Characterization of the Heptad Repeat Regions, HR1 and HR2, and Design of a Fusion Core Structure Model of the Spike Protein from Severe Acute Respiratory Syndrome (SARS) Coronavir...

ARTICLE *in* BIOCHEMISTRY · DECEMBER 2004

Impact Factor: 3.02 · DOI: 10.1021/bi049101q · Source: PubMed

Characterization of the Heptad Repeat Regions, HR1 and HR2, and Design of a Fusion Core Structure Model of the Spike Protein from Severe Acute Respiratory Syndrome (SARS) Coronavirus†

Yanhui Xu,‡ Jieqing Zhu,§ Yiwei Liu,‡ Zhiyong Lou,‡ Fang Yuan,⊥ Yueyong Liu,§ David K. Cole,⊥ Ling Ni,§ Nan Su,‡ Lan Qin,‡ Xu Li, Zhihong Bai,‡ John I. Bell,⊥ Hai Pang,‡ Po Tien,§ George F. Gao,*,§,⊥ and Zihe Rao*,‡

Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China, National Laboratory of Bio-Macromolecules, Institute of Biophysics, Beijing 100101, China, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China, and Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford University, Oxford OX3 9DU, United Kingdom

Received May 4, 2004; Revised Manuscript Received June 8, 2004

ABSTRACT: Severe acute respiratory syndrome coronavirus (SARS-CoV) is a newly emergent virus responsible for a worldwide epidemic in 2003. The coronavirus spike proteins belong to class I fusion proteins, and are characterized by the existence of two heptad repeat (HR) regions, HR1 and HR2. The HR1 region in coronaviruses is predicted to be considerably longer than that in other type I virus fusion proteins. Therefore the exact binding sequence to HR2 from the HR1 is not clear. In this study, we defined the region of HR1 that binds to HR2 by a series of biochemical and biophysical measures. Subsequently the defined HR1 (902−952) and HR2 (1145−1184) chains, which are different from previously defined binding regions, were linked together by a flexible linker to form a single-chain construct, 2-Helix. This protein was expressed in *Escherichia coli* and forms a typical six-helix coiled coil bundle. Highly conserved HR regions between mouse hepatitis virus (MHV) and SARS-CoV spike proteins suggest a similar three-dimensional structure for the two fusion cores. Here, we constructed a homology model for SARS coronavirus fusion core based on our biochemical analysis and determined the MHV fusion core structure. We also propose an important target site for fusion inhibitor design and several strategies, which have been successfully used in fusion inhibitor design for human immunodeficiency virus (HIV), for the treatment of SARS infection.

A new coronavirus has been identified as the distinct pathological entity responsible for severe acute respiratory syndrome (SARS), which infected more than 8000 people and killed 774 worldwide in 2003. The SARS-associated coronavirus (SARS-CoV) is not closely related to any of the previously characterized coronaviruses isolated from either humans or animals and has therefore been assigned to a new group, group IV, along with the previous three known groups (*1*, *2*).

The family Coronaviridae is a group of enveloped positive-stranded RNA viruses with the largest genome among the RNA viruses, and the members are characterized by 3−4 envelope proteins embedded on the surface of the viral envelope lipid membrane (*3*, *4*). The genome sequencing reveals that, as with other enveloped RNA viruses including the coronaviruses, SARS-CoV envelope spike (S) protein contains highly conserved heptad repeat regions (HR1 and HR2), which have been shown to be important in virus membrane fusion. Such heptad repeat regions have been successfully used as targets for virus entry/fusion inhibitors in a number of viruses (*5*, *6*). The existence of HR regions is also a characteristic of class I viral fusion proteins, which is different from class II (two classes, class I and II, have been classified at present) (*7*). It is generally believed that the envelope protein undergoes a series of conformational changes during the virus fusion process. The HR1 and HR2 regions are believed to be important modules/domains in this process and show different conformations in different fusion states (*6*). Under the current model, there are at least three conformational states of the envelope fusion protein. They are the prefusion native state, prehairpin intermediate state, and postfusion hairpin state (*6*, *8*). During these state transitions, HR1 and HR2 are exposed in an intermediate conformational state but bind to each other to form a coiled coil structure in the postfusion state. Therefore the in vitro introduced HR peptides compete with their endogenous HR counterparts in the intermediate state, preventing the transition into the formation of HR1/HR2 coiled coil bundle in the postfusion state. Some HR peptides or analogues have been shown to be potent inhibitors for virus fusion (*6*, *9*,

† This work was supported by Project 973 and 863 of the Ministry of Science and Technology of China (Grant Nos. 200BA711A12, G199075600, 2003CB514116, and 2001CB510001), The Key Project of the Knowledge Innovation Program of Chinese Academy of Sciences (CAS), and the National Natural Science Foundation of China (NSFC, Grant No. 30228025). G.F.G.'s stay at Tsinghua University was supported by Chunhui Project Scheme of Ministry of Education of China.

* To whom correspondence should be addressed. George F.Gao: tel 44-1865-228927; fax 44-1865-222901; e-mail george.gao@ndm.ox.ac.uk. Zihe Rao: tel 86-10-62771493; fax 86-10-62773145; e-mail raozh@xtal.tsinghua.edu.cn.

‡ Tsinghua University and Institute of Biophysics.

§ Chinese Academy of Sciences.

⊥ Oxford University.

10.1021/bi049101q CCC: $27.50
Published on Web 10/13/2004

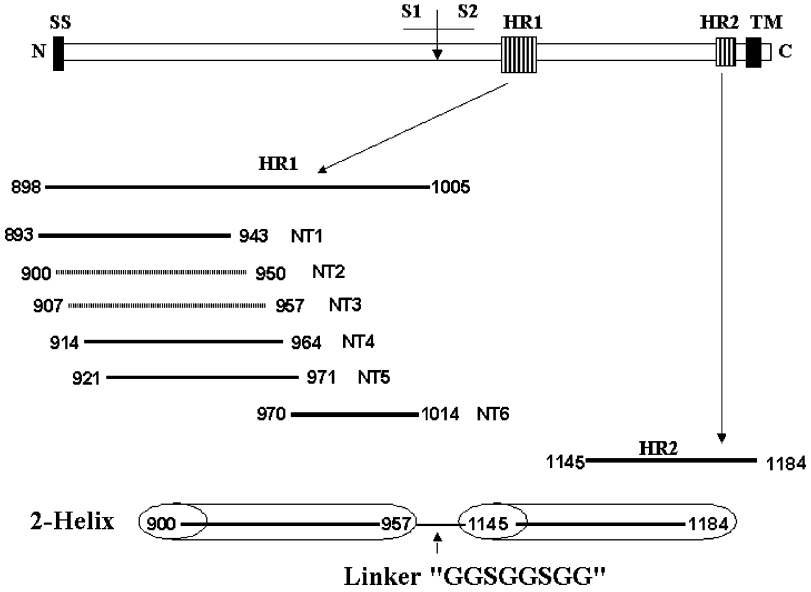

FIGURE 1: Schematic diagram of SARS-CoV spike protein with the location of structurally significant domains indicated. S1 and S2 are formed after proteolytic cleavage (vertical arrow) and are noncovalently linked. The enveloped protein has an N-terminal signal sequence (SS) and a transmembrane domain (TM) adjacent to the C terminus. S2 contains two HR (heptad repeat) regions (hatched bars), HR1 and HR2 as indicated. The residues of HR1 (898−1005) and HR2 (1145−1184) used in this study were derived from the LearnCoil-VMF prediction program (*20*). NT1−NT6 are six truncations derived form HR1 to test for the exact region that binds HR2 in HR1. The 2-Helix protein construct consists of HR2 and part of HR1, which is the major region binding HR2, connected by an eight amino acid linker as indicated.

10). This coiled coil bundle conformation is believed to be important for bringing two lipid membranes (cellular and viral) into proximity with each other, thus enabling the membrane fusion for virus entry. Membrane fusion is the key step for enveloped virus infection. The HR1/HR2 coiled coil bundle is called the virus fusion protein fusion core (*6*). Several crystal structures of fusion cores have been determined to date, including human immunodeficiency virus (HIV) gp41, human respiratory syncytial virus (HRSV) F, influenza virus HA, and Ebola virus GP (*8, 11*−*15*). Typical of this kind of structure, three copies of HR1 bind together to form a central trimeric core and three copies of HR2, which is shorter than HR1, surround this core. Because both HR1 and HR2 are α-helical in this fusion core, this structure is also referred to as a six-helix coiled coil bundle.

To investigate the structural basis of SARS-CoV fusion and entry and identify new fusion inhibitors, we have carried out the analysis of the SARS-CoV HR1 and HR2 fusion core. In this study, both HR1 and HR2 have been expressed as a GST-fusion proteins, and their binding regions have been analyzed. A single-chain polypeptide of HR1 and HR2 (linked by a flexible amino acid linker), named 2-Helix, has also been engineered and analyzed. The engineered 2-Helix shows that SARS-CoV S protein is a typical class I viral fusion protein, as observed recently for the murine coronavirus, mouse hepatitis virus (MHV) (*16*).

Although sequence similarity is not very high in other regions of spike proteins, HR regions between MHV and SARS-CoV spike proteins are highly conserved. Based on our biochemical analysis and determined MHV fusion core structure, a SARS coronavirus fusion core model was constructed. Strategies that have been successfully used in fusion inhibitor design for HIV could also be used in the design of SARS fusion inhibitors (*6, 9, 10, 17, 18*). Therefore, using our fusion core model, we have proposed important target sites for the fusion inhibitor design and several potential inhibitors that may be useful for treating SARS.

Tripet and co-workers have defined the binding regions of HR1 and HR2 as residues 916−950 of HR1 and residues 1151−1185 of HR2 (*19*). In contrast, our results show that the binding regions of HR1 and HR2 are residues 902−952 in HR1 and 1145−1184 in HR2. There are several significant differences between the two definitions of the fusion core. We will give a reasonable interpretation in the further discussion.

EXPERIMENTAL PROCEDURES

Gene Construction. The SARS-CoV S gene was cloned from SARS-CoV coronavirus strain BJ302 clone 1 (GenBank accession no. AY429072). The HR1 and HR2 regions were predicted using the computer program LearnCoil-VMF (*20*). The HR1 region covers amino acids 898−1005, whereas the HR2 covers amino acids 1145−1184 (Figure 1). To test the actual HR1 region bound to HR2, six additional HR1 constructs, named as NT1−NT6, were made (Figure 1). Their lengths are shown in Figure 1. All PCR-amplified fragments were cloned into pGEX-6P-1 GST-fusion expression vector (Pharmacia) by restriction enzyme sites *Eco*RI and *Xho*I (introduced by PCR primers). For the 2-Helix construct, the HR1 region used was derived from amino acids 900 to 957 and HR2 from 1145 to 1184. The 2-Helix construct was made by linking the HR1 and HR2 with an eight amino acid linker (GGSGGSGG, single amino acid abbreviation used here). All the cloned constructs were verified by direct DNA sequencing.

Protein Expression and Purification. All the relevant positive expression vectors were transformed into *Escherichia coli* strain BL21 (DE3) competent cells and single colony was inoculated into 2× YT medium containing 100 μg/mL ampicilin at 37 °C for overnight culture. Then the

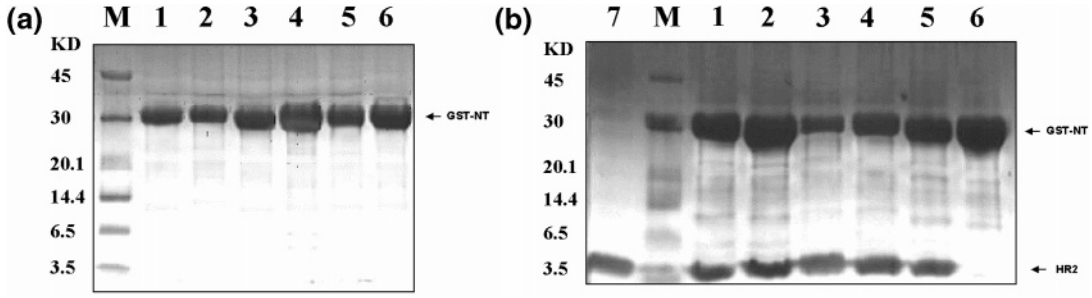

FIGURE 2: GST-pulldown experiments: (a) purified GST-fusion protein (NT1−NT6, lanes 1−6) for control; (b) lanes 1−6, pulldown results of GST-fusion protein NT1−NT6 with HR2, respectively; lane 7, HR2 for control. The samples were subjected to Tris-tricine 10% SDS−PAGE analysis. Lane M contains molecular weight standards in kDa.

overnight culture was transferred to new 2× YT medium for large-scale protein production at 37 °C. When the culture density (OD_{600}) reached 0.8, the culture was induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for an additional 10 h at 16 °C before the cells were harvested.

The harvested culture was centrifuged, and the bacterial cell pellet was resuspended in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) and homogenized by sonication. The lysate was centrifuged at 18 000*g* for 20 min at 4 °C, and the supernatant was loaded onto a glutathione Sepharose 4B column (Pharmacia). The protein-loaded column was then washed with 10 × column volume of PBS. After that, for GST-fusion protein, the protein was eluted with 10 mM reduced glutathione. For the GST-removed proteins, the GST-3C rhinovirus protease (kindly provided by Drs. J. Heath and K. Hudson) was added into the resin, and the mixture was incubated with gentle agitation for about 10 h at 4 °C. The target protein was eluted with 10 mL of PBS.

For SARS 2-Helix protein, expressed with a His-tag, the supernatants were loaded on a Ni−NTA column (QIAGEN) that was equilibrated with PBS. The contaminated protein was washed with washing buffer (1× PBS, 60 mM imidazole) over 10 column volumes, and the target protein was eluted with elution buffer (1× PBS, 500 mM imidazole) for 5 column volumes. The resultant 2-Helix protein was concentrated to a proper concentration by ultrafiltration (10 kDa cutoff). Proteins were analyzed by 10% Tricine SDS−PAGE.

Gel-Filtration Experiment. The targeted protein was loaded on a Superdex 75 column (Pharmacia) running on an Akta Explorer FPLC system (Amersham Pharmacia). The fractions of the peak were collected and run on 10% Tricine SDS−PAGE. The molecular weights of the peak fractions were estimated by comparison with the protein standard running on the same column.

GST-Pulldown Experiments. Purified GST-fusion protein (NT1−NT6) was mixed with GST-removed HR2 in PBS at 4 °C for 2 h, and then the mixture was loaded onto the glutathione-Sepharose 4B column. The column was thoroughly washed with 10 × column volume of PBS until no obvious proteins were seen and then eluted by 10 mM reduced glutathione. The collected peak was subjected to Tris-tricine SDS−PAGE analysis.

Circular Dichroism (CD) Spectroscopy. CD spectra were measured on a Jasco J-715 spectrophotometer in PBS. Wavelength spectra were recorded at 25, 50, 70, and 85 °C using a 0.1 cm path length cuvette. For the thermodynamic stability of the 2-Helix protein, it was measured at 222 nm by monitoring the CD signal in the temperature range from 25 to 90 °C with a 5 °C increase for each scan.

Chemical Cross-Linking. The gel-filtration purified 2-Helix proteins were dialyzed against cross-linking buffer (50 mM Hepes, pH 8.3; 100 mM NaCl) and concentrated to about 2 mg/mL by ultrafiltration (10 kDa cutoff). Proteins were cross-linked with ethylene glycol bis(succinimidyl succinate) (EGS) (Sigma). The reactions were incubated for 1 h on ice at concentrations of 0.1, 0.5, 1.5, and 5 mM EGS and then quenched with 50 mM glycine. Cross-linked products were analyzed under reducing conditions on 10% Tricine SDS−PAGE.

Model Building. The SARS fusion core structure model was built based on our recent crystal structure of MHV fusion core (31). All the amino acids of SARS core were substituted by those of MHV core. After several cycles of manual building and energy minimization, the model was built and the surface map was drawn by the program GRASP.

RESULTS AND DISCUSSION

All Expressed Polypeptides Are Soluble Proteins. As described in the Experimental Procedures, nine expression constructs of GST-fusion polypeptides were prepared in this experiment (Figure 1). They were full-length HR1, NT1−6 of HR1, HR2, and 2-Helix. All engineered constructs were verified by DNA sequencing, and the protein expression yield was high. All nine polypeptides were expressed in the supernatants of the sonicated cells, that is, they were soluble in their GST-fusion forms and easily purified by GST-affinity column (Figure 2a). In addition, all polypeptides were soluble even after the removal of the fusion partner, GST, with some of them showing aggregation at higher concentrations.

N-Terminal Moiety (Amino Acids 900−957) of the Predicted HR1 Binds HR2. From the LearnCoil-VMF program prediction, the HR1 region covers over 100 amino acids, and it is unlikely that the whole region would interact with HR2, though we previously showed that the in-vitro-prepared full-length HR1 and HR2 bound each other (*21*). Therefore the binding region(s) of HR1 to HR2 was examined. To do this, six additional truncated HR1 polypeptides (NT1−6) were expressed and purified in addition to the full-length HR1. GST pull-down experiments were carried out, and it was clear that NT1−5 bound to HR2 but NT6 did not (Figure 2b).

To refine the binding region and test the HR1−HR2 binding affinity, CD spectra at different temperatures were

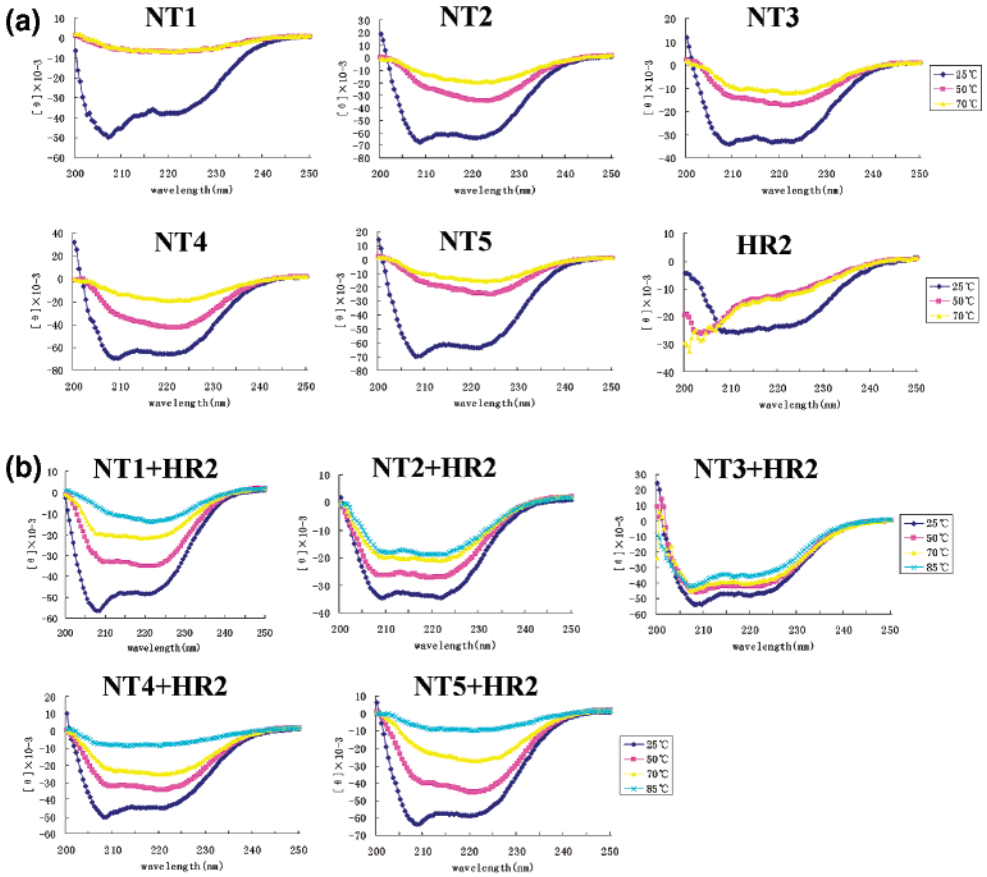

FIGURE 3: CD spectra for refining the binding region of HR1−HR2. Panel a shows CD spectra of NT1−NT5 and HR2 alone. Wavelength spectra were recorded at 25, 50, and 70 °C. Panel b shows CD spectra of mixtures of NT1−NT5 with HR2 in equal molar amounts. Wavelength spectra were recorded at 25, 50, 70, and 85 °C.

measured. As shown in Figure 3a, NT1−5 all possessed α-helical structure at 25 °C (typical double minima at 208 and 222 nm) but with different thermodynamic stabilities, as shown by the altered absorptions following increase of the temperature from 25 to 50 and 70 °C. HR2 itself also possesses α-helical structure at 25 °C but becomes unstructured when the temperature reaches 50 °C, indicating that the α-helix structure is not stable. Figure 3b shows CD spectra of five mixtures of NT1−5 with HR2 (in equal molar) at different temperatures. These complexes all possessed α-helical structure at 25 °C. Different thermodynamic stabilities were observed with altered absorptions following the increase of the temperature from 25 to 50, 70, and 85 °C. The complexes of NT2−HR2 and NT3−HR2 were more stable than the other three complexes. We can conclude that the N-terminal moiety (amino acids 900−957) of the predicted HR1 is the major region that binds HR2.

2-Helix Protein Forms Characteristic Stable Six-Helix Bundle. Based on previous idenfication of the region of HR1 that binds HR2, the HR1 and HR2 regions of SARS-CoV spike protein have been chosen in this study as described in Experimental Procedures. A schematic representation of the SARS 2-Helix construct of HR1, linker, and HR2 is shown in Figure 1. The single-chain construct of the fusion core HR1/HR2 was made as 2-Helix. As expected from our previous work on paramyxovirus fusion core 2-Helix proteins (*22*−*24*), the SARS 2-Helix polypeptide construct in this study was expressed at high level in the *E. coli* expression system. The protein has been shown to be soluble in either PBS or pH 8.0 Tris buffers and at concentrations of up to

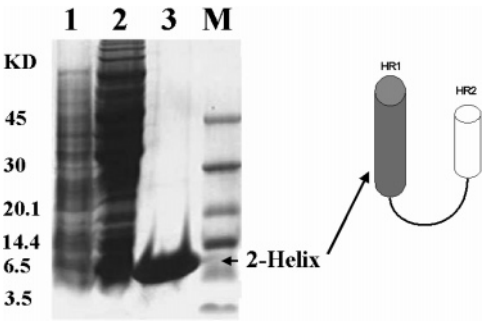

FIGURE 4: SDS−PAGE analysis of the expression and purification of the 2-Helix proteins of SARS-CoV S protein: lane 1, noninduced; lane 2, IPTG-induced (final concentration is 1 mM); lane 3, purified 2-Helix protein; lane M, molecular weight standards in kDa.

40 mg/mL. This implies that the protein has most likely been folded correctly.

The SARS 2-Helix proteins were expressed in *E. coli* as soluble proteins (Figure 4). The supernatants of the cell lysates were loaded on a Ni−NTA column, and the target proteins were purified by affinity chromatography. The purified proteins were further analyzed by gel filtration and chemical cross-linking for estimation of the molecular weight. The SARS 2-Helix protein was eluted between the corresponding position of 52 and 14 kDa (Figure 5a). In comparison, the computed molecular weight of 2-Helix (with 6 × histidine tag) is about 13 kDa, which indicates that the 2-Helix might form oligomers (about 40 kDa). The complex of full-length HR1 with HR2 was eluted just adjacent to 52

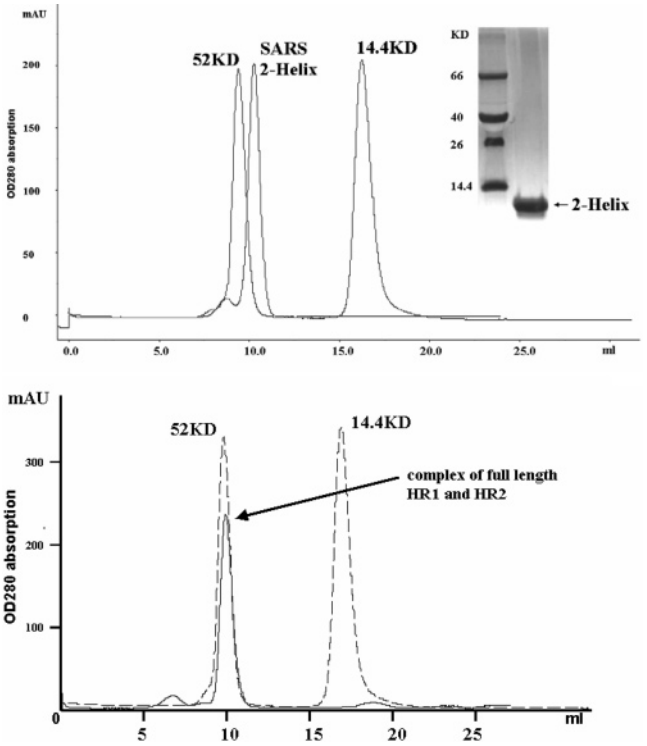

FIGURE 5: Gel-filtration analysis of the purified SARS 2-Helix protein and complex of HR1 and HR2. Panel a shows gel-filtration analysis of the purified SARS 2-Helix protein. The inset picture is of the proteins collected from the peak analyzed on Tris-tricine 10% SDS−PAGE. Panel b shows gel-filtration analysis of the complex of HR1 and HR2. The estimated molecular weight was determined in comparison with the standards (14 and 52 kDa).

kDa, also indicating that the fusion core might form oligomers (Figure 5B). Subsequent chemical cross-linking demonstrated that the SARS 2-Helix protein oligomer was, in fact, a trimer (Figure 6a), although the monomer/dimer bands could also be found in high concentrations of the cross-linker. In addition, the trimer content increased as the concentration of the chemical cross-linker was increased. Dimer and trimer bands could also be seen if the samples (reduced by SDS loading buffer) were not boiled at 100 °C. These results indicate that HR1 interacts with HR2 to form a stable trimeric complex.

The secondary structure of the 2-Helix protein was examined by CD spectroscopy as described in Experimental Procedures. The absorption curve showed that the 2-Helix protein had double minima at 208 and 222 nm (Figure 7a), consistent with a typical α-helix structure. The thermal stability test by CD spectroscopy showed that trimer formation of the 2-Helix was very stable in PBS (Figure 7b), indicating that the 2-Helix trimer represents the core structure of the postfusion state of the coiled coil bundle. The postfusion state has been studied in great detail for other viral fusion proteins (*6*, *25*−*29*).

SARS-CoV has been convincingly identified as a member of family Coronaviridae (*1*). Our results here show that their heptad repeat regions, HR1 and HR2, interact with each other to form a typical stable six-helix coiled coil bundle, which represents the fusion core of the fusion protein (*6*, *26*). The six-helix bundle is extremely stable to thermal denaturation. The fusion core studied here has a melting temperature of about 85 °C and could not be separated easily in SDS−PAGE unless boiled at 100 °C with SDS−PAGE loading

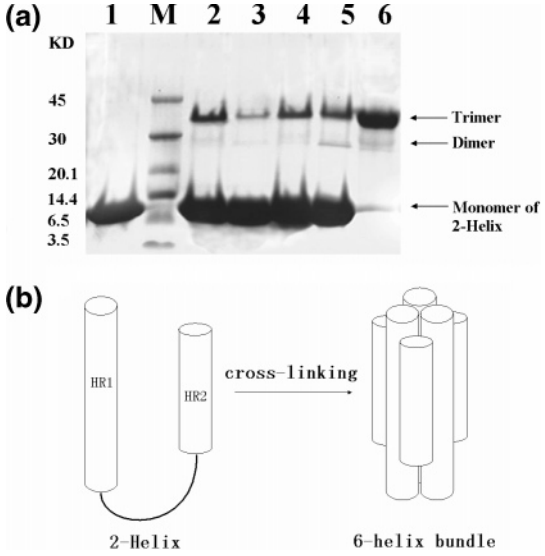

FIGURE 6: Chemical cross-linking of the SARS-CoV 2-Helix protein with different concentrations of chemical cross-linker, EGS (ethylene glycol bis(succinimidyl succinate) from Sigma). In panel a, cross-linked products were separated on Tris-tricine 10% SDS−PAGE followed by Commassie brilliant blue staining. Protein markers (M) are shown in kDa. Lanes 1 and 3−6 indicate the different concentrations of the EGS used (0, 0.1, 0.5, 1.5, and 5mM respectively). Lane 2 represents 2-Helix protein samples (reduced by SDS loading buffer) not boiled at 100 °C. Bands corresponding to monomer, dimer, and trimer are indicated. Panel b presents a schematic diagram of cross-linking.

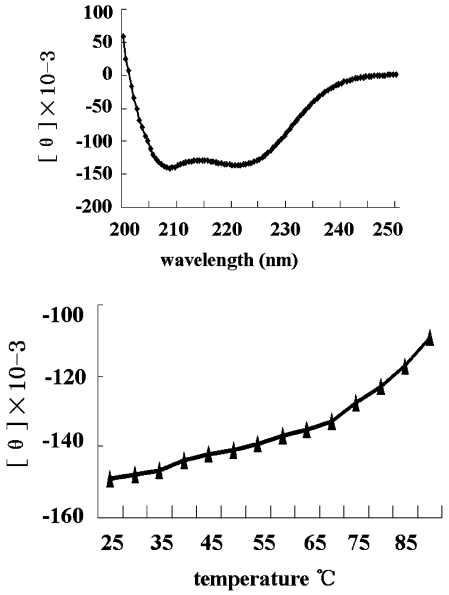

FIGURE 7: The CD spectra and thermal stability measurement of the SARS-CoV 2-Helix proteins in PBS. In panel a, the 2-Helix proteins show a typical α-helix secondary structure at 25 °C. In panel b, thermal stability was recorded at 222 nm. The T_m is over 90 °C as the plateau was not reached up to 90 °C.

buffer. Recent studies showed that peptides derived from HR1 region or HR2 region of SARS-CoV spike protein are potent inhibitors for viral entry (*21*,*30*), which is consistent with the observations made in this study. These properties suggest that the six-helix bundle has similar structure with other viral fusion cores such as HIV gp41 and influenza HA, which have been studied in great detail (*8*, *11*, *13*).

Model of SARS-CoV Fusion Core Structure and Potential Fusion Inhibitor Design. We have recently determined the

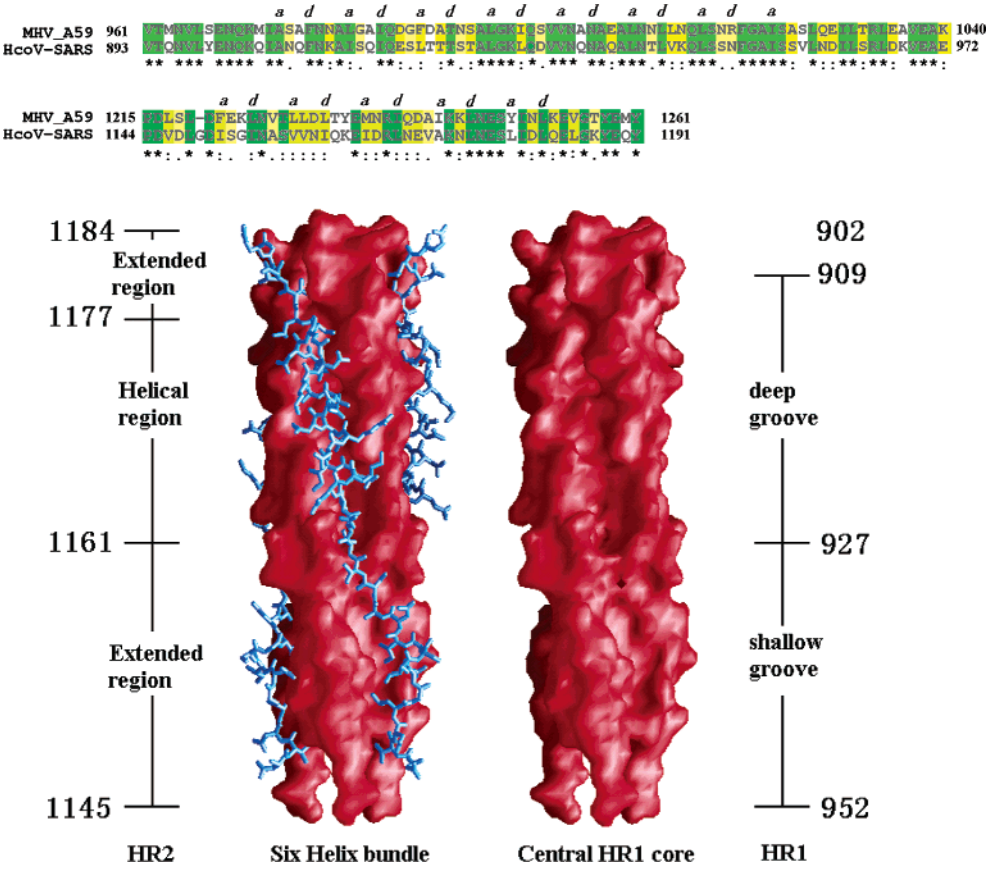

FIGURE 8: A model of the fusion core structure of SARS-CoV spike protein: (a) sequence alignment of HR1 and HR2 region between MHV and SARS-CoV spike proteins; (b) model of fusion core structure of SARS-CoV S protein. A surface map showing the hydrophobic grooves on the surface of three central HR1 is presented on the right side. Three HR2 helices pack against the hydrophobic groove in an antiparallel manner (left side). The helical regions in HR2 extended regions could be observed clearly. The figure on the right side shows the deep and relatively shallow grooves on the surface of the central HR1 coiled coil. The helical region of HR2 just packs against the deep groove and the extended region packs against shallow grooves. The deep groove would be an important target site for the fusion inhibitor design.

crystal structure of fusion core of the MHV S protein, which represents the first fusion core structure of any coronavirus (*31*). The structure reveals a central hydrophobic coiled coil trimer surrounded by three helices in an oblique, antiparallel manner. This structure shares significant similarity to both the low-pH-induced conformation of influenza hemagglutinin and fusion core of HIV gp41, indicating that the structure represents a postfusion state formed after several conformational changes. Our results further indicate that mechanisms for the viral fusion of coronaviruses are similar to those of influenza virus and HIV.

Here, we propose a SARS-CoV fusion core structure model (Figure 8b) as a six-helix antiparallel bundle based on the following evidence. First, the sequence alignment of HR1 and HR2 regions in spike proteins between MHV and SARS-CoV shows significant similarity in these heptad repeat regions (for HR1, identity is 60% and similarity is 91%; for HR2, identity is 35% and similarity is 85%) (Figure 8a). highly conserved HR1 and HR2 regions in coronavirus spike proteins suggest a similar three-dimensional structure between MHV and SARS-CoV fusion core. Second, residues in the a, d, e, and g positions in HR1 and the a and d positions in HR2 regions, which are very important for the coiled coil formation, are highly conserved (Figure 8a). Third, our previous biochemical analysis in this paper indicates that the 2-Helix construct (HR1 and HR2 with a short linker) of SARS-CoV spike protein is also a trimeric α-helix and is thermostable. Last, virus cell entry inhibition and cell−cell fusion inhibition experiments show that the HR2 region of SARS-CoV spike protein has a similar function to C peptides in HIV-1 in viral fusion inhibition (*21*,*30*). We can therefore conclude that the SARS-CoV spike protein fusion core has a three-dimensional structure similar to that of MHV.

In this model, the complex is a six-helix bundle comprising a trimer of 2-Helix. The six-helix bundle consists of a parallel trimeric coiled coil of three HR1 helices with three HR2 helices packing against the grooves formed on the surface of the HR1 core (Figure 8b). The N terminus of HR1 and the C terminus of HR2 are located at the same end of the six-helix bundles. In this pattern, the likely role of the six-helix bundle structure is to facilitate juxtaposition of the viral and cellular membranes by bringing the fusion peptide, which inserts into the cellular membrane, close to the transmembrane segment, which is anchored in the viral membrane. Analogous to the HIV-1 C peptides, the HR2 region of SARS-CoV spike protein is likely to function in a dominant-negative manner by binding to the transiently exposed hydrophobic grooves in the prehairpin intermediate and consequently blocking the formation of the fusion-active hairpin structures (*5*). A similar approach may be applied to identify inhibitors of SARS-CoV infection since there are currently no effective agents for the antiviral therapy of SARS infection.

Table 1: Potential Inhibitors for Viral Entry

no.	source[a]	residues	sequence
P1	HR2	1145−1184	DVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQE
P2	HR2	1160−1177	NIQKEIDRLNEVAKNLNE
P3	HR2	1160−1184	NIQKEIDRLNEVAKNLNESLIDLQE
P4	HR2 + spacer[b]	1160−1200	NIQKEIDRLNEVAKNLNESLIDLQE*LGKYEQYIKWPWYVWL*
P5	HR1	909−927	FNKAISQIQESLTTTSTAL
P6	HR1	902−927	QKQIANQFNKAISQIQESLTTTSTAL
P7	GCN4[c] + HR1	909−927	*GCN4*−FNKAISQIQESLTTTSTAL
P8	GCN4[c] + HR1	902−927	*GCN4*-QKQIANQFNKAISQIQESLTTTSTAL
5-helix[d]	HR1 + HR2		P6−linker-P3-linker-P6-linker-P3-linker-P6
D-peptide[e]	phage library		C−X-X−X-X−X-X−X-X−X-X−C

[a] HR1, QKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNF (902−952); HR2, DVDLGDISGINASVVNIQKEIDRL-NEVAKNLNESLIDLQE (1145−1184). [b] Spacer is the region between HR2 and the transmembrane domain. [c] The sequence of GCN4 is RMKQIEDKIEEIESKQKKIENEIARIKK (*9*). [d] P6-to-P3 linker, GGSGG; P3-to-P6 linker, GSSGG. (*9*). [e] D-peptide will be selected by mirror-image phage display using P7 as a target (*17*).

In the model of fusion core of SARS-CoV spike protein, residues 902−952 form the HR1 region and residues 1145−1184 form the HR2 region, which packs against the hydrophobic grooves on the surface of HR1. Residues 1161−1177 (IQKEIDRLNEVAKNLNE) form a typical α-helix structure, which packs against a deep groove consisting of residues 909−927 (FNKAISQIQESLTTTSTAL) in the HR1 region (Figure 8b). The other regions in HR2 form extended strands, which pack against the relatively shallow grooves on the surface of central HR1 coiled coil (Figure 8b). On the basis of our biochemical analysis and molecular modeling in this study, we can propose an important target site (deep groove) for the fusion inhibitor design. Several strategies, which have been successfully used in fusion inhibitor design for HIV may prove useful for treating SARS infection (*5, 6, 9, 10, 17, 18, 21, 30*).

HR1, HR2, and their derivatives could be potential inhibitors for viral entry. Table 1 shows several potential inhibitors that we proposed on the basis of our biochemical results and molecular modeling. P1−P4 are designed from the HR2 region and P5−P8 from the HR1 region. P1 is the full-length HR2, which we propose will have high inhibitory activity but will be sensitive to proteolytic degradation. P2 is a typical α-helix region in HR2, which packs against the deep groove on the surface of central HR1 coiled coil. The inhibitory activity of peptides derived from the HR2 region has been reported in recent published papers (*21*,*30*), which is consistent with the structure model. P3 and P4 contain P2 and are extended to include additional residues. P3 extends to the end of the HR2 region and P4 extends to residues near the transmembrane region, similar to T20 in C-peptide of HIV gp41 (*18*). P5 consists of residues forming a deep groove in the HR1 region, which we think is a strong potential binding site for small molecule inhibitors. P6 consists of residues from the N terminus of HR1 to the C terminus of P5, the end of the deep groove. Both P5 and P6 are expected to have low inhibitory activity because they are highly hydrophobic and will tend to aggregate in solution. P7 and P8 are derived from P5 and P6, respectively, with the addition of GCN4 at the N terminus to increase their solubility and inhibitory activity. 5-Helix and D-peptide are also potential inhibitors for viral fusion, which have been successfully identified in studies of HIV gp41 (*9, 17*). Biological experiments to verify the inhibitory activity of these peptides and potent inhibitor design are underway.

Comparison to Recent Results for the Binding Regions HR1 and HR2 in SARS Fusion Core. Recent studies by Tripet and co-workers have shown that the site of interaction between the HR1 and HR2 regions is between residues 916−950 of HR1 and residues 1151−1185 of HR2 (*19*). However, our results show that the binding regions of HR1 and HR2 are residues 902−952 in HR1 and 1145−1184 in HR2. The most significant difference between the two conclusions is that the fusion core that we defined includes residues 902−916 in HR1, which are particularly important for binding to HR2. These 14 residues form half of the deep groove for the binding of helical regions in HR2. Another difference is that the HR2 region that we defined includes residues 1145−1151, which also pack against the grooves formed by HR1 by means of hydrophobic interactions, although these seven residues form an extended strand rather than an α-helix. In our results, the HR1 chain is slightly longer than the HR2 chain but almost the same in length in the three-dimensional structure. Analogous to extended region of HR2 in SV5 F (*32*), the extended regions of HR2 in SARS fusion core pack against the central HR1 coiled coil, making HR1 (40 residues) have similar length as HR2 (51 residues).

Conclusions. Biochemical and structural analysis of the six-helix bundle provides a detailed picture of the fusion core structure of SARS spike protein and the underlying molecular mechanism for viral entry and fusion inhibition. This will also open a new avenue toward the structure-based fusion inhibitor design of peptides, or peptide analogues, for example, small molecules, targeting the newly emerging infectious disease, severe acute respiratory syndrome (SARS).

ACKNOWLEDGMENT

We thank Dr. Mark Bartlam for comments and critical reading.

REFERENCES

1. 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*, 1399−1404.
2. 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*, 1394−1399.
3. Cavanagh, D. (1983) Coronavirus IBV: structural characterization of the spike protein, *J. Gen. Virol. 64* (Part 12), 2577−2583.
4. Siddell, S., Wege, H., and Ter Meulen, V. (1983) The biology of coronaviruses, *J. Gen. Virol. 64* (Part 4), 761−776.
5. Chan, D. C., and Kim, P. S. (1998) HIV entry and its inhibition, *Cell 93*, 681−684.
6. Eckert, D. M., and Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem. 70*, 777−810.
7. Lescar, J., Roussel, A., Wien, M. W., Navaza, J., Fuller, S. D., Wengler, G., and Rey, F. A. (2001) The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH, *Cell 105*, 137−148.
8. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein, *Cell 89*, 263−273.
9. Eckert, D. M., and Kim, P. S. (2001) Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region, *Proc. Natl. Acad. Sci. U.S.A. 98*, 11187−11192.
10. Root, M. J., Kay, M. S., and Kim, P. S. (2001) Protein design of an HIV-1 entry inhibitor, *Science 291*, 884−888.
11. Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994) Structure of influenza haemagglutinin at the pH of membrane fusion, *Nature 371*, 37−43.
12. Weissenhorn, W., Carfi, A., Lee, K. H., Skehel, J. J., and Wiley, D. C. (1998) Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain, *Mol. Cell 2*, 605−616.
13. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41, *Nature 387*, 426−430.
14. Zhao, X., Singh, M., Malashkevich, V. N., and Kim, P. S. (2000) Structural characterization of the human respiratory syncytial virus fusion protein core, *Proc. Natl. Acad. Sci. U.S.A. 97*, 14172−14177.
15. Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milhollen, M. A., Pang, J. X., and Kim, P. S. (1999) Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2662−2667.
16. 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*, 8801−8811.
17. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A., and Kim, P. S. (1999) Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket, *Cell 99*, 103−115.
18. Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B., and Matthews, T. J. (1994) Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection, *Proc. Natl. Acad. Sci. U.S.A. 91*, 9770−9774.
19. 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*., in press.
20. Singh, M., Berger, B., and Kim, P. S. (1999) LearnCoil-VMF: computational evidence for coiled-coil-like motifs in many viral membrane-fusion proteins, *J. Mol. Biol. 290*, 1031−1041.
21. Zhu, J., Xiao, G., Xu, Y., Yuan, F., Zheng, C., Liu, Y., Yan, H., Cole, D. K., Bell, J. I., Rao, Z., Tien, P. and Gao, G. F. (2004) Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors, *Biochem. Biophys. Res. Commun. 319*, 283−288.
22. Zhu, J. Q., Zhang, C. W., Rao, Z., Tien, P., and Gao, G. F. (2003) Biochemical and biophysical analysis of heptad repeat regions from the fusion protein of Menangle virus, a newly emergent paramyxovirus, *Arch. Virol. 148*, 1301−1316.
23. Xu, Y., Gao, S., Cole, D. K., Zhu, J., Su, N., Wang, H., Gao, G. F., and Rao, Z. (2004) Basis for fusion inhibition by peptides: analysis of the heptad repeat regions of the fusion proteins from Nipah and Hendra viruses, newly emergent zoonotic paramyxoviruses, *Biochem. Biophys. Res. Commun. 315*, 664−670.
24. Zhu, J., Zhang, C. W., Qi, Y., Tien, P., and Gao, G. F. (2002) The fusion protein core of measles virus forms stable coiled-coil trimer, *Biochem. Biophys. Res. Commun. 299*, 897−902.
25. Skehel, J. J., and Wiley, D. C. (1998) Coiled coils in both intracellular vesicle and viral membrane fusion, *Cell 95*, 871−874.
26. Lamb, R. A., Joshi, S. B., and Dutch, R. E. (1999) The paramyxovirus fusion protein forms an extremely stable core trimer: structural parallels to influenza virus haemagglutinin and HIV-1 gp41, *Mol. Membr. Biol. 16*, 11−19.
27. Weissenhorn, W., Dessen, A., Calder, L. J., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1999) Structural basis for membrane fusion by enveloped viruses, *Mol. Membr. Biol. 16*, 3−9.
28. Bentz, J. (2000) Membrane fusion mediated by coiled coils: a hypothesis, *Biophys. J. 78*, 886−900.
29. Skehel, J. J., and Wiley, D. C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin, *Annu. Rev. Biochem. 69*, 531−569.
30. 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*, 938−947.
31. Xu, Y., Liu, Y., Lou, Z., Qin, L., Li, X., Bai, Z., Pang, H., Tien, P., Gao, G. F., and Rao, Z. (2004) Structural basis for coronavirus-mediated membrane fusion: Crystal structure of MHV spike protein fusion core, *J. Biol. Chem.*, M403760200.
32. Baker, K. A., Dutch, R. E., Lamb, R. A., and Jardetzky, T. S. (1999) Structural basis for paramyxovirus-mediated membrane fusion, *Mol. Cell 3*, 309−319.

BI049101Q