

Chapter 1

Study of Inhibitors Against SARS Coronavirus by Computational Approaches

**Kuo-Chen Chou, Dong-Qing Wei, Qi-Shi Du, Suzanne Sirois,
Hong-Bin Shen, and Wei-Zhu Zhong**

Abstract Called by many as the biology's version of Swiss army knives, proteases cut long sequences of amino acids into fragments and regulate most physiological processes. They are vitally important in life cycle and have become a main target for drug design. This Chapter is focused on a special protease that plays a key role in replicating SARS (Severe Acute Respiratory Syndrome) coronavirus, the culprit of SARS disease. The progresses reported here are mainly from various computational approaches, such as structural bioinformatics, pharmacophore modelling,

K.-C. Chou (✉)

Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; College of Life Science and Technology, Shanghai Jiaotong University, 800 Donglin Road, Shanghai, 200240, China; College of Life Science and Biotechnology, Guangxi University, Nanning, Guangxi, 530004, China; Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai, 200240, China
e-mail: kcchou@gordonlifescience.org

D.-Q. Wei

Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; College of Life Science and Technology, Shanghai Jiaotong University, 800 Donglin Road, Shanghai, 200240, China

Q.-S. Du

Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; College of Life Science and Biotechnology, Guangxi University, Nanning, Guangxi, 530004, China

S. Sirois

Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; Université du Québec à Montréal (UQAM), Chemistry Department, C.P. 8888 Succursale Centre-Ville, Montréal, Québec, Canada, H3C 3P8

H.-B. Shen

Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA; Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai, 200240, China

W.-Z. Zhong

Gordon Life Science Institute, 13784 Torrey Del Mar Drive, San Diego, CA 92130, USA

molecular docking, and peptide-cleavage site prediction, among others. It is highlighted that the compounds $C_{28}H_{34}O_4N_7Cl$, $C_{21}H_{36}O_5N_6$ and $C_{21}H_{36}O_5N_6$, as well as KZ7088, a derivative of AG7088, might be the promising candidates for further investigation, and that the octapeptides ATLQAIAS and ATLQAENV, as well as AVLQSGFR, might be converted to effective inhibitors against the SARS protease. Meanwhile, how to modify these octapeptides based on the “distorted key” theory to make them become potent inhibitors is explicitly elucidated. Also, a brief introduction is given for how to use computer-generated graphs to rapidly diagnose SARS coronavirus. Finally, a step-by-step protocol guide is given on how to use ProtIdent, a web-server developed recently, to identify the proteases and their types based on their sequence information alone. ProtIdent is a very user-friendly bioinformatics tool that can provide desired information for both basic research and drug discovery in a timely manner. With the avalanche of protein sequences generated in the post-genomic age, it is particularly useful. ProtIdent is freely accessible to the public via the web-site at <http://www.csbio.sjtu.edu.cn/bioinf/Protease/>.

Keywords SARS • coronavirus proteinase • KZ7088 • AG7088 • binding pocket • octapeptide inhibitors • distorted key theory • ProtIdent web server

Abbreviations CoV: Coronavirus; M^{pro}: Main proteinase; SARS: Severe acute respiratory syndrome

1.1 Introduction

Proteases, also termed as proteinases, or peptidases, are enzymes that are essential for all stages of life cycle, such as conception, birth, growth, ageing, and death of all organisms. Functioning as biology’s version of Swiss army knives, they cleave long sequences of amino acids into fragments, a process that is indispensable for the synthesis of all proteins, controlling protein composition, size, shape, turnover and ultimate destruction. Therefore, it can help us better understand the biology of life and the treatment of diseases to study proteases, their receptors and inhibitors.

A respiratory illness called “severe acute respiratory syndrome” (SARS) was first reported 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 ($>38^{\circ}\text{C}$ or 100.4°F) with symptoms such as headache, malaise, chilly, rigor, diarrhoea, 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%.

It is known that the culprit that causes SARS is the SARS coronavirus (sometimes shortened to SARS-CoV) (Peiris et al., 2003). It is also known that a key step for the replication of the culprit is cleaving the SARS-coronavirus polyproteins

by a special protease, the so-called SARS coronavirus main protease (SARS CoV Mpro). 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.

Progress in synthesis of novel test compounds for antiviral chemotherapy of SARS has been summarized by Kesel (Kesel, 2005), and that in drug discovery against SARS-CoV reported in a recent review (Wu et al., 2006). This chapter will focus on the progress mainly from the approaches of computer-aided drug discovery.

1.2 Binding Interactions

Based on the atomic coordinates of SARS-CoV Mpro (Anand et al., 2003), two binding models were developed (Chou et al., 2003). One is for the binding interaction of SARS-CoV Mpro with a compound called KZ7088 (Fig. 1.1a), and the other is for that with the octapeptide AVLQSGFR.

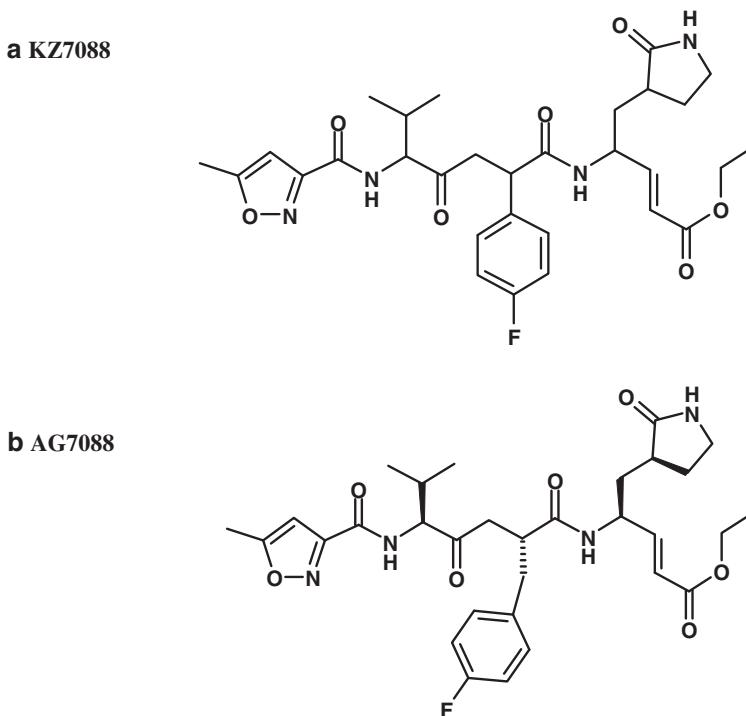


Fig. 1.1 Chemical structure of KZ7088: (a) a derivative of AG7088; (b) by removing $-CH_2$ from its fluorophenylalanine side chain (Reproduced from Chou et al., 2003. With permission)

KZ7088 (Chou et al., 2003) is a derivative of AG7088 (Fig. 1.1b). 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. 1.1, AG7088 has a p-fluorophenylalanine side chain (p-fluorobenzyl), which is too long (or bulky) to fit into the binding pocket of SARS-CoV Mpro. By removing $-CH_2$ from the side chain, the compound KZ7088 thus obtained could well fit into the binding pocket, as shown in Fig. 1.2. 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 ≤ 5 Å from a heavy atom of KZ7088. A similar binding pocket was defined for ATP in the Cdk5-Nck5a^{*}-ATP complex (Chou et al., 1999) and other studies (see, e.g., Zhou and Troy, 2003, 2005a, b; Chou, 2004b; Wei et al., 2006a; Wang et al., 2007a, b; Li et al., 2007a, b; Gu et al., 2009; Gong et al., 2009). The binding pocket of SARS-CoV Mpro for KZ7088 involved 23 residues, and the ligand was tethered to the enzyme by six hydrogen bonds, as detailed in (Chou et al., 2003). A series of follow-up discussions about the binding mechanism can be found in (Wu et al., 2006; Zhang and Yap, 2004; Samee, 2005; Clercq, 2006).

The binding interaction of SARS-CoV Mpro with the octapeptide AVLQSGFR is illustrated in Fig. 1.3, from which we can see that the octapeptide is tethered to Arg-40, His-41, Phe-185, Asp-187, and Gln-189 of SARS-CoV Mpro 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 $P_4P_3P_2P_1P_1'P_2'P_3'P_4'$ with the scissile bond located between the subsites P_1 and P_1' , as generally expressed by $P1\downarrow P1'$ (Schechter and Berger, 1967; Miller et al., 1989; Chou, 1993b). (2) The SARS coronavirus enzyme and several viral proteinases exhibit $Gln\downarrow$ (Ser, Ala, Gly) specificity (Anand et al., 2003). (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” (Chou, 1996), 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 proteinase.

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

1.3 Narrow down the Compounds-Searching Scope by Pharmacophore Approach

As mentioned above, both KZ7088 and the octapeptide AVLQSGFR are tethered to SARS-CoV Mpro by six hydrogen bonds. It is instructive to point out that analyzing the hydrogen bonding interactions of a receptor with its ligand often provided

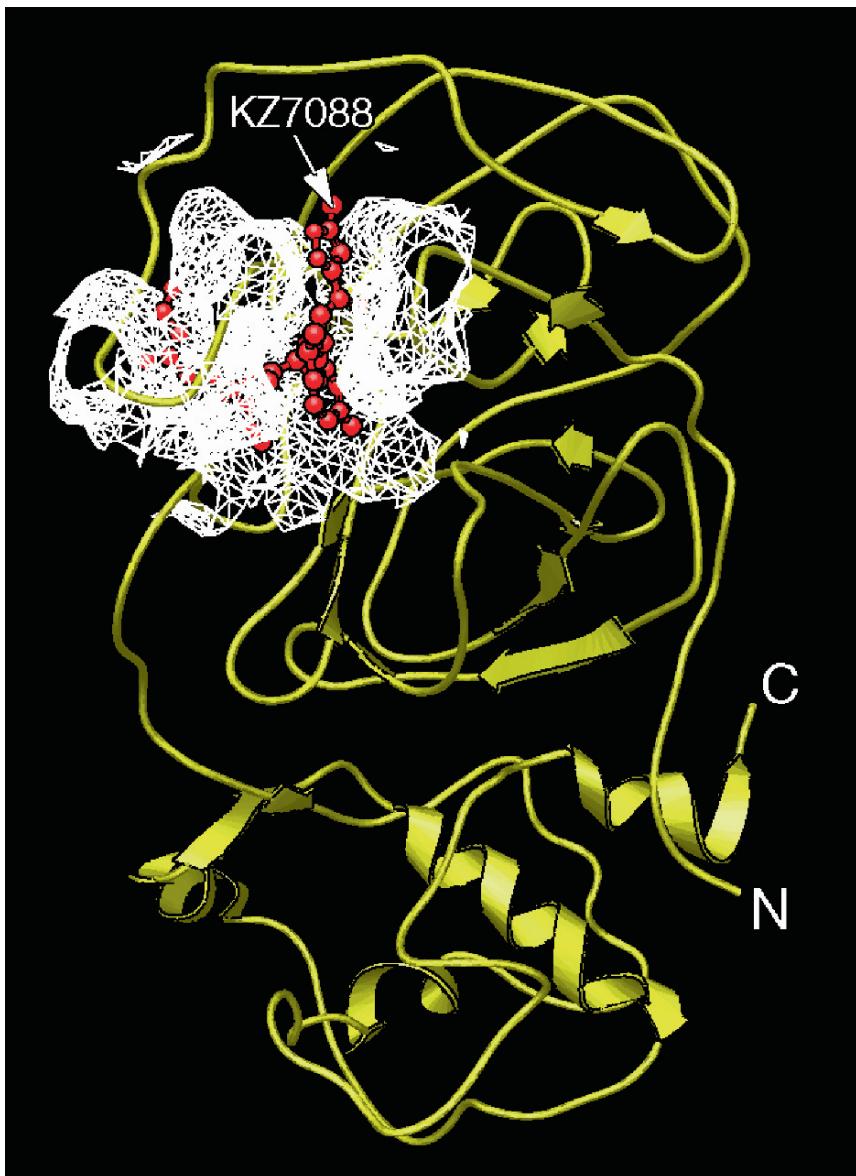


Fig. 1.2 An overall view of the complex obtained by docking KZ7088 to SARS-coronavirus main proteinase, where the SARS enzyme is colored yellow (in ribbon drawing), the KZ7088 colored red (in ball-and-stick drawing), and the white mesh represents the binding pocket for KZ7088 (Reproduced from Chou et al., 2003. With permission)

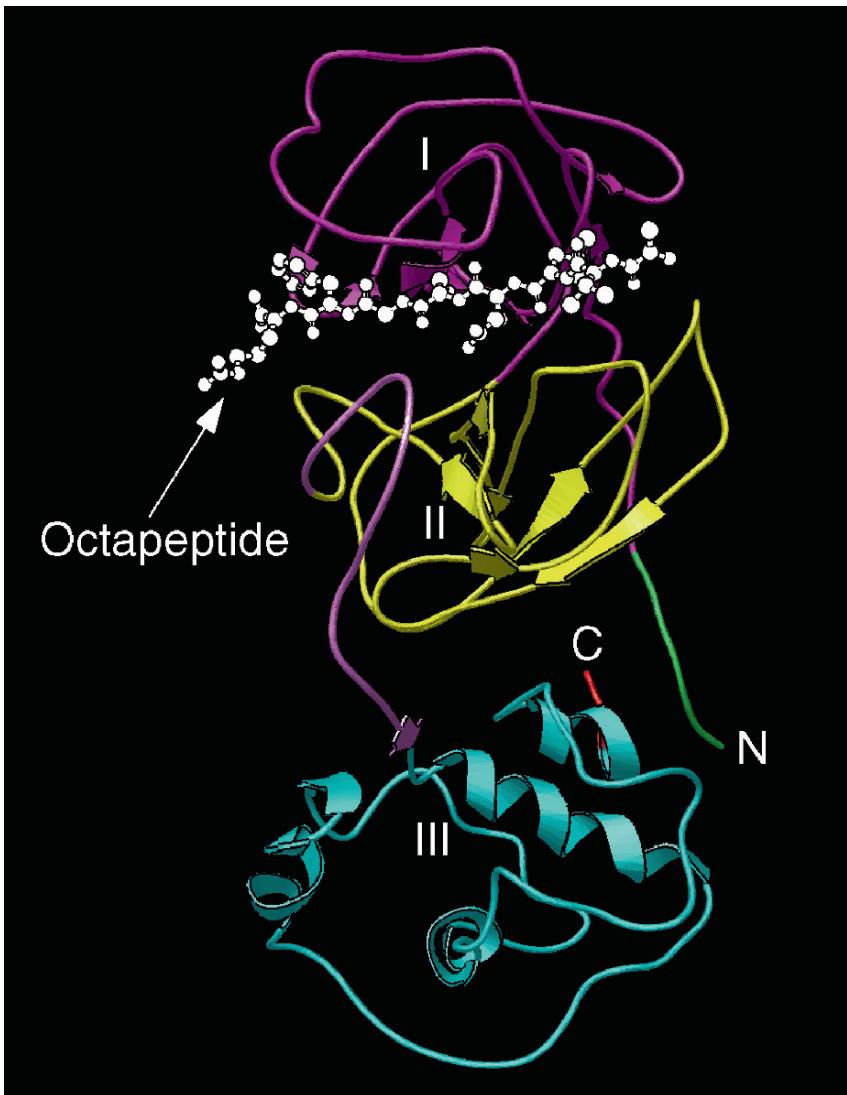


Fig. 1.3 An overall view of the complex obtained by docking the octapeptide AVLQSGFR to SARS-coronavirus main proteinase, where the octapeptide is coloured white (in ball-and-stick drawing), and the SARS enzyme is in ribbon drawing. As shown in the figure, the octapeptide is inserted into the cleft between domains I and II as expected (Reproduced from Chou et al., 2003. With permission)

useful insights for drug development (see, e.g., Chou, 2004a, b, 2005a, b; Chou & Howe, 2002; Wei et al., 2005, 2006a, b, 2007; Wang et al., 2007a, b, c, d, 2008; Li Y et al., 2007; Chou et al., 1997, 2000; Du et al., 2004, 2005a, 2007a, b; Zhang et al., 2006; Gao et al., 2007; Li L et al., 2007; Zheng et al., 2007; Guo et al., 2008; Schnell et al., 2008).

Based on the hydrogen bonding interactions of SARS-CoV M^{pro} with KZ7088 (Chou et al., 2003), a template of pharmacophore points of KZ7088 was generated (Sirois et al., 2004) as shown in Fig. 1.4, where the four hydrogen bond acceptors are coloured in cyan, the two hydrogen bond donors in purple, the aromatic group in green, and the volume constraints in grey. Based on the pharmacophore template, the virtual screening search operation was performed by Sirois and co-workers (Sirois et al., 2004) for the commercially and academically available compounds. It was found by pharmacophore searches 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 hydrogen bond acceptors and two hydrogen bond donors (Fig. 1.4). Moreover, each of the hit compounds was further evaluated for its drugability according to a novel score function, an equation formulated by 13 metrics that took into account physical (Lipinski, 2000), chemical and structural properties (Xu and Stevenson, 2000), as well the presence of undesirable functional groups (Oprea, 2000; Rishton, 1997). The selection of these properties was based on the work proposed by Baurin (Baurin, 2002). After such a drugability-evaluating procedure, it was found that 17% of the compounds 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 drugability was set at a maximum allowance of two violations, we obtained that only $0.07 \times (0.17 + 0.23 + 0.13) = 0.037$

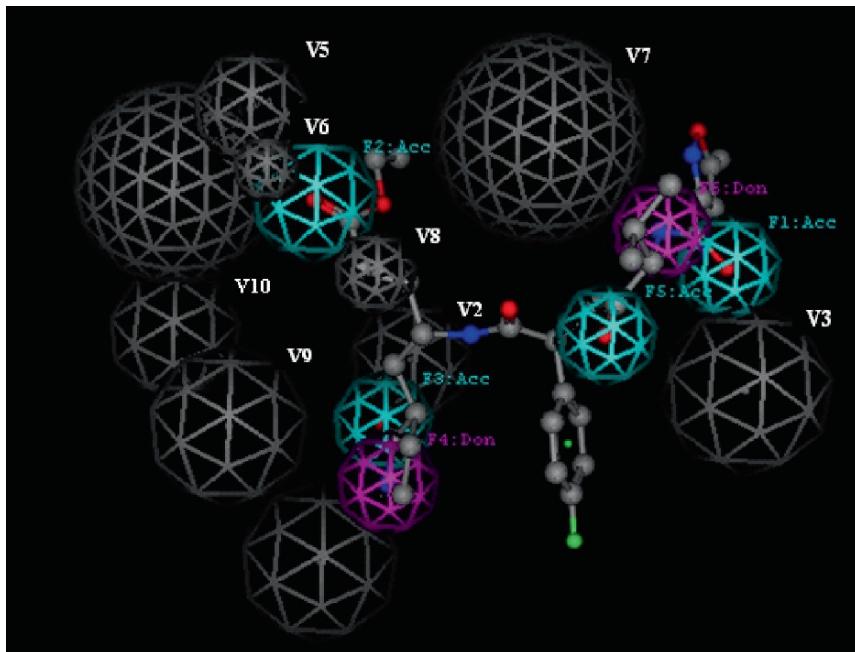


Fig. 1.4 A template of the pharmacophore points of KZ7088 derived from its binding interaction with SARS-CoV M^{pro} (Chou et al., 2003). The four H-b donors are colored in cyan, the two H-b donors in purple, the aromatic group in green, and the volume constraints in grey (Reproduced from Sirois et al., 2004. With permission)

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.

It is instructive to point out that, instead of choosing the hydrogen bond donors and acceptors for pharmacophore modelling as done in (Sirois et al., 2004), one can also choose some other features for modelling, such as aromatic ring, hydrophobic aromatic, hydrophobic aliphatic, positive charge, negative charge, hydrogen bond acceptor lipid, positive ionizable, and negative ionizable.

1.4 Winnow for Anti-SARS Drug Candidates by Molecular Docking Operation

From the 2,589 unique 3D hits with volume constrains found by Sirois et al. (2004), 1,386 drugable compounds were collected by Wei and co-workers (Wei et al., 2006) that both match the 3D (dimensional) KZ7088 pharmacophore points and satisfy the drugable rules with a scoring value ≥ 0.8 . It was found via similarity search and auto docking operation that only three compounds, i.e., $C_{28}H_{34}O_4N_7Cl$, $C_{21}H_{36}O_5N_6$, and $C_{21}H_{36}O_5N_6$ (Wei et al., 2006), were the most promising candidates worthy of further investigation.

Meanwhile, a thermodynamic analysis (Wei et al., 2006) indicated that the hydrogen bonds make an important contribution to the interactions of SARS-CoV Mpro with these compounds, which is quite consistent with the case of KZ7088 (Chou et al., 2003). In addition to hydrogen bonds, there are some other contributions to the interactions, such as desolvation, hydrophobic interaction, and electrostatic interaction. The hydrogen bonds 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. Therefore, in conducting rational drug design it is critic to identify the hydrophobic groups of the ligand and receptor and make sure that they are facing to each other upon binding. Also, it was observed by examining the Connolly surface (Connolly, 1993) of the complex of SARS-CoV Mpro with each of the three compounds that the steric complementarity is responsible for the micro-mechanical interlock at the 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 Mpro may provide useful insights for the development of potent inhibitors against the SARS enzyme.

Recently, Wang et al. performed a virtual screening for finding natural inhibitor against cathepsin-L for SARS therapy (Wang et al., 2007d).

1.5 Distorted-Key Theory and Peptide Inhibitors

Owing to their inability to readily crossing through membrane barriers such as the intestinal and blood-brain barriers, developing peptide drugs is limited by their poor metabolic stability and low bioavailability. 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 (Adessi and Soto, 2002). Some efforts have been made in an attempt to develop peptide inhibitors against the SARS-CoV M^{pro} (Chou et al., 2003; Du et al., 2004, 2005a, b, 2007b; Zhang et al., 2006). The development of peptide inhibitors against proteases was based on the “distorted key theory” (Chou, 1996), 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 favourable 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 chemical procedure, will completely lose its cleavability but it can still bind to the active site of an enzyme. Actually, the molecule thus modified can be compared 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. 1.5, where panel (a) shows an effective binding of a cleavable peptide to the active site of SARS CoV M^{pro}, while panel (b) shows that the peptide has become a non-cleavable one after 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^{pro}. It is instructive to point out that, even for the case of developing non-peptide inhibitors, the knowledge thus acquired can also provide useful insights for 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^{pro}. That is why during the process of developing inhibitors against HIV protease many algorithms were developed (Chou, 1993a, b; Poorman et al., 1991; Chou et al., 1993, 1996; Zhang and Chou, 1993; Thompson et al., 1995; Cai and Chou, 1998; You et al., 2005; Liang and Li, 2007; Rognvaldsson et al., 2007) for predicting the cleavable peptides by HIV protease. Very recently, a freely accessible web-server called HIVcleave (Shen and Chou, 2008) was also established for identifying the cleavage sites of proteins by HIV protease.

Stimulated by the distorted key theory for finding the peptide inhibitors against HIV protease, the similar bioinformatics tools were also developed (Gao et al., 2003; Chen et al., 2003) for finding the peptide inhibitors against SARS-CoV M^{pro}.

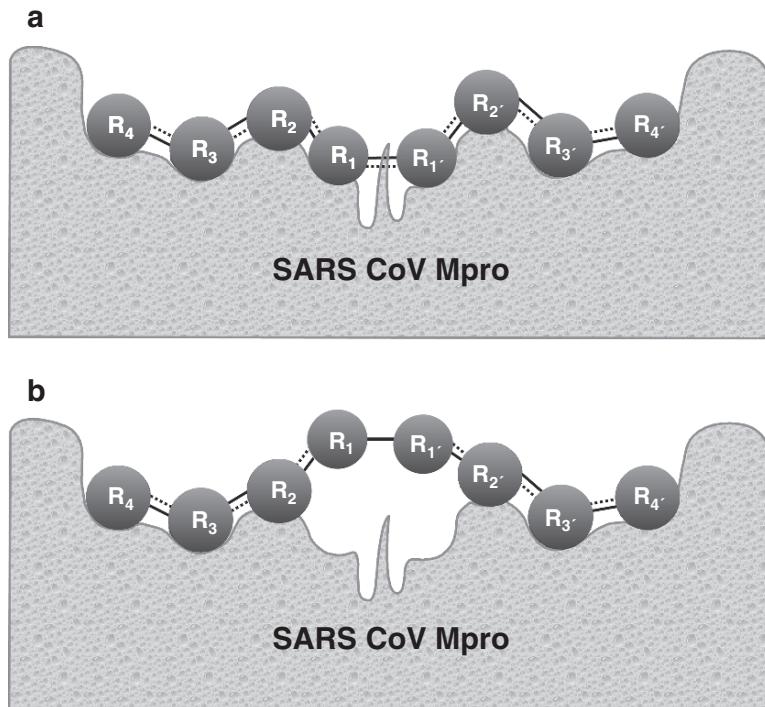


Fig. 1.5 Schematic illustration to show (a) a cleavable octapeptide is chemically effectively bound to the active site of SARS CoV Mpro, and (b) although still bound to the active site, the peptide has lost its cleavability after its scissile bond is modified to a strong hybrid peptide bond. The eight residues of the peptide is sequentially symbolized by R₄, R₃, R₂, R₁, R_{1'}, R_{2'}, R_{3'}, and R_{4'}, and their counter parts by S₄, S₃, S₂, S₁, S_{1'}, S_{2'}, S_{3'}, and S_{4'}, respectively (Adapted from Chou, 1996. With permission)

Du and co-worker (Du et al., 2005a) used these tools to search 36 SARS coronaviruses and found 396 cleavage sites, from which 11 cleavable octapeptides were extracted. In order to find the physical and chemical properties of the amino acids from positions R₄ to R_{4'}, Du et al. aligned the 11 cleavable peptides using the flexible molecular alignment technique (Du et al., 2007b). The results are given in Fig. 1.6, where panel (a) shows the lipophilicity and hydrophilicity of amino acid side chains of the 11 cleavable peptides, and panel (b) shows the pharmacophore types of atoms in the relevant amino acid side chains. As shown in the figure, residues at R₂ and R₄ are the most lipophilic; residues at R_{1'} and R_{3'} are partially lipophilic residues. On the other hand, residues at R₁ are the most hydrophilic, and residues at R₃ and R₄ are partially hydrophilic residues. The residues at R_{2'} could be lipophilic or hydrophilic. The size of side chain of R_{1'} is very small, and the volumes of side chains of residues R₃, R₂ and R_{3'} are larger than the others. Aromatic groups are on the positions R₂ and R_{3'}. Subsequently, the docking operations were performed for the 11 cleavable peptides with respect to SARS CoV M^{pro}. It was found that of the 11 cleavable peptides, NH₂-ATLQ↓AENV-COOH had the highest affinity with SARS CoV M^{pro}.

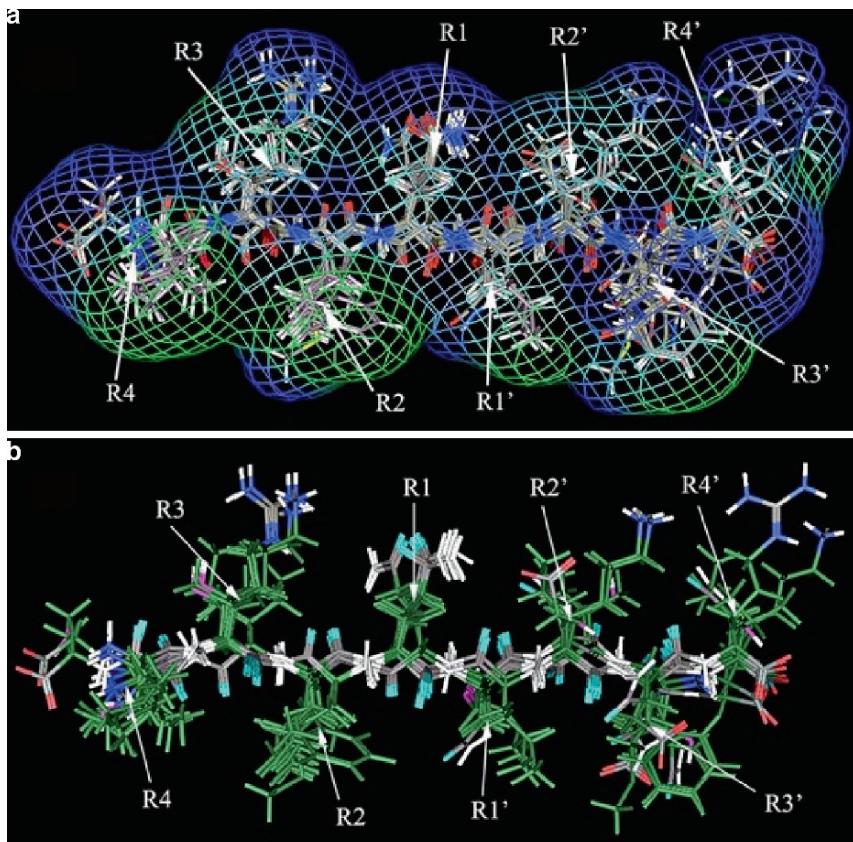


Fig. 1.6 The alignment of the 11 cleavable peptides. (a) Hydrophilic and hydrophobic surfaces of peptides. The green is for lipophilic regions and the blue is for hydrophilic regions. (b) Atomic pharmacophore types of peptides. Atoms are colored according to their pharmacophore type, donor is white, acceptor is cyan, acceptor and donor is magenta, cation or resonance cation is light blue, anion or resonance anion is red, hydrophobe is dark green, and gray is for other atoms

The binding interaction of the enzyme with $\text{NH}_2\text{-ATLQ}\downarrow\text{AENV-COOH}$ obtained by the docking operation is shown in Fig. 1.7.

Although a cleavable peptide can fittingly bind to the active site of a protease, it is not an inhibitor to the enzyme because it has a scissile peptide bond that will be eventually cleaved. However, a cleavable peptide can become an effective inhibitor after some proper chemical modification, as elaborated in Chou, 1996. Converting the cleavable peptides to stable and bio-available drugs is a hot research topic in medicinal chemistry (Chou, 1996, 2004; Du et al., 2004). Some possible chemical modification methods in this regard were analyzed (Du et al., 2007b). According to the distorted key theory (Chou, 1996), a cleavable oligopeptide can automatically become a competitive inhibitor if its scissile peptide bond between R_1 and $R_{1'}$ is replaced by a strong “hybrid peptide bond” (cf. Fig. 1.5). For the case of $\text{NH}_2\text{-ATLQ}\downarrow\text{AIAS-COOH}$, the position R_1 and $R_{1'}$ is occupied by amino acids Gln and

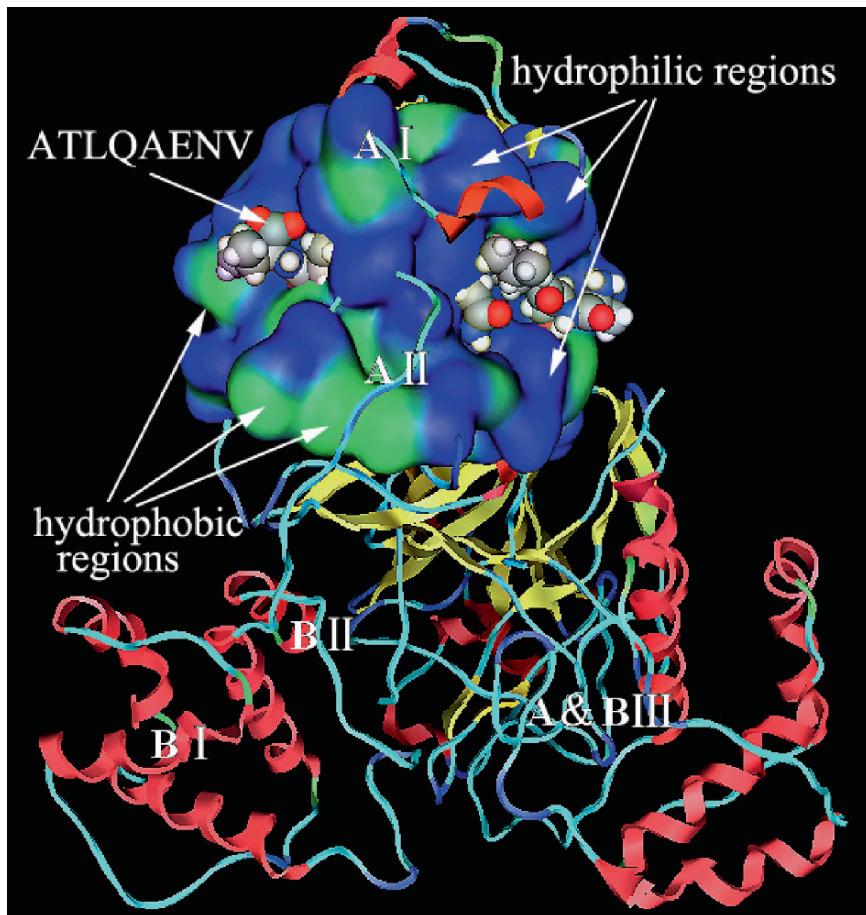


Fig. 1.7 The binding interaction obtained by docking the octapeptide ATLQAENV to SARS CoV M^{pro} . The green is for hydrophobic regions and the blue is for hydrophilic regions. The cleavable peptide P9 is shown as space filling model. The bioactive pocket is the cleavage cavity between domains I and II of protomer A

Ala. The peptide bond $CO \approx NH$ between Gln and Ala is considered as a pseudo conjugate π -bond (here we use $CO \approx NH$ to emphasize the conjugate π -bond), which is easily hydrolyzed by SARS CoV M^{pro} . If the group CO is replaced by CH_2 , or NH is replaced by CH_2 , the conjugate π -bond will become a strong single bond, and hence the hydrolyzation will become very difficult (Du et al., 2004; Szelke et al., 1982; Venkatesan and Kim, 2002). In general, the methylamino isostere is prepared by reduction of the amide bond with different reducing agents or by reductive amination of an amino aldehyde (Benkirane et al., 1996). The isostere has been incorporated into the ras farnesyl-protein transferase inhibitor that is a potential anticancer drug (Graham et al., 1994). Pseudo peptide analogs containing reduced amide isostere are also shown to inhibit C_{α} -N-C cleavage in the gag region of HIV-1 protease (Lowther et al., 1995). It was also observed by quantum chemistry calcula-

tions (Du et al., 2007b) that after CO was replaced by CH₂, the charge of the carbon atom in CH₂ turned to negative (-0.6572) from original a positive value (0.8612) in CO group, hence the nucleophilic attack by OH⁻ became impossible. Furthermore, the negative charge of N(NH) on the Ser side decreased to -0.2892 from -0.8190, and hence the electrophilic attack by H⁺ became more difficult as well.

On the other hand, the experiments by Gan et al. (Gan et al., 2006) indicated that the octapeptide AVLQSGFR proposed originally in (Chou et al., 2003) was indeed cleavable by SARS-CoV M^{pro} with a high bioactivity. Based on the crystal structure of SARS-CoV M^{pro} (Yang et al., 2003), the cleavage mechanism of the SARS-CoV M^{pro} on the octapeptide NH₂-AVLQ↓SGFR-COOH (Chou et al., 2003) was investigated using molecular mechanics (MM) and quantum mechanics (QM) (Du et al., 2005b). It has been observed that the catalytic dyad (His-41/Cys-145) site between domain I and II of the protease [4,5] attracts π-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_{δ1} 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 CH₂ or CF₂, or changing the NH of Ser to CH₂ or CF₂. According to the “distorted key theory” (Chou, 1996), the octapeptide thus modified might become an effective inhibitor against SARS-CoV M^{pro} or a potential drug candidate for SARS therapeutics.

1.6 Using Graphic Approaches to Detect SARS-CoVirus

By analyzing the two-dimensional cellular automata images (Wolfram, 1984, 2002) for 96 SARS coronavirus and 25 non-SARS coronavirus, Wang et al. (Wang et al., 2005) found a remarkable distinction between the SARS and non-SARS coronal viruses: all the images of the former were with a V-shaped cross-lines pattern (Fig. 1.8a) but all the images of the latter are with a parallel slash-lines pattern (Fig. 1.8b). They concluded that the V-shaped cross-lines image pattern might be deemed as a fingerprint for SARS coronal viruses for rapidly diagnosing SARS coronavirus.

Subsequently, Gao and co-workers (Gao et al., 2006) developed another graphic method for visualizing the characteristic of SARS-CoV. It 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.

1.7 Identification of Proteases and Their Types

Owing to their indispensable role in life cycle, proteases have become an important target for drug discovery. Different types of proteases have different action mechanisms and biological processes. With the avalanche of new protein sequences generated

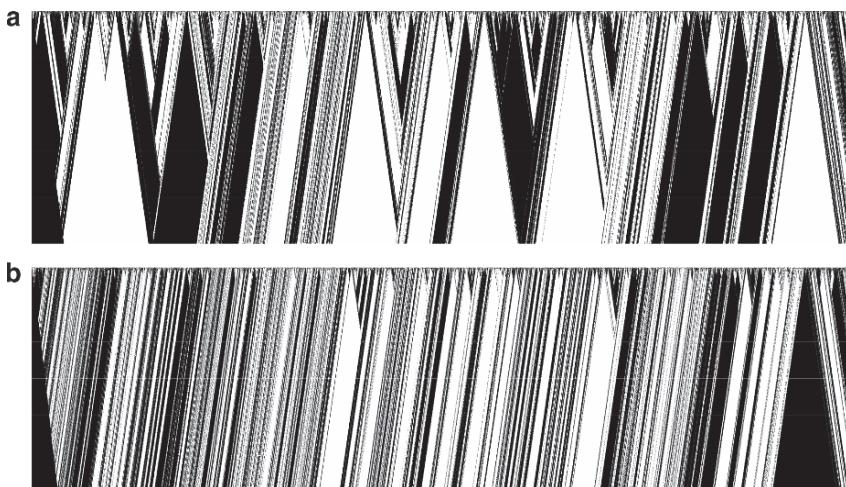


Fig. 1.8 The cellular automata images obtained by applying the modified Rule 184 (Wolfram, 1984, 2002) on (a) SARS coronavirus and (b) non-SARS coronavirus (Reproduced from Wang et al., 2005. With permission)

in the post-genomic age, it is highly desired to develop an automated method that can timely identify proteases and their types according to their sequence information alone.

Recently, a web server called ProtIdent (Protease Identifier) has been established at <http://www.csbio.sjtu.edu.cn/bioinf/Protease/>. ProtIdent is a two-layer predictor. The first layer is for identifying a query protein as protease or non-protease. If it is a protease, the process will automatically go to the second layer to further identify it among the following six functional types: (1) aspartic, (2) cysteine, (3) glutamic, (4) metallo, (5) serine, and (6) threonine. The success expectancy with ProtIdent in identifying a protein as protease or non-protease was about 92%, and that in identifying the protease type was about 96%.

As a user-friendly web server, ProtIdent is freely accessible to the public. Below, let us provide a step-by-step guide on how to use ProtIdent to get the desired results.

Step 1. Open the web page <http://www.csbio.sjtu.edu.cn/bioinf/Protease> and you will see the top page of the ProtIdent predictor on your computer screen, as shown in Fig. 1.9. Click on the Read Me to see a brief introduction about ProtIdent predictor and caveat in using it.

Step 2. Either type or copy and paste the query protein sequence into the input box at the center of Fig. 1.9. The input sequence should be in the single-letter amino acid code, as shown by clicking on the Example button right above the input box. For instance, if your input is the 306 amino acids of SARS coronavirus main protease unit taken from the grey-boxed and underlined segment in **Box 1.1**, the input screen should look like the screenshot of Fig. 1.10.

ProtIdent: identifying proteases and their types by fusing functional domain and sequential evolution information

| [Read Me](#) | [Data](#) | [Citation](#) |

Input the protein sequence ([Example](#)):

Contact @ [Hong-Bin](#)

Fig. 1.9 A screenshot to show the top page of the web server ProtIdent at <http://www.csbio.sjtu.edu.cn/bioinf/Protease>

Box 1.1 The full-length amino acid sequence of SARS coronavirus protease (SW: P59641) where the grey-boxed and underlined segment (from res.3241 to res.3546) is the main protease

```
MESLVGVNEKTHVQLSLPVLQVRDVVLVRGFGDSVEALSEAREHLKNGTCGLVELEKGV
LPQLEQPYVFIKRSDLSTNHGHKVVELVAEMDGIQYGRSGITLGVLPVHGETPIAYRN
VLLRKNGNKGAGGHSYGIDLKSYDLGDELGTDPIEDYEQNWNTKHGSGALRELTRELNGG
AVTRYVDNNFCGPDPYPLDCIKDFLARAGKSMCTLSEQLDYIESKRGVYCCRDHEHEIAW
FTERSDKSYEHQTPFEIKAFFDTFKGECPKFVFLNSKVVKVQPRVEKKTEGFMGRI
RSVYPVASPQECNNMHLSTLMKCNHCDEVWSQTCDFLKATCEHCGTENLVIEGPTTCGYL
PTNAVVKMPCPACQDPEIGPHEHSVADYHNHSNIETRLKGGTRCFGGCVFAYVGCYNR
AYWVPRASADIGSGHTGITGDNVETLNEDLLEILSRERVNINIVGDFHLDNEEVAILASF
SASTSAFIDTIKSLDYKSFKTIVESCGNYKVTGKPVKGAWNIGQQRSVLTPLCGFPSQA
AGVIRSFARTLDAANHSIPDLQRAAVTILDGISEQSRLRVAMVYTSDLLTNSVIIMAY
VTGGLVQQTSQWLSNLLGTTVEKLRPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIV
KGQIQVASDNIKDCVVKCFIDVNVNKALEMCIDQVTIAGAKLRSNLGEVFIAQSKGLYRQC
IRGKEQLQLLMPKKAPKEVTFLEGDSHTVLTSEEVVLKNGELEALETPVDSFTNGAIVG
TPVCVNGMLLEIKDKEQYCALSPGLLATNNVFRLKGGAPIKGVTGFEDTVWEVQGYKNV
RITFELDERVDKVLNEKCSVYTVESGTEVTEFACVVAEAVVKTLPVSDLLTNMGIDLDE
WSVATFYLFDDAGEENFSSRMYCSFYPPDEEEEDAECEEEDETCEHEYTEDDYQGL
PLEFGASAETVRVEEEEEEDWLDDTEQSEIEPEPEPTPEEPVNQFTGYLKLTDNVAIKC
VDIVKEAQSANPMIVVNAANILKHGGGVAGALNKATNGAMQKESDDYIKLNGPLTVGGS
CLLSGHNLAKKCLHVGPNLNAGEDIQLLKAAYENFSQDILLAPLLSAGIFGAKPLQLS
```

(continued)

Box 1.1 (continued)

QVCVQTVRTQVYIAVNDKALYEQVVMYLDNLKPRVEAPKQEPPNTEDSKTEEKSVVQK
 PVDVKPKIKACIDEVTTLLEETKFLTNKLLLFADINGKLYHDSQNMLRGEDMSFLEKDAP
 YMVGDVITSGDITCVVIPSKKAGGTTEMLSRALKKVPVDEYITTYPGQGCAGYTLEEAKT
 ALKKCKSAFYVLPEASEPNAKEEILGTVSWNLREMLAHAEETRKLMPICMDVRAIMATIQR
 KYKGKIKIQEGIVDYGVRFFFYTSKEPVASIITKLNLSNEPLVTMPIGYVTHGFNLEEAAR
 CMRSLKAPAVSVSSPDAVTTYNGYLTSSEHFVETVSLAGSYRDWSYSGQRTELG
 VEFLKRGDKIVYHTLESPVEFHLDGEVLSLKLKSLLSREVTKIVFTTDNTNLHTQL
 VDMSMTYQGQFGPTYLDGADVTKIKPHVNHEGKTFVLPSSDTLRSEAFYYHTLDESFL
 GRYMSALNHTKKWKFQVGGLTSIKWADNNCYLSSVLLAQQLEVKFNPALQEAYYRAR
 AGDAANFCALILAYSNKTVGELGDVRETMTHLLQHANLESAKRVLNVVCKHCGQKTTLT
 GVEAVMYMGTLSYDNLKTGVUSICVGRDATQYLVQQESSFVMMMSAPPAEYKLQQGTFLC
 ANEYTGNYQCGHYTHITAKETLYRIDGAHLTKMSEYKGPVTDVFYKETSYTITIKPVSYK
 LDGVTYTEIEPKLDGYKKDNAYYTEQPIDLVTQPLPNASFDFNFKLTCNTKFADDLNQ
 MTGFTKPKASRELSTVTPDNLGVVAIDYRHYSASFKKGAKLLHKPIVWHINQATTKTF
 KPNTWCRLCLWSTKPVDTNSNSFEVLADEDTQGMNDNLACESQQPTSEEVVENPTIQKVEIE
 CDVKTEEVGNVILKPSDEGVKTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIA
 THGIAAIINSVPWSKILAYVKPFLGQAAITTSNCAKRLAQRVFNNYMPYVFTLFFQLCTFT
 KSTSNSRIRASLPTTIAKNSVKSVAKLCLADAGINYVKSPKFSKLFTIAMWLLLSICLGS
 ICVTAAGVLLSNFGAPSYCNGVRELYLNNSNVTMDFCESFPCSICLGLDSDLSDSYPA
 LETIQVTISSYKLDLTLIGLAAEWVLAYMLFTKFFYLLGLSAIMQVFFGYFASHFISNSW
 LMWFIIISIVQMAPVSAMVRMYIFFASFYYIWKSYVHIMDGCTSSTCMCYKRNRATRVEC
 TTIVNGMKRSFYVYANGGRGFCKTHNWNCNLCDTFCGTGTFISDEVARDLSLQFKRPINP
 TDQSSYIVDSVAVKNGALHLYFDKAGQKTYERHPLSHFVNLDNLRANNTKGSPLINVIVF
 DGSKKCDESASKSASVYYSQLMCQPILLDQALVSDVGDSTEVSVKMFDAYVDTFSATFS
 VPMEKLKALVATAHSELAKGVALDGVLSFTVSAARQGVVDTDVDTKDVIECLKLSHHS
 EVTDGDCNNFMILTYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVKDYMSLSE
 QLRKQIRSAAKKNIPFRLLTCACTRQVNVITTKISLKGKIVSTCFKLMKATLLCVLA
 ALVCYIVMPVHTLSIHGDYTNIEIIGYKAIQDGVTVDIISTDDCFANKHAGFDAWFSQRGG
 SYKNDKSCPVVAAECTIFKDAMGKPVPYCYDTNLLEGSISYSELRPDTRYVLMGSIQ
 EYSDFATSACVLAEECTIFKDAMGKPVPYCYDTNLLEGSISYSELRPDTRYVLMGSIQ
 FPNTYLEGSVRVTTFDAEYCRHGTCERSEVGICLSTSGRWVNLNEHYRALSGVFCGVDA
 MNLIANIFTPLVQPGALDVSVASVAGGIIALTVTCAAYYFMKRRVGEYNHVVAANAL
 LFLMSFTILCLVPAYSLPVGVYSVFYLTFYFTNDVSFLAHLQWFAMFSPIVPFWITAI
 YVFCISLKHCHWFNNYLRKRVMFNGVTFTFEEAACHLAKALNDFNSNGADVLQPPQTSITSAVLQ
 SGFRKMAFPSKGVEGCMVQVTCGTTLNGWLDDTVYCPRHVICTAEDMLNPNEYEDLLIR
 KSNHSFLVQAGNVQLRVIGHSMQNCRLKVDTSNPKTPKYKFVRIQPGQTFSVLACYNG
 SPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGK
 FYGPVVDROTAQAAGTDITITLNVLAWLYAAVINGDRWFLNRFTTILNDFNLVAMKYN
E
PLTQDHVDILGPLSAQTGIAVLDMCAALKELLONGMNGRTILGSTILEDEFTPFDVVRQC
SGVTFQGKFKKIVKGTHHWMLLTFLTSLLILVQSTQWSLFFFVYENAFLPFTLGMIAAA
 CAMLLVKKHAFCLFLPLSLATVAYFNMVYMPASWVMRIMTWLELADTSLGYRLKDCV
 MYASALVLLILMTARTVYDDAARRVWLMNVITLVYKVYYGNALDQAISMWALVISVTSN
 YSGVTTIMFLARAIVFVCVEYYPLLFITGNTLQCIMLVYCFGLGYCCCCYFGLFCLLNRY
 FRLTGLVYDYLVSTQEFRYMSNSQGLLPPKSSIDAFKLNKLLGIGGKPCIKVATVQSKMS

Box 1.1 (continued)

DVKCTSVVLLSVLQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSLLSVLLSMQG
 AVDINRLCEEMLDNRATLQAIASEFSSLPSYAAYATAQEAEYEQAVANGDSEVVLLKKKS
 LNVAKSEFDRAAMQRKLEKMAQDQAMTQMYKQARSEDKRAKVTSAMQTMFLMLRKLDND
 ALNNIINNARDGCVPLNIPLTTAAKLMVVVPDYGTYKNTCDGNTFTYASALWEIQQVVD
 ADSKIVQLSEINMDNSPNLAWPLIVTALRANSAVKLQNNELPVALRQMSCAAGTTQTAC
 TDDNALAYYNNSKGGRFVLALLSDHQDLKWARFPKSDGTGTYTELEPPCRFTDTPKGPD
 KVKYLYFIKGLNNLNRMGVLSLAATVRLQAGNATEVPANSTVLSFCAFADPAKAYKD
 LASGGQPITNCVKMLCTHTGQAITYVPEANMDQESFGGASCCLYCRCHIDHPNPKGFC
 DLKGKYVQIPTTCANDPVGFTLRLNTVCTCGMWKGYGCSCDQLREPLMQSADASTFLNRV
 CGVSAARLTPCGTGTSTDVYRAFDIYNEKAGFAKFLKTNCRCFQEKEDEEGNLDSYFV
 VKRHTMSNYQHEETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQLTKYTMA_DL_VYAL
 RHFDEGNCDTLKEILVTYNCCDDDFNKDDWYDFVENPDLRVYANLGERVRSQLLTVQ
 FCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIDSYSSLLMPILTLTRAL
 AAESHMDADLAKPLIKWDLLYDFTEERLCLFDRYFKYWDQTYHPNCINCLDDRCILHCA
 NFNVLFSTVFPPTSFGPLVRKIFVGDGVFVVSTGYHFRELGVVHNDVNLHSSRLSFKEL
 LVYAADPAMHAASGNLLLDKRTTCFSVAALTNNAFQTVKPGFNKDFYDFAVSKGFFKE
 GSSVELKHFFAQDGNAISDYDYYRNLPTMCDIRQLLFFVVEVVDKYFDCYDGGCINAN
 QVIVNNLSDKSAGFPFNWKWGKARLYYDSMSYEDQDALFAYTKRNVIPIITQMNLYKAISAK
 NRARTVAGVSICSTMNRQFHQKLLKSIATRGATVVIGTSKFYGGWHNMLKTVYSDVET
 PHLMGWDPKCDRAMPNMLRIMASLVLARKHNTCCNLSHRFYRLANECAQVLSEMVMCGG
 SLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNIKIAKYVRNLQHRLYECL
 YRNDRVDFEVDEFYAYLRKHFSSMMILSDDAVVCYNSNYAAQGLVASIKNFKAVLYYQNN
 VFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVLPYDPSRILGAGCFVDDIVKTD
 GTLMIERFVSLAIDAYPLTKHPNQEYADVFHLYQYIRKLHDELTGHMLDMYSVMLTNDN
 TSRYWEPEFYEAMYTPHTVLQAVGACVLCNSQTSLRGACIRRPFLCCCYDHVISTSH
 KLVLSVNPYVCNAPGCDVTDVTQYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVG
 SDNVTDFNIAITCDWTNAGDYIILANTCTERLKLFAAETLKATEETFKLSYGIATREVLS
 DRELHLSWEVGKPRPPLRNRYVFTGYRVTKNNSVQIGEYTFEKGDYGDADVYRGTTTYKL
 NVGDYFVLTSHVMPLSATPLVQEHYVRTGLYPTLNISDEFSSNVANYQKVMGQKYST
 LQGPPGTGKSHFAIGLALYPPSARIVYTACSHAADVDALCEKALKYLPIDKCSRIIPARAR
 VECFDKFKVNSTLEQYVFCVTNVALPETTADIVFDEISMATNYDL SVNARLRAKHYYI
 GDPAQLPAPRTLLTKTLEPEYFNSVCRMKTIGPDMFLGTCRRCPAEIVDTVSALVYDN
 KLKAHKDKSAQCFKMFYKGVTIDVSSAINRPQIVGVREFLTRNPAPRKAVFISPYNQN
 AVASKILGLPTQTVDSQGSEYDYVIFTQTETAHSCNVNRFNVAITRAKIGILCIMSDR
 DLYDKLQFTSLEIPRRNVAUTLQAENVTGLFKDCSKIITGLHPTQAPTHLSVDIKFKTEGL
 CVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPNMFITREEAIRHVRRAWIGFDVEGCHATR
 DAVGTNPLQLGFSTGVNLVAVPTGYVDTENNEFTRVNAKPPGDQFKHLIPLMYKGLP
 WNVVRIKIVQMLSDLKGLSDRVFVLAHGFELTSMKYFVKIGPERTCCLCDKRATCFS
 TSSDTYACWNHHSVGFDVYNPMIDVQQWGFITGNLQSNHDQHCQVHGNNAHVASCDAIMTR
 CLAVHECFVKRVDWSVEYPII_GDELRVNSACRKVQHMVVKSALLADKFVPLHDIGNPKAI
 KCVPQAEVEWKFYDAQPCSDKAYKIEELFYSYATHHDKFTDGVCLFWNCVDRY PANIAV
 CRFDTRVLNSNLPGCDGGSLYVNKHAFHTPAFDKSAFTNLKLPFFYSDSPCESHGKQ
 VVSDIDYVPLKSATCITRCNLGGAVCRHHANEYRQYDAYNMMISAGFSLWIYKQFDTYN
 LWNTFTRLQSLENVAYNVVNKGHFDPHAGEAPVSIINNAVYTKVGDIDVEIFENKTLPV
 NVAFELWAKRNKPKVPEIKILNNLGVDIAANTVIWDXKREAPAHVSTIGVCTMTDIAKKP

(continued)

Box 1.1 (continued)

```

TESACSSLTVLFDCRVEGVQVDFRNARNGVLITEGSVKGLTPSKGPAQASVNGVTLIGES
VKTQFNKFKKVDCIIQQLPETYFTQSRSDFLEFKPRSQMETDFLELAMDEFIQRYKLEGYA
FEHIVYGFDSHGQLGLHLIGLAKRSQDSPLKLEDFIPMDSTVKNYFITDAQTGSSKCV
CSVIDLDDDFVEIJKSQQDLSVISVKVKVTDYAEISFMLWCKDGHVETFYPKLQASQAW
QPGVAMPNLYKMQRMLLEKCDLQNYGENAVIPKGIMMNVAKYTQLCQYLNTLAVPYNM
RVIHFGAGSDKGVAPGTAVLRQWLPTGTLVDSDLNDFSDADSTLGDCATVHTANKWD
LIISDMYDPRTKHVTKEKNSKEGFFTYLCGFIKQKLALGGSIAVKITEHSWNADLYKLMG
HFSSWWTAFTVTVNASSEAFLIGANYLGKPKEQIDGYTMHANYIFWRNTNPQLSSYSLF
DMSKFPLKLRGTAVMSLKENQINDMYSLLERKGRILIRENNRVVVSSDILVNN

```

ProtIdent: identifying proteases and their types by fusing functional domain and sequential evolution information

| [Read Me](#) | [Data](#) | [Citation](#) |

Input the protein sequence ([Example](#)):

```

SGFRKMAFPSGKVEGCMVQVTCGTTLNGLWLDLTVYCPRHVICTAEDMLNPNEYDLIR
KSNHSLFLVQAGNVQLRVIGHSMQNCLRLKVTDSNPKTPKVKFVRQPGQTFSVLACYNG
SPSGVYQCAMRPNH-TIKGSFLNGSCGSVGFNIDYDCVSFCYMHMELPTGVHAGTDELEGK
FYGPFPVDRQTAQAAAGTDTTITLNVLAWLWYAAVINGDRWFNLNRFTTLDNFNLVAMKYNYE
PLTQDHVDILGPLSAQQTGIAVLDMAALKELLQNGMNGRTILGSTILEDEFTPFDVVRQC
SGVTFQ

```

Contact @ [Hong-Bin](#)

Fig. 1.10 A screenshot to show the sequence of the SARS main protease entered into the input box of the ProtIdent web server

Step 3. Click on the Submit button and you will see in about 15 s that “the query protein is a protease” and it “belongs to Cysteine type”, as shown on the output screen of Fig. 1.11. If your input is the 7,073 amino acids of the full length SARS coronavirus protease as taken from **Box 1.1**, the same predicted result will be obtained although it will take much longer time (about 230 s) because the input sequence now is much longer.

Step 4. Click on the Citation button to find the relevant papers that document the detailed development and algorithm of ProtIdent.

Step 5. Click on the Data button to download the benchmark datasets used to train and test the ProtIdent predictor.

ProtIdent: identifying proteases and their types by fusing functional domain and sequential evolution information

| [Read Me](#) | [Data](#) | [Citation](#) |

Your input sequence(306aa) is:

```
>QuerySeq
SGFRKMAFPSSGVEGCMVQVTCGTTLNLWLDDTVYCPRHVICTAEDMLNPVNVEDLLIR
KSNHSLFLVQAGNVQLRIVIGHSMQNCLLRKVDTSNPKTPKYKFVRIQPGQTFSVLACYNG
SPSGVYQCAMRPNHTIKGSFLNGSGSVEFNIYDTCVSFCYMHMELPTGVHAAGTDLEGK
FYGFVVDQTAQAAGTDTITLNLWALWYAAVINGDRWFNLRTTILNDFNLVAMKYNYE
PLTQDHVDILQPLSAQTGLAVLDMAALKELLQNGMNGRILGSTILEDEFTPFDVVRQC
SGVTFQ
```

-----Predicted Results-----

	Predicted Results
<i>1st Layer</i>	the query protein is a protease
<i>2nd Layer</i>	belongs to Cysteine type

[Home Page](#)

Contact @ [Hong-Bin](#)

Fig. 1.11 A screenshot to show the output predicted by the ProtIdent web server

Caveat. To obtain the predicted result with the expected success rate, either the entire sequence of the query protein or the entire sequence of its main protease should be used as an input.

1.8 Conclusion

Since the outbreak of SARS in November 2002 in Southern China's Guangdong Province, considerable progress has been made in developing drugs for SARS therapy. This chapter is focused on the advances achieved mainly from the structural bioinformatics, pharmacophore modeling, molecular docking, peptide-cleavage site prediction, and other computational approaches. It is indicated that the compounds $C_{28}H_{34}O_4N_7Cl$, $C_{21}H_{36}O_5N_6$ and $C_{21}H_{36}O_5N_6$ (Wei et al., 2006), as well as KZ7088 (Chou et al., 2003), a derivative of AG7088, might be the promising candidates for further investigation. Meanwhile, it is elucidated according to the “distorted key” theory (Chou et al., 1996) that the octapeptides ATLQAENV and AVLQSGFR may be converted to effective inhibitors against the SARS enzyme. Also, how to modify these octapeptides and make them become potent inhibitors is suggested. Finally, a step-to-step guide on how to use ProtIdent, a web server for identifying proteases and their types (Chou and Shen, 2008), is presented that may be useful for people working in the area of Proteases in Biology and Disease.

References

- Adessi, C. and Soto, C. 2002, Converting a peptide into a drug: strategies to improve stability and bioavailability. *Curr Med Chem*, **9**, 963–978.
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. and Hilgenfeld, R. 2003, Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, **300**, 1763–1767.
- Baurin, N. 2002, Etude et développement de techniques QSAR pour la recherche de molécules d'intérêts thérapeutique. *Cribleage virtuel et Analyse de chmiothèque*, Université d'Orléans.
- Benkirane, N., Guichard, G., Briand, J.P. and Muller, S. 1996, Exploration of requirements for peptidomimetic immune recognition. Antigenic and immunogenic properties of reduced peptide bond pseudopeptide analogues of a histone hexapeptide. *J Biol Chem*, **271**, 33218–33224.
- Cai, Y.D. and Chou, K.C. 1998, Artificial neural network model for HIV protease cleavage sites in proteins. *Adv Eng Softw*, **29**, 119–128.
- Chen, L.L., Ou, H.Y., Zhang, R. and Zhang, C.T. 2003, ZCURVE_CoV: a new system to recognize protein coding genes in coronavirus genomes, and its applications in analyzing SARS-CoV genomes. *Biochem Biophys Res Commun*, **307**, 382–388.
- Chou, J.J. 1993a, Predicting cleavability of peptide sequences by HIV protease via correlation-angle approach. *J Protein Chem*, **12**, 291–302.
- Chou, K.C. 1993b, A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. *J Biol Chem*, **268**, 16938–16948.
- Chou, K.C. 1996, Review: prediction of HIV protease cleavage sites in proteins. *Analyt Biochem*, **233**, 1–14.
- Chou, K.C. 2004a, Molecular therapeutic target for type-2 diabetes. *J Proteome Res*, **3**, 1284–1288.
- Chou, K.C. 2004b, Review: structural bioinformatics and its impact to biomedical science. *Curr Med Chem*, **11**, 2105–2134.
- Chou, K.C. 2004c, Insights from modelling the 3D structure of the extracellular domain of alpha7 nicotinic acetylcholine receptor. *Biochem Biophys Res Commun*, **319**, 433–438.
- Chou, K.C. 2005a, Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. *J Proteome Res*, **4**, 1681–1686.
- Chou, K.C. 2005b, Modeling the tertiary structure of human cathepsin-E. *Biochem Biophys Res Commun*, **331**, 56–60.
- Chou, K.C. and Howe, W.J. 2002, Prediction of the tertiary structure of the beta-secretase zymogen. *Biochem Biophys Res Commun*, **292**, 702–708.
- Chou, K.C. and Shen, H.B. 2008, ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information. *Biochem Biophys Res Commun*, **376**, 321–325.
- Chou, K.C., Zhang, C.T. and Kezdy, F.J. 1993, A vector approach to predicting HIV protease cleavage sites in proteins. *Proteins: Struct Funct Genet*, **16**, 195–204.
- Chou, K.C., Tomasselli, A.L., Reardon, I.M. and Heinrikson, R.L. 1996, Predicting HIV protease cleavage sites in proteins by a discriminant function method. *Proteins: Struct Funct Genet*, **24**, 51–72.
- Chou, K.C., Jones, D. and Heinrikson, R.L. 1997, Prediction of the tertiary structure and substrate binding site of caspase-8. *FEBS Lett*, **419**, 49–54.
- Chou, K.C., Watenpaugh, K.D. and Heinrikson, R.L. 1999, A Model of the complex between cyclin-dependent kinase 5(Cdk5) and the activation domain of neuronal Cdk5 activator. *Biochem Biophys Res Commun*, **259**, 420–428.
- Chou, K.C., Tomasselli, A.G. and Heinrikson, R.L. 2000, Prediction of the tertiary structure of a caspase-9/inhibitor complex. *FEBS Lett*, **470**, 249–256.
- Chou, K.C., Wei, D.Q. and Zhong, W.Z. 2003, Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: ibid., 2003, 310, 675). *Biochem Biophys Res Commun*, **308**, 148–151.
- Clercq, E.D. 2006, Potential antivirals and antiviral strategies against SARS coronavirus infections. *Expert Rev Anti-infect Ther*, **4**, 291–302.

- Connolly, M.L. 1993, The molecular surface package. *J Mol Graph*, **11**, 139–141.
- Du, Q.S., Wang, S.Q., Wei, D.Q., Zhu, Y., Guo, H., Sirois, S. and Chou, K.C. 2004, Polyprotein Cleavage Mechanism of SARS CoV Mpro and Chemical Modification of Octapeptide. *Peptides*, **25**, 1857–1864.
- Du, Q.S., Wang, S., Wei, D.Q., Sirois, S. and Chou, K.C. 2005a, Molecular modelling and chemical modification for finding peptide inhibitor against SARS CoV Mpro. *Analyt Biochem*, **337**, 262–270.
- Du, Q.S., Wang, S.Q., Jiang, Z.Q., Gao, W.N., Li, Y.D., Wei, D.Q. and Chou, K.C. 2005b, Application of bioinformatics in search for cleavable peptides of SARS-CoV Mpro and chemical modification of octapeptides. *Med Chem*, **1**, 209–213.
- Du, Q.S., Wang, S.Q. and Chou, K.C. 2007a, Analogue inhibitors by modifying oseltamivir based on the crystal neuraminidase structure for treating drug-resistant H5N1 virus. *Biochem Biophys Res Commun*, **362**, 525–531.
- Du, Q.S., Sun, H. and Chou, K.C. 2007b, Inhibitor design for SARS coronavirus main protease based on “distorted key theory”. *Med Chem*, **3**, 1–6.
- Gan, Y.R., Huang, H., Huang, Y.D., Rao, C.M., Zhao, Y., Liu, J.S., Wu, L. and Wei, D.Q. 2006, Synthesis and activity assess of an octapeptide inhibitor designed for SARS coronavirus main proteinase. *Peptides*, **27**, 622–625.
- Gao, F., Ou, H.Y., Chen, L.L., Zheng, W.X. and Zhang, C.T. 2003, Prediction for proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Letters*, **553**, 451–456.
- Gao, L., Ding, Y.S., Dai, H., Shao, S.H., Huang, Z.D. and Chou, K.C. 2006, A novel fingerprint map for detecting SARS-CoV. *J Pharmaceut Biomed Analysis*, **41**, 246–250.
- Gao, W.N., Wei, D.Q., Li, Y., Gao, H., Xu, W.R., Li, A.X. and Chou, K.C. 2007, Agaritine and its derivatives are potential inhibitors against HIV proteases. *Med Chem*, **3**, 221–226.
- Gong, K., Li, L., Wang, J.F., Cheng, F., Wei, D.Q. and Chou, K.C. 2009, Binding mechanism of H5N1 influenza virus neuraminidase with ligands and its implication for drug design. *Med Chem*, in press.
- Graham, S.L., deSolms, S.J., Giuliani, E.A., Kohl, N.E., Mosser, S.D., Oliff, A.I., Pompliano, D.L., Rands, E., Breslin, M.J., Deana, A.A. et al. 1994, Pseudopeptide inhibitors of Ras farnesyl-protein transferase. *J Med Chem*, **37**, 725–732.
- Gu, R.X., Gu, H., Xie, Z.Y., Wang, J.F., Arias, H.R., Wei, D.Q. and Chou, K.C. 2009, Possible drug candidates for Alzheimer’s disease deduced from studying their binding interactions with alpha7 nicotinic acetylcholine receptor. *Med Chem*, in press.
- Guo, X.L., Li, L., Wei, D.Q., Zhu, Y.S. and Chou, K.C. 2008, Cleavage mechanism of the H5N1 hemagglutinin by trypsin and furin. *Amino Acids*, **35**, 375–382.
- Kesel, A.J. 2005, Synthesis of novel test compounds for antiviral chemotherapy of severe acute respiratory syndrome (SARS). *Curr Med Chem*, **12**, 2095–2162.
- Li, L., Wei, D.Q., Wang, J.F. and Chou, K.C. 2007a, Computational studies of the binding mechanism of calmodulin with chrysin. *Biochem Biophys Res Comm*, **358**, 1102–1107.
- Li, Y., Wei, D.Q., Gao, W.N., Gao, H., Liu, B.N., Huang, C.J., Xu, W.R., Liu, D.K., Chen, H.F. and Chou, K.C. 2007b, Computational approach to drug design for oxazolidinones as antibacterial agents. *Med Chem*, **3**, 576–582.
- Liang, G.Z. and Li, S.Z. 2007, A new sequence representation as applied in better specificity elucidation for human immunodeficiency virus type 1 protease. *Biopolymers*, **88**, 401–412.
- Lipinski, C.A. 2000, Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods*, **44**, 235–249.
- Lowther, W.T., Majer, P. and Dunn, B.M. 1995, Engineering the substrate specificity of rhizopuspepsin: the role of Asp 77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P1. *Protein Sci*, **4**, 689–702.
- Miller, M., Schneider, J., Sathyaranayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B. and Wlodawer, A. 1989, Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science*, **246**, 1149–1152.

- Oprea, T.I. 2000, Property distribution of drug-related chemical databases. *J Comput Aided Mol Des*, **14**, 251–264.
- Peiris, J.S., Chu, C.M., Cheng, V.C., Chan, K.S., Hung, I.F., Poon, L.L., Law, K.I., Tang, B.S., Hon, T.Y., Chan, C.S. et al. 2003, Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet*, **361**, 1767–1772.
- Poorman, R.A., Tomasselli, A.G., Heinrikson, R.L. and Kezdy, F.J. 1991, A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate data base. *J Biol Chem*, **266**, 14554–14561.
- Rishton, G.M. 1997, Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today*, **2**, 382–384.
- Rognvaldsson, T., You, L. and Garwicz, D. 2007, Bioinformatic approaches for modeling the substrate specificity of HIV-1 protease: an overview. *Expert Rev Mol Diagn*, **7**, 435–451.
- Samee, W. 2005, Severe acute respiratory syndrome coronavirus (SARS-CoV) protease inhibitors. *J Club Pharm Chem Pharmacognosy*, January, 1–12.
- Schechter, I. and Berger, A. 1967, On the size of the active site in protease. I. Papain. *Biochem Biophys Res Comm*, **27**, 157–162.
- Schnell, J.R. and Chou, J.J. 2008, Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*, **451**, 591–595.
- Shen, H.B. and Chou, K.C. 2008, HIVcleave: a web-server for predicting HIV protease cleavage sites in proteins. *Analyt Biochem*, **375**, 388–390.
- Sirois, S., Wei, D.Q., Du, Q.S. and Chou, K.C. 2004, Virtual Screening for SARS-CoV protease based on KZ7088 pharmacophore points. *J Chem Inf Comput Sci*, **44**, 1111–1122.
- Szelke, M., Leckie, B.J., Tree, M., Brown, A., Grant, J., Hallett, A., Hughes, M., Jones, D.M. and Lever, A.F. 1982, A potent new renin inhibitor. In vitro and in vivo studies. *Hypertension*, **4**, 59.
- Thompson, T.B., Chou, K.C. and Zheng, C. 1995, Neural network prediction of the HIV-1 protease cleavage sites. *J Theoret Biol*, **177**, 369–379.
- Venkatesan, N., Kim, B.H. 2002, Synthesis and enzyme inhibitory activities of novel peptide isosteres. *Curr Med Chem*, **9**, 2243–2270.
- Wang, J.F., Wei, D.Q., Li, L., Zheng, S.Y., Li, Y.X. and Chou, K.C. 2007a, 3D structure modeling of cytochrome P450 2C19 and its implication for personalized drug design. *Biochem Biophys Res Commun (Corrigendum: ibid, 2007, Vol357, 330)*, **355**, 513–519.
- Wang, J.F., Wei, D.Q., Lin, Y., Wang, Y.H., Du, H.L., Li, Y.X. and Chou, K.C. 2007b, Insights from modeling the 3D structure of NAD(P)H-dependent D-xylose reductase of *Pichia stipitis* and its binding interactions with NAD and NADP. *Biochem Biophys Res Comm*, **359**, 323–329.
- Wang, S.Q., Du, Q.S., Zhao, K., Li, A.X., Wei, D.Q. and Chou, K.C. 2007c, Virtual screening for finding natural inhibitor against cathepsin-L for SARS therapy. *Amino Acids*, **33**, 129–135.
- Wang, S.Q., Du, Q.S. and Chou, K.C. 2007d, Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases. *Biochem Biophys Res Comm*, **354**, 634–640.
- Wang, J.F., Wei, D.Q., Chen, C., Li, Y. and Chou, K.C. 2008, Molecular modeling of two CYP2C19 SNPs and its implications for personalized drug design. *Protein Pept Lett*, **15**, 27–32.
- Wang, M., Yao, J.S., Huang, Z.D., Xu, Z.J., Liu, G.P., Zhao, H.Y., Wang, X.Y., Yang, J., Zhu, Y.S. and Chou, K.C. 2005, A new nucleotide-composition based fingerprint of SARS-CoV with visualization analysis. *Med Chem*, **1**, 39–47.
- Wei, D.Q., Sirois, S., Du, Q.S., Arias, H.R. and Chou, K.C. 2005, Theoretical studies of Alzheimer's disease drug candidate [(2,4-dimethoxy) benzylidene]-anabaseine dihydrochloride (GTS-21) and its derivatives. *BBRC*, **338**, 1059–1064.
- Wei, D.Q., Du, Q.S., Sun, H. and Chou, K.C. 2006a, Insights from modeling the 3D structure of H5N1 influenza virus neuraminidase and its binding interactions with ligands. *Biochem Biophys Res Comm*, **344**, 1048–1055.
- Wei, D.Q., Zhang, R., Du, Q.S., Gao, W.N., Li, Y., Gao, H., Wang, S.Q., Zhang, X., Li, A.X., Sirois, S. et al. 2006b, Anti-SARS drug screening by molecular docking. *Amino Acids*, **31**, 73–80.

- Wei, H., Zhang, R., Wang, C., Zheng, H., Chou, K.C. and Wei, D.Q. 2007, Molecular insights of SAH enzyme catalysis and their implication for inhibitor design. *J Theoret Biol*, **244**, 692–702.
- Wolfram, S. 1984, Cellular automation as models of complexity. *Nature*, **311**, 419–424.
- Wolfram, S. 2002, *A New Kind of Science*. Wolfram Media, Champaign, IL.
- Wu, Y.S., Lin, W.H., Hsu, J.T. and Hsieh, H.P. 2006, Antiviral drug discovery against SARS-CoV. *Curr Med Chem*, **13**, 2003–2020.
- Xu, J. and Stevenson, J. 2000, Drug-like index: a new approach to measure drug-like compounds and their diversity. *J Chem Inf Comput Sci*, **40**, 1177–1187.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H. et al. 2003, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci USA*, **100**, 13190–13195.
- You, L., Garwicz, D. and Rognvaldsson, T. 2005, Comprehensive bioinformatic analysis of the specificity of human immunodeficiency virus type 1 protease. *J Virol*, **79**, 12477–12486.
- Zhang, C.T. and Chou, K.C. 1993, An alternate-subsite-coupled model for predicting HIV protease cleavage sites in proteins. *Protein Eng*, **7**, 65–73.
- Zhang, R., Wei, D.Q., Du, Q.S. and Chou, K.C. 2006, Molecular modeling studies of peptide drug candidates against SARS. *Med Chem*, **2**, 309–314.
- Zhang, X.W. and Yap, Y.L. 2004, Exploring the binding mechanism of the main proteinase in SARS-associated coronavirus and its implication to anti-SARS drug design. *Bioorg Med Chem*, **12**, 2219–2223.
- Zheng, H., Wei, D.Q., Zhang, R., Wang, C., Wei, H. and Chou, K.C. 2007, Screening for new agonists against Alzheimer's disease. *Med Chem*, **3**, 488–493.
- Zhou, G.P. and Troy, F.A., 2nd ed. 2003, Characterization by NMR and molecular modeling of the binding of polyisoprenols and polyisoprenyl recognition sequence peptides: 3D structure of the complexes reveals sites of specific interactions. *Glycobiology*, **13**, 51–71.
- Zhou, G.P. and Troy, F.A., 2nd ed. 2005a, NMR study of the preferred membrane orientation of polyisoprenols (dolichol) and the impact of their complex with polyisoprenyl recognition sequence peptides on membrane structure. *Glycobiology*, **15**, 347–359.
- Zhou, G.P. and Troy, F.A. 2005b, NMR studies on how the binding complex of polyisoprenol recognition sequence peptides and polyisoprenols can modulate membrane structure. *Curr Protein Peptide Sci*, **6**, 399–411.