ELSEVIER

The substrate specificity of SARS coronavirus 3C-like proteinase

Keqiang Fan [a,1], Liang Ma [a], Xiaofeng Han [a,b], Huanhuan Liang [a],
Ping Wei [a], Ying Liu [a], Luhua Lai [a,b,*]

[a] State Key Laboratory of Structural Chemistry of Stable and Unstable Species, College of Chemistry, Peking University, Beijing 100871, China
[b] Center for Theoretical Biology, Peking University, Beijing 100871, China

Abstract

The 3C-like proteinase of severe acute respiratory syndrome coronavirus (SARS) has been proposed to be a key target for structural based drug design against SARS. We have designed and synthesized 34 peptide substrates and determined their hydrolysis activities. The conserved core sequence of the native cleavage site is optimized for high hydrolysis activity. Residues at position P4, P3, and P3′ are critical for substrate recognition and binding, and increment of β-sheet conformation tendency is also helpful. A comparative molecular field analysis (CoMFA) model was constructed. Based on the mutation data and CoMFA model, a multiply mutated octapeptide S24 was designed for higher activity. The experimentally determined hydrolysis activity of S24 is the highest in all designed substrates and is close to that predicted by CoMFA. These results offer helpful information for the research on the mechanism of substrate recognition of coronavirus 3C-like proteinase.
© 2005 Elsevier Inc. All rights reserved.

Keywords: SARS; Coronavirus; 3C-like proteinase; substrate specificity; CoMFA

Human coronaviruses are major causes of upper respiratory tract illness in humans. A novel form of coronavirus has been identified as the major cause of severe acute respiratory syndrome (SARS), a disease that was rapidly spreading from southern China to several countries in 2003 [1,2]. Coronaviruses are members of positive-stranded RNA viruses featuring the largest viral RNA genomes up to date. The SARS coronavirus replicase gene encompasses two overlapping translation products, polyprotein 1a (~450 kDa) and 1ab (~750 kDa), which are conserved both in length and amino acid sequence to other coronavirus replicase proteins. Polyprotein 1a and 1ab are cleaved by the internally encoded 3C-like proteinase to release functional proteins necessary for virus replication. The SARS 3C-like proteinase is fully conserved among all released SARS coronavirus genome sequences and is highly homologous with other coronavirus 3C-like proteinases. Due to the functional importance of SARS 3C-like proteinase in the viral life cycle, it has been proposed to be a key target for structural based drug design against SARS [3].

The crystal structures of the 3C-like proteinases of human coronaviruses (HCoV), transmissible gastroenteritis virus (TGEV), and SARS-associated coronavirus (SARS-CoV) have been solved, all of which are dimeric [3–5]. The structure of coronavirus 3C-like proteinase contains three domains, the first two domains form a chymotrypsin fold, which is responsible for the catalytic reaction, and the extra helical domain of the enzyme plays an important role in controlling the association–dissociation equilibrium and regulating the activity and specificity of the enzyme [6]. The crystal structures of TGEV and SARS-CoV 3C-like proteinases in complex with a hexapeptidyl chloromethyl ketone inhibitor

---

have also been solved [3,4], which afford the structural insights into the substrate specificity of coronavirus 3C-like proteinase. As shown in the crystal structure of the TGEV 3C-like proteinase—inhibitor complex, the well-defined S1, S2, and S4 specificity pockets indicate residues at P1, P2, and P4 positions in the substrate are critical for substrate binding [3].

The substrate specificity of coronavirus 3C-like proteinase is determined mainly by the P1, P2, and P1′ positions [7]. The P1 position has a well-conserved Gln residue and P2 a hydrophobic one. Unlike other previously identified coronavirus 3C-like proteinases, which have Leu/Ile at position P2, SARS 3C-like proteinase also tolerates Phe, Val, and Met residues at P2 position. P1′ position requires a small aliphatic residue, usually Ser or Ala. Conserved substrate selectivity among different coronavirus 3C-like proteinases is determined using an HPLC-based peptide cleavage assay by Hegyi and Ziebuhr [7]. This result indicates the possibility of both the development of universally applicable 3C-like proteinase assay and the design of broad-spectrum inhibitors blocking all coronavirus 3C-like proteinases. Coronavirus 3C-like proteinases share the chymotrypsin fold part and similar substrate specificities with the 3C proteinases from other viruses like rhinovirus, picornavirus [3,8,9].

In order to study the substrate specificity of SARS 3C-like proteinase, we have cloned, expressed, and purified the protein, and studied its activity towards 11 peptides covering the 11 cleavage sites on the virus polyproteins. Our results confirm that purified SARS 3C-like proteinase is active towards the substrate peptides mapped from the cleavage sites on the polyprotein and reveals the relationship between the reaction activity and the secondary structure content of the substrates [10]. In the current study, we have studied the substrate requirements for SARS 3C-like proteinase binding using 34 truncated and mutated substrate peptides. This study helps one to understand the mechanism of substrate recognition and catalysis of coronavirus 3C-like proteinases.

Materials and methods

Peptides synthesis. The natural substrate peptide S12 was designed based on the N-terminal self-cleavage site of SARS 3C-like proteinase, with the sequence of: Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-CONH₂. Other substrates were designed by truncating the N- and C-terminals, or mutation of S12. All the designed substrate peptides were synthesized by solid-phase peptide synthesis using standard 9-fluorenylmethoxycarbonyl/*tert*-butyl strategy. Cleavage of the peptide from Rink resin and removal of all sidechain-protecting groups were achieved in trifluoroacetic acid solution. The crude peptide was purified by reversed-phase high performance liquid chromatography (RP-HPLC, Elite P200II, Dalian, China) on a Zorbax C18 semi-preparative column (9.4 by 250 mm, Agilent) with gradients of water/acetonitrile containing 0.1% trifluoroacetic acid. Peptide homogeneity and identity were analyzed by analytical RP-HPLC, and matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy, respectively.

Peptide cleavage. The reaction activities of the designed substrate peptides were determined using HPLC-based peptide cleavage assay. The C-terminal His-tagged SARS 3C-like proteinase was expressed and purified as previously described [10], which was shown to have comparable enzyme activities with the non-His-tagged protein [11]. Cleavage assay solution was incubated at room temperature, which contains 0.2–0.4 mM substrate peptide, 5.41, 7.41 or 27.05 μM enzyme, and 57 μM DTT in 40 mM Tris–HCl buffer, pH 7.3. Aliquots of reactions were removed every 10 or 60 min within 1–7 h, stopped by the addition of 0.1% trifluoroacetic acid aqueous solution, and analyzed by HPLC (LabPrep System, Gilson) on a Zorbax C18 analytic column (4.6 by 250 mm, Agilent). Cleavage products were resolved using a 15-min, 10–60% linear gradient of acetonitrile in 0.1% trifluoroacetic acid, as described previously [10]. k_{cat}/K_m was determined by plotting substrate peak area using the equation below:

$$\ln PA = C - (k_{cat}/K_m)_{app} c_E t, \tag{1}$$

where PA is the peak area of the substrate peptide, c_E is the total concentration of His-tagged 3C-like proteinase, and C is an experimental constant and averaged for two independent measurements.

Results

Substrate design and peptide cleavage assay

Previous studies have indicated that coronavirus 3C-like proteinases share a highly conserved substrate core sequence. We have studied the activities of SARS 3C-like proteinase towards 11 peptides covering the 11 cleavage sites on the SARS coronavirus polyproteins 1a and 1ab [10]. However, the 11 native cleavage sequences are not enough to offer detailed information about the residue specificity at each substrate position. Here we have designed 34 mutated and truncated peptides based on the N-terminal self-cleavage site of SARS 3C-like proteinase, among which 28 are octapeptides and the other 6 are shorter peptides. The native substrate S12 has the sequence of: Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-CONH₂. Different mutations are designed at position P5 to P1′.

The dimeric form of the SARS 3C-like proteinase has been proved to be the major form for biological activity by various researchers [6,10,12]. The apparent k_{cat}/K_m is highly enzyme concentration-dependent. In this study, the proteolytic activities of the substrates are determined at constant proteinase concentrations (5.41 μM for most octapeptides, 7.41 and 27.05 μM for truncated and P2 mutated substrates due to their low activities), using an HPLC-based peptide cleavage assay. The results are listed in Table 1, and only relative activities are used in further analysis (see Fig. 1).

The substrate core sequence (position P2 to P1′) of SARS 3C-like proteinase is highly conserved

The coronavirus 3C-like proteinase recognizes a highly conserved core sequence (position P2 to P1′) of Leu-Gln-(Ser/Ala) [3–5,7,13,14]. In this study, residues

Table 1
Enzyme cleavage activities of mutated substrate peptides

Substrate	Sequence[a]	Apparent k_{cat}/K_m ($mM^{-1}\,min^{-1}$)[b]	$(k_{cat}/K_m)_{rel}$
S12	SAVLQ ↓ SGF-CONH$_2$	2.13 ± 0.20	1.00
P5L	LAVLQ ↓ SGF-CONH$_2$	8.32 ± 0.25	3.90
P5T	TAVLQ ↓ SGF-CONH$_2$	7.80 ± 0.24	3.66
P5V	VAVLQ ↓ SGF-CONH$_2$	7.66 ± 0.30	3.59
P5A	AAVLQ ↓ SGF-CONH$_2$	7.10 ± 0.24	3.33
P4L	SLVLQ ↓ SGF-CONH$_2$	0.33 ± 0.04	0.15
P4T	STVLQ ↓ SGF-CONH$_2$	3.14 ± 0.12	1.47
P4V	SVVLQ ↓ SGF-CONH$_2$	5.21 ± 1.27	2.44
P3L	SALLQ ↓ SGF-CONH$_2$	1.85 ± 0.06	0.87
P3T	SATLQ ↓ SGF-CONH$_2$	2.53 ± 0.54	1.19
P3A	SAALQ ↓ SGF-CONH$_2$	0.40 ± 0.03	0.19
P3K	SAKLQ ↓ SGF-CONH$_2$	5.72 ± 0.45	2.68
P1′A	SAVLQ ↓ AGF-CONH$_2$	4.36 ± 0.19	2.04
P1′G	SAVLQ ↓ GGF-CONH$_2$	1.76 ± 0.09	0.83
P1′L	SAVLQ ↓ LGF-CONH$_2$	ND[c]	
S12	SAVLQ ↓ SGF-CONH$_2$	12.68 ± 0.60*	1.000
P2M	SAVMQ ↓ SGF-CONH$_2$	2.54 ± 0.07*	0.208
P2F	SAVFQ ↓ SGF-CONH$_2$	0.58 ± 0.10*	0.046
P2I	SAVIQ ↓ SGF-CONH$_2$	0.080 ± 0.011*	0.0063
P2V	SAVVQ ↓ SGF-CONH$_2$	0.070 ± 0.009*	0.0056
P2A	SAVAQ ↓ SGF-CONH$_2$	0.058 ± 0.008*	0.0046
P2R	SAVRQ ↓ SGF-CONH$_2$	ND[c,*]	
P1N	SAVLN ↓ SGF-CONH$_2$	ND[c,*]	
P1E	SAVLE ↓ SGF-CONH$_2$	ND[c,*]	
P1K	SAVLK ↓ SGF-CONH$_2$	ND[c,*]	
S21	TVVLQ ↓ SGF-CONH$_2$	8.48 ± 1.32	3.98
S22	TVTLQ ↓ SGF-CONH$_2$	5.58 ± 0.21	2.62
S23	VVTLQ ↓ SGF-CONH$_2$	5.18 ± 0.74	2.43
S24	TVKLQ ↓ AGF-CONH$_2$	9.18 ± 0.25	4.31

[a] Mutated residues are underlined. Cleavage sites are indicated by ↓.
[b] The concentration of SARS 3C-like proteinase is 5.41 μM for substrates without any label and that for substrates labeled with asterisk is 27.05 μM.
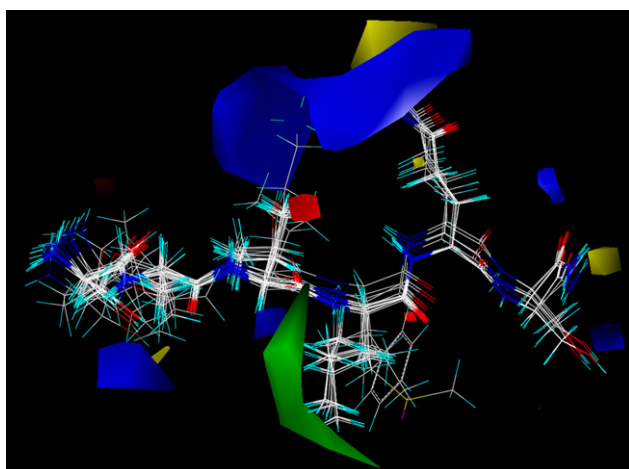[c] Not detectable in HPLC-based peptide cleavage assay.

Fig. 1. The superimposed 22 substrate structures and the contour plot of the CoMFA model. This result indicates that increasing positive charge at position P3 is favored (blue), and large hydrophobic residue at position P2 is favored (green), which is compatible with the crystal structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

at position P2, P1, and P1′ were mutated to amino acids with similar or different properties. The activities of mutated substrates have indicated that the conserved core

sequence reveals high hydrolytic activity. Mutation at these positions will decrease the substrate activity significantly, or completely abolish its activity, such as P2R, P1′L, and P1 mutants (see Table 1).

P1′ position mutants with small residues Ser, Ala, and Gly, which are commonly found in this position in 3C-like proteinase cleavage sites, have comparable substrate activity. When it is replaced by a large Leu, no substrate peptide is cleaved as observed by RP-HPLC. This confirms that small aliphatic residues are favored at P1′ position.

The single mutation of Gln-P1 to Glu or Asn abolishes the substrate activity completely, which is compatible with the absolute specificity for the Gln-P1 substrate residue. This result strongly suggests that the Gln-P1 is critical for substrate binding and cleavage.

P2 position requires a large hydrophobic residue as shown in the native cleavage sites. When Leu-P2 is replaced by Arg, a positive-charged large residue, no cleavage is observed in HPLC-based assay. These results confirm the hydrophobic interaction between the P2 residue and the S2 pocket of the enzyme indicated by crystal structure [3,5]. Mutations to small residues such as Ala also decrease the relative reaction

activity of the substrate significantly. Although Phe, Val, and Met are tolerant at P2 for SARS 3C-like proteinase [10], the mutation of Leu-P2 to other hydrophobic residues decreases the substrate reaction activity by a factor of 5–200 (see Table 1). The relative reaction activity of mutated substrates was significantly decreased while the P2 residue is replaced by β-branched Ile and Val. This result indicates that the steric hindrance effect is important in the residual specificity at P2 position.

The residual selectivity of the substrate at position P5, P4, and P3

Our previous research has shown that the β-strand-like conformation of the substrate is favored for substrate binding and cleavage [10]. In this study, we have further studied the residue specificity and its relationship with the residue secondary structure tendency via substrate mutation. Residues at position P5–P3 are mutated to Ala, Leu, Val, and Thr, which differ in their polarities and secondary structure tendencies. The relative activities of the mutants are listed in Table 1.

It is interesting that all the four mutants at position P5 have significant increment in their activities by a factor of 3.33–3.90. This indicates that Ser is not the best residue for position P5. Position P5 shows no specificity among residues that have different secondary structure tendency, but the decrement of residue polarity seems helpful for the substrate activity.

When Ala-P4 is replaced by residues with higher β-sheet tendency, such as Val and Thr, the substrate activity increases obviously; on the other hand, similar mutation at position P3 only causes little increment in substrate activity. It is notable that the native substrate S12 has an α-helix-preferred residue Ala at position P4 and a β-sheet-preferred residue Val at P3. Single mutation of Ala-P4 to Val or Thr will increase the β-sheet tendency of the substrate significantly due to the tandem β-sheet-preferred residues at position P4 and Val at P3; however, single mutation of Val-P3 to other residues does not increase the β-sheet-preference of the substrate due to no such cooperation being found with the α-helix-preferred Ala-P4. The possibility of conformational cooperativity with its neighbors is in agreement with the activity increment, which explains the different activities of mutants at position P4 and P3. This is supported by the multiply mutated substrates S21–S23, which are designed for increasing the β-sheet-preference of the substrate by tandem-placed β-sheet-preferred residues and have highest activities among the tested substrates in Table 1.

It is notable that the mutation of Val-P3 to Lys (substrate P3K) causes 2.68-fold increment in its hydrolysis activity. As shown in the crystal structure, the side chain of Glu-166 of the enzyme is exposed to solvent and is near the edge of the specific pocket S3. It is possible that the formation of the salt bridge between the side chain of Lys-P3 and that of Glu-166 of the enzyme is helpful for substrate binding. This result suggests that an additional positive charge at position P3 will be propitious in the substrate or inhibitor design.

The relative reaction activities of truncated substrates indicate position P4, P3, and P3′ are important for substrate recognition

The contribution of each residue to the substrate activity is further investigated using truncated substrate based on the native substrate peptide S12. The activities of the native octapeptide and 6 truncated substrates are determined by HPLC peptide cleavage assay and are listed in Table 2.

As shown in Table 2, the activity of the substrate decreases after truncation. Longer substrates, which have more residues binding to the proteinase specific pockets, are favored in enzyme binding and hydrolysis. However, the contributions of residues at different positions are significantly different. Deletion of Ala-P4 and Val-P3 causes significant decrease in the activity of the substrate by a factor of 11.4 and 7.08, respectively, while the activity of substrate S17 (deleting Ser-P5) remains 76.3% compared with that of native octapeptide S12. This result suggests that the residues at position P4 and P3 are important for substrate recognition and binding. On the other hand, the residues at C-terminal of the substrate seem to be less important. As shown in Table 2, deletion of Phe-P3′ and Gly-P2′ decreases the substrate activity by a factor of 5.03 and 1.15, respectively. Phe-P3′ has large contribution to substrate hydrolysis, which may arise from the potential hydrophobic interaction between the aromatic side chain and the enzyme.

Comparative molecular field analysis and application in predicting new substrate

The quantitated $\lg(k_{cat}/K_m)$ of 22 substrates was used for comparative molecular field analysis

Table 2
Enzyme cleavage activities of truncated substrate peptides

Substrate	Sequence[a]	Apparent k_{cat}/K_m (mM^{-1} min^{-1})[b]	$(k_{cat}/K_m)_{rel}$
S12	SAVLQ ↓ SGF-CONH$_2$	7.85 ± 0.30	1.00
S13	AVLQ ↓ SG-CONH$_2$	1.19 ± 0.03	0.152
S14	VLQ ↓ SG-CONH$_2$	0.148 ± 0.006	0.0189
S15	AVLQ ↓ S-CONH$_2$	1.01 ± 0.02	0.129
S16	VLQ ↓ S-CONH$_2$	0.131 ± 0.006	0.0167
S17	AVLQ ↓ SGF-CONH$_2$	5.99 ± 0.08	0.763
S18	LQ ↓ SG-CONH$_2$	0.013 ± 0.004	0.0017

[a] Cleavage sites are indicated by ↓.
[b] The concentration of SARS 3C-like proteinase is 7.41 μM.

(CoMFA) [15]. The complex structures of SARS 3C-like proteinase and substrates were generated based on the crystal structure of the enzyme (PDB code: 1UK4) using Sybyl 6.91, followed by energy minimization using 500 steps of steepest descents method. The structures of the 22 substrates were superimposed based on their backbone atoms for CoMFA. The best orientation was searched by all-orientation search strategy using AOS/APS program [16], to minimize the effect of the initial orientation of the substrates. CoMFA was performed using the QSAR module in Sybyl 6.91. The standard grid spacing of 0.2 nm was chosen. The steric and electrostatic field energies were calculated using an sp^3 carbon probe atom with $+1$ charge. The steric and electrostatic energy cutoffs were set to 125.4 kJ/mol.

The partial linear square analysis, with a number of components of 5 and a column filtering of 8.36 kJ/mol, resulted a leave-one-out cross-validated q^2 of 0.661 and a number of optimum components of 2. The full data partial linear square analysis using the optimum component of two resulted in conventional correlate coefficient $r^2 = 0.822$, standard error of estimation $s = 0.429$, and $F_{2,20} = 46.213$. The relative contributions of steric and electrostatic fields in the QSAR equation are 81.8% and 18.2%, respectively. This result suggests that steric interaction is the major force for substrate recognition.

Designing highly active substrate

A new substrate octapeptide was designed by combining all hydrolysis-preferred factors, including tandem-placed β-sheet-preferred residues at position P5 and P4, a positive-charged Lys at P3, and a Ala at P1′ (see Table 1). The substrate was synthesized and its activity was determined by HPLC. The substrate S24 shows the highest hydrolysis activity among all 34 designed substrates, which was 4.31 times compared with the native sequence S12.

This substrate was used to test the predicting ability of the CoMFA model. The $lg(k_{cat}/K_m)$ of S24 predicted by our CoMFA model was 0.58, which is close to the experiment result of 0.634. This result strongly suggests that the CoMFA model can be used to predict the activity of newly designed peptide substrates. It also confirms all previous discoveries derived from the substrate activities directly, especially the relationship between the hydrolysis activity and the substrate secondary structure preference.

Discussion

Previous researches on the coronaviruses 3C-like proteinases via sequence analysis have suggested a highly conserved core sequence (position P2 to P1′) of Leu-Gln-(Ser/Ala) [3–5,7,13,14]. All the 11 cleavage sites in SARS coronavirus have a conserved Gln at position P1; 8 of them have a Leu at P2; and 9 have Ala or Ser at P1′. Similar result has been reported in the sequence analysis of 77 cleavage sites of 3C-like proteinases in different coronaviruses [13]. In this study, mutation at these positions decreased the substrate activity significantly or completely abolished its activity. This indicates that the conserved core sequence reveals high hydrolytic activity. Both the highly conserved substrate sequence and the specific enzyme might be the result of a cooperated evolutional process.

The crystal structures of TGEV and SARS coronavirus 3C-like proteinases in complex with a hexapeptidyl chloromethyl ketone inhibitor [3,4] afford the structural insights into the substrate specificity of coronavirus 3C-like proteinase. As shown in the crystal structure, the hydrogen bond formed between the side-chain amide group of Gln-P1 and the imidazole of a conserved His in the enzyme is critical for substrate recognition and binding [3,4]. Mutation of Gln-P1 to Glu or Asn is tolerant in 3C-like proteinase in other viruses, although the substrate activity may be decreased [17,18]. However, our result and the highly conserved Gln-P1 in native cleavage sites indicate that the residue specificity at position P1 is more critical for coronavirus 3C-like proteinases.

P2 position requires a large hydrophobic residue, especially Leu, as shown in the native cleavage sites. Although Phe, Val, and Met are tolerant at P2 for SARS 3C-like proteinase [10], the mutation of Leu-P2 to other hydrophobic residues decreases the substrate reaction activity dramatically (see Table 1). It is notable that the relative reaction activity of mutated substrates decreases from 0.208, 0.046 to about 0.006 while the P2 residue is replaced by no-branched Met, β-phenyl-containing Phe or β-branched Ile and Val. This result indicates that the steric hindrance effect may be critical at this position.

The crystal structure of TGEV 3C-like proteinase indicates that the substrate binds in the shallow substrate-binding site at the surface of the enzyme, between the domain I and II in the chymotrypsin fold. An antiparallel sheet is formed by the substrate peptide (P5–P3), the strand eII (164–167), and loop (189–191) of the enzyme, and two hydrogen bonds are formed between the backbone amides of Ala-P4, Val-P3, and the backbone carbonyls of Glu-165, Ser-189 of the proteinase [3]. Our results about the truncated substrates suggest the two hydrogen bonds are critical for substrate binding. On the other hand, the C-terminal residues seem to be less important as N-terminal residues. This is consistent with previous study on human rhinovirus 3C protease by Cordingley et al. [19]. Phe-P3′ has large contribution to substrate hydrolysis, which may arise

from the potential hydrophobic interaction between the aromatic side chain and the enzyme, and was confirmed in the virtual screen of SARS 3C-like proteinase inhibitors [20].

The crystal structure also indicates that increased β-sheet tendency of the substrate will be helpful for substrate binding, which has been revealed by our previous work [10]. The native substrate S12 has an α-helix-preferred Ala at position P4 and a β-sheet-preferred Val at position P3. The possibility of conformational cooperation with its neighbors will explain the different activities of mutants at position P4 and P3. Single mutation of Ala-P4 to Val or Thr will increase the β-sheet tendency of the substrate significantly due to the tandem β-sheet-preferred residues at position P4 and Val at P3; however, single mutation of Val-P3 to other residues does not increase the β-sheet-preference of the substrate due to no such cooperation being found with the α-helix-preferred Ala-P4. These results suggest that the hydrolysis activity may increase further by an additional β-sheet-preferred residue at position P5.

Substrates S21–S23 are designed with continued β-sheet-preferred residues at position P5–P3. Although no secondary structure analysis is done for these substrates due to their low solubility, their high tendency to aggregation in aqueous solution implies their high β-sheet content. This offers further evidence for the relationship between the secondary structure tendency and the activity of the substrate of coronavirus 3C-like proteinase.

A CoMFA model was generated using the relative activities of 22 octapeptide substrates. The predicting ability was tested with the multiply mutated substrate S24, which has the highest activity among all the tested substrates. The $lg(k_{cat}/K_m)$ of S24 predicted by our CoMFA model is close to the experiment. This result strongly suggests that the CoMFA model is useful for predicting the activity of newly designed peptide substrates for SARS 3C-like proteinase.

In summary, we have designed and synthesized 34 peptide substrates for SARS 3C-like proteinase, and determined their hydrolysis activities by HPLC-based peptide cleavage assay. A CoMFA model was generated and its prediction ability is proved by a newly designed peptide substrate S24. The core sequence of the proteinase cleavage site is highly conserved and optimized for enzyme recognition and catalysis. Residues at position P4, P3, and P3′ are critical for substrate recognition and binding, and increment of β-sheet conformation tendency at position P4 and P3 is helpful for substrate binding and hydrolysis. In addition, a salt bridge between position P3 and Glu-166 of the enzyme will increase the activity of the substrate. These results offer helpful information for understanding the mechanism of substrate recognition and catalysis of coronavirus 3C-like proteinase and the rational drug design against coronavirus.

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, A. Berger, A.M. Burguiere, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J.C. Manuguerra, S. Muller, V. Rickerts, M. Sturmer, S. Vieth, H.D. Klenk, A.D. Osterhaus, H. Schmitz, H.W. Doerr, 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, P.E. Uu, Rollin, S.F. Dowell, A.E. Ling, C.D. Humphrey, W.J. Shieh, J. Guarner, C.D. Paddock, P. Rota, B. Fields, J. DeRisi, J.Y. Yang, N. Cox, J.M. Hughes, J.W. LeDuc, W.J. Bellini, L.J. Anderson, A novel coronavirus associated with severe acute respiratory syndrome, N. Engl. J. Med. 348 (2003) 1953–1966.

[3] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLPro) structure: basis for design of anti-SARS drugs, Science 300 (2003) 1763–1767.

[4] H.T. Yang, M.J. Yang, Y. Ding, Y.W. Liu, Z.Y. Lou, Z. Zhou, L. Sun, L.J. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z.H. Rao, The crystal structure of severe acute respiratory syndrome virus main proteinase and its complex with an inhibitor, Proc. Natl. Acad. Sci. USA 100 (2003) 13190–13195.

[5] K. Anand, G.J. Palm, J.R. Mesters, S.G. Siddell, J. Ziebuhr, R. Hilgenfeld, Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain, EMBO J. 21 (2002) 3213–3224.

[6] J.H. Shi, Z. Wei, J.X. Song, Dissection study on the SARS 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly-specific protease inhibitors, J. Biol. Chem. 279 (2004) 24765–24773.

[7] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, J. Gen. Virol. 83 (2002) 595–599.

[8] D.A. Matthews, W.W. Smith, R.A. Ferre, B. Condon, G. Budahazi, W. Sisson, J.E. Villafranca, C. Janson, H.E. McElroy, C.L. Gribskov, Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein, Cell 77 (1994) 761–771.

[9] J. Seipelt, A. Guarna, E. Bergmann, M. James, W. Sommergruber, I. Fita, T. Skern, The structures of picornaviral proteinases, Virus Res. 62 (1999) 159–168.

[10] K.Q. Fan, P. Wei, Q. Feng, S.D. Chen, C.K. Huang, L. Ma, B. Lai, J.F. Pei, Y. Liu, J.G. Chen, L.H. Lai, Biosynthesis, purification and substrate specificity of SARS coronavirus 3C-like proteinase, J. Biol. Chem. 279 (2004) 1637–1642.

[11] C.K. Huang, P. Wei, K.Q. Fan, Y. Liu, L.H. Lai, 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism, Biochemistry 43 (2004) 4568–4574.

[12] C.J. Kuo, Y.H. Chi, J.T.A. Hsu, P.H. Liang, Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate, Biochem. Biophys. Res. Commun. 318 (2004) 862–867.

[13] F. Gao, H.Y. Ou, L.L. Chen, W.X. Zheng, C.T. Zhang, Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes, FEBS Lett. 553 (2003) 451–456.

[14] J. Ziebuhr, G. Heusipp, S.G. Siddell, Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase, J. Virol. 71 (1997) 3992–3997.

[15] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[16] R.X. Wang, Y. Gao, L. Liu, L.H. Lai, All-orientation search and all-placement search in comparative molecular field analysis, J. Mol. Model 4 (1998) 276–283.

[17] M.E. Hardy, T.J. Crone, J.E. Brower, K. Ettayebi, Substrate specificity of the Norwalk virus 3C-like proteinase, Virus. Res. 89 (2002) 29–39.

[18] C. Wirblich, M. Sibilia, M.B. Boniotti, C. Rossi, H.J. Thiel, G. Meyers, 3C-like proteinase of rabbit hemorrhagic disease virus: identification of cleavage sites in the ORF1 polyprotein and analysis of cleavage specificity, J. Virol. 69 (1995) 7159–7168.

[19] M.G. Cordingley, P.L. Callahan, V.V. Sardana, V.M. Garsky, R.J. Colonno, Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro, J. Biol. Chem. 265 (1990) 9062–9065.

[20] Z.M. Liu, C.K. Huang, K.Q. Fan, P. Wei, H. Chen, S.Y. Liu, J.F. Pei, L. Shi, B. Li, K. Yang, Y. Liu, L.H. Lai, Virtual screening of novel non-covalent inhibitors for SARS-CoV 3C-like proteinase, J. Chem. Inf. Model. 45 (2005) 10–17.