An Overall Picture of SARS Coronavirus (SARS-CoV) Genome-Encoded Major Proteins: Structures, Functions and Drug Development

Shuai Chen[1], Haibin Luo[1], Lili Chen[1], Jing Chen[1], Jianhua Shen[1], Weiliang Zhu[1], Kaixian Chen[1], Xu Shen[1,2,*] and Hualiang Jiang[1,2,*]

[1]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Graduate School of the Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 201203 and [2]School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

Abstract: A severe atypical pneumonia designated as severe acute respiratory syndrome (SARS) by The World Health Organization broke out in China and menaced to more than other 30 countries between the end of the year 2002 and June of the year 2003. A novel coronavirus called severe acute respiratory syndrome coronavirus (SARS-CoV) has been recently identified as the etiological agent responsible for the infectious SARS disease. Based on extensively scientific cooperation and almost two-year's studies, remarkable achievements have been made in the understanding of the phylogenetic property and the genome organization of SARS-CoV, as well as the detailed characters of the major proteins involved in SARS-CoV life cycle. In this review, we would like to summarize the substantial scientific progress that has been made towards the structural and functional aspects of SARS-CoV associated key proteins. The progress focused on the corresponding key proteins' structure-based drug and vaccine developments has been also highlighted. The concerted and cooperative response for the treatment of the SARS disease has been proved to be a triumph of global public health and provides a new paradigm for the detection and control of future emerging infectious disease threats.

Key Words: Atypical pneumonia; SARS Coronavirus; SARS-CoV genome-encoded major proteins; structural and functional characterization; inhibitor design and screening; structure-based drug development.

INTRODUCTION

From the end of the year 2002 to the June of the year 2003, one severe epidemic disease called severe acute respiratory syndrome (SARS) broke out in China and quickly spread to more than 30 other countries (http://www.who.int/csr/sars/country/en/). In fact, the SARS disease ever severely menaced the worldwide population and totally more than 8400 patients and 789 deaths were reported by the World Health Organization (http://www.who.int/csr/sars/country/en/). This emergent status urged a tightly worldwide collaboration in dealing with such a disease. With many efforts in the extensively scientific cooperation, one previously unknown coronavirus designated as SARS coronavirus (SARS-CoV) was determined as the etiological agent of this severe infection and believed to be a novel human coronavirus possibly originated from non-human host [1].

The coronaviruses are a diverse group of enveloped, positive-stranded RNA viruses that cause respiratory and enteric diseases in humans and other animals [2]. Genome sequencing analysis has revealed that SARS-CoV involve the largest viral RNA genome known to date, encompassing 29,727 nucleotides predicted to contain 14 functional open reading frames (ORFs) [1]. The two large 5'-terminal ORFs, 1a and 1b, constitute the replicase gene encoding the proteins ($PL2^{pro}$, $3CL^{pro}$ and a series of non-structural proteins) required for viral RNA transcription and replication and the remaining twelve ORFs encode the four key structural proteins, the spike (S) protein, the nucleocapsid (N) protein, the membrane (M) protein and the small envelope (E) protein, and other eight accessory proteins that are not likely to be essential in tissue culture but may provide a selective advantage in the infected host [3]. The phylogenetic analyses of the major structural proteins indicated that SARS-CoV does not closely resemble any of the previously known three groups of coronaviruses [4]. Accordingly, the elucidation of the structural and functional characters of the major proteins encoded by SARS-CoV genome has become an alluring project, and the corresponding studies have been extensively reported. With the resolved crystal or homology modeling structures and corresponding biological function determinations of these major proteins, a series of structure-based drug and vaccine developments have been carried out by using the virtual screening and the *in vitro* putative inhibitors screening platforms. In this review, we would like to summarize the current understandings of the structural and functional aspects of some major SARS-CoV proteins. The corresponding drug and vaccine development for the treatment of the SARS disease has been also presented in this review.

*Address correspondence to these authors at the Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Shanghai Institutes for Biological Sciences, Graduate School of the Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 201203 Fax: 86-21-50806918; E-mail: xshen@mail.shcnc.ac.cn, hljiang@mail.shcnc.ac.cn

1381-6128/06 $50.00+.00

MAJOR PROTEINS ENCODED BY SARS-COV GENOME AND CORRESPONDING DRUG DEVELOPMENT

3C-Like Protease (3CL^{pro})

SARS-CoV genome has been predicted to contain 14 functional open reading frames (ORFs) for encoding the proteins required for virus replication and transcription [5]. Two large 5'-terminal ORFs, 1a and 1b, constitute the replicase gene that encodes two overlapping polyproteins, pp1a (around 450kDa) and pp1ab (around 750kDa) necessary for viral RNA synthesis. Proteolytic processing of the non-structural polyproteins is a vital step in the replication cycle of coronavirus and such processing is commonly performed by virus-genome encoded proteases [6,7]. Polyproteins pp1a and pp1ab are cleaved extensively by a papain-like cysteine protease (PL2^{pro}) and another chymotrypsin-like protease to yield a multi-subunits protein complex called "viral replicase-transcriptase" [5,8,9]. The chymotrypsin-like protease was named as 3C-like protease (3CL^{pro}) due to its distant relationship with the 3C proteases of picornaviruses [8]. Since 3CL^{pro} functions as a pivotal protease in coronavirus polyproteins processing and controls the activities of coronavirus replication complexes, it is also called main protease (M^{pro}). Due to its functional indispensability in coronavirus life cycle, SARS-CoV 3CL^{pro} has become an attractive target in discovering new anti-SARS agents [9].

The molecular cloning, expression, purification, and mass spectrometric characterization of SARS-CoV 3CL^{pro} have been reported [10], and sequence alignment indicated that SARS-CoV 3CL^{pro} shares high homology with the 3CL^{pro}s of other coronaviruses such as transmissible gastroenteritis coronavirus (TGEV) and human coronavirus (HCoV) 229E [9,11]. The X-ray crystal structures of the 3CL^{pro}s of TGEV, HCoV 229E and SARS-CoV revealed that the 3D structures of 3CL proteases are more conserved compared with their sequences [9,12]. Structurally (Fig. **1**), SARS-CoV 3CL^{pro} contains three domains: domains I and II (residues 8-101 and residues102-184) have six-stranded antiparallel β-barrel forming a chymotrypsin fold, which resemble the architecture of chymotrypsin and picornaviruses 3C proteases. The substrate-binding site is located in a cleft between these two domains, and residues 201-303 form an antiparallel globular cluster of five α-helices (domain III) connecting to domain II by a long loop region (residues 185-200). The 16-residue loop region has been implicated to play a role in substrate-binding as indicated from the reported biochemical data and the crystal structure [9,12-14]. The crystal structure analysis shows that SARS-CoV 3CL^{pro} forms a dimer with the two molecules oriented perpendicular to one another (Fig. **1**), similar to the TGEV and HCoV 3CL^{pro} structures [9,12,13]. The contact interface predominantly involves interactions between domain II of one monomer and the N-terminal residues of the other monomer. The N-terminus of one monomer (residues 1-7) is squeezed into between domains II, III of the parent molecule and domain II of the other monomer resulting in a number of highly specific interactions within the dimeric structure. The influence of N-terminal residues 1-7 deletion on the dimerization features of SARS-CoV 3CL^{pro} has been quantitatively evaluated by using biochemical and biophysical techniques [15]. The results revealed that the N-terminal deleted protease still remains a dimer/monomer mixture within a wide range of protein concentrations, which is similar to the full-length protease. The equilibrium dissociation constant (K_{diss}) of the N-terminal deleted protease dimer (262 μM) is very similar to that of the full-length protease dimer (227 μM) (Table **1**). However, the N-terminal deletion results in almost complete loss of enzymatic activity of the protease. Complementary molecular dynamics (MD) and docking simulations demonstrated that the N-terminal deleted protease dimer adopts a state different from that of the full-length protease dimer, which increases the angle between the two monomers and reduces the binding pocket that is not beneficial to the substrate binding (Fig. **2**). These results thereby suggested that the N-terminus is not indispensable for the protease dimerization, but may fix the dimer in its active state that is vital to the enzymatic activity [15]. Furthermore, domain III has been predicted to be potent in substrate recognition and responsible for positioning the N-terminus of one monomer to interact with the active site of the other monomer. The critical role of domain III in dimerization and enzymatic activity of SARS 3CL^{pro} has been also demonstrated [14]. As the crystal structures of 3CL^{pro} in different coronaviruses give the similar dimeric structures and nearly all the side chains of 3CL^{pro} involved in the formation of the dimer are conserved [9,12], it is believed that the dimer might be the biological functional form of 3CL^{pro} and the dimerization interfacial region has been suggested to be a possible target for rational drug design against SARS-CoV [14,16,17].

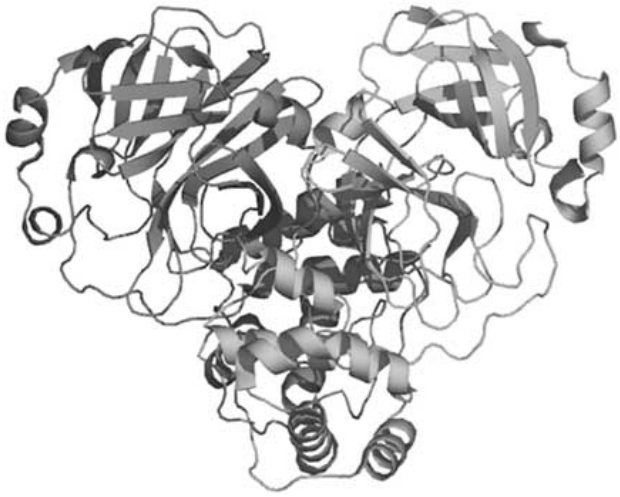

Fig. (1). The solved crystal structure of SARS-CoV 3CL^{pro} dimer.

More detailed structural characters of SARS-CoV 3CL^{pro} can be seen from reference [12].

The SARS-CoV replicase gene is encoded by two large 5'-proximal ORFs, 1a and 1b. Polyproteins ORF1a and ORF1b are connected by a ribosomal frameshift site, resulting in the translation of pp1a and a carboxyl-extended protein pp1ab. As a non-structural protein, SARS-CoV 3CL^{pro} cleaves pp1a and pp1ab precursors at 11 sites to release a

Table 1. Thermodynamic Parameters for SARS-CoV 3CL^{pro}'s Dimerization at 25°C and pH 7.5

SARS-CoV 3CL^{pro}	K_{diss}[a] (μM)	$\Delta H°_{dim}$[a] (kcal/mol)	$\Delta G°_{dim}$[b] (kcal/mol)	$\Delta S°_{dim}$[b] (cal/mol•K)
Full-length	227±34	-8.283±0.103	-4.968±0.083	-11.124±0.163
N-terminal deleted	262±15	-6.893±0.250	-4.879±0.033	-6.758±0.729

[a] From non-linear regression of calorimetric dilution data, fitted by a simple dimer-monomer dissociation model; [b]Calculated from the general thermodynamic equation, $\Delta G°_{dim} = RT \ln K_{diss} = \Delta H°_{dim} — T\Delta S°_{dim}$. Data obtained from Chen *et al.* [15]

series of non-structural proteins, including RNA-dependant RNA polymerase (RdRp), ATPase/helicase and other function-unidentified non-structural proteins (nsp's) [3,18]. These proteins are responsible for replication of the virus

Fig. (2). The dimerization state change of the N-terminal deleted SARS-CoV 3CL^{pro} dimer [redrawn from ref. 15].

Superposition of the N-terminal deleted SARS-CoV 3CL^{pro} dimer (monomers A and B are represented as black, respectively) with the full-length protease dimer (light gray). To clearly demonstrate the dimerization state change, the two monomer Bs of the full-length and N-terminal deleted proteases were superposed each other with the smallest root mean-square deviation (RMSD). The arc arrows represent the rotation direction of monomer A of the N-terminal deleted protease dimer.

genome, thereby producing nested transcripts that are indispensable for the synthesis of viral proteins. Similar to TGEV and HCoV 229E 3CL proteases, SARS-CoV 3CL^{pro} employs Cys-145 and His-41 as the catalytic dyad in the catalytic site but lacks the corresponding third catalytic site [8,16,19,20], which is an acidic residue in chymotrypsin. The substrate specificity of SARS-CoV 3CL^{pro} has been well defined and the cleavage site in polyproteins [(S, V, T, P)-X- (L, I, F, V, M)-Q↓ (S, A, G, N, C)] are highly conserved [18]. The P1 position of the substrate is exclusively occupied by glutamine, the bulky hydrophobic residues (mainly leucine/isoleucine) are dominant at the P2 position and Ser, Val, Thr, Pro are clearly favored at the P4 position. At the P1' position, small aliphatic residues (Ser, Ala, Gly, Asn, Cys) are found, of which the Ser content is more than 50%. There are no highly favored residues at the P3, P2', P3', P4' positions. In addition, the secondary structural studies for the putative substrates of SARS-CoV 3CL^{pro} have revealed that the substrates with more β-sheet like structures tend to be cleaved fast [16]. The *in vitro trans*-cleavage activity of SARS-CoV 3CL^{pro} was extensively determined by the substrate-analog peptide cleavage assay using the conventional RP-HPLC and fluorescence-based methods [16,20-22], *e.g.* the enzymatic activity of the protease at different pH values and temperatures was characterized in detail by fluorescence resonance energy transfer (FRET) techniques [22]. The results showed the substantial pH and temperature-triggered activity switch of SARS-CoV 3CL^{pro}, and the systematic site-directed mutagenesis analyses revealed that substitutions of His^{41}, Cys^{145}, and His^{163} resulted in complete loss of enzymatic activity, while replacement of Met^{162} with Ala exhibited strongly the increase of the enzymatic activity. This work provided valuable information in understanding the catalytic mechanism for SARS-CoV 3CL^{pro}, and this FRET-based assay might supply an ideal approach for SARS-CoV 3CL^{pro} putative inhibitors exploration. Furthermore, the crystal structure of SARS-CoV 3CL^{pro} has revealed considerable pH-dependant conformational changes, which correlates well with the varying *trans*-activities of 3CL^{pro} at different pH values [20, 22]. It is suggested that the reported low *in vitro* activity of SARS-CoV 3CL^{pro} in micromolar level is perhaps due to the low amount of the active dimeric form under the assay conditions, following the prediction that only the dimer is the active form of the protease. Recently, the characterization of *in vivo cis*-cleavage activity of the protease has been also presented by the cell-based *cis*-cleavage assay [23]. It is not clear whether SARS-CoV 3CL^{pro} cleaves itself in *cis* or *trans* from the replicase polyproteins precursor; however, once released, dimerization of SARS-CoV 3CL^{pro} is necessary for its *trans*-cleavage activity. The extensive interactions of one monomer's N-terminus with domain II of the other monomer seem to be the products of autoprocessing of precursor monomeric SARS-CoV 3CL^{pro} by a tailor-made mechanism along with formation of the dimer, which keeps the protease in an conformation suitable for the active site to subsequently act on other cleavage sites in the polyproteins pp1a and pp1ab during viral replication and assembly.

SARS-CoV 3CL^{pro} has supplied an attractive and key target for anti-SARS reagent discovery. According to the 3CL^{pro} structural models from the solved crystal structures of TGEV, HCoV and SARS-CoV, a number of potential in-

hibitors of SARS-CoV $3CL^{pro}$ have been proposed from the collections of synthetic compounds, natural products and approval drugs by using virtual screening [11,24-26]. *In vitro* fluorescence-based and *in vivo* cell-based assays have been developed for putative SARS-CoV $3CL^{pro}$ inhibitors screening [27-30]. To date, the potential inhibitors of SARS-CoV $3CL^{pro}$ identified from these studies include the HIV-1 protease inhibitors L-700 and 417 [24], the reverse transcriptase inhibitors calanolide A and nevirapine [31], the α-glucosidase inhibitor glycovir [31], the general antiviral ribavirin [31], an natural product sabadinine [25], KZ7088, a derivative of AG7088 (a drug to treat the common cold) [32], bifunctional aryl boronic acid compounds [28], the mercury-containing compounds thimerosal, phenylmercuric acetate and hexachlorophene, as well as zinc-conjugated compounds [33]. In addition, cinanserin, a well-characterized serotonin antagonist that has undergone preliminary clinical testing in humans in the 1960s, was also recently discovered by L. Chen, *et al.*, as one potent inhibitor of SARS-CoV $3CL^{pro}$ [34]. The extensively experimental evaluations demonstrated that both cinanserin and its hydrochloride could inhibit the bacterially expressed SARS-CoV $3CL^{pro}$ and HCoV 229E $3CL^{pro}$ by IC_{50} around 5μM (Fig. **3**). Antiviral activities for cinanserin and its hydrochloride were also evaluated in tissue culture assays, and the results revealed their strong inhibitions against coronavirus replication at nontoxic drug concentrations (Fig. **4**). This study has provided a new strategy in identifying new pharmacological activities for the known drugs [34].

Spike (S) protein

Coronavirus Spike (S) protein is a type I membrane glycoprotein that has an N-terminal ectodomain, a C-terminal hydrophobic anchor and an unusual cysteine-rich domain bridging the putative junction of the anchor and the cytoplasmic tail [8]. On the membrane-bound polysomes, S protein is translated and inserted into rough endoplasmic reticulum (RER), cotranslationally glycosylated, and transported to the Golgi complex. During the transport, S proteins are incorporated onto the maturing virus particles, which assembly and bud into a compartment that lies between the RER and Golgi. Virions are carried from Golgi to plasma membrane in the secretory vesicles, and are released from cells when virion-containing vesicles fuse with plasma membrane. The excess S proteins that are not incorporated onto virus particles are transported to the surface of plasma membrane. S protein can interact with a cellular receptor and mediate membrane fusion to allow viral entry into the susceptible target cells [35].

SARS-CoV S protein is 1255 amino acids long glycoprotein. It is predicted to possess a 13 amino acid signal peptide at the N-terminus, a single ectodomain (1182 amino acids) and a transmembrane region followed by a short cytoplasmic tail (28 residues) at the C-terminus [4] (Fig. **5A**). SARS-CoV S protein is translated as a large polypeptide, which is subsequently cleaved by virus-encoded or host-encoded proteases to produce two functional subunits, S1 and S2 (Fig. **5A**). S1 is known to be the peripheral fragment and S2 is the membrane-spanning fragment. Both the S1 and S2 subunits appear to cause cell fusion when expressed individually, this suggests that they are biologically fully active [8].

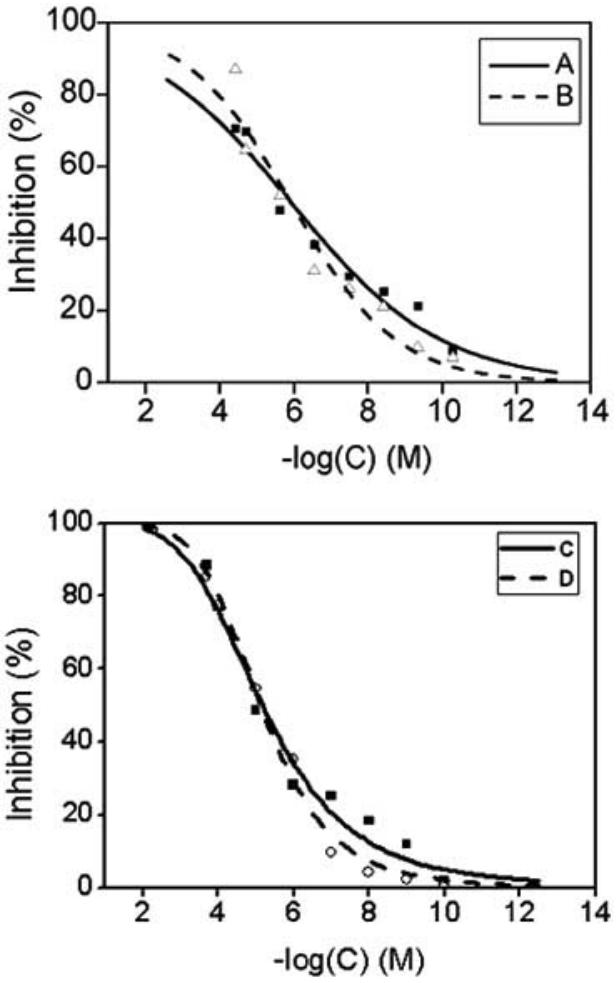

Fig. (3). Inhibitory activity of cinanserin and its hydrochloride on the proteolytic activity of SARS-CoV (A and B) and HCV 229E $3CL^{pro}$ (C and D).

Inhibition of cleavage was measured by FRET using a peptide substrate labeled with a pair of fluorogenic dyes. (A, C cinanserin; B, D cinanserin hydrochloride). Data obtained from Chen *et al.* [34].

The S proteins of coronaviruses including SARS-CoV, associate with cellular receptors to mediate infection against their target cells. Recently, Li *et al.* [36] used a straightforward coimmunoprecipitation method to precipitate the virus attachment S protein with lysates from Vero E6 cells that are susceptible to virus infection, and the coimmunoprecipitated proteins were analyzed by mass spectrometric analysis. The results showed that a metallopeptidase isolated from the SARS-CoV infected Vero E6 cells, called angiotensin-converting enzyme 2 (ACE2), could efficiently bind to the S1 domain of SARS-CoV S protein, and the soluble ACE2 form could block the association of S1 domain with Vero E6 cells. These results thereby indicate that ACE2 is a func-

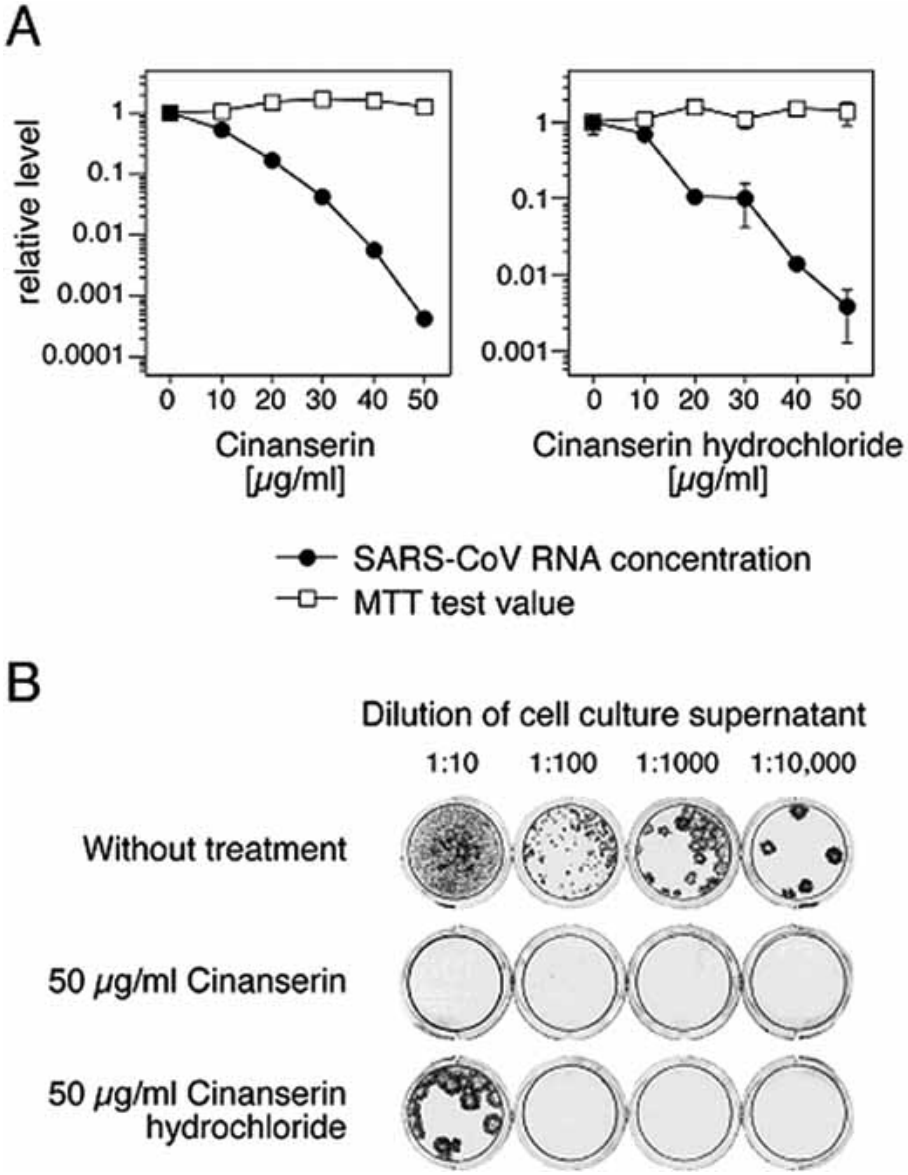

Fig. (4). Inhibition of SARS-CoV replication by cinanserin and cinanserin hydrochloride.

(A) Reduction of SARS-CoV RNA concentration in supernatant. Vero cells were infected with SARS-CoV at an MOI of 0.01 and virus RNA concentration was measured by real-time PCR. Influence of the compounds on cell viability was measured by MTT test. Virus RNA measurement and MTT test were performed with cells and supernatant, respectively, of the same cell culture well. Virus RNA concentration of untreated cells (4×10^7 RNA copies/ml) and the corresponding MTT value were defined as 1. Mean and range of duplicate testings are shown.

(B) Reduction of the concentration of infectious particles in supernatant. Supernatant of infected cells treated with 50 μg/ml compounds and of cells that were left untreated were harvested 2 days post infectionem and the virus titer was determined by immunofocusassay. The cell culture wells of the immunofocusassay inoculated with dilutions of the supernatant are shown. Data obtained from Chen *et al.* [34].

tional receptor of SARS-CoV [36,37]. It is found that the smaller S protein fragment of residues 327-510 or 318-490 did not bind to ACE2, and the 193-amino acid fragment (residues 318-510) could bind to ACE2 more efficiently than the full S1 domain (residues 12-672)[38]. Further binding inhibition assay suggested that the receptor-binding domain of SARS-CoV S protein is located between amino acid residues 303 and 537 [39]. Direct binding mapping experiment revealed that amino acids 270 to 510 of SARS-CoV S protein are required for the interaction of S protein with the receptor of the target cell [40]. The point mutation at Asp454 might abolish the association of the full S1 domain and the 193-residue fragment with ACE2, this demonstrates that Asp454 is critical for S protein interaction with ACE2 [38]. The fact that ACE2 acts as a partner of SARS-CoV S protein in mediating virus entry and cell fusion suggested that S protein might work as a major antigenic determinant that induces neutralizing antibody. The recombinant S protein could exhibit the antigenicity and ACE2-binding ability, and it should become a good candidate for further SARS vaccine

development and anti-SARS therapy [41]. Sui *et al.* [42] found that 80R human monoclonal antibody may be a useful viral entry inhibitor for the emergency prophylaxis and treatment of SARS, and that the ACE2-binding site of S1 domain could be an attractive target for subunit vaccine and drug development. The siRNAs could also effectively and specifically inhibit the gene expression of S protein in SARS-CoV-infected cells [43], which indicates that the interference of S protein expression could be a powerful tool for SARS-CoV inhibition. Yang *et al.* [44] found in a mouse model that a DNA vaccine encoding SARS-CoV S protein may induce T cell and neutralizing antibody responses and protective immunity. In addition, the soluble ACE2 and various fusion constructs or fragments of ACE2 could also serve as potent inhibitors for SARS-CoV infection *in vivo* [45]. In considering the essential role of the immunological fragment of the S protein (Ala251-His641, SARS_S1b) in SARS-CoV entering the host cells, the thermally induced and GuHCl-induced unfolding features of SARS_S1b was quantitatively characterized by Yu *et al*. [46], which revealed that the secondary structure of SARS_S1b has a relatively high thermal stability. Moreover, the secondary and three-dimensional structural predictions by homology modeling indicated that SARS_S1b folds as a globular-like structure by β-sheets and loops and two of the totally four tryptophans were located on the protein surface [46]. The percentage of α-helix (3%) and β-sheet (35%) for SARS_S1b obtained from the secondary structure prediction in this study is extremely close to the data reported by Spiga *et al.* [47], and the 3D model suggested that SARS_S1b is most likely to be an all-β-sheet globular protein, coinciding with the reported result that coronavirus S1 protein is the globular part of S protein [48]. Moreover, the three-dimensional model could be also used to explain the published results of S1/ACE-2 binding and immunizations, and afford a possible platform for further biological study and drug discovery targeting the S protein.

Two regions within S2 domain of SARS-CoV, named heptad repeat 1 and 2 (HR1, residues 892-1013 and HR2, residues 1145-1194) regions, exhibit a high degree of sequence conservation with other coronaviruses. HR1 and HR2 associate with each other into an antiparallel six-helix bundle revealed by the resolved crystal structure (Fig. **5B**) [49], with structural features typical of the other known class I fusion proteins [50,51]. The boundaries of the HR1/HR2 interaction are mapped to residues 896–972 in HR1 and residues 1142–1188 in HR2, possibly extending a few more residues at the N terminus of HR1 (up to residue 889) and the C terminus of HR2 (up to residue 1193, which is the predicted boundary of the S2 transmembrane domain); and these boundaries would position the fusion peptide in the region of residues 870–890 [51]. The HR1 domain of SARS-CoV S protein forms a continuous helical structure, which can be also fully formed in the absence of any interaction with the HR2 domain. This result is unlike HIV gp41, in which the N-peptides are unable to form a trimeric-coiled coil in the absence of the C-peptide [52], and is more similar to the influenza virus hemagglutinin [51]. After binding to the target cell, the transmembrane spike might change conformation by association between the HR1 and HR2 regions to form an oligomeric structure, leading to the fusion between the viral and target-cell membranes [50]. In addition, the crystal structure of S2 domain of SARS-CoV S protein [49] revealed that the fusion core is a six-helix bundle in which three HR2 helices pack against the hydrophobic grooves on the surface of central coiled coil formed by three parallel HR1 helices in an oblique antiparallel manner (Fig. **5B**). This structure shares significant similarity with the fusion core structure of mouse hepatitis virus (MHV) S protein and other viral fusion proteins, which suggests a conserved mechanism of membrane fusion in coronaviruses. The drug discovery strategy aimed at inhibiting viral entry by blocking hairpin formation that has been successfully used in human immunodeficiency virus 1(HIV-1) inhibitor development might be applicable to SARS-CoV inhibitor exploration on the basis of the identified structural information [49].

Nucleocapsid (N) Protein

The nucleocapsid (N) protein of SARS-CoV is a structural protein that primarily functions in recognizing a stretch of RNA that serves as packaging signal and leads to the formation of the ribonucleoprotein (RNP). It wraps the genomic RNA segment into a helical nucleocapsid that is further compacted into a core possibly with icosahedral symmetry. SARS-CoV N protein is a highly charged basic protein with 422 amino acids (range for other coronaviruses, 377 to 454) including seven successive hydrophobic residues near the middle of the protein. It shares a very low homology with the N proteins of other coronaviruses [1].

The major biochemical and thermodynamic features of SARS-CoV N protein was quantitatively studied by using the relevant biophysical and biochemical techniques [53]. The results showed that the recombinant SARS-CoV N protein is easy to denature at low temperature and with low concentration of denaturant, which indicates that this structural protein has low stability that might be critical for SARS-CoV function and stability [53,54]. SARS-CoV N protein tends to form oligomer *in vitro*, more probably dimer at low concentration, and it is suggested that the dimeric N protein might act as a basic functional unit *in vivo*. It has been known that N-N self-interactions may be necessary for subsequent formation of the nucleocapsid and assembly of the viral particles [55]. Noticeably, SARS-CoV N protein self-association may be important for initiation of RNP formation. As for the fragments involved in the formation of dimer or oligomer, however, there are two different opinions. He *et al.* [55] reported that a serine/arginine-rich (SR-rich) motif (SSRSSSRSRGNSR) between residues 184 and 196 is crucial for SARS-CoV N protein oligomerization since deletion of this region could completely abolish SARS-CoV N protein self-multimerization. Nevertheless, Surjit *et al.* [56] demonstrated that the C-terminal 209 amino-acid region constitutes the interaction domain responsible for self-association of SARS-CoV N protein to form dimers. While a recent work [57] has provided an incongruent result, which demonstrates that the C-terminal domain of SARS-CoV N protein (residues 283-422) has multimeric ability, although the full-length protein tends to form dimers. Further study revealed that the multimeric ability of the C-terminal domain could by weakened by the SR-rich motif

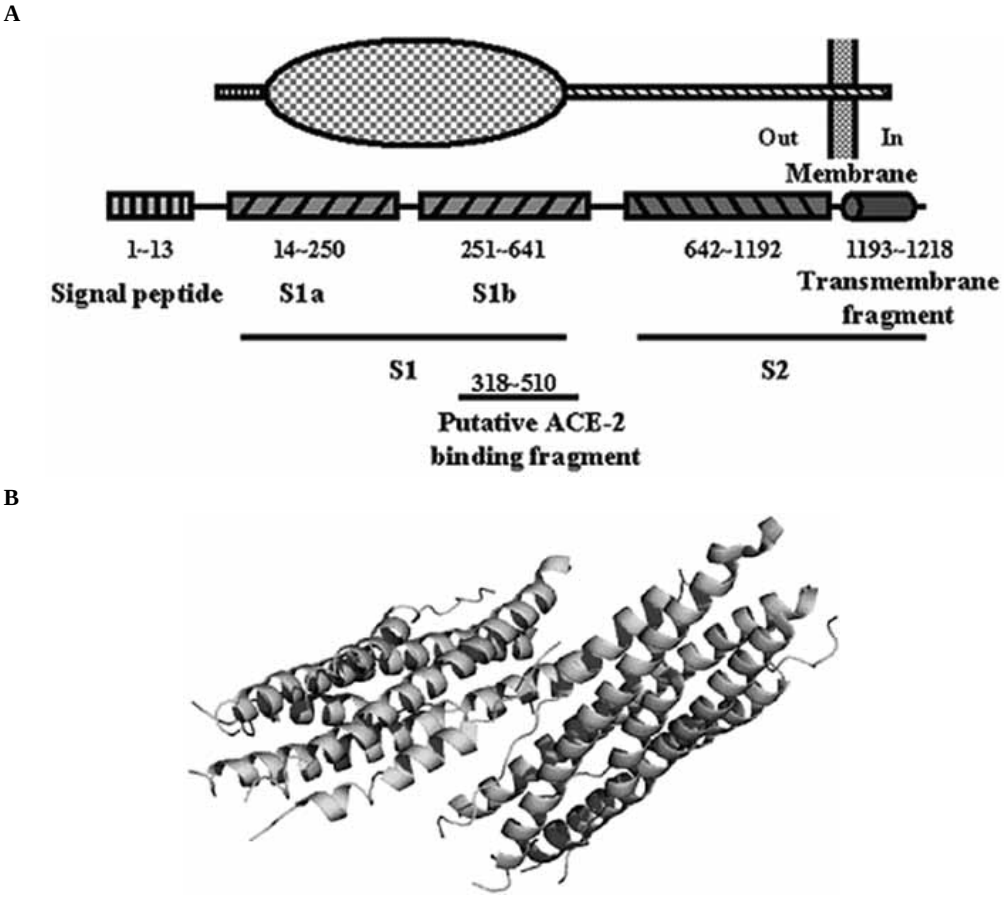

Fig. (5). The predicted overall structure of Spike (S) protein of SARS-CoV [redrawn from ref. 46].

(A). Structurally SARS-CoV S protein is predicted to have a 13 amino acid signal peptide at the amino-terminus, a single ectodomain (1182 amino acids) and a transmembrane region followed by a short cytoplasmic tail (28 residues) at the carboxyl-terminus. As being translated as a large polypeptide, the protein can be subsequently cleaved by virus-encoded or host-encoded proteases to produce two functional subunits, S1 and S2.

(B). The crystal structure of SARS-CoV S protein fusion core [redrawn from ref. 49].

interaction with the central region (residues 211-290), and suggested that the SR-rich motif might play an important role in the transformation of SARS-CoV N protein between the dimer and multimer for self-association or dissociation. Therefore, more attention should be paid in order to elucidate the oligomerization mechanism for SARS-CoV N protein considering these published conflicting results.

Recently, Luo *et al*. [58] reported that SARS-CoV N protein binds tightly to human cyclophilin A (CypA). CypA has ever been shown to play an important role in HIV infection. The Gag polyprotein of HIV-1 binds to most members of the cyclophilin family of peptidyl-peptide-prolyl isomerases [59]. However, of the 15 known human cyclophilins, only human CypA is integrated inside the viral core of HIV-1 by interacting specifically with the capsid domain (CA) of the Gag polyprotein [60], and human CypA subsequently performs an essential function in HIV-1 replication. Luo's result [58] indicated that SARS-CoV N protein has a binding affinity to human CypA by the equilibrium dissociation constant (K_D) ranging from 6 to 160nM. The probable binding sites of these two proteins were detected by modeling the three-dimensional structure of the SARS-CoV N/human CypA complex, from which the important interaction residue pairs between the proteins were deduced (Fig. **6**). Mutagenesis experiments further validated the binding model. Such presently observed SARS-CoV N/human CypA interaction model might provide a new hint for facilitating the understanding of another possible SARS-CoV infection pathway against the human target cell.

In addition, Luo *et al*. [61] also discovered another binding partner for SARS-CoV N protein. They found that SARS-CoV N protein exhibited high binding affinity against human heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1), which is related to the pre-mRNA splicing in the nucleus and translation regulation in the cytoplasm [62]. The results clearly demonstrated that SARS-CoV N protein could directly and specifically bind to human hnRNP A1 *in vitro* (Fig. **7**), and further *in vivo* yeast two-hybrid assays (Fig. **8A** and **8B**) indicated that such binding relates to the fragment (residues 161-210) of SARS-CoV N and the Gly-rich do-

main (residues 203-320) of hnRNP A1 [61]. It is suggested that both SARS-CoV N and hnRNP A1 proteins are possibly within the SARS-CoV replication/transcription complex and SARS-CoV N/human hnRNP A1 interaction might function in the regulation of SARS-CoV RNA synthesis.

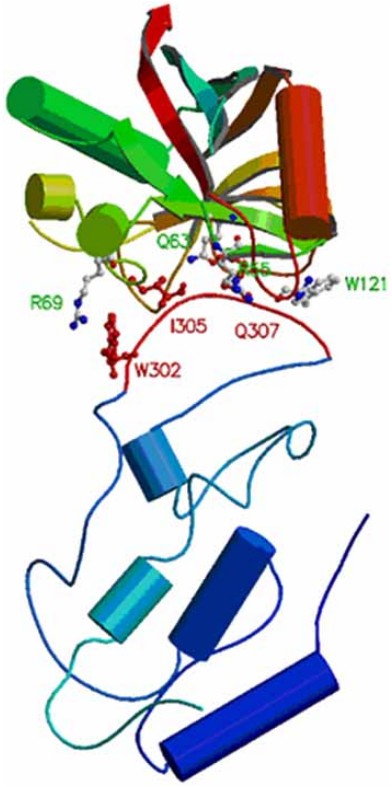

Fig. (6). Schematic representation of the 3D model of SARS-CoV N protein (residues 235-369)-human CypA complex [re-drawn from ref. 58]

The binding loop Trp302-Pro310 of SARS-CoV N protein was highlighted in red color. Resides shown in ball-and-stick model are essential in the SARS-CoV N-human CypA binding, and site-directed mutagenesis analyses were performed on these residues to validate this binding model. See [ref. 58] for further details.

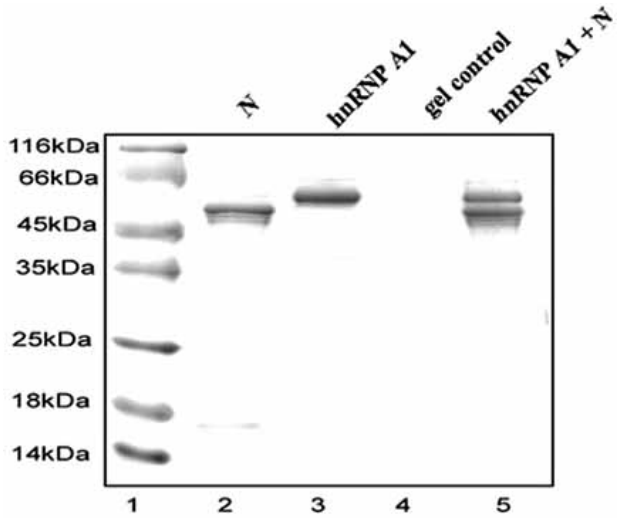

Fig. (7). SARS-CoV N protein/human hnRNP A1 interaction determined by GST pull-down Data obtained from Luo *et al.* [61].

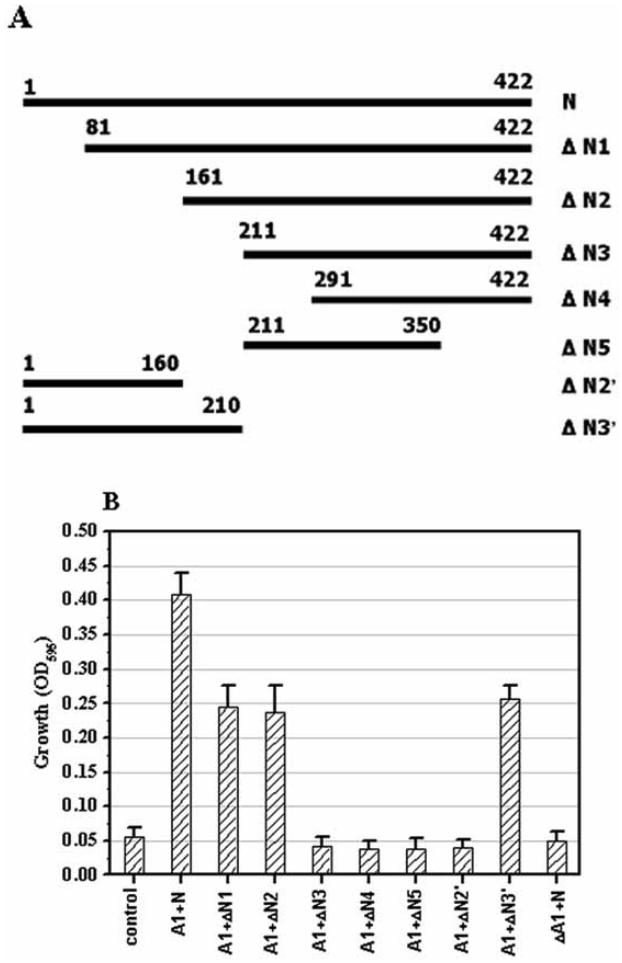

Fig. (8). Mapping the interaction domain of SARS-CoV N protein with human hnRNP A1.

(A) Schematic description of the truncated fragments and (B) the yeast two-hybrid assay results for SARS-CoV N protein/human hnRNP A1 interactions in their truncated and non-truncated forms. The empty vectors pGBKT7 and pGADT7 co-transformed were used as the negative control. Data obtained from Luo *et al.* [61].

A high affinity interaction has been reported to exist between the N protein and leader/intergenic RNA sequence in MHV[63]. The MHV N protein contains two RNA-binding domains that interact with the 3' RNA, one is located in the amino and the other in the carboxyl terminal regions. These two domains are part of the three conserved regions in the N proteins of coronaviruses. The smallest MHV N protein fragment that retains RNA-binding activity is a 55 amino acids segment containing residues 177-231, and AAUCYYA was identified to be the potential minimum ligand for MHV N protein [63]. Actually, SARS-CoV N

protein shares a high homology with MHV N protein regarding this domain, thus SARS-CoV N protein is proposed to interact with RNA through the same region even though there is little literature focusing on it until now. The structure of N-terminal RNA-binding domain (residues 49-178) of SARS-CoV N protein has been determined recently by nuclear magnetic resonance (NMR) spectroscopy [64]. The result showed that this domain consists of a five-stranded antiparallel beta-sheet with a folding topology distinct from other RNA-binding proteins (Fig. **9**). Even though the overall folding is very different from any known protein, the N-terminal domain of SARS-CoV N protein exhibits some similarities to the typical RNP-binding motif. The central strand of the beta-sheet, β3, contains the highly conserved sequence RWYFYYLGT, which is found in other reported coronaviruses N proteins [1,4]. The RNA-binding site of SARS-CoV N protein contains a high content of positively charged lysine and arginine residues on its surface, and the long flexible β hairpin with its positively charged surface may grasp RNA against the β sheet similar to the case in U1A RNP RNA-binding protein, where a highly positively charged loop between β2 and β3 and the face of the β-sheet are involved in RNA binding [64]. It is reported that the compound 6-amino-4-hydroxy-naphthalene-2-sulfonic acid, which might mimic the phosphate backbone of RNA, could interact with this arginines and/or lysines region of SARS-CoV N protein, thus serving as one possible lead compound that tightly binds to SARS-CoV N protein and subsequently inhibits its biological function [64].

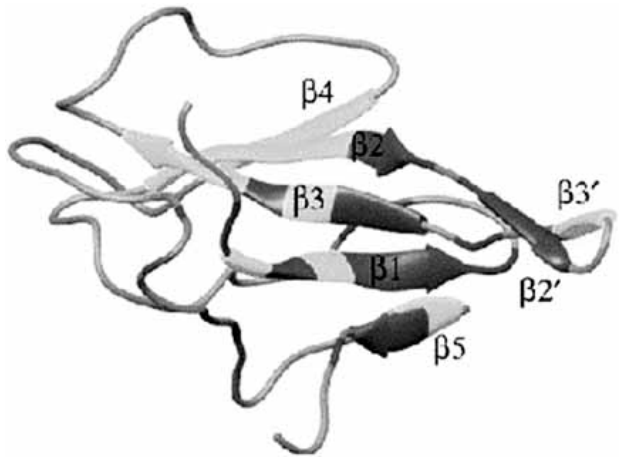

Fig. (9). The N-terminal RNA-binding domain (residues 49-178) of SARS-CoV N protein determined by NMR spectroscopy [redrawn from ref. 64].

See [ref. 64] for further details about the structure of the N-terminal RNA-binding domain of SARS-CoV N protein.

Membrane (M) Protein

The N-terminal domains of coronavirus membrane (M) proteins are localized on the viral surface and the C-terminus is exposed to the interior face of the virion. Furthermore, the predicted M protein of SARS-CoV contains a hydrophobic transmembrane domain (residues 12 to 37), which is similar to the membrane (M) proteins of other coronaviruses. A highly conserved amino acid sequence SMWSFNPE, immediately following this hydrophobic domain, exists in the SARS-CoV M protein [1].

The coronavirus M protein is the most abundant structural protein in virions and the key component in viral assembly and morphogenesis [65,66]. The N/M interactions that might be responsible for the encapsidation of the viral nucleocapsid into the budding virions were reported *in vitro* for MHV [67] and TGEV [68], and it was identified that the carboxyl terminuses of both M and N proteins are involved in this interaction. By employing mammalian two-hybrid system, He *et al.* [69] investigated the N/M protein-protein interactions in SARS-CoV. The results identified that the SARS-CoV N/M interaction does take place *in vivo* and a stretch of amino acids (residues 168–208) in SARS-CoV N protein may be critical for such protein–protein interactions. The same region has also been found to be required for SARS-CoV N protein multimerization [55], which suggests that this region might be crucial in maintaining correct conformation of SARS-CoV N protein for self-interaction and interaction with the M protein. While Luo *et al.* [70] reported a conflicting result that the C-terminal region (residues 351-422) of SARS-CoV N protein and the C-terminus (residues 197-221) of M protein might be involved in their interactions by using yeast two-hybrid and surface plasmon resonance techniques. Sequence analysis revealed that these two fragments are highly charged at neutral pH, suggesting that their interactions might be of electrostatic attraction.

Meanwhile, the M protein is also involved in the assembly of virus-like particle (VLP) together with the small envelope (E) protein [71], which would be vital for assembly and budding of the virions. In addition, the M proteins of coronaviruses are invariably glycosylated near the N terminus. Group 1 and group 3 coronaviruses are N-glycosylated, whereas those of group 2 viruses are O-glycosylated [72]. The predicted structure of SARS-CoV M protein has an NGT sequence profile near its N-terminus, which suggests that it might be also N-glycosylated at position 4 [1].

Small Envelope (E) Protein

Small envelope (E) proteins are structurally conserved within different coronavirus groups, yet exhibit little sequence similarity among these groups [1,73]. In general, coronavirus E proteins are small proteins (varying in size from 76 to 109 amino acids), with an unusually long hydrophobic stretch (25-30 residues) located in between hydrophilic N and C terminus (~8 and ~ 40 residues respectively). Raamsman *et al.* [74] and Corse and Machamer [75] characterized E proteins of MHV and infectious bronchitis virus (IBV), respectively. Both of those convincing studies revealed that E protein localizes in intracellular membrane with its C-terminal domain extending to cytoplasmic region in the infected cell and in the virion toward the interior. The hydrophobic N-terminal two-third region of the E protein is buried within the membrane. Maeda *et al.* [76] represented a detailed membrane topology model of MHV E protein, in which the E protein spans the membrane twice with its N-terminus near the cytoplasmic side, this indicates that no part of MHV E protein is exposed on the virion exterior or luminally in the infected cells.

SARS-CoV E protein has been successively expressed and purified from *E coli*, and its primary 3D model (Fig. **10**) according to the protein secondary structure prediction was built [77], the results showed that SARS-CoV E protein shares the same overall structural characteristics although only 17%-23% identical with other coronaviridea E proteins. The N-terminus consists of the first 11 amino acids, which seems to integrate into the membrane directly due to the lack of any predicted cleavage site in the SARS-CoV E protein sequence. The hydrophobic transmembrane (TM) segment ranges from residues 12 to 34 and adopts an α-helix conformation inserting into the lipid bilayer well. The two short β-sheets are composed of residues 45 to 51 (β-sheet I) and residues 55 to 61 (β-sheet II), respectively; β-sheet I seems to form hydrogen bonds with the surface of the lipid bilayer. Furthermore, the hydrophilic C-terminal tail (residues 62-76) is predicted to expose to the cytoplasmic side. The unusual length of the hydrophobic segment of SARS-CoV E protein has posed a problem with respect to assigning the topology to the protein. Arbely *et al*. [73] presented a detailed structural model for SARS-CoV E protein, which determined the topology of the protein and the effects upon the lipid bilayer thereof. The results showed that SARS-CoV E protein contains an unusually short palindromic transmembrane helical hairpin around a previously unidentified pseudo-center of symmetry, a structural feature that seems to be unique to SARS-CoV. In addition, the hairpin structure can deform lipid bilayer by increasing their curvature, which might provide a molecular explanation of the vital role for E protein in coronavirus budding.

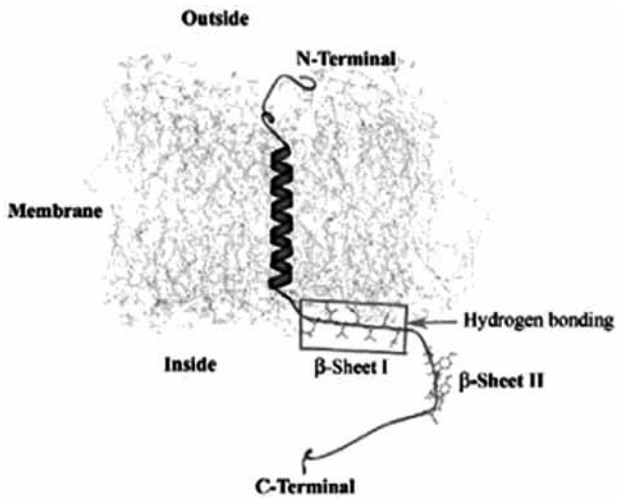

Fig. (10). A primary 3D model of SARS-CoV E protein.

The transmembrane segment was embedded in the POPC lipid bilayer, the two short β-sheets were represented by ball-and-stick model. Adapted from reference [77].

It has been reported that coronavirus E proteins have multiple biological functions. MHV E protein can induce apoptosis in E protein-expressing cells, and this MHV E protein-mediated apoptosis can be suppressed by a high level expression of Bcl-2 oncogene, indicating that the apoptosis pathway is caspase-dependent [78]. Meanwhile, E protein also plays an important role in some coronaviruses replication, such as TGEV [79] and MHV [80]. Another major biological function of E protein is about its pivotal association with coronavirus envelope assembly [71]. The formation of viral core, envelope and the assembly of virus particle would involve the specific interactions between the viral structural proteins and the host membrane components, among the virus structural proteins and between the viral proteins and viral RNAs. The extensive cellular studies on TGEV, MHV and IBV have revealed that coronavirus assembles at the pre-Golgi membranes of the intermediate compartment (IC) early in infection and in the rough endoplasmic reticulum (RER) at late period of the infection [81-83]. Unlike most of the other envelope RNA viruses, coronaviruses employ a nucleocapsid-independent strategy to drive virus particle assembly and budding. Coexpression of coronavirus M protein and E protein in intact cells was initially shown to be required for the production of virus-like particles (VLPs), while expression of M protein alone does not produce VLPs [84]. Furthermore, it was demonstrated that expression of E protein alone results in the release of E protein vesicles [85]. The crucial role of E protein in viral envelope assembly was also indicated by some other studies on MHV and TGEV [71, 86]. Although the studies of different coronavirus groups showed that possession of the M and E proteins is a minimal requirement for the assembly of viral particles, the demonstration that E and M proteins are sufficient for the assembly of SARS-CoV envelope is still imperative for the rational treatment of this deadly virus. A communication [87] reported the successful formation and isolation of SARS-CoV-like particles through the co-expression of E and M proteins in insect cells and these virus-like particles (VLPs) can further incorporate S proteins. Additionally, coronavirus E proteins are candidate members of virus-encoded proteins that form ion channels in considering their small size and hydrophobic nature. Wilson *et al*. [88] demonstrated that SARS-CoV E protein does form ion channels in planar lipid bilayer, and these channels are more selective for monovalent cations than monovalent anions. Meanwhile a recent study reported that SARS-CoV E protein could induce modification of cell membrane permeability [89], thereby favoring virus replication and promoting the translation of viral mRNAs during the virus infection due to the alterations of ion concentration in the cytoplasm of virus-infected cells. This result has evidently supported a critical role of E protein for the replication of coronaviruses.

Non-Structural Proteins (nsp's)

Analogous to other coronaviruses, the 5'-terminal 2/3 of the SARS-CoV genome involves the viral replicase gene (ORFS 1a and 1b), which encodes two large polyproteins, pp1a and pp1ab. Expression of pp1ab involves ribosomal frameshifting into the –1 frame just upstream of the ORF 1a translation termination codon. Proteolytic processing of these polyproteins are usually mediated by virus-encoded proteases thus producing a series of non-structural proteins (nsp), some of which are responsible for replication of the viral genome and transcription of a nested set of subgenomic mRNAs (sg mRNA) to express all the ORFs downstream of ORF1b [8]. SARS-CoV is predicted to have two proteases

for polyproteins processing, a papain-like cysteine protease (termed as $PL2^{pro}$, represented by nsp3) and a 3C-like protease (termed as $3CL^{pro}$ or M^{pro}, represented by nsp5)[5]. Unlike other coronaviruses, SARS-CoV does not have an ortholog of another papain-like protease ($PL1^{pro}$). The $PL2^{pro}$ and $3CL^{pro}$ of SARS-CoV are predicted to cleave the polyproteins at 3 and 11 sites respectively thus totally generating 16 non-structural proteins [5,90,91].

The structures and potential functions of several non-structural proteins of SARS-CoV have been characterized. The solved crystal structure of SARS-CoV nsp9 (Fig. **11**) showed that nsp9 is a single-stranded RNA-binding protein with a novel oligosaccharide/oligonucleotide fold-like structure [92]. In detail, the structure of SARS-CoV nsp9 has a compact corn-shaped β–barrel comprising seven β–strands and one α–helix, flanked by a C-terminal α-helix and N-terminal extension. Dynamic light scattering and analytical ultracentrifugation experimental results indicated that nsp9 exists as homodimer in solution at millimolar concentration grade [93], which is probably biologically important given the hydrophobic nature of the interacting surface and the conservation of amino acids involved in dimerization. Due to the crucial role of RNA-dependent RNA polymerase in the virus cycle and the success of polymerase inhibitors in the treatment of various viral infections [94,95], the homology model of SARS-CoV RdRp, represented by nsp12, was recently built for charactering the potential polymerization mechanism of SARS-CoV and putative inhibitor design (Fig. **12**) [96]. The result showed that the SARS-CoV nsp12 folds in a classic "right hand" shape with motifs in which the characteristic finger, palm and thumb sub-domains can be recognized. The finger domain consists of three helices, the RdRp unique helix-loop-helix and two stranded β–sheets. The palm domain is composed of two helices and the β-hairpin structure, which involves the two catalytic aspartates responsible for the nucleotide transfer reaction. The thumb domain consists essentially of two α-helices and a large loop, and the thumb and finger domains are interconnected with loops. The active center is buried in the center of the protein. Since SARS-CoV RdRp plays a pivotal role in virus replication and its inhibition will not cause undesirable side effects during therapy, it has been regarded as an excellent target for anti-SARS drug discovery. Based on the three-dimensional model, potential nucleoside analog and non-nucleoside inhibitors of SARS-CoV RdRp have been proposed by molecular docking [97].

Another non-structural protein that is likely to be important for viral replication is the SARS-CoV helicase (also called nsp13). Both Tanner *et al.* [98] and Ivanov *et al.* [99] reported that recombinant SARS-CoV nsp13 has multiple enzymatic activities, including both RNA and DNA duplex-unwinding activities with a 5' to 3' polarity, NTPase, dNTPase, and an RNA 5'-triphosphatase activity. These reported functional and structural properties of the SARS-CoV helicase may provide an opportunity for selective inhibitors design and the future development of suitable helicase-targeted anti-SARS drugs, considering that promising NTPase/helicase inhibitors have been tested for the treatment of herpes simplex virus [100,101] and hepatitis C viral (HCV) infections [102]. For the remaining SARS-CoV non-structural proteins produced from polyproteins pp1a and pp1ab, putative activities have been predicted based on the presence of functional domains in their sequences or by their structural similarities to other proteins, *e.g.* the N-terminal domain of SARS-CoV nsp14 has been predicted to be a 3'-to-5' exonuclease (ExoN) that belongs to the DEDD subfamily [90], a "nidovirus-specific" replicase domain in the C-terminal part of nsp15 is in association with a poly (U)-specific endoribonuclease (XendoU) [90], and SARS-CoV nsp16 has been confirmed as 2'-O-methyltransferase (2'-O-MT), which is supported by the presence of the conserved tetrad of residues K-D-K-E essential for mRNA cap-1 (mGpppNm) formation [103].

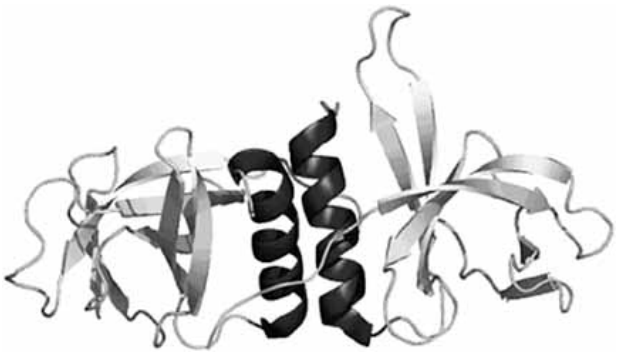

Fig. (11). Ribbon representation of the crystal structure of SARS-CoV nsp9.

The protein has a compact corn-shaped β–barrel comprised of seven β–strands and one α–helix flanking by a C-terminal α-helix and N-terminal extension, and exists as homodimer in the crystal. Adapted from reference [92].

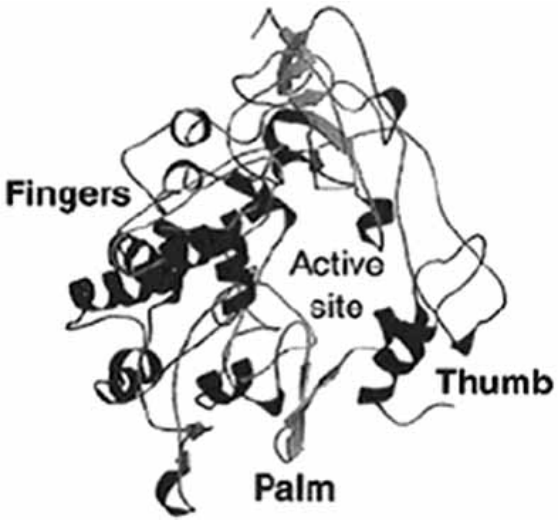

Fig. (12). Ribbon diagram of SARS-CoV RNA-dependent RNA polymerase (nsp12) [redrawn from ref. 96].

The characteristic fingers, palm and thumb domains can be recognized from the "right-hand"-like fold structure.

CONCLUSIONS

A highly contagious illness called severe acute respiratory syndrome (SARS) ever broke out in China and quickly

spread to more than 30 other countries almost two years ago. The rapid transmission and the high mortality rate (nearly 10%) make SARS a global threat. Although the SARS disease has been controlled by conventional measures such as rapid detection, infection control, isolation, quarantine, contact tracing, no efficacious therapy and drugs against SARS are available to date. SARS coronavirus (SARS-CoV) has been identified as the etiological agent responsible for this infectious disease, and SARS-CoV genome is predicted to contain 14 functional open reading frames (ORFs). The two large 5'-terminal ORFs, 1a and 1b, constitute the replicase gene encoding the proteins complex (*e.g.* the genome-encoded proteases and a series of non-structural proteins) required for viral RNA transcription and replication, and the remaining twelve ORFs encode the four key structural proteins, the spike (S) protein, the nucleocapsid (N) protein, the membrane (M) protein and the small envelope (E) protein, and several accessory proteins. Characterization of the structures and functions of these major proteins involved in SARS-CoV has become an imperative project, which can afford us an overall picture for understanding the assembly and maturation mechanism of SARS-CoV and exploring the feasible therapy and drugs for the treatment of this disease. Actually, extensive studies concerning these pivotal proteins encoded by SARS-CoV genome have been performed. This review gives a brief summary of the recently substantial scientific progress that has been made towards understanding the structural and functional characters of major proteins involved in SARS-CoV and the corresponding drug development. These studies would help us understand the molecular mechanism of SARS-CoV replication and infection, and provide valuable information for rational treatment with this disease.

ACKNOWLEDGEMENTS

This work was supported by Shanghai Basic Research Project from the Shanghai Science and Technology Commission (grant 02DJ14070, 03DZ19212, 03DZ19228), the National Natural Science Foundation of China (grants 2037 2069, 30525024), the State Key Program of Basic Research of China (grants 2004CB518905, 2002CB512802, 2002CB 512807), the 863 Hi-Tech Program (grants 2002AA233011, 2005AA233011, 2002AA104270), Sino-European Project on SARS Diagnostics and Antivirals (Proposal/Contract no.003831), and the special programs of oppugning SARS from the Ministry of Science and Technology, Chinese Academy of Sciences, National Natural Science Foundation of China and Shanghai Science and Technology Commission.

ABBREVIATIONS

SARS-CoV	=	Severe acute respiratory syndrome coronavirus
TGEV	=	Transmissible gastroenteritis coronavirus
HCoV	=	Human coronavirus
MHV	=	Mouse hepatitis virus
IBV	=	Infectious bronchitis virus
$3CL^{pro}$	=	3C-like protease
S protein	=	Spike glycoprotein
ACE2	=	Angiotensin-converting enzyme 2
N protein	=	Nucleocapsid protein
CypA	=	Cyclophilin A
hnRNP A1	=	Heterogeneous nuclear ribonucleoprotein A1
M protein	=	Membrane protein
VLP	=	Virus-like particle
E protein	=	Small envelope protein
nsp	=	Non-structural protein
RdRp	=	RNA-dependent RNA polymerase

REFERERNCES

References 104-106 are related articles recently published in Current Pharmaceutical Design.

[1] Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, *et al*. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003; 300: 1394-1399.

[2] Siddel SG. The Coronaviridae. New York: Plenum Press 1995.

[3] Ziebuhr J. Molecular biology of severe acute respiratory syndrome coronavirus. Curr Opin Microbiol 2004; 7: 412-419.

[4] Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, *et al*. The Genome sequence of the SARS-associated coronavirus. Science 2003; 300: 1399-1404.

[5] Thiel V, Ivanov KA, Putics A, Hertzig T, Schelle B, Bayer S, *et al*. Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol 2003; 84: 2305-2315.

[6] Ziebuhr J, Heusipp G, Siddell SG. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. J Virol 1997; 71: 3992-3997.

[7] Dougherty WG, Semler BL. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol Rev 1993; 57: 781-822.

[8] Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol 2000; 81: 853-879.

[9] Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase ($3CL^{pro}$) structure: basis for design of anti-SARS drugs. Science 2003; 300: 1763-1767.

[10] Sun HF, Luo HB, Yu CY, Sun T, Chen J, Peng SY, *et al*. Molecular cloning, expression, purification, and mass spectrometric characterization of 3C-like protease of SARS coronavirus. Protein Exp Purif 2003; 32: 302-308.

[11] Xiong B, Gui CS, Xu XY, Luo C, Chen J, Luo HB, *et al*. A 3D model of SARS-CoV 3CL proteinase and its inhibitors design by virtual screening. Acta Pharmacol Sin 2003; 24: 497-504.

[12] Yang HT, Yang MJ, Ding Y, Liu YW, Lou ZY, Zhou Z, *et al*. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci USA 2003; 100: 13190-13195.

[13] Anand K, Plam GJ, Mesters JR, Siddell SG, Ziebuhr J, Higenfeld R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J 2002; 21: 3213-3224.

[14] Shi JH, Wei Z, Song JX. Dissection study on the SARS 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: Defining the extra domain as a new target for design of highly-specific protease inhibitors. J Biol Chem 2004; 279: 24765-24773.

[15] Chen S, Chen LL, Tan JZ, Chen J, Du L, Sun T, *et al*. SARS Coronavirus $3CL^{pro}$ N-terminus is Indispensable for Proteolytic Activity but Not for Enzyme Dimerization: Biochemical and

Thermodynamic Investigation in Conjunction with Molecular Dynamics Simulations. J Biol Chem 2005; 280: 164-173.

[16] Fan KQ, Wei P, Feng Q, Chen SD, Huang CK, Ma L, *et al*. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. J Biol Chem 2004; 279: 1637-1642.

[17] Chou CY, Chang HC, Hsu WC, Lin TZ, Lin CH, Chang GG. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. Biochemistry 2004; 43: 14958-14970.

[18] Gao F, Ou HY, Chen LL, Zheng WX, Zhang CT. Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. FEBS Lett 2003; 553: 451-456.

[19] Hegyi A, Ziebuhr J. Conservation of substrate specificities among coronavirus main proteases. J Gen Virol 2002; 83: 595-599.

[20] Huang CK, Wei P, Fan KQ, Liu Y, Lai LH. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry 2004; 43: 4568-4573.

[21] Kuo CJ, Chi YH, Hsu JA, Liang PH. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. Biochem Biophys Res Commun 2004; 318: 862-867.

[22] Chen S, Chen LL, Luo HB, Sun T, Chen J, Ye F, *et al*. Enzymatic activity characterization of SARS coronavirus 3C-like protease ($3CL^{pro}$) by fluorescence resonance energy transfer (FRET) technique. Acta Pharmacol Sin 2005; 26: 99-107.

[23] Lin CW, Tsai CH, Tsai FJ, Chen PJ, Lai CC, Wan L, *et al*. Characterization of *trans*- and *cis*-cleavage activity of the SARS coronavirus $3CL^{pro}$ protease: basis for the *in vitro* screening of anti-SARS drugs. FEBS Lett 2004; 574: 131-137.

[24] Jenwitheesuk E, Samudrala R. Identifying inhibitors of the SARS coronavirus proteinase. Bioorg Med Chem Lett 2004; 13: 3989-3992.

[25] Toney JH, Navas-Martin S, Weiss SR, Koeller A. Sabadinine: A Potential Non-Peptide Anti-Severe Acute-Respiratory-Syndrome Agent Identified Using Structure-Aided Design. J Med Chem 2004; 47: 1079-1080.

[26] Zhang XW, Yap YL. Exploring the binding mechanism of the main proteinase in SARS-associated coronavirus and its implication to anti-SARS drug design. Bioorg Med Chem 2004; 12: 2219-2223.

[27] Kao RY, To PC, Ng WY, Tsui HW, Lee SW, Tsoi HW, *et al*. Characterization of SARS-CoV main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay. FEBS Lett 2004; 576: 325-330.

[28] Bacha U, Barrila J, Velazquez-Campoy A, Leavitt SA, Freire E. Identification of novel inhibitors of the SARS coronavirus main protease $3CL^{pro}$. Biochemistry 2004; 43: 4906-4912.

[29] Blanchard JE, Elowe NH, Huitema C, Fortin PD, Cechetto JD, Eltis LD, *et al*. High-Throughput screening identifies inhibitors of the SARS coronavirus main proteinase. Chem Biol 2004; 11: 1445-1453.

[30] Wu CY, Jan JT, Ma SH, Kuo CJ, Juan HF, Cheng YS, *et al*. Small molecules targeting severe acute respiratory syndrome human coronavirus. Proc Natl Acad Sci USA 2004; 101: 10012-10017.

[31] Lee VS, Wittayanarakul K, Remsungnen T, Parasuk V, Sompornpisut P, Chantratita W, *et al*. Structure and dynamics of SARS coronavirus proteinase: the primary key to the designing and screening for anti-SARS drugs. Sci Asia 2003; 29: 181-188.

[32] Sirois S, Wei DQ, Du QS, Chou KC. Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. J Chem Inf Comput Sci 2004; 44: 1111-1122.

[33] Hsu TA, Kuo CJ, Hsieh HP, Wang YC, Huang KK, Lin PC, *et al*. Evaluation of metal-conjugated compounds as inhibitors of 3CL protease of SARS-CoV. FEBS Lett 2004; 574: 116-120.

[34] Chen LL, Gui CS, Luo XM, Yang QG, Gunther S, Scandella E, *et al*. The old drug cinanserin is an inhibitor of the 3C-like proteinase of SARS coronavirus and strongly reduces virus replication *in vitro*. J Virol 2005; 79: 7095-7103.

[35] Han DP, Kim HG, Kim YB, Poon LL, Cho MW. Development of a safe neutralization assay for SARS-CoV and characterization of S-glycoprotein. Virology 2004; 326: 140-149.

[36] Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, *et al*. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 2003; 426: 450-454.

[37] Wang P, Chen J, Zheng A, Nie Y, Shi X, Wang W, *et al*. Expression cloning of functional receptor used by SARS coronavirus. Biochem Biophys Res Commun 2004; 315: 439-444.

[38] Wong SK, Li W, Moore MJ, Choe H, Farzan M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. J Biol Chem 2004; 279: 3197-3201.

[39] Xiao X, Chakraborti S, Dimitrov AS, Gramatikoff K, Dimitrov DS. The SARS-CoV S glycoprotein: expression and functional characterization. Biochem Biophys Res Commun 2003; 312: 1159-1164.

[40] Babcock GJ, Esshaki DJ, Thomas WD, Ambrosino DM. Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. J Virol 2004; 78: 4552-4560.

[41] Ho TY, Wu SL, Cheng SE, Wei YC, Huang SP, Hsiang CY. Antigenicity and receptor-binding ability of recombinant SARS coronavirus spike protein. Biochem Biophys Res Commun 2004; 313: 938-947.

[42] Sui J, Li W, Murakami A, Tamin A, Matthews LJ, Wong SK, *et al*. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. Proc Natl Acad Sci USA 2004; 101: 2536-2541.

[43] Zhang Y, Li T, Fu L, Yu C, Li Y, Xu X, *et al*. Silencing SARS-CoV Spike protein expression in cultured cells by RNA interference. FEBS Lett 2004; 560: 141-146.

[44] Yang ZY, Kong WP, Huang Y, Roberts A, Murphy BR, Subbarao K, *et al*. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. Nature 2004; 428: 561-564.

[45] Dimitrov DS. The secret life of ACE2 as a receptor for the SARS virus. Cell 2003; 115: 652-653.

[46] Yu C, Gui C, Luo H, Chen L, Zhang L, Yu H, *et al*. Folding of the SARS Coronavirus Spike Glycoprotein Immunological Fragment (SARS_S1b): Thermodynamic and Kinetic Investigation Correlating with Three-Dimensional Structural Modeling. Biochemistry 2005; 44: 1453-1463.

[47] Spiga O, Bernini A, Ciutti A, Chiellini S, Menciassi N, Finetti F, *et al*. Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein. Biochem Biophys Res Commun 2003; 310: 78-83.

[48] Kubo H, Takase-Yoden S, Taguchi F. Neutralization and fusion inhibition activities of monoclonal antibodies specific for the S1 subunit of the spike protein of neurovirulent murine coronavirus JHMV c1-2 variant. J Gen Virol 1993; 74(7): 1421-1425.

[49] Xu Y, Lou Z, Liu Y, Pang H, Tien P, Gao GF, *et al*. Crystal Structure of Severe Acute Respiratory Syndrome Coronavirus Spike Protein Fusion Core. J Biol Chem 2004; 279: 49414-49419.

[50] Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, *et al*. 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 2004; 363: 938-947.

[51] Ingallinella P, Bianchi E, Finotto M, Cantoni G, Eckert DM, Supekar VM, *et al*. Structural characterization of the fusion-active complex of severe acute respiratory syndrome (SARS) coronavirus. Proc Natl Acad Sci USA 2004; 101: 8709-8714.

[52] Lu M, Blacklow SC, Kim PS. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. Nat Struct Biol 1995; 2: 1075-1082.

[53] Luo H, Ye F, Sun T, Yue L, Peng S, Chen J, *et al*. *In vitro* biochemical and thermodynamic characterization of nucleocapsid protein of SARS. Biophys Chem 2004; 112: 15-25.

[54] Wang Y, Wu X, Wang Y, Li B, Zhou H, Yuan G, *et al*. Low stability of nucleocapsid protein in SARS virus. Biochemistry 2004; 43: 11103-11108.

[55] He R, Dobie F, Ballantine M, Leeson A, Li Y, Bastien N, *et al*. Analysis of multimerization of the SARS coronavirus nucleocapsid protein. Biochem Biophys Res Commun 2004; 316: 476-483.

[56] Surjit M, Liu B, Kumar P, Chow VT, Lal SK. The nucleocapsid protein of the SARS coronavirus is capable of self-association

through a C-terminal 209 amino acid interaction domain. Biochem Biophys Res Commun 2004; 317: 1030-1036.

[57] Luo HB, Ye F, Chen KX, Shen X, Jiang HL. SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. Biochemistry 2005; 44: 15351-15358.

[58] Luo C, Luo H, Zheng S, Gui C, Yue L, Yu C, *et al*. Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A. Biochem Biophys Res Commun 2004; 321: 557-565.

[59] Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell 1993; 73: 1067-1078.

[60] Braaten D, Luban J. Cyclophilin A regulates HIV-1 infectivity, as demonstrated by gene targeting in human T cells. EMBO J 2001; 20: 1300-1309.

[61] Luo HB, Chen Q, Chen J, Chen KX, Shen X, Jiang HL. The nucleocapsid protein of SARS coronavirus (SARS_N) has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1). FEBS Lett 2005; 579: 2623-2628.

[62] Shi ST, Huang P, Li HP, Lai MM. Heterogeneous nuclear ribonucleoprotein A1 regulates RNA synthesis of a cytoplasmic virus. EMBO J 2000; 19: 4701-4711.

[63] Nelson GW, Stohlman SA, Tahara SM. High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. J Gen Virol 2000; 81: 181-188.

[64] Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, *et al*. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. Biochemistry 2004; 43: 6059-6063.

[65] De Haan CA, Kuo L, Masters PS, Vennema H, Rottier PJ. Coronavirus particle assembly: primary structure requirements of the membrane protein. J Virol 1998; 72: 6838-6850.

[66] De Haan CA, Vennema H, Rottier PJ. Assembly of the coronavirus envelope: homotypic interactions between the M proteins. J Virol 2000; 74: 4967-4978.

[67] Kuo L and Masters PS. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J Virol 2002; 76: 4987-4999.

[68] Escors D, Ortego J, Laude H, Enjuanes L. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. J Virol 2001; 75: 1312-1324.

[69] He R, Leeson A, Ballantine M, Andonov A, Baker L, Dobie F, *et al*. Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res 2004; 105: 121-125.

[70] Luo HB, Wu DL, Shen C, Chen KX, Shen X, *et al*. Severe acute respiratory syndrome coronavirus membrane protein interacts with nucleocapsid protein mostly through their carboxyl termini by electrostatic attraction. Int J Biochem Cell Biol 2006; 38: 589-599.

[71] Baudoux P, Carrat C, Besnardeau L, Charley B, Laude H. Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes. J Virol 1998; 72: 8636-8643.

[72] De Haan CA, De Wit M, Kuo L, Montalto C, Masters PS, Weiss SR, *et al.* O-glycosylation of the mouse hepatitis coronavirus membrane protein. Virus Res 2002; 82: 77-81.

[73] Arbely E, Khattari Z, Brotons G, Akkawi M, Salditt T, Arkin IT. A highly unusual palindromic transmembrane helical hairpin formed by SARS coronavirus E protein. J Mol Biol 2004; 341: 769-779.

[74] Raamsman MJ, Locker JK, De Hooge A, De Vries AA, Griffiths G, Vennema H, *et al*. Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. J Virol 2000; 74: 2333-2342.

[75] Corse E and Machamer CE. Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. J Virol 2000; 74: 4319-4326.

[76] Maeda J, Repass JF, Maeda A, Makino S. Membrane topology of coronavirus E protein. Virology 2001; 281: 163-169.

[77] Shen X, Xue JH, Yu CY, Luo HB, Qin L, Yu XJ, *et al.* Small envelope protein E of SARS: cloning, expression, purification, CD determination, and bioinformatics analysis. Acta Pharmacol Sin 2003; 24: 505-511.

[78] An S, Chen CJ, Yu X, Leibowitz JL, Makino S. Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of E protein as an apoptosis inducer. J Virol 1999; 73: 7853-7859.

[79] Godet M, L'Haridon R, Vautherot JF, Laude H. TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. Virology 1992; 188: 666-675.

[80] Kuo L, Masters PS. The small envelope protein E is not essential for murine coronavirus replication. J Virol 2003; 77: 4597-4608.

[81] Risco C, Muntion M, Enjuanes L, Carrascosa JL. Two types of virus-related particles are found during transmissible gastroenteritis virus morphogenesis. J Virol 1998; 72: 4022-4031.

[82] Salanueva IJ, Carrascosa JL, Risco C. Structural maturation of the transmissible gastroenteritis coronavirus. J Virol 1999; 73: 7952-7964.

[83] Tooze J, Tooze S, Warren G. Replication of coronavirus MHV-A59 in sac- cells: determination of the first site of budding of progeny virions. Eur J Cell Biol 1984; 33: 281-293.

[84] Vennema H, Godeke GJ, Rossen JW, Voorhout WF, Horzinek MC, Opstelten DJ, *et al*. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. EMBO J 1996; 15: 2020-2028.

[85] Maeda J, Maeda A, Makino S. Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells. Virology 1999; 263: 265-272.

[86] Fischer F, Stegen CF, Masters PS, Samsonoff WA. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. J Virol 1998; 72: 7885-7894.

[87] Ho Y, Lin PH, Liu CY, Lee SP, Chao YC. Assembly of human severe acute respiratory syndrome coronavirus-like particles. Biochem Biophys Res Commun 2004, 318: 833-838.

[88] Wilson L, McKinlay C, Gage P, Ewart G. SARS coronavirus E protein forms cation-selective ion channels. Virology 2004; 330: 322-331.

[89] Liao Y, Lescar J, Tam JP, Liu DX. Expression of SARS-coronavirus envelope protein in *Escherichia coli* cells alters membrane permeability. Biochem Biophys Res Commun 2004; 325: 374-380.

[90] Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, *et al*. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 2003; 331: 991-1004.

[91] Tan YJ, Lim SG, Hong W. Characterization of viral proteins encoded by the SARS-coronavirus genome. Antiviral Res 2005; 65: 69-78.

[92] Egloff MP, Ferron F, Campanacci V, Longhi S, Rancurel C, Dutartre H, *et al*. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc Natl Acad Sci USA 2004; 101: 3792-3796.

[93] Sutton G, Fry E, Carter L, Sainsbury S, Walter T, Nettleship J, *et al*. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure (Camb) 2004; 12: 341-353.

[94] Carroll SS, Tomassini JE, Bosserman M, Getty K, Stahlhut MW, Eldrup AB, *et al*. Inhibition of hepatitis C virus RNA replication by 2'-modified nucleoside analogs. J Biol Chem 2003; 278: 11979-11984.

[95] Dhanak D, Duffy KJ, Johnston VK, Lin-Goerke J, Darcy M, Shaw AN, *et al*. Identification and biological characterization of heterocyclic inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. J Biol Chem 2002; 277: 38322-38327.

[96] Azzi A and Lin SX. Human SARS-coronavirus RNA-dependent RNA polymerase: activity determinants and nucleoside analogue inhibitors. Proteins 2004; 57: 12-14.

[97] Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. Nucleic Acids Res 2003; 31: 7117-7130.

[98] Tanner JA, Watt RM, Chai YB, Lu LY, Lin MC, Peiris JS, *et al*. The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases. J Biol Chem 2003; 278: 39578-39582.

[99] Ivanov KA, Thiel V, Dobbe JC, Van der MY, Snijder EJ, Ziebuhr J. Multiple enzymatic activities associated with severe acute respi-

ratory syndrome coronavirus helicase. J Virol 2004; 78: 5619-5632.

[100] Kleymann G, Fischer R, Betz UA, Hendrix M, Bender W, Schneider U, *et al*. New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. Nat Med 2002; 8: 392-398.

[101] Crute JJ, Grygon CA, Hargrave KD, Simoneau B, Faucher AM, Bolger G, *et al*. Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease. Nat Med 2002; 8: 386-391.

[102] Borowski P, Schalinski S, Schmitz H. Nucleotide triphosphatase/helicase of hepatitis C virus as a target for antiviral therapy. Antiviral Res 2002; 55: 397-412.

[103] Von Grotthuss M, Wyrwicz LS, Rychlewski L. mRNA cap-1 methyltransferase in the SARS genome. Cell 2003; 113: 701-702.

[104] Callanan M. Mining the probiotic genome: advanced strategies, enhanced benefits, perceived obstacles. Curr Pharm Des 2005; 11(1): 25-36.

[105] Gomez J, Nadal A, Sabariegos R, Beguiristain N, Martell M, Piron M. Three properties of the hepatitis C virus RNA genome related to antiviral strategies based on RNA-therapeutics: variability, structural conformation and tRNA mimicry. Curr Pharm Des 2004; 10(30): 3741-56.

[106] Martinez-Salas E, Fernandez-Miragall O. Picornavirus IRES: structure function relationship. Curr Pharm Des 2004; 10(30): 3757-67.