

Putative Receptor Binding Domain of Bat-Derived Coronavirus HKU9 Spike Protein: Evolution of Betacoronavirus Receptor Binding Motifs

Canping Huang,[†] Jianxun Qi,[‡] Guangwen Lu,[§] Qihui Wang,^{||} Yuan Yuan,[⊥] Ying Wu,[‡] Yanfang Zhang,[‡] Jinghua Yan,^{||} and George F. Gao^{*,†,‡,⊥,#,@}

[†]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing 102206, China

[‡]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[§]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

^{||}CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

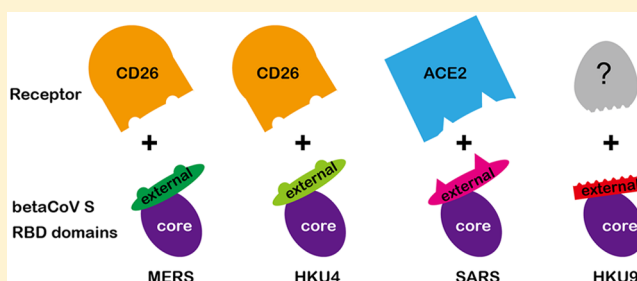
[⊥]School of Life Sciences, University of Science and Technology of China, Hefei, Anhui Province 230026, China

[#]Laboratory of Protein Engineering and Vaccines, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[@]Research Network of Immunity and Health (RNIH), Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China

ABSTRACT: The suggested bat origin for Middle East respiratory syndrome coronavirus (MERS-CoV) has revitalized the studies of other bat-derived coronaviruses with respect to interspecies transmission potential. Bat coronavirus (BatCoV) HKU9 is an important betacoronavirus (betaCoV) that is phylogenetically affiliated with the same genus as MERS-CoV. The bat surveillance data indicated that BatCoV HKU9 has been widely spreading and circulating in bats. This highlights the necessity of characterizing the virus for its potential to cross species barriers. The receptor binding domain (RBD) of the coronavirus spike (S) protein recognizes

host receptors to mediate virus entry and is therefore a key factor determining the viral tropism and transmission capacity. In this study, the putative S RBD of BatCoV HKU9 (HKU9-RBD), which is homologous to other betaCoV RBDs that have been structurally and functionally defined, was characterized via a series of biophysical and crystallographic methods. By using surface plasmon resonance, we demonstrated that HKU9-RBD binds to neither SARS-CoV receptor ACE2 nor MERS-CoV receptor CD26. We further determined the atomic structure of HKU9-RBD, which as expected is composed of a core and an external subdomain. The core subdomain fold resembles those of other betaCoV RBDs, whereas the external subdomain is structurally unique with a single helix, explaining the inability of HKU9-RBD to react with either ACE2 or CD26. Via comparison of the available RBD structures, we further proposed a homologous intersubdomain binding mode in betaCoV RBDs that anchors the external subdomain to the core subdomain. The revealed RBD features would shed light on the evolution route of betaCoV.



Coronaviruses are large, enveloped, and positive-stranded RNA viruses that can infect birds, animals, and humans.^{1,2} Taxonomically, these viruses are affiliated with the *Coronaviridae* family within the *Nidovirales* order.^{1,3} Ever since the 1930s when the first coronavirus of infectious bronchitis virus was isolated in chickens,⁴ coronaviruses have expanded into four genera, *Alpha-*, *Beta-*, *Gamma-*,³ and *Deltacoronavirus*.^{5,6} Of these, betacoronaviruses (betaCoVs) have attracted attention worldwide because of their pathogenic capacity and potential to cause a global pandemic of human infections^{7,8} and the widespread existence of an enormous number of species in bats.^{6,9–11} In 2002 and 2003, one representative betaCoV, the

severe acute respiratory syndrome coronavirus (SARS-CoV), first emerged in China^{12–15} and then rapidly spread to other countries, leading to >8000 cases of infection and >800 deaths.⁷ In 2012, another betaCoV, named the Middle East respiratory syndrome coronavirus (MERS-CoV),¹⁶ was identified first in Saudi Arabia.^{17,18} Despite the global efforts trying to control its transmission, MERS-CoV still spreads to affect multiple countries in the Middle East, Europe, North America, and

Received: August 2, 2016

Revised: September 26, 2016

Published: October 3, 2016

Asia, causing 1800 confirmed infections and at least 640 deaths as of June 23rd, 2016 (based on the latest statistical data released by the World Health Organization⁸). Meanwhile, a human-infective betaCoV of HKU1 was isolated from a patient with respiratory disease in Hong Kong.¹⁹ These unexpected outbreaks of betaCoV infection have posed a severe threat to global public health and led to enormous socioeconomic disruptions.

Phylogenetically, betaCoVs can be further categorized into four (A–D) evolutionary lineages/subgroups.^{1,3} SARS-CoV is a typical lineage B member, while MERS-CoV is grouped in lineage C.²⁰ Despite belonging to different subgroups, these two betaCoVs likely share similar interspecies transmission routes by “jumping” from their natural host(s) to an intermediate adaptive animal(s) and finally to humans.²¹ Current evidence clearly shows that SARS-CoV originated from bats^{9,22,23} and possibly adapted in civets or raccoon dogs²⁴ before it infected humans. Given the close phylogenetic relationship between MERS-CoV and a variety of bat-derived coronaviruses (BatCoV) (e.g., HKU4, HKU5,^{10,25} and those recently identified in the Middle East, Africa, Europe, and Asia^{26–31}), it is widely accepted that the current MERS epidemic represents another bat-to-human transmission event related to a betaCoV, though its intermediate host is shown, this time, to be dromedaries.^{32,33} Notably, two recent studies reported that BatCoV HKU4 could recognize human CD26, the MERS-CoV receptor,³⁴ as a functional entry receptor,^{35,36} indicating its potential adaptation for human infection. These continuously occurring yet unpredictable events of betaCoVs repeatedly crossing species barriers highlight the pressing necessity of studies of other members of the genus for the characteristics relevant to interspecies transmission.²¹

The coronavirus spike (S) protein, which is located on the envelope surface of the virion, functions to mediate receptor recognition and membrane fusion¹ and is therefore a key factor determining the virus tropism for a specific species.^{21,37} In most cases, coronavirus S will be further cleaved into S1 and S2 subunits, and the receptor binding capacity is allocated to the S1 subunit.¹ The receptor binding domain (RBD) of betaCoV that directly engages the receptor is commonly located in the C-terminal half of S1 [C-terminal domain (CTD)] such as in SARS-CoV,³⁸ MERS-CoV,^{39,40} and BatCoV HKU4,³⁵ though in rare cases such as with mouse hepatitis virus (MHV),⁴¹ the RBD region was identified in the S1 N-terminal domain (NTD). We previously characterized structurally the MERS-CoV RBD (MERS-RBD) as a relatively independent entity composed of a core and an external subdomain.³⁹ The latter subdomain, which is topologically an insertion between two scaffold strands of the core subdomain, presents a flat four-stranded β -sheet surface for contacting the CD26 receptor.³⁹ A similar topological arrangement of the core and external subdomains into a structural unit for receptor engagement was also observed in the SARS-CoV RBD (SARS-RBD).³⁸ Nevertheless, the SARS-RBD exhibits a unique loop-dominated external fold to recognize human angiotensin converting enzyme 2 (ACE2)⁴² as a receptor. These observations indicate that the homologous RBD regions of betaCoVs represent a key determinant in receptor adaptation and cross-species transmission.²¹

BatCoV HKU9 is a representative betaCoV of lineage D.¹¹ The virus was first identified in bats in 2007 by next-generation sequencing (NGS).¹¹ Though the isolation of live viruses has been unsuccessful thus far, its genomes are widespread in

different bat species.^{43–46} As its interspecies transmission potential is worrisome, the features of its S protein, especially of the homologous RBD region (HKU9-RBD), remain unknown. This would be an indispensable step in understanding the pathogenesis of BatCoV HKU9. In addition, the atomic structure of HKU9-RBD would provide requisite information for understanding the evolution of betaCoVs. It is notable that MERS-RBD and SARS-RBD share a conserved core structure but differ in the external fold for engaging different receptors.^{21,38,39} Sequence features of betaCoV RBDs clearly indicate that this scheme of subdomain arrangement might be expanded to the whole *Betacoronavirus* genus, regardless of the species. This notion was supported by our recent study of the BatCoV HKU4 RBD (HKU4-RBD) which exhibits a structure that quite resembles that of MERS-RBD.³⁵

In this study, we reported the structural and functional characterization of HKU9-RBD. The determined structure as expected contains a core subdomain homologous to those observed in other betaCoV RBD structures and an external subdomain that is mainly α -helical. This unique structural feature explains its inability to react with either human CD26 or ACE2, which is easily observed in our surface plasmon resonance (SPR) assay. Via comparison of available RBD structures, we further showed that the detailed interactions, anchoring the external subdomain to the core subdomain, share similar patterns in betaCoV RBDs. We believe the observed core/external interacting mode represents another structural feature in the S that is reserved during the evolution of betaCoVs, in addition to the conservation in the fold for the core subdomain. Our study therefore further supports the notion that betaCoV S originates from the same ancestor and divergently evolves mainly in the RBD external region to engage variant receptors, thereby preparing for potential interspecies transmission.

■ MATERIALS AND METHODS

Plasmid Construction. The plasmids used for protein expression were individually constructed by insertion of the coding sequences for HKU9-RBD (S residues S355–N521, GenBank accession number EF065513), MERS-RBD (S residues E367–Y606, GenBank accession number JX869050), SARS-RBD (S residues R306–F527, GenBank accession number NC_004718), human CD26 (residues S39–P766, GenBank accession number NP_001926), and human ACE2 (residues S19–D615, GenBank accession number BAJ21180) into the *EcoRI* and *XhoI* restriction sites of a previously modified pFastBac1 vector⁴⁷ that was engineered to include an N-terminal gp67 signal peptide coding sequence. For each protein, an engineered C-terminal hexahistidine tag was utilized to facilitate protein purification. To prepare mouse IgG Fc fragment (mFc)-fused proteins, the coding sequences of MERS-RBD, SARS-RBD, and HKU9-RBD were fused with the mFc sequence and then introduced into the pCAGGS vector.³⁵

Protein Expression and Purification. The proteins used for crystallization and SPR analysis were prepared with the Bac-to-Bac baculovirus expression system (Invitrogen) according to the manufacturer's instructions.⁴⁸ In brief, the verified pFastBac1 recombinant plasmid was transformed into the DH10Bac competent cells to generate the recombinant bacmid. The bacmid was then extracted and transfected into Sf9 cells to prepare the baculovirus stocks. Sf9 cells were further used to

amplify the baculoviruses, while High5 cells were used to express the protein.

The cell culture of High5 was collected 48 h postinfection. In total, 4 L of cell culture of each protein was collected and centrifuged at 6500 rpm for 1.5 h to remove cell debris. After the samples had been filtered with a 0.22 μm membrane, the supernatant was passed through two 5 mL HisTrap HP columns (GE Healthcare) to capture the individual protein of interest. For MERS-RBD, SARS-RBD, human CD26, and human ACE2, the bound proteins were detached from HisTrap with 20, 50, and 300 mM imidazole in 20 mM Tris-HCl and 150 mM NaCl buffer (pH 8.0). After sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) determination, fractions detached with 300 mM imidazole were pooled and further purified with a Superdex 200 column (GE Healthcare). For HKU9-RBD, the bound proteins were detached from HisTrap with 20, 50, and 300 mM imidazole in 20 mM HEPES and 150 mM NaCl buffer (pH 7.0). Fractions detached with 50 and 300 mM imidazole were pooled and dialyzed overnight against 5 L of 20 mM HEPES and 150 mM NaCl buffer (pH 7.0) to remove imidazole. The dialysates were concentrated and further purified with a Superdex 200 column (GE Healthcare). Each protein was stored in the buffer that was used for purification.

To prepare mFc-fused proteins with the mammalian cell expression system, the recombinant pCAGGS plasmids were confirmed with Sanger sequencing and then prepared with the EndoFree Maxi Plasmid Kit (Tiangen, Beijing, China). Each recombinant plasmid was transfected into 293T cells with 50 μg of plasmid DNA per T75 plate using polyethylimine (PEI, Polysciences Inc.). After being incubated for 5 h, the transfected cells were washed with PBS twice and then replaced with DMEM without serum. The cells were maintained for 3 days, and the supernatant was harvested and replaced with fresh DMEM medium and then maintained for an additional 4 days. The harvested supernatants were pooled and concentrated and then mixed with 2 volumes of 20 mM trisodium phosphate (pH 7.0). The mixture was passed through a 5 mL HiTrap Protein A HP prepacked column (GE Healthcare) to capture the individual protein of interest. After removal of impure proteins with 20 mM trisodium phosphate (pH 7.0), the bound protein was detached from the column with 100 mM glycine (pH 3.0). Each fraction was neutralized with 1 M Tris-HCl (pH 9.0). After SDS–PAGE determination, the detached fractions with the protein of interest were pooled and concentrated. The buffer of each protein was then changed to PBS (pH 7.0) for further experiments.

SPR Assay. The BIAcore experiments were performed at 25 $^{\circ}\text{C}$ using a BIAcore 3000 or BIAcore T100 machine with CM5 chips (GE Healthcare). For all the measurements, an HBS-EP buffer consisting of 10 mM HEPES (pH 7.4), 150 mM NaCl, and 0.005% (v/v) Tween 20 was used, and all proteins were exchanged into this buffer in advance. First, the HKU9-RBD, MERS-RBD, and SARS-RBD proteins expressed in insect cells were used for the SPR assay using a BIAcore 3000 machine. BSA (negative control), HKU9-RBD, MERS-RBD, and SARS-RBD proteins were immobilized on the chip at \sim 1000 response units (RU), according to the manufacturer's amine coupling chemistry protocol (GE Healthcare). Gradient concentrations of human CD26 (0, 19.5 to 5000 nM) or human ACE2 (0, 39 to 625 nM) were then passed over the chip surface. After each cycle, the sensor surface was regenerated via a short treatment with 10 mM NaOH. The equilibrium dissociation constants

(binding affinity and K_D values) were analyzed using BIA evaluation (BIAcore software). To exclude the possibility that HKU9-RBD could be nonfunctional because of immobilization or could be missing some important post-translational modifications on the protein, we purified the mFc-fused HKU9-RBD proteins in mammalian cells and assessed the abilities to bind CD26 or ACE2 proteins using a captured SPR method by a BIAcore T100 system. The CM5 chip was immobilized with the anti-mouse antibody for flow cells 1 and 2 (FC1 and FC2, respectively). The mFc-fused RBD proteins were then injected and captured on FC2, while FC1 was used as a negative control. Human CD26 or human ACE2 proteins were then injected, and the binding responses were measured. The immobilized anti-mouse antibody was regenerated with 10 mM glycine (pH 1.7) (GE Healthcare).

Crystallization. The crystallization trials were performed with 1 μL of protein being mixed with 1 μL of the reservoir solution and then equilibrating against 100 μL of the reservoir solution at 4 $^{\circ}\text{C}$ by the vapor-diffusion sitting-drop method. The initial crystallization was screened using the commercially available kits. Diffractive crystals of the HKU9-RBD protein were finally obtained under 0.1 M sodium citrate tribasic dihydrate (pH 7.0) and 12% PEG 20000 with a protein concentration of 2.2 mg/mL. Derivative crystals were obtained by soaking the crystals in the reservoir solution containing 1 mM $\text{KAuBr}_4 \cdot 2\text{H}_2\text{O}$ for 48 h at 4 $^{\circ}\text{C}$.

Data Collection, Integration, and Structure Determination. For data collection, all crystals were flash-cooled in liquid nitrogen after a brief soaking in the reservoir solution with the addition of 20% (v/v) glycerol. The diffraction data for the native (wavelength of 1.03906 \AA) and Au derivative crystals (wavelength of 1.03906 \AA) of HKU9-RBD were collected at Shanghai Synchrotron Radiation Facility (SSRF) BL17U. All data were processed with HKL2000.⁴⁹ The ice rings that form in the crystal flash cooling process were excluded from data processing, and the final overall completeness for the data set is 97.1%.

The structure of HKU9-RBD was determined by the SAD method. After location of Au sites by SHELXD⁵⁰ with the Au-SAD data, the identified positions were then refined and the phases were calculated with the SAD experimental phasing module of PHASER.⁵¹ The real space constraints were further applied to the electron density map in DM.⁵² The initial model was built with Autobuild in the PHENIX package.⁵³ Additional missing residues were added manually in COOT.⁵⁴ The final model was refined with phenix.refine in PHENIX⁵³ with energy minimization, isotropic ADP refinement, and bulk solvent modeling. The stereochemical qualities of the final model were assessed with MolProbity.⁵⁵ The Ramachandran plot distributions for the residues in the HKU9-RBD structure were 94.64, 5.36, and 0% for favored, allowed, and outlier regions, respectively. Data collection and refinement statistics are summarized in Table 1. All structural figures were generated using PyMol (<http://www.pymol.org>).

RESULTS

HKU9-RBD Does Not Bind the SARS-CoV or MERS-CoV Receptor. We first characterized the sequence of BatCoV HKU9 S by using a series of bioinformatic methods. This 1274-residue protein exhibits typical features of coronavirus S proteins (e.g., the presence of characteristic heptad repeats 1 and 2 in the S2 subunit), though the S1/S2 cleavage site potentially processed by furin-like proteases was not detected

Table 1. Data Collection and Refinement Statistics

	HKU9-RBD (Protein Data Bank entry 5GYQ)	Au derivative HKU9-RBD
Data Collection		
space group	P21	P1
wavelength (Å)	1.03906	1.03906
unit cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	42.7, 36.0, 62.9	36.0, 46.6, 57.3
α , β , γ (deg)	90.0, 102.7, 90.0	80.4, 88.8, 88.5
resolution ^a (Å)	50.00–2.10 (2.18–2.10)	50.00–2.48 (2.57–2.48)
no. of observed reflections	101588	52401
completeness (%)	97.1 (80.7)	97.7 (96.9)
redundancy	9.4 (9.4)	4.1 (3.7)
R_{merge}^b (%)	6.1 (15.4)	9.2 (39.0)
$I/\sigma I$	34.023 (12.057)	16.960 (4.637)
$CC_{1/2}$	0.998 (0.988)	0.986 (0.915)
Refinement		
resolution (Å)	41.7–2.10	
no. of reflections	10811	
completeness for range (%)	97.0	
$R_{\text{work}}/R_{\text{free}}^c$	0.1700/0.2006	
no. of atoms		
protein	1367	
water	128	
<i>B</i> factor (Å ²)		
protein	28.7	
water	34.1	
root-mean-square deviation		
bond lengths (Å)	0.003	
bond angles (deg)	0.820	
Ramachandran plot ^d (%)		
favored	94.64	
allowed	5.36	
outliers	0.00	

^aValues for the outermost resolution shell are given in parentheses.

^b $R_{\text{merge}} = \sum_i \sum_{hkl} |I_i - \langle I \rangle| / \sum_i \sum_{hkl} I_i$, where I_i is the observed intensity and $\langle I \rangle$ is the average intensity from multiple measurements. ^c $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the structure factor amplitudes from the data and the model, respectively. R_{free} is the *R* factor for a subset (5%) of reflections that were selected prior to refinement calculations and were not included in the refinement.

^dRamachandran plots were generated by using MolProbity.

(Figure 1A). Along the full-length protein, the amino acid sequence identity between BatCoV HKU9 S and other betaCoVSSs is rather limited (e.g., 27.9% identical to MERS-CoV S, 28.0% identical to HKU4 S, and 30.4% identical to SARS-CoV S). Nevertheless, we were able to identify the RBD region based on the characteristic cysteine residues of the core subdomain (Figure 1B), which were shown, in the thus-far available RBD structures,^{35,38,39,56} to form three conserved disulfide bonds stabilizing the core fold. The subsequent HKU9-RBD was allocated to the S region spanning residues

355–521 (Figure 1A). In comparison to other RBD sequences, the HKU9-RBD exhibits a comparable length in the core subdomain (Figure 1B) but is dramatically shortened in the external region (Figure 1C).

To test if HKU9-RBD could react with either SARS-CoV receptor ACE2⁴² or MERS-CoV receptor CD26,³⁴ the RBD and the receptor–ectodomain proteins were individually prepared in insect cells and purified to homogeneity. The ligand–receptor interaction was then characterized via SPR BIAcore by passing ACE2 or CD26 over the immobilized RBD proteins. As expected, potent interactions were observed for both the SARS-RBD–ACE2 ($K_D = 0.265 \mu\text{M}$) (Figure 2A) and MERS-RBD–CD26 ($K_D = 52.8 \text{ nM}$) (Figure 2B) binding pairs. The revealed kinetics were very similar to those reported previously,^{35,39} validating the integrity of our testing system. Under the same condition, however, neither ACE2 (Figure 2C) nor CD26 (Figure 2D) interacted with HKU9-RBD. To exclude the possibility that HKU9-RBD could be nonfunctional because of immobilization or because of the absence of some important post-translational modifications on the protein, we purified the mFc-fused HKU9-RBD proteins in mammalian (293T) cells and assessed the abilities to bind CD26 or ACE2 proteins using a captured SPR method. In the same way, there was no detectable binding of mFc-fused HKU9-RBD to ACE2 or CD26 (Figure 2G), while the mFc-fused SARS-RBD protein bound to ACE2 (Figure 2E) and the mFc-fused MERS-RBD bound to CD26 (Figure 2F) well. BatCoV HKU9, therefore, could utilize neither the SARS-CoV receptor nor the MERS-CoV receptor for cell entry. Rather, it must utilize a unique cellular receptor for entry.

Crystal Structure of HKU9-RBD. We further set out to investigate the structural features of HKU9-RBD via crystallography. The protein was successfully crystallized; a 2.1 Å data set was collected (Table 1), and the structure was determined by using the single-wavelength anomalous diffraction (SAD) method. The determined structure, with an R_{work} of 0.1700 and an R_{free} of 0.2006, contains a single molecule in the crystallographic asymmetric unit. Clear electron densities were traced for 176 consecutive HKU9-RBD residues, extending from S355 to A520. These amino acids fold into a compact structure, which can be further divided into two subdomains as shown schematically in the other RBD structures.^{35,38,39} The core subdomain comprises eight β -strands and six helices (α or 3_{10}). Five long strands ($\beta c1$ – $\beta c5$) are arranged in an antiparallel manner, forming the scaffold center of the core (core-center). This core-center sheet is further wrapped by the surface helices and loops. It is notable that the six helices (H1–H6) are sporadically distributed on the two sheet faces, thereby leading to an overall globular fold for the core subdomain. On one lateral side of the core-center sheet, the external subdomain covers the core like a hat, while on the distal opposite side, three small strands ($\beta p1$ – $\beta p3$) constitute a parallel peripheral sheet (core-peripheral), ensuring that the N- and C-termini of HKU9-RBD are in the proximity. As expected, the characteristic cysteine residues (Figure 1B) form three disulfide bonds in the core subdomain, further stabilizing the core structure from the interior. Of these, two (C357–C381 and C411–C517) are located in core-peripheral, contributing to the orientation of the RBD termini; one (C399/C452) resides in the core-center, linking strands $\beta c2$ and $\beta c4$ (Figure 3). Overall, the residue boundaries of the core subdomain observed in the structure are quite consistent with those deduced from the results of sequence alignment (Figure 1B).

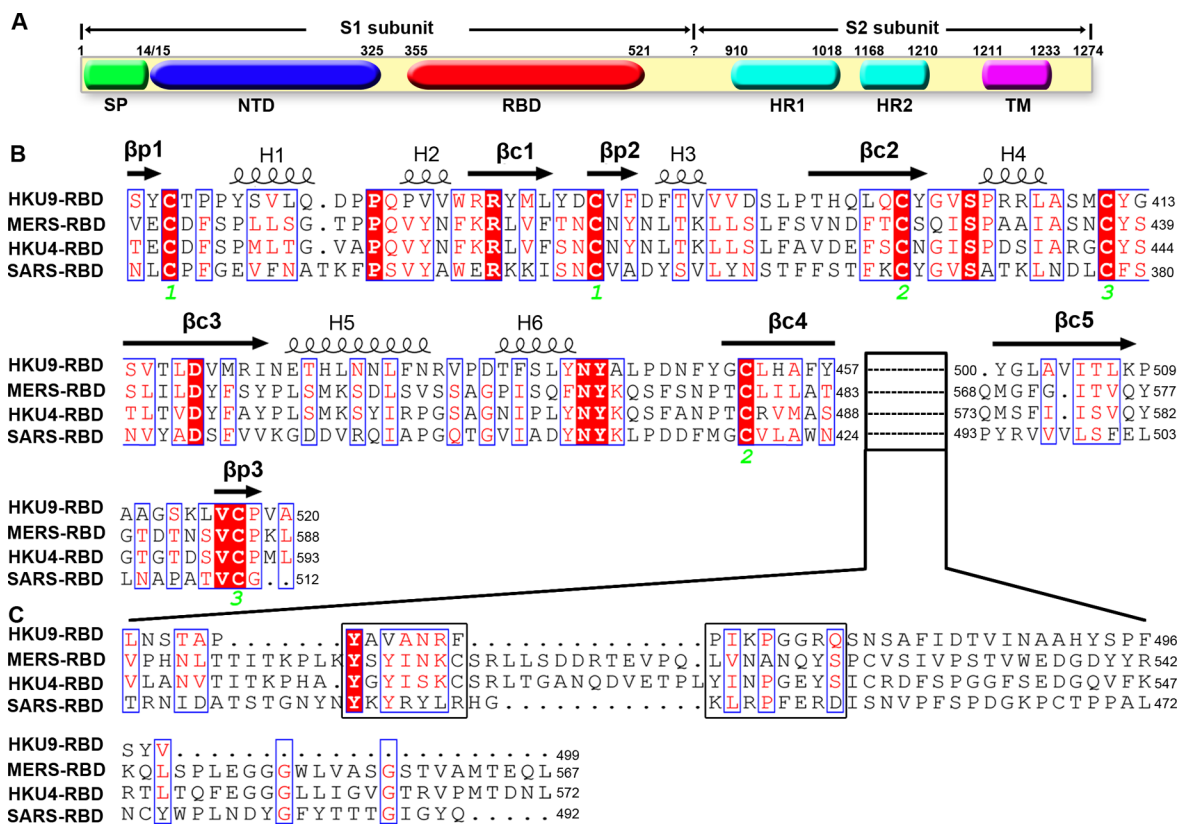


Figure 1. Sequence features of HKU9-RBD. (A) Schematic representation of BatCoV HKU9 S. The indicated domain elements were defined on the basis of either the pairwise sequence alignment results or the bioinformatics predictions. The signal peptide (SP), transmembrane domain (TM), and heptad repeats 1 and 2 (HR1 and HR2, respectively) were predicted by the SignalP 4.0 server, TMHMM server, and Learncoil-VMF program, respectively, while the N-terminal domain (NTD) and RBD were deduced by alignment with the N-terminal galectin-like domain of murine hepatitis virus S and MERS-RBD, respectively. The S1/S2 site potentially cleaved by furin-like proteases could not be ascertained and is therefore labeled with a question mark. (B and C) Structure-based alignment of the HKU9-, SARS-, MERS-, and HKU4-RBD sequences. The arrows and spiral lines indicate strands and helices, respectively. These secondary structure elements were labeled as illustrated in Figure 3. The conserved cysteine residues that form three disulfide bonds in the structures are marked with Arabic numerals 1–3. The core subdomain is conserved among the four RBD structures, but the external subdomain is structurally irrelevant. We therefore present the sequences separately. The two elements that anchor the external subdomain to the core subdomain are highlighted with black boxes. (B) Core subdomain sequence. (C) External subdomain sequence.

The external subdomain of HKU9-RBD consists of 42 residues from L458 to V499 (Figure 1C). These amino acids extend out of strand β 4 of the core-center sheet, first orient as a loop along the core subdomain like a clamp, then fold back to form a solvent-exposed α -helix (H1'), and finally proceed into core strand β 5 (Figure 3). This observed structure differs dramatically from those of SARS-RBD and MERS-RBD, which are shown to be devoid of any helical components in the external region.^{38,39} The unique external fold of HKU9-RBD could well explain its inability to bind either ACE2 or CD26.

Structural Conservation of the RBD Core Subdomain in BetaCoVs. Previously, three betaCoV RBD structures have been reported, including one lineage B structure (SARS-RBD³⁸) and two lineage C structures (MERS-RBD³⁹ and HKU4-RBD³⁵). These structures indicated an interspecies conservation in the core fold among betaCoVs.³⁹ BatCoV HKU9 is a representative member of betaCoV lineage D.¹¹ We therefore compared the currently available RBD structures with the HKU9-RBD structure determined in this study. As expected, a significant similarity was observed in the core subdomain (Figure 4A–D). Superimposition of the core structures revealed the root-mean-square deviation (rmsd) values ranging from 0.66 to 2.82 Å (Table 2), demonstrating the quite similar core folds (though with a low level of sequence

identity) among the four RBDs. The most conserved part was seen in the core-center sheet. This five-stranded scaffold element as well as the single interstrand disulfide bond is invariably reserved in all the structures. In the core-peripheral region, however, a small variance in strand composition is noted. In HKU9-RBD, it contains three short β -strands, arranged in a small parallel β -sheet. Both SARS-RBD and MERS-RBD retain two of these strands, whereas HKU4-RBD is devoid of any detectable strand elements in this region. Despite the observed difference in strand formula, the core-peripherals of the RBDs exhibit a similar orientation and present the same scheme in which the domain N-terminus is in the proximity of its C-terminus. Extra common features in core-peripheral strands lie in the two disulfide bonds in the region, which are structurally and topologically conserved in the four structures (Figure 4A–D).

In contrast to the core conservation, the external subdomains of the four RBDs are divergent in structures. HKU9-RBD presents a single H1' helix in the external region, whereas SARS-RBD is loop-dominated but contains two extra small β -strands. The external subdomains of MERS-RBD and HKU4-RBD, however, resemble each other and are predominantly a rigid β -sheet composed of four β -strands. Despite the structural irrelevance, the external subdomains are clearly topological

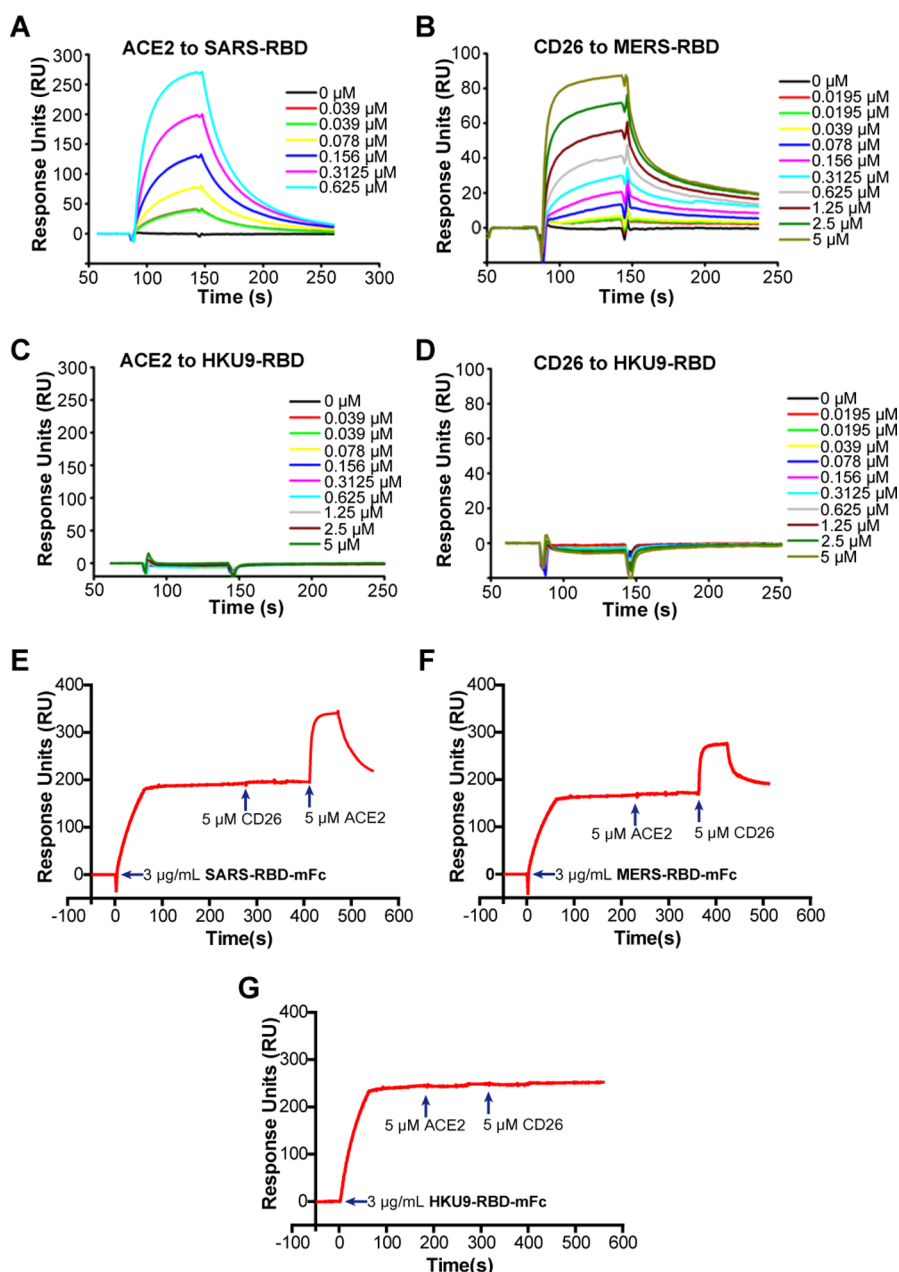


Figure 2. Characterization of HKU9-RBD by SPR assays. The indicated RBD proteins expressed by insect cells were immobilized on CMS chips and tested for the binding with gradient concentrations of human ACE2 or CD26 using a BIAcore 3000 machine. The recorded kinetic profiles are shown: (A) human ACE2 and SARS-RBD, (B) human ACE2 and HKU9-RBD, (C) human ACE2 and HKU9-RBD, and (D) human CD26 and HKU9-RBD. Clearly shown is the fact that HKU9-RBD does not bind either ACE2 or CD26, in the context of which SARS-RBD and MERS-RBD bind their respective receptors. Then we purified the mFc-fused HKU9-RBD proteins in mammalian (293T) cells and assembled the abilities to bind CD26 or ACE2 proteins using a captured SPR method by a BIAcore T100 system. The anti-mouse antibodies were immobilized on CMS chips. The mFc-fused RBD proteins were then captured ($3 \mu\text{g}/\text{mL}$ for 60 s) by the antibodies and tested for binding to human ACE2 or CD26. (E) The mFc-fused SARS-RBD (SARS-RBD-mFc) did not bind to CD26 but bound to ACE2 well. (F) The mFc-fused MERS-RBD (MERS-RBD-mFc) did not bind to ACE2 but bound to CD26 well. (G) The mFc-fused HKU9-RBD (HKU9-RBD-mFc) does not bind either ACE2 or CD26.

equivalents in these structures, being present as an insertion between two core-center strands (Figure 4A–D).

Homologous Interaction Mode Anchoring the External Subdomain to the Core Subdomain. By superimposing the available RBD structures, we unexpectedly identified two major elements in the external subdomain that could be well-aligned (Figure 5A). The first element (element 1) spans approximately seven residues (Y464–F470 in HKU9-RBD, Y438–R444 in SARS-RBD, Y497–C503 in MERS-RBD, and Y501–C507 in HKU4-RBD) (Figures 1C and 5C) and

proceeds along the core subdomain surface to be lodged between helices H2 and H6 (based on the secondary element definition of HKU9-RBD) (Figure 5A). The second element (element 2) contains eight amino acids (P471–Q478 in HKU9-RBD, K447–D454 in SARS-RBD, L517–S524 in MERS-RBD, and Y522–S529 in HKU4-RBD) (Figures 1C and 5C), extending as a curved loop covering helix H6 of the core subdomain (Figure 5A). It is interesting that these two elements are “saddled” upon the core helices, anchoring the external subdomain to the core subdomain (Figure 5A). We

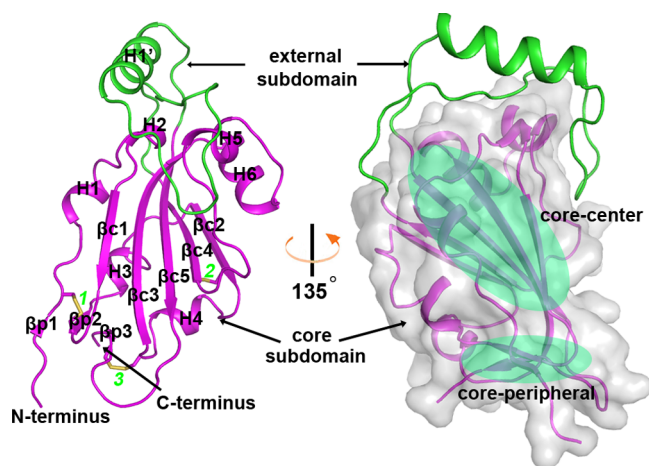


Figure 3. Crystal structure of HKU9-RBD. The core and external subdomains are colored magenta and green, respectively. The core subdomain is further divided into a center region (core-center) and a peripheral region (core-peripheral), which are encircled. The core-center strands and helices are labeled $\beta c1$ – $\beta c5$ and H1–H6, respectively, while the core-peripheral strands are marked $\beta p1$ – $\beta p3$. The disulfide bonds and the RBD termini are labeled. The core subdomain is further presented in a surface representation in the right panel to highlight the top positioning of the external subdomain like a hat.

therefore further explored the amino acid interaction details at this core–external interface.

Each element residue was scrutinized for both the side-chain orientation and the intersubdomain interactions. To facilitate the analyses and comparison, the two elements were assigned a position marker for each amino acid (a–g for element 1 and a–h for element 2) (Figure 5B,C). In element 1, residue a is invariably a tyrosine in the four RBDs. This amino acid orients its bulky side chain toward the core subdomain, providing strong hydrophobic contacts. An extra side-chain H-bond is also observed at this position in HKU9-RBD. Residue b extended away from the core surface and exhibited little conservation. The residue, however, invariably contributes to the subdomain anchoring by providing a main-chain H-bond. Following residue b, the amino acids are preferably facing toward the core at position c and spreading parallel to the core surface at position d. Multiple van der Waals (vdw) contacts and conserved main-chain H-bonds are observed at these two positions, respectively. A small discrepancy is seen in SARS-RBD, which orients its c residue outward for the bulky solvent region. The remaining three element 1 amino acids at positions e–g are distant from the core subdomain and completely solvent-exposed, therefore contributing little to the core–external interactions (Figure 5B,C).

In element 2, both residues a and d are oriented parallel to the surface of the core subdomain. The configuration allows the amino acids to provide apolar vdw contacts to strengthen core–external subdomain binding. A certain extent of amino acid conservation was observed at position d where a proline is favored to facilitate the turning of the loop. Following these two positions, residues b and e insert their side chains into two surface pockets of the core subdomain. At position b, the residue is conservatively hydrophobic and has a middle-sized side chain (Val/Ile/Leu). It is accommodated in a shallow apolar pocket, creating strong stacking forces bonding the core and external subdomains. In addition, the residue also

contributes a main-chain H-bond to the subdomain binding. For residue e, its accommodating pocket is deep and large, therefore allowing for amino acid variance at the position (Gly in HKU9- and HKU4-RBD, Phe in SARS-RBD, and Asn in MERS-RBD). In the four RBDs, this residue e invariably forms H-bonds with the core subdomain residue via the main-chain atom but may also provide side-chain H-bond interactions (e.g., in MERS-RBD) or multiple vdw contacts (e.g., in SARS-RBD). Extra core–external interactions in this region were further observed at position g, where the residue is oriented parallel to or toward the core subdomain and thereby contributes to the binding via hydrophobic and side-chain H-bond interactions. It is also of interest that these important interface residues of element 2 are regularly interspersed by amino acids at positions c, f, and h, which are solvent-exposed and rarely interact with the core subdomain (Figure 5B,C).

In summary, the four coronavirus RBD structures show homologous amino acid interaction patterns for intersubdomain binding. The binding relies mainly on two elements in the external subdomain, which are oriented similarly in these structures (Figure 5A). Despite the lower level of conservation in the element sequences, the side-chain orientation and the interaction modes (hydrophobic, vdw, or H-bond contacts) at each position are, in most cases, similar or homologous (Figure 5C).

DISCUSSION

Bats have been found to harbor the largest natural genetic pools for new coronaviruses or coronavirus genes. The origin of a majority of the betaCoVs could be traced back to bats;⁵⁷ e.g., a recent study isolated, in Chinese horseshoe bats, a live SARS-like coronavirus that can utilize the SARS-CoV receptor of ACE2 for cell entry,²² thereby providing the strongest evidence of the bat origin of this pandemic human pathogen. In addition, two studies reported the identification of gene fragments in bats that are almost identical to those of MERS-CoV,^{26,29} indicating that MERS-CoV likely also originates from bats. Noting the recent reports showing the adaptation of batCoV HKU4 for binding to human cells by recognizing CD26,³⁵ we believe that preparing for the unforeseeable events of potential interspecies transmission by other bat-derived betaCoVs is an urgent need. BatCoV HKU9 is an important lineage D betaCoV¹¹ and has been demonstrated to be widespread and circulating in different bat species.^{43–46} Noting the determinative role of the coronavirus S RBD in the process of crossing species barriers (as has been structurally illustrated in other coronaviruses^{35,38,39}), we characterized the structural and functional features of the homologous RBD protein of batCoV HKU9 S. The determined structure revealed a core subdomain that resembles those observed in SARS-, MERS-, and HKU4-RBD but a unique external subdomain that is composed of a single helix. Because the RBD external subdomain contains the key motifs [denoted the receptor binding motif (RBM)], interfacing with the receptor,^{35,38,39} the unique external fold of HKU9-RBD therefore is in accord with our functional data showing its inability to react with either ACE2 or CD26. Which host molecule could be recognized by HKU9-RBD as a functional cell entry receptor remains to be investigated. Nevertheless, taking into account the single helical component in the external subdomain of HKU9-RBD, we expect the RBM to be located on the solvent-access side of the helix, which might facilitate future attempts to identify the receptor.

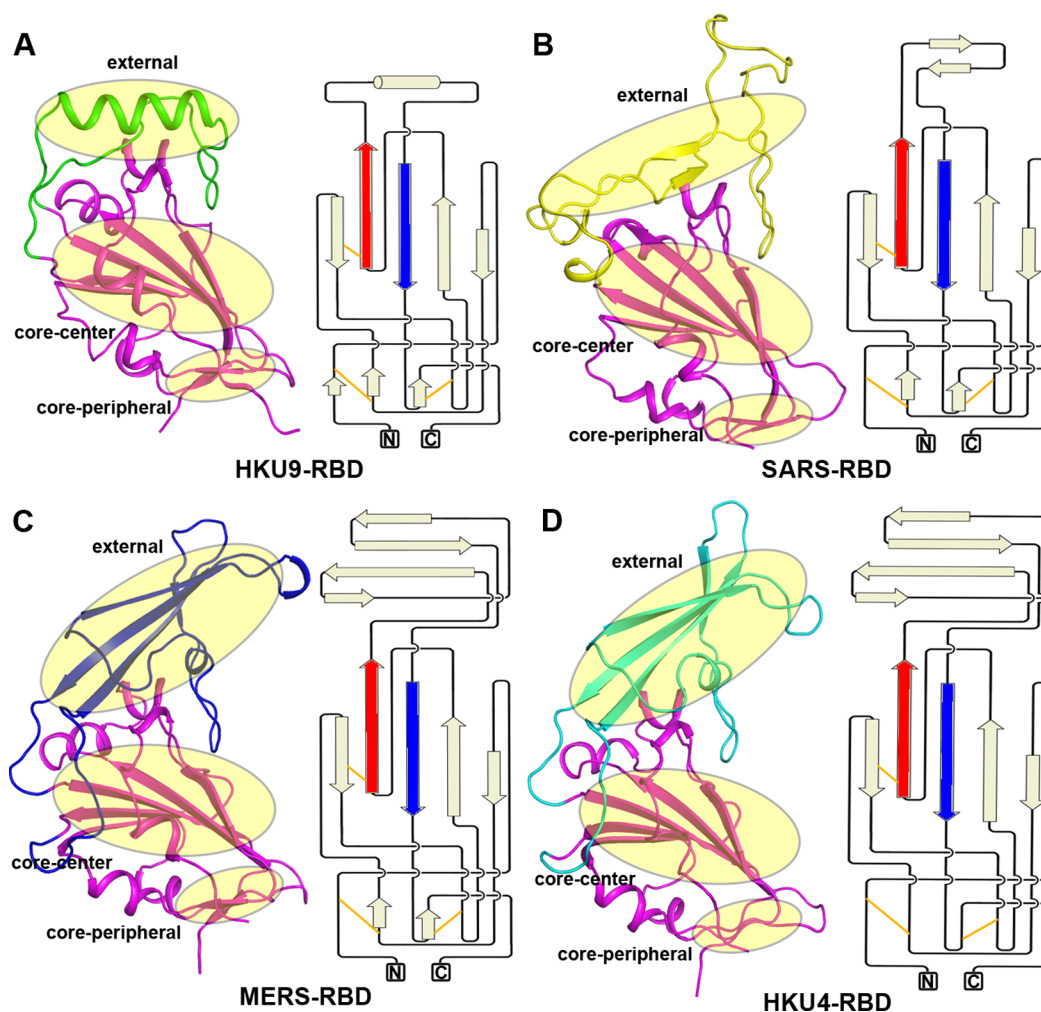


Figure 4. Structural and topological comparison of available betaCoV RBD structures. Four structures, including those of HKU9-, SARS-, MERS-, and HKU4-RBD, were oriented similarly and are presented as cartoons in parallel. The core-center, core-peripheral, and the external subdomain are encircled and highlighted in yellow. For each structure, the topological arrangement of the core-center and core-peripheral strands as well as of the external components is depicted. The core strands that flank the external subdomain are colored red and blue, respectively. Yellow lines indicate disulfide bonds. The N- and C-termini are highlighted: (A) HKU9-RBD, (B) SARS-RBD, (C) MERS-RBD, and (D) HKU4-RBD. The similarity in the topological arrangement of the external subdomain as an insertion between two core strands is illustrated.

Table 2. Statistics of the Core Subdomain Deviations among Available BetaCoV RBD Structures^a

	HKU9-RBD	MERS-RBD	SARS-RBD	HKU4-RBD
HKU9-RBD	–	2.07 Å (104 Cα atoms)	1.92 Å (100 Cα atoms)	1.37 Å (94 Cα atoms)
MERS-RBD		–	2.82 Å (75 Cα atoms)	0.66 Å (109 Cα atoms)
SARS-RBD			–	1.95 Å (82 Cα atoms)
HKU4-RBD				–

^aThe RBD core subdomain structures were superimposed onto each other by PyMol in a pairwise manner to calculate the rmsd values, which are listed in the table. The values in parentheses indicate the number of equivalent Cα atoms that were selected for rmsd calculations.

It should be noted that coronavirus RBDs are not necessarily located in the C-terminal half of the S1 subunit. Previous structural and mutagenesis data showed that the RBD of MHV S is located in the S1 N-terminal half (NTD).^{41,58} Nevertheless,

the current data seem to favor the notion that the CTD is prioritized over the NTD to function as the receptor binding entity, as the majority of the coronaviruses (e.g., SARS-CoV,³⁸ MERS-CoV,³⁹ batCoV HKU4,³⁵ human coronavirus NL63,⁵⁹ transmissible gastroenteritis virus,⁶⁰ etc.) harbor a CTD as the RBD. It is interesting that the current available betaCoV CTD/RBD structures^{35,38,39} all keep the N- and C-termini on the same side opposite from the location of the external subdomain. This arrangement mode would lead the S1 N-terminal half to being sterically underneath the C-terminal half, thereby projecting the CTD distant from the viral envelope for a trans-interaction with the receptors. Our structural study demonstrated that HKU9-RBD retains the same character and therefore stands a better chance of being the authentic receptor binding entity.

In comparison to SARS-, MERS-, and HKU4-RBD whose structures are available,^{35,38,39} HKU9-RBD differs in the external fold but reserves a resembled core subdomain structure. The most conserved part lies in the core-center sheet that is composed of five antiparallel strands and functions as the scaffold of the core subdomain. The sheet is sterically

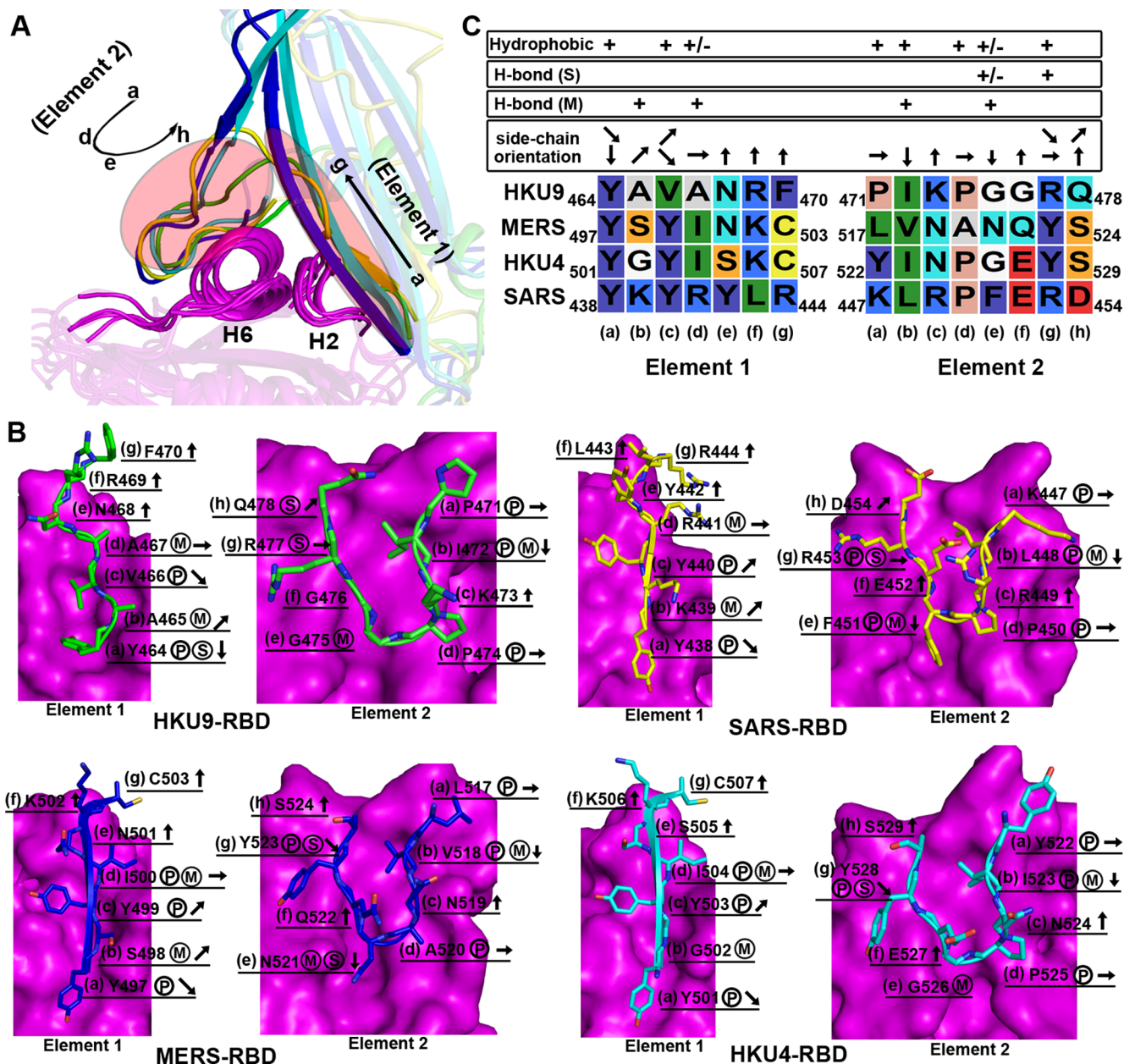


Figure 5. Homologous intersubdomain amino acid interactions anchoring the external subdomain to the core subdomain. (A) Superimposition of the betaCoV RBD (HKU9-RBD in green, SARS-RBD in yellow, MERS-RBD in blue, and HKU4-RBD in cyan) structures highlighting the external elements that can be well-aligned. These two elements, with seven (element 1) and eight (element 2) amino acids, respectively, engage mainly core subdomain helices H2 and H6 for the intersubdomain interactions. To facilitate comparison, the element residues were successively assigned a position marker (a–g for element 1 and a–h for element 2), which is highlighted. (B) Characterization of the element residues for their contributions to the intersubdomain binding. The two external elements are presented as cartoons, while the core subdomain is shown at the surface. At each position, the residue is marked sequentially with the position marker, the amino acid identity and numbering, the interacting mode/type, and the side-chain orientation. For the interaction mode, the hydrophobic or van der Waals interactions are indicated with encircled Ps, the side-chain H-bonds with encircled Ss, and main-chain H-bonds with encircled Ms. The side-chain orientations are indicated with arrows. (C) Summary of the intersubdomain interactions specified in panel B. The element sequences of the four RBDs are aligned and listed. A + indicates that a certain type of interaction is commonly observed at the position, while a +/- indicates that the interaction type is specific to some but not all of the four RBDs. The arrows mark the side-chain orientations.

and structurally conserved in all the RBD structures. Additional conserved elements include the core-center helices and core-peripheral structures. Nevertheless, these elements could vary in their secondary element compositions. For example, a recent study of the structure of HKU4-RBD³⁵ showed that its N-terminal-most part does not fold into a characteristic helix as observed in the structures of SARS-RBD³⁸ and MERS-RBD.³⁹ For core-peripheral, the number of strands was found to vary

from zero (as in HKU4-RBD) to three (as in HKU9-RBD) (Figure 4). This has added dramatic complexities to the nomenclature of the RBD secondary elements. The situation would be even worse were the external subdomains that could vary significantly in structure taken into account. We therefore suggest that the core-center strands and helices be designated as β cs (β c1– β c5) and Hs (H1, H2, etc.), respectively, and that the core-peripheral strands be designated as β ps (β p1, β p2,

etc.) and the external elements as H's (H1', H2', etc.) or β 's ($\beta 1'$, $\beta 2'$, etc.). This terminological strategy should be able to facilitate the comparison of homologous RBD structures and to reflect the fact that the external subdomain is topologically an insertion between two equivalent core-center strands.

The long evolutionary history, high mutation rates, and many genetic artifices of RNA viruses often lead to conundrums in the study of the origin of viruses. It would be even more difficult to track the evolutionary traces in the viral surface proteins that are normally under great evolutionary pressure. The evolutionary records, however, are more likely to be conserved in the tertiary structures than in the amino acid sequences. In betaCoVs, the interlineage sequence identity in the S RBD is rather limited. Nevertheless, we observed several conserved features in the betaCoV RBD structures. These include (1) a conserved core-center as the scaffold of the core subdomain, (2) a similar core-peripheral where the RBD termini are clinched in the proximity, (3) a similar topological arrangement of the external subdomain as an insertion between two core strands, and (4) a homologous intersubdomain binding mode anchoring the external subdomain to the core subdomain. The features indicate a common ancestor S protein that divergently evolves into different species. During evolution, the core subdomain is structurally reserved, whereas the external subdomain folds into variant structures to engage different receptors. It is also noteworthy that the aforementioned core features have been structurally validated in betaCoV lineages B (SARS-RBD), C (MERS- and HKU4-RBD), and D (HKU9-RBD) but not yet in lineage A. Structural studies of the equivalent S RBDs of the lineage A members should be conducted in the future.

AUTHOR INFORMATION

Corresponding Author

*Telephone: 86-10-64807688. Fax: 86-10-64807882. E-mail: gaof@im.ac.cn or gaofu@chinacdc.cn.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC, Grants 81461168030 and 81290342), the China National Grand S&T Special Project (Grant 2014ZX10004-001-006), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant XDB08020100). G.L. is supported in part by NSFC (Grants 31400154, 81522026, and 31570157) and the Sichuan Outstanding Youth Science & Technology Funding (Grant 2016JQ0001). G.F.G. is a leading principle investigator of the NSFC Innovative Research Group (Grant 81321063).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Assistance in diffraction data collection by the staff at the Shanghai Synchrotron Radiation Facility (SSRF beamline 17U) is acknowledged. We thank Dr. Zheng Fan (Institute of Microbiology, Chinese Academy of Sciences) and Yuanyuan Chen and Zhenwei Yang (Institute of Biophysics, Chinese Academy of Sciences) for their sophisticated technical support in the SPR assay.

REFERENCES

- (1) Fields, B. N., Knipe, D. M., and Howley, P. M. (2007) *Fields virology*, 5th ed., Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.
- (2) Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C., Zhou, J., Liu, W., Bi, Y., and Gao, G. F. (2016) Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24, 490–502.
- (3) International Committee on Taxonomy of Viruses and King, A. M. Q. (2012) *Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses*, Academic Press, London.
- (4) Schalk, A. F., and Hawn, M. C. (1931) An apparently new respiratory disease in baby chicks. *J. Am. Vet. Med. Assoc.* 78, 413–422.
- (5) Woo, P. C., Lau, S. K., Lam, C. S., Lai, K. K., Huang, Y., Lee, P., Luk, G. S., Dyrting, K. C., Chan, K. H., and Yuen, K. Y. (2009) Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J. Virol.* 83, 908–917.
- (6) 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., and 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.
- (7) Weinstein, R. A. (2004) Planning for epidemics—the lessons of SARS. *N. Engl. J. Med.* 350, 2332–2334.
- (8) World Health Organization (2014) Global Alert and Response (GAR): MERS-CoV summary updates (http://www.who.int/csr/disease/coronavirus_infections/archive_updates/en/).
- (9) Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Cramer, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S., and Wang, L. F. (2005) Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310, 676–679.
- (10) Woo, P. C., Lau, S. K., Li, K. S., Poon, R. W., Wong, B. H., Tsoi, H. W., Yip, B. C., Huang, Y., Chan, K. H., and Yuen, K. Y. (2006) Molecular diversity of coronaviruses in bats. *Virology* 351, 180–187.
- (11) 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., and 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, 1574–1585.
- (12) Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G., Peiris, M., Lim, W., Stohr, K., and Osterhaus, A. D. (2003) Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* 423, 240.
- (13) Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J., Anderson, L. J., and SARS Working Group (2003) A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1953–1966.
- (14) Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguier, A. M., Cinatl, J., Eickmann, M., Escriviou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H., and Doerr, H. W. (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, 1967–1976.
- (15) Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A. D., Drosten, C., Pallansch, M. A., Anderson, L. J., and

Bellini, W. J. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300, 1394–1399.

(16) 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., and Ziebuhr, J. (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J. Virol.* 87, 7790–7792.

(17) Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820.

(18) Bermingham, A., Chand, M. A., Brown, C. S., Aarons, E., Tong, C., Langrish, C., Hoschler, K., Brown, K., Galiano, M., Myers, R., Pebody, R. G., Green, H. K., Boddington, N. L., Gopal, R., Price, N., Newsholme, W., Drosten, C., Fouchier, R. A., and Zambon, M. (2012) Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill* 17, 20290.

(19) 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., and Yuen, K. Y. (2005) Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* 79, 884–895.

(20) Lu, G., and Liu, D. (2012) SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases. *Protein Cell* 3, 803–805.

(21) Lu, G., Wang, Q., and Gao, G. F. (2015) Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* 23, 468–478.

(22) Ge, X. Y., Li, J. L., Yang, X. L., Chmura, A. 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., Zhu, Y., Cramer, G., Zhang, S. Y., Wang, L. F., Daszak, P., and Shi, Z. L. (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503, 535–538.

(23) Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H., and Yuen, K. Y. (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14040–14045.

(24) Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. Y., Zheng, K., Lei, L. C., Chen, Q. X., Gao, Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan, H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F., Qin, P. Z., Li, L. H., Ren, Y. Q., Liang, W. J., Yu, Y. D., Anderson, L., Wang, M., Xu, R. H., Wu, X. W., Zheng, H. Y., Chen, J. D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L. Y., Li, H., Chen, F., Di, B., He, L. J., Lin, J. Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X. J., Spiga, O., Guo, Z. M., Pan, H. Y., He, W. Z., Manuguerra, J. C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y. X., Wu, C. I., and Zhao, G. P. (2005) Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2430–2435.

(25) Lau, S. K., Li, K. S., Tsang, A. K., Lam, C. S., Ahmed, S., Chen, H., Chan, K. H., Woo, P. C., and Yuen, K. Y. (2013) Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *J. Virol.* 87, 8638–8650.

(26) Ithete, N. L., Stoffberg, S., Corman, V. M., Cottontail, V. M., Richards, L. R., Schoeman, M. C., Drosten, C., Drexler, J. F., and Preiser, W. (2013) Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg. Infect. Dis.* 19, 1697–1699.

(27) Corman, V. M., Ithete, N. L., Richards, L. R., Schoeman, M. C., Preiser, W., Drosten, C., and Drexler, J. F. (2014) Rooting the phylogenetic tree of middle East respiratory syndrome coronavirus by

characterization of a conspecific virus from an african bat. *J. Virol.* 88, 11297–11303.

(28) Yang, L., Wu, Z., Ren, X., Yang, F., Zhang, J., He, G., Dong, J., Sun, L., Zhu, Y., Zhang, S., and Jin, Q. (2014) MERS-related betacoronavirus in *Vespertilio superans* bats, China. *Emerg. Infect. Dis.* 20, 1260–1262.

(29) Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., Alhakeem, R., Durosinloun, A., Al Asmari, M., Islam, A., Kapoor, A., Briese, T., Daszak, P., Al Rabeeah, A. A., and Lipkin, W. I. (2013) Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* 19, 1819–1823.

(30) De Benedictis, P., Marciano, S., Scaravelli, D., Priori, P., Zecchin, B., Capua, I., Monne, I., and Cattoli, G. (2014) Alpha and lineage C betaCoV infections in Italian bats. *Virus Genes* 48, 366–371.

(31) Annan, A., Baldwin, H. J., Corman, V. M., Klose, S. M., Owusu, M., Nkrumah, E. E., Badu, E. K., Anti, P., Agbenyega, O., Meyer, B., Oppong, S., Sarkodie, Y. A., Kalko, E. K., Lina, P. H., Godlevska, E. V., Reusken, C., Seebens, A., Gloza-Rausch, F., Vallo, P., Tschapka, M., Drosten, C., and Drexler, J. F. (2013) Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. *Emerg. Infect. Dis.* 19, 456–459.

(32) Azhar, E. I., El-Kafrawy, S. A., Farraj, S. A., Hassan, A. M., Al-Saeed, M. S., Hashem, A. M., and Madani, T. A. (2014) Evidence for camel-to-human transmission of MERS coronavirus. *N. Engl. J. Med.* 370, 2499–2505.

(33) Reusken, C. B., Haagmans, B. L., Muller, M. A., Gutierrez, C., Godeke, G. J., Meyer, B., Muth, D., Raj, V. S., Smits-De Vries, L., 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., Gortazar-Schmidt, C., Drosten, C., and Koopmans, M. P. (2013) Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect. Dis.* 13, 859–866.

(34) Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H., Muller, M. A., Dijkman, R., Muth, D., Demmers, J. A., Zaki, A., Fouchier, R. A., Thiel, V., Drosten, C., Rottier, P. J., Osterhaus, A. D., Bosch, B. J., and Haagmans, B. L. (2013) Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495, 251–254.

(35) Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., Ji, W., Li, Y., Wu, Y., Wang, J., Iwamoto, A., Woo, P. C., Yuen, K. Y., Yan, J., Lu, G., and Gao, G. F. (2014) Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 16, 328–337.

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

(37) Wang, Q., Wong, G., Lu, G., Yan, J., and Gao, G. F. (2016) MERS-CoV spike protein: Targets for vaccines and therapeutics. *Antiviral Res.* 133, 165–177.

(38) Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309, 1864–1868.

(39) 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., and Gao, G. F. (2013) Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500, 227–231.

(40) 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., and Wang, X. (2013) Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* 23, 986–993.

(41) Peng, G., Sun, D., Rajashankar, K. R., Qian, Z., Holmes, K. V., and Li, F. (2011) Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10696–10701.

(42) 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., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450–454.

- (43) Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., and Shi, Z. (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J. Virol.* 86, 4620–4630.
- (44) 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., and Yuen, K. Y. (2010) Coexistence of different genotypes in the same bat and serological characterization of Roussetus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J. Virol.* 84, 11385–11394.
- (45) Tao, Y., Tang, K., Shi, M., Conrardy, C., Li, K. S., Lau, S. K., Anderson, L. J., and Tong, S. (2012) Genomic characterization of seven distinct bat coronaviruses in Kenya. *Virus Res.* 167, 67–73.
- (46) Tong, S., Conrardy, C., Ruone, S., Kuzmin, I. V., Guo, X., Tao, Y., Niezgod, M., Haynes, L., Agwanda, B., Breiman, R. F., Anderson, L. J., and Rupprecht, C. E. (2009) Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* 15, 482–485.
- (47) Zhang, W., Qi, J., Shi, Y., Li, Q., Gao, F., Sun, Y., Lu, X., Lu, Q., Vavricka, C. J., Liu, D., Yan, J., and Gao, G. F. (2010) Crystal structure of the swine-origin A (H1N1)-2009 influenza A virus hemagglutinin (HA) reveals similar antigenicity to that of the 1918 pandemic virus. *Protein Cell* 1, 459–467.
- (48) Song, H., Qi, J., Khedri, Z., Diaz, S., Yu, H., Chen, X., Varki, A., Shi, Y., and Gao, G. F. (2016) An open receptor-binding cavity of hemagglutinin-esterase-fusion glycoprotein from newly-identified influenza D virus: basis for its broad cell tropism. *PLoS Pathog.* 12, e1005411.
- (49) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data. *Methods Enzymol.* 276, 307–326.
- (50) Uson, I., and Sheldrick, G. M. (1999) Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* 9, 643–648.
- (51) Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 57, 1373–1382.
- (52) Cowtan, K. D., and Zhang, K. Y. (1999) Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* 72, 245–270.
- (53) 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., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 213–221.
- (54) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 2126–2132.
- (55) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 12–21.
- (56) Chen, Y., Rajashankar, K. R., Yang, Y., Agnihothram, S. S., Liu, C., Lin, Y. L., Baric, R. S., and Li, F. (2013) Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *J. Virol.* 87, 10777–10783.
- (57) Shi, Z. (2010) Bat and virus. *Protein Cell* 1, 109–114.
- (58) Kubo, H., Yamada, Y. K., and Taguchi, F. (1994) Localization of neutralizing epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J. Virol.* 68, 5403–5410.
- (59) Wu, K., Li, W., Peng, G., and 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.
- (60) Reguera, J., Santiago, C., Mudgal, G., Ordone, D., Enjuanes, L., and Casasnovas, J. M. (2012) Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.* 8, e1002859.