# Structure-Based Drug Design and Structural Biology Study of Novel Nonpeptide Inhibitors of Severe Acute Respiratory Syndrome Coronavirus Main Protease

I-Lin Lu,[†,‡] Neeraj Mahindroo,[‡] Po-Huang Liang,[§] Yi-Hui Peng,[||] Chih-Jung Kuo,[§] Keng-Chang Tsai,[‡] Hsing-Pang Hsieh,[‡] Yu-Sheng Chao,[‡] and Su-Ying Wu*[,†,‡]

*Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan, Republic of China, Division of Biotechnology and Pharmaceutical Research, National Health Research Institutes, Miaoli, Taiwan, Republic of China, Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan, Republic of China, and Structural Biology Program, National Tsing Hua University, Hsinchu, Taiwan, Republic of China*

*Received February 22, 2006*

Severe acute respiratory syndrome coronavirus (SARS-CoV) main protease ($M^{pro}$), a protein required for the maturation of SARS-CoV, is vital for its life cycle, making it an attractive target for structure-based drug design of anti-SARS drugs. The structure-based virtual screening of a chemical database containing 58 855 compounds followed by the testing of potential compounds for SARS-CoV $M^{pro}$ inhibition leads to two hit compounds. The core structures of these two hits, defined by the docking study, are used for further analogue search. Twenty-one analogues derived from these two hits exhibited $IC_{50}$ values below 50 $\mu$M, with the most potent one showing 0.3 $\mu$M. Furthermore, the complex structures of two potent inhibitors with SARS-CoV $M^{pro}$ were solved by X-ray crystallography. They bind to the protein in a distinct manner compared to all published SARS-CoV $M^{pro}$ complex structures. They inhibit SARS-CoV $M^{pro}$ activity via intensive H-bond network and hydrophobic interactions, without the formation of a covalent bond. Interestingly, the most potent inhibitor induces protein conformational changes, and the inhibition mechanisms, particularly the disruption of catalytic dyad (His41 and Cys145), are elaborated.

## Introduction

Severe acute respiratory syndrome (SARS), a new respiratory disease caused by a novel coronavirus, SARS coronavirus (SARS-CoV),[1,2] spread rapidly all over the world in 2003 and infected more than 8000 people, resulting in approximately 800 deaths worldwide with mortality rates reaching over 40% in certain populations.[3,4] Developments of drugs and vaccines are vigorously being pursued, but these are still quite far from clinics.

SARS-CoV, an enveloped positive-strand RNA virus from the Coronaviridae family,[5] codes for two very large polyproteins, namely, pp1a (∼450 kDa) and pp1b (∼750 kDa), that mediate all the functions required for viral replication and transcription. To be functional, these polyproteins need to be processed by the 33.8 kDa main protease ($M^{pro}$), also called the 3C-like protease ($3CL^{pro}$).[6] For its important role in SARS-CoV maturation and infection, $M^{pro}$ has been suggested as a promising target for anti-SARS agent design.

The crystal structures of SARS-CoV $M^{pro}$ have been solved recently,[7−11] revealing that it forms a homodimer with three domains in each monomer. The antiparallel $\beta$-barrel structure of domains I and II is similar to other coronavirus proteases and forms a chymotrypsin-like fold responsible for catalytic reactions. The catalytic dyad residues His41 and Cys145 are located at the cleft between domains I and II. The third domain, C-terminal α-helical domain, is very diverse among the picornavirus and coronavirus $M^{pro}$. It has been reported that domain III existed as a stable dimer even at a very low concentration, indicating that this extra domain contributes to the dimerization of SARS-CoV $M^{pro}$ and therefore switches the enzyme from the inactive form (monomer) to active form (dimer).[12] Additionally, the N finger (residues N1∼N7) located in the same area also contributes to the dimerization of the two monomers. Availability of protein structures and the biological characteristics of SARS-CoV $M^{pro}$ provide insights on the substrate binding site, making it an attractive target for structure-based drug design in an effort to discover more potent and specific inhibitors against it.

Inhibitors of SARS-CoV $M^{pro}$ have been identified by various computational methods.[13−19] For examples, Liu et al.[14] and Dooley et al.[15] identified the inhibitors using 3D structure derived from molecular dynamic simulation of SARS-CoV $M^{pro}$ as a virtual screening target structure, while others used the pharmacophore model to predict potential inhibitors.[20,21] The discovery efforts by computer-aided drug design showed only a few cases of SARS-CoV $M^{pro}$ inhibition potency at micromolar range as confirmed by bioassay. These results indicate there is still a vacuum that needs to be filled to find more potent inhibitors against SARS-CoV $M^{pro}$.

Moreover, although a number of nonpeptide inhibitors of SARS-CoV $M^{pro}$ have been discovered, such as bifunctional arylboronic acids,[22] isatin derivatives,[23] polyphenols,[24] etacrynic acid analogues,[25] cinanserin,[26] and other chemically diverse small molecules,[15,27,28] the lack of structure biology information on these compounds and their interactions with SARS-CoV $M^{pro}$ further makes the design more difficult. All the published structures up-to-date are complexed with peptidyl inhibitors through covalent bonding to SARS-CoV $M^{pro}$.[7,8,10,11] Therefore, there is an urgent need to obtain the molecular insight of small molecule compounds to SARS-CoV $M^{pro}$ to design more potent and specific drugs against it.

In this study, we perform the structure-based virtual screening on a chemical database containing 58 855 compounds based on the 3D structure of SARS-CoV $M^{pro}$. Active compounds, selected from virtual screening approach and confirmed by the

* To whom correspondence should be addressed. For H.-P.H.: phone, 886-37-246166, ext 35708; fax, 886-37-586456; e-mail, hphsieh@nhri.org.tw. For S.-Y.W.: phone, 886-37-246166, ext 35713; fax, 886-37-586456; e-mail, suying@nhri.org.tw.

[†] National Defense Medical Center.

[‡] National Health Research Institutes.

[§] Academia Sinica.

[||] National Tsing Hua University.

10.1021/jm060207o CCC: $33.50 
Published on Web 08/01/2006

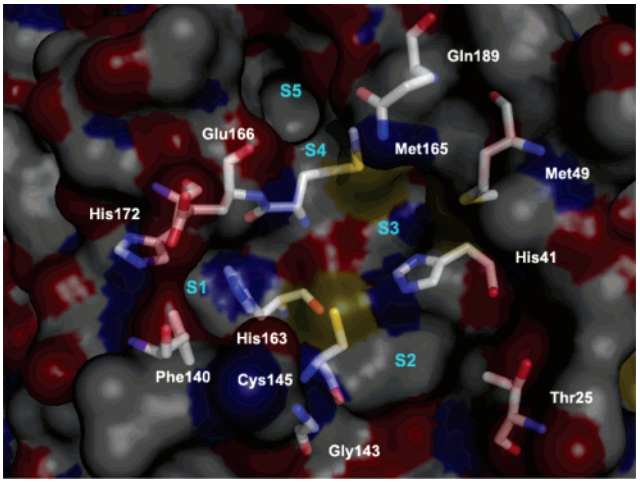



**Figure 1.** SARS-CoV $M^{pro}$ binding site.

bioassay, were taken as the templates to build the core structure for analogue search. The selected 42 analogues were again evaluated in a SARS-CoV $M^{pro}$ inhibition assay. Out of these analogues, 21 compounds showed inhibition activity against SARS-CoV $M^{pro}$ with $IC_{50}$ values less than 50 $\mu$M, with the most potent one showing 0.3 $\mu$M. Finally, the complex structures of potent inhibitors with SARS-CoV $M^{pro}$ were solved by X-ray crystallography to further study the SARS-CoV $M^{pro}$ inhibition mechanisms of these compounds.

## Results and Discussion

**Identification of Novel SARS-CoV $M^{pro}$ Inhibitors by Structure-Based Virtual Screening.** The structure of SARS-CoV $M^{pro}$ in complex with CMK, a substrate analogue (PDB ID 1UK4)[7] was used as the target to perform virtual screening on the Maybridge databases containing 58 855 small molecules. The binding site includes the catalytic center (His41 and Cys145) and several subsites, designated as S1 (His163, Glu166, Cys145 Gly143, His172, and Phe140), S2 (Cys145, His41, and Thr25), S3 (Met165, Met49, and His41), S4 (Met165 and Glu166) and S5(Gln189, Met165, and Glu166) (Figure 1). The program GOLD v2.1 (CCDC Software Limited, Cambridge, U.K.) was used to perform virtual screening. The docked molecules were first ranked by the fitness score of GOLDScore function to select the best pose from the 20 poses generated by GOLD, followed by resorting with the external hydrogen-bond energy term implemented in GOLDScore to rank the binding affinity. As GOLDScore scoring function has been optimized for the prediction of ligand binding positions as suggested by the user manual, it is reasonable to employ GOLDScore to predict the binding pose of the compounds. The best pose of each compound selected by GOLDScore was therefore retained for the further analysis. Since the H-bonding interactions are important for the ligand binding, as revealed by the protease−substrate complex structure, the best conformer of each compound was then further ranked by their H-bonding interactions with the protease. The top 50 compounds ranked by the external hydrogen-bond energy term, a subcomponent in GOLDScore, were then purchased and experimentally evaluated for their ability to inhibit SARS-CoV $M^{pro}$. Of these, two compounds were found to inhibit SARS-CoV $M^{pro}$ more than 50% at 10 $\mu$M (Figure 2). These two compounds, compound **1** [6-methoxy-3-nitro-2-(phenylsulfonyl)pyridine] and **2** (2-({[3-(4-chlorophenyl)-1,2,4-oxadiazol-5-yl]methyl}thio)-4,5-dihydro-1*H*-imidazol-3-ium chloride) (Figure 2), were then subjected to the second round of virtual screening.

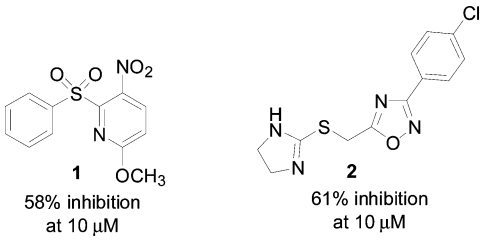



**Figure 2.** Structures of the hit compounds.

This is the first report of using single hydrogen-bond energy to rank and select compounds. This method could be applied to other proteins with H-bond-rich active sites or implemented at different stage of virtual screening to predict the H-bonding interactions with the protein.

**Identification of Core Structure and Analogue Search: (a) Docking Study of Compound 1.** Compound **1** (Figure 2) inhibited SARS-CoV $M^{pro}$ activity by 58% at 10 $\mu$M and was docked into the active site of the protease in the second run of virtual screening. The docking model (Figure 3) proposed that the benzene ring of compound **1** made strong hydrophobic interactions with the catalytic dyad, residues Cys145 and His41. The substituted nitro group of pyridine ring formed three H-bonds with His163, Cys145, and Ser144. In addition, the sulfone group was hydrogen-bonded with Ser144 and Gly143. As revealed in the docking model, the two rings (benzene and pyridine) together with the sulfone moiety made important interactions with the protein and were therefore identified as the scaffold for a further analogue search. Several criteria were applied in the analogue search (Figure 3). The two rings could be individually replaced by six-membered aryl or heteroaryl ring. The sulfone group, which functioned as a linker and made interactions with the protein, was retained in the core structure. To increase the analogue diversity, the substituents on the rings were not limited. A total of 151 compounds that fulfilled the above criteria were selected from the Maybridge database and were then filtered by use of molecular mass ($<$1000 Da) and structure diversity as the screens to remove the large and redundant compounds to a total of 28 compounds. These 28 compounds were then redocked to SARS-CoV $M^{pro}$ to exclude the compounds without important interactions with the protein. His163 and Glu166 are highly conserved residues among coronavirus main proteases, and the specific hydrogen-bond interactions between P1-Gln and these two residues result in the specificity for Gln at the P1 site. Keeping in view the importance and specific characteristics of the S1 site residues (Glu166 and His163) and the catalytic dyad (Cys145 and His41), the compounds without any interactions with Glu166, His163, Cys145, and His41 were withdrawn. Finally, 23 compounds were selected and 21 of them available commercially were purchased. Their ability to inhibit SARS-CoV $M^{pro}$ was evaluated in a bioassay. Out of 21 compounds, 12 compounds showed $IC_{50}$ values less than 50 $\mu$M (Table 1), and compound **3** (Table 1) exhibited the most potent inhibition with an $IC_{50}$ of 0.3 $\mu$M. It displayed a significantly improved potency over the initial hit, compound **1**, which makes it attractive to become a possible drug lead. Therefore, compound **3** was subjected to further characterization by structural biology studies.

**(b) Docking Study of Compound 2.** The other hit compound, **2**, (Figure 2), inhibited SARS-CoV $M^{pro}$ activity by 61% at 10 $\mu$M. The same strategy as described for compound **1** was employed to identify the core structure and search for analogues of compound **2**. The predicted model (Figure 3) showed that the dihydroimidazole ring of **2** fitted into the S2 hydrophobic

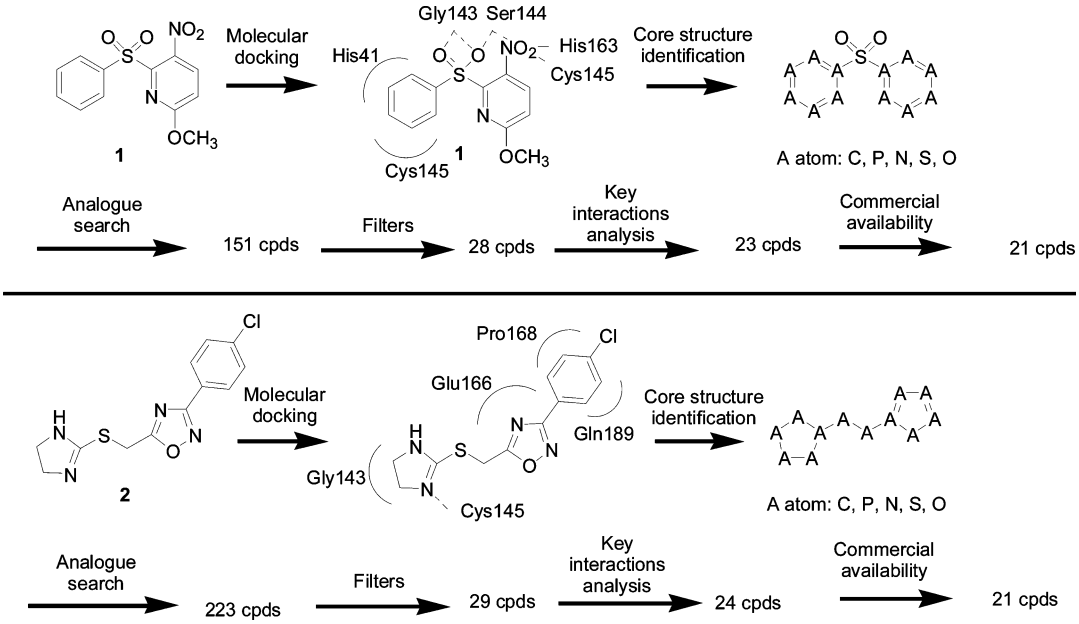



**Figure 3.** Identification of core structure and analogue search. The core structures of two initial hits, compounds **1** and **2**, are defined by docking studies and used for analogue search. Two filters, molecular weight and structure diversity, are applied after analogue search. Finally, the second round of docking study is applied to exclude the compounds without any interaction with the important residues, namely, Glu166, His163, Cys145, and His41.

pocket and the oxadiazole ring was close to the S1 pocket. Moreover, the chlorobenzene occupied the S4 and S5 subsites, which are solvent-accessible. The dihydroimidazole ring formed a H-bond with Cys145 and made close contacts with Gly143. The oxadiazole group formed hydrophobic interactions with Glu166 in the S1 site, while the chlorobenzene group made close contacts with Gln189 and Pro168. As the dihydroimidazole and oxadiazole had interactions with the key residues, Cys145 and Glu166, these two heterocyclic rings together with the linker were identified as the scaffold for the further analogue search. Several criteria were applied in the analogue search (Figure 3). The dihydroimidazole and oxadiazole rings could be individually replaced by a five-membered aryl or heteroaryl ring. The linker between dihydroimidazole and oxadiazole could be replaced by other linkers with a length equal to three C−C bonds to retain the relative position of the two rings. In view of cholorobenzene occupying the less specific S4 and S5 sites and to increase the analogue diversity, this part was kept flexible. A total of 223 compounds that fulfilled the above criteria were selected from the Maybridge database. The same filters and docking study as described for compound **1** were carried out to exclude the compounds with large molecular weight and lack of interactions with the important residues. Twenty-one compounds were finally purchased for the bioassay. Nine out of 21 compounds showed significant inhibition activity against SARS-CoV $M^{pro}$ with $IC_{50}$ values less than 50 $\mu$M. (Table 1). The analogues within this family were diverse. Of the nine active compounds, compound **15** (Table 1) showed potent inhibition with an $IC_{50}$ of 3 $\mu$M and was further studied by X-ray crystallography.

A flowchart representing various stages of the structure-based virtual screening, including the docking study and subsequent analogue search, is shown in Figure 4.

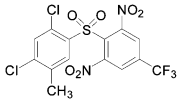
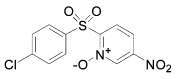
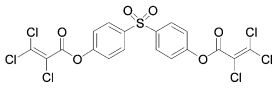
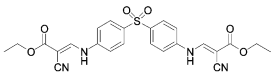
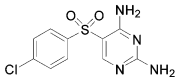
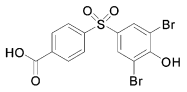
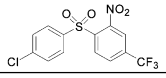
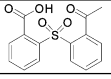
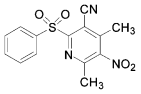
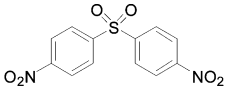
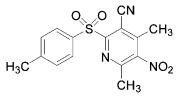
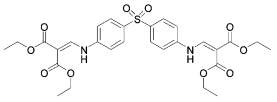
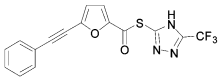
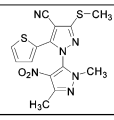
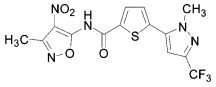
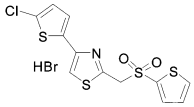
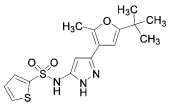
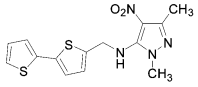
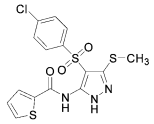
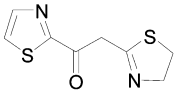
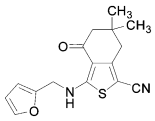
**Overall Structure of SARS-CoV $M^{pro}$.** Structural biology studies were carried out to elucidate the interactions of the potent inhibitors with SARS-CoV $M^{pro}$. The native structure, SARS-CoV $M^{pro}$/**3,** and SARS-CoV $M^{pro}$/**15** were solved to a resolution of 2.17, 1.86, and 1.97 Å, respectively (Table 2). The asymmetric unit contained only one monomer. The electron density maps of all residues of SARS-CoV $M^{pro}$ (residues 1−306) are clear except for the region of residues 45−48, which is flexible in all published structures.

The overall structure of the SARS-CoV $M^{pro}$ structure is very similar to the published structures except for residues 45−48 and the Asn142 residue. The flexible loop of residues 45−48 is located at the entrance of the active site and is flexible in all published structures. Its flexibility could probably allow the access of a ligand to the binding site of SARS-CoV $M^{pro}$. In contrast to the dramatic change in Asn142 upon ligand binding as described by Yang et al.,[8] Asn142 retains the same conformation in our native and complex SARS-CoV $M^{pro}$ structures. The conformation of Asn142 in our structures is the same as in the ligand-binding form described by Yang et al.

**Structure of SARS-CoV $M^{pro}$ in Complex with Compound 3.** As revealed in the crystal structure (Figure 5), **3** adopts a distinct binding mode compared to all the published structures.[7,8,10,11] It occupies the S3∼S5 pockets of SARS-CoV $M^{pro}$. The 2,4-dichloro-5-methylbenzene group inserts deep into the hydrophobic pocket consisting of residues Pro39, His41, Cys145, His163, His164, Phe181, Tyr182, and Phe185. The phenyl ring makes strong $\pi-\pi$ interactions with the side chain of His41, while the substituents, dicholoro and methyl groups, have close contacts with Cys145, His164, Pro39, and Leu27. Moreover, the 1,3-dinitro-5-(trifluoromethyl)benzene group forms intensive H-bonding interactions with the protein. One of the nitro groups forms a direct H-bond with the nitrogen on the side chain of His41 and two indirect H-bonds with Met49 and His41 via water molecule W75. The trifluoromethyl substituent forms a weak H-bond with Gln192 and has close contacts with Gln192, Gln189, Leu167, and Met165. In addition, the benzene group forms hydrophobic interactions with Met165. Moreover, the sulfone group makes H-bonding interactions with water molecule W261.

Upon the binding of **3** with the protein, the side chain of His41, which constitutes an important catalytic dyad with Cys145 of SARS-CoV $M^{pro}$, undergoes a dramatic conformational change. In the SARS-CoV $M^{pro}$, the imidazole group of His41 acts as a proton acceptor to make S$\gamma$ of Cys145 act as a nucleophile. The distance between His41 NE2 and Cys145 S$\gamma$

**Table 1.** Chemical Structures of SARS-CoV $M^{pro}$ Inhibitors with $IC_{50}$

| compd | Structure | $IC_{50}$[a] (μM) | compd | Structure | $IC_{50}$[a] (μM) |
|---|---|---|---|---|---|
| **Compound 1 family** | | | | | |
| **3** |  | 0.3±0.05 | **9** |  | 15±1.2 |
| **4** |  | 0.9±0.1 | **10** |  | 16±0.6 |
| **5** |  | 6±0.7 | **11** |  | 16±0.8 |
| **6** |  | 12±0.8 | **12** |  | 16±1.2 |
| **7** |  | 13±0.6 | **13** |  | 25±1.4 |
| **8** |  | 13±1.0 | **14** |  | 32±2.7 |
| **Compound 2 family** | | | | | |
| **15** |  | 3±0.2 | **20** |  | 18±1.2 |
| **16** |  | 5±0.3 | **21** |  | 18±0.9 |
| **17** |  | 10±0.7 | **22** |  | 20±1.7 |
| **18** |  | 15±1.5 | **23** |  | 40±3.3 |
| **19** |  | 16±2.0 | | | |

[a] Data are shown as mean ± SEM ($n = 3$).

is 3.69 Å in the native structure. Upon the binding of compound **3**, His41 moves away from Cys145 to a distance of 9.07 Å to accommodate the 2,4-dichloro-5-methylbenzene group of **3** sandwiched between Cys145 and His41. The movement of His41 completely blocks the catalytic dyad function and results in the inhibition of SARS-CoV $M^{pro}$ activity, which could provide the structural basis for the inhibition of **3** against the protease. The shift of His41 consequently results in the movement of its adjacent residue, Met49. In addition, Met165 also moves away to accommodate the nitro group on **3**.

The conformational change of the catalytic dyad is also seen in the inhibitor binding structure of caspase 1,[29] where the side chain of His237 is rotated from a +gauche to a trans conformation, creating a large hydrophobic pocket next to the P1 site. The benzene ring of the inhibitor forms strong $\pi-\pi$ interactions with the side chain of His237, leading to the inhibition against the protease.

Analysis of the SARS-CoV $M^{pro}$/**3** structure reveals that one of the nitro groups is close to the side-chain imidazole group of His41. The distance between the oxygen atom of nitro group and NE2 of His41 is about 2.87 Å. The nitro and the histidine imidazole group are both charge-bearing functional groups, as the nitro carries a negative charge and the histidine imidazole group carries a positive charge. The electrostatic interactions between the nitro and the histidine imidazole group are likely the major force responsible for triggering the dramatic conformational change of His41.

**Structure of SARS-CoV $M^{pro}$ in Complex with Compound 15.** The chemical structure of compound **15** can be subdivided into three groups for discussion of its interactions with the

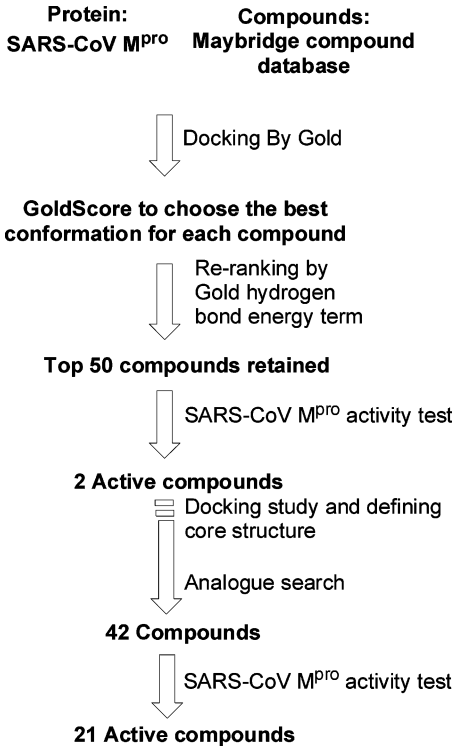



**Figure 4.** Flowchart of structure-based drug design for SARS-CoV M$^{pro}$.

**Table 2.** X-ray Data Collection and Structure Refinement

| | native | 3 | 15 |
|---|---|---|---|
| resolution (Å) | 20−2.17 | 30−1.86 | 30−1.97 |
| unit cell $C2(\alpha = \gamma = 90°)$ | | | |
| *a*, Å | 108.195 | 107.776 | 108.279 |
| *b*, Å | 82.419 | 82.777 | 82.107 |
| *c*, Å | 53.609 | 53.579 | 53.407 |
| $\beta$, deg | 104.98 | 104.931 | 104.66 |
| total reflections observed | 199 770 | 361 184 | 1 269 862 |
| unique reflections | 24 317 | 35 887 | 29 584 |
| multiplicity | 8.215 | 10.06 | 42.9 |
| $R_{merge}$, % (outer shell) | 5.3 (39.7) | 4.7 (51) | 4.5 (43.3) |
| $\langle I/\sigma(I)\rangle$ (outer shell) | 12.7 (1.89) | 23.9 (2.4) | 41.3 (4.5) |
| completeness, % (outer shell) | 97.7 (93.5) | 98.6 (99.8) | 99.1 (99.9) |
| $R_{work}$, % | 20.9 | 20.4 | 21.4 |
| $R_{free}$, % | 24.8 | 23.3 | 24.0 |
| RMS bonds, Å | 0.011 | 0.008 | 0.006 |
| RMS angles, deg | 1.328 | 1.619 | 1.337 |
| average *B* value | | | |
| protein | 35.181 | 30.421 | 40.156 |
| solvent | 41.496 | 41.465 | 48.303 |
| ligand | | 62.606 | 66.079 |

protein (Figure 6). The first group is the triazole group that inserts deep into the S2 pocket, making hydrophobic contacts with Cys145 and Asn142 and H-bonding interactions with the side chains of Cys145 and Asn142. The trifluoromethyl substituent on triazole group makes close contacts with the catalytic dyad residues, Cys145 and His41, and S1 pocket residues, Gly143 and Ser144. The second group is the furan group, which forms hydrophobic interactions with Glu166 and an indirect hydrogen bond via water molecule W16 with the main chain of Glu166. The third group, benzene, extends to SARS-CoV M$^{pro}$ S4 and S5 pockets and makes extensive hydrophobic interactions with the surrounding residues including Met165, Glu166, Gln189, Gln192, and Pro168. The oxygen atom of the carbothioate group, the linker connecting the first and second group, forms a H-bond with the side chain of Asn142 and two indirect H-bonds with the side chain of Glu166 and the main

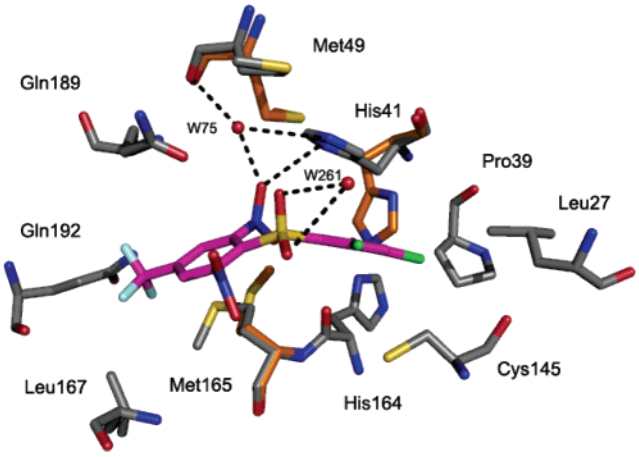



**Figure 5.** Structure of SARS-CoV M$^{pro}$ in complex with compound **3** (pink). The binding of compound **3** to SARS-CoV M$^{pro}$ induces the shift of imidazole group His41 (orange before inhibitor binding, gray after inhibitor binding), resulting in the collapse of catalytic dyad function. The H-bonding interactions are shown as dotted lines.

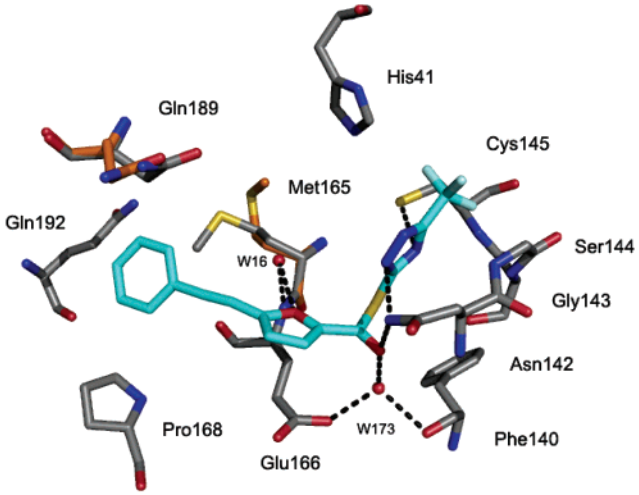



**Figure 6.** Structure of SARS-CoV M$^{pro}$ in complex with compound **15** (cyan). Compound **15** binds to the protein through hydrophobic interactions with the surrounding residues and makes the H-bonding interactions to Cys145 and Asn142. In addition, it forms indirect H-bonds to Glu166 and Phe140 through water molecule. Met165 and Gln189 are shifted to accommodate the benzene group of compound **15** (orange before inhibitor binding, gray after inhibitor binding). The H-bonding interactions are shown as dotted lines.

chain of Phe140 via water molecule W173. Compound **15** retains the important H-bonding interaction with Cys145, similar to the initial hit, **2**. However, the more intensive H-bond network with Asn142, water and Glu166 and additional hydrophobic interactions increase its potency to 3 $\mu$M.

Compared to the native protein structure, SARS-CoV M$^{pro}$ protein residues show no dramatic conformational change upon binding with **15** except for the residues Met165 and Gln189. Gln189 moves close to the benzene ring of **15** and consequently leads to the shift of its neighboring residue, Met165.

**Comparison of 3 and 15 to Other Complex Structures of SARS-CoV M$^{pro}$.** There have been several published structures of SARS-CoV M$^{pro}$ in complex with inhibitors till now.[7,8,10,11] All of them are peptide-like inhibitors bonding covalently with the protease. Compared to these complexed structures, compounds **3** and **15** bind noncovalently to SARS-CoV M$^{pro}$ and inhibit the protease in a distinct manner. To further explore the difference, the structure of SARS-CoV M$^{pro}$ bound with APE (azapeptide epoxides), a substrate-like inhibitor, was superim-

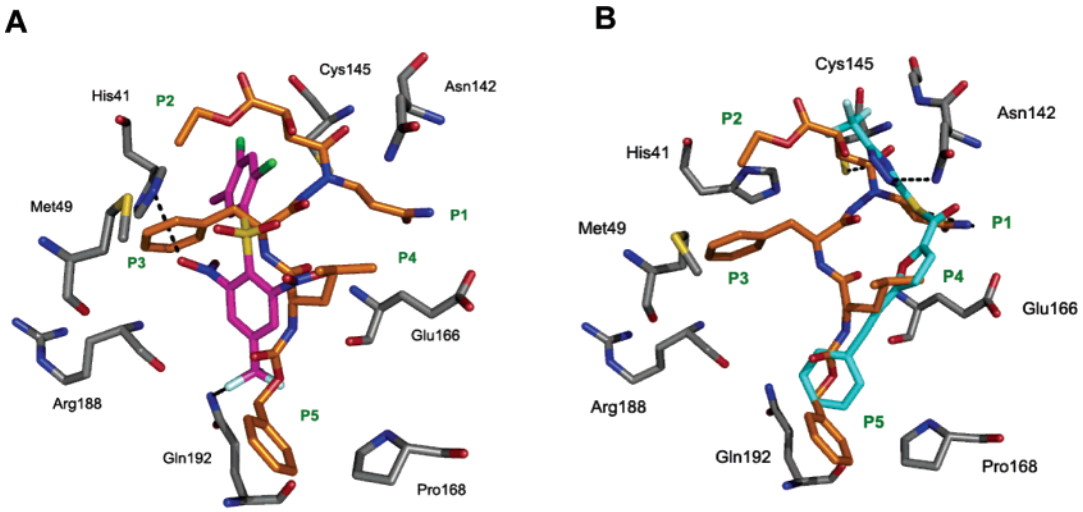



**Figure 7.** (A) Superimposition of the structures of **3** (pink) and APE (orange) in the binding pocket of SARS-CoV $M^{pro}$ (B) Superimposition of the structures of **15** (cyan) and APE (orange) in the binding pocket of SARS-CoV $M^{pro}$.

posed with the complex structures of **3** and **15** individually (Figure 7).

For compound **3** (Figure 7A), it occupies a similar position to P3∼P5 parts of APE. P3∼P5 of APE form hydrophobic interactions with Ala191, Pro168, and Met165 together with the H-bonding interactions with Glu166 and Gln189. These interactions are also observed in the structure of compound **3**, except for the H-bonds to Glu166 and Gln189. The most significant difference between these two structures is that the conformation of His41 in the APE complexed structure remains unchanged, whereas His41 shifts away from Cys145 upon the binding of compound **3**.

Superimposition of the complex structure of compound **15** (Figure 7B) with APE reveals that compound **15** binds to the P1, P2, P4, and P5 sites of APE. In this region, APE makes hydrophobic interactions with Glu166, Met165, Cys145, His41, His163, Pro168, and Gln192 and H-bonding interactions to Glu166, His163, and Gly143. In addition, it forms a covalent bond to Cys145. Although there is no covalent bond formed between compound **15** and Cys145, it does form one H-bond with Cys145 and two H-bonds with Asn142. The residues involved in hydrophobic interactions between compound **15** and protein are Gln189, Pro168, Glu166, Cys145, Gly143, Gln192, Met165, His41, Ser144, and Asn142. These hydrophobic interactions are similar to those of APE.

## Conclusion

In this study, novel nonpeptide inhibitors against SARS-CoV $M^{pro}$ are discovered by structure-based drug design, a combination of virtual screening, docking study, and analogue search. This strategy could successfully identify nonpeptide small molecules with inhibition in the nanomolar range. To our knowledge, compound **3** is the most potent inhibitor of SARS-CoV $M^{pro}$ discovered by the computer-aided drug design method, without chemical synthesis effort involved. Moreover, the structural biology studies reveal that two potent inhibitors, compounds **3** and **15**, adopt distinct binding modes as compared to other published structures. The shift of His41 away from Cys145 as observed in the SARS-CoV $M^{pro}$/**3** structure results in the complete loss of the catalytic dyad function of the protease, providing an insight into the inhibition mechanism against SARS-CoV $M^{pro}$. Moreover, the structure of SARS-CoV $M^{pro}$ in complex with compound **15** shows that the inhibitor forms H-bonding interactions to Cys145 instead of covalent bonding as seen in all published structures. Both binding modes reveal novel inhibition mechanisms for SARS-CoV $M^{pro}$ and could provide a rationale for the next generation of inhibitor design.

## Experimental Section

**Database Preparation.** The Maybridge (58 855 compounds) (Tintagel, Cornwall, U.K.) 2D compound database in SDF format were processed to remove salts and converted to 3D structures by the Insight II program module DB_CONVERT. Protonation states were assumed in the standard setting as suggested by DB_CONVERT. The Extended_Chains and Chair_Confs_Only parameters were set to off and the Rand_Chiral_Centers parameter was set to 0.

**Protein Preparation.** The crystal structure of the SARS-CoV $M^{pro}$ in complex with CMK (PDB code 1UK4) was used. The protonation states of residues were adjusted to the dominant ionic forms at pH 7.5. The bound inhibitor and water were removed in the docking run.

**Docking.** Docking was performed with GOLD version 2.1 (CCDC Software Limited, Cambridge, U.K.). The default parameter settings for library screening were used except the early-termination option was set to off. Residues within a radius of 10 Å around the S$\gamma$ atom of Cys145 were defined as the active site for docking study. Twenty genetic algorithm (GA) runs were carried out for each compound. For each GA run, the selection pressure was set to 1.1, and 100 000 GA operations were performed on a set of five islands with a population size of 100 individuals. The operator weights for crossover, mutation, and migration were set as the default values. Cutoff values of 2.5 Å for hydrogen bonds and 4.0 Å for van der Waals were applied to allow a few bad bumps and poor hydrogen bonds in the beginning of a GA run.

**SARS-CoV Main Protease Inhibition Assay.** SARS-CoV $M^{pro}$ inhibition assay was performed by fluorescence resonance energy transfer (FRET) based on the previous published procedure.[30,31] The gene of SARS-CoV $M^{pro}$ was amplified from whole viral genomic DNA by PCR and cloned into *Escherichia coli* expression vector pET32Xa/LIC. The recombinant protein was expressed in *E. coli* BL21 with a 6×-His tag. The protein was purified by Ni−NTA agarose column (Qiagen, Valencia, CA) and cleaved by FXa protease to remove the His tag. The purified SARS-CoV $M^{pro}$ has authentic sequence without extra amino acids, confirmed by N-terminal sequencing and mass spectrometry. FRET assay was performed at 25 °C in buffer containing 20 mM bis[(2-hydroxyethyl)amino]tris(hydroxymethyl)methane (pH 7.0). The fluorogenic substrate peptide (Dabcyl-KTSAVLQ-SGFRKME-Edans) cleaved by SARS-CoV $M^{pro}$ emitted fluorescence and the enhanced

fluorescence was monitored at 538 nm with excitation at 355 nm by use of a fluorescence plate reader. The $IC_{50}$ value of each inhibitor was measured in a reaction mixture containing 50 nM SARS-CoV $M^{pro}$, 6 $\mu$M fluorogenic substrate, and various concentrations of the inhibitor. The $IC_{50}$ value was obtained by plotting the initial velocities of the inhibited reactions against the different inhibitor concentrations by use of the following equation:

$$A[I] = A[0] \times \{1 - [[I]/([I] + IC_{50})]\}$$

where $A[I]$ is the enzyme activity with inhibitor concentration [I] and $A[0]$ is the enzyme activity without inhibitor.

**Protein Purification, Crystallization, and Structure Determination.** SARS-CoV $M^{pro}$ was expressed in *E. coli* BL21 host cell under the control of T7 promoter. The recombinant protein contained 6×-His tag and was first purified by Ni−NTA column. The His fusion protein was then cleavage by FXa protease to remove the tag and the mixture was loaded onto the second Ni−NTA column to obtain the pure protein. The purity of the protease was >95% pure as checked by SDS−PAGE, and the purified protein was subsequently concentrated for crystallization.

SARS-CoV $M^{pro}$ was crystallized in the absence and presence of the inhibitors. Crystals were grown by mixing 1.5 $\mu$L of protein solution [10.0 mg/mL in a buffer of 12 mM Tris-HCl (pH 7.5), 1 mM DTT, 120 mM NaCl, 0.1 mM EDTA and 7.5 mM $\beta$-mercaptoethanol] with 1.5 $\mu$L of well solution (6% PEG-6000, 2 mM DTT, and 0.1 M Mes, pH 6.0). For compound **3** and **15**, protein solutions were incubated with 2 mM compounds for 2 h on ice in advance. After 3∼7 days at 18 °C, tetragonal crystals grew to an average size of 0.2 mm. The crystals were soaked in a cryoprotectant solution of mother liquor with 20% glycerol for 30 s before being flash-frozen in liquid nitrogen.

Diffraction data were collected at two synchrotron radiation centers. The native crystal diffraction data were collected at NSRRC BL17B beamline. The SARS-CoV $M^{pro}$/**3** and SARS-CoV $M^{pro}$/**15** diffraction data were collected at Spring8 SP12B2 and NSRRC BL13B1 beamlines, respectively. All data were collected on an ADSC Quantum 4R CCD detector at 100 K. All data sets were scaled and integrated by HKL 2000.[32] Molecular replacement was performed by MOLREP[33] to solve the structures by use of the monomer of published SARS-CoV $M^{pro}$ structure (PDB code 1UK4, A chain) as the search model. The structures were then refined by REFMAC,[34] CNS,[35] and SHELX[36] together with several rounds of manual model-building in O.[37] All the figures were drawn by PyMOL (DeLano Scientific LLC, San Francisco, CA). The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 2GZ9, 2GZ7, and 2GZ8 for SARS-CoV $M^{pro}$ native protein, SARS-CoV $M^{pro}$/compound **3**, and SARS-CoV $M^{pro}$/compound **15**, respectively.

**Acknowledgment.** We thank Professor Lindsay Sawyer (Institute of Cell & Molecular Biology, University of Edinburgh, U.K.) and Professor. Malcolm D. Walkinshaw (Institute of Cell & Molecular Biology, University of Edinburgh, U.K.) for their fruitful discussions. We also thank Ms. Hsiao-Wen Edith Chu, Ms. Ashley Chen, Mr. Tai-Tsung Chen, and Ms. Hsiu-Hsiu Huang for their administrative support and the staff at beamlines NSRRC BL17B, SP12B2, and BL13B1 for technical assistance. We thank National Health Research Institutes and National Science Council of the Republic of China (Grants NSC 92-2751-B-400-001-Y and NSC93-2811-B-400-001) for financial support.

**Supporting Information Available:** Electron density maps ($2F_o - F_c$) of native protein, compounds **3** and **15**, and surrounding protein residues. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Drosten, C.; Gunther, S.; Preiser, W.; van der Werf, S.; Brodt, H. R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A.; Berger, A.; Burguiere, A. M.; Cinatl, J.; Eickmann, M.; Escriou, N.; Grywna, K.; Kramme, S.; Manuguerra, J. C.; Muller, S.; Rickerts, V.; Sturmer, M.; Vieth, S.; Klenk, H. D.; Osterhaus, A. D.; Schmitz, H.; Doerr, H. W. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **2003**, *348* (20), 1967−76.

(2) Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W.; Rollin, P. E.; Dowell, S. F.; Ling, A. E.; Humphrey, C. D.; Shieh, W. J.; Guarner, J.; Paddock, C. D.; Rota, P.; Fields, B.; DeRisi, J.; Yang, J. Y.; Cox, N.; Hughes, J. M.; LeDuc, J. W.; Bellini, W. J.; Anderson, L. J. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **2003**, *348* (20), 1953−66.

(3) Nie, Q. H.; Luo, X. D.; Zhang, J. Z.; Su, Q. Current status of severe acute respiratory syndrome in China. *World J. Gastroenterol.* **2003**, *9* (8), 1635−45.

(4) Donnelly, C. A.; Ghani, A. C.; Leung, G. M.; Hedley, A. J.; Fraser, C.; Riley, S.; Abu-Raddad, L. J.; Ho, L. M.; Thach, T. Q.; Chau, P.; Chan, K. P.; Lam, T. H.; Tse, L. Y.; Tsang, T.; Liu, S. H.; Kong, J. H.; Lau, E. M.; Ferguson, N. M.; Anderson, R. M. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* **2003**, *361* (9371), 1761−6.

(5) Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R.; Icenogle, J. P.; Penaranda, S.; Bankamp, B.; Maher, K.; Chen, M. H.; Tong, S.; Tamin, A.; Lowe, L.; Frace, M.; DeRisi, J. L.; Chen, Q.; Wang, D.; Erdman, D. D.; Peret, T. C.; Burns, C.; Ksiazek, T. G.; Rollin, P. E.; Sanchez, A.; Liffick, S.; Holloway, B.; Limor, J.; McCaustland, K.; Olsen-Rasmussen, M.; Fouchier, R.; Gunther, S.; Osterhaus, A. D.; Drosten, C.; Pallansch, M. A.; Anderson, L. J.; Bellini, W. J. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**, *300* (5624), 1394−9.

(6) Gorbalenya, A. E.; Donchenko, A. P.; Blinov, V. M.; Koonin, E. V. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Lett.* **1989**, *243* (2), 103−14.

(7) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (23), 13190−5.

(8) Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; Hilgenfeld, R.; Yuen, K. Y.; Wong, L.; Gao, G.; Chen, S.; Chen, Z.; Ma, D.; Bartlam, M.; Rao, Z. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **2005**, *3* (10), e324: 1742−52.

(9) Hsu, M. F.; Kuo, C. J.; Chang, K. T.; Chang, H. C.; Chou, C. C.; Ko, T. P.; Shr, H. L.; Chang, G. G.; Wang, A. H.; Liang, P. H. Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **2005**, *280* (35), 31257−66.

(10) Lee, T. W.; Cherney, M. M.; Huitema, C.; Liu, J.; James, K. E.; Powers, J. C.; Eltis, L. D.; James, M. N. Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.* **2005**, *353* (5), 1137−51.

(11) Ghosh, A. K.; Xi, K.; Ratia, K.; Santarsiero, B. D.; Fu, W.; Harcourt, B. H.; Rota, P. A.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *J. Med. Chem.* **2005**, *48* (22), 6767−71.

(12) Shi, J.; Wei, Z.; Song, J. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* **2004**, *279* (23), 24765−73.

(13) Jenwitheesuk, E.; Samudrala, R. Identifying inhibitors of the SARS coronavirus proteinase. *Bioorg. Med. Chem. Lett.* **2003**, *13* (22), 3989−92.

(14) Liu, Z.; Huang, C.; Fan, K.; Wei, P.; Chen, H.; Liu, S.; Pei, J.; Shi, L.; Li, B.; Yang, K.; Liu, Y.; Lai, L. Virtual screening of novel noncovalent inhibitors for SARS-CoV 3C-like proteinase. *J. Chem. Inf. Model.* **2005**, *45* (1), 10−17.

(15) Dooley, A. J.; Shindo, N.; Taggart, B.; Park, J. G.; Pang, Y. P. From genome to drug lead: Identification of a small-molecule inhibitor of the SARS virus. *Bioorg. Med. Chem. Lett.* **2006**, *16* (4), 830−3.

(16) Rajnarayanan, R. V.; Dakshanamurthy, S.; Pattabiraman, N. "Teaching old drugs to kill new bugs": structure-based discovery of anti-SARS drugs. *Biochem. Biophys. Res. Commun.* **2004**, *321* (2), 370−8.
(17) Du, Q. S.; Wang, S. Q.; Zhu, Y.; Wei, D. Q.; Guo, H.; Sirois, S.; Chou, K. C. Polyprotein cleavage mechanism of SARS CoV Mpro and chemical modification of the octapeptide. *Peptides* **2004**, *25* (11), 1857−64.
(18) Toney, J. H.; Navas-Martin, S.; Weiss, S. R.; Koeller, A. Sabadinine: a potential non-peptide anti-severe acute-respiratory-syndrome agent identified using structure-aided design. *J. Med. Chem.* **2004**, *47* (5), 1079−80.
(19) Zhang, X. W.; Yap, Y. L. Old drugs as lead compounds for a new disease? Binding analysis of SARS coronavirus main proteinase with HIV, psychotic and parasite drugs. *Bioorg. Med. Chem.* **2004**, *12* (10), 2517−21.
(20) Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1111−22.
(21) Xiong, B.; Gui, C. S.; Xu, X. Y.; Luo, C.; Chen, J.; Luo, H. B.; Chen, L. L.; Li, G. W.; Sun, T.; Yu, C. Y.; Yue, L. D.; Duan, W. H.; Shen, J. K.; Qin, L.; Shi, T. L.; Li, Y. X.; Chen, K. X.; Luo, X. M.; Shen, X.; Shen, J. H.; Jiang, H. L. A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening. *Acta Pharmacol. Sin.* **2003**, *24* (6), 497−504.
(22) Bacha, U.; Barrila, J.; Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E. Identification of novel inhibitors of the SARS coronavirus main protease $3CL^{pro}$. *Biochemistry* **2004**, *43* (17), 4906−12.
(23) Chen, L. R.; Wang, Y. C.; Lin, Y. W.; Chou, S. Y.; Chen, S. F.; Liu, L. T.; Wu, Y. T.; Kuo, C. J.; Chen, T. S.; Juang, S. H. Synthesis and Evaluation of Isatin Derivatives as Effective SARS Coronavirus 3CL Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15* (12), 3058−62.
(24) Chen, C. N.; Lin, C. P.; Huang, K. K.; Chen, W. C.; Hsieh, H. P.; Liang, P. H.; Hsu, J. T. Inhibition of SARS-CoV 3C-like Protease Activity by Theaflavin-3,3′-digallate (TF3). *Evid. Based Complement. Altern. Med.* **2005**, *2* (2), 209−215.
(25) Kaeppler, U.; Stiefl, N.; Schiller, M.; Vicik, R.; Breuning, A.; Schmitz, W.; Rupprecht, D.; Schmuck, C.; Baumann, K.; Ziebuhr, J.; Schirmeister, T. A New Lead for Nonpeptidic Active-Site-Directed Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus Main Protease Discovered by a Combination of Screening and Docking Methods. *J. Med. Chem.* **2005**, *48* (22), 6832−42.
(26) Chen, L.; Gui, C.; Luo, X.; Yang, Q.; Gunther, S.; Scandella, E.; Drosten, C.; Bai, D.; He, X.; Ludewig, B.; Chen, J.; Luo, H.; Yang, Y.; Yang, Y.; Zou, J.; Thiel, V.; Chen, K.; Shen, J.; Shen, X.; Jiang, H. Cinanserin Is an Inhibitor of the 3C-Like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and Strongly Reduces Virus Replication In Vitro. *J. Virol.* **2005**, *79* (11), 7095−103.
(27) Kao, R. Y.; Tsui, W. H.; Lee, T. S.; Tanner, J. A.; Watt, R. M.; Huang, J. D.; Hu, L.; Chen, G.; Chen, Z.; Zhang, L.; He, T.; Chan, K. H.; Tse, H.; To, A. P.; Ng, L. W.; Wong, B. C.; Tsoi, H. W.; Yang, D.; Ho, D. D.; Yuen, K. Y. Identification of Novel Small-Molecule Inhibitors of Severe Acute Respiratory Syndrome-Associated Coronavirus by Chemical Genetics. *Chem. Biol.* **2004**, *11* (9), 1293−9.
(28) Tsai, K. C.; Chen, S. Y.; Liang, P. H.; Lu, I. L.; Mahindroo, N.; Hsieh, H. P.; Chao, Y. S.; Liu, L.; Liu, D.; Lien, W.; Lin, T. H.; Wu, S. Y. Discovery of a Novel Family of SARS-CoV Protease Inhibitors by Virtual Screening and 3D-QSAR Studies. *J. Med. Chem.* **2006**, *49* (12), 3485−95.
(29) Shahripour, A. B.; Plummer, M. S.; Lunney, E. A.; Albrecht, H. P.; Hays, S. J.; Kostlan, C. R.; Sawyer, T. K.; Walker, N. P. C.; Brady, K. D.; Allen, H. J. Structure-Based design of nonpeptide inhibitors of interleukin-1[beta] converting enzyme (ICE, Caspase-1). *Bioorg. Med. Chem.* **2002**, *10* (1), 31−40.
(30) Wu, C. Y.; Jan, J. T.; Ma, S. H.; Kuo, C. J.; Juan, H. F.; Cheng, Y. S.; Hsu, H. H.; Huang, H. C.; Wu, D.; Brik, A.; Liang, F. S.; Liu, R. S.; Fang, J. M.; Chen, S. T.; Liang, P. H.; Wong, C. H. Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (27), 10012−7.
(31) Kuo, C. J.; Chi, Y. H.; Hsu, J. T.; Liang, P. H. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* **2004**, *318* (4), 862−7.
(32) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307−326.
(33) Vagin, A.; Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **1997**, *30* (6), 1022−1025.
(34) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D: Biol. Crystallogr.* **1997**, *53* (Pt. 3), 240−55.
(35) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D: Biol. Crystallogr.* **1998**, *54* (Pt. 5), 905−21.
(36) Sheldrick, G. M.; Schneider, T. R. SHELXL: High-resolution refinement. *Methods Enzymol.* **1997**, *277*, 319−343.
(37) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **1991**, *47* (Pt. 2), 110−9.

JM060207O