Send Orders of Reprints at Reprints@benthamscience.org

70 *The Open Conference Proccedings Journal,* **2012**, *3*, 70-80

Open Access

Two Rules on the Protein-Ligand Interaction

Xiaodong Pang[1,2,#], Linxiang Zhou[1,#], Mingjun Zhang[3], Lili Zhang[4], Lina Xu[4], Fang Xie[3], Long Yu[3] and Xinyi Zhang[1,2,*]

[1]*Surface Physics Laboratory, Fudan University, Shanghai*

[2]*Synchrotron Radiation Research Center, Fudan University, Shanghai*

[3]*State Key Laboratory of Genetic Engineering, Fudan University, Shanghai*

[4]*Rice University, Texas, USA*

Abstract: So far, we still lack a clear molecular mechanism to explain the protein-ligand interaction on the basis of electronic structure of a protein. By combining the calculation of the full electronic structure of a protein along with its hydrophobic pocket and the perturbation theory, we found out two rules on the protein-ligand interaction. One rule is that the interaction only occurs between the lowest unoccupied molecular orbitals (LUMOs) of a protein and the highest occupied molecular orbital (HOMO) of its ligand, not between the HOMOs of a protein and the LUMO of its ligand. The other rule is that only those residues or atoms located in both the LUMOs of a protein and a surface pocket of a protein are active residues or active atoms of the protein and the corresponding pocket is the ligand binding site. These two rules are derived from the structure of energy levels of a protein and could be one of important criterions of the drug design. They were validated on complex CypA/CsA and FKBP12/FK506. Furthermore, a tripeptide Ala-Gly-Pro (AGP) was predicted to bind to CypA as an application of our two rules. And the biological assays results show that AGP has the same order of binding affinity and inhibition of PPIase activity as CsA.

Keywords: Electronic structure, frontier molecular orbitals, HOMO, LUMO, peptidic inhibitor design, protein-ligand interaction.

INTRODUCTION

Understanding ruling principles of interactions between a target protein and a ligand is of paramount importance in drug discovery efforts. So far, in finding a real ligand for a given target protein, we are limited to experimental screening from a large number of small molecules, or through free energy calculations of assessing a ligand. However, we still lack a clear molecular mechanism to explain the protein-ligand interaction on the basis of electronic structure of a protein and guide novel molecules designing. Here we report two rules on the protein-ligand interaction, derived from full electronic structure calculation of a protein. These two rules may serve as a new means of drug design and exploring enzyme reaction.

The full electronic structure calculation of a protein has been a challenge, as the system of a protein is huge and complicated. We use the overlapping-dimer approximation (ODA) [1-3] and the extended negative factor counting (ENFC) [4] methods to calculate the electronic structure of a protein. These two methods were first developed by J. Ladik and Y. J. Ye *et al.* [1, 4] We further clarified and revised some parts to facilitate calculations of proteins with over 600 residues.

The ODA uses the divide and conquer algorithm, which works by recursively breaking down a complex problem into many subproblems of the same (or related) type of the same scale. These subproblems are independent of each other, and their characters are the same as the original problem. We can use the solutions of these small subproblems to get the solution of the original problem. First of all, the whole chain of a protein is divided into many dimers, each containing two amino acid residues, and then we can perform any quantum calculation methods for these dimers; finally, construct the full Hamiltonian matrix using the solution of each dimer and use the ENFC math method to get the energy eigenvalues and the orbital coefficients of the wave function of a protein. It should be noted that the surrounding environment of each dimer and the water environment of a protein are applied in the process.

We take the protein-ligand complex Cyclophilin A (CypA) and its ligand Cyclosporin A (CsA), and FKBP12 of the protein family FKBPs and its immunosuppressant FK506 as examples to validate our two rules, and the results obtained are in good agreement with experiments. Furthermore, as an application of our two rules, a tripeptide AGP was predicted to bind to CypA and the biological assays results show that AGP has the same binding affinity and inhibition of PPIase activity as CsA.

The system of a protein molecule is so large that the energy levels are extremely compact. Therefore, for a protein system, we need to extend the definition of the frontier

*Address correspondence to this author at the Surface Physics Laboratory, Fudan University, Shanghai; Tel: +86-21-65643522,
Fax: +86-21-65643626. E-mail: xy-zhang@fudan.edu.cn

[#] These authors contributed equally to this work

2210-2892/12

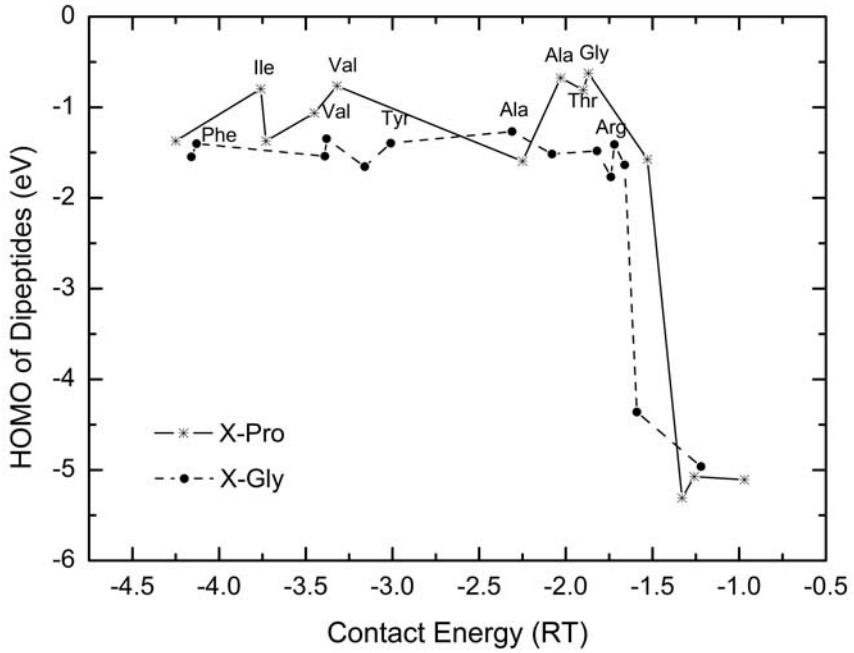

Fig. (1). The HOMO energies of dipeptide X-Pro and X-Gly. The dipeptide Gly-Pro has the highest HOMO energy among X-Pro. The dipeptide Ala-Gly has the highest HOMO energy among X-Gly.

orbitals to be a series of HOMOs and LUMOs, not just limited to be a single HOMO or LUMO. Ordinarily, the first 10 to 20 occupied molecular orbitals and the unoccupied ones are considered to be HOMOs and LUMOs, respectively. For example, protein CypA has a total of 4807 occupied molecular orbitals. We considered the first 10 highest occupied molecular orbitals to be its HOMOs and the first 10 lowest unoccupied molecular orbitals as its LUMOs. The 10 HOMOs comprise only 0.2% among total 4807 occupied molecular orbitals.

MATERIALS AND METHODS

Full Electronic Structure Calculation

We used ODA and a math method ENFC to calculate the electronic structure of protein CypA (165 residues) and FKBP12 (107 residues). After molecular dynamic (MD) simulation, we obtained the coordinates of the proteins and ligands at temperature of interest. We needed to transform the format of these pdb files into ODA format. Then, we followed the four steps of ODA mentioned above to calculate their eigenvalues, orbital coefficients of wave functions and density of state.

We used program ArgusLab [5] to calculate the HOMO and the LUMO of ligand CsA and FK506.

Designing of Peptide Inhibitor AGP

How to create this novel ligand AGP? The procedure is like Hidden Markov Chain's (HMM) as follows:

(1). Selecting an Initial Residue Pro

MJ matrix (the effective inter-residue contact energy evaluated by Miyazawa and Jernigan [6, 7]) is introduced to refine an active residue, with which the active residue Phe113 of CypA should has the lowest contact energy. Phe113 is taken as a target residue, and has the lowest contact energy with Pro (-4.25 RT) from the MJ matrix. In other words, Phe is the target of the highest probability to be bound by Pro. Therefore, residue Pro is the best option as the active residue of ligand.

Besides, the binding pocket is mainly composed of hydrophobic or non-polar residues. Therefore, the active part of the ligand fitting into the binding pocket should also be hydrophobic or non-polar. Pro happens to be a hydrophobic and non-polar residue.

(2). Taking Peptides as a Markov Chain to Calculate the HOMO

Suppose a peptides is composed of n residues, $S = S_1,S_2,S_3,...,S_n$. If we take it as a Markov Chain, then its probability is

$$P(S) = P(S_1)P(S_2 \mid S_1)P(S_3 \mid S_2)\cdots P(S_n \mid S_{n-1})$$

According to Bayes formula, we have

$$P(S_i \mid S_{i-1}) = P(S_{i-1}S_i)/P(S_{i-1})$$

Therefore, we can calculate the HOMO one pair residues by one pair residues.

First, we calculate HOMO of dipeptide X-Pro, where X represents one of the 20 residues. And, the HOMO vs. MJ contact energy is shown in Fig. (**1**). We can see that Gly-Pro, Ala-Pro, Val-Pro, Ile-Pro and Thr-Pro are of high HOMO energies, and the highest one is Gly-Pro. This can perfectly explain the suggestion proposed by Vajdos *et al.* [8] that CypA can function as a sequence-specific binding protein, with a strong preference for the sequence X-Gly-trans-Pro, which challenges the long-held assumption that CypA functions primarily as a non-specific proline isomerase.

Our two rules tell us that the potential ligand candidate should satisfy the following two conditions. 1) The higher the HOMO energy of a ligand candidate is the better, as the LUMO energy of the given target protein CypA is fixed. 2) The active atoms of a ligand candidate should be closed to

the active atoms of CypA, such as those of Phe113 and Phe60. Therefore, we select Gly-Pro from dipeptides X-Pro, and then fix Gly to calculate the HOMO of dipeptide X-Gly (Fig. **1**). Dipeptides Ala-Gly, Val-Gly, Tyr-Gly, Phe-Gly and Arg-Gly are of high HOMO energy, and the highest one is Ala-Gly. Therefore, the best tripeptide is Ala-Gly-Pro (AGP) as a combination of X-Gly and X-Pro dipeptides. It can explain the previous findings that CypA recognition of hexapeptides involves contacts with peptide residues Ala, Gly and Pro, and is independent of the context of longer sequences [8].

Docking

The program Autodock [9] was employed to dock the AGP initial conformation into the binding site of CypA. A grid box was centered in the ligand binding site and sized of 60x56x60 points to embrace all the residues making up the binding pocket using the default grid spacing 0.375. The active residue Phe113 of CypA was set as flexible one and the AGP was placed above the ligand binding site. We used the Larmarckian genetic algorithm (LGA) for docking with the number of energy evaluations of 2,500,000 and the maximum number of generation of 27000. And 100 LGA runs were performed.

Molecular Dynamic (MD) Simulation

Complex CypA/CsA and FKBP12/FK506 were performed MD simulation to obtain the ground state and the state of 300 K, respectively. Their starting coordinates were taken from the protein database bank [10] (PDB) with entry 1CWK and 1FKJ, respectively. The parallel MD program we used is NAMD [11] and all the molecules were solvated before the MD simulation.

However, for a ligand or other small molecules, there is often no topology and parameter information in the standard CHARMM's library. In that case, the web site HIC-UP [12] can help us to find the topology and parameter files for these small molecules. These files are X-PLOR version [13], not CHARMM version. The ligand FK506 was treated in this way.

For ligand CsA, a cyclic undecapeptide with the following primary structure: MeBmt1-Abu2-Sar3-MeLeu4-Val5-MeLeu6-Ala7-D-Ala8-MeLeu9-MeLeu10-MeVal11. Residues 1, 3, 4, 6, 9, 10 and 11 are N-methylated on the amide nitrogen. Here, we created its topology and parameter information based on existing topology information from other molecules.

The basic idea to create the topology file of CsA is as follows:

(1) For residues MeBmt1 and Abu2, of which both not belong to the 20 kinds of standard residues of protein, we need to write out their topology files. Details of creating these two topology files can be found at the book[6] from page 125 to 139.

(2) We used program PSFGEN [11] to build the protein data bank (PDB) file and protein structure file (PSF) of CsA by using a patch LINK to link MeBmt1 and MeVal11 since it's a cyclic molecule and a patch METH to methylate the amide nitrogen of resides 1, 3, 4, 6, 9, 10 and 11.

After we get the topology file of CsA, we can use the program PSFGEN to build its new PDB and PSF files. PSFGEN, a CHARMM version, can generate a revised PDB and PSF files with hydrogen atoms added, as well as the water environment

For protein CypA, consisting of 165 residues and 2053 atoms, we also used the PSFGEN to get its new PDB and PSF files.

Finally, we set up the configuration file to run the MD simulation. The parameters for CHARMM force field was selected as follows: exclude = scaled 4, 1-4 scaling = 1.0, switching = on, switchdist = 8 Å, cutoff = 12 Å, pairlestdist = 13.5. We minimized the CypA and CsA for 80ps, separately.

For ligand FK506, with chemical formula C44H69NO12, we get its topology and parameter files from HIC-UP server and run program XPLOR64 [13] to get its new PDB and PSF file. The XPLOR64, a XPLOR version, can generate a revised PDB and PSF files using topology file obtained from CNS [13].

For protein FKBP12, we also used XPLOR to generate its new PDB and PSF files.

We use the same MD parameters as CypA/CsA for FKPB12 and FK506 separately. They were firstly relaxed to near 20 K by running 20 ps. And then it was heated continuously from 20K to 290K with the temperature step 30 K, running 30 ps for each step. And we performed a smaller temperature step 5 K from 295 K to 300 K with 30 ps each step. Finally, run 50 ps at 300 K. The total time scale is 520 ps for the whole process using 16 CPUs. The same energy minimization procedure was applied on the complex system of CypA/AGP suggested by docking.

Proteins and Compounds

The full-length coding sequence of *CypA* was inserted in frame into plasmid pET28a (Clontech) to generate recombinant human CypA protein. HIS-CypA protein was expressed and purified (Fig. **2**). CsA, the substrate N-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanlilide (Suc-AAPF-pNA), α-chymotrypsin, and 2,2,2-trifluorethanal (TFE) were purchased from Sigma (St. Louis, MO). The peptides were purchased from Shanghai science peptide biological technology co., LTD.

Surface Plasmon Resonance (SPR) Analysis

Recombinant human CypA protein (10 μM) was immobilized on a carboxylmethylated dextran surface (CM5 chip from Biacore, Inc., Piscataway, NJ) in a buffer containing 10 mM sodium acetate (pH 4.0) using standard amine coupling chemistry procedure following the manufacturer's instructions. Flow cell 1 was used as the control surface and flow cell 2 contained 4000 resonance units (RU) of CypA protein (1 RU corresponds to 1 pg of protein per mm^2). All compounds were stocked in 100% dimethylsulfoxide at 10 mM and diluted at graded concentrations (0.3 to 10 μM) with PBS running buffer (pH 7.4) containing 5% dimethylsulfoxide. The temperature of the instrument was set to 25 °C. Each compound (50 μl) was injected sequentially by flow rate 30 μl/min. NaOH (20 mM) was used to regenerate the active surface between cycles. The 1:1 Langmuir binding model was used to analyze the equilibrium dissociation constant (K_D). For fast interactions,

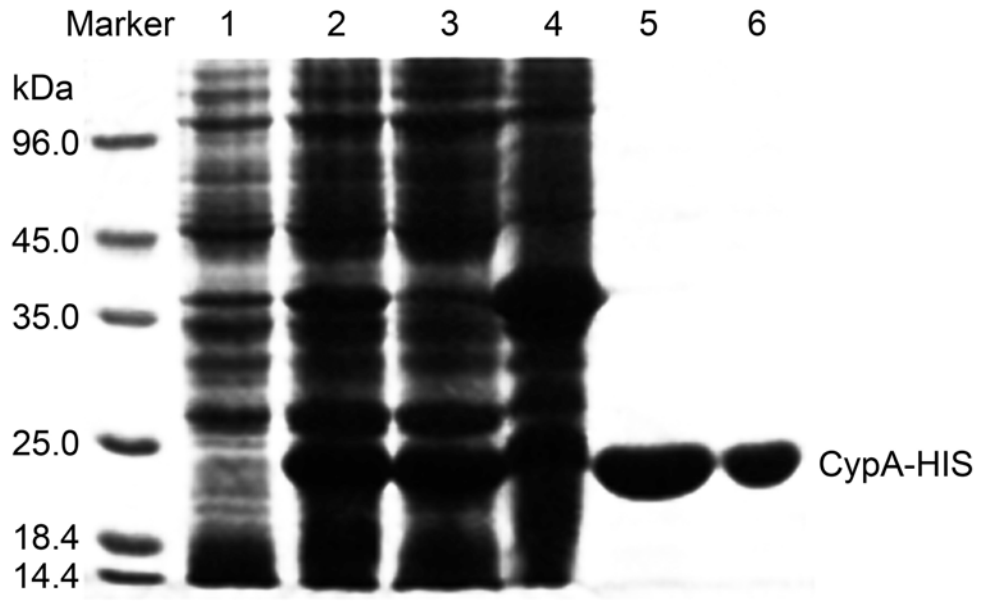

Fig. (2). Expression and purification of recombinant human CypA protein. HIS-CypA runs as a 23 kDa protein on SDS-PAGE. M: marker; lane 1: bacterial lysis before IPTG induction; lane 2: bacterial lysis after IPTG induction; lane 3: supernatane after centrifugation; lane 4: pellet after centrifugation; lane 5: CypA protein after purification by Ni-NTA beads (Novagen); lane 6: CypA protein dialyzed with PBS buffer.

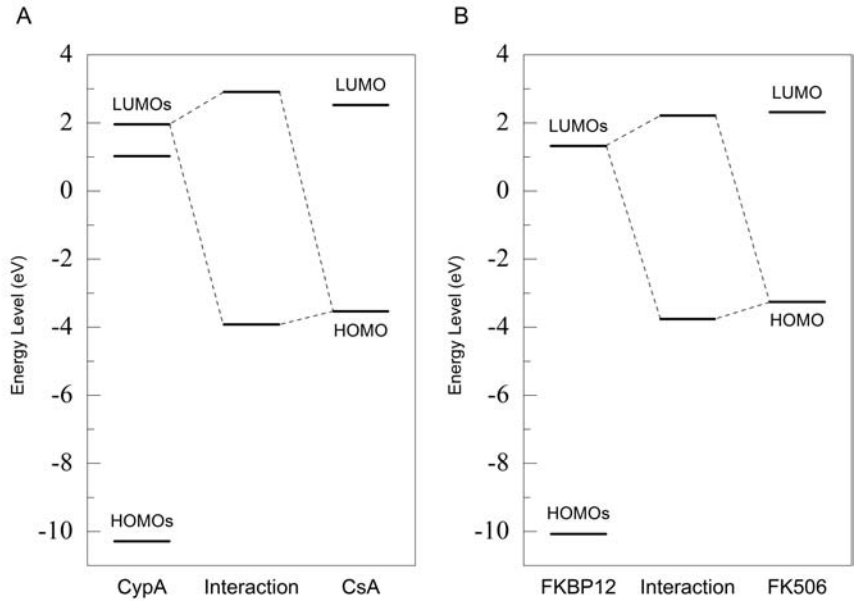

Fig. (3). Schematic diagrams of the energy level. (**A**) Protein CypA interacts with ligand CsA at ground state. (**B**) Protein FKBP12 interacts with ligand FK506 at 300 K. The structure of energy levels of a protein is that the LUMOs of a protein are always about 1.0 eV above the zero-energy-level, but its HOMOs are always far away below the zero-energy-level (about -10.0 eV). It is the very characteristics that result in the first rule.

the steady state model was used to determine the K_D values. All experiments were carried out in triplicate.

Enzyme Inhibition Assays

Inhibition of each compound on PPIase activity of CypA was determined at 4 °C based on the published method with some modifications [14]. Suc-AAPF-pNA was dissolved in TFE containing 480mM of LiCl to a concentration of 3 mM. Then, α-chymotrypsin was dissolved in 1 mM HCl to a concentration of 42 mg/ml. Each test compound was diluted in 94 μl of assay buffer (50 mM HEPES, 100 mM NaCl; pH 8.0 at 0 °C), and then mixed with 2 μl of CypA solution (5 μM). After pre-equilibrating for 3 h on ice, 2 μl of α-chymotrypsin solution and 2 μl of peptide substrate were added to the assay mixture. Absorbance at the wavelength of 390 nm was recorded on a Jasco V-550 spectrophotometer (Jasco, Inc., Easton, MD) for 20 s. Three independent experiments were performed for each test compound.

RESULTS AND DISCUSSION

First Rule of the Protein-Ligand Interaction.

Fig. (**3**) and Table **1** express the partial results of the frontier orbitals and orbital coefficients for CypA/CsA and

Table 1. The first 10 LUMOs of CypA and FKBP12, and the HOMO of CsA and FK506. The Definition of Occupancy at *nth* Residue is $a(n) = \sum_{j=1}^{m} \eta C_j^2(n) / \sum_{n=1}^{n=N} \sum_{j=1}^{m} \eta C_j^2(n)$, where C_j are the Orbital Coefficients. Those Residues that have *a(n)* < 0.05 are Neglected. The Wave Function Localizes on a Single Residue, Since the Occupancy of each Orbital is Larger than 99%. For CypA, Residues PHE60 and PHE113 are Active Residues; for FKBP12, Residues TRY26, PHE36, PHE46, TRP59, PHE82 and PHE99 are Active Residues; for CsA, Residue MVA11 is the Active Residue; for FK506, Atoms C2, C8, C9, C10, C11, N7, O3, O4 and O6 are Active Atoms

	Molecular Orbital	Energy Band (eV)	Position of Wave Function (Residues/Atoms)[a]	Occupancy
CypA	4817	1.959	PHE36	0.996
	4816	1.956	PHE60	0.996
	4815	1.954	PHE113	0.995
	4814	1.930	PHE145	0.998
	4813	1.880	PHE53	0.998
	4812	1.866	PHE67	0.996
	4811	1.850	PHE53	0.997
	4810	1.849	TYR79	0.998
	4809	1.690	TYR48	1.000
	4808	1.020	TRP121	0.996
FKBP12	3169	2.261	PHE48	0.993
	3168	2.141	TYR26	0.998
	3167	2.109	PHE36	0.995
	3166	2.101	PHE82	0.998
	3165	1.988	PHE15	0.999
	3164	1.932	PHE46	0.999
	3163	1.811	PHE48	0.998
	3162	1.754	PHE99	0.997
	3161	1.620	PHE36	0.992
	3160	1.328	TRP59	0.999
CsA	246	-3.519	MVA11	
FK506	841	-3.255	C2,C8,C9,C10, C11,N7,O3,O4,O6	

[a] Bold: active residue or active atom.

FKBP12/FK506 separately (see Tables **S1**, **S2** and **S3** for more details about the frontier orbitals).

We define the energy difference between a protein and its ligand as

$$DE_1 \equiv E_{proteinLUMO} - E_{ligandHOMO},$$

$$DE_2 \equiv E_{ligandLUMO} - E_{proteinHOMO},$$

where $E_{proteinLUMO}$ is the average value of the active residues and $E_{proteinHOMO}$ is the highest value of HOMOs, $E_{ligandLUMO}$ and $E_{ligandHOMO}$ are values of the LUMO and the HOMO of ligand, respectively.

The calculation result demonstrates as follows:

for CypA/CsA,

$$DE_1 = 1.96 - (-3.52) = 5.48eV < DE_2 = 2.52 - (-10.29) = 12.80eV,$$

for FKBP12/FK506,

$$DE_1 = 1.33 - (-3.26) = 4.59eV < DE_2 = 2.31 - (-10.07) = 12.38eV.$$

Especially, we have to point out that the results of CypA/CsA and FKBP12/FK506 are examples only. We calculated other several protein/ligand systems and obtained the same resulting relationship $DE_1 < DE_2$ as the above two systems without exception [15, 16]. Therefore, the $DE_1 < DE_2$ might be a generally correct relationship in protein-ligand systems.

According to the frontier orbital theory, the perturbation energy of interaction between two molecules mainly comes from the interaction between HOMO and LUMO. If the energy of a protein is different from the energy of its ligand, the effect of the interaction produces an energy split. As a result, the smaller the energy difference between two orbitals, the stronger the interaction. Whether the interaction between two molecules can form a bond, it is related to the symmetry of the molecular orbital. We can, therefore, deduce the first rule on the protein-ligand interaction:

The protein-ligand interaction only occurs between the lowest unoccupied molecular orbitals (LUMOs) of a protein and the highest occupied molecular orbital (HOMO) of its ligand, not between the HOMOs of a protein and the LUMO of its ligand.

Table 2. The Forming Atoms in the Pocket 19 of CypA. Only the Residues PHE60 and PHE113 are located on the LUMOs of CypA, and all other Residues on the LUMOs of CypA are not Located in other Pockets. So, the Pocket 19 is the Ligand Binding Site and the PHE113 and PHE60 are the Active Residues (From Table S4, we know that the Pocket 13 is the Ligand Binding Site of FKBP12 and Residues TRY26, PHE36, PHE46, TRP59, PHE82 and PHE99 are Active Residues)

Atom[a]	Res[b]	ID[c]	Atom	Res	ID	Atom	Res	ID	Atom	Res	ID
NH1	ARG	55	CA	ASN	102	C	GLY	74	HE22	GLN	111
HH11	ARG	55	C	ASN	102	O	GLY	74	**CB**	**PHE**	**113**
CE2	**PHE**	**60**	O	ASN	102	CA	GLY	75	**CG**	**PHE**	**113**
CZ	**PHE**	**60**	N	ALA	103	C	GLY	75	**CD1**	**PHE**	**113**
CG	MET	61	CB	ALA	103	O	GLY	75	**CD2**	**PHE**	**113**
SD	MET	61	CB	THR	107	CA	LYS	76	**CE1**	**PHE**	**113**
CG	GLN	63	O	THR	107	OE1	GLU	81	**CE2**	**PHE**	**113**
CD	GLN	63	N	GLY	109	O	GLU	81	**CZ**	**PHE**	**113**
OE1	GLN	63	CA	GLY	109	CG	LYS	82	CD1	LEU	122
NE2	GLN	63	N	SER	110	CE	LYS	82	HD13	LEU	122
HE21	GLN	63	N	GLN	111	CA	ALA	101	CD2	LEU	122
O	GLY	72	CB	GLN	111	CB	ALA	101	HD21	LEU	122
CA	THR	73	CG	GLN	111	C	ALA	101	HD22	LEU	122
C	THR	73	CD	GLN	111	O	ALA	101	CE1	HIS	126
O	THR	73	OE1	GLN	111	N	ASN	102			
CA	GLY	74	NE2	GLN	111	HE21	GLN	111			

[a] Bold: potential active atom. [b] Bold: active residue. [c] Bold: active residue ID.

Actually, it is the very characteristics of energy level structure of a protein that result in the first rule. The characteristics are that the LUMOs of a protein are always about 1.0 eV above the zero-energy-level, but its HOMOs are always far away below the zero-energy-level (about -10.0 eV). It is due to this kind of structure that the LUMOs of a protein are always near the HOMO of its ligand and the HOMOs of a protein are always far away from the LUMO of its ligand. That is why always there is the relationship $DE_1 < DE_2$ in the protein-ligand interaction. The first rule may be regarded as one of criteria of identifying a ligand. An interesting aside is that the wave function of a protein usually localizes on one or few residues (the occupancy item on Table **1**).

Second Rule of the Protein-Ligand Interaction

According to the first rule, as a necessary condition, the active residues or active atoms of a protein must be located on the LUMOs of the protein, but conversely, not every residue on the LUMOs is certainly the active residue. When a ligand interacts with its target protein, it must enter into a pocket on the surface of the protein. For the same reason, only the pocket which has its forming atoms located on the LUMOs is the ligand binding site. Therefore, on the basis of the full electronic structure of a protein along with the pocket calculation, we can derive the second rule that only those residues or atoms located in both the LUMOs of a protein and a surface pocket of a protein are active residues or active atoms of the protein and the corresponding pocket is the ligand binding site. This is a necessary and sufficient condition for being the active residues or active atoms, as well as the ligand binding site.

Now we can use this rule to check the active atoms of a protein. We selected CASTp method [17] to calculate pockets of protein. The CASTp is a geometry-based method of pocket detection, based on the alpha shape and discrete flow theory, and a related suite of programs. It can provide full descriptions of pockets on protein surface, including the atoms, residues, and volume of each pocket. For example, the CypA has a total of 19 pockets and the largest pocket is number 19. The FKBP12 has a total of 13 pockets. Table **2** and Table **S4** show the pocket 19 of CypA and pocket 13 of FKBP12, respectively. Tables **3**, **4** and Table **S5** describe active atoms and their orbital coefficients of protein CypA, ligand CsA and FK506, and protein FKBP12, respectively.

We can see that none but the residues PHE60 and PHE113 are located not only on the LUMOs of CypA, but also in the pocket 19 (compare Table **1** and Table **2**). All other residues on the LUMOs of CypA are not located in any other pocket. Therefore, we can say that the pocket 19 is the ligand binding site and the PHE113 and PHE60 are the active residues or that there are seven active atoms: PHE113: CG, CD1, CD2, CE1, CE2, and CZ (these six atoms form a quincunx in Fig. **4A**), and PHE60:CE2, as shown in Table **3**. Though the atoms PHE113: CB and PHE60: CZ are located on the LUMOs and in the pocket 19, their orbital coefficients are zero (Table **3**), thus they are not active atoms. Similarly, though the atoms PHE113: CA and PHE113: HB located on the LUMOs of CypA, they are not active atoms because they are not located in the pocket 19.

Table 3. Orbital Coefficient of the Active Residues PHE113 and PHE60 of CypA on the LUMOs. Those Atoms Whose Absolute Value of Orbital Coefficient is less than 0.1 are Neglected. Therefore, Atom CE2 of PHE60 and Atoms CG, CD1, CD2, CE1, CE2 and CZ of PHE113 are the Active Atoms. (AR: Active Residue; AO: Atomic Orbital; OC: Orbital Coefficient)

AR	Atom[a]	AO	OC
PHE60	CD1	2X	-0.440
		2Y	0.114
		2Z	-0.403
	CD2	2X	0.439
		2Y	-0.112
		2Z	0.398
	CE1	2X	0.434
		2Y	-0.112
		2Z	0.396
	CE2	2X	-0.457
		2Y	0.116
		2Z	-0.419
PHE113	CA	2S	-0.145
		2X	0.100
		2Z	0.102
	HB	1S	0.147
	CG	2X	0.211
		2Y	-0.494
		2Z	0.420
	CD1	2X	-0.153
		2Y	0.386
		2Z	-0.307
	CD2	2Y	0.121
	CE1	2Y	0.112
	CE2	2X	-0.149
		2Y	0.387
		2Z	-0.293
	CZ	2X	0.192
		2Y	-0.496
		2Z	0.384

[a] Bold: active atom.

Table 4. Orbital Coefficient of Active Atoms of CsA and FK506 on the HOMO. Those Atoms Whose Absolute Value of Orbital Coefficient is less than 0.1 are Neglected. The Residue MVA11 is the Active Residue of CsA with Four Active Atoms: N, CA, C and O. FK506 has Nine Active Atoms: C2, C8, C9, C10, C11, N7, O3, O4 and O6

CsA			FK506					
MVA11: N	2Pz	-0.144	C2	2Py	-0.123	N7	2S	0.100
MVA11: CA	2S	0.211	C8	2S	-0.151		2Px	-0.120
	2Px	0.160		2Py	-0.235		2Pz	0.123
	2Py	-0.313	C9	2Py	0.151	O3	2Px	-0.111
	2Pz	-0.154		2Pz	0.132		2Py	0.378
MVA11: C	2S	-0.380	C10	2S	0.147	O4	2Py	-0.451
	2Py	0.337		2Py	-0.117		2Pz	-0.441
	2Pz	-0.443		2Pz	-0.331	O6	2Pz	0.112
MVA11: O	2Py	-0.341	C11	2Pz	0.113			
	2Pz	0.362						

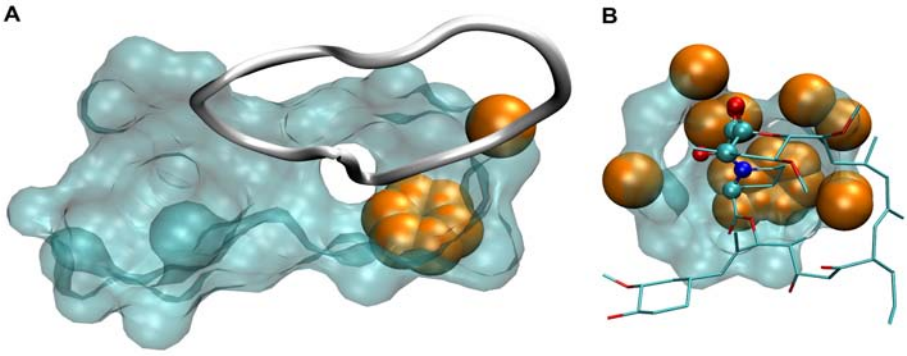

Fig. (4). Spatial configurations of binding pockets and active atoms. (**A**) The binding pocket of protein CypA and the backbone of ligand CsA are represented in transparent Surf model and Tube model, respectively. There are seven active atoms of CypA represented in VDW model (orange), PHE113: CG, CD1, CD2, CE1, CE2, CZ (these six atoms form a quincunx in the bottom of the pocket.) and PHE60: CE2; and four active atoms of CsA, MVA11: N, CA, C, O. The active residue MVA11 (the projecting part of the loop) of CsA clearly trends toward the active part of CypA. (**B**) The binding pocket of protein FKBP12 and ligand FK506 are represented in transparent Surf model and Lines model, respectively. There are nine active atoms of FK506 represented in CPK model and sixteen active atoms of FKBP12 represented in VDW model colored in orange (Table **4** and Table **S5**). The active atoms of FK506 (small colored balls) are embraced by the active atoms of FKBP12 (big orange balls). These pictures are generated by the program named VMD 1.8.6 [21].

It has been reported that CsA interacted with ARG55, PHE60, MET61, GLN63, GLY72, ALA101, ASN102, ALA103, GLN111, PHE113, TRP121, LEU122 and HIS126 of CypA [18]. The ligand binding pocket obtained by our method is number 19, which comprises ARG55, PHE60, MET61, GLN63, GLY72, THR73, GLY74, GLY75, LYS76, GLU81, LYS82, ALA101, ASN102, GLN111, PHE113, LEU122 and HIS126. They are in good agreement.

We can obtain the active residues of FKBP12 by the same way: TYR26, PHE36, PHE46, TRP59, PHE82 and PHE99 (compare Table **1** and Table **S4**), as well as the active atoms as shown in Table **S5**. Therefore, the pocket 13 is the ligand binding site of FKBP12.

X-ray experiment reported that the binding pocket of FKBP12 comprised the residues TYR26, PHE46, PHE99, VAL55, ILE56 and TRP59 [19]. The active residues we worked out are TYR26, PHE36, PHE46, TRP59, TRY82 and PHE99. They are also in good agreement.

The residue MVA11 of the ligand CsA is on the HOMO of CsA. Therefore, MVA11 is the active residue of CsA. Four atoms of residue MVA11 on the HOMO with the absolute value of its atomic coefficient larger than 0.1 are considered to be the active atoms of the CsA: MVA11:N, MVA11:CA, MVA11:C and MVA11:O. And the active atoms of FK506 are: C2, C8, C9, C10, C11, N7, O3, O4 and O6 (Table **4** and Fig. **4B**).

Fig. (**4**) shows the spatial configuration of the ligand binding pocket and the active atoms of CypA, CsA, FKBP12 and FK506. The active atoms of ligand are fitting so well with the active atoms of protein that indicates strong interactions between them (Fig. **4**). Especially, as NMR experiment pointed out [20], the CsA changed its configuration so that its active residue MVA11 (the projecting part of the loop as shown in Fig. **4A**) clearly trended toward the active part of CypA.

Peptide Inhibitor Ala-Gly-Pro

To verify the above two rules on the protein-ligand interaction, we designed a tripeptide Ala-Gly-Pro (AGP) to bind to CypA under guidance of the two rules (Please see the experimental section for more details about how and why AGP was selected). The calculation and biological experimental results showed perfect agreement with the two rules.

(1). Docking

The program Autodock [9] was employed to dock the AGP into the binding site of CypA. After docking, the complex system showed converged (Fig. **5**). The binding energy of docking was -7.25 Kcal/mol. And then energy minimization of 30ps was performed on the complex system. The structure of CypA/AGP complex remained stable and the AGP bound into the binding site perfectly as shown in Fig. (**6A**). Additionally, the hydrophobic parts of the ligand AGP were embraced by the hydrophobic residues of CypA and so did the non-hydrophobic part. It demonstrates that interactions between AGP and CypA are reasonable from a chemical view.

(2). Active Resides and Active Atoms

We recalculated the electronic structure of AGP and CypA using the coordinates obtained after energy minimization. The active residues and active atoms of the CypA and AGP and the ligand binding site were obtained as shown in Fig. (**6B**). The active residues of CypA are still the Phe60 and Phe113. There are two active residues (Pro, Gly) and seven active atoms of AGP, Pro: C, OT1, OT2, N, CA, HA and Gly: O. It has been reported through residue mutational experiments by Yoo *et al.* that residues Gly and Pro are the key determinants for CypA recognition [22]. Residues Pro and Gly are found to be active residues here and the active atoms localize mainly on the residue Pro, which is not only in excellent agreement with the previously findings. Furthermore, our work can identify the active atoms. The active atoms are located on right above the major active residue Phe113 of CypA (Fig. **6B**). The whole AGP fits into the binding pocket with perfect shape complementarity [23] as shown in Fig. (**6A**), which is also consistent with the previous experimental findings that the CypA active site is complementary to sequences containing

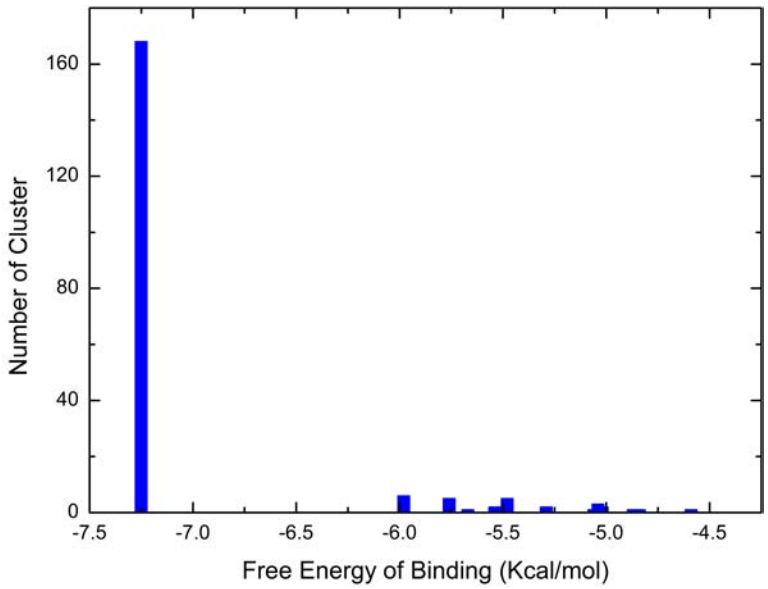

Fig. (5). The Conformers vs. the free energy of binding of peptide Ala-Gly-Pro. Peptide AGP is converged. The first cluster populates 168 conformers with the lowest binding energy of -7.25Kcal/mol. The convergence situation of all conformations is excellent.

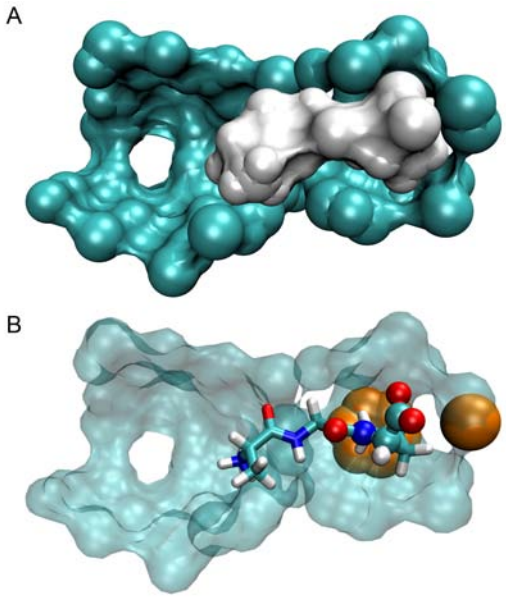

Fig. (6). Conformation of the CypA and tripeptide Ala-Gly-Pro (AGP) after energy minimization. (**A**) Both the ligand AGP (gray) and the binding site of CypA (bluish green) are represented in combined model of VDW and Surf. (**B**) The seven active atoms of AGP are represented in CPK model, Gly: C, Pro: N, CA, HA, C, OT1 and OT2, and the rest of AGP are represented in Licorice model. The seven active atoms of CypA (orange) are represented in VDW model. These pictures are generated by the program named VMD 1.8.6 [21].

the dipeptide Gly-trans-Pro [8]. The HOMO of the ligand AGP is -2.26 eV, whereas that of the CsA is -4.75 eV. The LUMO energy of the CypA is 2.00 eV (Phe113). The binding energy of docking AGP to CypA is -7.25 Kcal/mol. Theoretically, these results are excellent.

(3). Binding Affinity Determination and Inhibition of PPIase Activity of CypA

The binding affinity of the designed peptides to CypA was measured by the Surface Plasmon Resonance (SPR) with Biacore 3000 instrument (BiacoreAB Corporation, Uppsala, Sweden) as previously described [24, 25]. AGP was detected to bind to CypA in a concentration dependent manner with K_D value at 1.95×10^{-6} M (Fig. **7A**), while CsA, as a positive control, showed a K_D value at 6.42×10^{-6} M (Fig. **7B**).

Since AGP could bind to CypA, we sequentially measured its inhibition on the enzyme activity of CypA. The standard spectrophotometric method was applied to determine the inhibitory activity of the compounds on PPIase. During the assay, the rate constants for the *cis–trans* conversion were evaluated by fitting the data to the integrated first-order rate equation through nonlinear least-square analysis as follows: Inhibitory rate (%) = [CypA (dAbs/time) – compounds (dAbs/time)] / [CypA (dAbs/time) – control (dAbs/time)]. The inhibition results are shown in Fig. (**7**). As a positive control, CsA showed an inhibition of

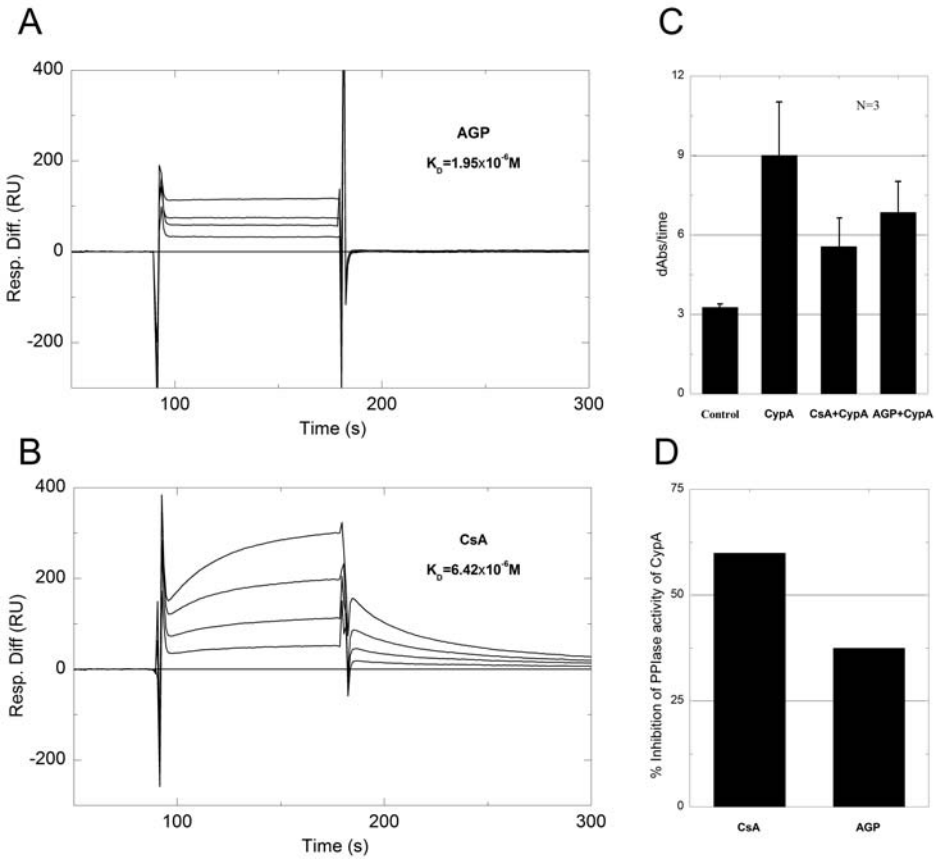

Fig. (7). Surface plasmon resonance analysis of the AGP/CypA and CsA/CypA interaction. Sensorgram for AGP and CsA binding to CypA surface on the CM5 sensor chip. Binding responses are shown for AGP and CsA injected at concentrations of 0.625, 1.25, 2.5, 5, and 10 μM (from bottom to top). The biosensor RUs are concentration dependent. The equilibrium constants (KD values) evaluating the protein-ligand binding affinities are denoted (**A**, **B**). CypA PPIase inhibitory activities of AGP and CsA at 1 μM. CypA was pre-incubated with 1 μM tested compounds, and the PPIase activity was evaluated by fitting the data to the integrated first-order rate equation through nonlinear least-square analysis (**C**, **D**). The value of dAbs/time represents the rate constant for the *cis–trans* conversion (**C**). The percent inhibition of the PPIase activity of AGP and CsA at 1 μM (**D**).

59.95% against the PPIase activity of CypA at 1μM, while AGP an inhibition of 37.47% at 1μM.

Our designed peptide AGP has the same order of binding affinity and inhibition of PPIase activity as CsA. Besides, AGP has no impact on cell proliferation and cell cycle (data not shown). Therefore, peptide AGP may be new inhibitor for the CypA.

CONCLUSIONS

Assuming we now have a new target protein, we can work out its LUMOs and pockets only from its coordinate file, and we can immediately obtain its ligand binding site, active residues and active atoms. If the HOMO energy of a small molecule is far away from the LUMOs energy of the target protein, we are sure it can't be a ligand.

These two rules enable us to have a clearer understanding of the mechanism of protein-ligand interaction on the basis of full electronic structure of a protein. The real reason which created these rules is due to the characteristics of energy level structure of a protein. They could be an important criterion for identifying or designing a ligand for a target protein. The first rule tells us the necessary condition for being the active residues or active atoms that they must locate on the LUMOs of a protein, and permits us to identify a ligand by checking its HOMO energy level. We can use the energy eigenvalues and orbital coefficients of the wave function to calculate many other characteristics of a protein. The second rule points out the necessary and sufficient condition for being the active residues/atoms and the ligand binding site. Therefore, it allows us to identify not only the ligand binding site, but also the active residues and active atoms of a protein. When detecting the active pocket, its volume and mouth-area can also help us identify or design a ligand.

As an application of our two rules, we predicted a tripeptide ligand AGP for the target receptor CypA. The peptide AGP was designed from scratch and is consistent with the previous experimental findings [8] indicating that the two rules we proposed can be criteria for guiding drug design. One may argue that the function of the designed peptide AGP is not surprising because this sequence (^{87}H***AGP***IA^{92}) has been reported to be essential for CypA recognition of the capsid domain of HIV-1 Gag polyprotein. However, another peptide Trp-Gly-Pro (WGP) was designed (after the current paper was finished) based on the rules [26], providing additional support for two rules. As demonstrated by virtual screening experiment, 20 tripeptide X-Gly-Pro and

20 tetrapeptide X-Ala-Gly-Pro were all docked to CypA. And three of the 40 peptides have lower binding energy than the positive control CsA. Of the three peptides, WGP was demonstrated to show comparable inhibitory ability as CsA, and is even able to inhibit HIV-1_{IIIB} infection. Therefore, with the two rules, we cannot only identify which residue dominates the interactions between a ligand and a protein, but also atoms which play key roles in the interactions.

We anticipate our method to be a starting point, which contributes to the study of the protein-ligand interaction and can assist the drug design on the basis of electronic structure, and provides us with a more detailed understanding of the mechanism of protein-ligand interaction.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

This work was supported by National Basic Research Program of China (no.2006CB504509), and the Project of the State Key Program of National Natural Science Foundation of China (Grant No. 10635060). The authors thank Ye Yuanjie, Wang Xun and Ye Ling for their kind help and the Modern Applied Mathematical Key Laboratory in Shanghai (the Department of Mathematics at Fudan University) and Shanghai Supercomputer Center (SSC) for providing the parallel computer.

SUPPORTING INFORMATION

Table **S1**, Table **S2**, Table **S3**, Table **S4** and Table **S5**.

REFERENCES

[1] Ladik, J.; Seel, M.; Otto, P.; Bakhshi, A.K. On the Electronic-structure and conduction properties of aperiodic DNA and proteins .1. Strategy and Methods of Investigations. *Chem. Phys.*, **1986**, *108*, 203-214.

[2] Bakhshi, A.K.; Otto, P.; Ladik, J.; Seel, M. On the Electronic-structure and conduction properties of aperiodic DNA and Proteins .2. electronic-structure of aperiodic DNA. *Chem. Phys.*, **1986**, *108*, 215-222.

[3] Otto, P.; Bakhshi, A.K.; Ladik, J.; Seel, M.; Chin, S. On the electronic-structure and conduction properties of aperiodic DNA and proteins .3. The Band Structures of Periodic Polypeptides. *Chem. Phys.*, **1986**, *108*, 223-231.

[4] Ye, Y.J. The extended negative factor counting method for tridiagonal block matrices with cross-links. *J. Math. Chem.*, **1993**, *14*, 121-139.

[5] Thompson, M.A.; Zerner, M.C. A Theoretical-examination of the electronic-structure and spectroscopy of the photosynthetic reaction center from rhodopseudomonas-viridis. *J. Am. Chem. Soc.*, **1991**, *113*, 8210-8215.

[6] Miyazawa, S.; Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, **1996**, *256*, 623-644.

[7] Miyazawa, S.; Jernigan, R.L. Estimation of effective interresidue contact energies from protein crystal-structures - quasi-chemical approximation. *Macromolecules*, **1985**, *18*, 534-552.

[8] Vajdos, F.E.; Yoo, S.H.; Houseweart, M.; Sundquist, W.I.; Hill, C.P. Crystal structure of cyclophilin A complexed with a binding site peptide from the HIV-1 capsid protein. *Protein Sci.*, **1997**, *6*, 2297-2307.

[9] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **1998**, *19*, 1639-1662.

[10] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.*, **2000**, *28*, 235-242.

[11] Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **2005**, *2*6, 1781-1802.

[12] Kleywegt, G.J.; Henrick, K.; Dodson, E.J.; van Aalten, D.M.F. Pound-wise but penny-foolish: How well do micromolecules fare in macromolecular refinement? *Structure*, **2003**, *11*, 1051-1059.

[13] Brunger, A.T.; Adams, P.D.; Clore, G.M.; DeLano, W.L.; Gros, P.; Grosse-Kunstleve, R.W.; Jiang, J.S.; Kuszewski, J.; Nilges, M.; Pannu, N.S.; Read, R.J.; Rice, L.M.; Simonson, T.; Warren, G.L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D*, **1998**, *54*, 905-921.

[14] Kofron, J.L.; Kuzmic, P.; Kishore, V.; Colonbonilla, E.; Rich, D.H. Determination of kinetic constants for peptidyl prolyl cis trans isomerases by an improved spectrophotometric assay. *Biochemistry (Mosc)*. **1991**, *30*, 6127-6134.

[15] Xu, L.N.; Zhang, L.L.; Gu, Z.; Zhou, L.X. Full electronic structure calculation of the biological activity in P173L enzyme. *Int. J. Quant. Chem.*, **2006**, *106*, 1544-1550.

[16] Zhou, L.X.; Shuai, J.W. *Protein computer simulation*. Xiamen Univ. Press, Xiamen, China, **2008**.

[17] Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic. Acids Res.*, **2006**, *34*, W116-W118.

[18] Ke, H.M.; Huai, Q. Crystal structure of cyclophilin and its partners. *Front. Biosci.*, **2004**, 9, 2285-2296.

[19] Vanduyne, G.D.; Standaert, R.F.; Karplus, P.A.; Schreiber, S.L.; Clardy, J. Atomic-structure of fkbp-fk506, an immunophilin-immunosuppressant complex. *Science*, **1991**, 252, 839-842.

[20] Theriault, Y.; Logan, T.M.; Meadows, R.; Yu, L.P.; Olejniczak, E.T.; Holzman, T.F.; Simmer, R.L.; Fesik, S.W. Solution structure of the cyclosporine-a cyclophilin complex by Nmr. *Nature*, **1993**, *361*, 88-91.

[21] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.*, **1996**, *14*, 33-&.

[22] Yoo, S.H.; Myszka, D.G.; Yeh, C.Y.; McMurray, M.; Hill, C.P.; Sundquist, W.I. Molecular recognition in the HIV-1 capsid/cyclophilin a complex. *J. Mol. Biol.*, **1997**, *269*, 780-795.

[23] Desjarlais, R.L.; Sheridan, R.P.; Seibel, G.L.; Dixon, J.S.; Kuntz, I.D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor-binding site of known 3-dimensional structure. *J. Med. Chem.*, **1988**, *31*, 722-729.

[24] Chen, S.A.; Zhao, X.M.; Tan, J.Z.; Lu, H.; Qi, Z.; Huang, Q.; Zeng, X.Z.; Zhang, M.J.; Jiang, S.B.; Jiang, H.L.; Yu, L. Structure-based identification of small molecule compounds targeting cell cyclophilin A with anti-HIV-1 activity. *Eur. J. Pharmacol.*, **2007**, *565*, 54-59.

[25] Thurmond, R.L.; Wadsworth, S.A.; Schafer, P.H.; Zivin, R.A.; Siekierka, J.J. Kinetics of small molecule inhibitor binding to p38 kinase. *Eur. J. Biochem.*, **2001**, *268*, 5747-5754.

[26] Pang, X.D.; Zhang, M.J.; Zhou, L.X.; Xie, F.; Lu, H.; He, W.; Jiang, S.B.; Yu, L.; Zhang, X.Y. Discovery of a potent peptidic cyclophilin A inhibitor Trp-Gly-Pro. *Eur. J. Med. Chem.*, **2011**, *46*, 1701-1705.

Received: March 13, 2012 Revised: September 09, 2012 Accepted: October 06, 2012