Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Structure of the S1 subunit C-terminal domain from bat-derived coronavirus HKU5 spike protein

Xue Han\textsuperscript{a,b}, Jianxun Qi\textsuperscript{a}, Hao Song\textsuperscript{a,c}, Qihui Wang\textsuperscript{d,e}, Yanfang Zhang\textsuperscript{a,b}, Ying Wu\textsuperscript{f}, Guangwen Lu\textsuperscript{g}, Kwok-Yung Yuen\textsuperscript{h,i,j}, Yi Shi\textsuperscript{a,b,e,k}, George F. Gao\textsuperscript{a,b,c,e,j,k,l,*}

\textsuperscript{a} CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China
\textsuperscript{b} University of Chinese Academy of Sciences, Beijing 100049, China
\textsuperscript{c} Research Network of Immunity and Health (RNIH), Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China
\textsuperscript{d} CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China
\textsuperscript{e} Shenzhen Key Laboratory of Pathogen and Immunity, Shenzhen Third People’s Hospital, Shenzhen 518112, China
\textsuperscript{f} School of Basic Medical Sciences, Wuhan University, Wuhan 430071, China
\textsuperscript{g} West China Hospital Emergency Department (WCHED), State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, and Collaborative Innovation Center of Biotherapy, Chengdu, Sichuan 610041, China
\textsuperscript{h} State Key Laboratory for Emerging Infectious Diseases, The University of Hong Kong, Pokfulam 999077, Hong Kong Special Administration Region
\textsuperscript{i} Department of Microbiology, The University of Hong Kong, Pokfulam 999077, Hong Kong Special Administration Region
\textsuperscript{j} Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou 310003, China
\textsuperscript{k} Center for Influenza Research and Early-warning (CASCIRE), Chinese Academy of Sciences, Beijing 100101, China
\textsuperscript{l} National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing 102206, China

\textbf{ARTICLE INFO}

Keywords:
MERS-CoV
BatCoV HKU5
CTD
Crystal structure
Evolution

\textbf{ABSTRACT}

Accumulating evidence indicates that MERS-CoV originated from bat coronaviruses (BatCoVs). Previously, we demonstrated that both MERS-CoV and BatCoV HKU4 use CD26 as a receptor, but how the BatCoVs evolved to bind CD26 is an intriguing question. Here, we solved the crystal structure of the S1 subunit C-terminal domain of HKU5 (HKU5-CTD), another BatCoV that is phylogenetically related to MERS-CoV but cannot bind to CD26. We observed that the conserved core subdomain and those of other betacoronaviruses (betaCoVs) have a similar topology of the external subdomain, indicating the same ancestor of lineage C betaCoVs. However, two deletions in two respective loops located in HKU5-CTD result in conformational variations in CD26-binding interface and are responsible for the non-binding of HKU5-CTD to CD26. Combined with sequence variation in the HKU5-CTD receptor binding interface, we propose the necessity for surveilling the mutation in BatCoV HKU5 spike protein in case of bat-to-human interspecies transmission.

1. Introduction

Coronaviruses (CoVs) are spherical enveloped viruses with single-positive-strand RNA genomes of ~30 kb in length, which is the largest among RNA viruses (Saif, 1993). CoVs are divided into four genera: alpha-, beta-, gamma-, and deltaCoVs (de Groot et al., 2013). BetaCoVs are further subdivided into four lineages/subgroups: A, B, C, and D (Chan et al., 2015). To date, both alpha- and betaCoVs are found to infect humans (Lu et al., 2015), causing subclinical or very mild symptoms and accounting for 10–15% of common colds (Heikkinen and Jarvinen, 2003). In addition, CoVs can also be life-threatening and have pandemic potential. The epidemic of severe acute respiratory syndrome coronavirus (SARS-CoV), which belongs to lineage B of the betaCoVs, originated in southern China in 2002 and spread to 28 countries, infecting over 8000 and leading to almost 800 related deaths (WHO, 2004). The outbreak of MERS-CoV, a member of the lineage C betaCoVs (Cotten et al., 2013; Zaki et al., 2012), has caused 1832 laboratory-confirmed cases since 2012, including 651 related deaths as of Nov. 28, 2016 (WHO, 2016). Unlike the SARS-CoV, which suddenly disappeared after a massive global disease control effort, especially in China, the number of MERS-CoV infections is still on the rise.

Mounting evidence indicates that CoVs circulating in bats (BatCoVs) are the gene sources of alphaCoVs and betaCoVs (W. Li et al., 2005; Woo et al., 2012), including SARS-CoV (Ge et al., 2013; Lau et al., 2005; W. Li et al., 2005). The data also underscore that bats are the likely natural reservoir for MERS-CoV (Annan et al., 2013; Ihthe et al., 2013; Memish et al., 2013; Wang et al., 2014; Yang et al., 2014). For instance, viral gene fragments identical or quite similar to
those of MERS-CoV have been reported in bats (Annan et al., 2013; Ithe et al., 2013; Memish et al., 2013). Moreover, parallel studies from our group and others show that BatCoV HKU4, grouped in lineage C with MERS-CoV, can also use human CD26 (hCD26; the receptor of MERS-CoV) for viral entry (Wang et al., 2014; Yang et al., 2014). In other words, two members in lineage C use the same human receptor. One has caused a human infection epidemic (MERS-CoV), and the other can utilize the same receptor (BatCoV HKU4) and has potential to infect humans. This highlights the necessity of surveillance for lineage C betaCoVs, including BatCoV HKU5, which was first sequenced in 2006 in Japanese pipistrelles (Pipistrellus abramus) (Woo et al., 2006) and is circulating in bats (Lau et al., 2013). Whether the virus has the potential to bypass the bat-human barrier needs to be evaluated.

CoV infections initiate with the virus binding to the host receptor. The envelope-impervious trimeric spike (S) protein plays a pivotal role in this process. S is further divided into two parts: S1, responsible for receptor binding, and S2, which initiates fusion (Belouzard et al., 2012; Dai et al., 2016; Kielland and Rey, 2006). S1 contains two relatively independent structures named the N-terminal domain (NTD) and C-terminal domain (CTD) based on their position. Most betaCoVs use the CTD as the receptor-binding domain (RBBD/CTD) except mouse hepatitis virus (MHV), which uses the NTD to bind the cellular receptor carcioembryonic-antigen-related cell-adhesion molecule 1 (CEACAM1) (Dai et al., 2016; Du et al., 2009; F. Li et al., 2005; Lu et al., 2013; Peng et al., 2011). Two of the RBBD/CTDs in lineage C betaCoVs (MERS-RBD/CTD and HKU4-RBD/CTD) bind to the same human receptor CD26 (hCD26) to initiate infection, and the two domains share high sequences identities (55%) in addition to high structural similarities, with a root mean square deviation (rmsd) of 1.14 (193 Ca) (Lu et al., 2013; Wang et al., 2014). Despite the similar sequence identities between HKU5-CTD and MERS-RBD/CTD (54%) or HKU4-RBD/CTD (57%), no detectable binding was found between HKU5-CTD and hCD26. The structural basis for this variation remains to be elucidated.

In this study, we determined the structure of HKU5-CTD. Similar to other solved structures, HKUS-CTD contains two subdomains: the core subdomain homologous to other CTDs in betaCoVs and the external subdomain, which resembles MERS-RBD/CTD and HKU4-RBD/CTD, indicating conservation of the external domain in lineage C. However, two deletions in HKUS-CTD lead to structural shifts in the hCD26-interaction interface and thereby result in its inability to bind this receptor. Our results suggest that the characteristic insertions in β4 and β5 among different lineages in betaCoVs result in different receptor engagement, thereby contributing for potential interspecies transmission.

2. Results

2.1. Overall structure of the HKU5-CTD

We first characterized the S protein of BatCoV HKU5 through bioinformatics analysis. BatCoV HKU5 S is composed of 1352 amino acids and exhibits typical features of CoVs S protein, including the predicted hydrophobic residue-rich HR1 and HR2 motifs and a similar concentration of hydrophobic amino acids to SARS-CoV fusion peptide (FP), internal fusion peptide (IFP), and pre transmembrane domain (PTM) (Gao et al., 2013; Mahajan and Bhattacharjya, 2015; Xu et al., 2004; Zhu et al., 2004). Like MERS-CoV S protein, a furin-like protease recognition motif is predicted at position R745/A746 (S1/S2), which separates the S1 and S2 subunits (Millet and Whittaker, 2014). In addition, a second furin cleavage site can be found at R884/S885, which resembles S2' in the MERS-CoV S protein (Millet and Whittaker, 2014), indicating that the priming process of BatCoV HKU5 S in human cells probably occur in the same way like MERS-CoV (Fig. 1A). Because most betaCoVs use their CTD to bind their respective receptors, we next focused on the evolutionary relationships of the CTDs. Consistent with the phylogenetic relationships, HKU5-CTD, HKU4-RBD/CTD, and MERS-RBD/CTD are grouped in one branch representing lineage C, while HKU1-CTD and MHV-CTD cluster together in lineage A. HKU9-CTD, a member in lineage D, is phylogenetically more related to SARS-RBD/CTD, which belongs to lineage B (Fig. 1B).

The HKU5-CTD was then purified, crystallized and the structure was successfully determined at a resolution of 2.1 Å, with clear electron densities tracing from Q389 to Q586. The structure, solved through the molecular replacement method, contains a single molecule in the crystallographic asymmetric unit, with an Rwork of 0.2160 and an Rfree of 0.2585 (Table 1). Like the other solved CTD structures of betaCoVs (Huang et al., 2016; Kirchdoerfer et al., 2016; F. Li et al., 2005; Lu et al., 2013; Walls et al., 2016; Wang et al., 2013, 2014), HKU5-CTD folds into two discrete subdomains, the core and the external (Fig. 2). The core subdomain contains a five-stranded anti-parallel scaffold center (core-center), which is decorated by five helices (αα310) and two small strands (βp1 and βp2) on the exterior. Three pairs of disulfide bonds help to stabilize the scaffold, namely C391-C415 and C445-C583, located in the peripheral region of the core subdomain (core-peripheral), and C433-C486 in the core-center, linking βc2 and β4. Notably, two antiparallel β strands, one of which is located in the C-terminus and the other forming the disulfide bond with the N-terminus, help to make keep two termini in proximity. In addition, traceable electron densities can be observed for two glycosylation modifications at N418 and N495, which form two protrusions at the core-peripheral region (Fig. 2).

The external subdomain of HKU5-CTD extends out of βc4 in the core-center, sequentially folds into two antiparallel β strands (β1’ and β2’), an α helix (H1’), and another two antiparallel β strands (β3’ and β4’), and finally proceed into βc5. Between β1’ and H1’, a pair of disulfide bonds (C511-C532) is formed to stabilize the external structure (Fig. 2).

2.2. Conserved core subdomain and variable external subdomain for betaCoVs S protein

To date, seven structures of CTDs covering all four lineages of betaCoVs have been solved. They are HKU1-CTD and MHV-CTD, belonging to lineage A (Kirchdoerfer et al., 2016; Walls et al., 2016), MERS-RBD/CTD (Lu et al., 2013; Wang et al., 2013), HKU4-RBD/ CTD (Wang et al., 2014), and HKU5-CTD grouped in lineage C (reported here), and SARS-RBD/CTD (F. Li et al., 2005) and HKU9-CTD (Huang et al., 2016) representing lineages B and D, respectively. All seven betaCoV CTD structures display a conserved core subdomain, with five antiparallel beta strands and a conserved disulfide bond between βc2 and βc4 (Fig. 3).

Despite the different combinations of α helices and β strands, the orientations of the secondary structures are conserved among CTDs in the core-peripheral region. In addition, two highly conserved disulfide bonds exist. One is formed between the N-terminus and the loop/β strand extended from βc1, and the other links the C-terminus and the loop/β strand proceeding to βc3. Thus, through the two disulfide bonds, both termini are brought into close proximity (Fig. 3). Although in the SARS-RBD/CTD electron density at the C-terminus is not clear enough to determine the structures (Fig. 3C), two conserved cysteines are present, indicating the probability of disulfide bond formation (Fig. 1C).

Opposite to the conserved core subdomain, the external subdomain varies considerably among different lineages. In lineage A, the external subdomain of MHV-CTD, which was obtained by density-guided homology modelling due to its large flexibility and poor quality of the density in this region, consists of four β strands and three small helices (PDB code: 3C3L) (Fig. 3A). HKU1-CTD is comprised of a large, variable loop with three inlaid β strands (Fig. 3B). The absence of clear
secondary structure from residues C476-F572 indicates the flexibility of this region (PDB code: 5I08). In lineage C, three CTDs show similar external folds, with rmsd ranging from 0.962 (HKU5-CTD vs. MERS-RBD/CTD (PDB code: 4KQZ)) to 1.178 (HKU5-CTD vs. HKU4-RBD/CTD (PDB code: 4QZV)). All external subdomains are strand-dominated structures with four anti-parallel β strands and expose a flat strand-face that is stabilized by a conserved disulfide bond (Fig. 3D-F). In lineage B, the SARS-RBD/CTD is dominated by a disulfide bond-stabilized flexible loop that connects two small β strands (PDB code: 2GHV) (Fig. 3C). In BatCoV HKU9, representing lineage D, the external subdomain only contains one large helix in this region (PDB code: 5GYQ) (Fig. 3G). Although their external subdomain structures differ, all CTDs in betaCoVs extend out from βc4 and proceed back to the core subdomain through βc5 (Fig. 3D-F). This, then, led to different receptor usage if the CTD is utilized as the RBD.
activated cell sorting (FACS) or surface plasmon resonance (SPR) (Fig. 4).

According to the two solved complex structures, four concentrations of residues in MERS-RBD/CTD and HKU4-RBD/CTD are involved in binding to hCD26. These residues located in four beta strands and two loops (β1'/β2' loop and β3'/β4' loop) in the external subdomain as well as H4 and H5 (for MERS-RBD/CTD) or H5 and H6 (for HKU4-RBD/CTD) positioned in the core subdomain and the loop connecting the two helices (Figs. 1C, 3E and 3F). However, half of these regions (β1'/β2' loop and β3') have deletions in HKU5-CTD (Fig. 1C). Due to these deletions, the orientations of two loops (marked 1 and 2, respectively, in Fig. 3D-F) in HKU5-CTD vary compared to MERS-RBD/CTD and HKU4-RBD/CTD, which leads to conformational shifts in HKU5-CTD at the hCD26-binding interface (Fig. 5A and E). The β1'/β2' loop in both MERS-RBD/CTD and HKU4-RBD/CTD inserts into the groove formed by two helices on the side and β strands on the bottom (Fig. 5B and F). Sixty-five (328 in total) and 49 (214 in total) van der Waals contacts, including 5 (16 in total) and 4 (13 in total) hydrogen bonds, are formed in MERS-RBD/CTD/hCD26 and HKU4-RBD/CTD/hCD26, respectively. In contrast, this loop in HKU5-CTD is tilted away by ~6 and 9 Å (Fig. 5B and F) compared to MERS-RBD/CTD and HKU4-RBD/CTD, respectively, which results in the loss of binding to hCD26 at this region.

Moreover, a six-residue deletion in β3' causes large discrepancies in the assemblies of β3', β4' and their connecting loop, compared with MERS-RBD/CTD and HKU4-RBD/CTD (Fig. 3D-F). In β3' of both hCD26-binding RBD/CTDs, the side chains of Y540 and R542 face the receptor, conferring a strong hydrophilic interaction between the ligand and the receptor. In contrast, in HKU5-CTD, the orientation of β3' is opposite. In addition, Y544 in HKU5-CTD likely sterically clashes Q286, which further pushes HKU5 away from hCD26 (Fig. 5C and G). In the other beta strand of β4', both MERS-RBD/CTD and HKU4-RBD/CTD form a large hydrophobic interaction patch with hCD26. In HKU5-CTD, aside from the shift of the β3'/β4' loop away from the receptor, the deletion of hydrophobic residues (e.g., W535) compared to MERS-RBD/CTD (Fig. 5D) and the substitution of hydrophilic residues (e.g., T553 and T555) instead of hydrophobic ones (I560 and V562 in HKU4-RBD/CTD) (Fig. 5H) likely inhibit HKU5-CTD binding to hCD26. In total, the conformational variations between HKU5-CTD and hCD26-binding RBD/CTDs explain the lack of hCD26 binding by HKU5-CTD. However, various deletions in HKU5-CTD loop 1 are present in nature (Fig. 6), and might contribute to evolve for receptor binding.

Table 1

| Data collection          | HKU5 CTD |
|--------------------------|----------|
| Wavelength (Å)           | 0.97915  |
| Space group              | P 21     |
| Cell dimensions          |          |
| α, β, γ (deg)            | 49.612, 212.659, 87.943 |
| Resolution (Å)           | 50.00–2.10 |
| Rwork                    | 0.104 (1.038) |
| Rfree                    | 0.052 (0.527) |
| I / σt                   | 15.355 (1.645) |
| CC1/2                    | 0.998 (0.801) |
| Completeness (%)         | 99.9 (99.9) |
| Redundancy               | 5.0 (4.8) |

Refinement

| Resolution (Å) | 37.80–2.10 |
| No. reflections | 104555     |
| Rwork / Rfree  | 0.2160/0.2585 |
| No. atoms      | 52.9       |
| Ligand/ion     | 527        |
| Water          | 45.2       |
| R.m.s. deviations |
| Bond lengths (Å) | 0.006     |
| Bond angles (deg) | 1.062    |
| Ramachandran plot |
| Favored (%)    | 95.73      |
| Allowed (%)    | 3.44       |
| Outliers (%)   | 0.83       |

* Values in parentheses are for highest-resolution shell.

2.3. Structural basis for HKU5-CTD not binding to CD26

Both MERS-RBD/CTD and HKU4-RBD/CTD bind to hCD26 to initiate infection. In addition, the structure of the HKU5-CTD displays a similar topology to the two RBD/CTDs in lineage C. Thus, we assayed for binding between HKU5-CTD and hCD26. However, consistent with previous results, no binding was detected, either by fluorescence-
3. Discussion

In this study, we solved the crystal structure of HKU5-CTD, which represents the seventh structure of a CTD belonging to a betaCoV. Like the other six CTDs (Huang et al., 2016; Kirchdoerfer et al., 2016; F. Li et al., 2005; Lu et al., 2013; Peng et al., 2011; Walls et al., 2016; Wang et al., 2013, 2014), there are two subdomains in HKU5-CTD, the core and the external. Despite the low residue conservation among CTDs (pair-to-pair amino acid identity ranging from 17.2% to 58.7%) and the core subdomains (pair-to-pair amino acid identity ranging from 16.6% to 66.7%) in the four lineages, the topology of the latter ones are highly conserved, with five anti-parallel β strands constituting the core-center and the same orientation of secondary elements in the core-peripheral (Fig. 3). This includes the same region of MHV, which uses the NTD of S1 to bind the receptor.

However, the external subdomains vary considerably among lineages. In lineage A, the MHV-CTD contains several β strands and inlaid helices, while the HKU1-CTD is comprised of loops and three small β strands. However, approximately 100 amino acids (C476-F572) are unclear at this region, likely due to their flexibility. SARS-RBD/CTD, in lineage B, is dominated by loops, which are stabilized by a disulfide bond and two anti-parallel β strands. Most CTD structures solved to date are in lineage C, and all three CTDs (MERS-RBD/CTD, HKU4-RBD/CTD, and HKU5-CTD) display conserved structures with β strand-forming platforms decorated with helices. In addition, a disulfide bond is conserved among CTDs in lineage C in the external subdomain. HKU9-CTD, a member of lineage D, is comprised of a helix that is clamped with loops. Although different structures and topologies exist among lineages, all external subdomains extend out from βc4 and proceed back to βc5 (Fig. 3), indicating that different insertions between βc4/βc5 during betaCoV evolution have conferred the betaCoVs with different properties, such as receptor usage, and thereby led to the parallel evolution of lineages.

The ligand–receptor interaction is a key factor determining the tissue tropism and host range of CoVs. For SARS-CoV, MERS-CoV, and BatCoV HKU4, the receptors are clear, and the complex structures demonstrate that the receptor mainly binds to the varied external subdomains of CTDs. Neutralizing antibodies against HCoV HKU1

Fig. 3. Topological diagrams of CTDs in betaCoVs. Structural and topological comparison of available betaCoV CTD structures. Seven structures, including those of MHV-CTD (PDB code: 3CJL), HKU1-CTD (PDB code: 5I08), SARS-RBD/CTD (PDB code: 2GHV), HKU5-CTD, HKU4-RBD/CTD (PDB code: 4QZV), MERS-RBD/CTD (PDB code: 4KQZ), and HKU9-CTD (PDB code: 5GYQ) were oriented similarly and are presented as cartoons in parallel. The conserved disulfide bonds are labeled in red lines, while the non-conserved ones are displayed with lines in accordance with the color of indicated external subdomain. Arrows and cylinders represent the strands and helices, respectively.
bind to the HKU1-CTD, not the HKU1-NTD (Qian et al., 2015), indicating that the CTD in HCoV HKU1 is most likely to be the RBD, though the protein receptor has yet to be identified (Chan et al., 2016; Huang et al., 2016). In our study, we found that although HKU5-CTD displays a similar structure and topology to MERS-RBD/CTD and HCoV HKU4-CTD, it does not bind to ACE2 or hCD26 (Huang et al., 2016). The structural basis for the lack of binding between HKU5-CTD and hCD26 is explored in Figure 5. The superimposition of the structures of HKU5-CTD and hCD26 binding-MERS-RBD/CTD or HKU4-RBD/CTD reveals the variations in the receptor binding interface of HKU5-CTD compared with MERS-RBD/CTD or HKU4-RBD/CTD. These variations are further delineated in Figure 5E-H for detailed structural shifts, respectively. The conserved core subdomains in HKU5-CTD, MERS-RBD/CTD, and HKU4-RBD/CTD are colored in grey, while the external subdomains of the three proteins are marked with orange, cyan, and wheat, respectively. The magenta represents hCD26.
HKU4-RBD/CTD, detailed structural analysis revealed variations at the hCD26-binding interface, which results in the loss of binding to this receptor. Thus, subtle distinctions in external subdomains could determine different receptor usage.

In addition to receptor binding, the priming process, which involves host proteases to liberate S2 and the fusion peptides from the otherwise covalently-linked S1 subunit, is another key factor affecting cell tropism and the entry route of CoVs. A two-step activation mechanism has been proposed for MERS-CoV entry (Millet and Whittaker, 2014). During the secretion of S protein, the proteolysis at S1/S2 occurs in the endoplasmic reticulum (ER)-Golgi compartments where furin is localized, while during virus entry into target cells, S2' is cleaved. The same proteolysis in S1/S2 and S2' is also essential for SARS-CoV infection, except that due to the lack of a furin-recognition site at S1/S2, SARS-CoV S remains uncleaved after biosynthesis and a diverse array of proteases are involved in this process (Millet and Whittaker, 2015; Simmons et al., 2011). In contrast, although BatCoV HKU4 can utilize hCD26 as a receptor, the proteolysis is stuck due to the lack of a protease site. Treatment of pseudovirus particles containing BatCoV HKU4 S protein with trypsin or importing the furin-recognition site into S protein enables the particles to infect hCD26-expressing cells (Wang et al., 2014; Yang et al., 2015), indicating the BatCoV HKU4 is less adapted to human cells. However, in the BatCoV HKU5 S protein, both furin-recognition sites are present, as in MERS-CoV. Accordingly, BatCoV HKU5 S is predicted to be cleaved at S1/S2 after biosynthesis and at S2' during virus entry. Furin is a ubiquitous protease and
expressed in nearly all cell lines. The presence of the two furin-recognition sites indicates that the priming process of BatCoV HKU5 is ready to occur.

BatCoV HKU5 has been circulating in bats (Lau et al., 2013). In an epidemiology study over a 7-year period (April 2005 to August 2012), 25% of alimentary specimens of Japanese pipistrelle bats (Pipistrellus abramus) collected from 13 locations in Hong Kong were positive for this virus (Lau et al., 2013), indicating that it might target gastrointestinal tissues. However, BatCoV HKU5 virus has not been isolated and cultured successfully, which is an obstacle to virus transmission research. The problem is largely due to a lack of suitable cell lines for BatCoV HKU5 virus. The emerging but puzzling question is whether this virus could infect humans or not.

An infectious clone of BatCoV HKU5 containing the ectodomain from the SARS-CoV S protein was constructed through reverse genetics and synthetic-genome design, and the recombinant virus replicates efficiently in cell culture and in young and aged mice (Agnihotram et al., 2014). In addition, the key proteins for virus replication, such as the 3C-like protease, polymerase, and exonuclease of BatCoV HKU5 display high amino acid sequence similarity to those in MERS-CoV, indicating that once the genome of BatCoV HKU5 is released into host cells, genome replication, virus particle assembly, and release can readily occur. Therefore, the receptor would be the last barrier for BatCoV HKU5 to infect humans. Our data show that BatCoV HKU5-CTD does not use hCD26 as a receptor, though it folds into a very similar structure as MERS-RBD/CTD and HKU4-RBD/CTD. In other words, the cellular receptor of BatCoV HKU5 is still a mystery that requires further study.

Evolutionally, BatCoV HKU5 S protein is more diverse than BatCoV HKU4 S protein (Lau et al., 2013), and various deletions in loop 1 (Fig. 3D and Fig. 6) have been sequenced. This indicates that BatCoV HKU5 is able to generate variants to occupy new ecological niches and might acquire the ability to bind to hCD26 by accumulating mutations and ultimately cause human respiratory infections like MERS-CoV and SARS-CoV. Accordingly, it is very important to perform long-lasting surveillance of BatCoV HKU5 evolution, especially the variety of S protein in the event that the virus breaks the inter-species and/or inter-tissue transmission barriers.

4. Materials and methods

4.1. Gene construction, protein expression, and purification

The coding region for HKU5-CTD (Q389-Q586) with a 6×His-tag at its C-terminus was cloned into the EcoRI and Xhol sites of pFastBac. HKU5-CTD, MERS-RBD/CTD, HKU4-RBD/CTD, and hCD26 were expressed and purified as previously reported (Lu et al., 2013; Wang et al., 2014). Briefly, the proteins were expressed in baculovirus-infected Hi5 cells. After 48 h of culturing, the supernatants were collected and purified through a 5 mL HisTrap HP column (GE Healthcare). The bound protein was eluted by a gradient of imidazole. Fractions containing the target protein as determined by SDS-PAGE were pooled and further subjected to gel filtration using a Superdex® 75 column (GE Healthcare) in a buffer composed of 20 mM Tris-HCl (pH 8.0) and 50 mM NaCl.

The Fc fusion protein used for cell staining was purified following a previously published method (Lu et al., 2013; Wang et al., 2014). In brief, the plasmid containing the target gene was transfected into HEK293T cells. After 3 days, the supernatants containing secreted protein were pooled and subjected to HiTrap ProteinA chromatography (5 mL, GE Healthcare). Protein was eluted with 0.1 M sodium citrate (pH 4.5) and further purified by gel filtration. The protein was finally buffer-exchanged into PBS, concentrated to ~1 mg/mL, and stored at ~80 °C before further usage.
References

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.D., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution and refinement. Acta. Biol. Crystallogr. Sect. D., Biol. Crystallogr. 66, 213–221.

Agnihotrim, S., Yount, B.L., Donaldson, E.F., Liu, S.W., Zheng, B.J., Jiang, S.B., 2009. The spike protein of Middle East respiratory syndrome coronavirus. J. Virol. 83, 1853–1861.

Annan, A., Baldwin, H.J., Corman, V.M., Klose, S.M., Owusu, M., Nkrumah, E.E., Badu, R., Yentima, S.A., Njoh, P., Mafman, J., Okafor, U., Meyer, B., Szeroka, Y.A., Lam, C.S., Lina, P.H., Goddelska, E.V., Reusken, C., Sebenne, A., Gloria-Rausch, F., Fallo, P., Tschapka, M., Drosten, C., Drexler, J.F., 2013. Human betacoronavirus 2c EMC/2013-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, 1011–1033.

Chan, C.M., Chu, H., Wang, X.Y., Bong, B.H.Y., Zhao, X.Y., Zhou, J., Yang, D., Leung, S.P., Chan, J.F.W., Yeung, M.L., Yan, J.H., Lu, G.W., Gao, G.F., Yuen, K.Y., 2016. Carnivore/insectivore antigen-related cell death molecule 5 is an important surface attachment factor that facilitates entry of Middle East respiratory syndrome coronavirus. J. Virol. 90, 9114–9127.

Chan, J.F.W., Lau, S.K., To, K.K.W., Cheng, V.C.C., Woo, P.C.Y., Yuen, K.Y., 2015. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. Clin. Microbiol. Rev. 28, 465–522.

Cotten, M., Lam, T.T., Watson, S., Palser, A.L., Petrova, V., Grant, P., Pybus, O.G., Rambaut, A., Yuan, Y., Pillay, D., Kellam, P., Nastoulid, E., 2013. Full-genome deep sequencing and phylogenetic analysis of novel human coronavirus. Emerg. Infect. Dis. 19, 736–742.

Dai, L., Wang, Q., Qi, J., Shi, S., Yan, J., Gao, J.F., 2016. Molecular basis of antibody-mediated neutralization and protection against flavivirus. IUBMB Life 68, 783–791.

Debreuyn, J.E., Emelyanov, D., 2012. Handling ligands with Coot. Acta Crystallogr. Sect. D., Biol. Crystallogr. 68, 425–430.

Du, L.Y., He, Y.X., Zou, Y.S., Liu, S.W., Zheng, B.J., Jiang, S.B., 2016. The spike protein of SARS-CoV - a target for vaccine and therapeutic development. Nat. Rev. Microbiol. 14, 27–37.

Gao, J., Lu, G., Qi, J., Li, Y., Wu, Y., Deng, Y., Geng, H., Li, H., Wang, Q., Xiao, H., Tian, W., Yan, J., Gao, G.F., 2013. Structure of the fusion core and inhibition of fusion by a humanized-repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. J. Virol. 87, 13134–13140.

Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., Zhang, Y.-J., Luo, C.-M., Tan, B., Wang, N., Zhub, Y., Hamam, F., Shi, Z.-Y., Wang, L.-F., Daskaz, P., Shi, Z.-L., 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 533–538.

de Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier, R.A.M., Galiano, L., Gorbalenya, A.E., Memish, Z.A., Perlman, S., Polack, P., Snijder, E.J., Stohler, M., Zuberbuhler, J., 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J. Virol. 87, 7790–7792.

Heinkin, T., Jarvinen, A., Lappalainen, H., Mattila, M., Peltola, H., Helminen, M., Hyypia, T., 1991. The coronavirus spike gene. Nature 351, 51–59.

Huang, C., Qi, J., Lu, G., Wang, Q., Yuan, Y., Wu, Y., Zhang, Y., Yan, J., Gao, G.F., 2016. Putative receptor binding domain of bat-derived coronavirus HKU9 spike protein: evolution of betacoronavirus receptor binding motif. Biochemistry 55, 5977–5988.

Huang, C., Qi, J., Du, L., Wang, Q., Liu, C., Wu, Y., Zuo, X., Zhang, G., Shi, Y., Shi, Z.-Y., Yang, Y., Gao, G.F., 2014. Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. J. Virol. 87, 8638–8650.

Li, B., Wang, Q., Peng, C., Zhai, X., Zhang, Z., Zhen, H., Zong, B., Zhou, Q., Jia, D., Wang, J., Li, J., 2015. Identification of the receptor-binding domain of the spike glycoprotein of human coronavirus HKU1. J. Virol. 89, 8816–8827.

Saif, L.J., 1993. Coronavirus immunogens. Vet. Microbiol. 37, 285–297.

Simmons, B., Gerber, S., Glowacka, I., Steffen, I., Chalup, C., Agudelo, J., Lu, K., Renneckamp, A.J., Hofmann, H., Bates, P., Pohlmann, S., 2011. Different host cell proteases activate the SARS-coronavirus spike protein for cell-cell and virus-cell fusion. Virology 413, 265–274.

Walls, A.C., Torborg, M.A., Bosch, B.J., Frenz, B., Rottler, P.J.M., DiMaio, F., Rey, F.A., Felsenfeld, P., 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimmer. Nat. Medicine 531, 118–120.

Arledge, K.C., Chen, Y.H., Zhang, L., Shi, Y., Yan, J., Gao, G.F., 2015. Bat-to-human: spike features determining ‘host jump’ make a hairpin. Nat. Rev. Microbiol. 4, 67–70.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Yuen, K.Y., Chan, K.H., Yuen, K.Y., 2006. Molecular diversity of coronaviruses in bats. Virology 351, 180–187.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 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.

Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., Gao, G.F., Rao, Z., 2004. Crystal structure of severe acute respiratory syndrome coronavirus spike protein core fusion. J. Biol. Chem. 279, 49414–49419.

Yang, Y., Du, L., Liu, C., Wang, Q., Wu, Y., Zhang, Z., Zhen, H., Wang, Q., Cao, G., Jia, D., Wang, J., Li, J., 2015. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of Middle East respiratory coronavirus. Proc. Natl. Acad. Sci. USA 111, 12516–12521.

Yang, Y., Li, D., Lou, J., Shi, S., Zhai, X., Baric, R.S., Li, F., 2015. Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. J. Virol. 89, 9119–9123.

Zhi, A.M., van Boven, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. 367, 1814–1820.

Zhu, J., Xiao, G., Xu, Y., Yuan, F., Zheng, C., Liu, C., Yang, H., Cole, D.K., Bell, J.J., Rao, Z., Tien, P., Gao, G.F., 2004. Following the rule: formation of the 6 helix bundle of the fusion protein from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. Biochem. Biophys. Res. Com. 319, 283–288.