LETTER

Molecular basis of binding between novel human coronavirus MERS–CoV and its receptor CD26

Guangwen Lu[1]*, Yawei Hu[2]*, Qihui Wang[1]*, Jianxun Qi[1]*, Feng Gao[3,4]*, Yan Li[1], Yanfang Zhang[1,5], Wei Zhang[1], Yuan Yuan[1,6], Jinku Bao[4], Buchang Zhang[2], Yi Shi[7], Jinghua Yan[1] & George F. Gao[1,5,6,7,8]

The newly emergent Middle East respiratory syndrome coronavirus (MERS-CoV) can cause severe pulmonary disease in humans[1,2], representing the second example of a highly pathogenic coronavirus, the first being SARS-CoV[3]. CD26 (also known as dipeptidyl peptidase 4, DPP4) was recently identified as the cellular receptor for MERS-CoV[4]. The engagement of the MERS-CoV spike protein with CD26 mediates viral attachment to host cells and virus–cell fusion, thereby initiating infection. Here we delineate the molecular basis of this specific interaction by presenting the first crystal structures of both the free receptor binding domain (RBD) of the MERS-CoV spike protein and its complex with CD26. Furthermore, binding between the RBD and CD26 is measured using real-time surface plasmon resonance with a dissociation constant of 16.7 nM. The viral RBD is composed of a core subdomain homologous to that of the SARS-CoV spike protein, and a unique strand-dominated external receptor binding motif that recognizes blades IV and V of the CD26 β-propeller. The atomic details at the interface between the two binding entities reveal a surprising protein–protein contact mediated mainly by hydrophilic residues. Sequence alignment indicates, among betacoronaviruses, a possible structural conservation for the region homologous to the MERS-CoV RBD core, but a high variation in the external receptor binding motif region for virus-specific pathogenesis such as receptor recognition.

The recent identification of a novel coronavirus, MERS-CoV—which, as of May 15th 2013, had infected 40 patients with a total of 20 fatalities—has drawn worldwide attention as a potential cause of a future pandemic[5]. Unlike most coronaviruses circulating in humans that only cause mild respiratory illness[6], MERS-CoV possibly represents a second reported coronavirus of severely high virulence after SARS-CoV, which caused over 8,000 infection cases globally in 2003, with more than 800 deaths[3]. The clinical manifestations of MERS-CoV infection include fever, cough, acute respiratory distress syndrome and, in some cases, accompanying renal failure[1,2], and are very similar to those caused by SARS-CoV. However, the novel coronavirus diverges from SARS-CoV in genomic sequence, and is much more closely related to the bat-derived HKU4 and HKU5 coronaviruses[7,8]. Consistent with phylogenetic analysis, MERS-CoV does not use the SARS-CoV receptor, angiotensin converting enzyme 2 (ACE2), as its entry receptor[9]; rather, a recent study showed that it uses human CD26 for this purpose[4]. CD26 is the third peptidase to be identified as a functional coronavirus receptor, the others being aminopeptidase N (ANPEP, also known as APN and CD13)[10,11] and ACE2 (ref. 12).

The recognition of CD26 by MERS-CoV is mediated by virus surface spike (S) protein[4]. As with other coronaviruses, the MERS-CoV S protein would be cleaved in host cells into S1 and S2 subunits (Fig. 1a). S1 engages the receptor[4] whereas S2, with typical sequence motifs homologous to those identified as the heptad repeats in class I enveloped

viruses[13–15], should mediate membrane fusion. The exploitation of the virus–receptor interaction and thus of the intervention strategies requires an atomic delineation of the receptor-binding properties of S1. On the basis of previous studies, the receptor attachment sites of coronavirus S1 subunits might locate to either the amino-terminal (such as in murine hepatitis virus[16]) or the carboxy-terminal (such as in, for example, SARS-CoV[17] and human coronavirus NL63 (ref. 18)) domain. We therefore tested individually the binding of MERS-CoV S1 and its N- and C-terminal-domain proteins to cell-surface-expressed CD26 molecules. The receptor-binding capacity was attributed to the C-terminal amino acids 367–606 of MERS-CoV S1 (Fig. 1b). We hereby referred to this domain as RBD. The potent interaction between MERS-CoV RBD and CD26 was further demonstrated by surface plasmon resonance assays, in which CD26 binds to MERS-CoV RBD with a dissociation constant (K_d) of about 16.7 nM (K_{on}, 1.79×10^5 M^{-1} s^{-1}; K_{off}, 2.99×10^{-3} s^{-1}), but does not bind to the RBD of SARS-CoV (Fig. 1c).

We crystallized MERS-CoV RBD and solved its structure at a resolution of 2.5 Å (Supplementary Table 1). Two molecules of essentially the same structure are present in the asymmetric unit. Each molecule contains 208 consecutive density-traceable amino acids from V381 to L588. A Dali[19] search within the Protein Data Bank (PDB) revealed clear structural homology between MERS-CoV RBD and SARS-CoV RBD (PDB code, 2DD8; Z score, 15.1). We therefore divided the MERS-CoV RBD structure into two subdomains: a core and an external β-sheet, using the structure of SARS-CoV RBD as a reference. The core subdomain reveals a five-stranded antiparallel β-sheet (β1, β3, β4, β5 and β10) in the centre. The connecting helices (four α-helices: α1–4 and two 3₁₀-helices: η1 and η2) and two small β-strands (β2 and β11) further decorate the sheet on both sides, together forming a globular fold. Three disulphide bonds, connecting C383 to C407, C425 to C478, and C437 to C585, respectively, stabilize the core-domain structure from the interior. At the solvent-exposed side, the RBD termini are clinched adjacent to each other (Fig. 2a, b). This subdomain fold is very similar to that of the SARS-CoV RBD core (a root mean squared deviation of 2.79 Å for 76 Cα pairs). Superimposition of the two structures reveals a well-aligned centre sheet and homologous peripheral helices and strands, although several intervening loops are observed to exhibit large conformational variance (Fig. 2c).

The external subdomain of MERS-CoV RBD is mainly a β-sheet structure with three large (β6, β8 and β9) and one small (β7) strand arranged in an antiparallel manner. It is anchored to the RBD core through the β5/6, β7/8 and β9/10 intervening loops, which touch the core subdomain like a clamp at both the top and bottom positions. Two small 3₁₀ helices (η3 and η4) and most of the connecting loops in this subdomain locate on the interior side of the sheet, hence exposing a flat exterior sheet-face to the solvent. Residues C503 and C526 form
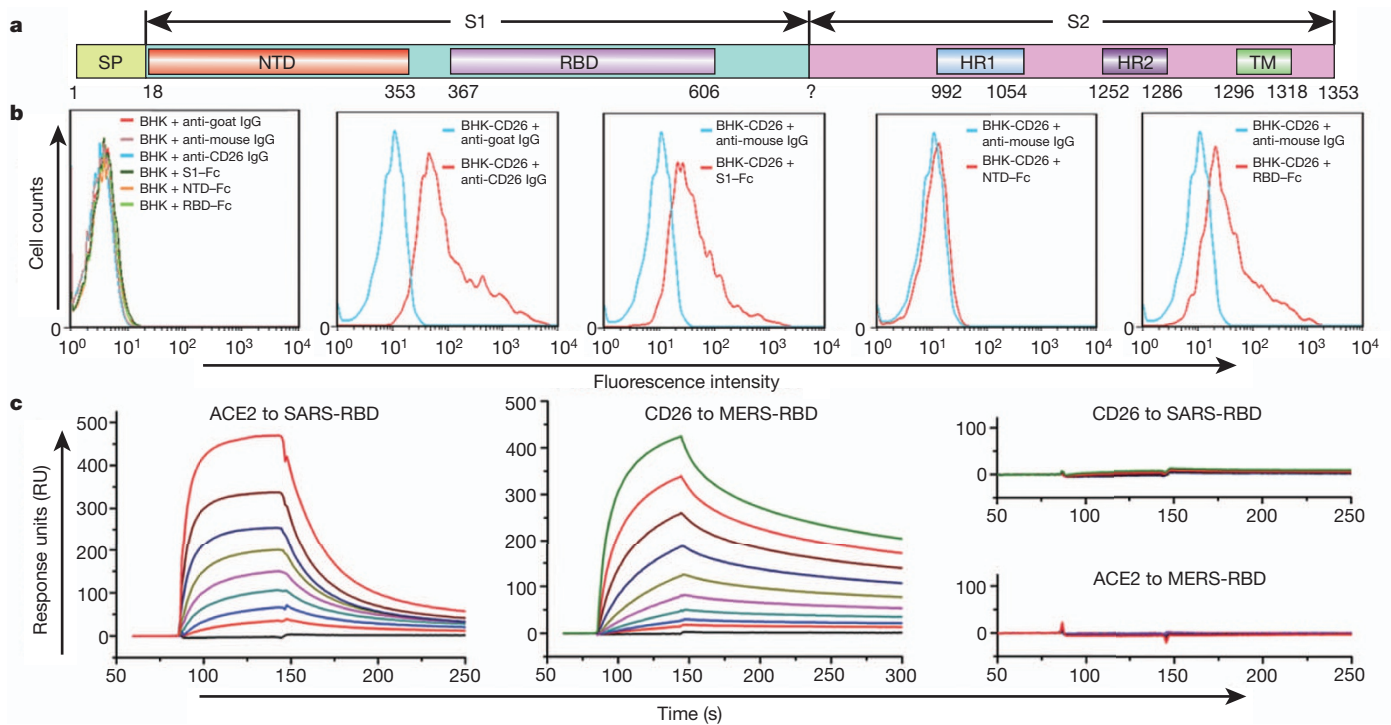
Figure 1 | Identification of the MERS-CoV RBD. a, A schematic representation of the MERS-CoV S protein. The N-terminal domain (NTD) and RBD are defined on the basis of a pairwise sequence alignment with the N-terminal galectin-like domain of murine hepatitis virus S and the RBD of SARS-CoV S, respectively. The remaining domain elements are bioinformatically defined on the basis of the web-server predictions (signal peptide (SP), SignalP 4.0 server; transmembrane domain (TM), TMHMM server; heptad repeats 1 and 2 (HR1 and HR2), Learncoil-VMF program). ? denotes the presumed/estimated S1/S2 cleavage site. A previous prediction[4] indicates cleavage between R751 and S752 with a 602-residue S2. However, a recent study[28] revealed a spike C-terminal domain (possibly S2) of ~100 kDa, indicating a cleavage site upstream of R751/S752. **b**, A flow cytometric assay of the Fc-fused S protein or its subdomain proteins involved in CD26 binding. Mock-transfected baby hamster kidney (BHK) cells or BHK cells transfected with CD26-expressing plasmid (BHK-CD26) were tested with the individual Fc-fusion proteins or an anti-CD26 antibody (anti-CD26 IgG). For each test, the secondary antibodies (anti-goat IgG or anti-mouse IgG) were used as the negative control. The profiles are shown. From left to right: BHK cells with the indicated Fc-fusion proteins or antibodies, BHK-CD26 with anti-CD26 antibody, BHK-CD26 with Fc-fused S1, BHK-CD26 with Fc-fused NTD, BHK-CD26 with Fc-fused RBD. **c**, A surface plasmon resonance assay characterizing the specific binding between CD26 and MERS-CoV RBD. The profiles are shown. Left, human ACE2 to SARS-CoV RBD; middle, CD26 to MERS-CoV RBD; top right, CD26 to SARS-CoV RBD; bottom right, human ACE2 to MERS-CoV RBD.

the fourth disulphide bond, linking the η3 helix to strand β6 (Fig. 2a, b). With no observable structure homology (Fig. 2c), the external subdomains of MERS-CoV and SARS-CoV RBDs are topological equivalents, both being present as an 'insertion' between the equivalent core-strands (strands β5 and β10 in MERS-CoV, and β6 and β9 in SARS-CoV) (Supplementary Fig. 1).

To elucidate the structural basis of the virus–receptor engagement, we further prepared the RBD–CD26 complex by *in vitro* mixture of the two proteins and then purification on a gel filtration column. Consistent with the high binding affinity between MERS-CoV RBD and CD26, the complex is easily obtainable and stable (Supplementary Fig. 2). The complex structure was solved at 2.7 Å resolution (Supplementary Table 1) with one RBD binding to a single CD26 molecule in the asymmetric unit. The receptor, as shown in previous reports[20,21], is composed of an eight-bladed β-propeller domain and an α/β hydrolase domain. MERS-CoV RBD binds to the side-surface of the CD26 β-propeller, recognizing blades IV and V and a small bulged helix in the blade-linker. As for the viral ligand, the entire receptor binding site locates in the external subdomain and to the solvent-exposed sheet-face, qualifying the subdomain as the receptor binding motif (RBM) (Fig. 3a). Overall, engagement of the receptor does not induce obvious conformational changes in RBM, although small structural variance could be observed for the tip-loops. The η2–α4 loop in the RBD core, however, unexpectedly exhibits a large conformational difference between the free and the bounded structures (Supplementary Fig. 3). We believe this is due to a crystal contact present in the free RBD structure, which is interrupted in the complex crystal by the engaging receptor.

CD26 is a type II transmembrane protein. It is present as a homodimer on the cell surface[20–22]. The dimerization of the peptidase relies on broad intermolecule contacts contributed by the hydrolase domain and the extended strands in blade IV of the β-propeller[20,21]. A lateral binding of MERS-CoV RBD to CD26 would therefore not disrupt CD26 dimerization. Accordingly, a similar U-shaped CD26 dimer could be generated by symmetry operations of the complex structure. The viral ligand locates at the membrane-distal tip of the dimer, corresponding well to a *trans* interaction between the virus and the receptor (Fig. 3b). Considering that the RBD N and C termini are on the same side distant from CD26, it is unlikely that the remaining S domains would contact the receptor molecule. The binding mode revealed by the complex structure is also in good accordance with a previous study showing that the virus–receptor interaction is independent of the peptidase activity of CD26 (ref. 4). The bound RBD is far away from interfering with either the substrate/product accessing tunnels or the catalytic centre[20,21] (Fig. 3b).

Overall, a surface area of 1203.4 and 1113.4 Å^2 in CD26 and MERS-CoV RBD, respectively, is buried by complex formation (Fig. 4a). Scrutinization of the binding interface reveals a group of hydrophilic residues at the site, forming a polar-contact (H-bond and salt-bridge) network. These interactions are predominantly mediated by the residue side chains (including RBD Y499 with CD26 R336, N501 with Q286, K502 with T288, D510 with R317, E513 with Q344, and D539 with K267), although CD26 L294 and RBD D510 are observed to contact RBD R542 and CD26 Y322, respectively, through the main-chain oxygen atom (Fig. 4b). In addition, the bulged helix in CD26
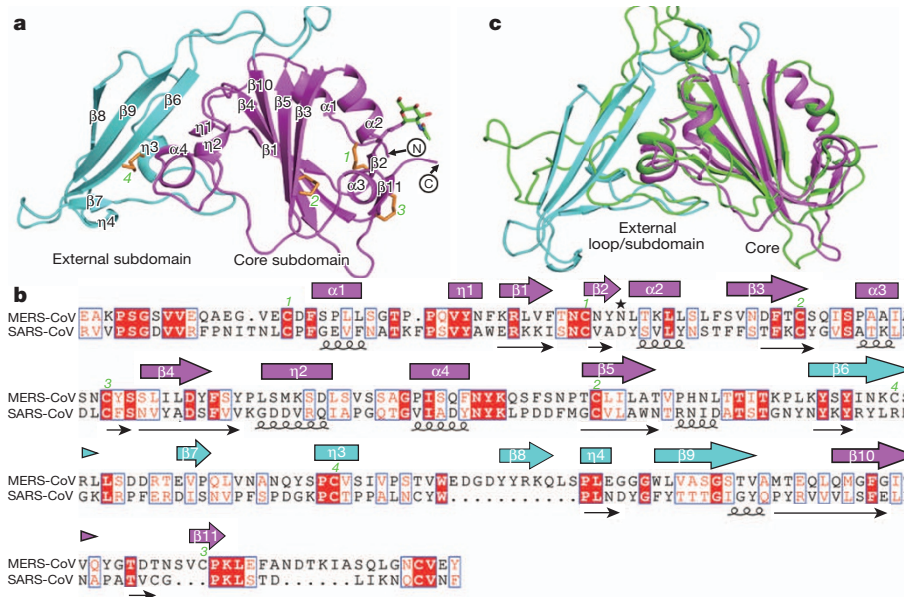
Figure 2 | The overall structure of MERS-CoV RBD. a, A cartoon representation of the RBD structure. The secondary structural elements are labelled according to their occurrence in sequence. The disulphide bonds (marked with Arabic numbers 1–4) and N-glycan linked to N410 are shown as orange and green sticks, respectively. Core subdomain, magenta; external subdomain, cyan. The N and C termini are labelled. **b**, An amino acid sequence alignment between MERS-CoV and SARS-CoV RBDs. The hollow boxes and arrows indicate α/3₁₀ helices and β-strands, respectively, and are coloured as in **a**. To facilitate comparison, the secondary-structure elements of SARS-CoV RBD (PDB code, 2DD8) are marked with spiral (helices) and arrow (strands) lines below the sequence. The cysteine residues that form disulphide bonds are labelled as in **a**, and residue N410 with a star. **c**, A structural alignment between MERS-CoV (magenta for core and cyan for external subdomains) and SARS-CoV (green) RBDs.

properly positions three hydrophobic residues A291, L294 and I295 into close proximity with the RBD amino acids Y540, W553 and V555, forming a hydrophobic centre at the interface (Fig. 4c). Further virus–receptor contacts include V341 and I346 of CD26 packing against P515 and the apolar carbon atoms of R511 and E513 in RBD (Fig. 4d), and a CD26 N229-linked carbohydrate moiety interacting with RBD amino acids W535 and E536 (Fig. 4e). Overall, the virus–receptor engagement is dominated by the polar contacts mediated by the hydrophilic residues, and mutations of those in RBD (six alanine substitutions and one Y499F mutation of the CD26-interacting amino acids) completely abrogated its interaction with CD26 (Supplementary Fig. 4). The features of these residue interactions are very similar to those mediating the interaction between adenosine deaminase (ADA) and CD26 (ref. 23). By a pairwise comparison, we unexpectedly found

that all those CD26 residues identified in the virus–receptor interface are also involved in ADA binding, indicating a competition between ADA and the virus for CD26 receptor. As the ADA–CD26 interaction is shown to induce co-stimulatory signals in T cells[22], this may indicate a possible manipulation of the host immune system by MERS-CoV through competition for the ADA-recognition site. It is also noteworthy that those CD26 residues involved in RBD binding are highly conserved between human and bat, with only two variations (I295T and R317Q), explaining the capability of MERS-CoV using bat CD26 for cell entry[4] (Supplementary Fig. 5).

Coronaviruses can be categorized into three main genera or groups (group 1 (alpha), group 2 (beta) and group 3 (gamma) coronaviruses)[24]. Both MERS-CoV and SARS-CoV belong to the betacoronavirus genus, but are classified into different lineage subgroups (subgroup 2b for
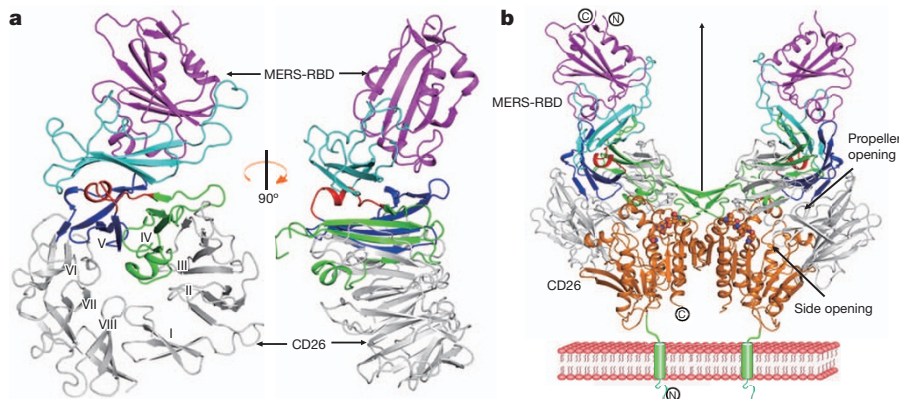

Figure 3 | The complex structure of MERS-CoV RBD bound to CD26. a, A cartoon representation of the complex structure. For clarity, only the β-propeller domain of CD26 (grey) is shown. Blades IV, V and the intervening IV/V linker that recognize RBD are highlighted in green, blue and red, respectively. The core subdomain and external RBM are coloured magenta and cyan, respectively. The right panel is yielded by clockwise rotation of the left panel along a longitudinal axis in the page-face. **b**, A symmetry-related CD26

dimer observed in the complex crystal. The two-fold axis is shown as an upright arrow. The transmembrane topology of CD26 is indicated with a modelled lipid-bilayer membrane. In CD26, the propeller and side openings indicated as the substrate entrance/exit tunnels are marked with arrows, and the catalytic triad residues are highlighted as spheres. Colour selections are the same as in **a**, and the CD26 α/β hydrolase domain is shown in orange. The N and C termini are labelled.
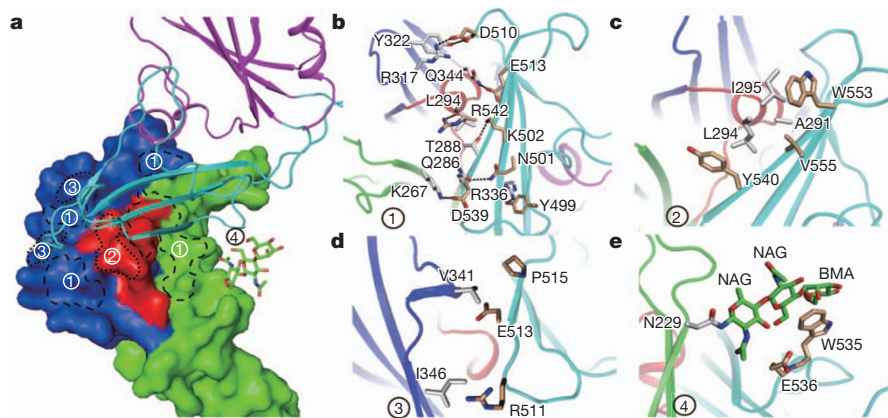
Figure 4 | The atomic interaction details at the binding interface. a, An overview of the binding interface. CD26 and RBD are shown in surface and cartoon representations, respectively, and are coloured as in Fig. 3. The carbohydrate moiety linked to CD26 N229 is shown as green sticks. The contacting sites (each allocated with an Arabic number 1–4) are further delineated in **b**–**e** for the amino acid interaction details. **b,** A strong polar-contact (H-bond and salt-bridge) network. **c, d,** The small patches of hydrophobic interactions. **e,** Contribution of the carbohydrate moiety. The residues involved are shown and labelled. NAG, *N*-acetyl-D-glucosamine; BMA, beta-D-mannose.

SARS-CoV and 2c for MERS-CoV)[8]. We noted that the spike sequences are of low identity among different subgroup members. For example, MERS-CoV and SARS-CoV S proteins show a sequence identity of less than 28%. Nevertheless, RBDs of the two coronaviruses are homologous for the core subdomain. Notably, the three interior disulphide bonds in the core are well-aligned for the steric positions in the two RBD structures and well-conserved in sequence among betacoronaviruses. Conversely, the external RBM region is highly variable in both length and residue composition (Supplementary Fig. 6). Consistently, no structural homology in this subdomain is observed between MERS-CoV and SARS-CoV. Yet it is this subdomain that engages cellular receptors. We therefore assume that betacoronaviruses probably have a similar core-domain fold in the S protein to present the external amino acids with divergent structures for viral pathogenesis, such as receptor recognition.

Our work presents the fifth structure of virus S protein–receptor complexes in the *Coronaviridae* family[16–18,25]. Taking into account both the RBD structure and the binding mode with receptors, MERS-CoV is related to SARS-CoV[17] (a single insertion functioning as RBM) but differs from porcine respiratory coronavirus[25] and NL63 (ref. 18) of alphacoronaviruses (multiple discontinuous RBMs) (Supplementary Fig. 7). Nevertheless, related structural topologies can still be observed in RBDs of these coronaviruses[26]. We noted that in the RBD–receptor complex structures of both MERS-CoV and porcine respiratory coronavirus the binding interfaces involve a receptor N-glycan. This might represent another cross-genus similarity in the *Coronaviridae* family, which supports a proposed common evolutionary origin of coronavirus S proteins[26]. It would therefore be interesting to investigate the contribution of the sugar moiety to the virus–receptor interaction for MERS-CoV in the future.

Vaccination remains the most useful measure to combat viral infection and transmission. A large number of antibodies show neutralization activity by targeting the RBD and thereby disrupting the virus–receptor engagement. Therefore, a properly folded RBD could be an ideal immunogen for vaccination, as demonstrated for SARS-CoV[27]. A recent report indeed shows the presence of S-specific neutralizing antibodies in MERS-CoV-infected patients[28]. It may be worth attempting to test the immunization effect of MERS-CoV RBD in the future.

METHODS SUMMARY

Protein expression, purification, crystallization and structure determination. Both His-tagged CD26 and MERS-CoV RBD proteins were expressed in insect High Five cells using the Bac-to-Bac baculovirus expression system (Invitrogen). The recombinant proteins were then purified via nickel-chelated affinity chromatography and gel filtration. Crystals were obtained by initial screening with the commercially available kits followed by optimization. The RBD structure was solved by single-wavelength anomalous diffraction and the complex structure by molecular replacement.

Full Methods and any associated references are available in the online version of the paper.

1. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367,** 1814–1820 (2012).
2. Bermingham, A. *et al.* 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 (2012).
3. World Health Organization. Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS) http://www.who.int/csr/sars/country/en/.
4. Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495,** 251–254 (2013).
5. World Health Organization. Novel coronavirus infection - update. http://www.who.int/csr/don/2013_05_15_ncov/en/ (2013).
6. Weiss, S. R. & Navas-Martin, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* **69,** 635–664 (2005).
7. van Boheemen, S. *et al.* Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* **3,** e00473–e12 (2012).
8. Lu, G. & Liu, D. SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases. *Protein Cell* **3,** 803–805 (2012).
9. Müller, M. A. *et al.* Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines. *mBio* **3,** e00515–e12 (2012).
10. Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357,** 420–422 (1992).
11. Delmas, B. *et al.* Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* **357,** 417–420 (1992).
12. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450–454 (2003).
13. Gao, G. F. in *Combating the Threat of Pandemic Influenza: Drug Discovery Approaches* (ed. Torrence, P.F.) 226–246 (John Wiley & Sons, 2007).
14. Skehel, J. J. & Wiley, D. C. Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* **95,** 871–874 (1998).
15. Zhu, J. *et al.* Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem. Biophys. Res. Commun.* **319,** 283–288 (2004).
16. Peng, G. *et al.* Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl Acad. Sci. USA* **108,** 10696–10701 (2011).
17. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309,** 1864–1868 (2005).
18. Wu, K., Li, W., Peng, G. & Li, F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl Acad. Sci. USA* **106,** 19970–19974 (2009).
19. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).

20. Rasmussen, H. B., Branner, S., Wiberg, F. C. & Wagtmann, N. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nature Struct. Biol.* **10,** 19–25 (2003).
21. Engel, M. *et al.* The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. *Proc. Natl Acad. Sci. USA* **100,** 5063–5068 (2003).
22. Gorrell, M. D., Gysbers, V. & McCaughan, G. W. CD26: a multifunctional integral membrane and secreted protein of activated lymphocytes. *Scand. J. Immunol.* **54,** 249–264 (2001).
23. Weihofen, W. A., Liu, J., Reutter, W., Saenger, W. & Fan, H. Crystal structure of CD26/dipeptidyl-peptidase IV in complex with adenosine deaminase reveals a highly amphiphilic interface. *J. Biol. Chem.* **279,** 43330–43335 (2004).
24. Lai, M. M., Perlman, S. & Anderson, L. J. in *Fields Virology* (ed. Knipe, D.M.) 1305–1336 (Lippincott Williams & Wilkins, 2007).
25. Reguera, J. *et al.* Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.* **8,** e1002859 (2012).
26. Li, F. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J. Virol.* **86,** 2856–2858 (2012).
27. Du, L. *et al.* The spike protein of SARS-CoV–a target for vaccine and therapeutic development. *Nature Rev. Microbiol.* **7,** 226–236 (2009).
28. Gierer, S. *et al.* The spike-protein of the emerging betacoronavirus EMC uses a novel coronavirus receptor for entry, can be activated by TMPRSS2, and is targeted by neutralizing antibodies. *J. Virol.* **87,** 5502–5511 (2013).

METHODS

Protein expression and purification. The proteins used for crystallization and surface plasmon resonance experiments were prepared with the Bac-to-Bac baculovirus expression system (Invitrogen). The coding sequences for MERS-CoV RBD (GenBank accession number JX869059, spike residues 367–606), SARS-CoV RBD (accession number NC_004718, spike residues 306–527), human CD26 (accession number NP_001926, residues 39–766) and human ACE2 (accession number BAJ21180, residues 19–615) were individually cloned into the pFastBac1 vector. For each construct, a previously described gp67 signal peptide sequence[29] was added to the protein N terminus for protein secretion, and a hexa-His tag was added to the C terminus to facilitate further purification processes. Transfection and virus amplification were conducted with Sf9 cells, and the recombinant proteins were produced in High Five cells. The cell culture was collected 48 h after infection and passed through a 5-ml HisTrap HP column (GE Healthcare). After removal of most of the impurities, the recovered proteins were then pooled and further purified on a Superdex 200 column (GE Healthcare). Finally, each collected protein was prepared in a buffer consisting of 20 mM Tris-HCl (pH 8.0) and 150 mM NaCl and concentrated to about 10 mg ml^{-1} for further use.

To obtain the complex of MERS-CoV RBD bound to CD26, the individual proteins were *in vitro* mixed at a molar ratio of 1:1 and incubated at 4 °C for about 2 h. The complex was then further purified on a Superdex 200 column, and concentrated to about 15 mg ml^{-1} for crystallization experiments.

To prepare the Fc chimaeric proteins, the fragment encoding MERS-CoV S1 (residues 1–751) or NTD (residues 1–353) or RBD (adding the S residues 1–17 of the signal peptide to its N terminus to facilitate protein secretion) was fused 5′-terminally to a fragment coding for the Fc domain of mouse IgG and ligated into the pCAGGS expression vector. A mutant RBD–Fc protein-expressing plasmid was also constructed by site-directed mutagenesis, for which the identified hydrophilic residues involved in CD26 binding were mutated simultaneously (Y499F; N501A, K502A, D510A, E513A, D539A and R542A). The expression plasmids were then transfected into HEK293T cells. The cell culture was collected 48 h after transfection and directly used in the flow cytometric assay.

Analytical gel filtration. MERS-CoV RBD, CD26 and their protein complex were individually prepared and adjusted to the same volume. The samples were then loaded onto a calibrated Superdex 200 column (GE Healthcare). The chromatographs were recorded and overlaid onto each other. The pooled proteins were analysed on a 12% SDS–PAGE gel and stained with Coomassie blue.

Surface plasmon resonance assay. The BiAcore experiments were carried out at room temperature (25 °C) using a BIAcore 3000 machine with CM5 chips (GE Healthcare). For all the measurements, an HBS-EP buffer consisting of 10 mM HEPES, pH 7.5, 150 mM NaCl, 3 mM EDTA and 0.005% (v/v) Tween-20 was used, and all proteins were exchanged to the same buffer in advance via gel filtration. The MERS-CoV RBD and SARS-CoV RBD proteins were immobilized on the chip at about 500 response units. Gradient concentrations of human CD26 (0, 5, 10, 20, 40, 80, 160, 320, 640 and 1,280 nM) or human ACE2 (0, 10, 20, 40, 80, 160, 320, 640 and 1,280 nM) were then used to flow over the chip surface. After each cycle, the sensor surface was regenerated via a short treatment using 10 mM NaOH. The binding kinetics were analysed with the software BIAevaluation Version 4.1 using the 1:1 Langmuir binding model.

Flow cytometric assay. For the surface expression of CD26, the full-length coding sequence was cloned into the pEGFP-C1 vector which yields a plasmid encoding a recombinant CD26 protein with an EGFP-tag fused to its N terminus. The plasmid was transfected into the CD26-negative BHK cells using lipo2000 (Invitrogen) according to the manufacturer's instructions. The cells were collected 48 h after transfection.

For staining, the mock-transfected BHK cells or the cells transfected with the CD26-expressing plasmid were suspended in PBS and incubated with the individual Fc-fusion protein culture or goat anti-CD26 IgG (R&D Systems) at room temperature for 1 h. The cells were then washed and further incubated at room temperature for about 0.5 h with anti-mouse or anti-goat secondary IgG antibodies (R&D Systems). After washing, the cells were analysed by flow cytometry with a

BD FACSCalibur machine. The cells incubated only with the secondary antibodies were used as the negative controls.

Crystallization. All the crystals were obtained by vapour-diffusion sitting-drop method with 1 μl protein mixing with 1 μl reservoir solution and then equilibrating against 100 μl reservoir solution at 18 °C. The initial crystallization screenings were carried out using the commercially available kits. The conditions that yield crystals were then optimized. Diffractable crystals of the free RBD protein were finally obtained in a condition consisting of 0.1 M ammonium tartrate dibasic, pH 7.0, and 12% PEG 3,350 with a protein concentration of 10 mg ml^{-1}. Derivative crystals were obtained by soaking RBD crystals for 24 h in mother liquor containing 2 mM KAuCl$_4$•2H$_2$O. The complex crystals were grown in 6% (v/v) 2-propanol, 0.1 M sodium acetate pH 4.5 and 26% PEG550 with a protein concentration of 15 mg ml^{-1}.

Data collection, integration and structure determination. For data collection, all crystals were flash-cooled in liquid nitrogen after a brief soaking in reservoir solution with the addition of 20% (v/v) glycerol. The native RBD data set was collected at the High Energy Accelerator Research Organization (KEK) BL1A (wavelength, 1.03818 Å), whereas the diffraction data for the Au derivative crystal (wavelength, 1.0382 Å) and the complex crystal (wavelength, 0.97930 Å) were collected at the Shanghai Synchrotron Radiation Facility (SSRF) BL17U. All data were processed with HKL2000 (ref. 30). Additional processing was performed with programs from the CCP4 suite[31].

The structure of RBD was determined by the single-wavelength anomalous diffraction (SAD) method. The Au sites were first located by SHELXD[32] for the Au-SAD data. The identified position were then refined and the phases were calculated with SAD experimental phasing module of Phaser[33]. The real space constraints were further applied to the electron density map in DM[34]. The initial model was built with Autobuild in Phenix package[35]. Additional missing residues were added manually in Coot[36]. The final model was refined with phenix.refine in the Phenix[35] with energy minimization, isotropic ADP refinement, and bulk solvent modelling. The complex structure was solved by molecular replacement module of Phaser[33], with the solved RBD structure and previously reported CD26 structure (PDB code, 2BGR) as the search models. The atomic model was completed with Coot[36] and refined with phenix.refine[35]. The stereochemical qualities of the final models were assessed with PROCHECK[37]. The Ramachandran plot distributions for the residues in the free RBD structure were 86.8, 11.8 and 1.4% for the most favoured, additionally and generously allowed regions, respectively. These values were 86.5, 13.1 and 0.5% for the RBD–CD26 complex structure. Data collection and refinement statistics are summarized in Supplementary Table 1. All structural figures were generated using Pymol (http://www.pymol.org).

Secondary-structure determination. The secondary structure determination was based on the ESPript[38] algorithm.

29. Zhang, W. *et al.* 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 (2010).
30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
31. Collaborative Computing Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
32. Usón, I. & Sheldrick, G. M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9,** 643–648 (1999).
33. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57,** 1373–1382 (2001).
34. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72,** 245–270 (1999).
35. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
36. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).
38. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15,** 305–308 (1999).