

---

# Rapid peptide-based screening on the substrate specificity of severe acute respiratory syndrome (SARS) coronavirus 3C-like protease by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

---

LING-HON MATTHEW CHU,<sup>1,4</sup> WAI-YAN CHOY,<sup>1,4</sup> SAU-NA TSAI,<sup>2</sup> ZIHE RAO,<sup>3</sup>  
AND SAI-MING NGAI<sup>1,2</sup>

<sup>1</sup>Molecular Biotechnology Program and <sup>2</sup>Department of Biology, The Chinese University of Hong Kong, Shatin, Hong Kong

<sup>3</sup>Laboratory of Structural Biology, Department of Biological Science and Biotechnology, Tsinghua University, 100084 Beijing, China

(RECEIVED December 6, 2005; FINAL REVISION December 6, 2005; ACCEPTED January 13, 2006)

## Abstract

Severe acute respiratory syndrome coronavirus (SARS-CoV) 3C-like protease (3CL<sup>Pro</sup>) mediates extensive proteolytic processing of replicase polyproteins, and is considered a promising target for anti-SARS drug development. Here we present a rapid and high-throughput screening method to study the substrate specificity of SARS-CoV 3CL<sup>Pro</sup>. Six target amino acid positions flanking the SARS-CoV 3CL<sup>Pro</sup> cleavage site were investigated. Each batch of mixed peptide substrates with defined amino acid substitutions at the target amino acid position was synthesized via the “cartridge replacement” approach and was subjected to enzymatic cleavage by recombinant SARS-CoV 3CL<sup>Pro</sup>. Susceptibility of each peptide substrate to SARS-CoV 3CL<sup>Pro</sup> cleavage was monitored simultaneously by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The hydrophobic pocket in the P2 position at the protease cleavage site is crucial to SARS-CoV 3CL<sup>Pro</sup>-specific binding, which is limited to substitution by hydrophobic residue. The binding interface of SARS-CoV 3CL<sup>Pro</sup> that is facing the P1' position is suggested to be occupied by acidic amino acids, thus the P1' position is intolerant to acidic residue substitution, owing to electrostatic repulsion. Steric hindrance caused by some bulky or  $\beta$ -branching amino acids in P3 and P2' positions may also hinder the binding of SARS-CoV 3CL<sup>Pro</sup>. This study generates a comprehensive overview of SARS-CoV 3CL<sup>Pro</sup> substrate specificity, which serves as the design basis of synthetic peptide-based SARS-CoV 3CL<sup>Pro</sup> inhibitors. Our experimental approach is believed to be widely applicable for investigating the substrate specificity of other proteases in a rapid and high-throughput manner that is compatible for future automated analysis.

**Keywords:** substrate specificity; SARS-CoV; protease; MALDI-TOF; mass spectrometry; synthetic peptide

---

Reprint requests to: Sai-Ming Ngai, Department of Biology, The Chinese University of Hong Kong, Shatin, Hong Kong; e-mail: smngai@cuhk.edu.hk; fax: + (852) 2603-5745.

<sup>4</sup>These authors contributed equally to this work.

**Abbreviations:** SARS, severe acute respiratory syndrome; SARS-CoV, SARS Coronavirus; 3CL<sup>Pro</sup>, 3C-like protease; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.052007306>.

Severe acute respiratory syndrome (SARS) is a life-threatening, infectious respiratory illness characterized by high fever, chills, rigor, myalgia, dry cough, and headache, which may further progress to a dramatic increase in lung opacity, dyspnea, and hypoxemia (Drosten et al. 2003; Lee et al. 2003). The etiological agent of SARS was identified as a novel type of human coronavirus, SARS-CoV (Drosten et al. 2003; Ksiazek et al.

2003; Peiris et al. 2003). Coronaviruses are large, enveloped, single-stranded positive-sense RNA viruses that replicate in the cytoplasm of the infected host cell (Siddell et al. 1983). Phylogenetic studies reveal SARS-CoV does not belong to any of the three known groups of the *Coronaviridae* family; thus, it is suggested that SARS-CoV defines a distinct fourth group of coronaviruses (Marra et al. 2003; Rota et al. 2003). The RNA genome of SARS-CoV is ~30 kb in length, with five major open reading frames (ORFs) encoding the replicase polyprotein, the spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein (Marra et al. 2003; Rota et al. 2003).

Investigations of various SARS-CoV proteins can facilitate the development of vaccines and drugs against the SARS disease. The pivotal role of SARS-CoV surface S protein in mediating viral entry sheds light on the generation of S protein-specific neutralizing antibodies (Gallagher and Buchmeier 2001; Choy et al. 2004), and some of the experimental vaccines have been tested in the murine viral replication model recently (Bisht et al. 2004; Yang et al. 2004). Meanwhile, the 33.8-kDa SARS-CoV main protease, also known as the SARS-CoV 3C-like protease (SARS-CoV 3CL<sup>pro</sup>) is being considered extensively as an attractive and promising target for anti-SARS drug design, owing to its important biological role in the viral life cycle (Anand et al. 2003).

SARS-CoV 3CL<sup>pro</sup> mediates extensive proteolytic processing of two overlapping replicase polyproteins, pp1a (486 kDa) and pp1ab (790 kDa), to yield the corresponding functional polypeptides that are essential for SARS-CoV replication and transcription processes (Herold et al. 1993; Ziebuhr et al. 1995; Rota et al. 2003). There are at least 11 conserved SARS-CoV 3CL<sup>pro</sup> cleavage sites on the SARS-CoV polyprotein pp1ab, in which the substrate specificity involves preferentially the Leu-Gln↓ sequence in the P2 and P1 positions at the cleavage site, respectively (Ziebuhr et al. 2000; Heygi and Ziebuhr 2002; Anand et al. 2003). SARS-CoV 3CL<sup>pro</sup> comprises three domains, including the catalytic domains I and II that form the N-terminal chymotrypsin fold, and the unique extra helical domain III in C-terminal that is important for regulating the activity and specificity of SARS-CoV 3CL<sup>pro</sup> (Shi et al. 2004). The reported X-ray crystal structures of the SARS-CoV 3CL<sup>pro</sup> allows the investigation of possible interactions of SARS-CoV 3CL<sup>pro</sup> with its substrates, which contributes to the development of specific protease inhibitors as potential anti-SARS drugs (Anand et al. 2003; Yang et al. 2003).

A better understanding of the specific enzyme substrate-binding mechanism during the SARS-CoV 3CL<sup>pro</sup> proteolytic cleavage process is required for the rational design of an effective protease inhibitor with strong binding affinity to SARS-CoV 3CL<sup>pro</sup>. To screen the substrate

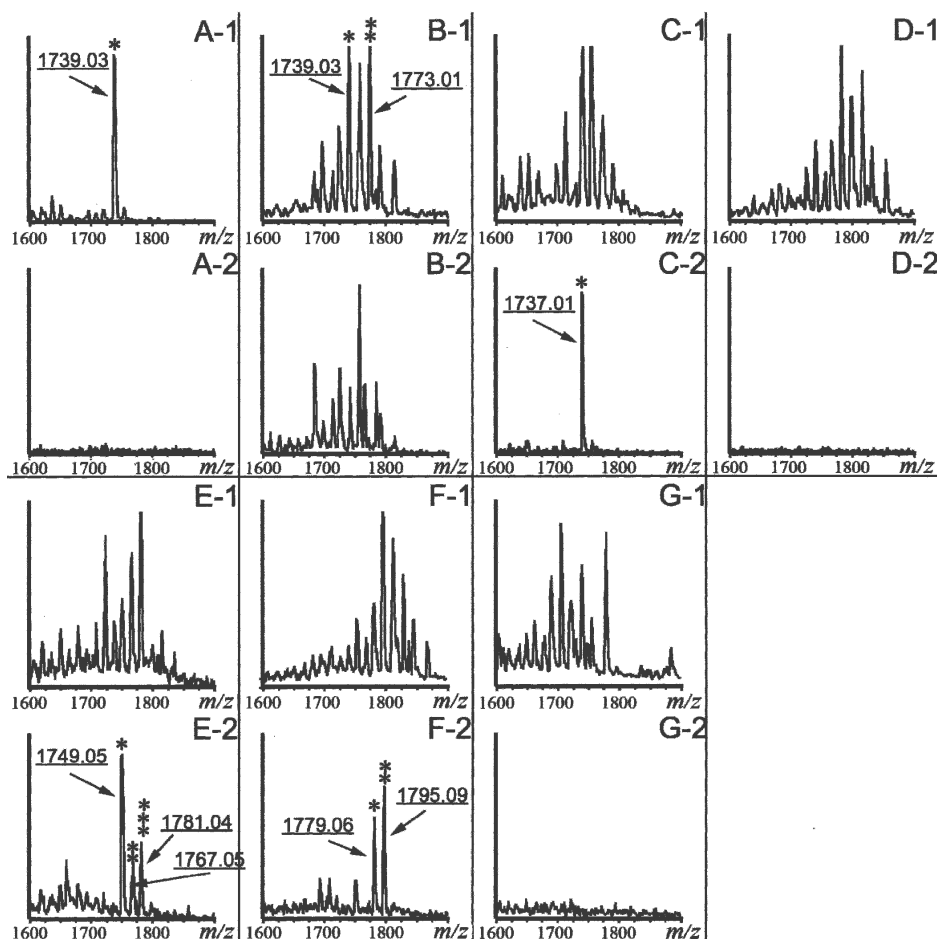
specificity of SARS-CoV 3CL<sup>pro</sup> in a rapid and high-throughput manner in contrast to the traditional tedious procedures, we applied the matrix-assisted laser desorption/ionization time-of-flight mass spectrometric (MALDI-TOF MS) analysis in combination with the novel “cartridge replacement” solid-phase peptide synthesis approach to investigate the biological significance of amino acid residues in the P2, P3, P4, P1', P2', and P3' positions that are flanking the conserved Gln residue in the P1 position at the SARS-CoV 3CL<sup>pro</sup> cleavage site (Schechter and Berger 1967; Fan et al. 2004, 2005). Owing to the sensitivity of MALDI-TOF MS, the protease cleavage progress of a mixture of different peptide substrates can be monitored simultaneously to generate comprehensive data on the substrate specificity of SARS-CoV 3CL<sup>pro</sup>. Therefore, all of the 20 possibilities of amino acid substitutions in a particular position of the peptide substrate can be easily investigated in the same reaction. Our study can provide insights into the molecular mechanism of the SARS-CoV 3CL<sup>pro</sup> cleavage process and reveal the feasibility of developing synthetic peptide-based protease inhibitors as potential drugs against SARS-CoV and other coronavirus infections. Furthermore, the rapid approach we described in this study could be widely applicable for monitoring the cleavage activities of other proteases, such as allowing the rapid screening of potential protease inhibitors of the human immunodeficiency virus (HIV) as the basis of drug development. In addition, automations are highly compatible and feasible with our experimental procedures to cope with the increasing demand of high-throughput analysis in proteomic research.

## Results

### *Substrate specificity preferences of SARS-CoV 3CL<sup>pro</sup>*

#### *Positive control of the SARS-CoV 3CL<sup>pro</sup> cleavage assay*

The PS01 peptide was amidated and acetylated at the C and N termini, respectively. The corresponding amino acid sequence of PS01 was based on one of the SARS-CoV 3CL<sup>pro</sup> cleavage sites on the SARS-CoV BJ01 polyprotein pp1ab (residues 3232–3247) and was used as the positive control in the enzymatic cleavage assay to assess the activity of the recombinant SARS-CoV 3CL<sup>pro</sup>. The original PS01 peptide peak with  $m/z$  value of 1739.03 before the enzymatic assay was absent from the mass spectrum after the cleavage reaction (Fig. 1) as it was being cleaved into two peptide fragments with their corresponding  $m/z$  values (data not shown), which validated the enzymatic activity of the recombinant SARS-CoV 3CL<sup>pro</sup> that we prepared.



**Figure 1.** Primary screening: Mass spectrometry analysis of SARS-CoV 3CL<sup>pro</sup> cleavage products. (A) Positive control. PS01 peptide peak (\*) was resolved before cleavage assay (A-1) and was absent from the mass spectrum after the reaction (A-2). (B) P2 position. For PS02 peptide substrates, all 20 peptide peaks were resolved on the mass spectrum before cleavage assay (B-1), and only the peptide peaks corresponding to Leu (\*) and Phe (\*\*) substitutions in the P2 position were absent, whereas all other peaks remained after the cleavage reaction (B-2). (C) P3 position. For PS03 peptide substrates, all 20 peptide peaks were resolved on the mass spectrum before cleavage assay (C-1), and only the peptide peak with Pro (\*) substitution in the P3 position remained after the reaction (C-2). (D) P4 position. For PS04 peptide substrates, all 20 peptide peaks were resolved on the mass spectrum before cleavage assay (D-1), which were all absent from the mass spectrum after the reaction (D-2). (E) P1' position. For PS05 peptide substrates, all 20 peptide peaks were resolved on the mass spectrum before cleavage assay (E-1), and the peptide peaks with Pro (\*), Asp (\*\*), and Glu (\*\*\*) in the P1' position remained after the reaction (E-2). (F) P2' position. For PS06 peptide substrates, all 20 peptide peaks were resolved on the mass spectrum before cleavage assay (F-1), and the peptide peaks with Pro (\*) and Ile/Leu (\*\*) in the P2' position remained after the reaction (F-2). (G) P3' position. For PS07 peptide substrates, all 20 peptide peaks were resolved on the mass spectrum before cleavage assay (G-1), which were all absent from the mass spectrum after the reaction (G-2).

### “Cartridge replacement” strategy

In this study, we used MALDI-TOF MS analysis in combination with the solid-phase peptide synthesis approach to examine the biological significance of amino acid residues in a total of six target positions at the SARS-CoV 3CL<sup>pro</sup> cleavage sites, including the P2, P3, and P4 positions at the amino side of the P1 position; and the P1', P2', and P3' positions at the carboxyl side of the P1 position (Table 1).

Based on the flexibility in changing the contents of each amino acid cartridge, we designed a novel peptide synthesis

strategy that we named as the “cartridge replacement” solid phase peptide synthesis. Our “cartridge replacement” strategy consists of two screening procedures: (1) primary screening and (2) secondary screening.

#### 1. Primary screening using the “cartridge replacement” strategy

By using the “cartridge replacement” strategy, the corresponding amino acid cartridge at the target position under investigation was deliberately replaced by another

**Table 1.** Amino acid sequence of the seven synthetic peptide substrates

Peptide substrate	Sequence <sup>a</sup>															
	P9	P8	P7	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'	P7'
PS01 <sup>b</sup>	T	S	I	T	S	A	V	L	Q	S	G	F	R	K	M	A
PS02	T	S	I	T	S	A	V	X <sub>20</sub> <sup>c</sup>	Q	S	G	F	R	K	M	A
PS03	T	S	I	T	S	A	X <sub>20</sub> <sup>c</sup>	L	Q	S	G	F	R	K	M	A
PS04	T	S	I	T	S	X <sub>20</sub> <sup>c</sup>	V	L	Q	S	G	F	R	K	M	A
PS05	T	S	I	T	S	A	V	L	Q	X <sub>20</sub> <sup>c</sup>	G	F	R	K	M	A
PS06	T	S	I	T	S	A	V	L	Q	S	X <sub>20</sub> <sup>c</sup>	F	R	K	M	A
PS07	T	S	I	T	S	A	V	L	Q	S	G	X <sub>20</sub> <sup>c</sup>	R	K	M	A

<sup>a</sup> The amino acids flanking the SARS-CoV 3CL<sup>pro</sup> cleavage sites are labeled from the amino terminus to the carboxyl terminus as follows: -P3-P2-P1 ↓ P1'-P2'-P3', as described previously (Schechter and Berger 1967).

<sup>b</sup> The control peptide substrate is based on the amino acid sequence of one of the SARS-CoV 3CL<sup>pro</sup> cleavage sites on the SARS-CoV BJ01 polyprotein pp1ab (residues 3232–3247).

<sup>c</sup> The corresponding amino acid at this position is substituted by a mixture of 20 standard amino acids in equal molar ratios.

cartridge containing a mixture of 20 standard amino acids in equal molar ratios during the solid-phase peptide synthesizing process so as to generate a mixed pool of 20 different peptide substrates with single amino acid residue alteration at the target position. Each batch of the mixed peptide substrates was subjected to in vitro enzymatic cleavage assay by recombinant SARS-CoV 3CL<sup>pro</sup> in a single reaction and then analyzed by MALDI-TOF MS to detect the susceptibility of each of the 20 peptides to cleavage by SARS-CoV 3CL<sup>pro</sup>.

The results from the primary screening of the in vitro enzymatic cleavage assay of the six batches of synthetic peptide substrates (namely PS02, PS03, PS04, PS05, PS06, and PS07) with amino acid substitutions in their corresponding target positions are shown in Figure 1 and summarized in Table 2. Before the enzymatic assay, the corresponding peaks of the 20 peptides in

each substrate pool were resolved on the mass spectrum with their exact *m/z* values, which demonstrated the validity of solid-phase peptide synthesis using the “cartridge replacement” approach to generate a pool of 20 peptides with single amino acid difference at the target position. The results also demonstrated the capability of MALDI-TOF MS to discriminate most of the peptide species with single amino acid difference.

*P2, P3, and P4 positions at the amino side of the SARS-CoV 3CL<sup>pro</sup> cleavage site*

For PS02 peptide substrates with amino acid substitutions in the P2 position, only the original peptide (*m/z* value of 1739.03) and the peptide with Phe substitution for Leu in the P2 position (*m/z* value of 1773.01) were cleaved, with their corresponding peptide peaks absent from the mass spectrum after the enzymatic assay,

**Table 2.** Summary of results from primary screening

Peptide substrate	Target position <sup>a</sup>	Corresponding <sup>b</sup> amino acid	Amino acid substitution <sup>c</sup>	
			Cleaved peptides	Uncleaved peptides
PS02	P2	Leu	Leu, Phe	Others <sup>d</sup>
PS03	P3	Val	Others <sup>e</sup>	Pro
PS04	P4	Ala	All <sup>f</sup>	—
PS05	P1'	Ser	Others <sup>e</sup>	Pro, Asp, Glu
PS06	P2'	Gly	Others <sup>e</sup>	Pro, Ile/Leu
PS07	P3'	Phe	All <sup>f</sup>	—

<sup>a</sup> The positions flanking the SARS-CoV 3CL<sup>pro</sup> cleavage sites are labeled from the N terminus to the C terminus as follows: -P3-P2-P1 ↓ P1'-P2'-P3', as described previously (Schechter and Berger 1967).

<sup>b</sup> The original amino acid residue in the position under investigation before the substitution with 20 standard amino acids.

<sup>c</sup> The identity of the amino acid in the particular peptide that substituted the residue at the position under investigation.

<sup>d</sup> All peptides other than those in the opposite column were not cleaved by SARS-CoV 3CL<sup>pro</sup>.

<sup>e</sup> All peptides other than those in the opposite column were cleaved by SARS-CoV 3CL<sup>pro</sup>.

<sup>f</sup> All 20 peptides were cleaved by SARS-CoV 3CL<sup>pro</sup>.

whereas all other peptides were not cleaved, with their corresponding peaks remaining. For PS03 peptide substrates with amino acid substitutions in the P3 position, only the peptide with Pro substitution for Val in the P3 position ( $m/z$  value of 1737.01) was not cleaved, with its corresponding peptide peak remaining in the mass spectrum after the enzymatic assay, whereas all other peptides were cleaved successfully. For PS04 peptide substrates with amino acid substitutions in the P4 position, all of the 20 peptides were cleaved completely, with their corresponding peaks absent from the mass spectrum after the enzymatic assay.

*P1', P2', and P3' positions at the carboxyl side of the SARS-CoV 3CL<sup>pro</sup> cleavage site*

For PS05 peptide substrates, only three peptides with Pro substitution ( $m/z$  value of 1749.05), Asp substitution ( $m/z$  value of 1767.02), and Glu substitution ( $m/z$  value of 1781.04) for Ser in the P1' position, respectively, were not cleaved. For PS06 peptide substrates, only two peptides with Pro substitution ( $m/z$  value of 1779.06) and Ile/Leu substitution ( $m/z$  value of 1795.09) for Gly in the P2' position, respectively, were not cleaved. For PS07 peptide substrates with amino acid substitutions in the P3' position, all of the 20 peptides were cleaved completely, with their corresponding peaks absent from the mass spectrum after the enzymatic assay.

*2. Secondary screening using the "cartridge replacement" strategy*

After the primary screening procedure, we generated the comprehensive cleavage profiles of the corresponding 20 peptide substrates of each target amino acid position simultaneously, based on the "cartridge replacement" strategy. However, there were peptides (carrying ambiguous residues like Ile/Leu and Gln/Lys that carry the identical or close masses) in which true cleavage might not be identified with absolute confidence. A secondary screening procedure was necessary, and we proceeded to the synthesis of target peptide substrates into defined batches, particularly for those peptides with identical or close molecular masses, which resulted in ambiguous peaks for detection on the mass spectra in the primary screening procedure.

The results of the *in vitro* enzymatic cleavage assay of five peptide batches (namely, PS02-1, PS02-2, PS02-3, PS06-I, and PS06-L) are shown in Figure 2 and summarized in Table 3. For PS02-1 peptide batches, only the original peptide with Leu in the P2 position ( $m/z$  value of 1739.03) was cleaved, with its corresponding peptide peaks absent from the mass spectrum after the enzymatic assay, whereas the other three peptides with Pro, Cys, and

Glu substitutions for Leu in the P2 position were not cleaved, with their corresponding peaks remaining. For PS02-2 and PS02-3 peptide batches, not all of the peptides were cleaved, with their corresponding peaks remaining. For PS06-I and PS06-L peptide batches, both the peptides with Ile substitution ( $m/z$  value of 1795.09) and Leu substitution ( $m/z$  value of 1795.09) for Gly in the P2' position, respectively, were not cleaved, with their corresponding peaks remaining.

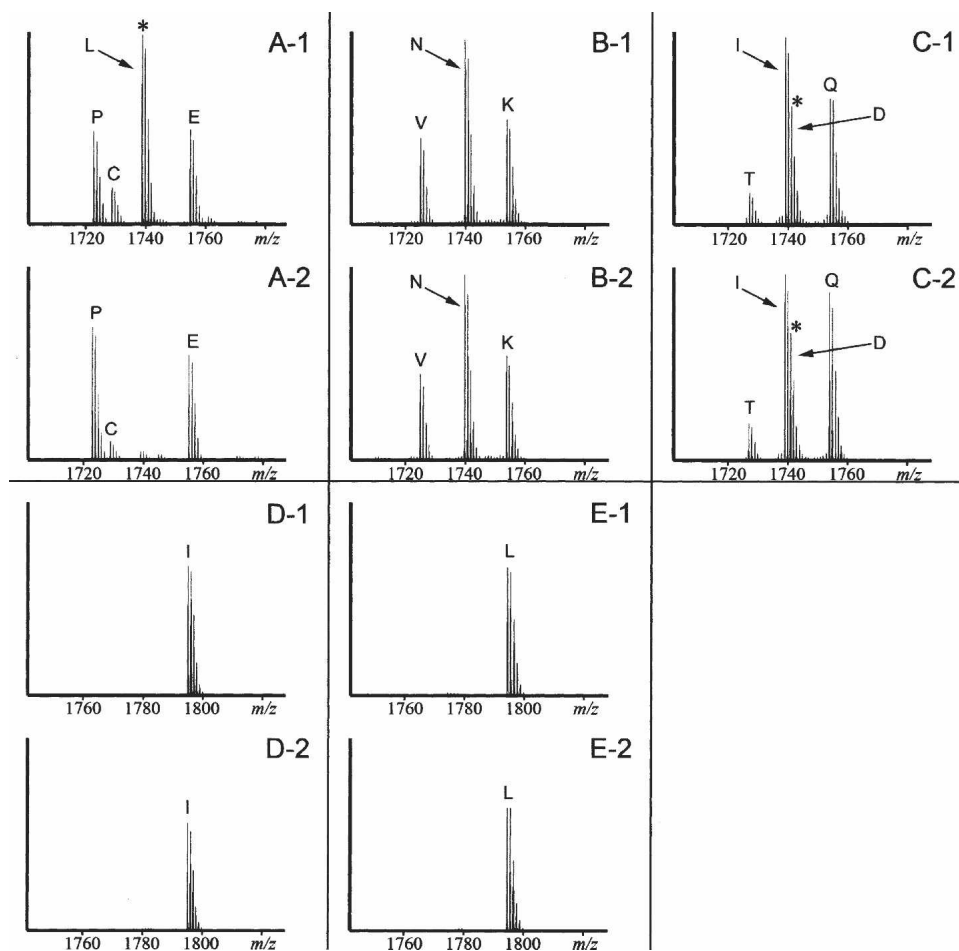
*Molecular docking of SARS-CoV 3CL<sup>pro</sup> and peptide substrates*

Based on the mass spectrometry results from the *in vitro* enzymatic cleavage assay (Table 2), four target positions including P2, P3, P1', and P2' positions located at the SARS-CoV 3CL<sup>pro</sup> cleavage sites were further analyzed by the Insight II molecular modeling platform, and the molecular docking results are shown in Figures 3 and 4. The control peptide, PS01, was docked to the predefined active site of SARS-CoV 3CL<sup>pro</sup>, with the conserved Gln residue located at close proximity to Cys145-His41 residues at the catalytic dyad of SARS-CoV 3CL<sup>pro</sup> (Fig. 4), in which this total energy data (182.74 kcal/mol) generated from energy minimization was used as the reference for comparing the docking results using other peptide substrates. This slightly positive total energy value is due to the relatively high repulsive van der Waals energy attributed by the narrow binding pocket on SARS-CoV 3CL<sup>pro</sup>.

For the PS02 peptide with Phe substitution in the P2 position, the total energy of 186.98 kcal/mol was similar to that of the PS01 control peptide (182.74 kcal/mol), which further confirmed the tolerance of Phe substitution at the P2 position at the cleavage site (Fig. 4). On the other hand, the total docking energy for PS03 peptide with Pro substitution in the P3 position (1028.33 kcal/mol); PS05 peptides with Pro substitution (1385.02 kcal/mol), Asp substitution (1384.69 kcal/mol) and Glu substitution (1384.91 kcal/mol) in the P1' position; and PS06 peptides with Pro substitution (1383.25 kcal/mol), Ile substitution (1381.35 kcal/mol), and Leu substitution (1387.60 kcal/mol) in the P2' position were at least five-fold higher than the total energy of the PS01 control peptide (182.74 kcal/mol). The relatively high total docking energy revealed the unfavorable cleavage reactions of these peptides by SARS-CoV 3CL<sup>pro</sup> (Fig. 4).

**Discussion**

We have introduced a rapid and high-throughput screening method to generate a comprehensive overview of the substrate specificity preferences of SARS-CoV 3CL<sup>pro</sup> via a combinatorial approach that uses MALDI-



**Figure 2.** Secondary screening: Mass spectrometry analysis of SARS-CoV 3CL<sup>PRO</sup> cleavage products. The monoisotopic peaks of each peptide can be clearly shown. (A) PS02-1 peptide batch. The four peaks of peptides with Pro (P;  $m/z = 1722.90$ ), Cys (C;  $m/z = 1728.86$ ), Leu (L;  $m/z = 1739.03$ ), and Glu (E;  $m/z = 1754.89$ ) substitutions were resolved on the mass spectrum before cleavage assay (A-1); only the peptide peaks corresponding to Leu (\*) substitutions were absent, whereas all other peaks remained after the reaction (A-2). (B) PS02-2 peptide batch. The three peaks of peptides with Val (V;  $m/z = 1724.92$ ), Asn (N;  $m/z = 1739.89$ ), and Lys (K;  $m/z = 1753.95$ ) substitutions were resolved on the mass spectrum before cleavage assay (B-1), which were all remaining after the reaction (B-2). (C) PS02-3 peptide batch. The four peaks of peptides with Thr (T;  $m/z = 1726.90$ ), Ile (I;  $m/z = 1739.03$ ), Asp (D;  $m/z = 1740.88$ ), and Gln (Q;  $m/z = 1753.91$ ) substitutions were resolved on the mass spectrum before cleavage assay (C-1), which all were remaining after the reaction (C-2). The 1740.88 monoisotopic peak with high intensity (\*) belongs to the mother peak of the peptide with Asp (D) substitution. (D) PS06-I peptide batch. The peak of peptide with Ile (I;  $m/z = 1795.09$ ) substitution was resolved on the mass spectrum before cleavage assay (D-1), which remained after the reaction (D-2). (E) PS06-L peptide batch. The peak of peptide with Leu (L;  $m/z = 1795.09$ ) substitution was resolved on the mass spectrum before cleavage assay (E-1); which remained after the reaction (E-2).

TOF MS and the novel “cartridge replacement” solid-phase peptide synthesis strategy. Research on the substrate specificity preferences of SARS-CoV 3CL<sup>PRO</sup> is gaining increasing attention, owing to the biological significance of SARS-CoV 3CL<sup>PRO</sup> in the viral life cycle and the feasibility to design specific SARS-CoV 3CL<sup>PRO</sup> inhibitors as potential anti-SARS drugs. It has been explicitly demonstrated that SARS-CoV 3CL<sup>PRO</sup> cleaves the coronavirus replicase polyproteins at no less than 11 conserved sites, preferentially involving the Leu-Gln↓ sequence in the P2 and P1 positions at the cleavage site,

respectively (Ziebuhr et al. 2000; Heygi and Ziebuhr 2002; Anand et al. 2003). To further extend our knowledge of the substrate specificity of SARS-CoV 3CL<sup>PRO</sup>, we aimed to comprehensively study the biological importance of amino acid residues in six selected positions (namely P2, P3, P4, P1', P2', and P3') flanking the conserved P1 position at the SARS-CoV 3CL<sup>PRO</sup> cleavage site by a rapid and high-throughput approach (Table 1).

From the primary screening procedure, the results of the PS03, PS04, PS05, and PS07 peptide substrates can be interpreted precisely. From the results of PS04 and

**Table 3.** Summary of results from secondary screening

Peptide batch	Amino acid substitution	Cleaved peptides	Uncleaved peptides
PS02-1	Pro, <sup>a</sup> Cys, <sup>b</sup> Leu, <sup>c</sup> Glu <sup>d</sup>	Leu	Pro, Cys, Glu
PS02-2	Val, <sup>a</sup> Asn, <sup>c,e</sup> Lys <sup>d</sup>	—	Val, Asn, Lys
PS02-3	Thr, <sup>b</sup> Ile, <sup>c</sup> Asp, <sup>c</sup> Gln <sup>d</sup>	—	Thr, Ile, Asp, Gln
PS06-I	Ile <sup>f</sup>	—	Ile
PS06-L	Leu <sup>f</sup>	—	Leu

<sup>a</sup>Pro ( $m/z$  value of 1722.90) and Val ( $m/z$  value of 1724.92) substitutions with close masses were separated into PS02-1 and PS02-2 peptide batches, respectively.

<sup>b</sup>Cys ( $m/z$  value of 1728.86) and Thr ( $m/z$  value of 1726.90) substitutions with close masses were separated into PS02-1 and PS02-3 peptide batches, respectively.

<sup>c</sup>Leu ( $m/z$  value of 1739.03), Asn ( $m/z$  value of 1739.89), and Ile ( $m/z$  value of 1739.03) substitutions with close masses were separated into PS02-1, PS02-2, and PS02-3 peptide batches, respectively.

<sup>d</sup>Glu ( $m/z$  value of 1754.89), Lys ( $m/z$  value of 1753.95), and Gln ( $m/z$  value of 1753.91) substitutions with close masses were separated into PS02-1, PS02-2, and PS02-3 peptide batches, respectively.

<sup>e</sup>Asp ( $m/z$  value of 1740.88) and Asn ( $m/z$  value of 1739.89) substitutions with close masses were separated into PS02-2 and PS02-3 peptide batches, respectively.

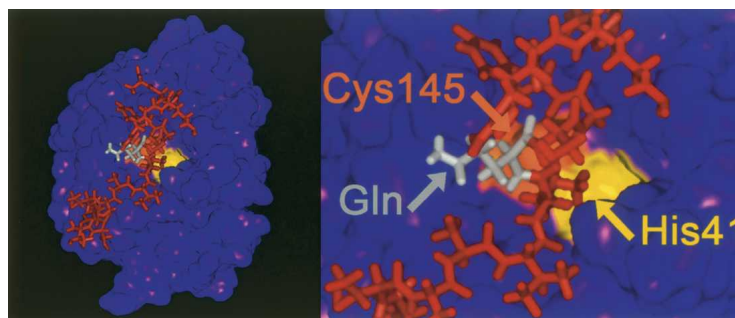
<sup>f</sup>Ile and Leu substitutions with identical masses ( $m/z$  value of 1795.09) were separated into PS06-I and PS06-L peptide batches, respectively.

PS07 peptide substrates (Fig. 1), all 20 peptide peaks were absent from the mass spectra after the SARS-CoV 3CL<sup>Pro</sup> cleavage reaction: The SARS-CoV 3CL<sup>Pro</sup> can cleave all of the 20 peptides regardless of the types of amino acid substitutions at the P3' or P4 positions. For PS03 and PS05 peptide substrates, the corresponding peptide peaks ( $m/z$  value of 1737.01 for PS03 and  $m/z$  values of 1749.05, 1781.04, and 1767.05 for PS05) retained on the mass spectra could be clearly resolved after the cleavage reaction (Fig. 1); the SARS-CoV 3CL<sup>Pro</sup> cannot cleave those peptides for which the non-overlapping  $m/z$  values were shown, and hence, those uncleaved peptides can be easily identified.

For PS02 peptide substrates, there were only two peptide peaks that vanished after the cleavage reaction.

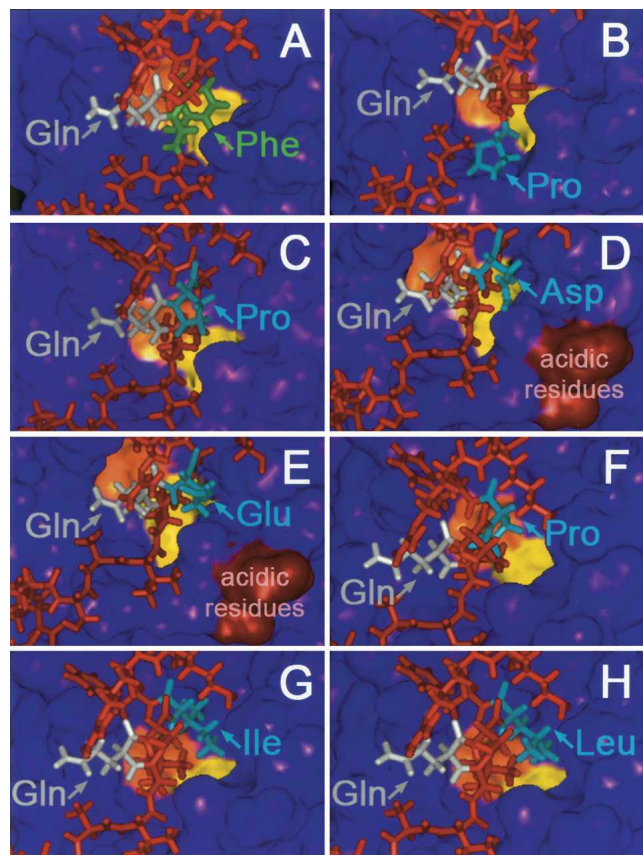
Although these two peaks can be identified as corresponding to their exact  $m/z$  values, several remaining peptide peaks with close masses introduced ambiguity in the mass spectra. In addition, for PS06 peptide substrates, there were only two peptide peaks found on the mass spectra after the cleavage reaction. One of the peptide peaks ( $m/z$  value of 1795.09) may contain two other peptide entities with exact  $m/z$  values, i.e., the Ile/Leu containing peptide. In fact, this result was also observed in one of the cleaved peaks of PS02 peptide substrates ( $m/z$  value of 1739.03). Therefore, a secondary screening procedure using the defined batch synthesis approach (Table 3) was used to solve the above problem.

In addition to the well-known conserved Gln residue in the P1 position at the SARS-CoV 3CL<sup>Pro</sup> cleavage site, the P2 position, which is exclusively occupied by Leu residue, also serves as another important determinant of substrate specificity. Our peptide cleavage results indicated that only the peptide substrate with Phe replacement in the P2 position was also favorable for SARS-CoV 3CL<sup>Pro</sup> cleavage (Fig. 1; Table 2), with similar total energy level (186.98 kcal/mol) in molecular docking when compared with the total energy of the positive control peptide (182.74 kcal/mol) with Leu in the P2 position. Also, the results of the three peptide batches (PS02-1, PS02-2, and PS02-3) in the secondary screening procedure further confirmed the results obtained from the primary screening procedure. The PS02-3 peptide batch showed that Ile was intolerant in the P2 position for SARS-CoV 3CL<sup>Pro</sup>. Our results demonstrated consistency with other findings (Fan et al. 2004, 2005), suggesting that the P2 position serves as an important hydrophobic pocket for SARS-CoV 3CL<sup>Pro</sup>-specific binding; thus, the peptide substrate with replacement of the original Leu with the hydrophobic Phe in the P2 position was still susceptible to SARS-CoV 3CL<sup>Pro</sup> cleavage (Fig. 4). From the results of a previous time course cleavage experiment (data not shown), we observed that the



**Figure 3.** Interaction between SARS-CoV 3CL<sup>Pro</sup> and PS01 control peptide as predicted by molecular docking. Overview (left) of the interaction between the SARS-CoV 3CL<sup>Pro</sup> (purple) and peptide substrate PS01 (red) and the zoom-in view (right) of the conserved Gln residue (gray) in the P1 position at the cleavage site of the PS01 control peptide being docked into the catalytic dyad of SARS-CoV 3CL<sup>Pro</sup>, which is composed of the Cys145 (orange) and His41 (yellow) residues.





**Figure 4.** Comparison between the molecular models of different selected peptide substrates docked to the active site of SARS-CoV 3CL<sup>pro</sup>. The Cys145 (orange) and His41 (yellow) residues in the catalytic dyad of SARS-CoV 3CL<sup>pro</sup> (purple) and the conserved Gln residue (gray) in the P1 position at the cleavage site of the peptide substrate (red) were shown. (A) PS02 peptide with Phe (green) in the P2 position was in close proximity to the active site of SARS-CoV 3CL<sup>pro</sup>. Peptide substrates that were unfavorable for SARS-CoV 3CL<sup>pro</sup> cleavage were deviated away from the active site of SARS-CoV 3CL<sup>pro</sup>, which include PS03 peptide with Pro (blue) in the P3 position (B); PS05 peptide with Pro (blue) in the P1' position (C); PS05 peptide with Asp (blue) in the P1' position that was in close proximity to the acidic residues (brown) at the active site (D); PS05 peptide with Glu (blue) in the P1' position that was in close proximity to the acidic residues (brown) at the active site (E); PS06 peptide with Pro (blue) in the P2' position (F); PS06 peptide with Ile (blue) in the P2' position (G); PS06 peptide with Leu (blue) in the P2' position (H).

SARS-CoV 3CL<sup>pro</sup> proteolytic cleavage rate of the positive control peptide was higher than that of the peptide with Phe substitution in the P2 position, which accounted for the dominant occurrence of Leu in the P2 position. Other studies also demonstrated that SARS-CoV 3CL<sup>pro</sup> favorably tolerates Phe and Val substitutions, and to a lesser extent, for Met and Ile substitutions in the P2 position (Fan et al. 2004, 2005), hence revealing the significance of the P2 hydrophobic pocket in the determination of SARS-CoV 3CL<sup>pro</sup> substrate specificity.

Furthermore, the P1' position, which is frequently occupied by Ser residue, also contributes to the substrate specificity of SARS-CoV 3CL<sup>pro</sup> considerably. Our peptide cleavage results revealed that the P1' position is highly unfavorable to the substitution by Pro, Asp, and Glu residues (Fig. 1; Table 2), which is consistent with molecular docking results showing their total energy levels being higher than that of the positive control peptide by more than fivefold. Our findings on the intolerance of Asp or Glu substitutions in the P1' position suggest that the acidic residues at the binding interface of the SARS-CoV 3CL<sup>pro</sup> active site are probably in close proximity and sterically complementary to the P1' position of the substrate, leading to electrostatic repulsions between the negatively charged acidic residues that hinder the substrate binding to SARS-CoV 3CL<sup>pro</sup>. Our molecular docking results also demonstrated that the P1' position is proximal to the Glu47 and Asp48 residues at the SARS-CoV 3CL<sup>pro</sup> active site (Fig. 4), which further confirmed our interpretations. Also, the unfavorable substitution by Pro in the P1' position could be accounted for by the steric hindrance arising from its cyclic side chain (Fig. 4), which obstructs the substrate binding to SARS-CoV 3CL<sup>pro</sup>. Previous studies also demonstrated that small aliphatic residues such as Ser, Ala, and Gly are favored in the P1' position (Fan et al. 2005).

Moreover, our results demonstrated that the substrate specificity of SARS-CoV 3CL<sup>pro</sup> are less dependent on the P2' and P3 positions at the cleavage site. The substitutions of Pro, Ile, or Leu in the P2' position and the substitution of Pro in the P3 position in the peptide substrates are unfavorable for SARS-CoV 3CL<sup>pro</sup> cleavage (Fig. 1; Table 2), in which these bulky Pro residue and  $\beta$ -branching Ile and Leu residues may cause steric hindrance to binding of SARS-CoV 3CL<sup>pro</sup> with its substrates (Fig. 4). The results of PS06-I and PS06-L peptide batches further confirmed the results of the primary screening: Neither Ile nor Leu is tolerant in the P2' position for SARS-CoV 3CL<sup>pro</sup>. On the other hand, the peptide cleavage results showed that the P3' and P4 positions have no effect on determining the substrate specificity preferences of SARS-CoV 3CL<sup>pro</sup> (Fig. 1; Table 2).

In contrast to providing quantitative measurements on the kinetic data of the interaction between SARS-CoV 3CL<sup>pro</sup> and the substrates, the scope of our studies mainly focuses on the description of a useful tool for rapid and comprehensive screening of substrate specificity for SARS-CoV 3CL<sup>pro</sup>, which is also applicable to the studies of other proteases. By combining the MALDI-TOF MS and synthetic peptide-based approaches, peptide-cleavage studies on the defined equal molar mixture of a batch of 20 peptide substrates before and after the protease cleavage reaction could be performed simultaneously on a single MALDI-TOF MS analysis. The MALDI-TOF mass spectrometer is



a delicate and sensitive instrument that enables clear and precise discrimination of different peptides, even with single amino acid difference. Therefore, each of the 20 peptide substrates with equal molar ratios in the same mixture can be monitored simultaneously on a single MALDI-TOF MS analysis before and after the protease cleavage reaction. Our approach is advantageous over the traditional methods that are restricted to testing single peptides with defined condition one at a time and the setup of a total of 20 identical experimental conditions for every analysis. After generating the comprehensive cleavage profile on the 20 peptides simultaneously based on the primary screening procedure of the “cartridge replacement” strategy, we can easily identify some of the biologically important residues taking part in the recognition process. Nevertheless, we employed the secondary screening procedure using the defined batch-synthesis approach to synthesize target peptides in different batches for exact identification purpose (ambiguous residues like Ile/Leu and Gln/Lys that carry identical or close masses). Our approach provides an option in which the resolution of the MALDI-TOF mass spectrometer instrument is not sensitive enough to distinguish between species with 1-Da mass difference. Also, MALDI-TOF MS analysis is tolerant to impurities; therefore, the reaction mixture can be analyzed directly after a simple desalting procedure, which reduces the workload of subsequent downstream purifications of cleavage products prior to analysis.

For the synthesis of peptide substrates, we deliberately applied the novel “cartridge replacement” strategy, which was based on the flexibility on modifying the solid-phase peptide synthesis process, in which each amino acid cartridge can be filled with different contents of amino acids in defined ratios, to generate the desired combinations of mixed peptide substrates according to experimental needs. By the incorporation of MALDI-TOF MS analysis and “cartridge replacement” peptide synthesis approach, we presented a useful tool that is widely applicable to studies that involve the monitoring of protease cleavage activity, which includes the rapid screening of potential protease inhibitors of other coronaviruses and the human immunodeficiency virus (HIV). On the basis of our studies, future experiments will be carried out to address the effectiveness of various synthetic peptide-based inhibitors in blocking the protease activity, with the ultimate aim of anti-SARS drug development.

In this study, we have combined the sensitivity, the high resolution of MALDI-TOF MS, and the flexibility of the “cartridge replacement” solid-phase peptide synthesis approach to generate SARS-CoV 3CL<sup>pro</sup> proteolytic cleavage data rapidly in a comprehensive manner. Our studies definitely provide insights into the full automation of such experimental procedures to meet the increasing demand of high-throughput and automated proteomic studies.

## Materials and methods

### *Protein expression and purification*

The coding sequence of SARS-CoV 3CL<sup>pro</sup> used for cloning was based on the SARS-CoV BJ01 strain complete genome sequence downloaded from GenBank (accession no. AY 278488; <http://www.ncbi.nlm.nih.gov/entrez>). The construction of the pGEX-3CL<sup>pro</sup> plasmid has been described previously by Dr. Rao's team at Tsinghua University (Yang et al. 2003). The pGEX-3CL<sup>pro</sup> plasmid was transformed into *Escherichia coli* strain BL21(DE3)-competent cells and plated on Luria-Bertani (LB) agar containing ampicillin (100 µg/mL)/chloramphenicol (25 µg/mL) and grown overnight at 37°C. A single colony was inoculated into 50 mL of LB-broth containing ampicillin (100 µg/mL) and grown overnight at 37°C with shaking. This overnight culture was then transferred into 5 L of LB-broth containing ampicillin (100 µg/mL) for large-scale protein expression at 37°C with shaking. When the culture was grown to an OD<sub>600nm</sub> of 0.6, it was induced with 0.1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and grown for an additional 6 h at 25°C. Cells were harvested by 20-min centrifugation at 3200g at 4°C, and the bacterial cell pellet was resuspended in lysis buffer containing 10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM dithiothreitol (DTT), 0.5 mM EDTA, and 0.2 mM phenylmethylsulfonyl fluoride (PMSF), and homogenized by sonication. The lysate was centrifuged at 48,000g for 45 min at 4°C, and the supernatant was loaded onto a Glutathione Sepharose 4B affinity column (Amersham Biosciences) equilibrated with the lysis buffer. The column was then washed with 10× column volume of lysis buffer. To cleave the GST-tag from the GST-3CL<sup>pro</sup> fusion protein, PreScission Protease (Amersham Biosciences) was added into the column, and the mixture was incubated overnight at 4°C. The SARS-CoV 3CL<sup>pro</sup> recombinant protein was eluted with the lysis buffer, concentrated to 10 mg/mL by Microcon Centrifugal Filter Devices (Millipore), and then stored at -80°C.

### *“Cartridge replacement” solid-phase peptide synthesis*

The sequence of peptide substrate used in this study was based on one of the SARS-CoV 3CL<sup>pro</sup> cleavage sites on the SARS-CoV BJ01 polyprotein p1ab (residues 3232–3247) (Table 1). The biological significance of amino acid residues in the P2, P3, P4, P1', P2', and P3' positions that are flanking the conserved Gln residue in the P1 position at the cleavage site was investigated one at a time using the novel “cartridge replacement” solid-phase peptide synthesis strategy. In a standard procedure of solid-phase peptide synthesis, each amino acid cartridge that contains a particular type of amino acid is placed side-by-side in the peptide synthesizer according to the target peptide sequence, and the peptide synthesis begins with the cartridge containing the first amino acid at the carboxyl terminus of the sequence. In this study, we utilized flexibility in changing the contents of each amino acid cartridge and designed a novel peptide synthesis strategy that we named as the “cartridge replacement” solid-phase peptide synthesis.

### *Primary screening*

Prior to peptide synthesis, the corresponding amino acid cartridge at the target position under investigation was replaced by another cartridge containing a mixture of 20 standard amino acids in equal molar ratio, such that a mixed pool

of 20 synthetic peptides differing by a single amino acid residue was generated as the end product after complete synthesis of the substrate sequence. Six sets of such peptide substrate pools together with the original peptide substrate were synthesized with this approach (Table 1) using a standard solid-phase peptide synthesis protocol as described previously (Choy et al. 2004).

### Secondary screening

The PS02 and PS06 peptide substrates require further analysis by a secondary screening. For the PS02 peptide substrates, the peptides with close masses were separated into three batches: PS02-1, PS02-2, and PS02-3. For the peptide synthesis of the PS02-1 batch, the corresponding Leu cartridge at the P2 position was replaced by a cartridge containing a mixture of Cys, Glu, Leu, and Pro in an equal molar ratio. For the peptide synthesis of the PS02-2 batch, the corresponding Leu cartridge at the P2 position was replaced by a cartridge containing a mixture of Asn, Lys, and Val in an equal molar ratio. For the peptide synthesis of the PS02-3 batch, the corresponding Leu cartridge at the P2 position was replaced by a cartridge containing a mixture of Asp, Gln, Ile, and Thr in an equal molar ratio. For the PS06 peptide substrates, only the Ile and Leu substitutions were further investigated and separated into two batches: PS06-I and PS06-L. For the peptide synthesis of PS06-I and PS06-L batches, the corresponding Gly cartridge at the P2' position was replaced by a cartridge containing Ile or Leu, respectively.

All of the peptide substrates were synthesized by the solid-phase technique with an in-house Applied Biosystems 433A Peptide Synthesizer on preloaded resins using standard 9-fluorenylmethoxycarbonyl (Fmoc) synthesis chemistry with piperidine deprotection and 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU)/N-hydroxybenzotriazole (HOBt) activation. The amino (NH<sub>2</sub>) termini of peptides were acetylated on the resin with acetic anhydride capping solution containing 0.5 M acetic anhydride, 0.125 M N,N-diisopropylethylamine (DIEA), 0.015 M N-hydroxybenzotriazole (HOBt) in N-methylpyrrolidone (NMP). The acetylated peptides were cleaved from the resins, and side-chain protecting groups were removed by cleavage solution containing 25 g/L ethanedithiol (EDT) and 50 g/L thioanisole in trifluoroacetic acid (TFA). The cleaved peptides were then precipitated with cold diethyl ether, washed, and lyophilized. The identities of the synthetic peptides were confirmed by mass spectrometry on an Ettan MALDI-TOF Pro Mass Spectrometer (Amersham Biosciences).

### Peptide cleavage assay and MS analysis

The *in vitro* peptide cleavage assay was performed in a total reaction volume of 25  $\mu$ L (pH 7.6), containing 1.5  $\mu$ M recombinant SARS-CoV 3CL<sup>pro</sup>, 0.5 mM sum total concentration of the peptide substrate batch, 20 mM Na<sub>2</sub>HPO<sub>4</sub>, 200 mM NaCl, 1 mM EDTA, and 1 mM DTT, with overnight incubation at 25°C. Cleavage products were desalted by ZipTip<sub>U-C18</sub> (Millipore), and the corresponding peptide peaks before and after the cleavage reactions were resolved by mass spectrometry on an Ettan MALDI-TOF Pro Mass Spectrometer (Amersham Biosciences) or an ABI 4700 MALDI TOF/TOF Analyzer (Applied Biosystems). For acquisition of mass spectra, 0.5- $\mu$ L aliquots of samples were dispensed onto the MALDI-plate, followed by a 0.5- $\mu$ L matrix solution (5 g/L  $\alpha$ -cyano-4-

hydroxycinnamic acid in 50% acetonitrile/0.1% trifluoroacetic acid). All spectra were acquired in a reflectron-positive ion mode and accumulated from 1000 laser shots with acceleration of 20 kV and pulsed extraction. Angiotensin III (*m/z* value of 897.523) and adrenocorticotrophic hormone (*m/z* value of 2465.199) were used as standards for internal calibration of MALDI-TOF MS spectra.

### Molecular docking

All three-dimensional molecular modeling studies were performed on an Insight II molecular modeling platform (Accelrys Software Inc.) run on Silicon Graphics Octane2 workstations (Silicon Graphics Inc.). The consistent valence force field (CVFF) was selected in all simulations. The X-ray crystal structure of the SARS-CoV 3CL<sup>pro</sup> (1UK4) was downloaded from the Protein Data Bank (PDB; <http://www.rcsb.org/pdb>) and used as a starting model. The catalytic active site of chain A in the SARS-CoV 3CL<sup>pro</sup> structure without the bound CMK peptide was used in the study. Hydrogen atoms were added for the structure of SARS-CoV 3CL<sup>pro</sup> chain A using the BUILDER module. Then, the force field potentials and partial charges were assigned. The resulting structure was subsequently subjected to energy minimization using the DISCOVER module and was used as the initial structures for molecular docking. Since the proteolytic activity SARS-CoV 3CL<sup>pro</sup> mainly relies on the active nucleophile Cys145 and the acid-base catalyst His41, the region containing all residues (10.0 Å radius) around this Cys-His dyad was selected as the target site and was used to set up as a grid with the DOCKING module. The nonbond energies of this region were precalculated on the grid. Different peptide substrates were then docked into the target site by manual docking using the DOCKING module. The best docking position was based on the docking energy of the peptide/SARS-CoV 3CL<sup>pro</sup> complex. Thousands of orientations of peptides were searched by maneuvering in the Cys-His dyad region of SARS-CoV 3CL<sup>pro</sup> manually until reaching the energy minimum. The resultant complex was subjected to energy minimization and molecular dynamics simulation using the DISCOVER module. A dielectric constant of 1.0 was used throughout the energy minimization and molecular dynamics simulation. The system was equilibrated at 300 K for 100 psec, during which time the potential energy of the system reached a stable value. A time step of 1 fsec was used to integrate the equation of motion. The final conformation of the structure was obtained through 5000 iterations of steepest descent energy minimization. Ultimately, the total docking energy between the peptide substrate and SARS-CoV 3CL<sup>pro</sup> in the energy-minimized complex was calculated using the DOCKING module. The docking energy included the van der Waals and electrostatic energies, which were the components of the intermolecular energy.

### Acknowledgments

We thank Dr. Zihé Rao's research team at Tsinghua University, in particular Haitao Yang, Xiaodong Zhou, and Xiaoyu Xue, for their generosity in providing us with the pGEX-3CL<sup>pro</sup> construct. This project is supported by the SARS Special RGC Grant and special SARS funding from GreaterChina Technology Group Ltd.

## References

- Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J.R., and Hilgenfeld, R. 2003. Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science* **300**: 1763–1767.
- Bisht, H., Roberts, A., Vogel, L., Bukreyev, A., Collins, P.L., Murphy, B.R., Subbarao, K., and Moss, B. 2004. Severe acute respiratory syndrome coronavirus spike protein expressed by attenuated vaccinia virus protectively immunizes mice. *Proc. Natl. Acad. Sci.* **101**: 6641–6646.
- Choy, W.Y., Lin, S.G., Chan, P.K., Tam, J.S., Lo, Y.M., Chu, I.M., Tsai, S.N., Zhong, M.Q., Fung, K.P., Wayne, M.M., et al. 2004. Synthetic peptide studies on the severe acute respiratory syndrome (SARS) coronavirus spike glycoprotein: Perspective for SARS vaccine development. *Clin. Chem.* **50**: 1036–1042.
- Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., et al. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**: 1967–1976.
- Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J., et al. 2004. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**: 1637–1642.
- Fan, K., Ma, L., Han, X., Liang, H., Wei, P., Liu, Y., and Lai, L. 2005. The substrate specificity of SARS coronavirus 3C-like proteinase. *Biochem. Biophys. Res. Commun.* **329**: 934–940.
- Gallagher, T.M. and Buchmeier, M.J. 2001. Coronavirus spike proteins in viral entry and pathogenesis. *Virology* **279**: 371–374.
- Hegyi, A. and Ziebuhr, J. 2002. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**: 595–599.
- Herold, J., Raabe, T., Schelle-Prinz, B., and Siddell, S.G. 1993. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**: 680–691.
- Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**: 1953–1966.
- Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G.M., Ahuja, A., Yung, M.Y., Leung, C.B., To, K.F., et al. 2003. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **348**: 1986–1994.
- Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., et al. 2003. The Genome sequence of the SARS-associated coronavirus. *Science* **300**: 1399–1404.
- Peiris, J.S., Lai, S.T., Poon, L.L., Guan, Y., Yam, L.Y., Lim, W., Nicholls, J., Yee, W.K., Yan, W.W., Cheung, M.T., et al. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**: 1319–1325.
- Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., et al. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**: 1394–1399.
- Schechter, I. and Berger, A. 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**: 157–162.
- Shi, J., Wei, Z., and Song, J. 2004. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: Defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* **279**: 24765–24773.
- Siddell, S., Wege, H., and Ter Meulen, V. 1983. The biology of coronaviruses. *J. Gen. Virol.* **64**: 761–776.
- 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.* **100**: 13190–13195.
- Yang, Z.Y., Kong, W.P., Huang, Y., Roberts, A., Murphy, B.R., Subbarao, K., and Nabel, G.J. 2004. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. *Nature* **428**: 561–564.
- Ziebuhr, J., Herold, J., and Siddell, S.G. 1995. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J. Virol.* **69**: 4331–4338.
- Ziebuhr, J., Snijder, E.J., and Gorbalenya, A.E. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**: 853–879.