

Why OppA protein can bind sequence-independent peptides? A combination of QM/MM, PB/SA, and structure-based QSAR analyses

Feifei Tian · Li Yang · Fenglin Lv ·
Xiaoli Luo · Yuzhu Pan

Received: 3 May 2010 / Accepted: 12 June 2010 / Published online: 27 June 2010
© Springer-Verlag 2010

Abstract Periplasmic oligopeptide-binding protein (OppA) is the initial receptor in the ATP-binding cassette (ABC) system of bacteria, which exhibits a broad specificity in binding oligopeptides without regard to sequence. Here, we present a computational study on the structural properties and energetic landscapes of OppA protein interacting with its cognate ligands on the basis of 28 structure/affinity-known OppA–tripeptide complexes. By employing a well-designed protocol that couples the hybrid quantum mechanical/molecular mechanical (QM/MM) scheme and the sophisticated Poisson–Boltzmann/surface area (PB/SA) solvent model together to analyze and decompose the energy components associated with the OppA–peptide binding, we demonstrate that the broad specificity of OppA-recognizing peptides is originated from a series of exquisite balances between the free energy contributions from, for example, the direct nonbonded interactions and indirect desolvation effects, the main chains and side chains, and the different residue positions of the tripeptide ligands. We also show that, in a framework of structure-based quantitative structure–activity relationship (SB-QSAR) methodology, the QM/MM–PB/SA-derived energy terms could be used as a good descriptor to characterize the interaction profile of OppA

with peptides and correlate pretty well with the experimentally measured affinities of the binding.

Keywords OppA protein · Peptide · QM/MM · PB/SA · SB-QSAR

Introduction

Bacteria take up many types of molecule using binding protein-dependent permeases, part of a much larger group of transport systems known as the ATP-binding cassette (ABC) transporters, one of the largest and most diverse protein families known (Higgins 1992). Unlike their eukaryotic counterparts, the bacterial ABC import systems also involve a globular protein family which acts as the initial receptor of the ligand to be brought into the cell. In this family, the one of the most characterized proteins is the periplasmic oligopeptide-binding protein (OppA), which is a 58.8 kDa periplasmic peptide receptor found in Gram-negative bacteria (Monnet 2003). Unlike other members of this protein family, OppA has a broad specificity in the peptide ligands that it binds. Early studies found that the OppA can carry peptides of two to five amino acid residues regardless of their size, charge, and polarity, and therefore has thousands of potential ligands including peptide-based antibiotics (Guyer et al. 1986). Later, the structure of OppA bound to diverse oligopeptide ligands solved by X-ray crystallography revealed a surprising complex mode by which the peptide ligands are enveloped between two lobes (I and III) of the protein and buried deeply in the protein interior (Fig. 1; Tame et al. 1994). The enclosure of ligands in this way is a hallmark of the periplasmic substrate-binding proteins and is normally associated with complementary interactions and exquisite specificity (Tame et al.

F. Tian · L. Yang (✉) · F. Lv (✉) · X. Luo · Y. Pan
College of Bioengineering and Key Lab of Biorheological
Science and Technology, Chongqing University,
Chongqing 400044, Chongqing, China
e-mail: liyang@cqu.edu.cn

F. Lv
e-mail: lufenglin001@yahoo.com.cn

F. Tian
Center for Heterocyclic Compounds, Department of Chemistry,
University of Florida, Gainesville, FL 32611, USA

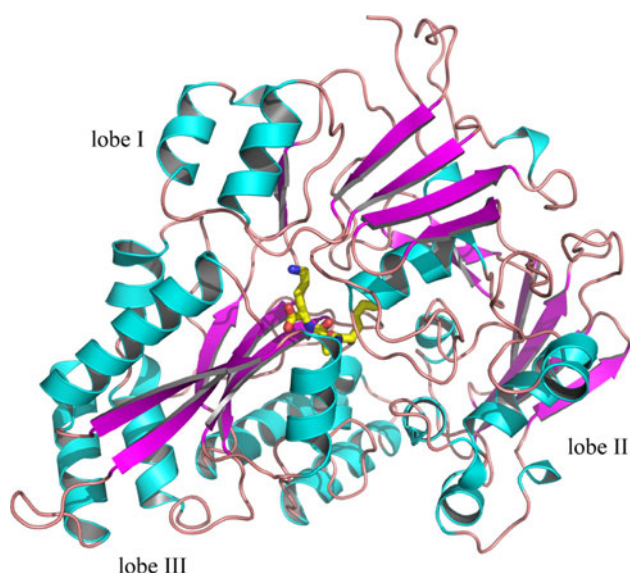


Fig. 1 Stereoview of OppA protein–peptide KAK complex (PDB: 1jet). It is shown that the tripeptide ligand is enveloped between the lobes I and III of the protein and buried deeply in the protein interior

1996). However, the exception of the OppA, which possesses a broad specificity, implies a larruping interaction mechanism of this protein with its peptide ligands.

According to the crystal structures of OppA–peptide complexes, the main chain of the bound peptides is in an extended conformation and forms parallel and antiparallel β -sheet interactions with the OppA that satisfy the hydrogen bonding capacity of the peptide backbones; the N-terminus of the peptides donates a salt bridge to the side chain of Asp419; the C-terminus of the tri-, tetra-, and pentapeptides forms buried salt bridges with the residues Arg413, His371, and Lys307, respectively; the side chains of the peptides are accommodated in voluminous hydrated pockets, where few direct contacts are made with the protein (Picon et al. 2000). This structure picture preliminarily unraveled the origin of the nonspecific nature of OppA binding to sequence-independent peptides: (1) the main chains of diverse peptide ligands share a common interacting mode with the protein receptor, and (2) the binding fluctuations over the capricious peptide side chains can be largely erased by the spacious, adaptable hydrated pockets.

Although the crystallographic information has shed light on the structural basis of OppA–peptide recognition, there still exist a number of problems that need to be addressed. In particular, the quantitative energetic knowledge regarding the free energy contributions of peptide chains and residue sites would be fundamentally valuable for understanding the thermodynamic behavior of this nonspecific binding, but it is completely unavailable from the crystal structures. Recently, Li et al. (2009) have proposed a protocol that incorporates empirical solvation model into hybrid quantum

mechanical/molecular mechanical (QM/MM) methodology to dissect and analyze the energetic profile associated with the binding of peptide ligand to protein receptor. This method has been successfully applied to investigate the roles of water molecules and anchor residues in HLA-A*0201–epitope recognition. In this study, the more rigorous Poisson–Boltzmann/surface area (PB/SA) strategy is used instead of empirical solvation model to accurately account for the indirect desolvation effect accompanied with the OppA–peptide binding, and the direct nonbonded interactions involved in this binding are described using the QM/MM scheme. In this way, the energy contributions of different structure moieties of the tripeptide ligands contained in 28 structure-known OppA–peptide complexes are decomposed and examined in detail. The decomposed energy terms are further correlated with the experimentally measured affinities using a structure-based quantitative structure–activity relationship (SB-QSAR) approach. On the basis of these calculations, we would give a systematic discussion around the thermodynamic nature of the broad specificity of OppA–peptide binding.

Materials and methods

Structures and affinities of OppA–peptide complexes

The crystal structures of OppA protein in complex with 28 tripeptides, which share a consensus motif Lys-Xxx-Lys comprising a variable central residue Xxx flanked by two lysines, were solved at the atomic resolution level by Tame et al. (Tame et al. 1994, 1995, 1996; Sleight et al. 1997, 1999; Davies et al. 1999). The central residues Xxx of the 28 tripeptides cover the complete spectrum of 20 coded and 8 non-coded amino acids. The thermodynamic data associated with the formation of these complexes were measured at a standard condition (50 mM sodium phosphate buffer at pH 7.0 and 298 K) using the isothermal titration calorimetry (Sleight et al. 1999; Davies et al. 1999). Here, the tripeptide sequences, PDB entries, binding free energies ΔG_{expl} , and logarithmic dissociation constants $\ln K_D$ of the 28 OppA–peptide complexes are collected in Table 1, which will be used as a basic data set for subsequent analyses.

Pretreatment of OppA–peptide complex structures

The crystal structures of the 28 OppA–tripeptide complexes were retrieved from the protein data bank (PDB) (Berman et al. 2000), from which the water molecules, metal ions, and other cofactors were removed manually and the hydrogen atoms were added using the programs REDUCE (for coded amino acids) (Word et al. 1999a) and I-INTERPRET (for non-coded amino acids) (Zhao et al.

Table 1 The relevant information and experimental thermodynamic data for the 28 OppA–tripeptide complexes

No.	PDB	Peptide	ΔG_{expl}^a	$\ln K_D^b$	No.	PDB	Peptide	ΔG_{expl}^a	$\ln K_D^b$
1	1jet	KAK	−9.82	4.03	15	1qka	KRK	−8.07	7.09
2	1b05	KCK	−9.70	4.32	16	1b51	KSK	−10.03	3.76
3	1b4z	KDK	−7.12	8.68	17	1b52	KTK	−9.70	4.33
4	1jeu	KEK	−9.29	5.01	18	1qkb	KVK	−10.01	3.81
5	1b40	KFK	−9.91	3.97	19	1jev	KWK	−9.36	4.87
6	1b3l	KGK	−8.03	7.17	20	1b58	KYK	−8.96	5.56
7	1b3f	KHK	−9.39	4.87	21	1b0h	K–Nap–K	−9.12	5.30
8	1b3g	KIK	−9.15	5.30	22	1b1h	K–Chx–K	−8.46	6.43
9	2olb	KKK	−7.55	7.97	23	1b2h	K–Hph–K	−9.63	4.53
10	1b9j	KLK	−8.10	7.00	24	1b3h	K–Nva–K	−10.68	2.71
11	1b32	KMK	−9.67	4.37	25	1b4h	K–Nle–K	−11.01	2.25
12	1b5i	KNK	−9.60	4.50	26	1b5h	K–Orn–K	−6.19	10.28
13	1b46	KPK	−7.19	8.56	27	1b6h	K–Dab–K	−7.43	8.14
14	1b5j	KQK	−10.13	3.61	28	1b7h	K–Dap–K	−8.19	6.88

^a The unit of ΔG is kcal/mol^b The unit of K_D is nM

2007). Subsequently, each complex structure was subjected to an energy minimization procedure with the AMBER03 force field (Duan et al. 2003). Briefly, the maximum number of minimization steps was set to 2,000; the first 300 steps were performed with the steepest descent algorithm and the rest of the steps were performed with the conjugate gradient algorithm. The solvent effect was described using the generalized Born model. The convergence criterion for the rms of the Cartesian elements of the energy gradient was set to 0.2 kcal/mol/Å, as recommended by Hou et al. (2008).

QM/MM analysis of OppA–peptide interactions

At present, one of the most widely applied strategies to implement hybrid QM/MM calculations is the ONIOM method (our own N-layered integrated molecular orbital and molecular mechanics) (Svensson et al. 1996), which enables different levels of theory to be applied to different parts of a molecule/system and combined to produce a consistent energy expression. The objective is to perform a high-level calculation on just a small part of the system and include the effects of the remainder at lower levels of theory, with the end result being of similar accuracy to a high-level calculation on the full system. Previously, the ONIOM-based QM/MM scheme has been broadly used to investigate the interaction profile of proteins with small molecules, including binding activity and nonbonding behavior (Alzate-Morales et al. 2009; Lu et al. 2009; Zhou et al. 2010a, b). However, for the larger ligands such as peptides and oligomers, high-level calculation is also difficult even when adopting the non-electron correlation

Table 2 The key protein residues in direct interaction with the three residue sites of the tripeptide ligand through salt bridges and hydrogen bonds, as characterized by Tame et al. (1994, 1995, 1996)

Tripeptide residue site	Interacting key protein residues	Interaction manner
Position 1	Cys417	Hydrogen bond
	Asp419	Salt bridge
Position 2	Glu32	Hydrogen bond
	Val34	Hydrogen bond
Position 3	Arg413	Salt bridge
	Gly415	Hydrogen bond

method in conjunction with the smallest basis set. Therefore, Li et al. recently suggested the use of semiempirical AM1 method—which has been demonstrated to be effective and efficient in the analysis of molecular systems involving weak nonbonded interactions (Dannenberg 1997)—to treat the high-level layer which, in this study, contained the tripeptide ligand and several key protein residues that were experimentally characterized to be fundamental in determining the binding of the OppA–peptide complex (Table 2; Fig. 2), while the remainder of this complex system was included in the MM layer and treated with a lower level of AMBER96 force field (Cornell et al. 1995), in which the partial atomic charges and force field parameters for the non-coded amino acid atoms were pre-obtained using restricted electrostatic potential (RESP) fitting (Comell et al. 1993) and generalized amber force field (GAFF) (Wang et al. 2004), respectively. Here, the hybrid ONIOM method was implemented in the Gaussian03 suite of programs (Frisch



Fig. 2 QM/MM partition scheme for the OppA–peptide complex. The tripeptide ligand (shown in *stick style*) and the key protein residues (shown in *stick and ball style*) that directly interact with the ligand are in the QM layer, while the remainder (shown in *ribbon style*) of this system is in the MM layer. The salt bridges and hydrogen bonds are represented by *solid* and *broken lines*, respectively

et al. 2003), and the electrostatic interactions between the two layers were treated in terms of mechanical embedding scheme to save computational cost.

The protein–peptide (or peptide segment) interaction energy (ΔE_{int}) in a complex system was predicted according to the strategy proposed by Zhou et al. (2009a). This was accomplished by performing single point energy calculation twice, one on the complex system (E_1) and other on the same system but its members (protein and peptide) were separated by several angstroms of distance each other (E_2). In this way, the interaction energy can be derived as $\Delta E_{\text{int}} = E_1 - E_2$.

PB/SA analysis of accompanied desolvation effects

Protein–ligand binding is a desolvation process, with which a substantial free energy change could be accompanied. To count the solvent effects associated with this process, the empirical solvent accessible surface area (SASA) model (Eisenberg and McLachlan 1986) and semi-empirical Poisson–Boltzmann/surface area (PB/SA) model (Kollman et al. 2000) are available. Previously, although Li et al. (2009) have successfully incorporated the SASA method into QM/MM to investigate the binding behavior of HLA-A*0201 protein with its peptide ligands and received satisfactory results for qualitative explanation, for the quantitative purpose as this study the relatively strict PB/SA method could be more reasonable choice.

In PB/SA procedure, the total desolvation free energy (ΔG_{dsolv}) was estimated from the polar (electrostatic)

desolvation energy (ΔG_{polar}) and the nonpolar desolvation energy ($\Delta G_{\text{nonpolar}}$). The polar part was calculated by finite difference solutions to the nonlinear PB equation, as implemented in the DelPhi program (Rocchia et al. 2001), with ionic strength 0.15 M, and dielectric constants 4 for solute and 80 for solvent. A grid spacing of 2.0 grid/Å, in which the longest linear dimension of the solute occupied 60% of the lattice, was used to determine the size of the cubic lattice, and the boundary potentials were set to the sum of Debye–Hückel values (Yan et al. 2003). The PARSE parameter set (Sitkoff et al. 1994) was used for protein and peptide atoms, and a probe radius of 1.4 Å to define the dielectric boundary. The nonpolar contribution was determined by summing up the weighted surface area of whole solute molecule, i.e. $\Delta G_{\text{nonpolar}} = \gamma \cdot \Delta A$, where $\gamma = 0.00542 \text{ kcal/mol } \text{\AA}^2$ (Kollman et al. 2000), and ΔA is the changes in surface area due to the OppA–peptide binding. We computed the solute’s surface area using Sanner’s algorithm implemented in the MSMS program (Sanner et al. 1996), with a solvent probe radius of 1.4 Å and the PARSE atomic radii parameters (Sitkoff et al. 1994).

SB-QSAR correlation of the energy terms with binding affinities

The SB-QSAR methodology (Zhou et al. 2007) was further employed to ascertain the intrinsic relationship between the calculated energy terms and the experimental binding affinities of OppA–peptide complexes. The first 20 samples as listed in Table 1, of which the central residues of peptide ligands are defined by coded amino acids, were adopted as training set for constructing SB-QSAR models, while the remaining 8 samples with non-coded amino acids at the central positions of the peptides were used as test set for validating the constructed models. Partial least squares (PLS) regression (Wold et al. 2001), which has been broadly applied in the QSAR community, was employed to linearly correlate the energy terms with binding affinities, and we implemented this algorithm using the ZP-explore program (Zhou et al. 2009b). For a constructed SB-QSAR model, its performance in the statistical viewpoint, could be measured quantitatively by the coefficients of determination of fitting (r), leave-one-out (LOO) cross-validation (q), and prediction (r_{pred}) on, respectively, training set, training set, and test set:

$$r = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i^{\text{fitting}})^2}{\sum_{i=1}^n (y_i - \bar{y}_{\text{tr}})^2}} \quad (1)$$

$$q = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i^{\text{cv}})^2}{\sum_{i=1}^n (y_i - \bar{y}_{\text{tr}})^2}} \quad (2)$$

$$r_{\text{pred}} = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i^{\text{pred}})^2}{\sum_{i=1}^m (y_i - \bar{y}_{\text{te}})^2}} \quad (3)$$

where $n = 20$ and $m = 8$ are the sample numbers of, respectively, training and test sets, y_i is the experimentally determined affinity of sample i , $\bar{y}_{\text{tr}}$ and $\bar{y}_{\text{te}}$ are the average values of all the y_i in, respectively, training and test sets, $\hat{y}_i^{\text{fitting}}$, $\hat{y}_i^{\text{cv}}$, and $\hat{y}_i^{\text{pred}}$ are the estimated affinities for sample i by, respectively, fitting, cross-validation, and prediction.

Results and discussion

Analysis of the different energy terms involved in OppA–peptide binding

The direct interaction energies ΔE_{int} and indirect desolvation effects ΔG_{dsolv} for the binding of 28 OppA–peptide complexes were calculated according to the procedures described in “Materials and methods”, and the resulting values are compiled in Table 3. It is evident that there have been intensive nonbonded contacts between the OppA protein and its peptide ligands, given by the interaction energies ΔE_{int} of most samples larger than -100 kcal/mol—which even achieve to the energy level of conventional covalent bonds. The strong interactions involved in these complexes are not unexpected, since OppA protein can form a number of hydrogen bonds and in particular salt bridges with the peptides, and these electrostatics-driven interactions could be quite effective at the low dielectric peptide-binding site deeply buried in the OppA protein interior where the electrostatic screening is considerably weak. Although the direct interaction of OppA with peptides is remarkable, a large part of this interaction energy are used to pay the unfavorable desolvation penalty due to the burial of polar and charged groups in the low dielectric protein interior. As can be seen, the desolvation free energies, ΔG_{dsolv} , are also noticeable and most of them have fallen into the range 60–100 kcal/mol, leading to a substantial discount for the favorable contribution of direct interactions to the binding of peptides to OppA.

The linearly fitting results, as shown in Fig. 3a and b, reveal that the correlations both between calculated interaction energy ΔE_{int} and experimental binding affinity $\ln K_D$ and between calculated desolvation energy ΔG_{dsolv} and $\ln K_D$ are quite modest, with their correlation coefficients r of, respectively, 0.134 and 0.173, indicating that neither nonbonded interaction nor desolvation effect can independently dominate the binding. Nevertheless, the sum of ΔE_{int} and ΔG_{dsolv} presents a good correlation with the $\ln K_D$, as suggested by a considerably high r value of 0.652

(Fig. 3c). The compatibility of $\ln K_D$ with ΔE_{int} plus ΔG_{dsolv} ($\Delta E_{\text{int}} + \Delta G_{\text{dsolv}}$), but not the individual ΔE_{int} or ΔG_{dsolv} , implies that the force driving the formation of OppA–peptide complexes is a compromise between the direct interaction and indirect solvent effect. This conclusion is consistent with previous crystallographic studies, which figured out that, in addition to intensive salt bridge and hydrogen bond networks, a large number of water molecules were also observed at the OppA–peptide interfaces (Tame et al. 1996; Rostom et al. 2000). Moreover, it is worth noting here that, though the sum of ΔE_{int} and ΔG_{dsolv} is correlated well with $\ln K_D$, which is deviated significantly from the experimentally determined binding energy ΔG_{expl} . This phenomenon is interesting but not difficult to explain, if considering that some energy terms such as conformational entropy and conformation change effect were not considered in the present calculation procedure. However, ignoring these additional terms would not essentially influence our conclusions, since their contributions could be roughly regarded as a constant.

Furthermore, we herein give a preliminary discussion on the apparent relationship between the calculated energy terms and the physicochemical properties of amino acids, and the further detailed analysis of the energy contributions from each residue position of tripeptide ligands will be addressed in next section. From Table 3, it can be seen that weaker interactions ΔE_{int} are commonly associated with the bigger nonpolar residues at the central position of the tripeptides. For example, the KKK, K-Hph-K, and K-Nap-K with hydrophobic and/or aromatic central residues possess the lowest ΔE_{int} values as -76.54 , -82.24 , and -87.63 kcal/mol, respectively. In contrast, the peptides with charged central residues, such as KKK, KRK, and K-Dab-K, exhibit a stronger interaction with the OppA protein. Visual analysis of complex structure models unraveled that both nonpolar and polar central residues, especially those of bulk, are in, more or less, steric contact and/or collision with the neighboring OppA atoms, leading to a slight distortion on their side-chain rotamers. However, this unfavorable effect could be paid by long-range electrostatic attractions of the central residues with protein context, if the central residues are charged, or could be counteracted by favorable desolvation effects of the central residues buried in protein interior, if the central residues are hydrophobic. In addition, although the charged residues could keep effective interaction with OppA, they inevitably incur severe desolvation penalty. This can be rationalized by the relatively large desolvation free energy ΔG_{dsolv} generally associated with the peptides possessing polar central residues, albeit the ΔG_{dsolv} difference between polar and nonpolar is not significant as compared to ΔE_{int} . Sum up, the more polar the residues are, the stronger/larger the electrostatic interaction/desolvation penalty is, and vice

Table 3 The calculated energy terms associated with the binding of the 28 OppA–tripeptide complexes

No.	PDB	Peptide	Whole peptide		Position 1		Position 2		Position 3	
			ΔE_{int}	ΔG_{dolv}	$\Delta E_{\text{int}}^{\text{p1}}$	$\Delta G_{\text{dolv}}^{\text{p1}}$	$\Delta E_{\text{int}}^{\text{p2}}$	$\Delta G_{\text{dolv}}^{\text{p2}}$	$\Delta E_{\text{int}}^{\text{p3}}$	$\Delta G_{\text{dolv}}^{\text{p3}}$
1	ljet	KAK	−117.63	79.01	−57.63	41.53	−32.21	−4.27	−32.78	37.67
2	1b05	KCK	−135.21	101.67	−48.25	48.86	−34.35	1.38	−46.43	46.64
3	1b4z	KDK	−125.69	96.37	−53.42	45.52	−18.76	10.55	−25.38	47.79
4	ljeu	KEK	−121.98	87.11	−67.32	37.04	−13.18	13.67	−36.57	38.67
5	1b40	KFK	−113.45	75.56	−53.54	46.78	−28.62	−14.32	−28.99	42.53
6	1b3l	KGK	−139.33	106.69	−68.53	50.12	−27.77	7.34	−25.32	44.10
7	1b3f	KHK	−125.49	90.02	−60.02	36.24	−16.39	12.17	−42.17	45.33
8	1b3g	KIK	−98.67	71.42	−48.43	46.10	−8.44	−16.45	−43.75	39.27
9	2olb	KKK	−146.79	117.13	−58.43	41.06	−32.26	28.76	−31.23	36.24
10	1b9j	KLK	−76.54	48.67	−41.62	33.66	−10.05	−9.69	−35.22	27.72
11	1b32	KMK	−114.07	72.79	−59.11	32.67	−21.76	1.78	−28.31	33.41
12	1b5i	KNK	−125.83	88.27	−53.43	35.46	−29.67	14.26	−37.85	30.78
13	1b46	KPK	−106.33	82.02	−55.98	45.78	−17.16	−4.22	−38.23	41.12
14	1b5j	KQK	−143.14	110.92	−49.21	49.57	−34.37	10.68	−47.28	45.31
15	1qka	KRK	−120.87	94.33	−56.37	34.67	−35.55	28.97	−31.87	33.46
16	1b5l	KSK	−139.66	95.21	−58.76	43.45	−22.44	4.68	−52.46	42.18
17	1b52	KTK	−158.72	124.32	−71.77	57.36	−21.64	16.32	−58.47	49.15
18	1qkb	KVK	−128.97	90.38	−51.32	45.63	−30.57	0.05	−42.03	39.68
19	ljev	KWK	−97.28	73.17	−45.74	46.12	−2.54	−26.75	−36.96	47.38
20	1b58	KYK	−106.54	77.09	−57.45	49.13	−24.53	−18.66	−31.74	42.67
21	1b0h	K–Nap–K	−87.63	62.53	−53.44	34.58	−12.56	−10.71	−26.28	37.21
22	1b1h	K–Chx–K	−132.69	103.33	−59.86	50.13	−18.98	2.58	−47.91	42.17
23	1b2h	K–Hph–K	−82.24	58.63	−46.05	30.64	−12.68	−9.19	−37.36	28.16
24	1b3h	K–Nva–K	−158.67	122.95	−58.45	55.79	−36.72	2.34	−54.33	52.33
25	1b4h	K–Nle–K	−126.45	87.98	−62.33	45.32	−21.63	−0.18	−45.47	39.64
26	1b5h	K–Orn–K	−113.68	90.79	−46.22	40.56	−36.21	28.54	−39.32	36.74
27	1b6h	K–Dab–K	−131.12	104.82	−54.21	46.46	−25.35	30.17	−42.79	42.18
28	1b7h	K–Dap–K	−148.76	120.81	−60.53	49.36	−22.76	28.36	−36.78	47.18

All units are in kcal/mol

versa. In this way, the total binding energies of different peptides, no matter their central residues are polar or nonpolar, bulky or small, and charged or uncharged, are basically in a similar level, giving rise to the OppA protein capable of binding sequence-independent peptides.

Decomposition of total energy terms into individual residue sites

In order to examine in detail the energy contributions of the three residue sites in the tripeptides, we here broke down the peptide bonds (and capped by hydrogen atoms) of the peptide ligands in complexes and separately computed the interaction energy and desolvation effect of each residue binding to the OppA protein. The protocol used for calculating the energy components for individual residues was the same as that for whole peptides, and the resulting

values are tabulated in Table 3. As can be seen, for most samples the sums of individual energy terms are not equal to the total energy values obtained from the whole peptides ($\Delta E_{\text{int}} \neq \Delta E_{\text{int}}^{\text{p1}} + \Delta E_{\text{int}}^{\text{p2}} + \Delta E_{\text{int}}^{\text{p3}}$ and $\Delta G_{\text{dolv}} \neq \Delta G_{\text{dolv}}^{\text{p1}} + \Delta G_{\text{dolv}}^{\text{p2}} + \Delta G_{\text{dolv}}^{\text{p3}}$). It could be attributed to the fact that the cooperation between the peptide residues was overlooked when the energy contributions were calculated using isolated residues. However, this discrepancy is not significant because the sums of individual terms are, albeit not equal to, roughly consistent with the total values.

Position 1

The N-terminal residues of all the 28 tripeptide ligands in our sample set are the charged lysine (Lys1). Crystal structures revealed that the backbone α -amino and side chain ϵ -amino of this residue can separately form a

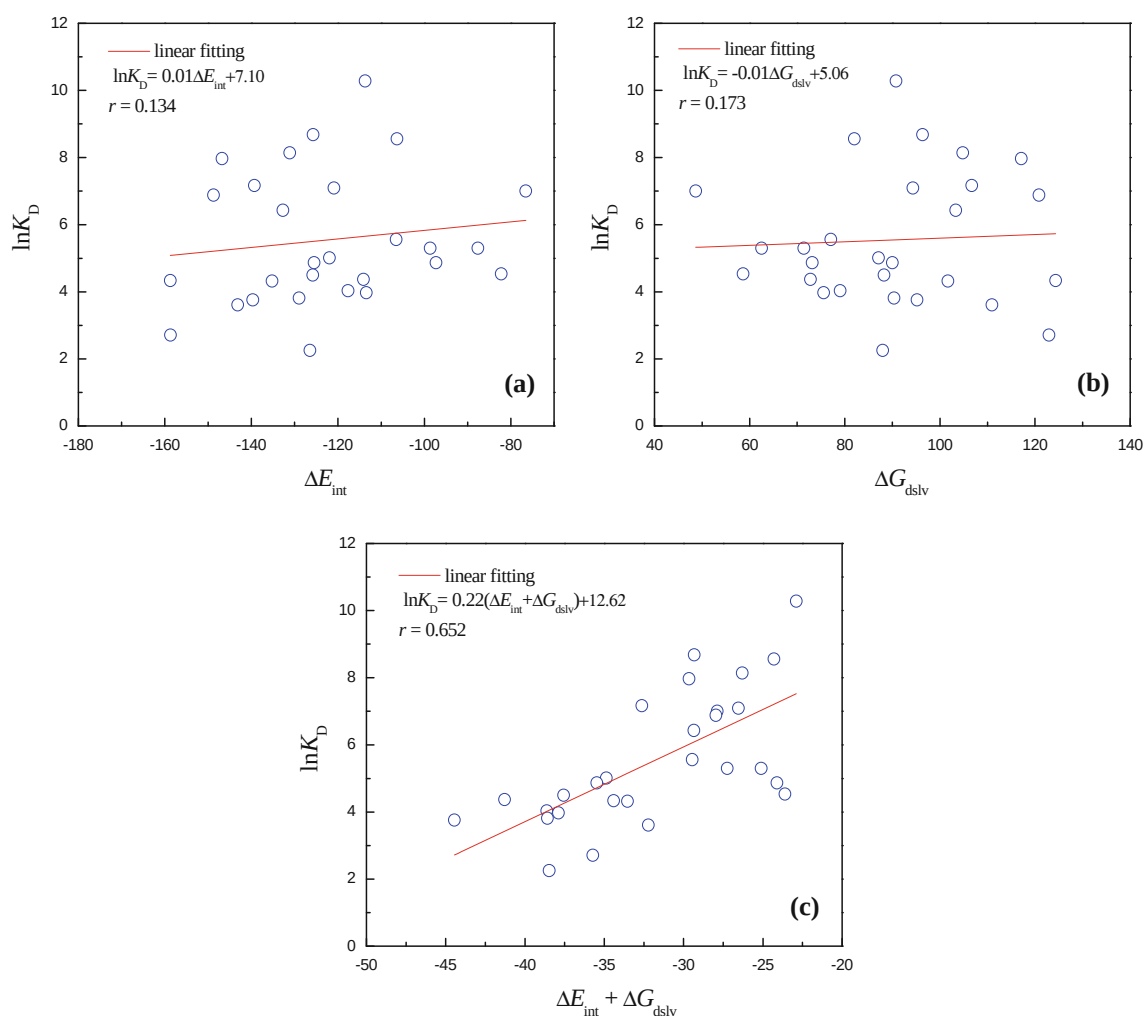


Fig. 3 Plots of experimental binding affinity $\ln K_D$ against calculated interaction energy ΔE_{int} (a), desolvation energy ΔG_{dslv} (b), and the sum of ΔE_{int} and ΔG_{dslv} (c)

well-defined salt bridge with the Asp419 of OppA protein. In addition, a hydrogen bond was observed between the backbone amides of tripeptide Lys1 and OppA Cys417 (Fig. 2; Sleight et al. 1997). According to the optimized complex models, these three nonbonded interactions of Lys1 with OppA are quite conserved and their configurations were basically unchanged during the MM minimization procedure, indicating that the energy contributions of Lys1 to the peptide binding in different complexes are significant but vary slightly. This could be rationalized by the relatively large average values (-55.62 and 43.56 kcal/mol) and considerably small standard deviations (7.14 and 7.01 kcal/mol) of the $\Delta E_{\text{int}}^{\text{p1}}$ and $\Delta G_{\text{dslv}}^{\text{p1}}$ over all samples. In this respect, we concluded here that, at least for the samples in our data set, the position 1 of tripeptide ligands performs the role as a stabilizer but not a specifier; it confers a large part of stabilization energy to the OppA–peptide binding but does not give rise to much judgment for different peptide ligands.

Position 3

Just like position 1, the position 3 of all sample peptides is occupied by the lysine residue (Lys3). Previous studies figured out that the conformations of Lys3 are shown to be somewhat more variable and disordered than that of Lys1, albeit this conformational fluctuation is insignificant (Tame et al. 1996). According to our calculations, the average values $\pm$ standard deviations of $\Delta E_{\text{int}}^{\text{p3}}$ and $\Delta G_{\text{dslv}}^{\text{p3}}$ are, respectively, -38.68 ± 8.77 and 40.60 ± 6.21 kcal/mol—they can be nearly completely counteracted by each other—manifesting that the Lys3 can only throw a slight effect on the binding. The less impact of Lys3 than Lys1 on the complex's formation could be at least in part attributed to its absence of a salt bridge with the OppA as compared to Lys1 (Fig. 2). However, this difference is partly erased from the total energy values due to the additional desolvation penalty. Sum up, if considering the solvent effect, Lys3 dominates neither stabilization nor specificity for the

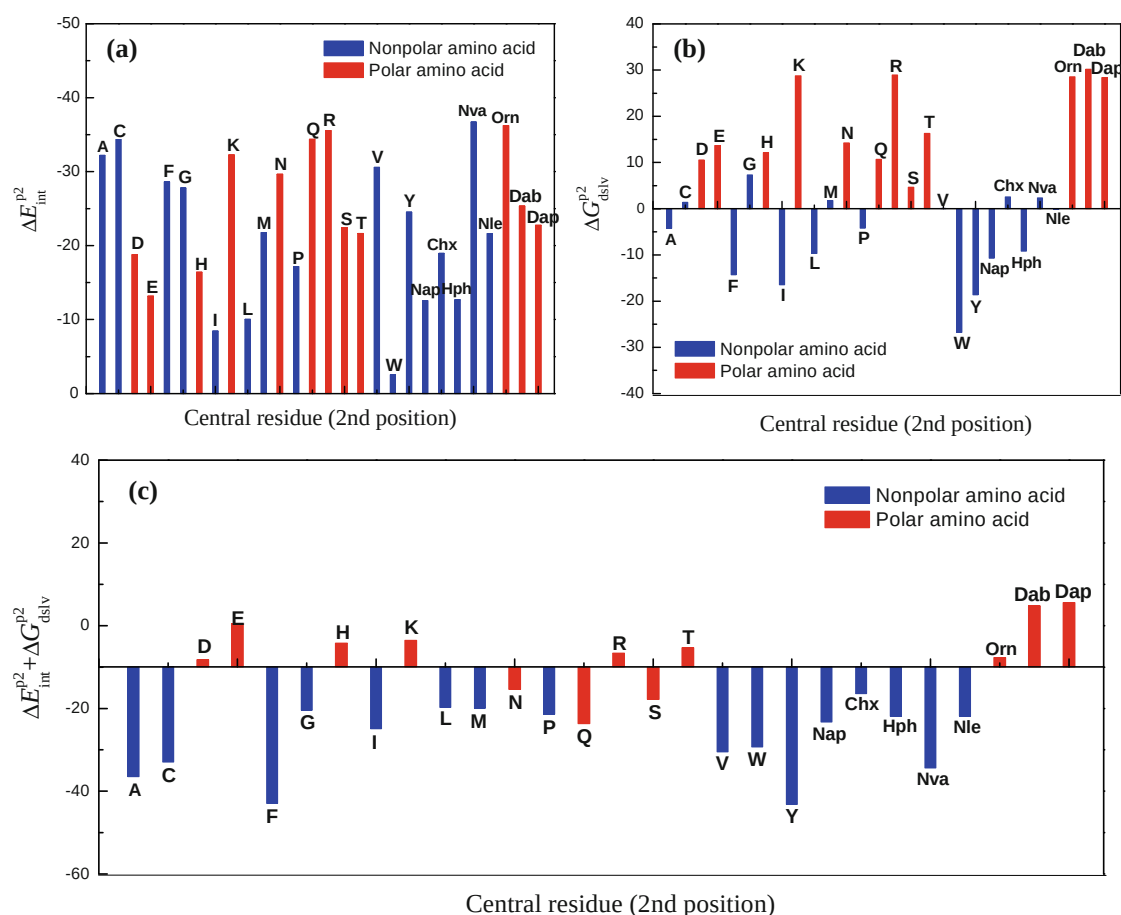


Fig. 4 Direct interaction energies ΔE_{int}^{p2} (a), indirect desolvation effects ΔG_{dslv}^{p2} (b), and sum binding energies $\Delta E_{int}^{p2} + \Delta G_{dslv}^{p2}$ (c) of different amino acid types present at the second position of the tripeptide ligands

peptides binding to OppA. Thus, it is more suitable to be viewed as an adaptor or filler in the complexes.

Position 2

The position 2 is completely variable in our sample panel, which presents 20 coded and 8 non-coded amino acids. Therefore, as might be anticipated, both direct interaction energies ΔE_{int}^{p2} and indirect desolvation effects ΔG_{dslv}^{p2} have a larger variance over all the samples than that at positions 1 and 3. From Fig. 4a it can be seen that, generally speaking, the polar and small, nonpolar amino acids exhibit stronger nonbonded interactions with context than those with bulky, nonpolar side chains, such as Ile, Leu, and Trp, which are electrostatically inactive but would introduce significant steric hindrance to the binding. For example, use of small probe technique (Word et al. 1999b) to detect atomic packing behavior in OppA–peptide KWK complex can apperceive intensive van der Waals contacts and collisions at the binding interface of OppA with the side chain of tryptophan residue in KWK (Fig. 5). This atomic overlapping, albeit is not severe, would more or less

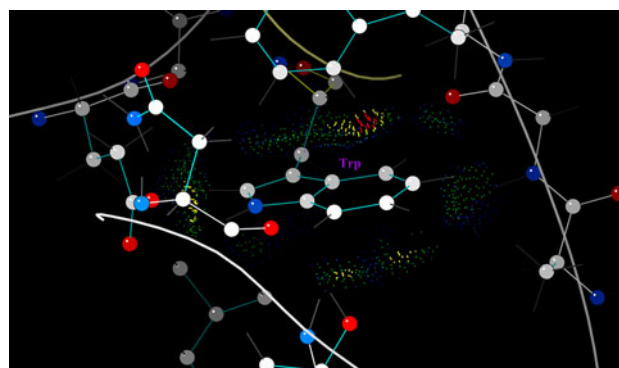


Fig. 5 Intensive steric contacts and collisions between the central residue Trp of tripeptide KWK and its neighboring OppA residues. The steric contacts and collisions were detected using PROBE program (Word et al. 1999b) and are shown in this picture as dots and spikes, respectively

undermine the compatibility between receptor and ligand and hence results in a low ΔE_{int}^{p2} value—as predicted to be -2.54 kcal/mol. Furthermore, the indirect solvent effect, as shown in Fig. 4b, is readily accordant with the polarity of amino acids; there are significant desolvation penalties

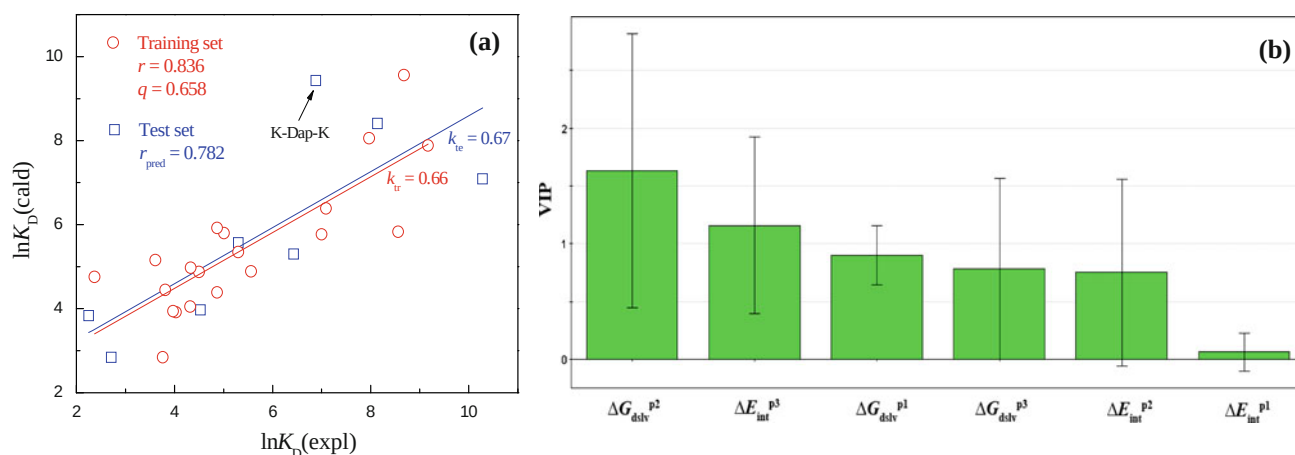


Fig. 6 **a** Plot of PLS-calculated against experimental binding affinities, $\ln K_D$. **b** Variable importance in the projection (VIP) of PLS model

($\Delta G_{\text{dslv}}^{\text{p}2} > 0$) for polar and charged residues due to wrapping them into the low-dielectric OppA interior during the binding processes, whereas the desolvation effects for most nonpolar amino acids, especially those bulky nonpolar amino acids, are favorable, given by their $\Delta G_{\text{dslv}}^{\text{p}2}$ values < 0 . Although $\Delta E_{\text{int}}^{\text{p}2}$ and $\Delta G_{\text{dslv}}^{\text{p}2}$ appear to be much variable, the fluctuations of $\Delta E_{\text{int}}^{\text{p}2}$ and $\Delta G_{\text{dslv}}^{\text{p}2}$ roughly present a complementary profile and, consequently, the sum binding energies at the position 2 ($\Delta E_{\text{int}}^{\text{p}2} + \Delta G_{\text{dslv}}^{\text{p}2}$) vary insignificantly through the entire sample set (Fig. 4c). In this way, the apparent contribution of position 2 to the specificity of OppA–peptide binding is reduced considerably.

SB-QSAR modeling and analyses

SB-QSAR methodology was further employed to statistically model the linear relationship of energy values with affinities for this OppA–peptide system. The independent variable matrix **X** consists of the six decomposed energy terms, $\Delta E_{\text{int}}^{\text{p}1}$, $\Delta G_{\text{dslv}}^{\text{p}1}$, $\Delta E_{\text{int}}^{\text{p}2}$, $\Delta G_{\text{dslv}}^{\text{p}2}$, $\Delta E_{\text{int}}^{\text{p}3}$, and $\Delta G_{\text{dslv}}^{\text{p}3}$, and the dependent variable vector **y** is the experimentally determined binding affinities, $\ln K_D$. The results stemming from PLS modeling between the **X** and **y** were prettily good. As shown in Fig. 6a, the calculated $\ln K_D$ agree well with corresponding experimental values. On the training set, the PLS model with four principal components (PCs) can explain a large ratio of variance for observed **y**, with coefficients of determination of fitting (r) and cross-validation (q) of 0.836 and 0.658, respectively; on the test set the performance of this model was also satisfactory with predictive r_{pred} of 0.782. All of these definitely claimed that (1) the calculated energy contributions by the combination of QM/MM and PB/SA methods exhibit a strong linear correlation with the experimental binding affinities, and (2) the PLS model based on the first 20 tripeptides comprised of coded amino acids could be used to reliably predict other

samples, no matter they are made up by coded or non-coded amino acids.

The largest deviation of calculated from experimental $\ln K_D$ value occurred on the test sample K-Dap-K, of which the central residue Dap possesses a small, charged side chain (CH_2NH_3^+). According to model's prediction, this peptide should present an active binding behavior to OppA because of the low steric hindrance and strong electrostatic attraction. However, its experimental affinity was not as high as the predicted value owing to some unknown factors, leading to the overestimation by PLS model.

Furthermore, variable importance in the projection (VIP) (Wold et al. 2001) gives a straightforward insight into the relative importance of these six energy terms involved in the PLS model (Fig. 6b). As can be seen, indirect desolvation effect seems to be more decisive than direct nonbonded interaction in the OppA–peptide binding, which is compatible with crystallographic information because the strong nonbonded interactions are mainly formed on the invariant peptide's backbones (Picon et al. 2000) and therefore provide only little judgment between the different peptide ligands. In addition, variation in desolvation effects, although which is relatively significant, is also largely discounted as the peptide's side chains are accommodated in a spacious, plastic hydrated pocket (Tame et al. 1994, 1995, 1996) that can partially cancel the specific solvent effects arisen from the capricious side chains of central residues.

Conclusions

Exploring the complicated binding behavior of flexible biopolymer ligands to their cognate receptor is definitely a challenge but can provide much valuable information about

thermodynamic properties associated with the recognition process. In this study, we attempt to understand the structural basis and energetic landscape of OppA protein interacting with its peptide ligands and to, at least in part, explain why the OppA protein can bind sequence-independent peptides? To achieve this, we employed a combination of ONIOM-based QM/MM computation and PB/SA solvent model to ascertain the binding mechanism of 28 structure/affinity-known OppA–tripeptide complexes. We demonstrate that: (1) in the tripeptides, the N-terminal residue confers significant stability but little specificity for the OppA–peptide binding; the C-terminal residue contributes neither stability nor specificity to the binding; and the variable central residue provides a large ratio of specificity for the binding, but which is considerably discounted if considering solvent effect. (2) Direct interactions are mainly formed on the invariant backbones of peptide ligands and thus contribute very little to the specificity, while the variation in desolvation effects due to the capricious side chains of central residues is erased largely by the voluminous, adaptable hydrated pocket. (3) The bulky central residues would incur intensive steric collisions from neighboring OppA atoms. This unfavorableness could be partly compensated from the favorable desolvation effect when the bulky residue is hydrophobic (non-polar) or the favorable long-range electrostatic attraction when the bulky residue is charged (polar).

Acknowledgments This work was supported by the Natural Science Foundation Project of Chongqing CSTC (Grant No. 2009BA5068) and the Innovative Talent Training Project, Third Stage of ‘211Project’, Chongqing University (Grant No. S-09104).

References

- Alzate-Morales JH, Caballero J, Jague AV, Nilo FDG (2009) Insights into the structural basis of N2 and O6 substituted guanine derivatives as cyclin-dependent kinase 2 (CDK2) inhibitors: prediction of the binding modes and potency of the inhibitors by docking and ONIOM calculation. *J Chem Inf Model* 49:886–899
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
- Comell WD, Cieplak P, Bayly CI, Kollman PA (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J Am Chem Soc* 115:9620–9631
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
- Dannenbergh JJ (1997) Hydrogen bonds: a comparison of semiempirical and ab initio treatments. *J Mol Struct* 401:279–286
- Davies TG, Hubbard RE, Tame JR (1999) Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA–peptide complexes. *Protein Sci* 8:1432–1444
- Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012
- Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. *Nature* 319:199–203
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (2003) Gaussian 03. Gaussian Inc., Wallingford
- Guyer CA, Morgan DG, Staros JV (1986) Binding specificity of the periplasmic oligopeptide-binding protein from *Escherichia coli*. *J Bacteriol* 168:775–779
- Higgins CF (1992) ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* 8:67–113
- Hou T, Zhang W, Case DA, Wang W (2008) Characterization of domain–peptide interaction interface: a case study on the amphiphysin-1 SH3 domain. *J Mol Biol* 376:1201–1214
- Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
- Li Y, Yang Y, He P, Yang Q (2009) QM/MM study of epitope peptides binding to HLA-A*0201: the roles of anchor residues and water. *Chem Biol Drug Des* 74:611–618
- Lu Y, Shi T, Wang Y, Yang H, Yan X, Luo X, Jiang H, Zhu W (2009) Halogen bonding—a novel interaction for rational drug design. *J Med Chem* 5:2854–2862
- Monnet V (2003) Bacterial oligopeptide-binding proteins. *Cell Mol Life Sci* 60:2100–2114
- Picon A, Kunji ER, Lanfermeijer FC, Konings WN, Poolman B (2000) Specificity mutants of the binding protein of the oligopeptide transport system of *Lactococcus lactis*. *J Bacteriol* 182:1600–1608
- Rocchia W, Alexov E, Honig B (2001) Extending the applicability of the nonlinear. Poisson–Boltzmann equation: multiple dielectric constants and multivalent ions. *J Phys Chem* 105:6507–6514
- Rostom RA, Tame JRH, Ladbury JE, Robinson CV (2000) Specificity and interactions of the protein OppA: partitioning solvent binding effects using mass spectrometry. *J Mol Biol* 296:269–279
- Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320
- Sitkoff D, Sharp KA, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 9:1978–1988
- Sleigh SH, Tame JRH, Dodson EJ, Wilkinson AJ (1997) Peptide binding in OppA, the crystal structures of the periplasmic oligopeptide binding protein in the unliganded form and in complex with lysyllysine. *Biochemistry* 36:9747–9758
- Sleigh SH, Seavers PR, Wilkinson AJ, Ladbury JE, Tame JRH (1999) Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J Mol Biol* 291:393–415

- Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) ONIOM: a multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $\text{Pt}(\text{P}(\text{t-Bu})(3))(2) + \text{H}_2$ oxidative addition. *J Phys Chem* 100:19357–19363
- Tame JRH, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF, Wilkinson AJ (1994) The structural basis of sequence-independent peptide binding by OppA protein. *Science* 264:1578–1581
- Tame JRH, Dodson EJ, Murshudov GN, Higgins CF, Wilkinson AJ (1995) The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure* 3:1395–1406
- Tame JRH, Sleight SH, Wilkinson AJ, Ladbury JE (1996) The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat Struct Biol* 3:998–1001
- Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Developing and testing of a general amber force field. *J Comput Chem* 25:1157–1174
- Wold S, Sjöström M, Eriksson L (2001) PLS regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* 58:109–130
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999a) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285:1735–1747
- Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC (1999b) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* 285:1711–1733
- Yan S, Wu M, Patel DJ, Geacintov NE, Broyde S (2003) Simulating structural and thermodynamic properties of carcinogen-damaged DNA. *Biophys J* 84:2137–2148
- Zhao Y, Cheng T, Wang R (2007) Automatic perception of organic molecules based on essential structural information. *J Chem Inf Model* 47:1379–1385
- Zhou P, Tian F, Li Z (2007) A structure-based, quantitative structure–activity relationship approach for predicting HLA-A*0201-restricted cytotoxic T lymphocyte epitopes. *Chem Biol Drug Des* 69:56–67
- Zhou P, Zou J, Tian F, Shang Z (2009a) Fluorine bonding—how does it work in protein–ligand interactions? *J Chem Inf Model* 49:2344–2355
- Zhou P, Tian F, Lv F, Shang Z (2009b) Comprehensive comparison of eight statistical modelling methods used in quantitative structure–retention relationship studies for liquid chromatographic retention times of peptides generated by protease digestion of the *Escherichia coli* proteome. *J Chromatogr A* 1216:3107–3116
- Zhou P, Lv J, Zou J, Tian F, Shang Z (2010a) Halogen–water–hydrogen bridges in biomolecules. *J Struct Biol* 169:172–182
- Zhou P, Ren Y, Tian F, Zou J, Shang Z (2010b) Halogen-ionic bridges: do they exist in the biomolecular world? *J Chem Theory Comput*. doi:[10.1021/ct100167w](https://doi.org/10.1021/ct100167w) (in press)