A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening[1]

XIONG Bin[2], GUI Chun-Shan[2], XU Xiao-Ying[2], LUO Cheng[2], CHEN Jing[2], LUO Hai-Bin[2], CHEN Li-Li[2], LI Guo-Wei[2], SUN Tao[2], YU Chang-Ying[2], YUE Li-Duo[2], DUAN Wen-Hu[2], Shen Jing-Kang[2], QIN Lei[3], SHI Tie-Liu[4*], LI Yi-Xue[3,4*], CHEN Kai-Xian[2*], LUO Xiao-Min[2*], SHEN Xu[2*], SHEN Jian-Hua[2*], JIANG Hua-Liang[2*]

[2]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201203; [3]Shanghai Center for Bioinformation Technology, Shanghai 201203; [4] Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200031, China

KEY WORDS severe acute respiratory syndrome (SARS); 3CL proteinase; inhibitors; molecular modeling; virtual screening; bioinformatics

ABSTRACT

AIM: To constructed a three-dimensional (3D) model for the 3C like (3CL) proteinase of SARS coronavirus (SARS_CoV), and to design inhibitors of the 3CL proteinase based on the 3D model. **METHODS:** Bioinformatics analyses were performed to search the homologous proteins of the SARS_CoV 3CL proteinase from the GenBank and PDB database. A 3D model of the proteinase was constructed by using homology modeling technique. Targeting to the 3D model and its X-ray crystal structure of the main proteinase (M^{pro}) of transmissible gastroenteritis virus (TGEV), virtual screening was performed employing molecular docking method to identify possible 3CL proteinase inhibitors from small molecular databases. **RESULTS:** Sequence alignment indicated that the SARS_CoV 3CL proteinase was extremely homologous to TGEV M^{pro}, especially the substrate-binding pocket (active site). Accordingly, a 3D model for the SARS_CoV 3CL proteinase was constructed based on the crystal structure of TGEV M^{pro}. The 3D model adopts a similar fold of the TGEV M^{pro}, its structure and binding pocket feature are almost as same as that of TGEV M^{pro}. The tested virtual screening indicated that 73 available proteinase inhibitors in the MDDR database might dock into both the binding pockets of the TGEV M^{pro} and the SARS_CoV 3CL proteinase. **CONCLUSIONS:** Either the 3D model of the SARS_CoV 3CL proteinase or the X-ray crystal structure of the TGEV M^{pro} may be used as a starting point for design anti-SARS drugs. Screening the known proteinase inhibitors may be an appreciated shortcut to discover anti-SARS drugs.

[1] Project supported by the 863 Hi-Tech Program, (No 2001AA235051, 2001AA235071 and 2001AA231111), the National Natural Science Foundation of China, (No 29725203, 20072042), the State Key Program of Basic Research of China, (No, 002CB512801 and 002CB512802), and the special programs of oppugning SARS from the Ministry of Science and Technology, Chinese Academy of Sciences and Shanghai Basic Research Project from the Shanghai Science and Technology Commission, (No 02DJ14070).

[*] Correspondence to Prof JIANG Hua-Liang, SHEN Jian-Hua, LUO Xiao-Min, SHEN Xu, CHEN Kai-Xian, LI Yi-Xue, and SHI Tie-Liu.

Received 2003-05-16 Accepted 2003-05-18

INTRODUCTION

Recently, severe acute respiratory syndrome (SARS) is being spread in many countries and places[1,2]. Therefore, discovering anti-SARS drugs is mostly significant. Genomic sequencing and bioinformatics analyses have addressed the important proteins that may be associated with the SARS coronaviruse (SARS_CoV) infection, including the polymerase, the spike (S) glycoprotein, the envelope (E) protein, the membrane (M) protein, the nucleocapsid (N) protein, and the 3C like (3CL) proteinase[1,2]. Theoretically, all these proteins can be used as targets to screen anti-SARS drugs. However, the 3CL proteinase is a preferred target for the task of discovering anti-SARS drugs in currently emergent situation of SARS spreading by following reasons: (1) the 3CL proteinase possibly plays an important role in the SARS_CoV replication as deduced from the function of the 3CL proteinases of other coronaviruses[3]; (2) numerous inhibitors of other 3CL proteinases are available and several of them are in clinic test, if some of them show anti-SARS activity, they can be developed as anti-SARS drugs rapidly; (3) 3CL proteinases is easily to be expressed (we have expressed this proteinase in *E coli* strain), and thus screening model can be established quickly; (4) homology modeling can be employed to construct the three-dimensional (3D) structure of this proteinase because highly homologous protein with X-ray crystal structure has been found (see discussion below), thereby structure-based drug design methods, such as virtual screening, can be applied to search active compounds from the compound databases.

In the present paper, we report a 3D structure model for the SARS_CoV 3CL proteinase constructed based on the X-ray crystal structure of the transmissible gastroenteritis virus (TGEV) main proteinase (M^{pro}). We also provide a virtual screening strategy for discovering anti-SARS drugs targeting to both the X-ray crystal structure of the TGEV M^{pro} and the modeled structure of the SARS_CoV 3CL proteinase.

MATERIALS AND METHODS

3D structure construction The sequence of 3CL proteinase of SARS was retrieved from the GenBank (GenBank protein ID NP_828863) (http://www.ncbi.nlm.nih.gov/). The Align123 module encoded in InsightII[4] was used in the pairwise sequence alignment. Align123 is a sequence alignment method developed based on the CLUSTAL W[5] program, combining with the secondary structural prediction, which may produce more precise alignment. According to the secondary structure information of M^{pro}, the sequence alignment was adjusted manually to obtain a fine alignment for 3D structure construction. The 3D model of the SARS 3CL proteinase was generated by using the MODELLER program[6] encoded in InsightII[4]. MODELLER uses a spatial restraint method to build up protein 3D models. The structure of each template proteins was used to derive spatial restraints expressed as probability density functions for each of the restrained features of the models. For the alignment of the SARS_CoV 3CL proteinase, MODELLER yielded three models, each of which contains three optimizing loop structures. The structure with the lowest violation score and lowest energy score was chosen as the candidate. Refine routine in the Homology module of InsightII[4] was used to adjust the positions of the side chains. Finally, the whole structural models were optimized using Amber force field[7] with the following parameters: a distance-dependent dielectric constant of 4.0, nonbonded cut-off 8 Å, Kollman-all-atom charges[7]. The structures were minimized by steepest descent first, then by conjugate gradient method to the energy gradient root-mean-square (RMS) <0.05 kcal· (mol· Å)$^{-1}$.

Several structural analysis softwares were adopted to check the structure quality. The Prostat module of InsightII[4] was used to analyze the properties of bonds, angles, and torsions. Profile-3D[8] program was used to check the structure and sequence compatibility.

Binding site mapping The major residues possibly compose the binding site of the SARS 3CL proteinase was identified by the sequence alignment with M^{pro} of TGEV and SiteID program encoded in Sybyl6.8 [9]. The surface structure of the binding pocket was constructed by using the MOLCAD module of Sybyl6.8[9].

Virtual screening Both the optimized 3D model of SARS 3CL proteinase and the X-ray crystal structure of TGEV M^{pro} were used as the targets for virtual screening on several databases, including MDL/ACD (http://www.mdli.com/), MDL/MDDR (http://www.mdli.com/), SPECS (http://www.specs.net/), the China Natural Product Database (CNPD) (http://www.neotrident.com/), compound sample database of the National Center for Drug Screening (http://www.screen.org.cn/). The virtual screening is being performed on the 64-processor SGI Origin 3800 supercomputer at the Drug Discovery and Design Center (http://www.dddc.ac.cn/) and the 392-processor Sunway-1

supercomputer at the Shanghai Supercomputer Center (http://www.ssc.net.cn/). So far the entire virtual screening has not been finished yet, here we just report the strategy and test results for the virtual screening against the SARS 3CL proteinnase 3D model and the X-ray crystal structure of TGEV M^{pro}.

No scoring function has been developed so far that may reliably and consistently predict a ligand-protein binding model and binding affinity[11]. Therefore, heuristic docking and consensus scoring strategies are often used in virtual screening; ie, different docking and scoring methods are applied to evaluate the screening results. In the present study, the program DOCK4.0[11,12] was employed for the primary screening. Residues around the catalytic center (around His41 and Cys145 for SARS_CoV 3CL proteinase and His41 and Cys144 for the TGEV M^{pro}) at radius of 6 Å was isolated for constructing the grids of docking screening, and the pocket composed by these residues was larger enough to include residues of the binding pocket. During the docking calculations, Kollman-all-atom charges[7] were assigned to the protein, and Geisterger-Hückel charges[13-15] were assigned to the small molecules in the small molecular databases due to lack of proper Kollman charges. The conformational flexibility of the compounds from the database was considered in the docking searching.

The DOCK suite of programs is designed to find possible orientations of a ligand in a "receptor" site[11]. The orientation of a ligand is evaluated with a shape scoring function and/or a function approximating the ligand-receptor binding energy. The shape scoring function is an empirical function resembling the van der Waals attractive energy. The ligand-receptor binding energy is taken to be approximately the sum of the van der Waals and electrostatic interaction energies. After the initial orientation and scoring evaluation, a grid-based rigid body minimization is carried out for the ligand to locate the nearest local energy minimum within the receptor binding site. The position and conformation of each docked molecule were optimized using single anchor search and torsion minimization method of DOCK4.0[11,12]. Thirty configurations per ligand building a cycle and 50 maximum anchor orientations were used in the anchor-first docking algorithm. All docked configurations were energy minimized using 100 maximum iterations and 1 minimization cycle.

Next, the top-1000 molecules for each database were selected for further analyses. These molecules were re-scored by Cscore[16] and the scoring function of AutoDock3.0[17]. Based on the second scoring result, 100 molecules were selected from each database according to above comprehensive scorings.

RESULTS AND DISCUSSION

Bioinformation analysis Recentely, various genomic sequences of SARS_CoV from different patients have been released in the GenBank, NCBI. Multi-alignment for the sequences of the 3CL proteinase of SARS resulted that only Asp189 was found to be substituted by Phe residue in one SARS strain (data are not shown). This minor difference may be caused by the sequencing error. This indicates that the 3CL proteinase is very conservative during the SARS mutations and can be used as a potential target for anti-SARS drug screening.

Sequence analysis and conserved residues identification were carried out among several virus genomes, including human coronavirus, murine hepatitis virus, porcine epidemic diarrhea virus, feline infectious peritonitis virus, avian infectious bronchitis virus, porcine teschovirus, and transmissible gastroenteritis virus (TGEV). Totally 43 BLAST hits were obtained. Sequence search from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/) for identifying the homologous proteins of the 3CL proteinase of SARS was performed by using the BLASTP program[18]. Only one protein was found with significant sequence homology to the SARS 3CL proteinase, the sequence identity is 43 %, positive 60 %, and gap 1 %. This protein is the main proteinase (M^{pro}) of TGEV. Therefore, the crystal structure of the M^{pro} of TGEV can be used as a template for modeling the 3D structure of SARS 3CL proteinase.

3D Model According to the bioinformatics analysis result, we constructed the three-dimensional (3D) structure model for the SARS 3CL proteinase based on the X-ray crystal structure of the M^{pro} of TGEV (PDB entry 1LVO) determinated by Anand *et al*[3]. The 3D model is shown in Fig 1, which superposes well with the X-ray crystal structure of the TGEV M^{pro}, the root-mean-square distance (RMSD) for the C^{a} atoms is about 0.34 Å. Because of sharing high homology, the SARS 3CL proteinase folds in a similar way of the M^{pro} of TGEV, containing three domains (Fig 1B). Domains I and II adopts chymotrypsin family fold, rich in β-sheets; domain III is a new folding found by Anand *et al*[3], containing five α helix strands.

Active site and binding pocket for the inhibi-

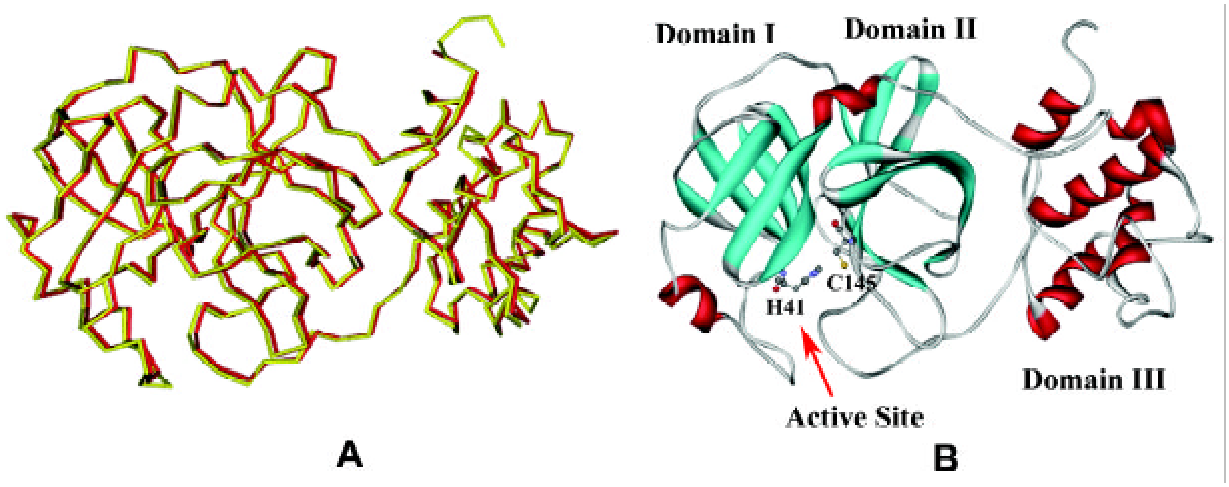

Fig 1. (A) Structure superposition of the modeled structure of SARS 3CL proteinase (yellow) with the X-ray crystal structure of the M^{pro} of TGEV (red). Only backbones are shown in this picture. (B) The solid ribbon representation of the structure model of SARS 3CL proteinase. The substrate-binding site is located at the deep cleft between domains I and II, and the active site is situated at the center of the cleft, the catalytic residues H41 and C145 are represented by ball-and-stick.

tors Sequence alignment (Fig 2) and structure superposition (Figs 1 and 3) between the SARS 3CL proteinase and the M^{pro} of TGEV indicated that His41 and Cys145 were the two catalytic residues. Possibly, in the substrate catalytic reaction by the SARS 3CL protinase, Cys145 acts as the nuclephilic attacking agent and His145 figures as a acid-base catalyst, identically to the Cyc144 and His41 of the TGEV M^{pro} [3]. The distance between the S atom of Cyc145 and the N^{e2} atom of His41 is 3.81 Å, about 0.3 Åshorter than the cysteine—

```
3clp: 1   SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIR 60
1lvo: 1   SGLRKMAQPSGLVEPCIVRVSYGNNVLNGLWLGDEVICPRHVIASDTTRV-INYENEMSS 59

3clp: 61  KSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTFSVLACYNG 120
1lvo: 60  VRLHNFSVSKNNVFLGVVSARYKGVNLVLKVNQVNPNTPEHKFKSIKAGESFNILACYEG 119

3clp: 121 SPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGK 180
1lvo: 120 CPGSVYGVNMRSQGTIKGSFIAGTCGSVGYVLENGILYFVYMHHLELGNGSHVGSNFEGE 179

3clp: 181 FYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE 240
1lvo: 180 MYGGYEDQPSMQLEGTNVMSSDNVVAFLYAALINGERWFVTNTSMSLESYNTWAKTNSFT 239

3clp: 241 PLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTPFDVVRQC 300
1lvo: 240 ELSS--TDAFSMLAAKTGQSVEKLLDSIVR-LNKGFGGRTILSYGSLCDEFTPTEVIRQM 296

3clp: 301 SGVTFQ 306
1lvo: 297 YGVNLQ 302
```

Fig 2. The sequence alignment of the SARS 3CL proteinase and the M^{pro} of TGEV. Residues composed the substrate-binding site are highlighted in red color.

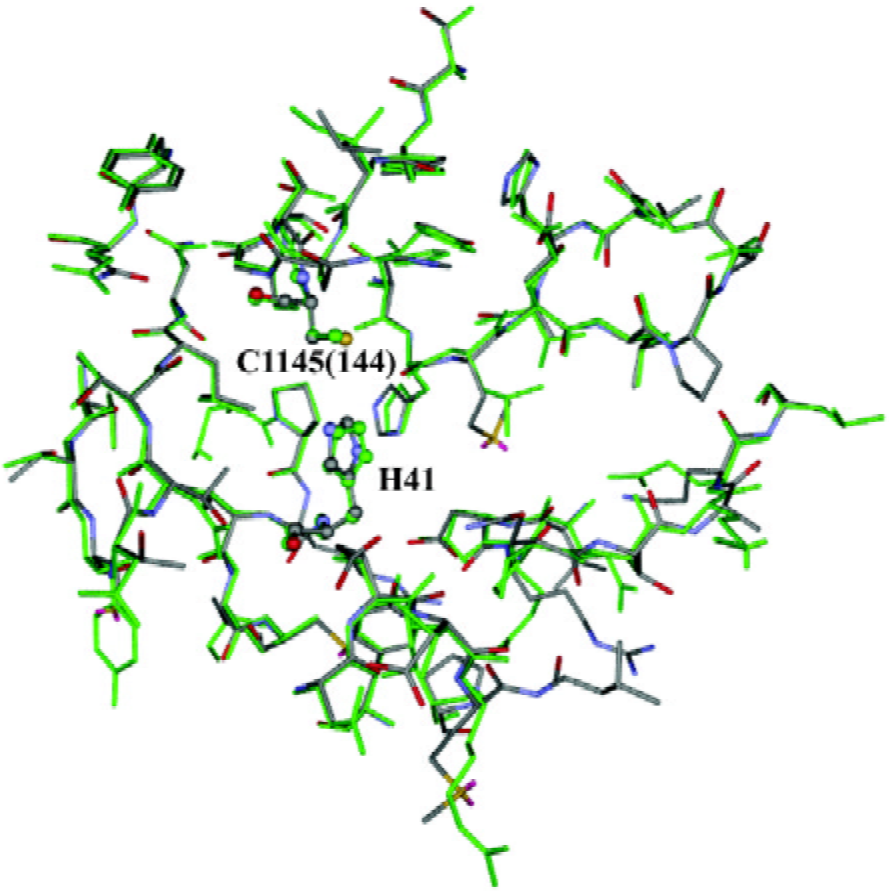

Fig 3. Superposition of the residues composed the substrate-binding pocket of the SARS 3CL proteinase (colored by atom type) with that of the TGEF M^{pro} (green). The two catalytic residues (histidine and cysteine) for both proteinases were represented by ball-and-stick.

histidine distance in TGEV M^{pro} [3]. Similar to the X-ray crystal structure of the TGEV M^{pro}, the substrate binding site is located in the deep cleft between domains I and II, lined by hydrophobic residues, and the active site is situated at the center of the cleft (Fig 1B). Residues around the two catalytic residues for each of proteinases at radius of 6 Å was isolated for constructing the substrate-binding pocket. The residues lined the binding pocket is very conservative (Fig 2). Structurally, residues composed the binding pockets of the two proteinases superpose well (Fig 3), the atomic RMSD between the two pockets is only 0.18 Å, indicating the spatial structure and shape of the substrate-binding pocket of the SARS_CoV 3CL proteinase is similar to that of the TGEV M^{pro}. Surface shapes for the two proteinases were constructed by using the MOLCAD program encoded in Sybyl6.8[9], which are shown in Fig 4. Fig 4 is impressed that the concave of the binding pocket of SARS 3CL proteinase resembles that of the TGEV M^{pro}, and the small molecule (Fig 4C), originally an inhibitor of cathepsins B and L, fits well into the two binding pockets in a similar way. This demonstrates again the conservation of the substrate-binding pockets of the two proteinases. Thus, probably some TGEV M^{pro} inhibitors and other cysteine proteinase inhibitors may inhibit the activity of the SARS_CoV 3CL proteinase, functioning as anti-SARS drugs.

Test virtual screening To test the virtual screening models and scoring strategy for discovering inhibitors of the SARS_CoV 3CL proteinase, we constructed a small database containing the inhibitors of different proteinases isolated from the MDDR database (http://www.mdli.com/). Totally, this small database contains 73 molecules. The docking and scoring results against the X-ray crystal structure of TGEV M^{pro} and the 3D model of SARS_CoV 3CL proteinase are listed in Tab 1. All the 73 molecules except molecule 4 can fit into both the binding pockets of TGEV M^{pro} and the 3D model

Tab 1. Scoring data for the test virtual screening targeting to the X-ray crystal structure of TGEV M^{pro} and the 3D model of SARS_CoV 3CL proteinase.

No	TGEV M^{pro}			SARS_CoV 3CL proteinase		
	Dock	CScore	Auto-Dock	Dock	CScore	Auto-Dock
1	-22.82	2	-3.68	-23.67	3	-3.97
2	-25.15	2	-3.21	-25.14	3	-3.36
3	-24.50	2	-3.60	-23.80	3	-3.85
4	-28.79	2	-1.40	-	-	-
5	-23.95	2	-3.69	-23.29	3	-4.12
6	-23.97	2	-2.99	-24.90	3	-3.52
7	-23.20	2	-3.06	-24.16	3	-3.57
8	-22.52	1	-3.29	-22.29	3	-3.06
9	-25.03	3	-3.57	-26.41	3	-3.84
10	-22.10	2	-3.91	-21.97	3	-3.94
11	-25.39	2	-3.10	-25.19	3	-2.79
12	-42.28	3	-2.48	-42.62	5	-3.65
13	-33.02	3	-4.18	-36.33	5	-5.12
14	-30.07	2	-3.47	-24.39	2	-1.98
15	-25.01	2	-1.9	-23.95	2	-1.83
16	-25.28	3	-3.58	-26.84	3	-4.09
17	-15.57	2	-2.04	-20	4	-4.77
18	-25.89	3	-2.62	-27.39	4	-2.74
19	-23.72	2	-1.84	-24.85	3	-2.65
20	-38.26	1	1.76	-44.09	2	1.2
21	-30.94	4	-4.01	-26.95	2	-0.54
22	-35.44	5	-4.48	-28.16	2	-2.77
23	-26.42	4	-5.51	-29.1	4	-7.45
24	-38.39	5	-6.98	-37.29	5	-6.17
25	-34.13	4	-4.88	-22.58	2	-2.21
26	-33.31	3	-2.43	-25.38	3	-0.53
27	-28.79	3	-2.75	-37.73	5	-6.07
28	-19.15	3	-4.09	-20.19	4	-4.16
29	-24.63	3	-2.66	-38.24	4	-7.29
30	-40.87	5	-4.98	-35.96	3	-3.68
31	-36.09	3	-4.18	-26.32	2	-0.21
32	-24.56	1	-0.2	-28.39	2	-1.33
33	-30.64	4	-2.91	-31	3	-2.91
34	-33.14	1	-1.89	-31.26	3	-3.22
35	-37.32	2	-4.18	-35.82	3	-4.98
36	-40.55	2	0.59	-32.78	1	3.77
37	-26.32	3	-3.74	-31.7	4	-5.75
38	-33.72	4	-7.88	-29.98	3	-6.13
39	-33.95	4	-4.3	-28.05	3	-2.45
40	-34.36	5	-6.12	-35.14	4	-6.84
41	-27.93	4	-2.56	-30.18	4	-3.47
42	-29.36	4	-2.97	-38.14	5	-4.92
43	-29.98	4	-5.09	-29.58	3	-4.4
44	-27.81	2	-2.78	-35.13	5	-4.35
45	-41.81	4	-4.87	-26.29	4	-2.95
46	-28.37	5	-4.71	-30.3	4	-5.58
47	-30.76	3	-2.95	-35.44	4	-5
48	-44.71	2	-3.6	-39.83	2	-2.18
49	-28.01	5	-4.09	-29.15	5	-5.42
50	-32.6	3	-1.31	-33.28	2	-2.54
51	-32.68	5	-4.62	-36.28	3	-4.91
52	-21.96	2	-2.49	-27.6	4	-5.09
53	-35.58	4	-4.87	-45.5	4	-5.73
54	-31.57	4	-5.99	-32.23	5	-5.8
55	-37.7	4	-5.24	-40.8	5	-6.8
56	-30.47	4	-5.86	-27.26	4	-4.47
57	-27.86	2	-2.5	-25.52	2	-2.43
58	-32.87	5	-6.22	-34.58	5	-7.66
59	-32.21	4	-3.89	-32.05	4	-4.66
60	-31.03	3	-2.81	-35.34	3	-4.9
61	-26.78	3	-3.2	-42.26	4	-6.35
62	-36.92	5	-5.42	-32.31	5	-3.54
63	-37.23	5	-6.06	-40.6	5	-5.41
64	-28.62	2	-0.59	-24.56	2	2.18
65	-28.57	4	-4.92	-29.38	5	-5.91
66	-25.07	3	-5.99	-25.13	3	-5.13
67	-30.83	4	-4.9	-28.95	4	-5.11
68	-31.9	4	-5.35	-27.72	3	-2.94
69	-27.04	3	-4.92	-23.21	3	-4.29
70	-26.99	3	-3.53	-34.13	4	-6.66
71	-36.08	4	-4.5	-31.77	5	-3.92
72	-28.27	2	-3.25	-26.12	4	-3.99
73	-30.35	4	-4.81	-30.86	5	-5.45

of SARS_CoV 3CL proteinase. Molecule 4 cannot dock into the binding pocket of SARS_CoV 3CL proteinase, however, the scoring of this molecule to the X-ray crystal structure of TGEV M^{pro} is also low. Most of the molecules bind to the two proteinases in a similar way as illustrated in Fig 4. The correlation between the scoring results against the X-ray crystal structure of TGEV M^{pro} and the 3D model of SARS 3CL proteinase is not so good, but show a same tendency. As an example, Fig 5 illustrates the correlation between the scoring results against the two targets resulted from DOCK screening, the correlation coefficient (R=0.65) is at the medium level. This demonstrates again the structural conservation of these two proteinases. The virtual

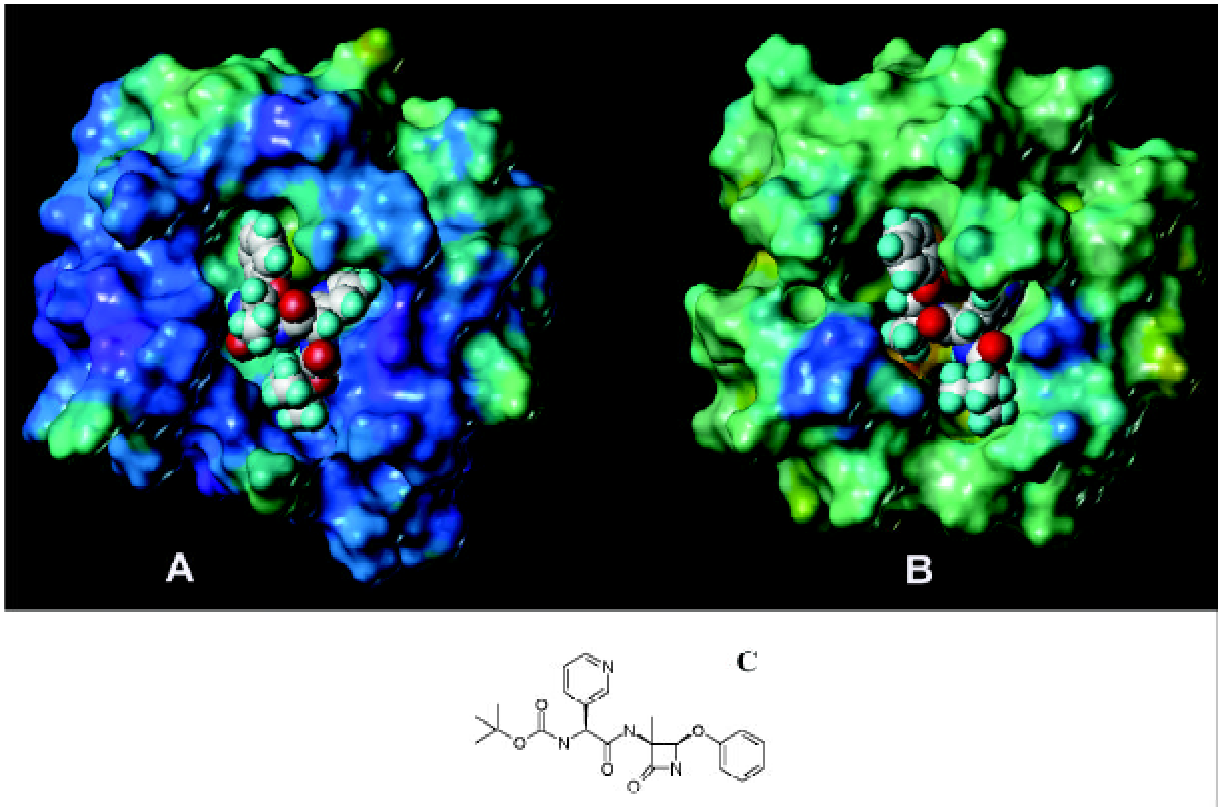

Fig 4. Surface features of the substrate-binding pockets of TGEV M^{pro} (A) and SARS 3CL proteinase (B). The surface color was loaded by the electrostatic properties. One small molecule, its chemical structure is shown in (C), produced by the virtual screening on the MDDR database, represented as CPK model, was docked into the binding pockets.

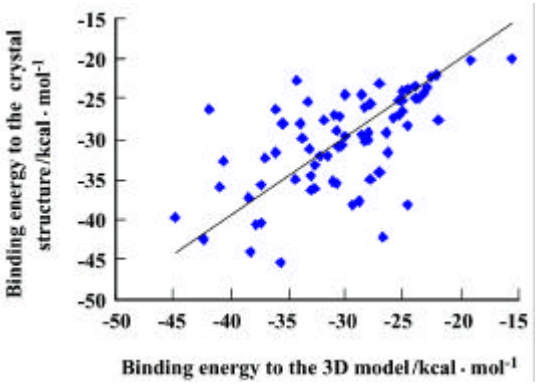

Fig 5. Correlation between the DOCK scoring results against the 3D model of SARS 3CL proteinase and the X-ray crystal structure of TGEV M^{pro}.

screening result implies that screening the inhibitors of available proteinases against the SARS_CoV 3CL proteinase may be an appropriate shortcut to discover anti-SARS drugs.

CONCLUSIONS

Sequence alignment analysis addressed that, among the 3C like proteinases with known X-ray crystal structures, TGEV M^{pro} is the most homologous of the SARS_CoV 3CL proteinase. Therefore, a 3D model of the SARS_CoV 3CL proteinase was constructed based on the crystal structure of TGEV M^{pro}. The modeled structure of the SARS_CoV 3CL proteinase is analogous to the crystal structure of TGEV M^{pro}. For these two proteinases, the substrate-binding pockets adopt a similar shape. Virtual screening indicated that numerous available proteinase inhibitors might bind to both proteinases, suggesting that screening the known proteinase inhibitors may be an appreciated way to discover anti-SARS drugs.

REFERENCES

1 Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science (Sciencexpress). 2003 May 1.

2 Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A. The genome sequence of the SARS-associated coronavirus. Science (Sciencexpress). 2003 May 1.
3 Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R, *et al*. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. EMBO J 2002; 21: 3213-24.
4 InsightII [molecular modeling package], Version 2000. Calif (USA): Molecular Simulations Inc.
5 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL_W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22: 4673-80.
6 Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993; 234: 779-815.
7 Cornell WD, Cieplak DP, Bayly CI, Gould IR, Merz KM, Ferguson DM, *et al*. A Second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995; 117: 5179-97.
8 Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991; 253: 164-70.
9 Sybyl [molecular modeling package], version 6.8. St Louis (MO): Tripos Associates; 2000.
10 Muegge I, Rarey M. Small molecule docking and scoring. In: *Reviews in Computational Chemistry*, Vol. *17*. Lipkowitz KB, Boyd DB, editors. New York: John Wiley & Sons; 2001. p1-60.
11 Ewing T, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 1997; 18: 1175-89.
12 Kuntz ID. Structure-based strategies for drug design and discovery. Science 1992; 257, 1078-82.
13 Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity-A rapid access to atomic charges, Tetrahedron 1980; 36: 3219-28.
14 Marsili M, Gasteriger J. p Charge distribution from molecular topology and p orbital electronegativity. Croat Chem Acta 1980; 53: 601-14.
15 Purcell WP, Singer JA. Brief review and table of semiempirical parameters used in the Hückel molecular orbital method. J Chem Eng Data 1967; 12: 235-46.
16 Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model 2002; 20: 281-95.
17 Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE. Automated docking using Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 1998; 19: 1639-62.
18 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215: 403-10.