PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2010, 37(8): 904~915
www.pibb.ac.cn

Computational Study of Binding Mode for N-substituted Pyrrole Derivatives to HIV-1 gp41*

CONG Xiao-Jing**, TAN Jian-Jun**, LIU Ming, CHEN Wei-Zu, WANG Cun-Xin***
(*College of Life Science and Bioengineering, Beijing University of Technology, Beijing* 100124, China)

Abstract Molecular docking, molecular dynamics (MD) simulation and molecular mechanics Poisson-Boltzmann surface area (MM-PBSA)/molecular mechanics Generalized Born surface area (MM-GBSA) analysis are applied to predict the binding mode of two N-substituted pyrrole derivate inhibitors to the hydrophobic pocket in HIV-1 envelope protein gp41. Taking into account the flexibility of the receptor, multiple receptor conformations are used in docking with the ligands, which results in several possible binding modes. MD simulations and MM-PBSA binding energy calculations are performed on all the binding modes to identify the most favorable binding estimate. The MM-PBSA results indicate that the binding is mainly driven by non-polar interactions, while polar interactions determine the orientation of the ligands binding into the target site. Further analysis reveals the key residues and ligand-receptor interactions which contribute significantly to the binding affinity. This study provides useful information for rational design and optimization of N-substituted pyrrole derivatives as HIV-1 fusion inhibitors.

Key words HIV-1 fusion inhibitor, gp41, molecular docking, molecular dynamics simulation, MM-PBSA/MM-GBSA
DOI: 10.3724/SP.J.1206.2010.00110

Human immunodeficiency virus type 1 (HIV-1) fusion inhibitors have been recognized as a new generation of anti-HIV-1 drugs that block viral entry into target cells. HIV-1 envelope glycoprotein (Env) transmembrane subunit gp41 plays a key role in the early steps of virus-cell fusion process and may serve as an important target for developing HIV-1 fusion inhibitors[1]. During viral entry, gp41 adopts a transient conformation known as "prehairpin intermediate" in which a highly conserved therapeutic target, namely the N-helix trimer, is exposed. In each of the grooves on the surface of the N-helix trimer, there is a hydrophobic pocket that accommodates conserved hydrophobic residues in the gp41 C-terminal heptad repeat regions (C-helix) to form a stable six-helix bundle. The formation of the six-helix bundle is a crucial step in the fusion process. Compounds binding to the gp41 N-helix and blocking the formation of six-helix bundle have inhibitory activity on gp41 mediating viral-cell membrane fusion[2], suggesting that the hydrophobic pocket in the N-helix is an attractive target for designing new anti-HIV drugs[3].

The first member of this new class of anti-HIV drugs, T-20 (Enfuvirtide, approved by the US FDA in 2003), is a synthetic peptide of 36 amino acid based on the sequence of the C-helix of gp41 and has been used to treat HIV/AIDS patients who have failed to respond to reverse transcriptase inhibitors and protease inhibitors. T-20 is believed to interact with the gp41 N-helix and block the six-helix bundle formation, thereby inhibits membrane fusion[4]. However, lack of oral availability and high production cost have limited its utility in clinical treatment[5]. Lately there has been considerable interest in developing effective small molecular inhibitors of gp41. Using the hydrophobic pocket in the N-helix trimer as target, a number of small molecular fusion inhibitors have been identified[6]. Biological assays have confirmed that these small molecules, as well as some small molecules discovered

*Th is work was supported by grants from The National Natural Science Foundation of China (30670497), National Basic Research Program of China (2009CB930200), Beijing Natural Science Foundation (5072002, 7082006) and Research Fund for the Doctorate Program (X0015001200801).
**These authors contributed equally to this work.
***Corresponding author.
Tel: 86-10-67392724, E-mail: cxwang@bjut.edu.cn
Received: April 5, 2010 Accepted: May 25, 2010

in other ways[6-7], are able to block the formation of six-helix bundle thus blocking the viral fusion. In 2006, Frey *et al*.[6] synthesized a five-helix protein that resembles the gp41 six-helix core except for the absence of one of the three out-layer C-helices. This protein is designed so that the one of the hydrophobic pockets in the N-helix trimer is exposed as drug target. Screening assays using this target identified a set of compounds that interact with the gp41 inner core and significantly reduce the envelope-mediated membrane fusion. In 2004, Jiang *et al*[7] discovered two pyrrole derivatives, NB-2 and NB-64 (Figure 1), which can inhibit HIV-1 entry by interfering with the formation of the gp41 six-helix bundle at concentrations lower than 10 mg/L ($IC_{50} \approx 15$ μmol/L). These two inhibitors have simple "drug-like" structures with small molecular mass and can be used as leads for designing new HIV-1 fusion inhibitors. Experimental data indicates that the carboxylate group in the two inhibitors is essential for their inhibitory activity but detailed binding mechanism and ligand-protein interactions are unknown. In a recent modeling work, Teixeira *et al*.[8] performed 2D, 3D-QSAR analysis on 23 pyrrole derivatives(including NB-64) as HIV-1 gp41 inhibitors. The CoMFA model attained in this work characterizes the fundamental features of the inhibitors for their inhibitory activities. The authors used molecular docking to associate the ligand-derived model with gp41 hydrophobic pocket, nevertheless, the result doesn't provide specific details of individual inhibitors interacting with gp41. Moreover, sufficient evidence is necessary for validating the arbitrary molecular docking result proposed by the authors.

We have previously proved with *in vitro* assays that NB-2/NB-64 could interact with the gp41 five-helix protein thus inhibiting the C-helix derived peptide C34 [4] to form a six-helix bundle with the five-helix protein (data unpublished). In this study, we use a protocol combining multi-conformation docking, molecular dynamics (MD) simulation and molecular mechanics Poisson-Boltzmann surface area(MM-PBSA)/ molecular mechanics Generalized Born surface area (GBSA) methods to investigate the binding modes and molecular interactions between the gp41 five-helix protein and these two inhibitors. Multi-conformation docking is employed to take into account the flexibility of the protein. To identify the correct binding mode and reveal the binding mechanism, MD simulations are carried out for the possible binding modes attained by docking. The binding energies are calculated using the MM-PBSA method and further decomposed through MM-GBSA analysis. The results help to investigate intermolecular interactions upon binding and to identify the structural features of the ligands that are associated with their inhibitory activities. This study provides good guidance for designing or discovering new inhibitors targeting HIV-1 gp41. Based on the findings of this work, we have designed and synthesized a series of novel N-substituted pyrrole analogues as potential HIV-1 fusion inhibitors(Wang C X, *et al*. De novo design of HIV-1 gp41 inhibitors. Journal of Beijing University of Technology).

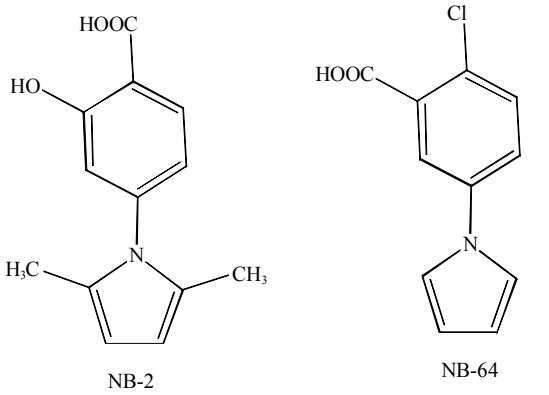

Fig. 1 Chemical structures of NB-2 and NB-64

1 Materials and methods

1.1 Preparation of the receptor and ligands

The molecular structures were prepared and minimized through a series of steps carried out in SYBYL 7.3[9]. The X-ray crystal structure of the gp41 core was retrieved from the Protein Data Bank (PDB code: 1aik [4]). The hydrophobic pocket in the gp41 N-helix [4] was used as the target site for molecular docking. Structural water and one of the C-helixes were removed from the six-helix bundle to expose the hydrophobic pocket. For the convenience of further analysis, the three N-helices were labeled as chains L, M, N, respectively and the remaining two C-helices, as chains D and E, respectively. The hydrophobic pocket is located between chain M and chain N. Hydrogen atoms were then added to the modified core structure (five helices) and optimized using Amber99 force field[10].

The structures of NB-2 and NB-64 were constructed in SYBYL 7.3. The carboxylic group in NB-2/NB-64 is set ionized (as expected in physiological

environment) and will be referred to as carboxylate group hereafter. The two carboxylate oxygen atoms are equivalent and the ligands both have a net charge of –1. The ligands were then assigned Gasteiger-Hückel charges and minimized with Tripos force field parameters through 1 000 steps of steepest descent minimization followed by 1 000 steps of conjugate gradient minimization.

1.2 MD simulation and sampling of the receptor

Given the particularly small size of NB-2/NB-64 as well as their special structural features, using the frozen receptor conformation for docking may fail to find the correct binding mode of these two ligands. We therefore performed a 2 ns MD simulation to explore the flexibility of the receptor. The MD simulation was carried out in AMBER 8.0 package [11] using Amber force field 03(ff03)[12]. A 10Å buffer of TIP3P[13] water molecules was added around the receptor. Na^+ ions were added as counterions to neutralize the system. Particle mesh Ewald (PME) method[14] was applied to treat long-range electrostatic interactions. The direct sum cutoff distance for long-range electrostatic and van der Waals interactions was set to 10.0Å. SHAKE algorithm [13] was applied for all the bonds involving hydrogen atoms. First, the water positions were minimized with the solute being restricted (restricting force=10 kcal/mol•Å^2). Then the solute was allowed to move and 2 750 steps of minimization (250 steps of steepest descent followed by 2 500 steps of conjugate gradient) were performed on the system. After gradually heating from 0 K to 300 K with the solute restricted, the 2 ns MD simulation without restriction was carried out at 300 K with a time step of 2.0 fs. Visualizing the MD trajectory in VMD, we find that the helical structure and the overall protein backbone are stable during the simulation time; whilst the side chains experience relatively large fluctuations. The root-mean-square deviations (RMSDs) of backbone atoms against the initial conformation are less than 1.5Å (Figure 2a). The RMSDs of residues which form the hydrophobic pocket have an average fluctuation of (1.7 ± 0.4)Å (Figure 2b). As observed in the MD trajectory, the shape of the pocket keeps changing in the first 140 ps then stays half-closed during the rest of the simulation. Clustering analysis was performed on all the structures in MD trajectory using the Jarvis Patrick method in Gromacs 3.0.1 [15]. The structures were first fitted against the initial conformation on backbone atoms; an RMSD matrix was then calculated for the side chains of residues which form the gp41 hydrophobic pocket. Using the RMSD matrix, the trajectory was clustered with an RMSD cutoff of 0.1Å. Middle structures of clusters which have more than one cluster member were selected as representative conformations of the receptor. As a result, five representative conformations are out of the first 140 ps MD trajectory and seven are out of the remainder. All twelve conformations as well as the initial receptor structure were used in the following multi-conformation molecular docking.

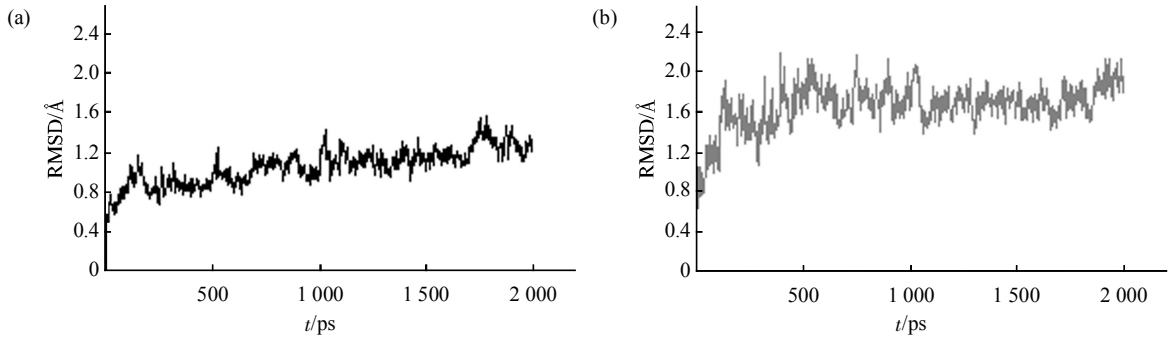

Fig. 2 RMSDs of (a) backbone of the receptor and (b) side-chain atoms of residues forming hydrophobic pocket against the initial receptor conformation with respect to the simulation time

1.3 Multi-conformation molecular docking

AutoDock 4[16] was used to carry out the docking in which the grid spacing was set to 0.375Å and each grid map consisted of 80×80×80 grid points in three dimensions. For each conformation of the receptor, grid center coordinates were set as the mean coordinates of atom CG1 in Val570 (chain M) and atom CH_2 in Trp571 (chain N). For ligands, all the single bonds outside the rings were set free to rotate in 20 degree increments. With the other options set as

default, 100 docking runs were performed during each docking experiment. The docking results were analyzed by cluster analysis. For each ligand, three binding modes which have much lower binding free energies and larger cluster populations than the rest modes were selected for further analysis.

1.4 MD simulations of the binding complexes

The binding modes attained from docking were used as initial complex conformations in MD simulations. Topology prep files for ligands were built with the general amber force field (gaff)[17] and RESP[18] charges by using Antechamber in AMBER 8.0. For each of the binding modes, 3 ns MD simulation was performed through the same procedure as described in section **2.2**. Snapshots were extracted from the trajectory at every 13 ps from 400 to 3 000 ps. Finally, 200 snapshots of each binding mode were used for the following binding energy calculation.

1.5 Binding energy calculations and energy decompositions

The binding energies were calculated by using molecular mechanics Poisson-Boltzmann/surface area (MM-PBSA) method[19] implemented in AMBER 8.0. MM-PBSA has been successfully used both to rank different ligands binding to a given site and to rank different binding modes of a single ligand[20]. In this method, the binding free energy(ΔG_{bind}) is evaluated as a sum of the changes in the molecular mechanics gas-phase binding energy (ΔE_{MM}), solvation free energy shift (ΔG_{solv}) and entropy contribution ($T\Delta S$):

$$\Delta G_{bind}=\Delta E_{bind}-T\Delta S, \quad (1)$$

$$\Delta E_{bind}=\Delta E_{MM}-\Delta G_{solv}, \quad (2)$$

$$\Delta E_{MM}=\Delta E_{intra}+\Delta E_{ele}+\Delta E_{vdw}, \quad (3)$$

$$\Delta G_{solv}=\Delta\Delta G_{PB}+\Delta\Delta G_{SA}, \quad (4)$$

Note that Eq. (1) is an approximation under the assumption of $\Delta H_{bind\ (gas)} \approx \Delta E_{MM}$ when the volume change during the binding process is negligible under constant temperature and pressure. The electrostatic solvation free energies used to evaluate $\Delta\Delta G_{PB}$ are calculated with the finite-difference solution to the Poisson-Boltzmann (PB) equation[21]. The non-polar part of the solvation energy ($\Delta\Delta G_{SA}$) is calculated using the function

$$\Delta G_{SA}=\gamma SA+b \quad (5)$$

where SA is the solvent accessible surface area[22] of the complex, receptor or ligand, determined with LOPO[23] method; the parameters γ and b, are set to 0.007 2 kcal/mol•Å^2 and 0.00 kcal/mol, respectively.

In this study, the single trajectory approach is applied to estimate the energies, which means the receptor and ligand geometries are taken from that of the complex, thus there is no internal energy (ΔE_{intra}) contribution to the net MM binding energy (ΔE_{MM}). Estimation of energies in this manner has been proven successful in many studies[24-25]. The separate trajectories approach in which the three trajectories of complex, free receptor and free ligand are used for energy calculation is deficient in practice due to sampling limitation and large fluctuations. For ligands binding to a same protein site, the entropy contributions are not remarkably different when the ligands are of similar structure. It has been shown in earlier studies that neglecting the entropy effects of a set of analogues binding to the same receptor results in good agreements between the calculated and experimental relative binding free energies[25-26]. Moreover, the normal-mode analysis, often used to estimate entropy changes, is computationally demanding and gives large errors for bio-molecular systems[27-28]. Therefore, the entropy contributions are omitted in this study since we are only interested in the relative order of binding affinities.

To investigate the energetic contributions of each residue to ligand binding in different binding modes, energy decomposition is performed using MM-GBSA approach wherein the solvation free energies are calculated with General Born (GB) model[29], and non-polar contributions to solvation are attained by the LCPO method[23]. The decomposition is carried out on a pairwise per-residue basis, which lists the binding energy items for each residue-ligand pair.

2 Results and discussion

2.1 Multi-conformation molecular docking results

The twelve conformations collected from the MD trajectory of the free receptor and the initial receptor conformation were used as receptor conformations in docking. In each docking experiment, the 100 docking runs were finally clustered and ranked according to the estimated free energy of binding. For each ligand, three binding modes which have much lower binding free energies and larger cluster populations than the other modes were attained for further analysis. The six binding modes of each mode are shown in Figure 3. As expected, none of the seven receptor structures in which the hydrophobic pocket is half-closed gives rational results for ligand binding.

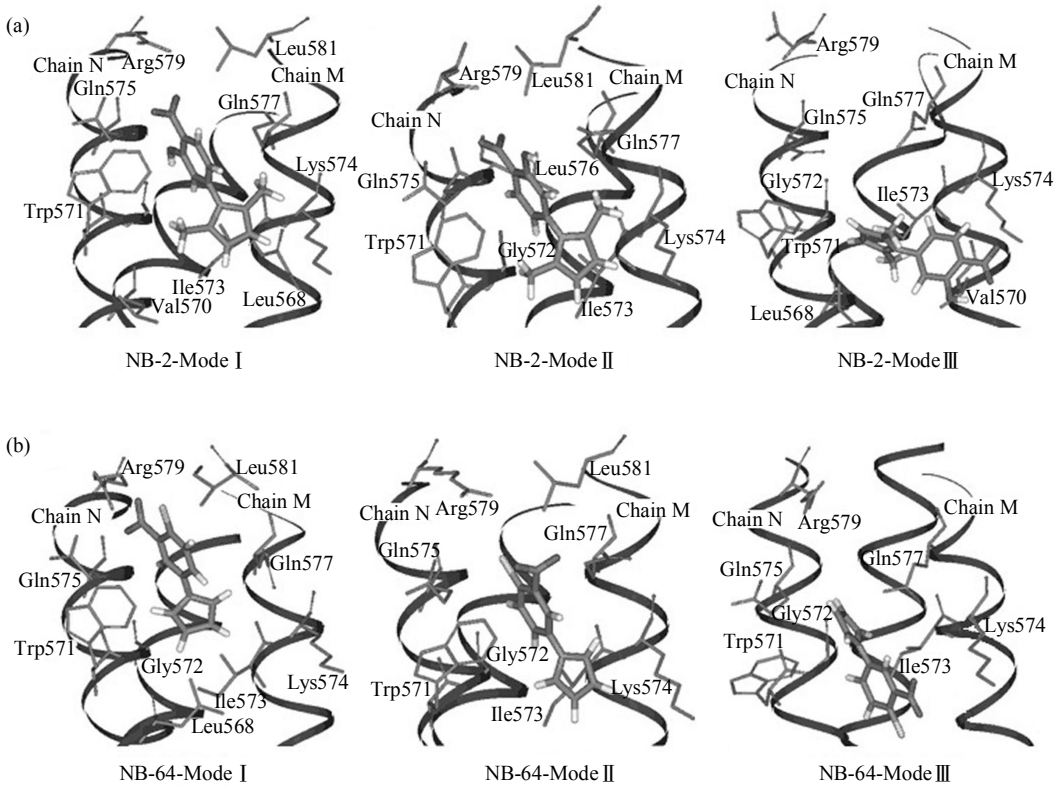

Fig. 3 Binding modes by docking NB-2/NB-64 into gp41 hydrophobic pocket
(a) Three binding modes for NB-2. (b) Three binding mode for NB-64.

From Figure 3a we can see that in Mode Ⅰ and Mode Ⅱ, NB-2 adopts similar orientations except for the positions of the hydroxyl group. In both modes, the whole NB-2 molecule fits very well into the hydrophobic pocket and the carboxylate group interacts with positively charged Arg579 (N). In Mode Ⅰ, the hydroxyl group is buried inside the pocket whereas in Mode Ⅱ, the hydroxyl group is placed in an opposite position. In Mode Ⅲ, the carboxylate group of NB-2 orients to the positively charged Lys574 (M); the pyrrole group occupies part of the hydropobic pocket while the phenyl group is partly outside of the pocket region.

In NB-64-Mode Ⅰ (Figure 3b), the carboxylate group interacts with Arg579(N) and the pyrrole group is deeply buried in the hydrophobic pocket. In NB-64-Mode Ⅱ, the chlorine atom points to Arg579(N) while the carboxylate group points out of the pocket. NB-64-Mode Ⅲ is similar to NB-2-Mode Ⅲ, with the carboxylate group orienting to Lys574 (M), but the ligand appears to have a better hydrophobic contact with the pocket due to the conformational change of the receptor.

2.2 MD trajectories analysis

2.2.1 RMSD analysis. During the 3ns MD simulations, energetic and structural properties were monitored for the six systems. The RMSDs and energies converged in all the six systems, which indicated well-behaved simulations. Figure 4 shows the RMSDs of ligand atoms and the backbone atoms in the receptor against initial complex conformations. As we can see, the receptor structures are relatively stable in all of the six simulation systems, while the ligands undergo different RMSD fluctuations. In NB-64-Mode Ⅲ, the large RMSD of ligand from 1 500 to 1 650 ps (Figure 4b) implies a remarkable change in position. We observed that the ligand experienced an inversion of orientation during this period of time. The carboxylate group, which previously interacted with Lys574 (M), transferred to the opposite side of the pocket. This indicates that the interactions between Lys574(M) and the carboxylate group are not strong enough to keep the ligand in position.

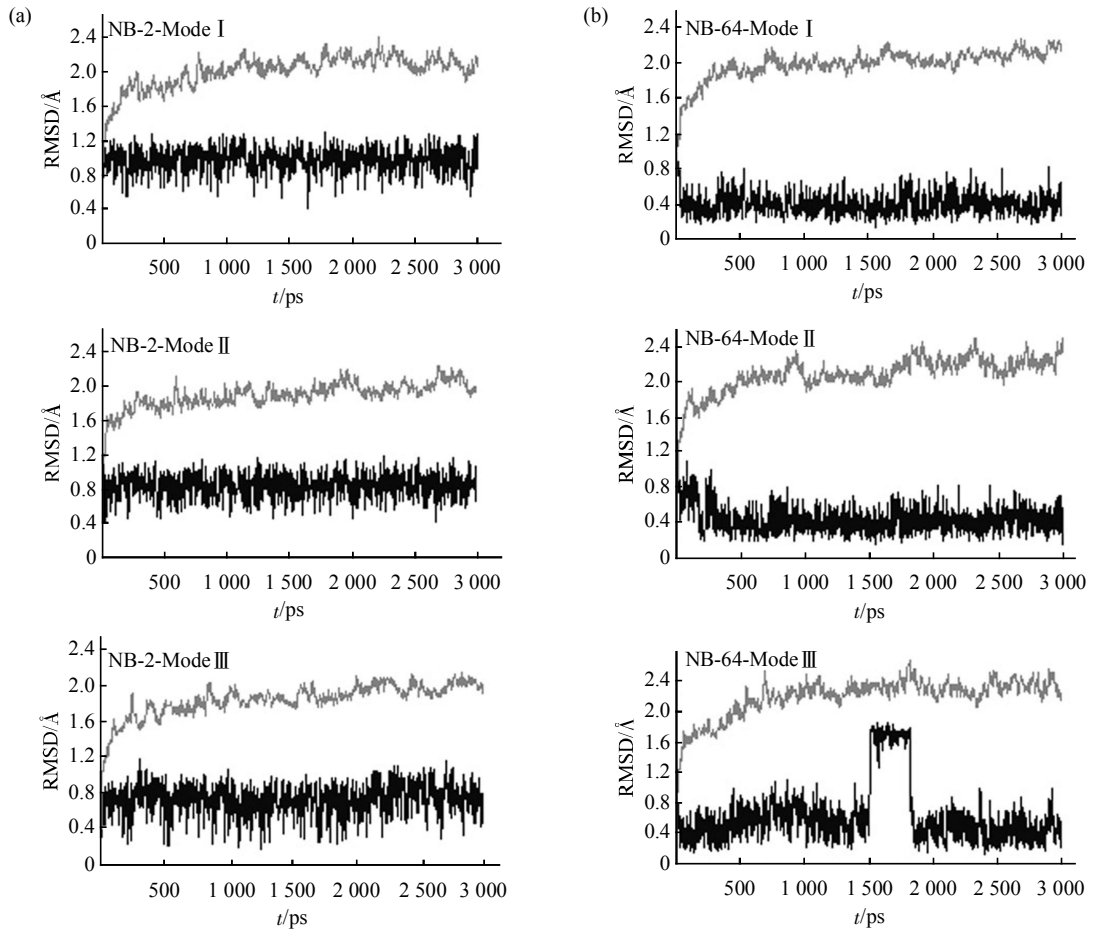

Fig. 4 RMSD of backbone of the receptor(grey) and the ligand(black) against the initial complex conformation of each binding mode

(a) Three binding modes for NB-2. (b) Three binding modes for NB-64.

2.2.2 Salt bridge analysis.

Salt bridge interaction is strong electrostatic interaction which stabilizes protein structure or protein-ligand complex. Salt bridge analysis provides useful information for investigating protein structure and protein-protein/protein-ligand interaction. The stability and energy of salt bridge interaction depend on the medium, thus the cutoff distance varies in different situations. Here, a cutoff distance of 3.2Å is used to estimate the salt bridges between the carboxylate group and charged residues in the receptor. The results are given in Table 1 as occupancies during the simulation time.

Among the three binding modes of NB-2, strong salt bridge interactions between carboxylate group and Arg579 were observed in Mode Ⅰ and Mode Ⅱ. In Mode Ⅰ, both the carboxylate oxygen atoms form salt bridges with Arg579, whereas in Mode Ⅱ, only one of the oxygen atoms forms a stable salt bridge with Arg579. In Mode Ⅲ, the carboxylate group has salt bridge interaction with Lys574 but is much less stable.

In NB-64-Mode Ⅰ, Arg579 forms a stable salt bridge with one of the carboxylate oxygen atoms and an unstable one with the other oxygen atom. In NB-64-Mode Ⅲ, weaker salt bridge interactions exist between the carboxylate group and Lys574. In NB-64-Mode Ⅱ, the carboxylate group is placed externally and interacts with water; correspondingly, no ligand-receptor salt bridge is observed in this mode.

Table 1 The occupancies of salt bridges between the carboxylate group and residues during the simulation time of different binding modes

Ligand	Mode	Residue: atom ID	Group: atom ID	Occupation/%	Distance/Å
NB-2	Ⅰ	Arg579 (N): NH2	Carboxylate: O2	62.50	2.88 ± 0.12
		Arg579 (N): NH1	Carboxylate: O1	45.80	3.00 ± 0.13
		Arg579 (N): NH2	Carboxylate: O1	30.20	2.83 ± 0.12
	Ⅱ	Arg579 (N): NH1	Carboxylate: O2	82.90	2.83 ± 0.12
		Arg579 (N): NH2	Carboxylate: O2	56.40	2.89 ± 0.13
	Ⅲ	Lys574 (M): NZ	Carboxylate: O2	26.20	2.85 ± 0.12
		Arg579 (N): NH1	Carboxylate: O2	68.40	2.92 ± 0.14
NB-64	Ⅰ	Arg579 (N): NH2	Carboxylate: O2	65.00	2.84 ± 0.13
		Arg579 (N): NH2	Carboxylate: O1	10.20	3.01 ± 0.12
	Ⅱ	–	–	–	–
	Ⅲ	Lys574 (M): NZ	Carboxylate: O2	35.60	2.84 ± 0.14
		Lys574 (M): NZ	Carboxylate: O1	3.00	3.01 ± 0.12

Cutoff distance for salt bridge is 3.2Å.

2.2.3 Hydrogen bonds analysis. Hydrogen bonds are essential in stabilizing ligand-receptor complexes. Table 2 lists the hydrogen bond occupancies and geometries in all the binding modes. In NB-2-Mode Ⅰ, the carboxylate group forms two hydrogen bonds with Arg579 (N) through the two oxygen atoms. In NB-2-Mode Ⅱ, only one of the two hydrogen bonds exists, with a similar stability. In NB-2-Mode Ⅲ, one hydrogen bond is formed between carboxylate group and Lys574(M) with a very shot lifetime. The hydroxyl group of NB-2 is a potential H-bond donor and acceptor, however, no hydrogen bond is formed between hydroxyl group and the receptor in any of the three binding modes. In NB-64-Mode Ⅰ, the carboxylate group also forms two hydrogen bonds with Arg579(N) but one of them is maintained in only 18.80% of the simulation time. In NB-64-Mode Ⅲ, hydrogen bonds between the carboxylate group and Lys574(M) exist for a short time. As mentioned above, the ligand had a large position deviation during the simulation and the carboxylate group couldn't interact well with Lys574(M). In NB-64-Mode Ⅱ, no hydrogen bond is formed, which is in accordance with the orientation of the carboxylate group in this mode.

Table 2 The occupancies of hydrogen bonds formed between the ligand and the receptor during the simulation time of different binding modes

Ligand	Mode	Donor Residue: atom ID	Acceptor Group: atom ID	Occupation/%	Distance/Å	Angle/(°)
NB-2	Ⅰ	Arg579 (N): NH2	Carboxylate: O2	66.40	2.91 ± 0.17	157.61 ± 11.16
		Arg579 (N): NH1	Carboxylate: O1	31.00	2.85 ± 0.15	156.74 ± 11.18
	Ⅱ	Arg579 (N): NH2	Carboxylate: O2	64.30	2.95 ± 0.20	144.86 ± 9.80
	Ⅲ	Lys574 (M): NZ	Carboxylate: O2	7.80	2.88 ± 0.17	155.79 ± 12.31
NB-64	Ⅰ	Arg579 (N): NH1	Carboxylate: O2	71.00	2.89 ± 0.20	151.83 ± 13.44
		Arg579 (N): NH2	Carboxylate: O1	18.80	3.17 ± 0.20	157.17 ± 10.08
	Ⅱ	–	–	–	–	–
	Ⅲ	Lys574 (M): NZ	Carboxylate: O2	20.20	2.90 ± 0.20	147.56 ± 15.16
		Lys574 (M): NZ	Carboxylate: O1	2.60	3.12 ± 0.22	144.56 ± 10.05

Hydrogen bond criteria are 3.5Å for donor-acceptor distance and 120.0° for donor-H-acceptor angle.

2.3 Binding energy calculations and energy decompositions

Binding energies were calculated for the snapshots collected from equilibrated trajectories of the simulation systems and the results are listed in Table 3. The mode that has the lowest binding energy

is proposed to be the most favorable binding mode. The most favorable binding mode of each ligand has a binding energy about 5 and 7 kcal/mol more than the second mode, respectively. The data in Table 3 shows that in all the binding modes, the non-polar interactions are the main driving forces for binding, wherein the van der Waals interactions make major contributions, implying they all have good hydrophobic contacts. In the two modes with relatively weak van der Waals interactions (NB-2-Mode Ⅲ and NB-64-Mode Ⅲ), the ligands cannot fit well in the hydrophobic pocket, as we have discussed in section **3.1**, which lead to weaker binding affinities. Non-polar solvation terms ($\Delta\Delta G_{SA}$), corresponding to the burial of solvent accessible surface area upon binding, contribute slightly to binding without noteworthy differences among different binding modes. We notice that the most favorable binding modes (NB-2-Mode Ⅰ and NB-64-Mode Ⅰ) have large negative gas-phase electrostatic energies (ΔE_{ele}) and are the only two modes that have favorable polar energies (ΔE_{polar}). This is mainly attributed to the negatively charged carboxylate group which forms strong salt bridges and hydrogen bonds with the residue Arg579(N).

Table 3 Binding energies of different binding modes for NB-2 and NB-64

Ligand	Mode	ΔE_{ele}	ΔE_{vdw}	ΔE_{MM}	$\Delta\Delta G_{PB}$	$\Delta\Delta G_{SA}$	ΔG_{solv}	ΔE_{polar} [1)]	$\Delta E_{non\text{-}polar}$ [2)]	ΔE_{bind}
NB-2	Ⅰ	−72.37	−20.45	−92.82	70.66	−3.90	66.77	−1.71	−24.35	−26.06
	Ⅱ	−44.23	−22.70	−66.93	49.77	−3.68	46.09	5.54	−26.38	−20.84
	Ⅲ	−5.04	−17.12	−22.16	11.52	−3.24	8.28	6.48	−20.36	−13.88
NB-64	Ⅰ	−45.26	−19.58	−64.84	43.74	−3.15	40.59	−1.52	−22.73	−24.25
	Ⅱ	−6.05	−23.93	−29.98	16.30	−3.47	12.83	10.25	−27.40	−17.15
	Ⅲ	14.86	−16.94	−2.08	−6.29	−2.97	−9.26	8.57	−19.91	−11.34

[1)] The polar ($\Delta E_{ele}+\Delta\Delta G_{PB}$) contributions; [2)] The non-polar ($\Delta E_{vdw}+\Delta\Delta G_{SA}$) contributions. All energies are averaged over 200 snapshots and are in kcal/mol.

Energy decomposition enables us to observe the energetic contributions of every residue in the receptor to ligand binding. Figure 5 illustrates the key residues for binding and their contributions in different binding modes. The remarkable difference in polar energy is mainly attributed to four residues: Lys574 (M), Gln577 (M), Gln575 (N) and Arg579 (N). Non-polar contributions come from the residues around the hydrophobic pocket, as shown in Figure 5c and d. The energy decomposition combined with binding energy calculation provides an insight into the receptor-ligand interactions in each binding mode.

Among the three binding modes of NB-2, Mode Ⅰ and Mode Ⅱ adopt the same orientation except for the location of hydroxyl groups. In Mode Ⅱ, NB-2 has a better hydrophobic interaction with the binding pocket than in Mode Ⅰ (implied by ΔE_{vdw} in Table 3), which leads to a favorable overall non-polar binding energy ($\Delta E_{non\text{-}polar}$). This difference, however, is slight when attributed to individual residues (Figure 5c). The gas-phase electrostatic interaction (ΔE_{ele}) of Mode Ⅱ is much weaker than that of Mode Ⅰ, although this effect is typically offset by the polar contribution of solvation free energy ($\Delta\Delta G_{PB}$). The positive ΔE_{polar} of Mode Ⅱ means the ligand has stronger polar interactions with the solvent (water molecules) than with the receptor. Thus, when the ligand transfers from the solvent to the binding pocket, the overall polar contribution is unfavorable. Figure 5a shows that the less favorable ΔE_{polar} of Mode Ⅱ arises from a weaker interaction between the ligand and Arg579 (N), which has been revealed by the previous salt bridge and hydrogen bond analysis. The positive ΔE_{polar} of NB-2-Mode Ⅲ implies unfavorable electrostatic energy for this binding mode, despite a considerable polar interaction between Lys574 (M) and the ligand (Figure 5a). In addition, the values of ΔE_{vdw} and $\Delta E_{non\text{-}polar}$ suggest weaker hydrophobic interactions in this mode.

In the case of NB-64, Mode Ⅰ is apparently superior with highly favorable polar interactions. As with NB-2-Mode Ⅰ, the polar contributions are mainly from Gln577 (M), Gln575 (N) and Arg579 (N), as we can see in Figure 5. In NB-64-Mode Ⅱ, the ligand has an excellent hydrophobic contact with the receptor; however, the large positive polar energy gives rise to a less favorable total binding energy. The polar

interactions between ligand and individual residues of NB-64-Mode Ⅱ are scarcely noticeable (Figure 5b), determined by the location of carboxylate group in this mode. NB-64-Mode Ⅲ is the only mode that has a positive electrostatic energy and the non-polar interaction is the weakest among all the modes. The gas-phase electrostatic contribution of Lys574 (M) is almost cancelled out by the corresponding polar energy of dissovation. These results suggest that this site is by no means a favorable site for NB-64 binding to gp41.

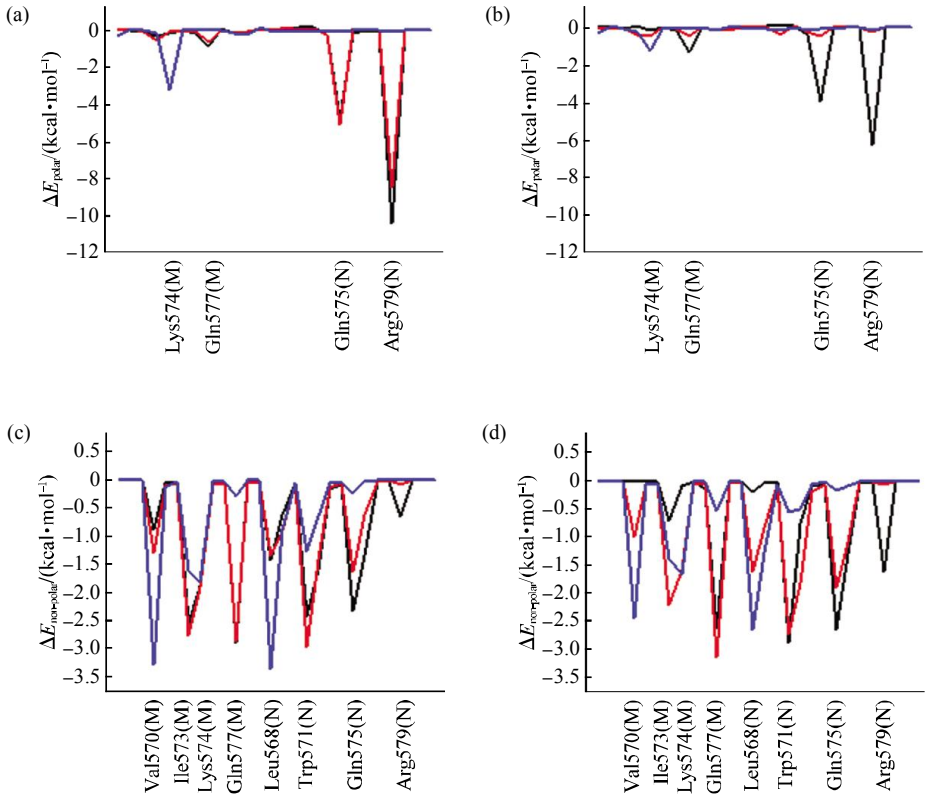

Fig. 5　Residues contributing distinctly to the binding affinity of different binding modes

(a) Polar energy of three binding modes of NB-2. (b) Polar energy of three binding modes of NB-64. (c) Non-polar energy of three binding modes of NB-2. (d) Non-polar energy of three binding modes of NB-64. —— : Mode Ⅰ; —— : Mode Ⅱ; —— : Mode Ⅲ.

The above analysis confirms that NB-2-Mode Ⅰ and NB-64-Mode Ⅰ are the favorable binding modes and may represent the real binding poses of NB-2/NB-64 to gp41. The two ligands bind to the receptor through very similar mechanics. Binding is mainly driven by non-polar interactions, especially the van der Waals interactions between the ligand and the residues that form the hydrophobic pocket. Gas-phase electrostatic energy is in favor of the binding but its contribution to the total binding energy is almost cancelled out by the large desolvation penalty. Jiang *et al*[7-8] propose that the negatively charged carboxylate group may orient either to Arg579(N) or to Lys574(M). In our result, the strong electrostatic interactions between the ligand and Arg579 (N), Gln575 (N), Lys574(M) indicate this is the favorite site for binding. Salt bridges and hydrogen bonds between the carboxylate group and Arg579 (N) account for the electrostatic interactions and are very important in stabilizing the ligand-receptor complex. This confirms a crucial role of the carboxylate group in binding, which has been suggested in prior experiments[7].

There are slight differences between the binding modes of the two ligands (Figure 6), due to their structural features. We can see in Figure 6a that the pyrrole ring with two methyl groups fits in parallel to the protein surface while the phenyl ring is placed perpendicularly in the narrow groove formed by Trp571 and Gln577. The pyrrole ring of NB-64, which is smaller in size, binds to the region between Trp571

and Gln577, ensuring the carboxylate group to have good interactions with Arg579. Conformational variation of the receptor can also be observed. In Figure 6b, the indolyl ring of Trp571 is approximately parallel to the phenyl ring of NB-64; a π-π stacking between these two rings may stabilize the binding complex. In the other binding modes, π-stacking interactions are not observed. However, the energetic contribution of Trp571 is almost the same in NB-64-Mode Ⅰ and Ⅱ (Figure 5b). In a recent evaluation of force fields descriptions for non-bonded interactions[30], the Amber force field gave noticeable errors in calculating stacking interaction energies. Therefore, the π-π stacking (if exists) in NB-64-Mode Ⅰ may have been underestimated, which would also lead to an underestimated relative binding affinity. The above result confirms that multi-conformation docking is necessary to reproduce the real binding behavior when a similar ligand-protein crystal structure is not available.

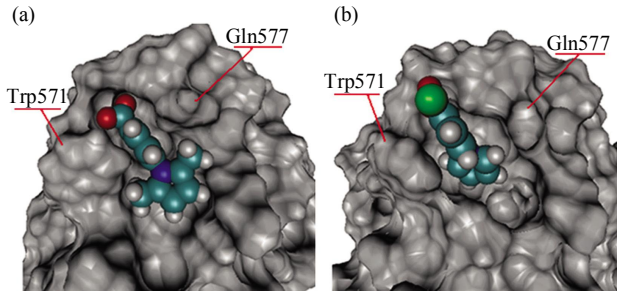

Fig. 6 Surface representation of the hydrophobic pocket in gp41 with the docked ligands
(a) NB-2 binding to gp41. (b) NB-64 binding to gp41.

3 Conclusion

By using multi-conformation docking, we took into account the receptor flexibility upon binding and attained several possible binding modes of NB-2/NB-64 to the gp41 hydrophobic pocket. 3 ns MD simulations were successfully applied to all of the binding complexes. Binding energy calculation and decomposition enable us to identify the favorable binding modes and important residues in the receptor. The two inhibitors have very similar binding orientations and ligand-receptor interactions. Non-polar interactions, especially van der Waals interactions are the main driving forces of binding. Polar interactions play an essential role in determining the ligand orientation, wherein the electrostatic interaction between Arg579 and the ligand makes the largest contribution. Salt bridges and hydrogen bonds are formed between Arg579 and the carboxylate group in ligand, which stabilizes the binding complex.

Liu *et al*.[31] has studied to assess gp41 interactions with biphenyl compound by using site-directed mutagenesis. In their study, each of the three amino acid residues that were suggested to interact with the compound in N36 was mutated to Ala one at a time. The results demonstratedance that interaction of the compound with N36 is highly dependent on gp41 residues Trp571 and Arg579. This show that our models are in accordance with their experiment[31]. At the same time, our results are highly in agreement with other simulation result[8, 32]. This means we can expect similar binding modes for the molecules of similar structures binding to gp41. It is therefore reasonable to use this predicted binding mode as a scaffold for designing new HIV-1 gp41 inhibitors.

Acknowledgement The authors thank Mr. Christopher C. Tait of Raffles Design Institute (Beijing) for editing assistance.

References

[1] Tan J J, Cong X J, Hu L M, *et al*. Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection. Drug Discov Today, 2010, **15**(5-6): 186-197

[2] Debnath A K, Radigan L, Jiang S B. Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. J Med Chem, 1999, **42**(17): 3203-3209

[3] Chan D C, Chutkowski C T, Kim P S. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. Proc Natl Acad Sci USA, 1998, **95**(26): 15613-15617

[4] Chan D C, Fass D, Berger J M, *et al*. Core structure of gp41 from the HIV envelope glycoprotein. Cell, 1997, **89**(2): 263-273

[5] Liu S W, Wu S G, Jiang S B. HIV entry inhibitors targeting gp41: From polypeptides to small-molecule compounds. Curr Pharm Des, 2007, **13**(2): 143-162

[6] Frey G, Rits-Volloch S, Zhang X Q, *et al*. Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. Proc Natl Acad Sci USA, 2006, **103**(38): 13938-13943

[7] Jiang S B, Lu H, Liu S W, *et al*. N-substituted pyrrole derivatives as novel human immunodeficiency virus type 1 entry inhibitors that interfere with the gp41 six-helix bundle formation and block virus fusion. Antimicrob Agents Chemother, 2004, **48**(11): 4349-4359

[8] Teixeira C, Barbault F, Rebehmed J, *et al*. Molecular modeling studies of N-substituted pyrrole derivatives-Potential HIV-1 gp41

inhibitors. Bioorg Med Chem, 2008, **16**(6): 3039-3048

[9] Fontenot D R, Den Hollander P, Vela E M, *et al*. Dynein light chain 1 peptide inhibits human immunodeficiency virus infection in eukaryotic cells. Biochem Biophys Res Commun, 2007, **363** (4): 901-907

[10] Cornell W D, Cieplak P, Bayly C I, *et al*. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. J Am Chem Soc, 1995, **117**(19): 5179-5197

[11] Case D A, Cheatham T E, Darden T, *et al*. The Amber biomolecular simulation programs. J Comput Chem, 2005, **26**(16): 1668-1688

[12] Duan Y, Wu C, Chowdhury S, *et al*. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem, 2003, **24**(16): 1999-2012

[13] Jorgensen W L, Chandrasekhar J, Madura J D, *et al*. Comparison of simple potential functions for simulating liquid water. J Chem Phys, 1983, **79**(2): 926-935

[14] Darden T, York D, Pedersen L. Particle mesh ewald-an N.log(N) method for ewald sums in large systems. J Chem Phys, 1993, **98**(12): 10089-10092

[15] Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model, 2001, **7**(8): 306-317

[16] Morris G M, Goodsell D S, Halliday R S, *et al*. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem, 1998, **19**(14): 1639-1662

[17] Wang J M, Wolf R M, Caldwell J W, *et al*. Development and testing of a general amber force field. J Comput Chem, 2004, **25**(9): 1157-1174

[18] Bayly C I, Cieplak P, Cornell W D, *et al*. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the resp model. J Phys Chem, 1993, **97**(40): 10269-10280

[19] Massova I, Kollman P A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov, 2000, **18**(1): 113-135

[20] Wang J M, Morin P, Wang W, *et al*. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc, 2001, **123**(22): 5221-5230

[21] Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science, 1995, **268**(5214): 1144-1149

[22] Sanner M F, Olson A J, Spehner J C. Reduced surface: An efficient way to compute molecular surfaces. Biopolymers, 1996, **38** (3): 305-320

[23] Weiser J, Shenkin P S, Still W C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comput Chem, 1999, **20**(2): 217-230

[24] Bao J, Zhang D W, Zhang J Z H, *et al*. Computational study of bindings of olive leaf extract (OLE) to HIV-1 fusion protein gp41. FEBS Lett, 2007, **581**(14): 2737-2742

[25] Yan C L, Xiu Z L, Li X H, *et al*. Comparative molecular dynamics simulations of histone deacetylase-like protein: Binding modes and free energy analysis to hydroxamic acid inhibitors. Proteins Struct Funct Bioinformat, 2008, **73**(1): 134-149

[26] Rafi S B, Cui G L, Song K, *et al*. Insight through molecular mechanics Poisson-Boltzmann surface area calculations into the binding affinity of triclosan and three analogues for FabI, the *E. coli* enoyl reductase. J Med Chem, 2006, **49**(15): 4574-4580

[27] Cheatham T E, Srinivasan J, Case D A, *et al*. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. J Biomol Struct Dyn, 1998, **16**(2): 265-280

[28] Kuhn B, Kollman P A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J Med Chem, 2000, **43**(20): 3786-3791

[29] Onufriev A, Bashford D, Case D A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins Struct Funct Bioinformat, 2004, **55**(2): 383-394

[30] Paton R S, Goodman J M. Hydrogen bonding and π-stacking: how reliable are force fields? a critical evaluation of force field descriptions of nonbonded interactions. J Chem Inf Model, 2009, **49**(49): 944-955

[31] Liu B, Joseph R W, Dorsey B D, *et al*. Structure-based design of substituted biphenyl ethylene ethers as ligands binding in the hydrophobic pocket of gp41 and blocking the helical bundle formation. Bioorg Med Chem Lett, 2009, **19**(19): 5693-5697

[32] Tan J J, Chen W Z, Wang C X. Investigating interactions between HIV-1 gp41 and inhibitors by molecular dynamics simulation and MM-PBSA/GBSA calculations. J Mol Struct Theochem, 2006, **766**(2-3): 77-82

HIV-1 跨膜蛋白 gp41 与 N-取代吡咯衍生物的结合模式研究 *

丛肖静 **　谭建军 **　刘　明　陈慰祖　王存新 ***

(北京工业大学生命科学与生物工程学院，北京 100124)

摘要　采用分子对接，分子动力学(MD)模拟和分子力学 / 泊松 - 波尔兹曼溶剂可有面积方法与分子力学 / 广义伯恩溶剂可及面积方法(MM-PBSA/MM-GBSA)，预测两种 N- 取代吡咯衍生物与 HIV-1 跨膜蛋白 gp41 疏水口袋的结合模式与作用机理. 分子对接采用多种受体构象，并从结果中选取几种可能的结合模式进行 MD 模拟，然后通过 MM-PBSA 计算结合能的方法识别最优的结合模式. MM-PBSA 计算结果表明，范德华相互作用是结合的主要驱动力，而极性相互作用决定了配体在结合过程中的取向. 进一步的结合能分解显示，配体的羧基与 gp41 残基 Arg579 的静电相互作用对结合有重要贡献. 上述工作为进一步优化 N- 取代吡咯衍生物类的 HIV-1 融合抑制剂建立了良好的理论基础.

关键词　HIV-1 融合抑制剂，跨膜蛋白 gp41，分子对接，分子动力学模拟，MM-PBSA/MM-GBSA

学科分类号　R914.3，O64　　**DOI**: 10.3724/SP.J.1206.2010.00110

* 国家自然科学基金(30670497), 国家重点基础研究发展计划(973)(2009CB930200), 北京市自然科学基金 (5072002, 7082006)和博士启动基金(X0015001200801)资助项目.

** 共同第一作者.

*** 通讯联系人.

Tel: 010-67392724, E-mail: cxwang@bjut.edu.cn

收稿日期：2010-04-05，接受日期：2010-05-25