UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS

Departamento de Biología Molecular

Structural insights into coronavirus binding to host aminopeptidase N and interaction dynamics

Ph.D Thesis

Gaurav Mudgal

Madrid, 2014

UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS

Departamento de Biología Molecular

Structural insights into coronavirus binding to host aminopeptidase N and interaction dynamics

Ph.D Thesis

Gaurav Mudgal

Madrid, 2014

ACKNOWLEDGEMENTS

At first, I would like to thank La Caixa foundation for providing excellent financial support in realizing my dream come true. The National Center of biotechnology, CNB deserves my special gratitude for the state of its state art research facilities, without which the dissertation wouldn't have seen the light of the day. I wish to thank Prof. Dr. Jose Maria Casasnovas for extending his valuable guidance and healthy criticisms throughout the period of the work which has laid the foundation for strategic thinking to my research intellect. I wish to also thank Dr. Juan José Berlanga from the Department of Molecular Biology, UAM to be my tutor. I express my deepest regards to Dr. Cesar Santiago for helping me all through with tools and techniques in structural biology, for encouraging thought provoking discussions and also for lending moral support. I also thank previous researchers this group, Dr. Juan Reguera, Dr. Carolina Ruiz, and Desiderio Ordono, all of whom contributed to the layout on coronavirus and aminopeptidaseN work. I thank Dr. Fernando Almazan and Dr. Luis Enjuanes for extending their guidance and facilities in virology.

I extend my warm regards to Meriem Echbarthi for sharing the doctoral journey in the laboratory, Nuria Cubells for great help with mammalian protein expression and initial how about of biochemical assays, Dr. Angela Ballesteros for inspiring the hope of successful completion of the work, Dr. Rosario Recarcha for being a good listener of my anxiousness to research and a very good person by heart. My company with Dr. Alan Patterson and Tadeo Moreno had worked de-stressors to me and most of the goods of Madrid is what I saw with them. I also thank Dr. Manuela Zonca, Laura, Susana Rodriguez for their unconditional love and care during the vicious topsy-turvy moments. I owe thanks to my mates, Dr. Aneesh Vijayan, and Dr. Abhimanyu Kumar Singh, for sharing their intellects during the coffee moments. I also thank the personnels at the CNB reception who beard with my works till late hours at very important verges of experiments and taking great care of the people.

Last but not the least, I am falling short of words thanking my little family, beloved mother Mrs. Lavang Lata Mudgal, who had been waiting since long to see me back home in India, my wife, Brajesh Mudgal, who gifted me with a child (Madhav) this same year as a bonus to my achievement of the doctorate, besides also packing me lunch boxes daily, and my two brothers Saurav Mudgal and Raghav Mudgal for believing in me.

I thank my father, Late Shri Brij Mohan Mudgal, for putting the strength in me to face the world, for believing in my potentials, and for nurturing the social being in me. Times can't be rewinded, but I believe he is always there in me, in my good deeds.

I dedicate this work to

my father, Late Shri Brij Mohan Mudgal,

and

my mother, Shrimati Lavanglata Mudgal.

ABBREVIATIONS

α	alpha
β	beta
γ	gamma
δ	delta
µ	micro
µg	micrograms
µl	microliter
µM	micromolar
AGE	agarose gel electrophoresis
APN	aminopeptidase N
approx	approximately
APS	ammonium per sulphate
BAC	bacterial artificial chromosome
BHK	baby hamster kidney
BS^3	bis(sulfosuccinimidyl)suberate
BSA	bovine serum albumin
CD	cluster of diffrentiation
cDNA	complementary deoxyribonucleic acid
CHO	chinese hamster ovary
CoV	coronavirus (s)
CPE	cytopathic effect
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
EDTA	ethylene diamine tetra acetate
ELISA	enzyme linked immunosorbent assay
E-MEM	Eagle's minimum essential medium
FACS	fluorescence activated cell sorting
Fc	fragment crystallizable
FCS	fetal calf serum
FITC	fluoresceinisothiocyanate
HA	Hemagglutinin A epitope
HEK	human embryonic kidney
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

hpi	hours post infection
i.e.	that is
ICTV	international committee on taxonomy of viruses
Ig	immunoglobulin
IPTG	isopropyl thiogalactoside
L	liter
LB	Luria-Bertani
LTRI	lower tract respiratory infections
M	molar
mAb	monoclonal antibody
mAU	milli absorbance unit
mg	milligram
ml	milliliter
mm	millimeter
mM	millimolar
MERS	middle east respiratory syndrome
MSX	methionine sulfoxime
PAGE	polyacrylamide gel electrophoresis
PFU	plaque forming units
PRCV	porcine respiratory coronavirus
rmsd	root mean square deviation
RNA	ribonucleic acid
SARS	severe acute respiratory syndrome
SDS	sodium dodecylsulphate
ST	swine testis
TEMED	tetramethylethylenediamine
TGEV	transmissible gastroenteritis virus
w.r.t	with respect to

INDEX

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Coronaviruses (CoV) are large enveloped RNA viruses of animals and humans associated mostly with enteric and respiratory diseases. These viruses were earlier considered more of a veterinary interest and were associated to humans with mild flu. The last decade brought outbreaks with high mortality rates caused by CoV transmission from animals to man. This remarkable cross-species transmission potential is related to CoV adaptation to a variety of cell surface molecules for entry into host cells. The CoV particles bear exposed spike (S) proteins in their envelope that attach to specific cell entry receptors, which determines CoV host cell range and tropism. CoV can recognize diverse entry receptors, but they preferentially use membrane bound ectoenzymes. A subset of CoV recognizes the cell surface aminopeptidase N (APN), a membrane-bound metalloprotease.

APN (CD13) is a "moonlighting" ectoenzyme linked to multiple functions such as angiogenesis, cell-cell adhesion and tumorogenesis. It cleaves neutral amino acid side chains from the N-terminus of oligopeptides, it is distributed in wide variety of tissues its expression is dysregulated in tumors. APN is an important target for cancer therapies and anti-inflammatory drug design, as well as a major CoV entry receptor.

This Thesis presents an extensive structural study on mammalian APN ectodomains as well as its function as CoV receptor. APN crystal structures revealed ectodomain architecture, dimerization and motions important for peptide hydrolysis and CoV recognition. Allosteric inhibition of APN catalysis can be mediated by the suppression of APN movements. Moreover, the crystal structure of a porcine CoV spike fragment bound to the pig APN (pAPN) ectodomain uncovers how CoV bind to its APN receptor. A protruding receptor-binding edge in the S penetrates in small APN cavities, a receptor-binding mode distinct from other CoV-receptor interactions. CoV specifically bind to the open APN conformation. Structure-guided studies identified key virus and receptor motifs at the CoV-APN binding interface and they show that the receptor-binding region is a major antigenic determinant in CoV binding to APN. CoV neutralizing antibodies target key receptor-binding residues, showing that they prevent CoV binding to the APN receptor and infection.

The Thesis provides a compelling view on CoV cell entry and neutralization, as well as important structural insights to understand the multifunctional APN protein.

RESUMEN

Los coronavirus (CoV) son virus ARN con envuelta responsables de enfermedades entéricas y respiratorias en animales y humanos. Estos virus fueron inicialmente considerados de interés veterinario y se asociaron con resfriados en humanos. La última década trajo infecciones de CoV con altas tasas de mortalidad, causadas por la transmisión de CoV animales a humanos. Este potencial de transmisión transversal entre especies está relacionado con la adaptación CoV al uso de diferentes moléculas de la superficie celular para la entrada en la célula huésped. Las partículas CoV contienen espículas (S) expuestas en la envuelta que adhieren las particulas virales a receptores para su penetración en la célula, que determina el tipo de células permisivas para CoV y su tropismo. Los CoV pueden reconocer diversos receptores para la entrada, pero preferentemente utilizan ectoenzimas. Un grupo de CoV reconoce la aminopeptidasa N (APN), una metaloproteasa unida a la membrana celular.

APN (CD13) es una ectoenzima que se ha relacionado múltiples funciones, tales como la angiogénesis, la adhesión célula-célula y la tumorogénesis. Esta proteína hidroliza aminoácidos N-terminales neutros en oligopéptidos, se encuentra en gran variedad de tejidos y su expresión aumenta en tumores. APN es una diana importante de terapias contra el cáncer y de fármacos anti-inflamatorios, así como un receptor para la entrada CoV en la célula.

Esta tesis presenta un extenso estudio estructural sobre los ectodominios de APN de mamíferos, así como sobre su función como receptor de CoV. Las estructuras cristalográficas de la APN revelaron la arquitectura modular de su ectodominio, dimerización y los movimientos del mismo, importantes para la catálisis y el reconocimiento de CoV. Se muestra la posible inhibición alostérica de la actividad catalítics de la APN mediante la supresión de los cambios conformacionales de la APN. Asimismo, la estructura cristalográfica de un fragmento de la espícula de un CoV porcino unido al ectodominio de la APN de cerdo, han identificado cómo los CoV se unen a su receptor. Una región expuesta en la proteína S del CoV penetra en pequeñas cavidades de la APN; este modo de unión al receptor es distinto de otras interacciones CoV-receptor descritas. Los CoV se unen específicamente a la conformación abierta de la APN. Estudios funcionales identificaron motivos clave para la unión del virus y receptor y muestran que la región de unión al receptor es un determinante antigénico en CoV, reconocido por anticuerpos neutralizantes, que inhiben la unión de los CoV a la APN y la infección. La Tesis proporciona una visión general sobre la entrada de CoV en el huésped y su neutralización, así como importantes puntos de vista estructurales para comprender las múltiples funciones de la APN.

CHAPTER 1

1. INTRODUCTION

1.1 Viruses

Viruses are the smallest (with exceptions), most abundant and highly diversified biological entities which can infect specifically (with exceptions) almost every life form, whether plant, bacteria, animals, or archaea. Nonliving, until a suitable living host is chosen, where is when they fit in the definition of a life form, and continue following the propensities of a living civilization i.e. to grow, multiply, and proliferate to long distances.

Viral infection is a complex multi-step phenomenon under which a virus recognizes and attaches to the host, invades and releases its progenies. In most viruses, recognition and attachment is triggered by the interaction between viral anchors and target host cell membrane receptors. Viral tissue tropism and host range are upgraded evolutionarily by '*specificity switching'* of virus-receptor complexes (Stehle and Casasnovas, 2009), which paves the way to more evolved strains with multiple infection mechanisms and broadened host coverage. Due to this ability to self-engineer, which favors their mechanistic remodeling, that the viral outbreaks are inevitably uncontrollable. The situation becomes graver when these smallest biological entities gallop their way through cross-species barriers by virtue of their remodeled tropism backed by specificity-switching. This has kept viruses shameless and medical research communities puzzled.

1.2 Coronaviruses (CoV)

The *Coronaviridae* is a large family of enveloped, positive single stranded RNA viruses with 27-32kb genome which imparts a high degree of genome plasticity and in part, adaptability and diversity to CoV. Spherical coronavirions (120-160 nm across) appear as peculiar crowns due to the presence of spike glycoproteins (Masters, 2006). They have a distinct replication mode that by use of template switching mechanism of transcription produces a set of sub-genomic mRNAs (Masters, 2006).

From human disease view scope, coronavirus studies were initially regarded as 'virology backwaters' (Cavanagh, 2005) and were more of veterinarians´ interest due to fatal diseases in animals associated mostly with meat and dairy supplies or domestics purposes. Human CoV infections form HCoV-229E and HCoV-OC43 only caused mild cough and fever. With the severe acquired respiratory syndrome (SARS) epidemic in 2003, the CoV gathered

spotlights from research communities hailing from almost all realms of disease biology. Following SARS, two more human CoV (HCoV) namely HCoV-HKU1 and HCoV-NL63 were identified. The current scenario projects CoV as important pathogens of animals (including humans) that cause enteric, respiratory, neuronal and/or hepatic diseases incurring heavy economic losses (Perlman, 1998, Weiss and Navas-Martin, 2005 , Masters, 2006, Enjuanes et al., 2008) and high mortalities as has been seen during SARS outbreaks a decade earlier (Perlman and Netland, 2009) and recently from the Middle east respiratory syndrome (MERS) (de Groot et al., 2013).

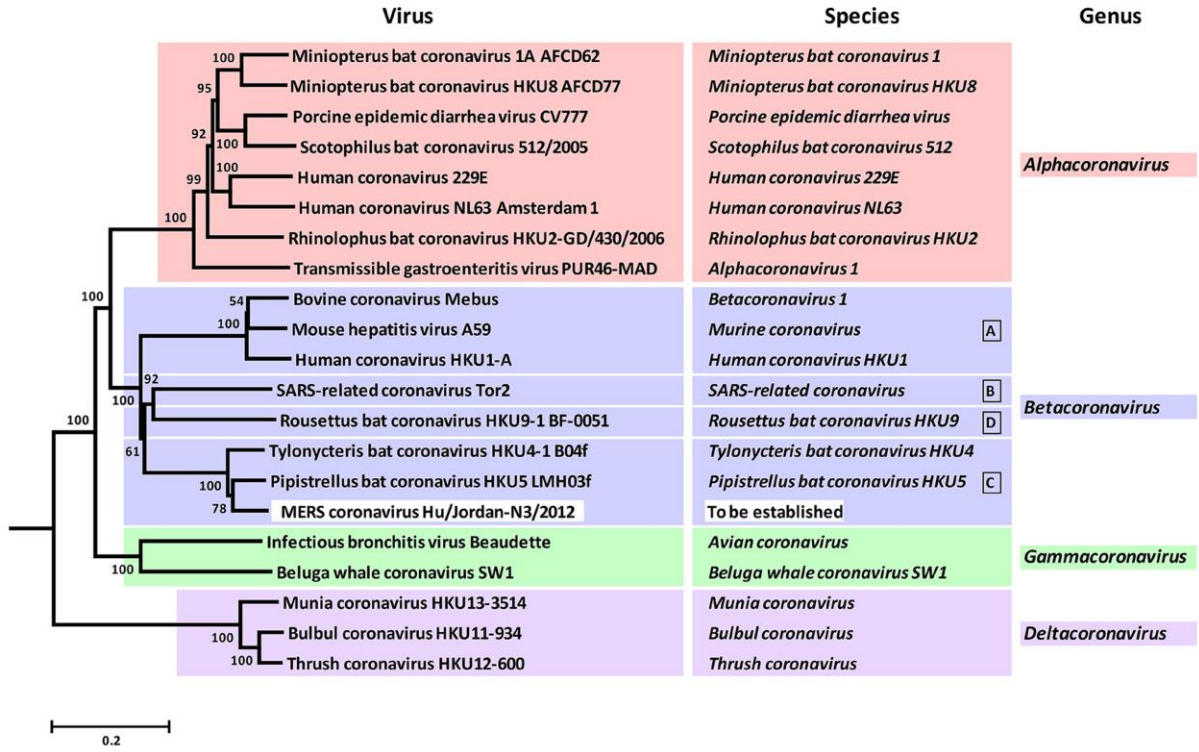


Figure 1.1. Phylogenetic relationship among CoV. A rooted neighbor-joining tree generated from amino acid alignments of *Coronaviridae*-wide conserved domains in *replicase* polyprotein 1ab of 21 CoV. Each of these is representative of a currently recognized CoV species; Equine torovirus Berne served as the out-group. Virus names are given with strain specifications; species and genus names are in italics as per convention. The tree shows the four main color coded monophyletic clusters, respective to genera *Alpha*-, *Beta*-, *Gamma*-, and *Deltacoronavirus.* Also indicated are *Beta*-CoV lineages A through D (corresponding to former CoV subgroups 2A through D). Bootstrap values (1,000 replicates) are indicated at branch points. The tree is drawn to scale (scale bar, 0.2 amino acid substitutions per site). Figure and legend adapted from latest report on Virus taxonomy by ICTV (Groot, 2011)

1.3 CoV Taxonomy

Order *Nidovirales* is shared together by families of coronaviruses, Arteriviruses and Roniviruses. Family *Coronaviridae* (a monophyletic cluster in the order *Nidovirales*) has been categorized in to sub-family *Coronavirinae* and *Torovirinae*. The sub-family *Coronavirinae* has been variously sieved originally into three groups based on serological properties of antigenic cross-reactivities (Lai, 2001).

The recent viral taxonomic convention (Groot, 2011) establishes CoV species on genetic basis in to three approved genera, *Alpha*-, *Beta*- and *Gamma*-CoV with a tentative new member, *Delta*-CoV (Fig 1.1). Demarcations among these genera are based on the rooted phylogeny and calculations of pair-wise evolutionary distances for the conserved domains in the CoV ***replicase*** polyprotein. Currently, complete genomic data are available from around 300 CoV. Accordingly, high ranking clusters corresponding to genus and sub-family levels are recognized. Besides, evidences suggest that both the *alpha*- and *beta*-CoV lineages originate from the bat gene pool (Woo et al., 2007, Lau et al., 2012, Lau et al., 2010). Typical prototypes of each include Transmissible GastroEntritis Virus (TGEV), HCoV-229E, HCoV-NL63 (genus *alpha*-CoV); Mouse Hepatitis Virus (MHV), HCoV-OC43, SARS-CoV, MERS-CoV (*Beta*-CoV) and Infectious Bronchitis Virus (IBV) (*Gamma*-CoV). The new, though officially unapproved genera, *Delta*-CoV has been compiled from findings on CoV in bird species (Groot, 2011).

1.3.1 Genus Alpha

This genus contains all CoV of the previously classified CoV phylogroup I. Albeit their very close phylogenetic relatedness, the CoV in this genus exhibit the presence of a unique cleavage product of their *replicase* polyprotein, nsp1, different in size and sequence from that in beta-CoV and absence of similar candidate sequence in *gamma*-CoV. These CoV also exhibit presence of commonly shared accessory gene for dispensable multispanning *alpha*-CoV specific membrane protein (αmp) which for some members is the only accessory protein and for some like the canine members there can be up to 6 accessory genes (Groot, 2011).

The *alpha*-CoV genus is a group of important animal and human viruses subdivided into several lineages (Groot, 2011). The *alpha*-CoV1 lineage comprises two types of canine (CCoV and CCoV-NTU336) and feline (FCoV and FIPV) CoV, porcine respiratory virus (PRCV) and transmissible gastroenteritis virus (TGEV); another lineage includes human CoV hCoV-229E and HCoV-NL63, and other members of the genus *alpha* are porcine epidemic

diarrhea virus (PEDV) and some bat CoV. The animal species victimized by the respectively specific *alpha*-CoV are all mammals. Humans (HCoV-229E and HCoV-NL63), Cats (Feline CoV: type I and type II), dogs (Canine CoV: type I and type II), pigs (TGEV, PRCV, PEDV), and Bats.

TGEV is the most extensively studied CoV of the species *alpha*-CoV1. TGEV affects mainly the enteric tract with the onset of severe diarrhea and dehydration but can also trigger Lower tract respiratory infections (LTRI), leading to about 90% death toll in newborn piglet (Cavanagh, 2005). The attachment of TGEV to sialic acid has also been indicated necessary for its enteric tropism (Groot, 2011). PRCV, a distinct homolog with a deletion of the N-terminal 250 residues of the S glycoprotein exhibits only respiratory tropism suggesting that the N-terminal domain is responsible for the TGEV enteric tropism.

HCoV-229E is the first discovered (around mid-1960s) human CoV and along with HCoV-OC43 (*beta* CoV) causes very mild colds in humans or pneumonia in immune-compromised patients (McIntosh, 2005). Another closely related *alpha*-CoV is the HCoV-NL63, which uses a distinct entry receptor, as presented below.

PEDV was first discovered in Europe but had become more problematic in Asian countries and now with current rediscoveries from US of a new strain, highly mutated at spike(Wang et al., 2014) to which vaccines (developed from previous strain CV777) are partially protective (Song and Park, 2012). PEDV causes severe diarrhea and dehydration in piglet. PEDV has a tropism distinct from TGEV and it cannot infect ST cells, a cell line highly susceptible to TGEV (Nam and Lee, 2010).

1.3.2 Genus Beta

This genus comprises CoV with distinct nspI sequence than *alpha*-CoV and contains four distinct lineages (Fig.1.1): A (**MHV**, HCoV-HKU1 and the *beta*1-CoV), B (**SARS-CoV**), C (Bt-CoV HKU4, HKU5 and **MERS-CoV**), and D (Bt-CoV HKU9).The most representative prototype CoV of this genus are HCoV-OC43, MHV and SARS-CoV, and the recently emerging MERS-CoV. Lineage A CoV harbour an extra short spike like glycoprotein, HE (Hemagglutinin Esterase).

HCoV-OC43 also causes common cold and pneumonia in elderly populations as well as severe LTRI in immuno-compromised patients such as those undergoing chemotherapies

and/or infected with HIV. It exists as four genotypes: A, B, C and D, wherein D aroused from natural recombinations over time (Lau et al., 2011).

MHV strains have been the best studied b*eta*-CoV before the SARS-CoV emerged especially in laboratory mouse both *in vitro* and *in vivo*. Certain MHV Strains cause inflammations in several mice organs like the neurotropic strains JHM and A59 responsible for acute encephalitis and chronic demyelination in survivors, the latter serving as a model to study multiple sclerosis. Both innate and adaptive immunity are crucial for host defense towards MHV (Bender and Weiss, 2010, Lane and Hosking, 2010).

SARS-CoV brought corona-virology to center of attention from research communities through its global epidemic reporting a fatality rate of 9.6% (2004). The SARS progresses with flu like symptoms with fever over 100ºF followed by shortening of breath and pneumonia. Long-term effects may include pulmonary fibrosis, osteoporosis, femoral necrosis and in some major depressive disorders. It can be more damaging from reactions within immune system called 'cytokine storm'(Perlman and Dandekar, 2005).

MERS-CoV, also called CoV-EMC for first isolate at the Erasmus Medical Center, is a human CoV that emerged in Saudi Arabia by 2012 a decade after SARS epidemic. Until May 2014, MERS has claimed 250 deaths from 700 registered human cases from 20 countries. Sharing 90% sequence identities with bat-CoV HKU4 and HKU5, it docks in lineage 2C of the *beta*-CoV. Within the lineage C, initial MERS isolates belonged to clade A cluster (EMC/2012 and Jordan N3/2012) and genetically distinct new cases belong to clade B (Chu et al., 2014).

1.3.3 Genus Gamma

This genus generally lacks the nspI fragment present in *alpha* and *beta*-CoV. **IBV** infects chickens and it was the first known CoV from 1930s, which exemplifies highly contagious CoV infection in fowls affecting enteric, respiratory, renal, and reproductive systems (Casais et al., 2001, Cavanagh, 2007). This brings huge annual losses to poultry farms running on egg and meat production (Cavanagh, 2007). Disease progresses by leading to loss of ciliary motion in respiratory epithelia, desquamation, mucus deposition, and necrosis. This ultimately culminates in respiratory distress followed by viremia to nephrons and tubules in the kidneys that become swollen. Its then extends to oviduct where it causes lesions in uterus and magnum and thereby hampers egg production.

1.4 CoV genome organization

CoV RNA genome is infectious, nonsegmented, single stranded and like most eukaryotes is 5’ capped and 3’ polyadenylated (Masters, 2006). The RNA genome (Fig 1.2D) starts at the 5’end with a leader RNA (L, 60-80 nt) followed by 5’ (200-600 nt) UTR and then by two overlapping ORF, 1a and 1b, which encompass 60% of the genome (approx. 20 kb). These encode for polyproteins *rep1a* and *rep1b* by means of ribosomal frameshifting (Holmes and Lai, 1996), which can be processed proteolytically to nonstructural proteins (Nsps), viral RNA dependent RNA polymerase (RdRp), helicase, papain like proteases ($PL1^{pro}$ and $PL2^{pro}$), and main protease (M^{pro}) to name a few (Groot, 2011). The non-structural genes are followed by CoV structural genes invariantly in the order 5’-SEMN-3’expressed from a 3'-coterminal nested set of subgenomic mRNAs (Fig. 1.2D). At least eight accessory genes are intercalated among the structural genes, which may be essential for natural infection though not *in vitro*. Finally a 3’ UTR (200-500 nt) is flanked by a polyadenylation signal. The UTRs share genome replication and encapsidation signals with adjacent and internal coding regions.

1.5 CoV structural proteins

Spike (S) is a large (17-20 nm, 1128-1427 aa peptide, mol. mass 150-180 kDa) petal shaped, type I membrane protein (Fig 1.2A and B) which anchors to host cell membrane receptors, induces membrane fusion and is also a major target of CoV neutralizing and hemagglutination-inhibiting antibodies (Masters, 2006, Cavanagh, 2005). It exists as a homotrimer in mature form with each monomer comprising several structural and functional regions. The globular N-terminal S1 region is highly variable among CoV species, mediates receptor recognition and is responsible for altered pathogenicity and antigenicity of the CoV (Masters, 2006). The membrane proximal stalk, S2, is more conserved and bears two heptad repeat regions (Fig 1.2B) with a coiled-coil structure involved in the virus-cell fusion reaction during virus entry into cells. The S1 and S2 regions form a single polypeptide that can be cleaved by host proteases and extent of cleavage depends on host type and varies among CoV. Spike from most CoV of the genus *alpha* are not cleavable though few as feline infectious peritonitis virus (FIPV) can induce cell-cell fusion (Lai, 2006, Masters, 2006).

Membrane protein (M) is a 218-263 aa long (25-30 kDa), type III glycoprotein with N-terminal ectodomain composed of either N-linked (IBV, TGEV, SARS-CoV) or O-linked (MHV) glycans. It has a long C-terminal endodomain that associates with the inner membrane leaflet to form a thick matrix like lattice, explaining the extraordinary thickness of

the CoV envelope. Vital roles in virus assembly like packaging genomic RNA into nucleocapsid are ascribed to M protein (Lai, 2006, Masters, 2006).

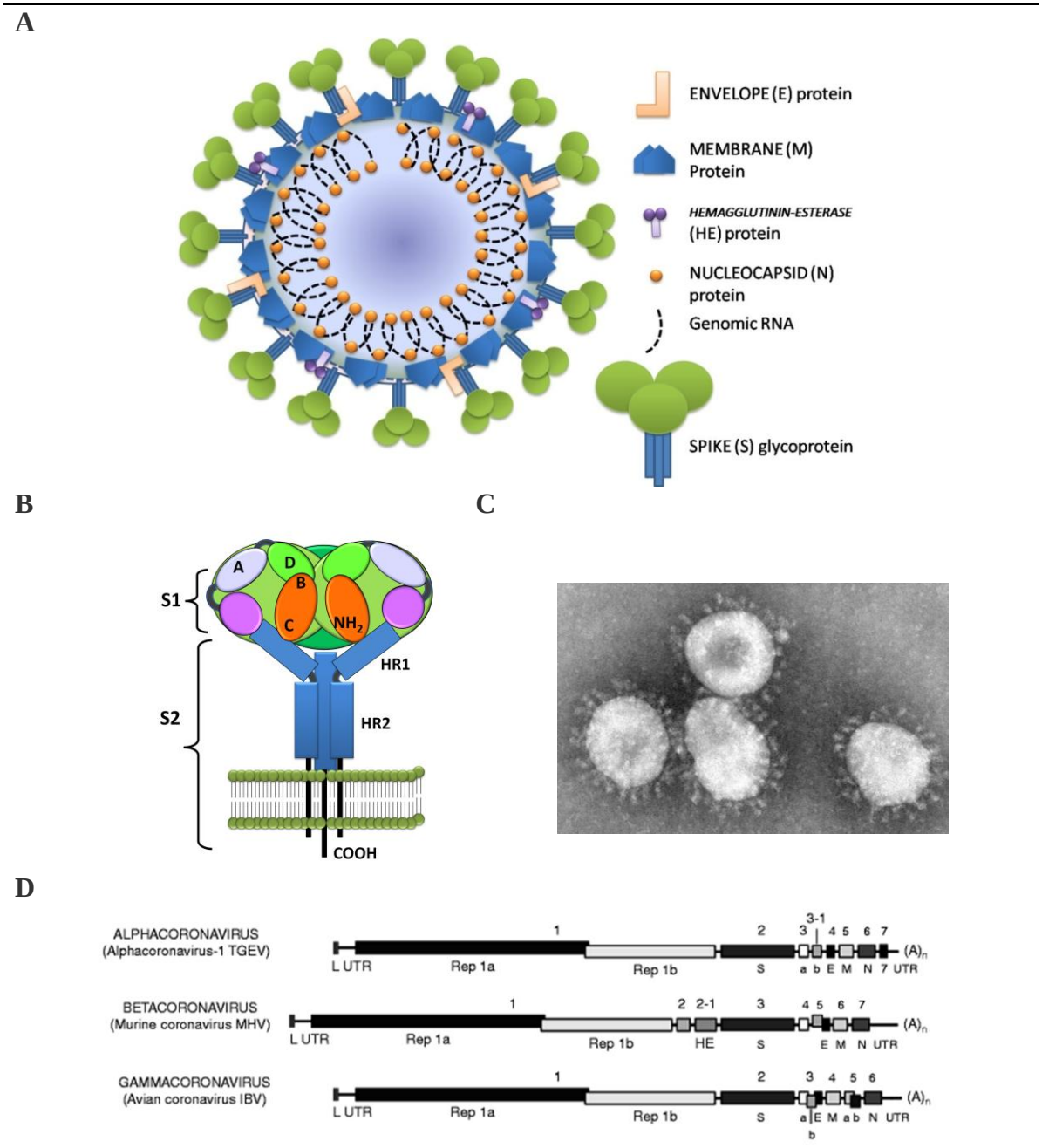


Figure 1.2. General structural and genomic organization in CoV. A. Schematic drawing showing morphological features of coronavirus particles and its envelope glycoproteins. **B.** Detailed scheme of Spike glycoprotein with the globular S1 portion and the elongated stem formed by the S2 region. The four modules and the antigenic sites C, B, D and A defined in the S1 of TGEV are shown (Reguera et al., 2011). The two heptad repeats (HR) of S2 are in rectangles. The N terminus of S1 and the antigenic site A is in close contact with S2 (Reguera et al., 2011). **C.** Electron micrograph of a CoV (the Centers for Disease Control and Prevention's Public Health Image Library (PHIL), identification number 4814). **D.** Genomic organization of the CoV prototypes respective to three approved genera of CoV.

Envelope protein (E) is a small 77-109 aa (8.4-12 kDa), integral membrane protein that forms pentamers and has cation-selective ion channel activity (Wilson et al., 2004). It is required in virion assembly and morphogenesis, and is a virulence factor in SARS-CoV.

Nucleocapsid Protein (N) is a 349-470 aa (50-60kDa), RNA binding phosphoprotein responsible for genome encapsidation, RNA synthesis, translation, exhibits RNA chaperone activity and act as type I interferon antagonist (Lai, 2006).

Hemagglutinin Esterase protein (HE) is a short spike-like glycoprotein found only in lineage A of the *beta*-CoV genus and in MHV has been known to interact with sialic acids. CoV HE proteins share 30% amino acid sequence similarity with that of influenza C virus. It may help initial adsorption of the CoV to the host cell membrane, but it cannot initiate infection in the absence of spike protein (Lai, 2006).

1.6 Immunogenicity of CoV envelope glycoproteins

The CoV S proves an important target for T cell response and epitopes located in its N-terminal portion elicit the production of virus-neutralizing antibodies. The S protein, when inoculated alone, can induce protective immunity (Cavanagh, 2005). For instance, the S1 subunit of IBV S protein has been used as an imunogen in various ways, either by direct removal from virus using either urea or non-ionic detergent (Cavanagh, 2005) or by expression in insect cells following a recombinant baculovirus vector (Song et al., 1998). Immunizations with these S1 preparations induced protective immune responses in about 80% chickens. Greater protection was reported by single oral administration of non-pathogenic fowl adenovirus expressing S1 subunit (Johnson et al., 2003). Similar protection of mice has been seen against lethal challenge with MHV, using its S expressed in adenovirus vector or recovered from purified virus (Wesseling et al., 1993, Daniel and Talbot, 1990). N terminal epitopes in the CoV M protein also induce neutralizing antibodies that function in the presence of the complement system (Woods et al., 1988). The N protein, a dominant antigen during infection, like the S, can also induce T cell responses (Boots et al., 1992). The HE protein solely found in group A *Beta*-CoV, also induce antibodies that prevent virion binding to O-acetylated sialic acids or inhibit sialate-O-acetylesterase activity. The Spike and HE ectodomains are highly variable, indicating vast antigenic drift that may be a result of probable RNA recombinations inter and intraspecies. Besides these, both structural and non structural proteins prove to be $CD4^+$ and $CD8^+$ T cell antigens.

1.7 Coronavirus infectious cycle

CoV infection cycle (Fig. 1.3) initiates with host recognition and S binding to the species-specific host cell membrane receptor. This entails a battery of events that deposit the viral nucleocapsid in to the host cell cytoplasm.

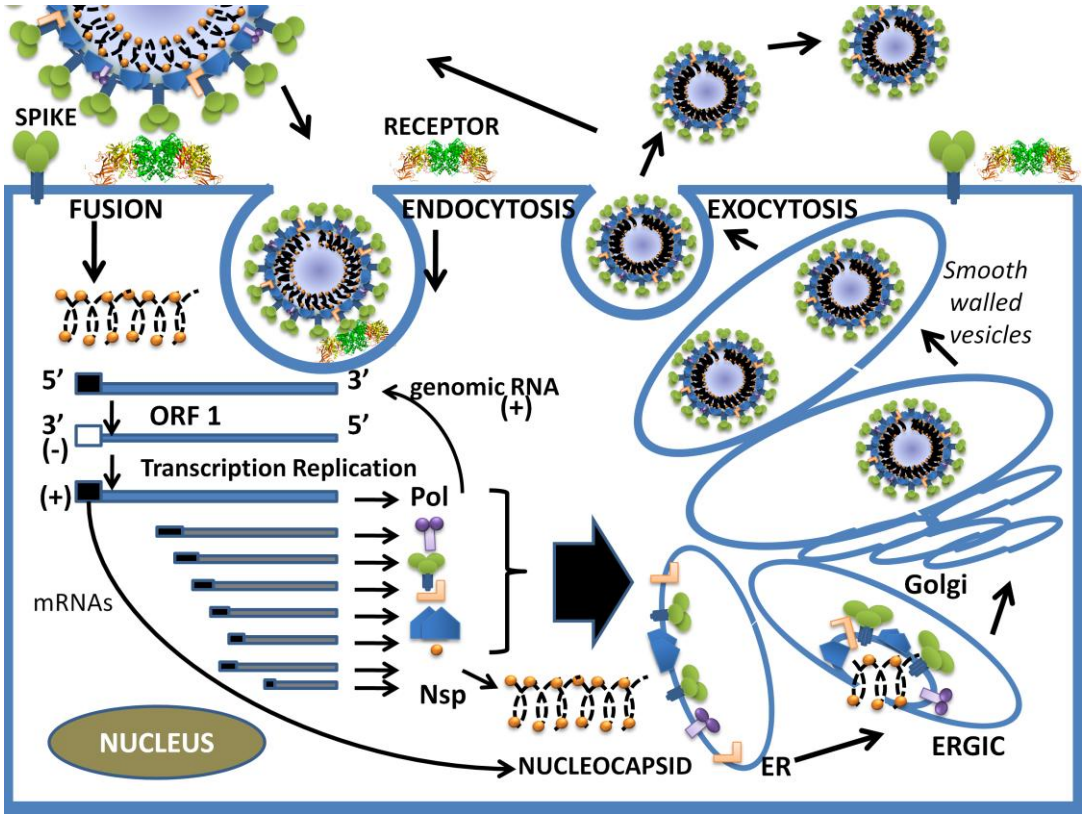


Figure 1.3. Coronavirus infection cycle. Infection initiates with recognition of specific host cell receptor (shown APN) by CoV spike (S) protein. Penetration and uncoating of CoV occurs following S mediated fusion of the viral envelope with host cell plasma membrane. Gene 1 of CoV genomic RNA translates into a polyprotein in turn processed to yield Pol proteins comprising the transcription-replicase complex. These use genomic RNA template to produce negative-stranded RNA, which in turn are synthesized into full length genomic RNA and subgenomic RNA. Each mRNA is translated to only the protein encoded by 5'-most open reading frame (ORF). These include structural proteins N, M, E, S and HE (shown as in Fig 1.2A) and several nonstructural proteins. The N protein and newly synthesized genomic RNA assemble in to helical nucleocapsids. All the structural proteins then assemble at the ER and prepare for CoV budding at the endoplasmic reticulum-golgi intermediate complex (ERGIC). The S and HE proteins are glycosylated, trimerized, and carried to golgi apparatus and along with M are incorporated to the maturing virions, which are released by exocytosis-like fusion of smooth walled vesicles with the plasma membrane

Enveloped viruses utilize a fusion process to penetrate the host cell membrane. The entry of CoV into the cell cytoplasm is preceded by the fusion of virus and host cell membranes following different pathways. Some CoV fuse directly with the cell membrane at neutral pH, as has been shown in some strains of MHV. Other CoV, such as TGEV (Hansen et al., 1998), SARS-CoV (Hofmann et al., 2004, Simmons et al., 2004, Yang et al., 2004) and HCoV-299E

(Nomura et al., 2004) require a low pH step for cell entry, and fusion occurs in endosomes after endocytosis of receptor-bound viruses. Following penetration into the host cell, the first Open Reading Frame of the positive stranded RNA genome translates into a huge polyprotein, which is proteolytically processed into a collection of Pol proteins that contribute to the a transcription-replicase complex (Fig. 1.3). Pol then uses a negative strand RNA intermediate to synthesize both new copies of complete genome and sub-genomic mRNA species. The latter are translated into structural and accessory proteins. The structural proteins are pinned into the endoplasmic reticulum (ER) wherefrom they tread to the endoplasmic reticulum-golgi intermediate compartment (ERGIC). N proteins encapsidate the progeny genomes to form the nucleocapsid, which assembles with the membrane bound components to form virions by budding through the ERGIC (Fig. 1.3). Excess of S and HE proteins skip incorporation into virions and are transported to the plasma membrane, where they assist cell-cell fusion or hemadsorption, respectively The last step culminates into export of virions to the plasma membrane in smooth walled vesicles and its release through exocytosis (Masters, 2006, Lai, 2006).

1.8 Factors affecting CoV cell entry

1.8.1 Membrane fusion

The CoV S is a class I viral fusion protein functionally similar to that of phylogenetically distant RNA viruses such as HIV, influenza and Ebola Virus (Bosch et al., 2003). Like the class I fusion proteins, the S2 region adopts a helical structure, and it harbours a fusion peptide upstream of a heptad repeat 1 (HR1) and a transmembrane domain downstream of HR2 (Fig. 1.4). As in other enveloped viruses, the initiation of the fusion reaction requires partial disassembly of the trimeric spikes and the exposure of the fusion peptide for binding to the host cell membrane. In the follow up, the two HR form a six helix bundle, a highly stable oligomer with 1.45 nm rod like structure (Bosch et al., 2003) . This bundle brings viral and cell membranes in close proximity and facilitates its fusion (Fig. 1.4).

In some CoV, anchoring of the S to its specific host cell receptor triggers a major conformational change in the S that triggers fusion. Other CoV, such as TGEV and SARS-CoV, fusion requires an acidic pH step. This difference in pH requirement is reflected in distinct modes of CoV cell entry (Masters, 2006). Some MHV strains fuse at the plasma membrane, while others enter via receptor-mediated endocytosis and fusion in the acidified endosomes. In MHV, the switch between these two cell entry modes have been ascribed to

residues in HR region of S2, while SARS-CoV could strictly revert to direct cell surface fusion in cells pretreated with protease at the earliest stages of infection. Receptor induced conformational changes in CoV S have been studied mostly in MHV (Masters, 2006). Highly virulent strain MHV4 (JHM) carries S in a metastable conformations with high fusogenic potential; it can mediate fusion among infected cells lacking the receptor (Gallagher and Buchmeier, 2001, Krueger et al., 2001, Nash and Buchmeier, 1996).

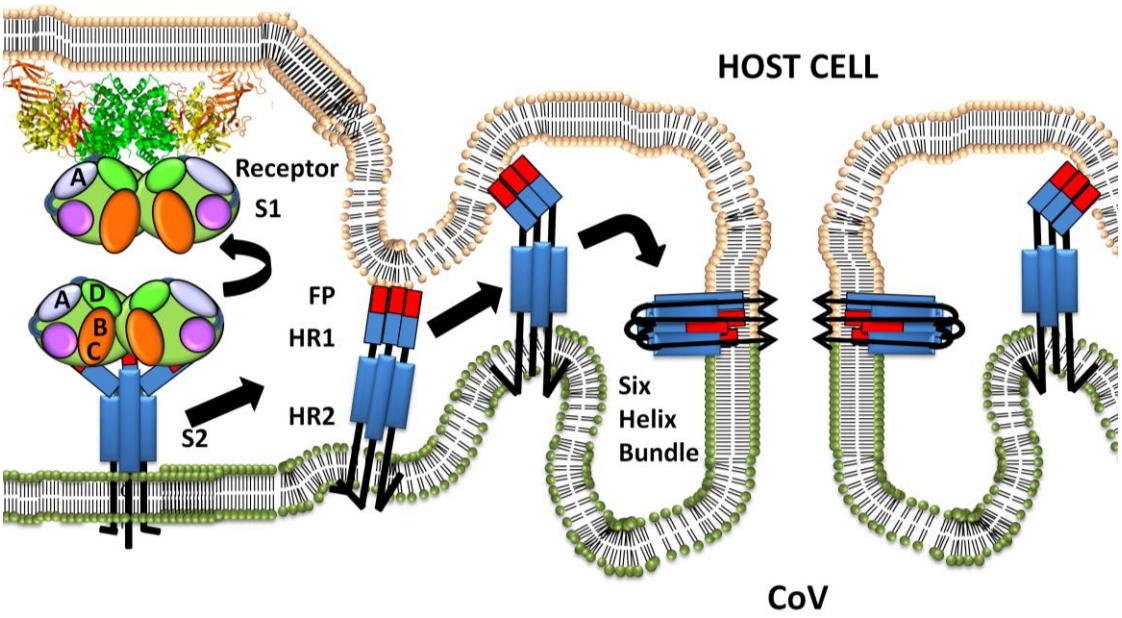


Figure 1.4. CoV-host cell membrane fusion. Steps of S protein conformational changes that may occur during membrane fusion. In the first step, receptor binding, pH reduction and/or (in case of MHV) S protein proteolysis induces dissociation of S1 from S. In the second step, the fusion peptide (FP) is intercalated into the host cell membrane (the fusion-intermediate stage). In the third stage, the part of the S protein nearest to the virus membrane refolds onto a heptad repeat 1 (HR1) core to form the six-helix bundle, which is the final post-fusion configuration of the S2 protein. Figure modified from a published source (Heald-Sargent and Gallagher, 2012)

The S protein in some CoV is cleaved into S1 and S2 to facilitate exposure of the fusion peptide at the N terminus of S2 region toward the cell membrane, but this not an absolute requirement for cell fusion in other CoV (Gallagher and Buchmeier, 2001, Keck et al., 1988, Yoo et al., 1991). Receptor-binding can mediate undefined changes in the S trimmers that initiate the fusion reaction. This has been explored extensively in MHV (Zelus et al., 2003). Cholesterol has been shown to influence differently in CoV that fuse on the cell surface or after endocytosis. Cholesterol supplementation enhances fusion, while its depletion inhibits entry (Thorp and Gallagher, 2004). This effect is not due to clustering of receptors into lipid rafts (Thorp and Gallagher, 2004, Choi et al., 2005). However cell-bound virions do cluster in

lipid rafts, indicating another factor that can be important for MHV fusion (Choi et al., 2005). HCoV-229E virions bound to its receptor redistributes from disperse pattern to clusters in caveolin-1-rich lipid rafts (Nomura et al., 2004).

1.8.2 Protease requirement

All *alpha*-CoV species have spike uncleaved after receptor binding. MHV spike does not need proteolytic cleavage for fusion (Stauber et al., 1993), but it seems more efficient in cleaved S. Protease treatment of SARS-CoV may render its entry into host cells directly on the cell surface without endocytosis (Matsuyama et al., 2005). S proteolysis can thus be required to allow membrane fusion among CoV and target host. For instance, TMPRSS2 colocalizing with the SARS-CoV receptor, cleaves the S and enhances viral entry (Glowacka et al., 2011). Despite the utilization of same receptor, HCoV-NL63 and SARS-CoV display distinct tropism and pathogenesis (Weiss and Navas-Martin, 2005). The lack of function of this protease for HCoV-NL63 entry explains partly the lower infectivity of this virus compared to the SARS-CoV (Krempl and Herrler, 2001). Cathepsin L, a protease that is active in low pH conditions, has been shown to activate HCoV- 229E spike-dependent fusion, leading to a cell entry mechanism similar to the SARS-CoV

Protease requirement may differ between cell-cell and virus-cell fusion. The former correlates with the extent and kinetics of S protein cleavage and can be enhanced by trypsin treatment (Frana et al., 1985). Cleavage varies among different cells indicating the implication of a host protease (Frana et al., 1985). MHV A59 isolate from persistently infected glial cells has altered cleavage site, containing RRADR instead of original RRAHR sequence, which implies a lag in cell-cell fusion infection spread in infected cells (Gombold et al., 1993). But studies with hepatocytes show that viral-cell membrane fusion was not inhibited, indicating that cleavage was not a requirement for this process. Role of furin or furin like protease has been established based on effect of their inhibitors on *in vitro* infection, which blocked both S protein cleavage and cell-cell fusion but not viral-cell fusion (de Haan et al., 2004). A cysteine rich region in the MHV spike operates in cell-cell fusion, mutation of which abrogates syncytia formation (Bos et al., 1995).

1.8.3 pH optima for fusion

Virus-cell fusion occur at neutral pH in some CoV, which indicates that such events can occur directly on cell surface, while other CoV variants require low pH and receptor mediated endocytosis. Use of lysosomotropic agents (ammonium chloride, chloroquine,

bafilomycin, dansylcadaverine) that neutralize acidic pH environments in endosomal compartments, uncovered that some MHV variants fuse in endosomes; other MVH strains can fuse at a neutral pH of 7.4 on the cell surface (Kooi et al., 1991, Weismiller et al., 1990). More recent studies with the MHV strain A59 have used confocal and electron microscopy besides lysosomotropic agents, as well as inhibitors of clathrin-dependent endocytosis (Eifart et al., 2007) and a novel β-galactosidase complementation assay. These studies show MHV enter by receptor mediated endocytosis (Burkard et al., 2014). The Beaudette strain of IBV, a gamma-CoV, shows 90% reduced productive infectivity with ammonium chloride treatment, indicating that fusion occurs soon after endocytosis and it is triggered by low pH. Entry could be strain specific, as no effect was seen with different IBV strains, indicating probable cell surface fusion (Li and Cavanagh, 1992).

1.9 Cross-species CoV transmission:

Pathogenicity of the CoV is considered species-dependent as is infection severity; infections usually occur on closely related natural hosts. This notion has been challenge by the recent CoV outbreaks caused by the SARS- and MERS-CoV. Epidemiologists believe that SARS virus originated in bats (natural reservoir hosts), which was transmitted to Himalayan palm civets, Chinese ferret badgers, and raccoon dogs (amplification and transmission hosts) and subsequently introduced into humans due to handling or consumption of these game animals in the Guangdong province of China. Similar bat to human spread has been accounted for HCoV-229E; cattle to human in case of HCoV-OC43 and cattle to dogs in Canine respiratory CoV (McIntosh et al., 1970, McIntosh et al., 1967).

Alarmed with the SARS event of 2003, molecular epidemiology surveyed two more human CoV, HCoV-HKU1 and HCoV-NL63, both causing common cold with latter leading to bronchitis in children (Woo et al., 2005). More studies revealed a new lineage of solely avian species (Thrush, Bulbul, and Munia-CoV) with relatives in Asian leopard, Chinese ferret badger (all mammals). This supports a new notion that bat coronaviruses serve as a reservoir of *alpha*-CoV and *beta*-CoV, while *gamma*-CoV and *delta*-CoV originate from birds (Patrick C. Y. Woo, 2012). However these hypotheses may still need a broader animal CoV sampling (Groot, 2011).

The emerging MERS-CoV is distinct from other endemic human CoV (HCoV-OC43 and HKU1) with a uniquely strong tropism towards non-ciliated (while most respiratory CoV target ciliated) bronchial epithelia, where it evades innate immunity and antagonizes IFN

production. Its human to human transmission requires a higher respiratory inoculum, as the virus doesn't infect more than 20% respiratory epithelia, so only people in close contact with the infected individual are prone to transmission (OMRAN, 2014). Though, until recently researchers revealed that the human to human transmission mode has been suspended by the virus, it could prove a *silence before the storms* as 20% of the individuals asymptomatic for infection still exhibited MERS-CoV antibodies in their blood. Preliminary isolate from Egyptian Tomb Bat genetically matched perfect with the index patient. Later, other CoV isolates from bat species (*Nycteris* in Ghana and *Pipistrellus* bats in Europe) were diagnosed to occupy 2C lineage among *beta*-CoV and were phylogenetically very related to MERS-CoV (Augustina Annan, 2013). This indicated that bats could be a natural reservoir of the MERS-CoV as in the case of SARS-CoV. However, latest reports from the Journal of Emerging Infectious Diseases, advocate camels (Chu et al., 2014) to be the natural niche, with 99.9% genetic match to human clade B MERS-CoV. It must have been transmitted to humans by either close contact with camels or drinking camels' milk. Supportive reports evidence 50% Omani and 14% Spanish camels have spike specific antibodies to MERS-CoV (Reusken et al., 2013) and they might have spread to different countries during imports. More recent studies document around 90% of Saudi Arabian camels positive for MERS-CoV, suggesting camels as the primary source animal for MERS-CoV, bats rather as ultimate reservoirs. These reports also confirm instances of direct camel to human transmissions of MERS-CoV (Memish et al., 2014a, Memish et al., 2014b). More recently, it came to news-boards that the MERS-CoV could be air borne: a research conducted by Abdulaziz University (Jeddah) on the air samples collected from the camel barns in Saudi Arabia.

Recurrently, CoV infections can cross species barrier and they originate from bats. The question is, why bats? There are globally 60 viral species associated with bats and surprisingly 59 of these are RNA viruses that can potentiate emerging and remerging cases of human infection. Bats (approx. 930 species) belong to the second largest order within the mammalian class (approx. 4600 species) after rodents (2000 species) and so cover 20% of the mammalian dynasty (Samson Wong, 2007). They are geographically ubiquitous (except poles and oceanic islands), with diverse diets (plants, insects, animals and a blood diet), and wide colonization range (10 to 200,000 individuals per colony). It is interesting how bats could harbor such huge diversity of CoV (and other infectious agents) and so lend a big hand in viral ecology and evolution and act as natural wildlife reservoirs for these viruses.

1.10 Coronaviruses receptors

Spike proteins are the primary receptor binding proteins on in the CoV, which specifically target a variety of cell surface molecules (Table 1.1) for entry into host cells. This interaction remains principal if not the only, determinant for CoV tissue tropism and host range (Masters, 2006). Non-permissive cells can become permissive to a CoV by the expression of the host receptor (Delmas et al., 1992). Swapping of spike among different CoV changes the host cell specificity, as it has been shown with FIPV and MHV (Kuo et al., 2000, Haijema et al., 2003); and among the distinct TGEV (Sanchez et al., 1999) and MHV strain (Navas et al., 2001, Phillips et al., 1999). In addition, the remarkable cross-species transmission potential of certain CoV is related to virus adaptation to the use of orthologous receptor molecules by subtle modification in their spike glycoproteins.

The S1 region of the spike is largely variable in sequence and length among CoV, and is specialized in recognition of cell surface receptors (Li, 2012, Masters, 2006). Receptor-binding domains (RBD) can be located at the N- and/or C-terminal moieties of the S1 region (Li, 2012, Peng et al., 2011). The S glycoprotein N-terminal domain (NTD) can function as a RBD; it can be the only S1 domain engaged in receptor recognition, or in conjunction with C-terminal RBD, can broaden tissue tropism of certain CoV. The NTD region adopts a galectin-like structure and it appears as a domain acquired from the host during evolution of CoV (Li, 2012, Peng et al., 2011). In most CoV, the major determinants of cell tropism are found in the C-terminal portion of the S1 region (Masters, 2006). These RBD can usually fold independently and adopt unique structures in CoV, so that they can be considered genuine CoV RBD (Peng et al., 2011, Li et al., 2005a, Wu et al., 2009, Chen et al., 2013). The CoV can use distinct entry receptor molecules, responsible for their broad host range and tissue tropism; however, they preferentially recognize ectoenzymes. The reason of this specificity is currently unknown.

Aminopeptidase N (APN) was the first *alpha*-CoV1 entry receptor described. It serves as receptor for CoV, like TGEV, PRCV, PEDV, FeCoV, FIPV, CCoV, and HCoV-229E over respiratory and enteric epithelia (Table 1.1). The APN RBD locates in the C-terminal portion of the S1 region. APN is type II membrane protein that forms dimers on the cell surface (see below). Other metalloproteases besides APN can also serve as CoV entry receptors. Angiotensin converting enzyme 2 (ACE2) is a zinc-carboxypeptidase involved in cardiac functions and acts as a receptor for SARS-CoV and HCoV-NL63. Both CoV bind similar receptor regions despite their structurally distinct RBD (Li et al., 2005a, Wu et al., 2009).

Catalytically inactive mutants for ACE2 do not exhibit any defect in SARS-CoV spike binding (Moore et al., 2004) or spike mediated syncytia formation (Li et al., 2003). The viruses recognize a single domain of the ACE2 protein and they do not bind to the catalytic site. Recently, a distinct ectoenzyme, dipeptidyl peptidase 4 (DPP4), was identified as the entry receptor of the MERS-CoV (Lu et al., 2013, Wang et al., 2013). DPP4, also called 'adenosine deaminase complexing protein 2' or CD26, is an exopeptidase that cleaves X-proline dipeptides from the N-terminus of polypeptides.

It serves as an antigenic enzyme engaged in signal transduction, apoptosis, and immune regulation and also works as a tumor suppressor (Pro and Dang, 2004, Havre et al., 2008). Its substrate peptides are growth factors, chemokines, neuropeptides, and vasoactive peptides (Chen, 2006). Similarly to APN, DPP4 is a type II membrane protein and forms non-covalent dimers on the cell surface. It is composed of two domains and MERS-CoV binds to a single domain that is distant from the cell membrane (Lu et al., 2013, Wang et al., 2013).

Carcinoembryonic antigen cell adhesion molecule 1 (CEACAM1), a424 aa glycoprotein of the immunoglobulin (Ig) superfamily, is used as entry receptor by MHV (Table 1.1). CEACAM 2 is also envisaged a weak receptor *in vitro* but not an alternate receptor *in vivo.* It is not a peptidase; MHV uses the spike NTD for binding to the N-terminal Ig-like domain of CEACAM1 (Lewicki and Gallagher, 2002, Tan et al., 2002). NTD of other CoV are also in recognition of cell surface sialic acids, which can be used as attachment factors or entry receptor molecules. TGEV NTD interact with an unknown mucin-like glycoprotein that contains sialic acid on the porcine intestinal brush border epithelia, responsible of its enterotropism (Krempl and Herrler, 2001, Schwegmann-Wessels et al., 2002). N-acetyl-9-O-neuraminic acid (Neu5,9Ac2) was found as an entry receptor for BCoV (Schultze and Herrler, 1992, Schultze et al., 1991b, Schultze et al., 1991a). Sialic acid residue α2,3- linked N- acetyl neuraminic acid is also involved in hemagglutination of RBC during IBV infection (Schultze et al., 1992, Winter et al., 2006). Spike-linked carbohydrates can be also used by CoV to attach to host cells. The SARS-CoV can bind to the lectin CD209L (also called LSIGN or DCSIGNR), although it is less efficient SARS-CoV receptor (Jeffers et al., 2004, Marzi et al., 2004, Yang et al., 2004).

CoVGenus	*Species*	*Receptor*
Alpha	• Alphacoronavirus 1 comprising:	
	Feline Coronavirus (FCoV) serotype 2	Aminopeptidase N
	Canine Coronavirus (CCoV) serotype 2	Aminopeptidase N
	Transmissible gastroenteritis virus (TGEV)	Aminopeptidase N/Sialic acid
	• Human coronavirus 229E	Aminopeptidase N
	• Human coronavirus NL63	ACE2
	• Porcine Epidemic Diarrhea Coronavirus (PEDV)	Aminopeptidase N
	• Rhinolophus bat coronavirus HKU2	
	• Scotophilus bat coronavirus 512/05	
	• Miniopterus bat coronavirus 1	
	• Miniopterus bat coronavirus HKU8	
Beta	• Betacoronavirus 1 comprising:	
	Bovine coronavirus (BCoV)	Neu 5,9 Ac2
	Human coronavirus OC43 (HCoV-OC43)	Neu 5,9 Ac2
	Equine coronavirus (ECoV)	
	Human enteric coronavirus (HECoV)	
	Porcine haemagglutinating encephalomyelitis virus (PHEV)	
	Canine respiratory coronavirus (CrCoV)	
	• Murine coronavirus comprising:	
	Existing species of **mouse hepatitis virus (MHV)**	CEACAM1
	Rat coronavirus	
	Puffinosis virus	
	• Human coronavirus HKU9	
	• Rousettus bat coronavirus HKU4	
	• Tylonycteris bat coronvirus HKU5	
	* **MERS-CoV**	DPP4
	• SARSr-CoV (SARS related Coronavirus) comprising	
	Human SARS-CoV	ACE2
	Rhinolophus bat viruses	
Gamma	• Avian coronavirus comprising:	Sialic acid
	IBV, Various coronaviruses infecting turkey, pheasant, duck, goose and pigeon	
	• Beluga Whale coronavirus SW1	
Delta	• Bulbul coronavirus HKU11	
	• Thrush coronavirus HKU12	
	• Munia coronavirus HKU13	

Table 1.1 Coronavirus genera, host and receptors. Table adapted from a published report (Belouzard et al., 2012) and modified by including MERS-CoV. Typical representatives have been highlighted in red.

1.11 Aminopeptidases N

APN (EC3.4.11.2) is 150-160 kDa, type II membrane metallopeptidase that belongs to the M1 family (Look et al., 1989, Luan and Xu, 2007). Also called CD13, it is a multifunctional protein winning the title of moonlighting enzyme. Many of its features were unveiled after its crosslinking with monoclonal antibodies, overexpression or silencing in variety of cell lines, gene-knockouts in animal models and/or by use of catalytic inhibitors. Many of these features are very well documented, though most still seek further studies (Mina-Osorio, 2008).

1.11.1 Structural features

APN coordinates a zinc ion at its active site through highly conserved consensus HEXXH and GXMEN motifs (Hooper, 1994) which are critical for APN's catalytic activity (Noren O, 1997, Look et al., 1989). APN exists as a heavily glycosylated (30% mol. mass) (Danielsen et al., 1982, Riemann et al., 1999) head-to-head homodimer on the cell surfaces, and the ectodomain protrudes about 10.5 nm from the cell membrane. Compared to other members of the M1 family, which usually are monomers, APN are dimers that hydrolyze broad range of peptides. Dimerization imparts stability to the APN as it stands harsh environments than the intracellular proteins. It may also exist in small proportions as functionally active soluble isoform. APN are ubiquitously found over membranes of the microvilla in small intestine, renal, intestinal, pulmonary epithelia and also on granulocytes, monocytes, fibroblasts, endothelial cells, pericytes of blood-brain barrier and synaptic membranes of the CNS (Delmas et al., 1994, Look et al., 1989, Olsen et al., 1988).

1.11.2 Catalytic functions

CD13 catalyses the removal of N-terminal amino acids from unsubstituted oligopeptides, amide or arylamide, with the exception of peptides with proline in the penultimate position. It prefers neutral amino acids and the order of preference for substrates is Ala>Phe>Tyr>Leu>Arg>Thr>Trp>Lys>Ser>Asp>His>Val. Its peptidase function cleaves individual amino acids in the intestinal lumen, peptides involved in MHC binding and also helps degradation of neurotransmitters at the synaptic junctions (Larsen et al., 1996, Matsas et al., 1985, Noren O, 1997). Distinct open and closed conformational states have been observed among the M1 family members and these movements are thought important for peptide catalysis (Kochan et al., 2011).

1.11.3 Receptor functions

APN acts as a receptor for most *alpha*-CoV1 subgenus, such as porcine (TGEV, PRCV, PEDV), feline (FIPV and FeCoV), canine (CCoV) and HCoV-229E. Feline APN can serve as a receptor for all APN receptive CoV, thus indicating cats being the origin of host's specificity in *alpha*-CoV, which has been related to different glycosylation pattern of CD13 among hosts (Tresnan and Holmes, 1998, Hegyi and Kolb, 1998). Accordingly, the host range of CoV primarily associates with receptor usage, while the disease type does not. CoV-receptor and APN enzymatic functions are considered independent. This has been indicated by the use of catalytic inhibitors of APN, which failed to influence CoV infectivity *in vitro* (Delmas et al., 1994, Moore et al., 2004).

1.11.4 Other moonlighting functions of APN

Ligand-binding to APN may results in its internalization, as seen with antibodies that mask its surface activities (Mina-Osorio et al., 2006). Many molecules or auxillary proteins viz., galectin 3, galecin 4, RECK, and tumor associated antigen L6, regulate APN cellular trafficking in order to either control its optimal functions or its cell signaling capacities independent to its receptor functions. Praising its roles in MAPK phosphorylation, calcium influxing, and cell adhesion, APN has been grouped into the list of signal regulators (SR). Besides, catalytic activity of both soluble and surface APN forms are correlated with invasiveness of various tumor cells, thus qualifying them as important indicators of many neoplastic disorders, including those of skin, colon, kidney, lung, stomach, bone and prostrate tumors (Kehlen et al., 2003, Fujii et al., 1995)**.** APN is a target for cancer chemotherapies; drugs that bind this protein have been studied for treating tumors, some of which are in clinical trials**.** APN is also involved in cell differentiation as evidenced from their altered expression in different cell lines which potentiates its use as a diagnostic marker for peculiar lymphoma and leukemic disorders (Nakase et al., 1996). This function is also correlated with catalysis possibly because of its processing of peptides that mediate the induction or secretion of growth factor involved in differentiation (Lohn et al., 1997). More recent interest has been envisaged on their role in stem cell differentiation (Chen et al., 2007). Other important functions that APN plays in cell biology relate to cell proliferation, programmed cell death, motility of tumor cells and spermatozoids, chemotaxis, antigen presentation, cholesterol crystallization and uptake, phagocytosis, angiogenesis and cell adhesion, and they have been faithfully considered with demanding details in recent literature (Mina-Osorio, 2008).

CHAPTER 2

2.0 OBJECTIVES

The following objectives were realized during the course of study:

2.1	**Expression and purification of the human and pig APN ectodomains and CoV S length variants.**
2.2	**Expression and purification of the CoV S length variants.**
2.3	**Determination of APN ectodomain structures.**
2.4	**Crystal structure of an alpha-CoV RBD bound to its APN receptor.**
2.5	**Identification of critical virus-receptor binding motifs in the TGEV-APN interface.**
2.6	**Influence of the APN conformation and peptidase activity in CoV binding and infection.**
2.7	**Characterization of the main antigenic site in the spike of TGEV.**

CHAPTER 3

3. MATERIALS AND METHODS

3.1 Cells and media

Bacterial cell Top10F' and DH10B cultures of *E.Coli* were used for competent cell preparation for routine gene cloning procedures, maintenance of plasmid and BACmid vectors respectively. LB media supplemented with appropriate antibiotic was used for bacterial cultures. Mammalian cells HEK293T cells were used in transient transfections to test soluble and surface expressions of glycoproteins. CHOLec 3.2.8.1 (CHOLec) cells were used to generate clones stably expressing recombinant soluble proteins intended for crystallization (Stanley, 1989, Casasnovas and Springer, 1995, Ordono et al., 2006). BHK21 and CHOK1cells were used in stable surface expression of proteins intended for carrying out protein interaction assays and were also accompanied by ST cells in virus infection experiments. Eagle's Minimum Essential Medium (E-MEM) supplemented with non-essential amino acids, sodium pyruvate, glutamic acid, asparagines and nucleoside (Sel-MEM), as well as 10% heat inactivated FCS was used as growth and maintenance media for tissue monolayer culture of adherent CHOLec and CHOK1 cells. For selection of their stable transfected CHO cell clones 25-30µM of L-methionine sulfoximine (MSX) selective drug was used in Sel-MEM supplemented with 5-10% dialyzed FCS. Dulbecco's Modified Eagle's Medium (DMEM) culture media supplemented with 10% heat inactivated FCS was used for adherent 293T, BHK and ST cells. For serum free 293T expression, transfectants were cultured in either serum free DMEM or OptiMEM-I medium (Life Technologies). Mammalian cell culture media was also supplemented with antibiotics, a penicillin/streptomycin mixture or gentamycin, in order to avoid bacterial contamination. Geneticin (G418) antibiotic was used to sort stable BHK-21 cell transfectants.

3.2 Protein Sequences

Accession numbers used for pig APN and human APN are P15145 and M22324, respectively. TGEV spike is Q0PKZ5. PRCV HOL87 S protein sequence is reported in published literature (Sanchez et al., 1992).

3.3 Recombinant protein expression and purification

Positional cloning was routinely performed using polymerase chain reaction (PCR) and gene specific sets of primers designed to anneal and amplify genes encoding protein of interest from full length cDNA templates. Restriction enzyme recognition sites were also designed to

flank the amplicons. These restriction sites were used to transfer cDNA among different expression vectors. Constructs intended for expression of soluble ectodomains, required a signal peptide for extracellular secretion. This was facilitated by cloning the cDNA amplicons into a pDisplay (Invitrogen) derived vector, pBSKS-D (Fig. 3.1A) which imparts a human IgK leader signal sequence followed by a Hemagglutinin A (HA) tag.

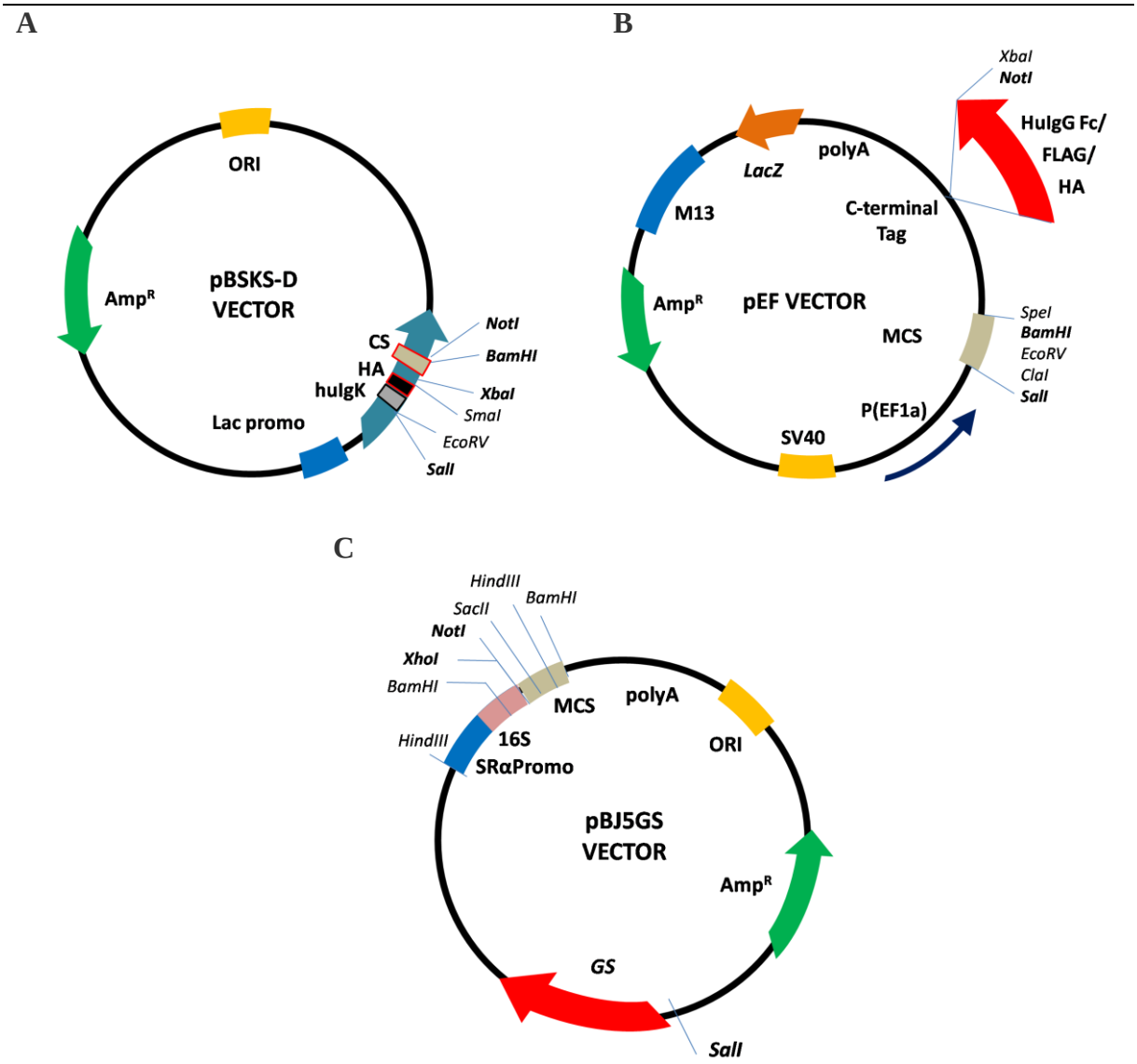


Figure 3.1. Mammalian expression vectors. **A.** pBSKS-D cloning vector derived from pDisplay by introducing its human IgK (huIgK) secretory signal and Hemagglutinin A (HA) epitope tag engineered into pBSKS+ plasmid (Stratagene).**B.** pEF vector derivatives of pEF-BOS (Mizushima and Nagata, 1990) containing SV40 origin of replication, human EF-1a promoter , polyadenylation site (polyA), regions of *LacZ'*, M13, and ampicillin resistance. C-terminal Tag (highlighted in red) included either human IgGFc (HuIgG Fc), HA or FLAG tag. **C.** pBJ5GS vector with *Glutamate synthetase* (GS) gene for selection of stably expressing CHO cell clone under glutamine starvation. Cloning *SalI-NotI* fragment from recombinant pEF-vector (in B) in to *XhoI-NotI* sites of pBJ5GS dissolves *SalI* and *XhoI* sites in pBJ5GS. Routinely used restriction enzymes within the Multiple Cloning Sites (MCS) are highlighted in bold. All vectors use ampicillin resistance (Amp^R) for bacterial selection.

Further, cloning involved transfer of *SalI-BamHI* fragment from recombinant pDisplay to different derivatives (Fig. 3.1B) of the vector pEF-BOS (Mizushima and Nagata, 1990). This introduced desired C-terminal tags for preparing fusion proteins which aided their antibody-affinity based purification. A thrombin recognition sequence was also introduced between the protein and the C-terminal tags. The recombinant pEF vectors were transfected into 293T cells using the calcium phosphate method for transient protein expression (Jordan et al., 1996). The presence of the proteins in the cell supernatants was monitored by ELISA at about 3 days post-transfection.

cDNAs with the required terminal tags were excised from respective recombinant pEF derivatives using *SalI-NotI* restriction site and were cloned into the unique *XhoI-NotI* site of the pBJ5-GS vector (Fig. 3.1C). Recombinant pBJ5GS vector were used for preparation of stable CHO-Lec clones following the glutamine synthetase expression system (Casasnovas and Springer, 1995, Ordono et al., 2006). Clones growing at 30 μM MSX and secreting the protein to the cell supernatants were selected by ELISA. Cells were subcloned by limiting dilution for selection of clones expressing high amount of proteins. Cell clones were grown in sterile plastic roller bottles (Greiner bio-one) for large scale protein production.

All the recombinant glycoproteins used in the study (Fig 3.2) were developed employing the above cloning and expression strategy. From a cDNA template encoding full length pig Aminopeptidase N (pAPN), specific fragments were PCR amplified using primers APN-N4 and APN-C2 (Table B, appendix) for the soluble ectodomain and primers APN-N3 and APN-Call for the surface expressible variant. Primers APN-N4 and hAPN-C1 were used for amplifying human APN (hAPN) ectodomain from a full length cDNA template (Look et al., 1989). Preparation of CoV spike length variants was based on the established modules in the S1 globular region and their studied interaction with receptor, pAPN (Reguera et al., 2011). Accordingly, CoV S length variants (Fig 3.2B) were produced using a cDNA template encoding full length spike (S) from TGEV (PUR46MAD) and PRCV (strain HOL87). TGEV S length variants, S1 (primer Cos-N1 and Cos-C1) and S3 (primer Cos-N1 and Cos-C3) included the RBD region which was also exclusively developed (primer Cos-N4 and Cos-C32). Similarly, generation of PRCV-SH1 used primers Cos-N1 and Cos-C1 while SH3 used primers Cos-N1 and Cos-C3 (Table B, appendix). Soluble APN ectodomains and TGEV S RBD harbored an N-terminal IgK leader followed by an HA tag (Fig 3.2A and B). At the C-terminus, soluble pAPN ectodomain was flanked by a Flag tag, while its surface expressible construct carried an HA tag. For secretion in cell supernatants, S1, S3 (TGEV S), SH1 and

SH3 (PRCV S) harbored endogenous signal peptide at the N terminus, while in case of APN and TGEV S RBD, this was facilitated by the human IgK leader sequence during cloning in pBSKS-D vector (Fig 3.1A). The S1, S3, SH1, and SH3 glycoproteins carried at their C-terminal either a Flag or HA tag. The RBD variant carried either a Flag tag (monovalent) or a human IgG-Fc protein (divalent) tag at the C-terminus.

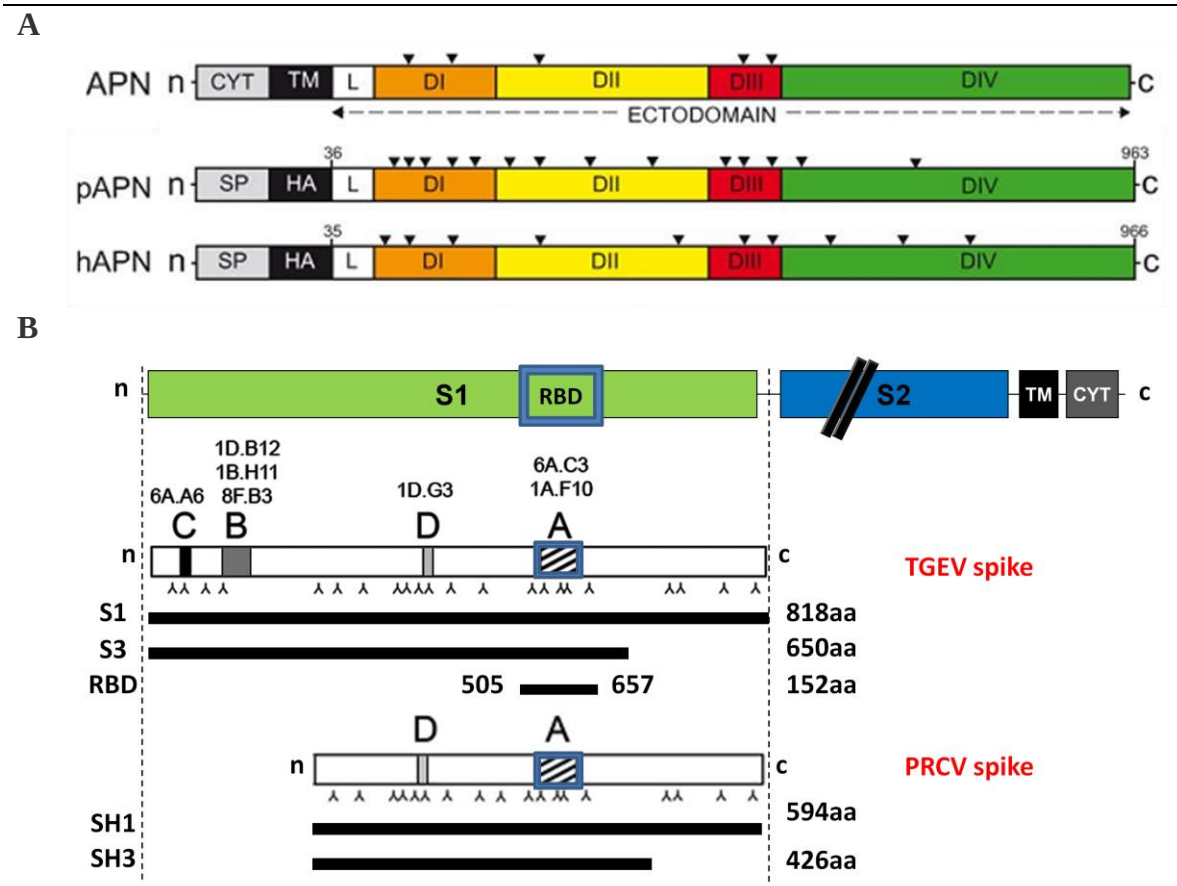


Figure 3.2. Schematic of designed genetic elements for expressing and purifying soluble proteins. A. Cartoons depict modular organization of full length APN (top) and modified soluble ectodomains from pAPN and hAPN (bottom). The N-terminal cytoplasmic (CYT) and transmembrane (TM) domains in APN were substituted by a signal peptide (SP) and an HA epitope. Colored boxes indicate domains I-IV with a white box (marked L) corresponding to a flexible polypeptide. Black triangles indicate of N-linked glycosylation sites. **B.** Cartoon (top) depicts general full length CoV spike with N-terminal S1 and C-terminal S2 stalk region (drawn smaller in size w.r.t S1 to fit the scope, marked by diagonal bars) followed by TM and CYT domains. Dotted vertical lines emphasize S1 region used to construct the length variants (black bars) S1, S3 and RBD from TGEV (PUR46MAD) S protein; SH1 and SH3 from PRCV (HOL87) S protein. Antigenic modules C, B, D, and A and antigenic sites shown for S1 in both CoV. Blue rectangles locate the receptor binding domain (RBD) on respective constructs.

Proteins secreted to culture supernatants were initially purified by affinity chromatography using monoclonal antibody against the epitopes either on protein of interest or the tagged proteins. For the APN this used 12AC5 mAb against the HA tag and for the CoV spike length variants, mAb 6AC3 was used which recognizes epitopes in the antigenic site A (Fig 3.2B).

These mAbs were coupled to CL4B-sepharose and packed on to gravity flow glass columns (Glass Econo-Column®, BIORAD) equilibrated in a buffer solution (10 mM Tris, 150 mM NaCl, pH 7.3). Cell supernatants (1-2 L) were passed through this column followed by 10 column volumes of wash buffer (10 mM Tris, 150 mM NaCl, pH 8.3). Proteins of interests were then recovered as fractions using an elution buffer contacting 20-40 mM glycine (pH 3.0 for 12AC5 and pH 2.8 for 6AC3 column). Protein elution was immediately followed by pH neutralization of the samples (using 1 mM Tris, pH8.0). Column was washed again followed by re-equilibration (10 mM Tris, 150 mM NaCl, pH 7.3). Following visualization on SDS-PAGE, apparently pure protein fractions were concentrated and further purified by size exclusion chromatography in HEPES-saline buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) and were again analyzed on SDS-PAGE gels. For size exclusion chromatography, sepharose columns (GE healthcare) fitted with the Akta-Purifier (GE heathcare) system.

3.4 Site directed mutagenesis (SDM)

SDM was performed to introduce point mutations in gene of interest contained in plasmids used as template and primers designed (Table B, Appendix). Quickchange kit (Stratagene) protocol was followed using either *pfu*I Turbo (Agilent Technologies) or *Nzy*DNAchange (NzyTech) polymerase. Briefly, PCR was performed as recommended and an aliquot was checked by electrophoresis to confirm amplification; *DpnI* treatment was done for 2 hours at 37^{o}C to eliminate template plasmid and amplified plasmid was transformed to Top10F´ competent cells. Transformed colonies were selected and plasmid DNA was prepared using a kit protocol (Qiagen) to verify the presence of intended mutagenesis by sequencing.

3.5 ELISA

ELISA was used as a binding assay for soluble proteins and was performed in 96 well plates (Nunc-immunosorp). Binding of anti-TGEV spike or tag specific mAbs to wild type and mutant CoV RBD was assayed using purified mAbs or hybridoma supernatants. The CoV RBD-Fc fusion protein produced in serum free supernatants was bound to plastic and mAb binding was monitored by optical density (OD_{492nm}). At least four RBD-Fc concentrations ranging from 1 to 10 µg/ml were used in duplicate and average binding was deduced after correcting the background binding. ELISA was also used as a test to determine purification efficiency in antibody affinity based purifications of soluble proteins.

3.6 Flow cytometry

Flow cytometry was used to characterize cell surface expression of APN, as well as its interaction with TGEV-RBD on the surface of 293T, BHK21, and CHOK1 cells. Briefly, cells washed with ice cold PBS, were detached from the plate surface using 0.02% EDTA solution. After brief sedimentation at 1400rpm, cells were resuspended in flow cytometry (FACS) buffer (0.5-1%BSA in PBS). Cells were counted on haemocytometer and plated in conical wells of a 96 well plate at a density of $5x10^4$ cells/ well. Binding procedures were performed at 4^oC. The TGEV-RBD-Fc fusion protein was added to wells with cells at the appropriate concentration in 20 µl of FACS buffer for binding to APN for 30 mins, followed by three centrifugal washes with FACS buffer. Subsequently, cells were stained with anti-human IgG-FITC Ab (Invitrogen) and incubated on ice for 30 mins followed by 3 washes prior to analysis in the flow cytometer.

APN expression on the cell surface was analyzed by FACS using either the anti-human APN mAb (WM15) or anti-HA mAb for pAPN (5 µg/ml). An anti-mouse IgG-FITC Ab (Invitrogen) was used for antibody detection.

3.7 Cell surface APN crosslinking and immuno-precipitation

Cell surface dimerization of wild type and pAPN mutants were analyzed by chemical crosslinking of cell surface proteins on transiently transfected 293T cells. Cells were transfected with recombinant pEF-HA vector with a cDNA encoding the membrane-bound pAPN protein and with a C-terminal HA peptide. Two days post-transfection, cells were washed twice with PBS and collected for flow cytometry analysis of pAPN expression and TGEV-RBD binding (Section 3.6). A water soluble and membrane-impermeable reagent bis[sulfosuccinyimidyl] suberate (BS^3, Pierce), was used to crosslink proteins on the cell surface according to manufacturer's instructions. Briefly, cells were treated with 3 mM BS^3 for (30-60 mins at 4ºC) and the reaction quenched with 20 mM Tris-HCl pH 7.5 for 15 mins. Cells were washed, sedimented, and resuspended in ice-cold lysis buffer (50 mM Tris-HCl pH 7.4, 1% Triton X-100, 300 mM NaCl, 1 mM DTT, 1 mM PMSF, 1 mM EDTA) by brief shaking. Solublized pAPN proteins were immuno-precipitated with anti-HA mAb, analyzed by SDS-PAGE followed by a western blot using the same mAb, an HRP-labeled secondary antibody, and the ECL plus detection kit (GE Healthcare).

3.8 APN catalytic activity assays

Commercially available leucine-paranitroanilide (L-pNA) substrate (Sigma) was used in a standard spectrophotomteric assay to characterize catalytic activity of soluble or cell surface aminopeptidase N. Assay was performed in 96 well plates with 100ul reaction volumes/well prepared in PBS pH 7.5. FPLC purified soluble APN (5µg/ml; 40nM) were added to wells in duplicates, either alone or with increasing amounts of soluble TGEV/PRCV S proteins. L-pNA substrate was added to the reactions at a final concentration of 1 mM at 4^{o}C. Plates were incubated at room temperature and OD at 405nm was measured at different instances to determine the amount of released pNA byproduct of APN catalysis. Background OD of wells without APN was subtracted to determine specific catalytic activity. Activity of cell surface expressed APN was similarly analyzed using 0.5x10^{5}cells/wells.

3.9 Generation of TGEV mutants

Methods towards engineering coronavirus genomes in bacterial artificial chromosomes (BACs) are well established (Almazan et al., 2000, Almazan et al., 2008) and have been followed in the research work of this thesis to develop and rescue TGEV mutants. A scheme of steps involved in developing the mutants has been depicted in Fig 3.3A. Firstly, a recombinant cloning vector pGEM-T, which carried a copy of the full length cDNA of TGEV S (PUR46MAD), was used as a template to generate S protein mutants (Section 3.4). Using the *Pac*I and *Mlu*I restriction enzymes, the mutated S fragment was transferred to an intermediate recombinant BAC plasmid, pBeloBAC11-TGEV$^{\Delta cla\mathrm{I}}$ **(**Fig. 3.3). This BAC plasmid carried an engineered cDNA copy of the full length TGEV genome, except an element (*Cla*I-*Cla*I fragment) which is toxic to bacterial culture. This modification (*Cla*I fragment) was separately maintained in another BAC construct, (pBAC-TGEV$^{cla\mathrm{I}}$) and introduced in to the intermediate BAC plasmid in the last cloning step to make the final full length infectious construct, pBeloBAC11-TGEV^{FL} (Fig. 3.3B). Mutations were verified by sequencing the resultant BAC plasmids using primers amplifying a stretch of S. Colony PCR and restriction digestion pattern analysis was used to verify the correct orientations of the inserts and the BAC stability throughout the steps in engineering of the final TGEV clone. Due to low copy number (1 or 2 per cell) and the need for suitable amounts of the recombinant BAC constructs midiprep plasmid DNA kit (Qiagen) was used with special recommendations for BAC system. For BAC and insert DNA purification, agarose gel extraction was done with the QIAEX II Gel extraction kit (Qiagen).

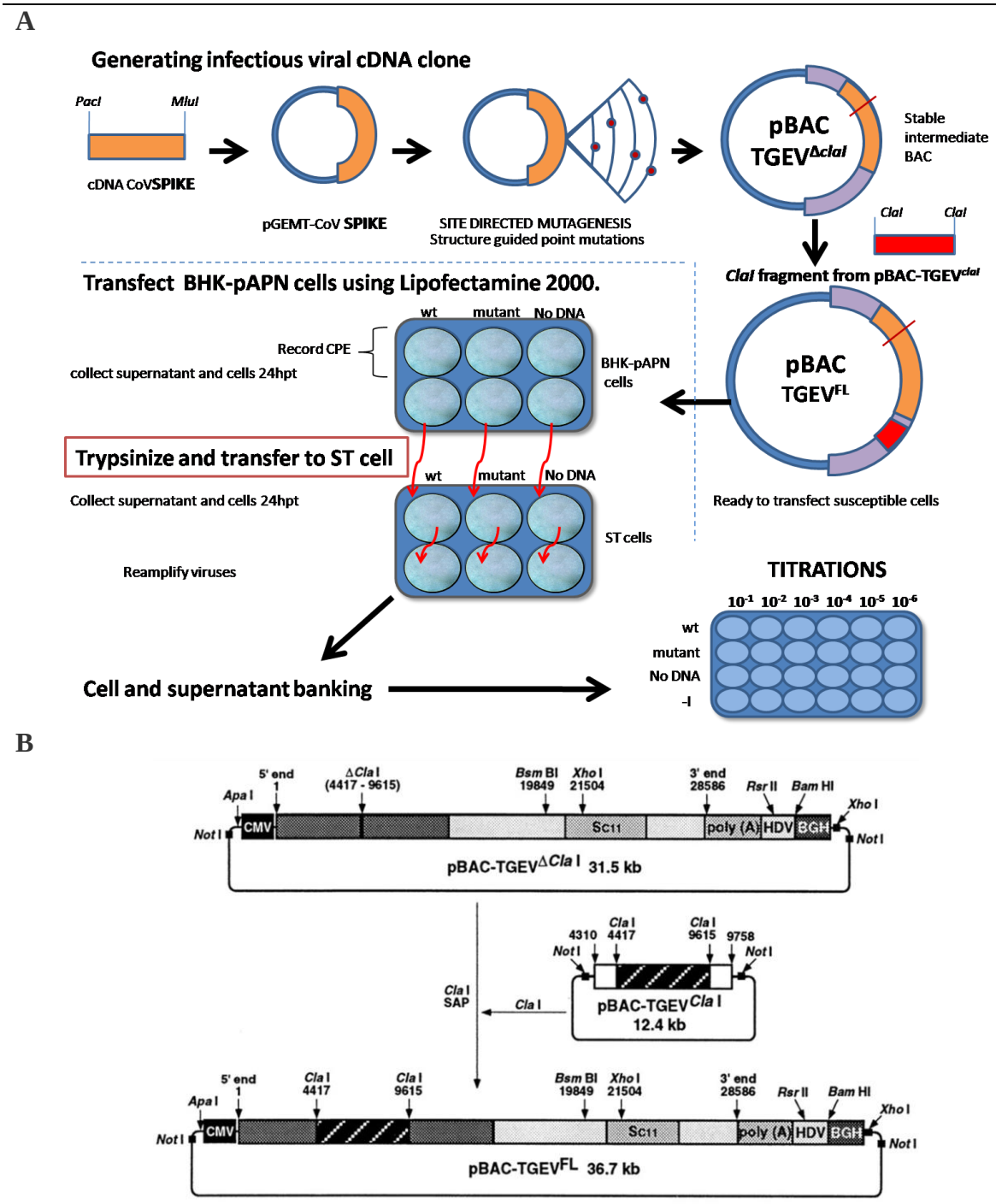


Figure 3.3. Generation and rescue of mutant TGEV. A Scheme of procedures followed to generate TGEV mutants starting from site directed mutagenesis of the S-cDNA, followed by its transfer to a stable intermediate pBAC-TGEV$^{\Delta ClaI}$ vector disarmed for a toxic fragment. A final infectious BAC vector is generated for transfection of BHK-pAPN cells for generation of recombinant TGEV with mutation in the S protein, Transfected BHK-pAPN cells are layered on fresh monolayer culture of ST cells for amplification. The virus titers are determined by plaque assays. **B.** Cartoons showing details of the cloning strategy of using the BAC vectors. Figure adapted from a published source (Almazan et al., 2000).

Rescue and characterization of the respective TGEV mutants, was carried out in mammalian cells. Highly pure preparations of pBeloBAC11-TGEV^{FL} constructs prepared with the Large

Construct kit (Qiagen) were transfected into monolayers of BHK-21 cells expressing pAPN on their cell surface (BHK-pAPN cells). The transfection reagent, Lipofectamine 2000 (Invitrogen) was used following established protocols (Almazan et al., 2008). After 6 hours of transfection, transfected cells were detached by brief trypsin treatment and layered on a ST cell monolayer. When a cytoptahic effect was seen (2-3 days post transfection), supernatants were collected for virus titrations using plaque formation assays.

3.10 Cytopathic effect (CPE) assays

CHOK1 cells were seeded on 96 well plate at a cell density of $5x10^4$ cells/well in E-MEM supplemented with 10% dialyzed FCS. The following day, cells were inoculated with increasing dilutions of the viral stock. CPE was observed 48 hours after infection, following fixing in 10% formaldehyde for 1 hour and cell staining using crystal violet solution. For quantification of viruses in supernatants, titrations utilized plaque formation assays on ST cell monolayers. ST cell monolayers were seeded on 24-well tissue culture plates and inoculated with serial dilutions of supernatants in DMEM supplemented with 2% FCS. After 90 minutes absorption, monolayers were washed with media and fed with semisolid overlay media (0.7% agar in media). Plates were incubated for 2 days followed by crystal violet staining for plaque counting.

3.11 Crystallography methods and tools

Crystallography of soluble APN ectodomains and their complexes with CoV S length variants followed the workflow shown in (Fig. 3.4).

3.11.1 Protein Concentration

After obtaining pure homogenous protein sample following gel filtration chromatography (Section 3.3), the sample fractions were pooled together for concentration in centrifugal concentrator units (Millipore, PAL life sciences). Samples were centrifuged in these concentrators at the recommended speed and10^oC in either SS34 rotor (Sorvall) or bench-top centrifuge (5810R, Eppendorf). Concentration procedure was continued to reach a protein sample of 15-20 mg/ml, which was used for setting up crystallization screening trials.

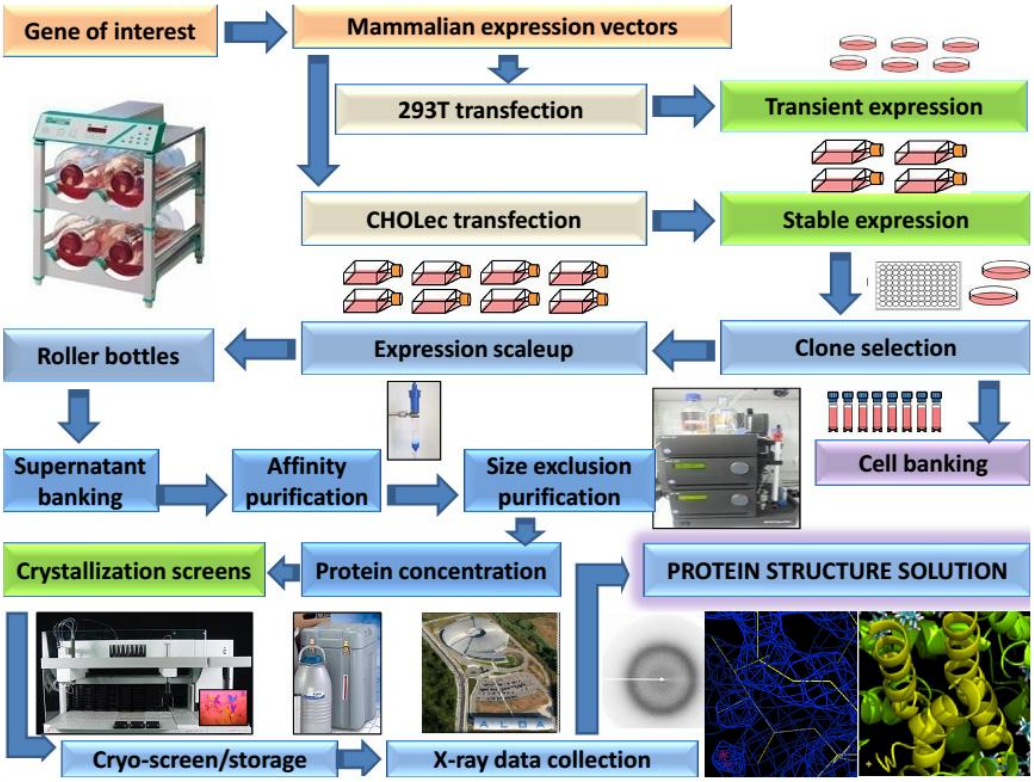


Figure 3.4. Strategic workflow for structure determination of APN and complexes with CoV S proteins. The flowchart points out important technical milestones to during protein preparation in mammalian cells, their crystallization and structure determination. The recombinant cDNA is cloned in a mammalian expression vector with a suitable tag (Fig 3.1). Protein expression is first tested by transient expression in HEK-293T cells. This is followed by stable expression and selection of CHOLec clone, which is then scaled up in roller bottle cultures for periodic collection of supernatants and media replenishment. Supernatants are run through affinity columns to purify proteins in the first round and later by a size exclusion step. The concentrated proteins are crystallized using several crystallization kits and hit conditions are repeated to improve crystals that are then exposed to X-rays at the synchrotron beamlines to perform diffraction data collection. Crystal structures are determined following different methodologies.

3.11.2 Crystallization screening

Commercially available crystallizations kits (Jena Biosciences and Hampton Research) were used in the screening of crystallization conditions for all proteins in this Thesis work. All trials were done with the sitting drop method using 96 well crystallization plates. Crystallization drops were set up using a liquid handling robot (TECAN). Fig. 3.5 shows a schematic diagram of sitting drop technique that allows vapour diffusion mediated crystallization of proteins (Bergfors 1999). For co-crystallization of protein complexes, equimolar mixtures of the interacting proteins were incubated overnight before setting up crystallization screens. Crystallization screening plates were incubated at constant room temperature of 22^{o}C. Various crystal forms were obtained for different proteins and complexes. The kit screen solutions giving protein crystals were usually modified for crystal improvement for performing X-ray diffraction.

3.11.3Crystal data collection and processing

X-ray diffraction data collection is performed with frozen crystals at -160^{o}C to prevent X-ray crystal damage. Crystals must be transferred to a cryoprotectant solution prior to its freezing (Rodgers, 1994). We usually dialyze the crystals against a solution that contains cryoprotectant mixed with the crystallization solution. These cryoprotectants include ethylene glycol, or glycerol. The amount of cryoprotectant was screened beforehand using a liquid nitrogen stream.

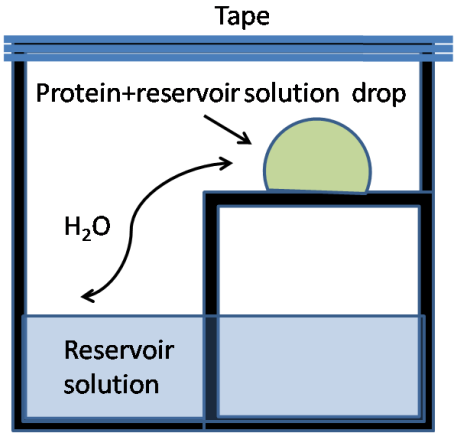


Figure 3.5. Sitting drop vapour diffusion method of protein crystallization. Equal volume of purified protein and reservoir solution is placed over a platform in the sub-well of a 96 well crystallization plate. The well contains the reservoir or crystallization solution which contains precipitants at higher concentration than in the drop. The phenomenon of crystallization occurs during gain of solution equilibria in the system which occurs by gradual vapour diffusion of water and/or precipitant from the drop solution. This renders the proteins supersaturated, and thereby crystallized with an appropriate crystallization solution (Bergfors 1999).

Using X-ray synchrotron sources at ESRF (Grenoble, France), Swiss-SLS facility and ALBA (Barcelona, Spain), diffraction data were collected for single protein crystals. When a crystal is exposed to an incident X-ray beam, the result is a diffraction pattern, which consists of millions of spots that constitute the raw data. Each of the signals is determined by three parameters that are necessary to obtain the electron density map. These parameters are the position, intensity and phase of the waves forming each signal. Diffraction data are first processed to determine the index (hkl values) and respective intensities of the diffraction spots. We used the MOSFLM (Leslie, 1992) and XDS (Kabsch, 1993) processing programs. The objective of these processing was to acquire the index (hkl values) and respective intensities of the diffraction spots. The data were subsequently scaled using the program SCALA (1994). The scaling process sets all diffractions measurements on a common scale and produces an internally consistent set of data (Evans, 2006).

3.11.4 Structure solution and refinement

Crystal structure determination needs the resolution of the phase one of the major drawbacks of X-ray crystallography. To solve this problem, the two techniques frequently employed are Multiple Isomorphous replacement (MIR) and Multiple-wavelength Anomalous dispersion (MAD). MIR is based on the incorporation of heavy metals in the crystallized protein without changing its native structure, and hence the crystals are isomorphic. Data for both the native and heavy atom incorporated crystals are collected and then the location of the heavy atom in the structure is determined using Patterson difference map. This helps in elucidating both the amplitude and the phase of the heavy atom. MAD (Hendrickson and Ogata, 1997) method however records anomalous diffraction at different wavelengths of the coherent X-rays at the synchrotron. Another method that does not require the use of heavy atoms is Molecular r Replacement (MR). This method makes use of the structural similarities shared by the members of a family of proteins. A known structure can be used to determine a similar crystal structure. The position and orientation of the unknown molecule within the unit cell can be determined using the MR method and thus solve its structure.

The results of phase and structure determination is a three dimensional maze of protein electron clouds known as the electron density map. Based on this map and with the amino acid sequence, an initial model of the protein structure is constructed, which after undergoing a series of refinement results in a final model. This latest model or protein structure consists of the Cartesian coordinates of each atom of the molecule models except the hydrogen atoms. An important property of the final model is the resolution, that is, the minimum distance can be defined. The unit of this parameter is the Amstrong (Å), and the lower the value will be higher the resolution, i.e., the electron density map will be of higher quality.

Following above methods, structure of a Seleno-methionine (Se-Met) derivative of pAPN was solved using MR and Single-wavelength Anomalous Dispersion (SAD) methods. This Se-Met pAPN protein was deglycosylated by *Endoglycosidase*-H treatment (pAPNeh). MR was carried out with the program PHASER (Read, 2001) using domain I, II and -III of the tricorn interacting factor F3 (TIF3, PDB code Z1W1) as search models. TIF3 domains share ~30% residue identity with the N-terminal three domains of the pAPN ectodomain. This gave a partial structure which was then completed following the MRSAD protocol implemented in the Auto-Rickshaw server (Panjikar et al., 2009). The final structure included two pAPN molecules in the asymmetric unit, which were then adjusted manually and refined with phenix.refine (Adams et al., 2010) using data extending to 2.5 Å resolution. The pAPN final

structure comprises residues 60-963 of its ectodomain and a zinc atom coordinated in the enzyme's active site. The pAPNeh structure formed the basis of MR applied to determine structure of the other pig and human APN ectodomain structure. Two ensembles including either DI-DIII or isolated DIV were used for structure determination with PHASER. Structures were refined using phenix.refine. In all APN structures, the N-terminal 25-30 residues were very disordered and hence were not included in the final models.

The structure of the co-crystallized complex that includes the PRCV-RBD bound to the pAPN was determined by MR using known pAPN structure (described above) and the TGEV RBD as search models. The TGEV-RBD structure was previously determined by the group in complex with the Fab fragment of the neutralizing 1AF10 mAb (PDB code 4F2M). MR solutions were obtained for the two pAPN molecules (chains A and B) in the asymmetric unit and for one RBD molecule (chain E). The three molecules were manually adjusted and refined using the phenix.refine program. The second RBD molecule (chain F) bound to pAPN molecule B was built manually into the electron density map. The N-terminal 282 residues in the PRCV SH3 variant were highly disordered or degraded during crystallization, and they are absent in the structure. The complex structure was refined with the phenix.refine applying solvent corrections, NCS, refinement of individual coordinates and atomic displacement parameters combined with TLS. The current model includes amino acids 60-963 of the pAPN ectodomain with the zinc metal ion in the catalytic site, and residues 283-426 of the PRCV S (homologous to TGEV S residues 507-650 that defined the TGEV RBD structure). All the residues are within the allowed regions of the Ramachandran plot.

CHAPTER 4

4. RESULTS

4.1 Expression and purification of soluble APN ectodomains and CoV spike protein.

4.1.1. Cloning and expression

APN and CoV S proteins are heavily glycosylated, so we have produced them in mammalian cells. The cDNA coding for the APN ectodomains, including residues 36-963 of the pig (pAPN) and 33-967 of the human (hAPN), were PCR amplified from their respective cDNA templates (Fig.3.2A). Similarly, the coding sequences for different length variants of TGEV and PRCV CoV S (Fig. 3.2B) were prepared using the respective cDNA (Sánchez et al., 1990). Recombinant cDNA amplicons for APN and a small TGEV spike length variant RBD (Fig. 3.2B) were individually introduced into a vector, pBSKS-D (Fig. 3.1A) bringing them in frame with an exogenous N-terminal IgK leader sequence (for cellular secretion) and an HA epitope. The PRCV S length variants (Fig. 3.2B), SH1 and SH3 (both containing the antigenic site A) contained an endogenous secretion signal peptide at their N-termini. The recombinant DNA were then transferred to vector that was derived from pEF-BOS (Fig. 3.1B) and cloned in frame with a Flag or HA epitope (monomeric), or the human IgG1 Fc region (Fc, dimeric) at the C-terminus. Thrombin recognition sequence was introduced between the protein and the C-terminal tags. S glycoprotein of PRCV exhibits 96% sequence identity to corresponding fragments in TGEV.

For transient protein expression, recombinant pEF vectors were transfected into HEK293T cells following the calcium phosphate method (Jordan et al., 1996). After 3-4 days of protein production in either DMEM supplemented with 10% FCS or serum free optiMEM media, protein concentrations were determined using ELISA with antibodies that bind to the CoV S or to the different tags.

Having confirmed expression in transiently transfected HEK293T cells, the recombinant cDNA in pEF vector were cut with *Sal*I-*Not*I enzymes and transferred into the unique *Xho*I and *Not*I sites in pBJ5GS vector (Fig 3.1C). These recombinant pBJ5GS constructs were used to develop stably expressing CHOLec cell clones following *glutamate synthetase* expression system. ELISA was performed to screen for protein expressing clones growing at 30µM

MSX. Protein expression was scaled up by culturing highly expressing clones in roller bottles, with periodic collection of supernatant and replenishing with fresh selection media.

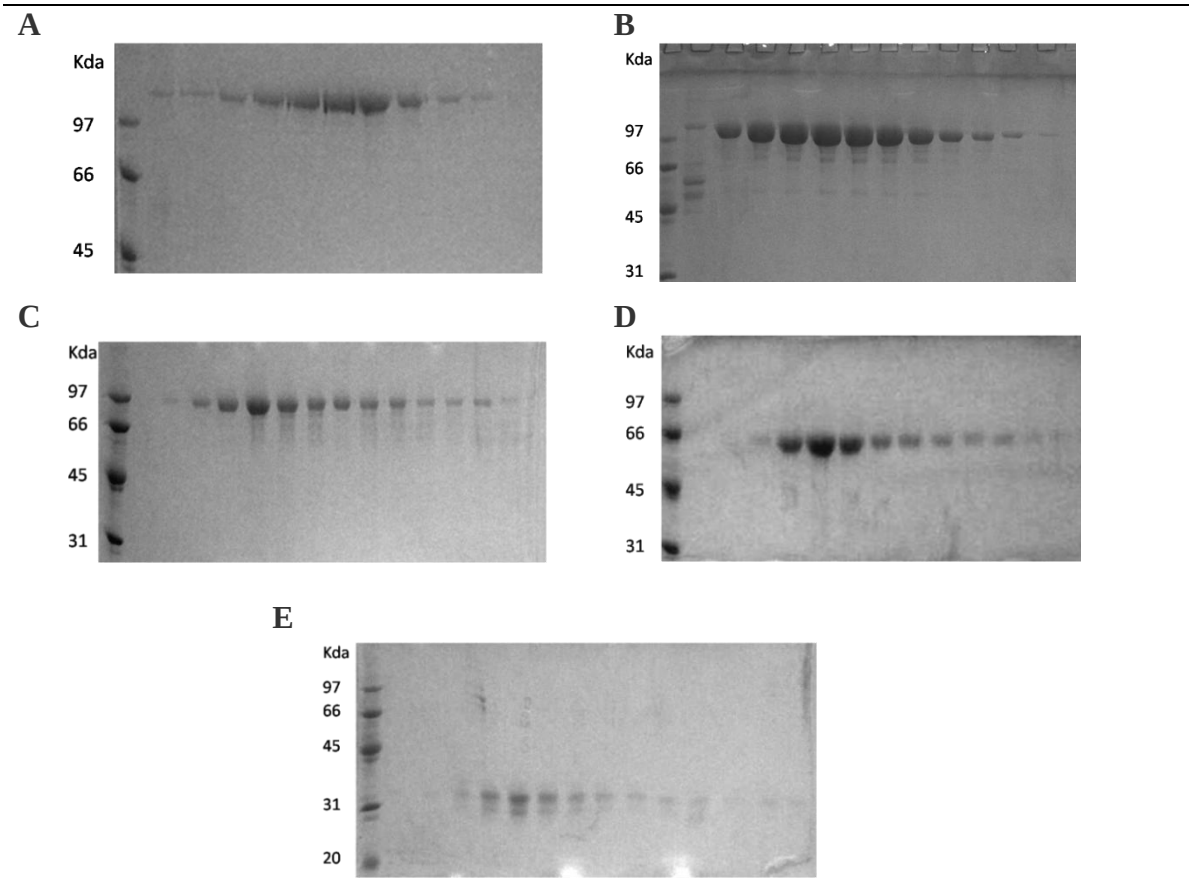


Figure 4.1. Analysis of affinity purified recombinant proteins. 10% SDS-PAGE of elution fractions collected from mAb-Sepharose columns. Loaded with supernatants from CHOLec cell transfectants expressing pAPN (**A**), hAPN (**B**), PRCV SH1 (**C**), PRCV SH3 (**D**) and TGEV S RBD (**E**). Size of the molecular weight markers loaded in the left track is shown.

4.1.2 Protein purification

Supernatants collected from CHOLec cell clone for respective proteins were first purified by affinity chromatography using mAb coupled to Sepharose (Fig. 4.1). TGEV RBD and PRCV SH1 and SH3 length variants were purified using the 6A.C3 mAb (against antigenic site A) while pAPN and hAPN were purified with the HA 12A.C5 mAb. After supernatant loading, the columns were washed with TBS pH 8.3 and the proteins were eluted with 20-40mM glycine buffer (pH 2.8 for 6A.C3 and pH 3.0 for 12A.C5 column) and quickly pH neutralized. The purification efficiency was determined by ELISA. Concentration of the protein in the elution fractions was determined by OD at 280 nm and the size and purity of the proteins by SDS-PAGE (Fig. 4.1).

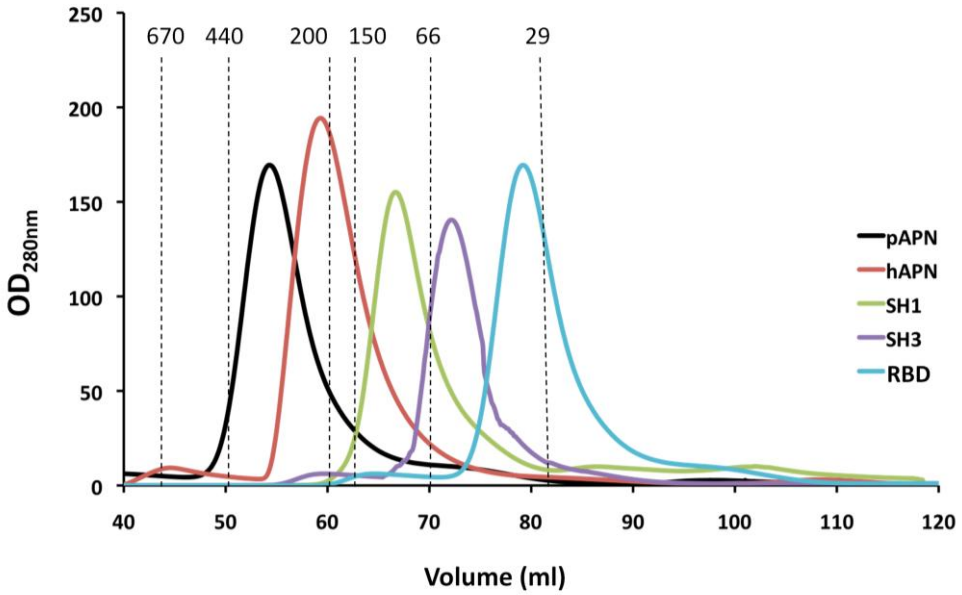


Figure 4.2. Size exclusion purification of recombinant proteins. Overlaid chromatograms recorded during purification of the indicated soluble proteins run through a Superdex 200 (16/60) column. Exclusion volumes and molecular weights (kD) of markers indicated.

Final purification of the affinity purified protein samples were performed following size exclusion chromatography on Superdex 200 column fitted to an "Akta-purifier" (GE healthcare) (Fig. 4.2). The mobile phase buffer included 20 mM HEPES, 150 mM NaCl (HBS), pH 7.5. Concentrations of the purified proteins were determined by OD at 280 nm.

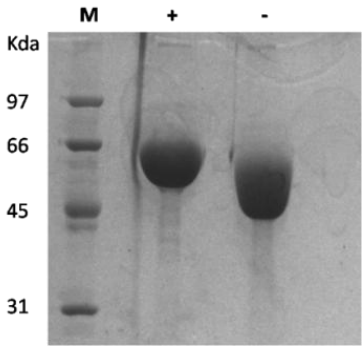


Figure 4.3. Deglycosylation of PRCV SH3 length variant. 10% SDS-PAGE of *Endoglycosidase*-H treated (+) and untreated (-) SH3 length variant. Protein was expressed in CHOLec cells. Size of the molecular weight markers loaded in the left track is shown.

Purified PRCV SH1 and SH3 depict a slightly higher molecular weight in SDS-PAGE than expected likely related to glycosyltaions (Fig. 4.1 and Fig. 4.4). Proteins were prepared in CHOLec cells so that they should bear high mannose N-linked glycosylation sensible to

Endoglycosidase H (Endo-H). This was verified by overnight Endo-H treatment at 30°C followed by SDS-PAGE analysis (Fig. 4.3). In the size exclusion chromatograms, the APN apparent molecular weight (Fig. 4.2) was about twice of their size in SDS-PAGE (Fig. 4.4), indicating protein dimerization in solution. The dimerization of the pAPN and d hAPN protein in solution was confirmed by analytical ultracentrifugation (data not shown).

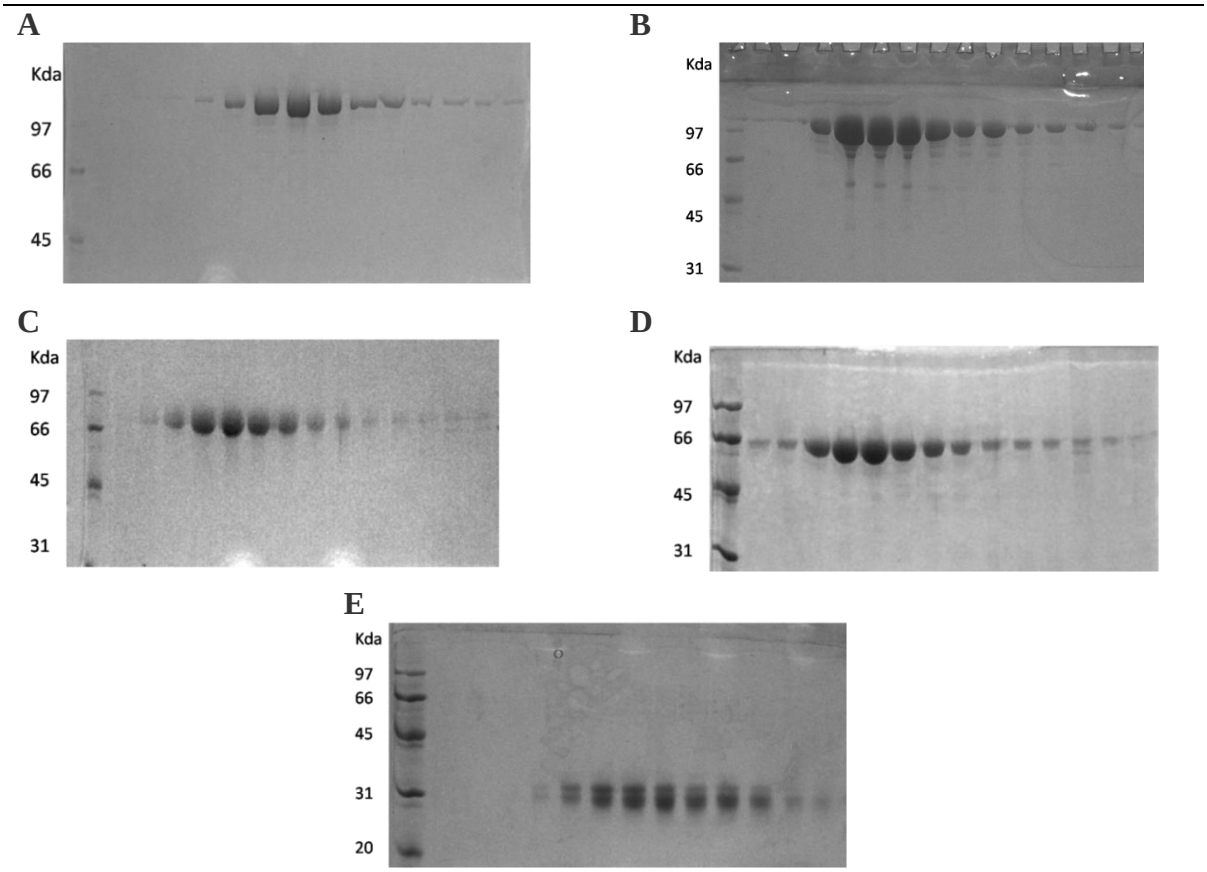


Figure 4.4. Analysis of the size exclusion purified recombinant proteins. 10% SDS-PAGE of elution fractions collected from size exclusion columns of affinity purified pAPN (**A**) and hAPN (**B**), SH1 (**C**), SH3 (**D**), and TGEV RBD (**E**). Size of molecular weight markers loaded in the left track is shown.

4.2 Crystallization of the APN ectodomain and Structure determination

4.2.1 Crystallization of the isolated APN ectodomains

Chromatograms fractions with homogenous and pure proteins (Fig. 4.4) were pooled and concentrated (~15 mg/ml) using centrifugal concentrator units. Both glycosylated and Endo-H treated pAPN and hAPN were used to set up screens for crystallization trials following vapor diffusion sitting drop method (Section 3.11).

Native glycosylated pAPN was crystallized at 21°C with a reservoir solution containing 20% PEG3350, 100 mM sodium acetate pH 5.6 (Fig. 4.5A). Crystals of Endo-H treated pAPN

(pAPNeh) and its seleno-methionine conjugate (pAPNeh-Se-Met) were obtained at 21^{o}C (Fig. 4.5B) in reservoir solution containing 5% PEG-1000 and 10%PEG-8000. Single crystals of glycosylated human APN were first obtained in a condition with 20% PEG10000, 100 mM HEPES, pH 7.5, but these crystals did not diffract X-rays.

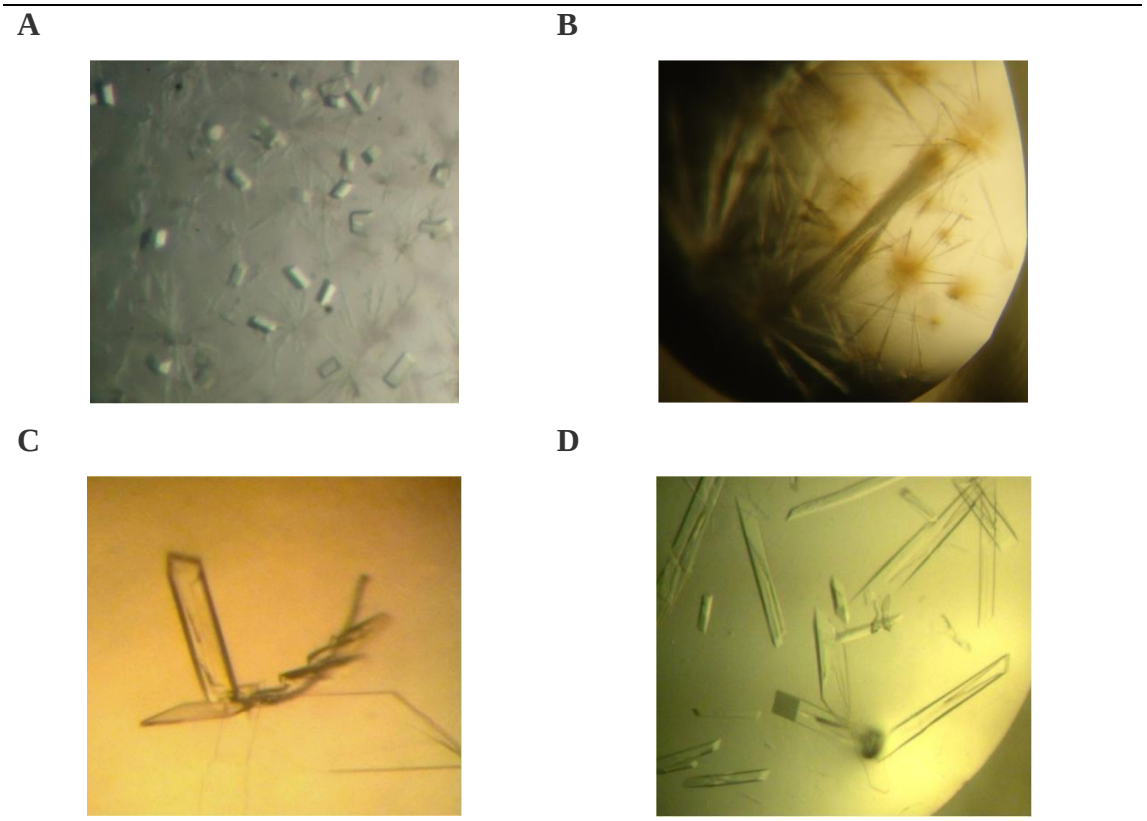


Figure 4.5. Crystallization of APN. A. Cuboidal crystals of native glycosylated pAPN. **B.** Needle-like crystals of Endo-H treated pAPN. **C and D.** Plate shaped and cuboidal crystals of glycosylated hAPN with 1mM curcumin.

Besides crystallizing APN alone, I carried out co-crystallization of hAPN and pAPN with catalysis inhibitors (bestatin, actinonin, and curcumin) and/or CNGRC peptide ligands, but diffracting crystals could not be obtained in most cases. Only human APN gave suitable crystals in a mixture with 1mM curcumin (Sigma) in two different crystallization conditions: i) 20% PEG4000, 200mM lithium sulphate, 100mM Tris-HCl pH8.5 (Fig. 4.5C) and ii) 20% PEG6000, 59mM imidazole-HCl pH8.0 (Fig. 4.5D). Crystals were frozen in crystallization solution containing 20% ethylene glycol as cryoprotectant and transported frozen to the synchrotron for diffraction data collection.

4.2.2 Crystallization of the pAPN in complex with TGEV and PRCV CoV S variants

Sitting drop method was used to prepare crystallization screens of pAPN in complex with SH1, SH3 and RBD length variants of TGEV and PRCV S protein. Complexes were prepared with combinations of either glycosylated and/or deglycosylated forms of both the proteins. Crystals appeared only in trials performed with an equimolar mixture of pAPN and RBD (both in glycosylated form) at a final concentration of 13 mg/ml, and in a crystallization solution of 20% PEG-4K, 200 mM lithium sulfate and 100 mM Tris buffer pH 8.5 (Fig. 4.6). Crystals were transferred to crystallization solution containing 20% ethylene glycol and frozen for diffraction data collection at the ID29 beamline.

Figure 4.6. Crystals of pAPN-SH3 complex. Photographs show multiple crystals (left), bouquet shaped crystals with possible high twinning (middle) which was optimized to single crystals (right) by seeding in a 24 well crystallization plate.

4.2.3 Structure determination

The crystal structure of a Seleno-methionine derivative of pAPN was solved first. Diffraction data at different wavelength was collected with some of the few crystals (~10) obtained with the pAPNeh-Se-Met protein in the ID14-4 beamline (ESRF). Crystals suffer from radiation damage, so that only the Se peak wavelength was collected with three crystals; one complete dataset including data to about 2.5Å was finally processed (Table 4.1). No clear Se peaks were observed in Patterson maps determined with anomalous data. Simultaneously, we started crystal structure determination of the pAPN ectodomain using the pAPNeh-Se-Met diffraction data and the homologous domains of the TIF3 following MR (Materials and Methods). DI-DII-DIII of TIF3 share 30-35% sequence identity with the homologous pAPN domains. This three-domain module was used as a search model and gave two distinct solutions in the molecular replacement search, which corresponded to the two molecules in the asymmetric unit. No solution was found with the TIF3 DIV, which is expected to be divergent from the pAPN DIV. To determine the structure of the complete pAPN

ectodomain we followed the MRSAD protocol of the Auto-Rickshaw server (Panjikar et al., 2009). The resulting model was adjusted manually to the electron density maps and refined with phenix.refine (Adams et al., 2010) using data to 2.5 Å resolution (Table 4.1). The pAPN final structure comprises residues 60-963 of its ectodomain and a zinc atom coordinated in the enzyme's active site.

Diffraction data for the native glycosylated pAPN and hAPN crystals were collected in the ID23-1 (ESRF) and PXII (SLS) beamlines (Table 4.1), whereas the crystal diffraction of the pAPN in complex with the PRCV-RBD was performed in the ID29 (ESRF) beamlines. Two independent ensembles including either DI-DIII or isolated DIV of pAPNeh were used for APN structure determination by MR using PHASER. The TGEV-RBD structure (PDB code 4F2M) was also used for determination of the pAPN-RBD structure such as described in Materials and Methods. Structures were refined using phenix.refine and adjusted manually. In all APN structures, the N-terminal 25-30 residues were very disordered and hence were not included in the final models. The PRCV RBD includes residues 283-426 of the PRCV S (homologous to TGEV S residues 507-650 that defined the TGEV RBD structure). The N-terminal 282 residues of the SH3 protein used for the crystallization of the complex were disorder or more likely degraded. On the portion corresponding to the RBD was defined in the electron density maps after structure refinement. All the residues are within the allowed regions of the Ramachandran plot in the crystal structures included in (Table 4.1).

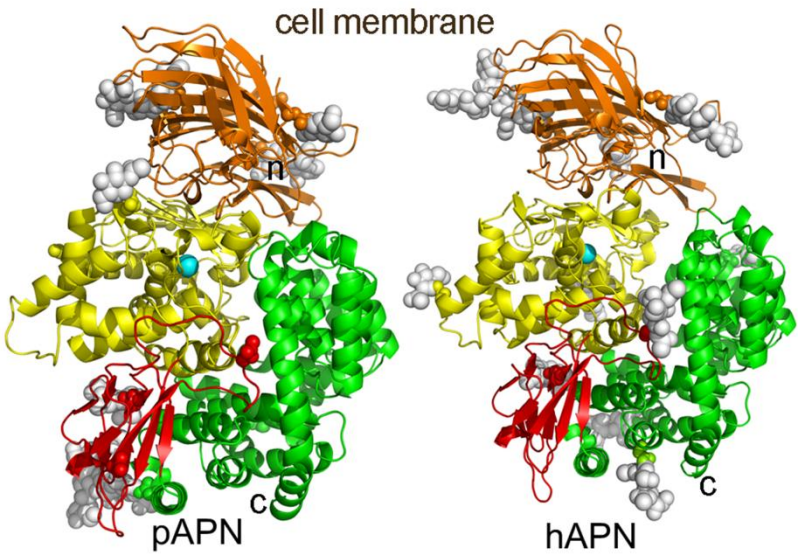
Statistics	pAPN	pAPNeh (Se-Met)	hAPN	pAPN-**RBD**
Data Processing				
Space group	P1	$P2_1$	$P2_12_12_1$	C2
Cell dimensions				
a, b, c (Å)	78.7, 78.8, 223.9	66.5, 215.7, 78.6	127.1, 168.9 ,244.2	220.86, 87.94, 176.91
α, β, γ (°)	99.7, 92.6, 111.3	90.0, 91.9, 90.0	90, 90, 90	90, 90.54, 90
Wavelength	0.87260	0.97950	1.00000	0.97914
Resolution (Å)	25-2.0	25-2.5	20-2.6	25-3.2
Unique reflections	322701	73986	160355	53840
Redundancy	2 (2.0)	4.1 (3.9)	4.2 (4.1)	3.0 (3.1)
Completeness (%)	97.5 (96.6)	97 (95.5)	99.79 (99.9)	96 (97)
R_{sym} or R_{merge}	87.5 (20.7)	10.5 (35.6)	5 (35.3)	5.7 (37.7)
I/ σ(I)	8.5 (3.6)	6 (1.8)	12.3 (2.1)	5.5 (2.1)
Refinement				
Resolution (Å)	25.0-2.0	25.0-2.5	20-2.6	25-3.2
R_{work} / R_{free}	17.14/20.9	19/22.9	18.11/21.5	20.1/24.5
No. of atoms				
Protein	28817	14292	29182	16697
Carbohydrates	726	252	1417	365
Ligands				
Ion (Zn)	4	2	4	2
Water	4899	381	707	3
Average B-factors				
Protein	13	20	55	95
Carbohydrates	32	37	101	125
Ligands				
Ions (Zn)	6	17	43	75
Water	26	20	47	60
R.m.s deviations				
Bond lengths (Å)	0.006	0.009	0.004	0.003
Bond angles (°)	1.022	1.28	0.77	0.880
Ramachandranplot[a]	92/4/0/0	92.4/7.1/0.4/0	97/2.8/0/0	81.7/16.1/2.2/0.0

Table 4.1. Data collection and refinement statistics. Values for highest-resolution shell are shown in parentheses. [a]Percentage of residues in most favored/allowed/generously allowed/disfavored regions are shown.

4.3 Crystal structures of APN ectodomains

The human and pig APN ectodomains exhibit a hook-like topology in crystal structures (Fig.4.7). The N-terminal HA tag and about 30 residues that follow it (Fig. 3.2A) were much disordered indicating a highly flexible region in the membrane proximal end. APN ectodomain is composed of four domains (Fig. 4.7B). The first domain (DI) exists as a twisted β-barrel domain of about 200 residues (Fig. 4.7A and Fig. 4.8). This domain assembles as a packing of one extended β-sheet against two shorter β-sheets with three and four β-strands (Fig. 4.7A). Domain II (DII) adopts a thermolysin-like fold. The DI β-sheet with four strands stacks on the top of a five stranded β-sheet of DII in such a way that the catalytic groove (in DII) is covered from the top (Fig.4.7 B). This groove is formed at the junction between the DII N and C-terminal subdomains (Fig. 4.7A), two long α-helices bearing residues that coordinate zinc ion are at the bottom of the groove. The third domain (DIII) is a small (90 aa) β-barrel domain heavily glycosylated in ectodomain's middle portion, at the turn of the hook (Fig.4.7B). DIII carries a conserved N-linked glycan on a long loop which is disordered in the pAPN. This glycan interacts with domain II and IV (Fig. 4.7B). With about 330 aa (Fig. 4.8), the fourth APN domain (DIV) is the largest among the four domains and is composed of alpha-helices. Albeit, being highly divergent among the M1 family of the peptidases, overall DIV conformation remains preserved. A double layer of helices in DIV adopts a circular configuration forming a bowl-like structure (Fig. 4.7B) (Addlagatta et al., 2008). The N and C-terminal portions of the domain IV contact DIII and they are bridged by the last, long α-helix (α29) of the ectodomain. DIV is an assembly of 8 repeats of a helix-turn-helix motif with either a HEAT or ARM conformation (Nguyen et al., 2011). In APN, DIV has 1 ARM and 7 HEAT repeats. The single ARM repeat is formed by helices α25, α26 and α27 (Fig. 4.7B) and its conformation varies between the hAPN and pAPN ectodomain structures (Fig. 4.8). The ARM repeat is the most variable region between hAPN and pAPN ectodomains, and can contact the peptide substrate bound to the active site (see below). The concave inner side of the DIV bowl faces the DII active site (Fig. 4.6A), enlarging the cavity below the coordinated zinc ion, where the substrate is accommodated during catalysis.

A



B

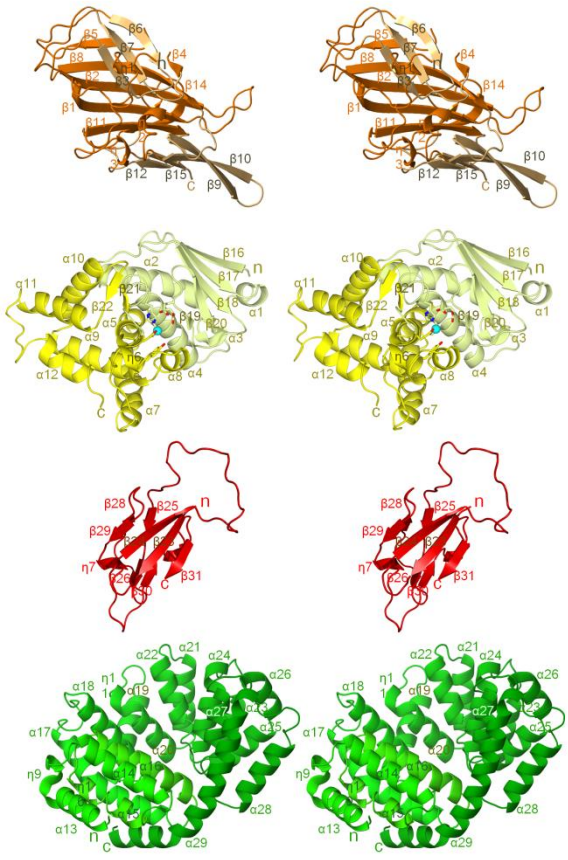


Figure 4.7. Crystal structures of APN ectodomains. **A** Ribbon diagrams of the glycosylated pAPN and hAPN structures with domains (D) in orange (N-terminal DI), yellow (DII), red (DIII) and green (C-terminal DIV), and the N-terminal end (n) near the cell membrane. Asn residues in N-linked glycosylation sites and modeled glycans (grey) in the structures are shown as spheres. The zinc metal ions at the active site in DII are shown as cyan spheres. The approximate location of N- and C-terminus of the structures, and the cell membrane are marked. **B.** Stereo view of the individual domains that compose the APN ectodomains with similar orientation as in A. The three β-sheets of DI and the two subdomains (N and C-terminal) of DII are shown with distinct colors. Helix and β-stands are numbered as in Fig. 4.8.

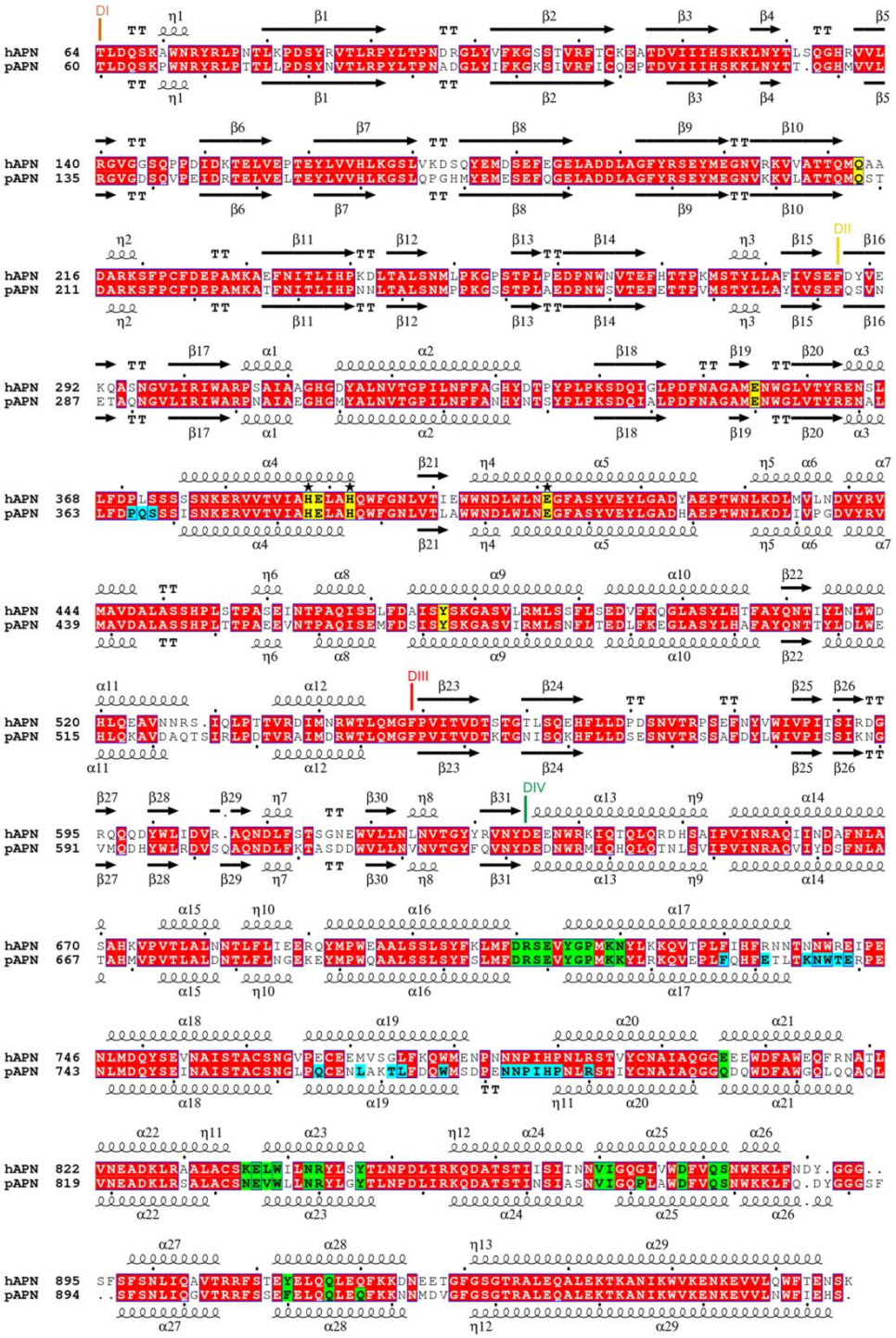


Figure 4.8. Alignment of human and pig APN ectodomains. Structure-based sequence alignment of the human and pig APN with conserved residues highlighted in red. The individual domains of the hAPN and pAPN structures were aligned with the program Modeller using a gap penalty of 3. Secondary structure elements are shown above and below the sequences. Residues involved in peptide binding and hydrolysis at the active site are in yellow, with the residues coordinating the zinc ions marked with an asterisk. Residues involved in protein dimerization and CoV binding are in green and blue, respectively. Each domain's first residue is labeled. Figure prepared with ESPript 3.0 (http://espript.ibcp.fr).

4.3.1 The dimeric conformation of the APN ectodomain

The dimeric APN assembly is well preserved in both hAPN and pAPN ectodomain crystal structures (Fig. 4.9A). The APN ectodomains behave as dimers in solution (Fig. 4.2). APN crystals must thus grow by contacts between preformed dimers as they appear in the crystals grown with different protein samples and crystallization conditions. The exposed convex side of DIV contacts the same face of another DIV in a neighboring APN molecule (Fig. 4.9A). Approximately 950 $Å^2$ of each monomer is buried at the dimer interface, indicative of a stable protein-protein interaction that engaged only DIV residues. The protruding α22-α23 inter-helical region of one monomer contacts the bottom of domain IV bowl of the other monomer, thus closing ~10Å pore towards the internal APN cavity (Fig. 4.9B). Dimerization must prevent substrate access to the catalytic active site through the domain IV pore, an entry point advocated for other monomeric M1 aminopeptidases (Addlagatta et al., 2008). A tryptophan residue conserved in mammalian APN (W839 in hAPN, Fig. 4.8) appears to be a key dimerization residue and contacts seven residues in the neighboring monomer. Dimerization interface and the involved residues are relatively conserved among pAPN and hAPN structures (Fig. 4.8 and Fig. 4.9C). Hydrophilic interactions on the dimerization surface periphery differ between hAPN and pAPN (Fig. 4.9C).

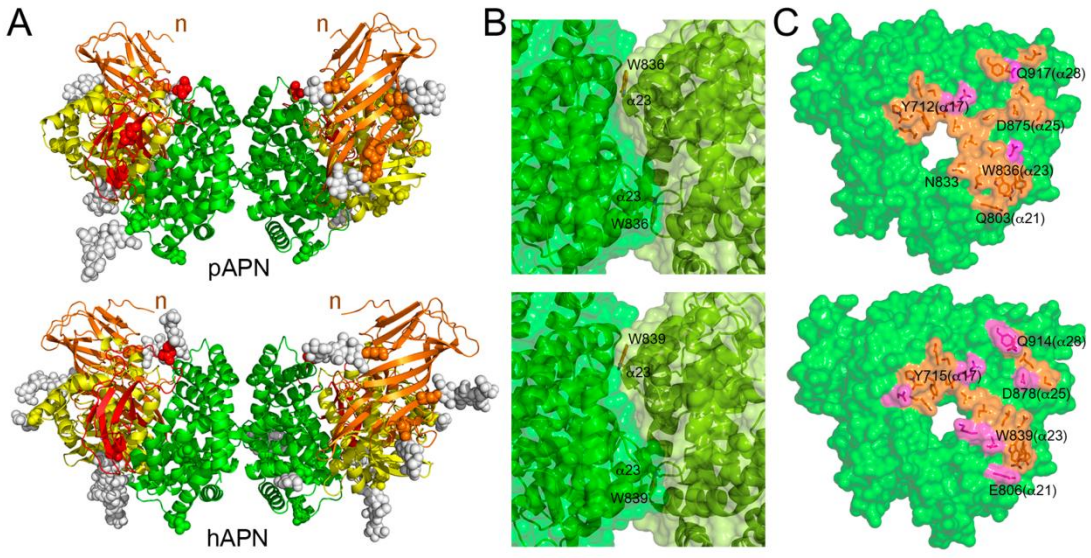


Figure 4.9. Conserved dimeric structures of the mammalian APN ectodomains. A. Ribbon representation of the dimeric pAPN and hAPN structures. Lateral and frontal views of the dimer interface are shown in **B** and **C**, respectively. Side chains of residues buried at the dimer interface are shown with sticks; residues involved in polar or non-polar interaction are with magenta or orange surfaces, respectively. Buried residues by dimerization in the different crystal structures have been shown in Table A (Appendix).

4.3.2 Demonstrating the relevance of the crystallographic APN homodimer

Mammalian APN has been reported to exist as membrane bound dimers on the cell surface (Hussain et al., 1981). Soluble hAPN and pAPN ectodomain dimerization was noticed before by comparing gel filtration (Fig. 4.2) with SDS-PAGE (Fig. 4.4). In all the APN crystal structures presented here the dimeric assembly of the ectodomain is well preserved and might be representative of the functional cell membrane dimers. To verify this crystallographic result, a mutation was introduced at the dimer interface of pAPN using site directed mutagenesis. This mutation substituted a threonine for valine at position V835 (Fig. 4.8) resulting in an N-linked glycosylation at Asn833 (N833glyc), near the center of the dimer interface (Fig. 4.9C). The soluble mutant was produced in HEK-293T cells and its size exclusion profile depicted a monomeric behavior with a molecular weight about half that of wild type pAPN dimer (Fig. 4.10A). The larger size of the N833glyc mutant compared to wild type pAPN in electrophoresis indicates additional glycosylation (Fig. 4.10A, inset).

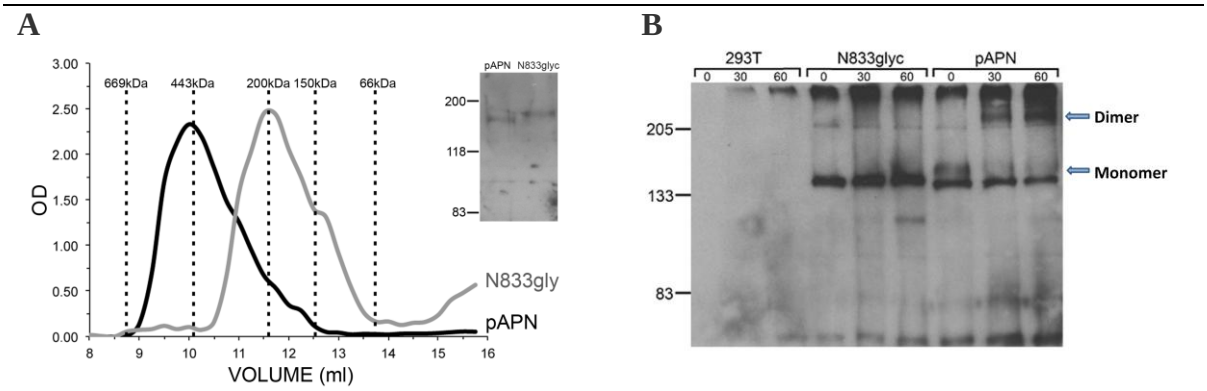


Figure 4.10. Biochemical and immune-analysis of the crystallographic APN dimer. A. Size exclusion chromatography of the 293T expressed soluble ectodomains of pAPN and a mutant (N833glyc) protein with an N-linked glycan in Asn833 at the dimer interface (Fig. 4.9C). Continuous lines show optical density (OD) at 280 nm for the elution volume, determined by ELISA with an anti-HA mAb. The inset shows immune-blot of the pAPN and the N833glyc mutant. Size (kDa) and elution volume/migration of molecular weight markers are indicated in both panels. **B.** Analysis of cell surface pAPN dimerization by chemical crosslinking. Immuno-blot of lysates of BS3 treated 293T expressing membrane bound pAPN wild type and mutant proteins or mock transfected cells (293T). Reaction time (min) is indicated at the top of each track, and migration of molecular weight markers (kDa) are on the left.

Besides, we also analyzed cell membrane pAPN dimerization by BS^3 crosslinking of surface wild type and N833glyc pAPN proteins (Fig.4.10B). A protein band consistent with pAPN dimers was only observed in crosslinked pAPN samples, indicating that inclusion of a glycan in the dimer interface prevents protein dimerization, and demonstrating the relevance of this region in APN homodimer formation.

4.4 The dynamic conformation of the APN ectodomain

Albeit the dimeric hAPN and pAPN assembly was preserved in various crystals, the conformations of the monomers differed among the crystal forms (Fig. 4.11A). The distance between the N-terminus of the ectodomains that formed the dimer increased from 95 to 131 Å in the structures (Fig. 4.11A). DI-DII are distant from DIV in the open conformation of the APN monomer (Fig. 4.11), and the zinc ion at the catalytic site is more accessible to the solvent. The DI-DII-DIII module swings 14^{o} towards DIV, closing the active site (Fig. 4.11B). The hAPN structure adopts an intermediate conformation in the crystals (Fig. 4.11A). The distance between the N-terminus of each monomer in the dimer is 115Å, and the angle difference of DII with the closed conformation in the monomers is approx. 6^{o} (Table 4.2). Other reported crystal structures of hAPN and pAPN also similarly showed closed and intermediate conformations respectively (Chen et al., 2012, Wong et al., 2012). The crystals provided different views of APN ectodomain dynamics. Closed ectodomains crystallized at low pH (<6.0), open at high pH (>8.0) and intermediate conformations crystallized at neutral pH (7.0-8.0) (Table 4.2). The contacts between the domain IV and other domains varied in the different APN conformations. Domain IV contacts with domain I or domain III changed markedly less (approx. 100-200 Å^2 in the buried surface) than with DII (1000Å^2); DII interaction with DIV thus mainly stabilized the closed APN conformation. There were no notable differences in the other interdomain contacts in the distinct APN conformations (Table 4.2).

APN dimerization mediated by DIV-DIV interactions is preserved among open, intermediate and closed APN conformations (Fig. 4.11A). On the cell surface, the DI-DII-DIII module of the ectodomain must swing over DIV, which is fixed by dimerization. The movement of this module must be facilitated by the flexibility of polypeptide of ~30 residues which links DI to the transmembrane domain. The length of this polypeptide must limit the interdomain movement shown here with APN, which is less pronounced than that reported for endoplasmic reticulum aminopeptidase 1 (ERAP-1) (22^{o}). The type of interdomain movement also differs between ERAP-1 and APN. The DIII-DIV module moves together relative to DI-DII module in ERAP-1, while DI-DII-DIII module swings over the fixed DIV in APN. Moreover, the hinge region in ERAP-1 is N-terminal in DIII, while in APN is N-terminal in DIV. The DI-DII-DIII module can pivot at the beginning of the first (α-13) or third (α-15) DIV helix (Fig. 4.9), which are perpendicular to the swing angle (Fig. 4.11B).

These differences in APN dynamics compared to other aminopeptidases probably relate to dimer formation in APN, which is not observed in other members of the M1 family.

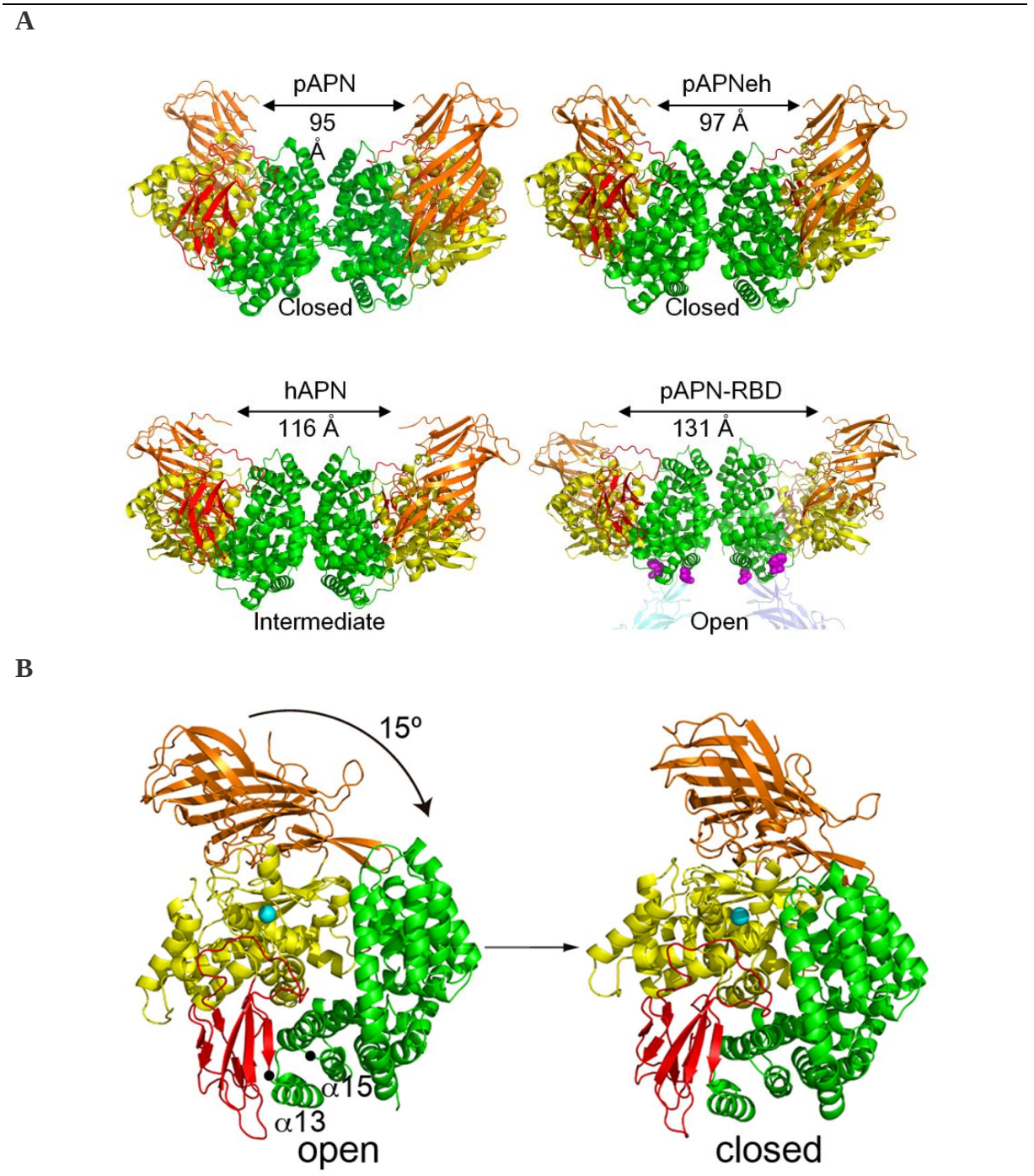


Figure 4.11. Dynamic conformation of mammalian APN ectodomains. A. Ribbon diagrams of the dimeric APN ectodomain structures. Fragments of the CoV RBD bound to the open pAPN are shown in cyan and blue, with the tyrosine and tryptophan residues at the RBD tip as spheres in magenta. Distances between monomers were determined from the N-terminus of the first β-strand in DI. **B.** Interdomain movement between the open and closed APN conformation. The arrow indicates the swing movement of the DI-DII-DIII module toward DIV upon ectodomain closing in each monomer of the APN dimer, with DIV fixed by dimerization. The hinge residues at the N-terminus of α13 and α15 in DIV are marked with circles. Conformations of the ectodomains are shown below the structures.

	Interdomains	Open (pAPN)	Intermediate (hAPN)	Closed (pAPN)
Interdomain surface (Å^2)	I-II	1713	1752	1726
	II-III	1244	1405	1341
	III-IV	1018	1141	1131
	I-IV	--	--	164
	II-IV	761	1130	1763
Interdomain angles		15	6	0
Crystallization pH		8.5	8.0	5.6

Table 4.2. Interdomain buried surface were computed with PISA server. Interdomain angles of the open and intermediate structures with respect to closed pAPN were determined after structure superposition based on DIV. Angles were computed based on the zinc ions at DII and the hinge residue at the C-terminus of DIV with the program coot.

4.4.1 APN dynamics in catalysis

The DII buried surface increases due to its interaction with DIV when the conformation changes from open to closed (Table 4.2), thus reducing accessibility of the active site cavity. It is already known that dynamics in M1 aminopeptidases are necessary for their catalytic activity and that closed conformations are believed to be their catalytically active forms (Addlagatta et al., 2006, Kochan et al., 2011). Though it remains still to be verified, how this structural switch connects to peptide processing. Conformation of the DII residues that participate in zinc coordination and peptide hydrolysis was conserved among the different APN structures presented here (Fig. 4.12A), indicating no changes in active site residues are required for catalysis.

Crystal structures with closed and intermediate APN conformations are available in complex with substrates locked at their catalytic site (Chen et al., 2012, Wong et al., 2012). To analyze peptide processing, we modeled an AngIV substrate crystallized with hAPN {Wong et al., 2012} in the active site of the open intermediate and closed APN structures shown here. In the closed pAPN, Phe893 is placed near the AngIV substrate and appears to lock the peptide in the active site for hydrolysis (Fig. 4.12B). Phe893 is found in the loop connecting α26 and α27 in single ARM repeat (Fig. 4.8) and in the closed structure fits over the substrate molecule hindering peptide translocation for further processing after hydrolysis. To proceed with product release from the active site, the lock needs to be released probably by DII displacement away from DIV. This may in turn facilitate removal of the Phe893 plug. APN dynamics hence must be indispensable for locking, processing and release of hydrolyzed peptide substrate at the active site groove.

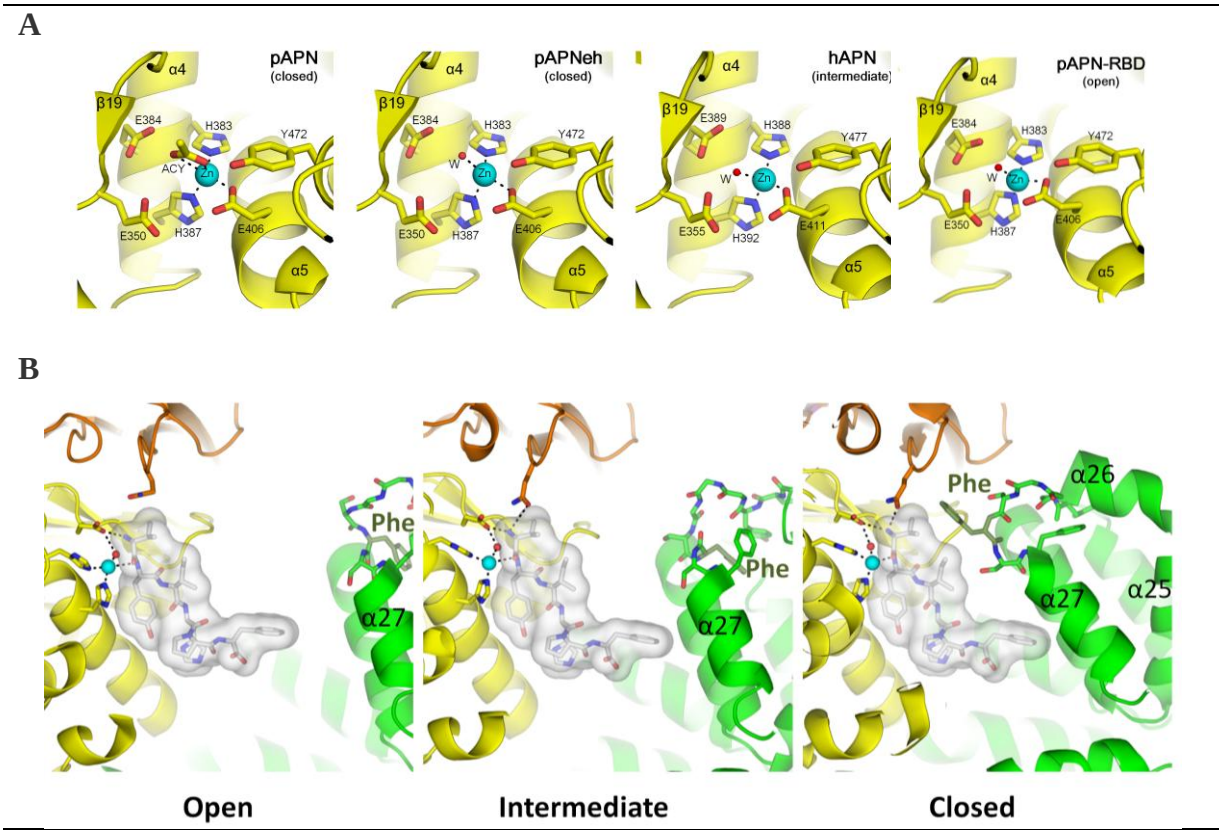


Figure 4.12. APN active site and ectodomain dynamics during catalysis. **A.** Ribbon diagram of the catalytic site of the four APN structures included in this report, with the side chains of DII residues involved in catalysis shown with sticks and the coordinated zinc ion as a cyan sphere. An acetate (ACY) was coordinated to the zinc in the pAPN structure, whereas the other structures had a water molecule (W, red sphere) that is involved in hydrolysis of the N-terminal residue of peptides substrates (Chen et al., 2012). **B.** Changes at the active site during APN ectodomain movements. The active site at DII (yellow) of the structures contains a modeled AngIV substrate coordinated to the zinc ion (cyan sphere). Side chains of APN active site residues are shown with sticks, whereas AngIV is with a gray surface and residues in sticks with carbons in grey. Nitrogens in blue and oxygens in red. Hydrogen bonds are dashed lines. The helices of the ARM repeat (α25-α27) of DIV (green) with the phenylalanine residue that contacts AngIV in the closed conformation are labeled. The crystal structure of the AngIV bound to the hAPN (PDB code 4FYS) was used to model it into the active site of the closed, intermediate and open structures by structural superposition based on DII.

4.5 The RBD structure in TGEV and related coronavirus

The RBD crystal structure of the TGEV S has been previously determined by our group in complex with a neutralizing antibody (PDB code 4F2M). TGEV S RBD adopts a β-barrel fold formed with two highly twisted β-sheets (Fig. 4.13A and C).

Each of these β-sheets is composed of five β-strands. The N and C terminal ends are on the same side of the domain (terminal side) presumably lying close to the other S protein domains. Opposite to the terminal side, two β-turns (β1-β2 and β3-β4) form the tip of the barrel (Fig. 4.13A and C). A bent at the β-strand 5, crosses both the β-sheets and exhibits a β-bulge (colored magenta in Fig. 4.13A and C) at N608. At one side of the β-barrel, all β-strands run antiparallel to each other (colored cyan in Fig. 4.12 A and C), while on the

opposite β-sheet, only β1, β3 and β7 run parallel (in blue). N-linked glycans cluster at one side of the β-barrel. N- and C-terminal ends of the RBD, where other S protein domains lie, are opposite to the tip of the β-barrel. One aromatic residues, tyrosine or tryptophan, protrude from each of the two β-turns at the RBD tip (Fig. 4.13A and B).

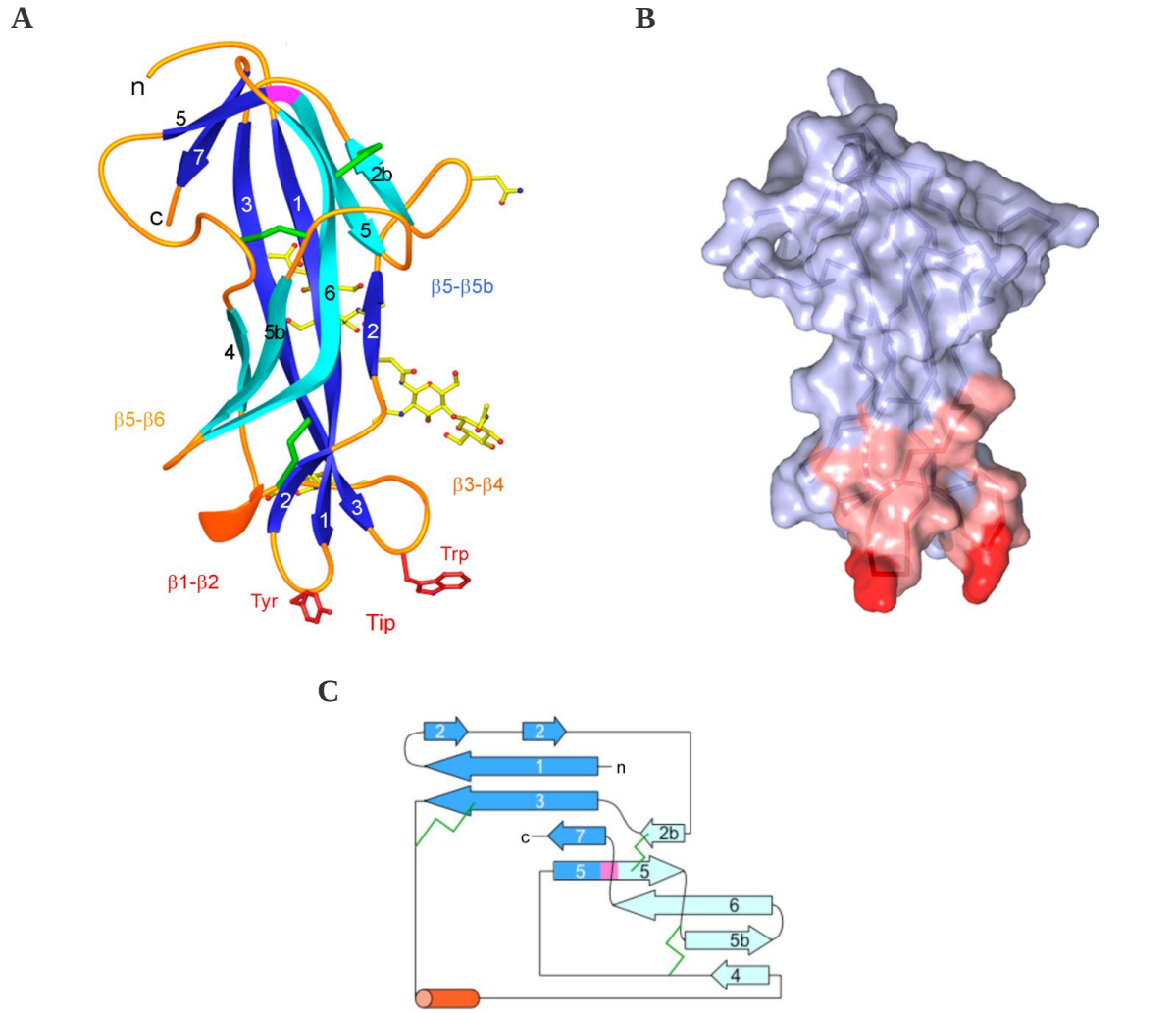


Figure 4.13. Structure of the RBD of TGEV S. A. Secondary structure elements of the RBD structures. β-strands are shown with arrows and colored in blue and cyan, a β-bulge at the β-strand 5 is shown in magenta, helix with a red cylinder, coils with black lines, and disulphide bonds with green lines. **B.** Ribbon diagram of RBD protein structure with β-strands in light or dark blue, coils in orange, and helix in red. A β-bulge at β-strand 5 is shown in magenta. N- and C-terminal ends on the terminal side of the structure are indicated in lowercase letters. The Asn residues at glycosylation sites and the attached glycans defined in the structure are shown as a ball-and-stick model, with carbons in yellow. Cysteine residues and disulfide bonds are shown as green cylinders. Side chains of the pAPN-binding Tyr and Trp residues in the loops at the β-barrel domain tip are shown in red. **C.** Surface representation of the TGEV RBD, residues engaged in direct receptor interaction in pink or red.

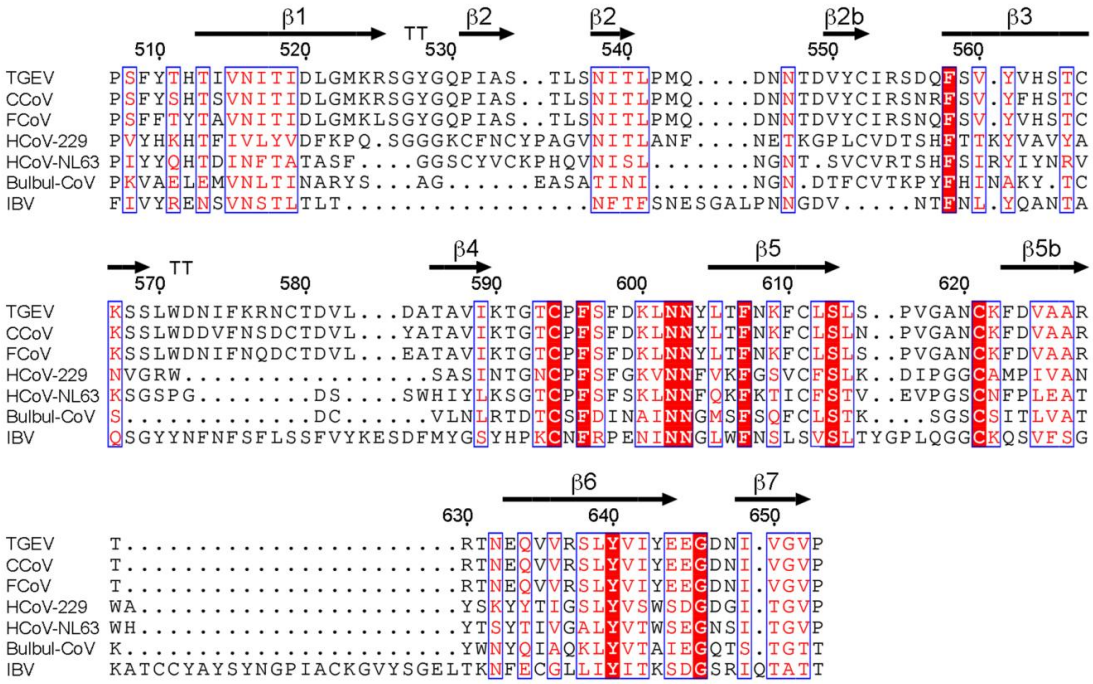


Figure 4.14. Sequence alignment of homologous CoV RBD. Alignment was carried out with the T-Coffee program (http://www.ebi.ac.uk/). RBD sequences of TGEV, canine and feline CoV (*Alphacoronavirus*1), human CoV (*Alphacoronavirus*), Bulbul-CoV (putative *Deltacoronavirus*) and IBV (*Gammacoronavirus*). Sequence of the TGEV RBD is numbered and its β-strands are shown with arrows. Residues in the two turns at the tip of the TGEV RBD β-barrel structure are indicated with a double T.

Sequence alignment (T-coffee program) of the RBD regions from members of *alpha*-CoV1 family (TGEV, CCoV, and FCoV) suggested a structure closely related to that of TGEV, including conformations of the loops (β1-β2 and β3-β4) at the β-barrel tip (Fig. 4.14). In case of the *alpha*-CoV HCoV-229E and HCoV-NL63 the β1-β2 and the β3-β4 loops at the RBD tip are different from the TGEV. The β1-β2 region has two Cys as in HCoV and lacks the exposed Tyr residue (Fig. 4.14). Moreover, the β3-β4 loop is seemingly shorter than that in TGEV. Sequence identities between the RBD of TGEV and IBV (*Gammacoronavirus*) or the Bulbul-CoV (tentative *Deltacoronavirus*) are relatively large (approximately 25%) and similarities are prominent mostly in the β-strands and at the RBD C-terminal half. These data suggest conservation of the RBD fold in *alpha*-CoV and g*amma*- or *Delta*-CoV.

4.6 Crystal structure of the PRCV RBD bound to the pAPN ectodomain

The structure of the pAPN crystallized with the PRCV SH3 protein showed a complex of the RBD bound to the pAPN ectodomain (Fig. 4.15). The asymmetric unit of the crystals contained two macromolecular RBD-pAPN complexes, with the pAPN ectodomain forming a dimer very similar to the others shown in section 4.3.1. On the contrary, the pAPN monomers

in the complex had an open conformation, different to the other APN crystal structures (Fig. 4.10). The PRCV RBD in the structure adopts a β-barrel fold like the TGEV RBD (Section 4.5). The N and C terminal ends of the RBD and remaining S are distant from pAPN (Fig. 4.15A), hence unlikely to contact the receptor molecule. Consistent with a cryo-EM structure of SARS-CoV S (Beniac et al., 2006), the PRCV RBD also must locate at viral-membrane distal region in the S. This should render receptor-binding edge of the CoV S to be accessible to the receptor for CoV binding. Each pAPN molecule was engaged by the tip of a single PRCV RBD molecule, which bears two exposed aromatic residues (Tyr and Trp) (Fig. 4.15A, in red)

The conformations of the receptor binding loops (β1-β2 and β3-β4) at the tips of the β-barrel domains in the structure are identical, suggesting very similar RBD-pAPN interaction in both complexes of the asymmetric unit. The virus–receptor interaction buries ~870 Å2 surface of the virus protein, 60% of which corresponds to the β1- β2 region (Fig. 4.15B) and 30% to the β3-β4 turn (Fig. 4.15C). The pAPN surface buried by the RBD was similar (~770 Å2) and included pAPN residues ranging from alpha helix 19 (α19) to 22 (α22) in DIV, and a few residues in DII (Fig. 4.8, Table 4.3).

The structure identified two critical Receptor-Binding Motifs (RBMs) at the pAPN-RBD binding interface:

RBM 1: This motif involved the end of the pAPN helix α19 and helix α21 contacting the β1-β2 region of the RBD (Fig. 4.15B). The tyrosine side chain (Tyr528 in TGEV) that protrudes from the β-turn in the RBD (Fig. 4.15B and Fig. 4.15D), is fully buried in the complex, locating between the first N-acetyl glucosamine (NAG7361) linked to pAPN Asn736, the end of helix α19, and the first half of the helix α21 (Fig. 4.15B). The hydroxyl group of the RBD Tyr528 was found hydrogen bonded to side chains of the pAPN residues Glu731 and Trp737, and contributed to virus-receptor binding specificity. The preceding residue, Gly527 of the RBD was at the pAPN proximal side of the beta turn, linked by hydrogen bonding to pAPN Asn736 main chain. At the opposite side, the RBD Gln530 side chain formed a network of hydrogen bonds with pAPN NAG7361 and Asn736 side chain (Fig. 4.15B). The N acetyl moiety of the glycan also interacted with the residues at the β2 and β6 strands (Fig. 4.15B, Table 4.3). The pAPN N-linked glycan (Asn736) and surrounding residues contacting β1-β2 region of the RBD in the structure were identified as one of the APN determinants of the CoV host range (Tusell et al., 2007).

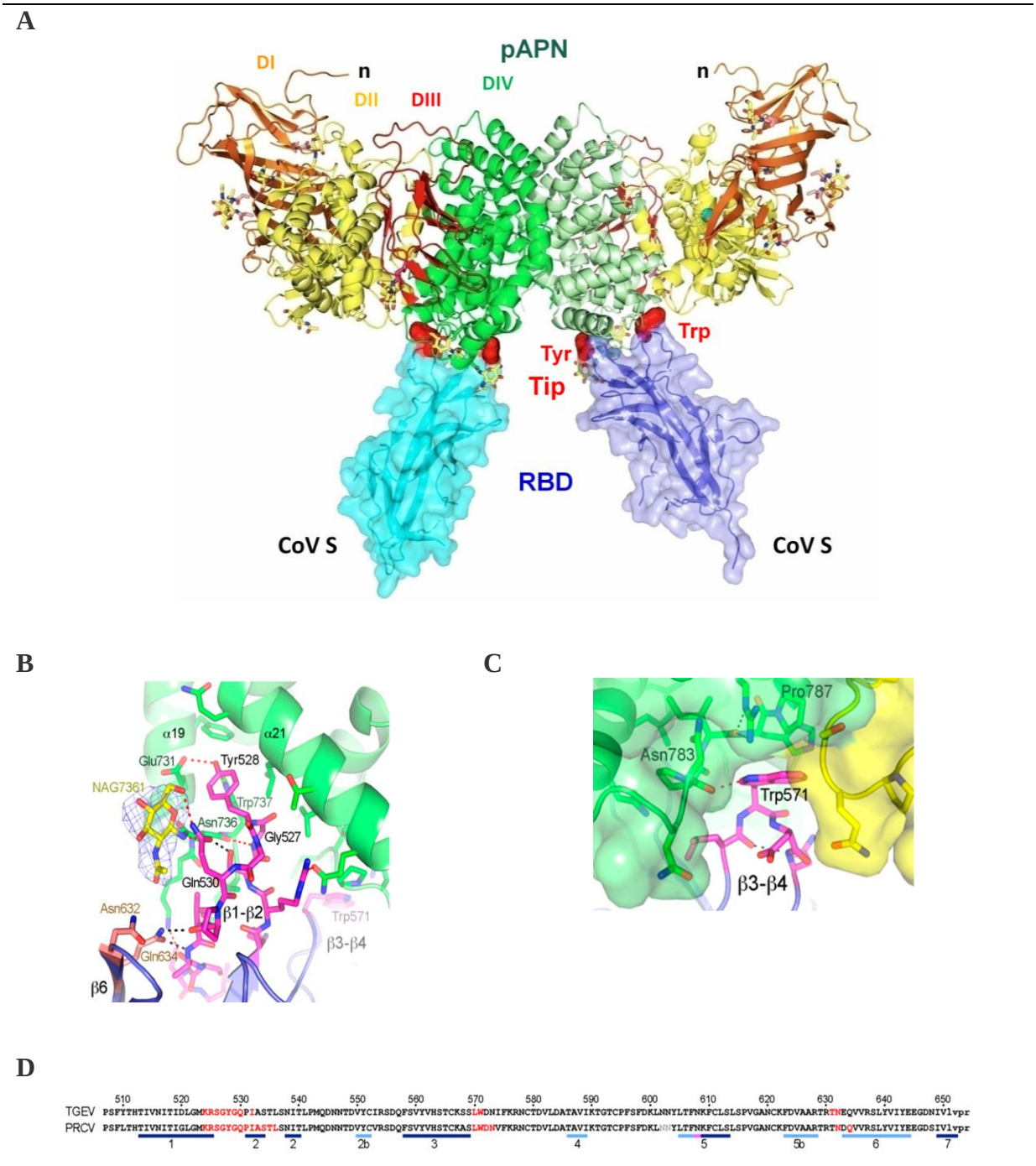


Figure 4.15. Crystal structure of PRCV RBD bound to the pAPN ectodomain. **A.** Dimeric PRCV RBD-pAPN complex in the crystals. pAPN molecules are show as ribbons representations with domains (D) in orange (N-terminal DI), yellow (DII), red (III) and green (C-terminal DIV) and n-terminal end (n) near the cell membrane. The RBD is shown s ribbon and surface drawings in blue and cyan, with the pAPN binding residues at the RBD tip in red. Glycans are shown as sticks with carbon in yellow and zinc ion at the pAPN catalytic site as a cyan sphere. **B.** RBM 1. RBD β1–β2 region with the exposed Tyr residue interacting with the pAPN. Side chains of RBD and pAPN residues engaged in the interaction are shown as sticks with carbons in magenta or green, respectively. NAG7361 glycan N-linked to pAPN Asn736 is shown with carbons in yellow and the electron density map, determined without the glycan, shown as a blue mesh contoured at 3 sigma. **C.** RBM 2. RBD β3–β4 region with the Trp residue interacting with the pAPN. **D.** Structure-based alignment of TGEV and PRCV RBD sequences. β-strands are marked with bars. Residue numbered for TGEV. In red, residues identified by the structure for 1AF10 mAb- (for TGEV) and pAPN receptor-binding. In B and C, RBD residues are numbered following the TGEV sequence shown in D, and intermolecular hydrogen bonds are shown as dashed red lines.

RBM 2: This region involved a β-turn at the start of the β3- β4 loop in the PRCV RBD (Fig. 4.15C and Fig. 4.15D). The unique RBD residue Trp571(in TGEV) protrudes at this β-turn and docks in a cavity formed by the coils preceding helices α22 (in domain IV) and α5 (in domain II) of pAPN (Fig. 4.15C, and Table 4.3). The bulky side chain of this RBD residue packed against pAPN residue His786 and Pro787 and its imino group hydrogen bonded with the main chain carbonyl group of the pAPN residue Asn783 (Fig. 4.15C).

Thus, RBD residues Tyr528 as well as Trp571 at the tip of the β-barrel in TGEV and PRCV appear to be central residues in the CoV-APN interaction; they contact many pAPN residues and contribute also to binding specificities by mediating polar interactions with the pAPN (Table 4.3).

RBD residues	pAPN residues
β1-β2 (RBM1)	
K524/300	T738
R525/301	T738/Q775
S526/302	N736/T738/Q775
G527/303	N736/W737/T738/L768/T771/L772
Y528/304	G727/E731/N736/W737/Q764/N767/L768/T771/NAG7361
G529/305	T771
Q530/306	N736/NAG7361
P531/307	NAG7361
I532/308	K735/N736/T738/NAG7361
A533/309	K735
S534/310	K735
T535/311	K735/E739
L536/312	E739
β3-β4 (RBM2)	
L570/346	Q775/N782/N783/P784
W571/347	P366/Q367/S36/N783/P784/I785/H786/P787/R790
D572/348	Q367
N573/349	T738
β5-β6 (RBM3)	
N632/408	NAG7361
Q634/410	NAG7361

Table 4.3. Intermolecular contacts in the PRCV RBD-pAPN complex structure. RBD and pAPN residues in close contact (≤5 Å) in the two complexes of the crystal asymmetric unit, computed with the program NCONT(1994). RBD residues from the β1–β2, β3–β4 and β5–β6 regions at the tip of the β-barrel domain are shown, with those engaged in hydrogen bonding in red. TGEV/PRCV numbering is given for the RBD residues.

4.7 Critical motifs for coronavirus attachment to APN and infection.

Crystal structures of APN and its complex with the RBD of PRCV defined the CoV-APN binding interface and RBM. Based on these crystallographic results we carried out an extensive research to identify critical residues for TGEV binding to APN and cell infection.

4.7.1 Identification of critical TGEV RBD residues for binding to cell surface pAPN

The crystal structure of the PRCV RBD –pAPN complex defined the virus-receptor binding interface. Subsequently, we characterized the contribution of residues to the interaction. We carried out site directed mutagenesis of RBD residues that contact the APN in the crystal structure of the RBD-pAPN complex (Table 4.3).

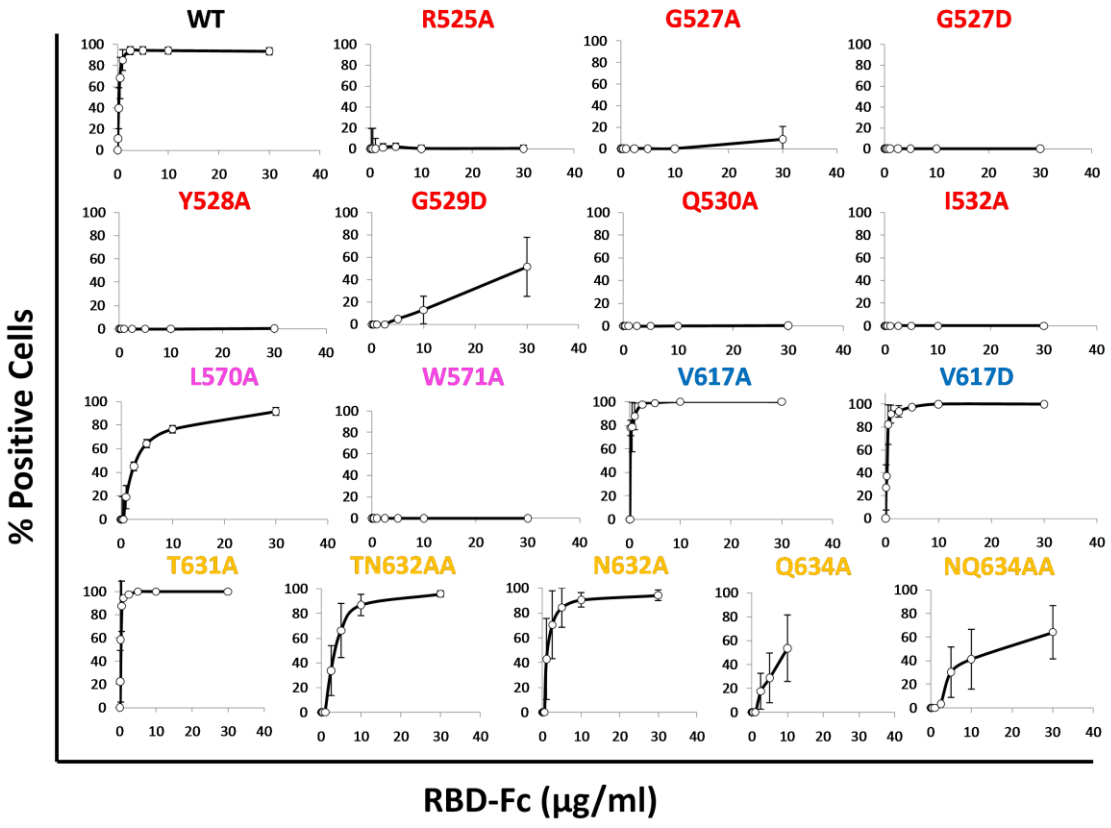


Figure 4.16. Binding of TGEV RBD-Fc proteins to pAPN on BHK cells. Binding determined by FACS. Plots represent the percentage of cells stained with the wild type and the indicated mutant RBD-Fc proteins for a range of protein concentrations. Mean and standard deviation for three experiments.

Mutated residues are in the β1-β2 and β3-β4 loops at the RBD tip, in the nearby β5-β6 loop and in the β5-β5b loop of the RBD, which is distant from the pAPN in the structure. Mutagenesis was done in the mammalian expression vector pEF/RBD-Fc, which express the

TGEV RBD regions fused to the Fc portion of human IgG1 (Materials and Methods). Binding of the RBD-Fc protein to cell surface pAPN was done by flow cytometry (FACS) with stably transfected BHK-pAPN cells. Binding was determined as percentage of cells stained with the wild type and mutant RBD-Fc proteins, using a range of protein concentration (Fig. 4.16). Alanine substitution of either the Tyr (Y528) or the Trp (W571) residues gave undetectable binding of the RBD-Fc protein at the highest concentration tested, which is about 100 times higher than the wild type BC_{50} (Table 4.4). Similar effect was observed only in residues closed to Y528 in the β1-β2 loop, except for the G529D mutation, which decreased 80 times the binding activity. Interestingly, mutation of L570, the residue preceding W571 in the β3-β4 turn, decreased about 10 times binding. Substitution of residues in the β5-β6 loop, nearby β1-β2 (Fig. 4.15B), also had a mild effect (Table 4.4). As expected, residues in the β5-β5b loop, which is distant from the pAPN-binding region in the structure of the complex, do not contribute to the interaction (Table 4.4).

RBD loop	*RBD-Fc mutant*	*BC_{50}(μg/ml)*
	Wild type	*0.318*
β 1-β2 (RBM1)	R525A	>30
	G527A	>30
	G527D	>30
	Y528A	>30
	G529D	24.335
	Q530A	>30
	I532A	>30
β 3- β4 (RBM2)	L570A	4.114
	W571A	>30
β5-β5b	V617A	0.423
	V617D	0.257
β5- β6 (RBM3)	T631A	0.230
	N632A	2,442
	TN632AA	4.227
	Q634A	8.235
	NQ634AA	14.855

Table 4.4. Critical TGEV RBD residues for binding to pAPN. RBD-Fc protein concentration staining 50% (BC_{50}) of BHK-pAPN cells in Figure 4.16 is shown as a measurement of the RBD binding activity. The RBD loops where residues locate are shown on the left.

4.7.2 Identification of TGEV RBD residues essential for cell infection

Following the RBD-pAPN binding assays, we determined the role of certain RBD residues in TGEV cell entry and infection. Recombinant TGEV (PUR46MAD) mutants were generated with mutations in the RBD region of the S, as illustrated before in Figure 3.3. After

confirmation of the mutation by DNA sequencing of the constructs in the pGEM-T vector, the mutated S fragments (4.514 kb) were transferred to an intermediate recombinant BAC plasmid, pBAC-TGEV$^{\Delta claI}$ following sequential digestions with *Pac*I and *Mlu*I restriction enzymes (Fig. 4.17) This recombinant BAC plasmid contained the cDNA copy of the TGEV genomic RNA, except a *Cla*I fragment (5,198 bp in the TGEV genome, Fig. 3.3) (Almazan et al., 2000).

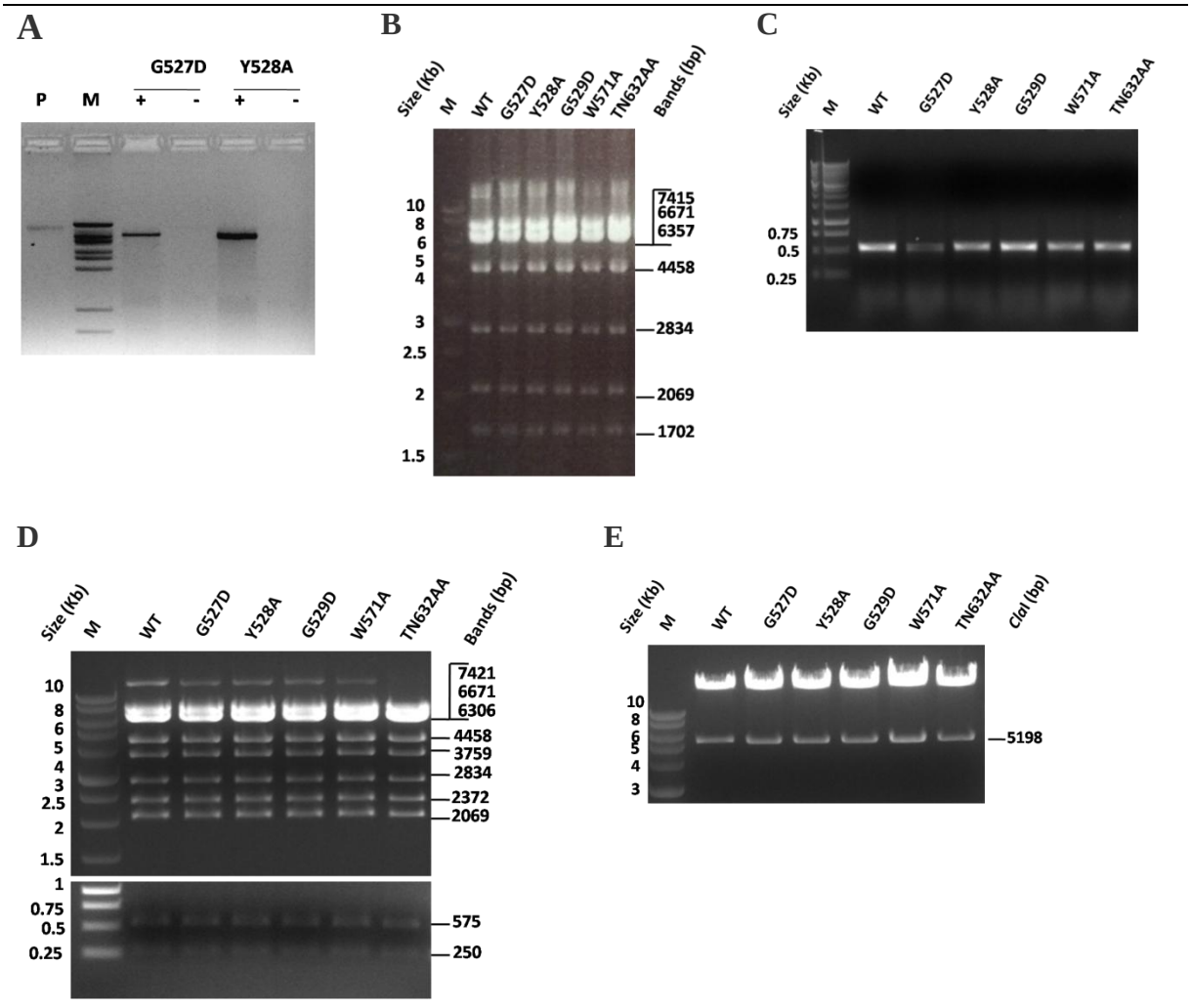


Figure 4.17. Generation of recombinant TGEV BAC clones with mutations in S. Agarose (0.8%-1%)gel electrophoresis. **A.** Site diercted mutagenesis PCR with (+) or without (-) DNA polymerase of the pGEM-T plasmid (P) carrying TGEV S (PUR46MAD). **B.** *Eco*RI-*Xho*I restriction patterns of recombinant pBAC-TGEV$^{\Delta ClaI}$ intermediate clones carrying original TGEV genome and mutant spike introduced via *Pac*I-*Mlu*I transfer. **C.** Colony PCR based selection of final pBAC-TGEV^{FL} clones carrying *Cla*I modification from an infectious cDNA clone. A 500bp amplicon infers correct orientation of the introduced modification. Analysis of genome stability is shown in D and E. **D.** *Eco*RI- *Xho*I restriction patterns of selected clones. **E.** *Cla*I insert release to confirm the maintainence of this modification in the final clones. Size (bp) of the DNA fragments expected from stable clones on the right, whereas the size (Kb) of molecular weight markers (M) on the left.

Stability of the recombinant BAC plasmid was verified by *Eco*RI-*Xho*I restriction pattern analysis (Fig. 4.17B) and DNA sequencing. In the last cloning step a *Cla*I fragment from an

infectious cDNA clone (pBAC-TGEV^{ClaI}) was introduced, resulting in the final construct (pBAC-TGEV^{FL}) for cell transfection. These procedures were followed for each of the intended mutants. Bacterial selection of these final clones is a critical step as their stability is highly sensitive to temperatures of bacterial cultures and orientation of the introduced *Cla*I modification. This *Cla*I modification was found toxic to the bacterial cultures, which exhibited a mixture of small and big colonies. As recommended in the established methods (Almazan et al., 2008), individual small colonies were selected for each mutant and stored frozen.

Colony PCR confirmed the correct orientation of the *Cla*I modification in the selected clones (Fig.4.17C). Low incubation temperature (30^{o}C) was used during bacterial amplification of selected final BAC preparations to minimize toxicity issues. *Eco*RI-*Xho*I restriction patterns verified the stability (Fig. 4.17D) and *Cla*I restriction proved the presence of the introduced modification in the pBAC-TGEV^{FL} (Fig. 4.17E). Large Construct kit method (Qiagen) facilitated the extraction of pure, genomic DNA free pBAC-TGEV^{FL} preparations. Monolayer of BHK-pAPN cells were transfected with at least 4 µg of pBAC-TGEV^{FL} for each construct (*wt* and mutants) using Lipofectamine-2000 (Invitrogen). After 6 hours, these cells were overlaid on fresh ST cell monolayer following brief trypsinyzation. In parallel, another set of transfection was done to visualize cytopathic effects on BHK21-pAPN monolayers. Supernatants were collected two days after transfection and used to infect ST cell monolayers for their titrations following standard plaque forming assays (Fig. 4.18A). Infectious virus units were determined for wild type and TGEV mutants (Fig. 4.18B).

Five TGEV mutants show lack or negligible cell infection, and include mutations in β1-β2 (G527D and Y528A at RBM1), β3-β4 (W571A at RBM2) and into the β5-β6 regions that contact RBM1 and the pAPN NAG (Fig. 4.15). Tyr528 appears the most critical APN-binding residue and no plaques have been observed for deletion of its side chain. Nonetheless, W571 is also a key APN-binding residue for TGEV as infectivity of the W571A mutant gave 8 log decrease in infectivity. We do not fully understand the large effect observed with the N632A and the NQ634AA, which did not give a very significant decrease in RBD binding to APN (Table 4.4). They could affect conformation of RBM1, but further experiments are required to confirm its involvement in TGEV infection. The other mutations had a lower effect (1-3 log) on TGEV cell infection, even those belonging to RBM1 affected significantly RBD binding to APN (Table 4.4). Among those residues Gln530 and Ile532 appear to contribute more than the others, even though they are not required for virus entry

and infection, as Tyr528. Interestingly, the G529D mutant was infectious, as previously shown for a TGEV mutant that escape to mAb neutralization (Gebauer et al., 1991, Delmas et al., 1990).

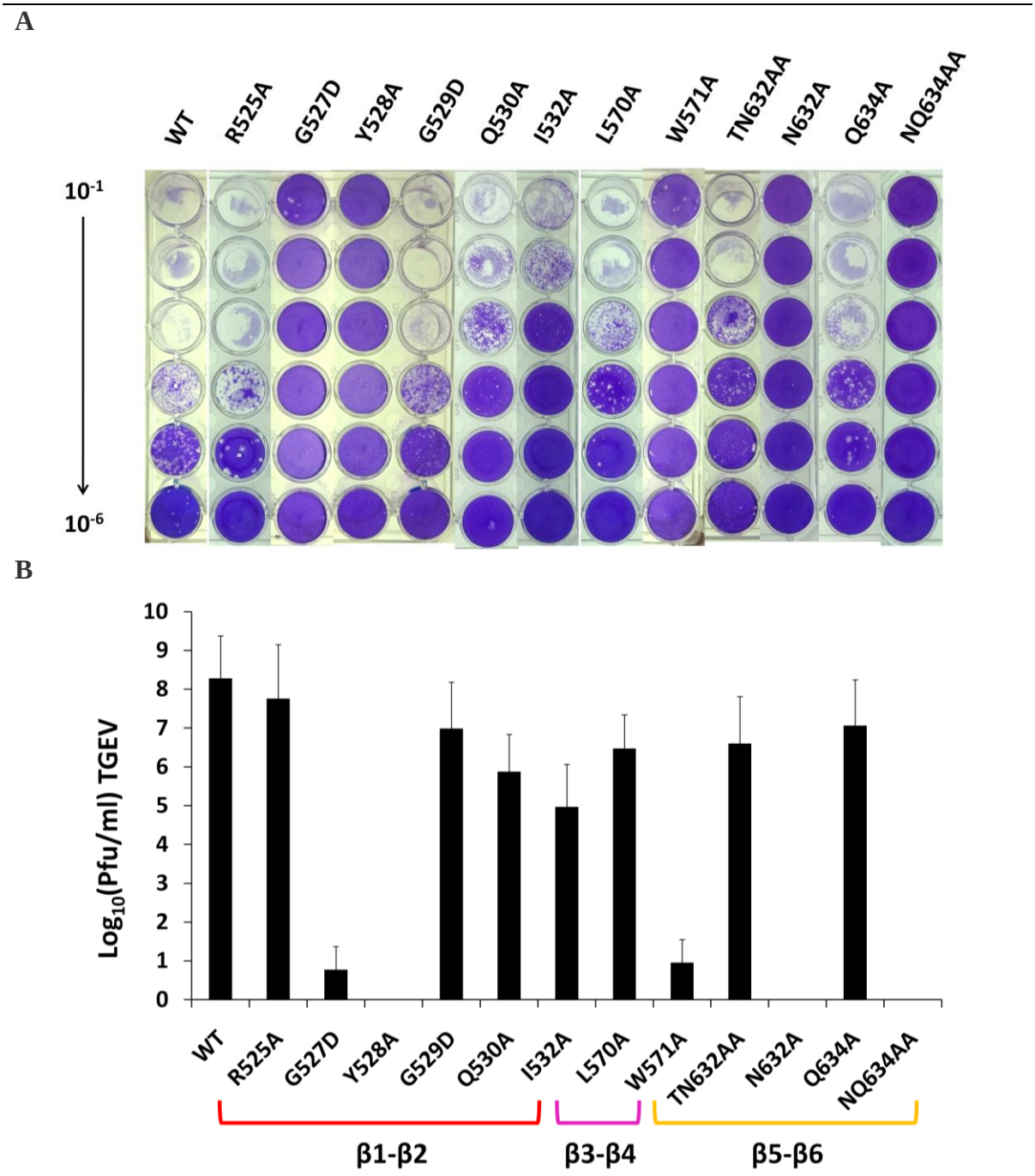


Figure 4.18. Cell infectivity profiling of the TGEV mutants. A. Supernatants from BHK-pAPN cell monolayers, transfected with BAC plasmid preparations of TGEV genome (pBAC-TGEV^{FL}) carrying structure guided mutations in TGEV S, were used in virus titration. ST cell monolayers seeded in 24 well tissue culture plates were inoculated serially with tenfold dilution of the virus stocks and plaques were counted by crystal violet staining (Section 3.10). **B.** Plaque forming assays (A) quantified virus tire (log_{10}Pfu/ml) in virus stocks of *wt* and mutants. RBD residues in which mutations locate are shown at the bottom.

4.7.3 An N-linked glycosylation in the pAPN ectodomain that is essential for PRCV RBD binding

In the PRCV RBD-pAPN crystal structure, the pAPN Asn 736 and the linked NAG carbohydrate bury the RBD Tyr528 and form a network of hydrogen bonds with residues in the RBD β1-β2 loop (Fig. 4.15, Table 4.3). To test the contribution of this N-linked glycan to pAPN DIV, we engineered two pAPN mutants lacking the glycosylation (N736A and T738V). We expressed membrane bound wild type and glycosylation pAPN mutants, and analyzed binding of RBD-Fc to these pAPN mutants. Flow cytometry data indicated negligible RBD binding of the glycosylation pAPN mutants (Fig. 4.19).

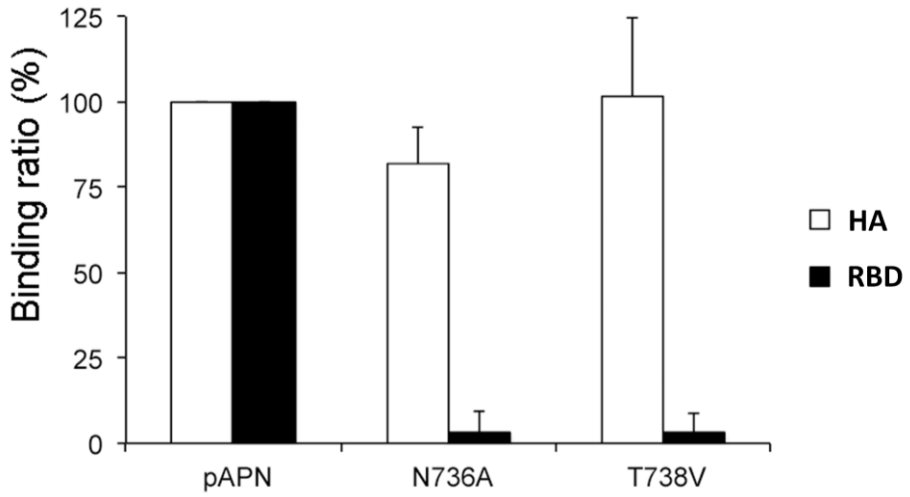


Figure 4.19. RBD binding to cell surface pAPN glycosylation mutants. Relative binding of the RBD-Fc protein and the anti-HA mAb to HA-tagged pAPN proteins with (pAPN) or without the glycan linked to Asn736 (N736A and T738V).). HA binding measures amount of cell surface protein. Mean and standard deviation for three experiments.

This is consistent with the findings on homologous glycan in feline APN (Tusell et al., 2007); its deletion similarly prevented cell infection by feline, canine and porcine CoV, all of which share the glycan-binding (Tyr528) residue (Fig. 4.14) at the β1-β2 turn (Fig. 4.15). Moreover, introduction of this glycan to human APN is known to sufficiently renders it a TGEV receptor (Tusell et al., 2007).

4.8 APN ectodomain dynamics in CoV binding

4.8.1 The TGEV RBD specifically recognizes the open pAPN ectodomain

The porcine CoV RBD binds to a pAPN ectodomains with the most open conformation reported for mammalian APN (Fig.4.11)(Reguera et al., 2012). The CoV-binding region is

distant from the catalytic site, on the opposite side of the ectodomain (Fig. 4.15). A critical APN-binding motif with a tryptophan in the CoV RBD penetrates a narrow cavity formed by DII and DIV (Fig. 4.20). Tryptophan aromatic side chain stacks onto pAPN DIV residues His786-Pro787, and is trapped by DIV residue Asn783-Pro784 on one side and DII residues Pro366-Gln367 on the other (Fig. 4.20). The main chain of DII residues is in close contact (3.9Å) with the tryptophan side chain, and its imino nitrogen forms a hydrogen bond with the Asn783 carbonyl. DIV-based superposition of the structure of open pAPN with bound RBD and that of closed pAPN showed a shift in the DII main chain region that contacts the RBD, this region collides (<3.0 Å) with the CoV tryptophan bound to the open APN (Fig. 4.20). Closing of the ectodomain would thus hinder penetration of the viral tryptophan between the pAPN DII and DIV domains.

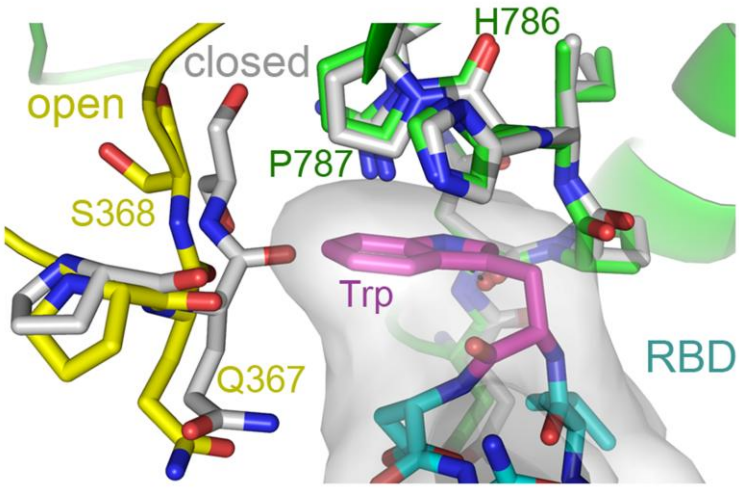


Figure 4.20. Conformation of the CoV binding cavity at the DII-DIV interface in the open and closed pAPN structures. Structures were superposed based on DIV. Ribbon diagrams of the open pAPN in complex with the porcine CoV RBD, with residues that contact the RBD in sticks with carbons in yellow (DII) and green (DIV). The same residues are shown for the superposed closed structure with carbons in grey. The RBD motif that penetrates the pAPN cavity is shown with a grey surface and residues in sticks with carbons in cyan or magenta (Trp). Nitrogens in blue and oxygens in red.

4.8.2 Drugs that bind the catalytic site inhibit TGEV RBD binding to pAPN

Metalloprotease inhibitors are mostly nonhydrolyzable drugs that act by binding the catalytic site. These ligands appear to restrict APN ectodomain movement, as shown by the reduction in number of APN conformation-specific mAb epitopes (Xu et al., 1997). Also, crystals of M1 aminopeptidases in complex with these drugs exhibit preferential closed state (Addlagatta et al., 2006, Kochan et al., 2011, Nguyen et al., 2011). Thus, these ligands would not only

compete with substrate for binding to the active site, but may also restrict aminopeptidase dynamics, a requisite for peptide hydrolysis and release.

CoV binding depends on APN dynamics and transition to an open conformation (Fig. 4.20). Restricted APN dynamics by active site-binding inhibitor ligands would favor an allosteric effect on CoV binding. To verify this hypothesis, the outcome of active site inhibitor drugs on TGEV RBD binding to cell surface expressed pAPN was studied (Fig. 4.21). In flow cytometry, binding of RBD-Fc fusion protein to pAPN expressing BHK cells alone or in presence of actinonin and/or bestatin was analysed (Fig. 4.21A). Both the drugs blocked RBD-Fc binding to APN and effect was found concentration-dependent (Fig. 4.21B). Actinonin has higher affinity for APN (IC_{50} ~1µM) than bestatin (IC_{50} ~20µM) and consequently inhibited RBD binding more effectively.

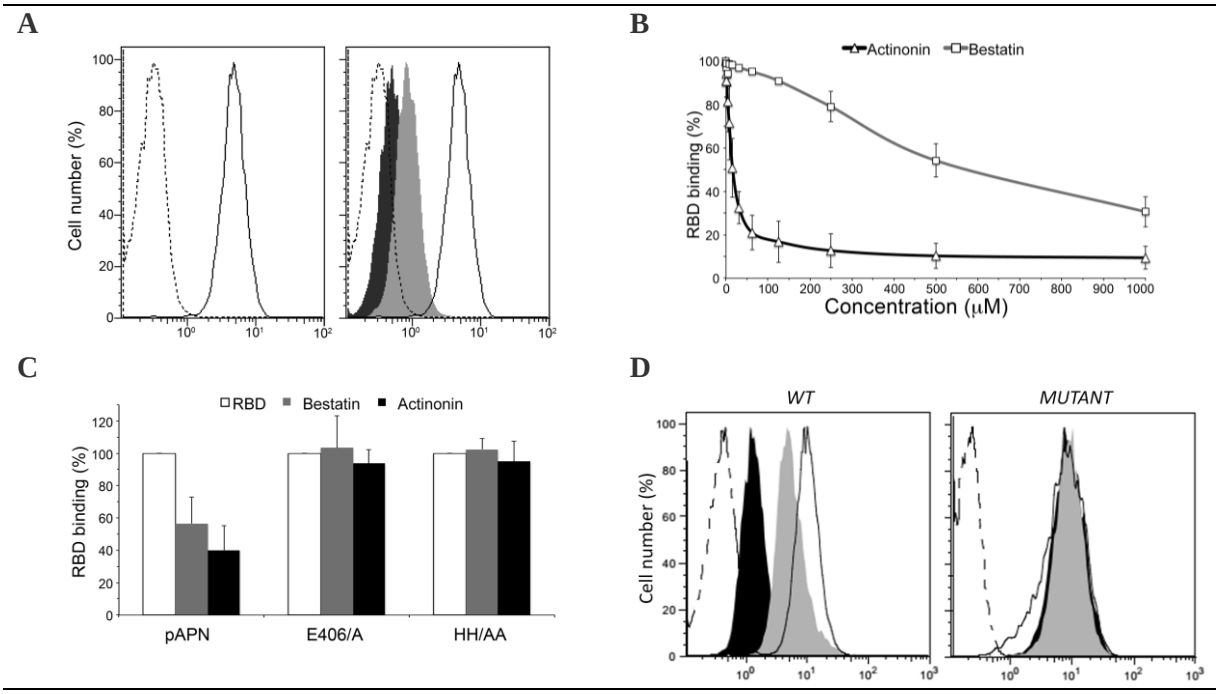


Figure 4.21. Effect of APN catalytic inhibitors on CoV RBD binding to the receptor. A. Flow cytometry monitoring binding of the RBD-Fc protein (1 µg/ml) to BHK-pAPN cells is shown as continuous line histograms on the left. Discontinuous histogram corresponds to an unrelated Fc fusion protein. The vertical dot line marks the threshold fluorescence to compute the percentage of stained cells with the RBD-Fc protein. On the right, overlay plot of histograms obtained with 1 µg/ml RBD-Fc in the absence (white) or presence of 500 µM bestatin (grey) or actinonin (black). **B.** Relative RBD-Fc binding to BHK-pAPN cells in the absence and presence of increasing concentration of bestatin and actinonin. Percentage of positive cells computed by flow cytometry was used to calculate the plotted ratios. Mean and standard deviation of three experiments. **C.** Relative RBD-Fc binding to 293T cells expressing wild type and catalytic mutant pAPN proteins. Binding was computed as in panel B with cell expressing similar amount of cell surface proteins, with (1 mM) and without drugs. **D**. RBD-Fc binding to CHOK1 cells expressing similar amounts of wild type (wt) and/or catalytically inactive HH/AA mutant in flow cytometry as in A.

To determine whether this is an allosteric effect of drug binding to the active site distant from the CoV-binding region, binding of RBD-Fc to two catalytically inactive pAPN mutants was

analyzed on surface of 293T cells, alone or at higher concentrations of both the drugs. Both drugs inhibited RBD-Fc binding to pAPN, although the effect in 293T was less pronounced than that in stably expressing BHK-pAPN cell. The drugs did not inhibit RBD-Fc binding to pAPN catalytic mutants (Fig. 4.21C). Besides, CHOK1 cells stably expressing catalytically defective pAPN mutant (Fig. 4.21D) do not show any inhibition of RBD-Fc binding under drug treatment. These results prove that active site binding drugs cause allosteric inhibition of TGEV RBD binding to pAPN, probably by restricting APN ectodomain opening.

4.8.3 Allosteric inhibition of pAPN catalysis by the TGEV RBD

CoV binds to the open APN and the structure data indicate it might lock the ectoenzyme in its open conformation (Fig 4.20). This would prevent the movements required for substrate entry and release during hydrolysis (Fig.4.20). To verify this hypothesis, the catalytic activity of soluble human and pig APN ectodomains was studied (Fig. 4.22) in presence of pig CoV spike fragments bearing the RBD (shown in Materials and Methods, Fig. 3.2).

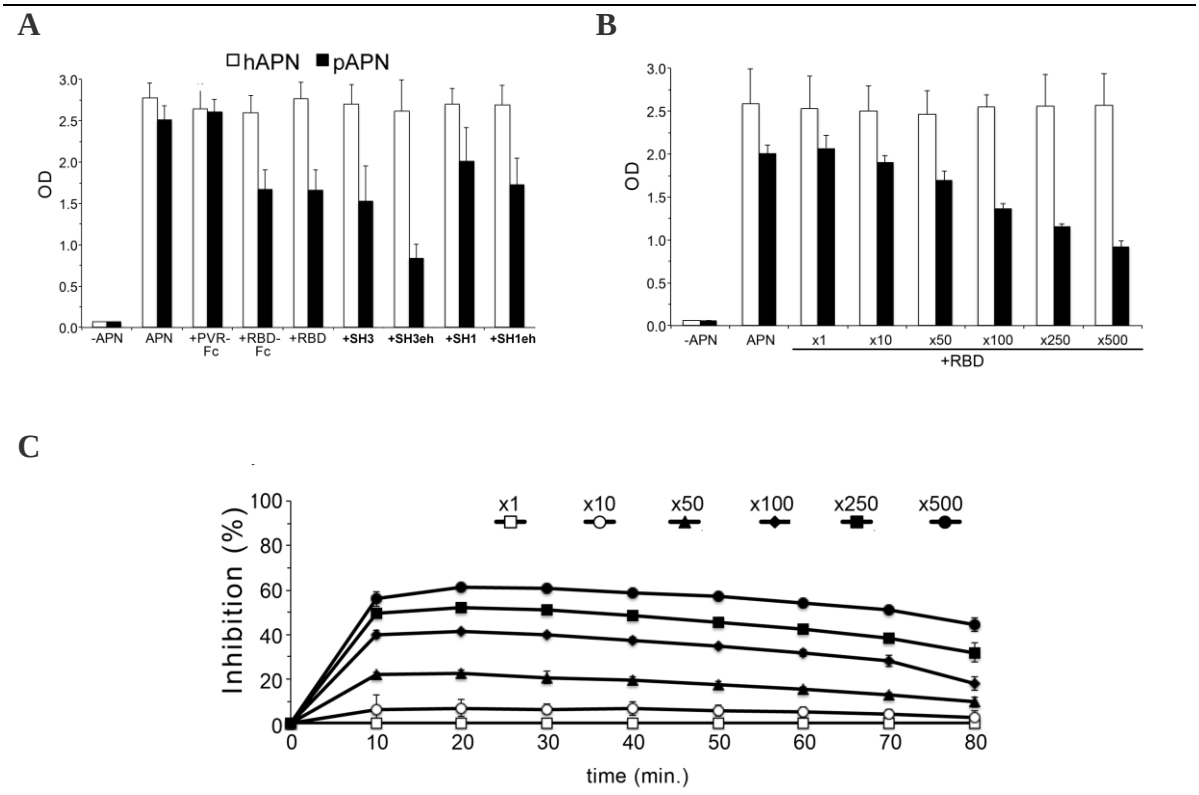


Figure 4.22. Allosteric inhibition of APN catalysis by CoV S proteins. APN catalysis was monitored for soluble hAPN and pAPN ectodomains by measuring the change in optical density (OD_{405nm}) from LpNA catalysis (Materials and Methods, Section 3.8). **A.** Activity determined for pAPN and hAPN (0.04 µM) alone or in the presence of the indicated porcine S proteins (10 µM), which specifically bind to pAPN(Reguera et al., 2011). An unrelated PVR-Fc fusion protein was used as control. **B.** Activity determined with increasing RBD/APN molar ratios. **C.** Inhibition of pAPN catalytic activity by the TGEV RBD with time.

The soluble spike proteins specifically inhibited pAPN catalysis, while hAPN catalysis was not affected (Fig. 4.22A). The TGEV RBD fragment was sufficient to inhibit catalysis of pAPN. Inhibition was directly dependent on RBD concentration and a high RBD/pAPN ratio was needed to achieve maximum catalytic inhibition (60%, Fig. 4.22B), which decreased slowly after 40 mins (Fig. 4.22C), probably due to slow RBD release from pAPN. The RBD thus can specifically prevent catalysis in only a fraction of APN molecules. These APN molecules are probably in open conformation as the catalytically active APN (closed) molecules would not allow RBD binding. The results suggest that the open APN is catalytically inactive and that blocking APN dynamics prevent catalysis.

4.9 Mutations in the pAPN catalytic site affect TGEV cell infection

The effect of APN enzymatic activity on CoV infection has not been studied in detail. A report with one TGEV strain shows lack of effect with transient transfected HEK239T cells, in experiments where the pAPN cell surface expression were not properly monitored for wild type and catalytically inactive mutants (Delmas et al., 1994). We have analyzed TGEV infection of CHOK1 cells expressing similar amounts of wild type and catalytically inactive pAPN on the cell surface.

Two sets of cell clones expressing similar amounts of wild type and the pAPN HH/AA catalytic mutant were selected by flow cytometry, based on similar RBD-Fc binding (not shown). The pAPN-HH/AA mutant lacks the histidine residues that coordinate the zinc atom at the active site (Fig. 4.12). The lack of enzymatic activity of this mutant in CHO cells was verified using the L-pNA substrate in the cell monolayers (Fig. 4.23A and B). TGEV (PUR46MAD) and PRCV (HOL87) were used to infect cell monolayer on 96 well plates at various dilutions and the cytopathic effect was determined (Fig. 4.23C and D). TGEV infectivity of CHO cells expressing the pAPN-HH/AA mutant was reduced (2.5 log) with respect to cell expressing wild type pAPN, whereas PRCV equally infected both cell types. The pAPN enzymatic activity can thus be important for cell infection by certain TGEV strains.

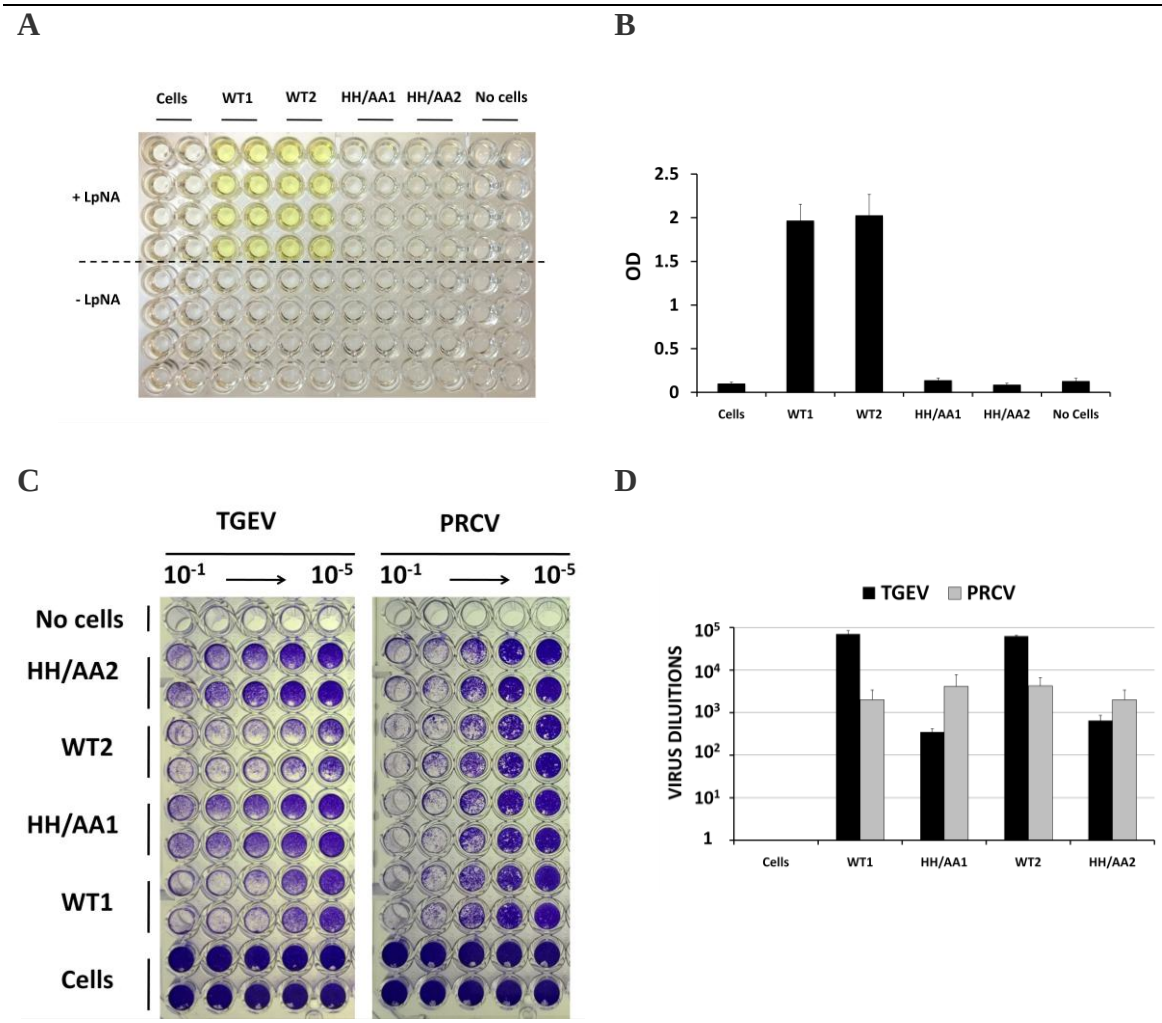


Figure 4.23. APN enzymatic activity and CoV infection. A. ELISA test monitoring the enzymatic activity of CHOK1 cells expressing wild type (WT) and the pAPN-HH/AA mutant (Materials and Methods, section 3.8). Two different pairs of WT and mutant pAPN clones (1 and 2) having similar cell surface expression are shown. Wells with untransfected CHOK1 cells included as control. **B.** OD_{405nm} of samples shown in A **C.** Cytopathic effect (CPE) of TGEV (PUR46MAD) and PRCV (HOL87) infection of the CHOK1 cells monolayers described in A. Virus stock dilutions are shown on the top. **D.** $TCID_{50}$ determined for CPE in C. Infectivity was monitored as OD_{620nm}, and dilution corresponding to 50% cell infection determined as 50% reduction of OD of uninfected cells.

4.10 The pAPN-binding region of the TGEV spike is the main determinant of antigenic site A

Monoclonal antibodies (mAb) that neutralize TGEV infections are known to recognize antigenic site A in the TGEV S (Fig. 3.2B), which has been mapped to the RBD (Gebauer et al., 1991, Sune et al., 1990, Delmas et al., 1990). Sites A includes overlapping but distinct epitopes of several neutralizing mAb, and it was divided into subsites Aa, Ab, and Ac (Gebauer et al., 1991). Previous work by the group showed that a mAb that belongs to site A, 1AF10, recognize the RBD tip (PDB code 4F2M) (Reguera et al., 2012).

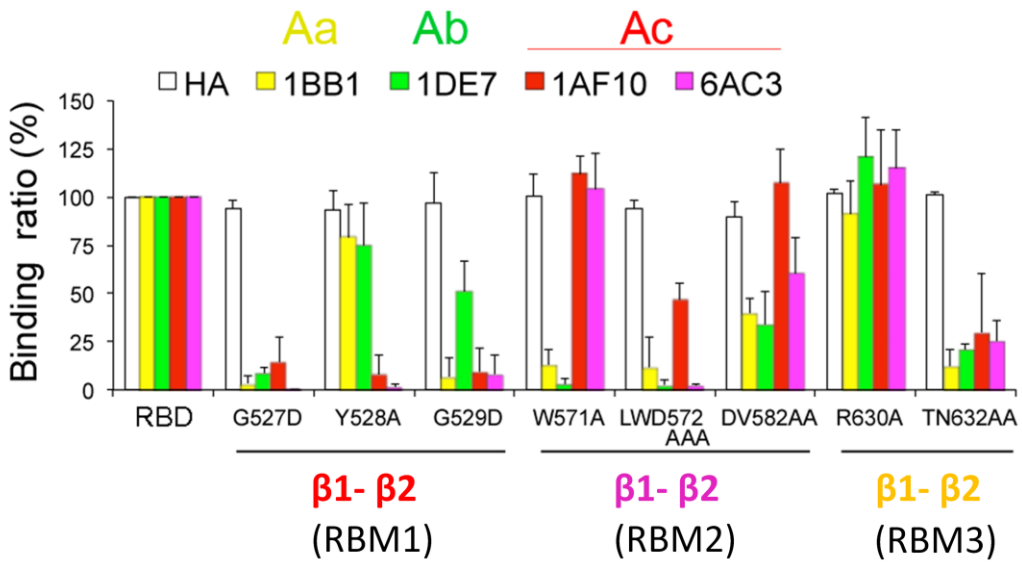


Figure 4.24.Identification of determinants of TGEV S antigenic site A. Normalized binding (%) of site A-specific mAbs to RBD mutants. mAb representative of the three A subsites were analyzed. HA mAb was used as a control. RBD regions in which mutations locate are shown (bottom).

To characterize further antigenic site A and epitopes of other TGEV neutralizing Ab, we analyzed by ELISA binding of several antigenic site A-specific mAbs to RBD mutants. Substitutions of pAPN-binding residues, G527D, Y528A and G529A in the RBM1 of the TGEV RBD (Fig. 4.15) abolished binding to Ac subsite-specific mAbs, 1AF10 and 6AC3 (Fig. 4.24). The Y528A mutant bound subsite Aa- (1BB1) and Ab (1DE7) specific mAbs and G529D mutant also recognized mAbs 1DE7. On the other hand, mutation of the other key receptor binding residue in RBM2 (Table 4.3, Fig. 4.15C), W571 did not affect RBD binding to the Ac subsite specific antibody, but that to Aa and Ab subsite antibodies was greatly reduced (Fig. 4.24).Mutations of three residues at the β3-β4 turn (LWD572A mutant) in the RBM2, greatly reduced its binding to 6AC3 mAb with only partial reduction in 1AF10 binding (Fig. 4.24). This infers that mAb 6AC3 recognizes a broader epitope, which thereby is responsible for its higher TGEV neutralization activity (Sune et al., 1990). Mutation in residues forming the β5-β6 hairpin (T631 and N632), which is close to the β1-β2 loop (RBM1), reduced binding with all site A specific antibody. This broad effect could be accounted for a conformational effect induced at the essential receptor-binding β1-β2 region.

Results from antibody binding assays with RBD mutants infer that site A epitopes encompass the RBD tip region, although some differences exist among the three a subsites. The RBD epitopes recognized by Aa-and Ab-subsite specific mAbs bear the exposed W571 residue at the β3-β4 loop (RBM2), whereas the epitopes for the Ac-specific mAbs center on Y528 in the

β1-β2 turn (RBM1). None of the tested mAbs simultaneously targeted the two aromatic side chains (Y528 and W571) at the tip of the RBD that bind pAPN (Fig. 4.15). Nonetheless, binding to a single RBM is sufficient to prevent TGEV binding to APN and neutralize infectivity

CHAPTER 5

5. DISCUSSIONS

The work presented here delivers compelling structural view of the APN ectodomain; it reveals its architecture, dynamic conformation, and provides new insights on its multiple functions related to peptide catalysis and CoV cell entry. APN is a major CoV receptor and one of the three metalloproteases that this virus family uses to attach to the host. We show how CoV attaches to the cell surface APN and the way CoV neutralizing antibodies prevent infection by inhibition of CoV-APN binding.

5.1 Mammalian expression system for structural biology of membrane protein

X-ray crystallography requires relatively large quantities of proteins for crystallization; glycoproteins must be produced using mammalian cell based expression systems (Aricescu et al., 2006). They have proved very effective, especially for the production and crystallization of secreted soluble glycoproteins (Jones et al., 1992, Ordono et al., 2006, Casasnovas et al., 1997, Wu et al., 1997). Accordingly, work in this Thesis was founded upon preliminary transient expression of proteins of interest into HEK-293T cells which verified their expression and correct folding using biochemical and immuno-assays. CHOLec 3.2.8.1 was the best suited cell line for the large scale production of proteins, due to the low complexity of glycosylation from this cell line ensured by a defective glycosylation pathway (Chang et al., 2007). The proteins secreted from stable CHOLec cell clones exhibit simple high mannose chains with only two N-acetyl glucosamine (NAG) and five mannose residues. This renders their easy deglycosylation by *Endo*-H, which leaves only a single NAG residue in the glycoprotein linked to the Asn residues and preserves glycoprotein conformation.

High expression level in CHOLec cell clones was achieved with the *glutamine synthetase* (GS) system, based on a vector that includes the GS and the gene of interest. Expression is done under glutamine starvation in presence of the methionine sulfoximine (MSX), a GS inhibitor that induces gene amplification and increases protein expression. Besides the large-scale production, the next important hurdle is the purity of the protein of interest due to its low abundance in the mammalian cell supernatants which also contain serum proteins. This was taken care by and affinity purification step from cell supernatants (Ordono et al., 2006) which gave a fairly pure protein in high yields to further downstream purification and use in crystallization screening trials.

5.2 Structural features of Aminopeptidase N (APN)

The presented crystal structures of APN depicted a domain architecture resembling that of related aminopeptidases (Kochan et al., 2011, Nguyen et al., 2011, Addlagatta et al., 2006), although the APN ectodomain is heavily glycosylated and forms dimers. DI, DII and DIII are more similar to homologous domains in M1 aminopeptidases than DIV. Human and pig APN domain structures are very similar, except for the DIV ARM repeat, which contains a highly flexible region whose conformation varies among APN structures. The APN ectodomain can adopt different conformations, related to interdomain movements. This inherent conformational dynamics determine APN multiple functions related to peptide hydrolysis and CoV cell entry (Fig. 5.1).

Mammalian APN are known to exist as stable dimers on the cell surface (Hussain et al., 1981). The dimeric APN state may enhance protein stability in extracellular environments. Dimerization was also found in soluble ectodomain prepared for crystallization. Different crystal structures exhibited this unique homodimerization in APN ectodomains mediated by a large DIV surface in contacting monomers. Introduction of a glycosylation motif at the dimerization interface disrupted dimerization both in soluble and cell surface forms of the pAPN. These results supported the involvement of domain IV in dimerization proving the relevance of the crystal structure dimers. This APN dimer is unique among M1 aminopeptidases (Kochan et al., 2011, Nguyen et al., 2011, Addlagatta et al., 2006), which is likely related to the distinct DIV conformation and sequence in APN. The DIV-'bowl' exhibits a pore in its bottom and it has been proposed as an entry point to small peptide substrates to access the catalytic center in other aminopeptidases (Addlagatta et al., 2006).This pore is plugged by APN dimerization, preventing its use in substrate entry to the active site.

In the structures, the APN monomers appear tilted in the dimer, but their N-termini point to the center of the dimer, likely locating nearby the transmembrane domains of the two monomers. Cell signaling feature of APN (Mina-Osorio, 2008) can be attributed to the organization of its membrane proximal regions and transmembrane segments and interaction of these domains with other cell surface proteins. These features can be influenced by APN ectodomain movements.

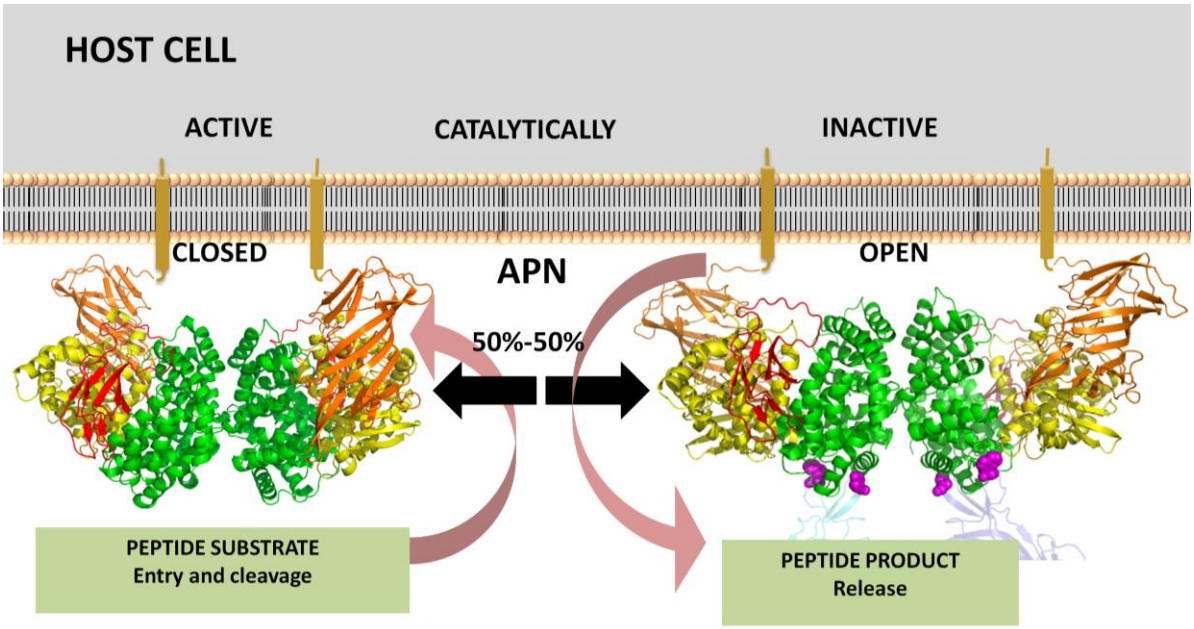


Figure 5.1. APN conformations and functions on the cell surface. The diagram discusses the prevalence of two conformational states of APN on the host cell surface. Substrate entry, locking and cleavage are favored by the catalytically active (closed) conformation while the inactive (open) conformation facilitates product release. Both the conformations are shown to exist in dynamic equilibrium.

5.2.1 The dynamic structure of mammalian APN

We have determined several APN dimeric crystal structures that show different conformational states of the ectodomains. This conformation dynamics is an intrinsic property of aminopeptidases, and it has been also described in ectoenzymes used as CoV receptors. The APN ectodomain movements are less pronounced and differ from that reported for other M1 aminopeptidases. These differences could be due to the APN dimeric conformation and its linkage to the cell surface. Dimerization only engages the DIV region, and we found that the dimer is preserved in closed, intermediate and open APN structures. APN DIV is locked by dimerization and thus cannot move as described for ERAP-1 or F3 (Kochan et al., 2011, Kyrieleis et al., 2005), proteins that do not exhibit dimerization. The fixed conformation of the APN dimer determines that the DI-DII-DIII module swings over DIV, with the hinge at the DIII C terminus at the beginning of DIV. Moreover, the swing angle is less prominent in APN (15°) than in its closely related ERAP-1 (22°), despite very similar closed state conformations. Displacement of the DI-DII-DIII APN module must be limited by the length of the flexible polypeptide that links DI to the transmembrane region, whose movement is restricted by membrane fluidity. Dimerization and linkage to the cell surface could be two important determinants of the APN dynamics described here. The extent of APN movement nonetheless appears to be sufficient for the release of the hydrolyzed N-

terminal residue, which is distant from DIV in the open or in the intermediate APN conformations. It is not clear the way each monomer in the dimer moves, whether their movement is random or synchronized, in the same or inverse directions. Experiments with monoclonal antibodies (Xu et al., 1997) and those shown here with the TGEV RBD (Fig. 4.21) suggest that ~50% of the molecules adopt different conformations; these data imply that APN monomer adopts a distinct conformation and its movement is probably independent.

The crystal structures reported here not only provide snapshots of APN dynamic conformation, but also guide experiments to confirm its importance in the functions of this aminopeptidase. APN exhibited open, intermediate, and closed conformations in the various structures presented here. The transition between a proteolytic active (closed) and an inactive (open) forms has been proposed important for the function of several M1 aminopeptidases (Kyrieleis et al., 2005, Kochan et al., 2011, Nguyen et al., 2011, Addlagatta et al., 2006). This dynamics is thought to be required for peptide hydrolysis and release from the aminopeptidase active site. The different APN structures do not show any change on the conformation of active site residues in DII (Fig. 4.12A), but they have significant differences in contacts between DII and DIV residues. The region that joins α26 and α27 in the DIV ARM repeat penetrates the active site groove in closed APN structures (Fig 4.12B).

After APN closure, a conserved phenylalanine in this region locks the substrate coordinated to the zinc ion, permitting hydrolysis. Further processing of the peptide requires removal of the phenylalanine "lock" by opening the APN ectodomain, facilitating N-terminal residue release and peptide translocation, both sterically hindered in the closed conformation. The inherent flexibility in the DIV ARM repeat might also enable substrate processing during catalysis.

5.2.2 APN dynamics and allostery

Studies towards realizing the effect of APN catalytic inhibitors on TGEV RBD-pAPN interaction and subsequent analysis of CoV S proteins on APN catalysis identified allosteric effects, likely related to restrictions in ectodomain movements (Fig. 5.2). Catalytic sites and CoV binding regions are quite distant in ectoenzymes serving as entry receptors (Li et al., 2005a, Wu et al., 2009, Reguera et al., 2012, Chen et al., 2013) and the respective functions are believed to be independent of each other (Delmas et al., 1994, Moore et al., 2004). Contrary to this notion, studies presented in this Thesis show that in APN agents that bind to

one region do prevent activities linked to the other. As mentioned above, the switch between an active (closed) and an inactive (open) form has been proposed essential for peptide hydrolysis in M1 aminopeptidases (Kyrieleis et al., 2005, Kochan et al., 2011, Nguyen et al., 2011, Addlagatta et al., 2006), but never demonstrated. Using different APN ligands that bind to distinct sites, we carried out several experiments to determine the functional relevance of the different APN conformations and its conformational switch. We show that binding of TGEV S proteins to open APN blocks catalysis, thus confirming the crucial role of APN dynamics in catalysis and demonstrating that the open APN structure is inactive in peptide hydrolysis.

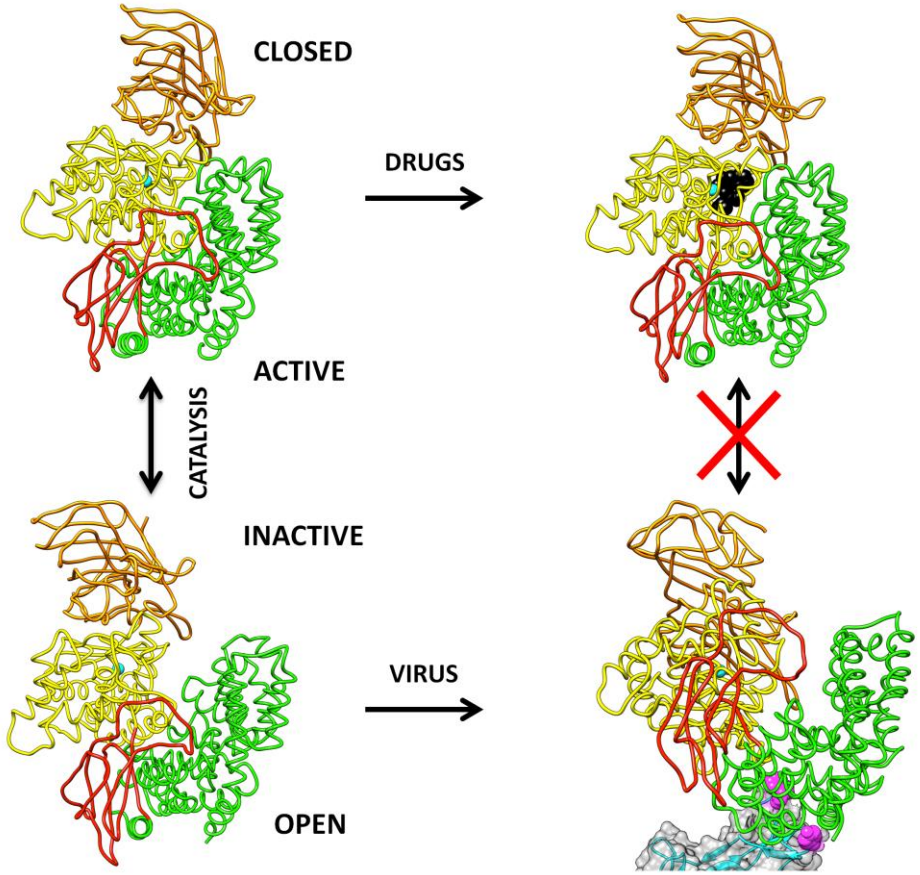


Figure 5.2. APN dynamics and allostery. Cartoons show APN with two functionally distinct conformational states. The open conformation is specifically recognized by CoV, whereas substrate hydrolysis must occur in the closed APN. The allosteric inhibition of APN functions shown here using viral proteins and drugs must be due to suppression of APN transient conformational states.

Drugs that bind to the active site promote closing of aminopeptidases (Addlagatta et al., 2006, Kochan et al., 2011, Nguyen et al., 2011), and they can thus be used to test its influence on CoV binding, which needs APN to open. These drugs inhibited binding of porcine CoV proteins to pAPN, probably by preventing opening of the ectodomain. The allosteric effects shown here with different ligands must be due to suppression of APN transient

conformational states (Fig. 5.2), such as that recently suggested by NMR studies of enzymes (Tzeng and Kalodimos, 2013). Blocking APN motions can prevent APN functions, and suggests a new approach for the development of drugs that target this protein. Inhibitors of ectodomain movements can bind to the active site, or interact with distant sites, such as shown here with the TGEV RBD fragment.

We have related two APN functions with different protein structures. APN peptide hydrolysis requires ectodomain closing as well as some opening for peptide release, whereas CoV specifically recognize the open APN conformation. These results show that the dynamic conformation of the APN ectodomain determines its multiple functions.

5.3 The CoV-APN interaction

The CoV S is a multifunctional glycoproteins that mediates both CoV attachment to cell surface molecules and membrane penetration by fusion. The N-terminal S1 region contains several domains that recognize cell entry receptors. In most CoV, the major determinants of cell tropism are found in the C-terminal portion of the S1 region (Masters, 2006, Gallagher and Buchmeier, 2001). These RBD have structures that are unrelated to host proteins, and they can thus be considered genuine CoV RBD. They recognize the APN, ACE2 or DDP4 ectoenzymes. APN is a cell entry receptor for most alpha-CoV. Our research showed how a group of CoV attaches to the cell surface APN metalloprotease for entry into host cells, and the way some CoV-neutralizing antibodies prevent infection. These studies are relevant for understanding receptor recognition in CoV, its evolution and the adaptation of this virus family to different hosts.

The RBD of porcine and related alpha-CoV adopt a similar β-barrel fold, with some structural divergences in its receptor-binding tip (Fig. 4.14). The RBD of CoV that bind APN have a tip with protruding receptor-binding motifs, which engage recessed surfaces on a membrane distal region of the receptor. PRCV and TGEV contain two main RBM, each composed by a loop that bears an exposed aromatic residue, tyrosine or tryptophan. RBM1 (β1-β2 region) has a tyrosine and it contacts with an α-helix and an N-linked glycan in the APN ectodomain. The inherent flexibility of this glycan might facilitate the initial contact of the CoV tyrosine with APN amino acids, and subsequent virus-receptor interactions could lock the bound Tyr between the glycan and an α-helix. Besides the tyrosine, other RBD residues contact the APN and they establish a network of interactions with the APN, including the N-linked NAG. The β5-β6 region at the RBD tip contact also with the glycan,

and it appears to stabilize the protruding conformation of the APN-binding β1-β2 loop. RBM2 at the beginning of the RBD β3-β4 loop contains a tryptophan residue that penetrates in a narrow cavity formed by the APN DII and DIV. The β3-β4 loop preserves the tryptophan in all the APN-binding CoV, in the alpha1-group and even in HCoV-229E, which has an RBD with a shorter loop (Fig. 4.14). This contact is responsible of the specific CoV binding to the open APN, which can accommodate the large tryptophan side chain. Other residues in RBM2 contact the APN in the crystal structure of the porcine RBD-pAPN complex, however they do not appear essential for either CoV S protein binding to APN or cell infection. In contrast, residues surrounding the tyrosine in RBM1 are also important for S protein binding to the APN, indicating that RBM1 could be a higher contribution than RBM2 to the stability of the CoV-APN interaction.

Receptor glycosylations are important determinants of CoV-receptor recognitions. They can favor or hinder CoV binding to cell entry receptors in certain species (Li et al., 2005a, Tusell et al., 2007), delimiting the CoV host range. A glycan linked to rat ACE2 Asn82 prevents its use as an efficient SARS-CoV receptor (Li et al., 2005a). Addition of an N-linked glycan to the human APN converts it into a TGEV entry receptor (Tusell et al., 2007). We show here that TGEV and PRCV specifically recognize an N-linked APN glycan. This glycan N-linked to amino acid Asn736 in pAPN is also conserved in canine and feline APN, as are the viral S protein residues that interact with this glycan in RBM1 (β1-β2 loop). This unique glycan-virus interaction must thus be conserved among the members of the species *alpha1*-CoV, in accordance with the glycan requirement reported for cell infection by CCoV, FCoV, and TGEV (Tusell et al., 2007). The lack of this glycan in human APN and the absence of the interacting Tyr residue in the β1–β2 region of HCoV-229E RBD (Fig. 4.8 and Fig. 4.14) imply distinct virus-APN local contacts in humans. Recent crystal structures show that the MERS-CoV RBD also bind to a glycan N-linked to DPPIV (Wang et al., 2013). Even though structures of MERS- and *alpha*-CoV RBD are very different, they both recognize N-linked glycosylations, showing this is a common feature and a unique characteristic of CoV-receptor interactions.

5.4 Conformation of the receptor-binding regions determine receptor-specificity in CoV

CoV RBD folding, conformation of the receptor binding motifs and subtle changes in these motifs determine receptor binding specificity and CoV host range. *Alpha*- and *beta*-CoV

RBD adopt two distinct folds, and they bind to different ectoenzymes. The *alpha*-CoV RBD have a β-barrel fold, whereas the RBD of the SARS- and MERS-CoV (beta-CoV) assumes a very different conformation (Li et al., 2005a, Chen et al., 2013). They are composed of two subdomains, the core and the receptor-binding subdomain, and both contain a single β-sheet. Even though alpha- and beta-CoV RBD are very different structurally, the HCoV-NL63 (*alpha*-CoV) and the SARS-CoV bind to the same protein, ACE2. The mode these CoV bind ACE2 shares similarities; they recognize overlapping ACE2 regions, including two helices and a unique □-turn in the virus-binding lobe of the receptor (Wu et al., 2009) (Li et al., 2005a). The ACE2-binding surfaces in both CoV are concave and exposed distant from the terminal end of the RBD. Virus use of recessed surfaces to bind to exposed receptor motifs is a strategy to hide conserved receptor-binding residues from antibodies (Casasnovas, 2013, Rossmann, 1989). These two CoV have highly distinct RBD, but their receptor-binding region evolved to recognize the same receptor molecule. HCoV-NL63 and PRCV/TGEV RBD have a similar β-barrel fold, but they use different entry receptors, ACE2 and APN, respectively.

The structures of PRCV RBD-APN (Fig. 4.15) and HCoV-NL63 RBD-ACE2 (Wu et al., 2009) complexes show reversed modes of CoV-receptor recognition, and identified the basis for the different receptor-binding specificity among alpha-CoV (Fig. 5.3). PRCV, TGEV and alpha1-CoV S bear RBD with bidentate and protruding tips that can penetrate in small cavities of the APN ectodomain; on the contrary, the HCoV-NL63 RBD tip has a “bowl”-like tip that cannot bind to APN, but recognizes exposed ACE2 motifs, following a similar receptor-binding strategy to *beta*-CoV. The conformation of the RBD tip thus dictates the receptor molecule alpha-CoV recognize (Fig. 5.3).

A B

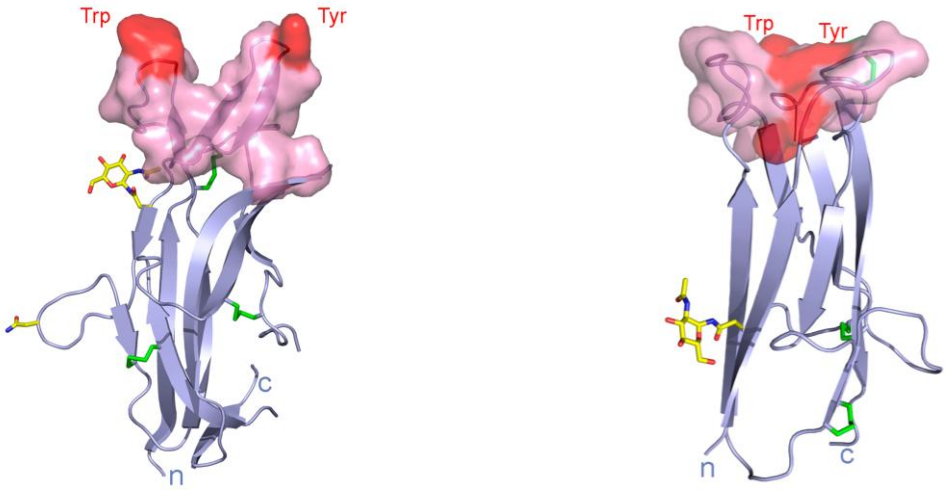


Figure 5.3. Conformation of receptor-binding regions in the RBD of alpha-CoV, a determinant of its receptor-binding specificity. Ribbon diagram of the PRCV (left) and hCoV-NL63 RBD (right) structures determined in complex with the APN (Fig. 4.15) and ACE2 (NL63, PDB ID 3KBH) (Wu et al., 2009), respectively. Receptor-binding surfaces in the RBD are shown in pink or red (tyrosine or tryptophan residues) and were generated by the RBD residues that contact the respective receptor molecules in the structures.

5.5 Immune neutralization of CoV TGEV and related alpha-CoV

Site A in the TGEV S is recognized by potent TGEV neutralizing mAb. In this Thesis we show that epitopes of several neutralizing mAb overlap with the receptor-binding region in the RBD, showing that the RBD tip is the major antigenic determinant in the envelope S protein of CoV that bind APN. All TGEV neutralizing mAb tested recognize TGEV residues that bind to APN; mAb bind either to the tyrosine or the tryptophan exposed in the RBD tip. The most potent mAb, 6AC3, targets the tyrosine, which we showed here as a key residue for TGEV binding to APN and infection.

The identification of the RBD tip as a major antigenic determinant in the TGEV S, demonstrate that the receptor-binding region in Alphacoronavirus is under selective pressure from the immune system. It is tempting to speculate that immune pressure on exposed receptor-binding residues in the CoV S could lead to conformational changes in receptor-binding edges of CoV RBD. This would result either in changes in the APN-recognition mode reported for HCoV-229E (Tusell et al., 2007), or important conformational changes in the RBD tip that lead to a receptor specificity switch for cell entry, as in HCoV-NL63 (Fig. 5.3). These data on CoV indicates that the humoral immune response can drive evolution of receptor recognition in viruses. Moreover, the identification of antigenic site determinants in

the *alpha*-CoV S protein, could guide the design of immunogens that boost immunity towards critical motifs for virus cell entry.

5.6 CoV selection of ectoenzymes as cell entry receptors

Alpha- and *beta*-CoV S glycoproteins contain RBD with distinct and unique folds, which are thought to have evolved from a common CoV RBD ancestor. They share an important common feature, the recognition of ectoenzymes. Some of these enzyme features must be essential for CoV entry into host cells. CoV preferential selection of ectoenzymes as cell entry receptors is unclear. It might be linked to their abundance on epithelial cells rather than on their peptidase function, which does not appear to be essential for CoV cell entry (Li et al., 2005b). Perhaps they cluster with other proteases that facilitate fusion (Glowacka et al., 2011). Here we analyzed porcine CoV infectivity in CHO cell clones expressing similar cell surface amounts of wild type and a catalytic mutant pAPN. Preliminary results included in this Thesis indicate decreased TGEV infection of cells expressing the catalytically inactive pAPN. This effect was strain specific and no effect on cell infectivity was observed on PRCV infection. PRCV is also known to form syncytia on cell monolayers; virus spreading by cell-cell fusion may not require APN enzyme activity. The differences may be also responsible of the distinct PRCV and TGEV tropism. Nonetheless, further studies are required to evaluate APN activity on CoV cell infection. The three cell surface CoV receptors described to date, ACE2, APN and DPP4, have distinct structures and functions, but their ectodomains share an inherent conformational flexibility (Xu et al., 1997, Boonacker and Van Noorden, 2003, Towler et al., 2004) that could assist in dissociation of the S1-S2 heterotrimer during cell entry. Trimeric spikes that bind simultaneously to several receptor molecules could disassemble by pulling forces generated during ectodomain movement. The conformation and dynamics of the APN ectodomain vary with the pH (Table 4.2), so that endosomal acidification can alter APN conformation during receptor-mediated endocytosis.

Movements in the APN ectodomain are essential for catalysis, CoV binding and cell infection. Blocking APN dynamics can thus prevent multiple functions associated with cancer and virus infections. Molecules directed to the active site that bridge APN DII and DIV could inhibit more efficiently APN functions; this approach open new avenues for drug design in APN and related aminopeptidases. Blocking ectodomain movements in the other CoV entry receptors could be an interesting strategy to prevent CoV infections.

CHAPTER 6

6. CONCLUSIONS

- Mammalian system such as HEK-293T and CHO cells provide a suitable platform for expression of membrane glycoproteins for use in protein crystallography.
- **Crystal structures of aminopeptidase N reveal ectodomain architecture, dimerization and conformation dynamics.**
- Domain IV mediates APN dimerization. Its divergence among M1-aminopeptidases explains the different dimerization properties among members of this family.
- **The APN dimer observed in several crystal structures represents a functional cell membrane dimer.**
- APN crystal structures revealed the dynamic conformation of its ectodomain and they identified three distinct APN conformations: Open (active site at DII is accessible), Intermediate and Closed (active site closed by DIV).
- Conformations of active site residues are preserved in the different APN structures.
- The dynamic motion of the APN ectodomain likely causes peptide hydrolysis.
- A conserved phenylalanine residue in a flexible DIV ARM repeat assists in locking and release of substrates during catalysis.
- The APN open conformation is catalytically inactive.
- **Allosteric inhibition of APN functions can be accomplished by restricting ectodomain dynamics.**
- **The crystal structure of the pig APN ectodomain has been determined in complex with the RBD of a porcine CoV of the genus alpha.**
- The RBD-APN complex structure defined a unique mode of receptor recognition among CoV; a protruding RBD tip contacts small cavities in a membrane-distal region of the APN ectodomain.
- One receptor-binding motif (RBM1) in porcine and related alpha-CoV is formed by a β-turn that exposes a tyrosine. This residue is fully buried at the virus-receptor interface, and it is essential for CoV binding to APN and infection.
- **RBM1 contacts a carbohydrate N-linked to the APN, which is an important determinant of the CoV-host specificity.**
- RBM2 is also formed by a β-turn that exposes a tryptophan. This residue is also critical for CoV binding to APN and infection.

- **RBM2 penetrates in a narrow cavity formed by APN DII and DIV. The size of this cavity limits CoV binding to APN in the open conformation.**
- **Conformation of the receptor-binding edge in the alpha-CoV RBD determines receptor-binding specificity.**
- TGEV-neutralizing antibodies target receptor-binding motifs in the S, and they preferentially bind either to the tyrosine in RBM1 or the tryptophan in RBM2.
- **The APN-binding region is the main antigenic determinant in the TGEV S.**
- **Immune pressure likely drives evolution of receptor-specificity in CoV.**
- APN enzymatic function can be important for cell entry and infection of some CoV strains.
- Targeting APN dynamics can be a new strategy towards more effective anticancer and anti-inflammatory agents, which should also prevent CoV infections.

BIBLIOGRAPHY

1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr,* 50**,** 760-3.

2004. *Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003* [Online]. WHO. http://www.who.int/csr/sars/country/table2004_04_21/en/ [Accessed 28/07/2014.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

ADDLAGATTA, A., GAY, L. & MATTHEWS, B. W. 2006. Structure of aminopeptidase N from Escherichia coli suggests a compartmentalized, gated active site. *Proc Natl Acad Sci U S A,* 103**,** 13339-44.

ADDLAGATTA, A., GAY, L. & MATTHEWS, B. W. 2008. Structural basis for the unusual specificity of Escherichia coli aminopeptidase N. *Biochemistry,* 47**,** 5303-11.

ALMAZAN, F., GALAN, C. & ENJUANES, L. 2008. Engineering infectious cDNAs of coronavirus as bacterial artificial chromosomes. *Methods Mol Biol,* 454**,** 275-91.

ALMAZAN, F., GONZALEZ, J. M., PENZES, Z., IZETA, A., CALVO, E., PLANA-DURAN, J. & ENJUANES, L. 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc Natl Acad Sci U S A,* 97**,** 5516-21.

ARICESCU, A. R., ASSENBERG, R., BILL, R. M., BUSSO, D., CHANG, V. T., DAVIS, S. J., DUBROVSKY, A., GUSTAFSSON, L., HEDFALK, K., HEINEMANN, U., JONES, I. M., KSIAZEK, D., LANG, C., MASKOS, K., MESSERSCHMIDT, A., MACIEIRA, S., PELEG, Y., PERRAKIS, A., POTERSZMAN, A., SCHNEIDER, G., SIXMA, T. K., SUSSMAN, J. L., SUTTON, G., TARBOUREICH, N., ZEEV-BEN-MORDEHAI, T. & JONES, E. Y. 2006. Eukaryotic expression: developments for structural proteomics. *Acta Crystallogr D Biol Crystallogr,* 62**,** 1114-1124.

AUGUSTINA ANNAN, H. J. B., VICTOR MAX CORMAN, STEFAN M. KLOSE, MICHAEL OWUSU, EVANS EWALD NKRUMAH, EBENEZER KOFI BADU, PRISCILLA ANTI, OLIVIA AGBENYEGA, BENJAMIN MEYER, SAMUEL OPPONG, YAW ADU SARKODIE, ELISABETH K.V. KALKO, PETER H.C. LINA, ELENA V. GODLEVSKA, CHANTAL REUSKEN, ANTJE SEEBENS, FLORIAN GLOZA-RAUSCH, PETER VALLO, MARCO TSCHAPKA, CHRISTIAN DROSTEN, AND JAN FELIX DREXLER 2013. Human Betacoronavirus 2c EMC/2012–related Viruses in Bats, Ghana and Europe. *Emerg Infect Dis,* 19**,** 456-459.

BELOUZARD, S., MILLET, J. K., LICITRA, B. N. & WHITTAKER, G. R. 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses,* 4**,** 1011-33.

BENDER, S. J. & WEISS, S. R. 2010. Pathogenesis of murine coronavirus in the central nervous system. *J Neuroimmune Pharmacol,* 5**,** 336-54.

BENIAC, D. R., ANDONOV, A., GRUDESKI, E. & BOOTH, T. F. 2006. Architecture of the SARS coronavirus prefusion spike. *Nat. Struct. Mol. Biol.,* 13**,** 751-2.

BERGFORS , T. 1999. *Protein Crystallization: Techniques, Strategies, and Tips,* La Jolla, California, International University Line.

BOONACKER, E. & VAN NOORDEN, C. J. F. 2003. The multifunctional or moonlighting protein CD26/DPPIV. *European Journal of Cell Biology,* 82**,** 53-73.

BOOTS, A. M., BENAISSA-TROUW, B. J., HESSELINK, W., RIJKE, E., SCHRIER, C. & HENSEN, E. J. 1992. Induction of anti-viral immune responses by immunization with

recombinant-DNA encoded avian coronavirus nucleocapsid protein. *Vaccine,* 10**,** 119-24.

BOS, E. C. W., HEIJNEN, L. E. O., LUYTJES, W. & SPAAN, W. J. M. 1995. Mutational Analysis of the Murine Coronavirus Spike Protein: Effect on Cell-to-Cell Fusion. *Virology,* 214**,** 453-463.

BOSCH, B. J., VAN DER ZEE, R., HAAN, C. A. M. & ROTTIER, P. J. M. 2003. The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. *J. Virol.,* 77**,** 8801-8811.

BURKARD, C., BLOYET, L. M., WICHT, O., VAN KUPPEVELD, F. J., ROTTIER, P. J., DE HAAN, C. A. & BOSCH, B. J. 2014. Dissecting Virus Entry: Replication-Independent Analysis of Virus Binding, Internalization, and Penetration Using Minimal Complementation of beta-Galactosidase. *PLoS One,* 9**,** e101762.

CASAIS, R., THIEL, V., SIDDELL, S. G., CAVANAGH, D. & BRITTON, P. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J Virol,* 75**,** 12359-69.

CASASNOVAS, J. M. 2013. Virus-receptor interactions and receptor-mediated virus entry into host cells. *Subcell Biochem,* 68**,** 441-66.

CASASNOVAS, J. M. & SPRINGER, T. A. 1995. Kinetics and thermodynamics of virus binding to receptor: Studies with rhinovirus, intercellular adhesion molecule-1 (ICAM-1), and surface plasmon resonance. *J Biol Chem.,* 270**,** 13216-13224.

CASASNOVAS, J. M., SPRINGER, T. A., LIU, J. H., HARRISON, S. C. & WANG, J. H. 1997. Crystal structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature,* 387**,** 312-5.

CAVANAGH, D. 2005. Coronaviridae: a review of coronaviruses and toroviruses. *In:* A. SCHMIDT, M. H. W. A. O. W. (ed.) *Coronaviruses with Special Emphasis on First Insights Concerning SARS.* Basel/Switzerland: Birkhäuser Verlag

CAVANAGH, D. 2007. Coronavirus avian infectious bronchitis virus. *Vet Res,* 38**,** 281-97.

CHANG, V. T., CRISPIN, M., ARICESCU, A. R., HARVEY, D. J., NETTLESHIP, J. E., FENNELLY, J. A., YU, C., BOLES, K. S., EVANS, E. J., STUART, D. I., DWEK, R. A., JONES, E. Y., OWENS, R. J. & DAVIS, S. J. 2007. Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure,* 15**,** 267-273.

CHEN, L., GAO, Z., ZHU, J. & RODGERS, G. P. 2007. Identification of CD13+CD36+ cells as a common progenitor for erythroid and myeloid lineages in human bone marrow. *Exp Hematol,* 35**,** 1047-55.

CHEN, L., LIN, Y. L., PENG, G. & LI, F. 2012. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proc Natl Acad Sci U S A,* 109**,** 17966-71.

CHEN, X. 2006. Biochemical properties of recombinant prolyl dipeptidases DPP-IV and DPP8. *Adv Exp Med Biol,* 575**,** 27-32.

CHEN, Y., RAJASHANKAR, K. R., YANG, Y., AGNIHOTHRAM, S. S., LIU, C., LIN, Y. L., BARIC, R. S. & LI, F. 2013. Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *J Virol,* 87**,** 10777-83.

CHOI, K. S., AIZAKI, H. & LAI, M. M. 2005. Murine coronavirus requires lipid rafts for virus entry and cell-cell fusion but not for virus release. *J Virol,* 79**,** 9862-71.

CHU, D. K., POON, L. L., GOMAA, M. M., SHEHATA, M. M., PERERA, R. A., ABU ZEID, D., EL RIFAY, A. S., SIU, L. Y., GUAN, Y., WEBBY, R. J., ALI, M. A., PEIRIS, M. & KAYALI, G. 2014. MERS coronaviruses in dromedary camels, Egypt. *Emerg Infect Dis,* 20**,** 1049-53.

DANIEL, C. & TALBOT, P. J. 1990. Protection from lethal coronavirus infection by affinity-purified spike glycoprotein of murine hepatitis virus, strain A59. *Virology,* 174**,** 87-94.

DANIELSEN, E. M., NOREN, O. & SJOSTROM, H. 1982. Co-and post-translational events in the biogenesis of pig small intestinal aminopeptidase N. *Tokai J Exp Clin Med,* 7 Suppl**,** 135-40.

DE GROOT, R. J., BAKER, S. C., BARIC, R. S., BROWN, C. S., DROSTEN, C., ENJUANES, L., FOUCHIER, R. A., GALIANO, M., GORBALENYA, A. E., MEMISH, Z. A., PERLMAN, S., POON, L. L., SNIJDER, E. J., STEPHENS, G. M., WOO, P. C., ZAKI, A. M., ZAMBON, M. & ZIEBUHR, J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol,* 87**,** 7790-2.

DE HAAN, C. A., STADLER, K., GODEKE, G. J., BOSCH, B. J. & ROTTIER, P. J. 2004. Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion. *J Virol,* 78**,** 6048-54.

DELMAS, B., GELFI, J., KUT, E., SJOSTROM, H., NOREN, O. & LAUDE, H. 1994. Determinants essential for the transmissible gastroenteritis virus-receptor interaction reside within a domain of aminopeptidase-N that is distinct from the enzymatic site. *J Virol,* 68**,** 5216-24.

DELMAS, B., GELFI, J., L'HARIDON, R., VOGEL, L. K., SJOSTROM, H., NOREN, O. & LAUDE, H. 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature,* 357**,** 417-420.

DELMAS, B., RASSCHAERT, D., GODET, M., GELFI, J. & LAUDE, H. 1990. Four major antigenic sites of the coronavirus transmissible gastroenteritis virus are located on the amino-terminal half of spike glycoprotein S. *J Gen Virol,* 71**,** 1313-23.

EIFART, P., LUDWIG, K., BOTTCHER, C., DE HAAN, C. A., ROTTIER, P. J., KORTE, T. & HERRMANN, A. 2007. Role of endocytosis and low pH in murine hepatitis virus strain A59 cell entry. *J Virol,* 81**,** 10758-68.

ENJUANES, L., DEDIEGO, M. L., ALVAREZ, E., DEMING, D., SHEAHAN, T. & BARIC, R. 2008. Vaccines to prevent severe acute respiratory syndrome coronavirus-induced disease. *Virus Research* 133**,** 45–62.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr,* 62**,** 72-82.

FRANA, M. F., BEHNKE, J. N., STURMAN, L. S. & HOLMES, K. V. 1985. Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: host-dependent differences in proteolytic cleavage and cell fusion. *J Virol,* 56**,** 912-20.

FUJII, H., NAKAJIMA, M., SAIKI, I., YONEDA, J., AZUMA, I. & TSURUO, T. 1995. Human melanoma invasion and metastasis enhancement by high expression of aminopeptidase N/CD13. *Clin Exp Metastasis,* 13**,** 337-44.

GALLAGHER, T. M. & BUCHMEIER, M. J. 2001. Coronavirus Spike Proteins in Viral Entry and Pathogenesis. *Virology,* 279**,** 371-374.

GEBAUER, F., POSTHUMUS, W. P. A., CORREA, I., SUÑÉ, C., SMERDOU, C., SÁNCHEZ, C. M., LENSTRA, J. A., MELOEN, R. H. & ENJUANES, L. 1991. Residues involved in the antigenic sites of transmissible gastroenteritis coronavirus S glycoprotein. *Virology,* 183**,** 225-238.

GLOWACKA, I., BERTRAM, S., MULLER, M. A., ALLEN, P., SOILLEUX, E., PFEFFERLE, S., STEFFEN, I., TSEGAYE, T. S., HE, Y., GNIRSS, K., NIEMEYER, D., SCHNEIDER, H., DROSTEN, C. & POHLMANN, S. 2011. Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome

Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol.*, 85**,** 4122-4134.

GOMBOLD, J. L., HINGLEY, S. T. & WEISS, S. R. 1993. Fusion-defective mutants of mouse hepatitis virus A59 contain a mutation in the spike protein cleavage signal. *J Virol*, 67**,** 4504-12.

GROOT, R. J. D., BAKER, S.C., BARIC, R., ENJUANES, L., GORBALENYA, A.E., HOLMES, K.V., PERLMAN, S., POON, L., ROTTIER, P.J.M., TALBOT, P.J., WOO, P.C.Y. & ZIEBUHR, J 2011. Family Coronaviridae *In:* A.M.Q. KING, M. J. A., E.B. CARSTENS & E.J. LEFKOWITZ (ed.) *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses.* . Oxford.: Elsevier.

HAIJEMA, B. J., VOLDERS, H. & ROTTIER, P. J. 2003. Switching species tropism: an effective way to manipulate the feline coronavirus genome. *J Virol*, 77**,** 4528-38.

HANSEN, G. H., DELMAS, B., BESNARDEAU, L., VOGEL, L. K., LAUDE, H., SJOSTROM, H. & NOREN, O. 1998. The coronavirus transmissible gastroenteritis virus causes infection after receptor-mediated endocytosis and acid-dependent fusion with an intracellular compartment. *J Virol*, 72**,** 527-34.

HAVRE, P. A., ABE, M., URASAKI, Y., OHNUMA, K., MORIMOTO, C. & DANG, N. H. 2008. The role of CD26/dipeptidyl peptidase IV in cancer. *Front Biosci*, 13**,** 1634-45.

HEALD-SARGENT, T. & GALLAGHER, T. 2012. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses*, 4**,** 557-80.

HEGYI, A. & KOLB, A. F. 1998. Characterization of determinants involved in the feline infectious peritonitis virus receptor function of feline aminopeptidase N. *J Gen Virol*, 79 (Pt 6)**,** 1387-91.

HENDRICKSON, W. A. & OGATA, C. M. 1997. [28] Phase determination from multiwavelength anomalous diffraction measurements. *In:* CHARLES W. CARTER, JR. (ed.) *Methods in Enzymology*. Academic Press.

HOFMANN, H., HATTERMANN, K., MARZI, A., GRAMBERG, T., GEIER, M., KRUMBIEGEL, M., KUATE, S., UBERLA, K., NIEDRIG, M. & POHLMANN, S. 2004. S protein of severe acute respiratory syndrome-associated coronavirus mediates entry into hepatoma cell lines and is targeted by neutralizing antibodies in infected patients. *J Virol*, 78**,** 6134-42.

HOLMES, K. V. & LAI, M. M. C. 1996. Coronaviridae: The viruses and their replication. *In:* FIELDS, B. N., KNIPE, D. M., HOWLEY, P. M., MELNICK, J. L., CHANOCK, R. M., ROIZMAN, B. & MONATH, T. P. (eds.) *Virology*. Third ed. New York: Raven Press.

HOOPER, N. M. 1994. Families of zinc metalloproteases. *FEBS Lett*, 354**,** 1-6.

HUSSAIN, M. M., TRANUM-JENSEN, J., NOREN, O., SJOSTROM, H. & CHRISTIANSEN, K. 1981. Reconstitution of purified amphiphilic pig intestinal microvillus aminopeptidase. Mode of membrane insertion and morphology. *Biochem J*, 199**,** 179-86.

JEFFERS, S. A., TUSELL, S. M., GILLIM-ROSS, L., HEMMILA, E. M., ACHENBACH, J. E., BABCOCK, G. J., THOMAS, W. D., JR., THACKRAY, L. B., YOUNG, M. D., MASON, R. J., AMBROSINO, D. M., WENTWORTH, D. E., DEMARTINI, J. C. & HOLMES, K. V. 2004. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci. U. S. A.*, 101**,** 15748-53.

JOHNSON, M. A., POOLEY, C., IGNJATOVIC, J. & TYACK, S. G. 2003. A recombinant fowl adenovirus expressing the S1 gene of infectious bronchitis virus protects against challenge with infectious bronchitis virus. *Vaccine*, 21**,** 2730-6.

JONES, E. Y., DAVIS, S. J., WILLIAMS, A. F., HARLOS, K. & STUART, D. I. 1992. Crystal structure at 2.8 [angst] resolution of a soluble form of the cell adhesion molecule CD2. *Nature,* 360**,** 232-239.

JORDAN, M., SCHALLHORN, A. & WURM, F. M. 1996. Transfecting Mammalian Cells: Optimization of Critical Parameters Affecting Calcium-Phosphate Precipitate Formation. *Nucleic Acids Research,* 24**,** 596-601.

KABSCH, W. 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography,* 26**,** 795-800.

KECK, J. G., SOE, L. H., MAKINO, S., STOHLMAN, S. A. & LAI, M. M. 1988. RNA recombination of murine coronaviruses: recombination between fusion-positive mouse hepatitis virus A59 and fusion-negative mouse hepatitis virus 2. *J Virol,* 62**,** 1989-98.

KEHLEN, A., LENDECKEL, U., DRALLE, H., LANGNER, J. & HOANG-VU, C. 2003. Biological significance of aminopeptidase N/CD13 in thyroid carcinomas. *Cancer Res,* 63**,** 8500-6.

KOCHAN, G., KROJER, T., HARVEY, D., FISCHER, R., CHEN, L., VOLLMAR, M., VON DELFT, F., KAVANAGH, K. L., BROWN, M. A., BOWNESS, P., WORDSWORTH, P., KESSLER, B. M. & OPPERMANN, U. 2011. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc Natl Acad Sci U S A,* 108**,** 7745-50.

KOOI, C., CERVIN, M. & ANDERSON, R. 1991. Differentiation of acid-pH-dependent and -nondependent entry pathways for mouse hepatitis virus. *Virology,* 180**,** 108-19.

KREMPL, C. & HERRLER, G. 2001. Sialic Acid Binding Activity of Transmissible Gastroenteritis Coronavirus Affects Sedimentation Behavior of Virions and Solubilized Glycoproteins. *J Virol,* 75**,** 844-849.

KRUEGER, D. K., KELLY, S. M., LEWICKI, D. N., RUFFOLO, R. & GALLAGHER, T. M. 2001. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J Virol,* 75**,** 2792-802.

KUO, L., GODEKE, G. J., RAAMSMAN, M. J., MASTERS, P. S. & ROTTIER, P. J. 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J Virol,* 74**,** 1393-406.

KYRIELEIS, O. J. P., GOETTIG, P., KIEFERSAUER, R., HUBER, R. & BRANDSTETTER, H. 2005. Crystal Structures of the Tricorn Interacting Factor F3 from Thermoplasma acidophilum, a Zinc Aminopeptidase in Three Different Conformations. *J Mol Biol,* 349**,** 787-800.

LAI, M. M. C., AND K. V. HOLMES. 2001. Coronaviridae: the viruses and their replication. *In:* D. M. KNIPE, P. M. H., D. E. GRIFFIN, R. A. LAMB, M. A. MARTIN, B. ROIZMAN, AND S. E. STRAUS (ed.) *Fields virology.* 4 ed. Philadelphia, PA.: Lippincott Williams & Wilkins.

LAI, M. P., S; ANDERSON, LJ 2006. *Fields Virology,* Philadelphia, Lippincott Williams and Wilkins.

LANE, T. E. & HOSKING, M. P. 2010. The pathogenesis of murine coronavirus infection of the central nervous system. *Crit Rev Immunol,* 30**,** 119-30.

LARSEN, S. L., PEDERSEN, L. O., BUUS, S. & STRYHN, A. 1996. T cell responses affected by aminopeptidase N (CD13)-mediated trimming of major histocompatibility complex class II-bound peptides. *The Journal of Experimental Medicine,* 184**,** 183-189.

LAU, S. K., LEE, P., TSANG, A. K., YIP, C. C., TSE, H., LEE, R. A., SO, L. Y., LAU, Y. L., CHAN, K. H., WOO, P. C. & YUEN, K. Y. 2011. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J Virol,* 85**,** 11325-37.

LAU, S. K., POON, R. W., WONG, B. H., WANG, M., HUANG, Y., XU, H., GUO, R., LI, K. S., GAO, K., CHAN, K. H., ZHENG, B. J., WOO, P. C. & YUEN, K. Y. 2010. Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J Virol,* 84**,** 11385-94.

LAU, S. K., WOO, P. C., YIP, C. C., FAN, R. Y., HUANG, Y., WANG, M., GUO, R., LAM, C. S., TSANG, A. K., LAI, K. K., CHAN, K. H., CHE, X. Y., ZHENG, B. J. & YUEN, K. Y. 2012. Isolation and characterization of a novel Betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits. *J Virol,* 86**,** 5481-96.

LESLIE, A. 1992. Jnt CCP4/ESF-EACMB. *Protein Crystallography Newsletters***,** 26.

LEWICKI, D. N. & GALLAGHER, T. M. 2002. Quaternary structure of coronavirus spikes in complex with CEACAM cellular receptors. *J Biol Chem*, 22, 19727-34.

LI, D. & CAVANAGH, D. 1992. Coronavirus IBV-induced membrane fusion occurs at near-neutral pH. *Archives of Virology,* 122**,** 307-316.

LI, F. 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol,* 86**,** 2856-8.

LI, F., LI, W., FARZAN, M. & HARRISON, S. C. 2005a. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science,* 309**,** 1864-8.

LI, W., MOORE, M. J., VASILIEVA, N., SUI, J., WONG, S. K., BERNE, M. A., SOMASUNDARAN, M., SULLIVAN, J. L., LUZURIAGA, K., GREENOUGH, T. C., CHOE, H. & FARZAN, M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature,* 426**,** 450-454.

LI, W., ZHANG, C., SUI, J., KUHN, J. H., MOORE, M. J., LUO, S., WONG, S. K., HUANG, I. C., XU, K., VASILIEVA, N., MURAKAMI, A., HE, Y., MARASCO, W. A., GUAN, Y., CHOE, H. & FARZAN, M. 2005b. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *Embo j,* 24**,** 1634-43.

LOHN, M., MUELLER, C., THIELE, K., KAHNE, T., RIEMANN, D. & LANGNER, J. 1997. Aminopeptidase N-mediated signal transduction and inhibition of proliferation of human myeloid cells. *Adv Exp Med Biol,* 421**,** 85-91.

LOOK, A. T., ASHMUN, R. A., SHAPIRO, L. H. & PEIPER, S. C. 1989. Human myeloid plasma membrane glycoprotein CD13 (gp150) is identical to aminopeptidase N. *J Clin Invest,* 83**,** 1299-307.

LU, G., HU, Y., WANG, Q., QI, J., GAO, F., LI, Y., ZHANG, Y., ZHANG, W., YUAN, Y., BAO, J., ZHANG, B., SHI, Y., YAN, J. & GAO, G. F. 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature,* 500**,** 227-231.

LUAN, Y. & XU, W. 2007. The structure and main functions of aminopeptidase N. *Curr Med Chem,* 14**,** 639-47.

MARZI, A., GRAMBERG, T., SIMMONS, G., MOLLER, P., RENNEKAMP, A. J., KRUMBIEGEL, M., GEIER, M., EISEMANN, J., TURZA, N., SAUNIER, B., STEINKASSERER, A., BECKER, S., BATES, P., HOFMANN, H. & POHLMANN, S. 2004. DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *J Virol,* 78**,** 12090-5.

MASTERS, P. S. 2006. The molecular biology of coronaviruses. *Adv. Virus Res.,* 66**,** 193-292.

MATSAS, R., STEPHENSON, S. L., HRYSZKO, J., KENNY, A. J. & TURNER, A. J. 1985. The metabolism of neuropeptides. Phase separation of synaptic membrane preparations with Triton X-114 reveals the presence of aminopeptidase N. *Biochem J,* 231**,** 445-9.

MATSUYAMA, S., UJIKE, M., MORIKAWA, S., TASHIRO, M. & TAGUCHI, F. 2005. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc Natl Acad Sci U S A,* 102**,** 12543-7.

MCINTOSH, K. 2005. Coronaviruses in the limelight. *J Infect Dis,* 191**,** 489-91.

MCINTOSH, K., DEES, J. H., BECKER, W. B., KAPIKIAN, A. Z. & CHANOCK, R. M. 1967. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc Natl Acad Sci U S A,* 57**,** 933-40.

MCINTOSH, K., KAPIKIAN, A. Z., TURNER, H. C., HARTLEY, J. W., PARROTT, R. H. & CHANOCK, R. M. 1970. Seroepidemiologic studies of coronavirus infection in adults and children. *Am J Epidemiol,* 91**,** 585-92.

MEMISH, Z. A., ASSIRI, A., ALHAKEEM, R., YEZLI, S., ALMASRI, M., ZUMLA, A., AL-TAWFIQ, J. A., DROSTEN, C., ALBARRAK, A. & PETERSEN, E. 2014a. Middle East respiratory syndrome corona virus, MERS-CoV. Conclusions from the 2nd Scientific Advisory Board Meeting of the WHO Collaborating Center for Mass Gathering Medicine, Riyadh. *Int J Infect Dis,* 24**,** 51-3.

MEMISH, Z. A., COTTEN, M., MEYER, B., WATSON, S. J., ALSAHAFI, A. J., AL RABEEAH, A. A., CORMAN, V. M., SIEBERG, A., MAKHDOOM, H. Q., ASSIRI, A., AL MASRI, M., ALDABBAGH, S., BOSCH, B. J., BEER, M., MULLER, M. A., KELLAM, P. & DROSTEN, C. 2014b. Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013. *Emerg Infect Dis,* 20**,** 1012-5.

MINA-OSORIO, P. 2008. The moonlighting enzyme CD13: old and new functions to target. *Trends Mol Med,* 14**,** 361-71.

MINA-OSORIO, P., SHAPIRO, L. H. & ORTEGA, E. 2006. CD13 in cell adhesion: aminopeptidase N (CD13) mediates homotypic aggregation of monocytic cells. *J Leukoc Biol,* 79**,** 719-30.

MIZUSHIMA, S. & NAGATA, S. 1990. PEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Research,* 18**,** 5322.

MOORE, M. J., DORFMAN, T., LI, W., WONG, S. K., LI, Y., KUHN, J. H., CODERRE, J., VASILIEVA, N., HAN, Z., GREENOUGH, T. C., FARZAN, M. & CHOE, H. 2004. Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. *J Virol,* 78**,** 10628-35.

NAKASE, K., KITA, K., SHIKU, H., TANAKA, I., NASU, K., DOHY, H., KYO, T., TSUTANI, H. & KAMADA, N. 1996. Myeloid antigen, CD13, CD14, and/or CD33 expression is restricted to certain lymphoid neoplasms. *Am J Clin Pathol,* 105**,** 761-8.

NAM, E. & LEE, C. 2010. Contribution of the porcine aminopeptidase N (CD13) receptor density to porcine epidemic diarrhea virus infection. *Vet Microbiol,* 144**,** 41-50.

NASH, T. C. & BUCHMEIER, M. J. 1996. Spike Glycoprotein-Mediated Fusion in Biliary Glycoprotein-Independent Cell-Associated Spread of Mouse Hepatitis Virus Infection. *Virology,* 223**,** 68-78.

NAVAS, S., SEO, S. H., CHUA, M. M., DAS SARMA, J., LAVI, E., HINGLEY, S. T. & WEISS, S. R. 2001. Murine coronavirus spike protein determines the ability of the virus to replicate in the liver and cause hepatitis. *J Virol,* 75**,** 2452-7.

NGUYEN, T. T., CHANG, S.-C., EVNOUCHIDOU, I., YORK, I. A., ZIKOS, C., ROCK, K. L., GOLDBERG, A. L., STRATIKOS, E. & STERN, L. J. 2011. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat Struct Mol Biol,* 18**,** 604-613.

NOMURA, R., KIYOTA, A., SUZAKI, E., KATAOKA, K., OHE, Y., MIYAMOTO, K., SENDA, T. & FUJIMOTO, T. 2004. Human coronavirus 229E binds to CD13 in rafts and enters the cell through caveolae. *J Virol,* 78**,** 8701-8.

NOREN O, S. H. A. O. J. 1997. Aminopeptidase N. *In:* KENNY AJ, B. C. (ed.) *In Cell-Surface Peptidases in Health and Disease.* Oxford: BIOS Scientific Publishers.

OLSEN, J., COWELL, G. M., KONIGSHOFER, E., DANIELSEN, E. M., MOLLER, J., LAUSTSEN, L., HANSEN, O. C., WELINDER, K. G., ENGBERG, J., HUNZIKER, W. & ET AL. 1988. Complete amino acid sequence of human intestinal aminopeptidase N as deduced from cloned cDNA. *FEBS Lett,* 238**,** 307-14.

OMRAN, E. K. A. A. A. 2014. Concerns Spread as New Saudi MERS Cases Spike. Available from: http://blogs.wsj.com/middleeast/2014/04/20/concerns-spread-as-new-saudi-mers-cases-spike/ [Accessed Apr 20, 2014 2014].

ORDONO, D., ENJUANES, L. & CASASNOVAS, J. M. 2006. Methods for preparation of low abundance glycoproteins from mammalian cell supernatants. *International Journal of Biological Macromolecules,* 39**,** 151-156.

PANJIKAR, S., PARTHASARATHY, V., LAMZIN, V. S., WEISS, M. S. & TUCKER, P. A. 2009. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr,* 65**,** 1089-97.

PATRICK C. Y. WOO, S. K. P. L., CAROL S. F. LAM, CANDY C. Y. LAU, ALAN K. L. TSANG, JOHN H. N. LAU, RU BAI, JADE L. L. TENG, CHRIS C. C. TSANG, MING WANGE, BO-JIAN ZHENG, KWOK-HUNG CHAN AND KWOK-YUNG YUEN 2012. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. *J Virol,* 86**,** 3995-4008.

PENG, G., SUN, D., RAJASHANKAR, K. R., QIAN, Z., HOLMES, K. V. & LI, F. 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc Natl Acad Sci USA,* 108**,** 10696-10701.

PERLMAN, S. 1998. Pathogenesis of coronavirus-induced infections. Review of pathological and immunological aspects. *Adv. Exp. Med. Biol.,* 440**,** 503-13.

PERLMAN, S. & DANDEKAR, A. A. 2005. Immunopathogenesis of coronavirus infections: implications for SARS. *Nat Rev Immunol,* 5**,** 917-27.

PERLMAN, S. & NETLAND, J. 2009. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol,* 7**,** 439-50.

PHILLIPS, J. J., CHUA, M. M., LAVI, E. & WEISS, S. R. 1999. Pathogenesis of chimeric MHV4/MHV-A59 recombinant viruses: the murine coronavirus spike protein is a major determinant of neurovirulence. *J Virol,* 73**,** 7752-60.

PRO, B. & DANG, N. H. 2004. CD26/dipeptidyl peptidase IV and its role in cancer. *Histol Histopathol,* 19**,** 1345-51.

READ, R. J. 2001. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr D Biol Crystallogr,* 57**,** 1373-82.

REGUERA, J., ORDONO, D., SANTIAGO, C., ENJUANES, L. & CASASNOVAS, J. M. 2011. Antigenic modules in the N-terminal S1 region of the transmissible gastroenteritis virus spike protein. *J Gen Virol,* 92**,** 1117-26.

REGUERA, J., SANTIAGO, C., MUDGAL, G., ORDONO, D., ENJUANES, L. & CASASNOVAS, J. M. 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog,* 8**,** e1002859.

REUSKEN, C. B. E. M., HAAGMANS, B. L., MÜLLER, M. A., GUTIERREZ, C., GODEKE, G.-J., MEYER, B., MUTH, D., RAJ, V. S., VRIES, L. S.-D., CORMAN, V. M., DREXLER, J.-F., SMITS, S. L., EL TAHIR, Y. E., DE SOUSA, R., VAN BEEK, J., NOWOTNY, N., VAN MAANEN, K., HIDALGO-HERMOSO, E., BOSCH, B.-J., ROTTIER, P., OSTERHAUS, A., GORTÁZAR-SCHMIDT, C., DROSTEN, C. & KOOPMANS, M. P. G. 2013. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *The Lancet Infectious Diseases,* 13**,** 859-866.

RIEMANN, D., KEHLEN, A. & LANGNER, J. 1999. CD13--not just a marker in leukemia typing. *Immunol Today,* 20**,** 83-8.

RODGERS, D. W. 1994. Cryocrystallography. *Structure,* 2**,** 1135-40.

ROSSMANN, M. G. 1989. The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. *J Biol Chem,* 264**,** 14587-90.

SAMSON WONG, S. L., PATRICK WOO ANDKWOK-YUNG YUEN 2007. Bats as a continuing source of emerging infections in humans. *Reviews in Medical Virology,* 17**,** 67–91.

SANCHEZ, C. M., GEBAUER, F., SUNE, C., MENDEZ, A., DOPAZO, J. & ENJUANES, L. 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology,* 190**,** 92-105.

SANCHEZ, C. M., IZETA, A., SANCHEZ-MORGADO, J. M., ALONSO, S., SOLA, I., BALASCH, M., PLANA-DURAN, J. & ENJUANES, L. 1999. Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. *J Virol,* 73**,** 7607-18.

SÁNCHEZ, C. M., JIMÉNEZ, G., M.D., L., CORREA, I., SUÑÉ, C., BULLIDO, M., GEBAUER, F., SMERDOU, C., CALLEBAUT, P., ESCRIBANO, J. M. & ENJUANES, L. 1990. Antigenic homology among coronaviruses related to transmissible gastroenteritis virus. *Virology,* 174**,** 410-417.

SCHULTZE, B., CAVANAGH, D. & HERRLER, G. 1992. Neuraminidase treatment of avian infectious bronchitis coronavirus reveals a hemagglutinating activity that is dependent on sialic acid-containing receptors on erythrocytes. *Virology,* 189**,** 792-4.

SCHULTZE, B., GROSS, H. J., BROSSMER, R. & HERRLER, G. 1991a. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *J Virol,* 65**,** 6232-7.

SCHULTZE, B. & HERRLER, G. 1992. Bovine coronavirus uses N-acetyl-9-O-acetylneuraminic acid as a receptor determinant to initiate the infection of cultured cells. *J Gen Virol,* 73 (Pt 4)**,** 901-6.

SCHULTZE, B., WAHN, K., KLENK, H. D. & HERRLER, G. 1991b. Isolated HE-protein from hemagglutinating encephalomyelitis virus and bovine coronavirus has receptor-destroying and receptor-binding activity. *Virology,* 180**,** 221-8.

SCHWEGMANN-WESSELS, C., ZIMMER, G., LAUDE, H., ENJUANES, L. & HERRLER, G. 2002. Binding of transmissible gastroenteritis coronavirus to cell surface sialoglycoproteins. *J Virol,* 76**,** 6037-43.

SIMMONS, G., REEVES, J. D., RENNEKAMP, A. J., AMBERG, S. M., PIEFER, A. J. & BATES, P. 2004. Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc Natl Acad Sci U S A,* 101**,** 4240-5.

SONG, C. S., LEE, Y. J., LEE, C. W., SUNG, H. W., KIM, J. H., MO, I. P., IZUMIYA, Y., JANG, H. K. & MIKAMI, T. 1998. Induction of protective immunity in chickens vaccinated with infectious bronchitis virus S1 glycoprotein expressed by a recombinant baculovirus. *J Gen Virol,* 79 (Pt 4)**,** 719-23.

SONG, D. & PARK, B. 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes,* 44**,** 167-75.

STANLEY, P. 1989. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol Cell Biol,* 9**,** 377-83.

STAUBER, R., PFLEIDERERA, M. & SIDDELL, S. 1993. Proteolytic cleavage of the murine coronavirus surface glycoprotein is not required for fusion activity. *J Gen Virol,* 74 (Pt 2)**,** 183-91.

STEHLE, T. & CASASNOVAS, J. M. 2009. Specificity switching in virus-receptor complexes. *Curr Opin Struct Biol,* 19**,** 181-8.

SUNE, C., JIMENEZ, G., CORREA, I., BULLIDO, M. J., GEBAUER, F., SMERDOU, C. & ENJUANES, L. 1990. Mechanisms of transmissible gastroenteritis coronavirus neutralization. *Virology,* 177**,** 559-69.

TAN, K., ZELUS, B. D., MEIJERS, R., LIU, J.-H., BERGELSON, J. M., DUKE, N., ZHANG, R., JOACHIMIAK, A., HOLMES, K. V. & WANG, J.-H. 2002. Crystal structure of murine sCEACAM1a[1,4]:a coronavirus receptor in the CEA family. *EMBO J.,* 21**,** 2076-2086.

THORP, E. B. & GALLAGHER, T. M. 2004. Requirements for CEACAMs and cholesterol during murine coronavirus cell entry. *J Virol,* 78**,** 2682-92.

TOWLER, P., STAKER, B., PRASAD, S. G., MENON, S., TANG, J., PARSONS, T., RYAN, D., FISHER, M., WILLIAMS, D., DALES, N. A., PATANE, M. A. & PANTOLIANO, M. W. 2004. ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis. *J Biol Chem,* 279**,** 17996-18007.

TRESNAN, D. B. & HOLMES, K. V. 1998. Feline aminopeptidase N is a receptor for all group I coronaviruses. *Adv Exp Med Biol,* 440**,** 69-75.

TUSELL, S. M., SCHITTONE, S. A. & HOLMES, K. V. 2007. Mutational analysis of aminopeptidase N, a receptor for several group 1 coronaviruses, identifies key determinants of viral host range. *J Virol,* 81**,** 1261-73.

TZENG, S.-R. & KALODIMOS, C. G. 2013. Allosteric inhibition through suppression of transient conformational states. *Nat Chem Biol,* 9**,** 462-465.

WANG, L., BYRUM, B. & ZHANG, Y. 2014. New variant of porcine epidemic diarrhea virus, United States, 2014. *Emerg Infect Dis,* 20**,** 917-9.

WANG, N., SHI, X., JIANG, L., ZHANG, S., WANG, D., TONG, P., GUO, D., FU, L., CUI, Y., LIU, X., ARLEDGE, K. C., CHEN, Y.-H., ZHANG, L. & WANG, X. 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res,* 23**,** 986-993.

WEISMILLER, D. G., STURMAN, L. S., BUCHMEIER, M. J., FLEMING, J. O. & HOLMES, K. V. 1990. Monoclonal antibodies to the peplomer glycoprotein of coronavirus mouse hepatitis virus identify two subunits and detect a conformational change in the subunit released under mild alkaline conditions. *J Virol,* 64**,** 3051-5.

WEISS, S. & NAVAS-MARTIN, S. 2005 Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiol. Mol. Biol. Rev.,* 69**,** 635–664.

WESSELING, J. G., GODEKE, G. J., SCHIJNS, V. E., PREVEC, L., GRAHAM, F. L., HORZINEK, M. C. & ROTTIER, P. J. 1993. Mouse hepatitis virus spike and

nucleocapsid proteins expressed by adenovirus vectors protect mice against a lethal infection. *J Gen Virol,* 74 (Pt 10)**,** 2061-9.

WILSON, L., MCKINLAY, C., GAGE, P. & EWART, G. 2004. SARS coronavirus E protein forms cation-selective ion channels. *Virology,* 330**,** 322-331.

WINTER, C., SCHWEGMANN-WESSELS, C., CAVANAGH, D., NEUMANN, U. & HERRLER, G. 2006. Sialic acid is a receptor determinant for infection of cells by avian Infectious bronchitis virus. *J Gen Virol,* 87**,** 1209-16.

WONG, A. H., ZHOU, D. & RINI, J. M. 2012. The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing. *J Biol Chem,* 287**,** 36804-13.

WOO, P. C., LAU, S. K., CHU, C. M., CHAN, K. H., TSOI, H. W., HUANG, Y., WONG, B. H., POON, R. W., CAI, J. J., LUK, W. K., POON, L. L., WONG, S. S., GUAN, Y., PEIRIS, J. S. & YUEN, K. Y. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol,* 79**,** 884-95.

WOO, P. C., WANG, M., LAU, S. K., XU, H., POON, R. W., GUO, R., WONG, B. H., GAO, K., TSOI, H. W., HUANG, Y., LI, K. S., LAM, C. S., CHAN, K. H., ZHENG, B. J. & YUEN, K. Y. 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J Virol,* 81**,** 1574-85.

WOODS, R. D., WESLEY, R. D. & KAPKE, P. A. 1988. Neutralization of porcine transmissible gastroenteritis virus by complement-dependent monoclonal antibodies. *Am J Vet Res,* 49**,** 300-4.

WU, H., KWONG, P. D. & HENDRICKSON, W. A. 1997. Dimeric association and segmental variability in the structure of human CD4. *Nature,* 387**,** 527-30.

WU, K., LI, W., PENG, G. & LI, F. 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc Natl Acad Sci U. S. A.,* 106**,** 19970-19974.

XU, Y., WELLNER, D. & SCHEINBERG, D. A. 1997. Cryptic and regulatory epitopes in CD13/aminopeptidase N. *Exp Hematol,* 25**,** 521-9.

YANG, Z. Y., HUANG, Y., GANESH, L., LEUNG, K., KONG, W. P., SCHWARTZ, O., SUBBARAO, K. & NABEL, G. J. 2004. pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J Virol,* 78**,** 5642-50.

YOO, D. W., PARKER, M. D., SONG, J., COX, G. J., DEREGT, D. & BABIUK, L. A. 1991. Structural analysis of the conformational domains involved in neutralization of bovine coronavirus using deletion mutants of the spike glycoprotein S1 subunit expressed by recombinant baculoviruses. *Virology,* 183**,** 91-8.

ZELUS, B. D., SCHICKLI, J. H., BLAU, D. M., WEISS, S. R. & HOLMES, K. V. 2003. Conformational changes in the spike glycoprotein of murine coronavirus are induced at 37 degrees C either by soluble murine CEACAM1 receptors or by pH 8. *J Virol,* 77**,** 830-40.

ZHANG, G., HU, Y., WANG, H., ZHANG, L., BAO, Y. & ZHOU, X. 2012. High incidence of multiple viral infections identified in upper respiratory tract infected children under three years of age in Shanghai, China. *PLoS One,* 7**,** e44568.

APPENDIX

Appendix

TABLE A: Residues at the crystallographic interface of APN dimer.

Monomer1	Mon. 2	pAPN 939	pAPNeH 908	pAPN-RBD 985	hAPN 962	Mon. 2	Monomer1
V835, W836	D707					D710	W839
V835, W836, N839	R708					R711	L838, W839, N842
W836	S709					S712	W839, R843, S846
W836, R840	E710					E713	W839, R843, S846
W836	Y712					Y715	W839
W836	P714					P717	E806
E834, W836	G713					G716	E806
Q803, D804	K717					N720	E806
G713	Q803					E806	G716, P717, N720
I868	N833					K836	K836, E837, I871
K716	E834					E837	K836
B867, I868, D707, R708	V835					L838	R711, V870
R708, S709, E710, V711, Y712, G713	W836					W839	D710, R711, S712, E713, Y715, G716, K719
R708	N839					N842	R711, S712
E710	R840					R843	E713
E710	Y844					Y847	E713
V835	V867					V870	L838
N833, V835, I868	I868					I871	K836, I871
P871	P871						
F910	Q875					D878	N842, Y913
F910	Q878					Q881	Y913
F910	S879					S882	Y913
D875, Q878, S879	F910					Y913	D878, Q881, S882, Y913
Q878, Q914	Q914					Q917	D878
Q917	Q917						

Buried surface areas (Å2) in the dimer interfaces of the APN structures are shown in the first row. APN residues from monomers 1 and 2 that interact in the dimers are on the same row. pAPN and hAPN interacting residues in the crystal structures are shown on the left and on the right, respectively. Contacts absent in specific structures are marked white in the central columns, whereas residues that mediate polar interactions are red.
Determined with the PISA server (http://www.ebi.ac.uk/msdsrv/prot_int/pistart.html).

TABLE B: Oligonucleotide primers used in the study

SNo.	Identity	Sequence (5'-3')
Aminopeptidase N		
	APN-N1	ATGGAATTCGCGTTCGACTACCTCTGG
	APN-N2	ATGTCTAGAGCGTTCGACTACCTCTGG
	APN-N3	TTGTCGACATGGCCAAGGGATTCTAC
	APN-N4	GGGGAGAAGAACAAGAATGCC-3´
	APN-C1	TATCTAGACCGGATCCACGCGGAACCAGGCTGTGCTCTATGACCA
	APN-C2	TCAGGATCCACGCGGAACCAG
	APN-C3	TTTCTAGAGCCACGCGGAACCAGTTCGGAGTCGAGGAGGAA
	APN-C4	TCAGGATCCGCTGTGCTCTATGAACCA
	hAPN-C1	TCATGCGGCCGCTCATTTGCTGTTTTCTGTGAA
	APN-Call	TCAGCGGCCGCTCAGGATCCACGCGGAACCAG
	APN-N738/A.	ACTCTCACTAAAGCCTGGACCGAGCGC
	APN-T7360V.	ACTAAAAACTGGGTCGAGCGCCCAGAA
	APN-E406A	GACCTGTGGCTGAATGCGGGCTTTGCCTCCTATG
	APN-HH/AA	GTCACTGTGATTGCTGCCGAGCTGGCCGCCCAGTGGTTTGGCAAC
	APNV835T	TGCAGCAACGAGACCTGGCTCCTGAAC
Coronavirus Spike		
	CoS-N1	TT GTCGACATGAAAAAACTATTTGTGG
	CoS-N2	ATGTCTAGAAATAACATTAAATGCTCT
	CoS-N3	ATGTCTAGAACTGCTAATTTGAATAAT
	CoS-N4	ATGTCTAGATTACTACCTAGCTTTTAC
	CoS-C1	TCAGGATCCACGCGGAACCAGATTACCAGTGCTAATTGG
	CoS-C2	TCAGGATCCACGCGGAACCAGGCCATTATAGTATTGAGC
	CoS-C3	TCAGGATCCACGCGGAACCAGCACTATGTTGTCTCCTTC
	CoS-C4	TCAGGATCCACGCGGAACCAAATTAAAAGATATACA
	CoS-C5	TCAGGATCCACGCGGAACCAGAATGACGTCTACAGTGTT
	Cos-C32	TCAGGATCCACGCGGAACCAGATTATCAGACGGTACACC
	R525A	CTTGGTATGAAGGCGAGTGGTTATGGT
	G527D.	ATGAAGCGTAGTGATTATGGTCAACCC
	Y528A.	CGTAGTGGTGCTGGTCAACCC
	G529D.	AGT GGTTATGATCAACCCATAGCC
	Q530A.	AGTGGTTATGGTGCACCCATAGCCTCA
	I532A.	TAATGTTGAGGCTGCGGGTTGACCATA
	L570A.	TGCAAAAGTTCTGCATGGGACAATATT
	W571A.	GCAAAAGTTCTTTAGCGGACAATATTTTTAAGC
	LWD572A.	TGC AAA AGT TCT GCC AAT ATT TTT AAG CGA
	DV582AA.	CGA AAC TGC ACG GCC GCT TTA GAT GCC ACA
	V617A	TCGTTGAGTCCTGATGGTGCTATTGT
	V617D	TCGTTGAGTCCTG**A**TGGTGCTAATTGT
	R630A.	GCT GCC CGT ACA GCT ACC AAT GAG CAG
	TN632AA.	GCC CGT ACA AGA **G**CC **GC**T GAG CAG GTT GTT
	N632A.	CGTACAAGAACCGCTGAGCAGGTTGTT
	Q634A.	AGAACCAATGAGGCGGTTGTTAGAAGT
	NQ634AA.	CGTACAAGAACCGCTGCGGCGGTTGTTAGAAGT
Miscellaneous		
	pEF-1a	TCAAGCCTCAGACAGTGGTTC
	pEF-R	CTCTCAAGGGTCCCAGGT
	ClaI-4829	AACAGACACACGTTCATGG
	ClaI-4310	CTTTTATCAGGGTGCTTTGG