



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Molecular modeling and chemical modification for finding peptide inhibitor against severe acute respiratory syndrome coronavirus main proteinase

Qishi Du^{a,e,*}, Shuqing Wang^a, Dongqing Wei^{a,b,e,*}, Suzanne Sirois^{b,c,d,e},
Kuo-Chen Chou^{a,d,*}

^a Tianjin Institute of Bioinformatics and Drug Discovery, Tianjin Normal University, Tianjin 300074, China

^b Center for Research in Molecular Modeling (CERMM), Concordia University, Montreal, Canada

^c Université du Québec à Montréal (UQAM) Chemistry Department, C.P. 8888, succursale Centre-Ville, Montreal, Que., Canada H3C 3P8

^d Immune Deficiency Treatment Center (IDTC), Montreal General Hospital, McGill University Health Center, 1650 Cedar Avenue, Montréal, Québec, H36 1A4, Canada

^e Gordon Life Science Institute, San Diego, CA 92130, USA

Received 28 August 2004

Available online 23 December 2004

Abstract

Severe acute respiratory syndrome (SARS) is a respiratory disease caused by a newly found virus, called SARS coronavirus. In this study, the cleavage mechanism of the SARS coronavirus main proteinase (M^{pro} or 3CL^{pro}) on the octapeptide $\text{NH}_2\text{-AVLQ}\downarrow\text{SGFR-COOH}$ was investigated using molecular mechanics and quantum mechanics simulations based on the experimental structure of the proteinase. It has been observed that the catalytic dyad (His-41/Cys-145) site between domains I and II attracts the 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. $\text{N}_{\delta 1}$ on the imidazole ring plays the acid–base catalytic role. Based on the “distorted key theory” [K.C. Chou, *Anal. Biochem.* 233 (1996) 1–14], the possibility to convert the octapeptide to a competent inhibitor has been studied. It has 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 . The octapeptide thus modified might become an effective inhibitor or a potential drug candidate against SARS.

© 2004 Elsevier Inc. All rights reserved.

Keywords: SARS; SARS CoV M^{pro} ; Inhibitor; Drug design; “Distorted key” theory

A new coronavirus (SARS CoV)¹ has been found as the etiological agent of the disease SARS [1–4]. SARS CoV is a positive-sense, single-stranded RNA virus with the largest viral RNA genome known to date [5,6]. The SARS CoV genome is composed of around 29,700

nucleotides. The replicase gene alone encompasses more than 21,000 nucleotides, 2/3 of the whole genome sequence. The replicase gene encodes two overlapping polyproteins, pp1a (4377 a.a.) and pp1ab (7072 a.a.). SARS CoV main proteinase (M^{pro}) cleaves the polyprotein at no less than 11 conserved sites, a process initiated by the enzyme’s own autolytic cleavage from pp1a and pp1ab [7,8]. The released polypeptides mediate all of the functions required for SARS viral replication and transcription. The functional importance of the M^{pro}

* Corresponding author.

E-mail addresses: duqishi@yahoo.com (Q. Du), dqwei_2000@yahoo.com (D. Wei), kchou@san.rr.com (K.-C. Chou).

¹ Abbreviations used: SARS CoV, severe acute respiratory syndrome coronavirus; a.a., amino acids; M^{pro} , main proteinase.

in the viral life cycle makes it an attractive target for the drugs design against SARS [7–10].

Anand et al. [7] suggested that the rhinovirus 3C^{pro} inhibitor AG7088 could well serve as a starting point for an anti-SARS drug based on the theoretical homology model of SARS CoV M^{pro}. Chou et al. [9] found the fitting problem of AG7088 to the binding pocket of SARS CoV M^{pro}, and they suggested its derivative KZ7088 as a better starting point. Sirois et al. [11] did virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. Chou et al. [9] also suggested an octapeptide NH₂-AVLQ↓SGFR-COOH that can be converted to an inhibitor of SARS CoV M^{pro} by some chemical modification. This octapeptide was synthesized, proved to be a cleavable peptide, and showed strong inhibiting activity in a tube test [12]. Du et al. [13] analyzed the cleavage mechanism of the octapeptide by SARS CoV M^{pro} and discussed the possible chemical modification of the octapeptide using molecular dynamical and quantum mechanical simulations based on the homology structure of SARS CoV M^{pro}, to convert it into a stable and effective drug. This progress on finding inhibitors against the SARS enzyme is also summarized in a recent review [10].

The experimental structure of SARS CoV M^{pro} was first determined by Rao's research group [8]. They provided several crystal structures of SARS CoV M^{pro} at different pH conditions and the complex structure of SARS CoV M^{pro} with ligand hexapeptidyl CMK [7,8]. The experimental structure and the theoretical homology structure of SARS CoV M^{pro} share many structural similarities, such as the catalytic active region and the Cys–His cleavage dyad [7,8]. However, there are several differences between the theoretical and the experimental structures. For examples, the experimental structure of SARS CoV M^{pro} is a dimer, the protomer A has catalytic activity, and the protomer B has no activity. However, the N finger of protomer B, which plugs in the active cleft of protomer A, plays an important role in both dimerization and maintenance of the active form of the enzyme [8]. These new findings from the experimental structure of SARS CoV M^{pro} provide a reliable basis for inhibitor design.

Method

The octapeptide NH₂-AVLQ↓SGFR-COOH shows strong inhibiting activity in vitro [12]. However, the use of native peptide for clinical applications has been hampered by their intrinsic metabolic instability (indigestion by enzymes) and poor absorption through membrane. A cleavable peptide of SARS CoV M^{pro} means that it possesses a competent binding with its receptor and a breakable peptide bond in the cleavage site. To convert a cleavable peptide into a stable and effective inhibitor,

the peptide has to be modified according to the “distorted key” theory [14,15]. One important step for the chemical modification is a good understanding of the cleavage mechanism, particularly the changes in chemical bonds [16–22].

The protease-susceptible sites in a given protein or peptide usually extend to an octapeptide region [14,15]. The corresponding amino acid residues are sequentially symbolized by eight subsites R₄, R₃, R₂, R₁, R₁', R₂', R₃', R₄' and the eight matching positions in the protease by S₄, S₃, S₂, S₁, S₁', S₂', S₃', S₄', respectively (see Fig. 3 of [15]). Occasionally, the susceptible sites in some proteins may contain one subsite less or one more. However, eight amino acid residues are the most common case. Although the protein being cleaved contains much more than eight amino acid residues, only an octapeptide segment fits in the active region of the protease. Therefore, our research will focus on the cleavability of octapeptides.

As shown in Fig. 1, the cleavage point is always on the peptide bond between R₁ and R₁'. In Fig. 1A we use the combination of a thin line and a dashed line to represent the conjugate π property in the peptide bond. In Fig. 1B the breakable peptide bond between R₁ and R₁' is converted to a strong “hybrid peptide bond” through a chemical modification and hence cannot be cleaved by the protease. The modified octapeptide can still bind to the active site, but it cannot be cleaved by the enzyme, just like a distorted key that can neither open the lock nor be pulled out. The octapeptide thus modified will automatically become a desired inhibitor candidate.

The octapeptide AVLQSGFR is the first suggested [9] based on the molecular structure of SARS CoV M^{pro} and has been proved cleavable by experiments. In this research we study the cleavage mechanism, the properties of the relevant chemical bonds, and the catalytic interactions between the octapeptides and the SARS CoV M^{pro} using molecular mechanical and quantum chemical simulations to provide useful insights for the chemical modification. The study was carried out in the following four steps: (1) docking the octapeptide to SARS CoV M^{pro} followed by an energy minimization for the complex structure using molecular mechanics, (2) computing the atomic charge distribution in the binding pocket for the energy-minimized structure using *ab initio* quantum mechanics, (3) computing the molecular energy, chemical bond properties, and atomic charges of the octapeptide with the background charge distribution [23] of SARS CoV M^{pro} using *ab initio* quantum chemistry, and (4) with the same background charge distribution, computing the molecular energy, chemical bond properties, and atomic charges of the modified octapeptide. The computations have been carried out with the molecular dynamics and standard *ab initio* quantum chemistry at the HF/6-31G* level [23].

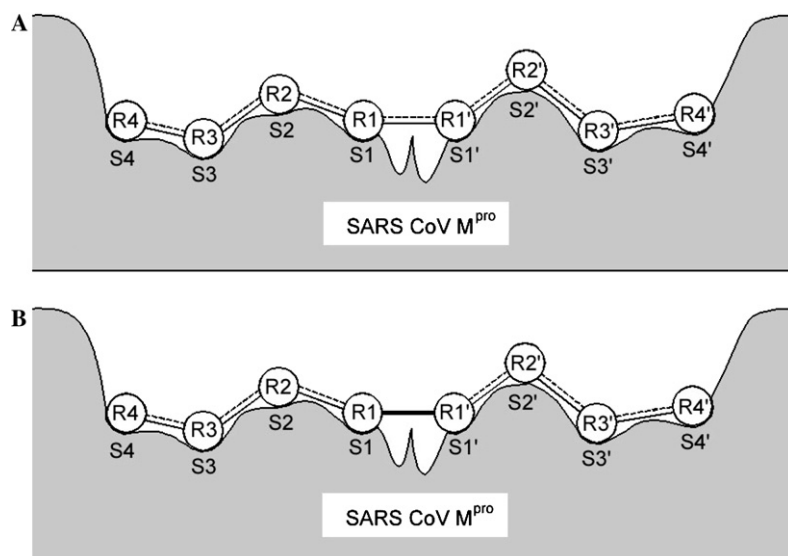


Fig. 1. Schematic illustration to show (A) a cleavable octapeptide is chemically effectively bound to the active site of SARS enzyme and (B) although still bound to the active site, the peptide has lost its cleavability after its scissile bond was modified from a hybrid peptide bond to a strong bond. Adapted from Chou [15] with permission.

Results

Docking and energy minimization

The experimental structure (1uj1.pdb) of SARS CoV M^{pro} [8] is used in the calculation and shown in Fig. 2A. Each protomer structure of the SARS CoV M^{pro} dimer has three domains. The two protomers A and B form a right angle and the join part is in the domain III. The N

finger of protomer B plugs into the domain II of protomer A and plays an important role in the catalytic activity. The PDB data of SARS CoV M^{pro} contains 300 hydration water molecules, which provide the solvation ion environment with some water molecules joining the catalytic reaction. In Fig. 2A the green signs “<” represent the solvation water molecules. The docking calculation between SARS CoV M^{pro} and the octapeptide AVLQSGFR was performed using the molecular

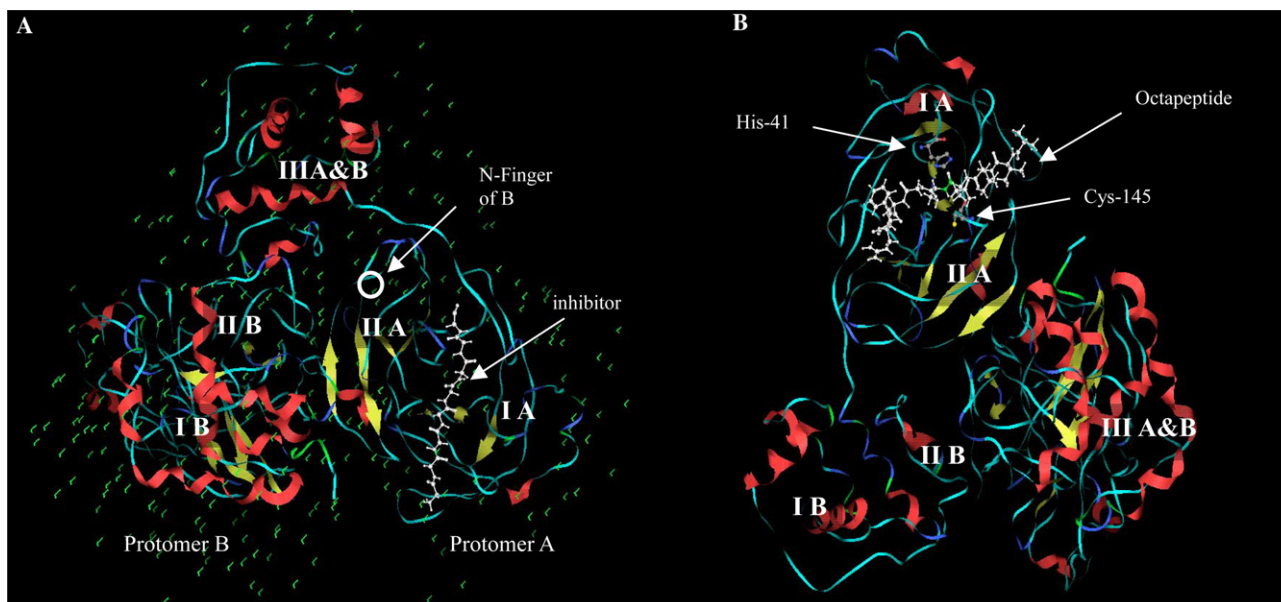


Fig. 2. (A) Crystal structure of SARS CoV M^{pro} consists of two protomers A and B. Each protomer has three domains and the two protomers form a right angle at the domain III. The N finger of protomer B plugs into the domain II of protomer A. The green signs “<” represent the solvation water molecules. (B) The complex structure obtained by docking the octapeptide to SARS CoV M^{pro} and a follow-up energy minimization. The catalytic dyad of His-41 and Cys-145 is marked by an arrow and the peptide bond to be cleaved of the octapeptide is shown in green.

operating environment program package [24]. A total of 25 docking conformations were obtained, and the one with the lowest energy was used for further energy minimizing calculation. Fig. 2B shows the complex structure of SARS CoV M^{PRO} with the ligand octapeptide. The positions of catalytic dyad His-41 and Cys-145 are indicated in Fig. 2B.

According to Anand et al. [7], Yang et al. [8], and Chou et al. [9], the catalytic active region is in the pocket between domains I and II of protomer A. During calculation the catalytic active region on which we were focusing contains the amino acid residues Cys-22, Gly-23, Thr-24, Thr-25, Leu-27, His-41, Val-42, Cys-44, Thr-45, Ala-46, Glu-47, Asp-48, Met-49, Leu-50, Asn-51, Pro-52, Tyr-54, Phe-140, Leu-141, Asn-142, Gly-143, Ser-144, Cys-145, Glu-163, His-164, Met-165, Glu-166, His-172, Asp-187, Arg-188, and Gln-189 from promoter A and the residues Phe-3, Arg-4, Lys-5, Met-6, and Ala-7 from promoter B. But in the practical calculation, the actual region involved contains much more than the above-listed amino acid residues. It includes all the amino acid residues within 5 Å surrounding the octapeptide [9,18].

Fig. 3A shows the catalytic active region of SARS CoV M^{PRO} and the location of octapeptide (white). The catalytic dyad of His-41 and Cys-145 is shown by a ball and stick drawing, the peptide bond Gln–Ser is in green, and the N finger in pink. Fig. 3B gives a close look at the bridge water molecule (W) between the SARS CoV M^{PRO} and the octapeptide. The distance O(W)–C(CO) is 4.70 Å, H₁(W)–N_{δ1}(His-41) is 2.39 Å, H₂(W)–N(NH) is 4.20 Å, and H(NH)–S(Cys-145) is 4.12 Å. The bridge

water molecule plays an important role in the proteolytic reaction. These results are derived from the complex structure obtained by following the aforementioned docking and energy-minimization procedures.

Hydrophobic–hydrophilic interaction

The hydrophilic and hydrophobic complementary interaction is important in the ligand–receptor interaction. According to the α sphere model [25,26], the active region of proteinase consists of a series of α spheres. Each interaction site between the ligand and receptor contains one or more α spheres. There must be at least one hydrophobic interaction site. Figs. 4A and B are the hydrophilic and hydrophobic surfaces of octapeptide AVLQSGFR and the catalytic active region of SARS CoV M^{PRO}, respectively. In the figures the green parts are hydrophobic surfaces and the blue parts are hydrophilic surfaces. There is good hydrophilic and hydrophobic complementary interaction between ligand and receptor: the hydrophilic parts of the ligand correspond to the hydrophilic parts of the receptor and the hydrophobic parts of the ligand correspond to the hydrophobic parts of the receptor. The octapeptide NH₂–AVLQ↓SGFR–COOH has four hydrophobic amino acid residues: R4(Ala), R3(Val), R3'(Phe), and R4'(Arg). Among them R4(Ala) and R3(Val) are exposed in solvent and R3'(Phe) and R4'(Arg) are covered by hydrophobic surfaces of the proteinase. The peptide bond to be cleaved between R1(Gln) and R1'(Ser) is just in the join point between the exposed part and the covered part.

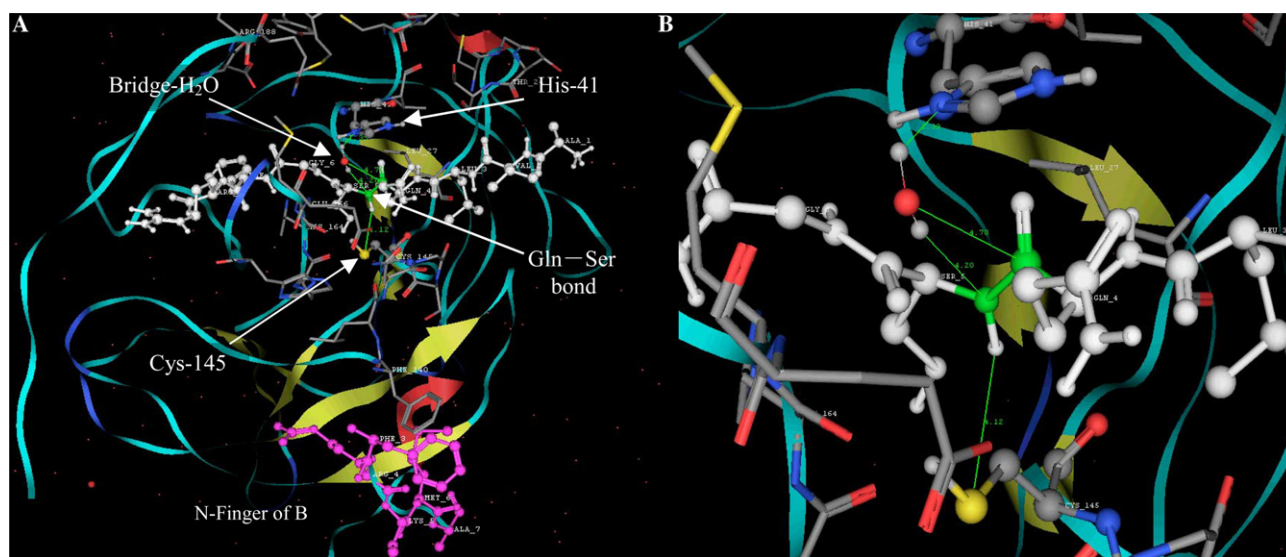


Fig. 3. (A) Catalytic active region of SARS CoV M^{PRO} and the location of octapeptide (white). The catalytic dyad of His-41 and Cys-145 is shown in ball and stick form, the peptide bond Gln–Ser is in green, and the N finger is in pink. (B) Bridge water molecule (W) between SARS CoV M^{PRO} and octapeptide. The distance O(W)–C(CO) is 4.70 Å, H₁(W)–N_{δ1}(His-41) is 2.39 Å, H₂(W)–N(NH) is 4.20 Å, and H(NH)–S(Cys-145) is 4.12 Å.

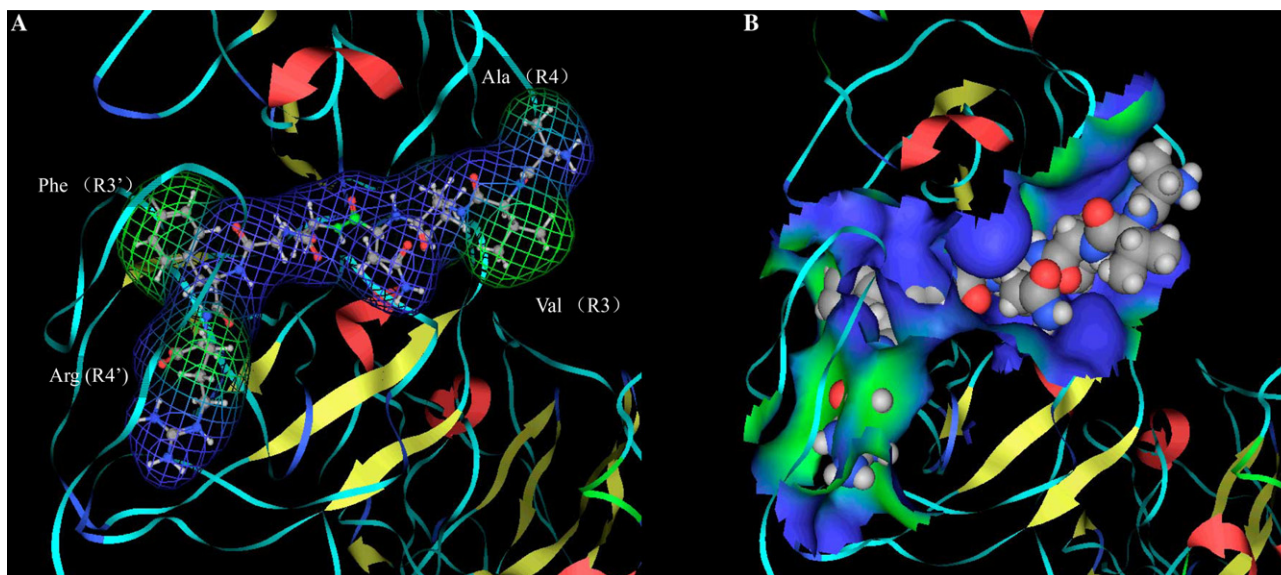


Fig. 4. (A) Hydrophilic and hydrophobic surfaces of octapeptide NH-AVLQ↓SGFR-COOH. The green parts are hydrophobic surfaces and the blue parts are hydrophilic surfaces. The four hydrophobic amino acid residues are R4(Ala), R3(Val), R3'(Phe), and R4'(Arg). (B) Hydrophilic and hydrophobic surfaces of SARS CoV M^{P_{ro}}. The hydrophobic residues R4(Ala) and R3(Val) are exposed in solvent, while R3'(Phe) and R4'(Arg) are covered by hydrophobic surfaces of the proteinase.

Quantum chemical calculations

The calculations were focused on the effect of SARS CoV M^{P_{ro}} on the peptide bond to be cleaved. How to properly treat a biomacromolecule and the solvent effects is a nontrivial problem in quantum chemical calculations. Electrostatic interaction plays the dominant role in ligand–receptor combinations and must be taken into account in the quantum mechanical study. In view of this, we treated the SARS CoV M^{P_{ro}} as a background with many point charges. The amino acid residues in the catalytic region between domains I and II of protomer A were divided into seven segments, and their atomic charges were separately calculated using the ab initio quantum chemical HF/6-31G* method. The 5 amino acid residues of the N finger of protomer B were taken into account also. The 76 amino acid residues in eight segments are listed in Table 1. There are 1148 atomic charges in the background, including all atoms in the catalytic cleft; the overlapped atoms are not counted.

Another difficult problem for quantum mechanical calculation for a biological system in solution is the solvent effect. In the PDB data [8] of SARS CoV M^{P_{ro}} there

are 300 solvation water molecules. These water molecules not only provide the solvation environment but also indicate that some water molecules are involved in the catalytic reaction. In our calculation the 300 solvation water molecules were treated as the background charges. Each water molecule has 3 point charges. Therefore, there are $1148 + 300 \times 3 = 2048$ background point charges.

In Table 2 we list the Mulliken atomic charges q^{Mull} and electrostatic potential equivalent charges q^{ESP} of His-41, which are obtained by fitting atomic charges to the electrostatic potential at the van der Waals surface. The atomic charges from semiempirical method AM1 are quite different from the results by ab initio HF/6-31G* [23]; in particular; the q^{ESP} charges of AM1 do not appear reasonable. Because the HF/6-31G* q^{ESP} charges reproduce the quantum mechanical electrostatic potential on the molecular surface, in this research we use q^{ESP} from the HF/6-31G* calculations to illustrate our points. The imidazole group of His-41 is a nucleophilic agent and its $\text{p}K = 6.0$, near the concentration of proton $[\text{H}^+]$ in pure water. The imidazole ring is both a proton donor and a receptor. It has good catalytic activity in the physiological condition [27]. It can be seen

Table 1
Distribution of amino acid residues in the active cleft of SARS-CoV M^{P_{ro}}: 76 amino acid residues and 1148 atoms are included

Domain I of A			Domain II of A				B
A1-1	A 1-2	A 1-3	A 2-1	A 2-2	A 2-3	A 2-4	B1-1
20–30	40–50	49–59	140–150	160–169	168–178	181–190	1–5

Table 2
Atomic charges of amino acid residue His-41 of SARS CoV M^{Pro}, in atomic unit *e*

Atom	AM1 q^{Mull}	AM1 q^{ESP}	HF/6-31G q^{Mull}	HF/6-31G q^{ESP}
N	-0.5314	0.5569	-0.9054	-0.4429
H(N)	0.2987	-0.5387	0.4113	0.1614
C $_{\alpha}$	-0.05334	2.0820	-0.0462	0.1192
H $_{\alpha}$	0.2611	-0.7777	0.3081	0.2461
C	0.3174	-1.9381	0.7638	0.3254
O	-0.3257	0.0303	-0.5750	-0.5518
C $_{\beta}$	-0.2302	-1.2490	-0.3088	-0.1595
H $_{\beta 3}$	0.2633	-0.1937	0.3268	0.2057
H $_{\beta 2}$	0.1628	0.1624	0.1903	0.0576
C $_{\gamma}$	-0.0590	1.7519	0.2424	0.2770
N $_{\delta 1}$	-0.2861	-1.7978	-0.8863	-0.4094
H $_{\delta 1}(+)$	0.4132	0.7184	0.4821	0.4318
C $_{\delta 2}$	-0.1669	1.5835	0.1266	0.0852
H $_{\delta 2}$	0.3390	-1.5772	0.3350	0.0992
C $_{\epsilon 1}$	0.0476	0.1000	0.5776	0.2349
H $_{\epsilon 1}$	0.3489	1.6428	0.3430	0.2178
N $_{\epsilon 2}$	-0.2709	-3.4887	-0.9061	-0.4587
H $_{\epsilon 2}$	0.4020	2.1826	0.46957	0.4604

from Table 2 that the negative charge of N $_{\delta 1}$ (-0.4094) and the positive charge of H $_{\delta 1}^+$ (0.4318) are large in magnitude and are close to the peptide bond Gln-Ser. Therefore, the proton H $_{\delta 1}^+$ attracts the π electron density from the peptide bond Gln-Ser so as to weaken this chemical bond. The polar nitrogen N $_{\delta 1}$ attracts one

hydrogen H $_1$ of bridge water and makes the water molecule dissociated, causing the nucleophilic attack of hydroxyl group OH $^-$ to the C(CO) and the electrophilic attack of H $^+$ to N(NH) on the two sides of peptide bond Gln-Ser. The sulfur atom S in Cys-145 forms a hydrogen bond with H(NH) and helps the proteolytic reaction (Fig. 3B).

Chemical modification of cleavable peptides

The peptide bond is considered a pseudo π bond. The six atoms in the peptide bond (Gln)C $_{\alpha}$ CO-NHC $_{\alpha}$ (Ser) are almost in a plane. Fig. 5A shows the electron density contour map of the peptide bond Gln-Ser in the gaseous phase on the π plane containing the carbonyl group CO of Gln and the NH group of Ser. The electron density contours surrounding atoms N(NH) and C(CO) form two triangles reminiscent of the hybrid orbits of the sp 2 type, and the 2p $_z$ orbitals, which are perpendicular to the plane and form a π_3^4 bond system. In Table 3, we list the calculated atomic charges of the six atoms on both sides of the peptide bond Gln-Ser obtained from the ab initio HF/6-31G* calculations in the gaseous phase with the background charges of SARS-CoV M^{Pro}. The negative charge of N(NH) of serine in the octapeptide increases to -0.8612 in the background of proteinase charges from the original value of -0.8534 in the gaseous phase. Similarly, the corresponding

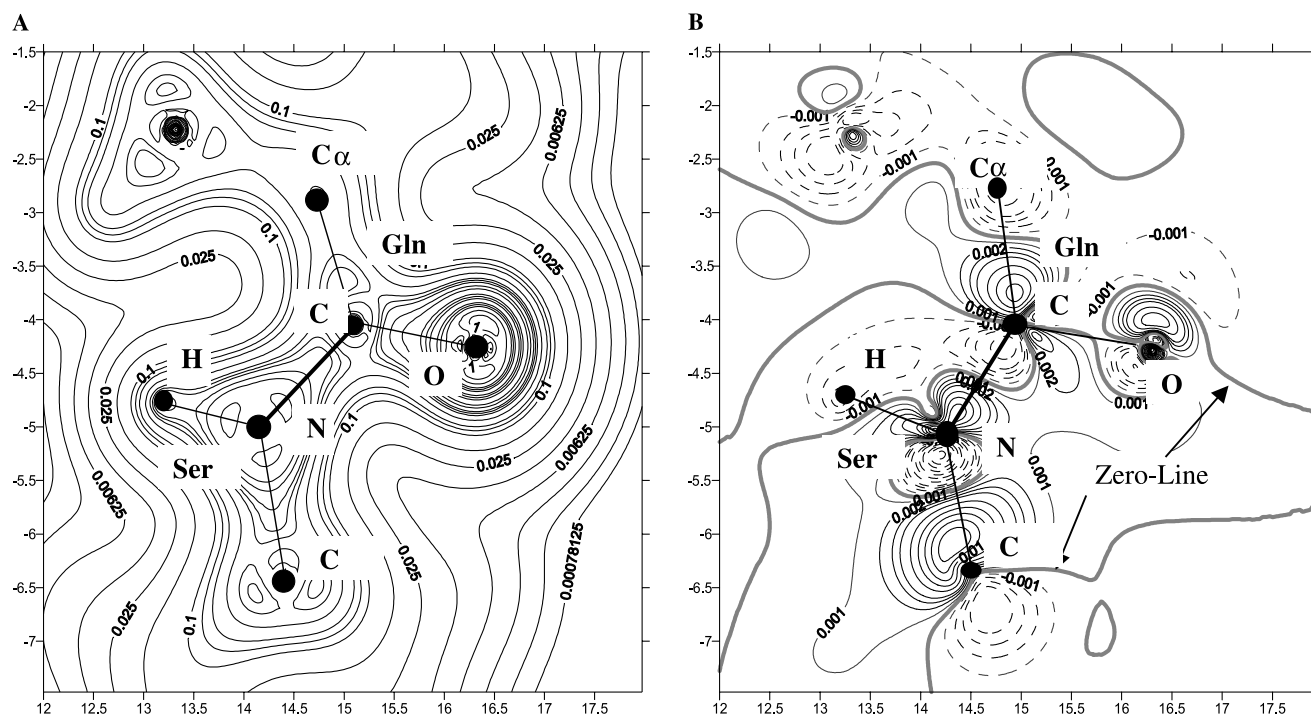


Fig. 5. (A) Electron density contour map of the peptide bond Gln-Ser in the π plane containing the carbonyl C and O atoms of Gln and N(NH) of Ser, as calculated in the gaseous phase. (B) Contour map of the differential electron density of peptide bond Gln-Ser in octapeptide AVLQSGFR, obtained by subtracting the electron density in the gaseous phase from the electron density in the proteinase background and solvent water.

Table 3

Atomic charges of relevant atoms in peptide bond Gln–Ser of octapeptide, in atomic unit e

Phase	Gln side			Ser side		
	C α	C(CO)	O(CO)	C α	N(NH)	H(NH)
$q^{\text{ESP}}(\text{Gas})$	−0.0133	0.8534	−0.6252	0.4124	−0.8050	0.3528
$q^{\text{ESP}}(\text{M}^{\text{pro}})$	0.0020	0.8612	−0.6219	0.3842	−0.8190	0.3497

positive charge of the carbonyl carbon C(CO) in glutamine of the octapeptide increases to 0.8190 from the original 0.8050. This is favorable to the electrophilic attack of OH[−] to C(CO) and the nucleophilic attack of H⁺ to N(NH) in the proteolytic reaction.

Fig. 5B is the contour map of differential electronic density of the peptide bond Gln–Ser in the octapeptide AVLQSGFR, obtained by subtracting the electron density in the gaseous phase from the electron density in the background [23] of SARS CoV M^{pro} and solvent water molecules. In Fig. 5B the bold gray line is the 0 value line where the electron densities in the gaseous phase and in the protease background are equal. The solid thin lines show the regions where the computed electron densities are greater in the SARS CoV M^{pro} background than in the gaseous phase, and the dashed thin lines show the areas where computed electron densities are smaller in the proteinase background than in the gaseous phase. We find that along the peptide bond between (Gln)C–N(Ser) the electron densities increase on the N(Ser) side and decrease on the C(Gln) side. This change is favorable for a nucleophilic attack of anion OH[−] to (Gln)C and for an electrophilic attack of cation H⁺ to N(Ser).

SARS CoV M^{pro} has very high cleavage specificity and is highly conserved in the coronavirus family [7]. The cleavage specificity of SARS CoV M^{pro} concentrates on the sites R2, R1, and R1'. The development of peptides as clinically useful drugs is limited by their poor stability and bioavailability, which is due in part to their inability to readily cross membrane barriers such as the intestinal and blood–brain barriers. Systematic chemical modification strategies that convert peptides

into drugs are an attractive research topic in current medicinal chemistry [28]. For the peptide inhibitor of proteinase, the chemical modification to cleavable octapeptide should focus on the scissile peptide bond between R₁ and R_{1'} to be cleaved by SARS CoV M^{pro}. A successful example shows that after the peptide bond CO=NH is replaced by reduced single bond CH₂–NH, the modified octapeptide rennin inhibitors have shown greatly increased affinity for rennin and strong resistance to enzyme hydrolysis [29].

In this study we suggest a similar modification to octapeptide AVLQSGFR. In the peptide bond (Gln)CO=NH(Ser) the carboxyl group CO of glutamine on subsite R₁ is replaced by CH₂ or CF₂ or the NH of serine on subsite R_{1'} is replaced by CH₂; then the π bond system is broken and the modified octapeptide AVLQSGFR may become a stable inhibitor for SARS-CoV M^{pro} and an effective drug candidate against SARS. Here we show the possibility of the chemical modification through computational modeling.

Table 4 lists the atomic charges of the six atoms on both sides of the hybrid peptide bond Gln–Ser after the chemical modification. Comparing with Table 3, we find that after CO is replaced by CH₂, the charge of the carbon atom in CH₂ turns negative (−0.6572) from originally a positive value (0.8612) in the CO group; hence the nucleophilic attack by OH[−] becomes impossible. Furthermore, the negative charge of N(NH) on the Ser side decreases to −0.2892 from −0.8190 in Table 3, and hence the electrophilic attack by H⁺ also becomes more difficult. Likewise, the modifications CO → CF₂ and NH → CH₂ will make the proteolytic reaction more difficult also.

Table 4

Atomic charges in peptide bond Gln–Ser after modification, in atomic unit e

	Gln side			Ser side		
	C α	C(CH ₂)	H(CH ₂)	C α	N(NH)	H(NH)
CO → CH ₂						
$q^{\text{ESP}}(\text{Gas})$	0.0770	−0.6756	0.1351	−0.1689	−0.2473	0.1910
$q^{\text{ESP}}(\text{M}^{\text{pro}})$	0.0867	−0.6572	0.1251	−0.2121	−0.2892	0.1807
CO → CF ₂						
$q^{\text{ESP}}(\text{Gas})$	0.0814	0.3679	−0.0625	0.2955	−0.6854	0.3143
$q^{\text{ESP}}(\text{M}^{\text{pro}})$	0.1182	0.3092	−0.0653	0.2597	−0.6483	0.3085
NH → CH ₂						
$q^{\text{ESP}}(\text{Gas})$	0.0311	0.8581	−0.5980	0.4387	−0.6707	0.1546
$q^{\text{ESP}}(\text{M}^{\text{pro}})$	0.0551	0.8437	−0.5940	0.4086	−0.6460	0.1511

Discussion

Searching for cleavable peptides is the first step in finding potential inhibitors against the proteinase. Even within a limited range of octapeptides, it is by no means an easy job because the possible different patterns of octapeptides are as many as $20^8 = 2.56 \times 10^{10}$. However, some algorithms, such as the vectorized sequence-coupled algorithm [14] and the discriminant function algorithm [15], can be successfully used to narrow the search scope. Meanwhile, similar methods that can be used to search the cleavable peptides by SARS enzyme have been developed [5,6].

The octapeptide NH₂-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-COOH is the first designed on the basis of the molecular structure of SARS-CoV M^{Pro} [9] and it has been experimentally demonstrated to be a cleavable peptide with high bioactivity [12]. A detailed study of the cleavage mechanism needs a great deal of experimental work. However, computational simulation may provide useful insights for an in-depth understanding of the mechanism. The cleavable octapeptide could be changed into an effective inhibitor of SARS-CoV M^{Pro} or a drug candidate against SARS after a proper chemical modification to replace the cleavable peptide bond with a non-cleavable bond. The octapeptide thus modified will lose its susceptibility although it can still tightly bind to the active site and become a competent inhibitor, just like the role of a distorted key in blocking the access of other keys to a lock, as elucidated by Chou [15] in a review paper.

Acknowledgments

This work was supported by grants from the Chinese National Science Foundation under Contract No. 20373048 and the Tianjin Commission of Sciences and Technology under Contract Nos. 033801911 and 023618211.

References

- [1] C. Drosten, S. Gunther, W. Preiser, S. van der Werf, H.R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A. Fouchier, Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *N. Engl. J. Med.* 348 (2003) 1967–1976.
- [2] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, A novel coronavirus associated with severe acute respiratory syndrome, *N. Engl. J. Med.* 348 (2003) 1953–1966.
- [3] K.F. Shortridge, Severe acute respiratory syndrome and influenza: virus incursions from southern China, *Am. J. Respir. Crit. Care Med.* 168 (2003) 1416–1420.
- [4] H.S. Miura, K. Nakagaki, F. Taguchi, N-terminal domain of the murine coronavirus receptor CEACAM1 is responsible for fusogenic activation and conformational changes of the spike protein, *J. Virol.* 78 (2004) 216–223.
- [5] L.-L. Chen, H.-Y. Ou, R. Zhang, C.-T. Zhang, 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 (2003) 382–388.
- [6] F. Gao, H.-Y. Ou, L.-L. Chen, W.-X. Zheng, C.-T. Zhang, Prediction for proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes, *FEBS Lett.* 553 (2003) 451–456.
- [7] K. Anand, J. Ziebuhr, P. Wadhvani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CL^{Pro}) structure: basis for design of anti-SARS drugs, *Science* 300 (2003) 1763–1767.
- [8] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, *Proc. Natl. Acad. Sci. USA* 100 (2003) 13190–13195.
- [9] K.-C. Chou, D.-Q. Wei, W.-Z. Zhong, Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS, *Biochem. Biophys. Res. Commun.* 308 (2003) 148–151 (Erratum: *ibid* 2003, Vol. 310, 675).
- [10] K.C. Chou, Review: structural bioinformatics and its impact to biomedical science, *Curr. Med. Chem.* 11 (2004) 2105–2134.
- [11] S. Sirois, D.Q. Wei, Q. Du, K.-C. Chou, Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points, *J. Chem. Inf. Comput. Sci.* 44 (2004) 1111–1122.
- [12] Y.R. Gan, Y.D. Huang, H. Huang, Synthesis and biological evaluation of a novel SARS-CoV M^{Pro} inhibitor, *Peptides* (2004), submitted.
- [13] Q.S. Du, S.Q. Wang, Y. Zhu, D.Q. Wei, H. Guo, S. Sirois, K.C. Chou, Polyprotein cleavage mechanisms of SARS CoVM^{Pro} and chemical modification of the octapeptide, *Peptide* 25 (2004) 1857–1864.
- [14] K.C. Chou, A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins, *J. Biol. Chem.* 268 (1993) 16938–16948.
- [15] K.C. Chou, Review: prediction of HIV protease cleavage sites in proteins, *Anal. Biochem.* 233 (1996) 1–14.
- [16] K.C. Chou, D. Jones, R.L. Heinrikson, Prediction of the tertiary structure and substrate binding site of caspase-8, *FEBS Lett.* 419 (1997) 49–54.
- [17] J.J. Chou, H. Matsuo, H. Duan, G. Wagner, Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment, *Cell* 94 (1998) 171–180.
- [18] K.C. Chou, K.D. Watenpaugh, R.L. Heinrikson, A Model of the complex between cyclin-dependent kinase 5(Cdk5) and the activation domain of neuronal Cdk5 activator, *Biochem. Biophys. Res. Commun.* 259 (1999) 420–428.
- [19] J.J. Chou, H. Li, G.S. Salvesen, J. Yuan, Solution structure of BID, an intracellular amplifier of apoptotic signaling, *Cell* 96 (1999) 615–624.
- [20] K.C. Chou, A.G. Tomasselli, R.L. Heinrikson, Prediction of the tertiary structure of a caspase-9/inhibitor complex, *FEBS Lett.* 470 (2000) 249–256.
- [21] K.C. Chou, W.J. Howe, Prediction of the tertiary structure of the beta-secretase zymogen, *Biochem. Biophys. Res. Commun.* 292 (2002) 702–708.
- [22] J. Zhang, C.H. Luan, K.C. Chou, Identification of the N-terminal functional domains of Cdk5 by molecular truncation and computer modeling, *Proteins Struct. Funct. Genet.* 48 (2002) 447–453.
- [23] J.B. Foresman, A. Frisch, *Exploring chemistry with electronic structure methods: a guide to using Gaussian*, Gaussian, Pittsburgh, PA, 1993.
- [24] MOE (Molecular Operating Environment), Chemical Computing Group Inc., Toronto, Canada, 1997–2004.

- [25] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, *J. Med. Chem.* 28 (1985) 849–856.
- [26] A. Miranker, M. Karplus, Functionality maps of binding sites: a multiple copy simultaneous search method, *Proteins Struct. Funct. Genet.* 11 (1991) 29–34.
- [27] L. Stryer, *Biochemistry*, third ed., W.H. Freeman, New York, 1998.
- [28] C. Adessi, C. Soto, Converting a peptide into a drug: strategies to improve stability and bioavailability, *Curr. Med. Chem.* 9 (2002) 963–978.
- [29] M. Szelke, B.J. Leckie, M. Tree, A. Brown, J. Grant, A. Hallett, M. Hughes, D.M. Jones, A.F. Lever, A potent new renin inhibitor. In vitro and in vivo studies, *Hypertension* 4 (1982) 59.