Structural and Molecular Basis for Multifunctional Roles of Cellular Proteases

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Yang Yang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Supervisor: Dr. Fang Li

September 2015

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Fang Li, for his continued support throughout my thesis research. His guidance and reliable mentorship inspired me to pursue my academic excellence during my graduate study. I also want to extend my sincere gratitude to all my committee members, Drs. Sundaram Ramakrishnan, W. Gibson Wood and Christopher A. Pennell, who provided me a lot of help and advices for improving my thesis work.

Additionally, I want to thank the scientists from our collaborating labs. Their great work and critical comments were very beneficial in my completion of my thesis work. The collaborators that have contributed to this work include the labs of Dr. Ralph S. Baric (The University of North Carolina at Chapel Hill), Dr. Shibo Jiang and Dr. Lanying Du (New York Blood Center), Dr. Kanagalaghatta R. Rajashankar (Cornell University), and Dr. Zhengli Shi (Wuhan Institute of Virology, Chinese Academy of Sciences). I also want to thank Dr. Jay Nix at Advance Light Source beamline 4.2.2 for assistance in x-ray data collection.

This work was supported in part by funding provided by NIH Grants R01AI089728 and R01AI110700. Computer resources were provided by the Basic Sciences Computing Laboratory of the University of Minnesota Supercomputing Institute.

Abstract

Cellular proteases were first known for their roles in processing peptides such as signal peptides. Recent discoveries have begun to unravel their multifunctional roles in many other biological processes including blood pressure regulation and virus infection. However, the knowledge gathered thus far is still limited regarding the underlying mechanisms for these functions. Therefore, my thesis research focuses on elucidating the structural and molecular basis for the multiple functions of several critical cellular proteases (aminopeptidase A (APA), dipeptidyl peptidase 4 (DPP4), proprotein convertases (PPC) and lysosomal cysteine proteases (LCP)).

APA is well known for its role in blood pressure regulation, but the lack of its structure is a major obstacle for designing APA-targeting agents for the treatment of hypertension, which affects about one third of US adults. My study determined the crystal structures of APA in complex with a variety of substrates and inhibitors, and elaborated the substrate specificity of APA in physiological contexts. The structural information sheds light on the blood pressure regulation by APA, and lays the foundation for developing novel APA inhibitors to treat hypertension.

DPP4 was recently identified as the receptor for the newly emerged Middle East respiratory syndrome coronavirus (MERS-CoV). To provide insights into the origin of MERS-CoV, I investigated the receptor usage and cell entry of MERS-CoV in comparison with HKU4, a related bat coronavirus. I discovered that DPP4 is also the

receptor for HKU4. However, HKU4 failed to infect human cells because of its inability to use human cellular protease (*e.g.* PPC and LCP) for cell entry.

PPC and LCP are critically involved in cell entry of many viruses. Here, I identified two residue differences between the spikes of MERS-CoV and HKU4 that account for their differential processes by human PPC and LCP, respectively. These results are critical for understanding the cross-species transmission of MERS-CoV, as well as for preventing the spread of related bat viruses.

Taken together, the work presented in this thesis contributes to the overall understanding of the underlying mechanisms for diverse functions of cellular proteases in various health problems, and provides rationales for combating related health-threatening diseases.

Table of Contents

List of Tables

List of Figures

Abbreviations

RAS	Renin-angiotensin system
ACE	Angiotensin-converting enzyme
APA	Aminopeptidase A
APN	Aminopeptidase N
*p*NA	*p*-nitroanilide
MIRAS	Multiple isomorphous replacement and anomalous scattering
MERS-CoV	Middle East respiratory syndrome coronavirus
SARS-CoV	Severe acute respiratory syndrome coronavirus
PRCV	Porcine respiratory coronavirus
MHV	Mouse hepatitis coronavirus
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1
ACE2	Angiotensin-converting enzyme 2
DPP4	Dipeptidyl peptidase 4
NTD	N-terminal domain
CTD	C-terminal domain
RBD	Receptor-binding domain
SIRAS	Single isomorphous replacement with anomalous signal
RBM	Receptor-binding motif
FBS	Fetal bovine serum
DMEM	Dulbecco's modified Eagle medium
PBS	Phosphate-buffered saline

PBST	Phosphate-buffered saline with Tween-20
RLU	Relative luciferase units
SDS	Sodium dodecyl sulfate
PAGE	Polyacrylamide gel electrophoresis
DTT	Dithiothreitol
hPPC	Human proprotein convertases
hLCP	Human lysosomal cysteine proteases
r.m.s.d.	Root mean square deviation

Summary of published manuscripts resulting from thesis work

Yang Y*, Liu C*, Du L, Jiang S, Shi Z, Baric RS, Li F (2015) Two mutations were critical for bat-to-human transmission of MERS coronavirus. ***J Virol*** 89: 9119-9123. (***Co-first authors**)

Yang Y*, Du L*, Liu C*, Wang L, Ma C, Tang J, Baric RS, Jiang S, Li F (2014) Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. ***Proc Natl Acad Sci USA*** 111: 12516-12521. (***Co-first authors**)

Ma C, Wang L, Tao X, Zhang N, **Yang Y**, Tseng CT, Li F, Zhou Y, Jiang S, Du L (2014) Searching for an ideal vaccine candidate among different MERS coronavirus receptor-binding fragments-The importance of immunofocusing in subunit vaccine design. ***Vaccine*** 32: 6170- 6176

Du L*, Zhao G*, **Yang Y***, Qiu H, Wang L, Kou Z, Tao X, Yu H, Sun S, Tseng CT, Jiang S, Li F, Zhou Y (2014) A conformation-dependent neutralizing monoclonal antibody specifically targeting receptor-binding domain in Middle East respiratory syndrome coronavirus spike protein. ***J Virol*** 88: 7045-7053. **(*Co-first authors)**

Chen Y*, Rajashankar KR*, **Yang Y***, Agnihothram SS, Liu C, Lin YL, Baric RS, Li F (2013) Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. ***J Virol*** 87: 10777-10783. (***Co-first authors**)

Yang Y, Liu C, Lin YL, Li F (2013) Structural insights into central hypertension regulation by human aminopeptidase A. ***J Biol Chem*** 288: 25638-25645.

Chapter One

Structural Insights into Central Hypertension Regulation by Human Aminopeptidase A

Yang Yang, Chang Liu, Yi-Lun Lin, Fang Li

J Biol Chem 288: 25638-25645 (2013).

1.1 Introduction

Approximately 1 billion people or ~26% of the adult population of the world are affected by hypertension[1]. Hypertension cannot be cured, but antihypertensive treatment can reduce cardiovascular morbidity and mortality. Despite the availability of many antihypertensive drugs, hypertension remains difficult to control. ~20% of the hypertensive population are "resistant hypertension" patients, who do not respond well to three or more different antihypertensive drugs used at the same time[2]. This proportion is even greater when patients who do not respond well to two antihypertensive drugs used concurrently are included. Additionally, current antihypertensive drugs have lower efficacy in patients of African ancestry[3,4] or patients with renal impairment or diabetes[5,6]. Consequently, the demand to develop new antihypertensive drugs with different mechanisms of action remains high.

Blood pressure is regulated through both the central and systemic renin-angiotensin systems (RASs)[7-9]. Some hypertensive patients show hyperactivity in the central RAS but low activity in the systemic RAS. Hence, these patients are resistant to treatments targeting the systemic RAS[10]. The same set of peptidases play essential roles in both the central and systemic RASs (Fig. 1.1)[11,12]. Renin (also known as angiotensinogenase) cleaves the N-terminal 12 residues from the protein angiotensinogen to produce the peptide angiotensin I. Angiotensin-converting enzyme (ACE) then cleaves the C-terminal two residues from angiotensin I to produce angiotensin II. Aminopeptidase A (APA) cleaves the N-terminal aspartate from angiotensin II to produce angiotensin III. Finally,

aminopeptidase N (APN) cleaves the N-terminal arginine from angiotensin III to produce angiotensin IV. Interestingly, angiotensin peptides play different roles in the central and systemic RASs: in the systemic RAS, angiotensin II is the main effector peptide that increases blood pressure, whereas in the central RAS, angiotensin III plays this role[13-15]. Renin and ACE in the systemic RAS are among the main targets for antihypertensive drugs[11]. Recently, APA in the central RAS has been suggested as a new target for antihypertensive drugs[16-21]. In several types of hypertension animal models, brain-targeting APA inhibitors, either directly injected into the brain or administered orally followed by brain distribution, have been shown to be effective in lowering blood pressure[16-19]. However, the design and development of this new class of antihypertensive drugs have been hindered by the lack of an atomic structure of APA. Like the crystal structures of renin and ACE[22,23], the crystal structure of APA can inform therapeutic discoveries and aid in the control and treatment of hypertension.

Mammalian APA is a member of the M1 family of zinc metalloenzymes and is widely expressed in many tissues, including the brain[20,24]. As a cell surface-anchored ectoenzyme, it preferentially cleaves acidic residues, either glutamate or aspartate, from the N-terminus of peptide substrates[25]. The catalytic mechanism of APA is believed to be similar to that of another zinc-dependent aminopeptidase in the same family, APN, whose structure our lab reported recently[26]. However, APA is unique among M1 family zinc metalloenzymes for favoring cleavage of acidic residues. Intriguingly, the substrate specificity of APA is modulated by calcium, a unique feature among all families of zinc

metalloenzymes. More specifically, calcium enhances the enzymatic activity of APA for cleavage of acidic residues while reducing its activity with basic residues[27-31]. The calcium-modulated substrate specificity of APA is deemed physiologically relevant because the concentrations of calcium that modulate APA activity are in the same range as those found in brain fluid (*i.e.* 1–2 mM)[32]. Mutagenesis studies have identified several residues in APA that may be involved in calcium modulation[27-31]. However, the structural mechanism for calcium-modulated substrate specificity of APA has remained a puzzle due to the lack of an atomic structure of APA. Understanding this mechanism will not only enrich our knowledge about the interesting APA enzymology but will also provide insights into central hypertension regulation by APA.

Here, I have determined the crystal structures of the human APA ectodomain by itself and in complex with amino acids or peptidomimetic inhibitors. These structures illustrate detailed interactions between APA and its ligands. I have also identified a calcium-binding site in APA and elucidated the structural basis for calcium-modulated APA activity. Additionally, this study reveals the structural basis for the different APA-inhibiting potencies of peptidomimetic inhibitors. Taken together, these results provide an understanding of the substrate specificity and calcium modulation of APA in central hypertension regulation and will guide the development of a new class of brain-targeting APA inhibitors to treat hypertension.

1.2 Materials and Methods

Reagents and Constructs

The synthetic substrates glutamyl-*p*-nitroanilide (*p*NA), aspartyl-*p*NA, arginine-*p*NA, phenylalanine-*p*NA, and leucine-*p*NA and the peptidomimetic inhibitors bestatin and amastatin were from Sigma-Aldrich. pFastBac (Invitrogen) was described previously[26].

Protein Preparation and Crystallization

The human APA ectodomain (residues 76–957; GenBank accession number AAA35522.1) was expressed and purified as described previously for porcine APN[26]. Briefly, the human APA ectodomain, containing an N-terminal honeybee melittin signal peptide and a C-terminal His_6 tag, was expressed in insect cells using the Bac-to-Bac expression system (Invitrogen), secreted into cell culture medium, and subsequently purified on a HiTrap Chelating HP affinity column (GE Healthcare) and a Superdex 200 gel filtration column (GE Healthcare). APA was concentrated to 10 mg/ml and stored in buffer containing 20 mM Tris (pH 7.4) and 200 mM NaCl. Crystallization of APA was set up using the sitting drop vapor diffusion method at 4 °C, with 2 µl of protein solution added to 2 µl of well buffer containing 1.6 M sodium citrate (pH 7.0). After 4 weeks, APA crystals were harvested in buffer containing 10% (v/v) glycerol and 1.6 M sodium citrate (pH 7.0) and flash-frozen in liquid nitrogen.

Heavy atom derivatives of APA crystals were prepared by soaking crystals overnight in buffer containing 1 mM heavy atom ($HgCl_2$ or K_2PtBr_6), 10% (v/v) glycerol, and 1.6 M

sodium citrate (pH 7.0). Crystals of APA-ligand complexes were prepared by soaking crystals overnight in buffer containing 8 mM ligand (amino acids or peptidomimetic inhibitors), 10% (v/v) glycerol, and 1.6 M sodium citrate (pH 7.0). Crystals of APA-ligand complexes in the presence of calcium were prepared by soaking crystals overnight in buffer containing 8 mM ligand, 20 mM $CaCl_2$, 10% (v/v) glycerol, and 1.6 M sodium citrate (pH 7.0).

Data Collection and Structure Determination

Data were collected at Advanced Light Source beamline 4.2.2 at a wavelength of 1.0003 Å and at 100 K for native crystals. The crystal structure was determined with a mercury derivative and a platinum derivative by the multiple isomorphous replacement and anomalous scattering (MIRAS) method, with data collection conducted at wavelengths of 1.0003 and 1.06883 Å. X-ray diffraction data were processed using HKL2000[33]. The program SOLVE[34] was used to find heavy atom sites and for MIRAS phasing. From a mercury-derivatized crystal, 14 mercury sites were identified. From a platinum-derivatized crystal, 12 platinum sites were identified. The program RESOLVE was used for solvent flattening and automated model building[35]. The model was improved by iterative cycles of manual building in Coot [36] and refinement with CNS[37]and Refmac[38]. The similes files of ligands were loaded onto CCP4 Module Sketcher to generate CIF (crystallographic information file) files. The generated CIF files were used in CCP4 refinement of APA-ligand complex structures. The CIF files were also converted to topology and parameter files in CNS using module cns_import_cif. The generated

topology and parameter files were used in CNS refinement of APA-ligand complex structures. Models of ligands were built based on $F_o - F_c$ omit maps calculated in the absence of the ligands. For the APA native model, 97% of residues are in the favored regions of the Ramachandran plot, and 0.23% of residues are in the disallowed regions.

Catalysis and Inhibition Assays

APA catalytic activities were determined with 10 nM APA and 1 mM aminoacyl-*p*NA (amino acid-*p*NA) in 100 µl of 25 mM Tris-HCl buffer at pH 7.0 in the presence or absence of 2 mM $CaCl_2$. The reactions were incubated at 37 °C for 30 min. Formation of product *p*NA was measured using a spectrophotometer and monitored at 405 nm.

APA inhibition assays were carried out using a concentration gradient of the peptidomimetic inhibitors. Formation of product *p*NA was measured every 10 min. The IC_{50} is defined as the concentration of each inhibitor that led to 50% of maximal APA turnover of Glu-*p*NA. The K_i values for the inhibition assay were determined from the IC_{50} using the Cheng-Prusoff equation: $K_i = IC_{50}/(1+[S]/K_m)$

1.3 Results

Structural Determinations of APA and APA-Ligand Complexes

Mammalian APA is a class II membrane protein with a short intracellular N-terminal tail, a single-pass transmembrane anchor, a small stalk, and a large C-terminal ectodomain[24]. The human APA ectodomain (residues 76–957) was expressed, purified, and crystallized in space group *P*6422 (a = 142.3 Å, b = 142.3 Å, and c = 237.3 Å), with one

molecule/asymmetric unit. The structure was determined by MIRAS using one mercury derivative and one platinum derivative. The final structural model was refined at 2.15 Å resolution (Fig. 1.2A and B). It contains residues 76–954 of the APA ectodomain (except for disordered loop 608–611) and 11 N-linked glycans. APA-ligand complexes were prepared by soaking each of the ligands (amino acids or peptidomimetic inhibitors) into APA crystals in the presence or absence of calcium. Their structures were determined by molecular replacement using the unliganded APA structure as the search template and were subsequently refined. Data and model statistics are listed in Table 1.1.

Overall Architecture

The human APA ectodomain has the same domain structures and arrangement as other M1 family zinc metalloenzymes (Figs. 1.2A and C, 1.3). It consists of four distinct domains: head (residues 633–954), side (residues 543–632), body (residues 294–542), and tail (residues 76–293) (Fig. 1.2A). The zinc-binding active site is located in a cavity created by the head, side, and body domains (Fig. 1.2B). The active site cavity of M1 family zinc metalloenzymes can alternate between an open conformation for substrate loading or product unloading and a closed conformation for catalysis. These conformational changes are through a hinge-opening movement between the head domain and the other domains, accompanied by positional changes of catalytic residues[26]. All of the catalytic residues in the current APA structure are in their catalytically active positions (Fig. 1.4), suggesting that the APA structure captured in this study is in a closed conformation. However, the active site cavity of APA in the closed conformation is

significantly narrower than that of mammalian APN in the closed conformation (Fig. 1.3A), suggesting that APA excludes large peptide substrates from accessing its active site and hence has a more restricted range of substrates compared with APN.

The APA ectodomain forms a novel homodimer. In the crystal, there is only one molecule of the APA ectodomain in each asymmetric unit. However, two molecules from adjacent asymmetric units form the homodimer through their head domains (Fig. 1.2C), burying ~1600 $Å^2$ of accessible surface between them. These structural data are consistent with the previous observation that full-length mammalian APA forms a homodimer on cell surfaces[25]. Interestingly, although APA and APN both form a homodimer through their head domains, different regions in their head domains are involved in dimer formation, leading to completely different dimer interfaces (Figs. 1.2C and 1.3B). Dimer formation can increase the stability of APA and APN, both of which are ectoenzymes and face harsher environments than monomeric intracellular zinc metalloenzymes in the same family. Furthermore, the head-to-head dimer arrangements in both APA and APN can facilitate the closed-to-open conformational changes, during which the dimer interfaces are maintained and the membrane anchors move along the cell membrane. The different dimer arrangements of APA and APN suggest that a conserved dimer interface is not a prerequisite for the above functional benefits associated with dimer formation.

Substrate Specificity

To investigate the substrate specificity of APA, I determined the crystal structures of

APA complexed with amino acids or peptidomimetic inhibitors (Figs. 1.5A and B, 1.6A and B). APA and APN form the same interactions with the main chain of the N-terminal residue (P1 residue) of their ligands (Fig. 1.4), suggesting that the two enzymes share a common catalytic mechanism despite their modest sequence similarity. However, the S1 pocket that accommodates the P1 side chain of ligands differs markedly between APA and APN, consistent with the different substrate specificities of the two enzymes. In general, the S1 pocket of APN is largely hydrophobic and prefers to accommodate hydrophobic amino acid side chains. In contrast, the S1 pocket of APA is polar and prefers to accommodate polar P1 side chains. In this study, I focused on the interactions between the S1 pocket of APA and the P1 side chains of ligands that are major determinants of the substrate specificity of APA.

The S1 pocket of APA is well suited to accommodate the side chains of acidic residues. The carboxylate side chain of bound glutamate forms a strong salt bridge with Arg-887 and a hydrogen bond with Thr-356 in the S1 pocket (Fig. 1.5A). These energetically favorable interactions are consistent with high APA activity on substrates with a P1 glutamate (Fig. 1.5C). Mutation R887A causes a dramatic decrease in APA activity on substrates with a P1 glutamate through removal of the salt bridge (Fig. 1.5D). Mutation T356V also decreases APA activity on substrates with a P1 glutamate through removal of the hydrogen bond. In the absence of calcium, aspartate could not be soaked into the APA crystals, consistent with the relatively low APA activity on substrates with a P1 aspartate in the absence of calcium. Presumably, the relatively short side chain of

aspartate would form a weak salt bridge with Arg-887, and thus, aspartate is less favored than glutamate by APA at the P1 substrate site.

The S1 pocket also accommodates the side chains of basic residues (Fig. 1.5C). The guanidinium group of bound arginine forms a salt bridge with the side chain of Asp-221, a hydrogen bond with the main chain carbonyl of Glu-223, and another hydrogen bond with the side chain of Thr-356 (Fig. 1.5B). Mutation T356V decreases APA activity on substrates with a P1 arginine through removal of the hydrogen bond with the Thr-356 side chain. Interestingly, compared with bound glutamate, the side chain of bound arginine points in a different direction, which reduces, but does not eliminate, charge repulsion with Arg-887 in the S1 pocket. Mutation R887A increases APA activity on substrates with a P1 arginine through elimination of the charge repulsion (Fig. 1.5D). These results are consistent with the observation that in the absence of calcium, arginine is less favored than glutamate, but more favored than aspartate, by APA at the P1 substrate site (Fig. 1.5C).

The S1 pocket discriminates against the side chains of neutral residues, as evidenced by low APA activity on substrates with a P1 leucine or phenylalanine (Fig. 1.5C). Accordingly, the amino acids leucine and phenylalanine were not observed bound to APA when soaked into the APA crystals. Instead, two peptidomimetic inhibitors, amastatin and bestatin, which contain a P1 leucine and phenylalanine, respectively, were successfully soaked into the APA crystals (Fig. 1.6A and B). The side chains of leucine

and phenylalanine both form hydrophobic interactions with the side chains of Phe-474 in the S1 pocket but also have unfavorable steric interactions with the polar residues in the S1 pocket (Fig. 1.6C and D). Steric hindrance from the P1 phenylalanine of bestatin even forces Arg-887 to adopt a different conformation (Fig. 1.6D). Therefore, residues with large neutral side chains are disfavored by APA at the P1 substrate site.

Calcium Modulation

To investigate the calcium-modulated substrate specificity of APA, I soaked calcium into the crystals of APA complexed with glutamate. The $F_o - F_c$ map showed clear additional electron density in the S1 pocket of APA, which we interpreted to be a calcium ion and two calcium-coordinating water molecules (Fig. 1.7A). Our interpretation was based on the following observations. First, in the absence of calcium, the proposed calcium-binding site is occupied by a water molecule (Fig. 1.7B). In the presence of calcium, the structural model was first refined with the water molecule occupying the proposed calcium-binding site, and the $F_o - F_c$ map was then calculated. The $F_o - F_c$ map contained significant positive electron density at this site, indicating a species more electron-rich than water. Second, in the absence of calcium, the water molecule occupying the proposed calcium-binding site is four-coordinate, forming hydrogen bonds with another water molecule, the Asp-221 side chain, the Glu-223 main chain carbonyl, and the bound glutamate (which is likely protonated due to the strong bifurcated salt bridge with Arg-887) (Fig. 1.7B). In the presence of calcium, two additional water molecules are observed, increasing the coordination number to six, consistent with

calcium replacing the water (Fig. 1.7A). Third, the geometry of the coordinated calcium is octahedral, consistent with the geometry for calcium coordination in other calcium-binding proteins[39,40]. Finally, previous biochemical data show that mutation of Asp-221 to other residues (*e.g.* alanine or asparagine) abolishes calcium modulation of APA activity[30], which is consistent with our structural data. Therefore, I conclude that the calcium-binding site is located in the S1 pocket of APA adjacent to the P1 side chain of its ligands.

As discussed above, in the absence of calcium, aspartate could not be observed in soaked APA crystals due to the relatively low binding affinity of aspartate. In contrast, in the presence of calcium, aspartate was observed following soaking into the APA crystals, consistent with the observation that calcium enhances the APA activity on substrates with a P1 aspartate (Fig. 1.5C). Using the same strategies that we used to identify the calcium-binding site in the glutamate-bound APA structure, I again identified the same calcium-binding site in the aspartate-bound APA structure (Fig. 1.7C).

The identified calcium-binding site provides a structural basis for the calcium-modulated APA activity. Calcium enhances the APA-binding affinities of glutamate and aspartate by ligating their carboxylate side chains (Fig. 1.7A and C). In contrast, arginine could not be observed following soaking into calcium-bound APA crystals. The positively charged side chain of arginine cannot bind effectively due to charge repulsion with calcium, although it could be observed in soaked APA crystals in the absence of calcium (Figs.

1.5B and 1.7D). Therefore, the calcium ion that binds to the S1 pocket of APA enhances the binding of acidic residues at the P1 substrate site through coordination of their side chains but reduces the binding of basic residues at the P1 substrate site due to charge repulsion with their positively charged side chains.

APA Inhibitors

Two peptidomimetic inhibitors, amastatin and bestatin, are universal inhibitors of M1 family zinc metalloenzymes. Here, I show that amastatin is a much more potent APA inhibitor than bestatin (Fig. 1.6F). By determining the crystal structures of APA complexed with each inhibitor, I elucidated the structural details of APA inhibition by these two inhibitors. First, as discussed above, the hydrophobic P1 side chains of both of the inhibitors have unfavorable steric interactions with the polar S1 pocket of APA. Due to its bulkier side chain, the P1 phenylalanine of bestatin has more severe steric clashes with the polar S1 pocket, forcing Arg-887 in the S1 pocket to adopt a different conformation (Fig. 1.6D). Second, amastatin contains four residues, whereas bestatin contains only two. The side chain of the C-terminal residue of amastatin (*i.e.* P3′ position) fits nicely into the S3′ subsite of APA, forming a strong bifurcated salt bridge with Arg-386 and a hydrogen bond with Asn-371 (Fig. 1.6E). In contrast, the shorter bestatin does not interact with the S3′ subsite of APA. These structural differences between the two inhibitors account for their different APA-inhibiting potencies.

My study also has implications for the mechanisms of action of APA inhibitors that were

developed prior to the structure determination of APA. For example, RB150 is a prodrug for the potent APA inhibitor EC33[18]. RB150 is orally bioavailable, can cross the blood-brain barrier, and is subsequently converted to EC33 in the brain (Fig. 1.8). EC33 has been shown to be an effective antihypertensive compound in animal models[18]. Interestingly, EC33 and glutamate share common structural features by containing a free N-terminal amine group and a negatively charged side chain of the same length. The free N-terminal amine group of EC33 likely binds to the APA active site in the same way as those of peptide substrates and analogs (Fig. 1.8). Although the side chains of EC33 and glutamate contain a sulfonic acid group and a carboxylate group, respectively, the former likely binds APA in the same way as the latter by forming salt bridges with Arg-887 and calcium in the S1 pocket. In addition, EC33 also contains a thiol group on the opposite end of the molecule to the amine group, in contrast to the carboxylate group of the C terminus of glutamate. It was hypothesized that the thiol group of EC33 directly interacts with the active site zinc[18], but this hypothesis will need to be tested by crystallographic studies of APA complexed with EC33. Nevertheless, EC33 likely functions as a substrate analog and competitively binds to the APA active site.

1.4 Discussion

Hyperactivity in the central RAS has been implicated in the development and maintenance of hypertension[7-9]. In particular, a high-level central RAS, accompanied by a low-level systemic RAS, has been described in some hypertensive patients who are resistant to treatments targeting the systemic RAS[10]. The major effector peptide for blood

pressure regulation in the brain is angiotensin III[13-15]. Angiotensin III is produced by the action of APA on angiotensin II and is converted to angiotensin IV by APN (Fig. 1.1)[11,12]. To achieve central hypertension regulation, it is critical that the two different enzymes control these two angiotensin-converting steps. Thus, the substrate specificity of APA is essential for central hypertension regulation. How APA specifically cleaves only the N-terminal aspartate from angiotensin II but does not additionally cleave the subsequent arginine following the formation of angiotensin III has been a mystery. More intriguingly, the substrate specificity of APA is modulated by calcium at physiologically relevant concentrations. This study has elucidated the structural mechanisms for calcium-modulated high APA activity on angiotensin II and low APA activity on angiotensin III, providing structural insights into central hypertension regulation by human APA.

In this study, I determined the crystal structure of human APA by itself and in complex with a variety of ligands in the absence and presence of calcium. In the absence of calcium, the polar S1 pocket of APA forms salt bridges and hydrogen bonds with the charged side chains of both acidic and basic residues at the N-terminus of ligands, but it has unfavorable steric interactions with the hydrophobic side chains of neutral residues. Thus, in the absence of calcium, the S1 pocket of APA has evolved to accommodate both acidic and basic residues, but not neutral residues, at the N-terminus of ligands. In fact, in the absence of calcium, APA has higher activity on substrates with an N-terminal arginine than on those with an N-terminal aspartate. I further identified a calcium-binding site in the S1 pocket. When bound to APA, calcium is coordinated by the carboxylate

side chain of glutamate or aspartate but leads to charge repulsion with the guanidinium group of arginine. Thus, APA-bound calcium enhances the binding of glutamate and aspartate to APA and reduces the binding of arginine to APA. Consequently, in the presence of calcium, APA has much higher activity on substrates with an N-terminal aspartate (*e.g.* angiotensin II) than on those with an N-terminal arginine (*e.g.* angiotensin III). These unique and elegant structural mechanisms ensure that under physiological conditions, APA cleaves only angiotensin II, but not angiotensin III, providing regulation to central hypertension.

This study also provides a structural platform for the design and development of novel APA inhibitors to treat hypertension. In several types of experimental and genetic hypertension animal models, orally active and brain-targeting APA inhibitors can modulate central RAS activity and reduce blood pressure without changing systemic RAS activity[16-19]. Furthermore, these APA inhibitors do not alter blood pressure in normotensive animals. The crystal structures of renin and ACE, two other peptidases in the RAS, have been successfully used in guiding rational drug design. The crystal structures of APA and its complexes with ligands can also be used in the rational design and development of novel antihypertensive drugs. More specifically, on the basis of the structural information of APA, we can design novel APA inhibitors with functional groups that explore and fit into different subsites in APA. In addition, using the structural framework of APA, we can also optimize and further develop existing APA inhibitors to enhance their potencies. To conclude, the structural information contained in this study

regarding APA holds promise for the design and development of a new class of antihypertensive drugs.

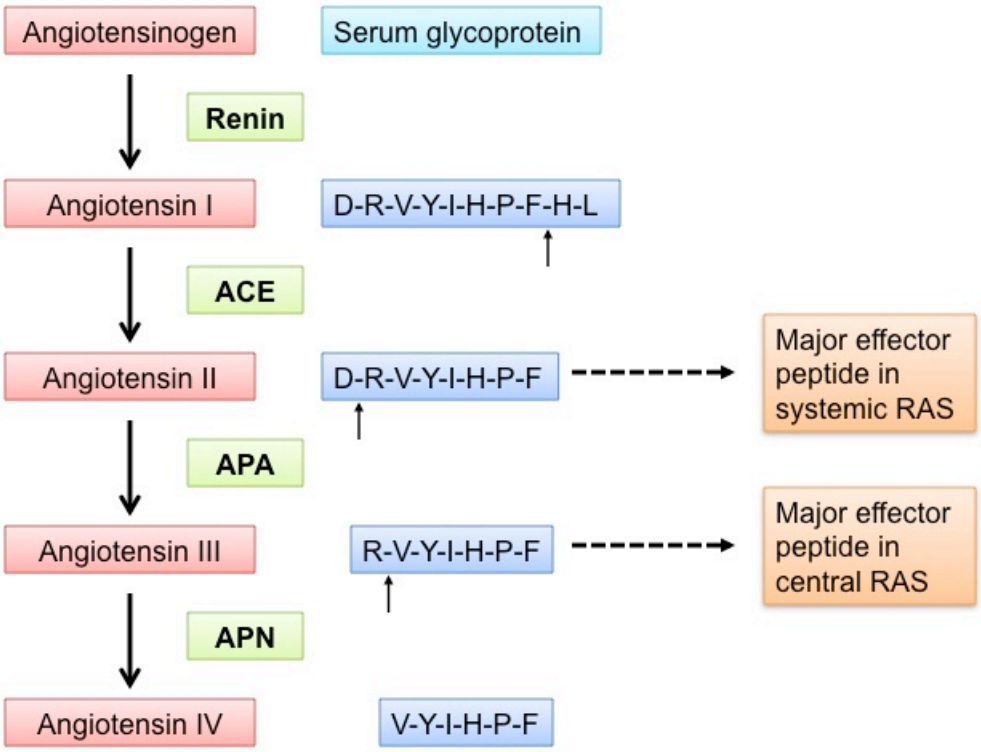

Figure 1.1 Central and systemic RASs that together regulate blood pressure.
In both the central and systemic RASs, renin cleaves angiotensinogen to produce angiotensin I, ACE converts angiotensin I to angiotensin II, APA converts angiotensin II to angiotensin III, and APN converts angiotensin III to angiotensin IV. Angiotensin II is the major effector peptide in the systemic RAS, whereas angiotensin III is the major effector peptide in the central RAS.

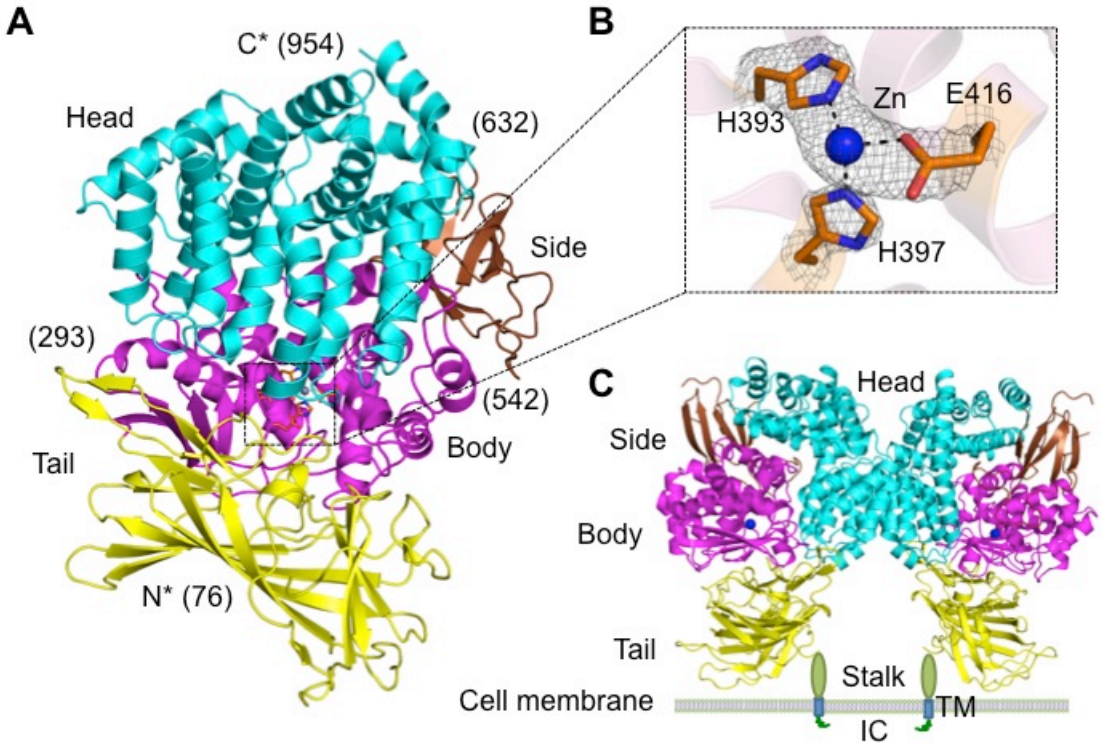

Figure 1.2 Overall structure of the human APA ectodomain.
A, the human APA ectodomain contains four domains: head (cyan), side (brown), body (magenta), and tail (yellow). Residue numbers indicating domain boundaries are in parentheses. B, $2F_o - F_c$ electron density map of the zinc-binding site (contoured at 1.5σ). C, structure of the dimeric APA ectodomain and model of dimeric full-length APA on the cell surface. Zinc is shown as a blue sphere. IC, intracellular tail; TM, transmembrane anchor

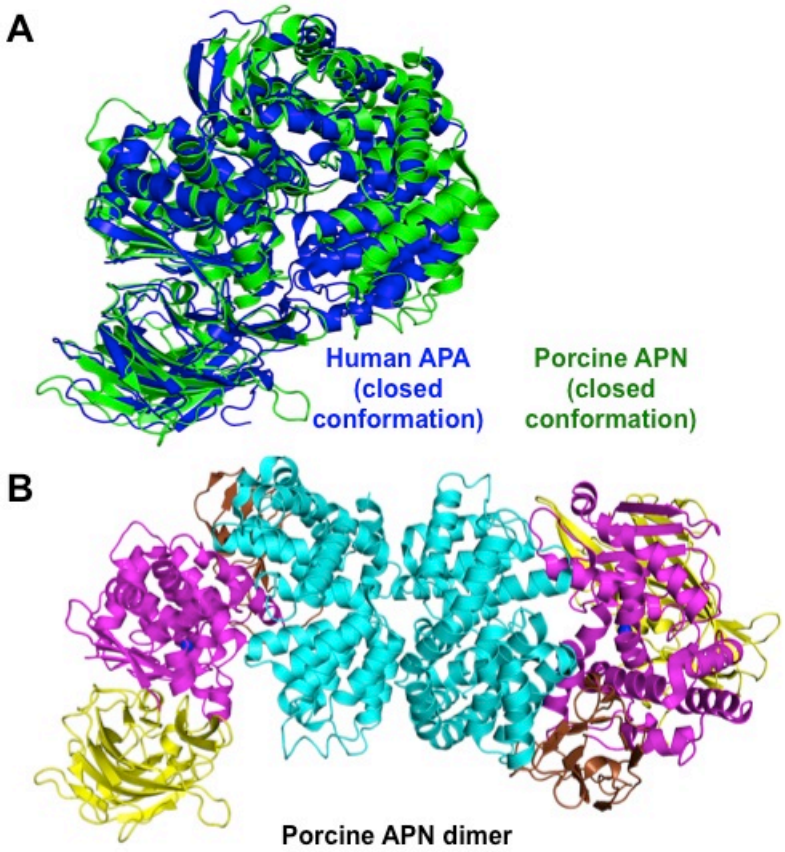

Figure 1.3 Structural comparisons of the conformation and dimer formation of porcine APN and human APA.
A, structural alignment of APA (blue) and APN (green; Protein Data Bank code 4FKE), both of which are in the catalytically active and closed conformation. B, structure of the dimeric porcine APN ectodomain.

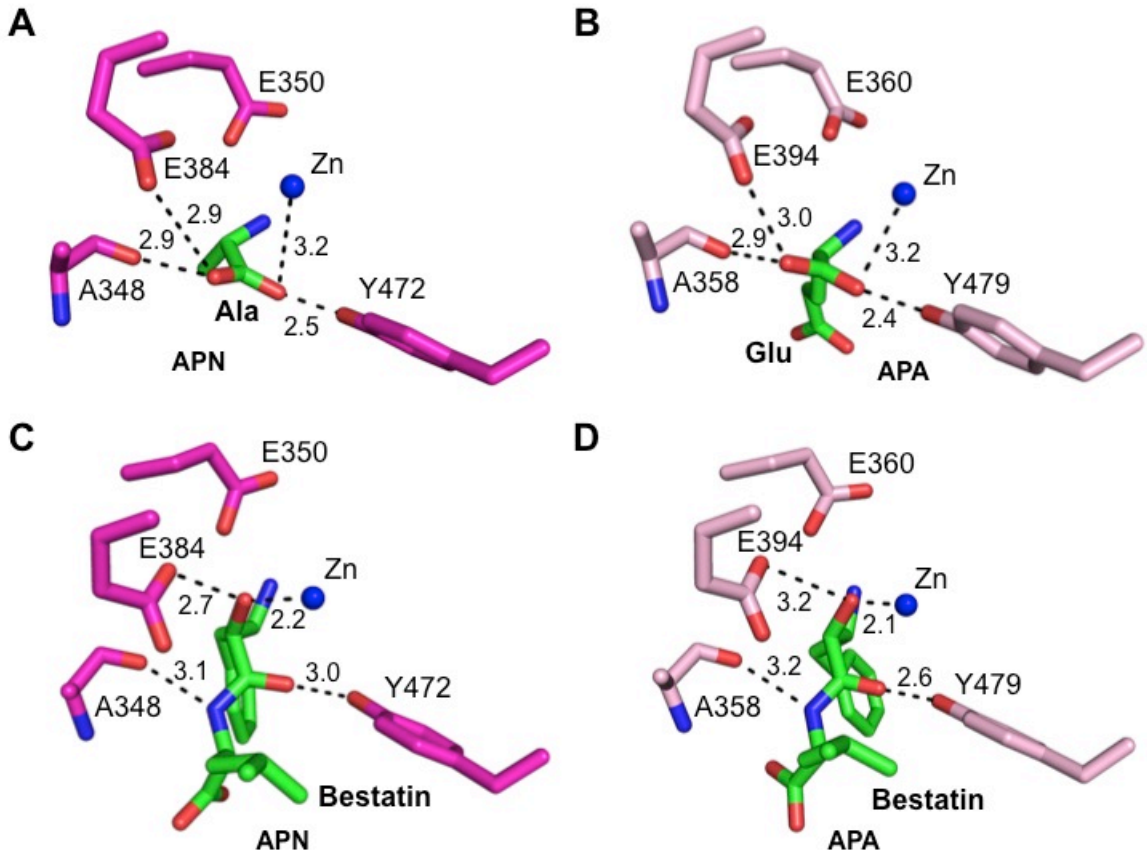

Figure 1.4 Structural comparisons of the active sites of porcine APN and human APA.

A, interaction between the catalytic residues of APN (magenta) and the main chain of bound alanine (green; Protein Data Bank code 4FKH). B, interaction between the catalytic residues of APA (pink) and the main chain of bound glutamate (green). C, interaction between the catalytic residues of APN and the main chain of bound bestatin (green; Protein Data Bank code 4FKK). D, interaction between the catalytic residues of APA and the main chain of bound bestatin (green).

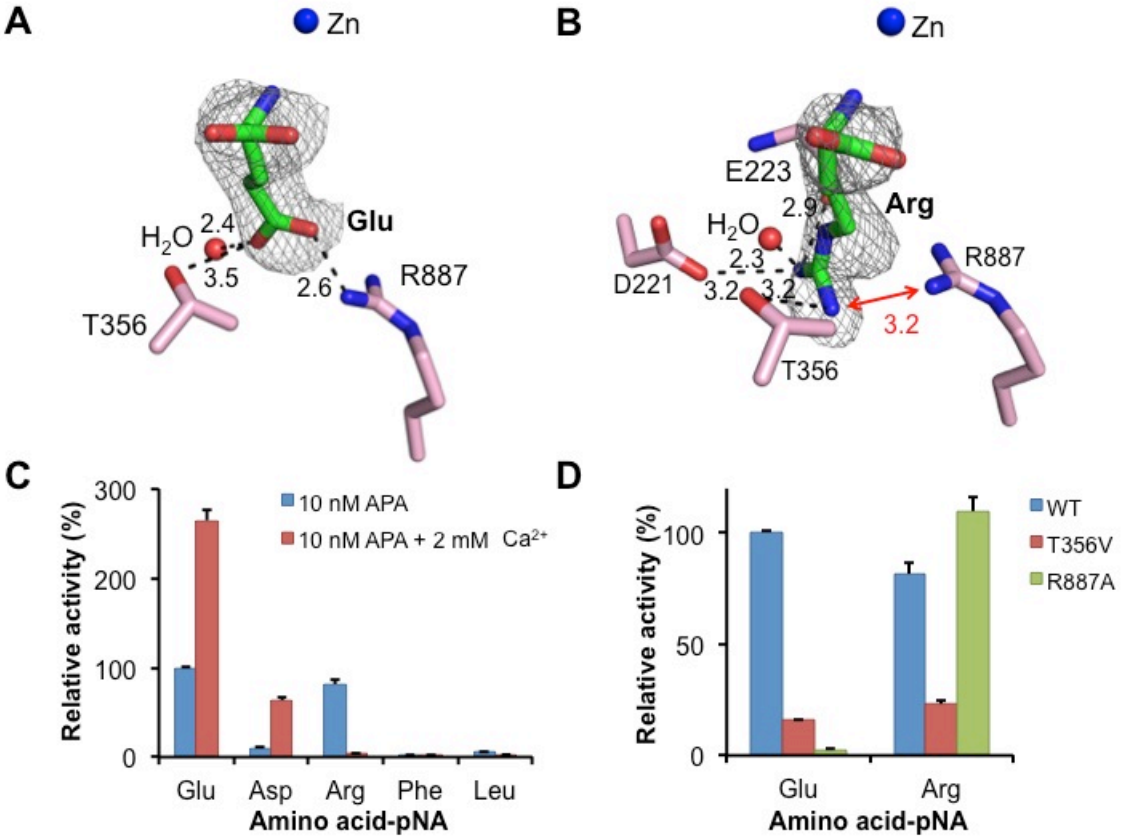

Figure 1.5 Substrate specificity of APA in the absence of calcium.
A, detailed interactions between a glutamate side chain and the S1 pocket of APA. B, detailed interactions between an arginine side chain and the S1 pocket of APA. APA residues are in pink, and ligands are in green. Electron density maps of ligands correspond to $F_o - F_c$ omit maps (contoured at 2.5σ) that were calculated in the absence of ligands. Models of APA-bound ligands were built based on these maps. Unit of distances is angstrom. C, relative catalytic activities of APA on different substrates (amino acid-*p*NA) in the absence or presence of calcium. The catalytic activity of APA on Glu-*p*NA in the absence of calcium was taken as 100%. Each bar shows the mean ± SEM (n = 3). D, relative catalytic activities of APA mutants on substrates. The catalytic activity of wild-type APA on Glu-*p*NA was taken as 100%.

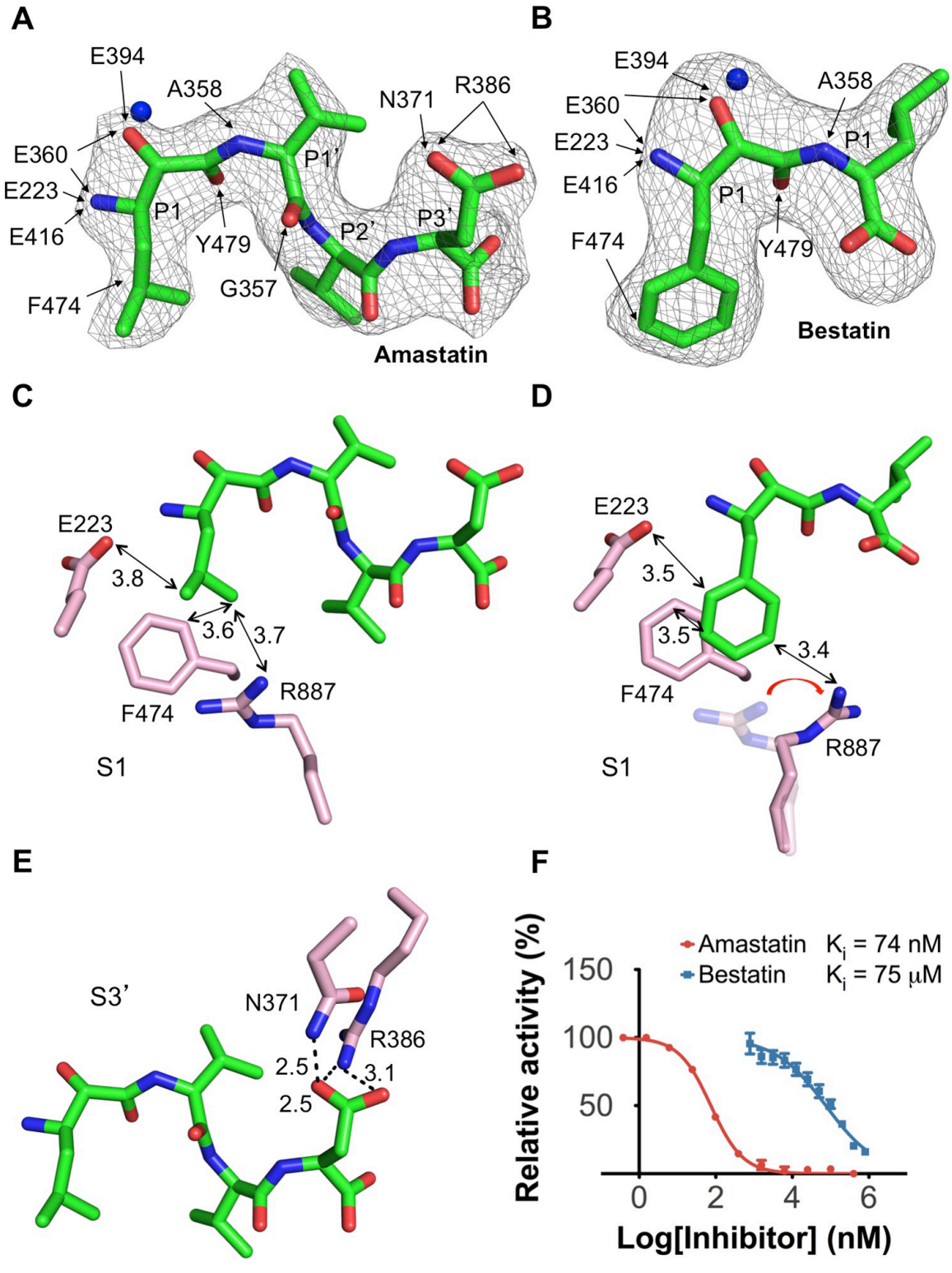
A
E394
A358
N371
R386
E360
P1'
E223
P1
E416
Y479
P2'
P3'
F474
G357
Amastatin
B
E394
A358
E360
E223
P1
E416
P1
F474
Y479
Bestatin
C
E223
3.8
3.6
3.7
F474
R887
S1
D
E223
3.5
3.5
3.4
F474
R887
S1
E
S3'
N371
R386
2.5
3.1
2.5
F
Amastatin Ki = 74 nM
Bestatin Ki = 75 μM
Relative activity (%)
150
100
50
0
2
4
6
Log[Inhibitor] (nM)

Figure 1.6 APA-inhibiting mechanisms of peptidomimetic inhibitors.
A, interactions between the inhibitor amastatin and APA. B, interactions between the inhibitor bestatin and APA. Detailed interactions between the inhibitors and APA residues are shown by arrows. Electron density maps of inhibitors correspond to $F_o - F_c$ omit maps (contoured at 2.5σ) that were calculated in the absence of inhibitors. Models of APA-bound inhibitors were built based on these maps. C, detailed interactions between the P1 side chain of amastatin and the S1 pocket of APA. D, detailed interactions between the P1 side chain of bestatin and the S1 pocket of APA. E, detailed interactions between the P3′ side chain of amastatin and the S3′ subsite of APA. F, inhibition of APA catalytic activity by amastatin and bestatin. The catalytic activity of APA on Glu-*p*NA in the absence of inhibitor was taken as 100%. All values are mean ± SEM (n = 3).

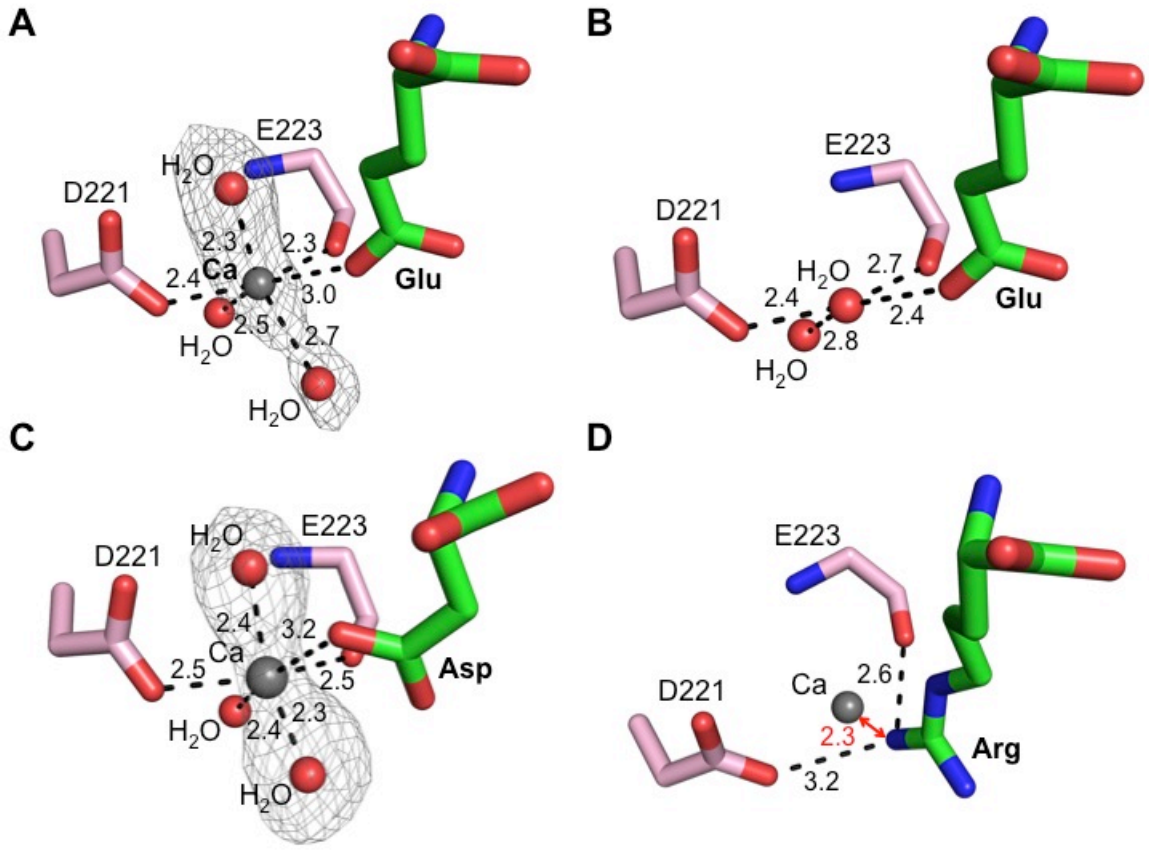

Figure 1.7 Calcium-modulated substrate specificity of APA.
A, observed calcium-binding site in the S1 pocket of glutamate-bound APA. The electron density map of calcium and two additional calcium-coordinating water molecules corresponds to the $F_o - F_c$ omit map (contoured at 3.5σ) that was calculated with a water molecule occupying the calcium-binding site and in the absence of the two additional water molecules. B, in the absence of calcium, a water molecule occupies the calcium-binding site in the S1 pocket of glutamate-bound APA. C, observed calcium-binding site in the S1 pocket of aspartate-bound APA. The electron density map was calculated in the same way as described for A, except that it was contoured at 3.0σ. D, modeled calcium-binding site in the S1 pocket of arginine-bound APA. In the presence of calcium, arginine was not observed upon soaking into APA crystals. Instead, a structural model was constructed in which calcium replaced the water molecule occupying the calcium-binding site as in Fig. 1.5B.

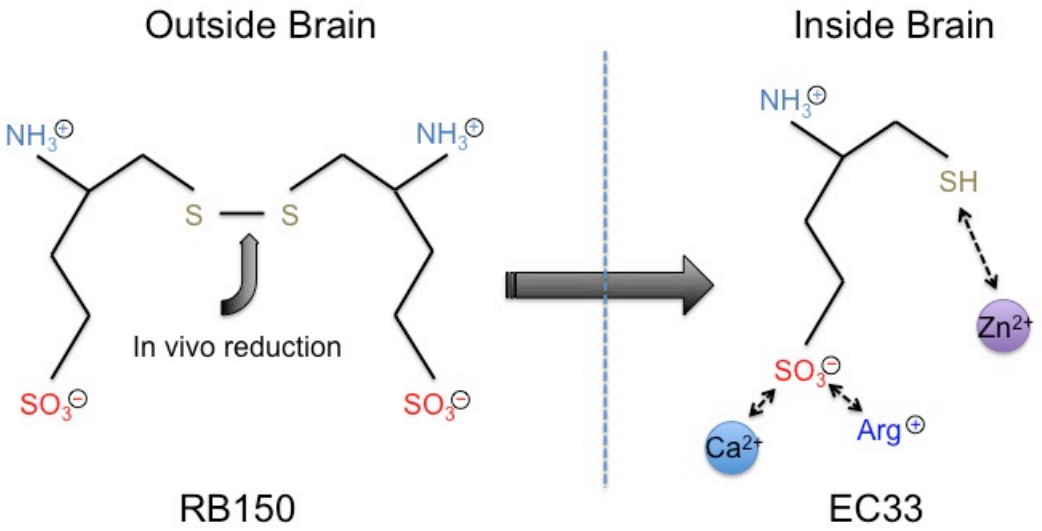

Figure 1.8 Possible mechanisms of action of RB150, a brain-targeting APA inhibitor.
RB150 is an orally bioavailable prodrug, can pass the blood-brain barrier, and is converted to its active form (EC33) inside the brain[18]. Potential interactions between EC33 and human APA are shown schematically.

	APA native	APA/ glutamate complex	APA/Ca/ glutamate complex	APA/Ca/ aspartate complex	APA/ arginine complex	APA/ amastatin complex	APA/ bestatin complex
Data Collection							
Space group	$P6_422$	$P6_422$	$P6_422$	$P6_422$	$P6_422$	$P6_422$	$P6_422$
Cell dimensions							
a, c (Å)	142.3	142.2	142.2	142.1	142.3	142.7	141.9
	237.3	237.1	237.1	237.4	237.2	237.8	237.1
γ (°)	120	120	120	120	120	120	120
Resolution (Å)	50-2.05	50-2.4	50-2.15	50-2.4	50-2.25	50-2.4	50-2.4
Total reflections	534840	336976	525986	403261	481998	609019	334475
Unique reflections	76159	49269	74275	54380	67214	53672	55096
Test-set reflections	3832	2518	3742	2750	3403	2713	2793
Wilson B-factor	35.90	42.53	31.44	45.48	32.96	34.62	43.00
R_{sym}	0.080	0.085	0.083	0.077	0.095	0.126	0.086
	(0.573)	(0.537)	(0.454)	(0.613)	(0.439)	(0.425)	(0.639)
$R_{p.i.m.}$	0.035	0.033	0.031	0.032	0.038	0.039	0.038
	(0.274)	(0.264)	(0.196)	(0.271)	(0.200)	(0.131)	(0.278)
I/σ	22.0	17.5	19.16	25.7	21.8	14.9	21.4
	(2.9)	(2.6)	(3.67)	(2.3)	(3.9)	(7.18)	(2.4)
Completeness (%)	99.4	88.7	97.3	97.9	99.8	98.2	99.7
	(99.6)	(90.8)	(93.9)	(98.9)	(98.3)	(100.0)	(99.7)
Redundancy	6.0	6.8	7.0	7.3	7.1	11.0	6.0
	(5.2)	(5.1)	(5.6)	(7.1)	(5.5)	(10.5)	(5.7)
Refinement							
Resolution (Å)	38.9-2.15	38.8-2.4	45.6-2.15	48.6-2.4	45.6-2.25	41.2-2.4	33.8-2.4
No. reflections	72327	49680	74275	54380	67214	53672	55096
R_{work}/R_{free}	0.173 /	0.167 /	0.183 /	0.152 /	0.167 /	0.210 /	0.170 /
	0.218	0.226	0.248	0.224	0.236	0.276	0.252
No. atoms	7993	7854	7945	7901	8011	7956	7853
Protein	7186	7153	7159	7159	7160	7151	7159
Carbohydrate	210	238	224	224	224	238	224
Ligand	0	10	10	9	12	33	22
Ion	1	1	2	2	1	1	1
Water	595	452	550	507	614	533	445
B-factors (Å^2)	58.7	61.5	56.9	64.6	59.7	56.1	65.4
Protein	56.9	59.6	55.2	62.8	57.5	54.2	63.6
Ligand		79.7	67.1	96.3	94.9	53.4	74.4
Ion	80.8	74.1	59.2	77.0	85.0	76.5	74.2
Water	58.8	60.6	58.6	62.8	61.9	55.2	64.4
R.m.s derivations							
Bond lengths (Å)	0.007	0.008	0.007	0.007	0.008	0.009	0.009
Bond angles (°)	1.239	1.297	1.265	1.201	1.245	1.313	1.282
Ramachandran plot							
Favored (%)	97	96	96	96	96	97	95
Outliers (%)	0.23	0.57	0.69	0.57	0.57	0.69	0.34

*Values in parentheses are for highest-resolution shell.

Table 1.1 Data collection and refinement statistics of APA and its complexes with substrates and inhibitors.
Values in parentheses are for the highest-resolution shell.

Chapter Two

Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus

Yang Yang*, Lanying Du*, Chang Liu*, Lili Wang, Cuiqing Ma, Jian Tang, Ralph S Baric, Shibo Jiang, Fang Li (***Co-first authors**)

Proc Natl Acad Sci USA 111: 12516-12521 (2014).

2.1 Introduction

Since the summer of 2012, a novel coronavirus, Middle East respiratory syndrome coronavirus (MERS-CoV), has emerged from the Middle East and spread all over the world. MERS-CoV infection often leads to acute pneumonia, renal failure, and even death. As of June 9th, 2015, MERS-CoV had infected 1218 people, with a fatality rate of ~37% (http://www.who.int/csr/don/09-june-2015-mers-korea/en/), and had demonstrated the capability for human-to-human transmission[41,42]. Alarmingly, coronavirus surveillance studies have suggested that MERS-CoV originated from bats. The genomic sequence of MERS-CoV is closely related to the sequences of two bat coronaviruses, HKU4 and HKU5[43-45], raising concerns over persistent bat-to-human cross-species transmission of the virus. Understanding the cross-species transmission mechanism of MERS-CoV is critical for evaluating long-term emerging disease potentials and for preventing and controlling the spread of bat-originated coronaviruses in humans. The clinical signs and epidemic patterns of MERS-CoV are reminiscent of the severe acute respiratory syndrome coronavirus (SARS-CoV), the etiological agent of the worldwide SARS epidemic in 2002-2003 that infected more than 8,000 people with a ~10% fatality rate[46,47]. MERS-CoV poses a significant threat to global health and economy.

Coronaviruses are enveloped and positive-stranded RNA viruses and can be divided into three major genera, α, β, and γ[48]. They mainly cause respiratory, gastrointestinal, and central nervous system diseases in mammals and birds. Coronaviruses recognize a wide range of receptors, including proteins and sialic acids[49]. Human NL63 respiratory

coronavirus (HCoV-NL63) from α-genus and SARS-CoV from β-genus both recognize angiotensin-converting enzyme 2 (ACE2) as their host receptor [50,51]. Porcine respiratory coronavirus (PRCV) and some other coronaviruses from α-genus recognize aminopeptidase N (APN)[52,53]. Mouse hepatitis coronavirus (MHV) from β-genus recognizes carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1)[54,55], although certain MHV strains also recognize heparan sulfate[56,57]. Some coronaviruses from each of the three genera recognize sugars[58-61]. MERS-CoV belongs to the β-genus and uses a cellular protease, human dipeptidyl peptidase 4 (DPP4), as its host receptor[62].

Receptor recognition has been established as an important determinant of the host range and tropism of coronaviruses[48,63]. An envelope-anchored trimeric spike protein mediates coronavirus entry into host cells by first binding to a host receptor through its S1 subunit and then fusing the host and viral membranes via its S2 subunit. The S1 subunit contains two independent domains, an N-terminal domain (NTD) and a C-terminal domain (CTD), both of which can potentially function as receptor-binding domains (RBD) (Fig. 2.3A)[49]. Specifically, coronavirus S1 CTDs can function as ACE2-, APN-, or heparan sulfate-binding RBDs, whereas S1 NTDs can function as CEACAM1- or sugar-binding RBDs. To date, crystal structures have been determined for a number of coronavirus RBDs by themselves or in complex with their host receptors, revealing how coronaviruses have evolved to recognize host receptors and thereby traffic between different species[64-68]. It is not known which one of MERS-CoV S1 domains is the DPP4-binding RBD or the tertiary structure of MERS-CoV RBD, both of which can provide insights into the

receptor recognition of MERS-CoV.

MERS-CoV spike protein shares 67% and 64% sequence similarity with the spike proteins in HKU4 and HKU5, respectively[45], raising the possibility that HKU4 and HKU5 may also use DPP4 as their receptor. However, previous studies showed that receptor recognition by coronaviruses is sensitive to residue changes in coronavirus RBDs[69]. For example, one or two residue changes in the RBD of another bat-originated coronavirus, severe acute respiratory syndrome coronavirus (SARS-CoV), can dictate whether SARS-CoV uses human angiotensin-converting enzyme 2 as its receptor for efficient human infections[64,69-71]. Therefore, investigating the receptors of HKU4 and HKU5 is important for understanding their potential threats to human health and for delineating the causes of MERS-CoV infections in humans.

In addition, coronavirus entry into host cells is limited not only by receptor recognition but also by membrane fusion. To fuse host and viral membranes, coronavirus spikes need to be cleaved at the S1/S2 boundary by host proteases[72-76]. The availability of these host proteases to coronaviruses and the specificities of these host proteases on coronavirus spikes can contribute to the host range and tropism of coronaviruses. Coronavirus-spike-processing host proteases may include proprotein convertases from the virus-producing cells (e.g., furin), proteases from the extracellular environment (e.g., elastase), proteases on the cell surface of virus-targeting cells (e.g., type 2 transmembrane serine protease, or TMPRSS2), and proteases in the lysosomes of virus-targeting cells[72,73] (e.g., cathepsin

L). Bat-originated and human-infecting coronaviruses MERS-CoV and SARS-CoV both use cell-surface protease TMPRSS2 and lysosomal protease cathepsins L for human cell entry, although proteases from the extracellular environment may also facilitate the cell entry process[72-79]. However, the mechanisms for human cell entry by bat coronaviruses are poorly understood, leaving a critical missing link in the bat-to-human transmission of coronaviruses. Therefore, examining the cell entry of HKU4 and HKU5 can provide novel knowledge of the causes of MERS-CoV (and SARS-CoV) infections in humans.

In collaboration with Dr. Yaoqing Chen and Dr. Kanagalaghatta R. Rajashankar, we identified a defined RBD in MERS-CoV spike S1 subunit binds human DPP4 with high affinity and determined its crystal structure. By investigating the receptor usage and cell entry of HKU4 and HKU5, I also elucidated the mechanisms by which bat coronaviruses adapt to human cells during cross-species transmission events. These results provide valuable knowledge that can be used to evaluate emerging disease potentials of bat coronaviruses and to prevent and control the future spread of bat coronaviruses in humans.

2.2 Materials and Methods

Cell lines

The HEK293T (human embryonic kidney), MRC-5 (human lung), and Tb1-Lu (*Tadarida brasiliensis* bat lung) cells were obtained from ATCC (www.atcc.org). Huh-7 (human liver) and Calu-3 (human lung) cells were kindly provided by Charles M. Rice at

Rockefeller University and Chien-Te K. Tseng at the University of Texas Medical Branch, respectively. These cell lines were maintained in Dulbecco's modified Eagle medium (DMEM), supplemented with 10% FBS, 2 mM L-glutamine, 100 units/ml penicillin, and 100 μg/ml streptomycin (Life Technologies).

Protein preparation and purification

The S1 domains of MERS-CoV (residues 367–588; GenBank accession number AFS88936.1), HKU4 (residues 372–593; GenBank accession number ABN10839.1), HKU5 (residues 375–586; GenBank accession number ABN10875.1), and SARS-CoV (residues 306–527; GenBank accession number NC_004718) were expressed and purified as previously described for the SARS-CoV C-domain[64,70,71]. Briefly, the coronavirus spike C-domains containing an N-terminal honeybee melittin signal peptide and a C-terminal His_6 tag was expressed in insect cells using the Bac-to-Bac expression system (Life Technologies), secreted into cell culture medium, and subsequently purified on a HiTrap Chelating HP affinity column (GE Healthcare) and a Superdex 200 gel filtration column (GE Healthcare). The proteins were concentrated to 10 mg/mL and stored in buffer containing 20 mM Tris (pH 7.4) and 200 mM NaCl. Fc-tagged coronavirus S1 domains were obtained by fusion of the human IgG_4 Fc region to the C terminus of the proteins and were expressed using the same procedure as for the His_6-tagged S1 domain. The Fc-tagged S1 domains were purified sequentially on a HiTrap Protein G HP column (GE Healthcare) and Superdex 200 gel filtration column. Human DPP4 (hDPP4) ectodomain (residues 39–766; GenBank accession number

NP_001926.2) and bat DPP4 (bDPP4) ectodomain (residues 36–760; GenBank accession number KC249974) containing an N-terminal human CD5 signal peptide and a C-terminal His_6 tag were expressed and purified using the same procedure as for His_6-tagged coronavirus S1 domains.

Pull-down assay using cell lysate

HEK293T cells were transfected with either empty pcDNA3.1(+) vector or pcDNA3.1(+) vector containing the human DPP4 gene. At 48 h post-transfection, cells were harvested and lysed in PBS with 300 mM NaCl and 0.25% SDS. Cell lysate was mixed with MERS-CoV C-domain. The protein complex was then precipitated with nickel-nitrilotriacetic acid (Ni-NTA) agarose (Thermo Scientific). DPP4 and MERS-CoV C-domain contain a C-terminal hemagglutinin (HA) tag and His_6 tag, respectively, and were detected by anti-HA and anti-His_6 antibodies (Santa Cruz Biotechnology), respectively.

Protein crystallization

Crystallization of the MERS-CoV C-domain was set up using the sitting drop vapor diffusion method, with 2 μl protein solution added to 2 μl reservoir buffer containing 27% (vol/vol) polyethylene glycol (PEG) 3350, 0.2 M $MgCl_2$, and 0.1 M bis-Tris (pH 7.0) at 20°C. After 2 weeks, crystals of the MERS-CoV C-domain were harvested in buffer containing 25% (vol/vol) ethylene glycol, 27% (vol/vol) PEG 3350, 0.2 M $MgCl_2$, and 0.1 M bis-Tris (pH 7.0) and flash-frozen in liquid nitrogen. Heavy-atom derivatives of MERS-CoV C-domain crystals were prepared by soaking crystals for 15 min in buffer

containing 1 M NaI, 25% (vol/vol) ethylene glycol, 27% (vol/vol) PEG 3350, 0.2 M $MgCl_2$, and 0.1 M bis-Tris (pH 7.0).

Data collection and structure determination

Data were collected at the Advanced Photon Source beamline 24-ID-C at 1.0716 Å for native crystals and 1.4586 Å for iodine-derivatized crystals. The crystal structure was determined using SIRAS (single isomorphous replacement with anomalous signal). X-ray diffraction data were processed using HKL2000[33]. Sixteen iodine sites were identified with the program HYSS[80] followed by phase refinement and solvent flattening with RESOLVE[35]. The model was built and refined with Refmac[38] at 2.13 Å to a final R_{work} and R_{free} of 14.9% and 20.6%, respectively. In the final model of the MERS-CoV C-domain, 98% of residues are in the favored regions of the Ramachandran plot, and 0% of residues are in the disallowed regions.

AlphaScreen protein-protein binding assay

The interactions between different coronavirus S1 domains and hDPP4 or bDPP4 were measured using AlphaScreen, as previously described[81,82]. Briefly, each of the Fc-tagged coronavirus S1 domains at 3 nM final concentrations was mixed with 100 nM His_6-tagged hDPP4 or bDPP4 in ½ AreaPlate (PerkinElmer) for 1 h at room temperature. AlphaScreen Nickel Chelate Donor Beads and AlphaScreen protein A acceptor beads (PerkinElmer) were added to the mixtures at final concentrations of 5 μg/ml. The mixtures were incubated at room temperature for 1 h and protected from light. The assay

plates were read in an EnSpire plate reader (PerkinElmer).

Pull-down assay using purified protein

Protein pull-down assay was carried out as previously described[83]. Briefly, 5 μg His_6-tagged hDPP4 or bDPP4 was mixed with each of the 5 μg Fc-tagged coronavirus S1 domains. The formed complex was precipitated using Ni-NTA agarose (Thermo Scientific). The DPP4 and coronavirus S1 domains were separated by SDS/PAGE and detected by anti-His_6 and anti-human IgG_4 Fc antibodies (Santa Cruz Biotechnology), respectively.

Dot blot binding assay

Dot blot binding assay was carried out as previously described[65]. Briefly, 10 μg each of the Fc-tagged coronavirus S1 domains was dotted onto a nitrocellulose membrane. The membranes were dried completely and blocked with 5% skim milk at 37 °C for 1 h. Next, 25 μg/ml His_6-tagged hDPP4 or bDPP4 was preincubated alone or with 20 μg/ml goat anti-hDPP4 polyclonal antibodies (R&D Systems) at 37 °C for 1 h, added to the membrane, and incubated at 4 °C overnight. The membrane was then washed five times with phosphate-buffered saline with Tween-20 (PBST), incubated with anti-His_6 mouse monoclonal IgG_4 HRP conjugate antibody (Santa Cruz biotechnology) at 37 °C for 2 h, and washed five times with PBST. Finally, the bound proteins were detected using ECL plus (GE Healthcare).

Inhibition of MERS-CoV-spike-mediated pseudovirus entry by HKU4 RBD

Pseudovirus cell entry assay was carried out as previously described[84]. Briefly, MERS-CoV-spike-pseudotyped retroviruses expressing a luciferase reporter gene were prepared by cotranfecting HEK293T cells with a plasmid carrying Env-defective, luciferase-expressing HIV-1 genome (pNL4-3.luc.R-E-) and a plasmid encoding MERS-CoV spike protein. The produced pseudovirus particles were harvested 72 h after transfection. HEK293T cells transiently expressing hDPP4 or bDPP4 were preincubated with gradient concentrations of each of the purified coronavirus S1 domains for 1 h at 37 °C and then infected by equal amounts of MERS-CoV-spike-pseudotyped retrovirus particles. After incubation for 5 h at 37 °C, medium was changed and cells were incubated for an additional 60 h. Cells were then washed with PBS and lysed. Aliquots of cell lysates were transferred to Optiplate-96 (PerkinElmer), followed by addition of luciferase substrate (Promega). Relative light units were measured using EnSpire plate reader (PerkinElmer).

HKU4-spike-mediated pseudovirus entry into human and bat cells

Retroviruses pseudotyped with MERS-CoV spike, HKU4 spike, or no spike (mock) were incubated with gradient concentrations of TPCK-treated trypsin (Sigma-Aldrich) for 10 min at 25 °C, and 100 μg/ml soybean trypsin inhibitor (Sigma-Aldrich) was then added to stop the proteolysis reactions. Trypsin-treated pseudoviruses were used to spin-infect HKE293T cells (transiently expressing hDPP4, bDPP4, or no DPP4), Huh-7 cells (transiently expressing TMPRSS2 or no TMPRSS2), Calu-3 cells, MRC-5 cells, or Tb1-Lu cells (transiently expressing hDPP4, bDPP4, or no DPP4) in 96-well plates at 1200 ×

g for 2 h at 4 °C. After incubation for 5 h at 37 °C, medium was changed. Cells were incubated for an additional 60 h, then lysed and measured for luciferase activity.

Neutralization of HKU4-spike-mediated pseudovirus entry by anti-hDPP4 antibodies

HEK293T cells expressing hDPP4, bDPP4, or no DPP4 were preincubated with 0, 2 μg/ml, or 20 μg/ml goat anti-hDPP4 polyclonal antibodies (R&D Systems) at 37 °C for 1 h and then infected by equal amounts of MERS-CoV-spike-pseudotyped retroviruses with no trypsin treatment or HKU4-spike-pseudotyped retroviruses after trypsin treatment. The infection procedure was the same as described earlier.

Protease processing of MERS-CoV- or HKU4-spike-packaged pseudoviruses

Retroviruses pseudotyped with MERS-CoV spike, HKU4 spike, or no spike (mock) were incubated with gradient concentrations of TPCK-treated trypsin (Sigma-Aldrich) for 10 min at 25 °C, and 100 μg/ml soybean trypsin inhibitor (Sigma-Aldrich) was then added to stop the proteolysis reactions. Trypsin-treated pseudoviruses were concentrated in an Amicon Ultracentrifuge filter unit with 100-kDa cutoff (EMD Millipore, Billerica, MA). The pseudovirus samples were then subjected to Western blot analysis. The His_6-tagged spikes were detected by antibody against their C-terminal His_6 tag.

Inhibition of pseudovirus entry into human cells by lysosome inhibitors

Huh-7 cells were preincubated with 8 or 40 mM NH_4Cl (Sigma-Aldrich) or 4 or 20 μM

E-64d (Sigma-Aldrich) at 37 °C for 1 h and then infected by HKU4- or MERS-CoV-spike-pseudotyped retroviruses that had been pretreated or not pretreated with trypsin. The infection procedure was the same as described earlier. Huh-7 cells infected by retroviruses pseudotyped with no spike (mock) were used as control.

2.3 Results

Identification and functional characterization of MERS-CoV RBD

Dr. Yaoqing Chen designed a construct of the MERS-CoV S1 C-domain (residues 367 to 588), based on the sequence alignment of MERS-CoV and SARS-CoV S1 subunits (Fig. 2.1). Although the two S1 subunits have low sequence similarity, sequence alignment was guided by conserved cysteines. Based on its crystal structure, SARS-CoV C-domain contains eight essential cysteines that form four disulfide bonds[64]. Our designed MERS-CoV C-domain also contains eight cysteines, seven of which are conserved between SARS-CoV and MERS-CoV (Fig. 2.1). The MERS-CoV C-domain was expressed and purified using the protocol that we used for the SARS-CoV C-domain[64,70,71]. The protein was a monomer in solution, based on the gel filtration chromatography profile.

I characterized the interactions between the recombinant MERS-CoV C-domain and hDPP4 on human cell surfaces, using pull down assay. The MERS-CoV C-domain pulled down hDPP4 from cells that express human DPP4 (Fig. 2.2). This result suggests that the recombinant MERS-CoV C-domain is the RBD that binds human DPP4.

Crystal structure of MERS-CoV RBD

Having identified the MERS-CoV RBD, we further determined its crystal structure (Fig. 2.3B and C). The MERS-CoV RBD was crystallized in space group $P2_12_12_1$, with 2 molecules per asymmetric unit. The structure was determined using an iodine derivative obtained by a short cryosoak[85] and refined at a 2.13 Å resolution (Table 2.1). The two copies of MERS-CoV RBD occupying the same asymmetric unit have a small buried interface (~500 $Å^2$), consistent with the observation during protein purification that the MERS-CoV RBD is a monomer in solution.

The crystal structure of the MERS-CoV RBD contains a core structure and an accessory subdomain (Fig. 2.3C). The core structure is a five-stranded antiparallel β-sheet with several short α-helices. The accessory subdomain lies on one edge of the core structure and consists of a four-stranded antiparallel β-sheet. The eight cysteines in the RBD form four disulfide bonds. Three of the disulfide bonds stabilize the core structure by connecting cysteine 383 to 407, 425 to 478, and 437 to 585; the remaining disulfide bond strengthens the accessory subdomain by connecting cysteine 503 to 526. The RBD also contains 2 glycans that are N-linked to Asn410 in the core structure and Asn487 in the accessory subdomain, respectively. We were able to trace all of the RBD residues (except for the N-terminal 12 residues) with an average B factor of 40.1 $Å^2$. Overall, the MERS-CoV RBD has a well-folded tertiary structure that is stabilized by disulfide bonds and glycans.

DPP4 is a functional receptor for HKU4 spike

Based on the sequence alignment of MERS-CoV, HKU4 and HKU5 spikes, I designed a construct of HKU4 S1 domain (residues 372–593) and a construct of HKU5 S1 domains (residues 375–586), which correspond to the DPP4-binding RBD in MERS-CoV spike (residues 367–588) (Fig. 2.4A). I expressed and purified the S1 domains of four viruses (HKU4, HKU5, MERS-CoV, and SARS-CoV), as well as DPP4 ectodomain from two species (human and bat), using procedures described previously[64,83]. The two bat species from which HKU4 and HKU5 were isolated belong to the same family, but different genera[43-45] (Fig. 2.4B). To understand coronavirus spike/DPP4 interactions, the DPP4 and coronavirus spike genes under investigation need to come from the same or similar bat species. Because the DPP4 genes from the above two bat species are unavailable, I chose the DPP4 gene from another bat species in the same genus as the one from which HKU5 was isolated (Fig. 2.4B). All of the recombinant proteins were purified to high homogeneity (Fig. 2.5) and subsequently used in the following biochemical studies.

Using three alternative approaches, I characterized the binding interactions between the S1 domain of each virus and DPP4 from each species. AlphaScreen protein-protein binding assay showed that whereas the MERS-CoV S1 domain bound hDPP4 with significantly higher affinity than it bound bDPP4, the corresponding HKU4 S1 domain bound bDPP4 slightly better than it bound hDPP4 (Fig. 2.6A). In contrast, the HKU5 S1 domain did not bind hDPP4 or bDPP4. As a control, the SARS-CoV S1 domain only bound its own receptor human angiotensin-converting enzyme 2, but not hDPP4 or

bDPP4. Pull-down assay revealed that hDPP4 was more efficient than bDPP4 at pulling down the MERS-CoV S1 domain from solution, whereas both hDPP4 and bDPP4 pulled down the HKU4 S1 domain efficiently (Fig. 2.6B). In contrast, neither hDPP4 nor bDPP4 pulled down the SARS-CoV or HKU5 S1 domain. Pseudovirus inhibition assay demonstrated that the HKU4 S1 domain, but not the HKU5 or SARS-CoV S1 domain, inhibited MERS-CoV-spike-mediated pseudovirus entry into HEK293T cells exogenously expressing hDPP4 or bDPP4. This indicates that the HKU4 S1 domain and MERS-CoV-spike-packaged pseudoviruses competed for the same DPP4 receptor on HEK293T cell surface (Fig. 2.6C). Taken together, these results showed that the HKU4 S1 domain, but not the HKU5 S1 domain, binds both hDPP4 and bDPP4, and that the MERS-CoV S1 domain binds hDPP4 significantly better than it binds bDPP4, whereas the HKU4 S1 domain binds bDPP4 slightly better than it binds hDPP4. These results strongly suggest that HKU4 spike, but not HKU5 spike, uses DPP4 as its receptor and that the HKU4 S1 domain is the DPP4-binding RBD.

MERS-CoV- and HKU4-spike-mediated viral entry into human cells

To confirm that DPP4 is the receptor for HKU4 spike, I investigated whether HKU4 spike could mediate viral entry into DPP4-expressing human cells. Because live HKU4 virus has never been successfully cultured, it is not an option to use live HKU4 virus in this study. Instead, I performed an HKU4-spike-mediated pseudovirus entry assay. To this end, retroviruses pseudotyped with HKU4 spike were used to enter HEK293T cells exogenously expressing either hDPP4 or bDPP4 on their surface. Surprisingly, HKU4

spike failed to mediate pseudovirus entry into these DPP4-expressing HEK293T cells (Fig. 2.7A). As a comparison, MERS-CoV spike efficiently mediated pseudovirus entry into hDPP4-expressing HEK293T cells, but less efficiently into bDPP4-expressing HEK293T cells (Fig. 2.7B). However, when treated with trypsin, HKU4-spike-packaged pseudoviruses entered both hDPP4- and bDPP4-expressing HEK293T efficiently. As a comparison, trypsin treatment had little effect on MERS-CoV-spike-mediated pseudovirus entry into DPP4-expressing HEK293T cells. To validate these findings, we performed two control experiments to ensure that HKU4-spike-mediated entry depended on both HKU4 spike and DPP4. First, after trypsin treatment, HKU4-spike-pseudotyped retroviruses could not enter HEK293T cells not expressing DPP4 (Fig. 2.7A). Second, retroviruses not pseudotyped with HKU4 or MERS-CoV spike could not enter DPP4-expressing HEK293T cells (Fig. 2.7C). Overall, these results support two conclusions: DPP4 is the functional receptor for HKU4 spike, and exogenous trypsin is needed for HKU4-spike-mediated, but not MERS-CoV-spike-mediated, pseudovirus entry into human cells.

To further investigate the cell entry mechanism of HKU4, I repeated the pseudovirus entry assay using Huh-7 cells (human liver cells), Calu-3 cells (human lung cells), and MRC-5 cells (human lung cells), all of which endogenously express hDPP4[79,84,86]. The results again showed that exogenous trypsin was needed for HKU4-spike-mediated, but not MERS-CoV-spike-mediated, pseudovirus entry into Huh-7, Calu-3, and MRC-5 cells (Fig. 2.8). Thus, our finding on the effects of trypsin on HKU4- and MERS-CoV-spike-

mediated entry can be extended to different types of human cells.

Neutralization of MERS-CoV- and HKU4-spike-mediated viral entry by DPP4-targeting antibody

Further confirmation that DPP4 is the receptor for HKU4 spike came from an examination of whether anti-hDPP4 polyclonal antibodies could competitively block the interactions between HKU4 spike and hDPP4. First, dot blot binding assay showed that the antibodies almost completely blocked the binding between HKU4 RBD and hDPP4 and significantly inhibited the binding between MERS-CoV RBD and hDPP4 (Fig. 2.9A). In contrast, the antibodies did not block the binding between HKU4 or MERS-CoV RBD and bDPP4, suggesting that the antibodies did not target bDPP4. Second, pseudovirus neutralization assay revealed that the antibodies efficiently blocked both HKU4- and MERS-CoV-spike-mediated entry into hDPP4-expressing HEK293T cells, but not bDPP4-expressing cells (Fig. 2.9B). These results demonstrated that anti-hDPP4 antibodies competed with HKU4 spike for the binding sites on hDPP4, further confirming that DPP4 is the functional receptor for HKU4 spike.

Activation of MERS-CoV and HKU4 spikes by human cellular proteases

What human cellular proteases process MERS-CoV spike, but not HKU4 spike? To address this question, I looked into all three stages in which human cellular proteases may play a role in coronavirus-spike-mediated cell entry. First, by performing Western blot on MERS-CoV- and HKU4-spike-packaged pseudoviruses, I found that MERS-CoV

spike, but not HKU4 spike, had been partially cleaved coming out of virus-producing HEK293T cells (Fig. 2.10). Moreover, HKU4 spike was cleaved by exogenous trypsin (Fig. 2.10). Thus, MERS-CoV spike, but not HKU4 spike, can be partially processed in virus-producing cells by human endogenous proteases. The processing may be achieved potentially by human proprotein convertases. Second, I carried out the pseudovirus entry assay using Huh-7 cells exogenously expressing human TMPRSS2 Huh-7 cells do not endogenously express TMPRSS2[87]. The results showed that human TMPRSS2 enhanced MERS-CoV-spike-mediated, but not HKU4-spike-mediated, pseudovirus entry (Fig. 2.11). Third, I explored the role of human lysosomal proteases in MERS-CoV- and HKU4-spike-mediated pseudovirus entry into Huh-7 cells, using two types of inhibitors: lysosomal acidification inhibitor NH_4Cl and lysosomal protease inhibitor E-64d which inhibits lysosomal cathepsins L, B, H, etc[88] (Fig. 2.12). MERS-CoV-spike-mediated entry was significantly inhibited by both NH_4Cl and E-64d, revealing the critical role of lysosomal proteases in MERS-CoV-spike-mediated entry (the higher-than-baseline level of pseudovirus entry in the presence of NH_4Cl or E-64d is probably a result of the prior partial processing of MERS-CoV spike during virus packaging and release). In contrast, these inhibitors showed little effect on HKU4-spike-mediated entry. Because HKU4 spike fails to transduce the same cell line whose lysosomal proteases activate MERS-CoV spike, HKU4 spike most likely cannot be activated by human lysosomal proteases. Overall, these results have demonstrated that MERS-CoV spike, but not HKU4 spike, can be processed by human cellular proteases from three different stages of virus infection: virus packaging and release, viral attachment to human cell surface, and viral

endocytosis.

MERS-CoV- and HKU4-spike-mediated viral entry into bat cells

Finally, I investigated whether HKU4 and MERS-CoV spikes can mediate pseudovirus entry into bat cells. I repeated pseudovirus entry assay, using Tb1-Lu cells (bat lung cells) that exogenously express hDPP4 or bDPP4. The results show that in the absence of trypsin, both HKU4 and MERS-CoV spikes efficiently mediated pseudovirus entry into Tb1-Lu cells, even though trypsin could further enhance HKU4-spike-mediated pseudovirus entry into Tb1-Lu cells (Fig. 2.13). In a control experiment, Tb1-Lu cells without exogenously expressed hDPP4 or bDPP4 failed to support MERS-CoV- or HKU4-spike-mediated pseudovirus entry, indicating a lack of endogenously expressed DPP4 in Tb1-Lu cells. These results revealed that in the presence of an appropriate receptor, activation by exogenous proteases is not a requirement for HKU4- or MERS-CoV-spike-mediated pseudovirus entry into bat cells.

2.4 Discussion

In this part of my thesis study, after identifying the MERS-CoV RBD and determining its crystal structure, I further characterized the receptor usage and cell entry of MERS-related bat coronaviruses HKU4 and HKU5 compared with bat-originated but human-infecting MERS-CoV. Because HKU4 and HKU5 had never infected human cells, this study provided an opportunity to understand human cell adaptations by MERS-CoV. Through comparative analysis of these viruses, this study not only has identified the

functional receptor for HKU4 but also has revealed the functional differences between MERS-CoV and HKU4 spikes in their adaptations to use human receptor and cellular proteases for human cell entry.

First, exploiting a combination of experimental approaches including AlphaScreen, protein pull down, dot blot hybridization, pseudovirus inhibition, pseudovirus entry, and antibody neutralization, this study has established DPP4 as the functional receptor for HKU4. Although HKU4 spike and bDPP4 are from two different bat species (Fig. 2.4B), the positive interaction between the two proteins suggests that HKU4 spike should also interact positively with bDPP4 from its own bat species. Indeed, my results show that both hDPP4 and bDPP4 are efficient receptors for HKU4 spike, although humans and bats are only remotely related species. In contrast, HKU5 spike does not use DPP4 as its receptor, despite sharing sequence similarity with MERS-CoV and HKU4 spikes. Again, this is unlikely because of the different bat species from which HKU5 and bDPP4 were isolated, but is a result of two deletions in the presumable DPP4-binding region of the HKU5 S1 domain (Fig. 2.4A). Moreover, whereas MERS-CoV spike binds hDPP4 with significantly higher affinity than it binds bDPP4, HKU4 spike slightly prefers bDPP4 over hDPP4. MERS-CoV's preference for hDPP4 over bDPP4 as its receptor is consistent with a recent finding[89]. Thus, it is likely that MERS-CoV spike has evolved to use hDPP4 efficiently but has become less effective in recognizing bDPP4, whereas HKU4 spike has not evolved adaptive mutations to promote efficient hDPP4 usage.

Second, using exogenous trypsin, lysosomal acidification and protease inhibitors, and pseudovirus entry into various types of human and bat cells, this study has revealed that human cellular proteases can activate MERS-CoV-spike-mediated, but not HKU4-spike-mediated, pseudovirus entry into human cells. These human cellular proteases act on different stages of MERS-CoV-spike-mediated pseudovirus entry and include human endogenous proteases (e.g., proprotein convertases) acting during virus packaging and release, human TMPRSS2 acting during viral attachment to human cell surface, and human lysosomal cathepsins acting during viral endocytosis. Exogenous proteases contribute to, but are not essential for, MERS-CoV entry into human cells exogenously expressing low-affinity DPP4 homologs[89]. In contrast, an exogenous protease is essential for HKU4-spike-mediated entry into human cells. However, both MERS-CoV and HKU4 spikes can mediate efficient pseudovirus entry into bat cells, suggesting that bat cellular proteases can process both MERS-CoV and HKU4 spikes. Because at this time, little is known about the functions of bat cellular proteases, future research will be needed to understand how coronavirus spikes mediate virus entry into bat cells. Nevertheless, this study indicates that MERS-CoV spike, but not HKU4 spike, has adapted to use human cellular proteases and thus has evolved the ability to mediate virus entry into human cells.

These findings provide critical insight into the current MERS-CoV infections and raises concerns over potential future bat coronavirus infections in humans. DPP4-recognizing bat coronaviruses such as HKU4 pose a threat to human health because their spikes can potentially adapt to use human DPP4 and human cellular proteases for human cell entry.

Viral adaptation to human receptors enhances viral attachment to human cells, whereas viral adaptation to human cellular proteases reduces viral dependence on the extracellular environment. Both adaptation processes can facilitate coronaviruses to expand their host ranges and tropisms. These evolutionary changes may have already taken place during the bat-to-human transmission of MERS-CoV. Recent findings suggest that MERS-CoV has been circulating in camels for some time[90-93]. Viral adaptations to camel cells could have been an intermediate step toward viral adaptations to human cells. Regardless of whether the transmission to humans came directly from bats or indirectly through camels, the eventual viral adaptations to human cells may have contributed to the increased replication, transmission, and pathogenesis of MERS-CoV in humans. Our findings enhance understanding of the likely causes of MERS-CoV infections in humans, the potential health risks associated with DPP4-recognizing bat coronaviruses, and the mechanisms of cross-species transmissions of animal viruses in general.

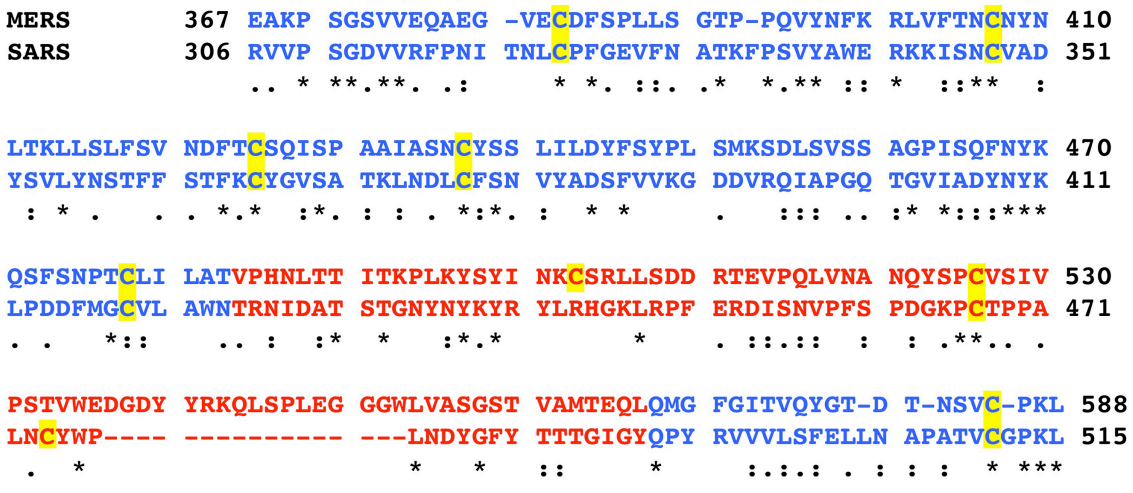

Figure 2.1 Sequence comparisons of MERS-CoV and SARS-CoV S1 C-domains. Sequence alignment of MERS-CoV and SARS-CoV C-domains. Residues corresponding to C-domain core structures are in cyan, and residues corresponding to the proposed RBM region are in red. Cysteine residues are highlighted. Asterisks indicate positions that have fully conserved residues; colons indicate positions that have strongly conserved residues; periods indicate positions that have weakly conserved residues.

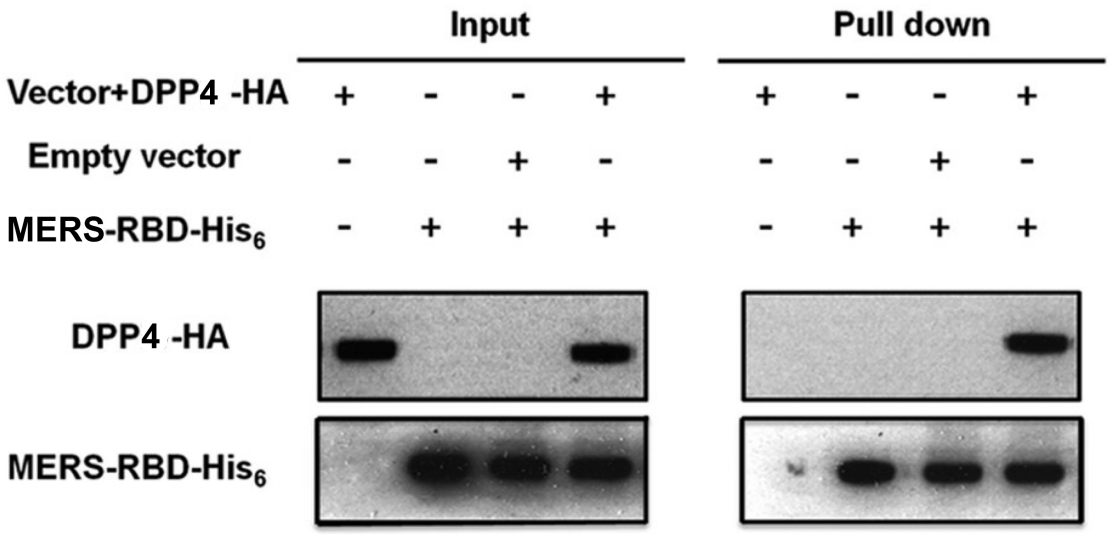

Figure 2.2 Interactions between the MERS-CoV S1 C-domain and human DPP4. Pulldown of cell surface DPP4 by MERS-CoV C-domain. 293T cells were transfected with either empty vector or vector containing DPP4-HA gene. At 48 h posttransfection, cells were harvested and lysed. Cell lysate was mixed with the MERS-CoV C-domain containing a C-terminal His_6 tag. The C-domain/DPP4 complex was then precipitated with Ni-NTA beads.

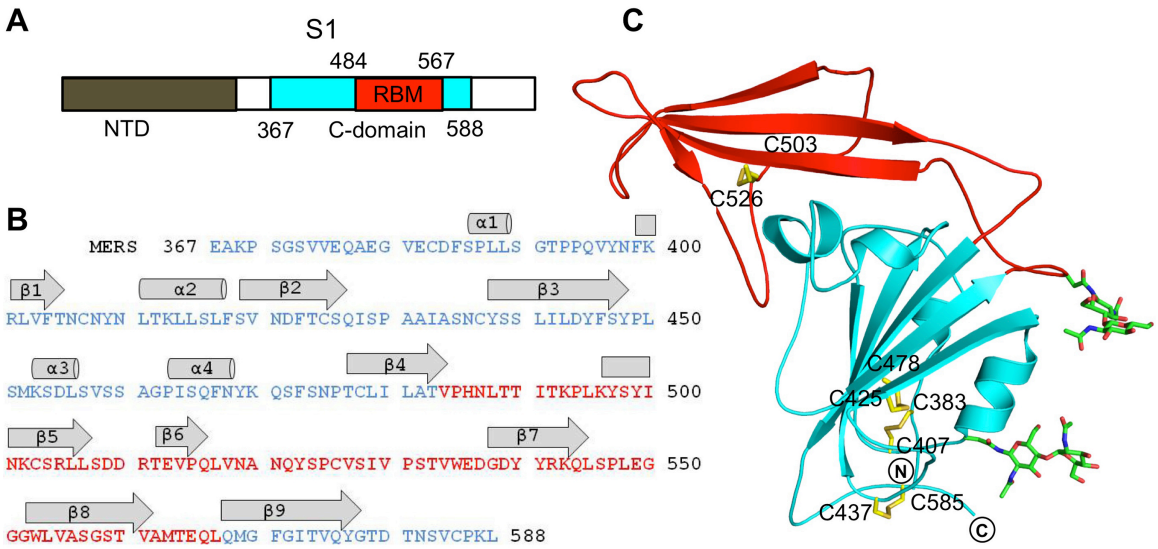

Figure 2.3 Crystal structure of the MERS-CoV S1 C-domain.
(A) Domain structure of MERS-CoV S1 subunit that contains an N-terminal domain (NTD) and a C-domain. The boundaries of the C-domain and proposed receptor-binding motif (RBM) of MERS-CoV were determined by sequence and structural comparisons with the SARS-CoV S1 subunit (see Fig. 2.1). (B) Sequence and secondary structure of the MERS-CoV C-domain. Helices are drawn as cylinders, and strands are drawn as arrows. The disordered region is shown as a dashed line. (C) Crystal structure of the MERS-CoV C-domain. The core structure is in cyan, and the proposed RBM region is in red, disulfide bond-linked cysteine residues are shown as sticks in yellow, and N-linked glycans are shown as sticks in green.

A

```
MERS               367 EAKP SGSVVEQAEG VECDFSPLLS GTPPQVYNFK RLVFTNCNYN 410
HKU4               372 EASA TGTFIEQPNA TECDFSPMLT GVAPQVYNFK RLVFSNCNYN 415
HKU5               375 EASP RGEFIEQATT QECDFTPMLT GTPPPIYNFK RLVFTNCNYN 418
                       **..  * .:**.    ****:*:*: *..* :**** ****:*****

MERS  LTKLLSLFSV NDFTCSQISP AAIASNCYSS LILDYFSYPL SMKSDLSVSS AGPISQFNYK 470
HKU4  LTKLLSLFAV DEFSCNGISP DSIARGCYST LTVDYFAYPL SMKSYIRPGS AGNIPLYNYK 475
HKU5  LTKLLSLFQV SEFSCHQVSP SSLATGCYSS LTVDYFAYST DMSSYLQPGS AGAIVQFNYK 478
      ******** * .:*:*  :**  ::* .***: * :***:*.  .*.* :  .* ** *  :***

MERS  QSFSNPTCLI LATVPHNLTT ITKPLKYSYI NKCSRLLSDD RTEVPQLVNA N-QYSPCVSIV 530
HKU4  QSFANPTCRV MASVLANVT- ITKPHAYGYI SKCSRLTGAN QDVETPLYIN PGEYSICRDFS 535
HKU5  QDFSNPTCRV LATVPQNLTT ITKPSNYAYL TECYKTSAYG KN---YLYNA PGAYTPCLSLA 536
      *.*:**** : :*:*  *:*  ****  *.*: .:* :  . . :     *       *: * . :

MERS  PSTVWEDGDY YRKQLSPLEG GGWLVASGST VAMTEQLQMG FGITVQYGTD TNSVCPKL 588
HKU4  PGGFSEDGQV FKRTLTQFEG GGLLIGVGTR VPMTDNLQMS FIISVQYGTG TDSVCPML 593
HKU5  SRGFS----- ---TKYQSHS DGELTTTGYI YPVTGNLQMA FIISVQYGTD TNSVCPMQ 586
      .  .              .. .* *   *    .:* :***. * *:*****. *:****
```

B

Bat	HKU4	HKU5	bDPP4
Order	Chiroptera	Chiroptera	Chiroptera
Family	Vespertilionidae	Vespertilionidae	Vespertilionidae
Subfamily	Vespertilioninae	Vespertilioninae	Vespertilioninae
Genus	Tylonycteris	Pipistrellus	Pipistrellus
Species	pachypus	abramus	pipistrellus

Figure 2.4 Coronavirus S1 domains and DPP4.
(A) Sequence alignment of the S1 domains from MERS-CoV, HKU4, and HKU5. Asterisks indicate positions that have fully conserved residues. Colons indicate positions that have strongly conserved residues. Periods indicate positions that have weakly conserved residues. DPP4-binding region in MERS-CoV RBD is in red. There are two deletions in the presumable DPP4-binding region of the HKU5 S1 domain. (B) Classification of bats from which the genes of HKU4, HKU5, and bDPP4 were isolated.

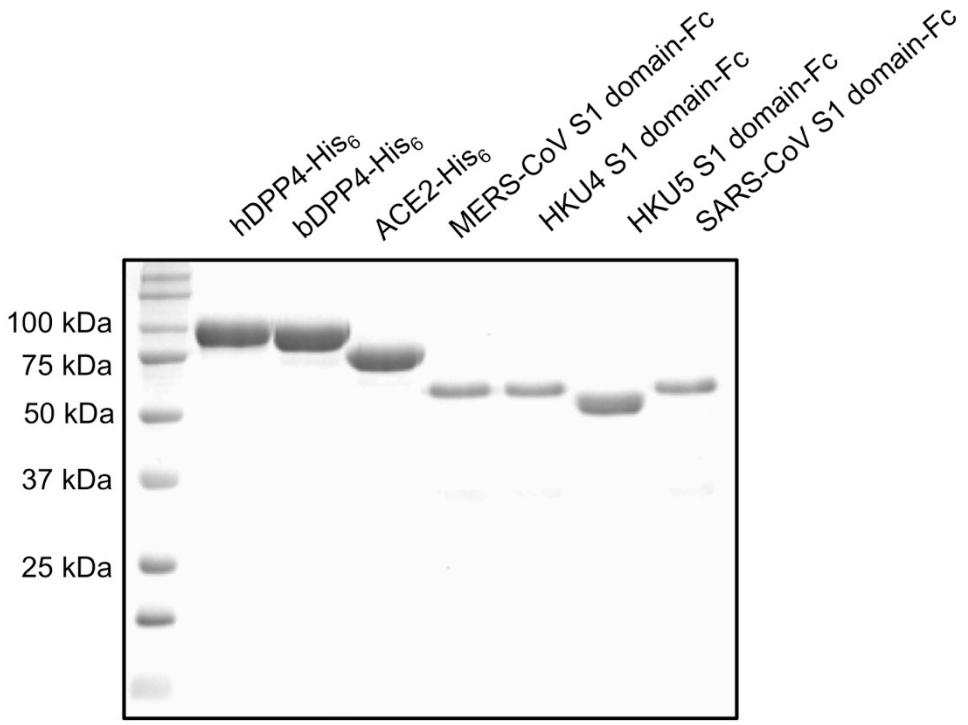

Figure 2.5 SDS/PAGE analysis of purified recombinant proteins used in this study. These proteins include His_6-tagged ectodomains of hDPP4 and bDPP4 and Fc-tagged S1 domains from MERS-CoV, HKU4, HKU5, and SARS-CoV. The gel was stained by Coomassie Blue.

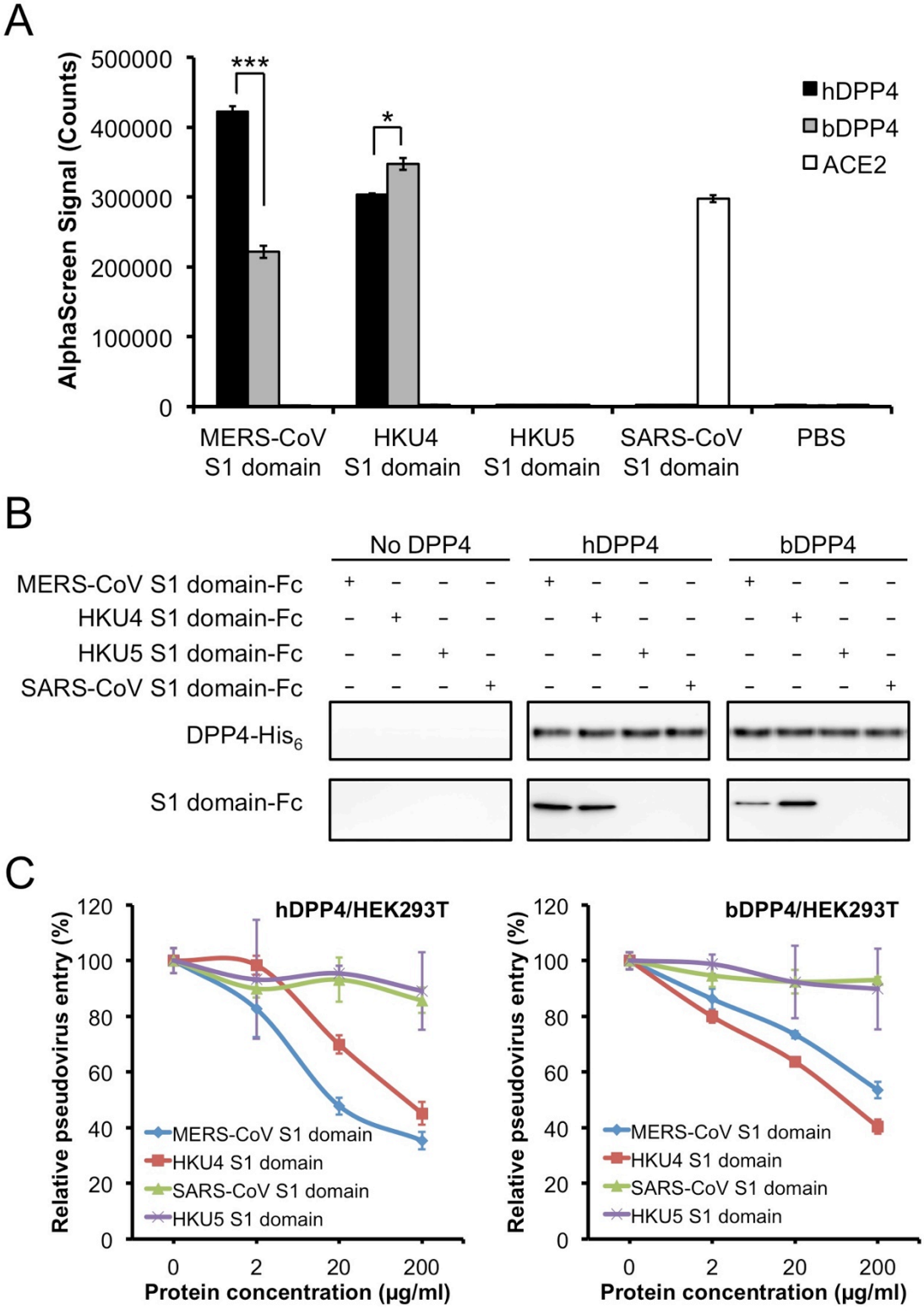
A
AlphaScreen Signal (Counts)
500000
400000
300000
200000
100000
0
***
*
hDPP4
bDPP4
ACE2
MERS-CoV S1 domain
HKU4 S1 domain
HKU5 S1 domain
SARS-CoV S1 domain
PBS
B
No DPP4
hDPP4
bDPP4
MERS-CoV S1 domain-Fc
HKU4 S1 domain-Fc
HKU5 S1 domain-Fc
SARS-CoV S1 domain-Fc
DPP4-His6
S1 domain-Fc
C
hDPP4/HEK293T
bDPP4/HEK293T
Relative pseudovirus entry (%)
120
100
80
60
40
20
0
MERS-CoV S1 domain
HKU4 S1 domain
SARS-CoV S1 domain
HKU5 S1 domain
0
2
20
200
Protein concentration (μg/ml)

Figure 2.6 Binding interactions between HKU4 spike and DPP4.
(A) AlphaScreen assay showing the direct binding interactions between the coronavirus spike S1 domains and hDPP4 or bDPP4. Binding affinity was characterized as AlphaScreen counts. Error bars indicate SEM (two-tailed t test, *P < 0.05, ***P < 0.001; n = 3). (B) Pull-down assay showing the direct binding interactions between the coronavirus S1 domains and hDPP4 or bDPP4. His6-tagged hDPP4 or bDPP4 was incubated with each of the Fc-tagged coronavirus S1 domains. The S1 domain/DPP4 complex was then precipitated with nickel-nitrilotriacetic acid beads. Binding affinity was characterized as Western blotting on the precipitated Fc-tagged coronavirus S1 domains. (C) Inhibition of MERS-CoV-spike-mediated pseudovirus entry by the HKU4 S1 domain. HEK293T cells expressing hDPP4 or bDPP4 were preincubated with gradient concentrations of purified coronavirus S1 domains and then infected by MERS-CoV-spike-packaged pseudoviruses expressing luciferase. The pseudovirus entry efficiency was characterized as luciferase activity accompanying the entry and normalized relative to the entry in the absence of any coronavirus S1 domain. Error bars indicate SEM (n = 3).

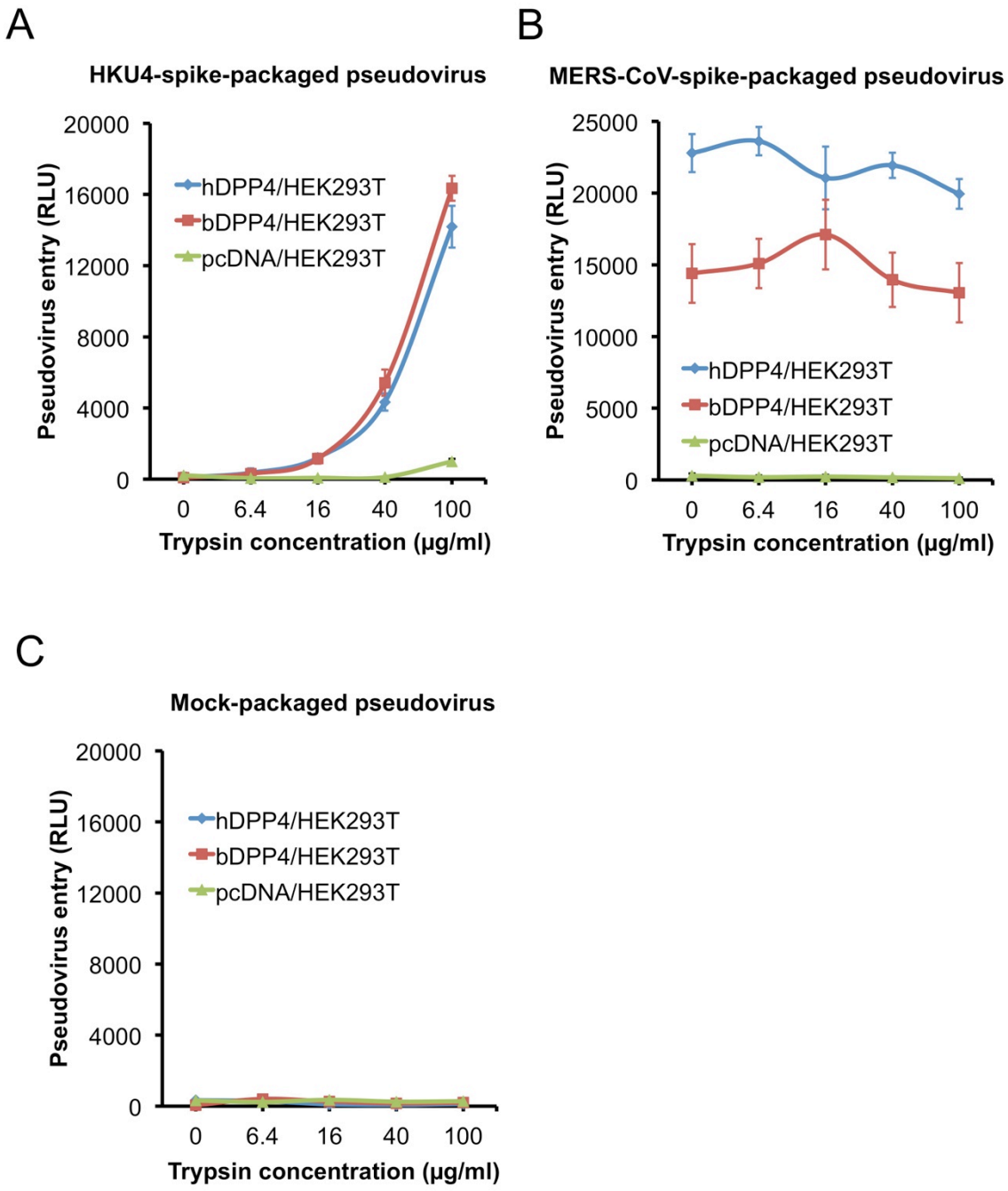
A
HKU4-spike-packaged pseudovirus
Pseudovirus entry (RLU)
20000
16000
12000
8000
4000
0
hDPP4/HEK293T
bDPP4/HEK293T
pcDNA/HEK293T
0
6.4
16
40
100
Trypsin concentration (μg/ml)
B
MERS-CoV-spike-packaged pseudovirus
Pseudovirus entry (RLU)
25000
20000
15000
10000
5000
0
hDPP4/HEK293T
bDPP4/HEK293T
pcDNA/HEK293T
0
6.4
16
40
100
Trypsin concentration (μg/ml)
C
Mock-packaged pseudovirus
Pseudovirus entry (RLU)
20000
16000
12000
8000
4000
0
hDPP4/HEK293T
bDPP4/HEK293T
pcDNA/HEK293T
0
6.4
16
40
100
Trypsin concentration (μg/ml)

Figure 2.7 HKU4- and MERS-CoV-spike-mediated pseudovirus entry into HEK293T cells.
Retroviruses pseudotyped with HKU4 spike (A), MERS-CoV spike (B), or no spike (mock) (C) were incubated with gradient concentrations of trypsin and used to enter HEK293T cells (exogenously expressing hDPP4, bDPP4, or no DPP4, respectively). The pseudovirus entry efficiency was characterized as luciferase activity accompanying the entry. Error bars indicate SEM (n = 4).

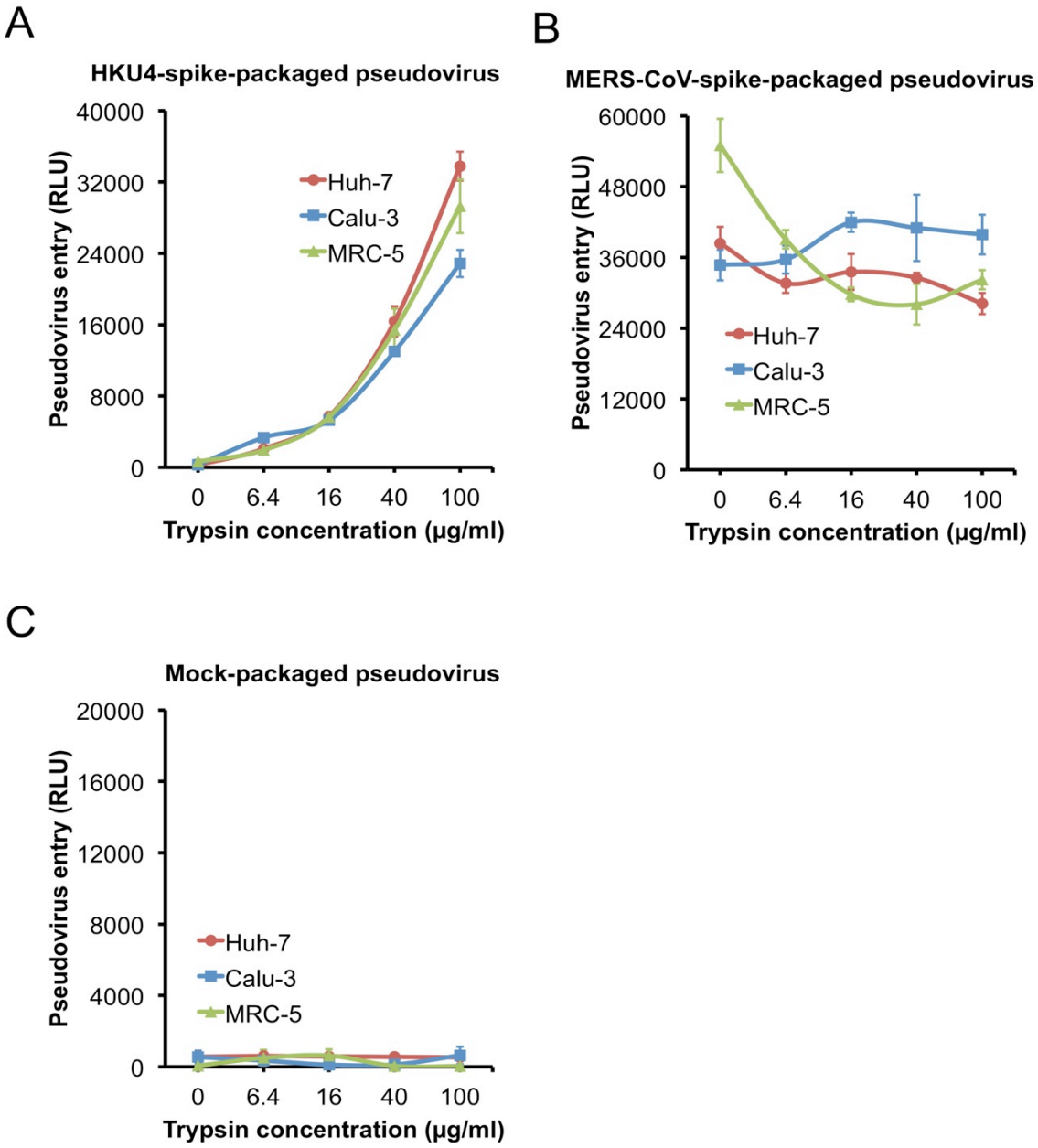

Figure 2.8 HKU4-spike-mediated pseudovirus entry into human cells endogenously expressing DPP4.
Retroviruses pseudotyped with HKU4 spike (A), MERS-CoV spike (B), or no spike (mock) (C) were incubated with gradient concentrations of trypsin and used to enter Huh-7 cells, Calu-3 cells, or MRC-5 cells. The pseudovirus entry efficiency was characterized as luciferase activity accompanying the entry. Error bars indicate SEM (n = 4)

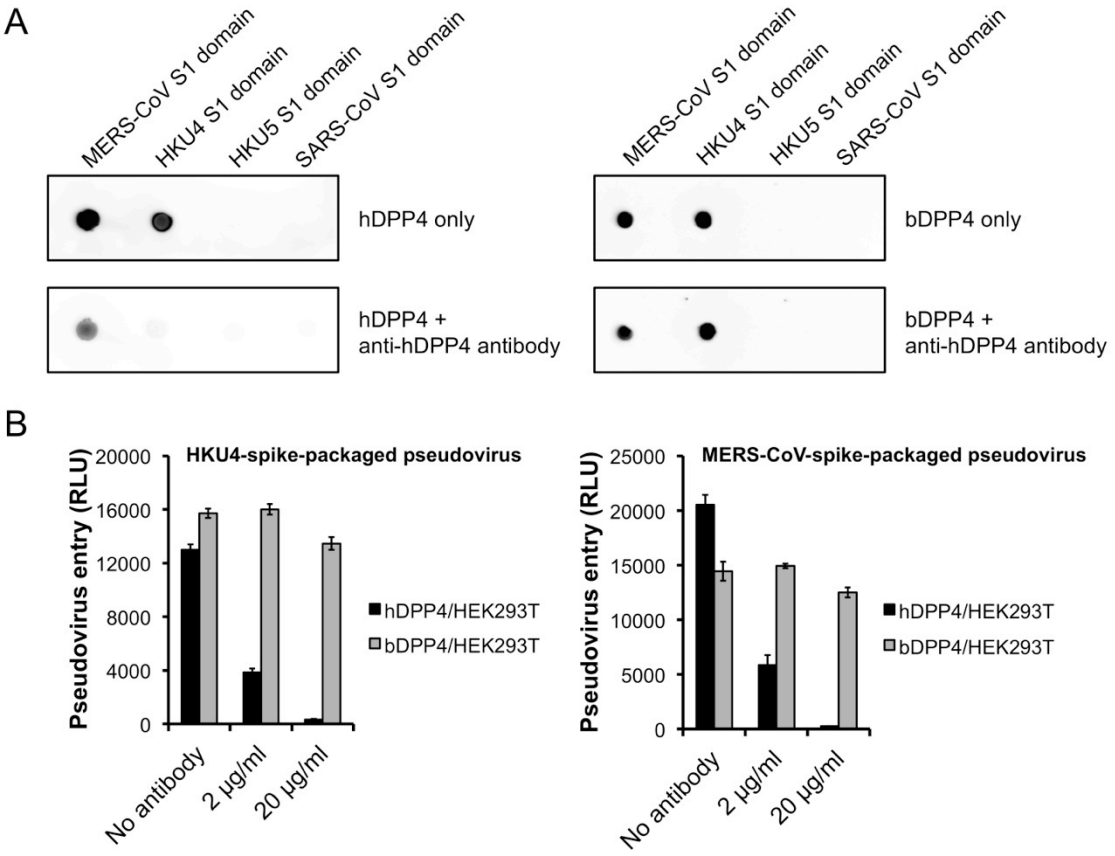

Figure 2.9 Anti-hDPP4 antibodies competitively block the interactions between HKU4 spike and hDPP4.

(A) Dot blot hybridization assay showing the competitive inhibition of HKU4 S1 domain/hDPP4 binding by anti-hDPP4 polyclonal antibodies. The Fc-tagged coronavirus S1 domain was dotted to the membrane first, and His_6-tagged hDPP4 or bDPP4 was added later to the membrane after preincubation alone or with anti-hDPP4 antibodies. Binding affinity was characterized by the ensuing enzymatic color reactions. (B) Neutralization of HKU4-spike-mediated pseudovirus entry by anti-hDPP4 antibodies. Anti-hDPP4 anti- bodies strongly neutralized HKU4-spike-mediated pseudovirus entry into hDPP4-expressing HEK293T cells after treatment with 100 μg/ml trypsin. The antibodies also strongly neutralized MERS-CoV-spike-mediated pseudovirus entry into hDPP4-expressing HEK293T cells without trypsin treatment. The anti- bodies did not neutralize pseudovirus entry into bDPP4-expressing HEK293T cells. Error bars indicate SEM (n = 4).

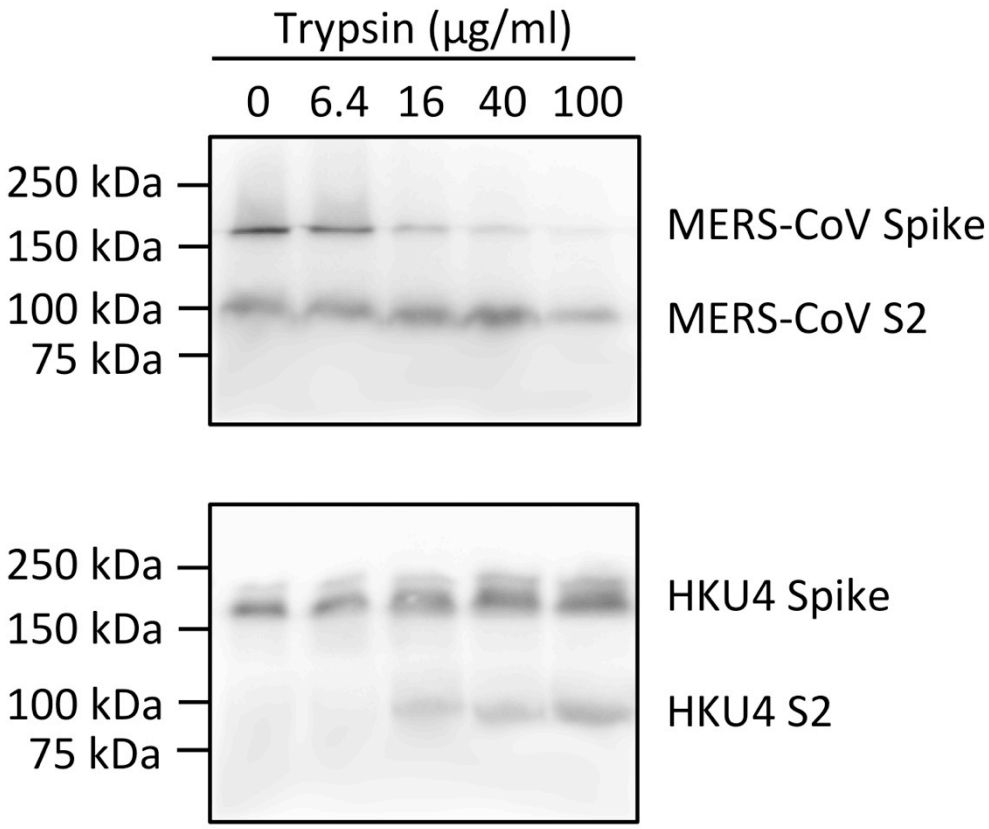

Figure 2.10 Protease processing of MERS-CoV-spike- or HKU4-spike-packaged pseudoviruses.
Retroviruses pseudotyped with MERS-CoV spike or HKU4 spike were treated with gradient concentrations of trypsin. The pseudovirus samples were then subjected to Western blot analysis. The His_6-tagged spikes were detected by antibody against their C-terminal His_6 tag.

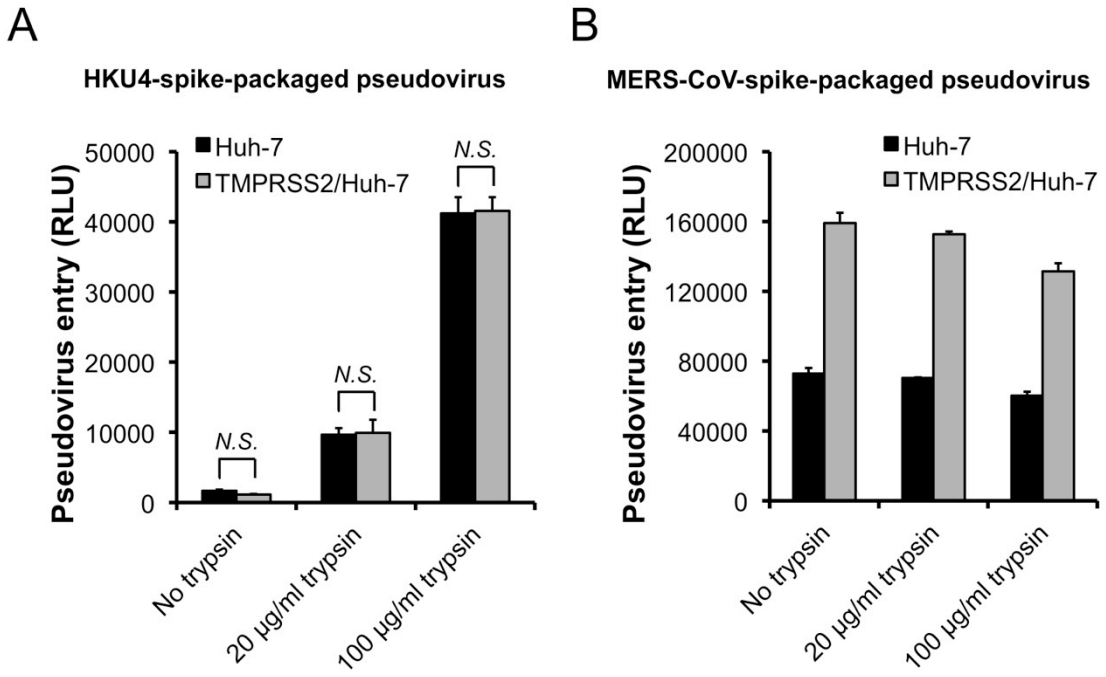

Figure 2.11 Effects of human TMPRSS2 on HKU4- or MERS-CoV-spike-mediated pseudovirus entry into human cells.
Retroviruses pseudotyped with HKU4 spike or MERS-CoV spike were pretreated with no trypsin, 20 μg/ml trypsin, or 100 μg/ml trypsin and used to enter Huh-7 cells or Huh-7 cells exogenously expressing human TMPRSS2. The pseudovirus entry efficiency was characterized as luciferase activity accompanying the entry. Error bars indicate SEM (two-tailed *t* test, *N.S.*, $P > 0.05$; $n = 4$).

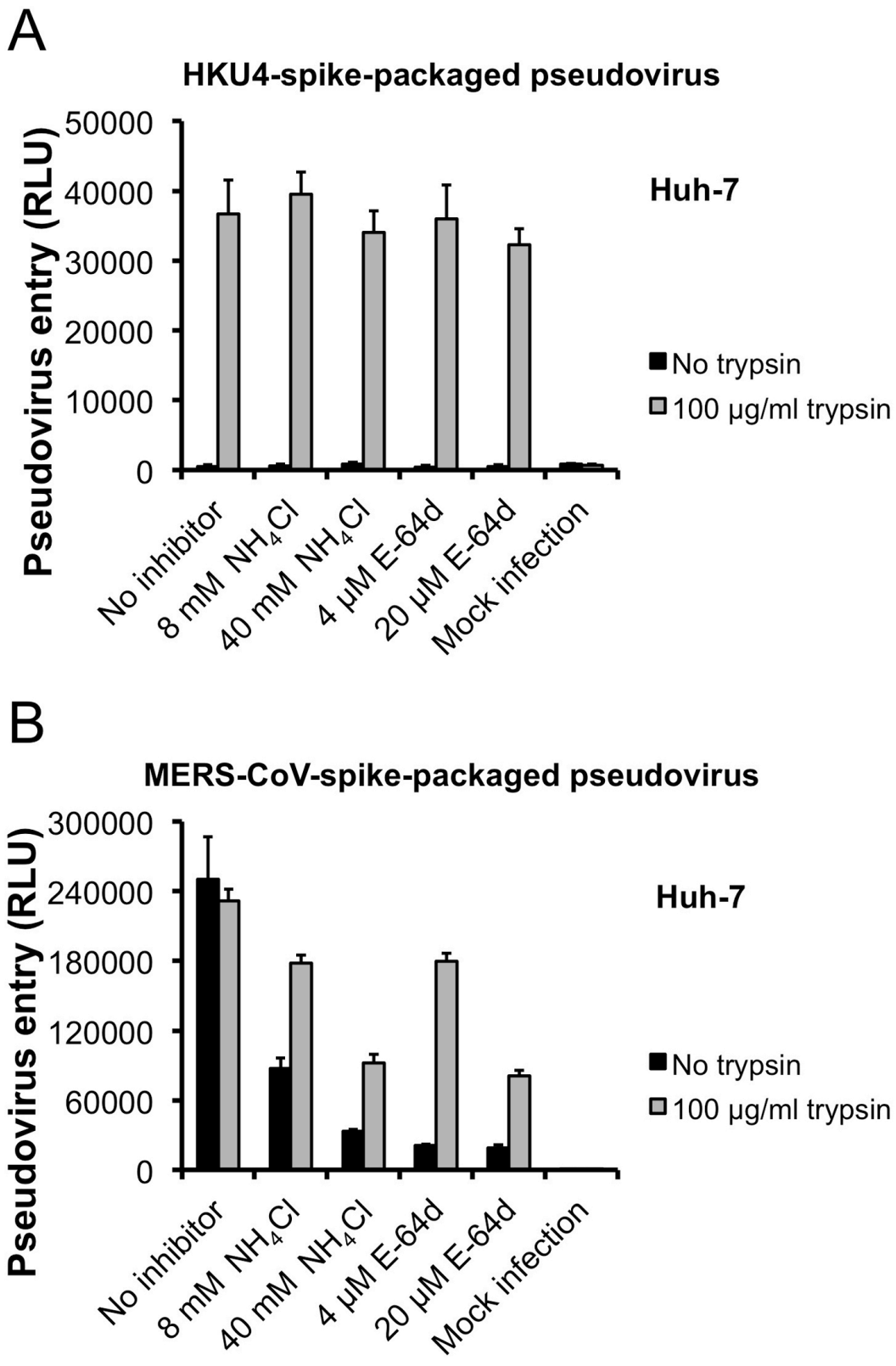
A
HKU4-spike-packaged pseudovirus
Pseudovirus entry (RLU)
50000
40000
30000
20000
10000
0
Huh-7
No trypsin
100 μg/ml trypsin
No inhibitor
8 mM NH4Cl
40 mM NH4Cl
4 μM E-64d
20 μM E-64d
Mock infection
B
MERS-CoV-spike-packaged pseudovirus
Pseudovirus entry (RLU)
300000
240000
180000
120000
60000
0
Huh-7
No trypsin
100 μg/ml trypsin
No inhibitor
8 mM NH4Cl
40 mM NH4Cl
4 μM E-64d
20 μM E-64d
Mock infection

Figure 2.12 Role of human lysosomal proteases in HKU4- and MERS-CoV-spike-mediated entry into human cells.
Huh-7 cells were first preincubated with lysosomal acidification inhibitor NH_4Cl or lysosomal protease inhibitor E-64d at the indicated concentrations. Then the cells were infected by HKU4- or MERS-CoV-spike-pseudotyped retroviruses that had been pretreated or not pretreated with 100 μg/ml trypsin. The pseudovirus entry efficiency was characterized as luciferase activity accompanying the entry. Error bars indicate SEM (n = 3).

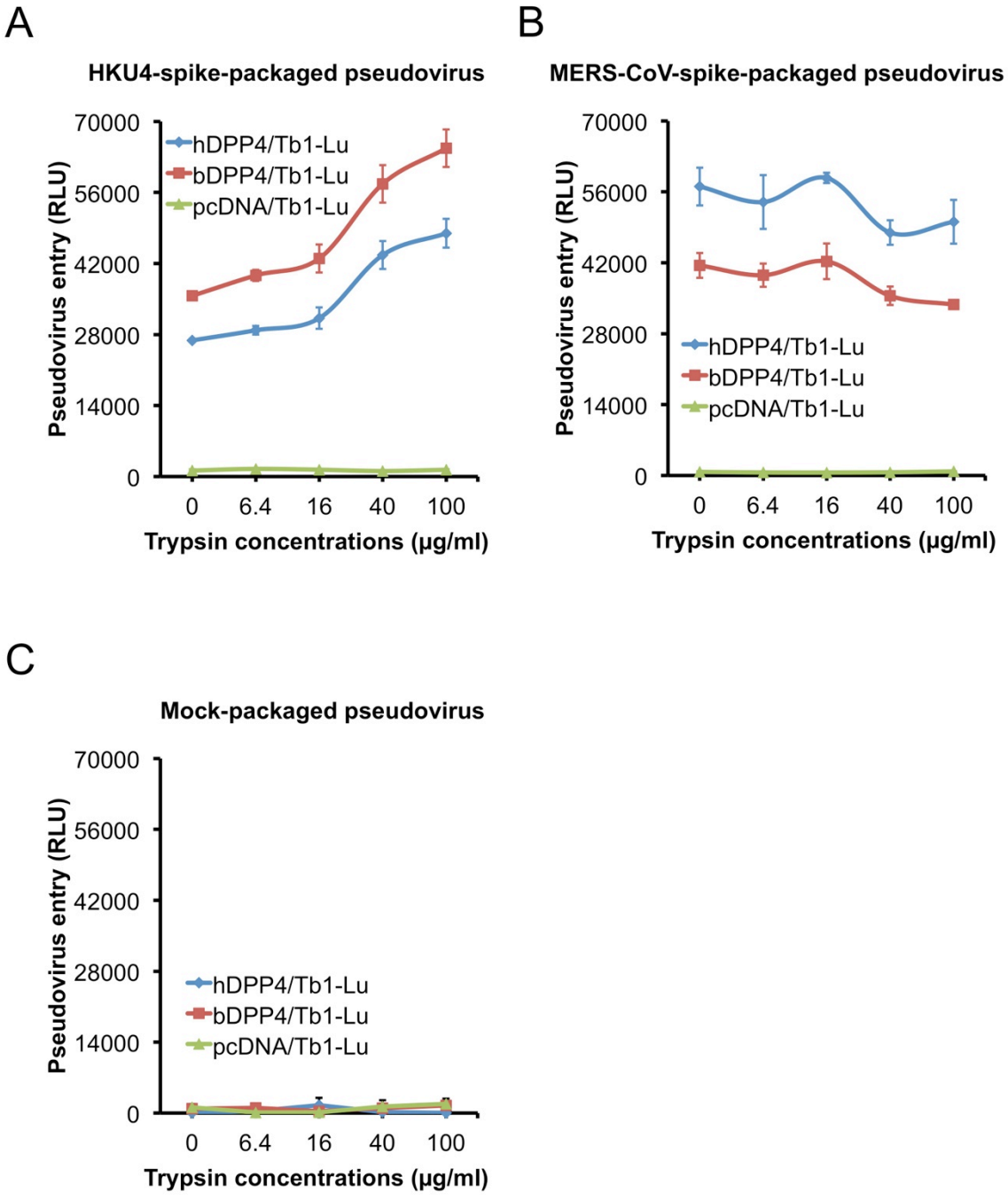
A
HKU4-spike-packaged pseudovirus
Pseudovirus entry (RLU)
70000
56000
42000
28000
14000
0
hDPP4/Tb1-Lu
bDPP4/Tb1-Lu
pcDNA/Tb1-Lu
0
6.4
16
40
100
Trypsin concentrations (μg/ml)
B
MERS-CoV-spike-packaged pseudovirus
Pseudovirus entry (RLU)
70000
56000
42000
28000
14000
0
hDPP4/Tb1-Lu
bDPP4/Tb1-Lu
pcDNA/Tb1-Lu
0
6.4
16
40
100
Trypsin concentrations (μg/ml)
C
Mock-packaged pseudovirus
Pseudovirus entry (RLU)
70000
56000
42000
28000
14000
0
hDPP4/Tb1-Lu
bDPP4/Tb1-Lu
pcDNA/Tb1-Lu
0
6.4
16
40
100
Trypsin concentrations (μg/ml)

Figure 2.13 HKU4- and MERS-CoV-spike-mediated pseudovirus entry into bat cells.
Retroviruses pseudotyped with HKU4 spike (A), MERS-CoV spike (B), or no spike (mock) (C) were incubated with gradient concentrations of trypsin and used to enter Tb1-Lu bat cells (exogenously expressing hDPP4, bDPP4, or no DPP4, respectively). The pseudovirus entry efficiency was characterized as luciferase activity accompanying the entry. Error bars indicate SEM (n = 4).

	MERS-RBD native	NaI derivative
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	45.361, 108.065, 124.287	45.822, 108.817, 124.328
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50-2.13 (2.17-2.13)	50-2.32 (2.36-2.32)
R_{sym} or R_{merge}	0.065 (0.459)	0.119 (0.574)
$I/\sigma I$	31.8 (2.0)	18.3 (2.0)
Completeness (%)	97.8 (62.3)	98.5 (90.5)
Redundancy	6.9 (4.3)	4.3 (2.7)
Refinement		
Resolution (Å)	49.6-2.13	
No. reflections	34770	
R_{work}/R_{free}	0.149 / 0.206	
No. atoms	3701	
Protein	3261	
Ligand	120	
Water	320	
B-factors (Å^2)	43.3	
Protein	40.1	
Ligand	104.3	
Water	53.1	
R.m.s derivations		
Bond lengths (Å)	0.008	
Bond angles (°)	1.33	

Table 2.1 Data collection and refinement statistics of MERS-CoV receptor-binding-domain.
Values in parentheses are for the highest-resolution shell.

Chapter Three

Two mutations were critical for bat-to-human transmission of MERS coronavirus

Yang Yang*, Chang Liu*, Lanying Du, Shibo Jiang, Zheng-Li Shi, Ralph S Baric, Fang Li (co-first authors)

J Virol 89: 9119-9123 (2015).

3.1 Introduction

Coronaviruses infect host cells through the activities of a viral-envelope-anchored spike protein[94,95]. The coronavirus spike is a member of the "class I" viral membrane fusion proteins that also include the virus-surface spike proteins from Ebola virus, influenza virus, and HIV [96-99]. The ectodomain of the coronavirus spike consists of a receptor-binding subunit S1 and a membrane-fusion subunit S2 (Fig. 3.1). The coronavirus spike is synthesized as a single chain precursor, and forms a clove-shaped trimer during the molecular maturation process[100,101]. To infect host cells, S1 binds to a host receptor for cell attachment[102], and S2 subsequently undergoes a dramatic conformational change to fuse the viral and host membranes[72-76]. As I described in Chapter 2, the conformational change of S2 requires proteolytic cleavage at the S1/S2 boundary by one or more host proteases. The availability of these host proteases to coronaviruses is a critical determinant of viral tropism and pathogenesis[72-76].

Bats are considered the natural reservoir of Middle East respiratory syndrome coronavirus (MERS-CoV) because a number of coronaviruses, including HKU4, have been isolated from bats and are genetically closely related to MERS-CoV[43-45,103]. MERS-CoV transmitted from bats to humans either directly or indirectly through intermediate hosts camels[90,91]. In the previous chapter, I have showed that HKU4 and MERS spikes both recognize host receptor dipeptidyl peptidase 4 (DPP4) for viral attachment to host cells[104]. Furthermore, only MERS-CoV spike, but not HKU4 spike, mediates viral entry into human cells because the former, but not the latter, can be activated by human

endogenous proteases (i.e. human proprotein convertases (hPPC), cell-surface proteases, and lysosomal cysteine proteases (LCP)). Here I further identified two residue differences between the two spikes that account for their different capability to mediate viral entry into human cells. These results have revealed a likely evolutionary pathway for the emergence of MERS-CoV as a human pathogen.

3.1 Materials and Methods

Cell lines and plasmids

HEK293T, MRC-5, Tb1-Lu and Huh-7 cells were cultured as described in Chapter 2. RSKT (*Rhinolophus sinicus* bat kidney) cell line was established in-house[105], and cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 100 units/ml penicillin, and 100 μg/ml streptomycin (Life Technologies).

The genes of MERS-CoV spike (GenBank accession number AFS88936.1) with a C-terminal C9 tag, HKU4 spike (GenBank accession number ABN10839.1) with a C-terminal C9 tag, human DPP4 (GenBank accession number NP_001926.2) with a C-terminal HA tag, and bat DPP4 (GenBank accession number KC249974) with a C-terminal HA tag were each cloned into pcDNA3.1(+) vector (Life Technologies). Mutations in MERS-CoV spike and HKU4 spike were carried out using site-directed mutagenesis and validated by automated DNA sequencing (University of Minnesota Genomics Center).

Pseudovirus production

MERS-CoV- or HKU4-spike-pseudotyped retroviruses were generated as previously described[104]. Briefly, HEK293T cells were co-transfected with a plasmid carrying Env-defective, luciferase-expressing HIV-1 genome (pNL4-3.luc.R-E-) and a plasmid encoding MERS-CoV spike or HKU4 spike using Lipofectamine 3000 reagent (Life Technologies) according to the manufacturer's instructions. Supernatants containing pseudoviruses were harvested 72 h after transfection, and centrifuged at 1200 × g for 10 min to remove cell debris.

Western blot analysis of HKU4 and MERS-CoV pseudoviruses

HKU4 and MERS-CoV pseudoviruses were packaged in HEK293T cells as describe above. Seventy-two hours after transfection, retroviruses pseudotyped with MERS-CoV or HKU4 spike were harvested and concentrated using Amicon Ultracentrifuge filter units with a 100-kDa cutoff (EMD Millipore). The pseudovirus particles were then subjected to Western blot analysis. The C9-tagged spikes were detected using anti-C9 tag monoclonal antibody. The expression levels of HKU4 and MERS-CoV spikes in pseudovirus particles were quantified using ImageJ software (National Institutes of Health).

Pseudovirus entry into human and bat cells

MERS-CoV- or HKU4-spike-mediated pseudovirus entry was carried out as previously described[104]. Briefly, retroviruses pseudotyped with wild type or mutant MERS-CoV

spike, wild type or mutant HKU4 spike, or no spike (mock) were used to spin-infect HKE293T cells transiently expressing human DPP4, RSKT cells, Tb1-Lu cells transiently expressing bat DPP4, Huh-7 cells, or MRC-5 cells in 96-well plates at 1200 × g for 2 h at 4°C. The medium was replaced with serum-free DMEM (pre-warmed to 37°C) in the presence or absence of 40 μg/ml TPCK-treated trypsin (Sigma-Aldrich). After incubating at 37°C for 10 min, DMEM supplemented with 75 μg/ml soybean trypsin inhibitor (Sigma-Aldrich) was added to neutralize trypsin. Cells were incubated at 37°C for 12 h, and medium was replaced with fresh DMEM. 48 h later, cells were washed with PBS and lysed. Aliquots of cell lysates were transferred to Optiplate-96 plate (PerkinElmer), followed by addition of luciferase substrate (Promega). Relative luciferase units (RLU) were measured using EnSpire plate reader (PerkinElmer). The expression levels of the spikes, all of which contain a C-terminal C9 tag, were measured by Western blot and used to normalize the entry efficiencies[106].

Inhibition of pseudovirus entry into human and bat cells using protease inhibitors

For proprotein convertase inhibition, HEK293T cells were co-transfected with plasmid pNL4-3.luc.R-E- and a plasmid encoding MERS-CoV spike or HKU4 spike. Five hours post-transfection, medium was replaced with DMEM supplemented with 10 μM or 50 μM dec-RVKR-CMK (Enzo Life Sciences). Pseudoviruses were harvested 72 h after transfection and used to infect HEK293T cells transiently expressing human DPP4.

For lysosomal cysteine protease inhibition, HEK293T cells transiently expressing human DPP4, RSKT cells or Tb1-Lu cells transiently expressing bat DPP4 were pre-incubated with 2 μM or 10 μM E-64d (Sigma-Aldrich) at 37°C for 1 h, and then spin-infected by HKU4-spike-pseudotyped or MERS-CoV-spike-pseudotyped retroviruses at 1200 × g for 2 h at 4°C. The medium was replaced with DMEM (pre-warmed to 37°C) in the presence or absence of E-64d (2 μM or 10 μM). Cells were then incubated at 37°C for 12 h, and medium was replaced with fresh DMEM. 48 h later, cells were lysed and measured for luciferase activity.

Western blot analysis of spike cleavage by lysosomal cysteine proteases

Lysosomal cysteine proteases were prepared as previously described[107-109]. Briefly, HEK293T cells or Tb1-Lu cells were transfected with plasmids encoding MERS-CoV spike or HKU4 spike using Lipofectamine LTX transfection reagent (Life Technologies) according to the manufacturer's instructions. 48 h after transfection, cells were harvested, washed with PBS and lysed by sonication in 100 mM sodium acetate, 1 mM dithiothreitol (DTT), 0.1% Triton X-100, pH 5.0, supplemented with 100 μM camostat mesylate (Sigma-Aldrich), 5 μM pepstatin A (Sigma-Aldrich), 1 mM EDTA (Sigma-Aldrich) and 1 mM EGTA (Sigma-Aldrich) to inhibit serine-, aspartate- and metallo-proteases, respectively. Cell lysates were then incubated at 37°C for 30 min, and subjected to Western blot analysis. The C9-tagged spikes were detected using anti-C9 tag monoclonal antibody (Santa Cruz Biotechnology).

3.2 Results

Identification of two protease motifs in MERS-CoV and HKU4 spikes

By comparing the sequences of MERS-CoV and HKU4 spikes, I identified two regions in MERS-CoV spike that serve as human protease target motifs, but differ from the corresponding regions in HKU4 spike (Fig. 3.1). The first region has been established as a signature motif for human proprotein convertases (hPPC motif)[110,111], but the critical Arg748 in MERS-CoV spike corresponds to Ser746 in HKU4 spike, which deviates from the hPPC motif. The second region has been established as a signature motif for human lysosomal cysteine proteases (hLCP motif)[112,113], but the motif Ala763-Phe764-Asn765-His766) (AFNH) in MERS-CoV spike corresponds to Asn762-Tyr763-Thr764-Ser765 (NYTS) in HKU4 spike, which likely introduces an N-linked glycosylation site and may potentially block the access of human lysosomal cysteine proteases. Therefore, these residue differences in the hPPC and hLCP signature motifs between MERS-CoV and HKU4 spikes may affect the spikes' capability to mediate viral entry into host cells, and thus may hold the key to the host ranges of the two viruses.

Re-engineering two protease sites in HKU4

To evaluate the potential genetic changes required for HKU4 to infect human cells, I re-engineered HKU4 spike, aiming to build its capacity to mediate viral entry into human cells. To this end, I introduced two single mutations, S746R and N762A, into HKU4 spike. The S746R mutation was expected to restores the hPPC signature motif in HKU4 spike, whereas the N762A mutation likely disrupts the N-linked glycosylation site in the

hLCP signature motif in HKU4 spike. To confirm that the S746R mutation restored the hPPC motif, I produced retroviruses pseudotyped with HKU4 spike (hereafter referred to as HKU4 pseudoviruses) in human cells, and showed that HKU4 spike containing the S746R mutation was partially cleaved during the molecular maturation process, whereas wild type HKU4 spike remained intact (Fig. 3.2A). Confirming that the N762A mutation disrupted the N-linked glycosylation site in the hLCP motif was technically challenging because of the large size and heavy glycosylation of HKU4 spike (trimeric HKU4 spike has 78 predicted N-linked glycosylation sites and a total molecular weight of ~530 kDa). Nevertheless, I managed to identify a slight downward shift of HKU4 spike with N762A mutation using Western blot analysis (Fig. 3.2B), consistent with successful removal of the N-linked glycosylation site.

To further confirm removing the N-linked glycosylation site made HKU4 spike cleavable by hLCP, I first expressed HKU4 spike on the surface of HEK293T cells, and prepared hLCP by lysing HEK293T cells and inhibiting the activities of the other human endogenous proteases (serine-, aspartate- and metallo-proteases)[107-109]. Subsequently, I treated HEK293T-cell-surface HKU4 spike with hLCP, and performed Western blot to detect the cleavage state of wild type and mutant HKU4 spikes (Fig. 3.3A). Wild type HKU4 spike remained intact, whereas mutant HKU4 spike bearing the re-engineered hLCP motif underwent significant cleavage. Thus the re-engineered hLCP motif allowed HKU4 spike to be processed by hLCP

HKU4-spike-mediated entry into human cells

I then examined the capability of mutant HKU4 spikes to mediate viral entry into human cells. HKU4 pseudoviruses were packaged in HEK293T cells (human embryonic kidney cells), and were used to infect HEK293T target cells that exogenously express human DPP4 receptor. In the absence of exogenous protease trypsin, HKU4 pseudoviruses bearing either the re-engineered hPPC motif or the re-engineered hLCP motif were able to enter HEK293T cells, whereas HKU4 pseudoviruses bearing both of the re-engineered human protease motifs entered HEK293T cells as efficiently as when activated by exogenous trypsin (Fig. 3.4A). In contrast, wild type HKU4 pseudoviruses failed to enter HEK293T cells. Similar results were obtained when two other human cell lines were used: Huh-7 (human liver cells) and MRC-5 (human lung cells) both of which endogenously express human DPP4 receptor (Fig. 3.5A and B). Therefore, the re-engineered hPPC and hLCP motifs enabled HKU4 spike to be activated by human endogenous proteases, allowing HKU4 pseudoviruses to bypass the need for exogenous proteases to enter human cells. These results reveal that HKU4 spike only needs two single mutations at the S1/S2 boundary to gain the full capacity to mediate viral entry into human cells.

To confirm the functions of the re-engineered human protease motifs in HKU4 spike, I used protease inhibitors to probe the human proteases that activate HKU4-spike-mediated viral entry into human cells. Human proprotein convertase inhibitor, Dec-RVKR-CMK, and lysosomal cysteine protease inhibitor, E-64d, both blocked human cell entry of

HKU4 pseudoviruses bearing the re-engineered hPPC motif and the re-engineered hLCP motif, respectively (Fig. 3.4B). This result verified that the gained capability of HKU4 pseudoviruses to enter human cells was due to the re-engineered human protease motifs in HKU4 spike. Taken together, these data confirm that the two mutations in the re-engineered human protease signature motifs allow HKU4 spike to be cleaved by human proprotein convertases and lysosomal cysteine proteases, respectively.

MERS-CoV-spike-mediated entry into human cells

To investigate the functions of the two human protease motifs in MERS-CoV-spike-mediated viral entry into human cells, I re-engineered MERS-CoV spike in the opposite way, aiming to reduce or eliminate its capability to mediate viral entry into human cells. To this end, I introduced a single mutation R748S and a triple mutation A763N/F764Y/N765T into MERS-CoV spike. The R748S mutation destroyed the hPPC signature motif in MERS-CoV spike (Fig. 3.2A), whereas the triple mutation A763N/F764Y/N765T successfully added an N-linked glycosylation site to the hLCP signature motif in MERS-CoV spike (Fig. 3.2B). When expressed on the surface of HEK293T cells, MERS-CoV spike bearing the mutant hLCP motif resisted digestion when incubated with human lysosomal cysteine proteases (Fig. 3.3B). To prevent the cleavage during the molecular maturation process, MERS-CoV spike also contained the mutant hPPC motif, such that the hLCP motif became the only potential cleavage site in the spike. MERS-CoV pseudoviruses bearing either the mutated hPPC motif or the mutated hLCP motif demonstrated decreased capability to enter three types of human

cells (Fig. 3.4C for HEK293T cells; Fig. 3.5C and D for Huh-7 and MRC-5 cells, respectively). MERS-CoV pseudoviruses bearing both of the mutated human protease motifs were unable to enter human cells. Exogenous trypsin was able to fully rescue the capability of mutant MERS-CoV pseudoviruses to enter human cells. Moreover, PPC inhibitor and LCP inhibitor both blocked human cell entry of wild type MERS-CoV pseudoviruses, but had no effect on human cell entry of MERS-CoV pseudoviruses bearing the mutant hPPC motif or the mutant hLCP motif, respectively (Fig. 3.4D). Thus, the mutations in the hPPC and hLCP motifs together eliminate MERS-CoV-spike mediated viral entry into human cells. These results demonstrate that the two functional human protease motifs in MERS-CoV spike played a critical role in the bat-to-human transmission of MERS-CoV.

HKU4- and MERS-CoV-spike-mediated entry into bat cells

After examining HKU4- and MERS-CoV-spike mediated viral entry into human cells, I investigated how these spikes mediate viral entry into bat cells. Because of the low transfection efficiency of bat cells, I was unable to package pseudoviruses in bat cells. Instead, I packaged HKU4 and MERS-CoV pseudoviruses in HEK293T cells and subsequently performed HKU4- and MERS-CoV-spike-mediated pseudovirus entry into two types of bat cells: RSKT and Tb1-Lu cells (Fig. 3.6). Wild type HKU4 pseudoviruses entered bat cells efficiently, whereas HKU4 pseudoviruses bearing the re-engineered hLCP motif (which removed the N-linked glycosylation site) demonstrated enhanced capability to enter bat cells (Fig. 3.6A and B). On the other hand, wild type MERS-CoV

pseudoviruses entered bat cells efficiently, whereas MERS-CoV pseudoviruses bearing the mutant hLCP motif (which added the N-linked glycosylation site) demonstrated reduced, but still significant, capability to enter bat cells (Fig. 3.6C and D). Moreover, most of HKU4 and MERS-CoV-spike-mediated pseudovirus entry into bat cells could be blocked by the LCP inhibitor, also suggesting that bat lysosomal cysteine proteases activate coronavirus spikes bearing a glycosylated hLCP motif (Fig. 3.6). These results indicate that unlike human lysosomal cysteine proteases, bat lysosomal cysteine proteases are capable of recognizing and cleaving efficiently the hLCP motif containing a glycosylation site. The functional differences between human and bat lysosomal cysteine proteases account for HKU4 spike's capability to mediate viral entry into only bat cells, but not human cells.

3.3 Discussion

Understanding the molecular mechanisms for cross-species transmissions of viruses is critical for evaluating their emerging disease potentials and for preventing and controlling their spread in human populations. In this part of my thesis research, I examined the different cell entry activities of the spike proteins from human-infecting MERS-CoV and a closely related bat coronavirus HKU4. Although MERS-CoV and HKU4 spikes share a high sequence homology and recognize the same host DPP4 receptor, only MERS-CoV spike, but not HKU4 spike, mediates viral entry into human cells. This study revealed that introduction of two single mutations, S746R and N762A, into HKU4 spike at the S1/S2 boundary fully instilled its capability to mediate viral entry into human cells.

MERS-CoV spike already contained both of these mutations, explaining why MERS-CoV is capable of infecting human cells. Thus these two mutations in the spike are essential for MERS-CoV to transmit from bats to humans by allowing MERS-CoV spike to be activated by human cellular proteases. Viral adaptation to human cellular proteases is critical for viral infection of human cells because human cellular proteases, particularly lysosomal proteases, are more reliable sources than some extracellular proteases to activate virus entry. Previous research also identified two mutations in SARS-CoV spike that led SARS-CoV to transmit from palm civets to humans[64,69,70,114]. These mutations increased SARS-CoV spike's capability to bind human receptor angiotensin-converting enzyme 2. Thus, different entry factors appear to have played the most critical roles in the cross-species transmission of MERS-CoV and SARS-CoV: adaption to human cellular proteases by MERS-CoV and adaption to human receptor by SARS-CoV. Although MERS-CoV spike might also need to adapt to human DPP4 receptor upon infecting human cells[104,115], such adaptations might only have incremental effects on the infectivity of MERS-CoV in human cells. In contrast, the two mutations adaptive to human cellular proteases transformed MERS-CoV spike from completely lacking to fully owning the capacity to mediate viral entry into human cells, and thus they likely played the most critical role in the bat-to-human transmission of MERS-CoV.

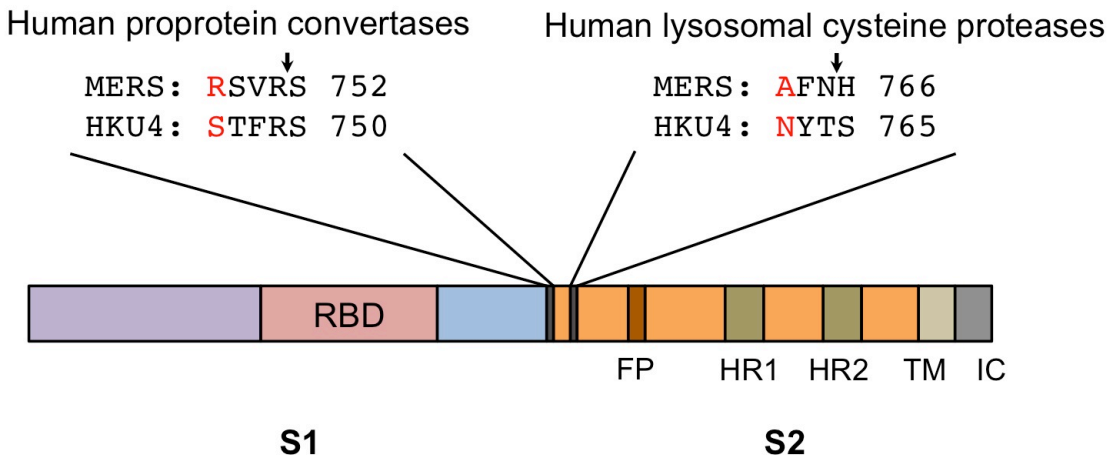

Figure 3.1 Domain structure of MERS-CoV and HKU4 spike proteins.
The spikes contain a receptor-binding S1 subunit, a membrane-fusion S2 subunit, a transmembrane anchor (TM), and an intracellular tail (IC). S1 contains the receptor-binding domain (RBD) that binds DPP4 receptor. S2 contains the fusion peptide (FP), heptad repeat 1 (HR1), and heptad repeat 2 (HR2), all of which are essential structural elements for the membrane fusion process. The S1/S2 boundary in MERS-CoV spike contains two established human protease motifs that are recognized by human proprotein convertases (hPPC) and lysosomal cysteine proteases (hLCP), respectively. Sequence alignments of these regions in MERS-CoV and HKU4 spikes are shown, with the critical residue differences labeled in red. Arrows indicate the cleavage sites in MERS-CoV spike by human proteases.

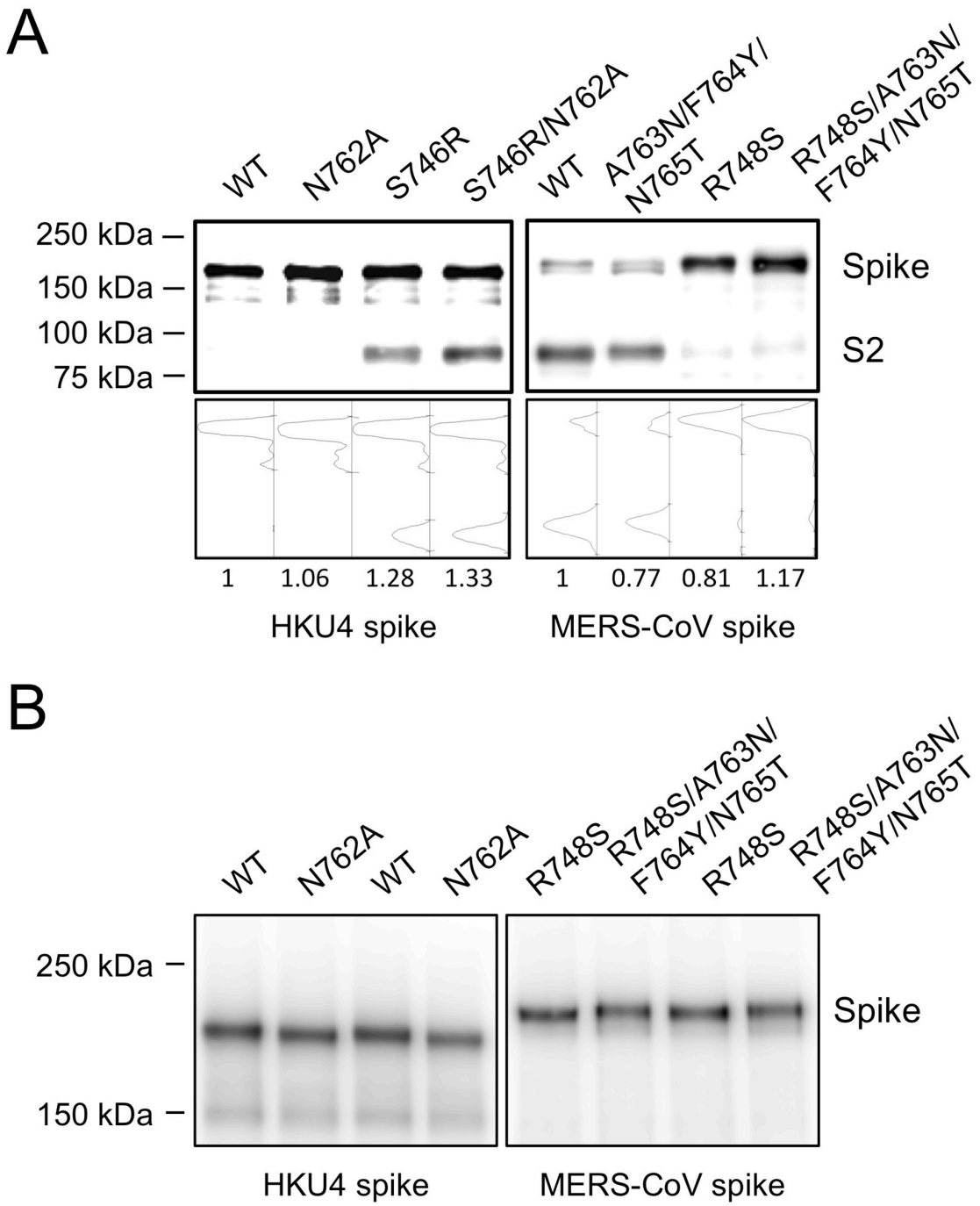
A
WT
N762A
S746R
S746R/N762A
WT
A763N/F764Y/
N765T
R748S
R748S/A763N/
F764Y/N765T
250 kDa
150 kDa
100 kDa
75 kDa
Spike
S2
1
1.06
1.28
1.33
1
0.77
0.81
1.17
HKU4 spike
MERS-CoV spike
B
WT
N762A
WT
N762A
R748S
R748S/A763N/
F764Y/N765T
R748S
R748S/A763N/
F764Y/N765T
250 kDa
150 kDa
Spike
HKU4 spike
MERS-CoV spike

Figure 3.2 Characterization of two protease motifs in MERS-CoV and HKU4 spike proteins.
(A) Western blot analysis of HKU4 and MERS-CoV spikes in pseudovirus particles. The incorporations of wild type (WT) and mutant HKU4 and MERS-CoV spikes into pseudovirus particles were measured by Western blot using antibody against their C-terminal C9 tags. Plots below the Western blot images correspond to quantifications of the band intensities from the cleaved and uncleaved spikes combined. The numbers below the plots indicate the relative amount of spikes incorporated into pseudovirus particles compared to wild type HKU4 and MERS-CoV spikes, respectively. All quantifications were done using ImageJ software (National Institutes of Health). (B) Glycosylation state of HKU4 spike at the hLCP motifs. HEK293T cells exogenously expressing wild type (WT) and mutant HKU4 spikes were lysed and subjected to western blot analysis. To improve the separation of large molecular weight spikes, 3-8% NuPAGE Tris-Acetate Gels (Life Technologies) were used for gel electrophoresis. To improve the accuracy of the result, each of the mutant and wild type spikes was run in two lanes that alternated between samples. Compared with wild type HKU4 spike, the downward shift in the band of HKU4 spike bearing the mutated hLCP motif (i.e., mutation N762A) is consistent with the removal of glycosylation. Compared with MERS-CoV spike bearing the re-engineered hPPC motif (i.e., mutation R748S), the upward shift in the band of MERS-CoV spike bearing both the re-engineered hPPC and hLCP motifs (i.e., mutations R748S/A763N/F764Y/N765T) indicates the introduction of glycosylation.

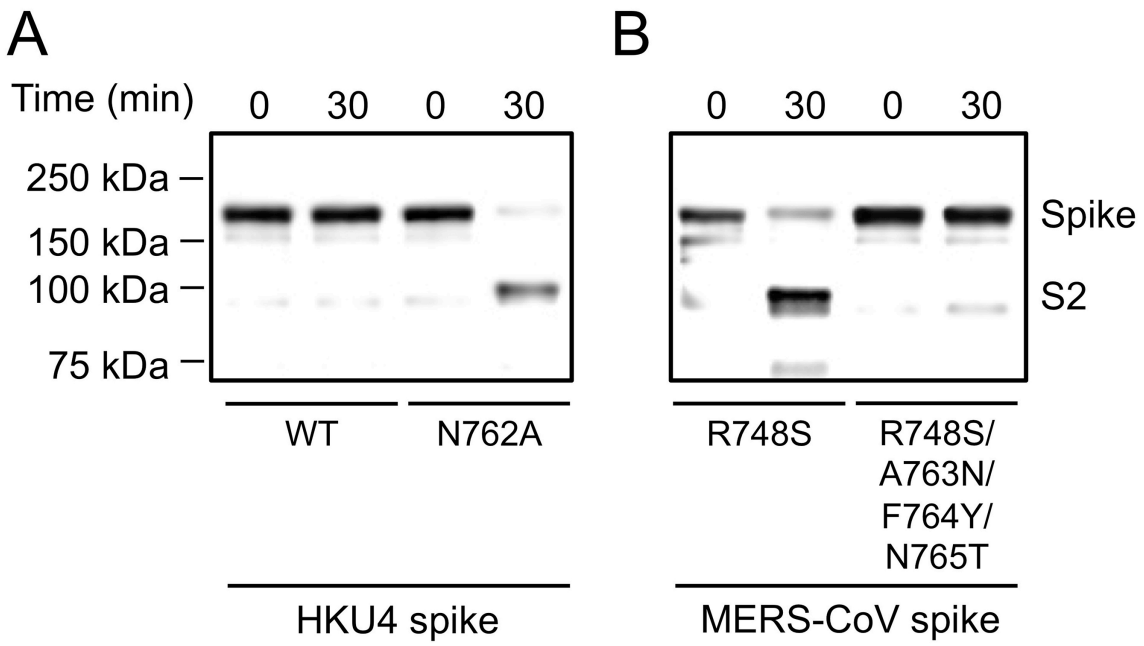

Figure 3.3 Processing of MERS-CoV and HKU4 spikes by human and bat lysosomal cysteine proteases.
(A) HKU4 spike bearing no mutation or the re-engineered hLCP motif (N762A) was transiently expressed on the surface of HEK293T cells, and subsequently incubated with human lysosomal cysteine proteases; (B) MERS-CoV spike bearing either the mutant hPPC motif alone (R748S) or both the mutant hPPC and hLCP motifs (R748S/A763N/F764Y/N765T) was transiently expressed on the surface of HEK293T cells, and subsequently incubated with human lysosomal cysteine proteases; The mutant hPPC motif (R748S) was included in MERS-CoV spike to prevent its cleavage by human proprotein convertases during the molecular maturation process. After incubation with human or bat lysosomal cysteine proteases at pH 5.0 and 37°C for 30 min, both HKU4 and MERS-CoV spikes with a C-terminal C9 tag were subjected to Western blot using anti-C9-tag antibodies.

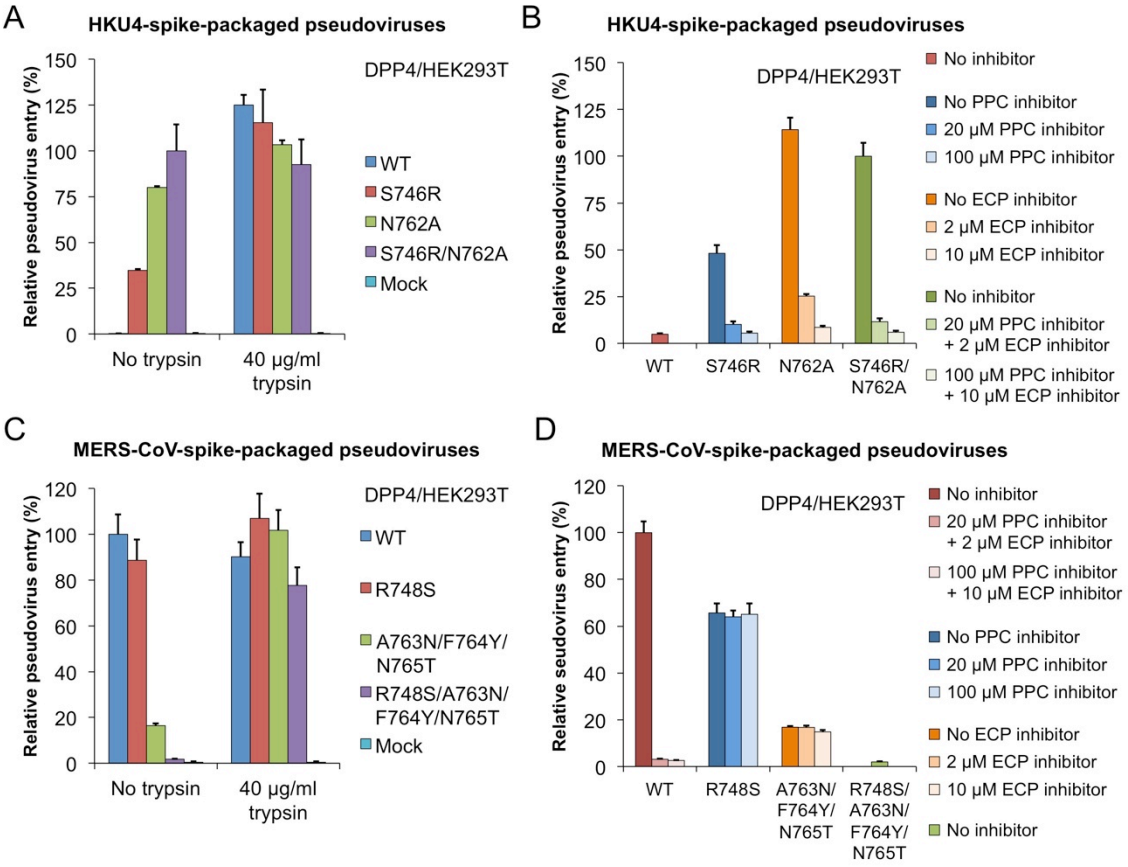
A
HKU4-spike-packaged pseudoviruses
Relative pseudovirus entry (%)
150
125
100
75
50
25
0
No trypsin
40 μg/ml trypsin
DPP4/HEK293T
WT
S746R
N762A
S746R/N762A
Mock
B
HKU4-spike-packaged pseudoviruses
Relative pseudovirus entry (%)
DPP4/HEK293T
WT
S746R
N762A
S746R/ N762A
No inhibitor
No PPC inhibitor
20 μM PPC inhibitor
100 μM PPC inhibitor
No ECP inhibitor
2 μM ECP inhibitor
10 μM ECP inhibitor
No inhibitor
20 μM PPC inhibitor + 2 μM ECP inhibitor
100 μM PPC inhibitor + 10 μM ECP inhibitor
C
MERS-CoV-spike-packaged pseudoviruses
Relative pseudovirus entry (%)
120
100
80
60
40
20
0
No trypsin
40 μg/ml trypsin
DPP4/HEK293T
WT
R748S
A763N/F764Y/ N765T
R748S/A763N/ F764Y/N765T
Mock
D
MERS-CoV-spike-packaged pseudoviruses
Relative seudovirus entry (%)
DPP4/HEK293T
WT
R748S
A763N/ F764Y/ N765T
R748S/ A763N/ F764Y/ N765T
No inhibitor
20 μM PPC inhibitor + 2 μM ECP inhibitor
100 μM PPC inhibitor + 10 μM ECP inhibitor
No PPC inhibitor
20 μM PPC inhibitor
100 μM PPC inhibitor
No ECP inhibitor
2 μM ECP inhibitor
10 μM ECP inhibitor
No inhibitor

Figure 3.4 HKU4 and MERS-CoV-spike-mediated pseudovirus entry into HEK293T cells.
(A) HKU4 pseudoviruses bearing no mutation, re-engineered hPPC motif (S746R), re-engineered hLCP motif (N762A), or both of the re-engineered motifs (S746R/N762A) were used to infect HEK293T cells exogenously expressing human DPP4. The infections were carried out in the presence or absence of exogenous trypsin. The pseudovirus entry efficiencies were characterized as luciferase activity accompanying the entry, and normalized against the relative amount of spikes incorporated into pseudovirus particles (Fig. 3.2A). The pseudovirus entry mediated by HKU4 spike bearing both of the re-engineered motifs was taken as 100%. (B) HEK293T cells were pre-treated with the following protease inhibitors before being infected by different HKU4 pseudoviruses: proprotein convertase inhibitor (dec-RVKR-CMK) for HKU4 pseudoviruses bearing the re-engineered hPPC motif (S746R), lysosomal cysteine protease inhibitor (E-64d) for HKU4 pseudoviruses bearing the re-engineered hLCP motif (N762A), and both of the inhibitors for HKU4 pseudoviruses bearing both of the mutations. (C) MERS-CoV pseudoviruses bearing no mutation, mutated hPPC motif (R748S), mutated hLCP motif (A763N/F764Y/N765T), or both of the mutated motifs (R748S/A763N/F764Y/N765T) were used to infect HEK293T cells exogenously expressing human DPP4. The pseudovirus entry mediated by wild type MERS-CoV spike was taken as 100%. (D) HEK293T cells were pre-treated with the following protease inhibitors before being infected by different MERS-CoV pseudoviruses: proprotein convertase inhibitor (dec-RVKR-CMK) for MERS-CoV pseudoviruses bearing the mutated hPPC motif (R748S), lysosomal cysteine protease inhibitor (E-64d) for MERS-CoV pseudoviruses bearing the mutated hLCP motif (A763N/F764Y/N765T), and both of the inhibitors for wild type MERS-CoV pseudoviruses. Error bars indicate SEM (n = 4).

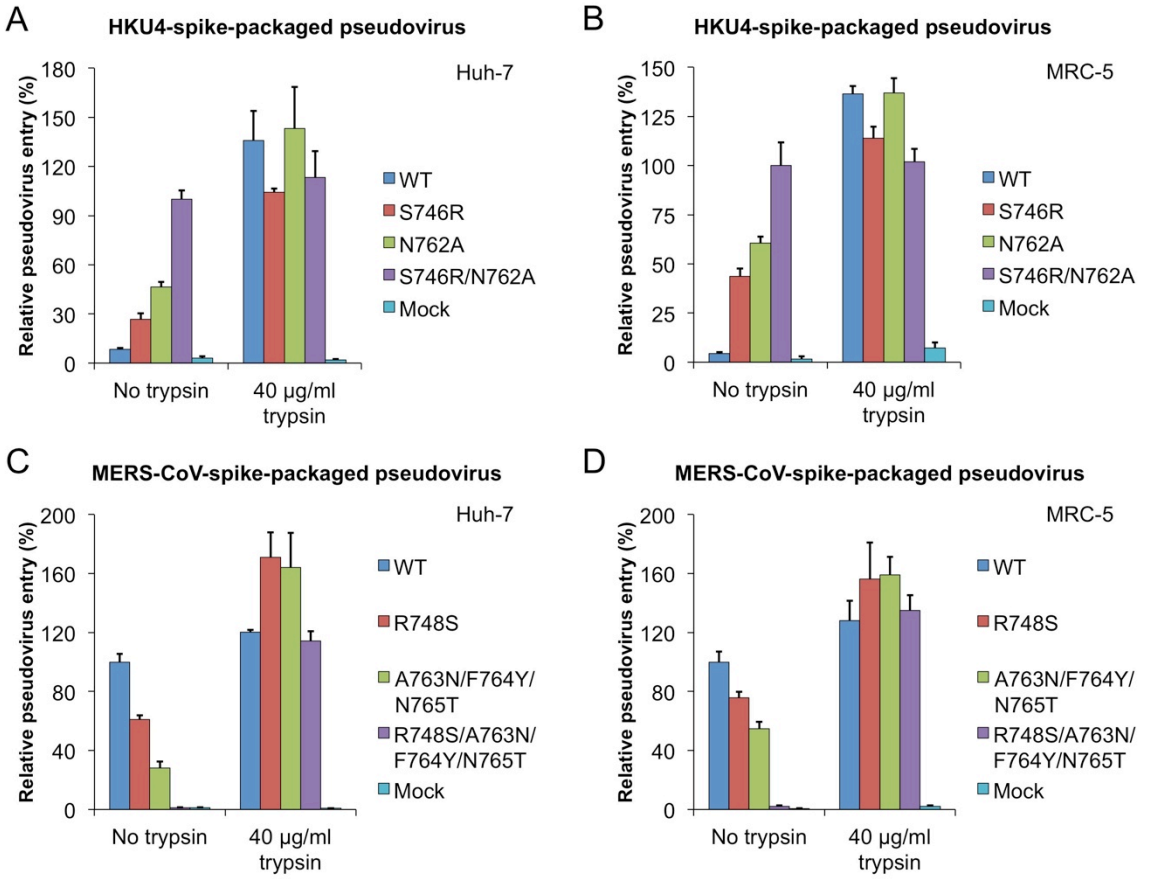

Figure 3.5 HKU4 and MERS-CoV-spike-mediated pseudovirus entry into Huh-7 and MRC-5 cells.

(A) HKU4 pseudoviruses bearing no mutation, re-engineered hPPC motif (S746R), re-engineered hLCP motif (N762A), or both of the re-engineered motifs (S746R/N762A) were used to infect Huh-7 cells endogenously expressing human DPP4. The pseudovirus entry efficiencies were characterized as luciferase activity accompanying the entry, and normalized against the relative amounts of spikes incorporated into pseudovirus particles (Fig. 3.2A). The pseudovirus entry mediated by HKU4 spike bearing both of the re-engineered motifs was taken as 100%. (B) Same as in panel (A), except that MRC-5 cells were used. (C) MERS-CoV pseudoviruses bearing no mutation, mutated hPPC motif (R748S), mutated hLCP motif (A763N/F764Y/N765T), or both of the mutated motifs (R748S/A763N/F764Y/N765T) were used to infect Huh-7 cells endogenously expressing human DPP4. The pseudovirus entry mediated by wild type MERS-CoV spike was taken as 100%. (D) Same as in panel (C), except that MRC-5 cells were used. The infections were carried out in the presence or absence of exogenous trypsin. Error bars indicate SEM (n = 4).

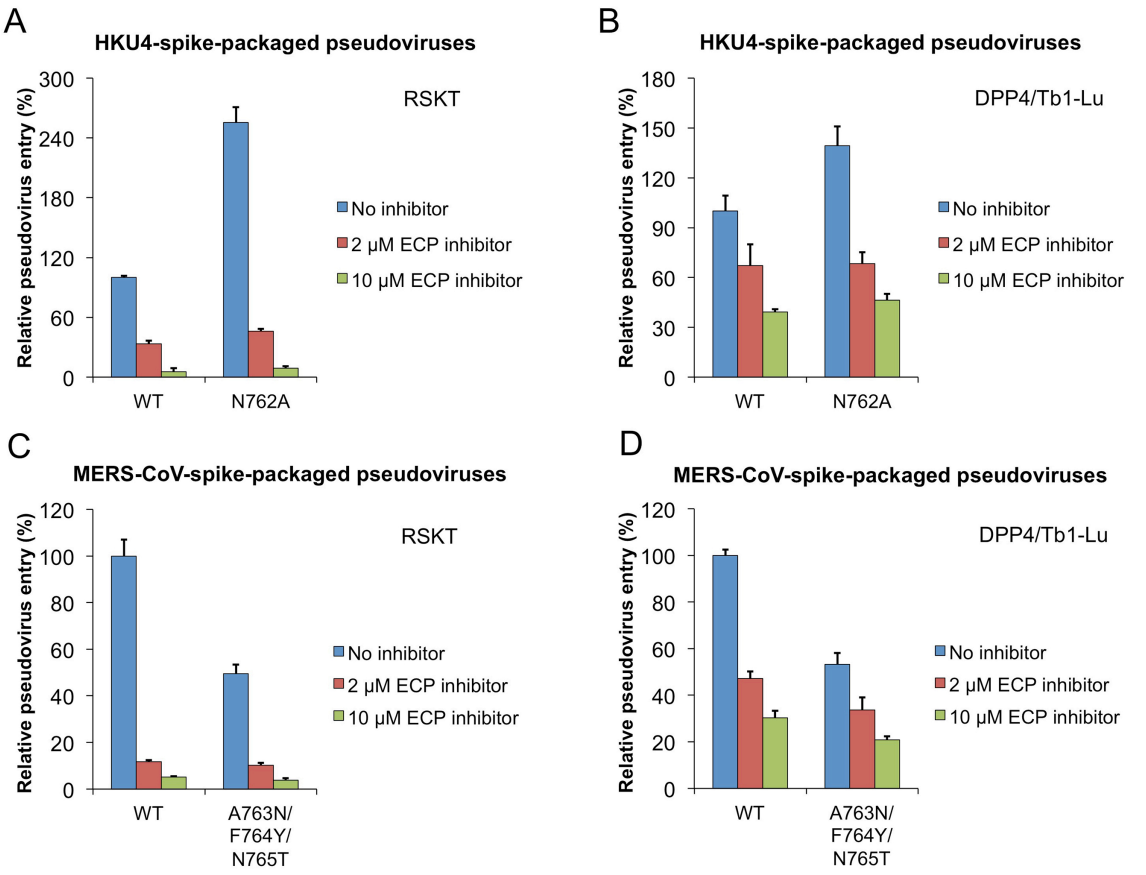

Figure 3.6 HKU4 and MERS-CoV-spike-mediated pseudovirus entry into bat cells. (A) HKU4 pseudoviruses bearing no mutation or re-engineered hLCP motif (N762A) and were prepared in HEK293T cells, and then used to infect RSKT cells endogenously expressing bat DPP4 (Rhinolophus sinicus bat kidney cells). (B) MERS-CoV pseudoviruses bearing no mutation or the mutant hLCP motif (A763N/F764Y/N765T) were prepared in HEK293T cells, and then used to infect RSKT cells. RSKT cells were pre-treated with indicated concentrations of LCP inhibitor (E-64d) before being infected by pseudoviruses. (C) HKU4 pseudoviruses bearing no mutation or re-engineered hLCP motif (N762A) were used to infect Tb1-Lu cells (Tadarida brasiliensis bat lung cells) exogenously expressing bat DPP4 (GenBank accession number KC249974). (D) MERS-CoV pseudoviruses bearing no mutation or the mutant hLCP motif (A763N/F764Y/N765T) were used to infect Tb1-Lu cells. Tb1-Lu cells were pre-treated with LCP inhibitor (E-64d) before being infected by pseudoviruses.
The pseudovirus entry efficiencies were characterized as luciferase activity accompanying the entry, and normalized against the relative amount of spike proteins incorporated into pseudovirus particles (Fig. 3.2A). The pseudovirus entry mediated by wild type HKU4 spike (for panel A and C) or wild type MERS-CoV spike (for panel B and D) in the absence of the inhibitor was taken as 100%. Error bars indicate SEM (n = 4).

Chapter Four

Conclusions

Since the discovery of double helix and deciphering of the genetic code in 1950s and 1960s, DNA was identified as the molecule that carries hereditary traits. It is a widely accepted concept that DNA is the "Secret of Life". But it is protein that embodies this genetic information from DNA and builds up living things. All types of living organisms depend on the existence of proteins. Proteins make it possible for us to see, hear, taste, smell, feel, experience, think and move. Proteins catalyze the life-essential chemical reactions in the human body, and protect us from pathogen attacks. However, proteins that have served their purpose must be degraded so that their constituent amino acids building blocks can be recycled for the synthesis of new proteins. Proteases are the scavengers in protein degradation. They play an important part in the metabolic activity of the body: life would be impossible without them.

By digestion proteins, proteases participate in all aspects of biological activities. They influence DNA replication and transcription, contribute to the processing of cellular information, regulate cell proliferation and differentiation, and affects inflammation and immunity. The body performs digestion at two levels–outside and inside the cell. Digestions outside the cells are accomplished by extracellular proteases and cell surface anchored proteases whose active sites are located in the extracellular space. The concentration of protein in a cell is regulated by a fine balance of synthesis and degradation by lysosomal proteases. Aside from degradation, the other major role of proteases in the body is in the activation and maturation of proteins from their inactive precursors. Proprotein convertases (PPC) are a family of proteases that participate in this

activation process. Cellular proteases sometimes can be hijacked by viruses. With the aid of cellular proteases in host cells, virus can bind to the cells and invade cells efficiently.

My thesis research focused on cellular proteases involved in blood pressure regulation and coronavirus infection. APA, a cell-surface anchored protease, play a pivotal role in blood pressure regulation by regulating the metabolism of vasoactive angiotensin peptides. By solving the atomic structure of APA and elucidating the substrate specificity of APA in the physiological context, my work sheds light on the biological functions of APA, and provides a structural framework for development of APA-targeting drugs for antihypertensive therapy.

Virus infection poses a constant threat to human health. Coronavirus has attracted the attention from scientific community as well as the public in the last decade due to the outbreak of SARS-CoV and MERS-CoV. MERS-CoV utilizes a cell surface protease, DPP4, for cell attachment; and exploits PPC and lysosomal proteases for efficient host cell entry during its infection process. By characterizing the receptor binding and cell entry mechanisms of MERS-CoV and its relatives in bat, this study advocates a possible cross-species transmission route of MERS-CoV, and provides novel insights in therapeutic discoveries for anti-MERS treatment.

Due to their ubiquitous existence in all kinds of tissues and organs, and prevalence in multiple biological processes, proteases are a major focus of attention for the

pharmaceutical industry as potential drug targets to treat many life-threating diseases. My thesis research brings new perspectives into understanding the diverse roles of cellular proteases, and contributes to therapeutic discoveries for the treatment of hypertension and virus infection.

Bibliography

1 Kearney, P. M. *et al.* Global burden of hypertension: analysis of worldwide data. *Lancet* **365**, 217-223, doi:10.1016/s0140-6736(05)17741-1 (2005).

2 Calhoun, D. A. *et al.* Resistant hypertension: diagnosis, evaluation, and treatment. A scientific statement from the American Heart Association Professional Education Committee of the Council for High Blood Pressure Research. *Hypertension* **51**, 1403-1419, doi:10.1161/hypertensionaha.108.189141 (2008).

3 Radevski, I. *et al.* Antihypertensive monotherapy with nisoldipine CC is superior to enalapril in black patients with severe hypertension. *American journal of hypertension* **12**, 194-203 (1999).

4 Saunders, E. *et al.* A comparison of the efficacy and safety of a beta-blocker, a calcium channel blocker, and a converting enzyme inhibitor in hypertensive blacks. *Archives of internal medicine* **150**, 1707-1713 (1990).

5 Matchar, D. B. *et al.* Systematic review: comparative effectiveness of angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers for treating essential hypertension. *Annals of internal medicine* **148**, 16-29 (2008).

6 Saine, D. R. & Ahrens, E. R. Renal impairment associated with losartan. *Annals of internal medicine* **124**, 775 (1996).

7 Basso, N., Ruiz, P., Mangiarua, E. & Taquini, A. C. Renin-like activity in the rat brain during the development of DOC-salt hypertension. *Hypertension* **3**, Ii-14-17 (1981).

8 Ganten, D., Hermann, K., Bayer, C., Unger, T. & Lang, R. E. Angiotensin synthesis in the brain and increased turnover in hypertensive rats. *Science (New York, N.Y.)* **221**, 869-871 (1983).

9 Guyenet, P. G. The sympathetic control of blood pressure. *Nature reviews. Neuroscience* **7**, 335-346, doi:10.1038/nrn1902 (2006).

10 Bakris, G., Bursztyn, M., Gavras, I., Bresnahan, M. & Gavras, H. Role of vasopressin in essential hypertension: racial differences. *Journal of hypertension* **15**, 545-550 (1997).

11 Zaman, M. A., Oparil, S. & Calhoun, D. A. Drugs targeting the renin-angiotensin-aldosterone system. *Nature reviews. Drug discovery* **1**, 621-636, doi:10.1038/nrd873 (2002).

12 Oparil, S. & Haber, E. The renin-angiotensin system (first of two parts). *The New England journal of medicine* **291**, 389-401, doi:10.1056/nejm197408222910805 (1974).

13 Zini, S. *et al.* Identification of metabolic pathways of brain angiotensin II and III using specific aminopeptidase inhibitors: predominant role of angiotensin III in the control of vasopressin release. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 11968-11973 (1996).

14 Reaux, A. *et al.* PC18, a specific aminopeptidase N inhibitor, induces vasopressin release by increasing the half-life of brain angiotensin III. *Neuroendocrinology* **69**, 370-376, doi:54439 (1999).

15 Zini, S. *et al.* Inhibition of vasopressinergic neurons by central injection of a specific aminopeptidase A inhibitor. *Neuroreport* **9**, 825-828 (1998).
16 Bodineau, L. *et al.* Orally active aminopeptidase A inhibitors reduce blood pressure: a new strategy for treating hypertension. *Hypertension* **51**, 1318-1325, doi:10.1161/hypertensionaha.107.098772 (2008).
17 Reaux, A. *et al.* Aminopeptidase A inhibitors as potential central antihypertensive agents. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13415-13420 (1999).
18 Fournie-Zaluski, M. C. *et al.* Brain renin-angiotensin system blockade by systemically active aminopeptidase A inhibitors: a potential treatment of salt-dependent hypertension. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7775-7780, doi:10.1073/pnas.0402312101 (2004).
19 Marc, Y. *et al.* Central antihypertensive effects of orally active aminopeptidase A inhibitors in spontaneously hypertensive rats. *Hypertension* **60**, 411-418, doi:10.1161/hypertensionaha.112.190942 (2012).
20 Zini, S. *et al.* Aminopeptidase A: distribution in rat brain nuclei and increased activity in spontaneously hypertensive rats. *Neuroscience* **78**, 1187-1193 (1997).
21 Mizutani, S. *et al.* New insights into the importance of aminopeptidase A in hypertension. *Heart failure reviews* **13**, 273-284, doi:10.1007/s10741-007-9065-7 (2008).
22 Natesh, R., Schwager, S. L., Sturrock, E. D. & Acharya, K. R. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* **421**, 551-554, doi:10.1038/nature01370 (2003).
23 Sielecki, A. R. *et al.* Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5 A resolution. *Science (New York, N.Y.)* **243**, 1346-1351 (1989).
24 Li, L., Wu, Q., Wang, J., Bucy, R. P. & Cooper, M. D. Widespread tissue distribution of aminopeptidase A, an evolutionarily conserved ectoenzyme recognized by the BP-1 antibody. *Tissue antigens* **42**, 488-496 (1993).
25 Benajiba, A. & Maroux, S. Purification and characterization of an aminopeptidase A from hog intestinal brush-border membrane. *European journal of biochemistry / FEBS* **107**, 381-388 (1980).
26 Chen, L., Lin, Y. L., Peng, G. & Li, F. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17966-17971, doi:10.1073/pnas.1210123109 (2012).
27 Rozenfeld, R., Iturrioz, X., Maigret, B. & Llorens-Cortes, C. Contribution of molecular modeling and site-directed mutagenesis to the identification of two structural residues, Arg-220 and Asp-227, in aminopeptidase A. *The Journal of biological chemistry* **277**, 29242-29252, doi:10.1074/jbc.M204406200 (2002).
28 Rozenfeld, R., Iturrioz, X., Okada, M., Maigret, B. & Llorens-Cortes, C. Contribution of molecular modeling and site-directed mutagenesis to the identification of a new residue, glutamate 215, involved in the exopeptidase

specificity of aminopeptidase A. *Biochemistry* **42**, 14785-14793, doi:10.1021/bi034358u (2003).
29 Claperon, C. *et al.* Asp218 participates with Asp213 to bind a Ca2+ atom into the S1 subsite of aminopeptidase A: a key element for substrate specificity. *The Biochemical journal* **416**, 37-46, doi:10.1042/bj20080471 (2008).
30 Goto, Y., Hattori, A., Mizutani, S. & Tsujimoto, M. Asparatic acid 221 is critical in the calcium-induced modulation of the enzymatic activity of human aminopeptidase A. *The Journal of biological chemistry* **282**, 37074-37081, doi:10.1074/jbc.M707251200 (2007).
31 Goto, Y., Hattori, A., Ishii, Y., Mizutani, S. & Tsujimoto, M. Enzymatic properties of human aminopeptidase A. Regulation of its enzymatic activity by calcium and angiotensin IV. *The Journal of biological chemistry* **281**, 23503-23513, doi:10.1074/jbc.M603191200 (2006).
32 Jones, H. C. & Keep, R. F. Brain fluid calcium concentration and response to acute hypercalcaemia during development in the rat. *The Journal of physiology* **402**, 579-593 (1988).
33 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326 (1997).
34 Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta crystallographica. Section D, Biological crystallography* **55**, 849-861 (1999).
35 Terwilliger, T. C. Maximum-likelihood density modification. *Acta crystallographica. Section D, Biological crystallography* **56**, 965-972 (2000).
36 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132, doi:10.1107/s0907444904019158 (2004).
37 Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D, Biological crystallography* **54**, 905-921 (1998).
38 Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta crystallographica. Section D, Biological crystallography* **55**, 247-255, doi:10.1107/s090744499801405x (1999).
39 Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. *Nature* **405**, 647-655, doi:10.1038/35015017 (2000).
40 Findeisen, F. & Minor, D. L., Jr. Structural basis for the differential effects of CaBP1 and calmodulin on Ca(V)1.2 calcium-dependent inactivation. *Structure (London, England : 1993)* **18**, 1617-1631, doi:10.1016/j.str.2010.09.012 (2010).
41 Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. & Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New England Journal of Medicine* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).

42 de Groot, R. J. *et al.* Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *Journal of virology* **87**, 7790-7792, doi:10.1128/jvi.01244-13 (2013).
43 Annan, A. *et al.* Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. *Emerg Infect Dis* **19**, 456-459, doi:10.3201/eid1903.121503 (2013).
44 Holmes, K. V. & Dominguez, S. R. The new age of virus discovery: genomic analysis of a novel human betacoronavirus isolated from a fatal case of pneumonia. *mBio* **4**, e00548-00512, doi:10.1128/mBio.00548-12 (2013).
45 Lau, S. K. *et al.* Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *Journal of virology* **87**, 8638-8650, doi:10.1128/jvi.01055-13 (2013).
46 Ksiazek, T. G. *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *The New England journal of medicine* **348**, 1953-1966, doi:10.1056/NEJMoa030781 (2003).
47 Peiris, J. S. *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319-1325 (2003).
48 Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nature reviews. Microbiology* **7**, 439-450, doi:10.1038/nrmicro2147 (2009).
49 Li, F. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *Journal of virology* **86**, 2856-2858, doi:10.1128/jvi.06882-11 (2012).
50 Hofmann, H. *et al.* Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7988-7993, doi:10.1073/pnas.0409465102 (2005).
51 Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454, doi:10.1038/nature02145 (2003).
52 Delmas, B. *et al.* Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* **357**, 417-420, doi:10.1038/357417a0 (1992).
53 Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357**, 420-422, doi:10.1038/357420a0 (1992).
54 Dveksler, G. S. *et al.* Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. *Journal of virology* **65**, 6881-6891 (1991).
55 Williams, R. K., Jiang, G. S. & Holmes, K. V. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 5533-5536 (1991).

56 Watanabe, R., Matsuyama, S. & Taguchi, F. Receptor-independent infection of murine coronavirus: analysis by spinoculation. *Journal of virology* **80**, 4901-4908, doi:10.1128/jvi.80.10.4901-4908.2006 (2006).
57 Watanabe, R., Sawicki, S. G. & Taguchi, F. Heparan sulfate is a binding molecule but not a receptor for CEACAM1-independent infection of murine coronavirus. *Virology* **366**, 16-22, doi:10.1016/j.virol.2007.06.034 (2007).
58 Cavanagh, D. & Davis, P. J. Coronavirus IBV: removal of spike glycopolypeptide S1 by urea abolishes infectivity and haemagglutination but not attachment to cells. *The Journal of general virology* **67 (Pt 7)**, 1443-1448 (1986).
59 Schultze, B., Cavanagh, D. & Herrler, G. Neuraminidase treatment of avian infectious bronchitis coronavirus reveals a hemagglutinating activity that is dependent on sialic acid-containing receptors on erythrocytes. *Virology* **189**, 792-794 (1992).
60 Schultze, B., Gross, H. J., Brossmer, R. & Herrler, G. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *Journal of virology* **65**, 6232-6237 (1991).
61 Schwegmann-Wessels, C. & Herrler, G. Sialic acids as receptor determinants for coronaviruses. *Glycoconjugate journal* **23**, 51-58, doi:10.1007/s10719-006-5437-9 (2006).
62 Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251-254, doi:10.1038/nature12005 (2013).
63 Li, W. *et al.* Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. *Journal of virology* **80**, 4211-4219, doi:10.1128/jvi.80.9.4211-4219.2006 (2006).
64 Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864-1868, doi:10.1126/science.1116480 (2005).
65 Peng, G. *et al.* Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10696-10701, doi:10.1073/pnas.1104306108 (2011).
66 Peng, G. *et al.* Crystal structure of bovine coronavirus spike protein lectin domain. *The Journal of biological chemistry* **287**, 41931-41938, doi:10.1074/jbc.M112.418210 (2012).
67 Reguera, J. *et al.* Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS pathogens* **8**, e1002859, doi:10.1371/journal.ppat.1002859 (2012).
68 Wu, K., Li, W., Peng, G. & Li, F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19970-19974, doi:10.1073/pnas.0908837106 (2009).
69 Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral research* **100**, 246-254, doi:10.1016/j.antiviral.2013.08.014 (2013).

70 Li, F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *Journal of virology* **82**, 6984-6991, doi:10.1128/jvi.00442-08 (2008).
71 Wu, K., Peng, G., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *The Journal of biological chemistry* **287**, 8904-8911, doi:10.1074/jbc.M111.325803 (2012).
72 Belouzard, S., Millet, J. K., Licitra, B. N. & Whittaker, G. R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4**, 1011-1033, doi:10.3390/v4061011 (2012).
73 Heald-Sargent, T. & Gallagher, T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* **4**, 557-580, doi:10.3390/v4040557 (2012).
74 Simmons, G., Zmora, P., Gierer, S., Heurich, A. & Pohlmann, S. Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. *Antiviral research* **100**, 605-614, doi:10.1016/j.antiviral.2013.09.028 (2013).
75 Simmons, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 11876-11881, doi:10.1073/pnas.0505577102 (2005).
76 Simmons, G. *et al.* Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 4240-4245, doi:10.1073/pnas.0306446101 (2004).
77 Gierer, S. *et al.* The spike protein of the emerging betacoronavirus EMC uses a novel coronavirus receptor for entry, can be activated by TMPRSS2, and is targeted by neutralizing antibodies. *Journal of virology* **87**, 5502-5511, doi:10.1128/jvi.00128-13 (2013).
78 Qian, Z., Dominguez, S. R. & Holmes, K. V. Role of the spike glycoprotein of human Middle East respiratory syndrome coronavirus (MERS-CoV) in virus entry and syncytia formation. *PloS one* **8**, e76469, doi:10.1371/journal.pone.0076469 (2013).
79 Shirato, K., Kawase, M. & Matsuyama, S. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *Journal of virology* **87**, 12552-12561, doi:10.1128/jvi.01890-13 (2013).
80 Grosse-Kunstleve, R. W. & Adams, P. D. Substructure search procedures for macromolecular structures. *Acta crystallographica. Section D, Biological crystallography* **59**, 1966-1973 (2003).
81 Du, L. *et al.* A conformation-dependent neutralizing monoclonal antibody specifically targeting receptor-binding domain in Middle East respiratory syndrome coronavirus spike protein. *Journal of virology* **88**, 7045-7053, doi:10.1128/jvi.00433-14 (2014).

82 Cassel, J. A., Blass, B. E., Reitz, A. B. & Pawlyk, A. C. Development of a novel nonradiometric assay for nucleic acid binding to TDP-43 suitable for high-throughput screening using AlphaScreen technology. *Journal of biomolecular screening* **15**, 1099-1106, doi:10.1177/1087057110382778 (2010).
83 Chen, Y. *et al.* Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *Journal of virology* **87**, 10777-10783, doi:10.1128/jvi.01756-13 (2013).
84 Zhao, G. *et al.* A safe and convenient pseudovirus-based inhibition assay to detect neutralizing antibodies and screen for viral entry inhibitors against the novel human coronavirus MERS-CoV. *Virology journal* **10**, 266, doi:10.1186/1743-422x-10-266 (2013).
85 Dauter, Z., Dauter, M. & Rajashankar, K. R. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta crystallographica. Section D, Biological crystallography* **56**, 232-237 (2000).
86 Tao, X. *et al.* Bilateral entry and release of Middle East respiratory syndrome coronavirus induces profound apoptosis of human bronchial epithelial cells. *Journal of virology* **87**, 9953-9958, doi:10.1128/jvi.01562-13 (2013).
87 Bertram, S. *et al.* TMPRSS2 and TMPRSS4 facilitate trypsin-independent spread of influenza virus in Caco-2 cells. *Journal of virology* **84**, 10016-10025, doi:10.1128/jvi.00239-10 (2010).
88 Tamai, M. *et al.* In vitro and in vivo inhibition of cysteine proteinases by EST, a new analog of E-64. *Journal of pharmacobio-dynamics* **9**, 672-677 (1986).
89 Barlan, A. *et al.* Receptor variation and susceptibility to Middle East respiratory syndrome coronavirus infection. *Journal of virology* **88**, 4953-4961, doi:10.1128/jvi.00161-14 (2014).
90 Reusken, C. B. *et al.* Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *The Lancet. Infectious diseases* **13**, 859-866, doi:10.1016/s1473-3099(13)70164-6 (2013).
91 de Wit, E. & Munster, V. J. MERS-CoV: the intermediate host identified? *The Lancet. Infectious diseases* **13**, 827-828, doi:10.1016/s1473-3099(13)70193-2 (2013).
92 Azhar, E. I. *et al.* Evidence for camel-to-human transmission of MERS coronavirus. *The New England journal of medicine* **370**, 2499-2505, doi:10.1056/NEJMoa1401505 (2014).
93 Chu, D. K. *et al.* MERS coronaviruses in dromedary camels, Egypt. *Emerging infectious diseases* **20**, 1049-1053, doi:10.3201/eid2006.140299 (2014).
94 Bosch, B. J., van der Zee, R., de Haan, C. A. & Rottier, P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *Journal of virology* **77**, 8801-8811 (2003).
95 Spaan, W., Cavanagh, D. & Horzinek, M. C. Coronaviruses: structure and genome expression. *The Journal of general virology* **69 (Pt 12)**, 2939-2952 (1988).

96 Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annual Review of Biochemistry* **69**, 531-569 (2000).
97 Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature* **289**, 366-373 (1981).
98 Eckert, D. M. & Kim, P. S. Mechanisms of viral membrane fusion and its inhibition. *Annual Review of Biochemistry* **70**, 777-810 (2001).
99 Harrison, S. C. Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698, doi:10.1038/nsmb.1456 (2008).
100 Li, F. *et al.* Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *Journal of virology* **80**, 6794-6800, doi:10.1128/jvi.02744-05 (2006).
101 Beniac, D. R., Andonov, A., Grudeski, E. & Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology* **13**, 751-752, doi:10.1038/nsmb1123 (2006).
102 Li, F. Receptor Recognition Mechanisms of Coronaviruses: a Decade of Structural Studies. *J Virol* **89**, 1954-1964, doi:10.1128/jvi.02615-14 (2015).
103 Tang, X. C. *et al.* Prevalence and genetic diversity of coronaviruses in bats from China. *Journal of virology* **80**, 7481-7490, doi:10.1128/jvi.00697-06 (2006).
104 Yang, Y. *et al.* Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci U S A* **111**, 12516-12521, doi:10.1073/pnas.1405889111 (2014).
105 Ge, X. Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-538, doi:10.1038/nature12711 (2013).
106 Wu, K. *et al.* A Virus-Binding Hot Spot on Human Angiotensin-Converting Enzyme 2 Is Critical for Binding of Two Different Coronaviruses. *Journal of Virology* **85**, 5331-5337, doi:10.1128/jvi.02274-10 (2011).
107 Creasy, B. M., Hartmann, C. B., White, F. K. & McCoy, K. L. New assay using fluorogenic substrates and immunofluorescence staining to measure cysteine cathepsin activity in live cell subpopulations. *Cytometry A* **71**, 114-123, doi:10.1002/cyto.a.20365 (2007).
108 Vaithilingam, A. *et al.* A simple methodology to assess endolysosomal protease activity involved in antigen processing in human primary cells. *BMC Cell Biol* **14**, 35, doi:10.1186/1471-2121-14-35 (2013).
109 Barrett, A. J. & Kirschke, H. Cathepsin B, Cathepsin H, and cathepsin L. *Methods Enzymol* **80 Pt C**, 535-561 (1981).
110 Gierer, S. *et al.* Inhibition of proprotein convertases abrogates processing of the middle eastern respiratory syndrome coronavirus spike protein in infected cells but does not reduce viral infectivity. *The Journal of infectious diseases* **211**, 889-897, doi:10.1093/infdis/jiu407 (2015).
111 Millet, J. K. & Whittaker, G. R. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike

protein. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15214-15219, doi:10.1073/pnas.1407087111 (2014).

112 Bosch, B. J., Bartelink, W. & Rottier, P. J. Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *J Virol* **82**, 8887-8890, doi:10.1128/jvi.00415-08 (2008).

113 Biniossek, M. L., Nagler, D. K., Becker-Pauly, C. & Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. *J Proteome Res* **10**, 5363-5373, doi:10.1021/pr200621z (2011).

114 Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *The EMBO journal* **24**, 1634-1643, doi:10.1038/sj.emboj.7600640 (2005).

115 Wang, Q. *et al.* Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell host & microbe* **16**, 328-337, doi:10.1016/j.chom.2014.08.009 (2014).