Amino Acids (2007) 33: 129–135
DOI 10.1007/s00726-006-0403-1
Printed in The Netherlands

Amino Acids

Virtual screening for finding natural inhibitor against cathepsin-L for SARS therapy

S.-Q. Wang[1], **Q.-S. Du**[2,5], **K. Zhao**[1], **A.-X. Li**[3], **D.-Q. Wei**[4,5], and **K.-C. Chou**[5]

[1] College of Pharmaceuticals and Biotechnology, Tianjin University, Tianjin, China
[2] Institute of Bioinformatics and Drug Discovery, Tianjin Normal University, Tianjin, China
[3] Tianjin Wujing Medical Institute, Tianjin, China
[4] College of Life Science and Technology, Shanghai Jiaotong University, Shanghai, China
[5] Gordon Life Science Institute, San Diego, CA, U.S.A.

Received April 25, 2006
Accepted August 8, 2006
Published online September 29, 2006;

Summary. Recently Simmons et al. reported a new mechanism for SARS virus entry into target cells, where MDL28170 was identified as an efficient inhibitor of CTSL-meditated substrate cleavage with IC_{50} of 2.5 nmol/l. Based on the molecule fingerprint searching method, 11 natural molecules were found in the Traditional Chinese Medicines Database (TCMD). Molecular simulation indicates that the MOL376 (a compound derived from a Chinese medicine herb with the therapeutic efficacy on the human body such as relieving cough, removing the phlegm, and relieving asthma) has not only the highest binding energy with the receptor but also the good match in geometric conformation. It was observed through docking studies that the van der Waals interactions made substantial contributions to the affinity, and that the receptor active pocket was too large for MDL21870 but more suitable for MOL736. Accordingly, MOL736 might possibly become a promising lead compound for CTSL inhibition for SARS therapy.

Keywords: Severe acute respiratory syndrome (SARS) – MDL28170 – KZ7088 – Molecular simulation – Docking – Structural bioinformatics

1. Introduction

Severe acute respiratory syndrome (SARS) is an acute respiratory disease caused by a newly identified coronavirus (SARS-CoV) (Rota et al., 2003), the result of a zoonosis from a highly related animal coronavirus (Guan et al., 2003). It continues to be potential for further zoonotic transmission events, leading to the reintroduction of SARS-CoV into the human population. Previous studies for drug discovery against SARS were mainly targeting on 3C-like proteinase (namely main proteinase, CoV M^{pro}) of SARS-coronavirus, which is the key step in the transcription and replication of SARS-CoV (Anand et al., 2003; Chou et al., 2003; Yang et al., 2003; Chou, 2004a). Anand et al. (2003) suggested that the inhibitor AG7088 of rhinovirus CoV M^{pro}, could serve as a starting point for anti-SARS drug based on the three dimensional structure of the proteinase. Owing to the fitting problem of AG7088 to the binding pocket of SARS CoV M^{pro}, Chou et al. (2003) suggested its derivative KZ7088 as a better starting point. Du et al. (2005) developed a very promising approach to find polypeptide inhibitors against SARS-CoV M^{pro} by searching for the cleavage sites in proteins by the SARS enzyme. So far all efforts in discovering drug against SARS are associated with protease inhibitors targeting on the SARS CoV M^{pro}.

Recently a new mechanism for entry of SARS CoV into target cells was reported by Simmons et al. (2005). They hypothesized that cellular factors sensitive to ammonium chloride, such as pH-dependent endosomal proteins, may have great effect on mediating SARS-CoV entry. In their study the requirements for proteases in activation of viral infectivity and the effect of protease inhibitors on SARS-CoV infection are examined. Their study results are consistent with a model in which SARS-CoV employs a unique three-step procedure in membrane fusion, involving receptor binding and induced conformational changes in S glycoprotein followed by cathepsin L (CTSL) proteolysis and activation of membrane fusion within endosomes.

The new progress provides an alternate target, cathepsin L, for drug design against SARS. A highly throughput screening of a library of pharmacologically active compounds was performed by Simmons et al. and MDL28170

(Simmons et al., 2005) was identified as an efficient inhibitor of CTSL-meditated substrate cleavage with high bioactivity IC50 of 2.5 nmol/l. Their research results demonstrate that MDL28170 can inhibit endosomal protease activity of SARS virus in the entering process of virus into the target cell. Therefore, MDL28170 can serve as a promising template for fusion inhibitor design targeting on CTSL.

In this study MDL28170 is used as the template in database searching based on the 3D-structure of CTSL for discovery of natural product drug candidates from the Chinese traditional herb medicines. The newly developed database, the Traditional Chinese Medicines Database (TCMD) (Zhou et al., 2004), was used in the database searching, which is a part of the research project on modernization of Chinese traditional medicines for finding out novel active herb compounds with low-toxicity using advanced drug design techniques.

2. Materials and methods

Materials

Structure of cathepsin L for docking

As cathepsin L is one of the major cysteine proteases, generally found throughout mammalian cell types and believed to be the primarily function in normal protein breakdown (Mary and McGrath, 1999). Cathepsin L has also been implicated in a number of disease processes, such as bone resorption and tumor metastases (Kakegawa, 1993). Because of the newly found functions of cathepsin L for proteolysis and activation of membrane fusion in SARS-CoV infection (Simmons et al., 2005), potent and selective inhibitors of the cysteine protease are expected to be good candidates for therapeutic agents of SARS.

The structure of cathepsin L (PDB code 1ICF) shares the common fold pattern of papain-like cysteine proteases (Guncar et al., 1999). The structure consists of two domains, left (L-) domain and right (R-) domain, reminiscent of a closed book with the spine in front. The domains separate on the top in a 'V'-shaped active site cleft, in the middle of which are residues Cys25 and His163, one from each domain, forming the catalytic dyad of the enzyme (Fig. 1).

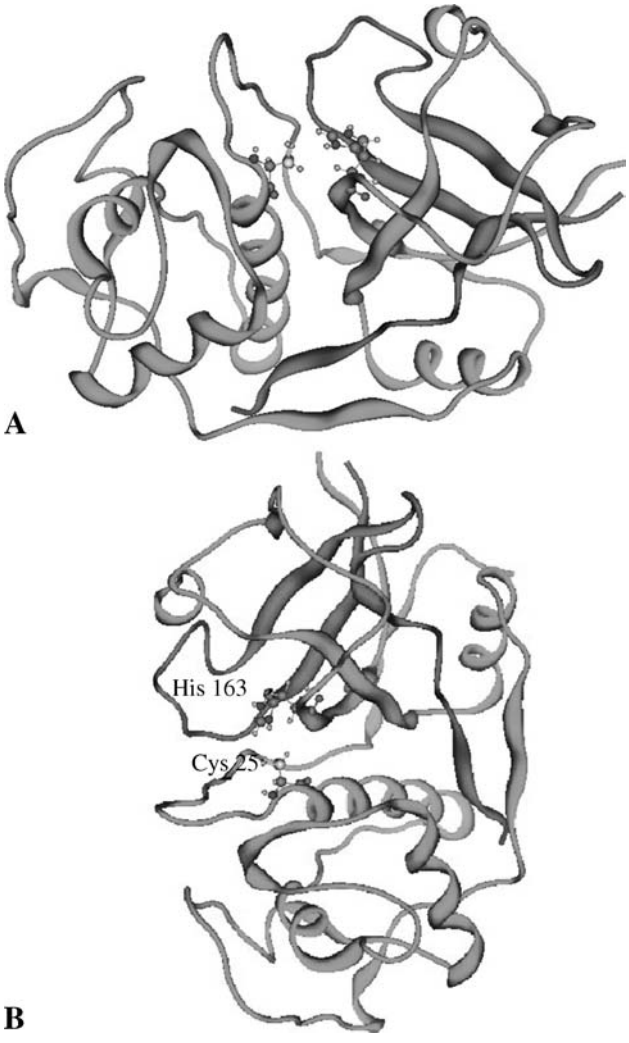

Fig. 1. Schematic representation of cathepsin L (1ICF). Side chains of the catalytic residues Cys25 and His163 are shown. The green flat ribbon represents heavy chain A, the red is light chain B. **A** Standard view, along the interdomain interface and the active site cleft of cathepsin L, with L-domain and R-domain on the left and right. **B** Side view, perpendicular to the standard view. (For interpretation of the references to color in the figure, the reader is referred to the web version of this paper under www.springerlink.com)

MDL21870

The chemical structure of MDL21870 (Jeffrey et al., 1995), an amino-terminal blocked dipeptide aldehyde, is displayed in Fig. 2, which was once employed as the inhibitor of calpain to protect rat erythrocyte membrane associated cytoskeletal proteins from proteolytic degradation (IC50 = 1 μM) (Shujaath et al., 1998) and which can also inhibit ß-amyloid processing (Jeffrey et al., 1995). The assay results of Simmons et al. (Simmons et al., 2005) suggest that MDL28170 can also inhibit CTSL-meditated substrate cleavage with IC50 of 2.5 nmol/l during SARS virus entry. Therefore, MDL28170 may become an initial potent candidate of antiviral inhibitor against SARS-CoV.

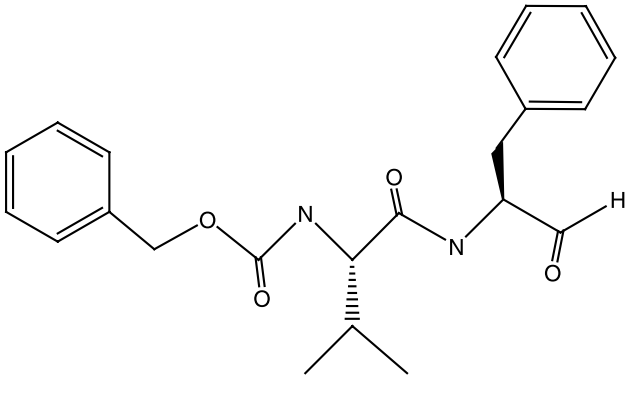

Fig. 2. Structure of MD28170

Methods

Structural bioinformatics (Chou, 2004a) has been applied to timely derive the 3D structures of some functionally important proteins (see, e.g., Chou, 2004b, c), helping to understand their action mechanisms (see, e.g.,

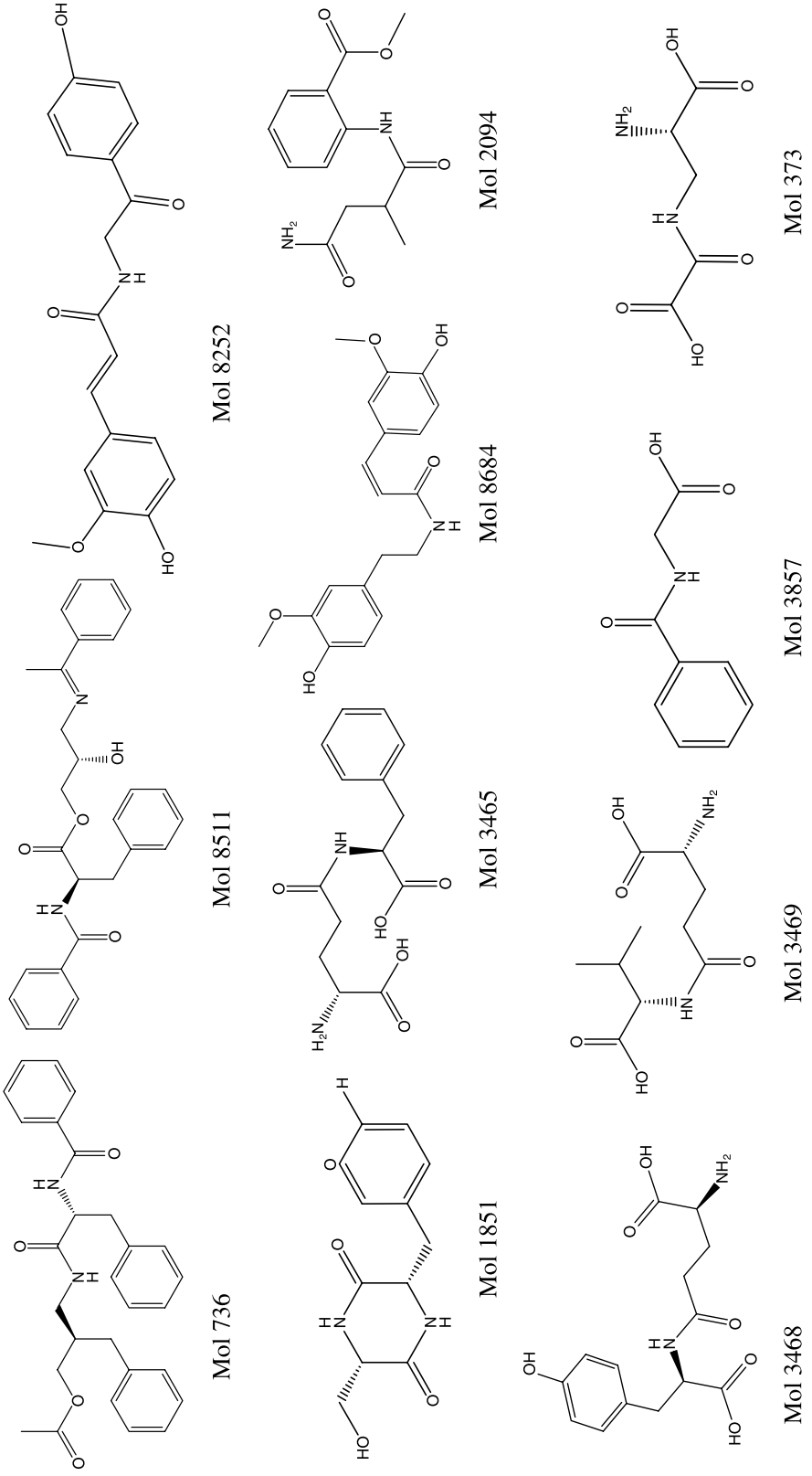

Fig. 3. The molecular structures of 11 compounds satisfied the "rule of 5" and found in the database-search

Chou et al., 1999; Chou and Howe, 2002; Chou, 2005a), and stimulating the course of drug discovery (see, e.g., Chou, 2004d, e; 2005b; Wei et al., 2006). In this study, the approach of structural bioinformatics is to use for finding inhibitors for SARS therapy.

Similarity search

The first step in similarity searching is to compute the fingerprints of all the molecules in the molecular database, which are represented by a set of features derived from the molecular structures. The molecular features are calculated based on the topology of the chemical graphs and the 3D conformations of molecules, and expressed by the following equation:

$$\mathrm{U} = \{\text{is-aromatic, has-ring, has-C, has-N, has-O, has-S, has-P, has-halogen}\} \quad (1)$$

In Eq. (1) the notation is-aromatic represents aromatic ring, has-ring is saturated ring, has-C, has-N, has-O, has-S, has-P and has-halogen represent the number of carbon, nitrogen, oxygen, sulfur, phosphorus and halogen, respectively. The 8 features in Eq. (1) form a subset U. Each term in Eq. (1) takes value 0 or 1, therefore, there are total $2^8 = 256$ possible fingerprints.

In this study MACSS Structural Keys (Sheridan et al., 1996; Bron and Martin, 1996) is used to calculate the fingerprints of 10458 natural product molecules in the database. In the second step a metric is used to compare the fingerprints with the template. In this study the metric Tanimoto Coefficient is used, which is a number between 0 and 1, where 0 means "maximally dissimilar" and 1 means "maximally similar". The Tanimoto coefficient is defined by the following equation,

$$\#\mathrm{AB}/(\#\mathrm{A} + \#\mathrm{B} - \#\mathrm{AB}) \quad (2)$$

where #A denotes the number of features in fingerprint A and #B is the number of features in fingerprint B, and #AB denotes the number of common features possessed by both A and B.

Docking and scoring

Computational docking operation is a useful vehicle for investigating molecular binding interactions (see, e.g., Zhou and Troy, 2003; 2005a, b). Here the well known docking program DOCK (Ewing et al., 2001) is used in this study with the MPI parallelization, exhaustive orientation searching, GB/SA solvation scoring and chemical/pharmacophore screening (Kuntz et al., 2005). In the calculations of binding energies Dock 5.30 program is performed using Amber force field,

$$E_{\mathrm{total}} = E_{\mathrm{vdw}} + E_{\mathrm{ele}} \quad (3)$$

where E_{total} is the binding energy of the ligand with receptor, E_{vdw} is van der Waals interaction energy, and E_{ele} is the electrostatic interaction energy. In the calculations the ligand molecule is flexible, but receptor is rigid. Then the binding energies are scored based on the E_{total} values.

3. Results and discussion

In the database search a total of 26 compounds with the initial overlap 60% was found in TCMD, more than 10458 natural product molecules; the search was conducted by using MDL28170 as the template molecule based on the molecular similarity described in Section 3. All the 26 molecules are filtrated according to the drug discovery "rule of 5" (Lipinski et al., 1997); 15 compounds are deleted and the molecular structures of 11 compounds are shown in Fig. 3, which are satisfied the "rule of 5".

The docking calculation is performed by Dock 5.3 program using Amber99 force field. The amino acid residues of the 1ICF within 6 Å are involved; the ligand is flexible but the receptor is rigid. The binding energies and favorable conformations in the active pocket of CTSL of the 11 molecules and template MDL21870 are computed, and the binding energies are ranked in Table 1. As shown in Table 1, two molecules mol 736 and MDL21870 have the higher scores. The docking binding energy of MOL736 is even bigger than that of MDL21870. Therefore, it is a possible inhibitor candidate to cathepsin L.

The compound MOL736 (see Fig. 4) is from a Chinese medicine herb, Aurantiamide acetate derived from the

Table 1. The rank of docking binding energy of ligands to 1ICF receptor (kcal/mol)

Ligand (no.)	E_{total}	E_{vdw}	E_{ele}
736	−50.767	−43.356	−7.411
MDL21870	−47.887	−41.014	−6.873
8511	−46.407	−41.282	−5.125
8684	−43.641	−37.151	−6.490
3468	−43.157	−31.901	−11.256
8252	−42.746	−34.812	−7.934
3465	−41.030	−30.515	−10.515
2094	−36.356	−26.339	−10.017
3469	−34.340	−25.982	−8.358
373	−32.794	−20.140	−12.654
1851	−32.027	−26.641	−5.386
3857	−28.371	−20.807	−7.564

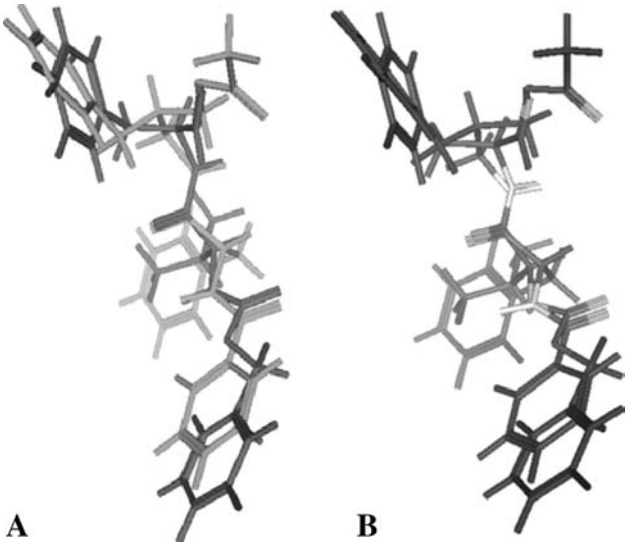

Fig. 4. **A** Alignment of the two ligands colored with chain types, where red and green represent MDL28170 and MOL736, respectively. **B** The alignment of the two ligands colored with pharmacophore types, white for hydrogen bond donor, cyan for hydrogen bond receptor, and dark green for hydrophobic atoms, respectively

Artemisia annua (Zhou et al., 2004). This herb medicine has the therapeutic efficacy on the human body such as antitracheitis, relieving cough, removing the phlegm, relieving asthma and so on (Sheng et al., 2004).

The flexible alignment for the two ligands are performed based on a set of predetermined features in order to find the common features with which the two ligands share, such as shape, log P (Octanol/Water) values, aromatic groups, hydrogen bond donors and acceptors. Figure 4A and B show the flexible alignment for two ligands colored according to pharmacophore types and chain types. The colors used for pharmacophore are as follows: white for hydrogen bond donor, cyan for hydrogen bond receptor and dark green for hydrophobic atoms. As shown in Fig. 4, the two ligands have highly chemical similarity, especially the hydrophobic sides, aromatic groups, the hydrogen bond donors and acceptors, and carbonyl and amidogen.

In the ligand-receptor docking calculations, the most favorable conformation for each ligand is selected from 100 conformations with lower energies, as shown in Fig. 5. Hydrogen bonds are supposed to make important contributions to the interactions between the ligand and receptor (see, e.g., Chou et al., 2000, 2003; Chou, 2004d, e). The docking results indicate there are two hydrogen bonds for MDL21870 ligand to the 1ICF receptor: Figure 5A shows the hydrogen bonds between ligand MDL28170 and 1ICF. One hydrogen bond is between a carbonyl oxygen of

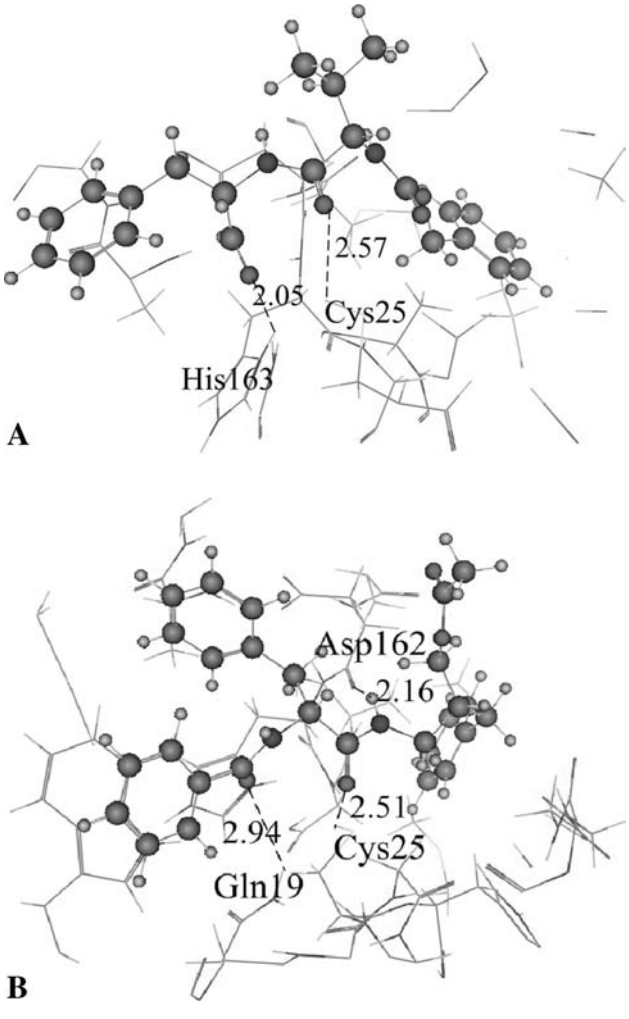

Fig. 5. The close view of the hydrogen bonds between receptor 1ICF and two ligands, **A** MDL28170, **B** MOL736. The molecule with the ball-line represents the ligand, some residues of the receptor close to ligand are line, where the red, grey, blue, yellow and light-gray colors represent oxygen, carbon, nitrogen, sulfur and hydrogen, respectively. (For interpretation of the references to color in the figure, the reader is referred to the web version of this paper under www.springerlink.com)

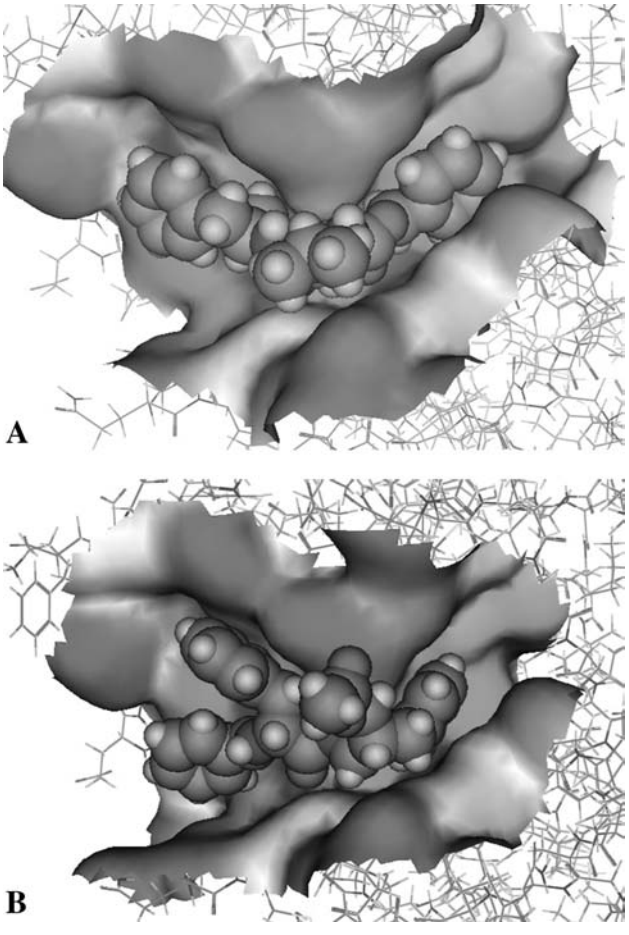

Fig. 6. Molecular surfaces of the ligands and the active pocket of receptor 1ICF. **A** Surface of MDL28170, **B** surface of MOL736. The molecular surface was rendered with the pocket attribute. The constituents of the pocket are defined by those residues within a distance 6 Å from each ligand. The green is for hydrophobic, blue for hydrophilic, and red represents exposed area, respectively. (For interpretation of the references to color in the figure, the reader is referred to the web version of this paper under www.springerlink.com)

ligand and H(N) of His-163 (2.05 Å), the other is between another carbonyl oxygen of ligand and the H(S) of Cys-25(2.57 Å). Figure 5B shows the hydrogen bond interaction between MOL736 ligand and 1ICF. There are three H-bonds. The first H-bond is carbonyl oxygen of the ligand to H(S) of Cys-25 (2.57 Å), the second is H(N) of MOL736 to carbonyl oxygen of Asp-162 of receptor (2.16 Å), and the last is the another carbonyl oxygen of ligand to H(N) of Gln19(2.94 Å).

As shown in Table 1 the van der Waals interaction energy makes the largest proportion in the affinity which means the contributions of van der Waals interaction is the key factor in determining the binding energy. Figure 6 is the plot of the molecular surface of the residues close to the ligands. The constituents of the active pocket (Chou et al., 1999) are defined by the residues within a distance 6 Å from each ligand. The surfaces were rendered with the pocket attributes, green for hydrophobic, blue for hydrophilic, and red for exposed surfaces, respectively. As shown in Fig. 6B, MOL736 has better matches with hydrophobic areas in the active pocket, especially the two hydrophobic aromatic rings embedded in the active pocket in the left domain are more matchable than MDL28170. It indicates that the left pocket space may be too large for MDL28170. This observation indicates that this hydrophobic domain of the receptor needs a bigger hydrophobic group. These findings would be helpful for future rational drug design.

4. Conclusion

The combination of the Traditional Chinese Medicine Database (TCMD) and the modern three dimensional drug design technique provides a powerful tool for new drug discovery and may find natural product drugs with high efficient and lower toxicity. The newly identified proteinase CTSL is an important target in drug design for therapeutic intervention of SARS-CoV infection. Because the function of CTSL is in the membrane fusion step, the inhibitors of CTSL may prevent the virus from entering the target cells. Therefore, it may be a better target than SARS-CoV M^{pro}. Eleven Chinese herb compounds are screened from TCMD based on the structure of template MDL28170. Among them MOL736 has the highest similarity and lowest binding energy. The compound MOL736 (chemical name Aurantiamide acetate) is derived from a Chinese medicine herb, Artemisia annua (Zhou et al., 2004). This plant is known as a herb medicine with many therapeutic efficacies, such as antitracheitis, relieving cough, removing the phlegm, relieving asthma and so on (Sheng et al., 2004). Therefore, it is a hopeful drug candidate against SARS.

Acknowledgment

This work is supported by the Chinese National Science Foundation (NSFC) under the contract number 20373048.

References

Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R (2003) Coronavirus main proteinase (3CL(pro)) structure: Basis for design of anti-SARS drugs. Science 300: 1763–1767

Bron RD, Martin YC (1996) Use of structure-activity data to compare structure-based clustering: methods and descriptors for use in compound selection. J Chem Info Comput Sci 36: 572–584

Chou KC (2004a) Review: Structural bioinformatics and its impact to biomedical science. Curr Med Chem 11: 2105–2134

Chou KC (2004b) Modelling extracellular domains of GABA-A receptors: subtypes 1, 2, 3, and 5. Biochem Biophys Res Commun 316: 636–642

Chou KC (2004c) Insights from modelling three-dimensional structures of the human potassium and sodium channels. J Proteome Res 3: 856–861

Chou KC (2004d) Insights from modelling the tertiary structure of BACE2. J Proteome Res 3: 1069–1072

Chou KC (2004e) Insights from modelling the 3D structure of the extracellular domain of alpha7 nicotinic acetylcholine receptor. Biochem Biophys Res Commun 319: 433–438

Chou KC (2005a) Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. J Proteome Res 4: 1681–1686

Chou KC (2005b) Modeling the tertiary structure of human cathepsin-E. Biochem Biophys Res Commun 331: 56–60

Chou KC, Howe WJ (2002) Prediction of the tertiary structure of the beta-secretase zymogen. Biochem Biophys Res Commun 292: 702–708

Chou KC, Watenpaugh KD, Heinrikson RL (1999) A Model of the complex between cyclin-dependent kinase 5(Cdk5) and the activation domain of neuronal Cdk5 activator. Biochem Biophys Res Commun 259: 420–428

Chou KC, Tomasselli AG, Heinrikson RL (2000) Prediction of the tertiary structure of a caspase-9/inhibitor complex. FEBS Lett 470: 249–256

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 (Erratum: ibid., 2003, Vol. 310, 675)

Du QS, Wang SQ, Jiang ZQ, Gao WN, Li Y, Wei DQ, Chou KC (2005) Application of Bioinformatics in search for cleavable peptides of SARS CoV Mpro and chemical modification of octapeptides. Med Chem 1: 209–213

Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) Dock 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 15: 411–428

Graham S, Dhaval NG, Andrew JR, Jacqueline DR, Scott LD, Paul B (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. Proc Natl Acad Sci OSA 102: 11876–11881

Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ et al (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302: 276–278

Guncar G, Pungercic G, Klemencic I, Turk V, Turk D (1999) Crystal structure of Mhc class II associated P41 II fragment bound

to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. EMBO J 18: 793–803

Jeffrey H, Diana Q, Zhong ZY, Barbara C (1995) Inhibition of /I-amyloid formation identifies proteolytic precursors and SubceUular site of catabolism. Neuron 14: 651–659

Kakegawa H (1993) Particapation of cathepsin L in bone resorption. FEBS Lett 321: 247–250

Kuntz ID, Demetri TM, Lang PT (2005) DOCK 5.3 User Manual. University of California

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23: 2–25

Mary E, McGrath ME (1999) The lysosomal cysteine proteases. Annu Rev Biophys Biomol Struct 28: 181–204

Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH et al (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300: 1394–1399

Sheng XG, Wu J, Liu H, Zhang JI (2004) Studies on the components of the fatty acid of Artemisia annua. Grassland Turf 4: 68–70

Sheridan RP, Miller MD, Underwood DJ, Kearsley SK (1996) Chemical similarity using geometric atom pair descriptors. J Chem Info Comput Sci 36: 128–136

Shujaath M, Michael RA, Jeffery SW, Philippe B (1998) Biochem Biophys Res Commun 157: 1117–1123

Wei DQ, Du QS, Sun H, Chou KC (2006) Insights from modeling the 3D structure of H5N1 influenza virus neuraminidase and its binding interactions with ligands. Biochem Biophys Res Commun 344: 1048–1055

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 ZH (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 GP, Troy FA 2nd (2003) Characterization by NMR and molecular modeling of the binding of polyisoprenols and polyisoprenyl recognition sequence peptides: 3D structure of the complexes reveals sites of specific interactions. Glycobiology 13: 51–71

Zhou GP, Troy FA 2nd (2005a) NMR study of the preferred membrane orientation of polyisoprenols (dolichol) and the impact of their complex with polyisoprenyl recognition sequence peptides on membrane structure. Glycobiology 15: 347–359

Zhou GP, Troy FA (2005b) NMR studies on how the binding complex of polyisoprenol recognition sequence peptides and polyisoprenols can modulate membrane structure. Curr Protein Peptide Sci 6: 399–411

Zhou JJ, Xie GR, Yan XJ (2004) Traditional chinese medicines: molecular structures, natural sources and applications. Chemical Industry Press, China

Authors' address: Professor Qi-Shi Du, Institute of Bioinformatics and Drug Discovery, Tianjin Normal University, Tianjin 300073, China Fax: +86-22-2354-0187, E-mail: lifescience@san.rr.com