Synthetic Peptide Studies on Spike Glycoprotein and 3C-like Protease of the Severe Acute Respiratory Syndrome (SARS) Coronavirus:

Perspective for SARS Vaccine and Drug Development

CHOY Wai Yan

A Thesis Submitted in Partial Fulfilment

of the Requirements for the Degree of

Master of Philosophy

in

Molecular Biotechnology

June 2005

Thesis committee

Supervisor: Professor NGAI Sai Ming

(Department of Biology, The Chinese University of Hong Kong)

Internal examiner: Professor JIANG Liwen

(Department of Biology, The Chinese University of Hong Kong)

Internal examiner: Professor TSUI Kwok-wing Stephen

(Department of Biochemistry, The Chinese University of Hong Kong)

External examiner: Professor ZHU Guang

(Department of Biochemistry, The Hong Kong University of Science and Technology)

Statement

All experimental works reported in this thesis were performed by the author, unless stated the otherwise.

CHOY Wai Yan

Abstract

Abstract of thesis entitled:

Synthetic Peptide Studies on Spike Glycoprotein and 3C-like Protease of the Severe Acute Respiratory Syndrome (SARS) Coronavirus: Perspective for SARS Vaccine and Drug Development

Submitted by CHOY Wai Yan

for the degree of Master of Philosophy

at The Chinese University of Hong Kong in June 2005

The epidemic of severe acute respiratory syndrome (SARS) swept the world in early 2003. Encountering this newly emerged infectious disease, the world was once caught off guard and overwhelmed with panic-stricken atmosphere, until the novel severe acute respiratory syndrome coronavirus (SARS-CoV) was identified to be the etiological agent of SARS. SARS-CoV contains a 30 kb RNA genome, with five major open reading frames encoding replicase polyproteins, spike (S) glycoprotein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein. A better understanding of some important viral proteins of SARS-CoV can shed light onto the

development of vaccines and drugs against SARS. In this study, we are particularly interested in identifying the antigenic regions on S glycoprotein and characterizing the substrate specificity of 3C-like protease ($3CL^{pro}$) of SARS-CoV.

SARS-CoV infection is mediated by S glycoprotein through receptor binding and membrane fusion between the virion and the host cell. Here, we focused on using synthetic peptides for developing antibodies against SARS-CoV, which aimed to block viral invasion by eliciting an immune response specific to the native SARS-CoV S glycoprotein. Synthetic peptides designed according to the surface regions of SARS-CoV S glycoprotein were used to immunize both rabbits and monkeys. Antisera were analyzed by ELISA and tested for antibody specificity against SARS-CoV by immunofluorescent confocal microscopy. Four of our six synthetic peptides (S2, S3, S5, and S6) were capable of eliciting SARS-CoV-specific antibodies, which may represent some of the biologically active sequences of the immunogenic regions on the SARS-CoV S glycoprotein. The study provides novel insights for future development of synthetic peptide vaccines against SARS.

Proteolytic processing of replicase polyproteins mediated by SARS-CoV $3CL^{pro}$ is essential for SARS-CoV replication and transcription; which makes SARS-CoV $3CL^{pro}$ a promising target for anti-SARS drug development. Here, we presented a rapid and high-throughput screening method to study the substrate specificity of SARS-CoV $3CL^{pro}$. Six target amino acid positions (P2, P3, P4, P1', P2' and P3') flanking the SARS-CoV $3CL^{pro}$ cleavage site were investigated. Each batch of mixed peptide substrates with twenty different 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 $3CL^{pro}$. Susceptibility of each peptide substrate to SARS-CoV $3CL^{pro}$ cleavage was monitored simultaneously by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer. The study generates a comprehensive overview of SARS-CoV $3CL^{pro}$ substrate specificity, which serves as the design basis of synthetic peptide-based SARS-CoV $3CL^{pro}$ inhibitors.

The development of effective vaccines and drugs is the ultimate goal in this anti-SARS combat. Our synthetic peptide-based approach is definitely a rapid initial step for providing useful information for SARS research.

摘要

嚴重急性呼吸系統綜合症(簡稱 SARS)疫潮於 2003 年年初橫掃全球。正當全球面對著這新出現的傳染病而感到不知所措與人心惶惶之際，一種全新的冠狀病毒名爲嚴重急性呼吸系統綜合症冠狀病毒(簡稱 SARS-CoV)被確認爲 SARS 的致病源。SARS-CoV 擁有一組包含三萬個城基的核糖核酸基因組，其五個主要的可讀框分別編碼複製酶多聚蛋白、刺突(S)糖蛋白(簡稱 S 糖蛋白)、包膜(E)蛋白、基質膜(M)蛋白，以及核衣殼(N)蛋白。進一步解析 SARS-CoV 內的一些重要蛋白，實在有助於研製抗 SARS 的疫苗和藥物。是次研究中，我們尤其感興趣於辨認 SARS-CoV S 糖蛋白的抗原決定簇及剖析 SARS-CoV 3C 樣蛋白水解酶(簡稱 $3CL^{pro}$)的底物特異性。

SARS-CoV S 糖蛋白通過與宿主細胞受體結合，及促使病毒胞膜與宿主細胞膜融合，以介導病毒的感染。這裡我們著重於利用合成肽，引發針對天然 SARS-CoV S 糖蛋白的免疫應答，繼而產生抗體以阻擋病毒入侵。我們把對應 SARS-CoV S 糖蛋白的表面部分而設計的合成肽，注射到兔和猴體內。提取動物的抗血清後，我們分別進行酶聯免疫吸附測定(ELISA)及免疫螢光共軛焦顯微鏡分析(Immunofluorescent confocal microscopy)，以檢測這些抗體對 SARS-CoV 的特異性。我們一共測試了六條合成肽，並發現其中有四條(分別爲 S2、S3、S5，及 S6)能產生針對抗 SARS-CoV 的特異性抗體，這很有可能代表 SARS-CoV S 糖蛋白上具有免疫原性的其中一些生物活性序列。因此，這項研究成果實在有助於日後研發抗 SARS 的合成肽疫苗。

SARS-CoV $3CL^{pro}$ 對複製酶多聚蛋白的水解，在病毒的複製和轉錄過程中發揮著重要的功用，因而被視爲極具潛力的抗 SARS 藥靶。這裡我們介紹一個快速及高生產量的篩檢方法，以探討 SARS-CoV $3CL^{pro}$ 的底物特異性。我們重點研究位於 SARS-CoV $3CL^{pro}$ 的切割位點旁側的六個目標氨基酸位置 (分別爲 P2、P3、P4、P1'、P2'，及 P3')。我們採用「卡式盒替換 (Cartridge replacement)」的方法，把目標位置原本的殘基，置換爲二十種不同的氨基酸，得以合成出混合肽底物。接著，我們利用重組 SARS-CoV $3CL^{pro}$ 對每一批混合肽進行酶促切割，並以基質輔助雷射解吸附 / 電離 - 飛行時間質譜儀分析(MALDI-TOF mass spectrometry)即時監察混合物中每一種肽底物對於被 SARS-CoV $3CL^{pro}$ 切割的感受性。這項研究爲 SARS-CoV $3CL^{pro}$ 的底物特異性提供了一個全面性的概觀，並作爲研製針對 SARS-CoV $3CL^{pro}$ 的合成肽抑制劑的設計藍本。

研製出有效的疫苗和藥物是我們在這場抗 SARS 戰役中的終極目標。我們這次以合成肽爲本的策略，絕對是能快捷對 SARS 研究提供有用資訊的第一步。

Acknowledgements

The two-year graduate studies in Molecular Biotechnology program at The Chinese University of Hong Kong was a special gift from God, which adequately prepared me for taking up scientific research as my life-long career. Here, I would like to express my gratitude to a vast number of people who contributed to the success of this thesis.

I wholeheartedly thank Ice, my supervisor, for his guidance and encouragement throughout these years. It was great to learn new stuffs on peptide chemistry, bioinformatics and proteomics from him; and I appreciated the opportunity to work on MALDI-TOF mass spectrometer, peptide synthesizer, and HPLC in his EG08 laboratory. Furthermore, I was given the chance to publish my results and to attend conferences, which further broadened my horizon in biology. It was indeed my pleasure to conduct my thesis under his supervision.

Thanks are given to all my labmates in EG08 laboratory, whom together created a warm, caring, and cheerful environment for me to enjoy my research work. I am obliged to our technician, Helen,

for her invaluable assistance and technical help on my project. I also appreciated the support and suggestions from Helen, Karen, Mandy, Matthew, and MingQi.

I also acknowledge Kiwi, the secretary of Molecular Biotechnology program, for her helpful reminders on the seminars and thesis.

My earnest gratitude is extended to my parents Rosa and Igino, and my brother Vincent, for their everlasting love, understanding, endurance and care; especially in times when I was so busy and impatience, they comforted me and provided me with the warmest shelter.

Last but not the least, I am deeply grateful to my beloved husband, Jackie, for his immense love, concern, and trust. He is always my best companion, supporter and teacher, even in times when I was frustrated and tired. Moreover, I am gratified for his generous help in some of the graphics in this thesis.

I appreciate the opportunity provided by The Croucher Foundation Scholarships 2005/2006, for me to pursue Ph.D. degree in University of California, Berkeley, after my M.Phil. studies here in The Chinese University of Hong Kong.

This project is supported by the SARS Special RGC Grant and special SARS funding from GreaterChina Technology Group Ltd.

CHOY Wai Yan

2005 June

General abbreviations

$3CL^{pro}$	3C-like protease
ACE2	Angiotensin-converting enzyme 2
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
CoV	Coronavirus
Cys	Cysteine
Da	Dalton
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
fs	Femtosecond
g	Gram
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
h	Hour
His	Histidine
IgG	Immunoglobulin G
Ile	Isoleucine
kb	Kilobase
kDa	KiloDalton

kg	Kilogram
L	Liter
Leu	Leucine
Lys	Lysine
μL	Microliter
μM	Micromole
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
Met	Methionine
mg	Microgram
min	Minute
mL	Milliliter
M protein	Membrane protein
nm	Nanometer
N protein	Nucleocapsid protein
ORF(s)	Open reading frame(s)
Phe	Phenylalanine
Pro	Proline
ps	Picosecond
RNA	Ribonucleic acid
RT-PCR	Reverse transcription-polymerase chain reaction
SARS	Severe acute respiratory syndrome
Ser	Serine
SARS-CoV	Severe acute respiratory syndrome coronavirus
S glycoprotein	Spike glycoprotein
SPPS	Solid phase peptide synthesis
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
WHO	World Health Organization

Abbreviations of chemicals

ACN	Acetonitrile
AngIII	Angiotensin III
ACTH	Adrenocorticotropic hormone
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
DIEA	Diisopropylethylamine
DMF	Dimethyl formamide
DTT	Dithiothreitol
EDT	Ethanedithiol
Fmoc	N^{α}-9-fluorenylmethyloxycarbonyl
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBT	1-hydroxybenzotriazole
IPTG	Isopropyl-1-thio-s-D-galactopyranoside
KLH	Keyhole limpet hemocyanin
NMP	*N*-methylpyrrolidone
PBS	Phosphate buffered saline
PMSF	Phenylmethylsulfonyl fluoride
TFA	Trifluoroacetic acid

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Severe acute respiratory syndrome (SARS) - An overview

1.1.1 Epidemiology of SARS

Severe acute respiratory syndrome (SARS) was the first life threatening and highly contagious viral epidemic of the 21^{st} century. The worldwide public health was seriously threatened by the SARS outbreak, and the World Health Organization (WHO) issued a heightened global health alert on March 15, 2003 about this mysterious pneumonia of unknown etiology, after cases identified in China's Guangdong Province, Hong Kong, Vietnam, Singapore and Canada.

In retrospect, SARS was transmitted out from China's Guangdong Province on February 21, 2003 by an infected 65-year-old medical professor who spent a single night on the ninth floor of the

Metropole Hotel in Hong Kong, and at least 12 other secondary-infected guests at the hotel's ninth floor thus became the seeds of SARS transmission around the world via international air travel routes. During the period from November 1, 2002 to July 31, 2003, there was a cumulative total of 8098 probable SARS cases with 774 deaths reported from 29 countries (Table 1.1).

SARS was successfully contained within 4 months since the first recognition in mid-March 2003, in which WHO reported the last human chain of transmission of SARS had been broken on July 5, 2003 (Stadler *et al.*, 2003).

1.1.2 Clinical presentation of SARS

Clinical symptoms

SARS is an atypical form of acute pneumonia. The most common symptom in nearly all SARS patients was high fever with a body temperature of over 38^{o}C, often associated with other symptoms including chills, rigors, dry cough, headache, lethargy[1], malaise[2], myalgia[3], and gastrointestinal symptoms such as diarrhea (Booth *et al.*, 2003; Donnelly *et al.*, 2003; Leung *et al.*, 2003; Peiris *et al.*, 2003; Tsang *et al.*, 2003). These initial symptoms thus resembled common symptoms of other existing forms of atypical pneumonia.

[1]Abnormal drowsiness

[2]A vague feeling of bodily discomfort

[3]Muscular pain or tenderness

Table 1.1: The number of reported SARS cases from different countries during the period from November 1, 2002 to July 31, 2003.

Country	Cumulative no. of case(s)	No. of deaths
Australia	6	0
Canada	251	43
China	5327	349
France	7	1
Germany	9	0
Hong Kong	1755	299
India	3	0
Indonesia	2	0
Italy	4	0
Kuwait	1	0
Macao	1	0
Malaysia	5	2
Mongolia	9	0
New Zealand	1	0
Philippines	14	2
Republic of Ireland	1	0
Republic of Korea	3	0
Romania	1	0
Russian Federation	1	0
Singapore	238	33
South Africa	1	1
Spain	1	0
Sweden	5	0
Switzerland	1	0
Taiwan	346	37
Thailand	9	2
United Kingdom	4	0
United States	29	0
Vietnam	63	5
	Total: 8098	Total: 774

(Modified from http://www.sarsreference.com/sarsref/epidem.htm)

Laboratory findings

Characteristic changes in peripheral blood counts which included leucopenia[4] (in particular lymphopenia[5]) and thrombocytopenia[6] were common in SARS patients (Ksiazek *et al.*, 2003; Kuiken *et al.*, 2003; Poutanen *et al.*, 2003). Remarkable reduction in both the CD4[7] and CD8[8] T cell counts was noted during the early phase of illness (Cui *et al.*, 2003; Wong *et al.*, 2003*b*). Other common laboratory findings in SARS patients included elevated levels of lactate dehydrogenase (LDH), C-reactive protein[9], aspartate or alanine aminotransferases, and creatine phosphokinase (Booth *et al.*, 2003; Lee *et al.*, 2003; Peiris *et al.*, 2003; Poutanen *et al.*, 2003; Tsang *et al.*, 2003). A prolonged activated partial-thromboplastin[10] time and elevated D-dimer[11] level were also noted in a substantial number of patients (Lee *et al.*, 2003).

Chest radiography and high-resolution computed tomography findings

Chest radiography and high-resolution computed tomographic lung scan played an important role in the early diagnosis of SARS. At the

[4]An abnormal reduction of the white blood cell count
[5]An abnormal reduction in the number of lymphocytes in the blood
[6]An abnormal reduction in the number of platelets in the blood
[7]A glycoprotein predominantly found on the surface of helper T cells
[8]A glycoprotein predominantly found on the surface of cytotoxic T cells
[9]A globulin that appears in the blood in certain acute inflammatory conditions
[10]A protease that converts prothrombin to thrombin in the early stages of blood clotting
[11]A blood test for the detection of thrombosis

onset of fever, 70-80% of the SARS patients had abnormal chest radiographs on presentation (Booth *et al.*, 2003; Wong *et al.*, 2003*a*). Lung infiltrates were often shown in early chest radiography, and patchy ground-glass opacification of the lung parenchyma was detected readily by computed tomographic scans (Wong *et al.*, 2003*a*). Progression of lung infiltrates with the increase in air-space opacities and multiple patchy consolidations of both lungs were shown for SARS patients with clinical deterioration, which required intubation and mechanical ventilation (Tsang *et al.*, 2003; Tsui *et al.*, 2003; Wong *et al.*, 2003*a*).

1.1.3 Diagnostic tests of SARS

Laboratory confirmation of SARS was based on the virus isolation, viral RNA detection, or antibody detection of the etiological severe acute respiratory syndrome coronavirus (SARS-CoV).

The presence of SARS-CoV can be detected by inoculating SARS-permissive cells such as Vero cells with clinical specimens[12] to propagate SARS-CoV *in vitro*. Reverse transcription-polymerase chain reaction (RT-PCR) and nucleic acid sequence-based amplification (NASBA) could be used to detect SARS-CoV-specific RNA in clinical specimens. Conventional RT-PCR assays require post amplification product processing such as gel electrophoresis and southern blot

[12]Blood, serum, sputum, oropharyngeal materials, nasopharyngeal materials or stool

hybridization, that are time-consuming and require multiple handling steps of the PCR products (Drosten *et al.*, 2003; Peiris *et al.*, 2003; Yam *et al.*, 2003). On the other hand, the real-time monitoring of accumulating amplicon by real-time PCR or NASBA enables quantitative and highly sensitive measurement, which is more preferable for the detection of SARS-CoV-specific RNA (Mackay *et al.*, 2002; Emery *et al.*, 2004).

Antibodies that are produced in response to SARS infection can be detected by enzyme-linked immunosorbent assay (ELISA), immunofluorescent microscopy, and neutralization test. ELISA and immunofluorescent microscopy detect a mixture of immunoglobulin M (IgM) and immunoglobulin G (IgG) antibodies in the SARS patient sera that specifically bind to the recombinant SARS-CoV proteins and SARS-CoV-infected cells fixed on a microscope slide, respectively (Ksiazek *et al.*, 2003; Shi *et al.*, 2003; Wu *et al.*, 2004*b*). Neutralization test is regarded as the most specific and the best correlate of immunity among all the serological tests for SARS-CoV, since it detects and quantifies SARS-CoV-specific immunoglobulins in the SARS patient sera by measuring their ability to neutralize the infectivity of SARS-CoV on cell culture (Wu *et al.*, 2004*b*). Since all these diagnostic tests for SARS-CoV available so far have limitations, special caution is required for result interpretations, in order to avoid false sense of security based on false negative test results.

1.1.4 Treatment of SARS

To date, there is no specific drug formulation against SARS-CoV infection. During the SARS outbreak in 2003, the mainstream therapeutic treatment of SARS included broad-spectrum antibiotics, antiviral agents, and immunomodulatory therapy. Suspected cases of SARS were initially prescribed with broad-spectrum anti-bacterial agents against typical bacterial causes of acute community-acquired or nosocomial[13] pneumonia. These broad-spectrum antibiotics included fluoroquinolones, cephalosporins and macrolide (Groneberg *et al.*, 2005). The antibiotic therapy could be withdrawn once SARS-CoV was identified in the SARS suspected case. Antiviral therapies included the administration of ribavirin[14], oseltamivir phosphate[15], lopinavir-ritonavir co-formulation[16], and human interferons (Booth *et al.*, 2003; Cinatl *et al.*, 2003; Lee *et al.*, 2003; Peiris *et al.*, 2003; Poutanen *et al.*, 2003; Tsang *et al.*, 2003; Tsui *et al.*, 2003; Groneberg *et al.*, 2005). Also, corticosteroids were widely used for immunomodulatory therapy to reduce immune hyperactivity due to cytokine dysregulation and induction of inflammatory mediators with diffuse alveolar damage (Ho *et al.*, 2003; Nicholls *et al.*, 2003).

[13]Originating in the hospital

[14]A nucleoside analog with broad-spectrum antiviral activity

[15]A neuraminidase inhibitor for the treatment of both influenza A and B viruses

[16]A protease inhibitor preparation used to treat HIV

1.2 Severe acute respiratory syndrome coronavirus (SARS-CoV)

1.2.1 The etiological agent of SARS

During the initial outbreak of SARS in early March 2003, identification of the causative agent of the newly emerging infectious SARS was once an urgent and challenging task faced by the researchers worldwide. On March 17, 2003, WHO called upon 11 leading laboratories around the world to setup a multicenter research network for identifying the unknown etiology of SARS. In the beginning of the investigations, infectious pathogens such as Chlamydia (Drosten *et al.*, 2003), rhinoviruses (Ksiazek *et al.*, 2003) and paramyxoviruses (Poutanen *et al.*, 2003) that cause interstitial pneumonia[17] were identified in some SARS patients, and considered as potential causative agents of SARS. Nonetheless, the presence of these pathogens could not be confirmed in all SARS cases (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Peiris *et al.*, 2003).

By unprecedented global collaborative efforts, the causative agent of SARS was finally identified to be a novel human coronavirus known as severe acute respiratory syndrome coronavirus (SARS-CoV) in late March 2003. Different laboratories were able to isolate SARS-CoV from Vero E6 cells and fetal rhesus kidney cells

[17]Chronic lung disease that affects the interstitial tissue of lung

(FRhK-4) that were inoculated with clinical specimens[18] collected from SARS patients (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Peiris *et al.*, 2003). Electron-microscopic examination of culture supernatants revealed characteristic ultrastructural features of coronaviruses (Ksiazek *et al.*, 2003; Peiris *et al.*, 2003). Furthermore, monkeys infected with the isolated virus showed similar symptoms to human cases of SARS (Fouchier *et al.*, 2003). Hence, after confirming the casual relationship between the virus and SARS according to the Koch's postulates (Fouchier *et al.*, 2003), WHO announced that this novel, previously unknown coronavirus, SARS-CoV, as the etiological agent of SARS.

1.2.2 The coronaviruses

Coronavirus is a genus of animal virus which belongs to the family *Coronaviridae*. Coronaviruses are large, enveloped, single-stranded positive sense RNA viruses that replicate in the cytoplasm of animal host cells (Siddell *et al.*, 1983). The genomic size of coronaviruses ranges from approximately 27 to 32 kb. The size of each coronavirus particle is about 100 to 140 nm in diameter. Most of the coronaviruses show distinct crown (*'corona'* in Latin) surface appearance which is attributed to the 20 nm projections of spike (S) glycoprotein from the virion surface.

[18]Sputum, oropharyngeal and nasopharyngeal materials

Previously, coronaviruses have been divided into three groups based on serological cross-reactivity and genomic sequence homology (Siddell *et al.*, 1983; Gonzalez *et al.*, 2003). Mammalian viruses are found in groups 1 and 2, and avian viruses are found only in group 3. Group 1 includes canine coronavirus (CCV), feline coronavirus (FIPV), human coronavirus 229E (HCoV-229E), porcine epidemic diarrhea virus (PEDV), transmissible gastroenteritis virus (TGEV), and human coronavirus-NL63 (HCoV-NL63). Group 2 includes bovine coronavirus (BCV), human coronavirus OC43 (HCoV-OC43), mouse hepatitis virus (MHV) and rat coronavirus (RCV). Group 3 includes avian infectious bronchitis virus (AIBV) and turkey coronavirus (TCoV) (Siddell *et al.*, 1983; Gonzalez *et al.*, 2003).

In the past, human coronaviruses found in both group 1 (HCoV-229E) and group 2 (HCoV-OC43) were associated only with mild upper respiratory tract diseases (Makela *et al.*, 1998), whereas the novel SARS-CoV appears to be the first human coronavirus responsible for severe disease in humans. Phylogenetic analysis indicates that SARS-CoV does not belong to any of the above three groups, which contain all other known coronaviruses, including the human coronaviruses. This suggests that SARS-CoV did not arise by recombination or mutation of the known coronaviruses; hence, the novel SARS-CoV defines a distinct fourth group of coronaviruses (Marra *et al.*, 2003; Rota *et al.*, 2003).

1.2.3 Genome of SARS-CoV

SARS-CoV contains a polyadenylated RNA genome of approximately 30 kb, which is organized into 13 to 15 open reading frames (ORFs) (Marra *et al.*, 2003; Rota *et al.*, 2003; Thiel *et al.*, 2003). Its overall genome organization is typical of other coronaviruses: 5' - replicase - spike (S) - envelope (E) - membrane (M) - nucleocapsid (N) - 3' (Figure 1.1).

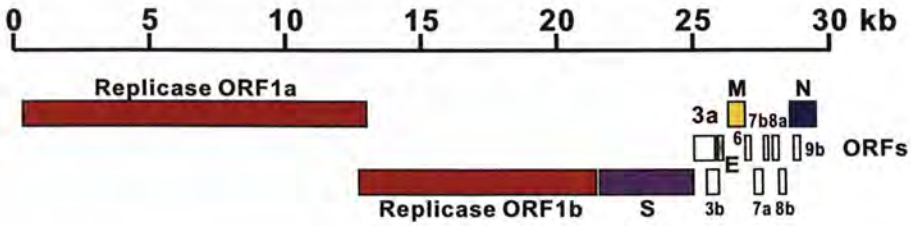

Figure 1.1: Genome organization of SARS-CoV.

The organization of SARS-CoV genome: 5' - replicase - spike (S) - envelope (E) - membrane (M) - nucleocapsid (N) - 3' (Modified from Thiel *et al.*, 2003).

The overlapping replicase ORF1a and ORF1b comprises two-thirds of the SARS-CoV genome, which are located at the 5' region downstream of the predicted RNA leader sequence (nucleotides 1-71) and an untranslated region (UTR) spanning 192 nucleotides. The replicase genes encode a number of proteins required for viral replication and transcription via proteolytic cleavage of a large polyprotein; more details on replicase polyproteins are discussed in section 1.4.

Structural proteins S, E, M and N are encoded by four ORFs located at the 3' region of the SARS-CoV genome, in which they are crucial to host cell entry, virion morphogenesis and release (Marra *et al.*, 2003) (Figure 1.2).

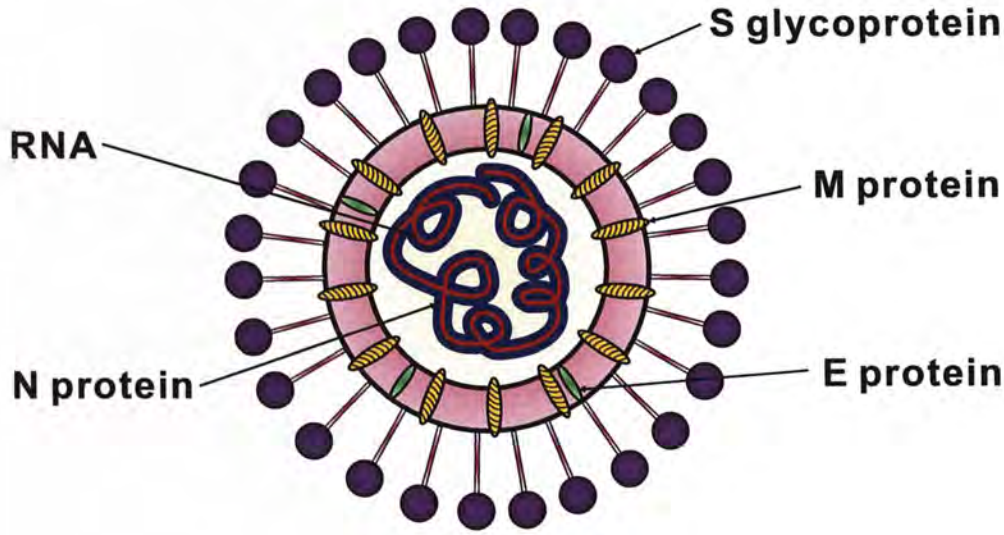

Figure 1.2: Structural proteins of SARS-CoV.

Structural proteins of SARS-CoV include spike (S) glycoprotein, envelope (E) protein, membrane (M) protein and nucleocapsid (N) protein. (Modified from Groneberg *et al.*, 2005).

The nucleocapsid (N) protein and the membrane (M) protein interact to form a spherical core, whereas the spike (S) glycoprotein, and the envelope (E) protein constitute the viral envelope (Marra *et al.*, 2003; Rota *et al.*, 2003; Thiel *et al.*, 2003). During the assembly of viral particles, the helical nucleocapsid is formed by the binding of N protein to a defined packaging signal on viral RNA.

Interactions between the E protein, M protein and the nucleocapsid trigger the budding event of the virion; and S glycoprotein is incorporated into the viral envelope which mediates the release of mature viral particles (Garoff *et al.*, 1998).

S glycoprotein plays an important role in binding to host-specific receptors and facilitating viral entry and fusion; more details on S glycoprotein are discussed in section 2.1 (Lai and Cavanagh, 1997; Gallagher and Buchmeier, 2001; Bosch *et al.*, 2003; Holmes, 2003; Hofmann and Pohlmann, 2004). N protein is responsible for providing nuclear-import signals, interfering cellular process, viral replication and RNA packaging (Kuo and Masters, 2002; Hiscox *et al.*, 2001). Previous studies indicated that the N protein from other strains of coronaviruses can cause massive hepatocellular necrosis[19] in the host body (Ning *et al.*, 2003). E protein is involved in virion envelope formation and the generation of virus-like particles when coexpressing with M protein (Vennema *et al.*, 1996), and also the induction of apoptosis (An *et al.*, 1999).

In addition to these proteins, the genome of SARS-CoV also contains several accessory genes that encode a number of non-structural proteins (Marra *et al.*, 2003; Rota *et al.*, 2003).

[19]The localized death of living cells

1.3 Spike (S) glycoprotein of SARS-CoV

The distinct and characteristic crown appearance on the surface of the SARS-CoV is attributed to the membrane projections of the spike (S) glycoprotein from the viral envelope (Schmidt *et al.*, 1987; Ksiazek *et al.*, 2003). The open reading frame encoding for S glycoprotein of SARS-CoV is 3768 nucleotides in length, comprising 12.67% of the total genome. The S glycoprotein is the largest structural viral protein in SARS-CoV and it is composed of 1255 amino acids with an estimated molecular mass of 139.12 kDa (Choy *et al.*, 2004).

There are 23 potential N-linked glycosylations and 69 potential sites of phosphorylation, including 37 serine, 15 threonine, and 17 tyrosine residues (Choy *et al.*, 2004) as predicted by sequence analysis. The S glycoprotein of SARS-CoV demonstrates a low sequence similarity level of 20% to 27% pairwise amino acid identity with the previous identified S glycoproteins of other known coronaviruses (Rota *et al.*, 2003; Choy *et al.*, 2004). The SARS-CoV S glycoprotein is a type I intergral membrane glycoprotein that consists of a leader (residues 1-14), an ectodomain (residues 15-1190), a transmembrane domain (residues 1191-1227), and a short intracellular tail (residues 1227-1255) rich in conserved cysteine residues (Rota *et al.*, 2003).

1.3.1 Functions of SARS-CoV S glycoprotein

The S glycoprotein of SARS-CoV is a typical type I viral fusion protein that plays a significant role in mediating the viral binding to species-specific host cell receptors, which further triggers a fusion event between the viral envelope and a cellular membrane to facilitate viral entry (Lai and Cavanagh, 1997; Gallagher and Buchmeier, 2001; Bosch *et al.*, 2003; Holmes, 2003; Hofmann and Pohlmann, 2004). Thus, it serves as the major inducer of neutralizing antibodies. Besides, the SARS-CoV S glycoprotein is also responsible for SARS-CoV pathogenesis, virulence and host cell tropism (Leparc-Goffart *et al.*, 1998; Sanchez *et al.*, 1999; Phillips *et al.*, 2002).

In fact, the S glycoprotein of SARS-CoV is composed of two subunits: the surface S1 subunit and the transmembrane S2 subunit. In contrast to the S glycoproteins of other coronaviruses such as mouse hepatitis virus (MHV), bovine coronavirus (BCV), avian infectious bronchitis virus (AIBV) and human coronavirus OC43 (HCoV-OC43) that undergo proteolytic cleavage into the two non-covalently associated S1 and S2 subunits (Stern and Sefton, 1982; Sturman and Holmes, 1984; Frana *et al.*, 1985; Jackwood *et al.*, 2001), there is no typical cleavage motif being identified in the SARS-CoV S glycoprotein (Rota *et al.*, 2003; He *et al.*, 2005). On the other hand, this issue remains controversial as other studies have demonstrated the cleavage of the SARS-CoV S glycoprotein in virus

infected cell lysate (Wu *et al.*, 2004*c*). Nevertheless, the S1 and S2 subunits of the SARS-CoV S glycoprotein can be predicted *in silico* via sequence alignment with the S1 and S2 subunits of the cleaved S glycoproteins of other coronaviruses, in which they are predicted to be localized at the regions of amino acid residues 17-680 and 681-1255, respectively (Rota *et al.*, 2003; Spiga *et al.*, 2003). The S1 subunit harbors the receptor-binding domain which defines the host range of the virus (Cavanagh *et al.*, 1986; Taguchi, 1995), and it is suggested to be the major antigenic determinants for eliciting immune response in the hosts (Cavanagh *et al.*, 1986; Correa *et al.*, 1990; Delmas *et al.*, 1990; Kapczynski *et al.*, 2003).

The S2 subunit is the stem region of the S glycoprotein that contains a putative fusion peptide and two highly conserved heptad repeats (HR1 and HR2), which are responsible for the fusion between viral and cellular membranes (Bosch *et al.*, 2003). During the virus fusion process, the HR1 region at the amino terminus and HR2 region at the carboxyl terminus undergo a series of conformational changes, from the pre-fusion state to the pre-hairpin intermediate state, and then to the final post-fusion hairpin state (Eckert and Kim, 2001). Previous studies demonstrated that three HR1 helices interact with three HR2 helices in an anti-parallel manner to form a stable six-helix coiled-coil bundle structure (Bosch *et al.*, 2004; Ingallinella *et al.*, 2004; Liu *et al.*, 2004; Tripet *et al.*, 2004; Xu

et al., 2004*b*), which resembles the fusion-active core of the mouse hepatitis virus (MHV) S glycoprotein (Xu *et al.*, 2004*a*) and human immunodeficiency virus type 1 (HIV-1) gp41 from the envelope glycoprotein (Chan *et al.*, 1997). This trimer of hairpins fusion core facilitates the juxtaposition of the viral and cellular membranes into close proximity, destabilizes the membrane lipid bilayers, which ultimately leads to the fusion and release of the viral nucleocapsid into the host cell (Tripet *et al.*, 2004).

1.3.2 Receptors for S glycoprotein of SARS-CoV

The coronavirus S glycoprotein mediates viral infection of receptor-expressing target cells (Lai and Cavanagh, 1997; Gallagher and Buchmeier, 2001; Bosch *et al.*, 2003; Holmes, 2003; Hofmann and Pohlmann, 2004), thus the identification of receptors for S glycoproteins can provide insights into the molecular mechanisms of coronavirus entry, pathogenesis and host cell tropism.

Although several S glycoprotein receptors were previously identified for other coronaviruses, such as the aminopeptidase N (APN) glycoproteins for human coronavirus 229E (HCoV-229E) and murine carcinoembryonic antigen cell adhesion molecule 1 (CEACAM1) for mouse hepatitis virus (MHV) (Yeager *et al.*, 1992; Wentworth and Holmes, 2001; Marra *et al.*, 2003); SARS-CoV S glycoprotein does not utilize any of these receptors, which is accounted by its low

sequence similarity with other coronavirus S glycoproteins (Marra *et al.*, 2003; Rota *et al.*, 2003).

Angiotensin-converting enzyme 2 (ACE2)

By using the immunoprecipitation and mass spectrometry approach, angiotensin-converting enzyme 2 (ACE2) (Donoghue *et al.*, 2000; Tipnis *et al.*, 2000) was first identified by Li's group to be a functional receptor for the S glycoprotein of SARS-CoV (Li *et al.*, 2003); which is consistent with the independent finding of Wang's group using the flow cytometry and expression cloning approach (Wang *et al.*, 2004). ACE2 is a type I integral membrane protein that is composed of 805 amino acids (Donoghue *et al.*, 2000; Tipnis *et al.*, 2000). It is a zinc-containing carboxypeptidase that converts angiotensin I to angiotensin II in the renin-angiotensin system, which plays an essential role in the regulation of heart function by maintaining blood pressure homeostasis (Crackower *et al.*, 2002; Eriksson *et al.*, 2002).

A 193-amino-acid receptor binding domain (RBD) of the SARS-CoV S glycoprotein is demonstrated to bind ACE2 with high affinity and efficiently block S glycoprotein-mediated SARS-CoV infection (Wong *et al.*, 2004). The high expression level of ACE2 in lung, kidney, heart and the gastrointestinal systems (Harmer *et al.*, 2002; Komatsu *et al.*, 2002) can account for the isolation of SARS-CoV

from each of these target tissues in SARS-CoV infected humans and animals. Several lines of evidence suggested ACE2 as the primary physiologically relevant receptor for SARS-CoV infection, and may play a critical role in SARS-CoV replication in the infected host cells (Li *et al.*, 2003; Hamming *et al.*, 2004; Nie *et al.*, 2004).

CD209L

In addition to the functional receptor ACE2, a different human cellular glycoprotein, CD209L (also called L-SIGN, DC-SIGNR and DC-SIGN2), was claimed to be an alternative receptor for the S glycoprotein of SARS-CoV (Jeffers *et al.*, 2004). It is demonstrated that human CD209L interacts specifically with purified recombinant SARS-CoV S glycoprotein (Jeffers *et al.*, 2004). Although CD209L was shown to enhance S glycoprotein-mediated SARS-CoV infection of ACE2-expressing permissive cells in *trans*, CD209L expression does not facilitate SARS-CoV S-driven infection of non-permissive cells in the absence of ACE2 (Marzi *et al.*, 2004; Yang *et al.*, 2004*a*).

1.4 3C-like protease ($3CL^{pro}$) of SARS-CoV

1.4.1 Extensive proteolytic processing of SARS-CoV replicase polyproteins

Common to other coronaviruses, SARS-CoV replicates in the cytoplasm of the infected host cells. The 5'-proximal overlapping replicase ORF1a and ORF1b are translated into two large replicase polyproteins pp1a and pp1ab (Marra *et al.*, 2003; Rota *et al.*, 2003). It is predicted that the expression of ORF1b-encoded region of pp1ab involves a translation read-through by a -1 ribosomal frameshift mechanism upstream of the ORF1a translation termination codon (Thiel *et al.*, 2003). The 486-kDa polyprotein pp1a contains a papain-like protease ($PL2^{pro}$), a 3C-like protease ($3CL^{pro}$), two putative membrane proteins MP1 (nsp4) and MP2 (nsp6), and several proteins of unknown function. The 790-kDa polyprotein pp1ab contains a helicase domain (nsp13) with ATPase and DNA duplex unwinding activities, a RNA polymerase (nsp12), an exonuclease (nsp14), an endoribonuclease (nsp15), and a methyltransferase (nsp16) (Schmidt-Mende *et al.*, 2001).

The pp1a and pp1ab polyproteins undergo extensive processing by viral proteases to yield 16 nonstructural protein products (nsp1 to nsp16) that assemble into a membrane-bound viral replication complex, for mediating both genome replication and transcription

of subgenomic mRNAs (Ziebuhr *et al.*, 2000; Thiel *et al.*, 2003). SARS-CoV $PL2^{pro}$ cleaves the N-proximal region of pp1a and pp1ab at three sites, whereas SARS-CoV $3CL^{pro}$ cleaves the central and C-proximal polyproteins regions at 11 conserved sites (Ziebuhr *et al.*, 2000; Thiel *et al.*, 2003).

1.4.2 SARS-CoV $3CL^{pro}$

The 33.8-kDa SARS-CoV $3CL^{pro}$, which is also known as the SARS-CoV main protease (M^{pro}), 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). The reported X-ray crystal structures of the SARS-CoV $3CL^{pro}$ allow the investigation of possible interactions of SARS-CoV $3CL^{pro}$ with its substrates, which contribute to the development of specific protease inhibitors as potential anti-SARS drugs (Anand *et al.*, 2003; Yang *et al.*, 2003) (Figure 1.3).

SARS-CoV $3CL^{pro}$ is a homodimer with two perpendicularly-arranged subunits. Each monomer of SARS-CoV $3CL^{pro}$ comprises three structural domains including the catalytic domain I (residues 8-101) and domain II (residues 102-184) that form the N-terminal chymotrypsin fold, and the unique C-terminal extra α-helical domain III (residues 201-301) that is important for regulating the activity and specificity of SARS-CoV $3CL^{pro}$ by controlling the

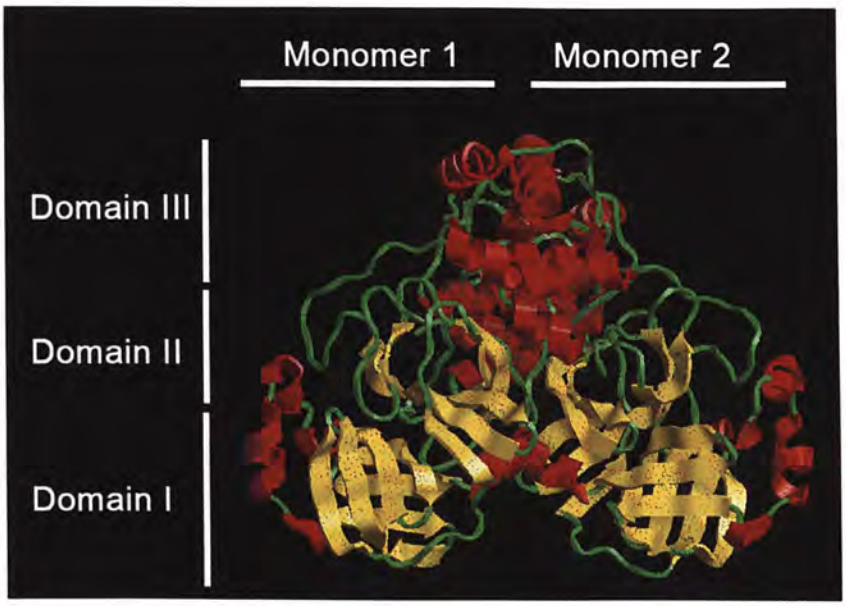

Figure 1.3: Dimer structure of SARS-CoV 3CL^{pro}.

SARS-CoV 3CL^{pro} is a homodimer, in which each monomer comprises structural domains I, II, and III. (Modified from Yang *et al.*, 2003).

dimerization of the enzyme (Shi *et al.*, 2004). Although SARS-CoV 3CL^{pro} resembles a chymotrypsin-like two-β-barrel fold, it employs the active nucleophile Cys145 and the acid-base catalyst His41 at the catalytic dyad instead of the classical Ser-His-Asp catalytic triad, by lacking an acidic residue (Thiel *et al.*, 2003).

1.4.3 Substrate specificity of SARS-CoV 3CL^{pro}

Based on sequence analysis (Figure 4.1), it is demonstrated that SARS-CoV 3CL^{pro} cleaves the replicase polyproteins at no less than

11 conserved sites, in which the substrate specificity involves preferentially the Leu-Gln ↓ sequence in P2 and P1 positions[20] at the cleavage site, respectively (Ziebuhr *et al.*, 2000; Hegyi and Ziebuhr, 2002; Anand *et al.*, 2003). Previous studies on the $3CL^{pro}$ of other coronaviruses suggested the conservation of the P1 (Gln), P2 (Leu, Ile, Val, Phe, Met), P4 (Ser, Thr, Val, Pro, Ala), P1' (Ser, Ala, Gly, Asn, Cys) positions at the cleavage site in determining the $3CL^{pro}$ substrate specificity (Ziebuhr *et al.*, 2000). In addition to the *in silico* sequence analyses of the SARS-CoV $3CL^{pro}$, 'wet-lab' experiments are necessary to provide a more comprehensive and valid interpretation on the substrate specificity preference of SARS-CoV $3CL^{pro}$, for a better understanding on the specific enzyme-substrate binding mechanism during the SARS-CoV $3CL^{pro}$ proteolytic cleavage process.

[20] According to the naming system suggested by Berger and Schechter (Berger and Schechter, 1970), the position (P) on the substrate is counted from the point of cleavage, i.e. P1 is the position just before the cleavage site ↓, with P1, P2, ... etc. located at N-terminal to the cleavage site; and P1', P2', ... etc. located at C-terminal to the cleavage site.

1.5 Combating SARS - Vaccine and drug development

1.5.1 Vaccine development against SARS

The development of safe and effective vaccines and drugs is the ultimate goal in this combat against SARS. Currently, mainstream anti-SARS vaccine candidates include inactivated SARS-CoV vaccines, DNA vaccines, and attenuated viral-vectored vaccines; nonetheless, only the inactivated SARS-CoV vaccines are tested in clinical trial (He *et al.*, 2004; Marshall and Enserink, 2004; Zhou *et al.*, 2005).

Inactivated SARS-CoV vaccines were prepared by treating SARS-CoV infected Vero cell lysates with β-propiolactone or formaldehyde; to date, these inactivated SARS-CoV vaccines have been evaluated for immunogenicity, safety and protectivity in mice, rabbits and rhesus monkeys (He *et al.*, 2004; Marshall and Enserink, 2004; Zhou *et al.*, 2005). Besides, DNA vaccines are studied extensively as potential SARS vaccine candidates. DNA plasmids encoding the SARS-CoV spike (S) glycoprotein or nucleocapsid (N) protein were synthesized using human-preferred codon, so as to optimize gene expression and to minimize the chance of recombination with endogenous coronaviruses. Neutralizing antibody responses, CD4 and CD8 T cells-mediated responses, and protective immunity were induced in mice that were immunized with DNA plasmids (Kim *et al.*, 2004;

Yang *et al.*, 2004*b*). In addition, attenuated viral-vectored vaccines such as the highly attenuated modified vaccinia virus Ankara (MVA) and attenuated parainfluenza virus (BHPIV3) containing the gene encoding the codon-optimized SARS-CoV S glycoprotein, were used to immunize mice and African green monkeys, which demonstrated reduction of SARS-CoV replication and complete protection against shedding of SARS-CoV, respectively (Bisht *et al.*, 2004; Bukreyev *et al.*, 2004).

1.5.2 Drug development against SARS

Owing to the important biological role of SARS-CoV $3CL^{pro}$ in the viral life cycle, it is being considered extensively as an attractive and promising target for anti-SARS drug design (Anand *et al.*, 2003). Efforts in screening different potential inhibitors of SARS-CoV $3CL^{pro}$ are underway worldwide.

To date, a number of potential candidates of SARS-CoV $3CL^{pro}$ inhibitors were identified from different compound and drug libraries, such as antiviral drugs[21] already in human clinical use, Food and Drug Administration (FDA) approved compounds, National Cancer Institute (NCI) small molecules database, MDL Drug Data Report (MDDR) database, MDL Available Chemicals Directory (ACD) database, LeadQuest database, Maybridge database, International

[21]Drugs for treating other coronaviruses and HIV-1

Species Information System (ISIS) database, MicroSource Genesis Plus Collection database, Chinese Medicine Database (TCMD), Marine Natural Products Database (MNPD), ... etc (Hsu *et al.*, 2004; Kao *et al.*, 2004; Rajnarayanan *et al.*, 2004; Wu *et al.*, 2004*a*; Yamamoto *et al.*, 2004; Liu and Zhou, 2005). Most of these SARS-CoV $3CL^{pro}$ inhibitor candidates were proposed using *in silico* molecular modeling and virtual screening approaches, in which only a few of them were experimentally tested for inhibitory activity.

Nelfinavir, a widely used human immunodeficiency virus type 1 (HIV-1) protease inhibitor, was demonstrated to strongly inhibit SARS-CoV replication (Yamamoto *et al.*, 2004). Two calpain[22] inhibitors also served as potent inhibitors of SARS-CoV replication (Barnard *et al.*, 2004). In addition, bifunctional aryl boronic compounds (specifically targeting the cluster of serine residues[23] near the SARS-CoV $3CL^{pro}$ active site) (Bacha *et al.*, 2004), metal-conjugated compounds (e.g. 1-hydroxypyridine-2-thione zinc) (Hsu *et al.*, 2004), keto-glutamine analogs (with a phthalhydrazido group at the α-position) (Jain *et al.*, 2004), and several synthetic peptide substrates (Yang *et al.*, 2003) were shown to be effective at inhibiting SARS-CoV $3CL^{pro}$.

[22]Calcium-activated cytoplasmic cysteine proteases

[23]Ser139, Ser144, and Ser147

1.6 Project objectives of this thesis

In this project, we mainly focused on using synthetic peptide-based approach to study two of the SARS-CoV proteins, including SARS-CoV S glycoprotein and SARS-CoV $3CL^{pro}$.

1.6.1 Synthetic Peptide Studies on SARS-CoV S glycoprotein

The potential receptor-binding site of SARS-CoV should lie within the surface region of S glycoprotein, and it would be a good target for the development of synthetic peptide-based vaccines against SARS-CoV. To date, there is no specific treatment for SARS, and the control of SARS-CoV infection by vaccination is not yet available. In this study, we focused on the use of biologically active synthetic peptides for viral protein neutralization, which aimed to block the viral invasion by eliciting an immune response that could specifically recognize and neutralize the S glycoprotein of SARS-CoV. Potential surface regions of SARS-CoV S glycoprotein served as the design basis of our peptides, which can be predicted *in silico* by sequence analyses on the secondary structure, hydrophobicity and posttranslational modifications. Synthetic peptides were used to immunize both rabbits and monkeys. Antisera were then analyzed by ELISA and tested for antibody specificity against SARS-CoV by immunofluorescent confocal microscopy. Our study can

also provide novel insights into future vaccine development against SARS-CoV by the synthetic peptide-based approach and help to achieve better understanding of SARS to allow better preparation for possible recurrences of SARS.

1.6.2 Synthetic Peptide Studies on SARS-CoV 3CL^{pro}

A better understanding of the specific enzyme-substrate binding mechanism during the SARS-CoV 3CL^{pro} proteolytic cleavage process is required for the rational design of an effective protease inhibitor with strong binding affinity to SARS-CoV 3CL^{pro}. To screen the substrate specificity of SARS-CoV 3CL^{pro} 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 (MALDI-TOF) mass spectrometric analysis in combination with the '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 P1 position at the SARS-CoV 3CL^{pro} cleavage site. Owing to the sensitivity of MALDI-TOF mass spectrometer, the protease cleavage reaction of a mixture of different peptide substrates can be monitored simultaneously to generate comprehensive data on the substrate specificity of SARS-CoV

$3CL^{pro}$. Therefore, all the twenty 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 SARS-CoV $3CL^{pro}$ 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 here is 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 type 1 (HIV-1) 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 analyses in proteomic research.

♦ **End of chapter.**

Chapter 2

Materials and Methods

2.1 Synthetic peptide studies on SARS-CoV S glycoprotein

2.1.1 Bioinformatics analyses of SARS-CoV S glycoprotein

The genome sequence of SARS-CoV used for peptide design in this study was based on the SARS-CoV CUHK-Su10 complete genome sequence [GenBank accession no. AY282752] downloaded from GenBank[1]. The correct open reading frame (ORF) of SARS-CoV spike (S) glycoprotein was determined by ORF Finder[2]. Predicted amino acid sequence of SARS-CoV S glycoprotein was compared with those of different coronaviruses belonging to the three existing groups.

[1] http://www.ncbi.nlm.nih.gov/entrez

[2] http://www.ncbi.nlm.nih.gov/gorf/gorf.html

Group 1 representatives included canine coronavirus (CCV) [GenBank accession no. AY342160], feline coronavirus (FIPV) [GenBank accession no. NC_007025], human coronavirus 229E (HCoV-229E) [GenBank accession no. AF304460], porcine epidemic diarrhea virus (PEDV) [GenBank accession no. AF353511], transmissible gastroenteritis virus (TGEV) [GenBank accession no. NC_002306], and human coronavirus NL63 (HCoV-NL63) [GenBank accession no. NC_005831]. Group 2 representatives included bovine coronavirus (BCV) [GenBank accession no. AF220295], human coronavirus OC43 (HCoV-OC43) [GenBank accession no. AY903460], mouse hepatitis virus (MHV) [GenBank accession no. AF201929], and rat coronavirus (RCV) [GenBank accession no. AF207551]. Group 3 representatives included avian infectious bronchitis virus (AIBV) [GenBank accession no. M95169] and turkey coronavirus (TCoV) [GenBank accession no.AY342357]. Multiple sequence alignment and phylogenetic tree were generated by ClustalX (Ver. 1.83) software with Gonnet protein comparison matrix. The phylogenetic tree was displayed by TreeView (Ver. 1.6.6) software. NetNGlyc 1.0 server[3] was used to predict potential N-glycosylation sites, and NetPhos 2.0 Server[4] was used to predict potential phosphorylation

[3]NetGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) is a neural network-based program for predicting potential N-glycosylation sites, that attempts to discriminate glycosylated sequons from non-glycosylated ones.

[4]NetPhos (http://www.cbs.dtu.dk/services/NetPhos/) is a neural network-based program for predicting potential phosphorylation sites at serine, threonine or tyrosine residues.

sites (Blom *et al.*, 1999). The secondary structure and hydrophobicity analyses on the sequence of the SARS-CoV S glycoprotein were performed with DNASIS MAX (Ver. 2.0) software (MiraiBio) based on Chou and Fasman algorithm and Hopp and Woods algorithm, respectively (Chou and Fasman, 1978; Hopp and Woods, 1981).

2.1.2 Peptide design and molecular modeling

Peptide sequences that correspond to the potential surface regions of the SARS-CoV S glycoprotein were designed based on the combined data from the above protein analyses (Table 2.1).

Table 2.1: Amino acid sequences of the six synthetic peptides[a].

Peptide	Amino acid positions	Amino acid sequence	No. of amino acids	Molecular mass, Da	pI
S1	75-96	TFGNPVIPFKDGIYFAATEKSN	22	2416.6	6.74
S2	229-251	TNFRAILTAFSPAQDIWGTSAAA	23	2409.5	6.51
S3	573-593	ISPCSFGGVSVITPGTNASSE	21	2010.1	3.25
S4	1120-1140	YDPLQPELDSFKEELDKYFKN	21	2618.8	4.10
S5	788-820	LPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYG	33	3754.2	9.53
S6	1002-1030	ASANLAATKMSECVLGQSKRVDFCGKGYH	29	3072.4	9.05

[a] Provisional patent application number: 60/487,396 (Filing date: July 14, 2003).

The three-dimensional structures of these peptides in the solvated state were simulated on an Insight II molecular modeling platform (Accelrys) running on Silicon Graphics Octane 2 workstations. The

peptide models were constructed using the BIOPOLYMER module and then 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 300K for 100 picoseconds (ps) during which time the potential energy of the system reached a stable value. A time step of 1 femtosecond (fs) was used to integrate the equation of motion. The final conformation of the structure was obtained through 5000 iterations of steepest descent energy minimization.

2.1.3 Solid phase peptide synthesis (SPPS)

The six designed peptides were synthesized by a solid-phase technique with an Applied Biosystems 433A Peptide Synthesizer on amide resins using N^{α}-9-fluorenylmethyloxycarbonyl (Fmoc) synthesis (Figure 2.1) with 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate/1-hydroxybenzotriazole (HBTU / HOBt) coupling (Figure 2.2).

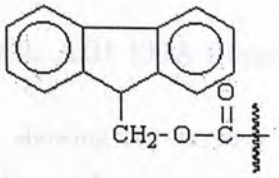

Figure 2.1: Fmoc-protecting group

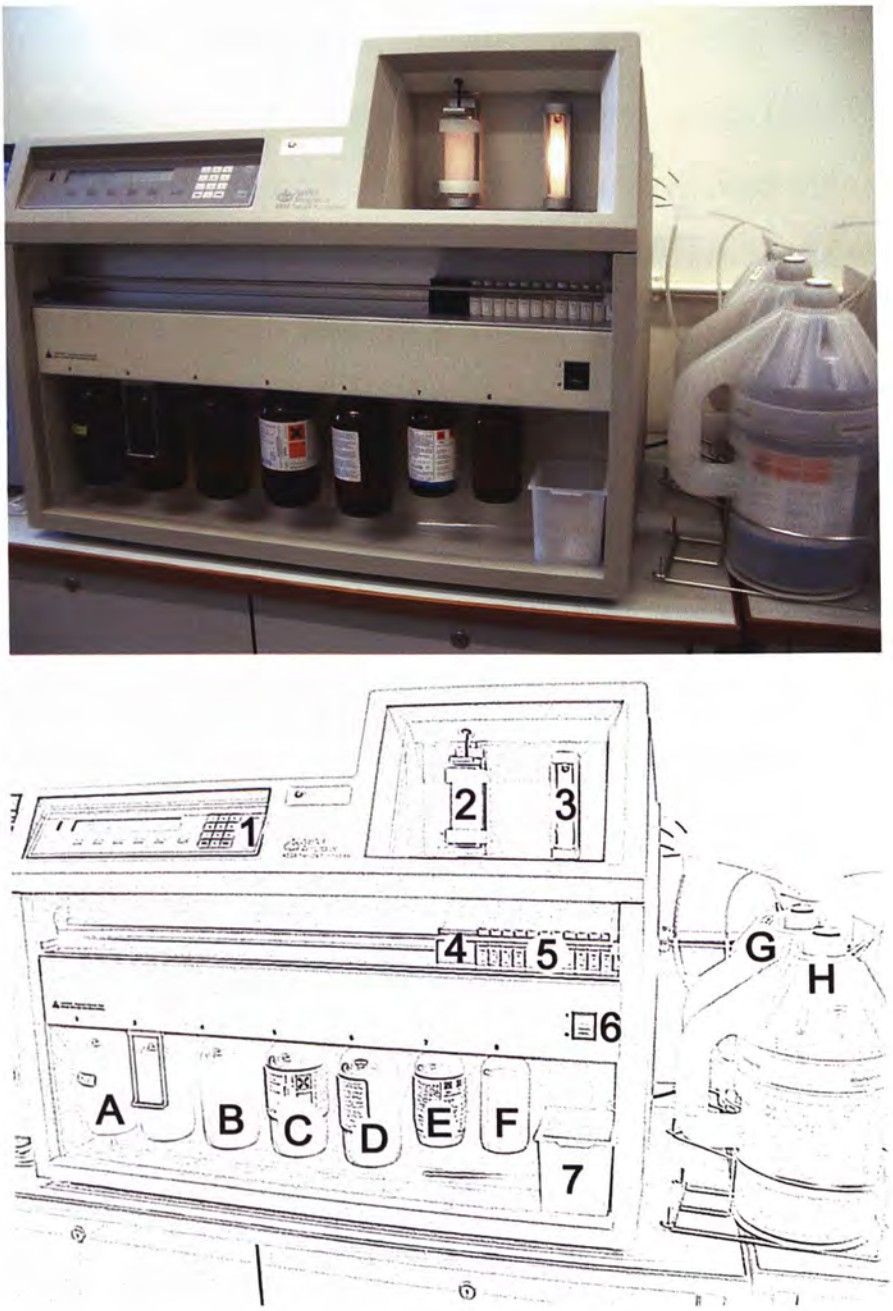

Figure 2.2: ABI 433A Peptide Synthesizer

Front View of the instrument, showing **(1)** keypad, **(2)** reaction vessel, **(3)** activator vessel, **(4)** pusher block, **(5)** amino acid cartridges, **(6)** powerswitch, and **(7)** cartridge disposal box. Reagents and solvents include **(A)** piperidine, **(B)** acetic anhydride in DMF, **(C)** HBTU/HOBT, **(D)** methanol, **(E)** diisopropylethylamine (DIEA) in NMP, **(F)** dicyclohexylcarbodiimide (DCC) in NMP, **(G)** dichloromethane (DCM), and **(H)** *N*-methylpyrrolidone (NMP).

The amino (N) termini of peptides were acetylated on the resin with 150 g/L acetic anhydride in dimethyl formamide (DMF). 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 peptides were then precipitated with cold diethyl ether, washed, and lyophilized. The cleaved peptides were purified with Agilent Technologies 1100 Series analytical and preparative HPLC systems using linear gradients formed from the solvent systems A (50 mL/L trifluoroacetic acid in H_2O) and B (0.5 mL/L trifluoroacetic acid in acetonitrile (ACN)). The identities of purified peptides were confirmed by mass spectrometry on an Ettan^{TM} matrix-assisted laser desorption/ionization time-of-flight Pro Mass Spectrometer (Amersham Biosciences).

2.1.4 Peptide conjugation

Keyhole limpet hemocyanin (KLH) was dissolved in phosphate-buffered saline (PBS) at pH 7.0 to a final concentration of 5 g/L. Approximately 1 mg of synthetic peptide was added to 1 mL of the KLH solution (i.e., approximately 200-300 mol peptides/mol of KLH). The mixture was sonicated in a water bath for 30 min; 5

mg of *N*-hydroxysuccinimide, 10 mg of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride were then added, and the mixture was stirred at room temperature for 30 min. The mixture was loaded on a HiTrap^{TM} Desalting Column (Amersham Biosciences), and the eluate was collected as the KLH-conjugated peptide solution.

2.1.5 Immunization in rabbits and monkeys [5]

Before the immunization of rabbits and monkeys, preimmune sera were collected and confirmed to be SARS-CoV negative by real-time quantitative RT-PCR as described previously (Ng *et al.*, 2003). In the primary injection, 0.5 mL of complete Freund's adjuvant was added to 0.5 mL of purified antigen with a peptide concentration of 1 g/L (either conjugate-free peptide or KLH-conjugated peptide) and emulsified. Each rabbit (male New Zealand White; weight, 2-3 kg) and monkey (male *Macaca fascicularis*[6]; weight, 2-3 kg) was immunized with 1 mL of emulsion by subcutaneous injection at five different sites on day 0 (Figure 2.3).

The rabbits and monkeys were then boosted by subcutaneous injection of an emulsion containing 0.5 mL of incomplete Freund's

[5]This part was conducted with the assistance of our collaborator Dr. Lin Shu-Guang's team in Research Center of Medical Sciences at Guangdong Provincial People's Hospital.

[6]The monkey strain *Macaca fascicularis* was chosen in this study because it has been demonstrated previously as a potent host of SARS-CoV that developed SARS symptoms after SARS-CoV infection (Fouchier *et al.*, 2003).

Figure 2.3: Immunization in *Macaca fascicularis.*

A: Monkey strain *Macaca fascicularis* was used for immunization to raise antibodies.
B: A monkey was being injected with the peptide antigen emulsion.

adjuvant and 0.5 mL of purified antigen on days 14 and 28. The first batch of rabbit and monkey antisera was collected 1 week after the second immunization.

2.1.6 ELISA analysis

Titrations of both the rabbit and monkey antisera against the corresponding peptide or peptide conjugate were determined in duplicate by ELISA and the results were averaged. The purified antigens of the six synthetic peptides (either conjugate-free peptide or KLH-conjugated peptide) were diluted to 10 mg/L in phosphate buffered saline (PBS); 100 μL of the diluted antigen solution was added to the corresponding well of a 96-well microplate and incubated at 4^{o}C

overnight. The wells were washed three times with 0.5 mL/L Tween 20 in PBS; 100 μL of PBS containing 50 mL/L skim milk and 0.5 mL/L Tween 20 was then added to each well and incubated at 37°C for 2 h for blocking. The rabbit and monkey antisera against the synthetic peptides were diluted in a twofold series to 1:25, 1:50, 1:100, 1:200, 1:400, 1:800, 1:1600, and 1:3200 in PBS containing 50 mL/L skimmed milk and 0.5 mL/L Tween 20, and 100 μL of diluted antiserum was added to each well. The plate was incubated at 37°C for 1 h. After plates were washed three times with PBS containing 0.5 mL/L Tween 20, bound antibody was measured by use of either alkaline phosphatase-conjugated goat anti-rabbit IgG or alkaline phosphatase-conjugated goat anti-monkey IgG (100 μL/well of 1:500 dilution of IgG in PBS containing 50 mL/L skimmed milk and 0.5 mL/L Tween 20). After incubation at 37°C for 1 h and three washes with PBS containing 0.5 mL/L Tween 20, 100 μL of developing solution (5 mg of *p*-nitrophenyl phosphate added to 5 mL of substrate buffer containing 50 mL of diethanolamine, 50 mg of $MgCl_2 \cdot H_2O$, and 97.5 mg of NaN_3; adjusted to pH 9.7 with NaOH and brought to 500 mL with MilliQ H_2O) was added to each well. Absorbance was monitored by a microplate reader at 415 nm. Both the rabbit and monkey preimmune sera were used as negative controls in this assay.

2.1.7 Immunofluorescent confocal microscopy [7]

The rabbit and monkey antisera against the synthetic peptides were diluted to 1:40-fold with PBS, and 10 μL of each diluted serum was added to a well of the slide that was coated with SARS-CoV-infected African green monkey kidney Vero cells and incubated at 37^{o}C for 1 h. The slide was washed three times with PBS and air dried. In each well of the slide, 10 μL of the diluted conjugate (1:40) and Evan Blue (1:400) were added, and the slide was incubated at 37^{o}C for 1 h. The slide was washed three times with PBS, air dried, and mounted. The mounted slide was observed with a confocal microscope. Noninfected African green monkey kidney Vero cells and the preimmune sera of both the rabbits and monkeys were used as negative controls in this assay.

[7]This part was conducted with the assistance of our collaborator Dr. Chan Paul Kay-Sheung's team in Department of Microbiology at The Chinese University of Hong Kong.

2.2 Synthetic peptide studies on SARS-CoV 3CL^{pro}

2.2.1 Protein expression and purification

The coding sequence of SARS-CoV 3CL^{pro} used for cloning was based on the SARS-CoV BJ01 strain complete genome sequence [GenBank accession no. AY278488] downloaded from GenBank[8]. The construction of the pGEX-3CL^{pro} plasmid[9] has been described previously (Yang *et al.*, 2003). The pGEX-3CL^{pro} plasmid was transformed into *Escherichia coli* (*E. coli.*) BL21(DE3) competent cells. Cultures were grown at 37^{o}C in 250 mL Luria-Bertani (LB) medium containing ampicillin (100 mg/L) until the cell density reached an absorbance of 0.6 at 600 nm, and the culture was induced with 0.1 mM isopropyl-1-thio-s-D-galactopyranoside (IPTG) at 25^{o}C for 6 h. After induction, cells were harvested by centrifugation at 4000 rpm for 20 min at 4^{o}C. The bacterial pellet was resuspended in buffer A (pH 7.5) containing 10 mM Tris-HCl, 500 mM NaCl, 1 mM dithiothreitol (DTT), 0.5 mM EDTA, and 0.2 mM phenylmethylsulfonyl fluoride (PMSF). After cell lysis by sonication, the cell lysate was centrifuged at 20000 rpm for 45 min at 4^{o}C. Supernatant was loaded onto a Glutathione Sepharose^{TM} 4B affinity column (Amersham Biosciences) equilibrated with buffer A.

[8]http://www.ncbi.nlm.nih.gov/entrez

[9]The pGEX-3CL^{pro} plasmid was provided by our collaborator Dr. Rao Zihe's team in Department of Biological Science and Biotechnology at Tsinghua University.

GST-tag was cleaved from the GST-3CL^{pro} fusion protein by on-column proteolytic digestion with PreScission^{TM} Protease (Amersham Biosciences) overnight at 4^{o}C. The SARS-CoV 3CL^{pro} recombinant protein was eluted with buffer A, concentrated to 10 g/L by Microcon® Centrifugal Filter Devices (Millipore) and then stored at -20^{o}C. The purity of protein samples were analyzed by discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) on a 12% gel as described (Laemmli, 1970), followed by Coomassie blue staining. Protein concentrations were determined by the Bradford protein assay reagent (Bio-Rad).

2.2.2 Solid phase peptide synthesis (SPPS)

The sequence of peptide substrate used in this study was based on one of the SARS-CoV 3CL^{pro} cleavage sites on the SARS-CoV BJ01 polyprotein pp1ab (residues 3232-3247) (Table 2.2).

The corresponding amino acid at the target position under investigation was deliberately substituted by 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^{pro} in a single reaction and then analyzed by MALDI-TOF

Table 2.2: Amino acid sequences of the seven synthetic peptide substrates.

Peptide substrate	Sequence[a]															
	P9	P8	P7	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'	P7'
PS01[b]	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_{20}[c]	Q	S	G	F	R	K	M	A
PS03	T	S	I	T	S	A	X_{20}[c]	L	Q	S	G	F	R	K	M	A
PS04	T	S	I	T	S	X_{20}[c]	V	L	Q	S	G	F	R	K	M	A
PS05	T	S	I	T	S	A	V	L	Q	X_{20}[c]	G	F	R	K	M	A
PS06	T	S	I	T	S	A	V	L	Q	S	X_{20}[c]	F	R	K	M	A
PS07	T	S	I	T	S	A	V	L	Q	S	G	X_{20}[c]	R	K	M	A

[a] The position (P) on the substrate is counted from the point of cleavage, i.e. P1 is the position just before the cleavage site ↓, with P1, P2, ... etc. located at N-terminal to the cleavage site; and P1', P2', ... etc. located at C-terminal to the cleavage site (Berger and Schechter, 1970).

[b] The control peptide substrate is based on the amino acid sequence of one of the SARS-CoV $3CL^{pro}$ cleavage site on the SARS-CoV BJ01 polyprotein pp1ab (residues 3232-3247).

[c] The corresponding amino acid at this position is substituted by a mixture of 20 standard amino acids in equal molar ratios.

mass spectrometry to detect the susceptibility of each of the 20 peptides to cleavage by SARS-CoV $3CL^{pro}$ (Table 2.2).

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 P1 position at the cleavage site were investigated one at a time using the 'cartridge replacement' solid-phase peptide synthesis strategy. Prior to peptide synthesis, the corresponding amino acid cartridge (Figure 2.4) at the target position under investigation was replaced by a 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 (Figure 2.5).

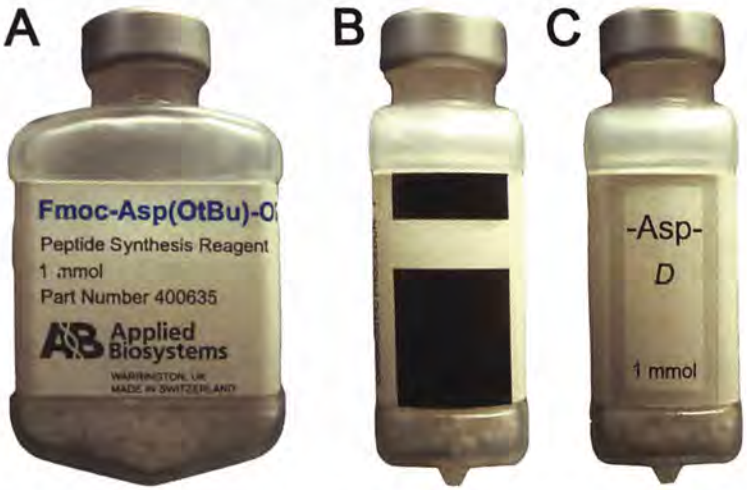

Figure 2.4: An amino acid cartridge.

A: Front view of an amino acid cartridge with label, showing this cartridge contains Asp. **B:** Left-side view showing the barcode, for identification of different amino acids by the peptide synthesizer. **C:** Right-side view with label (Asp).

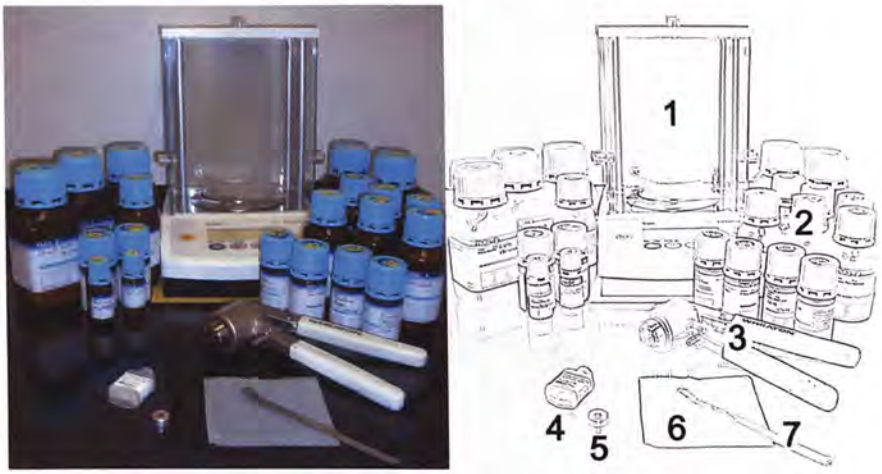

Figure 2.5: 'Cartridge replacement' solid phase peptide synthesis

Twenty kinds of amino acids are weighed and mixed in equal molar ratios. Here shows the **(1)** electronic balance, **(2)** twenty different amino acid powder in their corresponding bottles, **(6)** weighing paper, and **(7)** spatula. The mixture of amino acids is transferred to an empty amino acid cartridge **(4)**, and a cap **(5)** is sealed to the mouth of cartridge with a cap crimper **(3)**.

Six sets of such peptide substrate pools together with the original peptide substrate were synthesized with this approach (Table 2.2) using standard protocol as described in section 2.1.3.

2.2.3 Peptide cleavage assay

The *in vitro* peptide cleavage assay was performed in a total reaction volume of 25 μL at pH 7.6, containing 1.5 μM recombinant SARS-CoV $3CL^{pro}$, 0.5 mM peptide substrate, 20 mM Na_2HPO_4, 200 mM NaCl, 1 mM EDTA, and 1 mM DTT, with overnight incubation at 25^{o}C. Cleavage products were desalted by Reversed-Phase Zip-Tip® (Millipore) and the corresponding peptide peaks were resolved by mass spectrometry on an Ettan^{TM} MALDI-TOF

Pro Mass Spectrometer (Amersham Biosciences) (Figure 2.6). Internal calibration was performed using the angiotensin III (AngIII) and adrenocorticotropic hormone (ACTH) standards.

Figure 2.6: MALDI-TOF mass spectrometry.

A: Desalted cleavage products were loaded on to the 96-well MALDI-TOF sample plate. **B:** Sample plate was inserted to the tray of MALDI-TOF mass spectrometer. **C:** Front view of the MALDI-TOF mass spectrometer.

2.2.4 Molecular docking

All three-dimensional molecular modeling studies were performed on an Insight II molecular modeling platform (Accelrys) running on Silicon Graphics Octane 2 workstations. The consistent valence forcefield (CVFF) was selected in all simulations. The X-ray crystal structure of the SARS-CoV $3CL^{pro}$ (1UK4) was downloaded from the Protein Data Bank (PDB)[10] and used as a starting model. The catalytic active site of chain A in the SARS-CoV $3CL^{pro}$ structure without the bound CMK peptide was used in the study. Hydrogen atoms were added for the structure of SARS-CoV $3CL^{pro}$ chain A using the BUILDER module. Then, the forcefield 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 of SARS-CoV $3CL^{pro}$ mainly relies on the active nucleophile Cys145 and the acid-base catalyst His41, the region containing all residues (radius of 10.0 Å) 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.

[10] http://www.rcsb.org/pdb/

The best docking position was based on the docking energy of the peptide/SARS-CoV $3CL^{pro}$ complex. Thousand of orientations of peptides were searched by maneuvering it in the Cys-His dyad region of SARS-CoV $3CL^{pro}$ 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 300K for 100 picoseconds (ps) during which time the potential energy of the system reached a stable value. A time step of 1 femtosecond (fs) 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 $3CL^{pro}$ 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.

♦ **End of chapter.**

Chapter 3

Results

3.1 Synthetic peptide studies on SARS-CoV S glycoprotein

3.1.1 General features and structural analyses of the S glycoprotein

The open reading frame (ORF) encoding for the spike (S) glycoprotein of SARS-CoV is 3768 nucleotide in length, comprising 12.67% of the total genome. The S glycoprotein is the largest structural viral protein in SARS-CoV and is composed of 1255 amino acids with an estimated molecular weight of 139.12 kDa; its isoelectric point (pI) is 5.48. There are 23 potential sites of N-linked glycosylation (Figure 3.1), and 69 potential sites of phosphorylation, including 37 serine, 15 threonine and 17 tyrosine residues (Figure 3.2).

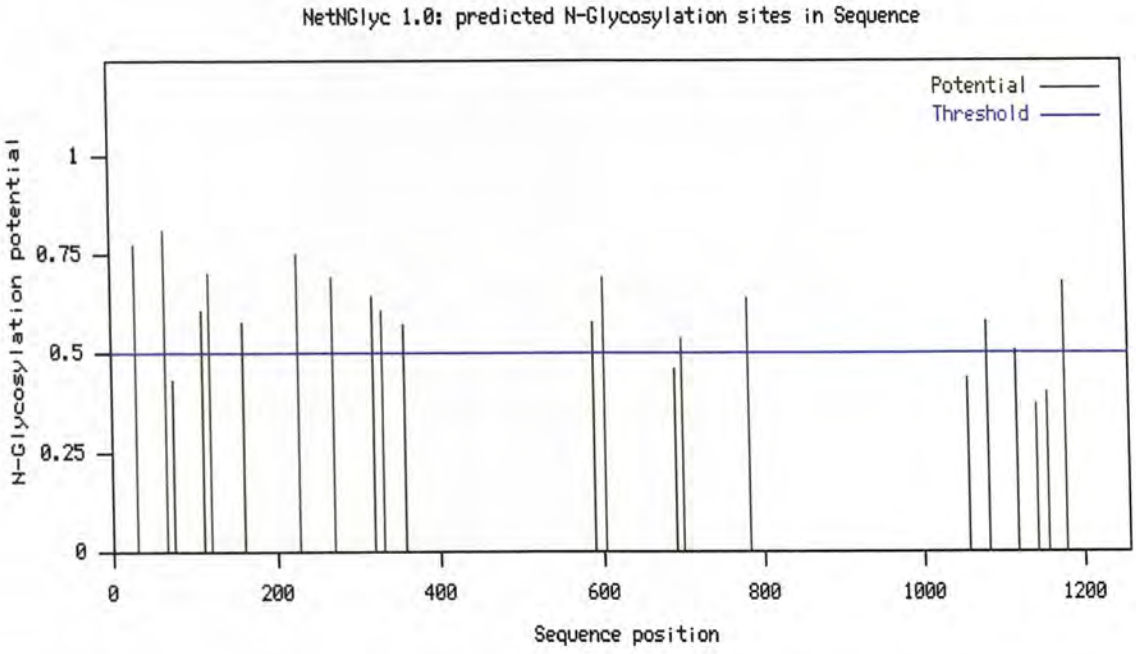

Figure 3.1: Predicted N-linked glycosylation sites on SARS-CoV S glycoprotein.

Twenty-three N-glycosylation sites on the S glycoprotein of SARS-CoV were predicted. NetGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) is a neural network-based program for predicting potential N-glycosylation sites, that attempts to discriminate glycosylated sequons from non-glycosylated ones. The graph illustrates predicted N-glycosylation sites across the protein sequence (x-axis: sequence position from N-terminal to C-terminal). A position with a potential (vertical green lines) crossing the threshold (horizontal blue line at 0.5) is predicted to be glycosylated (as long as it occurs in the required sequon Asn-Xaa-Ser/Thr without Proline at Xaa).

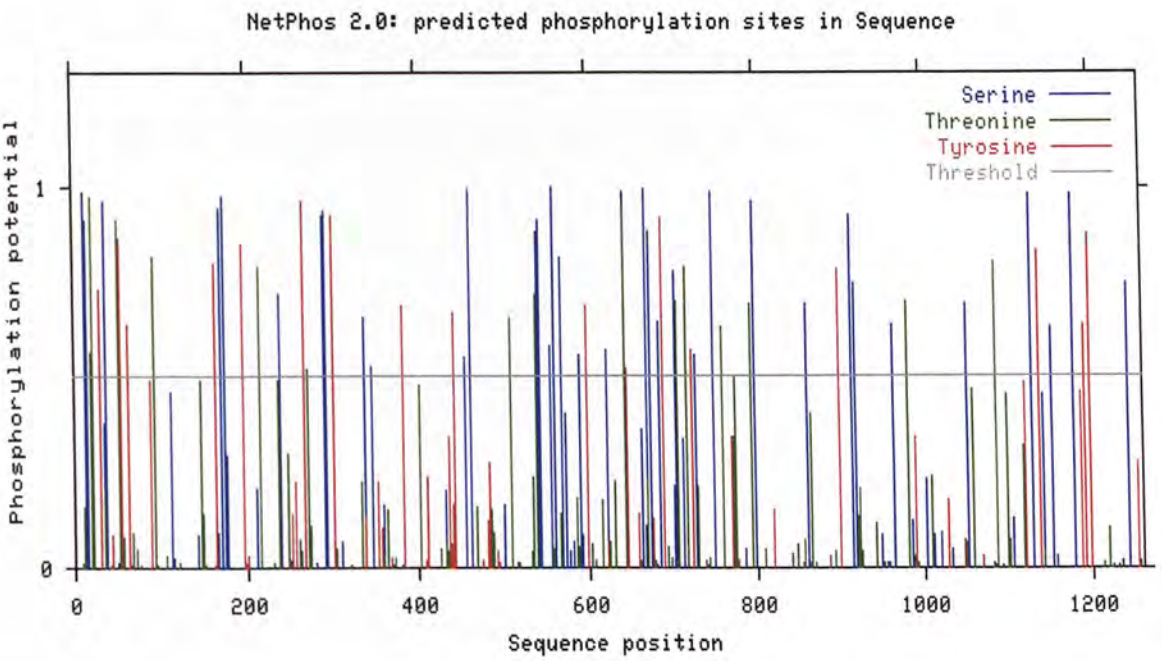

Figure 3.2: Predicted phosphorylation sites on SARS-CoV S glycoprotein.

Sixty-nine phosphorylation sites on the S glycoprotein of SARS-CoV, including 37 serine, 15 threonine, and 17 tyrosine residues were predicted. NetPhos (http://www.cbs.dtu.dk/services/NetPhos/) is a neural network-based program for predicting potential phosphorylation sites at serine, threonine or tyrosine residues in protein sequences. The graph illustrates predicted phosphoryaltion sites across the protein sequence (x-axis: sequence position from N-terminal to C-terminal). A position with a potential (vertical lines) crossing the threshold (horizontal gray line at 0.5) is predicted to be phosphorylated (Blom *et al.*, 1999).

The secondary structure and the hydrophobicity of the SARS-CoV S glycoprotein were predicted using DNASIS MAX (Ver. 2.0) software (Figure 3.3).

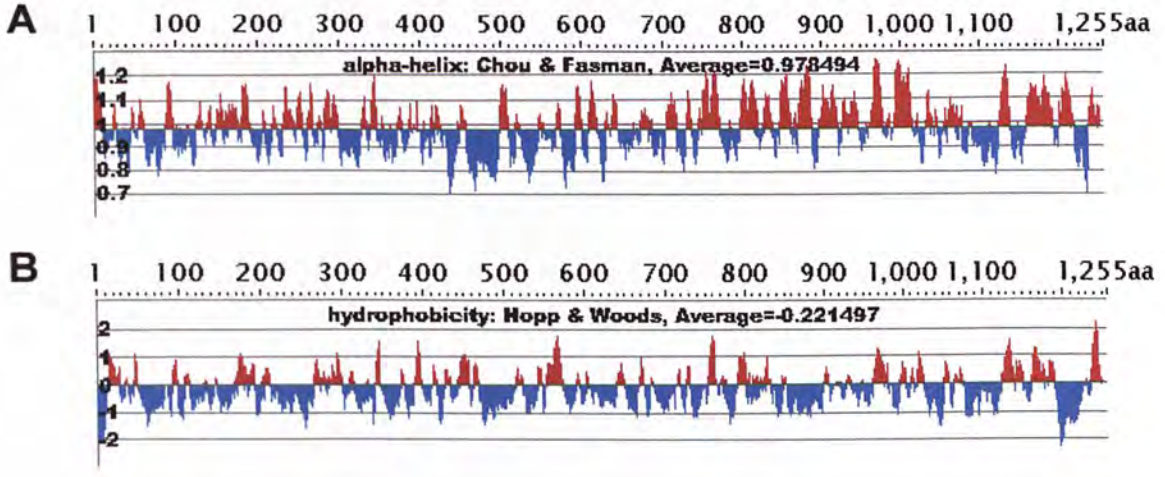

Figure 3.3: Sequence analyses of SARS-CoV S glycoprotein.

Secondary structure **(A)** and hydrophobicity **(B)** predictions on SARS-CoV S glycoprotein.

The S glycoprotein of SARS-CoV demonstrates less than 50% identity of the previous identified S glycoproteins of other known coronavirus. Phylogenetic analysis using multiple sequence alignment of different coronavirus S glycoproteins suggested that SARS-CoV defines a distinct fourth group of coronaviruses (Figure 3.4).

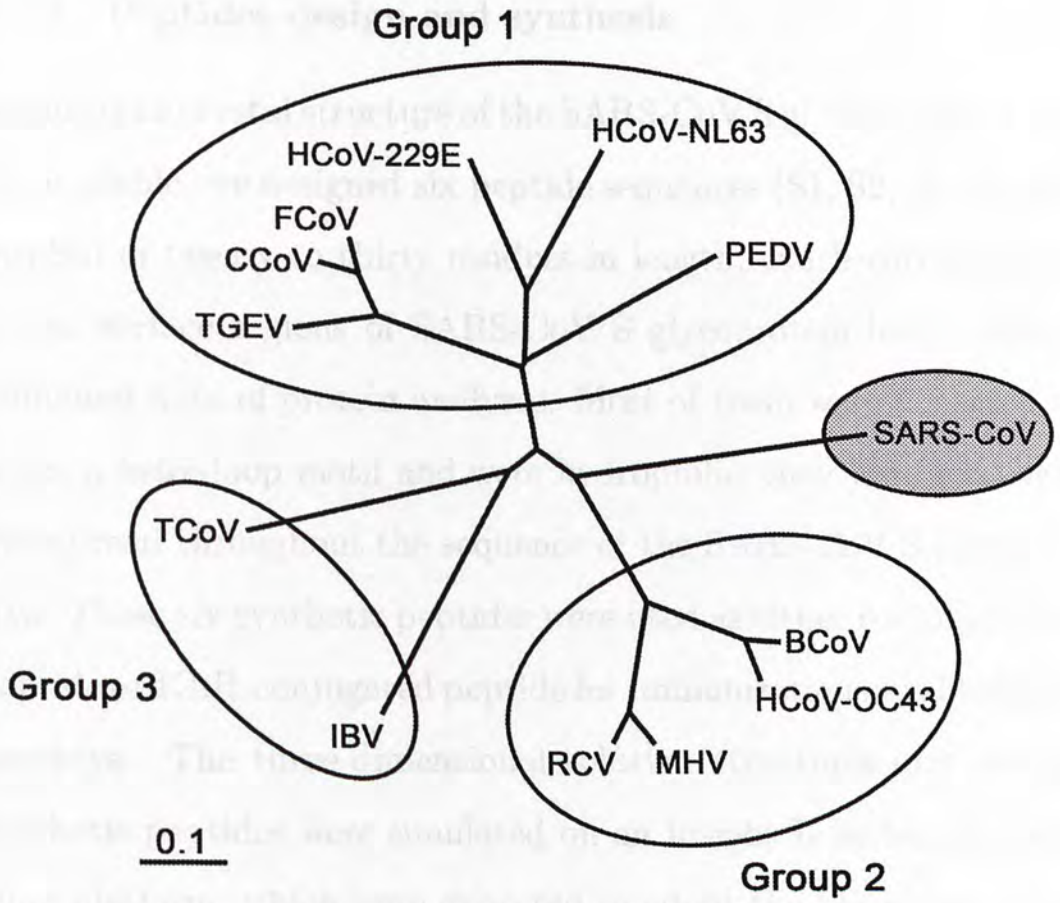

Figure 3.4: Phylogenetic relationships of SARS-CoV S glycoprotein to those of other coronaviruses.

Unrooted neighbor-joining tree of S glycoproteins of 13 coronavirus species suggested that SARS-CoV defined a distinct fourth group of coronaviruses. Predicted amino acid sequence of SARS-CoV S glycoprotein was compared with those of different coronaviruses belonging to the three existing groups [Group 1: canine coronavirus (CCV) feline coronavirus (FIPV), human coronavirus 229E (HCoV-229E), porcine epidemic diarrhea virus (PEDV), transmissible gastroenteritis virus (TGEV) and human coronavirus NL63 (NL63). Group 2: bovine coronavirus (BCV), human coronavirus OC43 (HCoV-OC43), mouse hepatitis virus (MHV), and rat coronavirus (RCV). Group 3: avian infectious bronchitis virus (AIBV) and turkey coronavirus (TCoV).] using ClustalX (Ver. 1.83) program. Branch lengths are proportional to genetic divergence. (The scale bar indicates the number of amino acid substitutions per site.)

3.1.2 Peptides design and synthesis

Because the crystal structure of the SARS-CoV S glycoprotein is not yet available, we designed six peptide sequences (S1, S2, S3, S4, S5, and S6) of twenty to thirty residues in length, which corresponded to the surface regions of SARS-CoV S glycoprotein based on the combined data of protein analyses. Most of them were expected to adopt a helix-loop motif and were hydrophilic; they were relatively widespread throughout the sequence of the SARS-CoV S glycoprotein. These six synthetic peptides were used as either conjugate-free peptide or KLH-conjugated peptide for immunization of rabbits and monkeys. The three-dimensional solution structures of these six synthetic peptides were simulated on an Insight II molecular modeling platform, which were expected to adopt the helix-loop motifs at their relatively stable energy states (Figure 3.5).

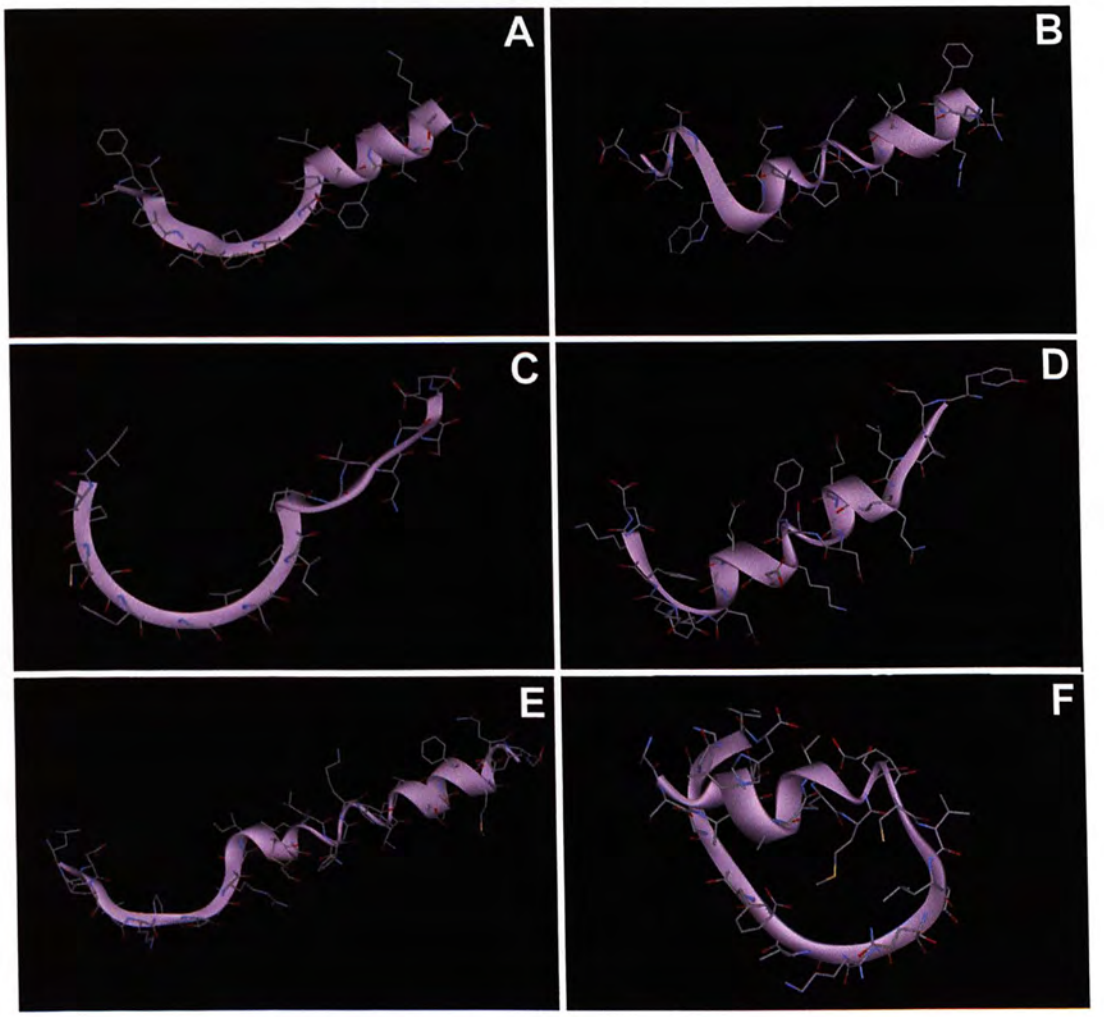

Figure 3.5: Predicted three-dimensional structures of the six synthetic peptides.

A: S1 peptide; **B:** S2 peptide; **C:** S3 peptide; **D:** S4 peptide; **E:** S5 peptide; and **F:** S6 peptide.

3.1.3 ELISA analysis and immunofluorescent confocal microscopy

Prior to the immunization of rabbit and monkey, sera were collected and were confirmed to be SARS-CoV negative by real-time quantitative RT-PCR. The first batch of the rabbit and monkey antisera against the six synthetic peptides was collected 1 week after the second immunization, and was tested for antibody specificity against the corresponding peptide (either conjugate-free peptide or KLH-conjugated peptide) by ELISA analysis. As shown in Figure 3.6 and Figure 3.7, all of these antisera were specific to their corresponding peptide antigens. Both the rabbit and monkey preimmune sera were used as negative controls in this assay.

These rabbit and monkey antisera were tested for antibody specificity against SARS-CoV by immunofluorescent confocal microscopy. The antisera were incubated with African green monkey kidney Vero cells infected by SARS-CoV obtained from a SARS-CoV infected patient. Negative results (Figure 3.8) in immunofluorescent confocal microscopy are indicated by a complete black microscopic field, whereas positive results (Figure 3.9) are indicated by the presence of green fluorescent signals. The results of immunofluorescent confocal microscopy are summarized in Table 3.1.

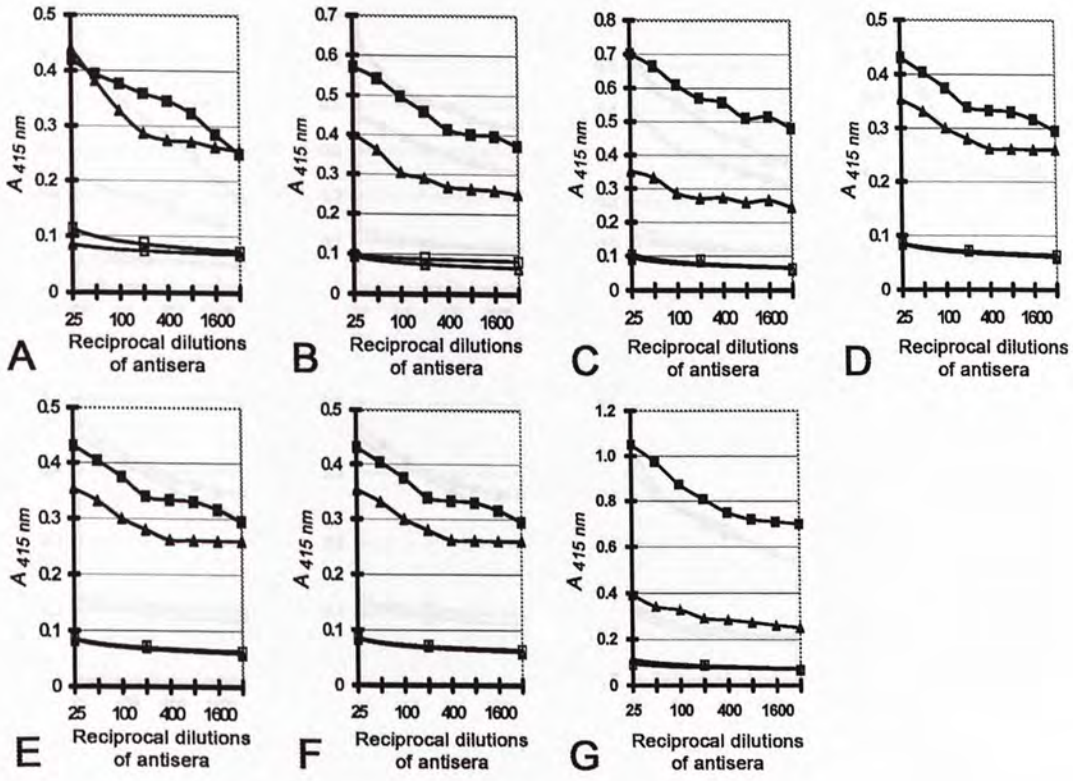

Figure 3.6: Titration of rabbit serum samples diluted from 1:25 in a twofold series against constant antigen concentration of 10 mg/L.

Anti-peptide antiserum is indicated by ▲ with the corresponding preimmune serum, indicated by △ , as negative control. Anti KLH-conjugated peptide antiserum is indicated by ■ , and the corresponding preimmune serum is indicated by □ as negative control. **A:** R_S1 antiserum (▲) and R_S1_KLH antiserum (■); **B:** R_S2 antiserum (▲) and R_S2_KLH antiserum (■); **:C:** R_S3 antiserum (▲) and R_S3_KLH antiserum (■); **D:** R_S4 antiserum (▲) and R_S4_KLH antiserum (■); **E:** R_S5 antiserum (▲) and R_S5_KLH antiserum (■); **F:** R_S6 antiserum (▲) and R_S6_KLH antiserum (■); **G:** R_MIX antiserum (▲) and R_MIX_KLH antiserum (■).

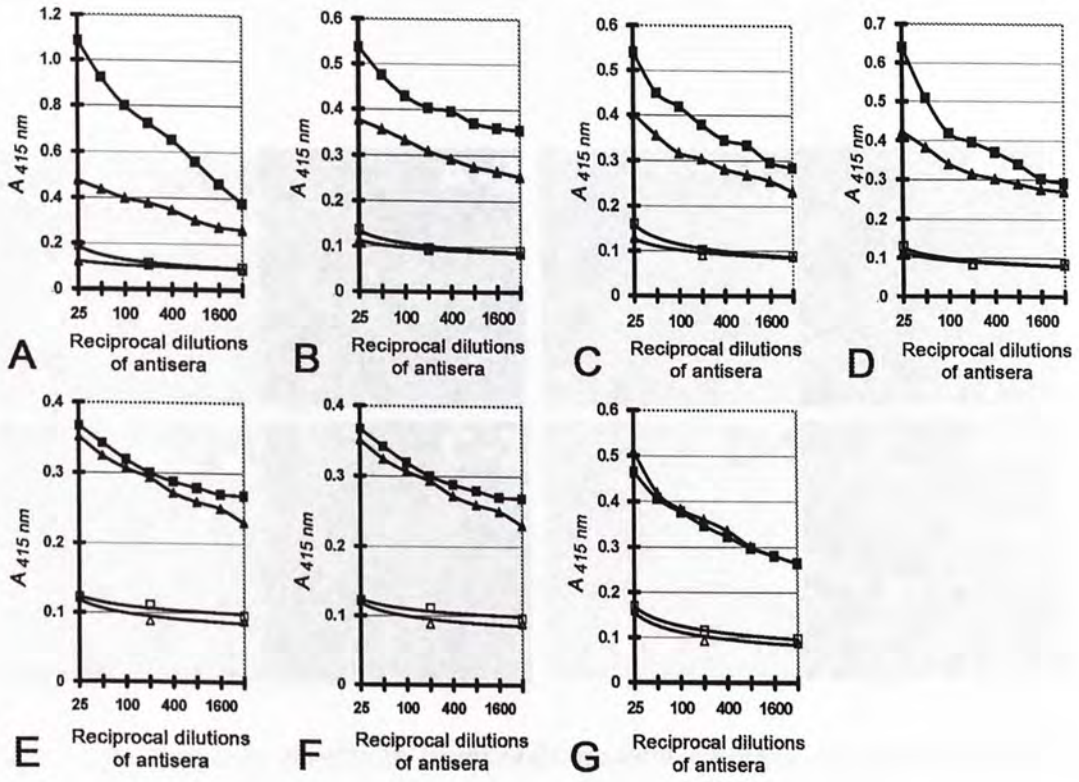

Figure 3.7: Titration of monkey serum samples diluted from 1:25 in a twofold series against constant antigen concentration of 10 mg/L.

Anti-peptide antiserum is indicated by ▲ with the corresponding preimmune serum, indicated by △ , as negative control. Anti KLH-conjugated peptide antiserum is indicated by ■ , and the corresponding preimmune serum is indicated by □ as negative control. **A:** M_S1 antiserum (▲) and M_S1_KLH antiserum (■); **B:** M_S2 antiserum (▲) and M_S2_KLH antiserum (■); **:**C: M_S3 antiserum (▲) and M_S3_KLH antiserum (■); **D:** M_S4 antiserum (▲) and M_S4_KLH antiserum (■); **E:** M_S5 antiserum (▲) and M_S5_KLH antiserum (■); **F:** M_S6 antiserum (▲) and M_S6_KLH antiserum (■); **G:** M_MIX antiserum (▲) and M_MIX_KLH antiserum (■).

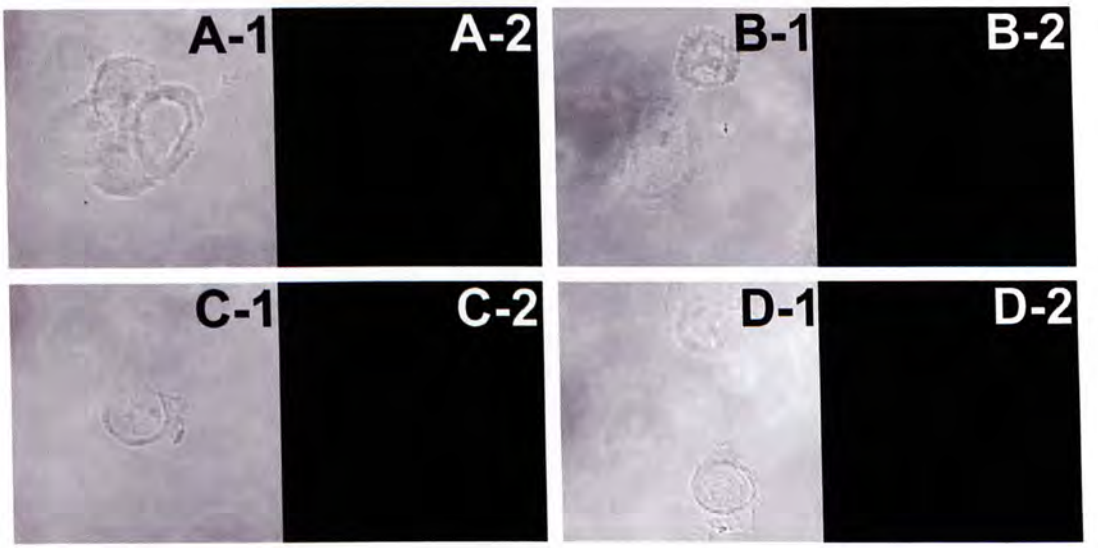

Figure 3.8: Negative results in immunofluorescent confocal microscopy are indicated by the absence of green fluorescent signals.

A: negative control using rabbit R_S1 preimmune serum, observed under light microscope **(A-1)** and confocal microscope **(A-2)**. Similar results were observed in all other rabbit preimmune sera. **B:** negative control using monkey M_S1 preimmune serum, observed under light microscope **(B-1)** and confocal microscope **(B-2)**. Similar results were observed in all other monkey preimmune sera. **C:** negative control using SARS patient serum on noninfected Vero cells, observed by light microscope **C-1** and confocal microscope **(C-2)**. **D:** negative result in confocal microscopy was shown for R_S1 antiserum, observed under light microscope **(D-1)** and confocal microscope **(D-2)**. Similar results were observed in other antisera with negative results in confocal microscopy.

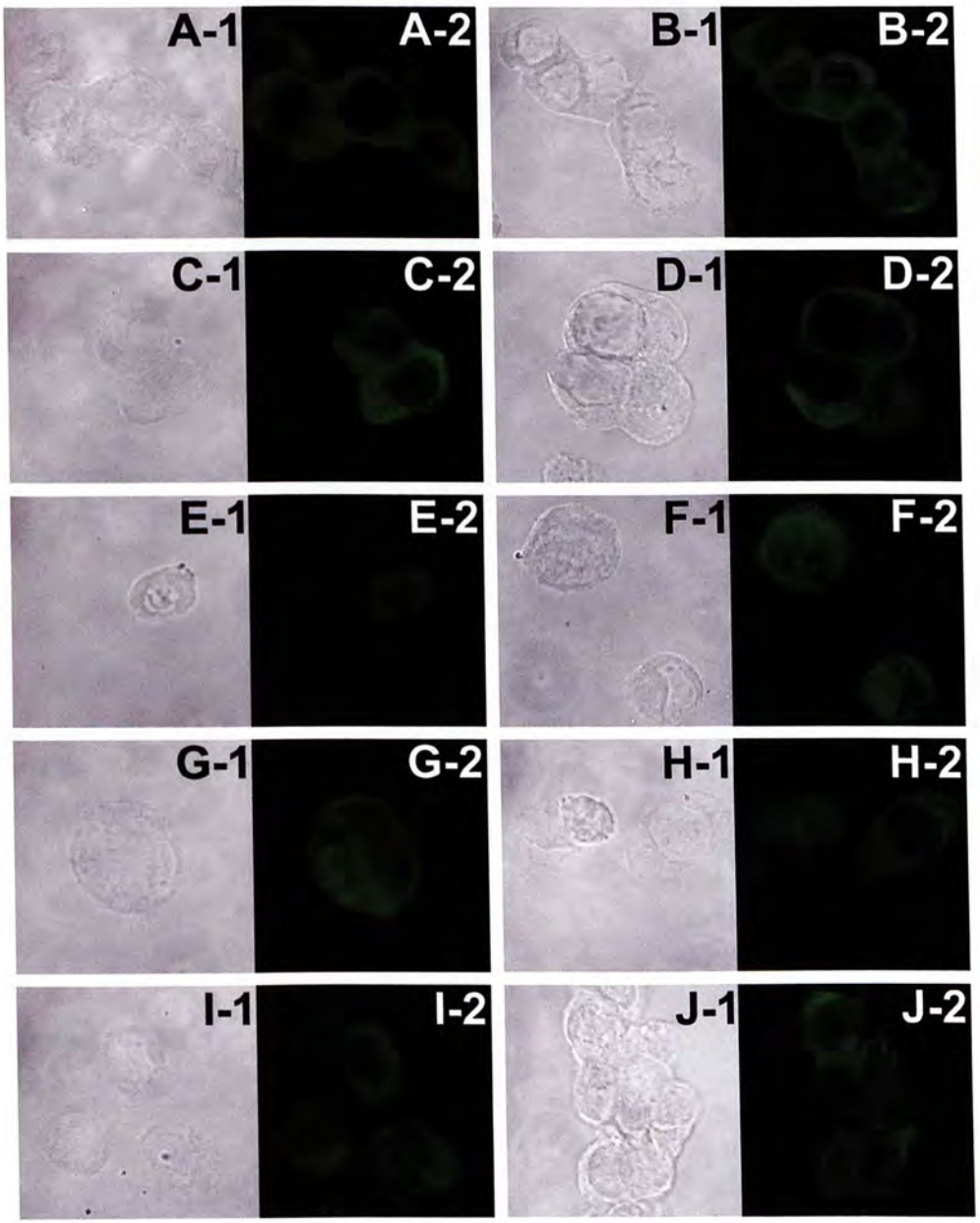

Figure 3.9: Positive results in immunofluorescent confocal microscopy are indicated by the presence of green fluorescent signals in the cytoplasm of the African green monkey kidney Vero cells.

A: positive control using SARS patient serum. Positive results in confocal microscopy were shown for the following rabbit and monkey antiserum samples: R_S2_KLH antiserum **(B)**; R_S5 antiserum **(C)**; R_S5_KLH antiserum **(D)**; R_S6_KLH antiserum **(E)**; M_S3 antiserum **(F)**; M_S6 antiserum **(G)**; M_S6_KLH antiserum **(H)**; M_MIX antiserum **(I)**; and M_MIX_KLH antiserum **(J)**. Each sample was observed under light microscope **(1)** and confocal microscope **(2)**, respectively.

Table 3.1: Results of immunofluorescent confocal microscopy.

Antigen used for immunization	Immunization in rabbits		Immunization in monkeys	
	Antiserum	Result of confocal microscopy	Antiserum	Result of confocal microscopy
S1 peptide	R_S1	-	M_S1	-
S1-KLH conjugate	R_S1_KLH	-	M_S1_KLH	-
S2 peptide	R_S2	-	M_S2	-
S2-KLH conjugate	R_S2_KLH	+	M_S2_KLH	-
S3 peptide	R_S3	-	M_S3	+
S3-KLH conjugate	R_S3_KLH	-	M_S3_KLH	-
S4 peptide	R_S4	-	M_S4	-
S4-KLH conjugate	R_S4_KLH	-	M_S4_KLH	-
S5 peptide	R_S5	+	M_S5	-
S5-KLH conjugate	R_S5_KLH	+	M_S5_KLH	-
S6 peptide	R_S6	-	M_S6	+
S6-KLH conjugate	R_S6_KLH	+	M_S6_KLH	+
Mix peptide[a]	R_MIX	-	M_MIX	+
Mix-KLH conjugate[a]	R_MIX_KLH	-	M_MIX_KLH	+

[a] Mix peptide and Mix-KLH conjugate were prepared by mixing the six synthetic peptides S1, S2, S3, S4, S5 and S6 in equal amounts.

For rabbit antisera, positive results were observed in samples R_S2_KLH, R_S5_KLH, and R_S6_KLH, which were elicited by the antigens: S2-KLH conjugate, S5-KLH conjugate and S6-KLH conjugate respectively. For monkey antisera, positive results were observed in M_S3, M_S6, M_S6_KLH, M_MIX, and M_MIX_KLH, which were elicited by the antigens: S3 conjugate-free peptide, S6 conjugate free peptide, S6-KLH conjugate, MIX conjugate-free peptide and MIX-KLH conjugate, respectively. Non-infected African green monkey kidney Vero cells and the preimmune sera of both the rabbits and monkeys were used as negative controls in this assay (Figure 3.8). The positive results indicated that most of the SARS-CoV was localized abundantly in the cytoplasm of the infected cells, which further confirmed the replication of SARS-CoV in the cytoplasm of animal host cells (Figure 3.9).

3.2 Synthetic peptide studies on SARS-CoV 3CL^{pro}

3.2.1 Substrate specificity of SARS-CoV 3CL^{pro}

In this study, we employed the MALDI-TOF mass spectrometric analysis in combination with the 'cartridge replacement' solid-phase peptide synthesis approach to examine the biological significance of amino acid residues in total six target positions at the SARS-CoV 3CL^{pro} cleavage sites, including the P2, P3, P4 positions at the amino side of P1 position; and the P1', P2', P3' positions at the carboxyl side of P1 position (Table 2.2).

Recombinant SARS-CoV 3CL^{pro} was expressed in the *E. coli.* expression system, and the purity of the protein samples were analyzed by SDS-PAGE. As shown in Figure 3.10, the GST tag-removed SARS-CoV 3CL^{pro} recombinant protein was prepared with high purity, which was then subjected to the *in vitro* enzymatic cleavage assay with the peptide substrates. The cleavage results were monitored by MALDI-TOF mass spectrometric analysis.

The PS01 peptide, with its amino acid sequence based on one of the SARS-CoV 3CL^{pro} cleavage sites on the SARS-CoV BJ01 polyprotein pp1ab (residues 3232-3247), was used as the positive control in the enzymatic cleavage assay to assess the activity of the recombinant SARS-CoV 3CL^{pro}. The original PS01 peptide peak with m/z value of 1739.03 resolved before the assay, was absent from

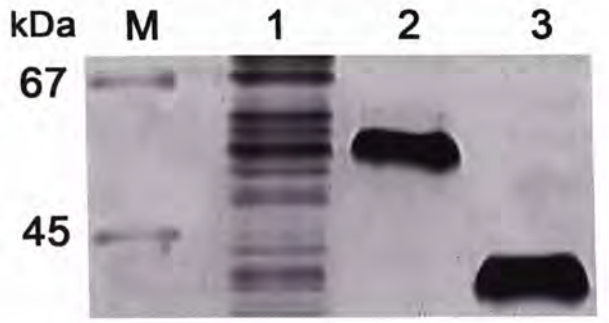

Figure 3.10: SDS-PAGE analysis of SARS-CoV 3CL^{pro} recombinant proteins.

Lane 1: soluble fraction of *E. coli.* transformed with pGEX-3CL^{pro} plasmid. **Lane 2:** purified GST-3CL^{pro} fusion protein after GST affinity chromatography (62.26 kDa). **Lane 3:** purified and concentrated SARS-CoV 3CL^{pro} recombinant protein without GST tag (33.85 kDa). All of the above samples were directly subjected to 12% SDS-PAGE with the size control by the protein molecular weight marker (**Lane M**).

the mass spectrum after the reaction (Figure 3.11), as it was cleaved by SARS-CoV 3CL^{pro}into smaller peptide fragments; which validated the enzymatic activity of the recombinant SARS-CoV 3CL^{pro} that we prepared. The results of the *in vitro* enzymatic cleavage assay of the six batches of synthetic peptide substrates (PS02, PS03, PS04, PS05, PS06 and PS07) with amino acid substitutions in their corresponding target positions are summarized in Table 3.2.

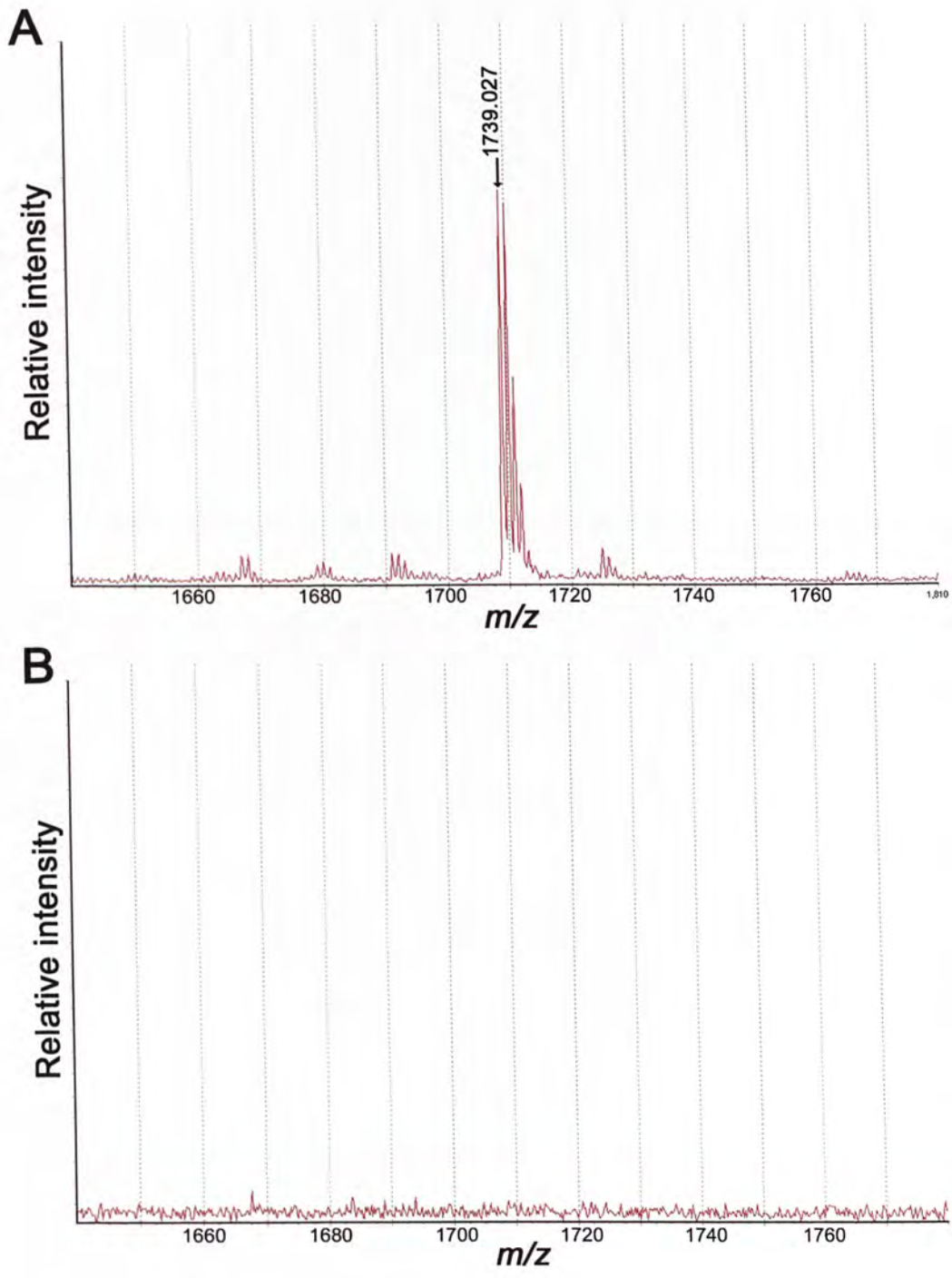

Figure 3.11: Cleavage result of PS01 control peptide.

PS01 control peptide peak (*) was resolved on the mass spectrum before the SARS-CoV 3CL^{pro} cleavage assay **(A)**; and was absent from the mass spectrum after the assay **(B)**.

Table 3.2: Results of the *in vitro* peptide cleavage assay.

Peptide substrate	Target Position[a]	Corresponding amino acid[b]	Amino acid substitution[c]	
			Cleaved peptides	Uncleaved peptides
PS02	P2	Leu	Ile/Leu, Phe	others[d]
PS03	P3	Val	others[e]	Pro
PS04	P4	Ala	all[f]	---
PS05	P1'	Ser	others[e]	Pro, Asp, Glu
PS06	P2'	Gly	others[e]	Pro, Ile/Leu
PS07	P3'	Phe	all[f]	---

[a] The position (P) on the substrate is counted from the point of cleavage, i.e. P1 is the position just before the cleavage site ↓, with P1, P2, ... etc. located at N-terminal to the cleavage site; and P1', P2', ... etc. located at C-terminal to the cleavage site (Berger and Schechter, 1970).

[b] The original amino acid residue in the position under investigation before the substitution with 20 standard amino acids.

[c] The identity of the amino acid in the particular peptide that substituted the residue at the position under investigation.

[d] All peptides other than those in the opposite column were not cleaved by SARS-CoV $3CL^{pro}$.

[e] All peptides other than those in the opposite column were cleaved by SARS-CoV $3CL^{pro}$.

[f] All the 20 peptides were cleaved by SARS-CoV $3CL^{pro}$.

Before the enzymatic assay, the corresponding peaks of the 20 peptides in each substrate pool were all resolved on the mass spectra 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 mass spectrometer to discriminate peptide species with single amino acid difference.

P2, P3, and P4 positions at the amino side of the SARS-CoV 3CL^{pro} cleavage site

For PS02 peptide substrates, only the original peptide (m/z value of 1739.04) and the peptide with Phe substitution for Leu in P2 position (m/z value of 1773.05) were cleaved, with their corresponding peptide peaks absence from the mass spectrum after the enzymatic assay; whereas all other peptides were not cleaved, with their corresponding peaks remained (Figure 3.12). For PS03 peptide substrates, only the peptide with Pro substitution for Val in P3 position (m/z value of 1737.01) was not cleaved, with its corresponding peptide peak remained after the enzymatic assay; whereas all other peptides were cleaved (Figure 3.13). For PS04 peptide substrates with amino acid substitutions in P4 position, all the 20 peptides were cleaved completely, with their corresponding peaks absence from the mass spectrum after the enzymatic assay (Figure 3.14).

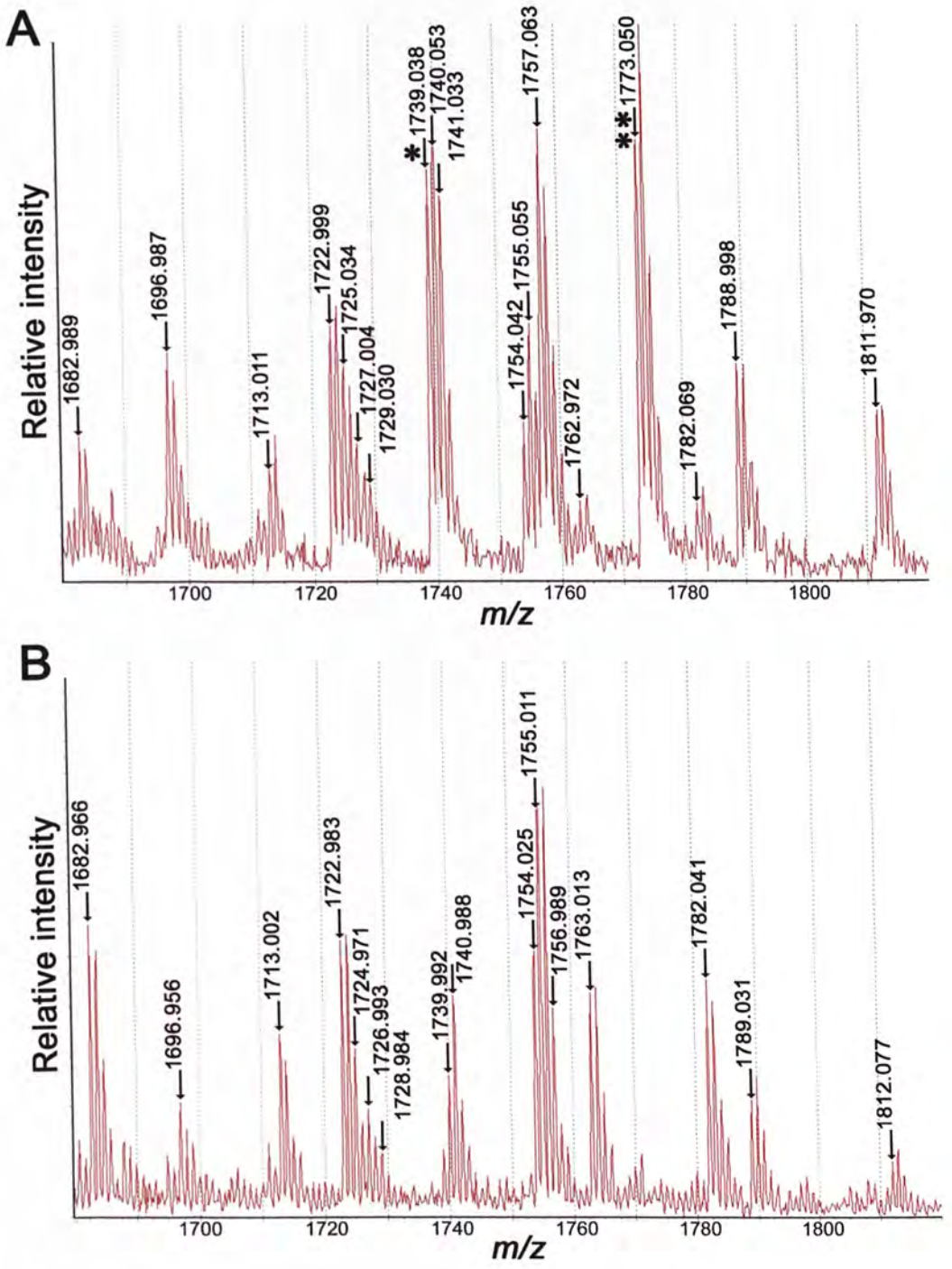

Figure 3.12: Cleavage result of PS02 peptide substrates.

PS02 peptide substrates comprised a mixture of 20 different peptides with single amino acid alteration in P2 position (Leu) were subjected to SARS-CoV $3CL^{pro}$ cleavage assay. All 20 peptides were resolved on the mass spectrum before the assay **(A)**; and only the peptide peaks corresponding to Ile/Leu **(*)** and Phe **(**)** substitutions in P2 position were absent, whereas all other peaks remained after the assay **(B)**.

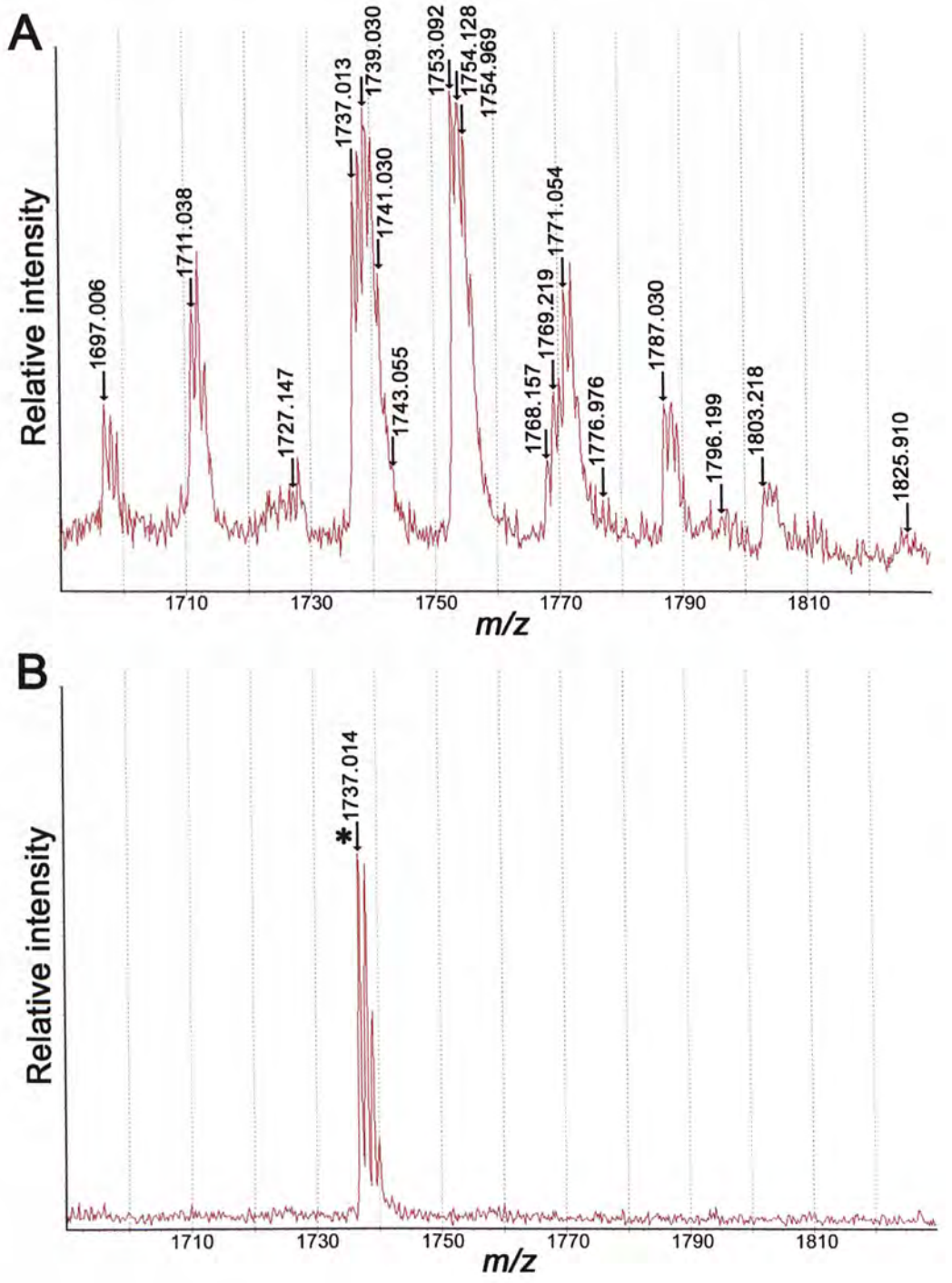

Figure 3.13: Cleavage result of PS03 peptide substrates.

PS03 peptide substrates comprised a mixture of 20 different peptides with single amino acid alteration in P3 position (Val) were subjected to SARS-CoV 3CL^{pro} cleavage assay. All 20 peptides were resolved on the mass spectrum before the assay **(A)**; and only the peptide peak with Pro **(*)** substitution in P3 position remained after the assay **(B)**.

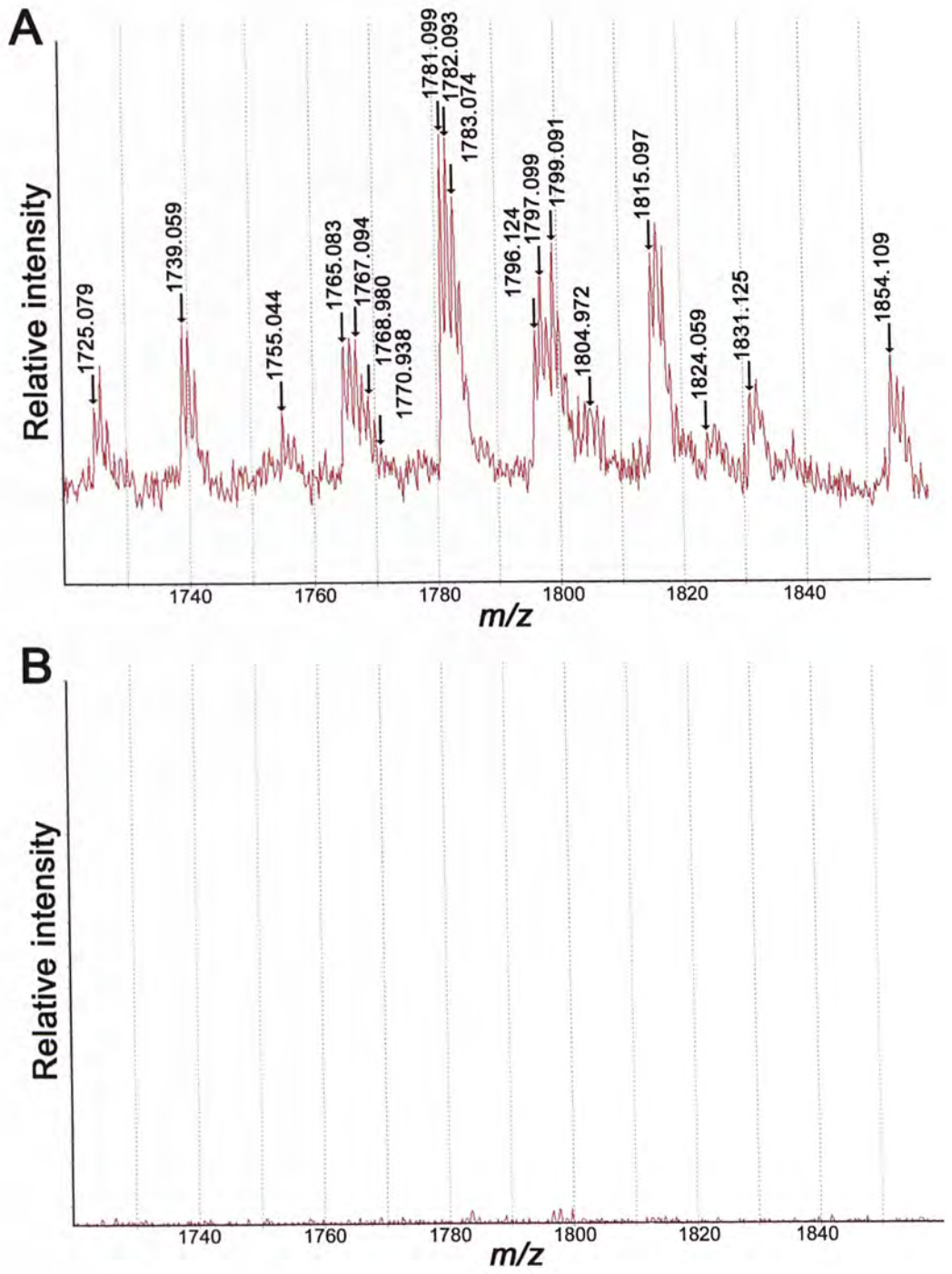

Figure 3.14: Cleavage result of PS04 peptide substrates.

PS04 peptide substrates comprised a mixture of 20 different peptides with single amino acid alteration in P4 position (Ala) were subjected to SARS-CoV $3CL^{pro}$ cleavage assay. All 20 peptides were resolved on the mass spectrum before the assay **(A)**; which were all absent from the mass spectrum after the assay **(B)**.

P1', P2', and P3' positions at the carboxyl side of the SARS-CoV 3CL^{pro} cleavage site

For PS05 peptide substrates, only three peptides with Pro substitution (m/z value of 1749.05), Asp substitution (m/z value of 1767.05) and Glu substitution (m/z value of 1781.04) for Ser in P1' position respectively were not cleaved (Figure 3.15). 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 P2' position respectively were not cleaved (Figure 3.16). For PS07 peptide substrates with amino acid substitutions in P3' position, all the 20 peptides were cleaved completely, with their corresponding peaks absence from the mass spectrum after the enzymatic assay (Figure 3.17).

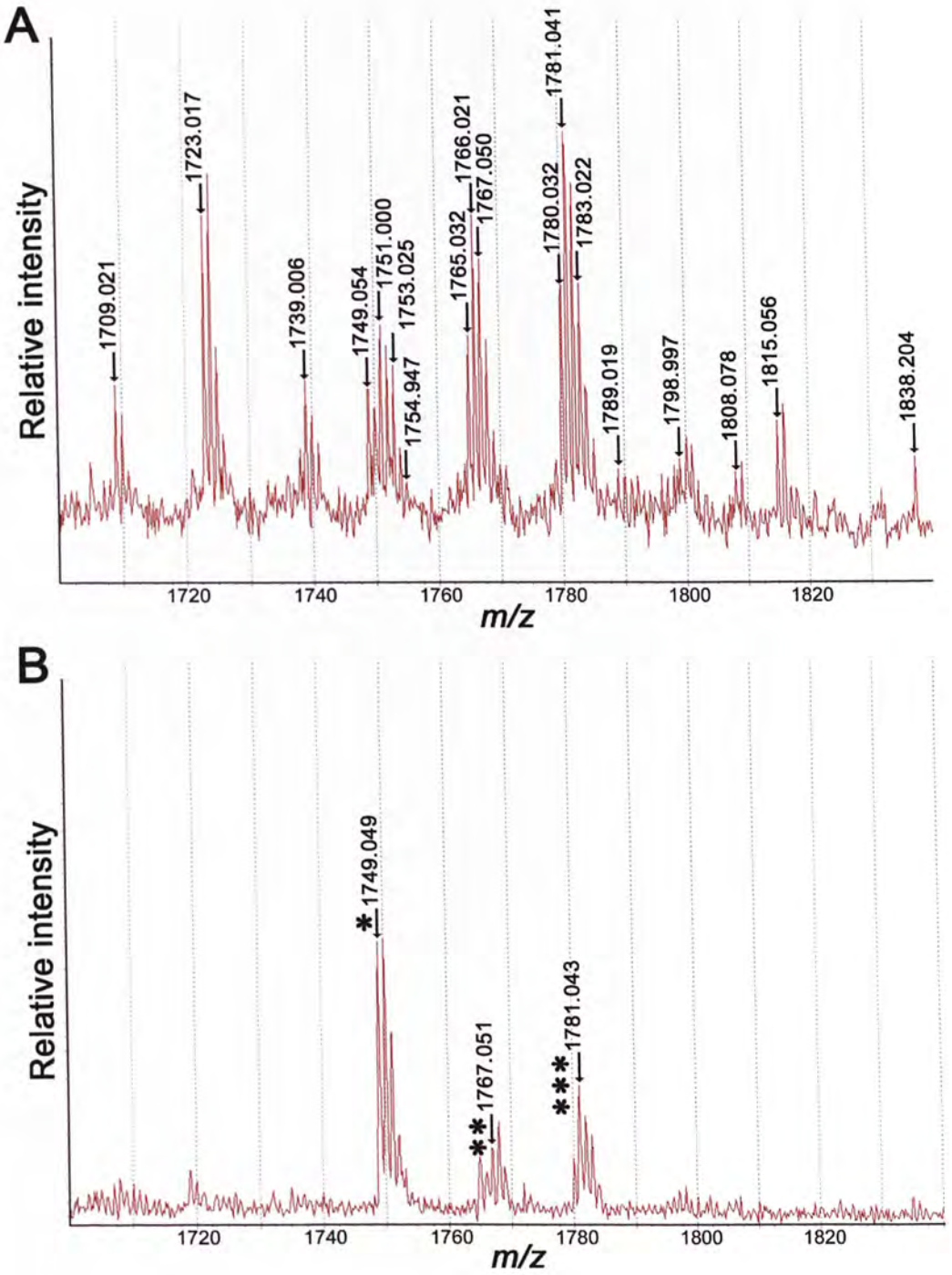

Figure 3.15: Cleavage result of PS05 peptide substrates.

PS05 peptide substrates comprised a mixture of 20 different peptides with single amino acid alteration in P1' position (Ser) were subjected to SARS-CoV $3CL^{pro}$ cleavage assay. All 20 peptides were resolved on the mass spectrum before the assay **(A)**; and the peptide peaks with Pro **(*)**, Asp **(**)** and Glu **(***)** in P1' position remained after the assay **(B)**.

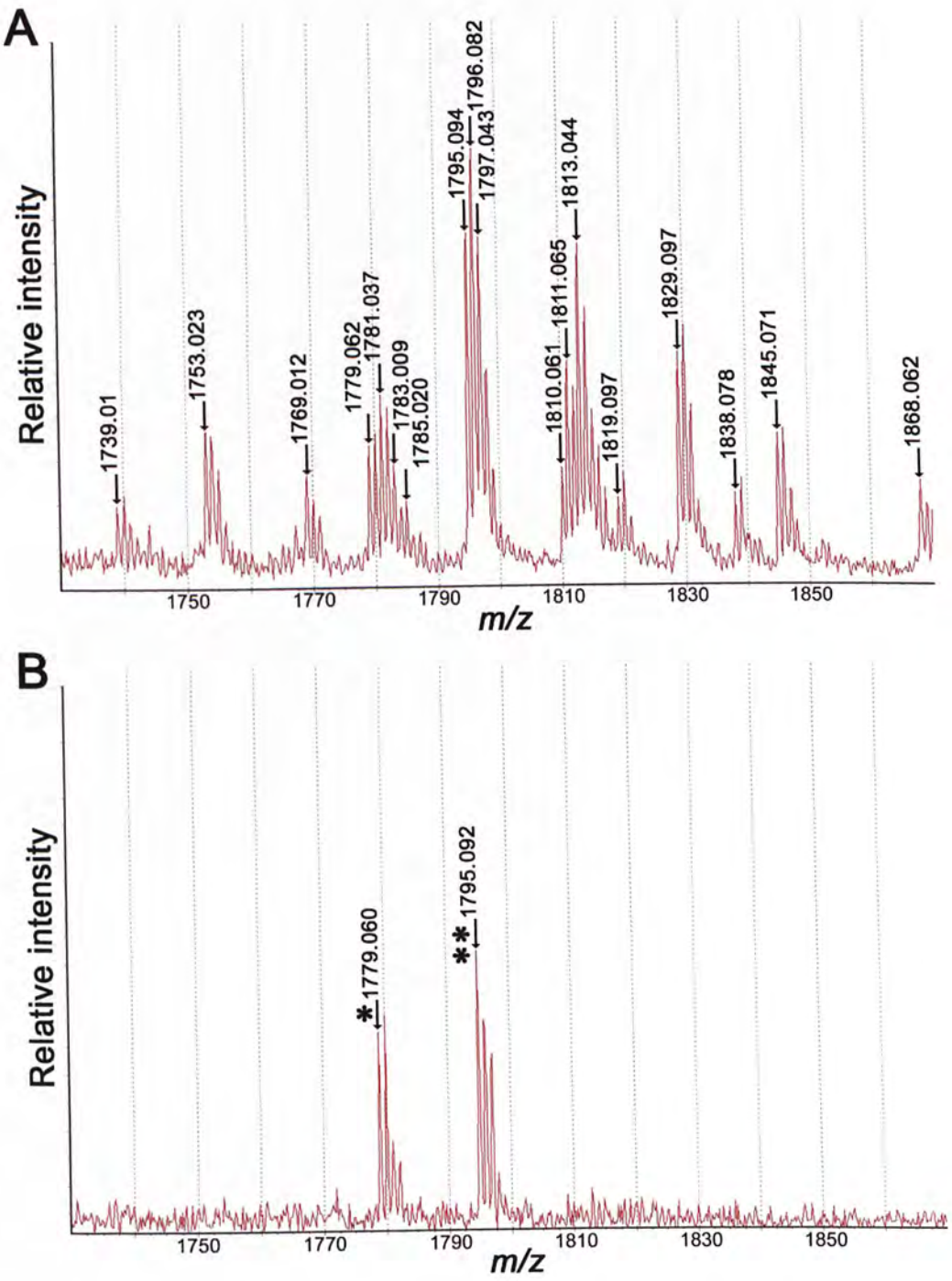

Figure 3.16: Cleavage result of PS06 peptide substrates.

PS06 peptide substrates comprised a mixture of 20 different peptides with single amino acid alteration in P2' position (Gly) were subjected to SARS-CoV 3CL^{pro} cleavage assay. All 20 peptide peaks were resolved on the mass spectrum before the assay **(A)**; and the peptide peaks with Pro **(*)** and Ile/Leu **(**)** in P2' position remained after the assay **(B)**.

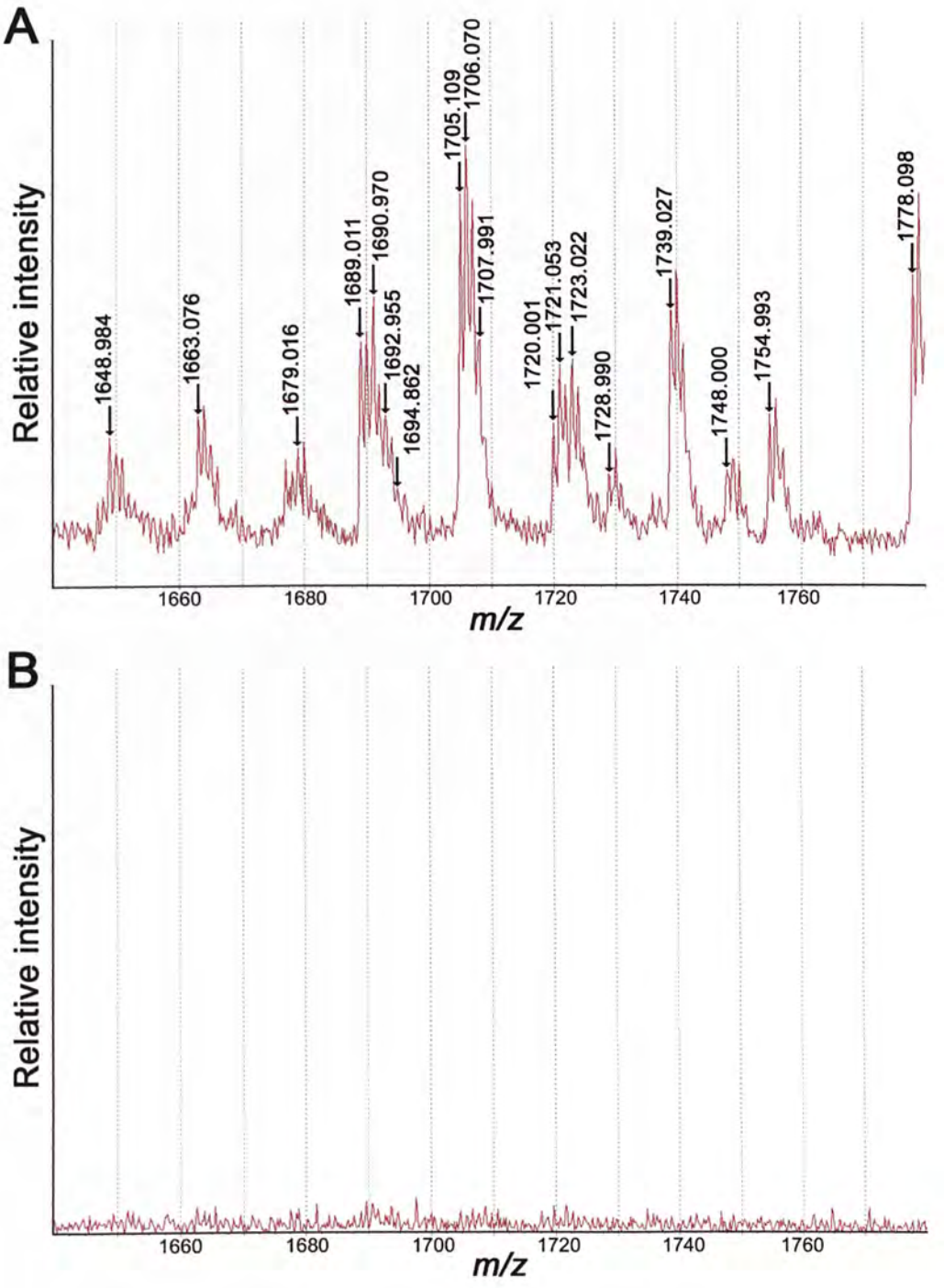

Figure 3.17: Cleavage result of PS07 peptide substrates.

PS07 peptide substrates comprised a mixture of 20 different peptides with single amino acid alteration in P3' position (Phe) were subjected to SARS-CoV $3CL^{pro}$ cleavage assay. All 20 peptides were resolved on the mass spectrum before the assay **(A)**; which were all absent from the mass spectrum after the assay **(B)**.

3.2.2 Molecular docking of SARS-CoV 3CL^{pro} and peptide substrates

Based on the mass spectrometry results from the *in vitro* enzymatic cleavage assay (Table 3.2), four target positions including P2, P3, P1' and P2' positions located at the SARS-CoV 3CL^{pro} cleavage sites were further analyzed by the Insight II molecular modeling platform, and the molecular docking results are shown in Figures 3.18 and 3.19. The control peptide, PS01, was docked to the predefined active site of SARS-CoV 3CL^{pro}, with the conserved Gln residue located at close proximity to Cys145-His41 residues at the catalytic dyad of SARS-CoV 3CL^{pro} (Figure 3.18); in which this energy data generated from energy minimization was used as the reference for comparing with the docking results using other peptide substrates.

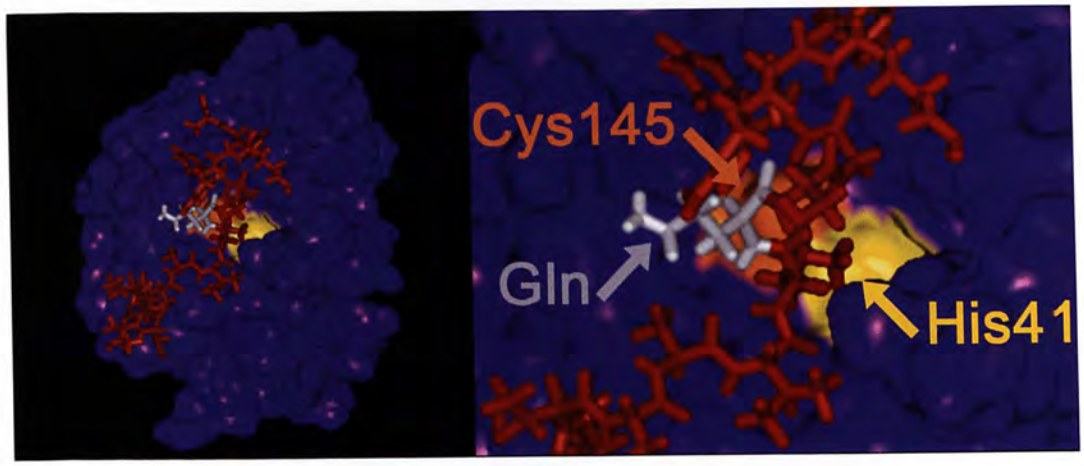

Figure 3.18: Interaction between SARS-CoV $3CL^{pro}$ and PS01 control peptide as predicted by molecular docking.

Overview **(left)** of the interaction between the SARS-CoV $3CL^{pro}$ *(purple)* and peptide substrate PS01 *(red)*; and the zoom-in view **(right)** of the conserved Gln residue *(gray)* in P1 position at the cleavage site of PS01 control peptide being docked into the catalytic dyad of SARS-CoV $3CL^{pro}$ which is composed of the Cys145 *(orange)* and His41 *(yellow)* residues.

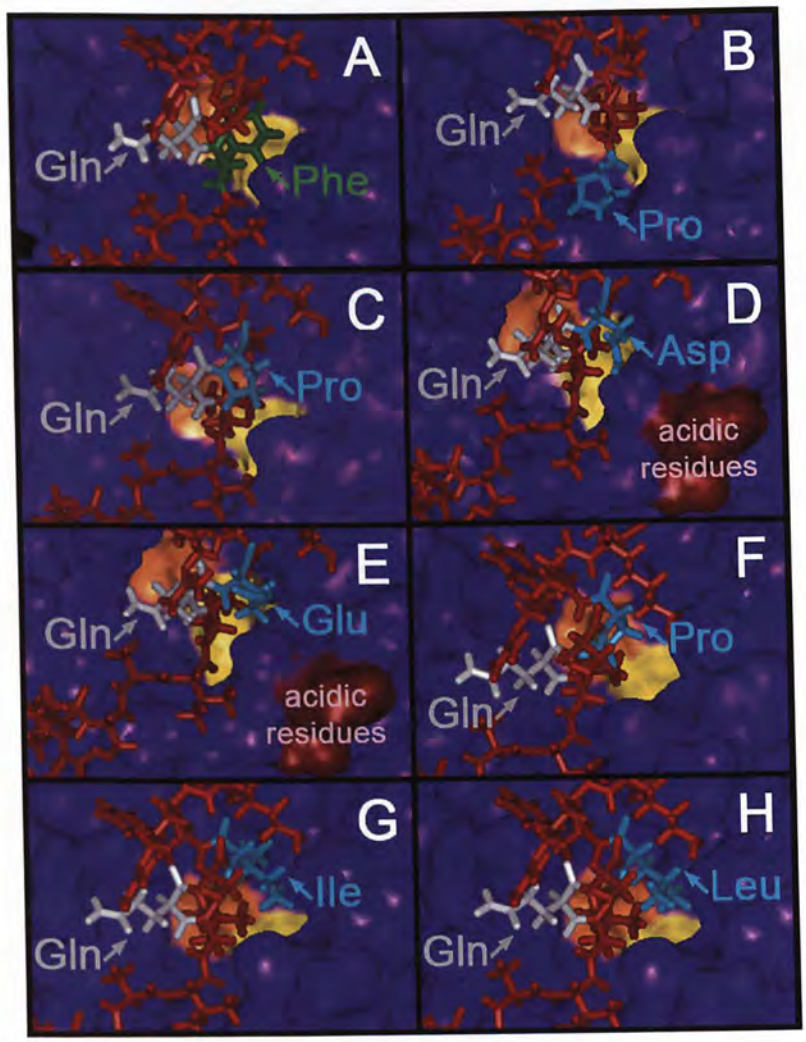

Figure 3.19: Comparison between the molecular models of different selected peptide substrates docked to the active site of SARS-CoV $3CL^{pro}$.

The Cys145 *(orange)* and His41 **(yellow)** residues in the catalytic dyad of SARS-CoV $3CL^{pro}$ **(purple)** and the conserved Gln residue **(gray)** in P1 position at the cleavage site of the peptide substrate (red) were shown. **A:** PS02 peptide with Phe *(green)* in P2 position was in close proximity to the active site of SARS-CoV $3CL^{pro}$. Peptide substrates which were unfavorable for SARS-CoV $3CL^{pro}$ cleavage were deviated away from the active site of SARS-CoV $3CL^{pro}$, which included: **B:** PS03 peptide with Pro *(blue)* in P3 position; **C:** PS05 peptide with Pro *(blue)* in P1' position; **D:** PS05 peptide with Asp *(blue)* in P1' position that was in close proximity to the acidic residues *(brown)* at the active site; **E:** PS05 peptide with Glu *(blue)* in P1' position that was in close proximity to the acidic residues *(brown)* at the active site; **F:** PS06 peptide with Pro *(blue)* in P2' position; **G:** PS06 peptide with Ile *(blue)* in P2' position; **H:** PS06 peptide with Leu *(blue)* in P2' position.

For PS02 peptide with Phe substitution in P2 position, the energy level was similar to that of the PS01 control peptide, which further confirmed the tolerance of Phe substitution at the P2 position at the cleavage site (Figure 3.19). On the other hand, the docking results for PS03 peptide with Pro substitution in P3 position, PS05 peptides with Pro, Asp and Glu substitutions in P1' position, and PS06 peptides with Pro, Ile, and Leu substitutions in P2' position, showed relatively high energy levels during docking studies when compared to the PS01 control, revealed the unfavorable cleavage reactions of these peptides by SARS-CoV $3CL^{pro}$ (Figure 3.19).

♦ **End of chapter.**

Chapter 4

Discussion

4.1 Synthetic peptide studies on SARS-CoV S glycoprotein

4.1.1 Synthetic peptides elicited SARS-CoV specific antibodies

In this study, we designed a total of six peptide sequences, S1-S6 (Table 2.1), that corresponded to the surface regions of SARS-CoV S glycoprotein based on the combined bioinformatics data from protein analyses. These six synthetic peptides were used to immunize both rabbits and monkeys. The results of immunofluorescent confocal microscopy indicated that four of the six synthetic peptides, i.e., S2, S3, S5, and S6, raised antibodies that could specifically recognize SARS-CoV, which suggested the efficacy of using synthetic peptides as antigens to elicit SARS-CoV specific immune responses (Figure 3.1).

In a parallel study carried out at Tsinghua University using recombinant protein approaches, similar results were obtained. Pang's team[1] at Tsinghua University has generated a series of truncated recombinant protein analogs (SG3, analog of residues 73-882; S10, analog of residues 730-1150; and S22, analog of residues 894-1192) of the SARS-CoV S glycoprotein to characterize the structural and functional properties of the S glycoprotein. Their immunoassays indicated that the SG3, S10, and S22 analogs of the S glycoprotein were recognized by the IgG in sera from recovered SARS patients. Their SG3 analog is suggested to be the most specific immunogenic region on the S glycoprotein. Our work clearly demonstrated that the immunogenic region of their SG3 analog (residues 735-882) could be further delineated to residues 788-820, which corresponds to our S5 peptide (Table 2.1). In addition, our S6 peptide (residues 1002-1030; Table 2.1) could be another minimum and significant immunogenic region of their S10 and S22 analogs of S glycoprotein. In addition, free S6 peptide is capable of inducing specific antibodies in monkeys.

[1]Pang H. 'Initial postgenomic studies of SARS coronavirus'. Presented at Advances in Protein Sciences, The Croucher Foundation Advanced Study Institute, December 15-20, 2003, Hong Kong.

4.1.2 Factors affecting the specificity and antigenicity of synthetic peptides

There are several important factors contributing to the ability of the four synthetic peptides, i.e., S2, S3, S5, and S6, to elicit the production of SARS-CoV-specific antibodies.

Strategies in peptide design

Peptide design based on protein sequence is a critical and formidable task in this study. Choices of peptide sequences were deliberately targeted to the potential surface region of the SARS-CoV S glycoprotein as predicted by *in silico* combined protein sequence analyses (Figure 3.3). Based on the assumption that the surface region of protein being exposed to the solvent is likely to be hydrophilic, we located the relatively hydrophilic stretches of the SARS-CoV S glycoprotein sequence using the Hopp and Woods algorithm (Hopp and Woods, 1981) as potential targets for peptide design. Secondary structure is another major consideration in the design of biologically active peptides. The three dimensional conformation assumed by the synthetic peptide in solvated state explicitly determines both the specificity and antigenicity of the peptide to elicit immune responses. The peptide backbone and polar side chains tend to exhibit a driving force to be solvated, in which the length of 20 to 30

amino acid residues is enough for a peptide to adopt a stable helical structure in solution. Chou and Fasman algorithm was used to predict possible helical stretches of the SARS-CoV S glycoprotein sequence (Chou and Fasman, 1978). Most of the designed peptides were expected to adopt an α-helix motif to provide rigid and stable structures to enhance antigenicity (Figure 3.5).

From bioinformatics analyses, it is noticed that the SARS-CoV S glycoprotein is heavily glycosylated and phosphorylated (Figures 3.1 and 3.2). These regions in the SARS-CoV S glycoprotein sequence with possible posttranslational modifications are undesirable choices for peptide design, because the antigenicity at these regions is merely contributed by the non-specific sugar or phosphate moieties rather than the more characteristic structural conformation alone. To generate peptide antigens with unique structural conformations to elicit specific immune responses, we deliberately avoided these potentially glycosylated or phosphorylated regions in the peptide design to ensure specificity and efficacy of our peptide antigens.

To further enhance the antigenicity of these peptides, later work may include introducing several conformational constraints to these peptides to increase structural rigidity. Local conformation constraints can be introduced to restrict the mobility of single amino acid residue in the peptide, such as by replacing a hydrogen adjacent to a rotatable bond with a methyl group. Cyclization of linear

peptides can be performed by 'head to tail' formation of amide bond between the amino (N) and carboxyl (C) terminals, disulfide bridge formation between two Cys residues, or lactam bridge formation between Asp/Glu and Lys residues. Besides, from the *in silico* energy minimization studies of our peptides, the predicted loop size carried on the peptide may reflect its antigenicity; thus, the loop structures of these peptides could be further adjusted and strengthened to optimize the corresponding antigenicity (Figure 3.5).

Mimicking the charge states in native protein

Since our designed peptides are derived from the internal sequence of SARS-CoV S glycoprotein, the presence of charges should be masked in order to mimic the charge states of the alpha amino and carboxyl groups in the native protein. Thus, we performed acetylation at the amino (N) terminal and amidation at the carboxyl (C) terminal during solid-phase peptide synthesis, so that the peptides can resemble the surface region of the native S glycoprotein with a higher fidelity.

Peptide conjugation to carrier protein

Due to the relatively small sizes of synthetic peptides, these peptides alone may not be antigenic enough to elicit host immune response. Therefore, it is suggested to conjugate these synthetic peptides to

a suitable carrier protein to enhance both the antigenicity and stability. Keyhole limpet hemocyanin (KLH) was used as the carrier protein for conjugation to the synthetic peptides in this study. KLH is a large copper-containing protein that is isolated from the mollusc *Megathura crenulata* with no homology with vertebrate proteins. Other common carrier proteins include bovine serum albumin (BSA), rabbit serum albumin (RSA), and bovine thyroglobulin. Our results suggested that the KLH-conjugated peptides generally elicit better and more specific production of antibodies that recognize SARS-CoV than conjugate-free peptides.

4.1.3 Next step towards vaccine development

On the basis of our studies, which demonstrated that four of the synthetic peptides, i.e., S2, S3, S5, and S6, could elicit the production of SARS-CoV-specific antibodies, future experiments will be carried out to address the protective nature of these antibody responses in monkeys. This can be done by vaccinating noninfected monkeys with these synthetic peptides. After exposure of these vaccinated monkeys to live SARS-CoV, their health status could be monitored to determine the efficacy of these synthetic peptides as potential SARS vaccines. The monkey strain *Macaca fascicularis* would be a good choice for this animal test because this strain has previously been demonstrated as a potent host of SARS-CoV that developed

SARS symptoms after SARS-CoV infection (Fouchier *et al.*, 2003). Because of the high infectiousness of SARS-CoV, a biosafety P3/P4 level laboratory is required whenever performing any experiment using live SARS-CoV to prevent unpredictable outbreaks of SARS.

4.1.4 A synthetic peptide-based approach

The development of safe and effective vaccines is necessary for sustained control of SARS. To date, vaccination against SARS is not available. Despite the efforts in SARS vaccine development are underway globally, only the inactivated SARS-CoV vaccines are tested in clinical trial in China (He *et al.*, 2004; Marshall and Enserink, 2004; Zhou *et al.*, 2005). Nonetheless, the major drawbacks of this inactivated SARS-CoV vaccines include the risk of incomplete inactivation of the vaccine virus, the induction of harmful immune hyperactivity and inflammatory responses after the immunization, and also the unpredictable laboratory accidents during vaccine preparation. In this study, we focused on using synthetic peptides for viral protein neutralization. Some of the major advantages of synthetic peptide vaccines are that they are relatively non-toxic and chemically defined, and there would not be the presence of infectious agents or foreign protein, thus they demonstrate absolute biosafety in both manufacturing and use. Moreover, they can be synthesized

in high yields with high purity and they are stable at ambient temperature.

With the advance in peptide chemistry, a synthetic peptide-based approach could be a rapid initial step for providing useful information for SARS research. Synthetic peptides have high flexibility in responding and confronting to newly emerging subtypes, in which new peptide antigens could be synthesized rapidly once the coding sequence of that particular subtype is known, without the need of generating the full-length protein. The results of synthetic peptide studies enable preliminary information about the possible antigenic determining sites on the native viral protein to be generated, which provide insights for generating a recombinant protein analog that could better mimic the conformation of the epitope regions. Moreover, the biophysical properties and binding kinetics of these synthetic peptide candidates could be systematically characterized to allow elucidation of the structural and functional relationships of different SARS-CoV proteins. Our experimental data are informative and could advance the understanding of SARS.

4.2 Synthetic peptide studies on SARS-CoV 3CL^{pro}

In this study, we have introduced a rapid and high throughput screening method to generate a comprehensive overview of the substrate specificity preferences of SARS-CoV 3CL^{pro}, via a combinatory strategy that employs MALDI-TOF mass spectrometry and 'cartridge replacement' solid-phase peptide synthesis. Research on the substrate specificity preferences of SARS-CoV 3CL^{pro} is gaining increasing attention, owing to the biological significance of SARS-CoV 3CL^{pro} in the viral life cycle and the feasibility to design specific SARS-CoV 3CL^{pro} inhibitors as potential anti-SARS drugs.

It has been demonstrated that SARS-CoV 3CL^{pro} cleaves the coronavirus replicase polyproteins at no less than 11 conserved sites, involving the Leu-Gln ↓ sequence in P2 and P1 positions at the cleavage site, respectively (Ziebuhr *et al.*, 2000; Hegyi and Ziebuhr, 2002; Anand *et al.*, 2003). To further extend our knowledge on the substrate specificity of SARS-CoV 3CL^{pro}, we aimed to comprehensively study the biological significance 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^{pro} cleavage site by a rapid and high-throughput approach (Table 2.2).

4.2.1 A comprehensive overview of the substrate specificity of SARS-CoV 3CL^{pro}

P2 position

In addition to the well-known conserved Gln residue in P1 position at the SARS-CoV 3CL^{pro} cleavage site, P2 position which is exclusively occupied by Leu residue, also serves as another important determinant of substrate specificity. Since it is well known that Leu residue predominantly occupies the P2 position (Fan *et al.*, 2004; Fan *et al.*, 2005), the absent peak of Ile/Leu (*) after the assay (Figure 3.12) is suggested to be the control peptide having the original Leu in P2 position. The exact discrimination of the ambiguous Ile and Leu residues yet requires further experiments involving the use of 2 different peptides in separate enzymatic cleavage assays.

Our peptide cleavage results indicated that the peptide substrate with Phe replacement in P2 position was favorable for SARS-CoV 3CL^{pro} cleavage (Figure 3.12), with similar energy level in molecular docking when compared to the positive control peptide with Leu in P2 position. Our results demonstrated consistency with other findings (Fan *et al.*, 2004; Fan *et al.*, 2005), suggesting that P2 position serves as an important hydrophobic pocket for SARS-CoV 3CL^{pro} specific binding; thus the peptide substrate with replacement of the original Leu with the more hydrophobic Phe in P2 position was still

susceptible to SARS-CoV 3CL^{pro} cleavage. Other studies demonstrated that SARS-CoV 3CL^{pro} favorably tolerates Phe and Val substitutions, and to a lesser extent for Met substitution in the P2 position (Fan *et al.*, 2004; Fan *et al.*, 2005); hence revealing the significance of the P2 hydrophobic pocket in the determination of SARS-CoV 3CL^{pro} substrate specificity.

P1' position

Furthermore, the P1' position which is frequently occupied by Ser residue, also contributes to the substrate specificity of SARS-CoV 3CL^{pro} considerably. Our peptide cleavage results revealed P1' position is highly unfavorable to the substitution by Pro, Asp and Glu residues (Figure 3.15), which is consistent with the relatively high energy data generated by molecular docking studies.

Our findings on the intolerance of Asp or Glu substitutions in P1' position suggests that the acidic residues at the binding interface of the SARS-CoV 3CL^{pro} active site are probably in close proximity and sterically complementary to P1' position of the substrate, leading to electrostatic repulsions between the negatively-charged acidic residues that hinder the substrate binding to SARS-CoV 3CL^{pro}. Our molecular docking results also demonstrated that the P1' position is proximal to the Glu47 and Asp48 residues at the SARS-CoV

3CL^{pro} active site (Figure 3.19), which further confirmed our interpretations. Besides, the unfavorable substitution by Pro in P1' position could be accounted by the steric hindrance arising from its cyclic side chain (Figure 3.19), which obstructs the substrate binding to SARS-CoV 3CL^{pro}. Previous studies also demonstrated that small aliphatic residues such as Ser, Ala and Gly are favored in P1' position (Fan *et al.*, 2004; Fan *et al.*, 2005).

P3, P4, P2' and P3' positions

Moreover, our results demonstrated the substrate specificity of SARS-CoV 3CL^{pro} are less dependent on the P2' and P3 positions at the cleavage site. The substitutions of Pro, Ile or Leu in the P2' position (Figure 3.16) and the substitution of Pro in the P3 position (Figure 3.13) in the peptide substrates are unfavorable for the SARS-CoV 3CL^{pro} cleavage, in which these bulky Pro residue and β-branching Ile and Leu residues may cause steric hindrance to binding of SARS-CoV 3CL^{pro} with its substrates (Figure 3.19). On the other hand, the peptide cleavage results showed that the P3' position (Figure 3.17) and P4 position (Figure 3.14) have no effects on determining the substrate specificity preferences of SARS-CoV 3CL^{pro}.

4.2.2 Sequence comparison between SARS-CoV 3CL^{pro} cleavage sites

As discussed in section 4.2.1, our interpretations on the SARS-CoV 3CL^{pro} substrate specificity can be summarized as follows: (1) Gln residue is a necessary requirement in P1 position at the conserved cleavage site; (2) P2 position is a hydrophobic pocket, that tolerates the replacement with hydrophobic amino acids; (3) P1' position is located proximal to the acidic residues at the SARS-CoV 3CL^{pro} active site, thus does not tolerate substitution of acidic amino acids; (4) Bulky Pro residue that leads to steric hindrance is not unfavorable for substitution in P3 and P2' positions, and β-branching Ile and Leu residues are unfavorable for substitution in P2' position; (5) P4 and P3' positions exert no significant effects on determining the substrate specificity of SARS-CoV 3CL^{pro}.

Since the results generated in this study were based on one of the SARS-CoV 3CL^{pro} cleavage site on the SARS-CoV pp1ab polyprotein (residues 3232-3247) which may not necessarily represent all other ten cleavage sites, thus we attempt to compare the amino acid sequence of this studied cleavage site with the other ten cleavage sites to correlate their consensus of substrate specificity. The sequences of the 11 potential SARS-CoV 3CL^{pro} cleavage sites on SARS-CoV BJ01 polyprotein pp1ab were predicted by NetCorona

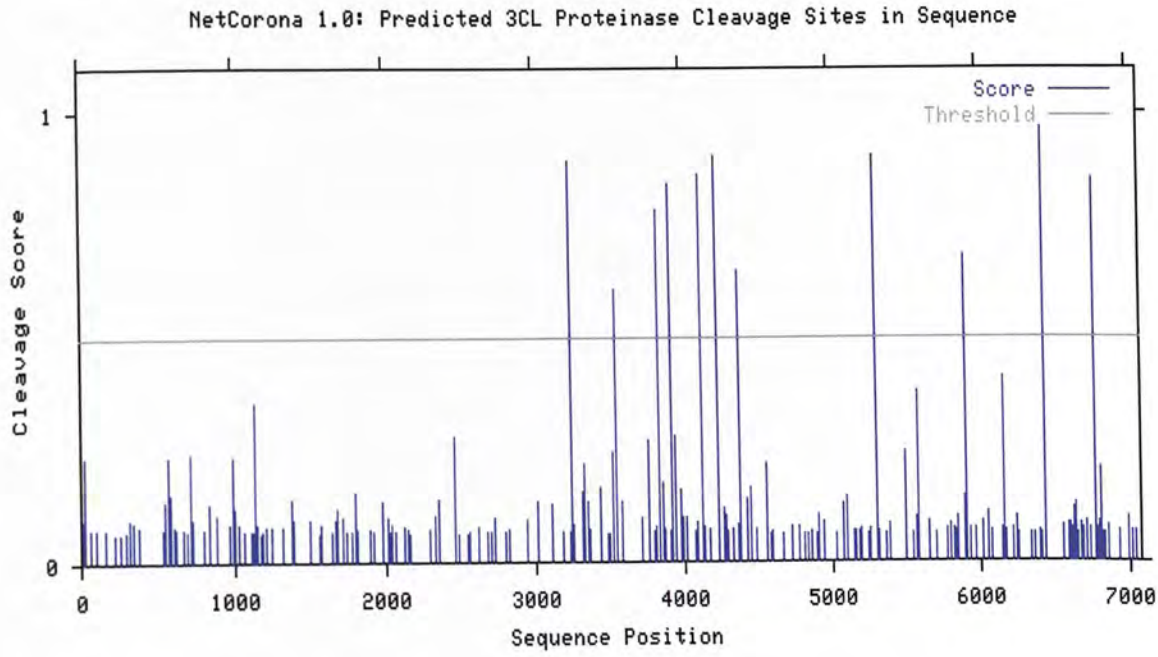

Figure 4.1: Predicted SARS-CoV $3CL^{pro}$ cleavage sites on SARS-CoV BJ01 polyprotein pp1ab.

Eleven 3C-like protease cleavage sites were predicted on the SARS-CoV polyprotein pp1ab. NetCorona (http://www.cbs.dtu.dk/services/NetCorona/) predicts coronavirus 3C-like protease cleavage sites using artificial neural networks on amino acid sequences. The graph illustrates predicted cleavage sites across the protein sequence (x-axis: sequence position from N-terminal to C-terminal). A position with a score (vertical blue lines) crossing the threshold (horizontal gray line at 0.5) is predicted to be possible cleavage sites (Kiemer *et al.*, 2004).

Table 4.1: Amino acid sequences of the 11 cleavage sites of SARS-CoV $3CL^{pro}$.

Cleavage site[a]	Amino acid position		Sequence							Protein product[c]
	Start	End	P4	P3	P2	P1	P1'	P2'	P3'	
1[b]	3237	3243	A	V	L	Q	S	G	F	nsp2
2	3543	3549	V	T	F	Q	G	K	F	nsp3
3	3833	3839	A	T	V	Q	S	K	M	nsp4
4	3916	3922	A	T	L	Q	A	I	A	nsp5
5	4114	4120	V	K	L	Q	N	N	E	nsp6
6	4227	4233	V	R	L	Q	A	G	N	nsp7
7	4366	4372	P	L	M	Q	S	A	D	nsp9
8	5298	5304	T	V	L	Q	A	V	G	nsp10
9	5899	5905	A	T	L	Q	A	E	N	nsp11
10	6426	6432	T	R	L	Q	S	L	E	nsp12
11	6772	6778	P	K	L	Q	A	S	Q	nsp13

[a] SARS-CoV $3CL^{pro}$ cleavage sites on SARS-CoV BJ01 polyprotein pp1ab.

[b] In this study, our peptide substrates are based on cleavage site 1 sequence.

[c] After the proteolytic processing by SARS-CoV $3CL^{pro}$ at the cleavage site, the corresponding non-structural protein (nsp) is yielded (Gao *et al.*, 2003).

1.0 Server[2] that is illustrated in Figure 4.1 (Kiemer *et al.*, 2004), and sequence comparison is shown in Table 4.1 (Gao *et al.*, 2003).

The conserved P1 position is occupied by a Gln residue for all the 11 cleavage sites with no exceptions. The P2 position is predominantly occupied by the Leu residue, and occasionally being occupied by other hydrophobic amino acids, namely Phe, Val and Met for cleavage sites 2, 3 and 7, respectively. In addition to Ser, other amino acids such as Gly, Ala, and Asn are found in P1' position at the 11 cleavage sites, and none of these sites contains the cyclic Pro or acidic Asp and Glu in this position. The P3 position is commonly occupied by Val, Thr, Lys, or Arg residues, and no cyclic Pro is found in this position. Although all the cleavage sites do not contain the cyclic Pro in P2' position, β-branching amino acids are found to occupy P2' position in two of the cleavage sites, i.e., Ile and Leu for cleavage sites 4 and 10, respectively. In this study, we demonstrated that P2' position does not favor the substitution of Ile and Leu for pp1ab cleavage site 1 (Table 4.1) owing to steric hindrance brought about by these β-branching amino acids. The tolerance of Ile or leu in P2' position at cleavage sites 4 and 10 can be accounted by the favorable overall structures of these 2 cleavage sites, are sufficient to mask the negative effects of the β-branching amino acids in P2' position, hence, the SARS-CoV $3CL^{pro}$ can still

[2]NetCorona (http://www.cbs.dtu.dk/services/NetCorona/) predicts coronavirus 3C-like protease cleavage sites using artificial neural networks on amino acid sequences.

bind and cleave effectively at these sites without affected by the P2' position. Further experiments using the peptide substrates based on the sequence of these 2 cleavage sites can be conducted to compare with the results in this study which employed the sequence of cleavage site 1. Different kinds of amino acids including the cyclic Pro are occupying the P3 and P4' position, suggesting that these 2 positions are less important to determine the SARS-CoV $3CL^{pro}$ substrate specificity.

By comparing the sequences of the 11 SARS-CoV $3CL^{pro}$ cleavage sites on the polyprotein pp1ab (Table 4.1), we demonstrated that our results on the sequence preference of cleavage site 1 is generally consistent with the other 10 cleavage sites, thus suggesting the feasibility in using this approach to generate a comprehensive overview of substrate specificity which can be generally applied to other cleavage sites.

4.2.3 A rapid and high throughput approach to screen protease substrate specificity

In contrast to providing quantitative measurements on the kinetic data of the interaction between SARS-CoV $3CL^{pro}$ and the substrates, the scope of our studies mainly focus on the description of a useful tool for rapid and comprehensive screen of substrate specificity for SARS-CoV $3CL^{pro}$, which is also applicable to the studies

of other proteases. By combining the MALDI-TOF mass spectrometry and synthetic peptide-based approaches, peptide cleavage studies on the defined equal molar mixture of a batch of twenty peptide substrates before and after the protease cleavage reaction could be performed simultaneous on a single mass spectrometric analysis.

The resolution of MALDI-TOF mass spectrometer is good enough to resolve the twenty different peptides based on their m/z values, even for those peptides with 1-Da mass difference. Although one inevitable limitation of MALDI-TOF mass spectrometry is the inability to distinguish between species of identical mass (i.e., Ile and Leu) or of almost the same mass (i.e., Gln and Lys); other amino acid residues with only 1-Da mass difference (e.g. Ile/Leu, Asn, Asp, Gln, Glu) can be clearly separated and resolved on the mass spectrum as side-by-side individual peaks as shown in Section 3.2.1. Our approach is advantageous over the traditional methods that are restricted to testing single peptide with defined condition each at a time, and the setup of a total of twenty identical experimental conditions for every analysis. After generating the comprehensive cleavage profile on the 20 peptides simultaneously in the first place based on our strategy, we can easily identify some of the important residues for pinpoint investigations. Hence, the next step can be proceed to the synthesis of target peptides in different batches if the separation of ambiguous residues (e.g. Ile/Leu, Gln/Lys) for

exact identification purpose is necessary. Certainly, if the cleavage result of each peptide (or a smaller number of mixed peptides in different batches) is to be analyzed separately, the resolution of the MALDI-TOF mass spectrometer instrument is not necessary to be as high as to distinguish between species with 1-Da mass difference. Nonetheless, the beauty of high resolution MALDI-TOF mass spectrometry is the ability to separate peaks with small mass difference (even as small as 1-Da), which enables us to generate a comprehensive cleavage profile on the 20 peptides rapidly, without the need of independent synthesis and analysis of each peptide one at a time. Besides, MALDI-TOF mass spectrometric analysis is tolerant to impurities; therefore, the reaction mixture can be analyzed directly after 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 'cartridge replacement' strategy which was based on the flexibility on modifying the solid-phase peptide synthesis (SPPS) 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 mass spectrometric 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 include the rapid screening of potential protease inhibitors of other coronaviruses and the human immunodeficiency virus type 1 (HIV-1). 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 of MALDI-TOF mass spectrometry and the flexibility of solid-phase peptide synthesis approach to generate SARS-CoV $3CL^{pro}$ proteolytic cleavage data rapidly in a comprehensive manner. Our studies provide insights on the full automation of these experimental procedures to meet the increasing demand of high-throughput and automated proteomic studies.

♦ **End of chapter.**

Bibliography

An,S., Chen,C.J., Yu,X., Leibowitz,J.L. and Makino,S. (1999) Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of e protein as an apoptosis inducer. *J Virol,* **73** (9), 7853–9.

Anand,K., Ziebuhr,J., Wadhwani,P., Mesters,J.R. and Hilgenfeld,R. (2003) Coronavirus main proteinase (3clpro) structure: basis for design of anti-sars drugs. *Science,* **300** (5626), 1763–7.

Bacha,U., Barrila,J., Velazquez-Campoy,A., Leavitt,S.A. and Freire,E. (2004) Identification of novel inhibitors of the sars coronavirus main protease 3clpro. *Biochemistry,* **43** (17), 4906–12.

Barnard,D.L., Hubbard,V.D., Burton,J., Smee,D.F., Morrey,J.D., Otto,M.J. and Sidwell,R.W. (2004) Inhibition of severe acute respiratory syndrome-associated coronavirus (sarscov) by calpain inhibitors and beta-d-n4-hydroxycytidine. *Antivir Chem Chemother,* **15** (1), 15–22.

Berger,A. and Schechter,I. (1970) Mapping the active site of papain

with the aid of peptide substrates and inhibitors. *Philos Trans R Soc Lond B Biol Sci,* **257** (813), 249–64.

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 U S A,* **101** (17), 6641–6.

Blom,N., Gammeltoft,S. and Brunak,S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol,* **294** (5), 1351–62.

Booth,C.M., Matukas,L.M., Tomlinson,G.A., Rachlis,A.R., Rose,D.B., Dwosh,H.A., Walmsley,S.L., Mazzulli,T., Avendano,M., Derkach,P., Ephtimios,I.E., Kitai,I., Mederski,B.D., Shadowitz,S.B., Gold,W.L., Hawryluck,L.A., Rea,E., Chenkin,J.S., Cescon,D.W., Poutanen,S.M. and Detsky,A.S. (2003) Clinical features and short-term outcomes of 144 patients with sars in the greater toronto area. *Jama,* **289** (21), 2801–9.

Bosch,B.J., Martina,B.E., Van Der Zee,R., Lepault,J., Haijema,B.J., Versluis,C., Heck,A.J., De Groot,R., Osterhaus,A.D. and Rottier,P.J. (2004) Severe acute respiratory syndrome coronavirus (sars-cov) infection inhibition using spike protein heptad

repeat-derived peptides. *Proc Natl Acad Sci U S A,* **101** (22), 8455–60.

Bosch,B.J., van der Zee,R., de Haan,C.A. and Rottier,P.J. (2003) The coronavirus spike protein is a class i virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol,* **77** (16), 8801–11.

Bukreyev,A., Lamirande,E.W., Buchholz,U.J., Vogel,L.N., Elkins,W.R., St Claire,M., Murphy,B.R., Subbarao,K. and Collins,P.L. (2004) Mucosal immunisation of african green monkeys (cercopithecus aethiops) with an attenuated parainfluenza virus expressing the sars coronavirus spike protein for the prevention of sars. *Lancet,* **363** (9427), 2122–7.

Cavanagh,D., Davis,P.J., Darbyshire,J.H. and Peters,R.W. (1986) Coronavirus ibv: virus retaining spike glycopolypeptide s2 but not s1 is unable to induce virus-neutralizing or haemagglutination-inhibiting antibody, or induce chicken tracheal protection. *J Gen Virol,* **67 (Pt 7)**, 1435–42.

Chan,D.C., Fass,D., Berger,J.M. and Kim,P.S. (1997) Core structure of gp41 from the hiv envelope glycoprotein. *Cell,* **89** (2), 263–73.

Chou,P.Y. and Fasman,G.D. (1978) Prediction of the secondary

structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol,* **47**, 45–148.

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., Waye,M.M., Tsui,S.K., Ng,K.O., Shan,Z.X., Yang,M., Wu,Y.L., Lin,Z.Y. and Ngai,S.M. (2004) Synthetic peptide studies on the severe acute respiratory syndrome (sars) coronavirus spike glycoprotein: perspective for sars vaccine development. *Clin Chem,* **50** (6), 1036–42.

Cinatl,J., Morgenstern,B., Bauer,G., Chandra,P., Rabenau,H. and Doerr,H.W. (2003) Treatment of sars with human interferons. *Lancet,* **362** (9380), 293–4.

Correa,I., Gebauer,F., Bullido,M.J., Sune,C., Baay,M.F., Zwaagstra,K.A., Posthumus,W.P., Lenstra,J.A. and Enjuanes,L. (1990) Localization of antigenic sites of the e2 glycoprotein of transmissible gastroenteritis coronavirus. *J Gen Virol,* **71 (Pt 2)**, 271–9.

Crackower,M.A., Sarao,R., Oudit,G.Y., Yagil,C., Kozieradzki,I., Scanga,S.E., Oliveira-dos Santos,A.J., da Costa,J., Zhang,L., Pei,Y., Scholey,J., Ferrario,C.M., Manoukian,A.S., Chappell,M.C., Backx,P.H., Yagil,Y. and Penninger,J.M. (2002) Angiotensin-converting enzyme 2 is an essential regulator of heart function. *Nature,* **417** (6891), 822–8.

Cui,W., Fan,Y., Wu,W., Zhang,F., Wang,J.Y. and Ni,A.P. (2003) Expression of lymphocytes and lymphocyte subsets in patients with severe acute respiratory syndrome. *Clin Infect Dis,* **37** (6), 857–9.

Delmas,B., Rasschaert,D., Godet,M., Gelfi,J. and Laude,H. (1990) Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike glycoprotein s. *J Gen Virol,* **71 (Pt 6)**, 1313–23.

Donnelly,C.A., Ghani,A.C., Leung,G.M., Hedley,A.J., Fraser,C., Riley,S., Abu-Raddad,L.J., Ho,L.M., Thach,T.Q., Chau,P., Chan,K.P., Lam,T.H., Tse,L.Y., Tsang,T., Liu,S.H., Kong,J.H., Lau,E.M., Ferguson,N.M. and Anderson,R.M. (2003) Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in hong kong. *Lancet,* **361** (9371), 1761–6.

Donoghue,M., Hsieh,F., Baronas,E., Godbout,K., Gosselin,M., Stagliano,N., Donovan,M., Woolf,B., Robison,K., Jeyaseelan,R., Breitbart,R.E. and Acton,S. (2000) A novel angiotensin-converting enzyme-related carboxypeptidase (ace2) converts angiotensin i to angiotensin 1-9. *Circ Res,* **87** (5), E1–9.

Drosten,C., Gunther,S., Preiser,W., van der Werf,S., Brodt,H.R.,

Becker,S., Rabenau,H., Panning,M., Kolesnikova,L., Fouchier,R.A., Berger,A., Burguiere,A.M., Cinatl,J., Eickmann,M., Escriou,N., Grywna,K., Kramme,S., Manuguerra,J.C., Muller,S., Rickerts,V., Sturmer,M., Vieth,S., Klenk,H.D., Osterhaus,A.D., Schmitz,H. and Doerr,H.W. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med,* **348** (20), 1967–76.

Eckert,D.M. and Kim,P.S. (2001) Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem,* **70**, 777–810.

Emery,S.L., Erdman,D.D., Bowen,M.D., Newton,B.R., Winchell,J.M., Meyer,R.F., Tong,S., Cook,B.T., Holloway,B.P., McCaustland,K.A., Rota,P.A., Bankamp,B., Lowe,L.E., Ksiazek,T.G., Bellini,W.J. and Anderson,L.J. (2004) Real-time reverse transcription-polymerase chain reaction assay for sars-associated coronavirus. *Emerg Infect Dis,* **10** (2), 311–6.

Eriksson,U., Danilczyk,U. and Penninger,J.M. (2002) Just the beginning: novel functions for angiotensin-converting enzymes. *Curr Biol,* **12** (21), R745–52.

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** (3), 934–40.

Fan,K., Wei,P., Feng,Q., Chen,S., Huang,C., Ma,L., Lai,B., Pei,J.,

Liu,Y., Chen,J. and Lai,L. (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3c-like proteinase. *J Biol Chem,* **279** (3), 1637–42.

Fouchier,R.A., Kuiken,T., Schutten,M., van Amerongen,G., van Doornum,G.J., van den Hoogen,B.G., Peiris,M., Lim,W., Stohr,K. and Osterhaus,A.D. (2003) Aetiology: koch's postulates fulfilled for sars virus. *Nature,* **423** (6937), 240.

Frana,M.F., Behnke,J.N., Sturman,L.S. and Holmes,K.V. (1985) Proteolytic cleavage of the e2 glycoprotein of murine coronavirus: host-dependent differences in proteolytic cleavage and cell fusion. *J Virol,* **56** (3), 912–20.

Gallagher,T.M. and Buchmeier,M.J. (2001) Coronavirus spike proteins in viral entry and pathogenesis. *Virology,* **279** (2), 371–4.

Gao,F., Ou,H.Y., Chen,L.L., Zheng,W.X. and Zhang,C.T. (2003) Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing sars-cov genomes. *FEBS Lett,* **553** (3), 451–6.

Garoff,H., Hewson,R. and Opstelten,D.J. (1998) Virus maturation by budding. *Microbiol Mol Biol Rev,* **62** (4), 1171–90.

Gonzalez,J.M., Gomez-Puertas,P., Cavanagh,D., Gorbalenya,A.E.

and Enjuanes,L. (2003) A comparative sequence analysis to revise the current taxonomy of the family coronaviridae. *Arch Virol,* **148** (11), 2207–35.

Groneberg,D.A., Poutanen,S.M., Low,D.E., Lode,H., Welte,T. and Zabel,P. (2005) Treatment and vaccines for severe acute respiratory syndrome. *Lancet Infect Dis,* **5** (3), 147–55.

Hamming,I., Timens,W., Bulthuis,M.L., Lely,A.T., Navis,G.J. and van Goor,H. (2004) Tissue distribution of ace2 protein, the functional receptor for sars coronavirus. a first step in understanding sars pathogenesis. *J Pathol,* **203** (2), 631–7.

Harmer,D., Gilbert,M., Borman,R. and Clark,K.L. (2002) Quantitative mrna expression profiling of ace 2, a novel homologue of angiotensin converting enzyme. *FEBS Lett,* **532** (1-2), 107–10.

He,Y., Lu,H., Siddiqui,P., Zhou,Y. and Jiang,S. (2005) Receptor-binding domain of severe acute respiratory syndrome coronavirus spike protein contains multiple conformation-dependent epitopes that induce highly potent neutralizing antibodies. *J Immunol,* **174** (8), 4908–15.

He,Y., Zhou,Y., Siddiqui,P. and Jiang,S. (2004) Inactivated sars-cov vaccine elicits high titers of spike protein-specific antibodies that block receptor binding and virus entry. *Biochem Biophys Res Commun,* **325** (2), 445–52.

Hegyi,A. and Ziebuhr,J. (2002) Conservation of substrate specificities among coronavirus main proteases. *J Gen Virol,* **83** (Pt 3), 595–9.

Hiscox,J.A., Wurm,T., Wilson,L., Britton,P., Cavanagh,D. and Brooks,G. (2001) The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. *J Virol,* **75** (1), 506–12.

Ho,J.C., Ooi,G.C., Mok,T.Y., Chan,J.W., Hung,I., Lam,B., Wong,P.C., Li,P.C., Ho,P.L., Lam,W.K., Ng,C.K., Ip,M.S., Lai,K.N., Chan-Yeung,M. and Tsang,K.W. (2003) High-dose pulse versus nonpulse corticosteroid regimens in severe acute respiratory syndrome. *Am J Respir Crit Care Med,* **168** (12), 1449–56.

Hofmann,H. and Pohlmann,S. (2004) Cellular entry of the sars coronavirus. *Trends Microbiol,* **12** (10), 466–72.

Holmes,K.V. (2003) Sars-associated coronavirus. *N Engl J Med,* **348** (20), 1948–51.

Hopp,T.P. and Woods,K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A,* **78** (6), 3824–8.

Hsu,J.T., Kuo,C.J., Hsieh,H.P., Wang,Y.C., Huang,K.K., Lin,C.P., Huang,P.F., Chen,X. and Liang,P.H. (2004) Evaluation of

metal-conjugated compounds as inhibitors of 3cl protease of sars-cov. *FEBS Lett,* **574** (1-3), 116–20.

Ingallinella,P., Bianchi,E., Finotto,M., Cantoni,G., Eckert,D.M., Supekar,V.M., Bruckmann,C., Carfi,A. and Pessi,A. (2004) Structural characterization of the fusion-active complex of severe acute respiratory syndrome (sars) coronavirus. *Proc Natl Acad Sci U S A,* **101** (23), 8709–14.

Jackwood,M.W., Hilt,D.A., Callison,S.A., Lee,C.W., Plaza,H. and Wade,E. (2001) Spike glycoprotein cleavage recognition site analysis of infectious bronchitis virus. *Avian Dis,* **45** (2), 366–72.

Jain,R.P., Pettersson,H.I., Zhang,J., Aull,K.D., Fortin,P.D., Huitema,C., Eltis,L.D., Parrish,J.C., James,M.N., Wishart,D.S. and Vederas,J.C. (2004) Synthesis and evaluation of keto-glutamine analogues as potent inhibitors of severe acute respiratory syndrome 3clpro. *J Med Chem,* **47** (25), 6113–6.

Jeffers,S.A., Tusell,S.M., Gillim-Ross,L., Hemmila,E.M., Achenbach,J.E., Babcock,G.J., Thomas,W. D.,J., Thackray,L.B., Young,M.D., Mason,R.J., Ambrosino,D.M., Wentworth,D.E., Demartini,J.C. and Holmes,K.V. (2004) Cd209l (l-sign) is a receptor for severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A,* **101** (44), 15748–53.

Kao,R.Y., To,A.P., Ng,L.W., Tsui,W.H., Lee,T.S., Tsoi,H.W. and Yuen,K.Y. (2004) Characterization of sars-cov main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay. *FEBS Lett,* **576** (3), 325–30.

Kapczynski,D.R., Hilt,D.A., Shapiro,D., Sellers,H.S. and Jackwood,M.W. (2003) Protection of chickens from infectious bronchitis by in ovo and intramuscular vaccination with a dna vaccine expressing the s1 glycoprotein. *Avian Dis,* **47** (2), 272–85.

Kiemer,L., Lund,O., Brunak,S. and Blom,N. (2004) Coronavirus 3clpro proteinase cleavage sites: possible relevance to sars virus pathology. *BMC Bioinformatics,* **5** (1), 72.

Kim,T.W., Lee,J.H., Hung,C.F., Peng,S., Roden,R., Wang,M.C., Viscidi,R., Tsai,Y.C., He,L., Chen,P.J., Boyd,D.A. and Wu,T.C. (2004) Generation and characterization of dna vaccines targeting the nucleocapsid protein of severe acute respiratory syndrome coronavirus. *J Virol,* **78** (9), 4638–45.

Komatsu,T., Suzuki,Y., Imai,J., Sugano,S., Hida,M., Tanigami,A., Muroi,S., Yamada,Y. and Hanaoka,K. (2002) Molecular cloning, mrna expression and chromosomal localization of mouse angiotensin-converting enzyme-related carboxypeptidase (mace2). *DNA Seq,* **13** (4), 217–20.

Ksiazek,T.G., Erdman,D., Goldsmith,C.S., Zaki,S.R., Peret,T., Emery,S., Tong,S., Urbani,C., Comer,J.A., Lim,W., Rollin,P.E., Dowell,S.F., Ling,A.E., Humphrey,C.D., Shieh,W.J., Guarner,J., Paddock,C.D., Rota,P., Fields,B., DeRisi,J., Yang,J.Y., Cox,N., Hughes,J.M., LeDuc,J.W., Bellini,W.J. and Anderson,L.J. (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med,* **348** (20), 1953–66.

Kuiken,T., Fouchier,R.A., Schutten,M., Rimmelzwaan,G.F., van Amerongen,G., van Riel,D., Laman,J.D., de Jong,T., van Doornum,G., Lim,W., Ling,A.E., Chan,P.K., Tam,J.S., Zambon,M.C., Gopal,R., Drosten,C., van der Werf,S., Escriou,N., Manuguerra,J.C., Stohr,K., Peiris,J.S. and Osterhaus,A.D. (2003) Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet,* **362** (9380), 263–70.

Kuo,L. and Masters,P.S. (2002) Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J Virol,* **76** (10), 4987–99.

Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage t4. *Nature,* **227** (5259),

680–5.

Lai,M.M. and Cavanagh,D. (1997) The molecular biology of coronaviruses. *Adv Virus Res,* **48**, 1–100.

Lee,N., Hui,D., Wu,A., Chan,P., Cameron,P., Joynt,G.M., Ahuja,A., Yung,M.Y., Leung,C.B., To,K.F., Lui,S.F., Szeto,C.C., Chung,S. and Sung,J.J. (2003) A major outbreak of severe acute respiratory syndrome in hong kong. *N Engl J Med,* **348** (20), 1986–94.

Leparc-Goffart,I., Hingley,S.T., Chua,M.M., Phillips,J., Lavi,E. and Weiss,S.R. (1998) Targeted recombination within the spike gene of murine coronavirus mouse hepatitis virus-a59: q159 is a determinant of hepatotropism. *J Virol,* **72** (12), 9628–36.

Leung,W.K., To,K.F., Chan,P.K., Chan,H.L., Wu,A.K., Lee,N., Yuen,K.Y. and Sung,J.J. (2003) Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection. *Gastroenterology,* **125** (4), 1011–7.

Li,W., Moore,M.J., Vasilieva,N., Sui,J., Wong,S.K., Berne,M.A., Somasundaran,M., Sullivan,J.L., Luzuriaga,K., Greenough,T.C., Choe,H. and Farzan,M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus. *Nature,* **426** (6965), 450–4.

Liu,B. and Zhou,J. (2005) Sars-cov protease inhibitors design using virtual screening method from natural products libraries. *J Comput Chem,* **26** (5), 484–90.

Liu,S., Xiao,G., Chen,Y., He,Y., Niu,J., Escalante,C.R., Xiong,H., Farmar,J., Debnath,A.K., Tien,P. and Jiang,S. (2004) Interaction between heptad repeat 1 and 2 regions in spike protein of sars-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet,* **363** (9413), 938–47.

Mackay,I.M., Arden,K.E. and Nitsche,A. (2002) Real-time pcr in virology. *Nucleic Acids Res,* **30** (6), 1292–305.

Makela,M.J., Puhakka,T., Ruuskanen,O., Leinonen,M., Saikku,P., Kimpimaki,M., Blomqvist,S., Hyypia,T. and Arstila,P. (1998) Viruses and bacteria in the etiology of the common cold. *J Clin Microbiol,* **36** (2), 539–42.

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., Cloutier,A., Coughlin,S.M., Freeman,D., Girn,N., Griffith,O.L., Leach,S.R., Mayo,M., McDonald,H., Montgomery,S.B., Pandoh,P.K., Petrescu,A.S., Robertson,A.G., Schein,J.E., Siddiqui,A., Smailus,D.E., Stott,J.M., Yang,G.S., Plummer,F., Andonov,A., Artsob,H., Bastien,N., Bernard,K.,

Booth,T.F., Bowness,D., Czub,M., Drebot,M., Fernando,L., Flick,R., Garbutt,M., Gray,M., Grolla,A., Jones,S., Feldmann,H., Meyers,A., Kabani,A., Li,Y., Normand,S., Stroher,U., Tipples,G.A., Tyler,S., Vogrig,R., Ward,D., Watson,B., Brunham,R.C., Krajden,M., Petric,M., Skowronski,D.M., Upton,C. and Roper,R.L. (2003) The genome sequence of the sars-associated coronavirus. *Science,* **300** (5624), 1399–404.

Marshall,E. and Enserink,M. (2004) Medicine. caution urged on sars vaccines. *Science,* **303** (5660), 944–6.

Marzi,A., Gramberg,T., Simmons,G., Moller,P., Rennekamp,A.J., Krumbiegel,M., Geier,M., Eisemann,J., Turza,N., Saunier,B., Steinkasserer,A., Becker,S., Bates,P., Hofmann,H. and Pohlmann,S. (2004) Dc-sign and dc-signr interact with the glycoprotein of marburg virus and the s protein of severe acute respiratory syndrome coronavirus. *J Virol,* **78** (21), 12090–5.

Ng,E.K., Hui,D.S., Chan,K.C., Hung,E.C., Chiu,R.W., Lee,N., Wu,A., Chim,S.S., Tong,Y.K., Sung,J.J., Tam,J.S. and Lo,Y.M. (2003) Quantitative analysis and prognostic implication of sars coronavirus rna in the plasma and serum of patients with severe acute respiratory syndrome. *Clin Chem,* **49** (12), 1976–80.

Nicholls,J.M., Poon,L.L., Lee,K.C., Ng,W.F., Lai,S.T., Leung,C.Y.,

Chu,C.M., Hui,P.K., Mak,K.L., Lim,W., Yan,K.W., Chan,K.H., Tsang,N.C., Guan,Y., Yuen,K.Y. and Peiris,J.S. (2003) Lung pathology of fatal severe acute respiratory syndrome. *Lancet,* **361** (9371), 1773–8.

Nie,Y., Wang,P., Shi,X., Wang,G., Chen,J., Zheng,A., Wang,W., Wang,Z., Qu,X., Luo,M., Tan,L., Song,X., Yin,X., Chen,J., Ding,M. and Deng,H. (2004) Highly infectious sars-cov pseudotyped virus reveals the cell tropism and its correlation with receptor expression. *Biochem Biophys Res Commun,* **321** (4), 994–1000.

Ning,Q., Lakatoo,S., Liu,M., Yang,W., Wang,Z., Phillips,M.J. and Levy,G.A. (2003) Induction of prothrombinase fgl2 by the nucleocapsid protein of virulent mouse hepatitis virus is dependent on host hepatic nuclear factor-4 alpha. *J Biol Chem,* **278** (18), 15541–9.

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., Cheng,V.C., Chan,K.H., Tsang,D.N., Yung,R.W., Ng,T.K. and Yuen,K.Y. (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet,* **361** (9366), 1319–25.

Phillips,J.J., Chua,M.M., Rall,G.F. and Weiss,S.R. (2002) Murine coronavirus spike glycoprotein mediates degree of viral spread,

inflammation, and virus-induced immunopathology in the central nervous system. *Virology,* **301** (1), 109–20.

Poutanen,S.M., Low,D.E., Henry,B., Finkelstein,S., Rose,D., Green,K., Tellier,R., Draker,R., Adachi,D., Ayers,M., Chan,A.K., Skowronski,D.M., Salit,I., Simor,A.E., Slutsky,A.S., Doyle,P.W., Krajden,M., Petric,M., Brunham,R.C. and McGeer,A.J. (2003) Identification of severe acute respiratory syndrome in canada. *N Engl J Med,* **348** (20), 1995–2005.

Rajnarayanan,R.V., Dakshanamurthy,S. and Pattabiraman,N. (2004) "teaching old drugs to kill new bugs": structure-based discovery of anti-sars drugs. *Biochem Biophys Res Commun,* **321** (2), 370–8.

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., Tong,S., Tamin,A., Lowe,L., Frace,M., DeRisi,J.L., Chen,Q., Wang,D., Erdman,D.D., Peret,T.C., Burns,C., Ksiazek,T.G., Rollin,P.E., Sanchez,A., Liffick,S., Holloway,B., Limor,J., McCaustland,K., Olsen-Rasmussen,M., Fouchier,R., Gunther,S., Osterhaus,A.D., Drosten,C., Pallansch,M.A., Anderson,L.J. and Bellini,W.J. (2003) Characterization of a novel

coronavirus associated with severe acute respiratory syndrome. *Science,* **300** (5624), 1394–9.

Sanchez,C.M., Izeta,A., Sanchez-Morgado,J.M., Alonso,S., Sola,I., Balasch,M., Plana-Duran,J. and Enjuanes,L. (1999) Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. *J Virol,* **73** (9), 7607–18.

Schmidt,I., Skinner,M. and Siddell,S. (1987) Nucleotide sequence of the gene encoding the surface projection glycoprotein of coronavirus mhv-jhm. *J Gen Virol,* **68 (Pt 1)**, 47–56.

Schmidt-Mende,J., Bieck,E., Hugle,T., Penin,F., Rice,C.M., Blum,H.E. and Moradpour,D. (2001) Determinants for membrane association of the hepatitis c virus rna-dependent rna polymerase. *J Biol Chem,* **276** (47), 44052–63.

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** (23), 24765–73.

Shi,Y., Yi,Y., Li,P., Kuang,T., Li,L., Dong,M., Ma,Q. and Cao,C. (2003) Diagnosis of severe acute respiratory syndrome (sars) by detection of sars coronavirus nucleocapsid antibodies in an

antigen-capturing enzyme-linked immunosorbent assay. *J Clin Microbiol,* **41** (12), 5781–2.

Siddell,S., Wege,H. and Ter Meulen,V. (1983) The biology of coronaviruses. *J Gen Virol,* **64 (Pt 4)**, 761–76.

Spiga,O., Bernini,A., Ciutti,A., Chiellini,S., Menciassi,N., Finetti,F., Causarono,V., Anselmi,F., Prischi,F. and Niccolai,N. (2003) Molecular modelling of s1 and s2 subunits of sars coronavirus spike glycoprotein. *Biochem Biophys Res Commun,* **310** (1), 78–83.

Stadler,K., Masignani,V., Eickmann,M., Becker,S., Abrignani,S., Klenk,H.D. and Rappuoli,R. (2003) Sars–beginning to understand a new virus. *Nat Rev Microbiol,* **1** (3), 209–18.

Stern,D.F. and Sefton,B.M. (1982) Coronavirus proteins: biogenesis of avian infectious bronchitis virus virion proteins. *J Virol,* **44** (3), 794–803.

Sturman,L.S. and Holmes,K.V. (1984) Proteolytic cleavage of peplomeric glycoprotein e2 of mhv yields two 90k subunits and activates cell fusion. *Adv Exp Med Biol,* **173**, 25–35.

Taguchi,F. (1995) The s2 subunit of the murine coronavirus spike protein is not involved in receptor binding. *J Virol,* **69** (11), 7260–3.

Thiel,V., Ivanov,K.A., Putics,A., Hertzig,T., Schelle,B., Bayer,S., Weissbrich,B., Snijder,E.J., Rabenau,H., Doerr,H.W., Gorbalenya,A.E. and Ziebuhr,J. (2003) Mechanisms and enzymes involved in sars coronavirus genome expression. *J Gen Virol,* **84** (Pt 9), 2305–15.

Tipnis,S.R., Hooper,N.M., Hyde,R., Karran,E., Christie,G. and Turner,A.J. (2000) A human homolog of angiotensin-converting enzyme. cloning and functional expression as a captopril-insensitive carboxypeptidase. *J Biol Chem,* **275** (43), 33238–43.

Tripet,B., Howard,M.W., Jobling,M., Holmes,R.K., Holmes,K.V. and Hodges,R.S. (2004) Structural characterization of the sars-coronavirus spike s fusion protein core. *J Biol Chem,* **279** (20), 20836–49.

Tsang,K.W., Ho,P.L., Ooi,G.C., Yee,W.K., Wang,T., Chan-Yeung,M., Lam,W.K., Seto,W.H., Yam,L.Y., Cheung,T.M., Wong,P.C., Lam,B., Ip,M.S., Chan,J., Yuen,K.Y. and Lai,K.N. (2003) A cluster of cases of severe acute respiratory syndrome in hong kong. *N Engl J Med,* **348** (20), 1977–85.

Tsui,P.T., Kwok,M.L., Yuen,H. and Lai,S.T. (2003) Severe acute respiratory syndrome: clinical outcome and prognostic correlates. *Emerg Infect Dis,* **9** (9), 1064–9.

Vennema,H., Godeke,G.J., Rossen,J.W., Voorhout,W.F.,

Horzinek,M.C., Opstelten,D.J. and Rottier,P.J. (1996) Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *Embo J,* **15** (8), 2020–8.

Wang,P., Chen,J., Zheng,A., Nie,Y., Shi,X., Wang,W., Wang,G., Luo,M., Liu,H., Tan,L., Song,X., Wang,Z., Yin,X., Qu,X., Wang,X., Qing,T., Ding,M. and Deng,H. (2004) Expression cloning of functional receptor used by sars coronavirus. *Biochem Biophys Res Commun,* **315** (2), 439–44.

Wentworth,D.E. and Holmes,K.V. (2001) Molecular determinants of species specificity in the coronavirus receptor aminopeptidase n (cd13): influence of n-linked glycosylation. *J Virol,* **75** (20), 9741–52.

Wong,K.T., Antonio,G.E., Hui,D.S., Lee,N., Yuen,E.H., Wu,A., Leung,C.B., Rainer,T.H., Cameron,P., Chung,S.S., Sung,J.J. and Ahuja,A.T. (2003*a*) Thin-section ct of severe acute respiratory syndrome: evaluation of 73 patients exposed to or with the disease. *Radiology,* **228** (2), 395–400.

Wong,R.S., Wu,A., To,K.F., Lee,N., Lam,C.W., Wong,C.K., Chan,P.K., Ng,M.H., Yu,L.M., Hui,D.S., Tam,J.S., Cheng,G. and Sung,J.J. (2003*b*) Haematological manifestations in patients

with severe acute respiratory syndrome: retrospective analysis. *Bmj,* **326** (7403), 1358–62.

Wong,S.K., Li,W., Moore,M.J., Choe,H. and Farzan,M. (2004) A 193-amino acid fragment of the sars coronavirus s protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem,* **279** (5), 3197–201.

Wu,C.Y., Jan,J.T., Ma,S.H., Kuo,C.J., Juan,H.F., Cheng,Y.S., Hsu,H.H., Huang,H.C., Wu,D., Brik,A., Liang,F.S., Liu,R.S., Fang,J.M., Chen,S.T., Liang,P.H. and Wong,C.H. (2004*a*) Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc Natl Acad Sci U S A,* **101** (27), 10012–7.

Wu,H.S., Chiu,S.C., Tseng,T.C., Lin,S.F., Lin,J.H., Hsu,Y.H., Wang,M.C., Lin,T.L., Yang,W.Z., Ferng,T.L., Huang,K.H., Hsu,L.C., Lee,L.L., Yang,J.Y., Chen,H.Y., Su,S.P., Yang,S.Y., Lin,S.Y., Lin,T.H. and Su,I.S. (2004*b*) Serologic and molecular biologic methods for sars-associated coronavirus infection, taiwan. *Emerg Infect Dis,* **10** (2), 304–10.

Wu,X.D., Shang,B., Yang,R.F., Yu,H., Ma,Z.H., Shen,X., Ji,Y.Y., Lin,Y., Wu,Y.D., Lin,G.M., Tian,L., Gan,X.Q., Yang,S., Jiang,W.H., Dai,E.H., Wang,X.Y., Jiang,H.L., Xie,Y.H., Zhu,X.L., Pei,G., Li,L., Wu,J.R. and Sun,B. (2004*c*) The spike

protein of severe acute respiratory syndrome (sars) is cleaved in virus infected vero-e6 cells. *Cell Res,* **14** (5), 400–6.

Xu,Y., Liu,Y., Lou,Z., Qin,L., Li,X., Bai,Z., Pang,H., Tien,P., Gao,G.F. and Rao,Z. (2004*a*) Structural basis for coronavirus-mediated membrane fusion. crystal structure of mouse hepatitis virus spike protein fusion core. *J Biol Chem,* **279** (29), 30514–22.

Xu,Y., Lou,Z., Liu,Y., Pang,H., Tien,P., Gao,G.F. and Rao,Z. (2004*b*) Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core. *J Biol Chem,* **279** (47), 49414–9.

Yam,W.C., Chan,K.H., Poon,L.L., Guan,Y., Yuen,K.Y., Seto,W.H. and Peiris,J.S. (2003) Evaluation of reverse transcription-pcr assays for rapid diagnosis of severe acute respiratory syndrome associated with a novel coronavirus. *J Clin Microbiol,* **41** (10), 4521–4.

Yamamoto,N., Yang,R., Yoshinaka,Y., Amari,S., Nakano,T., Cinatl,J., Rabenau,H., Doerr,H.W., Hunsmann,G., Otaka,A., Tamamura,H., Fujii,N. and Yamamoto,N. (2004) Hiv protease inhibitor nelfinavir inhibits replication of sars-associated coronavirus. *Biochem Biophys Res Commun,* **318** (3), 719–25.

Yang,H., Yang,M., Ding,Y., Liu,Y., Lou,Z., Zhou,Z., Sun,L., Mo,L.,

Ye,S., Pang,H., Gao,G.F., Anand,K., Bartlam,M., Hilgenfeld,R. and Rao,Z. (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A,* **100** (23), 13190–5.

Yang,Z.Y., Huang,Y., Ganesh,L., Leung,K., Kong,W.P., Schwartz,O., Subbarao,K. and Nabel,G.J. (2004*a*) ph-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through dc-sign. *J Virol,* **78** (11), 5642–50.

Yang,Z.Y., Kong,W.P., Huang,Y., Roberts,A., Murphy,B.R., Subbarao,K. and Nabel,G.J. (2004*b*) A dna vaccine induces sars coronavirus neutralization and protective immunity in mice. *Nature,* **428** (6982), 561–4.

Yeager,C.L., Ashmun,R.A., Williams,R.K., Cardellichio,C.B., Shapiro,L.H., Look,A.T. and Holmes,K.V. (1992) Human aminopeptidase n is a receptor for human coronavirus 229e. *Nature,* **357** (6377), 420–2.

Zhou,J., Wang,W., Zhong,Q., Hou,W., Yang,Z., Xiao,S.Y., Zhu,R., Tang,Z., Wang,Y., Xian,Q., Tang,H. and Wen,L. (2005) Immunogenicity, safety, and protective efficacy of an inactivated

sars-associated coronavirus vaccine in rhesus monkeys. *Vaccine,* **23** (24), 3202–9.

Ziebuhr,J., Snijder,E.J. and Gorbalenya,A.E. (2000) Virus-encoded proteinases and proteolytic processing in the nidovirales. *J Gen Virol,* **81** (Pt 4), 853–79.