Progress in Computational Approach to Drug Development Against SARS

Kuo-Chen Chou*,1,2, Dong-Qing Wei1,2, Qi-Shi Du1,3, Suzanne Sirois1,4 and Wei-Zhu Zhong1

1Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA
2College of Life Science and Technology, Shanghai Jiaotong University, 800 Donglin Road, Shanghai, 200240, China
3Institute of Bioinformatics and Drug Discoveries, Tianjin Normal University, Tianjin 300074, China
4Université du Québec à Montréal (UQAM), Chemistry Department, C.P., 8888 Succursale Centre-Ville, Montréal, Québec, Canada, H3C 3P8

Abstract: Since the outbreak of SARS (severe acute respiratory syndrome) in November 2002 in Southern China’s Guangdong Province, considerable progress has been made in the development of drugs for SARS therapy. The present mini review is focused on the area of computer-aided drug discovery, i.e., the advances achieved mainly from the approaches of structural bioinformatics, pharmacophore modeling, molecular docking, peptide-cleavage site prediction, and other computational means. It is highlighted that the compounds C28H34O4N7Cl, C21H36O5N6 and C21H36O5N6 (Wei et al., Amino Acids, 2006, 31: 73-80), as well as KZ7088 (Chou et al. Biochem. Biophys. Res. Commun., 2003, 308: 148-151), a derivative of AG7088, might be the promising candidates for further investigation, and that the octapeptides ATLQAIA and ATLQAENV, as well as AVLQSGF, might be converted to effective inhibitors against the SARS enzyme. Meanwhile, how to modify these octapeptides based on the “distorted key” theory to make them become potent inhibitors is explicitly elucidated. Finally, a brief introduction is given for how to use computer-generated graphs to rapidly diagnose SARS coronavirus.

Key words: SARS, Coronavirus proteinase, KZ7088, AG7088, Binding pocket, Octapeptide substrate, “distorted key” mechanism.

INTRODUCTION

SARS (severe acute respiratory syndrome) is a respiratory illness. Reported first in Asia in November 2002, the illness spread to more than two dozen countries in North America, South America, Europe, and Asia within only a few months. Patients suffering from SARS usually begin having a high fever (>38 °C or 100.4 °F) with symptoms such as headache, malaise, chilly, rigor, diarrhea, and body aches, followed by developing a dry (non-productive) cough and having trouble breathing that might be accompanied by or progress to hypoxia, a condition in which there is insufficient oxygen reaching body tissues. Most patients developed pneumonia with a fatality rate around 15%

SARS is caused by a previously unrecognized coronavirus, called SARS-coronavirus. The virus is mainly spread by close person-to-person contact, by direct contact with infectious materials, and by respiratory droplets produced when an infected person coughs or sneezes. In addition to the “person-to-person contact spread”, “infectious materials spread”, and “droplet spread”, SARS-coronavirus might also be spread through the air, the so-called “airborne spread” or by other ways that are not quite known yet. Although the SARS global outbreak of 2003 was contained, it is possible that the disease could re-emerge, causing even greater disaster because the viruses might occur in many different mutated forms.

Threatened by such a severe disease, many efforts have been made by scientists from various areas in order to provide useful knowledge and technology for helping conducting rational drug design and finding effective drugs against SARS. Considerable progresses have been achieved recently in this regard. Progress in synthesis of novel test compounds for antiviral chemotherapy of SARS has been summarized by Kesel [1], and that in drug discovery against SARS-CoV reported in a recent review [2]. This mini review will focus on the progress mainly from the approaches of computer-aided drug discovery.

BINDING INTERACTIONS OF SARS-CORONAVIRUS PROTEASE WITH ITS LIGANDS

Many evidences indicate that the SARS-coronavirus exists in SARS patients, suggesting that the virus is the culprit of SARS. It is also known that the process of cleaving the SARS-coronavirus polyproteins by a special proteinase, the so-called SARS coronavirus main proteinase (SARS CoV Mpro), is a key step for the replication of the “culprit”. The functional importance of the protease in the viral life cycle has made it an attractive target for developing drugs against SARS. To conduct the rational, or structure-based, drug design, a key step is to understand the binding interaction of SARS-CoV Mpro with its ligands.

Based on the atomic coordinates of SARS-CoV Mpro [3], two enzyme-ligand complexes were developed [4]. One is the complex by docking a compound called KZ7088 (Fig. 1a) to SARS-CoV Mpro, and the other is that by docking the octapeptide AVLQSGF to the same enzyme.

KZ7088 [4] is a derivative of AG7088 (Fig. 1b). The latter was developed by Pfizer Inc. and is currently in clinical
trials for the treatment of rhinovirus, a pathogen that can cause the common cold. As shown in Fig. 1, AG7088 has a $p$-fluorophenylalanine side chain ($p$-fluorobenzyl), which is too long (or bulky) to fit into the binding pocket of SARS-CoV M$^{pro}$. Accordingly, KZ7088 with a modified side chain by removing -CH$_2$ could well fit into the binding pocket, as shown in Fig. 2, where the SARS-coronavirus main proteinase is in ribbon drawing, the KZ7088 in ball-and-stick drawing, and the white mesh represents the binding pocket of SARS-CoV M$^{pro}$ for KZ7088. The constituents of the binding pocket are defined by those residues that have at least one heavy atom (i.e., other than hydrogen) with a distance $\leq$ 5 Å from a heavy atom of KZ7088. A similar binding pocket was defined for ATP in the Cdk5-Nck5a*-ATP complex [5] that has later proved quite useful in identifying the functional domains as well as stimulating the relevant truncation experiments [6] and other studies (see, e.g., [7-10]). The binding pocket involved 23 residues, and the ligand was tethered to the enzyme by six hydrogen bonds, as detailed in [4]. A series of follow-up discussions about KZ7088 can be found in some recent papers (see, e.g., [2,11-13]).

The binding interaction derived by docking the octapeptide AVLQSGFR to SARS-CoV M$^{pro}$ is shown in Fig. 3, where the SARS-coronavirus main proteinase is in ribbon drawing, and the octapeptide is in ball-and-stick drawing. The octapeptide was tethered to Arg-40, His-41, Phe-185, Asp-187, and Gln-189 of the enzyme by six hydrogen bonds. The binding interaction mode had the important implications in stimulating rationally designing drugs for SARS therapy due to the following considerations. (1) The protease-susceptible sites in proteins usually extend to an octapeptide,
as generally formulated by $\text{PPPPPPP}$, with the scissile bond located between the subsites $P_1$ and $P_3$, as generally expressed by $P_1 \downarrow P_3$. (2) The SARS coronavirus enzyme and several viral proteinases exhibit Gln $\downarrow$ (Ser, Ala, Glv) specificity [17]. (3) According to the “lock-and-key” mechanism in enzymology, the octapeptide cleavable by the SARS protease must have a good fit for binding to the active site. However, such a peptide, after a modification of its scissile bond with some simple routine procedure, will completely lose its cleavability but it can still bind to the active site. Actually, the molecule thus modified can be compared to a “distorted key” [18], which can be inserted into a lock but can neither open the lock nor be pulled out from it, spontaneously becoming an ideal competitive inhibitor against the SARS protease.

Stimulated by the above binding interaction modes, a series of follow-up studies were carried out, as described below.

PHARMACOPHORE MODELING AND VIRTUAL SCREENING

Pharmacophore modeling can provide valuable insight into ligand-receptor interactions. It can also be used in 3D (dimensional) database searching for finding potentially biologically active compounds and providing new research ideas and directions for drug-discovery projects. The term pharmacophore was originally defined as the 3D arrangement of atoms - or groups of atoms - responsible for the biological activity of a drug molecule. Pharmacophore models are constructed based on compounds of known biological activity and are refined as more data are acquired in an iterative process. The models can be used for optimizing a series of known ligands or, alternatively, they can be used to search molecular databases in order to find new structural categories, to wit: a process known as virtual screening.

Based on the hydrogen bonding interactions between KZ7088 and SARS-CoV M$^{\text{pro}}$ [4], a template of pharmacophore points of KZ7088 was generated [19] as shown in Fig. 4, where the four H-b donors are colored in cyan, the two H-b donors in purple, the aromatic group in green, and the volume constrains in grey. With the pharmacophore template, the virtual screening search operation was carried out by Sirois et al. [19] for both commercial and academic available compounds. The software MOE (Group, 2002) from Chemical Computing Group was used to conduct pharmacophore searches (virtual screening) of potential hits. It was found that, of the 3.6 millions of compounds screened, 0.07% were with the score satisfying five of the six pharmacophore points as defined by the four H-bond acceptors and two H-bond donors (Fig. 4). Moreover, each of the hit compounds was further evaluated for its druggability according to a novel score function, an equation formulated by 13 metrics that took into account physical [20], chemical and structural properties [21], as well the presence of undesirable functional groups [22,23]. The selection of these properties was based on the work proposed by Baurin [24]. After such a druggability-evaluating procedure, it was found that 17% of the compound thus retrieved from the first hit screen had a perfect score of 1.0, 23% with one violation of druggable rule, 13% with two violations, and 47% with more than two violations. If the criterion for druggability was set at a maximum allowance of two rule violations, we obtained that only 0.0037% of the total compounds screened are worthy of further tests for their activities. These findings would significantly narrow down the search scope for potential compounds, saving substantial time and money. Finally, the featured templates derived from the pharmacophore study would also be very useful for guiding the design and synthesis of effective drugs for SARS therapy.

Although the hydrogen bond donor and acceptor were chosen for pharmacophore modeling in [19], one can also choose some other features for modeling, such as aromatic ring, hydrophobic aromatic, hydrophobic aliphatic, positive charge, negative charge, hydrogen bond acceptor lipid, positive ionizable, and negative ionizable.

The pharmacophore approach bears the following advantages: (1) avoiding doing docking study one by one for thousands of different compounds that would otherwise be both time-consuming and costly; (2) providing useful information of the chemical functionality and orientation to help make better choice for synthesizing compounds; (3) gaining insight into common chemical features that are considered responsible for ligand-receptor binding affinity; (4) significantly reducing the scope of compounds for further studies; and (5) developing useful featured templates for faster lead finding and optimization.

ANTI-SARS DRUG SCREENING BY MOLECULAR DOCKING

From the 2,589 unique 3D hits with volume constrains found by Sirois et al. [19], 1,386 druggable compounds were collected by Wei et al. [25] that both match the 3D KZ7088 pharmacophore points and satisfy the druggable rules with a scoring value $\geq 0.8$. To further narrow down the investigation scope for molecular docking, a similarity search of the 1,386 compounds thus obtained were performed by using KZ7088 [4,26] as the template molecule. It was found thru a series of auto docking procedures that three compounds, i.e., C$_{26}$H$_{33}$O$_2$N$_2$Cl, C$_{26}$H$_{36}$O$_2$N$_6$, and C$_{21}$H$_{30}$O$_3$N$_6$ [25], were the most promising candidates for further investigation.

It has been derived through a thermodynamic analysis for the above docking studies that the hydrogen bonds make an important contribution to the interactions between these compounds with the SARS enzyme, which is quite in consistence with the case of KZ7088 [4]. Besides hydrogen bonds, there are some other contributions to the interactions, such as desolvation, hydrophobic interaction, and electrostatic interaction. The hydrogen bond may play the role as an “anchor”, determining the spatial position of the compounds in the binding pocket, thus facilitating the hydrophobic interactions and electrostatic interactions. Accordingly, in conducting rational drug design it is critic to identify the hydrophobic groups of the ligand and receptor and ensure that they are facing to each other upon binding. Also, it was observed thru examining the Connolly surface [27] of the SARS-CoV M$^{\text{pro}}$ with these compounds that the steric complementarity is responsible for the micro-mechanical interlock at molecular level. In other words, it is the steric complementarity between the ligand and the collagen receptor site that plays the role of the primary driving force for mechanical interlocking.
The main features shared by the aforementioned three potential inhibitors as well as the information of the involved side chains of SARS-CoV M\textsuperscript{pro} may provide useful insights for the development of potent inhibitors against SARS enzyme.

**PEPTIDE INHIBITORS**

The development of peptides as clinically useful drugs is limited by their poor metabolic stability and low bioavailability, which is partially due to their inability to readily crossing thru membrane barriers such as the intestinal and blood-brain barriers. On the other hand, however, peptide drugs are of low toxicity to human body than organic compounds, and hence systematic chemical modification strategies that convert peptides into drugs are an attractive research topic in current medicinal chemistry [28]. Some efforts have been made in an attempt to develop peptide inhibitors against the SARS-CoV M\textsuperscript{pro} [4,29-33]. The development of peptide inhibitors against proteases was based on the “distorted key theory” [18], as can be illustrated as follows.

According to the “lock-and-key” mechanism in enzymology, a protease-cleavable peptide must satisfy the substrate specificity, i.e. a good match for binding to the active site. Here, the phrase of “good match” should be understood in a broad sense rather than a narrow geometric sense; i.e., it means a favorable chemical-group-disposition for the binding of a substrate to the active site of an enzyme and the catalytic reaction thereof. However, such a peptide, after a modification of its scissile bond with some simple chemical routine procedure, will completely lose its cleavability but it can still tightly bind to the active site of an enzyme. In view of this, we can liken the derivative molecule thus obtained to a “distorted key”, which can be inserted into a lock but can neither open the lock nor be pulled out from it. That is why a molecule modified from a cleavable peptide can spontaneously become a competitive inhibitor against the enzyme. An illustration about using the concept of “distorted key” to find peptide inhibitor for the SARS enzyme is given in Fig. 5, where panel (a) shows an effective binding of a cleavable peptide to the active site of SARS CoV M\textsuperscript{pro}, while panel (b) shows that the peptide has become a non-cleavable one after

**Fig. (4).** A template of the pharmacophore points of KZ7088 derived from its binding interaction with SARS-CoV M\textsuperscript{pro} [4]. The four H-b donors are colored in cyan, the two H-b donors in purple, the aromatic group in green, and the volume constrains in grey. Reproduced from Sirois et al. [19] with permission. (For interpretation of the references to color of the figure, the reader is referred to the web version of this paper).
its scissile bond is modified although it can still bind to the active site. Such a modified peptide, or “distorted key”, will automatically become an inhibitor candidate against SARS CoV M\textsuperscript{pro}. It is instructive to point out that, even for developing non-peptide inhibitors, the knowledge acquired thru the above process can also provide useful insights about the key binding groups, proper microenvironment, fitting conformation, among many other subtle requirements. Therefore, it will significantly reduce the search scope and expedite the process for finding the desired peptide inhibitors if a computational method is available by which one can discern what kind of peptides can be cleaved by SARS CoV M\textsuperscript{pro}. That is why during the process of developing inhibitors against HIV protease many bioinformatics algorithms were proposed for predicting the cleavable peptides by HIV protease [16,34-40], as summarized in a review paper [18].

For the SARS enzyme case, the similar bioinformatics tools were also developed [41,42]. Du \textit{et al.} [31] used these tools to search 36 SARS coronaviruses and found 396 cleavage sites, from which 11 cleavable octapeptide were extracted. It was found via a statistical analysis that two octapeptides, NH\textsubscript{2}-ATLQ\textsubscript{4}AIAS-COOH and NH\textsubscript{2}-ATLQ\textsubscript{4}AE NV-COOH, were singled out as the most promising candidates converting into the inhibitors of SARS-CoV M\textsuperscript{pro}. Actually, the two octapeptides are very similar to NH\textsubscript{2}-AVLQ\textsubscript{4}SGFR-COOH according to their pharmacophores [32]. The chemical modification was focused on the scissile peptide bond between R\textsubscript{4} and R\textsubscript{1} (Fig. 5). The quantum chemical study showed that after the peptide bond CO=NH between R\textsubscript{4} and R\textsubscript{1} was replaced by a single bond, such as CH\textsubscript{2}-NH, CF\textsubscript{2}-NH, or CO-CH\textsubscript{2}, the cleavage of such a modified octapeptide by SARS-CoV M\textsuperscript{pro} was very difficult [29]. This is supported by the fact that, after the peptide bond CO=NH of an octapeptide cleavable by renin (also known as angiotensinogenase, a circulating enzyme released mainly by juxtaglomerular cells of the kidneys in response to low blood volume or low body NaCl content) was replaced by a single bond CH\textsubscript{2}-NH, the octapeptide became strongly resistant to the enzyme hydrolysis although its affinity to the enzyme was increased [43,44].

Meanwhile, AVLQSGFR, the first octapeptide introduced in this area by Chou \textit{et al.} [4], was experimentally demonstrated to be cleavable by the SARS enzyme with a high bioactivity by Gan \textit{et al.} [45]. It has been observed that the hydrophilic and hydrophobic complimentary interaction is important in the ligand-receptor interaction [30]. Shown in Fig. 6\textit{a,b} is the hydrophilic surface (blue) and hydrophobic surface (green) of octapeptide AVLQSGFR and the active site region of SARS CoV M\textsuperscript{pro}, respectively. There is a good match between the two; i.e., the hydrophilic parts of the ligand match with the hydrophilic parts of the receptor, and the hydrophobic parts of the ligand with the hydrophobic
parts of the receptor. The octapeptide NH$_2$-AVLQ$_1$SGFR-COOH has four hydrophobic amino acid residues: R$_4$(Ala), R$_3$(Val), R$_3'$(Phe), and R$_4'$(Arg). Among them R$_4$(Ala) and R$_3$(Val) are exposed in solvent, R$_3'$(Phe) and R$_4'$(Arg) are covered by hydrophobic surfaces of the proteinase. The peptide bond to be cleaved between R$_1$(Gln) and R$_1'$(Ser) is just at the join point between the exposed part and the covered part. Furthermore, based on the crystal structure of SARS-CoV M$^\text{pro}$ [46], the cleavage mechanism of the SARS-CoV M$^\text{pro}$ on the octapeptide NH$_2$-AVLQ$_1$SGFR-COOH [4] was investigated using molecular mechanics (MM) and quantum mechanics (QM) [30]. It has been observed that the catalytic dyad (His-41/Cys-145) site between domain I and II of the protease [3,4] attracts $\pi$-electron density from the peptide bond Gln-Ser, increasing the positive charge on C(CO) of Gln and the negative charge on N(NH) of Ser, so as to weaken the Gln-Ser peptide bond. The catalytic functional group is the imidazole group of His-41 and the S in Cys-145. N$_3$ on imidazole ring plays the acid-base catalytic role. It has also been found that the chemical bond between Gln and Ser will become much stronger and no longer cleavable by the SARS enzyme after either changing the carbonyl group CO of Gln to CH$_2$ or CF$_2$, or changing the NH of Ser to CH$_2$ or CF$_2$. According to the “distorted key theory” [18], the octapeptide thus modified might become an effective inhibitor or a potential drug candidate against SARS.

**DETECTION OF SARS-CORONAVIRUS**

It has been revealed [47] by conducting an extensive analysis of the 2-dimensional cellular automata images [48,49] for 96 SARS coronavirus and 25 non-SARS coronavirus that all the images of the former were with a V-shaped cross-lines pattern (Fig. 7a) but all the images of the latter are with a parallel slash-lines pattern (Fig. 7b), a remarkable distinction between the SARS and non-SARS coronal viruses. Therefore, the V-shaped cross-lines image pattern can be deemed as a token for SARS coronal viruses and used to rapidly diagnose SARS coronavirus for both basic research in laboratories and practical application in clinics.

Also, an *in silico* method was developed for visualizing the characteristic of SARS-CoV is suggested [50]. The characteristic thus revealed is unique for SARS-CoV, and can be regarded as the fingerprint map of SARS-CoV for diagnostic usage. The fingerprint map method has the merits of clear visibility and reliability that can serve as a complementary clinical tool for detecting SARS-CoV, particularly for the cases where the results obtained by the conventional methods are uncertain or conflicted with each other.

**CONCLUSION**

KZ7088 [4], C$_{28}$H$_{34}$O$_4$N$_7$Cl, C$_{21}$H$_{36}$O$_5$N$_6$ and C$_{21}$H$_{36}$O$_5$N$_6$ [25] are the four promising compounds that might become drug candidates for SARS therapy. AVLQSGFR, ATLQAIS and ATLQAENV are the three octapeptides that can be converted into potent inhibitors against the SARS enzyme. These findings were acquired thru a series of studies by combining the approaches of structural bioinformatics, pharmacophore modeling, virtual screening, molecular docking, peptide-cleavage site prediction, and the “distorted-key” theory [18]. Some of the findings have already been experimentally proved (see, e.g., [45]).

**ABBREVIATIONS**

SARS = Severe acute respiratory syndrome  
CoV = Coronavirus  
M$^\text{pro}$ = Main proteinase

---

**Fig. (6).** Illustration to show the hydrophilic and hydrophobic surface of (a) octapeptide NH-AVLQ$_1$SGFR-COOH, and (b) SARS CoV M$^\text{pro}$. The hydrophobic surfaces are colored in green and the hydrophilic surfaces in blue. In panel (a) the four hydrophobic amino acid residues are R$_4$(Ala), R$_3$(Val), R$_3'$(Phe), and R$_4'$(Arg). In panel (b) the hydrophobic residues R$_4$(Ala) and R$_3$(Val) are exposed in solvent, while R$_3'$(Phe) and R$_4'$(Arg) are covered by hydrophobic surfaces of the proteinase. Reproduced from Du *et al.* [30] with permission. (For interpretation of the references to color of the figure, the reader is referred to the web version of this paper).
REFERENCES

[27] Connolly, M. L. J. Mol. Graph., 1993, 11, 139-141.


