complexes were measured by the PCA scores. Then, a PLS analysis was carried out to correlate the *x*-variables and the *y*-variable and yielded initial PLS models of varying dimensionality. In this step, the *x*-variables showing no variation in the complexes (the sum of the squares of derivation from the average value is smaller than $10^{-7}$ kcal mol$^{-1}$ ($4.2 \times 10^{-7}$ kJ mol$^{-1}$)) were removed and about 800 *x*-variables were maintained to build the initial PLS models. The *x*-variables were not scaled, and tests of scaling procedures showed that they did not result in better models. Leave-one-out cross-validation $Q^2$ values were in the range of 0.5−0.6. To remove the noisy variables and improve the predictive abilities of the PLS models, an *x*-variable selection procedure consisting of a D-optimal preselection and a fractional factorial design (FFD) was performed for up to five latent variables. The D-optimal preselection removed nearly half of the *x*-variables without affecting model quality, and the FFD further removed a few *x*-variables while retaining uncertain variables. Less than 400 *x*-variables were left to build the final PLS models. These models have significantly higher $Q^2$ values and slightly higher $R^2$ values than the initial PLS models. The results of leave-one-out cross-validation are given in Table 2. Tests were performed with leave-two-out cross-validation and with cross-validation on five randomly chosen groups, and slightly lower $Q^2$ values were obtained.

## Results and Discussion

**PCA.** First, a PCA of the matrix of computed energy terms for the 28 complexes was performed. The score plot of the first two PCs (PC1 and PC2) is shown in Figure 3. Most complexes cluster together in this plot

**Table 2.** Predictive Performances of the COMBINE Models[a]

| data set | no. complexes | #LV | $R^2$ | $Q^2$ | SDEP (kJ/mol) | constant $C$ (kJ/mol) |
|---|---|---|---|---|---|---|
| $\Delta H$ | 28 | 3 | 0.76 | 0.64 | 8.36 | −86.90 |
| | | *4* | *0.83* | *0.66* | *8.15* | *−131.40* |
| | | 5 | 0.85 | 0.62 | 8.60 | −108.50 |
| $T\Delta S$ | 28 | 3 | 0.78 | 0.68 | 6.55 | −32.35 |
| | | *4* | *0.83* | *0.73* | *6.02* | *−53.37* |
| | | 5 | 0.86 | 0.74 | 5.88 | −40.36 |
| $\Delta G$ | 23[b] | 2 | 0.74 | 0.61 | 2.84 | −49.55 |
| | | *3* | *0.83* | *0.74* | *2.34* | *−75.87* |
| | | 4 | 0.85 | 0.73 | 2.39 | −68.60 |

[a] #LV is the number of latent variables of the model. The models with the optimum number of latent variables are shown in italic. $Q^2$ is the cross-validated predictive performance and is given by $Q^2 = 1 - [\sum_{i=1}^{n}(y_{\exp(i)} - y_{\mathrm{pred}(i)})^2]/[\sum_{i-1}^{n}(y_{\exp(i)} - \langle y_{\exp}\rangle)^2]$ where $y_{\mathrm{pred}(i)}$ corresponds to the value of the quantity predicted with the model for complex $i$, $y_{\exp(i)}$ is the experimental value of the quantity for complex $i$, and $\langle y_{\exp}\rangle$ is the average experimental value of the quantity for the complete set of $n$ complexes. $R^2$ is the equivalent of $Q^2$ calculated for fitting. SDEP is the standard deviation in cross-validated prediction and is given by SDEP $= \{[\sum_{i=1}^{n}(y_{\exp(i)} - y_{\mathrm{pred}(i)})^2]/n\}^{1/2}$. The constant $C$ is as given in eq 1 for each COMBINE model. [b] The five outliers in the $\Delta G$ models are the complexes **3**, **4**, **6**, **10**, and **13** with central residues D, E, G, L, and P, respectively.

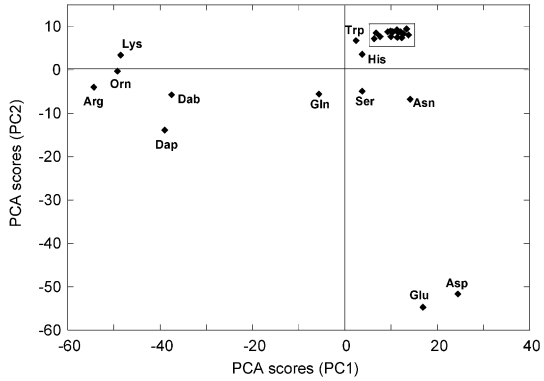



**Figure 3.** Score plot of the first (PC1) and second (PC2) PCs for the 28 complexes. Each complex is labeled by the name of the central residue R2.

and have positive scores for both PC1 and PC2. The exceptions are the three ligands with central residues Gln, Ser, and Asn, as well as the seven ligands with charged central residues. The ligands with negatively charged central residues (Asp and Glu) and the ligands with positively charged central residues (Lys, Orn, Dab, Dap, and Arg) are separated distinctly by both PC1 and PC2, indicating their significant differences in the energy space defined by the 1042 energy variables. PC2 appears to separate the positively charged central residues by size. In PC3 and PC4 (not shown), the distribution of the complexes is different and cannot easily be related to physicochemical properties such as charge and size. As in PC1 and PC2, the peptides with nonpolar central residues tend to cluster together in PC3 and PC4. For brevity, the unique name of the central residue of the tripeptide was used to represent the ligand and the complex in the following sections.

**PLS Models.** PLS analysis to correlate the computed energy terms with $\Delta H$, $T\Delta S$, and $\Delta G$ values was performed as described in the Materials and Methods section. The statistical parameters of the PLS models for $\Delta H$, $T\Delta S$, and $\Delta G$ correlations are given in Table 2. Both $\Delta H$ and $T\Delta S$ models are predictive in cross-validation of models constructed for all 28 complexes. On the other hand, when the 28 complexes are used to derive a model for $\Delta G$, the predictive ability is poor ($Q^2 < 0.2$). A predictive model for $\Delta G$ is, however, obtained when constructed using only 23 complexes, with the five excluded complexes being Asp, Glu, Leu, Gly, and Pro. The optimal dimensionalities of the $\Delta H$, $T\Delta S$, and $\Delta G$ models are four, four, and three latent variables, respectively. The relatively smaller standard deviation in the error of predictions (SDEP) value in the $\Delta G$ model than in both $\Delta H$ and $T\Delta S$ models corresponds with the smaller variation of the experimental $\Delta G$ values than those of $\Delta H$ and $T\Delta S$ values. The predicted $\Delta H$, $T\Delta S$, and $\Delta G$ values at the optimal dimensionalities are listed in Table 1 and plotted in Figures 4−6, respectively.

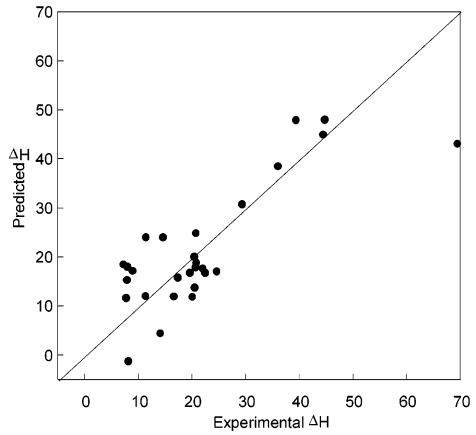



**Figure 4.** Plot of experimental vs predicted $\Delta H$ values (kJ/mol) for the 28 complexes from leave-one-out cross-validation at four latent variables.

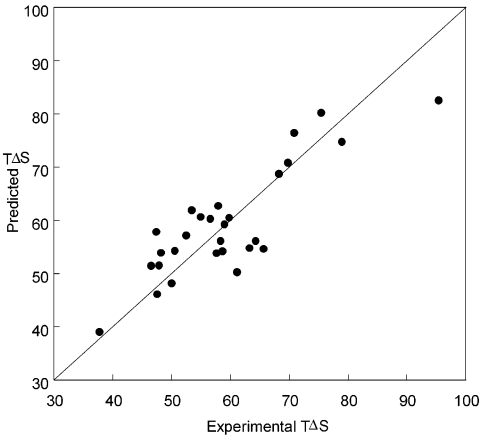



**Figure 5.** Plot of experimental vs predicted $T\Delta S$ values (kJ/mol) for the 28 complexes from leave-one-out cross-validation at four latent variables.

The constant terms in the three models are all negative and more negative than any predicted $\Delta H$, $T\Delta S$, or $\Delta G$ value. This means that together, the variable terms in the models disfavor binding. They do this to varying extents depending on the identity of the central residue. This can be interpreted as the constant term representing the favorable binding of the tripeptide, excluding the effect of the central residue. Only the variation of the central residue is then accounted for in the COMBINE models.

The constant terms at the optimal dimensionalities are in good agreement with the Gibbs equation $\Delta G =$

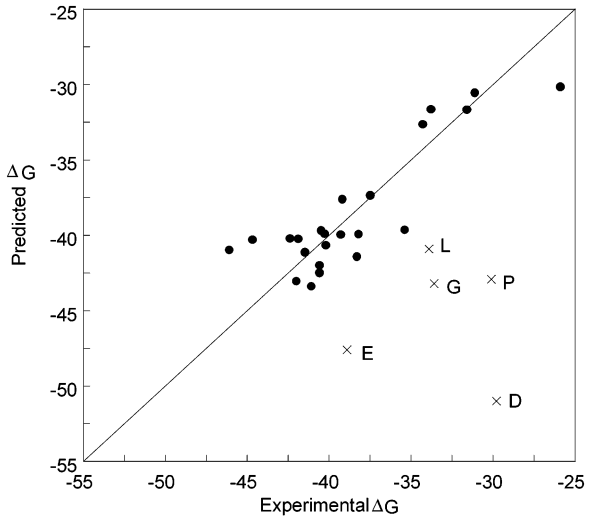



**Figure 6.** Plot of experimental vs predicted $\Delta G$ values (kJ/mol) for the 28 complexes from leave-one-out cross-validation at three latent variables. The external predicted values of the five outliers in the $\Delta G$ model are shown by crosses and labeled by the one letter names of the central residues.

$\Delta H - T\Delta S$; i.e., there is only a 2 kJ/mol deviation between the constant term in the $\Delta G$ model (−76 kJ/mol) and the difference in constant terms of the $\Delta H$ and $T\Delta S$ models (−131 − −53 = −78 kJ/mol). Therefore, we also computed estimates of $\Delta G$ for each complex by computing values of $\Delta H - T\Delta S$ from predictions of the $\Delta H$ and $T\Delta S$ models at four latent variables. These $\Delta H - T\Delta S$ values are listed in Table 1 for comparison with the $\Delta G$ model. Although both $\Delta H$ and $T\Delta S$ models are predictive in cross-validation for the complete set of 28 complexes, the predicted $\Delta H - T\Delta S$ values show poor correlation with the experimental $\Delta G$ values, with a correlation coefficient of only 0.18 for the 28 complexes. Even for the 23 complexes included in the $\Delta G$ model, the predicted $\Delta H - T\Delta S$ values have a much lower correlation coefficient (0.61) than the directly predicted $\Delta G$ values do (0.86).

As shown in Table 1, the outliers for the $\Delta H - T\Delta S$ prediction differ from those for the $\Delta G$ model. The five complexes with largest the $\Delta G - (\Delta H - T\Delta S)$ error values are Orn (13.5 kJ/mol), Asp (10.4 kJ/mol), Pro (9.4 kJ/mol), Nle (8.3 kJ/mol), and Gly (8 kJ/mol), of which only Asp, Gly, and Pro are outliers in the $\Delta G$ model. Orn shows the greatest deviation and also has a large prediction error in the $\Delta H$ model. In fact, Orn can be considered an outlier in the three experimental observables: it has the largest $\Delta H$ and $T\Delta S$ values and smallest $\Delta G$ values among the 28 complexes, and there are significant gaps between Orn and the next closest complex for each of these values. This means that the chemometric procedure tends to predict Orn to be more like the other complexes than it is.

In general, the discrepancy between the $\Delta H - T\Delta S$ and the $\Delta G$ predictions may be due to the larger SDEP values of the $\Delta H$ and $T\Delta S$ models relative to the $\Delta G$ model. This is due to enthalpy−entropy compensation: $\Delta G$ values cover a relatively small range and are given by the subtraction of quantities of larger absolute magnitude that vary over a larger range. Therefore, the subtraction of predicted $T\Delta S$ values from predicted $\Delta H$ values does not yield as accurate $\Delta G$ values as those obtained in the directly derived $\Delta G$ model.

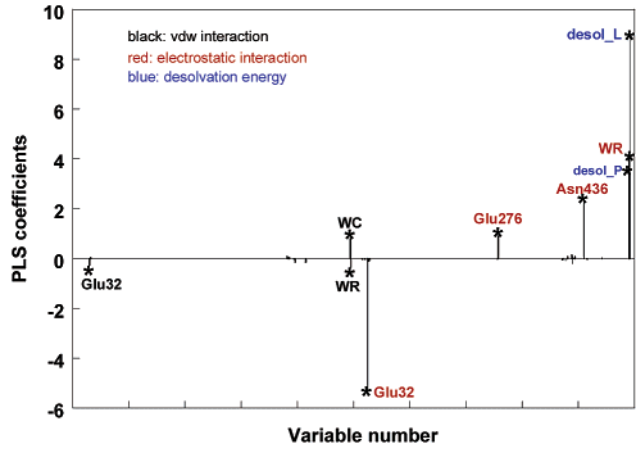



**Figure 7.** Plot of PLS normalized coefficients of variables in the $\Delta H$ model. The residues and variables with significant coefficients are labeled.

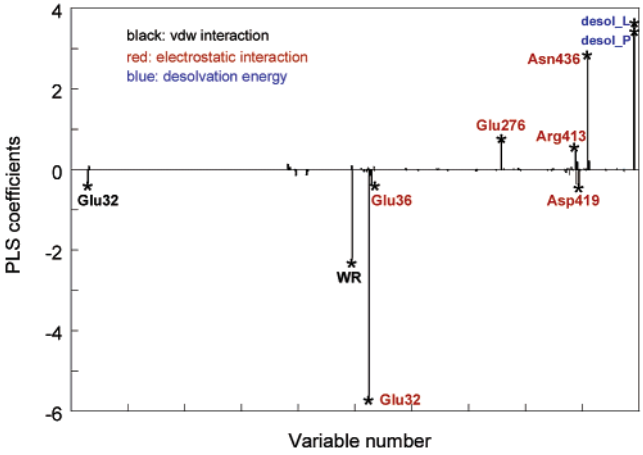



**Figure 8.** Plot of PLS normalized coefficients of variables in the $T\Delta S$ model. The residues and variables with significant coefficients are labeled.

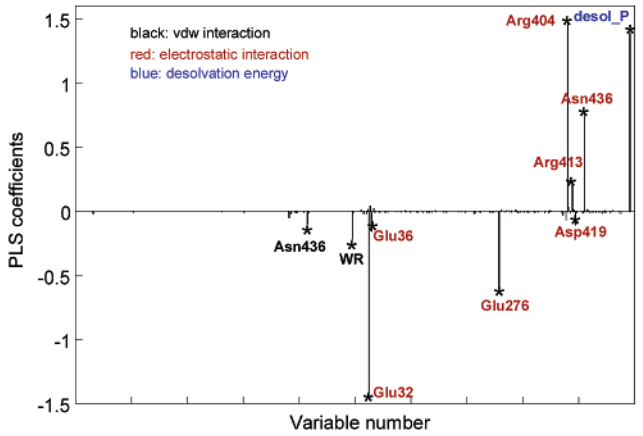



**Figure 9.** Plot of PLS normalized coefficients of variables in the $\Delta G$ model. The residues and variables with significant coefficients are labeled.

**Identification of Important Energy Contributions.** To investigate the contributions of the individual energy descriptors to the PLS models, the PLS normalized coefficients were computed and these are plotted in Figures 7 −9 for $\Delta H$, $T\Delta S$, and $\Delta G$ models, respectively. It can be seen that only a few energy descriptors

significantly contribute to the models. In other words, the differences in binding affinities of the 28 tripeptides can largely be explained by a few protein residues and their interactions, along with electrostatic desolvation energies.

In the $\Delta H$ model, the significant variables are the ligand and the protein electrostatic desolvation energies; the Coulombic interaction energies of protein residues Glu32, Glu276, and Asn436 and the water region WR; and the Lennard−Jones energies of protein residue Glu32 and the water regions WC and WR. These energy descriptors, with the exception of the Lennard−Jones energy of WC and the Coulombic energy of WR, are also highlighted as the most significant variables in the $T\Delta S$ model. These energy variables have coefficients of the same sign in both models. This means that computed $\Delta H$ and $T\Delta S$ values vary in the same direction with variation of the structures of the peptide complexes. This is consistent with the striking enthalpy−entropy compensation observed in OppA−peptide binding.

In the $\Delta G$ model, the significant variables are the protein desolvation energy; the Coulombic interaction energies of protein residues Glu32, Glu36, Glu276, Arg404, Arg413, Asp419, and Asn436; and the Lennard−Jones energies from protein residue Asn436 and the water region WR. The ligand desolvation energy, despite varying from 495 to 697 kJ/mol over the data set, has no significant contribution to the $\Delta G$ model, although it does contribute to the $\Delta H$ and $T\Delta S$ models. In fact, the ligand desolvation energy is significant in the initial $\Delta G$ model before the FFD variable selection. Its absence in the final model may be due to several factors. First, it might come from cancellation effects due to entropy−enthalpy compensation. Second, it might be due to the incomplete data set used for the $\Delta G$ model. Third, it may be due to the fine balance between terms that contribute to the ligand desolvation energy. For example, the shorter chain Dpp peptide has a larger desolvation energy than the Lys peptide. This is unexpected when considering chain length and the exposure of the charge on the end of the side chain but can be understood when taking into account the effect on ligand desolvation of differences in charge−charge interactions within the peptides. Finally, it may be because of the water molecules in the binding pocket. Their number varies according to the particular ligand, and they therefore affect relative ligand desolvation energies and may be able to compensate ligand desolvation penalties. The lack of dependence of relative binding affinity on ligand desolvation for this set of peptides may be one reason that OppA can bind to such a diversity of peptide sequences.

In contrast, the protein desolvation energy contributes to the model with the second largest positive coefficient, indicating that the desolvation of the protein strongly disfavors binding. In general, a larger ligand causes a greater desolvation cost to the protein than a small ligand does. This is due to its larger volume of low dielectric that is brought close to the protein. Interfacial water molecules have a minor effect on the variation of protein desolvation energy with different ligands as they do not perturb the proximity of the peptide backbone to the protein. Therefore, one strategy to strengthen the binding of a tripeptide KXK to OppA is to reduce the size of residue X.

Arg 404 lies opposite residue X of the peptide. The Coulombic interaction from Arg404 is a significant variable in the $\Delta G$ model with a large PLS coefficient, but surprisingly, it is absent in both the $\Delta H$ and the $T\Delta S$ models. In fact, Arg404 is significant in the initial models of both $\Delta H$ and $T\Delta S$ but was removed from these models during the FFD variable selection, this being the main change in significant variables after variable selection. The fact that Arg404 is important for the $\Delta G$ model but not the final $\Delta H$ and $T\Delta S$ models could be due to the different data sets used for these models. The $\Delta H$ and $T\Delta S$ models are derived from the set of 28 complexes, which includes complexes with peptides with both positively and negatively charged residues. The $\Delta G$ model, on the other hand, is derived from a data set from which peptides with negatively charged residues are excluded. Consequently, the PLS and variable selection procedure may be able to detect a predictive relation between the Arg 404 Coulombic interactions and peptides with neutral or positively charged middle residues (as for the $\Delta G$ model), but this is obscured (probably due to modeling or experimental inaccuracies) when peptides with negatively charged residues are present (as for the $\Delta H$ and $T\Delta S$ models).

The importance of electrostatic desolvation energy terms to the predictive ability of the COMBINE models is different for the three thermodynamic parameters. Exclusion of these terms from the models had the greatest effect on the $T\Delta S$ model, resulting in a drop in the $Q^2$ value of about 0.3 and a reduction in the optimum number of latent variables from four to three. It had a modest effect on the $\Delta G$ model, resulting in a drop in the $Q^2$ value of about 0.1 and a reduction in the optimum number of latent variables from three to two, and it had almost no effect on the predictive ability of the $\Delta H$ model (with the optimum number of latent variables staying constant at four). The electrostatic desolvation terms are free energy terms that implicitly include the entropic effects of water upon binding. All other terms included in COMBINE analysis are interaction energies with no explicit entropic component. Therefore, it is physically reasonable that the electrostatic desolvation terms are important for representing entropic effects and have the greatest influence on the $T\Delta S$ model. This also highlights the importance of including electrostatic desolvation terms in COMBINE analysis of systems demonstrating enthalpy−entropy compensation, with differences in binding entropy over the set of complexes analyzed. This is supported by the importance of electrostatic desolvation energies observed in COMBINE models for the binding affinity of a series of inhibitors to human immunodeficiency virus (HIV)-1 protease.[20] HIV-1 protease−inhibitor binding has been shown in recent calorimetry experiments[33] to be predominantly entropy driven with entropy−enthalpy compensation over a similar range to that observed in OppA−peptide binding.

Of the seven protein residues highlighted in the $\Delta G$ model, Glu32, Glu36, Arg404, Arg413, and Asp419 are in the first shell of the binding site whereas Asn436 and Glu276 are located behind Glu32 and Arg404, respectively (see Figure 2). Glu276 forms a stable salt link to

Arg404. Asn436 has no direct interaction with Glu32 in the conformation B adopted in most complexes but forms a hydrogen bond with the oxygen O$\epsilon$2 of Glu32 of conformation A in complexes Arg, Lys, Orn, His, and Trp. Glu32 with Asn436 and Arg404 with Glu276 flank the central pocket of the binding site, which notably discriminates against positively charged central residues by Coulombic interactions, as Glu32 and Glu276 show negative coefficients and Arg404 and Asn 436 show positive coefficients in Figure 7.

It appears that the favorable charge−charge interaction between Glu32 and a positively charged ligand disfavors binding. The crystal structures of the complexes show that the oxygen O$\epsilon$1 of Glu32 is always hydrogen bonding to the nitrogen N$\epsilon$ of His405, and this interaction is hypothesized to stabilize the bound (closed) form of OppA since it is not present in the unbound form. Of the five positively charged ligands (Arg, Lys, Orn, Dab, and Dap), three (Arg, Lys, and Orn) form hydrogen bonds with O$\epsilon$2 of Glu32 and Glu32 adopts conformation A in these complexes. If, as in most complexes, Glu32 adopted conformation B for these positively charged ligands, O$\epsilon$2 of Glu32 would point away from the peptide to a water molecule and O$\epsilon$1 of Glu32 would be within hydrogen-bonding distance of the side chain terminal nitrogen of Arg, Lys, or Orn. This would mean that the peptide would cause strong hydrogen-bonding competition with N$\epsilon$ of His405. To protect the stability of the bound (closed) form of OppA, Glu32 is rotated from conformation B to conformation A, allowing ligand−hydrogen bonding to O$\epsilon$2 of Glu32. This conformational change is assumed to be energetically unfavorable. But even so, the positive charges of the central residues of the peptides must still compete with His405. One favorable side effect of the conformational change is the release of one bound water molecule from the binding site (region WL). This water release likely results in entropy gain, but this is not strong enough to provide compensation. Consequently, the generally energetically favorable charge−charge interaction with Glu32 finally leads to an unfavorable effect on binding affinity due to the less favorable position of Glu32.

WR is the only water region highlighted in the $\Delta G$ model, and the negative Lennard−Jones energy coefficient of WR indicates that very close contact between the ligand and the two water molecules of WR favors binding. Therefore, this term suggests that a large X residue in the KXK peptide would be more favorable than a small one. This is in opposition to the effect of the protein electrostatic desolvation term, which indicated that a small X residue would be more favorable and reflects the fact that many terms combine and compensate to produce the overall binding affinity. The roles of the other two water regions, WL and WC, are small in the $\Delta G$ model, although they show more variation in interaction energy and number of water molecules than WR. WR makes a similar Lennard−Jones contribution to the $T\Delta S$ model. For the $\Delta H$ model, WR again makes Lennard−Jones interactions representing packing with the peptide but also contributes through electrostatic interactions.

To assess the importance of the water region interaction terms for the COMBINE models, we also built models excluding these terms. These models showed very small reductions in predictive performance (<0.1 units for $Q^2$). However, the optimal number of latent variables reduced from four to three for both $\Delta H$ and $T\Delta S$ models. While, at first sight, this might seem advantageous (similar predictive ability with fewer adjustable parameters), closer inspection shows that there are notable changes in the PLS coefficients, which make physically based understanding of the models derived without water interaction terms more difficult. For example, the PLS coefficient for the Coulombic interaction with Glu276 changed sign in the $T\Delta S$ model from positive with water terms included to negative with water terms excluded. It disappeared in the $\Delta H$ model with water terms excluded but retained a negative coefficient in the $\Delta G$ model. This results in the situation in which entropically favorable interactions with Glu276 are unfavorable to free energy, which is surprising given the entropically driven binding observed experimentally. This indicates that energy terms included in the COMBINE models are implicitly representing physical interactions that are not explicitly modeled, including those of the interfacial water when water interaction terms are excluded. Consequently, COMBINE models with similar $Q^2$ values are obtained with and without water interaction terms.

**Outliers in the Model for Binding Free Energy.** As shown in Table 2 and Figure 6, there are five outliers in the $\Delta G$ model: Asp, Glu, Leu, Gly, and Pro. Their binding affinities are significantly overpredicted in the $\Delta G$ model. This indicates that there must be some unfavorable energetic contribution missing from the COMBINE analysis. Considering that only the bound conformations of the peptides were taken into account in the COMBINE analysis, this overprediction of binding affinity could arise from the energetics of peptide conformational changes upon binding. This seems feasible for Asp, Glu, Leu, and Gly but not Pro.

**(i) Asp and Glu.** It is likely that the negatively charged Asp and Glu make favorable charge−charge interactions in the Lys-Asp(Glu)-Lys tripeptides in solution.[34] These intramolecular interactions will be disrupted upon binding, and this will have an unfavorable effect on binding affinity.

**(ii) Leu.** Analysis of rotamers with the InsightII software showed that the side chain of the bound Leu in the crystal structure (with $\chi_1 = -80°$ and $\chi_2 = 83°$) does not occupy any of the rotameric states in the rotamer library. It is the only case in which the central residue in the peptide−OppA complexes adopts a nonrotameric side chain conformation. The conformational change of the Leu peptide to this less stable nonrotameric form upon binding from solution disfavors binding. Potential of mean force calculations of the relative free energies of leucine rotamers indicates that the conformation observed in the OppA crystal structure is about 6−12 kJ/mol above the most energetically favorable Leu rotamers.[35] Thus, this rotamer change upon binding could account for the 7 kJ/mol overprediction of binding affinity for the Leu peptide.

**(iii) Gly.** As compared to other amino acids, Gly has more backbone freedom in solution. The loss of this conformational freedom upon binding leads to a greater

entropic penalty on binding affinity for Gly than for other amino acids.

**(iv) Pro.** In contrast to Gly, Pro has less backbone freedom than other amino acids due to the side chain making a ring by covalent attachment to the backbone N atom. The backbone of the bound Pro is in a standard conformation with its ring in the commonly observed UP conformation. So, the fact that Pro is an outlier remains difficult to explain by peptide conformational freedom loss or unfavorable intramolecular energy changes upon binding.

Because of the lack of experimental data about the structures of the unbound tripeptides in solution, it is not possible to provide further proof or quantitative calculations for the above explanations. Addition of an energy variable to describe loss of side chain entropy upon binding by the change in the number of rotamers available did not improve the model. No attempt was made to model the change in intramolecular energy upon binding as, in the absence of experimental data on their solution conformations, this would require extensive conformational sampling of the peptides in water.

## Concluding Remarks

COMBINE analysis has been used to correlate structural and thermodynamic quantities ($\Delta H$, $T\Delta S$, and $\Delta G$) for OppA−peptide binding. The predictive models derived demonstrate that the COMBINE method is applicable to estimation of $\Delta H$ and $T\Delta S$ values, as well as $\Delta G$ values, despite the problems posed by enthalpy−entropy compensation. Interfacial water molecules, often ignored in methods of estimating binding affinity, are explicitly treated in COMBINE analysis and contribute to the QSAR models.

This study provides the first quantitative QSAR model for OppA−peptide binding. It reveals interactions important for differences in binding affinity, including the contribution of protein electrostatic desolvation. It should thus provide a useful guide for further design studies on the OppA system. Furthermore, this study points to the applicability of COMBINE analysis for studying the binding thermodynamics of other protein−peptide complexes.

**Acknowledgment.** We are grateful to Dr. G. Cruciani for providing the GOLPE program and Dr. A. R. Ortiz for providing the COMBINE program. T.W. was the recipient of a DAAD postdoctoral fellowship. We thank the Klaus Tschira Foundation for financial support. We thank Peter Winn and John Ladbury for critical reading of an earlier version of the manuscript.

**Note Added after ASAP Posting.** This manuscript was released ASAP on 9/21/2002 with an error in Table 1 and with minor typographical and formatting errors in Outliers in the Model for Binding Free Energy section. The correct version was posted on 10/17/2002.

## References

(1) Guyer, C. A.; Morgan, D. G.; Staros, J. V. Binding specificity of the periplasmic Oligopeptide-binding protein from *Escherichia coli*. *J. Bacteriol*. **1986**, *168*, 775−779.

(2) Goodell, E. W.; Higgins, C. F. Uptake of cell-wall peptides by *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol*. **1987**, *169*, 3861−3865.

(3) Hammond, S. M.; Claesson, A.; Jansson, A. M.; Larsson, L. G.; Pring, B. G.; Town, C. M.; Ekstrom, B. A new class of synthetic antibacterials acting on lipopolysaccharide biosynthesis. *Nature* **1987**, *327*, 730−732.

(4) Fraser, C. M.; Casjens, S.; Huang, W. M.; Sutton, G. G.; et al. Genomic sequence of a Lyme disease spirochaete, *Borrelia Burgdorferi*. *Nature* **1997**, *390*, 580−586.

(5) Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R. H. Crystallographic and Calorimetric Analysis of Peptide Binding to OppA Protein. *J. Mol. Biol*. **1999**, *291*, 393−415.

(6) Davies, T. G.; Hubbard, R. E.; Tame, J. R. H. Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes. *Protein Sci*. **1999**, *8*, 1432−1444.

(7) Sleigh, S. H.; Tame, J. R. H.; Dodson, E. J.; Wilkinson, A. J. Peptide binding in OppA, the crystal structures of the periplasmic oligopeptide binding protein in the unliganded form and in complex with lysyllysine. *Biochemistry* **1997**, *36*, 9747−9758.

(8) Tame, J. R. H.; Dodson, E. J.; Murshudov, G.; Higgins, C. F.; Wilkinson, A. J. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* **1995**, *3*, 1395−1406.

(9) Tame, J. R. H.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol*. **1996**, *3*, 998−1001.

(10) Ladbury J. E. Just add water! The effect of water on the specificity of protein−ligand binding sites and its potential application to drug design. *Chem. Biol*. **1996**, *3*, 973−980.

(11) Rostom, A. A.; Tame, J. R. H.; Ladbury, J. E.; Robinson, C. V. Specificity and interactions of the protein OppA: partitioning solvent binding effects using mass spectrometry. *J. Mol. Biol*. **2000**, *296*, 269−279.

(12) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des*. **1994**, *12*, 243−256.

(13) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: a simplified potential approach. *J. Med. Chem*. **1999**, *42*, 791−804.

(14) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev*. **1993**, *93*, 2395−2417.

(15) Resat, H.; Marrone, T. J.; McCammon, J. A. Enzyme−inhibitor association thermodynamics: explicit and continuum solvent studies. *Biophys. J*. **1997**, *72*, 522−532.

(16) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem*. **2001**, *44*, 3417−3423.

(17) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem*. **1995**, *38*, 2681−2691.

(18) Wade, R. C. Derivation of QSARs using 3D structural models of protein−ligand complexes by COMBINE analysis. In *Rational Approaches to Drug Design: 13th European Symposium on Quantitative Structure−Activity Relationships*; Holtje, H.-D., Sippl, W., Eds.; Prous Science S. A.: Barcelona, 2001; pp 23−28.

(19) Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C. Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J. Med. Chem*. **1997**, *40*, 1136−1148.

(20) Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J. Med. Chem*. **1998**, *41*, 836−852.

(21) Pastor, M.; Gago, F.; Cruciani, G. Comparative binding energy (COMBINE) analysis on a series of glycogen phosphorylase inhibitors: comparison with GRID/GOLPE models. In *Molecular Modeling and Prediction of Bioactivity*; Gundertofte, K., Jorgensen, F. S., Eds.; Kluwer: New York, 2000; pp 329−330.

(22) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem*. **2001**, *44*, 961−971.

(23) Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F. 3D-QSAR methods on the basis of ligand receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 Ligands. *J. Comput.-Aided Mol. Des*. **2000**, *14*, 341−353.

(24) Kmunicek, J.; Luengo, S.; Gago, F.; Ortiz, A. R.; Wade, R. C.; Damborsky, J. Comparative Binding Energy (COMBINE) analysis of the substrate specificity of haloalkane dehalogenase from *xanthobacter autotrophicus* GJ10. *Biochemistry* **2001**, *40*, 8905−8917.

(25) Tomic, S.; Nilsson, L.; Wade, R. C. Nuclear receptor-DNA binding specificity: A COMBINE and Free-Wilson QSAR analysis. *J. Med. Chem.* **2000**, *43*, 1780−1792.
(26) Vriend, G. WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52−56.
(27) Hooft, R. W. W.; Sander, C.; Vriend, G. Positioning hydrogen atoms by optimizing hydrogen bond networks in protein structures. *Proteins* **1996**, *26*, 363−376.
(28) Cornell, W. D.; Cieplak, P.; Payly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(29) *InsightII*, version 2000; Molecular Simulations Inc.: San Diego, CA.
(30) *AMBER: Assisted Model Building with Energy Refinement*, version 6.0; Department of Pharmaceutical Chemistry, University of California: San Francisco, 1999.
(31) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. Electrostatics and diffusion of molecules in solution: simulations with the university of Houston brownian dynamics program. *Comput. Phys. Commun.* **1995**, *91*, 57−95.
(32) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.
(33) Todd, M. J.; Luque, I.; Valazquez-Campoy, A.; Freire, E. Thermodynamic basis of resistance to HIV-1 protease inhibition: calorimetric analysis of the V82F/I84V active site resistant mutant. *Biochemistry* **2000**, *39*, 11876−11883.
(34) Molano, A.; Segura, C.; Guzman, F.; Lozada, D.; Patarroyo, M. E. In human malaria protective antibodies are directed mainly against the Lys-Glu ion pair within the Lys-Glu-Lys motif of the synthetic vaccine SPf66. *Parasite Immunol.* **1992**, *14*, 111−124.
(35) Wade, R. C.; McCammon, J. A. Binding of an antiviral agent to a sensitive and a resistant human rhinovirus. Computer simulation studies with sampling of amino acid side-chain conformations. 1. Mapping the rotamers of residue 188 of viral protein 1. *J. Mol. Biol.* **1992**, *225*, 679−696.

JM020900L