Interaction-mediating sequences within class I viral fusion glycoproteins : their roles in viral infection and in applications

Zhang, Simin

2015

https://hdl.handle.net/10356/65839

https://doi.org/10.32657/10356/65839

NANYANG TECHNOLOGICAL UNIVERSITY

**INTERACTION-MEDIATING SEQUENCES WITHIN
CLASS I VIRAL FUSION GLYCOPROTEINS:
THEIR ROLES IN VIRAL INFECTION AND IN APPLICATIONS**

ZHANG SIMIN

SCHOOL OF BIOLOGICAL SCIENCES

2015

INTERACTION-MEDIATING SEQUENCES WITHIN
CLASS I VIRAL FUSION GLYCOPROTEINS:
THEIR ROLES IN VIRAL INFECTION AND IN APPLICATIONS

ZHANG SIMIN

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

The thesis work has been conducted within a joint PhD programme of
Doctor of Philosophy between Nanyang Technological University of
Singapore and Karolinska Institutet, Stockholm, Sweden

2015

ACKNOWLEDGEMENTS

I am grateful to my friends working outside of science, Zhong Hua, Huang Min, Xiao Han and Elsie, you endured hours and hours of me talking about my projects and my twisted humor of comparing the viral proteins to lollipops. I thank you for knowing these consequences, yet still caring enough to ask me the question, 'How is your project going?'. Ming Wei and James, how I miss the days when we were in Sweden! You have always offered me fresh perspectives on things, which never failed to inspire me both in life and work. Last but most certainly not least, Nick, thank you, for your friendship, support and encouragement.

Words are limited, but not my appreciation.

最后这段话是给妈妈和爸爸的。从小到大，你们给了我最多的爱，最细致的关怀，最及时的鼓励，最大的包容，和最好的教育。谢谢你们。在我的记忆里，你们总在鼓励自己做到最好，总是在奋斗，为自己也是为家人。这本博士毕业论文是我二十三年在学校读书的一个总结。在我看来，不尽完善，做的还不够多也还不够好。可是我想说，在你们身上，从这个过程里，我学会了坚韧和信念的重要。言传身教，耳濡目染，谢谢你们让我有机会在这本论文里写下这段话。

CONTENTS

SUMMARY

Class I viral fusion glycoproteins facilitate fusion of the viral envelope with cell membranes and entry of the virus into the cell, through extensive short sequence-specific interactions. Regions mediating these interactions include the N-terminal hydrophobic fusion peptide, a pair of extended 4,3-hydrophobic heptad repeats (HRs), a membrane-active membrane proximal external region (MPER), a hydrophobic transmembrane domain and the cytoplasmic tail region. In particular, the anti-parallel binding of the C-terminal HR to the central N-HR trimeric coiled-coil forms the 6-helix bundle fusion core. These interaction-mediating sequences are generally well preserved sequentially and structurally, allowing their peptidyl analogues to be developed as antiviral therapeutics and/or research reagents (e.g. HR-derived peptides). Novel targets for the development of antiviral drugs and viral detection reagents are required when facing drug-resistant viral strains, viral pathogens without effective and/or economical treatment, and newly emerging viral pathogens. This thesis focuses on the systematic identification of novel interaction-mediating sequences within Class I viral fusion glycoproteins, and the investigation of their involvements in viral replication as well as their potential applications in diagnosis and anti-viral interventions.

In Paper I, peptide array scanning identified 34 spike (S) protein-derived peptides that bound to the S protein of severe acute respiratory syndrome-associated coronavirus (SARS-CoV). These putative self-binding peptides contain five core octapeptide consensus sequences, among which the octapeptide GINITNFR was predicted to form β-zipper-mediated amyloid-like fibrils. The peptide C6 containing this sequence was subsequently shown to oligomerize and form amyloid-like fibrils. The potential of C6 to conduct β-zipper-mediated interactions was further applied to detect the S protein expression by immunofluorescence staining. The peptide array scanning in Paper I used the S protein ectodomain without the MPER and beyond. Using chemical crosslinking and immunofluorescence staining, in Paper II we could show that the S protein MPER could oligomerize and further heteromerize with the N-terminal internal fusion peptide (IFP). The MPER-derived peptides also inhibited the coronavirus entry in a dose-dependent manner, potentially through disrupting the MPER-mediated interactions. The ability of peptides derived from the MPER in inhibiting viral entry and infection was subsequently studied in Paper III, in the context of HIV-1. The antiviral activities of the HIV-1 Env MPER-derived peptides were abrogated upon Ala substitution of the Trp residues or deletion of the C-terminal cholesterol-interacting motif. Unexpectedly, Ala substitutions of the Trp residues within HIV-1 Env significantly elevated the biosynthesis of another viral structural protein, the p55/Gag, which led to enhanced viral particle release. In Paper IV, besides the MPER we identified the signal sequence of HIV-1 Env as another region that could negatively regulate the expression of p55/Gag. The HIV-1 Env signal sequence, which mediates the co-translational translocalization of nascent Env polypeptide into the endoplasmic reticulum, inhibited the viral protein expression and production, probably at a post-ER-targeting stage. N-terminal truncations of the Env signal sequence significantly elevated the intracellular and intraviral levels of late viral proteins and the proviral genome transcription in a time- and dose-dependent fashion. Moreover, the truncations suppressed the HIV-1 promotor (LTR)-driven expression of a reporter protein, suggesting that the Env signal sequence inhibited viral genome transcription through LTR-dependent interactions.

This thesis focused on three novel interaction-mediating sequences within two Class I viral fusion glycoproteins, which could regulate the viral infectivity, at both viral entry and assembly, through protein-protein, protein-lipid, and/or protein-nucleic acids interactions. These sequences and the interactions that they are meditating could be further targeted by their peptidyl analogues for viral detection and/or inhibition.

LIST OF PUBLICATIONS

I. Si Min Zhang, Ying Liao, Tuan Ling Neo, Yanning Lu, Ding Xiang Liu, Anders Vahlne, James P. Tam. **Identification and applications of self-binding zipper-like sequences in SARS-CoV spike protein.** *Manuscript.*

II. Ying Liao[‡], Si Min Zhang[‡], Tuan Ling Neo, James P. Tam. **Tryptophan-dependent membrane interaction and heteromerization with internal fusion peptide by membrane proximal external region of SARS-CoV spike protein**. *Biochemistry. 2015; 54 (9): 1819−1830.* [‡]These authors contributed equally.

III. Si Min Zhang, Alenka Jejcic, James P. Tam, Anders Vahlne. **Membrane-active Sequences within gp41 Membrane Proximal External Region (MPER) Modulate MPER-containing Peptidyl Fusion Inhibitor Activity and the Biosynthesis of HIV-1 Structural Proteins.** *PLoS ONE. 2015; 10(7): e0134851.*

IV. Si Min Zhang, Anders Vahlne, Alenka Jejcic. **The gp160 Signal Sequence Suppresses HIV-1 Replication in an LTR-Dependent Manner.** *Manuscript.*

LIST OF ABBREVIATIONS

6-HB	Six helix bundle
aa	Amino acid
ACE2	Angiotensin converting enzyme 2
AIDS	Acquired immunodeficiency syndrome
Ala	Alanine
APOBEC3	Apolipoprotein B mRNA editing complex 3
C	Conserved regions
CAT	Chloramphenicol acetyltransferase
CD	Circular dichroism
CRAC	Cholesterol recognition amino acid consensus
-DC	C-terminal dimer
DMEM	Dulbecco modified Eagle medium
-DN	N-terminal dimer
Env	Envelope
env	Envelope gene
ER	Endoplasmic reticulum
FIV	Feline immunodeficiency virus
FDA	Food and Drug Administration
FP	Fusion peptide
Gag	Group specific antigen
gag	Group specific antigen gene
Gag-Pol	Gag-polymerase
gag-pol	Gag-polymerase gene
Glut	Glutaraldehyde
HAART	Highly active antiretroviral therapy
HIV-1	Human immunodeficiency virus type 1
HR	Heptad repeat
IFP	Internal fusion peptide
LTR	Long terminal repeat
MBP	Maltose binding protein

MERS-CoV	Middle East respiratory syndrome coronavirus
MHC-1	histocompatibility complex class I
MPER	Membrane proximal external region
Nef	Negative regulatory factor protein
nef	Negative regulatory factor gene
ORF	Open reading frame
PBD	Pocket binding domain
Phe	Phenyloalanine
R	Repeated
RBD	Receptor binding domain
RBM	Receptor binding motif
Rev	Regulator of virion proteins
rev	Regulator of virion proteins gene
RRE	Rev response element
RT-qPCR	Real time quantitative reverse transcription polymerase chain reaction
S	Spike
SARS	Severe acute respiratory syndrome
SARS-CoV	SARS-associated coronavirus
SPase	Signal peptidase
ss	Signal sequence
Tat	Viral transcription transactivator
tat	Viral transcription transactivator gene
TFE	2,2,2-Trifluoroethanol
ThT	Thioflavin T
TM	Transmembrane
Trp	Tryptophan
Tyr	Tyrosine
U3	Unique, 3' end
U5	Unique, 5' end
UTR	Untranslated region

V	Variable regions
Vif	Virion infectivity factor
vif	Virion infectivity factor gene
Vpr	Viral protein R
vpr	Viral protein R gene
Vpu	Viral protein U
vpu	Viral protein U gene
WT	Wild type

1 INTRODUCTION

1.1 VIRAL FUSION GLYCOPROTEINS

Enveloped viruses deliver their contents into target cells through envelope-plasma membrane fusion, which is mediated by the viral fusion glycoproteins embedded in the viral envelopes. During this process, the fusion glycoproteins conduct extensive protein-protein and protein-membrane interactions through a common step-wise scheme that involves the adoption of an essential 'trimer-of-hairpin' tertiary conformation [1-3]. Upon being triggered by receptor binding, and for some viruses by a change in pH, the metastable fusion-competent viral fusion glycoproteins undergo extensive structural rearrangements. They progressively transform from the prefusion conformation to a trimeric pre-hairpin intermediate that connects the opposing membranes through a hydrophobic fusion peptide/loop, and then to the final rod-like trimer-of-hairpin conformation where three C-terminal regions pack onto a central N-terminal trimeric core [4, 5]. These structural rearrangements and extensive interactions mediated by the viral fusion glycoprotein ultimately provide the free energy for the juxtaposition of the opposing membranes and the thereafter formation of the membrane stalk, the hemifusion intermediate, and eventually fusion pores.

Despite the common mechanism in mediating membrane fusion, the viral fusion glycoproteins differ in their orientation on the viral surface, their major secondary structure, oligomeric state, location of the fusion peptide/loop, and most importantly, the structure of their trimer-of-hairpin inner cores [4]. Based on these differences, they are divided into three classes (Table 1). The Class I viral fusion glycoproteins encompass the envelope proteins of many medically important viruses including influenza virus [6], human immunodeficiency virus type 1 (HIV-1) [7] and severe acute respiratory syndrome-associated coronavirus (SARS-CoV) [8]. They are mostly synthesized as type I transmembrane precursor proteins and later oligomerize into trimeric peplomers, which are anchored perpendicularly from the plasma membrane, and after budding the viral envelope, through a single-pass C-terminal transmembrane domain. Based on the proteolytic cleavage site, the Class I viral fusion glycoproteins can be divided into two functional subunits, namely a surface subunit that interacts with the viral receptor and a fusion subunit that contains two Leu/Ile zipper heptad repeat (HR) regions. During viral entry, the HRs of the fusion protein trimers are responsible for forming a six-helix-bundle (6-HB) trimer-of-hairpin fusion core, where the three helical C-terminal HRs pack onto the grooves of a central coiled-coil formed by the three N-terminal HRs. The Class II viral fusion glycoproteins mostly contain surface proteins of the flaviviruses, alphaviruses and bunyaviruses [9, 10]. The Class II viral fusion glycoproteins are presented along the viral surfaces as dimers. During membrane fusion, their internal fusion loops induce the structural realignment into trimer-of-hairpins consisting of mostly β-structure core. Finally, Class III viral fusion glycoproteins have combined features of the Class I and II

fusion glycoproteins and are found on rhabdoviruses, herpesviruses and alphabaculoviruses [4].

The Class I viral fusion glycoproteins are the focus of this thesis, and their fusion mechanisms will be introduced in detail in the following sections with the prototype proteins, the SARS-CoV spike (S) protein and the HIV-1 envelope (Env) protein.

Table 1. Viral fusion glycoprotein classification (Adapted from [4]).

Features	Class I	Class II	Class III
Examples	Influenza HA, HIV-1 Env	TBEV E, SFV E1/E2	VSV G, HSV-1 gB
Orientation with respect to viral membranes	Perpendicular (spike-like)	Parallel (close to the viral membrane)	VSV G, perpendicular HSV-1 gB, unknown
Major secondary structure (of the fusion subunit)	α-helix	β-sheet	α-helix and β-sheet
Prefusion oligomerization state	Trimer	Dimer	VSV G, trimer
FP location	Near the N-terminus of the fusion subunit	Internal	Internal (bipartite)
Fusion-active oligomerization state	Trimer	Trimer	Trimer
Fusion core structure	Trimer-of-hairpin (6-HB)	Trimer-of-hairpin (mainly β structure)	Trimer-of-hairpin (central α-helical coiled coil accompanied with significant β-structure)

Influenza HA, influencza hemagglutinin protein; TBEV E, tick-borne encephalitis virus envelope protein; SFV E1/E2, Semliki forest virus envelope protein E1/E2; VSV G, vesicular stomatitis virus G protein; HSV-1 gB, herpes simplex virus glycoprotein B.

1.2 THE CLASS I FUSION GLYCOPROTEIN OF SARS-COV

1.2.1 Severe acute respiratory syndrome (SARS)

In November 2002, a large outbreak of an atypical and severe pneumonia named the severe acute respiratory syndrome, or SARS, was first reported in Guangdong, China. Person-to-person transmission *via* infectious respiratory droplets and international travelling of infected individuals quickly led the illness to spread through 25 countries and five continents [11-13]. Starting as an influenza-like syndrome with fatigue, rigors and high fevers, SARS could progress to an atypical pneumonia in two-thirds of the infected individuals. The destruction of the alveolar and bronchial epithelial cells, and the increased production of pro-inflammatory cytokines and chemokines may collectively result in respiratory failure, the most common cause of death among

the infected. Besides the upper and lower respiratory tract symptoms, extra-pulmonary manifestations, such as liver and renal impairment, bradycardia, hypotension and lymphopenia, were also reported in SARS patients [14]. Eventually, SARS resulted in 8,437 infected individuals and 813 deaths, between November 2002 and July 2003, with a fatality rate of 9 to 12% and higher than 50% among elders over the age of 65 [15].

1.2.2 SARS-CoV

1.2.2.1   Identification of SARS-CoV

In April 2003, based on the results from serologic analysis, molecular analysis and microscopic studies, a new member from the Betacoronavirus genus, Coronavirinae subfamily and the Coronaviridae family was identified as the causative agent of SARS and was named the SARS-associated coronavirus (SARS-CoV) [16-19]. Coronaviruses are enveloped, positive-stranded RNA viruses that can infect a variety of vertebrates with respiratory and enteric diseases. They were primarily known for their adverse effects on economically important live-stocks, a picture that has been changed since the identification of SARS-CoV, and more recently another deadly human pathogen, the Middle East respiratory syndrome coronavirus (MERS-CoV) [20].

1.2.2.2   SARS-CoV particle morphology and structure

SARS-CoV shares similar viral morphology, genome structure and replication cycle with other coronaviruses. The enveloped SARS-CoV is coated with three structural proteins, the spike (S) glycoprotein, the membrane glycoprotein and the envelope protein [17]. The S proteins form the clubbed-shaped trimers on the viral surface, giving the SARS-CoV the characteristic crown-like (corona) appearance (Figure 1A-B) [17]. The membrane protein interacts with the fourth structural protein, the nucleocapsid protein, which coats the viral RNA genome (Figure 1C).
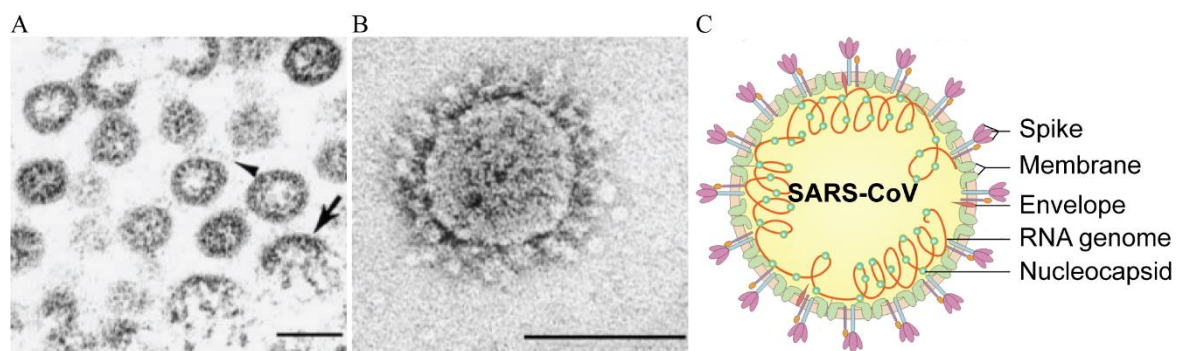

Figure 1. SARS-CoV viral morphology and structure. (A) SARS-CoV particles containing club-shaped peplomer structure on the viral surface (arrowhead) are aligned long the rough endoplasmic reticulum (ER) (arrow) as the virus bud into the cisternae [17]. **(B)** SARS-CoV shows an internal helical nucleocapsid-like structure and club-shaped surface projections. The

bar in (A) and (B) represented 100 nm [17]. **(C)** Schematic representation of the SARS-CoV viral structure.

1.2.2.3 SARS-CoV genome

The SARS-CoV genome has the characteristic features of a coronavirus genome [21] (Figure 2). The ~27 kb, 5'-capped and 3'-polyadenylated RNA genome has a 5' leader sequence, a 5' untranslated region, two protein-coding regions and a second 3' untranslated regions [22, 23]. The first protein-coding region contains two overlapping open reading frames that are translated into a single polyprotein by a -1 ribosomal frameshift. This polyprotein is further processed into the proteins required for transcription and replication. The second region encodes the four structural proteins and also contains the accessory genes.
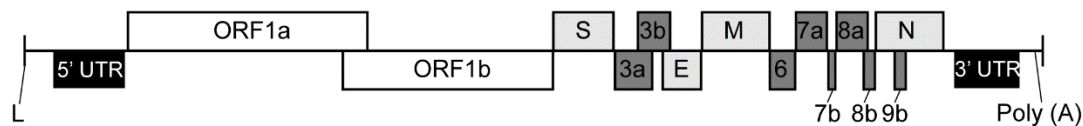
Figure 2. Schematic representation of the genome structure of SARS-CoV. Black boxes represent the 5' and 3' untranslated regions (UTR), followed by the two overlapping open reading frame (ORF) 1a and 1b in white boxes. The grey boxes represent the coding regions for the structural proteins, the S, envelope (E), membrane (M) and nucleocapsid (N) proteins. The dark grey boxes represent the ORF encoding the accessory proteins. The positions of the leader sequence (L) and poly(A) tract are indicated.

1.2.2.4 SARS-CoV replication cycle

The replication cycle of SARS-CoV is initiated by the receptor recognition by the S protein and thereby the attachment of the virus onto the target cells [21, 24, 25] (Figure 3). The angiotensin converting enzyme 2 (ACE2) is the major viral receptor and is present in the tissues, which are susceptible to SARS-CoV infection. They include the heart, lung, kidney, and gastrointestinal tract [26, 27]. Although less effectively, SARS-CoV also can use CD209L/L-SIGN to enter the host cells [28]. The S protein-receptor interaction triggers the endocytosis of viral particles into early endosomes, where the low pH and proteolytic processing of the S protein by cathepsin L activates the Class I viral fusion protein machinery within the S protein, leading to the fusion of the envelope with the cell membrane and the release of the viral content into the cytoplasm [29-31]. The two 5' overlapping open reading frames within the viral genomic RNA are then translated into a viral replicase polyprotein. Auto-cleavage of the viral replicase polyprotein leads to the assembly of the viral replication complex, which facilitates the synthesis of the nested subgenomic mRNA transcripts and the viral genome replication [32]. The subgenomic mRNA is then

translated into viral structural proteins, i.e. the S protein, membrane protein, envelope protein and the nucleocapsid protein. The nucleocapsid protein further associates with the viral genomic RNA to form the nucleocapsid. The nucleocapsid then acquires the viral envelope by budding through the intracellular membranes between the endoplasmic reticulum (ER) and the Golgi apparatus, where the structural proteins translated in the ER are converged. Finally, the virions travel through the Golgi apparatus where they receive glycosylation modifications of the viral glycoproteins, and eventually reach the cell surface where they are released *via* exocytosis.
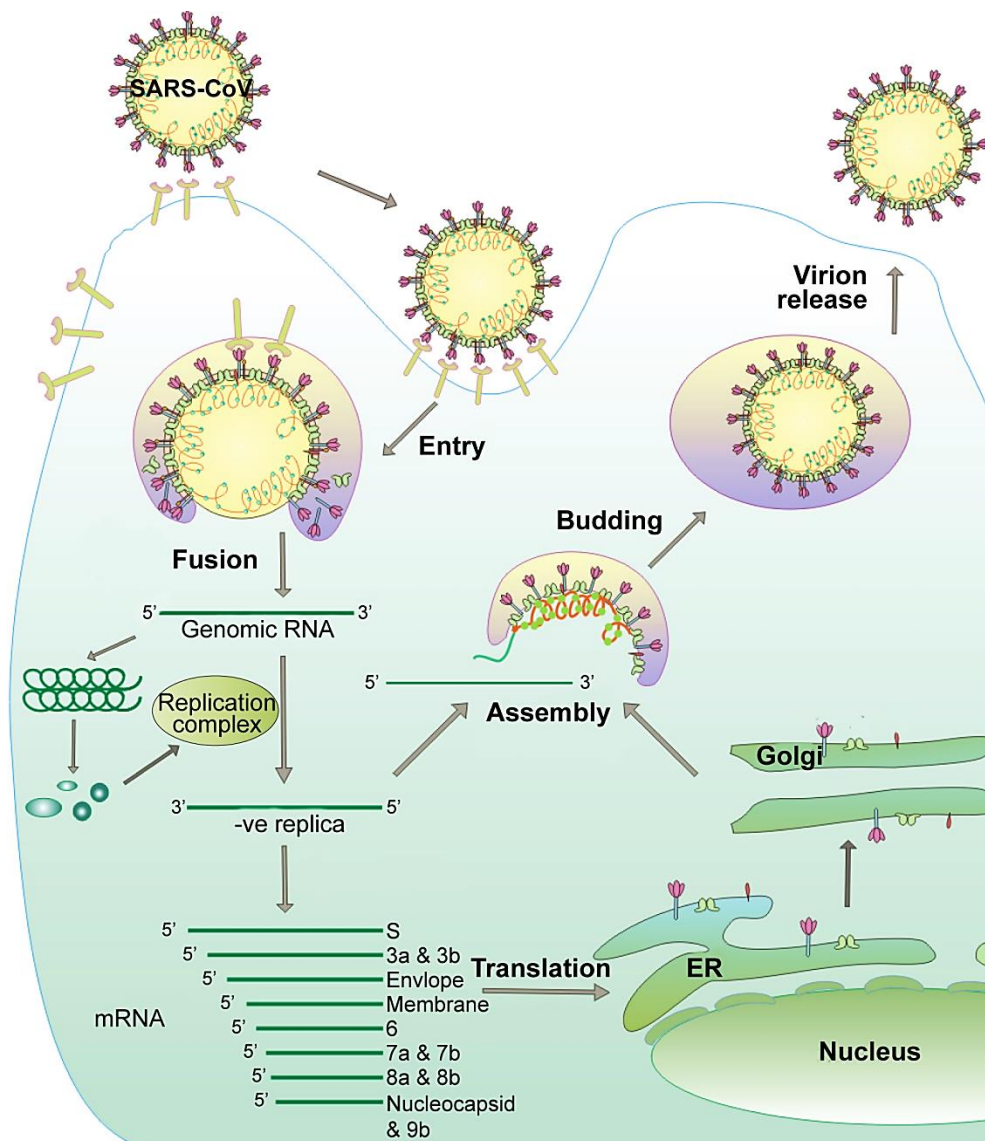
Figure 3. Schematic representation of the coronavirus replication cycle.

1.2.2.5 Anti-SARS treatment

Ribavirin, a guanosine analogue previously used for the treatment of hepatitis C virus, another RNA virus, was prescribed alone or in combination with corticosteroid or the

two HIV-1 protease inhibitors, lopinavir and ritonavir [13]. However, the efficacy of this treatment remains to be confirmed, due to the lack of controlled clinical trials and the inconclusive results generated from *in vitro* antiviral assays using different cell lines [14]. Traditional Chinese medicine was also prescribed in mainland China, but no systematic studies were conducted to evaluate the treatment outcome. Nevertheless, treatment with interferon-α and/or –β have shown some antiviral efficacy. Rational design of anti-SARS drugs have also made some progress with the study on cyclosporin A, a 'pan-coronavirus inhibitor' that have exhibited antiviral effects against a large range of coronaviruses [33].

1.2.3 SARS-CoV S protein

1.2.3.1 SARS-CoV S protein biosynthesis and domains

The SARS-CoV S protein is a 1255 amino acid (aa) long, Class I viral fusion glycoprotein synthesized following the secretory pathway. The N-terminal signal peptide (1-13 aa) of the S protein directs its co-translational translocation into the rough ER, where it is trimerized and extensively glycosylated on 23 putative N-glycosylation sites [23, 34, 35]. Upon the correct folding and trimerization in the ER, the S protein precursor is transported to Golgi apparatus, where it acquires the EndoH-resistant complex N-glycans [34]. Eventually, trimers of the S proteins are positioned on the viral membrane with the bulky N-terminal parts projecting out of the virion, which form the characteristic club-shaped trimeric peplomers as observed by electron microscopy (Figure 4A).

Unlike the spike proteins of most coronaviruses, SARS-CoV S protein is not proteolytically processed in the Golgi [34]. Instead, it is cleaved and activated during the viral entry, by the endosomal enzymes cathepsin L after its Thr678 or by trypsin at Arg792 [36, 37]. Based on the cathespsin L cleavage site, the S protein is arbitrarily divided into the surface S1 subunit and the transmembrane S2 subunit (Figure 4B). The S1 subunit is composed of a signal peptide (1-13 aa) at its N terminus, followed by a receptor binding domain (RBD) (318-510 aa) containing the receptor binding motif (RBM) (424-494 aa) [24, 38-40]. The S2 subunit contains two N-terminal hydrophobic domains, a fusion peptide (FP) domain (770-788 aa), an internal fusion peptide (IFP) domain (873-888 aa); two heptad repeat (HR) regions containing Leu/Ile zipper sequences, the HR1 (892-1013 aa) and HR2 (1145-1190 aa); a hydrophobic membrane proximal external region (MPER), a transmembrane (TM) domain and a C-terminal cytoplasmic tail [41-45].
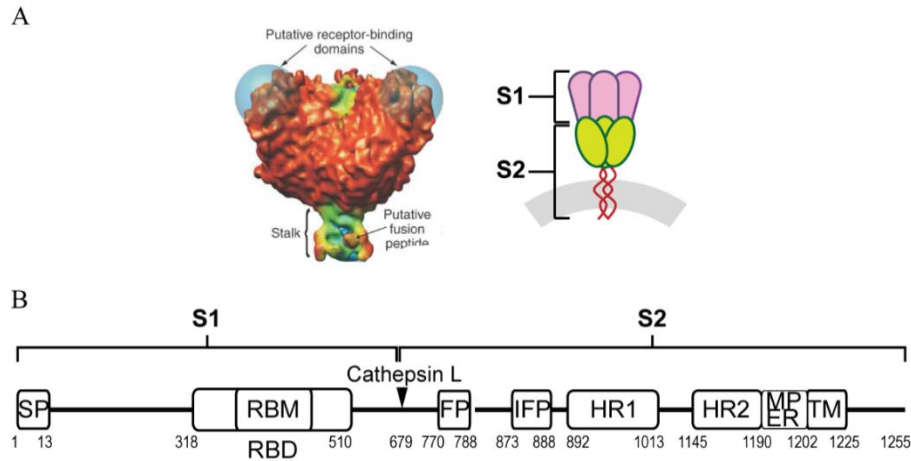
Figure 4. SARS-CoV S protein architecture and domains. (A) Atomic resolution structure reconstructed from cryo-electron microscope images [46] (left), and the schematic representation (right) of the prefusion S protein trimers **(B)** Schematic representation of the S protein domains.

1.2.3.2   S protein-mediated viral entry

The S protein mediates the envelope-cell membrane fusion and SARS-CoV entry through a Class I viral fusion mechanism [8, 47]. To execute fusion between the two closely positioned membranes, lipid bilayers need to be mixed and remodeled to form a hemifusion stalk, hemifusion diaphragm and eventually a fusion pore [48] (Figure 5). However, the membrane lipid bilayer is stabilized by a powerful hydrophobic effect, which acts against any structural changes. Any membrane remodeling would require energy investment, which comes from thermal fluctuations of the membrane itself and/or specialized protein involvement.
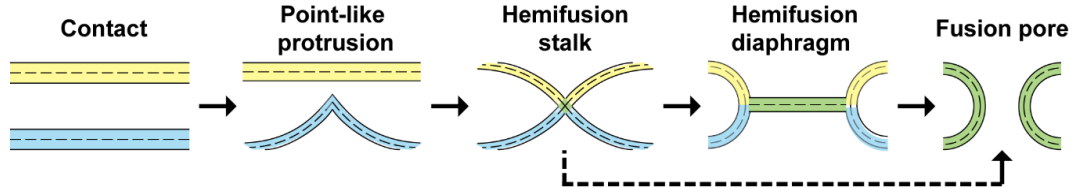
Figure 5. Schematic representation of the membrane fusion process.

The Class I viral fusion glycoproteins facilitate the membrane fusion through extensive protein-protein and protein-membrane interactions, which involve the formation of a 6-HB fusion core (Figure 6). Firstly, the RBD within the S1 subunit recognizes and interacts with the primary viral receptor ACE2 and induces the endocytosis of the virions [38]. Within the endosomes, the S protein will be activated by the low pH and/or the proteolytic cleavage by the endosomal enzymes such as cathepsin L [49]. This leads to structural rearrangements within the S protein and the exposure of the FP region situated at the S2 subunit N-terminus. Subsequently, it is

believed that the exposed FP inserts into the target cell membrane, leading the S protein to adopt a pre-hairpin intermediate where the two HRs are extended and transiently exposed. It has been suggested that the HR2 adopts a trimeric helical structure at this stage [50]. The pre-hairpin intermediate would slowly but eventually resolve into a more energetically favorable state by folding the C-terminal HR2 anti-parallel onto the N-terminal HR1. This leads to the formation of the 6-HB fusion core, where each of the HR2 helix binds to each of the three grooves of the trimeric HR1 coiled-coil. It is believed that the formation of the 6-HB provides the free energy to juxtapose the opposing membranes that are kept separated by the hydrophobic effect, which generates a locally anhydrous environment facilitating the subsequent lipid mixing and membrane fusion. By the end of the lipid mixing, according to the Class I membrane fusion model, the N- and the C-termini of the S proteins are located on the same side of the fused membrane in the post-fusion conformation.
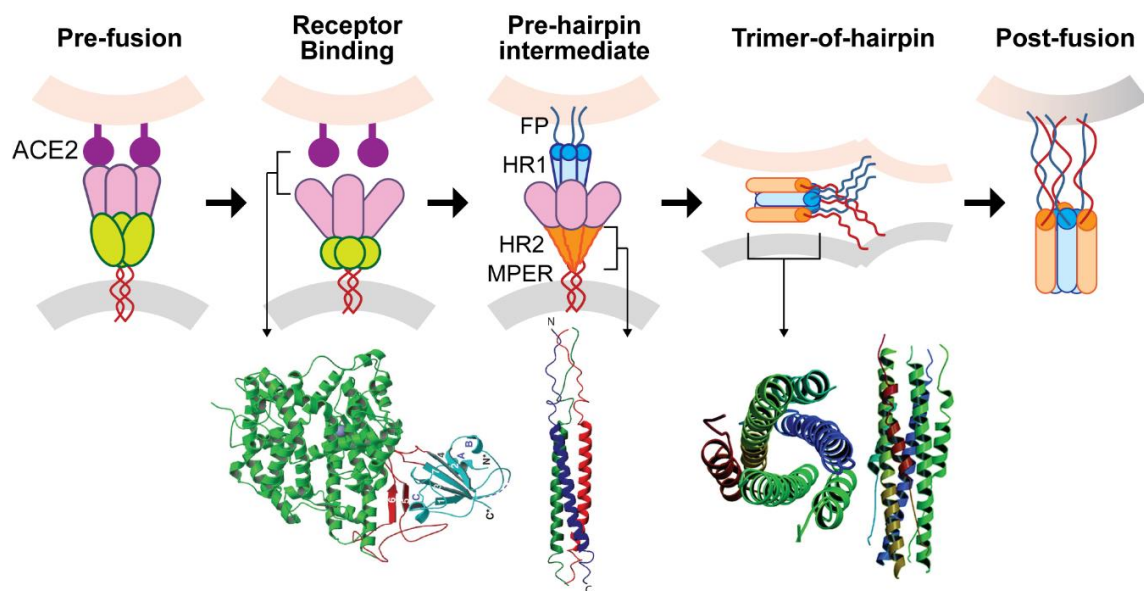
Figure 6. SARS-CoV S protein mediates the viral entry and membrane fusion *via* the Class I viral fusion mechanism. The entry scheme is presented with the structures of the ACE2-RBD complex [38], the prefusion HR2 (1141-1193 aa) [50] and the trimer-of-hairpin fusion core (HR1: 898-1005 aa, HR2: 1145-184 aa) [47]. Within the ACE2-RBD structure, green, ACE2; red, RBM; blue, the rest of the RBD. Within the trimeric HR2 structure, red, blue and green each represents a HR2 helix.

1.3 THE CLASS I FUSION GLYCOPROTEIN OF HIV-1

Although HIV-1 mainly infects CD4+ T lymphocytes, it can also infect macrophages, monocytes and dendritic cells. As the HIV-1 infection progresses, it depletes and exhausts the CD4+ T lymphocytes, thereby weakening the immune system, allowing opportunistic infections to occur, eventually resulting in a life-threatening

immunodeficiency disease named the acquired immunodeficiency syndrome or AIDS.

1.3.1 Acquired immunodeficiency syndrome (AIDS)

1.3.1.1 The AIDS epidemic

Since 1982, when AIDS was first officially recognized and named by the U.S. Centers for Disease Control and Prevention, there has been approximately 36 million AIDS-related deaths, and presently 35.3 million people are living with HIV-1 infection [51]. In 2012, the UNAIDS reported 2.3 million new HIV infections and 1.6 million AIDS deaths globally [51]. The epidemic has substantial effects in sub-Saharan Africa, where 24.7 million people were estimated to live with HIV in 2013. More recently, HIV-1 infection has become a major concern also in East Asia, Central Asia and Eastern Europe.

1.3.1.2 HIV transmission and pathogenesis

HIV-1 was isolated and identified as the etiological agent of AIDS between 1983 and 1984 [52-56]. In 1986, HIV-2, a closely related virus to HIV-1, was identified as the second AIDS-causing agent [57]. So far, most of the HIV infections worldwide are with HIV-1, while the HIV-2 infections are confined mainly to West Africa.

HIV can infect the host through free viral particles and/or infected cells in the blood, semen, vaginal fluids and breast milk. It is mostly transmitted through sexual intercourse, intravenous drug abuse and by mother-to-child transmission. Although the rate of progression varies greatly among infected individuals, an HIV infection generally progresses through three stages, the primary/acute infection, the asymptomatic infection, and finally AIDS [58]. The acute HIV infection stage generally lasts 2 to 10 weeks after the initial infection. During this period, HIV replicates rapidly and acutely depletes the CD4+ T lymphocytes in the gut-associated lymphoid tissues, and the infected individuals usually experience no symptom or a brief flu-like illness [59]. Subsequently, the host immune system responds and generates anti-HIV antibodies and cytotoxic lymphocytes, which subsides the viral load to a lower stable level and leads the infection to the second asymptomatic stage. At this clinically latent stage, the infected individual experiences minimal symptoms and could remain apparently healthy for an average of 10 years [60]. However, the CD4+ T cells counts continuously drop and thus the immune system gradually becomes weaker. When CD4+ T lymphocytes counts drop to fewer than 200 cells /µl, symptomatic opportunistic infections appear and the HIV-1 infections enter the last stage AIDS.

1.3.2 HIV-1

1.3.2.1 HIV-1 classification

HIV-1 is classified in the Lentiviruses genus of the Orthoretrovirinae subfamily of the Retroviridae, which is characterized by a long latency period and a slow onset of symptoms. Among the four HIV-1 groups, M, N, O and P, Group M is the major circulating group responsible for the global epidemic. It is divided into more than 8 subtypes that could further generate between-subgroup recombinants [61].

1.3.2.2 HIV-1 viral genome, proteins and morphology

Each HIV-1 viral particle contains two copies of a 9.2 kb single-stranded and positive-sensed RNA genome (Figure 7), which is reversely transcribed and integrated into the host genome upon viral infection [62]. The resulting provirus within the host genome is flanked by two long terminal repeats (LTRs) each composed of an U3 (unique, 3' end), a R (repeated) and an U5 (unique, 5' end) region [63]. The 5' LTR directs the transcription and the generation of the viral RNA transcripts, while the 3' LTR serves as the transcription termination signal [64]. Transcription of the viral genome generates the full-length viral RNA transcripts that either remain unspliced or spliced to different extent through an alternative splicing mechanism (Figure 7).

The viral mRNA transcripts eventually give rise to fifteen viral proteins. The ~9kb full-length mRNA transcript is used to express p55/Gag (group-specific antigen) and, upon a -1 ribosomal shift, the p160/Gag-Pol (Gag-polymerase) polyproteins. The ~4kb incompletely spliced mRNA transcripts, depending on the splicing sites, are used to express the Env (envelope protein), Vif (virion infectivity factor), Vpr (virion protein R), Vpu (virion protein U) and the one-exon Tat (trans-activator of transcription) proteins. The ~1.8kb completely spliced mRNA transcripts are used to express the Rev (regulator of virion protein expression), Nef (negative regulatory factor) and the two-exon Tat.
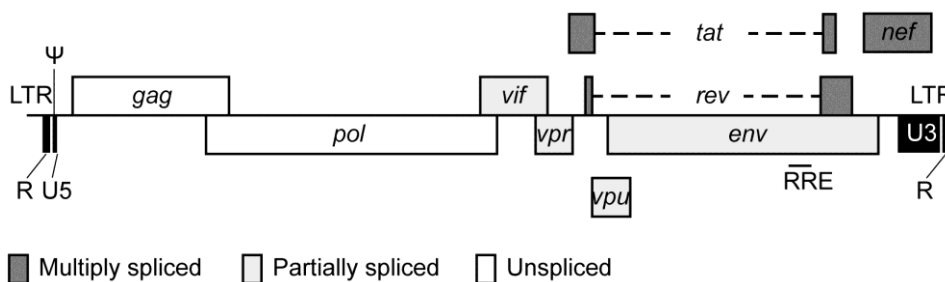
Figure 7. HIV-1 RNA genome structure. Transcription of the HIV-1 proviral genome within the host genome generates the full-length viral mRNA transcripts, which either remain unspliced or spliced to different extent through alternative splicing. ORFs encoding different

viral proteins are colored according to the extent of splicing. The packaging signal (Ψ) that regulates the packaging of the RNA genome into the viral capsid is depicted.

The fifteen HIV-1 proteins are grouped into the structural, the accessory and the regulatory proteins. The viral structural proteins are expressed as the precursor Env, p55/Gag, and p160/Gag-Pol proteins. They are the essential parts of the HIV-1 viral particles. When the viral particles are first assembled, arrays of the p55/Gag polyproteins are connected with the viral envelope, which gives the immature viral particle a characteristic doughnut shape when observed by electron microscopy (Figure 8A) [65]. During or soon after viral budding, the viral protease within p160/Gag-pol auto-cleaves itself and thereafter processes the p55/Gag into the matrix, capsid, nucleocapsid and p6, and cleaves the p160/Gag-pol additionally into the reverse transcriptase and integrase [66-68]. This leads to the rearrangement of the p55/Gag-derived structural proteins and the maturation of the viral particles, characterized by a cone-like structure within the viral particle (Figure 8A, B) [65, 69, 70]. Within the mature HIV-1 particle, two copies of the viral RNA genome are each encapsulated by nucleocapsid proteins and are further dimerized. The resulting nucleocapsid, together with the viral integrase and reverse transcriptase, are subsequently surrounded by a cone-shaped structure composed of the capsid proteins and a spherical shell composed of the matrix protein. The viral capsid is eventually enveloped by the host-derived viral lipid bilayers, which is studded with the trimeric Env proteins composed of non-covalent heterodimers of the surface gp120 protein and the transmembrane gp41 protein.
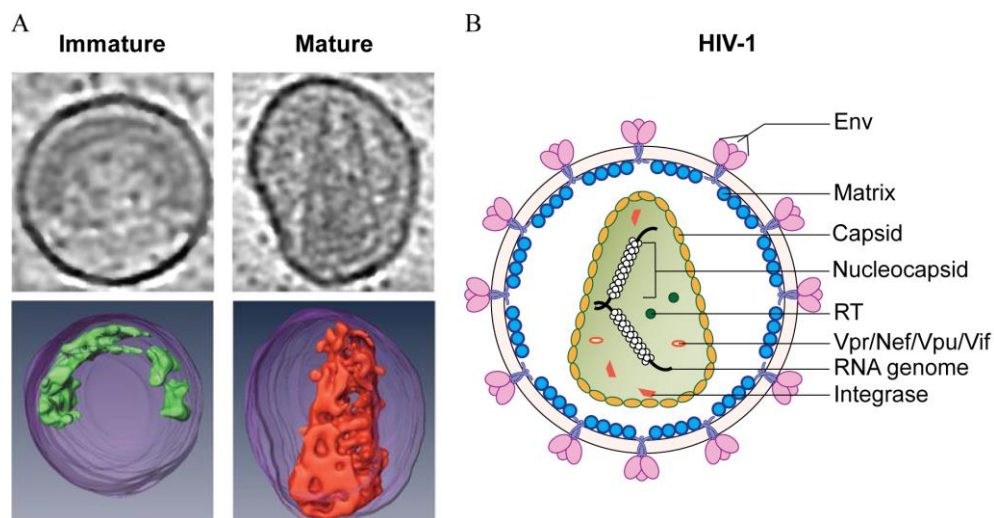
Figure 8. HIV-1 viral morphology. (A) HIV-1 morphological change during viral maturation [65]. Top row, single tomographic slices through the particle volume showing the complete views of the capsid structures; bottom row, full 3D segmentation of the entire particle volume. Purple, membrane; green, immature capsid; red, mature capsid **(B)** Schematic representation of the mature HIV-1 virion.

The regulatory proteins include Tat and Rev. Tat is essential for activating the transcription of the HIV-1 provirus. Rev regulates the post-transcriptional alternative splicing of the viral transcripts.

The accessory proteins, Vpu, Vif, Vpr and Nef, regulate the viral infectivity through interacting with various cellular proteins, which are further elaborated in the next section 'HIV-1 replication cycle'. Vpr, Nef and variable amounts of Vpu and Vif have been detected in the HIV-1 virions [71-74].

1.3.2.3   HIV-1 replication cycle

The HIV-1 replication cycle is initiated by the recognition of the viral receptors, CD4, by the surface gp120 unit of the Env proteins (Figure 9) [75, 76]. This induces structural rearrangement within gp120, allowing its interaction with the co-receptor, CCR5 or CXCR4 [77-82]. This in turn introduces further conformational changes, which exposes the transmembrane gp41 unit of Env and activates the Class I fusion mechanism within gp41 [83, 84]. It ultimately leads to the fusion between the viral and the cellular membranes, and the release of the viral capsid into the host cells. Within the released nucleocapsid, the viral RNA is reversely transcribed by the viral reverse transcriptase [85]. The viral reverse transcriptase lacks proofreading ability and its error-prone reverse transcription introduces mutations within the HIV-1 provirus and results in a high genetic variability of HIV-1 [86]. After the reverse transcription, a pre-integration complex is formed by the double-stranded viral DNA, the matrix protein, the integrase and some host factors and is transported into the host nucleus [87, 88]. Within the nucleus, the viral DNA is integrated into the host genome by the viral integrase.
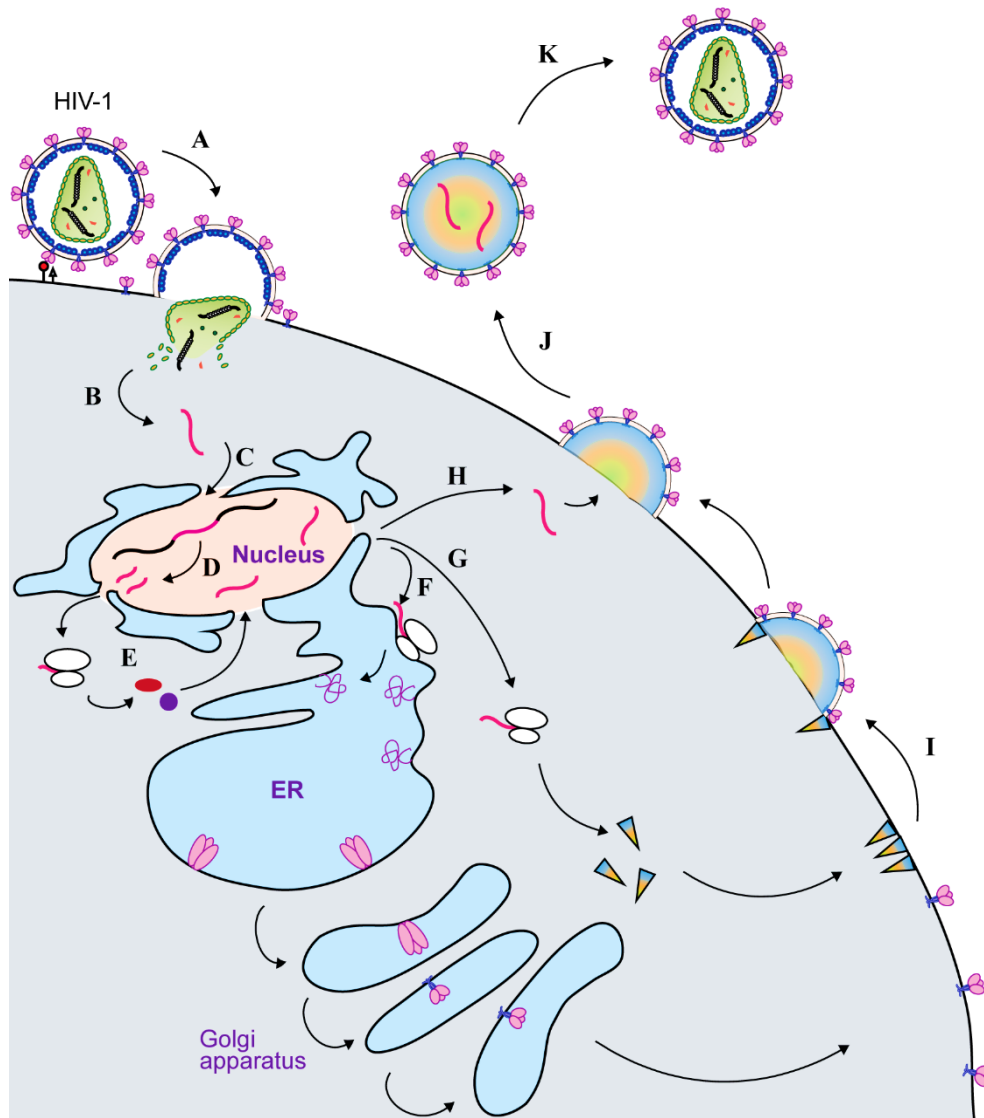
Figure 9. HIV-1 replication cycle. (A) Env binding to the receptor CD4 and the co-receptor CCR5/CXCR4 initiates the HIV-1 replication cycle. **(B)** Receptor binding triggers series of structural rearrangement within Env, which lead to the fusion between the viral and cell membranes, and to the release of viral capsid into the host cells. Within the released viral capsid, the viral RNA genome is reversely transcribed into double-stranded DNA. **(C)** The synthesized viral DNA, together with the matrix, integrase and host factors, is transported into the host nucleus, where the viral DNA is integrated into the host genome. **(D)** Provirus genome is transcribed at a baseline level. **(E)** The multiply spliced mRNA transcripts are then exported to the cytoplasm for the translation of viral protein Tat, Rev and Nef. Tat and Rev are then transported back to the nucleus, where Tat boosts the provirus transcription and Rev rescues the mRNA transcripts from multiple splicing. **(F)** The partially or un-spliced mRNA is used for the synthesis of Env, **(G)** p55/Gag and p160/Gag-pol and for the **(H)** viral assembly. **(I)** The synthesized structural proteins Env, p55/Gag and p160/Gag-pol converge onto the viral assembly site. **(J) — (K)** During, or soon after viral budding, the protease autocleaves from the p160/Gag-pol, leading to the viral maturation.

After integration, baseline levels of viral proviral mRNA are transcribed, multiply spliced and used to express the early viral proteins Tat, Rev and Nef. Tat and Rev are then transported back to the nucleus. Tat binds to the transactivation response region located within the 5' LTR region of the nascent viral mRNA, which greatly enhances

the processivity of the RNA polymerase II and hence boosts the transcription of the HIV-1 provirus [89-92]. Meanwhile, the Rev protein targets the Rev response element (RRE) within the *env* gene, binds to the partially or un-spliced RNA transcripts, and rescues them from multiple splicing by facilitating their export out from the nucleus [93].

The unspliced RNA transcript is packaged as the genomic material into the viral particles and is also used to express the p55/Gag and p160/Gag-pol proteins by the cytosolic ribosomes. The p55/Gag and p160/Gag-Pol are then transported to the viral assembly site on the cell surface lipid rafts [94, 95]. The other structural protein Env precursor gp160 is translated from a partially spliced RNA species by the rough ER-associated ribosomes. In the ER, Env undergoes glycosylation, structural isomerization, and trimerization prior to its export to the Golgi apparatus [96, 97]. In the Golgi, the Env gp160 is further proteolytically processed into the non-covalent heterodimer gp41/gp120 [98], which is then transported to the viral assembly site to join the p55/Gag and p160/Gag-pol proteins [99, 100]. Both direct and/or indirect interactions between gp41 and p55/Gag components have been suggested to initiate the incorporation of Env into the budding viral particles [101]. During, or shortly after, viral budding, the viral protease auto-cleaves from the p160/Gag-pol and leads to the structural rearrangement within and the maturation of the HIV-1 virions.

Although not always required for viral infection *in vitro*, the viral accessory proteins, Vpu, Vif, Vpr and Nef, enable the virus to overcome the host restrictions and evade the host antiviral responses *in vivo* through interacting with various cellular proteins [102, 103]. Vpu interacts with the ER-residing CD4 molecules and targets it for ubiquitine-mediated proteosomal degradation [104-106]. This prevents the nascent Env molecules from binding to the ER-residing CD4 prematurely, which could inhibit the Env from being incorporated into the budding virions. Vpu has also been found to down-regulate and counteract the host restriction factor tetherin. Tetherin is a transmembrane protein that could be simultaneously anchored onto both the cellular and the viral membranes and thereby tether the budded virions onto the cellular membrane and prevent their release for further infection [107]. Vif counteracts another host restriction factor, the apolipoprotein B mRNA editing complex 3G protein or simply APOBEC3G, by targeting it to ubiquitine-mediated degradation and preventing its incorporation into the virions [108-110]. APOBEC3G exerts several anti-viral activities [102, 111]. It attacks single-stranded DNA by converting cytidine to uridine, which results in guanosine-to-adenosine hypermutations within proviruses and the subsequent production of non-infectious viral particles. Furthermore, APOBEC3G could inhibit the processivity of the reverse transcriptase and upregulate the natural killer cell-mediated lysis of the infected cells. The third accessory protein Nef enables the viral infection to evade cytotoxic T lymphocytes and promotes viral budding, through its interactions with the host clathrin adaptor proteins. Nef stabilizes the interaction between the clathrin adaptor proteins with the histocompatibility

14

complex class I (MHC-1), which leads to the mislocalization and subsequent degradation of MHC-1 along the endolysosomal pathway [112-114]. Nef can also concurrently interact with the clathrin adaptor protein and the cytoplasmic tail of CD4, which facilitate the down-regulation of CD4 from the cell surface [115, 116]. This inhibits the retention of virions on the cell surface through Env-CD4 interaction, and therefore promotes virus release. Finally, Vpr promotes viral infectivity through multiple mechanisms, including enhancing HIV-1 replication in lymphocyte- and monocyte-derived cell lines, inducing host cell cycle arrest and promoting the nuclear translocation of the HIV-1 pre-integration complex [103].

1.3.2.4   *Anti-retroviral therapy*

In 2012, UNAIDS reported a 33% decrease in the annual new HIV-1 infection cases as compared to that in 2001. Also, in 2012, the number of AIDS deaths declined to 1.6 million, from 2.3 million in 2005. This is largely owing to the development and availability of antiretroviral medications. So far, there are 24 Food and Drug Administration (FDA)-approved antiretroviral drugs, which are categorized by their targets into protease inhibitors, nucleoside/nucleotide analogue reverse transcriptase inhibitors, non-nucleotide analogue reverse transcriptase inhibitors, entry inhibitors and the integrase inhibitors [117]. The HIV-1 reverse transcriptase is error-prone and could introduce one to ten mutations per copy per replication cycles. This results in an enormous genetic diversity and the emergence of drug-resistant HIV-1 quasispecies [118]. In fact, the HIV-1 quasispecies with reduced susceptibility to any one or two drugs likely exist within patients even before therapy [119]. In light of this, it is recommended to include three different antiretroviral drugs in the treatment regimen named HAART (highly active antiretroviral therapy) or lately just ART. HAART effectively reduces the likelihood of selecting virus clones that are resistant to all three antiretroviral drugs. However, HAART cannot eliminate HIV infection. To suppress the viral replication, the patients are required to maintain a lifetime drug adherence. Besides non-adherence, virological failure could be attributed to poor drug tolerability and drug interactions that could inhibit the antiretroviral efficacy, which collectively lead to the evolution of drug resistance and therefore there is a continuous need for new anti-HIV treatments [117].

1.3.3   HIV-1 envelope protein

1.3.3.1   *HIV-1 Env protein biosynthesis and domains*

HIV-1 Env protein is a ~860 aa type I transmembrane protein which is translated from a singly spliced bicistronic *vpu/env* mRNA [120, 121]. Its synthesis and maturation follow the host cellular secretory pathway. Nascent Env is translated with an N-terminal signal sequence (ss) region (1-30 aa), which directs the co-translational translocation of nascent Env into the rough ER [122]. Different from the common ss that is cleaved off during or upon the protein translocation and sent for degradation,

the Env ss stays within the nascent Env polyprotein for a prolonged period. This unusually delayed cleavage, which is catalyzed by an ER resident signal peptidase (SPase), has been found to strongly contribute to the slow folding process of the nascent Env [122, 123]. Specifically, the presence of the Env ss correlates with the interactions between nascent Env and various chaperone proteins in the ER, such as Bip and calnexin [122-125]. Upon the correct disulfide bond isomerization and folding of Env, the ss is liberated, while the Env/gp160 assembles predominantly into trimers and are transported to the Golgi apparatus [97, 126-128]. In the Golgi, the Env/gp160 receives further glycosylation modification and proteolytic processing by furin or furin-like proteases. Furin processes the Env/gp160 at the highly conserved poly-basic cleavage site K/R-X-K/R-R, resulting in the surface unit gp120 and the transmembrane unit gp41 (Figure 10A) [98, 129]. This process is essential for the viral infectivity. The two units, gp120 and gp41, remain non-covalently associated and are then transported to the cell surface for viral budding [130, 131].

The error-prone HIV-1 reverse transcriptase introduces high level of mutations during the viral replication cycles. Based on the sequential variation between the Env of different HIV-1 strains, the surface unit gp120 is divided into five variable (V) regions interspersed by multiple conserved (C) regions (Figure 10B) [132, 133]. The Env transmembrane unit gp41 has a typical organization of a Class I viral fusion glycoprotein, composed of an N-terminal fusion peptide, two HRs, the MPER, the TM domain and a long cytoplasmic tail (Figure 10B).
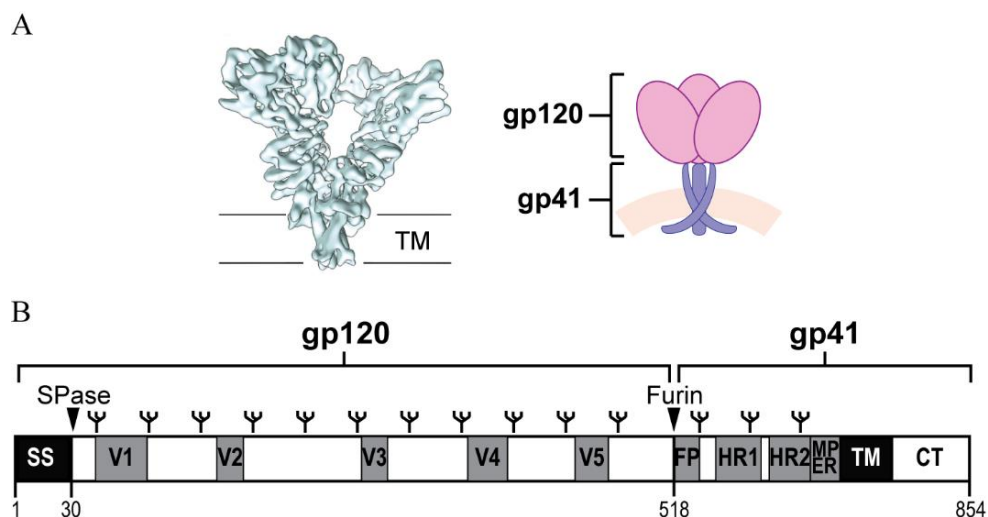
Figure 10. HIV-1 Env protein architecture and domains. (A) Atomic resolution structure reconstructed from electron microscope images [134] (left), and the schematic representation (right) of the prefusion Env protein trimers. (B) Schematic representation of the Env protein domains. Cleavage sites by the signal peptidase (SPase) and by furin are indicated by the arrowheads. The numbering of the domains is based on the HIV-1 NL4-3 strain.

The Env protein mediates the HIV-1 entry through a Class I viral fusion mechanism, similar to that of the S protein-mediated SARS-CoV entry (Figure 6). Structurally, the surface unit gp120 is composed of an inner domain, a β-sheet 'bridging sheet' domain, and an outer domain (Figure 11A) [135]. Elements of all three domains constitute a deep pocket of roughly 10 Å that interacts with the viral receptor CD4 molecule [135]. Recognition of the viral receptor introduces extensive structural rearrangement within the gp120, priming its interaction with the co-receptors, the chemokine receptors CCR5 or CXCR4. The V3 loop within gp120 is essential for the chemokine co-receptor interaction and determines the viral tropism [136, 137]. It is believed that the chemokine receptor recognition introduces further conformational changes within Env, leading the exposure of the N-terminal FP within gp41 [82, 138-140]. The exposed FP then inserts into the target membrane, and induces the formation of the pre-hairpin intermediate state and subsequently the 6-HB fusion core between the HR1 and HR2 regions (Figure 11B) [7, 97, 141].
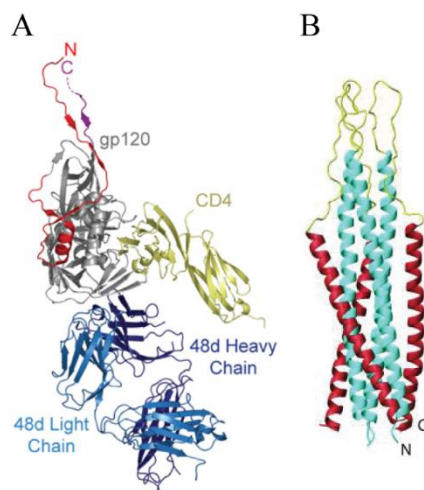
Figure 11. HIV-1 Env structure during viral entry. (**A**) Structure of the gp120, in complex with CD4 and the antibody fragment of 17b [135]. (**B**) Structure of the putative HIV-1 fusion core [141].

1.4   THE MPER OF CLASS I VIRAL FUSION GLYCOPROTEINS

Class I viral fusion glycoproteins mediate the membrane fusion through forming the 6-HB fusion cores, which releases the free energy to approximate the opposing membranes separated by hydrophobic effects. The formation of 6-HB also predisposes MPER, the region following the HR2, to an increasingly lipidic environment. MPER is a conserved hydrophobic pre-TM domain among the Class I viral fusion glycoproteins [43, 142-146]. Its unusual enrichment in aromatic residues, particularly tryptophan (Trp), indicates its involvement in lipid mixing and membrane remodeling [147-150].

Furthermore, the MPER of several Class I viral fusion glycoproteins, including the SARS-CoV S protein and HIV-1 gp41, contains a sequence similar or identical to the

cholesterol recognition amino acid consensus (CRAC) motif, -L/V-(X)(1-5)-Y-(X)(1-5)-R/K, where (X)(1-5) represents 1-5 of any amino acid [151, 152]. Cholesterol, together with sphingolipids are the key components of lipid rafts, functional membrane microdomains that are enriched with membrane-associated proteins [153]. While the viral entry through lipid rafts is facilitated by local concentration of viral receptors, cholesterol increases the membrane rigidity that might impede the membrane fusion process [154-157]. The MPER could counteract this effect and facilitates membrane fusion through CRAC-cholesterol interactions.

1.4.1 SARS-CoV S protein MPER

The S protein MPER (1191-1201 aa) is well conserved among coronaviruses (Table 2). It contains three conserved Trp residues, and its N-terminus together with the upstream HR2 sequence constitutes a CRAC motif LGKYEQYIK (Table 2). Truncations of the S protein MPER and/or alanine (Ala) substitutions of its Trp residues abrogate viral infectivity [158-160]. Biophysical studies using synthetic SARS-CoV S protein MPER-derived peptides further suggest that the MPER facilitates the membrane fusion and the coronavirus entry through membrane partitioning and perturbation [142]. The formation of the 6-HB fusion core in the S protein would also predispose the regions flanking the two HRs, the N-terminal FPs and the C-terminal MPER, into proximity. The potential interactions between these regions have been suggested previously. However, discrepancies exist in the results generated by using artificial lipid monolayers under different setting [161].

Table 2. MPER is highly conserved among coronaviruses fusion glycoproteins.

Genus	Virus	MPER sequence
Betacoronavirus	BCoV	YINLKDIGTYEYYVK**W**P**W**YV**W**L---LIGFAGVAMLVLLF[a]
	HCoV-OC43	YINLKDIGTYEYYVK**W**P**W**YV**W**L---LICLAGVAMLVLLF
	HEV	YINLKDIGTYEYYVK**W**P**W**YV**W**L---LIGLAGVAMLVLLF
	MHV	YINLKDVGTYEMYVK**W**P**W**YV**W**L---LIGLAGVAVCVLLF
	SARS-CoV	LIDLQELGKYEQYIK**W**P**W**YV**W**LGF-IAGLIAIVMVTILL
Alphacoronavirus	TGEV	LVNLEWLNRIETYVK**W**P**W**YV**W**L---LIGLVVIFCIPLLL
	CCoV	LVNLEWLNRIETYVK**W**P**W**YV**W**L---LIGLVVIFCIPLLL
	FCoV	LVDLEWLNRIETYVK**W**P**W**YV**W**L---LIGLVVIFCIPLLL
	PEDV	LVDLEWLNRVETYIK**W**P**W**YV**W**L---IIVIVLIFVVSLLV
	HCoV-229E	LVDLKWLNRVETYIK**W**P**W**WV**W**L---CISVVLIFVVSMLL
Gammacoronavirus	IBV	LIDLEKLSILKTYIK**W**P**W**YV**W**LAIAFATIIFILILGWVF
Deltacoronavirus	PorCoV HKU15	LVDLEWLNRVETYLK**W**P**W**YI**W**LAIALALIAFVTILITIF

[a]MPER sequences are shaded and the conserved Trp residues in bold.
BCoV, bovine coronavirus; HCoV-OC43, human coronavirus OC43; HEV, porcine hemagglutinating encephalomyelitis virus; MHV, murine hepatitis virus; TGEV,

transmissible gastroenteritis virus; CCoV, canine coronavirus; FCoV, feline coronavirus; PEDV, porcine epidemic diarrhea virus; HCoV-229E, human coronavirus 229E; IBV, avian infectious bronchitis virus; PorCoV HKU15, Porcine coronavirus HKU15.

1.4.2 HIV-1 Env MPER

HIV-1 Env MPER is essential for viral entry and infectivity, by mediating the formation and expansion of the lytic pores for viral content delivery [162, 163]. The HIV-1 Env MPER is also a hot-spot for immunological studies. Three epitopes of the broadly neutralizing antibodies, 2F5, 4E10 and Z13e, have been identified in this region [15, 164]. This also makes the HIV-1 Env MPER a relatively well characterized region, as compared to the MPER of other Class I viral fusion glycoproteins.

The Env MPER contains two highly conserved membrane-active sequence elements, the five Trp residues and a C-terminal CRAC motif LWYIK, which both contribute to the activity of MPER in membrane fusion (Table 3) [133, 164]. Trp residues are essential for the MPER-derived peptides to partition into, fuse, and/or cause leakage in artificial lipidic vesicles [145, 165-167]. Both the viral envelope and the viral entry site, i.e. plasma membrane lipid rafts, are enriched in cholesterol. High local concentration of cholesterol induces the formation of ordered membrane domains, which usually hinders water-membrane partitioning and subsequent assembly of lytic pores. The MPER CRAC motif LWYIK counteracts this effect by interacting specifically with cholesterol, sequestering it into ordered domains, and enabling the MPER-derived peptides to partition deeper into cholesterol-enriched membranes (~50% cholesterol). In fact, the membrane destabilization effect of the MPER-derived peptides could be enhanced by introducing cholesterol into lipid membranes, in a dose-dependent manner [167, 168].

Table 3. HIV-1 Env MPER is highly conserved [133].

Group : subtype	MPER sequence
M:A	AQVQQEQNEKKLLE**LKEWASIWNWLDITKWLWYIK**IAIIIVGA[a,b]
M:B	SQIQQEKNEKELLE**LDKWASLWNWFSITQWLWYIK**IFIMIVGG
M:C	SQNQQEQNEKDLLA**LDKWQNLWSWFDITNWLWYIK**IFIMIVGG
M:D	SQNQQEKNEKDLLE**LKQWANLWNWFTISNWLWYIK**IFIMIVGG
O	SQIQQEKNEQDLLA**LDKWANLWNWFDITNWLWYIK**IFIMIVGG

[a]MPER sequences are shaded and the conserved Trp residues in bold. [b]The CRAC motif is underlined.

Besides its membrane reactivity, the MPER could also facilitate the membrane fusion through self-oligomerization and heteromerization with the N-terminal FP domain, both in a Trp-dependent manner [166, 168-171]. Particularly, the post-6-HB formation of the MPER-FP complex is believed to serve as a continuum structure to the 6-HB, which could provide extra free energy for juxtaposing the opposing membranes, enhance the local hydrophobicity, and enable the lipid flow.

With the increase of local environment lipidity, Env MPER is also believed to undergo structural transition. Under aqueous condition, higher temperature and/or higher pH (~6), MPER is believed to contain high percentage of extended β- or turn-motifs [172, 173]. When it is associated with lipidic vesicles, the MPER-containing constructs adopt a solvent-exposed N-terminal α-helix and a membrane-embedded C-terminal α-helix, separated by a center hinge region [173, 174].

1.5 APPLICATIONS OF THE INTERACTION-MEDIATING SEQUENCES WITHIN THE CLASS I VIRAL FUSION GLYCOPROTEINS

Viral fusion glycoproteins are important targets for therapeutic interventions. Entry and fusion inhibitors inhibit the first steps of viral infection and replication and, therefore, are considered more effective in viral inhibition, as compared to antivirals targeting the subsequent steps of viral replication. However, viral fusion glycoproteins are usually heavily glycosylated and could be constantly evolving, camouflaging their functionally important regions from immune surveillance and rendering antiviral targeting difficult. The interaction-mediating regions or short sequences within the Class I viral fusion glycoprotein, for example the RBD, HRs, and MPER, conduct extensive infection-required protein-protein and/or protein-ligand interactions. To maintain the interactions and viral infectivity, these sequences are usually well conserved and potentially exposed, as evidenced by the identification of multiple broadly neutralizing antibodies that target regions including the CD4-binding site, HR1, and MPER [164, 175-181]. These characteristics render these interaction-mediating sequences attractive targets for the development of therapeutics, vaccines and research reagents [38, 47, 182-187]. The current viral entry/fusion inhibitors mainly focus on two interactions, the RBD-receptor interaction leading to the viral attachment, and the HR1-HR2 interaction leading to the formation of the 6-HB fusion core.

1.5.1.1 Receptor/Co-receptor recognition

Many molecules have been shown to be potent in inhibiting the recognition of HIV-1 to CD4 molecule or to the chemokine receptors CCR5 and CXCR4 *in vitro* [188]. Early attempts to find entry inhibitors include a recombinant CD4 molecule and the monoclonal antibody ibalizumab, which inhibit the viral attachment by directly targeting the gp120-CD4 interaction or by preventing the subsequent gp120-CCR5/CXCR4 interaction, respectively. So far, the CCR5 antagonist Maraviroc has

shown the most promising antiviral efficacy and was approved by FDA in 2007 for clinical use [189-191]. However, Maraviroc usage has been shown to induce the emergence of CXCR4-using viruses, which in turn have been correlated with decreasing CD4+ T lymphocyte count and disease progression [189, 192].

1.5.1.2 HR1-HR2 interaction

The development of antiviral fusion inhibitors has focused on the HRs and the 6-HB fusion core, the key structure within the Class I viral fusion glycoproteins that mediate the membrane fusion [186, 187, 193-196]. The most studied examples include the fusion inhibitors derived from the HRs of HIV-1 gp41. It is believed that the peptidyl derivative of either HR could selectively recognize and bind to the other HR within the fusion protein transiently exposed at the pre-hairpin intermediate stage, thereby blocking the 6-HB and fusion pore formation in a dominant negative manner. In support of this viral inhibition model, the peptidyl fusion inhibitor C34, which is derived from the gp41 HR2, formed 6-HB with the HR1-drived peptides [197]. The formation of the 6-HB between C34 and the HR1-derived peptides is stabilized by two mechanisms. The C34 helix can bind along the grooves of the central HR1 coiled-coil [197]. This mechanism has inspired ways to enhance the anti-viral efficacy of C34. Salt bridges and/or sequence replacements have been introduced into C34 to enhance its helicity and association with Env HR1 to form a stable 6-HB, which thereby improves its anti-viral efficacy [185, 198-200]. The formation of a stable 6-HB and the potency of C34 are also contributed by the interaction between the a pocket binding domain (PBD) within C34 and a hydrophobic cavity located at the HR1 C-terminus [197]. However, the PBD-cavity interaction and a stable 6-HB is not necessary for an effective viral inhibition, as evidenced by the development of T20/Enfuvirtide (Trimeris, Inc. and Roche) which is the only licensed antiviral fusion inhibitor for clinical use. T20 is derived from the aa 638-673 of HR2 and lacks the PBD [187]. Consequently, it cannot inhibit viral entry through forming a stable 6-HB with the viral HR1 [201]. Instead, the additional MPER sequence, WNWF, compensates the loss of PBD by conferring T20 with the ability to bind to the membrane, to interfere with the oligomerization of the fusogenic gp41, and to inhibit the post-lipid-mixing fusion pore formation [202]. Deletion or substitution of the partial MPER sequence abrogated the antiviral effect by increasing its IC_{50} over 1000-fold [187, 201]. Logically, both PBD and the lipid-binding MPER sequence are included in the next generation of fusion inhibitor, T1249. As compared to T20, T1249 has a higher potency, longer half-life, and broader antiviral spectrum [203]. However, the clinical development of T1249 was halted due to formulation issues. While most of the peptidyl fusion inhibitors derived from the HRs of the Class I viral fusion glycoprotein did not reach clinic use, their sequence-specific recognition of the HRs to form 6-HB have made them attractive research reagent in mechanistic studies and vaccine development.

1.5.1.3 MPER

HIV-1 Env MPER is capable to oligomerize under lipidic environment, and it also interacts with the N-terminal FP regions putatively post-6-HB formation [168]. The N-terminal MPER sequence WNWF has further been shown to be essential for the antiviral activity of T20, putatively by conferring T20 with membrane reactivity. It was logical to expect that the peptidyl derivative of MPER could bind to the viral MPER and/or other interacting ligands and hence halt fusion pore formation and viral entry. The high level of sequence conservation of the MPER further makes it an attractive target for the development of novel peptidyl fusion inhibitors. Peptide C8, a MPER-derived peptide of feline immunodeficiency virus (FIV) could inhibit the infection of 12 primary FIV strains with IC_{50} values of 0.03-0.63 ug/ml [204]. Retro-inversion of C8 by reversing the sequence and replacing all the residues with D-amino acids further enhanced the peptide stability while preserving the antiviral activity [205]. However, the anti-viral mechanistic study of the MPER or MPER-containing peptides have mainly focused on the N-terminal sequence and/or the Trp residues, and the potential contribution of the CRAC motif in viral inhibition has not been discussed.

1.5.1.4 The future of peptidyl antivirals derived from the Class I viral fusion glycoproteins

Despite of the progress and success in developing entry/fusion inhibitors, they are constantly challenged by the emergence of drug-resistant viral strains, poor bioavailability and formulation issues, which collectively contribute to patient non-adherence and virological failures [206-209]. These issues could be potentially circumvented by employing structure-based design to enhance the helicity of the HR-derived peptides that promotes their formation of 6-HB with the viral HRs [185, 200, 210], by attaching lipid-interacting groups such as fatty acid tail and cholesterol to concentrate the drugs to viral entry sites [211-213], by conjugating the peptides to a serum-stable molecules to enhance their stability [214], or by introducing the protease-resistant D-amino acids to the peptides [205, 215, 216].

Nevertheless, novel targets for developing antivirals and biochemical reagents are always needed, particularly when facing genetically evolving viral pathogens such as HIV-1, pathogens without effective treatment such as SARS-CoV, and emerging pathogens such as MERS-CoV. This thesis focuses on identifying novel interaction-mediating sequences within Class I viral fusion glycoproteins, investigating their potential involvements in viral infectivity, and on developing these sequences as anti-viral therapeutics and virus detection agents.

2 AIMS OF THE THESIS

The main objective of this thesis was to identify novel interaction-mediating sequences within Class I viral fusion glycoproteins, to study their implications in viral entry and infectivity, and to potentially apply them for viral detection and therapeutic interventions. Specifically,

- to identify the self-binding sequences that interacted through self-complementary zipper-like motifs within the SARS-CoV S protein and to develop them as virus detection reagents; (**Paper I**)

- to investigate the putative interaction(s) mediated by the SARS-CoV S protein MPER during viral entry; (**Paper II**)

- to study the invovlement of the membrane-active sequence elements within the HIV-1 Env MPER during viral entry and assembly, specifically in the design of MPER-based fusion inhibitors and in the biosynthesis of viral structural proteins; (**Paper III**)

- and to examine the post-ER-targeting influence of the HIV-1 Env ss in the viral replication. (**Paper IV**)

3 METHODOLOGICAL CONSIDERATIONS

In this section, general considerations are presented about materials and methods. Details about reagents and methods are described in the separate papers (I-IV).

3.1 FRAGMENT-BASED PEPTIDE ARRAY SCANNING

In peptide array scanning, rationally chosen or combinatorially generated peptide libraries are synthesized onto individual spots on a planar solid support (e.g. a membrane), and are then probed with solutions of a target protein [217, 218]. By synthesizing and probing peptides in the spatially addressable spots in parallel, peptide array scanning provides a high-throughput means to identify protein-protein or protein-ligand interactions. It has a broad spectrum of applications including antibody-peptide binding assays, proteomic studies and diagnostic assays [217-224]. Previously, peptide array scanning has also been applied to identify the protein regions that are involved in homotypic interactions or self-association, where overlapping peptides were probed with the cognate protein in solutions [225-228]. These studies inspired the choice of peptide array scanning to identify self-binding peptides within the S protein of SARS-CoV. To identify the self-binding peptides, overlapping S protein-derived peptides were synthesized on a PEG-derivatized cellulose membrane and then probed with a HA-tagged S protein ectodomain named S1188HA. The peptide spots that retained S1188HA were then identified by anti-HA chemiluminescence (**Paper I**).

A peptide array library can be prepared on a solid support either by the so-called SPOT technology [229, 230] or by photolithography [231]. The former is widely used and was also employed in this study. The SPOT technology involves delivery of small amounts of reagents to each spot position, which allows the individual peptides to be locally synthesized within each spot by solid-phase synthesis. The N-terminal amino acid is anchored onto the membrane and then the N-terminus to C-terminus elongation is performed using the Fmoc/tBu-chemistry. The simplicity of the SPOT technology in the simultaneous synthesis and anchorage of the peptide arrays, however, precludes the purification or identification of the synthesized peptides. Peptides up to 15 aa could be SPOT-synthesized with comparable purity as by the solid-phase peptide synthesizers, i.e. 70 to 90% purity [232]. Generally, the peptide purity decreases with increasing peptide length, although with optimization peptides of 34 aa could also be SPOT-synthesized up to 65% purity [233]. Nevertheless, to minimize differential S protein-binding efficiency due to different purities and/or amounts of peptides in different spots, the peptide array libraries in **Paper I** were constructed to contain peptides of 15 to 20 aa. Peptide arrays offer semi-quantitative readouts, allowing avidity of the peptide-protein interaction to be deduced from the spot intensities. Still, the identified interaction-mediating peptides need to be confirmed through other biochemical means. Hence, the S protein-binding peptides

identified by peptide array were synthesized by regular solid phase synthesis using Fmoc chemistry and confirmed *via* pull down assays employing S1188HA.

3.2 AMYLOID-LIKE FIBRILLATION PREDICTION AND BIOCHEMICAL DETERMINATION

Cross-β spine steric zipper motif, or simply steric β-zipper, is the central motif in mediating the formation of amyloid-like fibrils. Steric β-zipper peptides derived from the amyloidogenic proteins have been shown to bind to the fibrillation site specifically and to effectively obstruct the protein aggregation [234]. To apply sequence specific interaction, offered by the steric β-zipper, to viral detection, potential steric β-zipper-like sequences within the S protein were first predicted by the 3D profile method in **Paper I**.

The 3D profile method was first described as a general scheme to predict the potential of a given aa sequence to fold into a certain 3D structure [235]. By considering the local environment of each residue, and not just the isolated aa properties, the 3D profile method is a structure-based prediction algorithm. In this method, each residue within a known 3D structure is classified into one of 18 environmental classes, which summarize the residue's buried/inaccessible area, side chain area that is covered by polar atoms, and local secondary structure. By studying a library of known protein 3D structures, the occurrences or preferences of each aa in different environmental classes are then obtained. Thereby, the propensity of a sequence to adopt a given structure could be deduced from the likelihood of the residue at a certain position of that sequence to be present in the specific environmental class at that position in the given structure. Following the same principle, the 3D profile method was developed to predict amyloidogenic peptides. Specifically, it predicts the propensities of hexamer peptides to adopt the cross-β spine steric zipper structure of the amyloidogenic peptide NNQQNY [236-239]. Using this method, the prediction accuracy has been reported to be over 80% [239]. Nevertheless, biochemical studies such as Congo Red- and Thioflavin T (ThT)-binding spectroscopy studies are needed to confirm the abilities of the positively predicted peptides to form amyloid-like fibrils.

The ability to bind to Congo Red is considered to be one of the definitive characteristics of an amyloid-like fibril [240]. Upon binding to Congo Red, amyloid-like fibrils further exhibit apple green birefringence under polarized light. The Congo Red birefringence assay has been extensively applied to detect amyloid-like fibrils. However, many things including human hair could exhibit birefringence under polarized light and this assay could be subjective [240]. The Congo Red spectroscopic assay, which relies on the absorbance difference at 540 nm upon Congo Red-binding, offers a more objective means to identify amyloid-like fibrils [240]. The ThT-binding spectroscopic assay relies on the binding of amyloid-like fibrils to ThT, and can be further used to estimate the kinetics of the amyloid-like fibrillation [240]. The ThT-

binding and Congo Red spectroscopic assays were employed in **Paper I** to confirm the amyloid-like fibrillation of S protein-derived peptides.

3.3 CIRCULAR DICHROISM SPECTROSCOPY

An optically active or chiral molecule absorbs the left-handed circularly polarized light and the right-handed circularly polarized light differently, a phenomenon termed circular dichrism (CD). CD spectroscopy measures the CD of a given molecule over a range of wavelengths, and is widely used to determine the secondary structures of peptides/proteins. Different peptide/protein secondary structures have their characteristic CD spectra. For example, an α-helix exhibits dual minima at 208 nm and 222 nm and a single maximum at 193 nm, while a β-sheet exhibits a single minimum at 218 nm and a maximum at 195 nm. Most often, peptides or proteins contain a mixture of secondary structures and each structural component could be estimated by fitting the obtained CD spectra with the CD spectra of molecules with known structure. In **Papers I, II** and **III**, to estimate their secondary structures, the peptides were subjected to CD spectroscopy. Their CD spectra were subsequently de-convoluted with the CDSSTR de-convolution algorithm on Dichroweb, where the goodness-of-fit parameter, or NRMSD (normalized root mean square deviation), was set to be equal to or below 0.15 [241-243].

In **Papers II** and **III**, MPER-derived peptides were studied. Considering the membrane-proximal location of the MPER within the Class I viral fusion glycoproteins and their potential interaction with the membranes, secondary structures of the MPER-derived peptides were studied in a hydrophobic environment. Although the specific mechanism is still in dispute, 2,2,2-trifluoroethanol (TFE) has been known to stabilize native peptide structures and has therefore been considered to be a model system for a membrane-like environment [244]. Therefore, in **Papers II** and **III** the MPER-derived peptides were dissolved in buffers supplemented with increasing concentrations of TFE.

3.4 CHEMICAL CROSSLINKING

Protein-protein interactions could be stably organized multi-protein complexes or transient encounters. Transient protein-protein interactions are difficult to detect and define. Introduction of stable chemical linkage between the transiently interacting proteins is a direct way to demonstrate unambiguously a protein-protein interaction. Chemical crosslinkers contain two or more reactive ends that are capable of covalently linking two proteins when in close proximity. Thus, crosslinkers serve as useful tools in identifying transient protein-protein interactions [245]. Chemical crosslinkers can be divided into several classes based on their reaction groups, differences in the chemical specificities of their two reactive ends, their water-solubility and ability to be selectively activated. Glutaraldehyde is an amine-reactive homobifunctional crosslinker, meaning that it contains two identical reactive ends,

which react with the ε-amino groups of lysine residues and the primary amines on the N-termini of the target protein. It is a popular crosslinking reagent, as it is soluble in aqueous solvents and is able to form stable inter- and intra-subunit covalent bonds. In **Papers I** and **II**, glutaraldehyde was used to crosslink the S protein-derived peptides, to study the potential self-oligomerization of these peptides and the interaction between the S protein MPER and IFP. Peptides were incubated with glutaraldehyde for 1 h at room temperature. The reactions were quenched by 100 mM glycine. Peptide oligomers were analyzed by gel electrophoresis and Western blot. The relatively high prevalence of amine groups within proteins and peptides might result in non-specific chemical crosslinking by glutaraldehyde. To exclude connections between proteins that were in proximity but not reacting with one-another, negative control peptides were included in the crosslinking experiments.

3.5 WESTERN BLOT

Western blot is a standard peptide/protein analysis method. In Western blot, the proteins are separated by gel electrophoresis and then transferred to a membrane, which thereafter is blocked with blocking solution usually containing a mixture of unrelated proteins. The proteins to be analyzed are stained with target-specific antibodies [246]. The antibody binding is detected with a reporter molecule conjugated either directly onto the target-protein-specific primary antibody, or onto the species-specific secondary antibody targeting the primary antibody. The most common reporter molecule is horseradish peroxidase, which oxidizes a chemiluminescent reagent in the detection solution such as luminol to 3-aminophthalate, thereby releasing light for detection. In **Papers I, II, III** and **IV**, Western blot was used for the detection of biotinylated peptides, HA-tagged S protein ectodomain S1188HA, β-actin and HIV-1 viral proteins including Env, Vif, p55/Gag, p24/Capsid, p160/Gag-Pol and Nef. Protein concentrations of the observed bands in the Western blots may be quantified with high reproducibility by carefully establishing standard curves [247]. However, due to the low-dynamic range of detection and the difficulty to determine the detection limit, Western blot band densities are largely regarded as semi-quantitative. In this thesis, the protein/peptide band intensities were quantified by ImageJ, and were further normalized to the band intensity of a cellular control protein, such as β-actin or that of the WT viral protein.

3.6 PSEUDOTYPED HIV-1 SYSTEM AND THE VIRAL INHIBITION ASSAY

A pseudotyped virus is constructed by combining incomplete but complementing viral vectors from two different viruses, or where one of the vectors is mutated but from the same virus [248]. When a mutated envelope/capsid protein is introduced into a pseudotyped virus, it alters the host tropism or changes the stability of the virus particles [248]. Pseudotyped virus particles from the transfected cells can enter and infect cells. The progeny pseudotyped viral particles generated by the infected cells,

however, do not contain the vector(s) encoding the viral structural protein(s), rendering them replication incompetent, i.e. non-infectious [248]. This single-replication-cycle property of pseudotyped virus is particularly useful for studying the viral entry process and to test the anti-viral efficacy of entry/fusion inhibitors, being independent from other steps of the viral replication cycle [249].

In **Paper III**, pseudotyped HIV-1 particles were used to test the Env MPER-derived peptides as entry/fusion inhibitors. The pseudotyped HIV-1 particles were constructed by co-transfecting HLtat cells [250, 251] with pNLHIVΔuΔss and pNL1.5EU+. The NL4-3 strain proviral construct pNLHIVΔuΔss contains a mutation in the *vpu* start codon and a truncation of the Env ss region, which prevent the expression of functional Vpu and Env and the production of infectious viral particles [252, 253]. Meanwhile, pNL1.5EU+ contains the LTR-driven exon 1 and exon 5 of HIV-1(NL4-3) transcript that encode the viral Vpu, Env and Nef proteins [121]. Expression of Rev from pNLHIVΔuΔss boosts the expression of Vpu and Env from the unspliced transcript of pNL1.5EU+, which complement the viral-like particles produced from pNLHIVΔuΔss and enable them to enter and infect the CD4+, CCR5+ target cells [121]. The infectivity of the resulting pseudotyped HIV-1 particles were confirmed in the CD4+, CCR5+ TZM-bl cells that contain copies of LTR-driven luciferase and ß-galactosidase genes [208, 254-257].

The minimal region required for autonomous HIV-1 RNA packaging during viral assembly is mapped to a region covering the first 350-400 nt of the genome, spanning the LTR and the 5' part of *gag* gene [258-262]. This means that even if pNL1.5EU+ contains the LTR promoter and the viral packaging (ψ) signal, its unspliced Env-encoding transcript under the LTR promoter is unlikely to be packaged into the budding virions effectively. Furthermore, production of genomic RNA dimers, which is essential for infectious virus production, requires two complete copies of the genomic RNA [263]. Taken together, the pseudo-typed HIV-1 viral particles generated in this study could unlikely generate Env-positive infectious progeny virus particles. The single-replication-cycle property of the constructed pseudotyped HIV-1 particles was confirmed by their loss of infectivity when the culture supernatant of the infected TZM-bl cells was re-applied to a new dish of TZM-bl cells.

In the viral inhibition assays of this thesis, effective concentrations of peptides inhibiting 50% (IC_{50}) and 80% (IC_{80}) of the viral infectivity were estimated from the dose-response curves generated with the sigmoidal dose-response non-linear regression model on Prism GraphPad software.

3.7 REAL TIME QUANTITATIVE POLYMERASE CHAIN REACTION (RT-QPCR)

Polymerase chain reaction (PCR) was first introduced in 1983 by Kari Mullis [264]. It has been a powerful tool in the selective amplification, quantification and sequential

analysis of DNA samples. In **Paper IV**, a PCR method that reports the quantity of the amplified product in real time (real time (RT)-qPCR) was employed to quantify the full-length HIV-1 viral transcripts.

For the RT-qPCR, total RNA was extracted from the cell lysate, reverse-transcribed into cDNA, and then amplified in the presence of a pair of target gene-specific primers and a DNA-binding reporter dye/probe. The DNA-binding reporter dye/probe provides the basis of real-time detection of the amplified product. It is a stretch of gene-specific nucleotides linked with a reporter fluorescent dye on one end and a quencher on the other, commercially known as the TaqMan probe. When intact, the quencher molecule inhibits the fluorescence from the reporter. The gene-specific nucleotides are chosen so that the probe will bind within the intended amplicon of the target gene cDNA. During the amplification, as the polymerase machinery proceeds to the probe-binding site, the quencher molecule will be cleaved off the probe. This results in an increase of the fluorescence signal that is detected by the machine. Alternatively, a SYBR green system-based dye, which binds only to the double-stranded DNA amplification product, could be used in place of the TaqMan probe. While SYBR green is a cheaper way to set up the RT-qPCR experiments, TaqMan offers lower false positive signals. In this thesis, RT-qPCR was performed with the TaqMan chemistry.

The quantification can be done with two different methods, an absolute or a relative quantification. The absolute quantification determines the copy number of the target gene transcript directly, but requires the construction of standard curves. The relative quantification describes the changes in the copy number of the target gene transcript in relation to that of a house-keeping or reference gene transcript. It is more convenient to set up, but requires comparable polymerase reaction efficiencies between different genes. In this thesis, the full-length HIV-1 transcript was quantified using the relative expression analysis based on the $2^{-\Delta\Delta Ct}$ method, with the human glyceraldehyde 3-phosphate dehydrogenase (hGAPDH) as the reference gene [265].

4 RESULTS

4.1 PAPER I

IDENTIFICATION AND APPLICATION OF SELF-BINDING ZIPPER-LIKE SEQUENCES IN SARS-CoV SPIKE PROTEIN.

STUDY BACKGROUND

Self-binding peptides containing zipper-like sequences, such as the Leu/Ile zipper sequence within the coiled coils and the cross-β spine steric zippers within the amyloid-like fibrils, could bind to the protein-of-origin through forming sequence-specific zipper-like motifs. They represent an attractive target for various biochemical applications, despite their length-dependent and modest self-binding affinities. The cross-β spine steric zipper, or steric β-zipper, is a complementary, dry, steric-zipper interface formed between the tightly interdigitating side chains of two facing peptide β-sheets of the same sequence [238, 266-271]. It is the central motif in mediating peptide/protein to form amyloid-like fibrils and can be readily predicted using the 3D profile method [235-237, 239]. Previously, peptides containing this motif have been shown to inhibit fibrillation effectively [234]. The study in **Paper I** combines the predictive and peptide scanning methods to identify the self-binding sequences containing the cross-β spine steric zipper-like motif within the SARS-CoV S protein.

Identification of the putative self-binding peptides within the S protein

Thirty-four putative S protein-binding peptides of six clusters were identified, by scanning a peptide array composed of S protein-derived overlapping 20-mer peptides with S1188HA, an HA-tagged recombinant ectodomain of the S protein (aa 1-1188) (Figure 12A) [155]. Five of the six clusters of S protein-binding peptides each contained a core octapeptide consensus sequence, subsequently GINITNFR (aa 225-232) in Cluster I derived from the region upstream to the RBD, VLTPSSKR (aa 537-544) in Cluster II from the region immediately following RBD, LNRALSGI (aa 745-752) in Cluster III from the region upstream of the FP, AMQMAYRF (aa 881-888) in Cluster IV from the second fusion peptide region IFP, and, finally, RLITGRLQ (aa 977-984) in Cluster V from the HR1 (Figure 12B-F). The eight peptides of the sixth cluster R covering the β-1, -3, -4, -5 and -6 of the seven β sheets within the RBD, interacted with S1188HA with relatively high affinities but did not yield a core consensus sequence (Figure 12B) [38].
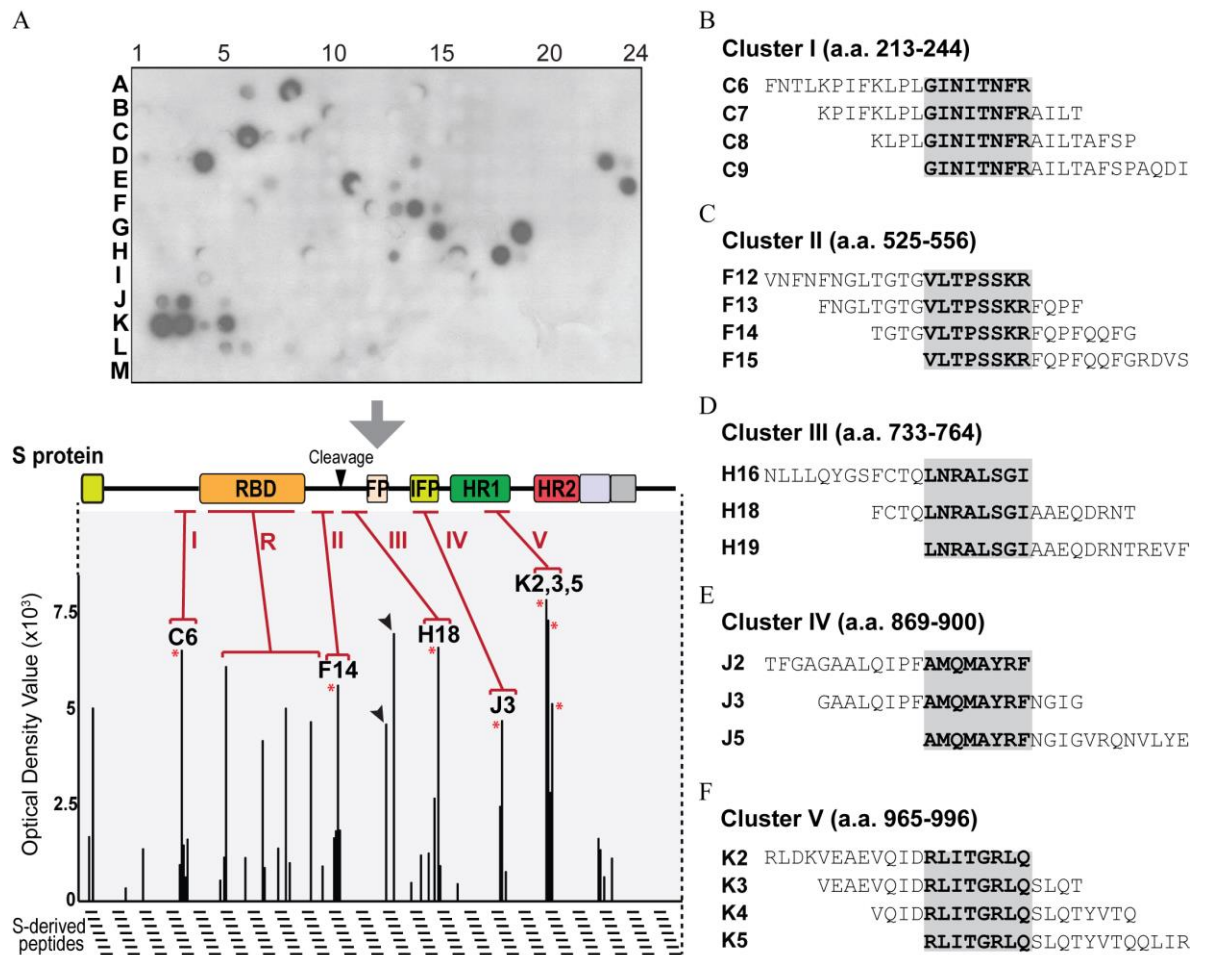
Figure 12. Peptide array scanning for putative self-binding sequences within the S protein. (**A**) The S protein-binding peptides were identified by probing the S protein-derived peptide array with Sf9 cell-expressed S1188HA. Positive peptide-protein interactions were quantified as peaks in optical density value (o.d.v.). Representative peptide(s) of each cluster (I, II, III, IV, V and R) and the false positive peptides were indicated by asterisk the arrow heads, respectively. (**B**) The core sequence (shaded and bold) of Cluster I, (**C**) II, (**D**) III, (**E**) IV and (**F**) V.

Peptide C6 containing the Cluster I core sequence self-interacted and formed amyloid-like fibrils

The 3D profile prediction [237, 239] indicated that the hexapeptide GINITN within the Cluster I core sequence had a high propensity of fibrillation through the steric β-zipper motif (Figure 13C). The mechanism of the interaction between S1188HA and the Cluster I representative peptide C6 was further elucidated. Despite SARS-CoV S protein is heavily glycosylated with 23 potential N-linked glycosylation sites [272], C6 pulled down N-deglycosylated S1188HA from the Sf9 cell lysate as effectively as pulling down the untreated S1188HA (Figure 13A). This suggests that the C6-S protein interaction was likely mediated by sequence-specific interactions. Furthermore, C6 trimerized, as did the control peptide GG38, which was derived from the trimerizing HR2, under increasing concentrations of the chemical crosslinker

glutaraldehyde (Glut) (Figure 13B) [46, 50, 273]. Finally, peptide C6 was subjected to CD spectroscopy for secondary structure determination and to amyloid-like fibrillation assays (Figure 13C). De-convolution of the CD spectrum of C6 revealed a relatively high β-strand content (~36%) under aqueous conditions (Figure 13C). C6 further formed amyloid-like fibrils as determined by an increase in absorbance at 540 nm upon binding to Congo Red, and an increase in florescence emission at 482 nm upon binding to ThT (Figure 13D-E). The data obtained suggest that part of C6 could form steric β-zippers, the common structural motif within amyloid-like fibrils [236, 238]. They further indicated that C6 could self-interact with the same sequence within the S protein through a β-zipper-based interaction.
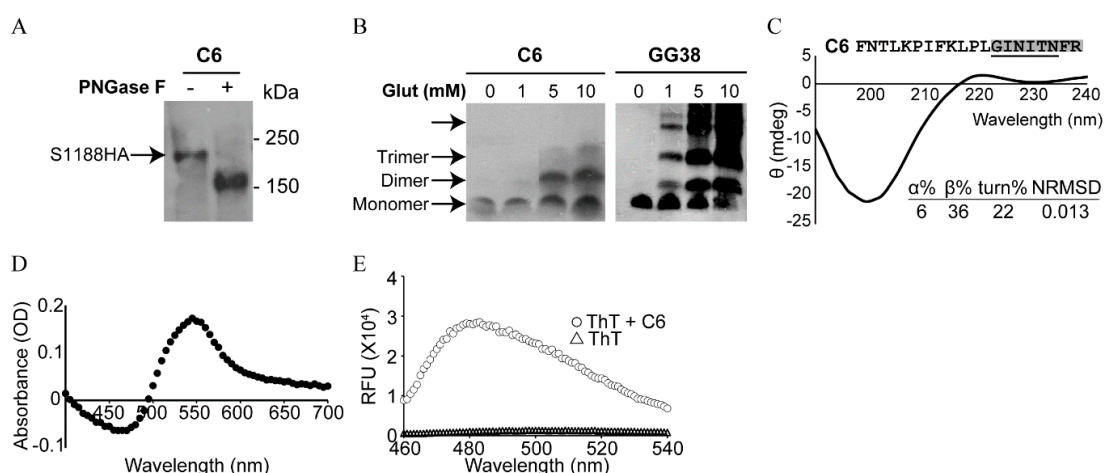
Figure 13. Peptide C6 containing cluster I core sequence could self-interact. (A) Peptide C6 pulled down N-deglycosylated S1188HA expressed in Sf9 cells. **(B)** Peptide C6 trimerized under chemical crosslinking using glutaraldehyde (Glut). **(C)** Cluster I core sequence (shaded) within peptide C6 was predicted to contain an amyloid-like fibril-forming hexapeptide (underlined). Peptide C6 exhibited relatively high β-strand content in ddH$_2$O. NRMSD, normalized root mean square deviation. **(D)** Peptide C6 (10mg/ml) in DMSO was accessed for amyloid formation by Congo Red and by **(E)** Thioflavin T (ThT) spectroscopic assays.

Self-interacting peptide C6 detected the S protein expression and homodimerization enhanced its affinity to S1188HA

Taking advantage of the sequence-specific self-interacting property of the steric β-zipper motif, C6 was used to detect the S protein expression *via* immunofluorescence staining. In the 293T cells transiently expressing an HA-tagged full-length S protein, the detection signal of peptide C6 at 2μM followed by the Dylight 488-conjugated NeutrAvidin^{TM} overlapped with the signal of an rabbit-raised anti-HA antibody as the primary antibody followed by an Alexa Fluor 594-conjugated anti-rabbit secondary antibody (Figure 14A). However, the overlapping detection signal by C6 and the anti-HA antibody could be due to an interaction between C6 and the anti-HA antibody. To

exclude this possibility, the 293T cells, expressing the un-tagged full-length S protein (S1255) were either stained with peptide C6 at 2µM followed by Dylight 488-conjugated NeutrAvidin™ or with a conventional mouse-raised anti-S protein monoclonal antibody followed by an Alexa Fluor 488-conjugated anti-mouse secondary antibody. The staining pattern and intensity of C6 and anti-S antibody were similar (Figure 14B).

Homodimerization of C6 through a lysine-based linker further enhanced the C6-S1188HA interaction avidity (Figure 14C-D). In a pull down assay, the dimeric C6 peptide, C6-D, pulled down approximately 2.5 times higher amount of S1188HA than C6 did.
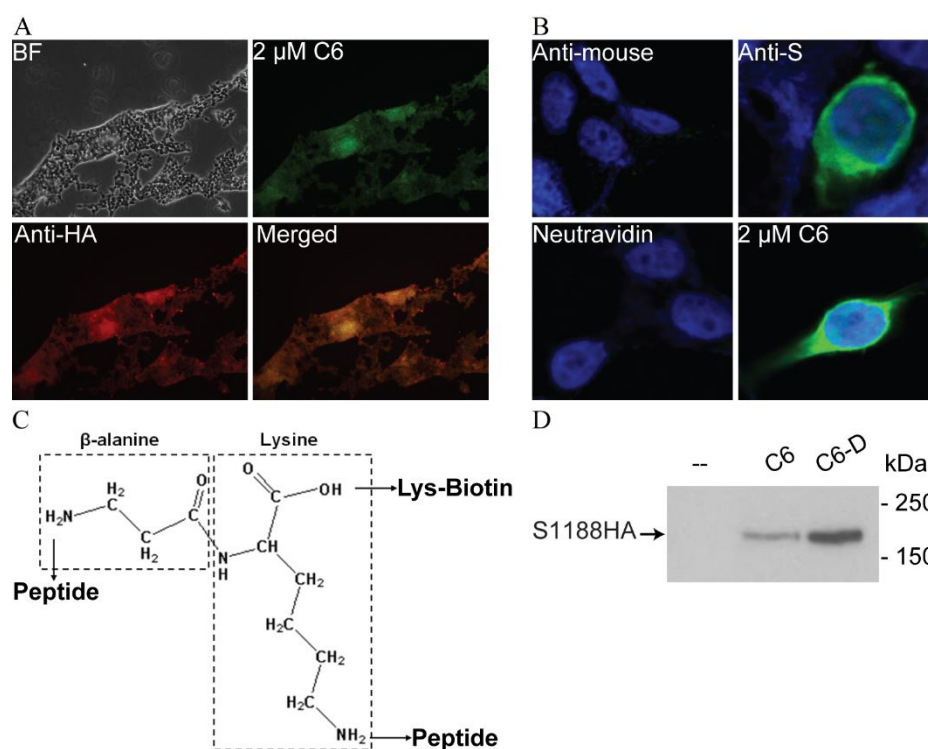
Figure 14. Peptide C6 detected S protein expression and the C6-S protein interaction could be enhanced through dimerization. (A) The detection signal of C6 overlapped with that of anti-HA antibody in the 293T cells transiently expressing an HA-tagged full-length S protein. **(B)** Immunoflorescence staining of 293T cells transiently expressing an untagged full-length S protein using C6 yielded similar staining intensity and pattern as when using an anti-S monoclonal antibody. **(C)** Lys-based linker molecule for the synthesis of C6-D. **(D)** C6-D pulled down approximately double amount of S1188HA than C6 did in pull down experiment.

Effects of peptide lengths in identifying helix-mediated interactions

The HR-derived peptides, which have been previously shown to oligomerize, were under-represented in the peptide array scanning. S1188HA pull-down experiment using twenty-five HR2 peptides of different lengths suggest that the 20-mer peptides

in the peptide arrays could not attain high-affinity helix-mediating interactions such as between the HR sequences. The recombinantly-produced HR2 peptides, MBP-n, were constructed with an N-terminal maltose binding protein (MBP) followed by n (11 to 35) aa of HR2 sequences. MBP-30 to MBP-35 containing near-full-length HR2 sequences of over 30 amino acids pulled down S1188HA with high efficiency, with the exception of MBP-31(Figure 15). Further truncation diminished the pull down activity of the MBP-n, with the exceptions being MBP-16 to MBP-21 containing approximately half of the HR2 sequences. However, they were significantly less efficient in pulling down S1188HA than the near-full-length HR2 peptides.
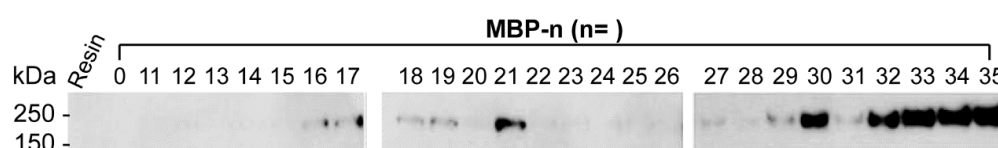
Figure 15. S1188HA pull-down experiment with MBP-n constructs. S1188HA in SF9 cell lysate was pulled down by MBP-HR2-n fusion proteins (n = 11-35), with the resin only or MBP only served as the negative controls.

4.2 <u>PAPER II</u>

TRYPTOPHAN-DEPENDENT MEMBRANE INTERACTION AND HETEROMERIZATION WITH INTERNAL FUSION PEPTIDE BY MEMBRANE PROXIMAL EXTERNAL REGION OF SARS-COV SPIKE PROTEIN

STUDY BACKGROUND

During viral entry mediated by Class I viral fusion glycoproteins, the formation of the 6-HB fusion core approximates the regions flanking HR1 and 2. In SARS-CoV S protein, they are the FP and IFP, upstream to HR1, and the MPER, downstream of HR2. Peptide array scanning for S protein sequences involved in homotypic interactions in **Paper I** identified three IFP-derived peptides. Meanwhile, the FP-derived peptides were only identified upon heat denaturation of the S protein, suggesting different positioning and potential interaction partners of these two hydrophobic regions within the S protein. Furthermore, peptide scanning in **Paper I** used the S protein ectodomain, S1188HA, excluding the MPER and beyond. In **Paper II**, the potential of SARS-CoV S protein MPER to oligomerize and heteromerize with the FP regions was examined, and the roles of the highly conserved Trp residues in maintaining these processes were further investigated.

34

Trp residues were essential for the oligomerization of the S protein MPER

Among the peptides derived from the SARS-CoV S protein MPER, the peptide designated M-wt contained the wild type (WT) MPER sequence, and the peptides M-3W3A and M-3W3F contained MPER sequences with triple Trp→Ala or Trp→phenylalanine (Phe) substitutions (Table 4).

Table 4. Peptides derived from the coronavirus S proteins.

Peptides	Virus	Region (aa)	Sequence
M-wt	SARS-CoV	MPER (1187-1202)	KYEQYIK**W**P**W**YV**W**LGF
M-3W3A	SARS-CoV	MPER (1187-1202)	KYEQYIK**A**P**A**YV**A**LGF
M-3W3F	SARS-CoV	MPER (1187-1202)	KYEQYIK**F**P**F**YV**F**LGF
M-ibv	IBV	MPER (1186 -1202)	LKTYIKWPWYVWLAIAF
FP19	SARS-CoV	FP (770-788)	MYKTPTLKYFGGFNFSQIL
IFP23	SARS-CoV	IFP (866-888)	AGWTFGAGAALQIPFAMQMAYRF
TE20	SARS-CoV	(743-762)	TQLNRALSGIAAEQDRNTRE

Under chemical crosslinking using glutaraldehyde, the WT MPER-derived peptide M-wt formed trimers and higher-order oligomers, while Ala-substituted peptide M-3W3A did not (Figure 16). Phe-substituted peptide M-3W3F retained the ability to oligomerize, suggesting that aromatic side chains were essential for the MPER-MPER interaction. However, instead of a discrete order of oligomers as in M-wt, M-3W3F formed a continuous band of oligomers.
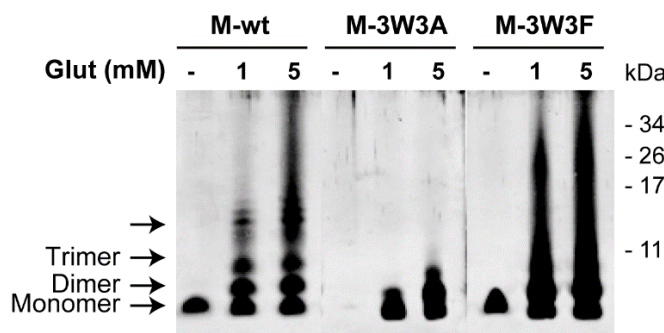

Figure 16. The S protein MPER-derived peptides oligomerized. Oligomerization of M-wt, M-3W3A and M-3W3F were accessed under 0, 1 or 5 mM glutaraldehyde (Glut).

Trp residues were essential for the heteromerization between the MPER- and IFP-derived peptides

Addition of M-wt to the biotinylated IFP-derived peptide IFP23b (Table 4) induced the formation of peptide dimer, trimer and higher-order oligomers, under chemical crosslinking (Figure 17A). Meanwhile, no peptide oligomers were formed when co-incubating M-wt with the biotinylated FP-derived peptide FP19b, nor with peptide TE20b, which was derived from an FP-proximal region (Figure 17A). This induced peptide oligomer formation was abrogated in the IFP23b/M-3W3A mixture, and restored in the IFP23b/M-3W3F mixture (Figure 17B).
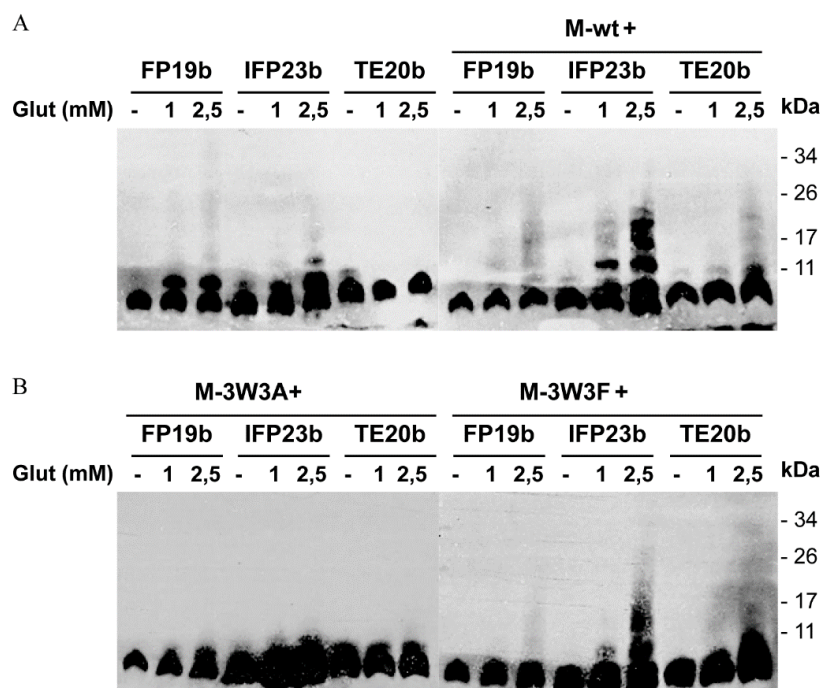
Figure 17. The S protein MPER-derived peptide M-wt heteromerized with the IFP-derived IFP23b, in a Trp-dependent manner. (A) Addition of M-wt induced the formation of hetero-oligomers between M-wt and IFP23b, as accessed by chemical crosslinking using glutaraldehyde (Glut). **(B)** Substitutions of the Trp residues within M-wt with Ala inhibited the formation of M-wt-IFP23b hetero-oligomer.

The MPER of various Class I fusion glycoproteins have been characterized as a membrane-active region, capable of membrane partition and lipid mixing [43, 142-144, 158-161, 166, 274]. Here, immunofluorescence staining of the cells that were treated with biotinylated MPER-derived peptides revealed that M-wtb signal co-localized with the lipid raft marker GM1 gangliosides (Figure 18A). The Ala substitutions abrogated the lipid-raft-binding of peptide M-3W3Ab and the Phe substitutions partially rescued the phenotype (Figure 18A).

Furthermore, co-application of M-wt and IFP23b to Vero E6 cells facilitated the recruitment of IFP23b, which was rarely detected when applied alone, to the cell membrane lipid rafts (Figure 18B). It suggests that the MPER- and IFP-derived peptides could form non-covalent complexes under a relatively physiological-relevant

condition as compared to chemical crosslinking. No increase in the IFP23b signal was observed in cells treated with IFP23b/M-3W3A, confirming the inability of IFP23b to bind to M-3W3A (Figure 18B). In contrast, an increase of the IFP23b signal was observed in cells treated with IFP23b/M-3W3F, highlighting that M-3W3F recovered the IFP23b-binding ability (Figure 18B).
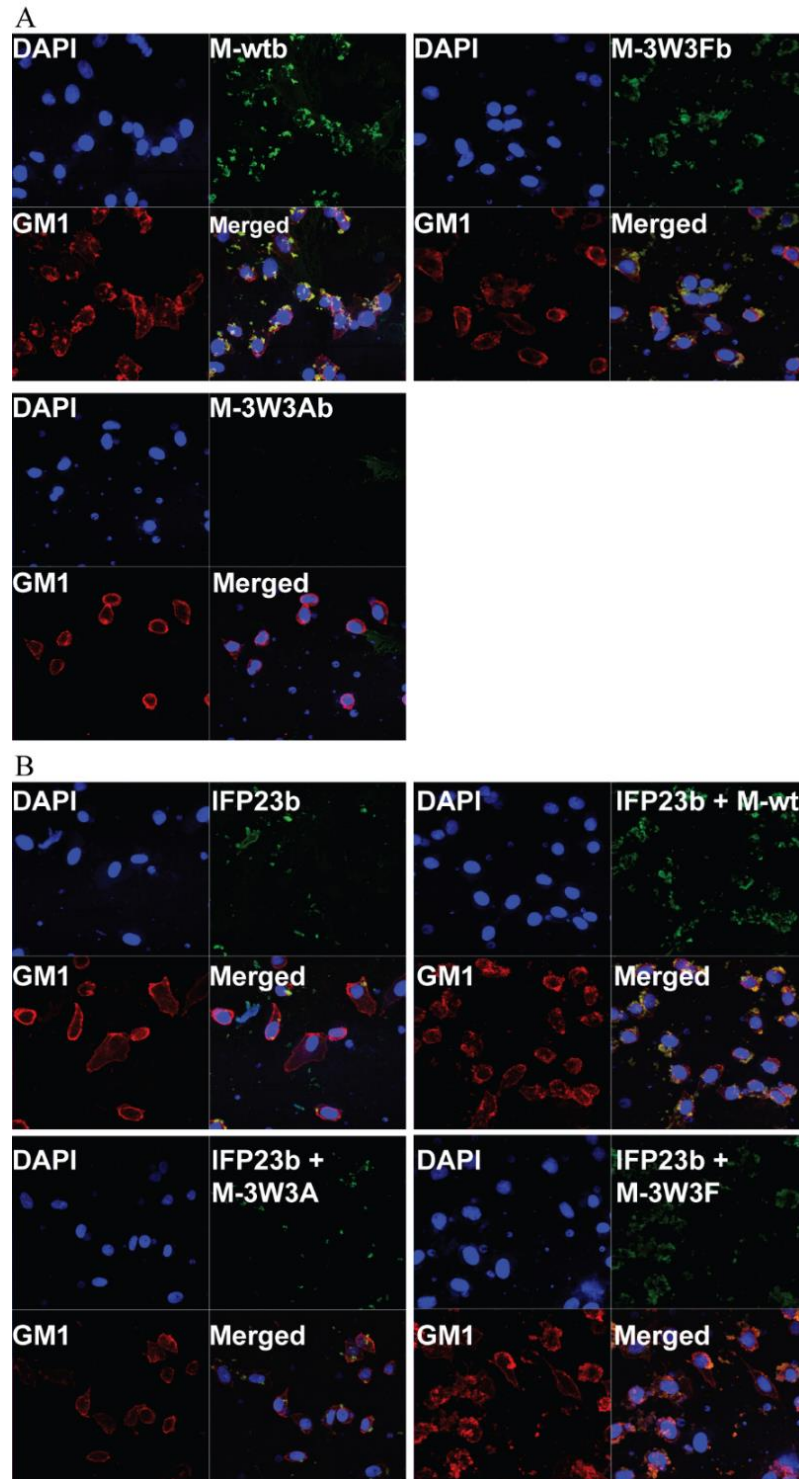
**Figure 18. Immunofluorescence staining confirms the Trp-dependent heteromerization between M-wt and IFP23b. (A)** Exogenously applied M-wtb bound to plasma membrane lipid rafts but not its Ala-substituted analogue M-3W3Ab. **(B)** M-wt enabled the

translocalization of IFP23b onto the plasma membrane lipid rafts, potentially through the M-wt-IFP23 interaction.

Synthetic MPER-derived peptides inhibited coronavirus entry and infection in a dose-dependent manner

M-wt inhibited the entry and infection of IBV, which contains a highly similar S protein MPER sequence as in SARS-CoV, in a dose-dependent manner (Figure 19). The peptide derived from the IBV MPER, M-ibv (Table 4), exhibited a similar level of inhibition. Meanwhile, the triple Trp→Ala substitutions in M-wt abrogated the antiviral efficacy of the peptide, it rather promoted the viral infectivity (Figure 19).
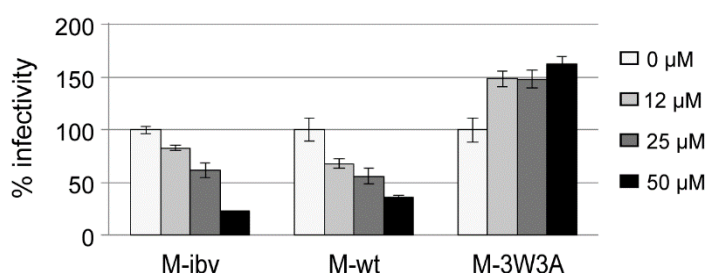
Figure 19. M-ibv and M-wt inhibited IBV-Luc infection in Vero E6 cells in a dose-dependent manner.

4.3   PAPER III

MEMBRANE-ACTIVE SEQUENCES WITHIN GP41 MEMBRANE PROXIMAL EXTERNAL REGION (MPER) MODULATE MPER-CONTAINING PEPTIDYL FUSION INHIBITOR ACTIVITY AND THE BIOSYNTHESIS OF HIV-1 STRUCTURAL PROTEINS

STUDY BACKGROUND

In **Paper II**, SARS-CoV S protein MPER-derived peptides inhibited the coronavirus entry and infection in a dose-dependent manner. **Paper III** further explored the anti-viral application of the MPER sequence but in the context of HIV-1. Previously, a partial N-terminal HIV-1 MPER sequence (WNWF) has been suggested to contribute to the anti-viral activities of T20 and T1249, two anti-HIV-1 fusion inhibitors, through conferring the peptides with membrane-reactivity [187, 201, 202, 275-277]. **Paper III** investigated the roles of two highly conserved membrane-active sequential elements, the local enrichment of Trp and the C-terminal CRAC motif LWYIK [133, 151, 152, 163, 165-167, 278, 279] (Figure 20), in the anti-viral activities of the short MPER-derived fusion inhibitor peptides. The fusion glycoproteins are synthesized as

type I transmembrane proteins and are matured through the secretory pathway, a process that involves multiple membrane-interactive steps. In **Paper III**, we further studied the roles of the membrane-active sequences of the HIV-1 Env MPER during the biosynthesis of the viral structural proteins.
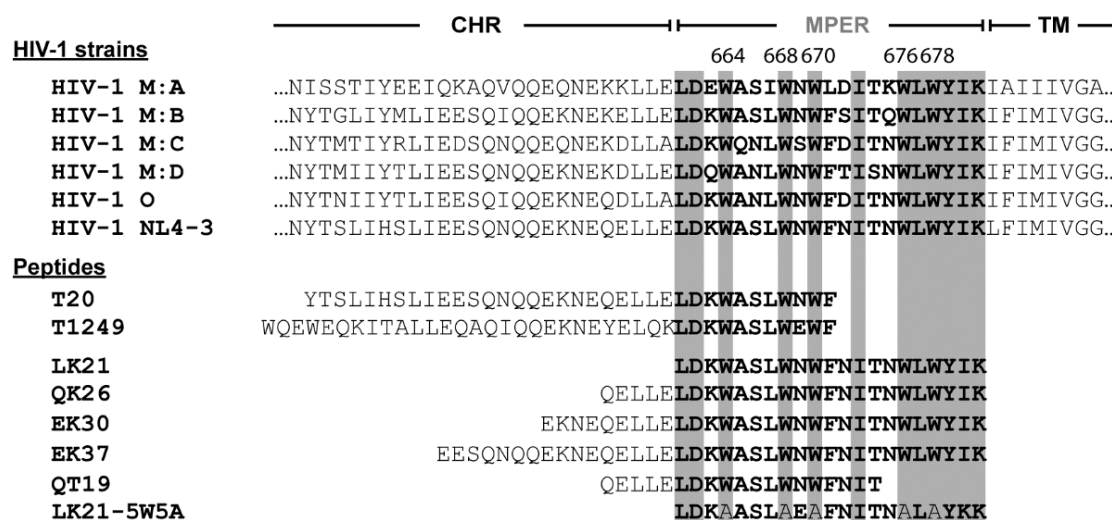
Figure 20. Sequence alignment of the MPER-derived fusion inhibitor peptides and the Env MPER from different groups of HIV-1.

The MPER-derived peptides required the membrane-active sequences to inhibit the pseudo-HIV-1 (NL4-3) infection

At 8.0 µM and 12.3 µM peptide LK21, containing exclusively the MPER sequence, inhibited viral entry and thereby infectivity by 50% and 80%, respectively (Figure 20) (Table 5). N-terminal extension of LK21 and, thus, inclusion of the HR2 residues, generating the peptide QK26, decreased the IC_{50} and IC_{80} values. Further N-terminal extension, resulting in EK30 and EK37 did not enhance the anti-viral efficacy further. In contrast, deletion of the C-terminal sequence, NWLWYIK, containing the CRAC motif LWYIK from the most active peptide QK26, abrogated the antiviral activity (Table 5). The abrogation of antiviral efficacy was also observed when the five Trp residues within peptide LK21 were substituted by Ala (Table 5).

Table 5. Effective concentrations of the MPER-derived peptides to inhibit 50% (IC_{50}) or 80% (IC_{80}) of the pseudotyped HIV-1 infection.

Peptides	IC_{50} (μM)	IC_{80} (μM)
LK21	8.0	12.3
QK26	3.9	8.8
EK30	5.9	18.5
EK37	3.1	12.6
QT19	>33.3	
LK21-5W5A	>33.3	

N-terminal dimerization of the MPER-containing peptides selectively enhanced viral inhibition

Addition of the HR2 hydrophilic residues to the MPER peptidyl analogue, LK21, greatly enhanced the solubility of the resulting peptides QK26, EK30 and EK37, allowing for peptide dimerization to possibly enhance their anti-viral efficacies. N-terminal dimerization of the MPER-containing peptides, EK30 and EK37, through a serine- and lysine-based linker showed beneficiary effect on their antiviral potencies. The N-terminal dimers (-DN), EK30-DN and EK37-DN, had IC_{50} values of 1.1 μM and 1.2 μM, an increase of the potency by 5.2- and 2.3-fold compared to the monomeric EK30 and EK37, respectively (Table 6). In contrast, the C-terminal dimers (-DC) of EK37 and QK26, EK37-DC and QK26-DC, had their IC_{50} values elevated by 1.4- and 1.3-fold, compared to the respective monomeric peptides (Table 6).

Table 6. Effective concentrations of the monomeric and dimeric MPER-containing peptides to inhibit 50% (IC_{50}) or 80% (IC_{80}) of the pseudotyped HIV-1 infection.

Peptides	IC_{50} (μM)	IC_{80} (μM)
QK26	2.4	6.3
QK26-DC	3.2	8.7
EK30	5.7	16.8
EK30-DC	2.6	>33.3
Ek30-DN	1.1	4.9
EK37	2.8	11.6
EK37-DC	4.0	>33.3
EK37-DN	1.2	5.1

Trp residues maintained the structural plasticity of the MPER peptidyl analogue, LK21

The secondary structures of the peptides LK21 and LK21-5W5A were determined by circular dichroism, along with the α-helical control peptide HR2-S derived from the SARS-CoV S protein HR2 region [280]. 10% – 40% of TFE was used to mimic the increase in environmental lipidity, which the MPER undergoes during membrane fusion. In the CD spectroscopy study, LK21 exhibited α-helical contents of 41% at 10% TFE, increasing up to 71% when with 40% TFE (Figure 21A), while LK21-5W5A exhibited an α-helical content of 84% starting from 20% TFE (Figure 21B).
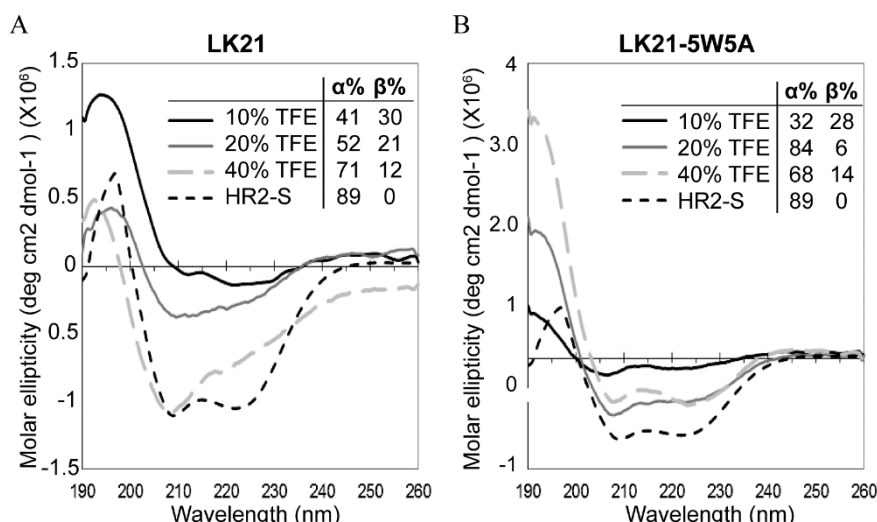
Figure 21. Secondary structural contents of (A) LK21 and (B) LK21-5W5A in increasingly hydrophobic environment.

Ala substitutions of the MPER Trp residues affected viral structural protein expression

The importance of Trp residues in maintaining the membrane reactivity and the correct structure of the MPER sequence indicated that the MPER Trp residues might modulate the Env structure and/or its interaction with plasma membranes during the biosynthesis of Env. Ala substitutions of the all five Trp, the three N-terminal Trp (W664, W668 and W670) or the two C-terminal Trp (W676 and W678) of the MPER within a pseudotyped HIV-1 (NL-43), decreased the intracellular Env levels to 60% of that in HIV(WT) at 48 h post-transfection (Figure 22A).

More interestingly, Ala substitutions of the three N-terminal MPER Trp residues significantly elevated the expression of p55/Gag, resulting in a higher p24/capsid level. At 48 h post-transfection, intracellular p55/Gag and p24 levels in cells producing HIV(W3A) were 93 % and 230% higher compared to HIV(WT), while the expression of two other viral proteins Vif and Nef stayed unaffected (Figure 22A).

Furthermore, the precipitated HIV(W3A) viral particles contained 200 % and 310 % higher levels of p55/Gag and p24/capsid, respectively, as compared to HIV(WT) particles that were precipitated from equal volume of cell culture supernatant (Figure 22B). This also indicated that HIV(W3A) viral particle production was increased as a result of the higher intracellular p55/Gag expression.

Figure 22. Ala substitutions of the MPER Trp residues affected the expression and viral incorporation of Env and p55/Gag. (A) At 48 h post-transfection, lysates of cells expressing HIV(WT), HIV(W5A), HIV(W3A) and HIV(W2A) were studied for the levels of Env gp41, p24/capsid, Vif, Nef and β-actin. **(B)** Env gp41, p55/Gag and p24/capsid levels in HIV(WT), HIV(W5A), HIV(W3A) and HIV(W2A) that were precipitated from equal volume of the cell culture supernatant in A. Un-transfected HLtat cells served as negative control in A and B.

4.4 PAPER IV

THE GP160 SIGNAL SEQUENCE SUPPRESSES HIV-1 REPLICATION IN AN LTR-DEPENDENT MANNER

STUDY BACKGROUND

In **Paper III**, Ala substitutions and possible structural rearrangement within the Env MPER induced a significant increase in both intracellular and intraviral content of p55/Gag. **Paper IV** further demonstrated that another Trp-rich Env region, the N-

terminal signal peptide (ss) (Figure 23), could regulate the HIV-1 replication and the expression of viral proteins, particularly p55/Gag.
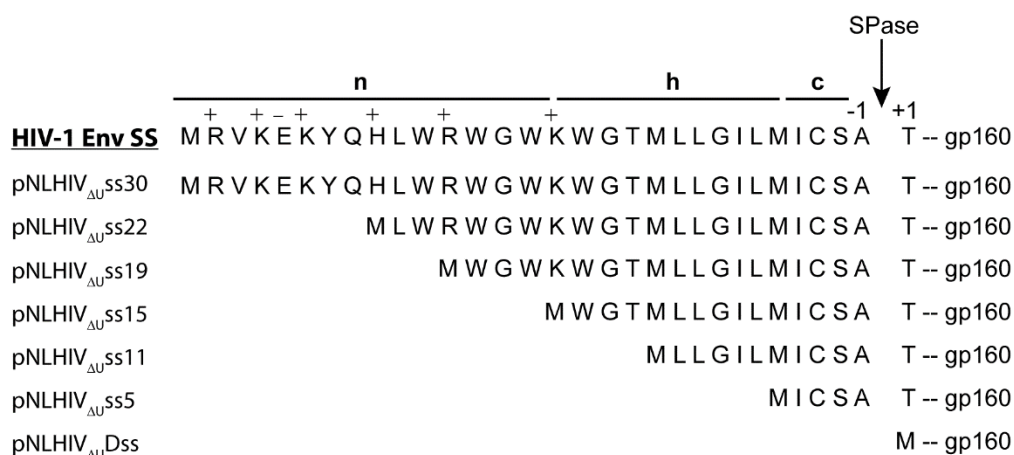
Figure 23. Schematic representation of the HIV-1 Env ss and the proviral vectors containing WT or mutant Env ss. The amino acid sequence of the native Env ss is shown with indications of its hydrophilic (n), hydrophobic (h) and C-terminal (c) regions, the cleavage site for signal peptidase (SPase) as well as the positively (+) and negatively charged residues.

The Env ss suppressed the p24/capsid expression and viral particle production

Expression of the six mutant HIV-1 proviral constructs HIVΔuss22, HIVΔuss19, HIVΔuss15, HIVΔuss11, HIVΔuss5 and HIVΔuΔss that contained 22 aa, 19 aa, 15 aa, 11 aa, 5 aa and 0 aa of Env ss, respectively (Figure 23), elevated intracellular and extracellular p24/capsid levels and enhanced the viral particle production as compared to the WT HIV-1, or HIVΔuss30 (Figure 24A-B). This suppressive effect of WT Env ss on p24/capsid expression became less pronounced from 18 h to 48 h post-transfection.
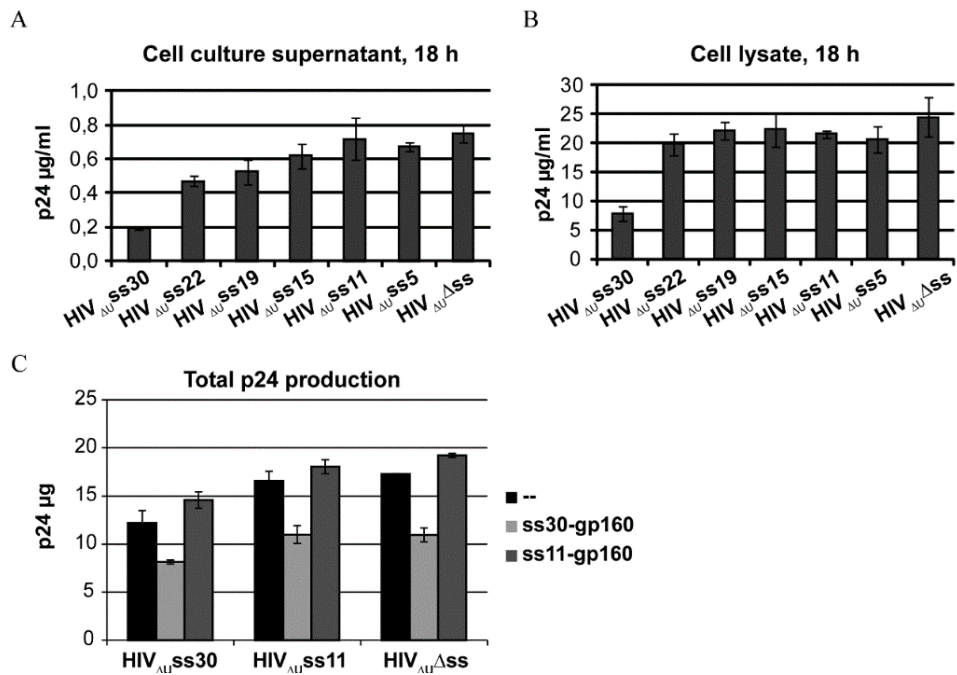
Figure 24. The Env ss suppressed the viral p24/capsid expression. (A) HIV-1 containing truncated ss had higher extracellular and **(B)** intracellular p24/capsid levels than WT HIV-1. **(C)** Co-expression of ss30-gp160, but not ss11-gp160, suppressed the p24 production from HIVΔuss30, HIVΔuss11, and HIVΔuΔss.

Introducing a copy of WT Env ss *in trans* enhanced the suppression of p24/capsid expression, further showing that the suppressive effect of Env ss is additive. At 18 h post-transfection, the expression of the WT ss from ss30-gp160, but not the 11 aa-long ss from ss11-gp160, suppressed the total p24/capsid production from HIVΔuss30, HIVΔuss11, or HIVΔuΔss in HeLa-tat-III cells (Figure 24C). The suppression of p24/capsid production was most pronounced when the WT Env ss was expressed from both HIVΔuss30 and the ss30-gp160 (Figure 24C).

The Env ss suppressed the expression of p55/Gag, p160/Gag-Pol and Vif

The Env ss was found to suppress the p24/capsid levels through inhibiting the expression, but not the proteolytic maturation, of its precursor protein p55/Gag and p160/Gag-Pol. At 48 h post-transfection, the intracellular p55/Gag and p160/Gag-Pol levels were higher in cells expressing HIVΔuΔss co-transfected with ss11-gp160 than co-transfected with ss30-gp160. Also the intracellular levels of another late HIV-1 protein, Vif, was up-regulated when the cells were co-transfected with HIVΔuΔss and ss11-gp160, as compared to the condition where HIVΔuΔss was co-transfected with the WT ss30-gp160 (Figure 25).
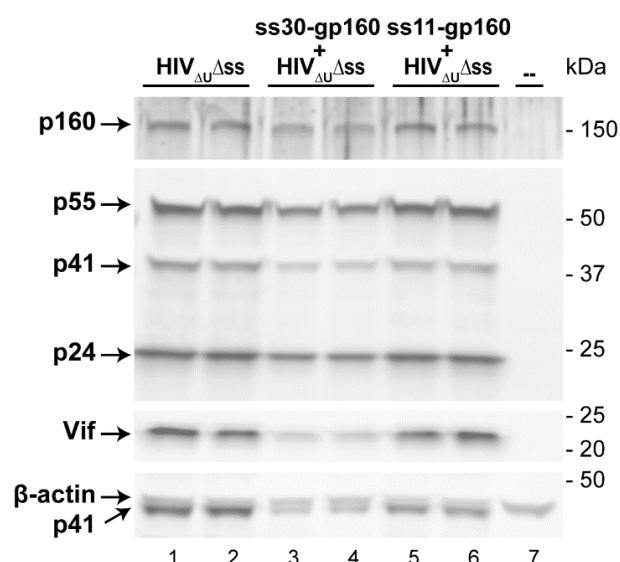
Figure 25. The Env ss suppressed the expression of Vif and the p24/capsid precursor proteins, p55/Gag and p160/Gag-Pol.

The Env ss downregulated the full-length viral RNA transcripts in a time-, dose- and LTR-dependent manner

The WT Env ss was found to suppress the full-length viral RNA transcript in a time-, dose- and LTR-dependent manner. In cells that were co-transfected with HIVΔuΔss and constant amounts of Env-expressing plasmids, the full-length viral RNA transcript in the presence of either ss30-gp160 or ss11-gp160 started out to be equal at 8 h post-transfection (Figure 26A). However, at 24 h and 48 h post-transfection, the RNA levels in the presence of ss30-gp160 were lower than with ss11-gp160 by about 13% and 30%, respectively (Figure 26B-C). When keeping the harvest time-point at 24 h post-transfection and increasing the amount of Env-expressing plasmids from 0.25 to 0.5 and then to 0.75 μg, the full-length viral RNA transcript in the presence of ss30-gp160 was 14%, 25% and finally 50% lower than that of ss11-gp160 (Figure 26D-F). Furthermore, the expression of an LTR-driven CAT reporter molecule, in combination with HIVΔuΔss and ss30-gp160, was lower than its expression in combination with HIVΔuΔss and ss11-gp160 (Figure 26G). No significant difference in cell viability or total protein concentration was observed.
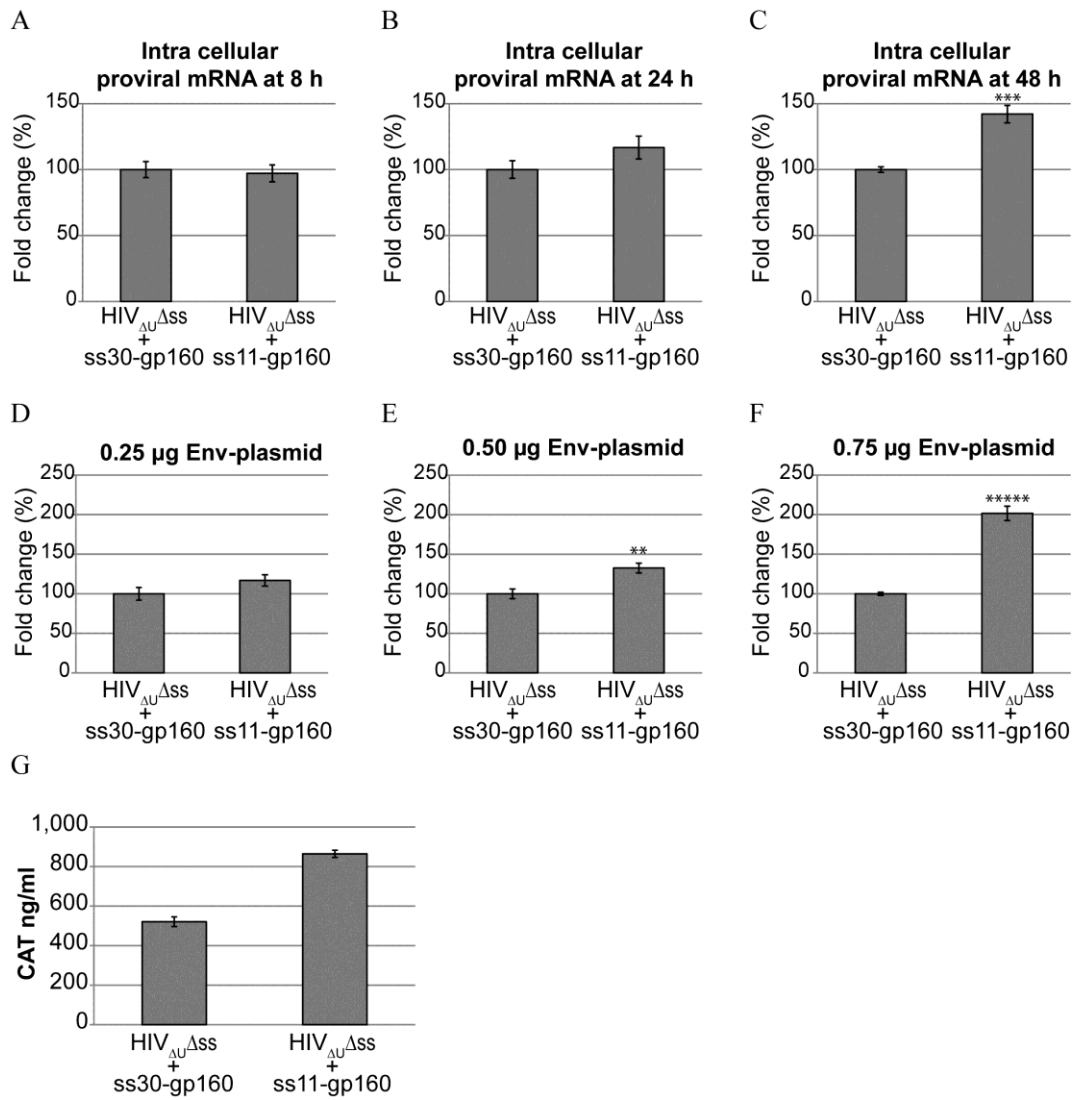
Figure 26. The Env ss suppressed the full-length viral RNA transcript levels. (A) In HLtat cells expressing HIVΔuΔss and ss30-gp160 or ss11-gp160, full-length viral transcript levels in the presence of ss30-gp160 were increasingly lower than those of ss11-gp160 from 8 h, **(B)** 24 h then **(C)** 48 h post-transfection. **(D)** When increasing the Env-expressing plasmid from 0.25 µg, **(E)** 0.5 µg then to **(F)** 0.75 µg, full-length viral transcript levels in the presence of ss30-gp160 was increasingly lower than those of ss11-gp160. ** $p < 0.01$, *** $p < 0.001$, ***** $p < 0.00001$. **(G)** The WT Env ss expression inhibited the LTR-driven expression of the reporter molecule CAT enzyme.

5 DISCUSSION

The viral fusion glycoproteins of enveloped viruses mediate the first steps of establishing a viral infection. They are considered important targets for both viral inhibition and detection. However, they are usually heavily glycosylated, camouflaging their functionally important regions from the immune system and from anti-viral interventions. Furthermore, they are often mutating to escape immunological attack and drug targeting. Nevertheless, some sequences within the fusion glycoproteins, which are involved in infection-required protein-protein and protein-ligand interactions, are usually exposed and well preserved both structurally and functionally. Their structural and functional stabilities have further allowed peptide analogues of these sequences to be used for *in vitro* studies of virus-host interactions as well as development of anti-viral therapeutics [38, 47, 182-187].

Class I viral fusion glycoproteins are each composed of a receptor-recognizing unit and a membrane-fusing unit. Within the Class I viral fusion glycoproteins, the most well-studied interaction-mediating sequences are the HRs within the membrane-fusing transmembrane unit, which are instrumental for the formation of fusion core [4]. Peptidyl analogues of the HRs, so far, have been the most successful peptidyl fusion inhibitors [186, 187, 193-196]. They inhibit the viral entry putatively through dominant-negative binding to the viral HRs, thereby preventing the HR heteromerization to form the fusion core. Examples include T20, a clinically approved anti-HIV-1 fusion inhibitor, which contains the sequence of both HR2 and MPER of gp41 [187]. However, in drug-experienced patients, HIV-1 is constantly evolving, which leads to the development of virus strains that are resistant to HR fusion inhibitors such as T20 [208, 209]. This warrants the search for other interaction-mediating sequences within Class I glycoproteins, which could be targets for drug intervention. Furthermore, for viruses such as SARS-CoV, there is no effective treatment. Identifying such new target sequences would be useful for both the development of new antivirals and for the study of viral entry mechanism.

In this thesis, three interaction-mediating sequences of two Class I viral fusion glycoproteins, the SARS-CoV S protein and the HIV-1 Env, were studied within the regions spanning from the extreme N-terminal signal sequence to the C-terminal MPER. The results suggested that they could promote viral entry at different stages, where they interact with different proteins and ligands such as the membranes. Their identification and potential involvements in the viral replication process extend our current knowledge of the Class I viral fusion glycoproteins-mediated viral entry and infection.

Recognition and interaction with the viral receptor on the target cells is the first step of the viral entry process meditated by the Class I viral fusion proteins. In SARS-CoV S protein, a self-interacting sequence (C6, **Paper I**) containing a putative cross-β spine steric zipper motif GINITNFR was identified in the region spanning aa 213 to

232. The self-interaction ability of peptide C6 could offer some insight into the mechanisms of the SARS-CoV receptor recognition and its inhibition. The C6 sequence is located upstream to the RBD that interacts with the main viral receptor, ACE2. A previous study using pseudotyped virus has shown that point mutations within this region (aa 215-232) could render the virus incapable of infecting the ACE2+ target cells without affecting the S protein expression [281]. Furthermore, a peptide derived from this region inhibits the ACE2-mediated viral entry in a dose-dependent manner [281]. With the region containing the C6 sequence being mapped to the exposed accessible region on the S protein model structure (PDB: 1T7G) [282], it is possible that the C6 regions between neighboring S proteins could self-interact along the viral entry process. Possible scenarios include that the self-interaction of the C6 regions may prepare the virus to attach onto the ACE2. Alternatively, the self-interaction of the C6 region happens post-receptor-binding and leads to additional conformation changes within the S protein, which are essential for the subsequent steps of viral entry. Exogenous application of the C6 peptide could obstruct the self-interaction of this region and, hence, inhibit the viral receptor binding or affect the S protein conformational change following receptor binding. SARS-CoV primarily utilizes the ACE2 to enter the target cells, but it can also establish infection through interacting with DC/L-SIGN [28]. Most interestingly, the C6 region, specifically its N-glycosylation site at N227, is also essential for the DC/L-SIGN-mediated viral entry and it is critical for the virus to achieve zoonotic transmission [28, 283, 284]. This raises the possibility that the C6 region may also participate in a glycan-mediated interaction with DC/L-SIGN and, thereby, facilitate the SARS-CoV entry into human cells.

The receptor recognition by a Class I viral fusion glycoprotein prompts the structural changes within its transmembrane membrane-fusing unit, resulting in the exposure and insertion of the N-terminal FP into the target cell membrane, which extends the two HRs following the FP. At this pre-fusion stage, both of the HRs are believed to be in a trimeric state. Peptides derived from the MPER of the SARS-CoV S protein, the hydrophobic region following the trimeric C-terminal HR, was shown to be capable of oligomerizing (**Paper II**). Furthermore, substitutions of the three Trp residues within the S protein MPER-derived peptide with the aliphatic amino acid Ala abrogated the peptide oligomerization, while with the aromatic amino acid Phe restored the phenotype (**Paper II**). This raises the possibility that the MPER-derived peptide or the S protein MPER could oligomerize through a zipper-like motif held together by multiple aromatic residues, i.e. an aromatic amino acid zipper (AAAZ) motif (Figure 27). The AAAZ motif is generated through the hydrophobic interaction (e.g. π-π interaction) between the interdigitating aromatic ring side-chains. Such hydrophobic interactions between aromatic amino acids including Trp, Phe and tyrosine (Tyr), have been reported to stabilize peptide or protein structures and to induce their oligomerization [285]. Due to the essentialness of the Trp residues in its

48

oligomerization, the S protein MPER-derived peptide likely oligomerized through a Trp zipper, where the indole ring side-chains of the Trp residues of the neighboring MPER-derived peptides were stacked at various angles [286-288]. Previously, a peptide derived from the HIV-1 Env MPER has also been reported to be able to oligomerize [169]. MPER is an envelope-proximal region and contains multiple membrane-active elements such as Trp and lipid-interacting motifs, which are instrumental for lipid mixing and the envelope-membrane fusion during the entry of SARS-CoV and HIV. It is puzzling how the MPER is folded into the viral fusion proteins so that it is not activated prior to the actual membrane fusion, as such premature activation of the MPER would disrupt the virus envelope. The answer might lie in the ability of the MPER to trimerize, as a continuation to the trimeric stalk-like C-terminal HR. It is possible that the oligomerization of MPER, maintained by the Trp residues, serves as a safety mechanism to preclude the membrane-active sequence elements from interacting with the envelope prematurely.
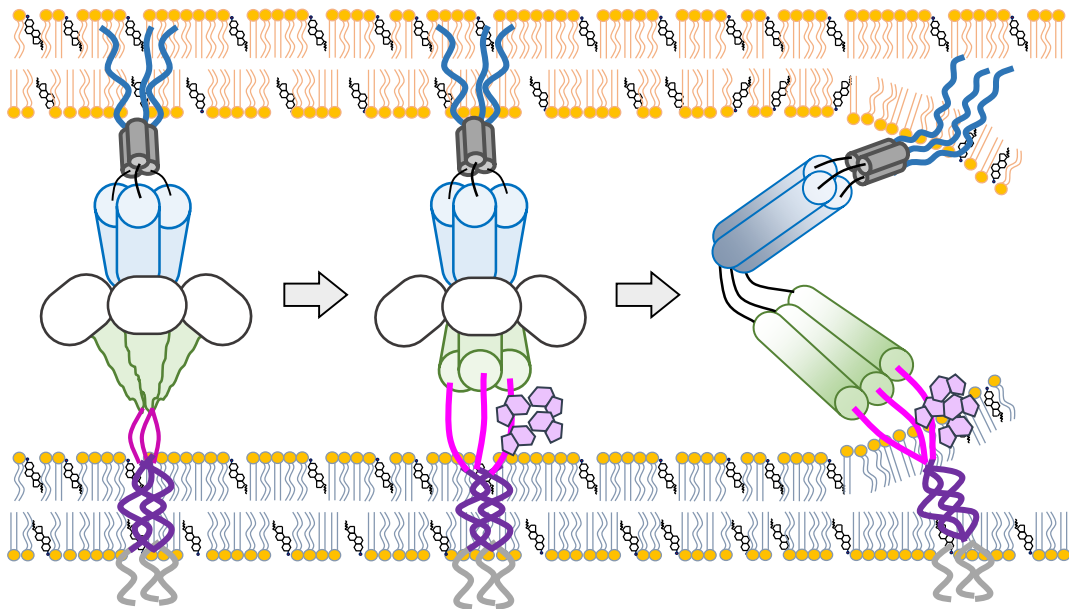
Figure 27. Schematic representation of the Class I viral fusion protein-mediated viral entry, at the pre-fusion stage. At the pre-fusion stage, post-receptor binding and prior to 6-HB formation, we propose that the MPER on neighboring fusion proteins trimerize *via* an AAAZ motif formed between the aromatic residues within MPER. Grey cylinder, FP or IFP; blue cylinder, N-HR; green cone/cylinder, C-HR; magenta, MPER; purple, TMD; and grey line, cytoplasmic tail. The Trp residues within MPER are highlighted by the pentagons in purple. The viral envelope is in blue and the target cell membrane in orange.

As the viral entry process proceeds, the extended HRs will slowly come together to form the 6-HB fusion core, where the C-terminal HR trimers dissociate and each monomer binds anti-parallel onto the trimeric N-terminal HR coiled-coil. The formation of the 6-HB fusion core is believed to be the key conformational change in bringing the viral envelop and cell membrane close for lipid mixing and fusion.

Concurrently, the formation of 6-HB also approximates the FP and MPER, the highly hydrophobic regions upstream to the N-terminal HR and downstream to the C-terminal HR, respectively. The high hydrophobicity of both the MPER and FP indicates that there may be an interaction between these two regions. In this study, we could show that peptides derived from the SARS-CoV S protein MPER interacted selectively with a peptide derived from the IFP, which is the region immediately N-terminal to HR1. The SARS-CoV S protein contains two hydrophobic regions upstream to the N-terminal HR, the FP and downstream thereof IFP [41, 42]. Previously, the potential interactions between the MPER, FP and IFP of the S protein have been investigated in artificial lipid membranes and/or vesicles [161], but these studies have generated inconsistent data. An MPER-derived peptide could cooperatively disrupt lipid membranes with an IFP-derived peptide but not an FP-derived peptide. Meanwhile, accessibility experiments suggest that an MPER-derived peptide interacted with an FP-derived peptide, but not one derived from the IFP [161]. In **Paper II**, the selective MPER-IFP interaction was confirmed both by biochemical crosslinking and by immunofluorescence staining on cellular membranes. The CD spectrum of an equimolar mixture of the MPER- and IFP-derived peptides shifted 4 nm to the left from the theoretical CD profile averaged from the individual profiles of either peptide, further suggesting that there is a weak MPER-IFP interaction existing in solution. The preference of the MPER to interact with the IFP but not with the FP may be owing to the closer physical proximity between the MPER and IFP upon the formation of the fusion core in the S protein, when the FP is directed to insert into the target membrane. Previously, a similar interaction has been reported between the HIV-1 Env MPER and FP. In a mechanistic study, it was shown that inclusion of the MPER and FP sequences to synthetic peptides derived from the N- and C-terminal HRs, respectively, enhanced the stability of the 6-HB, which may provide additional free energy to approximate the opposing membranes [152, 169, 170] (Figure 28).
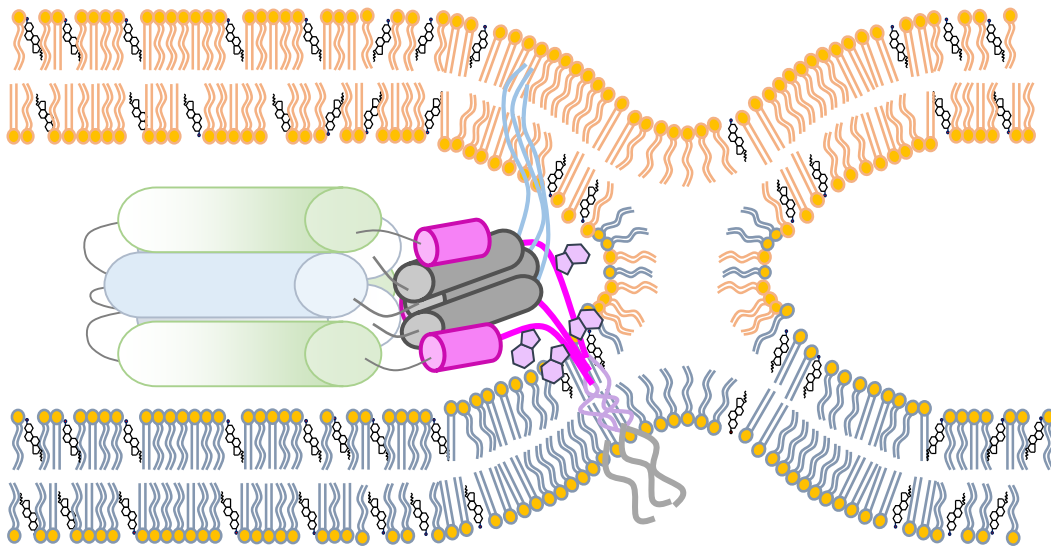
Figure 28. Schematic representation of the formation of 6-HB fusion core, in conjunction with the MPER-FP interaction. Upon the formation of 6-HB, we propose that the MPER will interact with the FP, or the immediate hydrophobic region upstream of the N-HR. The MPER-FP interaction strengthens the 6-HB and facilitates the approximation of the opposing membranes. The increase in local lipidity could induce the entire or part(s) of the MPER to adopt helical conformation, which exposes its membrane-active sequence elements, such as Trp, for lipid mixing. The color scheme in this figure follows that in Figure 27.

As the opposing membrane approaches one-another, the local lipidity increases. In both a previous study and as observed in **Papers II** and **III**, an increase in the environmental hydrophobicity or lipidity could induce an increase of helical contents within the peptides derived from the MPER of both HIV-1 Env and SARS-CoV S proteins [172, 173]. This lipidity-induced structural transformation may facilitate the exposure of the membrane-active sequence elements within the MPER, which thereafter participate in lipid mixing (Figure 28). Indeed, in association with lipid membrane, a peptide derived from the HIV-1 Env MPER exhibited a bended two-helix conformation, where the tilted N-terminal α-helix lies on the membrane, while the flat C-terminal helical segment is embedded in the membrane putatively for lipid mixing. Furthermore, both the structural transformation of the MPER-derived peptide, and its interactions with the N-terminal hydrophobic region and the lipid membranes required essential participation of the Trp residues (**Papers II** and **III**) [166]. This is in agreement with the previous observations that the Trp→Ala substitutions in the MPER abrogated the entry/infectivity of SARS-CoV and HIV-1 [158, 162, 163]. We propose that the interaction between the MPER and the immediate hydrophobic region upstream to the N-HR, together with the lipidity-induced MPER structural transformation, may collectively serve as a general, Trp-dependent mechanism among the Class I viral fusion glycoproteins in promoting the membrane fusion (Figure 28).

While we have studied the effects of Class I viral fusion glycoproteins on the viral entry process, we could also identify regions that affected viral infectivity during the post-entry steps of viral replication. Previous studies have suggested that the HIV-1 Env expression down-regulates the steady-state intracellular levels of p55/Gag protein, at both protein level through the Env cytoplasmic tails and at the RNA level through acting on the Rev response elements within *env* [289, 290]. In **Papers III** and **IV**, we observed two more Env regions that regulated the expression of p55/Gag, the MPER and the ss. Ala substitutions of the Trp residues within the Env MPER moderately decreased the Env expression, while significantly elevating the intracellular p55/Gag levels (**Paper III**). Contradictory to the previous studies on the relationship between Env and p55/Gag expressions, the extent of the increase in p55/Gag levels did not correlate with the decrease in Env expression. This suggests that the Env MPER could regulate p55/Gag expression in a yet-to-be defined mechanism.

Besides the MPER, the ss of Env was shown to down-regulate the p55/Gag level and to suppress viral production, unrelated to the Env expression level. In **Paper IV**, N-terminal truncation of the Env ss elevated the expression levels of viral proteins, p55/Gag, p160/Gag-pol and Vif. The positively charged n region (MRVKEKYQHLWRWGWK, aa 1-16) within the Env ss functions to retard the protein folding and maturation in the ER [122, 123]. However, according to our observation and the studies carried out by other groups, the basic residues of this region appear to be entirely dispensable for the Env expression or maturation [122, 291]. The N-terminal 8 aa are completely dispensable for adequate Env expression, and truncation of the first 19 aa does not affect the ER-targeting and maturation of the mutant Env protein [252]. These observations exclude the possibility that the elevation of viral production upon N-terminal truncation of the Env ss was due to an altered Env production. Instead, by studying the expression of an LTR-driven reporter molecule in the presence of either N-terminally truncated or WT ss, we could show that the WT Env ss suppressed the viral replication through the inhibition of the LTR-dependent viral genome transcription.

As compared to the ss within the ER-residing Env, which is likely constrained temporospatially, the signal peptide that is cleaved off the ss has bigger freedom to conduct the necessary interactions for suppressing the viral replication. The potential involvement of the Env signal peptide is supported by the observation that the Env signal peptide is not degraded immediately, but processed and further released into the cytosol [292]. A processed signal peptide fragment containing the N-terminal 23 aa may further antagonize the cell signaling molecule calmodulin (CaM) [292]. CaM is a calcium-sensor protein that can promiscuously interact with hundreds of proteins and thereby regulates numerous cellular functions and metabolic pathways [293-299]. In relation to HIV-1, a CaM-binding site has been identified *in vitro* within Tat [300]. The Tat CaM-binding sequence (QVCFITKALGISYGRKKRRQR) further overlaps

substantially with the Tat viral RNA-binding motif RKKRRQRRR (a.a.49 – 57 in strain NL4-3) [300-302]. This raises the possibility that the Tat-LTR interaction and/or proviral transcription could be influenced by the Tat-CaM interaction, which in turn could be linked to the signal peptide-CaM interaction. Alternatively, the Env signal peptide or its fragments could translocate into the nucleus and directly influence the LTR-driven proviral transcription. Signal peptides derived from the Env of some other retroviruses have been shown to accumulate in the nucleus or at the viral budding sites to exert various post-ER-targeting activities, [303-309]. However, prediction using online software failed to identify a nuclear localization signal (NLS) within the HIV-1 Env ss. The lack of ss-specific antibody, owing to unsuccessful attempts to immunize rabbits with peptides derived from the Env ss, further hampered our ability to localize the Env signal peptide in HIV-1 infected cells.

The putative Env signal peptide-induced suppression of the HIV-1 production demonstrated a time- and dose-dependency. This indicates that the suppression was at a late stage of the viral replication cycle when sufficient level of Env and, hence, signal peptide would have accumulated. This could be part of a negative feedback mechanism, which would favor sustainability over sporadic viral production and protect the host cells from fast exhaustion.

Besides regulating the viral entry and/or viral replication, all three interaction-mediating sequences studied here, the C6 region within SARS-CoV S protein, the MPERs within S protein and HIV-1 Env, and the Env ss; are located to surface exposed accessible regions with high levels of sequence conservation. These characteristics make them promising targets for therapeutic interventions and for diagnostic recognition of the virus. In **Papers I, II** and **III**, peptidyl analogues of the C6 region and MPER were studied for their use as diagnostic tool and as fusion inhibitors, respectively. We could show that peptide C6 at 2 μM detected the S protein expression in an immunofluorescence assay, with comparable staining pattern and signal as a conventional anti-S protein antibody (**Paper I**). This ability of peptide C6 to mimic an antibody in detecting protein expression opens up its potential to be further developed as a peptide-based diagnostic and research tool. In **Papers II** and **III**, the synthetic MPER-derived peptides showed dose-dependent inhibition of the entry and infection of coronavirus and HIV-1, putatively through dominant negative binding to the fusion glycoproteins, thereby, obstructing the MPER to oligomerize or heteromerize with the FP. Additionally, the membrane-active sequence elements, i.e. the aromatic residues and cholesterol-interacting CRAC motif, could anchor the MPER-derived peptides to lipid rafts, the viral entry sites of both SARS-CoV and HIV-1(**Paper II**) [187, 201, 202, 275-277]. Elimination of either Trp residues or the CRAC motif within the MPER-derived peptides abrogated the anti-viral efficacy of the peptides (**Paper III**). This suggests that the local enrichment of the MPER-derived peptides to the viral entry sites may additionally contribute to their anti-viral efficacy [187, 201, 202, 275-277].

Peptide multimerization, which increases the interaction sites on peptides, is an effective way to enhance the avidities of peptide-ligand interactions. It was employed in this study to enhance the protein detection sensitivity of peptide C6 (**Paper I**) and the viral inhibition efficacies of the Env MPER-derived peptides (**Paper III**). Albeit effective in enhancing the C6-S protein interaction, C6 dimerization also increased the propensity of the C6 sequence to self-aggregate and form amyloid-like fibrils (**Paper I**). This prevented the dimeric C6 peptide to be applied to detect the S protein expression *in vitro*. This could be potentially circumvented by including a β-sheet-breaker residue such as Pro into the peptide sequence. This method has been shown to effectively reduce the aggregation propensity of self-interacting peptides, without sacrificing their ability to specifically bind to their respective target protein [234, 310, 311]. Furthermore, multimerization of a peptide to enhance its interaction with the cognate protein, such as the interaction between the MPER-derived peptides and HIV-1 Env, should be applied with structural consideration. In **Paper III**, peptide dimerization up to 5-fold enhanced the antiviral efficacies of HIV-1 Env-derived peptides containing the entire MPER and a very short HR2 sequence. Notably, the N-terminal, as compared to C-terminal, dimerization of the MPER-derived peptides was more effective in enhancing the viral inhibition of the peptides. The better anti-viral efficacies of the N-terminal dimers are in agreement with their resemblance of the active 'tripod' conformation of Env MPER, where the MPER N-termini trimerize and the C-termini stay separated and membrane-bound [145, 312]. In contrast, C-terminal dimerization disfavors the MPER-derived peptides to mimic the active conformation of the Env MPER. Moreover, C-terminal dimerization could further constrain the peptide C-termini and prevent the CRAC motif from interacting with membranes and anchoring the peptides to the viral entry sites.

This study focused on three interaction-mediating sequences within two Class I viral fusion glycoproteins. They are the C6 sequence upstream of the RBD within the SARS-CoV S protein, the MPER within both the S protein and HIV-1 Env, and the ss of Env. We could show that they regulate the viral infectivity at different stages of the viral replication, and surprisingly not just confined to the viral entry process. Their potentials in diagnostic and therapeutic applications have also being studied and confirmed with their peptide analogues.

6 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The main findings of this thesis are:

- The identifiaction of a self-interacting sequence GINITNFR upstream of the RBD within the SARS-CoV S protein, which could potentially form cross-β spine steric zipper and which further could be applied to detect the S protein expression *in vitro*.

- The identification of the Trp-dependent oligomerization of a SARS-CoV S protein MPER-derived peptide and its selective heteromerization with an IFP-derived peptide.

- The peptides derived from the SARS-CoV S protein or HIV-1 Env MPER could inhibit viral entry in a dose-dependent manner, where the membrane-active sequential elements within the MPER-derived peptides, the Trp residues and/or the cholesterol-interacting motifs, played essential roles.

- The peptide-protein interaction avidity and the anti-viral efficacy could be effetively enhanced through peptide multimerization.

- HIV-1 Env has profound regulatory roles on viral replication and on viral protein expression, as observed in the mutant viruses containing Env with an Ala-substituted MPER or a N-terminally truncated signal sequence. Particularly, the HIV-1 Env signal sequence could downregulate the viral production in a time-, dose- and LTR-dependent manner.

Rapid and economic generation of practical detection reagent is instrumental for the diagnosis and treatment of viral infections, particularly for emerging pathogens. Antibodies have been invaluable research, diagnostic and therapeutic tools, due to their potent and specific interaction with the antigenic epitopes. However, the requirement of an immunogenic molecule and the stringent storage and production conditions limit the field applications of antibodies. These issues could be partially circumvented by identifying self-binding peptides from the protein-of-interest, such as GINITNFR within the SARS-CoV S protein. The sequence-specific peptide-protein interactions could potentially enable these self-binding peptides to mimic functions of antibodies and to be further developed as alternatives to antibodies. Nevertheless, future studies are needed to develop effective strategies to improve the interaction avidity and specificity between the self-binding peptides and the protein-of-interest. According to previous research and the work reported on in this thesis, MPER within the Class I viral fusion glycoproteins facilitate the membrane fusion with its flexible structure, dynamic oligomerization states, and multiple interaction

partners including both protein domains and lipid molecules. All of these interactions require essential participation of the conserved Trp residues. To integrate and relate the observed findings to viral entry and inhibition in a temporal fashion, detailed structural study under biologically relevant environments and in the context of the whole protein are needed. Finally, the mechanism through which the HIV-1 Env signal sequence suppressed the viral production requires detailed studies, for this interesting effect to be exploited therapeutically.

7 REFERENCES

1. Lee, K.K., *Architecture of a nascent viral fusion pore.* EMBO J, 2010. **29**(7): p. 1299-311.

2. Yang, L. and H.W. Huang, *Observation of a membrane fusion intermediate structure.* Science, 2002. **297**(5588): p. 1877-9.

3. Leikin, S.L., et al., *Membrane fusion: overcoming of the hydration barrier and local restructuring.* J Theor Biol, 1987. **129**(4): p. 411-25.

4. White, J.M., et al., *Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme.* Crit Rev Biochem Mol Biol, 2008. **43**(3): p. 189-219.

5. Harrison, S.C., *Viral membrane fusion.* Virology, 2015. **479-480**: p. 498-507.

6. Skehel, J.J. and D.C. Wiley, *Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin.* Annual Review of Biochemistry, 2000. **69**(1): p. 531-569.

7. Chan, D.C., et al., *Core structure of gp41 from the HIV envelope glycoprotein.* Cell, 1997. **89**(2): p. 263-73.

8. Bosch, B.J., et al., *The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex.* Journal of Virology, 2003. **77**(16): p. 8801-8811.

9. Kuhn, R.J., et al., *Structure of dengue virus: implications for flavivirus organization, maturation, and fusion.* Cell, 2002. **108**(5): p. 717-25.

10. Lescar, J., et al., *The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH.* Cell, 2001. **105**(1): p. 137-48.

11. Varia, M., et al., *Investigation of a nosocomial outbreak of severe acute respiratory syndrome (SARS) in Toronto, Canada.* CMAJ, 2003. **169**(4): p. 285-92.

12. Lee, N., et al., *A major outbreak of severe acute respiratory syndrome in Hong Kong.* N Engl J Med, 2003. **348**(20): p. 1986-94.

13. Hilgenfeld, R. and M. Peiris, *From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses.* Antiviral Res, 2013. **100**(1): p. 286-95.

14. Cheng, V.C., et al., *Clinical management and infection control of SARS: lessons learned.* Antiviral Res, 2013. **100**(2): p. 407-19.

15. Zwick, M.B., et al., *Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1.* J Virol, 2005. **79**(2): p. 1252-61.

16. Peiris, J.S., et al., *Coronavirus as a possible cause of severe acute respiratory syndrome.* Lancet, 2003. **361**(9366): p. 1319-25.

17. Ksiazek, T.G., et al., *A novel coronavirus associated with severe acute respiratory syndrome.* N Engl J Med, 2003. **348**(20): p. 1953-66.

18.   Drosten, C., et al., *Identification of a novel coronavirus in patients with severe acute respiratory syndrome.* N Engl J Med, 2003. **348**(20): p. 1967-76.

19.   Kuiken, T., et al., *Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome.* Lancet, 2003. **362**(9380): p. 263-70.

20.   Zaki, A.M., et al., *Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia.* N Engl J Med, 2012. **367**(19): p. 1814-20.

21.   Stadler, K., et al., *SARS--beginning to understand a new virus.* Nat Rev Microbiol, 2003. **1**(3): p. 209-18.

22.   Marra, M.A., et al., *The Genome sequence of the SARS-associated coronavirus.* Science, 2003. **300**(5624): p. 1399-404.

23.   Rota, P.A., et al., *Characterization of a novel coronavirus associated with severe acute respiratory syndrome.* Science, 2003. **300**(5624): p. 1394-9.

24.   Wong, S.K., et al., *A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2.* The Journal of biological chemistry, 2004. **279**(5): p. 3197-201.

25.   Li, W., et al., *Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus.* Nature, 2003. **426**(6965): p. 450-4.

26.   Harmer, D., et al., *Quantitative mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme.* FEBS Lett, 2002. **532**(1-2): p. 107-10.

27.   Farcas, G.A., et al., *Fatal Severe Acute Respiratory Syndrome Is Associated with Multiorgan Involvement by Coronavirus.* Journal of Infectious Diseases, 2005. **191**(2): p. 193-197.

28.   Jeffers, S.A., et al., *CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus.* Proc Natl Acad Sci U S A, 2004. **101**(44): p. 15748-53.

29.   Simmons, G., et al., *Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry.* Proc Natl Acad Sci U S A, 2005. **102**(33): p. 11876-11881.

30.   Wang, H., et al., *SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway.* Cell Res, 2008. **18**(2): p. 290-301.

31.   Huang, I.C., et al., *SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells.* The Journal of biological chemistry, 2006. **281**(6): p. 3198-203.

32.   Snijder, E.J., 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**(5): p. 991-1004.

33.   Pfefferle, S., et al., *The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors.* PLoS Pathog, 2011. **7**(10): p. e1002331.

34.   Nal, B., et al., *Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E.* J Gen Virol, 2005. **86**(Pt 5): p. 1423-34.

35.   Krokhin, O., et al., *Mass spectrometric characterization of proteins from the SARS virus: a preliminary report.* Mol Cell Proteomics, 2003. **2**(5): p. 346-56.

36.   Bosch, B.J., W. Bartelink, and P.J.M. Rottier, *Cathepsin L Functionally Cleaves the Severe Acute Respiratory Syndrome Coronavirus Class I Fusion Protein Upstream of Rather than Adjacent to the Fusion Peptide.* Journal of Virology, 2008. **82**(17): p. 8887-8890.

37.   Belouzard, S., V.C. Chu, and G.R. Whittaker, *Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5871-5876.

38.   Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor.* Science, 2005. **309**(5742): p. 1864-8.

39.   Babcock, G.J., et al., *Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor.* J Virol, 2004. **78**(9): p. 4552-60.

40.   Xiao, X., et al., *The SARS-CoV S glycoprotein: expression and functional characterization.* Biochem Biophys Res Commun, 2003. **312**(4): p. 1159-64.

41.   Sainz, B., Jr., et al., *Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein.* J Virol, 2005. **79**(11): p. 7195-206.

42.   Guillen, J., et al., *A second SARS-CoV S2 glycoprotein internal membrane-active peptide. Biophysical characterization and membrane interaction.* Biochemistry, 2008. **47**(31): p. 8214-24.

43.   Guillen, J., et al., *Identification of the membrane-active regions of the severe acute respiratory syndrome coronavirus spike membrane glycoprotein using a 16/18-mer peptide scan: implications for the viral fusion mechanism.* J Virol, 2005. **79**(3): p. 1743-52.

44.   Petit, C.M., et al., *Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion.* Virology, 2005. **341**(2): p. 215-30.

45.   Du, L., et al., *The spike protein of SARS-CoV [mdash] a target for vaccine and therapeutic development.* Nat Rev Micro, 2009. **7**(3): p. 226-236.

46.   Beniac, D.R., et al., *Architecture of the SARS coronavirus prefusion spike.* Nat Struct Mol Biol, 2006. **13**(8): p. 751-2.

47.   Xu, Y., et al., *Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core.* J Biol Chem, 2004. **279**(47): p. 49414-9.

48.   Chernomordik, L.V. and M.M. Kozlov, *Protein-lipid interplay in fusion and fission of biological membranes.* Annu Rev Biochem, 2003. **72**: p. 175-207.

49.   Hofmann, H. and S. Pohlmann, *Cellular entry of the SARS coronavirus.* Trends Microbiol, 2004. **12**(10): p. 466-72.

50.   Hakansson-McReynolds, S., et al., *Solution Structure of the Severe Acute Respiratory Syndrome-Coronavirus Heptad Repeat 2 Domain in the Prefusion State.* Journal of Biological Chemistry, 2006. **281**(17): p. 11965-11971.

51.   HIV/AIDS, J.U.N.P.o., *UNAIDS report on the global AIDS epidemic 2013.* 2013.

52.   Schupbach, J., et al., *Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS.* Science, 1984. **224**(4648): p. 503-5.

53.   Popovic, M., et al., *Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS.* Science, 1984. **224**(4648): p. 497-500.

54.   Sarngadharan, M.G., et al., *Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS.* Science, 1984. **224**(4648): p. 506-8.

55.   Gallo, R.C., et al., *Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS.* Science, 1984. **224**(4648): p. 500-3.

56.   Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.

57.   Clavel, F., et al., *Isolation of a new human retrovirus from West African patients with AIDS.* Science, 1986. **233**(4761): p. 343-6.

58.   Walker, B.D. and X.G. Yu, *Unravelling the mechanisms of durable control of HIV-1.* Nat Rev Immunol, 2013. **13**(7): p. 487-498.

59.   Cohen, M.S., et al., *Acute HIV-1 Infection.* New England Journal of Medicine, 2011. **364**(20): p. 1943-1954.

60.   Moss, J.A., *HIV/AIDS Review.* Radiol Technol, 2013. **84**(3): p. 247-67; quiz p 268-70.

61.   Requejo, H.I., *Worldwide molecular epidemiology of HIV.* Rev Saude Publica, 2006. **40**(2): p. 331-45.

62.   Frankel, A.D. and J.A. Young, *HIV-1: fifteen proteins and an RNA.* Annu Rev Biochem, 1998. **67**: p. 1-25.

63.   Muesing, M.A., et al., *Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus.* Nature, 1985. **313**(6002): p. 450-8.

64.   Klaver, B. and B. Berkhout, *Comparison of 5' and 3' long terminal repeat promoter function in human immunodeficiency virus.* J Virol, 1994. **68**(6): p. 3830-40.

65.   Woodward, C.L., S.N. Cheng, and G.J. Jensen, *Electron Cryotomography Studies of Maturing HIV-1 Particles Reveal the Assembly Pathway of the Viral Core.* Journal of Virology, 2015. **89**(2): p. 1267-1277.

66. Wiegers, K., et al., *Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites.* J Virol, 1998. **72**(4): p. 2846-54.

67. Morikawa, Y., et al., *In vitro processing of human immunodeficiency virus type 1 Gag virus-like particles.* Virology, 2000. **272**(2): p. 366-74.

68. Kaplan, A.H., M. Manchester, and R. Swanstrom, *The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency.* J Virol, 1994. **68**(10): p. 6782-6.

69. von Schwedler, U.K., et al., *Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly.* EMBO J, 1998. **17**(6): p. 1555-68.

70. Briggs, J.A., et al., *Structural organization of authentic, mature HIV-1 virions and cores.* EMBO J, 2003. **22**(7): p. 1707-15.

71. Singh, S.P., et al., *Virion-associated HIV-1 Vpr: variable amount in virus particles derived from cells upon virus infection or proviral DNA transfection.* Virology, 2001. **283**(1): p. 78-83.

72. Welker, R., et al., *Human immunodeficiency virus type 1 Nef protein is incorporated into virus particles and specifically cleaved by the viral proteinase.* Virology, 1996. **219**(1): p. 228-36.

73. Sova, P., et al., *Vif is largely absent from human immunodeficiency virus type 1 mature virions and associates mainly with viral particles containing unprocessed gag.* J Virol, 2001. **75**(12): p. 5504-17.

74. Cohen, E.A., et al., *Human immunodeficiency virus vpr product is a virion-associated regulatory protein.* Journal of Virology, 1990. **64**(6): p. 3097-3099.

75. Dalgleish, A.G., et al., *The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus.* Nature, 1984. **312**(5996): p. 763-7.

76. Klatzmann, D., et al., *T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV.* Nature, 1984. **312**(5996): p. 767-8.

77. Alkhatib, G., et al., *CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1.* Science, 1996. **272**(5270): p. 1955-8.

78. Doranz, B.J., et al., *A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors.* Cell, 1996. **85**(7): p. 1149-58.

79. Deng, H., et al., *Identification of a major co-receptor for primary isolates of HIV-1.* Nature, 1996. **381**(6584): p. 661-6.

80. Choe, H., et al., *The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates.* Cell, 1996. **85**(7): p. 1135-48.

81. Feng, Y., et al., *HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor.* Science, 1996. **272**(5263): p. 872-7.

82.   Moscoso, C.G., et al., *Quaternary structures of HIV Env immunogen exhibit conformational vicissitudes and interface diminution elicited by ligand binding.* Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6091-6.

83.   Kozak, S.L., et al., *CD4, CXCR-4, and CCR-5 dependencies for infections by primary patient and laboratory-adapted isolates of human immunodeficiency virus type 1.* J Virol, 1997. **71**(2): p. 873-82.

84.   Broder, C.C. and R.G. Collman, *Chemokine receptors and HIV.* J Leukoc Biol, 1997. **62**(1): p. 20-9.

85.   Arhel, N.J., et al., *HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore.* EMBO J, 2007. **26**(12): p. 3025-37.

86.   Preston, B.D., B.J. Poiesz, and L.A. Loeb, *Fidelity of HIV-1 reverse transcriptase.* Science, 1988. **242**(4882): p. 1168-71.

87.   Miller, M.D., C.M. Farnet, and F.D. Bushman, *Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition.* J Virol, 1997. **71**(7): p. 5382-90.

88.   Sherman, M.P. and W.C. Greene, *Slipping through the door: HIV entry into the nucleus.* Microbes Infect, 2002. **4**(1): p. 67-73.

89.   Dayton, A.I., et al., *The trans-activator gene of the human T cell lymphotropic virus type III is required for replication.* Cell, 1986. **44**(6): p. 941-7.

90.   Fisher, A.G., et al., *The trans-activator gene of HTLV-III is essential for virus replication.* Nature, 1986. **320**(6060): p. 367-71.

91.   Berkhout, B., R.H. Silverman, and K.T. Jeang, *Tat trans-activates the human immunodeficiency virus through a nascent RNA target.* Cell, 1989. **59**(2): p. 273-82.

92.   Marcello, A., M. Zoppe, and M. Giacca, *Multiple modes of transcriptional regulation by the HIV-1 Tat transactivator.* IUBMB Life, 2001. **51**(3): p. 175-81.

93.   Suhasini, M. and T.R. Reddy, *Cellular proteins and HIV-1 Rev function.* Curr HIV Res, 2009. **7**(1): p. 91-100.

94.   Finzi, A., et al., *Productive human immunodeficiency virus type 1 assembly takes place at the plasma membrane.* J Virol, 2007. **81**(14): p. 7476-90.

95.   Nguyen, D.H. and J.E. Hildreth, *Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts.* J Virol, 2000. **74**(7): p. 3264-72.

96.   Land, A., D. Zonneveld, and I. Braakman, *Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage.* FASEB J, 2003. **17**(9): p. 1058-67.

97.   Lu, M., S.C. Blacklow, and P.S. Kim, *A trimeric structural domain of the HIV-1 transmembrane glycoprotein.* Nat Struct Biol, 1995. **2**(12): p. 1075-82.

98.   McCune, J.M., et al., *Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus.* Cell, 1988. **53**(1): p. 55-67.

99. Checkley, M.A., B.G. Luttge, and E.O. Freed, *HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation.* J Mol Biol, 2011. **410**(4): p. 582-608.

100. Ako-Adjei, D., M.C. Johnson, and V.M. Vogt, *The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles.* J Virol, 2005. **79**(21): p. 13463-72.

101. Tedbury, P.R. and E.O. Freed, *The role of matrix in HIV-1 envelope glycoprotein incorporation.* Trends Microbiol, 2014. **22**(7): p. 372-8.

102. Collins, D.R. and K.L. Collins, *HIV-1 Accessory Proteins Adapt Cellular Adaptors to Facilitate Immune Evasion.* PLoS Pathog, 2014. **10**(1): p. e1003851.

103. Malim, M.H. and M. Emerman, *HIV-1 accessory proteins--ensuring viral survival in a hostile environment.* Cell Host Microbe, 2008. **3**(6): p. 388-98.

104. Schubert, U., et al., *CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway.* J Virol, 1998. **72**(3): p. 2280-8.

105. Margottin, F., et al., *A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif.* Mol Cell, 1998. **1**(4): p. 565-74.

106. Fujita, K., S. Omura, and J. Silver, *Rapid degradation of CD4 in cells expressing human immunodeficiency virus type 1 Env and Vpu is blocked by proteasome inhibitors.* J Gen Virol, 1997. **78 (Pt 3)**: p. 619-25.

107. Neil, S.J., T. Zang, and P.D. Bieniasz, *Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu.* Nature, 2008. **451**(7177): p. 425-30.

108. Sheehy, A.M., et al., *Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein.* Nature, 2002. **418**(6898): p. 646-50.

109. Jager, S., et al., *Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection.* Nature, 2012. **481**(7381): p. 371-5.

110. Yu, X., et al., *Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.* Science, 2003. **302**(5647): p. 1056-60.

111. Malim, M.H., *APOBEC proteins and intrinsic resistance to HIV-1 infection.* Philos Trans R Soc Lond B Biol Sci, 2009. **364**(1517): p. 675-87.

112. Roeth, J.F., et al., *HIV-1 Nef disrupts MHC-I trafficking by recruiting AP-1 to the MHC-I cytoplasmic tail.* J Cell Biol, 2004. **167**(5): p. 903-13.

113. Schwartz, O., et al., *Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein.* Nat Med, 1996. **2**(3): p. 338-42.

114. Collins, K.L., et al., *HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes.* Nature, 1998. **391**(6665): p. 397-401.

115. Ross, T.M., A.E. Oran, and B.R. Cullen, *Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral Nef protein.* Curr Biol, 1999. **9**(12): p. 613-21.

116. Lama, J., A. Mangasarian, and D. Trono, *Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner.* Curr Biol, 1999. **9**(12): p. 622-31.

117. Arts, E.J. and D.J. Hazuda, *HIV-1 Antiretroviral Drug Therapy.* Cold Spring Harbor Perspectives in Medicine, 2012. **2**(4): p. a007161.

118. Mansky, L.M. and H.M. Temin, *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.* J Virol, 1995. **69**(8): p. 5087-94.

119. Coffin, J.M., *HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy.* Science, 1995. **267**(5197): p. 483-9.

120. Haffar, O.K., D.J. Dowbenko, and P.W. Berman, *Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes.* J Cell Biol, 1988. **107**(5): p. 1677-87.

121. Schwartz, S., et al., *Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs.* J Virol, 1990. **64**(11): p. 5448-56.

122. Li, Y., et al., *Control of expression, glycosylation, and secretion of HIV-1 gp120 by homologous and heterologous signal sequences.* Virology, 1994. **204**(1): p. 266-78.

123. Li, Y., et al., *The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding.* Virology, 2000. **272**(2): p. 417-28.

124. Li, Y., et al., *Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9606-11.

125. Earl, P.L., B. Moss, and R.W. Doms, *Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein.* J Virol, 1991. **65**(4): p. 2047-55.

126. Schawaller, M., et al., *Studies with crosslinking reagents on the oligomeric structure of the env glycoprotein of HIV.* Virology, 1989. **172**(1): p. 367-9.

127. Pinter, A., et al., *Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1.* J Virol, 1989. **63**(6): p. 2674-9.

128. Earl, P.L., R.W. Doms, and B. Moss, *Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein.* Proc Natl Acad Sci U S A, 1990. **87**(2): p. 648-52.

129. Hallenberger, S., et al., *Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gpl60.* Nature, 1992. **360**(6402): p. 358-361.

130. Yang, X., et al., *Role of the gp120 inner domain β-sandwich in the interaction between the human immunodeficiency virus envelope glycoprotein subunits.* Virology, 2003. **313**(1): p. 117-125.

131. Helseth, E., et al., *Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein.* Journal of Virology, 1991. **65**(4): p. 2119-2123.

132. Starcich, B.R., et al., *Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS.* Cell, 1986. **45**(5): p. 637-648.

133. Choisy, M., et al., *Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes.* J Virol, 2004. **78**(4): p. 1962-70.

134. Mao, Y., et al., *Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer.* Proceedings of the National Academy of Sciences, 2013. **110**(30): p. 12438-12443.

135. Kwong, P.D., et al., *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.* Nature, 1998. **393**(6686): p. 648-659.

136. Speck, R.F., et al., *Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop.* J Virol, 1997. **71**(9): p. 7136-9.

137. Cocchi, F., et al., *The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockade of infection.* Nat Med, 1996. **2**(11): p. 1244-7.

138. Liu, J., et al., *Molecular architecture of native HIV-1 gp120 trimers.* Nature, 2008. **455**(7209): p. 109-13.

139. Zhu, P., et al., *Distribution and three-dimensional structure of AIDS virus envelope spikes.* Nature, 2006. **441**(7095): p. 847-52.

140. Zanetti, G., et al., *Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ.* PLoS Pathog, 2006. **2**(8): p. e83.

141. Weissenhorn, W., et al., *Atomic structure of the ectodomain from HIV-1 gp41.* Nature, 1997. **387**(6631): p. 426-30.

142. Sainz, B., Jr., et al., *The aromatic domain of the coronavirus class I viral fusion protein induces membrane permeabilization: putative role during viral entry.* Biochemistry, 2005. **44**(3): p. 947-58.

143. Giannecchini, S., et al., *The membrane-proximal tryptophan-rich region in the transmembrane glycoprotein ectodomain of feline immunodeficiency virus is important for cell entry.* Virology, 2004. **320**(1): p. 156-166.

144. Saez-Cirion, A., et al., *Pre-transmembrane sequence of Ebola glycoprotein. Interfacial hydrophobicity distribution and interaction with membranes.* FEBS Lett, 2003. **533**(1-3): p. 47-53.

145. Saez-Cirion, A., et al., *Structural and functional roles of HIV-1 gp41 pretransmembrane sequence segmentation.* Biophys J, 2003. **85**(6): p. 3769-80.

146. Jeetendra, E., et al., *The membrane-proximal region of vesicular stomatitis virus glycoprotein G ectodomain is critical for fusion and virus infectivity.* J Virol, 2003. **77**(23): p. 12807-18.

147. von Heijne, G. and C. Manoil, *Membrane proteins: from sequence to structure.* Protein Eng, 1990. **4**(2): p. 109-12.

148. Landolt-Marticorena, C., et al., *Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins.* J Mol Biol, 1993. **229**(3): p. 602-8.

149. de Kruijff, B., *Cholesterol as a target for toxins.* Biosci Rep, 1990. **10**(2): p. 127-30.

150. Gasset, M., et al., *Influence of cholesterol on gramicidin-induced HII phase formation in phosphatidylcholine model membranes.* Biochim Biophys Acta, 1988. **939**(1): p. 79-88.

151. Li, H. and V. Papadopoulos, *Peripheral-type benzodiazepine receptor function in cholesterol transport. Identification of a putative cholesterol recognition/interaction amino acid sequence and consensus pattern.* Endocrinology, 1998. **139**(12): p. 4991-7.

152. Vincent, N., C. Genin, and E. Malvoisin, *Identification of a conserved domain of the HIV-1 transmembrane protein gp41 which interacts with cholesteryl groups.* Biochim Biophys Acta, 2002. **1567**(1-2): p. 157-64.

153. Simons, K. and E. Ikonen, *Functional rafts in cell membranes.* Nature, 1997. **387**(6633): p. 569-72.

154. Popik, W., T.M. Alce, and W.C. Au, *Human immunodeficiency virus type 1 uses lipid raft-colocalized CD4 and chemokine receptors for productive entry into CD4(+) T cells.* J Virol, 2002. **76**(10): p. 4709-22.

155. Lu, Y., D.X. Liu, and J.P. Tam, *Lipid rafts are involved in SARS-CoV entry into Vero E6 cells.* Biochem Biophys Res Commun, 2008. **369**(2): p. 344-9.

156. Cambi, A., et al., *Microdomains of the C-type lectin DC-SIGN are portals for virus entry into dendritic cells.* J Cell Biol, 2004. **164**(1): p. 145-155.

157. Glende, J., et al., *Importance of cholesterol-rich membrane microdomains in the interaction of the S protein of SARS-coronavirus with the cellular receptor angiotensin-converting enzyme 2.* Virology, 2008. **381**(2): p. 215-21.

158. Lu, Y., et al., *Importance of SARS-CoV spike protein Trp-rich region in viral infectivity.* Biochem Biophys Res Commun, 2008. **371**(3): p. 356-60.

159. Howard, M.W., et al., *Aromatic amino acids in the juxtamembrane domain of severe acute respiratory syndrome coronavirus spike glycoprotein are important for receptor-dependent virus entry and cell-cell fusion.* J Virol, 2008. **82**(6): p. 2883-94.

160. Corver, J., et al., *Mutagenesis of the transmembrane domain of the SARS coronavirus spike glycoprotein: refinement of the requirements for SARS coronavirus cell entry.* Virology journal, 2009. **6**: p. 230.

161. Guillén, J., P.K.J. Kinnunen, and J. Villalaín, *Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2008. **1778**(12): p. 2765-2774.

162. Munoz-Barroso, I., et al., *Role of the membrane-proximal domain in the initial stages of human immunodeficiency virus type 1 envelope glycoprotein-mediated membrane fusion.* J Virol, 1999. **73**(7): p. 6089-92.

163. Salzwedel, K., J.T. West, and E. Hunter, *A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity.* Journal of Virology, 1999. **73**(3): p. 2469-80.

164. Muster, T., et al., *A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1.* J Virol, 1993. **67**(11): p. 6642-7.

165. Suarez, T., et al., *The pre-transmembrane region of the human immunodeficiency virus type-1 glycoprotein: a novel fusogenic sequence.* FEBS Lett, 2000. **477**(1-2): p. 145-9.

166. Suarez, T., et al., *Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fusion.* Journal of Virology, 2000. **74**(17): p. 8038-47.

167. Shnaper, S., et al., *The C- and the N-terminal regions of glycoprotein 41 ectodomain fuse membranes enriched and not enriched with cholesterol, respectively.* J Biol Chem, 2004. **279**(18): p. 18526-34.

168. Saez-Cirion, A., et al., *Sphingomyelin and cholesterol promote HIV-1 gp41 pretransmembrane sequence surface aggregation and membrane restructuring.* J Biol Chem, 2002. **277**(24): p. 21776-85.

169. Lorizate, M., et al., *Membrane-transferring sequences of the HIV-1 Gp41 ectodomain assemble into an immunogenic complex.* J Mol Biol, 2006. **360**(1): p. 45-55.

170. Bellamy-McIntyre, A.K., et al., *Functional links between the fusion peptide-proximal polar segment and membrane-proximal region of human immunodeficiency virus gp41 in distinct phases of membrane fusion.* J Biol Chem, 2007. **282**(32): p. 23104-16.

171. Noah, E., et al., *The membrane proximal external region of the HIV-1 envelope glycoprotein gp41 contributes to the stabilization of the six-helix bundle formed with a matching N' peptide.* Biochemistry, 2008. **47**(26): p. 6782-92.

172. Tulip, P.R., et al., *Conformational plasticity in an HIV-1 antibody epitope.* J Phys Chem B, 2010. **114**(23): p. 7942-50.

173. Coutant, J., et al., *Both lipid environment and pH are critical for determining physiological solution structure of 3-D-conserved epitopes of the HIV-1 gp41-MPER peptide P1.* FASEB J, 2008. **22**(12): p. 4338-51.

174. Sun, Z.Y., et al., *HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane.* Immunity, 2008. **28**(1): p. 52-63.

175. Stiegler, G., et al., *A potent cross-clade neutralizing human monoclonal antibody against a novel epitope on gp41 of human immunodeficiency virus type 1.* AIDS research and human retroviruses, 2001. **17**(18): p. 1757-65.

176. Zwick, M.B., et al., *Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41.* J Virol, 2001. **75**(22): p. 10892-905.

177.   Corti, D. and A. Lanzavecchia, *Broadly neutralizing antiviral antibodies.* Annu Rev Immunol, 2013. **31**: p. 705-42.

178.   Burton, D.R., et al., *Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody.* Science, 1994. **266**(5187): p. 1024-7.

179.   Burton, D.R., et al., *A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10134-7.

180.   Sabin, C., et al., *Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41.* PLoS Pathog, 2010. **6**(11): p. e1001195.

181.   Pietzsch, J., et al., *Human anti-HIV-neutralizing antibodies frequently target a conserved epitope essential for viral fitness.* J Exp Med, 2010. **207**(9): p. 1995-2002.

182.   Simmons, G., et al., *Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4240-5.

183.   Guillen, J., et al., *Interaction of a peptide corresponding to the loop domain of the S2 SARS-CoV virus protein with model membranes.* Molecular membrane biology, 2009. **26**(4): p. 236-48.

184.   Tripet, B., et al., *Structural characterization of the SARS-coronavirus spike S fusion protein core.* J Biol Chem, 2004. **279**(20): p. 20836-49.

185.   He, Y., et al., *Design and Evaluation of Sifuvirtide, a Novel HIV-1 Fusion Inhibitor.* Journal of Biological Chemistry, 2008. **283**(17): p. 11126-11134.

186.   Borrego, P., et al., *An ancestral HIV-2/simian immunodeficiency virus peptide with potent HIV-1 and HIV-2 fusion inhibitor activity.* AIDS, 2013. **27**(7): p. 1081-90.

187.   Wild, C.T., et al., *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, 1994. **91**(21): p. 9770-4.

188.   Kuritzkes, D.R., *HIV-1 entry inhibitors: an overview.* Curr Opin HIV AIDS, 2009. **4**(2): p. 82-7.

189.   Fatkenheuer, G., et al., *Subgroup analyses of maraviroc in previously treated R5 HIV-1 infection.* N Engl J Med, 2008. **359**(14): p. 1442-55.

190.   Gulick, R.M., et al., *Maraviroc for previously treated patients with R5 HIV-1 infection.* N Engl J Med, 2008. **359**(14): p. 1429-41.

191.   Fatkenheuer, G., et al., *Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1.* Nat Med, 2005. **11**(11): p. 1170-2.

192.   Waters, L., et al., *The impact of HIV tropism on decreases in CD4 cell count, clinical progression, and subsequent response to a first antiretroviral therapy regimen.* Clin Infect Dis, 2008. **46**(10): p. 1617-23.

193. Wang, X.J., et al., *Characterisation and evaluation of antiviral recombinant peptides based on the heptad repeat regions of NDV and IBV fusion glycoproteins.* Virology, 2011. **416**(1-2): p. 65-74.

194. Medinas, R.J., D.M. Lambert, and W.A. Tompkins, *C-Terminal gp40 Peptide Analogs Inhibit Feline Immunodeficiency Virus: Cell Fusion and Virus Spread.* Journal of Virology, 2002. **76**(18): p. 9079-9086.

195. Young, J.K., et al., *Interaction of Peptides with Sequences from the Newcastle Disease Virus Fusion Protein Heptad Repeat Regions.* Journal of Virology, 1999. **73**(7): p. 5945-5956.

196. Yao, Q. and R.W. Compans, *Peptides corresponding to the heptad repeat sequence of human parainfluenza virus fusion protein are potent inhibitors of virus infection.* Virology, 1996. **223**(1): p. 103-12.

197. Chan, D.C. and P.S. Kim, *HIV entry and its inhibition.* Cell, 1998. **93**(5): p. 681-4.

198. Naito, T., et al., *SC29EK, a peptide fusion inhibitor with enhanced alpha-helicity, inhibits replication of human immunodeficiency virus type 1 mutants resistant to enfuvirtide.* Antimicrob Agents Chemother, 2009. **53**(3): p. 1013-8.

199. Nishikawa, H., et al., *Electrostatically constrained alpha-helical peptide inhibits replication of HIV-1 resistant to enfuvirtide.* Int J Biochem Cell Biol, 2009. **41**(4): p. 891-9.

200. Dwyer, J.J., et al., *Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide-resistant virus.* Proc Natl Acad Sci U S A, 2007. **104**(31): p. 12772-7.

201. Liu, S., et al., *Different from the HIV fusion inhibitor C34, the anti-HIV drug Fuzeon (T-20) inhibits HIV-1 entry by targeting multiple sites in gp41 and gp120.* J Biol Chem, 2005. **280**(12): p. 11259-73.

202. Kliger, Y., et al., *Mode of action of an antiviral peptide from HIV-1. Inhibition at a post-lipid mixing stage.* J Biol Chem, 2001. **276**(2): p. 1391-7.

203. Naider, F. and J. Anglister, *Peptides in the treatment of AIDS.* Curr Opin Struct Biol, 2009. **19**(4): p. 473-82.

204. Giannecchini, S., et al., *Antiviral activity and conformational features of an octapeptide derived from the membrane-proximal ectodomain of the feline immunodeficiency virus transmembrane glycoprotein.* J Virol, 2003. **77**(6): p. 3724-33.

205. D'Ursi, A.M., et al., *Retroinverso analogue of the antiviral octapeptide C8 inhibits feline immunodeficiency virus in serum.* J Med Chem, 2003. **46**(10): p. 1807-10.

206. Abram, M.E., et al., *Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication.* Journal of Virology, 2010. **84**(19): p. 9864-9878.

207. Bebenek, K., et al., *Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase.* J Biol Chem, 1989. **264**(28): p. 16948-56.

208. Wei, X., et al., *Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy.* Antimicrob Agents Chemother, 2002. **46**(6): p. 1896-905.

209. Greenberg, M.L. and N. Cammack, *Resistance to enfuvirtide, the first HIV fusion inhibitor.* The Journal of antimicrobial chemotherapy, 2004. **54**(2): p. 333-40.

210. Eggink, D., et al., *Resistance of Human Immunodeficiency Virus Type 1 to a Third-Generation Fusion Inhibitor Requires Multiple Mutations in gp41 and Is Accompanied by a Dramatic Loss of gp41 Function.* Journal of Virology, 2011. **85**(20): p. 10785-10797.

211. Esposito, C., et al., *Physicochemical characterization of a peptide deriving from the glycoprotein gp36 of the feline immunodeficiency virus and its lipoylated analogue in micellar systems.* Biochim Biophys Acta, 2006. **1758**(10): p. 1653-61.

212. Zhang, H.Y., et al., *Process development of TRI-999, a fatty-acid-modified HIV fusion inhibitory peptide.* Organic Process Research & Development, 2008. **12**(1): p. 101-110.

213. Ingallinella, P., et al., *Addition of a cholesterol group to an HIV-1 peptide fusion inhibitor dramatically increases its antiviral potency.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5801-6.

214. Stoddart, C.A., et al., *Albumin-conjugated C34 peptide HIV-1 fusion inhibitor: equipotent to C34 and T-20 in vitro with sustained activity in SCID-hu Thy/Liv mice.* J Biol Chem, 2008. **283**(49): p. 34045-52.

215. Welch, B.D., et al., *Potent D-peptide inhibitors of HIV-1 entry.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 16828-33.

216. Berkhout, B., D. Eggink, and R.W. Sanders, *Is there a future for antiviral fusion inhibitors?* Curr Opin Virol, 2012. **2**(1): p. 50-9.

217. Volkmer, R., *Synthesis and application of peptide arrays: quo vadis SPOT technology.* Chembiochem, 2009. **10**(9): p. 1431-42.

218. Volkmer, R., V. Tapia, and C. Landgraf, *Synthetic peptide arrays for investigating protein interaction domains.* FEBS Lett, 2012. **586**(17): p. 2780-6.

219. Andresen, H., et al., *Peptide microarrays with site-specifically immobilized synthetic peptides for antibody diagnostics.* Sensors and Actuators B: Chemical, 2006. **113**(2): p. 655-663.

220. Emili, A.Q. and G. Cagney, *Large-scale functional analysis using peptide or protein arrays.* Nat Biotechnol, 2000. **18**(4): p. 393-7.

221. Katz, C., et al., *Studying protein-protein interactions using peptide arrays.* Chem Soc Rev, 2011. **40**(5): p. 2131-45.

222. Reingewertz, T.H., et al., *Mapping the Vif-A3G interaction using peptide arrays: a basis for anti-HIV lead peptides.* Bioorg Med Chem, 2013. **21**(12): p. 3523-32.

223. Tjernberg, L.O., et al., *Arrest of beta-amyloid fibril formation by a pentapeptide ligand.* J Biol Chem, 1996. **271**(15): p. 8545-8.

224. Fernandez, L., et al., *Design and application of GB virus C (GBV-C) peptide microarrays for diagnosis of GBV-C/HIV-1 co-infection.* Anal Bioanal Chem, 2013. **405**(12): p. 3973-82.

225. Ghosh, J.G. and J.I. Clark, *Insights into the domains required for dimerization and assembly of human alphaB crystallin.* Protein Sci, 2005. **14**(3): p. 684-95.

226. Hepojoki, J., et al., *Interactions and oligomerization of hantavirus glycoproteins.* J Virol, 2010. **84**(1): p. 227-42.

227. Kaukinen, P., A. Vaheri, and A. Plyusnin, *Mapping of the regions involved in homotypic interactions of Tula hantavirus N protein.* J Virol, 2003. **77**(20): p. 10910-6.

228. Rigter, A., et al., *Mapping of possible prion protein self-interaction domains using peptide arrays.* BMC Biochem, 2007. **8**: p. 6.

229. Frank, R. and H. Overwin, *SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes.* Methods Mol Biol, 1996. **66**: p. 149-69.

230. Frank, R., *The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications.* J Immunol Methods, 2002. **267**(1): p. 13-26.

231. Fodor, S.P., et al., *Light-directed, spatially addressable parallel chemical synthesis.* Science, 1991. **251**(4995): p. 767-73.

232. Wenschuh, H., et al., *Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides.* Biopolymers, 2000. **55**(3): p. 188-206.

233. Przezdziak, J., et al., *Probing the ligand-binding specificity and analyzing the folding state of SPOT-synthesized FBP28 WW domain variants.* Chembiochem, 2006. **7**(5): p. 780-8.

234. Hard, T. and C. Lendel, *Inhibition of amyloid formation.* J Mol Biol, 2012. **421**(4-5): p. 441-65.

235. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure.* Science, 1991. **253**(5016): p. 164-70.

236. Sawaya, M.R., et al., *Atomic structures of amyloid cross-beta spines reveal varied steric zippers.* Nature, 2007. **447**(7143): p. 453-7.

237. Thompson, M.J., et al., *The 3D profile method for identifying fibril-forming segments of proteins.* Proc Natl Acad Sci U S A, 2006. **103**(11): p. 4074-8.

238. Nelson, R., et al., *Structure of the cross-beta spine of amyloid-like fibrils.* Nature, 2005. **435**(7043): p. 773-8.

239. Goldschmidt, L., et al., *Identifying the amylome, proteins capable of forming amyloid-like fibrils.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3487-92.

240. Nilsson, M.R., *Techniques to study amyloid fibril formation in vitro.* Methods, 2004. **34**(1): p. 151-60.

241.   Sreerama, N. and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set.* Anal Biochem, 2000. **287**(2): p. 252-60.

242.   Whitmore, L. and B.A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W668-73.

243.   Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases.* Biopolymers, 2008. **89**(5): p. 392-400.

244.   Buck, M., *Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins.* Q Rev Biophys, 1998. **31**(3): p. 297-355.

245.   Kluger, R. and A. Alagic, *Chemical cross-linking and protein–protein interactions—a review with illustrative protocols.* Bioorganic Chemistry, 2004. **32**(6): p. 451-472.

246.   Renart, J., J. Reiser, and G.R. Stark, *Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure.* Proc Natl Acad Sci U S A, 1979. **76**(7): p. 3116-20.

247.   Taylor, S.C., et al., *A defined methodology for reliable quantification of Western blot data.* Molecular biotechnology, 2013. **55**(3): p. 217-26.

248.   Sanders, D.A., *No false start for novel pseudotyped vectors.* Curr Opin Biotechnol, 2002. **13**(5): p. 437-42.

249.   Canki, M., et al., *Highly productive infection with pseudotyped human immunodeficiency virus type 1 (HIV-1) indicates no intracellular restrictions to HIV-1 replication in primary human astrocytes.* J Virol, 2001. **75**(17): p. 7925-33.

250.   Felber, B.K., C.M. Drysdale, and G.N. Pavlakis, *Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein.* J Virol, 1990. **64**(8): p. 3734-41.

251.   Schwartz, S., et al., *Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1.* J Virol, 1990. **64**(6): p. 2519-29.

252.   Jejcic, A., *Altering HIV-1 envelope glycoprotein maturation and its effects on viral infectivity.*, in *Department of laboratory medicine*2011, Karolinska Institutet.

253.   Adachi, A., et al., *Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone.* J Virol, 1986. **59**(2): p. 284-91.

254.   Derdeyn, C.A., et al., *Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120.* Journal of Virology, 2000. **74**(18): p. 8358-67.

255.   Takeuchi, Y., M.O. McClure, and M. Pizzato, *Identification of gammaretroviruses constitutively released from cell lines used for human*

immunodeficiency virus research. Journal of Virology, 2008. **82**(24): p. 12585-8.

256. Platt, E.J., et al., *Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1.* Journal of Virology, 2009. **83**(16): p. 8289-92.

257. Platt, E.J., et al., *Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1.* J Virol, 1998. **72**(4): p. 2855-64.

258. Russell, R.S., C. Liang, and M.A. Wainberg, *Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably?* Retrovirology, 2004. **1**: p. 23.

259. Parolin, C., et al., *Analysis in human immunodeficiency virus type 1 vectors of cis-acting sequences that affect gene transfer into human lymphocytes.* J Virol, 1994. **68**(6): p. 3888-95.

260. McBride, M.S., M.D. Schwartz, and A.T. Panganiban, *Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation.* J Virol, 1997. **71**(6): p. 4544-54.

261. Kaye, J.F., J.H. Richardson, and A.M. Lever, *cis-acting sequences involved in human immunodeficiency virus type 1 RNA packaging.* J Virol, 1995. **69**(10): p. 6588-92.

262. Helga-Maria, C., M.L. Hammarskjold, and D. Rekosh, *An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA.* J Virol, 1999. **73**(5): p. 4127-35.

263. Moore, M.D. and W.S. Hu, *HIV-1 RNA dimerization: It takes two to tango.* AIDS Rev, 2009. **11**(2): p. 91-102.

264. Mullis, K.B. and F.A. Faloona, *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.* Methods Enzymol, 1987. **155**: p. 335-50.

265. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.

266. Reches, M., Y. Porat, and E. Gazit, *Amyloid Fibril Formation by Pentapeptide and Tetrapeptide Fragments of Human Calcitonin.* Journal of Biological Chemistry, 2002. **277**(38): p. 35475-35480.

267. Azriel, R. and E. Gazit, *Analysis of the Minimal Amyloid-forming Fragment of the Islet Amyloid Polypeptide: AN EXPERIMENTAL SUPPORT FOR THE KEY ROLE OF THE PHENYLALANINE RESIDUE IN AMYLOID FORMATION.* Journal of Biological Chemistry, 2001. **276**(36): p. 34156-34161.

268. Balbach, J.J., et al., *Amyloid Fibril Formation by Aβ16-22, a Seven-Residue Fragment of the Alzheimer's β-Amyloid Peptide, and Structural Characterization by Solid State NMR†.* Biochemistry, 2000. **39**(45): p. 13748-13759.

269. Tenidis, K., et al., *Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties.* J Mol Biol, 2000. **295**(4): p. 1055-1071.

270. von Bergen, M., et al., *Assembly of τ protein into Alzheimer paired helical filaments depends on a local sequence motif (306VQIVYK311) forming β structure.* Proceedings of the National Academy of Sciences, 2000. **97**(10): p. 5129-5134.

271. Balbirnie, M., R. Grothe, and D.S. Eisenberg, *An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid.* Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2375-80.

272. Ying, W., et al., *Proteomic analysis on structural proteins of Severe Acute Respiratory Syndrome coronavirus.* Proteomics, 2004. **4**(2): p. 492-504.

273. Beniac, D.R., et al. *Conformational reorganization of the SARS coronavirus spike following receptor binding: implications for membrane fusion.* PLoS One, 2007. **2**, e1082 DOI: 10.1371/journal.pone.0001082.

274. Vishwanathan, S.A. and E. Hunter, *Importance of the membrane-perturbing properties of the membrane-proximal external region of human immunodeficiency virus type 1 gp41 to viral fusion.* Journal of Virology, 2008. **82**(11): p. 5118-26.

275. Eron, J.J., et al., *Short-term safety and antiretroviral activity of T-1249, a second-generation fusion inhibitor of HIV.* The Journal of infectious diseases, 2004. **189**(6): p. 1075-83.

276. Liu, S., et al., *HIV gp41 C-terminal heptad repeat contains multifunctional domains. Relation to mechanisms of action of anti-HIV peptides.* J Biol Chem, 2007. **282**(13): p. 9612-20.

277. Veiga, A.S., et al., *HIV fusion inhibitor peptide T-1249 is able to insert or adsorb to lipidic bilayers. Putative correlation with improved efficiency.* J Am Chem Soc, 2004. **126**(45): p. 14758-63.

278. Epand, R.F., B.G. Sayer, and R.M. Epand, *The tryptophan-rich region of HIV gp41 and the promotion of cholesterol-rich domains.* Biochemistry, 2005. **44**(14): p. 5525-31.

279. Apellániz, B., et al., *Cholesterol-Dependent Membrane Fusion Induced by the gp41 Membrane-Proximal External Region–Transmembrane Domain Connection Suggests a Mechanism for Broad HIV-1 Neutralization.* Journal of Virology, 2014. **88**(22): p. 13367-13377.

280. Celigoy, J., B. Ramirez, and M. Caffrey, *SARS-CoV heptad repeat 2 is a trimer of parallel helices.* Protein Science, 2011. **20**(12): p. 2125-9.

281. Guo, Y., et al., *Identification of a new region of SARS-CoV S protein critical for viral entry.* J Mol Biol, 2009. **394**(4): p. 600-5.

282. Spiga, O., et al., *Molecular modelling of S1 and S2 subunits of SARS coronavirus spike glycoprotein.* Biochem Biophys Res Commun, 2003. **310**(1): p. 78-83.

283. Han, D.P., M. Lohani, and M.W. Cho, *Specific asparagine-linked glycosylation sites are critical for DC-SIGN- and L-SIGN-mediated severe*

acute respiratory syndrome coronavirus entry. Journal of Virology, 2007. **81**(21): p. 12029-39.

284. Yang, Z.Y., et al., *pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN.* J Virol, 2004. **78**(11): p. 5642-50.

285. Liu, J., et al., *Conformational transition between four and five-stranded phenylalanine zippers determined by a local packing interaction.* J Mol Biol, 2006. **361**(1): p. 168-79.

286. Cochran, A.G., N.J. Skelton, and M.A. Starovasnik, *Tryptophan zippers: stable, monomeric beta -hairpins.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5578-83.

287. Liu, J., et al., *Atomic structure of a tryptophan-zipper pentamer.* Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16156-61.

288. Jager, M., et al., *A cross-strand Trp-Trp pair stabilizes the hPin1 WW domain at the expense of function.* Protein Science, 2007. **16**(10): p. 2306-2313.

289. Chan, W.E. and S.S. Chen, *Downregulation of human immunodeficiency virus type 1 Gag expression by a gp41 cytoplasmic domain fusion protein.* Virology, 2006. **348**(2): p. 418-29.

290. Lopez, C.S., et al., *RRE-dependent HIV-1 Env RNA effects on Gag protein expression, assembly and release.* Virology, 2014. **462-463**: p. 126-34.

291. Ellerbrok, H., et al., *Functional tolerance of the human immunodeficiency virus type 1 envelope signal peptide to mutations in the amino-terminal and hydrophobic regions.* J Virol, 1992. **66**(8): p. 5114-8.

292. Martoglio, B., R. Graf, and B. Dobberstein, *Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin.* EMBO J, 1997. **16**(22): p. 6636-45.

293. Marshall, C.B., et al., *Calmodulin and STIM proteins: Two major calcium sensors in the cytoplasm and endoplasmic reticulum.* Biochem Biophys Res Commun, 2015. **460**(1): p. 5-21.

294. Chin, D. and A.R. Means, *Calmodulin: a prototypical calcium sensor.* Trends Cell Biol, 2000. **10**(8): p. 322-8.

295. Hoeflich, K.P. and M. Ikura, *Calmodulin in action: diversity in target recognition and activation mechanisms.* Cell, 2002. **108**(6): p. 739-42.

296. Ishida, H. and H.J. Vogel, *Protein-peptide interaction studies demonstrate the versatility of calmodulin target protein binding.* Protein Pept Lett, 2006. **13**(5): p. 455-65.

297. Osawa, M., et al., *A novel target recognition revealed by calmodulin in complex with Ca2+-calmodulin-dependent kinase kinase.* Nat Struct Biol, 1999. **6**(9): p. 819-24.

298. Vetter, S.W. and E. Leclerc, *Novel aspects of calmodulin target recognition and activation.* Eur J Biochem, 2003. **270**(3): p. 404-14.

299. Yamniuk, A.P. and H.J. Vogel, *Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides.* Molecular biotechnology, 2004. **27**(1): p. 33-57.

300. McQueen, P., et al., *Tat peptide-calmodulin binding studies and bioinformatics of HIV-1 protein-calmodulin interactions.* Proteins, 2011. **79**(7): p. 2233-46.

301. Weeks, K.M. and D.M. Crothers, *RNA recognition by Tat-derived peptides: interaction in the major groove?* Cell, 1991. **66**(3): p. 577-88.

302. Hauber, J., M.H. Malim, and B.R. Cullen, *Mutational analysis of the conserved basic domain of human immunodeficiency virus tat protein.* J Virol, 1989. **63**(3): p. 1181-7.

303. Lindemann, D., et al., *A particle-associated glycoprotein signal peptide essential for virus maturation and infectivity.* J Virol, 2001. **75**(13): p. 5762-71.

304. Wilk, T., et al., *Specific interaction of a novel foamy virus Env leader protein with the N-terminal Gag domain.* J Virol, 2001. **75**(17): p. 7995-8007.

305. Ruggieri, A., et al., *Human endogenous retrovirus HERV-K(HML-2) encodes a stable signal peptide with biological properties distinct from Rec.* Retrovirology, 2009. **6**(1): p. 17.

306. Caporale, M., et al., *The signal peptide of a simple retrovirus envelope functions as a posttranscriptional regulator of viral gene expression.* J Virol, 2009. **83**(9): p. 4591-604.

307. Hoch-Marchaim, H., et al., *The leader peptide of MMTV Env precursor localizes to the nucleoli in MMTV-derived T cell lymphomas and interacts with nucleolar protein B23.* Virology, 2003. **313**(1): p. 22-32.

308. Dultz, E., et al., *The signal peptide of the mouse mammary tumor virus Rem protein is released from the endoplasmic reticulum membrane and accumulates in nucleoli.* J Biol Chem, 2008. **283**(15): p. 9966-76.

309. Mertz, J.A., et al., *Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus.* J Virol, 2005. **79**(23): p. 14737-47.

310. Soto, C., et al., *Inhibition of Alzheimer's amyloidosis by peptides that prevent beta-sheet conformation.* Biochem Biophys Res Commun, 1996. **226**(3): p. 672-80.

311. Wood, S.J., et al., *Prolines and amyloidogenicity in fragments of the Alzheimer's peptide beta/A4.* Biochemistry, 1995. **34**(3): p. 724-30.

312. Reardon, P.N., et al., *Structure of an HIV-1–neutralizing antibody target, the lipid-bound gp41 envelope membrane proximal region trimer.* Proceedings of the National Academy of Sciences, 2014. **111**(4): p. 1391-1396.