Contribution of Solvation Energy in Protein-Peptide Recognition Systems

LI, Fei*,[a](李菲) LI, Wei[b](李惟) SHEN, Jia-Cong[a](沈家骢)

[a]*Key Laboratory for Supramolecular Structure and Spectroscopy of Ministry of Education, Jilin University, Changchun, Jilin 130023, China*

[b]*Department of Molecular Biology, College of Life Science, Jilin University, Changchun, Jilin 130023, China*

The contribution of solvation energy to guiding molecular recognition for six rigid protein-peptide systems had been evaluated by the variation in the number of the identified native-like configurations and in the driving force of specific interaction resulting from the addition of the explicit solvation term in the force field function. The AMBER force field energy and the total energy including the force field energy and the WZS solvation energy were calculated for sampled configurations. The results obtained by the calculations of both force field and total energies were compared with each other. It suggests that the contribution of solvation energy is important to guiding the specific recognition of the systems in which the ligands possess larger hydrophobic or aromatic residues while the protein receptors provide the active surfaces with hydrophobic property.

Keywords Contribution of solvation energy, protein-peptide interaction, rigid docking systems

Introduction

Many of the functions performed by biological molecules depend on appropriate interactions with each other. Molecular recognition is the process by which intermolecular forces act to bring about a productive collision between molecules; it is inherently a dynamic and stochastic process. Of the forces to guiding specific recognition of molecules, however, the contribution of each component varies with different receptor-ligand systems. Therefore, an evaluation to the contribution of each interaction component in molecular recognition would be particularly important to a better understanding of mechanism of molecular recognition.

The solvation interaction, including hydrophobic interaction, on which we focus our attention in this study, was usually ignored in the calculation of empirical energies and some scoring functions that were used to evaluate accessible conformations in docking calculation.[1-3] However, many ligands of interest, including some examined here, possess polar groups that point out into solvent and apolar groups that locate on the hydrophobic protein surface as they associate with protein. In this case, solvation interaction would play an important role in guiding molecular recognition.

In this article, we evaluate the contribution of solvation interaction for six rigid protein-peptide systems from two aspects: the efficiency of solvation energy to identifying native-like configurations and the change in the driving force of specific recognition arising from solvation energy. The six test systems are classified into two types in which the properties of interfaces are rather different. One corresponds to the complexes in which the protein receptors favor specifically to the hydrophobic groups in the binding sites and the ligands provide larger hydrophobic and aromatic residues. The other corresponds to the complexes in which the protein receptors specify the positively charged groups of Lys or Arg in the

* E-mail: lifei2001@263.net

Received July, 14, 2000; revised January 17, 2001; accepted March 5, 2001.

Project supported by the National Natural Science Foundation of China (No. 29734130).

binding sites and the ligands provide Lys or Arg residue. The difference in interface property leads to the difference in the contribution of solvation energy to molecular recognition.

Method and test systems

1. *Binding energy*

WZS *solvation binding energy*

For the rigid-body docking, only the intermolecular binding energy is necessary to determine the configuration of protein-ligand complex. The intramolecular energy is unchanged in the docking. Therefore, if each amino-acid residue is classified into atomic groups or molecular fragments based on hydrophobic character, the solvation binding energy ΔE_{sol}, the solvation energy difference between receptor-ligand complex and two isolated molecular components, can be expressed as

$$\Delta E_{sol} = \sum_i \sigma_i \left[\sum_{l}^{ligand} (A_{il} - A_{il}(0)) + \sum_{r}^{receptor} (A_{ir} - A_{ir}(0)) \right] \quad (1)$$

where A_{il} and A_{ir} are the solvent-accessible surface areas of ligand and receptor atoms of protein-ligand complex, respectively, for the ith atomic group or molecular fragment type. $A_{il}(0)$ and $A_{ir}(0)$ are the solvent-accessible surface areas of isolated ligand and isolated receptor, respectively. The sums in parentheses are over all ligand and receptor atoms of the ith atomic group or molecular fragment type, respectively.

The assignment of the solvation parameter depends on the classification method of amino-acid residue.[4-9] In the WZS solvation model,[8] 20 amino-acid residues are classified into a number of molecular fragment types. The solvation parameters are trained following the back-propagation learning approach[10] by maximizing the energy difference between compact non-native and native structures of a selected group of globular proteins.

In this study, 20 amino-acid residues were classified into 38 molecular fragment types. The solvation parameters were trained by 150 non-native protein structures generated by MC method from 27 native protein structures. The solvent-accessible surface area was calculated by a fast algorithm[11,12] using united-atom radii assigned to nine atom types: C, 0.170 nm; C(H), 0.180 nm; C(H_2), 0.190 nm; C(H_3), 0.200 nm; N(H), 0.150 nm; N(H_2, H_3), 0.160 nm; O(C = O), 0.140 nm; O(H), 0.150 nm; S, 0.185 nm. The spherical surface around each united atom was represented by 240 evenly distributed test points and the radius of a spherical water molecule was set to 0.14 nm.

AMBER *force field energy*

AMBER force field energy was calculated based on the AMBER-type potential function.[1] In practice, we have assumed the bond lengths, bond angles and ring conformations found in the crystal structures. Moreover, the torsion angles were also fixed as the same to the crystal structures for the rigid systems. Therefore, the force field binding energy ΔE_{ff} can be approximated by

$$\Delta E_{ff} = \sum_{i}^{ligand} \sum_{j}^{receptor} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + C \frac{q_i q_j}{\varepsilon(r) r_{ij}} \right) \quad (2)$$

where A_{ij} and B_{ij} are the nonbonded parameters; q_i and q_j are the AMBER partial charges on the receptor and ligand atoms, respectively; $\varepsilon(r)$ is the distant-dependent dielectric function. In this work, $\varepsilon(r) = r$ was used to mimic the protein interior. We do not use an explicit H-bond term of AMBER. We assume that contribution due to H-bond is included in the electrostatic term.

Total binding energy

The total binding energy ΔE_{ff+sol} consists of the AMBER force field and WZS solvation binding energies

$$\Delta E_{ff+sol} = \Delta E_{ff} + \Delta E_{sol} \quad (3)$$

2. *Test systems*

Six protein-peptide complexes were used as the test systems. The continuum-binding fragments in close contact with protein surfaces were selected from the ligands of the corresponding crystallographic complexes in the Protein Data Bank as the peptide components in the test systems. The structural characteristics of these test systems are listed in Table 1. Asp(E189) and Gly(E226) in trypsin are exchanged with each other in trypsin vari-

ant. The component molecules were treated as separate entities by using their respective atomic coordinates within the complexes. The crystallographic waters were removed before calculation and the hydrogen atoms were added.

Table 1 Test systems and their structural characteristics

PDB code	Resolution (nm)	Name in paper	Protein	$R_g(E)^a$ (nm)	Peptide fragment	R_g^b (nm)
1acb	0.20	1acb-pep	chymotrypsin	1.6359	ValThrLeuAspLeuArg (I43-I48)[c]	1.6300
1cho	0.18	1cho-pep	chymotrypsin	1.6308	CysThrLeuGluTyrArg (I16-I21)[c]	1.6251
1brc	0.25	1brc-pep	trypsin variant	1.6220	ProCysArgAlaMetIle (I13-I18)[c]	1.6148
1tab	0.23	1tab-pep	trypsin	1.6063	CysThrLysSerMetPro (I24-I29)[c]	1.5989
4er4	0.21	4er4-pep	pepsin	1.9561	ProHisProPheHisLeu ValIleHis(I1-I9)[c]	1.9372
5sga	0.18	5sga-pep	protease *A	1.4312	ProAlaProTyr(I1-I4)[c]	1.4304

[a] The crystallographic radius of gyration of isolated protein; [b] The crystallographic radius of gyration of protein-peptide complex; [c] The number of the peptide in the crystallographic ligand.

3. *Sampling*

Random sampling for each ligand was performed in a box that encloses the binding pocket of protein. The conformations of proteins and ligands were remained rigid in sampling. The sizes of the boxes are $1.2\times1.1\times1.7$ nm^3, $1.7\times0.8\times1.9$ nm^3, $1.3\times1.2\times1.7$ nm^3, $1.0\times1.2\times1.7$ nm^3, $1.4\times1.7\times2.3$ nm^3 and $1.1\times1.1\times0.9$ nm^3 for 1acb-pep, 1cho-pep, 1brc-pep, 1tab-pep, 4er4-pep and 5sga-pep, respectively. A three dimensional grid was constructed with steps of 0.05 nm in each dimension. There are 6 degrees of freedom for the rigid docking. Three translating degrees of freedom are the x, y, z coordinates of the grid points on which the nitrogen atom at N end of peptide is moved randomly. Three rotational degrees of freedom represent the orientation of peptide relative to protein. All the rotational angle spaces were divided with the steps of 1 degree.

In the procedure of sampling, we used the radius of gyration of complex, R_g, calculated with the three-dimensional coordinates of all non-hydrogen atoms in complex, as a criterion of packing compactness to screen the sampled configurations. The sampled configurations with R_g larger than $R_g(E)$, the radius of gyration of isolated protein, were removed in this stage, otherwise, accepted. The steric match restrain was further imposed after the compactness filter. Leonard-Jones 6-12 potential was used to evaluate the steric match between atoms. The tolerance of atomic overlapping was assigned as 1255 kJ/mol for all the systems. The configurations with the potential larger than the given tolerance were removed.

All the configurations passing the compactness and steric match filters were collected for the calculation of energies. The RMSD (Root-Mean Square Deviation) between docked configuration and the crystal structure was calculated by Kabsch method[13,14] using all non-hydrogen atoms.

Results

Fig. 1—Fig. 6 plot ΔE_{ff} vs. RMSD and ΔE_{ff+sol} vs. RMSD of the six test systems for the sampled configurations that have passed the compactness and steric match filters. In some of these figures, we can find an energetic value under which all the configurations are native-like while above which the configurations distributed over the entire RMSD range. This energetic value can be considered as a critical value of binding energy. The recognition is directive and specific if the binding energy of recognition system is lower than the critical value, on the opposite, the recognition process occurs randomly. We define this critical energy as $\Delta E_{ff}(C)$

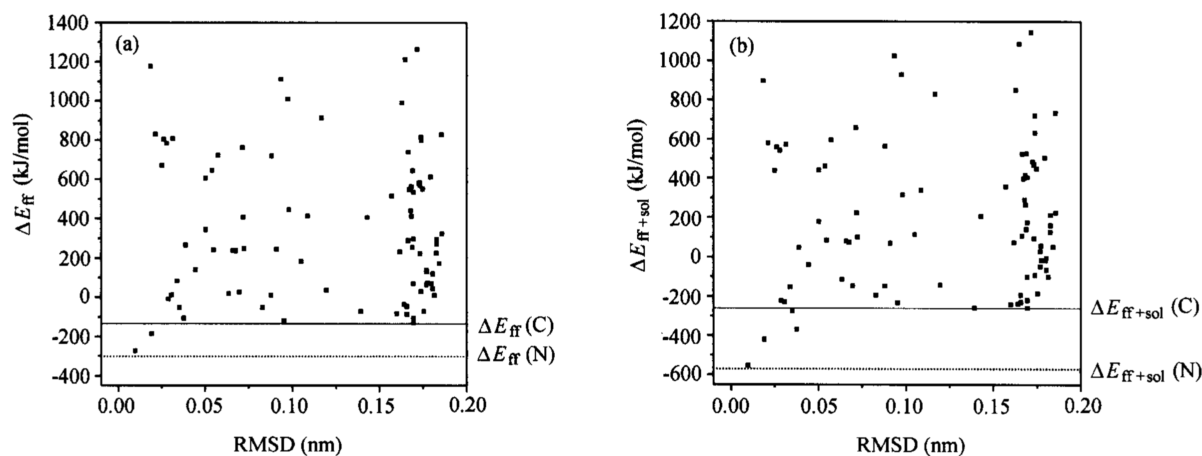

Fig. 1 Plots of ΔE_{ff} vs. RMSD (a) and ΔE_{ff+sol} vs. RMSD (b) for the rigid 1acb-pep system.

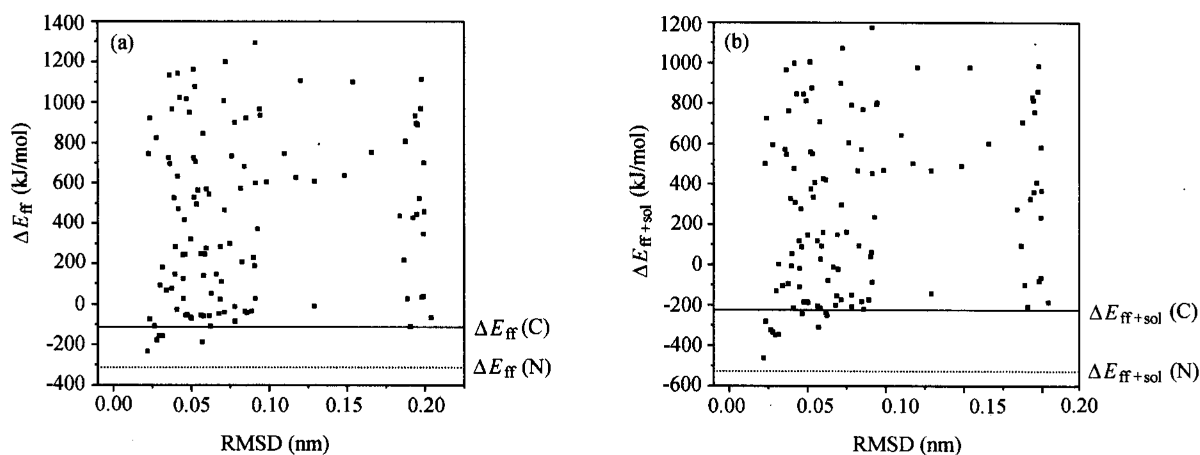

Fig. 2 Plots of ΔE_{ff} vs. RMSD (a) and ΔE_{ff+sol} vs. RMSD (b) for the rigid 1cho-pep system.

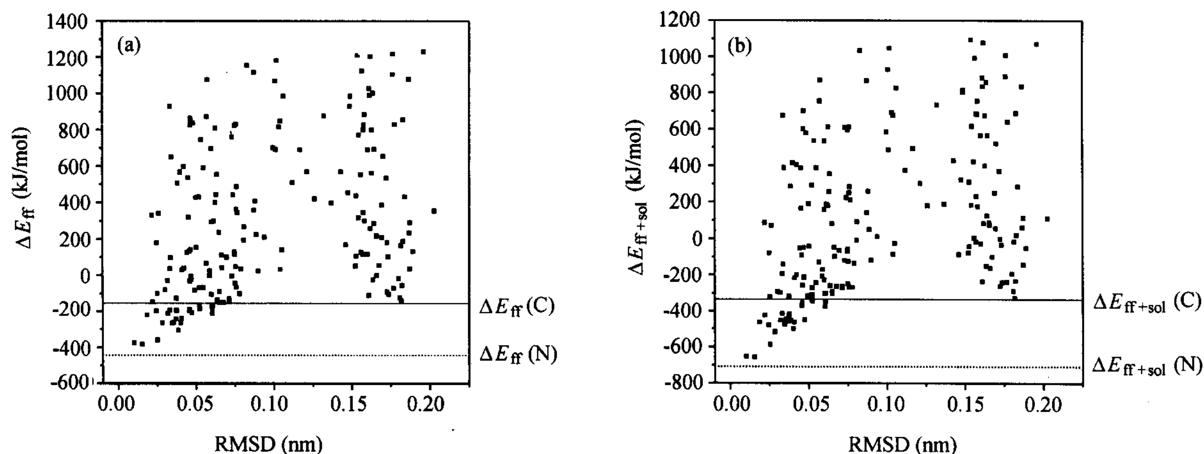

Fig. 3 Plots of ΔE_{ff} vs. RMSD (a) and ΔE_{ff+sol} vs. RMSD (b) for the rigid 1brc-pep system.

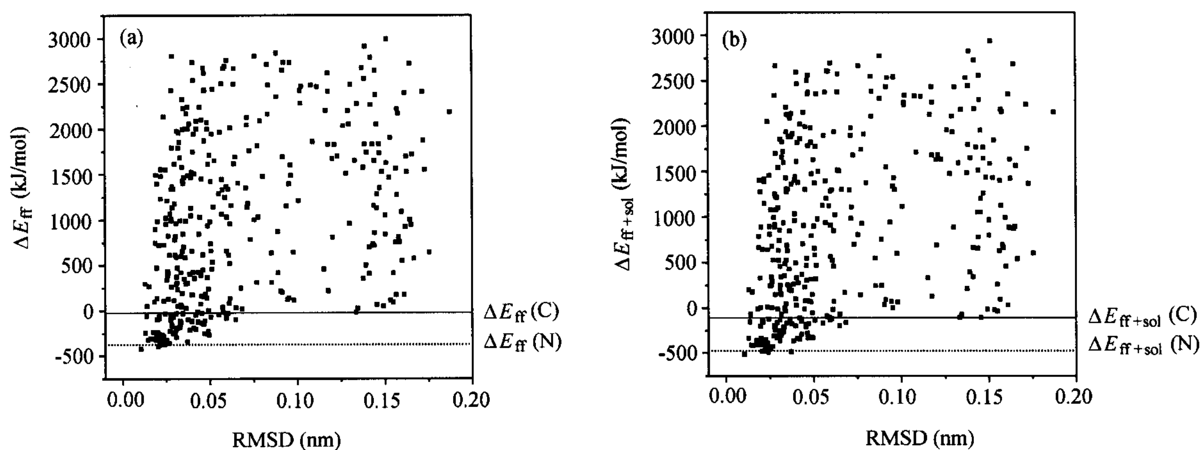

Fig. 4 Plots of ΔE_{ff} vs. RMSD (a) and ΔE_{ff+sol} vs. RMSD (b) for the rigid 1tab-pep system.

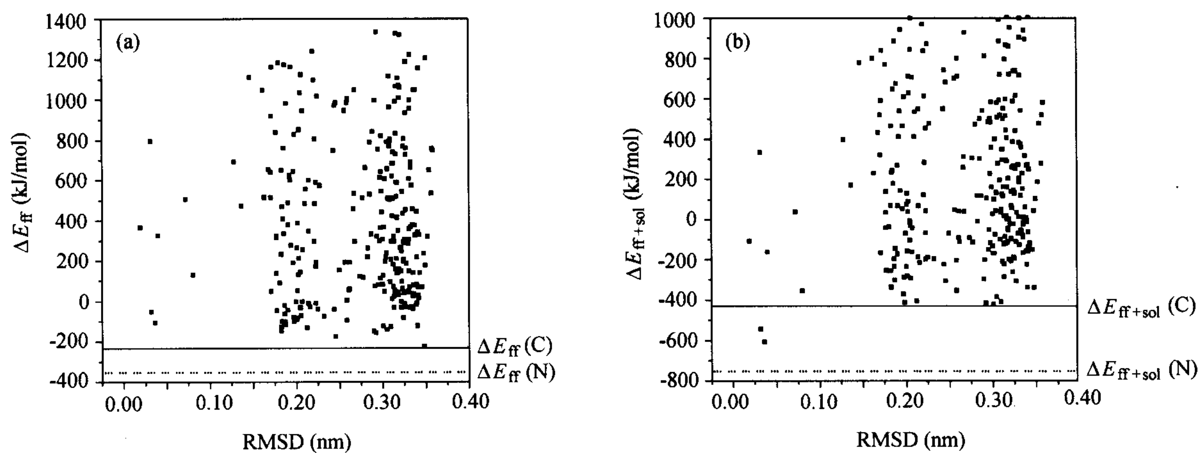

Fig. 5 Plots of ΔE_{ff} vs. RMSD (a) and ΔE_{ff+sol} vs. RMSD (b) for the rigid 4er4-pep system.

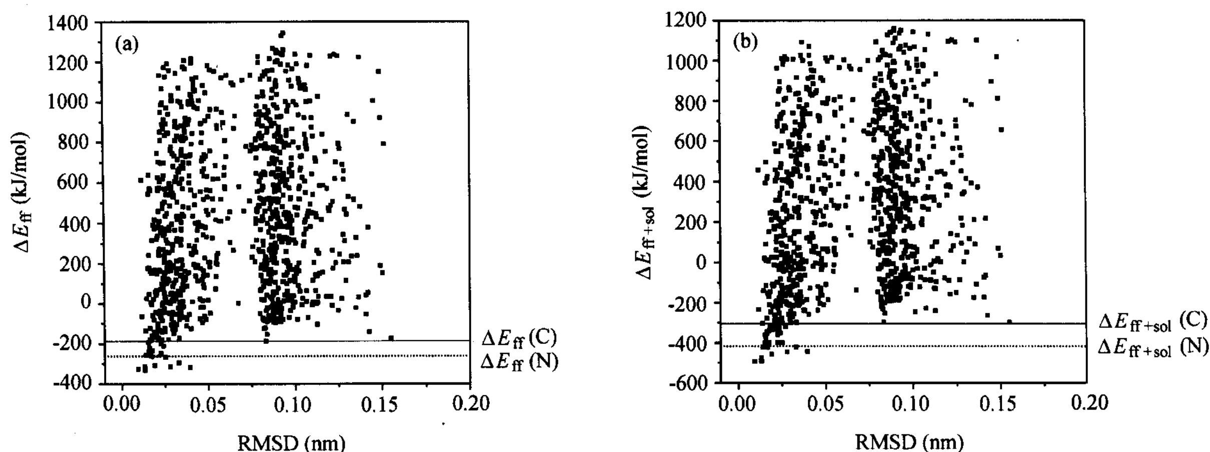

Fig. 6 Plots of ΔE_{ff} vs. RMSD (a) and ΔE_{ff+sol} vs. RMSD (b) for the rigid 5sga-pep system.

and $\Delta E_{ff+sol}(C)$ for the binding energy without the explicit solvation term and the total binding energy including the explicit solvation term, respectively. The amount of the configurations under this energy value is defined as the number of the native-like configurations identified by the binding energy. By contrast, the binding energies corresponding to the native structure are referred to $\Delta E_{ff}(N)$ and $\Delta E_{ff+sol}(N)$. The specific recognition process can happen only when $\Delta E_{ff+sol}(N)$ is lower the $\Delta E_{ff+sol}(C)$ (negative value).

Two factors are important to the occurrence of specific recognition. One is the effective collision between two molecular components under influence of intermolecular forces, which can be characterized by the number of the native-like configurations below $\Delta E_{ff+sol}(C)$, represented by n_{ff+sol}. The other is the energy difference between the specifically orientated state and the critical state (or the driving force that guides the recognition process specifically to the native structure), which can be characterized by the difference between $\Delta E_{ff+sol}(N)$ and $\Delta E_{ff+sol}(C)$, represented by $\Delta\Delta E_{ff+sol}$. The related data can be found in Fig. 1—Fig. 6. Therefore, we can evaluate the contribution of solvation energy from the proportion of solvation contributions to n_{ff+sol} and $\Delta\Delta E_{ff+sol}$, represented by $p(n_{sol})$ and $p(\Delta\Delta E_{sol})$, respectively.

$p(n_{sol})$ is defined as follows

$$p(n_{sol}) = \frac{n_{ff+sol} - n_{ff}}{n_{ff+sol}} \quad (4)$$

where n_{ff} represents the number of the native-like configurations below $\Delta E_{ff}(C)$. The another parameter $p(\Delta\Delta E_{sol})$ is expressed as

$$p(\Delta\Delta E_{sol}) = \frac{(\Delta\Delta E_{ff+sol}) - (\Delta\Delta E_{ff})}{(\Delta\Delta E_{ff+sol})}$$
$$\Delta\Delta E_{ff+sol}) < 0 \text{ and } \Delta\Delta E_{ff}) < 0 \quad (5)$$

where $\Delta\Delta E_{ff}$ is the separation between $\Delta E_{ff}(N)$ and $\Delta E_{ff}(C)$. These characteristic values and some related data are listed in Table 2.

Table 2 Data of n_{ff+sol}, n_{ff}, n, $p(n_{sol})$, $\Delta\Delta E_{ff+sol}$, $\Delta\Delta E_{ff}$ and $p(\Delta\Delta E_{sol})$ for the rigid test systems

Samples	1acb-pep	1cho-pep	1brc-pep	1tab-pep	4er4-pep	5sga-pep
n	87	114	189	211	280	956
n_{ff}	2	5	26	70	0	18
n_{ff+sol}	4	10	24	72	2	29
$p(n_{sol})$	0.5	0.5	-0.083	0.028	1	0.38
$\Delta\Delta E_{ff}$(kJ/mol)	-167	-201	-289	-356	-117	-71
$\Delta\Delta E_{ff+sol}$(kJ/mol)	-310	-301	-372	-368	-314	-113
$p(\Delta\Delta E_{sol})$	0.46	0.33	0.22	0.033	0.63	0.37

As shown by $p(n_{sol})$ in Table 2, the proportion of the solvation contribution to the number of the native-like configurations identified by the total binding energy ΔE_{ff+sol} is evidently larger for 4er4-pep, 1acb-pep, 1cho-pep and 5sga-pep systems than that for 1brc-pep and 1tab-pep systems. For 4er4-pep system, all the number of the native-like configurations identified by ΔE_{ff+sol} origins from the addition of the solvation energy in the force field energy. The 50% of the number of the native-like configurations identified by ΔE_{ff+sol} come from the contribution of the solvation energy for 1acb-pep and 1cho-pep systems and 38% for 5sga-pep system. This implies that the participation of the solvation interaction in the force field interaction leads to more effective collision of molecules for the four systems. Therefore, it would be inferred that the probability of forming native structure would increase for the four systems if two molecular components approach by the way of hydrophobic match. On the contrary, little difference between n_{ff+sol} and n_{ff} is observed for 1brc-pep and 1tab-pep systems. The values of $p(n_{sol})$ are nearly zero, meaning that the contribution of solvation interaction to the probability of effective collision is insignificant.

In the other hand, $\Delta E_{ff}(N)$ values of all the test systems are lower than $\Delta E_{ff}(C)$, $\Delta E_{ff+sol}(N)$ values are lower than $\Delta E_{ff+sol}(C)$ either. This indicates that all the binding energies are favorable to the formation of the native structure. However, the separations between

$\Delta E_{ff+sol}(N)$ and $\Delta E_{ff+sol}(C)$ are larger than those between $\Delta E_{ff}(N)$ and $\Delta E_{ff}(C)$ for all the test systems (see $\Delta\Delta E_{ff+sol}$ and $\Delta\Delta E_{ff}$ in Table 2 and Figs. 1—6), *i.e.*, the binding energy including the explicit solvation energy term is more favorable to the formation of the native structures. Nevertheless, the contribution of the solvation energy to the driving force $\Delta\Delta E_{ff+sol}$, represented by $p(\Delta\Delta E_{sol})$, is different for different test systems. The proportion of the partial driving force generated by the solvation binding energy to the total driving force is 63%, 46%, 37%, 33%, 22% and 0.03% for 4er4-pep, 1acb-pep, 5sga-pep, 1cho-pep, 1brc-pep and 1tab-pep systems, respectively. Obviously, the contribution of solvation energy to $\Delta\Delta E_{ff+sol}$ is rather important for the former four systems, while less important for 1brc-pep system and even negligible for 1tab-pep system.

Discussion

Summing the results of $p(n_{sol})$ and $p(\Delta\Delta E_{sol})$ for the test systems, we find that the proportion of solvation contribution to the effective collision and proportion to the driving force of the specific recognition are larger for 4er4-pep, 1acb-pep, 1cho-pep and 5sga-pep systems, while are smaller or nearly zero for 1brc-pep and 1tab-pep systems. The difference in the solvation contribution for various systems depends largely on the difference in the surface properties of ligands and their receptors. For 4er4-pep, 1acb-pep, 1cho-pep and 5sga-pep systems, the larger contribution of the solvation energy to determining the structures of the complexes is attributed to larger hydrophobic or aromatic groups of ligands penetrating the active cavities that are just specific to the hydrophobic or aromatic groups. While, the active site of trypsin in 1tab-pep or trypsin variant in 1brc-pep is a negatively charged group (Asp(E189) for trypsin and Asp(E226) for trypsin variant) locating on the bottom of the cavity that interacts specifically with positively charged Arg(I) or Lys(I) of ligand and forms a salt bridge. This kind of binding mode results in a severe decrease of solvation contribution both to the driving force and the number of the effective collision in the specific recognition process of 1brc-pep and 1tab-pep systems.

The correlation of the solvation contribution with the properties of protein and its ligand can be seen in detail by comparing two chymotrypsin and two trypsin systems. It has been shown that the proportion of the number of the effective collision generated by solvation energy to the total number of the effective collision is the same for 1acb-pep and 1cho-pep systems ($p(n_{sol})$ are all 0.5). While, the contribution of solvation energy to ΔE_{ff+sol} is different for the two systems. As we have seen from Table 1 that the main difference in the two chymotrypsin systems rests on some residues of ligands other than the specific residues (Leu(I45) for 1acb-pep and Leu(I18) for 1cho-pep). Therefore, it is inferred that the difference in the properties of these residues results in the difference in $p(\Delta\Delta E_{sol})$. Comparing the six-peptides of two chymotrypsin systems we find that two larger and apolar residue Val(I43) and Leu(I45) in the six-peptide of 1acb-pep are replaced by a smaller and apolar Cys(I16) and an aromatic Tyr(I20) in the six-peptide of 1cho-pep. The former is more favorable to the solvation binding energy than the latter. As a result, $p(\Delta\Delta E_{sol})$ is larger for 1acb-pep system than that for 1cho-pep system.

The similar results are also obtained for two trypsin systems. $p(n_{sol})$ values are nearly the same, while, $p(\Delta\Delta E_{sol})$ values are different for 1brc-pep and 1tab-pep systems. $p(\Delta\Delta E_{sol})$ of 1brc-pep is larger than that of 1tab-pep. The difference in $p(\Delta\Delta E_{sol})$ is also generated mainly from the ligand residues. For 1brc-pep, all the residues of six-peptide except the specific group Arg (I15) are apolar. However, there are two polar residues beside the specific residue Lys(I26) in the six-peptide of 1tab-pep. More apolar residues in the ligand of 1brc-pep result in larger $p(\Delta\Delta E_{sol})$.

An effective collision between a protein and its ligand depends on an appropriate orientation of two molecules under intermolecular forces. While, the appropriate orientation of two molecules is determined predominantly by the properties of the specific groups of ligands and the structures and properties of their receptors. For two chymotrypsin systems, the protein receptors are the same and the specific residues of ligands are also the same. Similarly, for two trypsin systems, their protein receptors are nearly the same in structure and property and their specific residues of ligands are also similar in property. It is just the similarity in protein structures and specific groups of ligands that leads to the nearly equal contribution of solvation energy to the number of the effective collision between 1acb-pep and 1cho-pep and

between 1brc-pep and 1tab-pep.

Now we can drown following conclusions: (1) The solvation binding energy is important in guiding molecular recognition of 4er4-pep, 1acb-pep, 1cho-pep and 5sga-pep systems, while it is less important for 1brc-pep system and even negligible for the 1tab-pep system; (2) The proportion of the contribution of the solvation binding energy to molecular recognition depends primarily on the properties of specific interaction groups of ligands and the surface properties of their receptors. (3) For the same type of recognition systems, more hydrophobic residues in the vicinity of specific groups are favorable to solvation binding energy.

Acknowledgements

The calculation obtained the support of the State Key Laboratory of the Chemical Computation.

References

1 Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C; Ghio, C.; Alagona G.; Profeta, S.; Weiner, P. *J. Am. Chem. Sci.* **1984**, *106*, 765.
2 Clark, K. P.; Ajay *J. Comput. Chem.* **1995**, *16*, 1210.
3 Jonnes, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
4 Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
5 Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.
6 Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.
7 Ponnuswamy, P. K. *Prog. Biophys. Mol. Biol.* **1993**, *59*, 57.
8 Wang, Y.; Zhang, H.; Scott, R. A. *Protein Sci.* **1995**, *4*, 1402
9 Wang, Y.; Zhang, H.; Li, W.; Scott, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 709.
10 Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* **1986**, *323*, 533.
11 Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351.
12 Scott, M. L. G.; Kenneth, M. M. *J. Comput. Chem.* **1993**, *14*, 349.
13 Kabsch, W. *Acta Crystallogr. A* **1976**, *32*, 922.
14 Kabsch, W. *Acta Crystallogr. A* **1978**, *34*, 827.

(E200007138 SONG, J.P.; DONG, L.J.)