

## Anti-SARS drug screening by molecular docking

D.-Q. Wei<sup>1,2,4</sup>, R. Zhang<sup>1</sup>, Q.-S. Du<sup>1,4</sup>, W.-N. Gao<sup>1</sup>, Y. Li<sup>1</sup>, H. Gao<sup>1</sup>, S.-Q. Wang<sup>1</sup>,  
X. Zhang<sup>1</sup>, A.-X. Li<sup>1</sup>, S. Sirois<sup>3,4</sup>, and K.-C. Chou<sup>1,4</sup>

<sup>1</sup> College of Life Science and Technology, Shanghai Jiaotong University, Shanghai, China

<sup>2</sup> Tianjin Institute of Bioinformatics and Drug Discovery (TIBDD), Tianjin Normal University, Tianjin, China

<sup>3</sup> Chemistry Department, Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

<sup>4</sup> Gordon Life Science Institute, San Diego, California, U.S.A.

Received September 5, 2005

Accepted February 1, 2006

Published online May 22, 2006; © Springer-Verlag 2006

**Summary.** Starting from a collection of 1386 druggable compounds obtained from the 3D pharmacophore search, we performed a similarity search to narrow down the scope of docking studies. The template molecule is KZ7088 (Chou et al., 2003, *Biochem Biophys Res Commun* 308: 148–151). The MDL MACCS keys were used to fingerprint the molecules. The Tanimoto coefficient is taken as the metric to compare fingerprints. If the similarity threshold was 0.8, a set of 50 unique hits and 103 conformers were retrieved as a result of similarity search. The AutoDock 3.011 was used to carry out molecular docking of 50 ligands to their macromolecular protein receptors. Three compounds, i.e., C<sub>28</sub>H<sub>34</sub>O<sub>4</sub>N<sub>7</sub>Cl, C<sub>21</sub>H<sub>36</sub>O<sub>5</sub>N<sub>6</sub>, and C<sub>21</sub>H<sub>36</sub>O<sub>5</sub>N<sub>6</sub>, were found that may be promising candidates for further investigation. The main feature shared by these three potential inhibitors as well as the information of the involved side chains of SARS CoV Mpro may provide useful insights for the development of potent inhibitors against SARS enzyme.

**Keywords:** SARS CoV Mpro – KZ7088 – Molecular docking – Similarity search – Inhibitor design

### Introduction

The outbreak of severe acute respiratory syndrome (SARS) creates need of anti-SARS drug (Drosten et al., 2003; Ksiazek et al., 2003; Shortridge, 2003; Miura et al., 2004). A new coronavirus (SARS CoV), a positive-sense, single-stranded RNA virus (Chen et al., 2003; Gao et al., 2003; Du et al., 2005a, b), has been found as the etiological agent of the disease SARS. SARS CoV main proteinase (Mpro) cleaves the polyprotein at no less than 11 conserved sites, a process initiated by the enzyme's own autolytic cleavage from pp1a and pp1ab (Anand et al., 2003; Yang et al., 2003). The released polypeptides mediate all of the functions required for SARS viral replication and transcription. The functional importance of the Mpro

in the viral life cycle makes it an attractive target for the drugs design against SARS (Anand et al., 2003; Yang et al., 2003; Chou et al., 2003, 2004).

Anand et al. (2003) proposed a homology model of the Mpro based on the experimental Mpro structures of human coronavirus (HCoV) and porcine transmissible gastroenteritis virus (TGEV) complex. The experimental structure of SARS CoV Mpro was first determined by Yang et al. (2003). They provided several crystal structures of SARS CoV Mpro at different pH conditions and the complex structure of SARS CoV Mpro with ligand hexapeptidyl CMK (Anand et al., 2003; Yang et al., 2003). The experimental structure and the theoretical homology structure of SARS CoV Mpro share many structural similarities, such as the catalytic active region and the Cys-His cleavage dyad (Anand et al., 2003; Yang et al., 2003). However, there are several differences between the theoretical and the experimental structures. For examples, the experimental structure of SARS CoV Mpro is a dimer, the protomer A has catalytic activity, and the protomer B has no activity. However, the N finger of protomer B, which plugs in the active cleft of protomer A, plays an important role in both dimerization and maintenance of the active form of the enzyme (Yang et al., 2003). These new findings from the experimental structure of SARS CoV Mpro provide a reliable basis for inhibitor design. Meanwhile, it was suggested (Anand et al., 2003) that suggested that the rhinovirus 3Cpro inhibitor AG7088 could well serve as a starting point for an anti-SARS drug based on the theoretical homology model of SARS CoV Mpro.

Chou et al. (2003) and Chou (2004) found the fitting problem of AG7088 to the binding pocket of SARS CoV Mpro, and they suggested its derivative KZ7088 as a better starting point. By conducting the virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points, Sirois et al. (2004, 2005) found 7083 unique hits without volume constraints and 2589 unique hits with volume constraints. Virtual screening has the advantages of being less expensive and easier to perform than the real experiments in searching for lead compounds; i.e., calculations could be performed on compounds that are not yet purchased or synthesized (Rayer, 2002). Therefore, it was widely used to find initial lead structures from a large collection of compounds. Besides the studies of Sirois et al. (2004), other groups also reported work of screening the inhibitor of SARS (Xiong, 2003). In the present study, by means of systematic search and molecular docking, we filtered the 2,589 candidates retrieved by Sirois et al. (2004), and found some interesting results as reported below.

## Materials and methods

### *Similarity search*

From the 2589 unique hits found by Sirois et al. (2004), we obtained a collection of 1386 druggable compounds that both match the 3D KZ7088 pharmacophore points and satisfy the druggable rules with a scoring value  $\geq 0.8$ . To further narrow down the investigation scope for molecular docking, a similarity search of the 1,386 compounds were performed by using KZ7088 (Chou et al., 2003; Chou, 2004) as the template molecule. The compounds thus obtained were placed in a database. The fingerprints (Sheridan et al., 1996; Brown and Martin, 1996) of the temperate molecule and each of these compounds in the database were calculated. The MDL MACCS keys were used to fingerprint the molecules. The MACCS fingerprint consists of a set of indicators showing whether each of the 166 MACCS keys was found to be present in a given molecule. The fingerprint is stored internally as a vector of indices, where the presence of an index in the vector indicates the presence of the corresponding substructure in the molecule. Once a fingerprint is derived from a chemical structure, a metric is needed to compare the fingerprints. A common metric is the Tanimoto coefficient (Morris et al., 1988), which is a number between 0 and 1, where 0 means "maximally dissimilar" and 1 means "maximally similar". If the similarity threshold was assigned as 0.8, a set of 50 unique hits and 103 conformers were retrieved as an outcome of the similarity search.

### *Molecular docking*

AutoDock 3.011 is a suitable software for performing automated docking of ligands to their macromolecular protein receptors. The individual components of the program include AutoTors, AutoGrid, and AutoDock. AutoTors defines which bonds in the ligand are rotatable, affecting the degrees of freedom (DOF) of the ligand and hence the complexity of computations. AutoGrid pre-calculates a three-dimensional grid of interaction energies based on the macromolecular target using the AMBER force field. AutoDock performs the task of the docking. First, the ligand moves randomly in any one of six degrees of freedom, namely 3 translation degrees and 3 rotation degrees (Chou et al., 1982; Chou and Scheraga, 1982), and the energy of the new ligand "state" is calculated. If the energy of the new state is lower than

that of the old state, the new one is automatically accepted as the next step in docking as the procedure used in (Chou and Carlacci, 1991; Chou, 1992).

### Preparation of ligands and protein

Since ligands are not peptides, Gasteiger charge was assigned and then non-polar hydrogens were merged. The rigid roots of each ligands were defined automatically instead of picking manually. The amide bonds were made non-rotatable. The experimental structure of SARS CoV Mpro was extracted from the RCSB Protein Data Bank whose PDB code is 1UJ1. The Mpro is a dimer where the protomer A has the catalytic activity while the protomer B has not. Accordingly, protomer B was removed. The Kollman charges were added to each atoms of the remained protomer A (Yang et al., 2003).

### Grid generation

N-H, O-H and S-H hydrogen bonds were modeled by choosing the proper 12-10 Lennard-Jones parameters. The grid box was centered in the catalytic active region between domain I and II of the Mpro including 23 amino acid residues according to (Gao et al., 2003): Cys-22, Gly-23, Thr-24, Thr-25, Leu-27, His-41, Val-42, Cys-44, Thr-45, Ala-46, Glu-47, Asp-48, Met-49, Leu-50, Asn-51, Pro-52, Tyr-54, Phe-140, Leu-141, Asn-142, Gly-143, Ser-144, Cys-145, Glu-163, His-164, Met-165, Glu-166, His-172, Asp-187, Arg-188, and Gln-189. The spacing between grid points was 0.375 angstroms.

### Docking

The GA-LS search algorithm (Lamarckian genetic algorithm) was chosen to search for the best conformers. During the docking process, a maximum of 50 conformers was considered for each compound (the default is 10 conformers). The initial position of each ligand was on the center of the box and oriented randomly. The parameters were set using the software ADT (Autodock Tool Kit) on PC which is associated with Autodock 3.0.3. The calculations of autogrid and autodock were performed on Sun Enterprise 10,000 Server with 16 CPUs and 64GB memory.

### Screening criteria

There are two kinds of free energies output by Autodock. One is the predicted binding free energy that includes the intermolecular energy and torsional free energy, and the other the docking energy that includes the intermolecular and intramolecular energies. The former is only reported at the end of a docking operation while the latter is used for selecting better individuals of population during a docking run (Morris et al., 1988). We used both energies, respectively, as the criterion for ranking. Moreover, there are some other factors that need to be examined for checking the screening results, such as the ligand's location, hydrophobic effects, steric complementarity, and size of the ligand. Therefore, after ranking the 20-top conformers according to their low energies, the ligands were further carefully checked according to the aforementioned factors as well. Furthermore, the intermolecular hydrogen bonds and electrostatic interaction, whose effects have already been counted in the binding energy, were also investigated in order to find useful information for drug design.

## Results and discussion

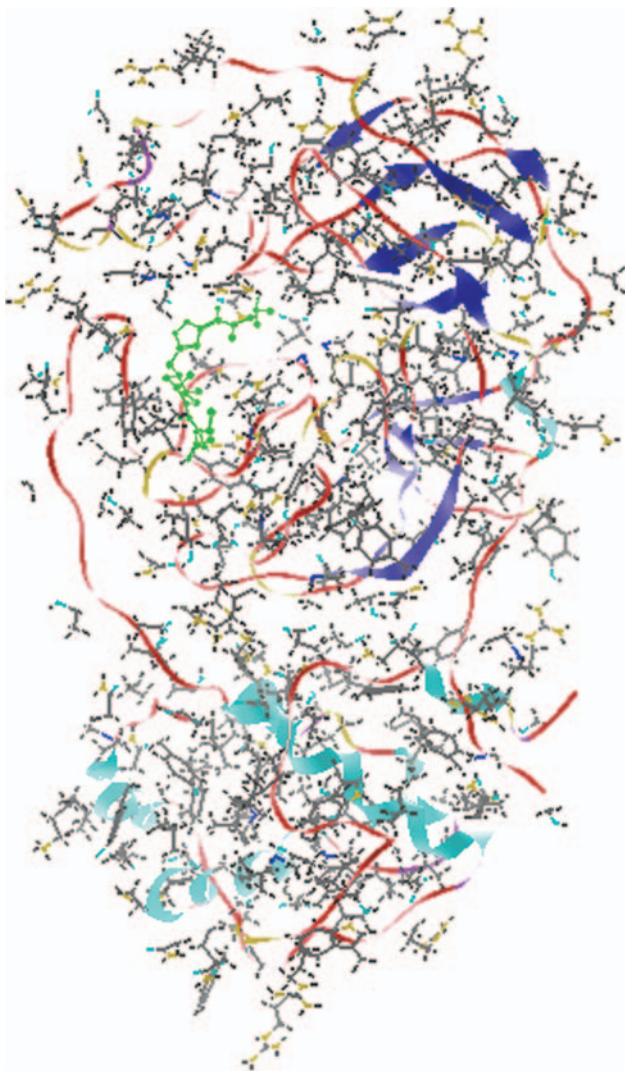
As shown in Rank 1 of Table 1, the docking results are ranked according to the ascent of the docking energies of the 50 conformers for each of the ligands investigated. It was found by closely viewing the top-20 compounds

**Table 1.** Docking results ranked by different types of energies in which names of the ligands are denoted by the sequence numbers of search results. Rank 1: rank by docked energy; Rank 2: rank by binding energy

Name	Formula	Lowest docking free energy (kcal/mol)	Lowest binding free energy (kcal/mol)	Rank 1	Rank 2
1	C <sub>30</sub> H <sub>39</sub> O <sub>7</sub> N <sub>4</sub> F	-16.97	-11.40	18	6
2	C <sub>18</sub> H <sub>17</sub> O <sub>6</sub> N <sub>2</sub> Cl	-7.12	-5.02	42	27
3	C <sub>22</sub> H <sub>29</sub> O <sub>4</sub> N <sub>5</sub>	-13.83	-8.43	29	17
4	C <sub>28</sub> H <sub>34</sub> O <sub>4</sub> N <sub>7</sub> Cl	-16.81	-12.13	19	5
5	C <sub>28</sub> H <sub>37</sub> O <sub>5</sub> N <sub>7</sub>	-13.87	-9.33	28	13
6	C <sub>25</sub> H <sub>38</sub> O <sub>4</sub> N <sub>8</sub>	-13.74	-8.69	30	16
7	C <sub>21</sub> H <sub>36</sub> O <sub>5</sub> N <sub>6</sub>	-22.64	-18.62	8	1
8	C <sub>28</sub> H <sub>34</sub> O <sub>6</sub> N <sub>7</sub> Cl	-12.23	-4.51	35	33
9	C <sub>33</sub> H <sub>42</sub> O <sub>8</sub> N <sub>6</sub>	-15.96	-8.31	22	18
10	C <sub>35</sub> H <sub>42</sub> O <sub>8</sub> N <sub>6</sub>	-16.75	-9.86	20	12
11	C <sub>31</sub> H <sub>37</sub> O <sub>4</sub> N <sub>8</sub> Cl	-16.43	-9.96	21	11
12	C <sub>21</sub> H <sub>36</sub> O <sub>5</sub> N <sub>6</sub>	-23.05	-18.30	7	2
13	C <sub>31</sub> H <sub>43</sub> O <sub>5</sub> N <sub>9</sub>	-18.08	-10.12	14	9
14	C <sub>28</sub> H <sub>34</sub> O <sub>6</sub> N <sub>7</sub> Cl	-15.28	-7.64	26	20
15	C <sub>28</sub> H <sub>40</sub> O <sub>7</sub> N <sub>8</sub>	-15.52	-8.75	25	15
16	C <sub>34</sub> H <sub>38</sub> O <sub>7</sub> N <sub>4</sub>	-11.85	-5.94	37	23
17	C <sub>30</sub> H <sub>43</sub> O <sub>8</sub> N <sub>7</sub>	-12.61	-4.81	33	28
18	C <sub>95</sub> H <sub>110</sub> O <sub>44</sub> N <sub>8</sub>	-17.91	-7.82	15	19
19	C <sub>31</sub> H <sub>42</sub> O <sub>7</sub> N <sub>4</sub>	-15.81	-10.20	24	8
20	C <sub>27</sub> H <sub>32</sub> O <sub>6</sub> N <sub>4</sub> F <sub>4</sub>	-5.99	-6.52	45	21
21	C <sub>70</sub> H <sub>93</sub> O <sub>13</sub> N <sub>18</sub>	-23.37	-4.10	6	37
22	C <sub>78</sub> H <sub>134</sub> O <sub>18</sub> N <sub>24</sub>	-27.13	-3.09	1	39
23	C <sub>56</sub> H <sub>75</sub> O <sub>13</sub> N <sub>13</sub>	-17.72	-4.77	16	29
24	C <sub>52</sub> H <sub>86</sub> O <sub>17</sub> N <sub>12</sub>	-22.42	-5.35	9	25
25	C <sub>76</sub> H <sub>118</sub> O <sub>23</sub> N <sub>24</sub>	-12.28	10.70	34	43
26	C <sub>61</sub> H <sub>88</sub> O <sub>14</sub> N <sub>18</sub>	-17.58	0.89	17	41
27	C <sub>73</sub> H <sub>95</sub> O <sub>13</sub> N <sub>18</sub>	-27.08	-5.47	2	24
28	C <sub>62</sub> H <sub>81</sub> O <sub>11</sub> N <sub>12</sub>	-5.80	25.87	46	50
29	C <sub>64</sub> H <sub>83</sub> O <sub>12</sub> N <sub>17</sub>	-21.97	-4.28	10	36
30	C <sub>55</sub> H <sub>71</sub> O <sub>10</sub> N <sub>11</sub> S <sub>2</sub>	-24.07	-10.80	5	7
31	C <sub>58</sub> H <sub>78</sub> O <sub>16</sub> N <sub>12</sub>	-15.83	-9.29	23	14
32	C <sub>63</sub> H <sub>87</sub> O <sub>19</sub> N <sub>13</sub>	-13.31	-5.13	31	26
33	C <sub>82</sub> H <sub>120</sub> O <sub>11</sub> N <sub>12</sub>	-13.91	-1.04	27	40
34	C <sub>12</sub> H <sub>9</sub> O <sub>5</sub> N <sub>2</sub> Br	-5.68	-4.30	48	35
35	C <sub>22</sub> H <sub>25</sub> O <sub>6</sub> N <sub>3</sub>	-18.29	-14.66	13	3
36	C <sub>11</sub> H <sub>19</sub> O <sub>7</sub> N <sub>2</sub> P	-8.68	-5.98	39	22
37	C <sub>57</sub> H <sub>81</sub> O <sub>18</sub> N <sub>12</sub> F <sub>3</sub>	-19.80	-4.54	12	32
38	C <sub>55</sub> H <sub>82</sub> O <sub>17</sub> N <sub>12</sub>	-25.30	-12.41	4	4
39	C <sub>61</sub> H <sub>88</sub> O <sub>22</sub> N <sub>18</sub>	-10.09	10.56	38	42
40	C <sub>94</sub> H <sub>148</sub> O <sub>31</sub> N <sub>32</sub>	-6.60	24.46	43	48
41	C <sub>70</sub> H <sub>99</sub> O <sub>17</sub> N <sub>25</sub>	-26.19	-4.40	3	34
42	C <sub>89</sub> H <sub>135</sub> O <sub>25</sub> N <sub>25</sub>	-11.99	15.95	36	46
43	C <sub>75</sub> H <sub>102</sub> O <sub>27</sub> N <sub>20</sub>	-5.77	24.87	47	49
44	C <sub>66</sub> H <sub>93</sub> O <sub>25</sub> N <sub>19</sub>	-4.56	20.57	49	47
45	C <sub>50</sub> H <sub>80</sub> O <sub>14</sub> N <sub>14</sub> S	-19.91	-4.76	11	30
46	C <sub>79</sub> H <sub>114</sub> O <sub>24</sub> N <sub>18</sub> S	-7.86	15.63	40	45
47	C <sub>143</sub> H <sub>230</sub> O <sub>37</sub> N <sub>42</sub> S <sub>7</sub>	-0.96	12.50	50	44
48	C <sub>20</sub> H <sub>18</sub> O <sub>5</sub> N <sub>3</sub> Br	-13.02	-10.12	32	10
49	C <sub>23</sub> H <sub>27</sub> O <sub>9</sub> N <sub>2</sub> F	-7.67	-3.82	41	38
50	C <sub>12</sub> H <sub>8</sub> O <sub>5</sub> N <sub>2</sub> Cl <sub>2</sub>	-6.23	-4.76	44	31

docked to Mpro that some conformers are outside the active region concerned, and some even far away from the pocket. When checking some other low docking energy conformers of the same ligands, such as ligand

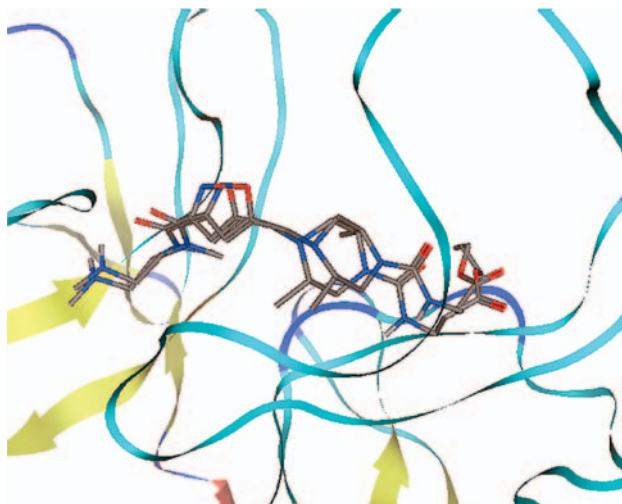
30, 36, 62, 29, 33, 38, we found that many of them are not buried in the protein but exposed outside. These molecules are large in size and cannot stay comfortably in the incommensurable pocket.



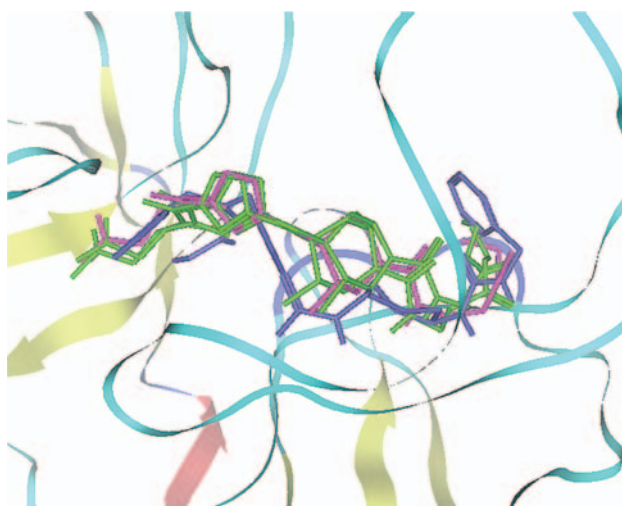
**Fig. 1.** An overall view of the complex obtained by docking ligand 7 to SARS CoV Mpro, where the backbone of the SARS enzyme is shown in ribbon drawing, its side chains in line drawing colored by atom type, the ligand in ball-and-stick drawing colored in green

In contrast, if ranking the docking results according to the binding free energy which includes the torsional term as shown in Rank 2 of Table 1, it was found that most of the top-20 ligands interacted quite well with the receptor in the pocket. These top 20 ligands in Rank 2 are smaller in size than the top 20 ligands in Rank 1. An example is given in Fig. 1 to show the complex of docking ligand 7 to SARS Mpro.

A comparison of the results between Ranks 1 and 2 suggests that the binding free energy is more reliable as a criterion for the virtual screening via molecular docking. It seems that the torsional term, i.e. the unfavourable entropy of ligand binding, plays an important role in ligand-receptor interaction. Especially for some large-sized li-



**Fig. 2.** A plot of docked conformations of ligand 7 in active site where the backbone of protein is shown in flat ribbon. The two conformations of ligand 7 are in stick drawing in which carbon atoms are colored in dark grey, polar hydrogen in white, oxygen in red and nitrogen in blue

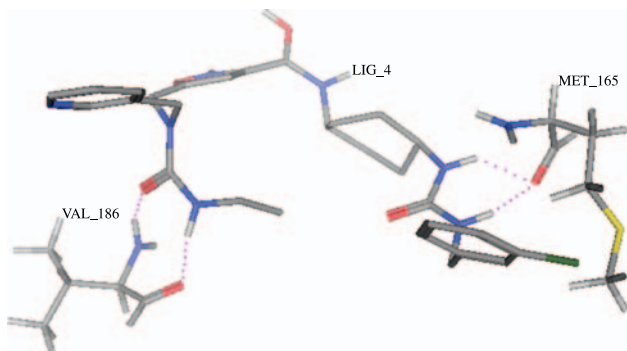


**Fig. 3.** A plot of different conformations with low energies extracted from 3 different ligands (7, 12, 4) which ranked high in Table 1. The marks in this plot are the same with the ones in Fig. 2 except that ligand 12 is rendered in purple and ligand 4 in blue. Their differences increase at the right ends of the molecules, which might be caused by the fact interactions are less specific

gands which suffer from more loss of torsional freedom upon binding. If the volume of the ligand does not allow it get into the pocket, the corresponding interaction energy would be meaningless.

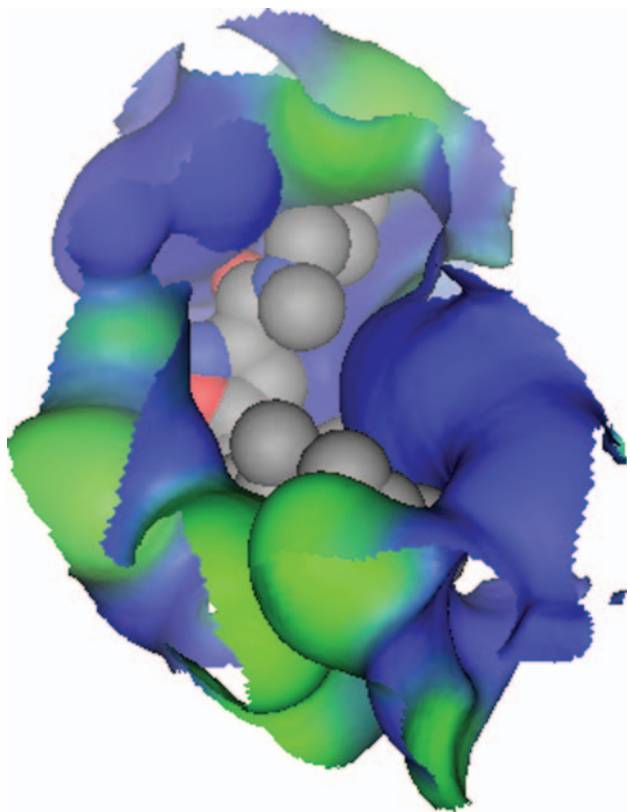
The recurrence of identical conformations of one ligand means that it fits very well to the pocket and is likely to be a good inhibitor candidate. In the output dlg file of a docking run, clusters of conformations are sorted out according to their RMDS values. We found a cluster



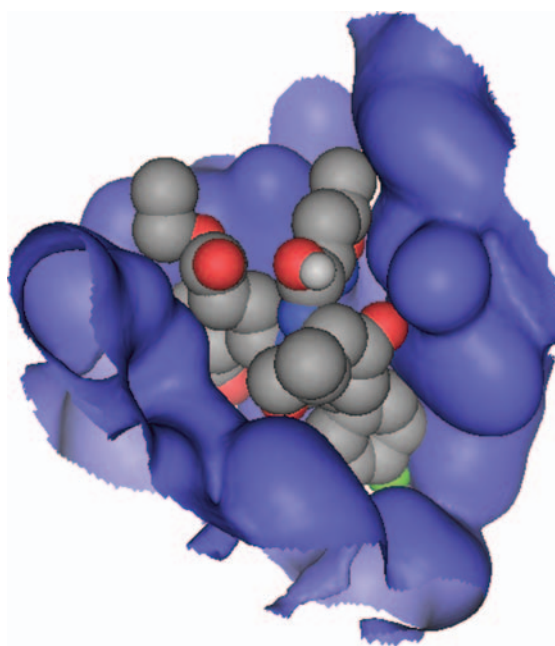


**Fig. 4.** Hydrogen bonds between ligand 4 and the involved residues of SARS CoV Mpro. The purple dots denote the hydrogen bonds and LIG4 denotes ligand4 and VAL\_186 and MET\_165 stand for the involved residues respectively. All the atoms are shown in stick drawing and colored by atom type in which hydrogen is colored white, carbon gray, oxygen red, nitrogen blue, chlorine green and sulfur yellow

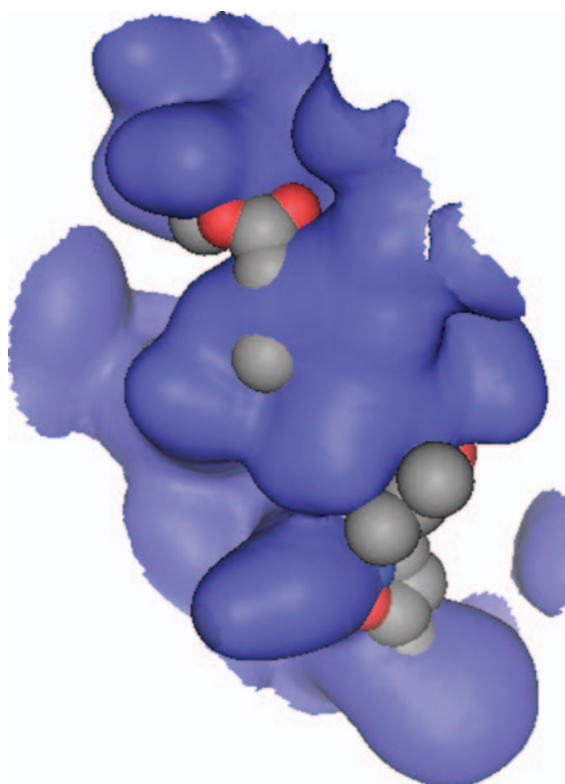
of two conformations in ligand 7 with the cluster's RMDS tolerance set to 1.0 and some other conformations identical to that two, as shown in Fig. 2. The first-ranked



**Fig. 5.** Hydrophobic interactions between ligand7 and SARS Mpro. Hydrophobic residues of SARS Mpro are colored in green and hydrophilic ones in blue. The ligand is colored by atom type in which carbons are in gray, hydrogens in white, oxygens in red and nitrogens in blue. Carbon atoms face the hydrophobic areas neatly while oxygen and nitrogen atoms face hydrophilic areas. The solvent excluded surface are buried in each other



**Fig. 6.** Connolly surface of ligand-SARS CoV Mpro complex. The surface of SARS CoV Mpro is colored blue and ligand 4 colored by atom type



**Fig. 7.** Connolly surface of complex between ligand 27 and SARS CoV Mpro which shows steric hindrances between them. Connolly surface is colored blue and the ligand is colored by atom type. As shown, some atoms of ligand have penetrated the molecular surface of Mpro causing strong steric hindrance. Note that this surface is atom scalar surface different from solvent excluded surface

ligand 7 according to the order of Rank 2 is a probably a potent inhibitor.

Also, a comparison for the binding structures among the best conformations of ligands 7, 12 and 4 (top three from ligand 7, top two from ligand 12, and top one from ligand 4) indicates that they are quite similar to each other, especially in the five-atom ring linked to an alkyl as shown in Fig. 3. These findings may also be useful for structure-based drug design.

The hydrogen bonds make important contributions to the interactions between ligand and receptor. They are observed in several of the conformations of the ligands whose complexes with Mpro were visually examined. In best four conformations of complexes formed by ligand 4 and Mpro, there are four, three, four, two hydrogen bonds formed, respectively. Those four hydrogen bonds involve Met-165, two involve Thr-175, two involve Val-186 and one involves His-164, His-172, Tyr-182. An example of hydrogen bonding between ligand 4 and the participating residues of Mpro is shown in Fig. 4. And there are also some hydrogen bonds formed between ligands 7, 12 and Mpro respectively in which Gly-174 occurs for five times, His-164, Arg-40 and Tyr-182 twice and Thr-175, His-172 and Leu-50 once. From the frequency of the residue's occurrence in the formation of hydrogen bonding, His-164 plays an important role, which may be due to the polar hydrogens of its imidazole ring being a perfect hydrogen bond donor. And Gly-174, Met-165, and Thr-175 also take part in the hydrogen bonding with relatively high frequency. This information could be helpful for our drug design in which the potential drug should interact well with these residues, especially with His-164. His-164 is of great concern in anti-SARS drug design which is believed to be one of the catalytic dyad (Anand et al., 2003; Du et al., 2005a, b). Therefore it is important to have this imidazole ring locked such that its catalytic function is depressed. It was found there were 6 hydrogen bonds between KZ7088 and nearby residues of the receptor (Chou et al., 2003). Among all the ligands screened we have not found anyone any formed more than 4 hydrogen bonds.

In comparison with the aforementioned active sites, some residues involved in the hydrogen bonds such as Gly-174, Thr-175 and Arg-40 are out of this set. Nonetheless they are quite close to the residues of the active site.

Besides hydrogen bonds, there are some other contributions to the total Gibbs free energy induced by the bonding, such as desolvation, hydrophobic interaction, and electrostatic interaction. The hydrogen bond may

act as an "anchor", determining the 3D position of the ligand in the binding pocket, thus facilitating the hydrophobic interactions and electrostatic interactions

In the output of Autodock's calculations, the potential energy of van der Waals interaction and electrostatic interaction is reported separately. So we selected a few groups of molecules with the same number of atoms but with obvious difference in energy whose van der Waals potentials and electrostatic potentials were added up. In the same group, the electrostatic interaction energy varies from one to another but it is too small in comparing with the total free energy. As we have calculated, the van der Waals energy plays a main part of the total energy. There are also some other factor which should be responsible for the remaining difference of energy, varies from 1 to 3 kcal/mol. Then the desolvation term emerges as a possibility which is included in autodock's scoring function. As for the torsional free energy, it depends on the number of torsions of each molecule and how the molecule is aligned in the pocket.

The solvent-excluded surface of the ligand-receptor complex is examined and shown in Fig. 5. The hydrophobic areas of the ligand and Mpro are complement with respect to each other as we can see in the Fig. 5. Thus the entropy of desolvation upon binding increases through the depart of water molecules from the molecular surface. The binding does not take place in vacuum but in solvation, so water is deeply involved and played an important role.

The study of hydrophobic interaction is relevant to the protein-folding problem (Anfinsen, 1973) and hydrophobicity (Tanford, 1962). The model of a folded protein with the hydrophobic amino acid side chains in the interior forming an oil drop, implying that as the protein folds the hydrophobic side chains were preferentially buried away from the external solvent. In our study, the process of the ligand-receptor surface buried in each other is just like the case of protein folding. So in further rational drug design it is critic to identify the hydrophobic groups of the ligand and receptor and ensure that they are facing each other upon binding.

What we have discussed above is thermodynamics of ligand-receptor binding. Moreover, the other side of this binding – kinetics should also be taken into consideration. The kinetics of binding process is mainly determined by steric complementarity (Vaidyanathan et al., 2001; Chou and Zhou, 1982; Chou, 1975, 1976; Zhou et al., 1982; Chou and Jiang, 1974; Zhou and Zhong, 1982). It is the steric complementarity that makes the cavity site acces-

sible for the ligand. The Connolly surface of ligand-SARS Mpro is examined and shown in Fig. 6, where ligand 4 forms four hydrogen bonds with SARS Mpro, and fully occupies the cavities and lies comfortably in the receptor's pocket.

To give evidence for the key role of steric complementarity, a poor steric complementarity or steric hindrance is shown in Fig. 7. Although there are three hydrogen bonds in the complex, its Gibbs free energy is still 10 kcal/mol higher than that of ligand 4. That means steric complementarity is supposed to be responsible for the micro-mechanical interlock at molecular level which is suggested by Vaidyanathan et al. (2001). The primary driving force for mechanical interlocking is the steric complementarity between the ligand and the collagen receptor site. Once mechanical interlocking is facilitated by steric complementarity, other binding interactions may also occur. Therefore our further drug design or modification of the lead compounds must first take the steric configuration into account.

## Conclusions

Three compounds,  $C_{28}H_{34}O_4N_7Cl$ ,  $C_{21}H_{36}O_5N_6$ , and  $C_{21}H_{36}O_5N_6$ , i.e., ligands 4, 7 and 12 of Table 1, may be promising candidates for further modification and structure-based drug design. Further studies are currently under way in our laboratories.

## Acknowledgements

We thank Professor Arthur J. Olson for his kindness in offering us the AutoDock 3.0.3 program. This work was supported by the Tianjin Commission of Science and Technology under the contract No. 033801911 and the Special Fund for Intensive Computation from the Tianjin Commission of Education.

## References

- Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300: 1763–1767
- Anfinsen CG (1973) Principles that govern folding of protein chains. *Science* 181: 223–230
- Brown RD, Martin YC (1996) Use of structure-activity data to compare structure-based clustering. *Methods and descriptors for use in compound selection*. *J Chem Inf Comput Sci* 36: 572–584
- Chen L-L, Ou H-Y, Zhang R, Zhang C-T (2003) ZCURVE\_CoV: a new system to recognize protein coding genes in coronavirus genomes, and its applications in analyzing SARS-CoV genomes. *Biochem Biophys Res Commun* 307: 382–388
- Chou KC (1975) Studies on the enzyme kinetics of the cavity-active site. *Acta Biochimica Biophys Sin* 7: 95–103
- Chou KC (1976) The kinetics of the combination reaction between enzyme and substrate. *Sci Sin* 19: 505–528
- Chou KC (1992) Energy-optimized structure of antifreeze protein and its binding mechanism. *J Mol Biol* 223: 509–517
- Chou KC (2004) Review: structural bioinformatics and its impact to biomedical science. *Curr Med Chem* 11: 2105–2134
- Chou KC, Carlacci L (1991) Simulated annealing approach to the study of protein structures. *Protein Eng* 4: 661–667
- Chou KC, Jiang SP (1974) Studies on the rate of diffusion-controlled reactions of enzymes. *Sci Sin* 17: 664–680
- Chou KC, Pottle M, Nemethy G, Ueda Y, Scheraga HA (1982) Structure of beta-sheets: origin of the right-handed twist and of the increased stability of antiparallel over parallel sheets. *J Mol Biol* 162: 89–112
- Chou KC, Scheraga HA (1982) Origin of the right-handed twist of beta-sheets of poly-L-valine chains. *Proc Natl Acad Sci USA* 79: 7047–7051
- Chou KC, Wei DQ, Zhong WZ (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. *Biochem Biophys Res Commun* 308: 148–151
- Chou KC, Zhou GP (1982) Role of the protein outside active site on the diffusion-controlled reaction of enzyme. *J Am Chem Soc* 104: 1409–1413
- Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 348: 1967–1976
- Du Q, Wang S, Wei DQ, Sirois S, Chou KC (2005a) Molecular modelling and chemical modification for finding peptide inhibitor against SARS CoV Mpro. *Anal Biochem* 337: 262–270
- Du QS, Wang SQ, Jiang ZQ, Gao WN, Li YD, Wei DQ, Chou KC (2005b) Application of bioinformatics in search for cleavable peptides of SARS-CoV Mpro and chemical modification of octapeptides. *Med Chem* 1: 209–213
- Gao F, Ou H-Y, Chen L-L, Zheng W-X, Zhang C-T (2003) Prediction for proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Lett* 553: 451–456
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348: 1953–1966
- Miura HS, Nakagaki K, Taguchi F (2004) N-terminal domain of the murine coronavirus receptor CEACAM1 is responsible for fusogenic activation and conformational changes of the spike protein. *J Virol* 78: 216–223
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1988) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J Comput Chem* 9: 1639–1662
- Rayer M (2002) Geometric problems and algorithms in computer-aided molecular design. In: Thomas L (ed) *Bioinformatics from genomes to drugs*, vol. I. Wiley-VCH, Weinheim, pp 318–319
- Sheridan RP, Miller MD, Underwood DJ, Kearsley SK (1996) Chemical similarity using geometric atom pair descriptors. *J Chem Info Comput Sci* 36: 128–136
- Shortridge KF (2003) Severe acute respiratory syndrome and influenza: virus incursions from southern China. *Am J Respir Crit Care Med* 168: 1416–1420
- Sirois S, Wei DQ, Du Q, Chou KC (2004) Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. *J Chem Inf Comput Sci* 44: 1111–1122
- Sirois S, Hatzakis GE, Wei DQ, Du Q, Chou KC (2005) Assessment of chemical libraries for their druggability. *Comput Biol Chem* 29: 55–67
- Tanford C (1962) Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J Am Chem Soc* 84: 4240–4274

- Vaidyanathan J, Vaidyanathan TK, Yadav P (2001) Collagen-ligand interaction in dentinal adhesion: computer visualization and analysis. *Linaras Biomat* 22: 2911–2920
- Xiong B, Gui CS, Xu XY, Luo C, Chen J, Luo HB, Chen LL, Li GW, Sun T, Yu CY, Yue LD, Duan WH, Shen JK, Qin L, Shi TL, Li YX, Chen KX, Luo XM, Shen X, Shen JH, Jiang HL (2003) *Acta Pharmacol Sin* 24: 497
- Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H, Gao GF, Anand K, Bartlam M, Hilgenfeld R, Rao Z (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci USA* 100: 13190–13195
- Zhou GQ, Zhong WZ (1982) Diffusion-controlled reactions of enzymes. A comparison between Chou's model and Alberty-Hammes-Eigen's model. *Eur J Biochem* 128: 383–387
- Zhou ZP, Li TT, Chou KC (1982) The flexibility during the juxtaposition of reacting groups and the upper limits of enzyme reactions. *Biophys Chem* 14: 277–281
- 
- Authors' address:** Dong-Qing Wei, College of Life Science and Technology, Shanghai Jiaotong University, 800 Dongchuan Road, Minhang District, Shanghai 200240, China,  
Fax: +86-21-3420-4717, E-mail: dqwei@sjtu.edu.cn