Spike (S) protein is the outermost component of the virion, and is crucial for the entry of coronaviruses into host cells. S protein is type I viral protein and responsible for the attachment of virus to host cells [1–3] and for instigating the fusion of the virus envelope with cell membrane. All coronavirus S2 proteins have a highly conserved ten-residue sequence Y(V/I)KWPW(W/Y)VWL, which is rich in aromatic amino acids with 3 to 4 tryptophan (Trp) residues. The last five residues of this region probably form the beginning of the membrane-spanning domain which is also called transmembrane (TM) domain, [4,5]. This region is so called Trp-rich region, membrane-proximal external region (MPER), proximal-membrane region or pretransmembrane region (preTM) etc; in this study it is referred as Trp-rich region.

The function of this Trp-rich region is intriguing and remains unknown. Similarly Trp-rich regions also exist in the TM of all lentiviruses, although they may differ somewhat with regard to the number of Trp residues contained, the length of the sequence in which these are interspersed, the properties of the other amino acids present, and the distances within the linear sequence of the Trp-residues between themselves and with the putative membrane-spanning domains.

According to current models for HIV entry [6–10], the high preference of Trp residues for residing at the external face of membranes [11–15], the proximal-membrane Trp-rich region appears to reside on the envelope surface at the membrane–water interface of the lipid bilayer. Following activation, the Trp-rich region is believed to undergo sequential conformational changes, from a reverse turn to an amphipathic helical structure capable of intimately interacting with the viral membrane through certain Trp residues and, concomitantly, with the cell membrane through others [16]. With their relatively bulky indole side chains, the Trp residues, most likely synergizing with the fusion peptide [12], would then destabilize both membranes and drive the energetically unfavorable lipid merging. This interaction permits the formation and expansion of the fusion pore in the late stages of the entry process [17,18]. Here, we examined the putative functions of the Trp-rich region of SARS-CoV S protein.

Materials and methods

Reagents. Calcium chloride (CaCl2), sodium chloride (NaCl), potassium chloride (KCl), disodium hydrogen phosphate anhydrous (Na2HPO4), potassium phosphate monobasic (KH2PO4) and 4-(2-hydroxyethyl)-1-piperazineethane-sulphonic acid (HEPES) were all Ultra grade and purchased from Sigma–Aldrich. Rabbit anti-S antibody was kindly provided by Prof. Ding Xiang Liu (Institute of Molecular and Cellular Biology of Singapore).

Cell cultures. Vero E6 cells (Vero C1008 [ATCC CRL-1586™]) and 293T cells (293T/17 [ATCC CRL-11268]), were maintained in DMEM with l-glutamine (Gibco), supplemented with 10% new born calf serum and 50 μg/ml penicillin–streptomycin (Gibco).

Plasmids. pNL4-3Luc+Env Vpr and pcDNA3.1-OPT9-S were kindly provided by Prof. Zhang Linqi (Aaron Diamond AIDS Research Center).
Global mutations

OPT9-S which contains codon-optimized S gene, a series of plasmids containing abbreviations

Name/ Sequence introduced in Trp-rich region of S protein

WY4A, WY5A, and F1202A) substantially decrease infectivity by >90%

residues are bold and underlined.

2. W3A W1194A + W1196A + W1199A, and F1202A) substantially decrease infectivity by >90%

abrogate infectivity, while single- and double-substitution with Ala

Table 1

<table>
<thead>
<tr>
<th>Name/ sequences</th>
<th>Description</th>
<th>Sequence (aa 1190–1204)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild-type</strong></td>
<td>Control</td>
<td>QYIKWPYWVWL G</td>
</tr>
<tr>
<td>1. <strong>Global mutations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WY5A</td>
<td>W1194A + W1196A + Y1197A + W1199A + F1202A</td>
<td>QYIKAPAVALGI</td>
</tr>
<tr>
<td>WY4A</td>
<td>W1194A + W1196A + Y1197A + W1199A</td>
<td>QYIKAPAVALGI</td>
</tr>
<tr>
<td>W3A</td>
<td>W1194A + W1196A + W1199A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>W2A-4/6</td>
<td>W1194A + W1196A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>W2A-4/9</td>
<td>W1194A + W1199A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>W2A-6/9</td>
<td>W1196A + W1199A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>2. <strong>Single mutations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1194A</td>
<td>W1194A</td>
<td>QYIKAPAYVWL</td>
</tr>
<tr>
<td>W1196A</td>
<td>W1196A</td>
<td>QYIKAPAYVWL</td>
</tr>
<tr>
<td>Y1197A</td>
<td>Y1197A</td>
<td>QYIKPWA VWL</td>
</tr>
<tr>
<td>W1199A</td>
<td>W1199A</td>
<td>QYIKPWPYV</td>
</tr>
<tr>
<td>Fl202A</td>
<td>Fl202A</td>
<td>QYIKPWPYV</td>
</tr>
</tbody>
</table>

Note: Wild-type aromatic residues are shown in bold and italic. Mutated aromatic residues are bold and underlined.

Table 2

<table>
<thead>
<tr>
<th>Name/ sequences</th>
<th>Description</th>
<th>Sequence (aa 1190–1204)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild-type</strong></td>
<td>Control</td>
<td>QYIKWPYWVWL G</td>
</tr>
<tr>
<td>1. <strong>Global mutations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WY5A</td>
<td>W1194A + W1196A + Y1197A + W1199A + F1202A</td>
<td>QYIKAPAVALGI</td>
</tr>
<tr>
<td>WY4A</td>
<td>W1194A + W1196A + Y1197A + W1199A</td>
<td>QYIKAPAVALGI</td>
</tr>
<tr>
<td>W3A</td>
<td>W1194A + W1196A + W1199A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>W2A-4/6</td>
<td>W1194A + W1196A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>W2A-4/9</td>
<td>W1194A + W1199A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>W2A-6/9</td>
<td>W1196A + W1199A</td>
<td>QYIKAPAYVLF</td>
</tr>
<tr>
<td>2. <strong>Single mutations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1194A</td>
<td>W1194A</td>
<td>QYIKAPAYVWL</td>
</tr>
<tr>
<td>W1196A</td>
<td>W1196A</td>
<td>QYIKAPAYVWL</td>
</tr>
<tr>
<td>Y1197A</td>
<td>Y1197A</td>
<td>QYIKPWA VWL</td>
</tr>
<tr>
<td>W1199A</td>
<td>W1199A</td>
<td>QYIKPWPYV</td>
</tr>
<tr>
<td>Fl202A</td>
<td>Fl202A</td>
<td>QYIKPWPYV</td>
</tr>
</tbody>
</table>

Note: Wild-type aromatic residues are shown in bold and italic. Mutated aromatic residues are bold and underlined.

Results and discussion

To investigate the importance of Trp-rich region in SARS-CoV viral entry, S proteins with mutations in Trp-rich region were prepared. All clones containing different mutated S genes were selected and confirmed by sequencing.

The expression of S protein mutants in 293T cells was detected by rabbit anti-S protein antibody. The results showed that mutants W1194A, W1199A, W2A-4/6, W2A-4/9, WY4A, and WY5A (see Table 2 for abbreviations) were expressed at high level, whereas others, such as the expression of Y1197A and W3A were detected at low level. No protein expression was detected under our experimental condition for W1196A, F1202A, and W2A-6/9 (Fig. 3).

SARS-CoV pseudotyped with S protein and its mutants was used to determine S protein-mediated infectivity. Pseudotyped retrovirus containing S protein mutants were first generated using a global site-directed mutagenesis and then Ala-scan of aromatic amino acids in the Trp-rich region to determine the positional importance of each Trp, Phe or Tyr. To restore the aromaticity of Ala-mutants, we performed rescue experiments using a Phe-scan. The results show that global substituted mutants, tri-, tetra- and penta-substitution with Ala (WY5A, WY4A, and W3A) completely abrogate infectivity, while single- and double-substitution with Ala (W2A-4/6, W2A-4/9, W2A-6/9, W1194A, W1196A, Y1197A, W1199A, and F1202A) substantially decrease infectivity by >90% (Table 3). On the other hand, Phe-substituted mutants are able to restore 10–25% infectivity comparing to the wild-type (Table 4).

These results suggest that the aromatic residues of the Trp-rich region of S protein are essential for SARS-CoV infectivity.

Coronavirus entry is mediated by type I viral envelope S protein. S protein of SARS-CoV is responsible for receptor-binding and membrane fusion. S protein contains several functional domains (Fig. 1) to support its correct folding and conformation which are crucial for its function. The Trp-rich region is absolutely conserved in members of coronaviruses (Fig. 2) and highly conserved in other RNA viruses such as HIV, FIV, and EbO. The importance of the Trp-
The rich region in vaccine and therapeutics development has been demonstrated in HIV research as its epitopes are recognized by broadly neutralizing antibodies from human sera. It may play important roles during the whole fusion process.

Many laboratories have shown that the Trp-rich region plays an important role in HIV infection [11,12,19]. More importantly, this exposed region contains epitopes to which broadly neutralizing antibodies could be raised [11–13]. Indeed, Sainz et al. [4] found that the Trp-rich region of coronavirus type I viral fusion protein induces membrane permeabilization to facilitate the viral entry. Epand et al. [20] claimed that the Trp-rich region of HIV gp41 could promote the formation of cholesterol-rich domains which could facilitate membrane fusion.

Results obtained from our former work [21] show that cholesterol-enriched microdomains, known as lipid rafts, are required for SARS-CoV entry. The cryoEM structures of HIV and SIV reveal that the Trp-rich region forms the “feet” of a tripod-like structure of the spike proteins [22], anchoring the legs (HR1 and HR2 regions) firmly on the lipid rafts of viral surface. In the post-fusion state, the two heptad-repeat regions (HR1 and HR2) form a 6-helix bundle, positioning the fusion peptide closely to the Trp-rich region to drive apposition and subsequent fusion of viral and cell membranes. Because of their similarity in fusion mechanism of SARS-CoV and HIV, we hypothesized that the highly conserved Trp-rich region in SARS-CoV S protein plays a similar role in viral entry as in HIV by anchoring the tripod-like structure of S protein on a lipid rafts-like environment on the viral surface. Such a mechanism of anchoring S protein promotes clustering of S proteins, an effect which facilitates docking of host receptors, and subsequently membrane fusion by merging the cholesterol-enriched microdomains of both virus and host membranes.

### Table 3

<table>
<thead>
<tr>
<th>Mutation on S protein</th>
<th>Name</th>
<th>Infectivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td></td>
<td>100*</td>
</tr>
<tr>
<td>Penta-mutation</td>
<td>WY5A</td>
<td>0</td>
</tr>
<tr>
<td>Tetra-mutation</td>
<td>WY4A</td>
<td>0</td>
</tr>
<tr>
<td>Tri-mutation</td>
<td>W3A</td>
<td>0</td>
</tr>
<tr>
<td>Double-mutation</td>
<td>W2A-4/6</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>W2A-4/9</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>W2A-6/9</td>
<td>0.12</td>
</tr>
<tr>
<td>Single-mutation</td>
<td>W1194A</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>W1196A</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Y1197A</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>W1199A</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>F1202A</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*The infectivity of SARS-CoV pseudotyped with wild-type S protein was set as 100%; the infectivity of each SARS-CoV pseudotyped with S protein mutants was compared with wild-type pseudovirus.

### Table 4

<table>
<thead>
<tr>
<th>Position</th>
<th>Infectivity (%)</th>
<th>Infectivity (%) restored by Phe-mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>A</td>
</tr>
<tr>
<td>1194</td>
<td>100</td>
<td>0.55</td>
</tr>
<tr>
<td>1196</td>
<td>100</td>
<td>0.61</td>
</tr>
<tr>
<td>1197</td>
<td>100</td>
<td>0.59</td>
</tr>
<tr>
<td>1199</td>
<td>100</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*The infectivity of SARS-CoV pseudotyped with wild-type S protein was set as 100%; the infectivity of each SARS-CoV pseudotyped with S protein mutants was compared with wild-type pseudovirus.

### Table 3

<table>
<thead>
<tr>
<th>Mutation on S protein</th>
<th>Name</th>
<th>Infectivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td></td>
<td>100*</td>
</tr>
<tr>
<td>Penta-mutation</td>
<td>WY5A</td>
<td>0</td>
</tr>
<tr>
<td>Tetra-mutation</td>
<td>WY4A</td>
<td>0</td>
</tr>
<tr>
<td>Tri-mutation</td>
<td>W3A</td>
<td>0</td>
</tr>
<tr>
<td>Double-mutation</td>
<td>W2A-4/6</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>W2A-4/9</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>W2A-6/9</td>
<td>0.12</td>
</tr>
<tr>
<td>Single-mutation</td>
<td>W1194A</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>W1196A</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Y1197A</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>W1199A</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>F1202A</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*The infectivity of SARS-CoV pseudotyped with wild-type S protein was set as 100%; the infectivity of each SARS-CoV pseudotyped with S protein mutants was compared with wild-type pseudovirus.

### Fig. 1

Schematic representative of SARS-CoV S protein. Modeling and predictive analysis of SARS-CoV S protein. SP, signal peptide; RBD, receptor-binding domain; RBM, receptor-binding motif; FP, fusion peptide; HR, heptad repeat; TM, transmembrane domain; CP, cytoplasmic domain. The number of residues of each region corresponds to their positions in S protein of SARS-CoV. The Trp-rich region is indicated in purple box. The italic red color labeled-residues in Trp-rich region are targeted for mutation in this study.

### Fig. 2

Conserved motifs in Coronavirus S protein. Alignment of the C-terminal region of the SARS-CoV and reference coronavirus S proteins. The black box indicates the amino acid sequence Y(V/I)KWPW(Y/W)VWL which is a conserved motif in all three coronavirus groups and SARS-CoV. This region is referred as Trp-rich region in this study.
et al. and Epand et al. [20,27–30] reported that a consensus 
sterol-binding motifs found in the S protein sequences. Vincent 
Y-(X)(1–5)-R/K-, in which (X)(1–5) represents 1–5 residues of 
sterol-rich regions of membranes has the pattern –L/V-(X)(1–5)-
sequence found in a group of proteins that sequester to choles-
terol-at-interface surface. Saez-Cirion et al.[23] reported that gp41
bind the membrane interface, adopted a \( \alpha \)-helical structure. This 
pre-transmembrane sequence might target membranes inherently 
prone to destabilization of the membrane. This group also reported 
that sphingomyelin and cholesterol promote HIV-1 gp41 Trp-rich 
region surface aggregation and membrane restructuring [24].

Furthermore, Guillen et al.[25] identified three membrane
regions of SARS-CoV S protein using a 16/18-mer peptide scan;
one of them is located in the Trp-rich region. Peptides corresponding
to this region exert a dramatic effect on leakage for different 
membrane models, suggesting that this region might be involved 
in the promotion of the membrane destabilization required for fusion, 
as well as in fusion pore formation and enlargement.

CryoEM data [22] show that Trp-rich region form the “feet”
interact with the lipid–water interface. New predictive 
approaches including computation of interfacial affinity and the 
corresponding hydrophobic moments also suggest that this 
region is functionally segmented into two consecutive subdo-
mains: one amphipathic at the N-terminal side and one fully 
interfacial at the C-terminus [26]. The N-terminal subdomain 
would extend \( \alpha \)-helices from the preceding carboxyl-terminal 
heptad repeat (HR2) and provide, at the same time, a hydropho-
bic-at-interface surface. Saez-Cirion et al. [23] reported that gp41
Trp-rich peptides have the ability to oligomerize and insert into 
the viral membrane interface. In our study, mutation of three or 
more aromatic residues in this region that functions as “feet”
in pre-fusion state would irreversibly change its hydrophobic mo-
ments and its ability to anchor stably on the membrane to sup-
port the fusion process. Mutation of one or two aromatic amino 
acid residues in this region partially affects the “feet”-like con-
formation and virus infectivity.

In the post-fusion state, removing the bulky indole side chains 
and its weakly charged amine of the Trp residues in the Trp-rich 
region could disrupt its role in synergizing with the fusion peptide [12]
to destabilize both membranes and drive the energetically unfavorable 
lipid–lipid merging that permits the formation and expansion of 
the fusion pore in the late stages of the entry process [17,18].

The fusion process is likely aided by the presence of choles-
terol-binding motifs found in the S protein sequences. Vincent 
et al. and Epad et al. [20,27–30] reported that a consensus 
sequence found in a group of proteins that sequester to choles-
terol-rich regions of membranes has the pattern \(-1/U-(X)(1–5)-
Y-(X)(1–5)-R/K-\), in which \((X)(1–5)\) represents 1–5 residues of 
any amino acid. This sequence can induce formation of choles-
terol-rich domains. The HIV-1 fusion protein gp41 has a LWWK
motif, consistent with this consensus sequence for this protein

Why is this Trp-rich region so important for the infection of 
SARS-CoV and other type I viruses? The role of the Trp-rich region 
in participating in the clustering of gp41 monomer within the HIV-
1 envelope has been shown by Saez-Cirion et al. We thank Prof. Zhang Linqi (Aaron Diamond AIDS Research Cen-
ter, Rockefeller University, New York, New York 10016) for providing plasmids: pNL4-3LucEnv/Vpr– and pCDNA3.1-OPT9-S.

References

[1] T.M. Gallagher, M.J. Buchmeier, Coronavirus spike proteins in viral entry and 
[2] M. Godet, J. Grosclaude, B. Delmas, H. Laude, Major receptor-binding and 
neutralization determinants are located within the same domain of the 
transmissible gastroenteritis virus (coronavirus) spike protein, J. Virol. 68 (1994)
8008–8016.
receptor-binding site within the amino-terminal 330 amino acids of the 
domain of the coronavirus class I viral fusion protein induces membrane 
permeabilization: putative role during viral entry, Biochemistry 44 (2005)
947–958.
Penaranda, B. Bankamp, K. Maher, M.H. Chen, S. Tong, A. Tamin, L. Lowe, M.
Kieff, P.E. Rollin, A. Sanchez, S. Lefkowitz, M. Oakland, M. Olsen-Rasmussen, R.
Foucher, S. Gunther, A.D. Osterhaus, C. Drosten, M.A. Pallen, J.L. Anderson, W.J.
1394–1395.
HIV-1 membrane fusion mechanism: structural studies of the interactions 
[9] I. Munoz-Barroso, S. Durell, K. Sakaguchi, E. Appella, R. Blumenthal, Dilution of 
the human immunodeficiency virus-1 envelope glycoprotein fusion pore 
revealed by the inhibitory action of a synthetic peptide from gp41, J. Cell Biol.
Durell, R. Blumenthal, The HIV Env-mediated fusion reaction, Biochem. 
rich region of the HIV glycoproteins, gp41, forms a well-defined helix in 
interacting sequences within the ectodomain of the human immunodeficiency 
virus type 1 envelope glycoprotein: putative role during viral fusion, J. Virol.
region of the human immunodeficiency virus type 1 glycoprotein: a novel 
[14] W.C. Wimley, S.H. White, Experimentally determined hydrophobicity scale for 
Cortese, R. Bazzo, J.W. Shaver, A. Pessi, Structural analysis of the epitope of the 
anti-HIV antibody 2FS sheds light into its mechanism of neutralization and 
413–423.
[18] S.G. Peisajovich, Y. Shi, Viral fusion proteins: multiple regions contribute to 


