Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2

Renhong Yan1,2, Yuanyuan Zhang2,2, Yaning Li2,2, Lu Xia1,2, Yingying Guo1,2, Qiang Zhou1,2●

Angiotensin-converting enzyme 2 (ACE2) is the cellular receptor for severe acute respiratory syndrome coronavirus (SARS-CoV) and the new coronavirus (SARS-CoV-2) that is causing the serious coronavirus disease 2019 (COVID-19) epidemic. Here, we present cryo-electron microscopy structures of full-length human ACE2 in the presence of the neutral amino acid transporter B0AT1 with or without the receptor binding domain (RBD) of the surface spike glycoprotein (S protein) of SARS-CoV-2, both at an overall resolution of 2.9 angstroms, with a local resolution of 3.5 angstroms at the ACE2-RBD interface. The ACE2-B0AT1 complex is assembled as a dimer of heterodimers, with the collectrin-like domain of ACE2 mediating homodimerization. The RBD is recognized by the extracellular peptide domain of ACE2 mainly through polar residues. These findings provide important insights into the molecular basis for coronavirus recognition and infection.

Structural determination of the ACE2-B0AT1 complex

Full-length human ACE2 and B0AT1, with Strep and FLAG tags on their respective N termini, were coexpressed in human embryonic kidney (HEK) 293F cells and purified through tandem affinity resin and size exclusion chromatography. The complex was eluted in a single monodisperse peak, indicating high homogeneity (Fig. 1A). Details of cryo-sample preparation, data acquisition, and structural determination are given in the materials and methods section of the supplementary materials. A three-dimensional (3D) reconstruction was obtained at an overall resolution of 2.9 Å from 418,140 selected particles. This immediately revealed the dimer of heterodimers’ architecture (Fig. 1B). After applying focused refinement and C2 symmetry expansion, the resolution of the extraacellular domains improved to 2.7 Å, whereas the TM domain remained at 2.9-Å resolution (Fig. 1B, figs. S1 to S3, and table S1).

The high resolution supported reliable model building. For ACE2, side chains could be assigned to residues 19 to 768, which contain the PD (residues 19 to 615) and the CLD (residues 616 to 768), which consists of a small extracellular domain, a long linker, and the single TM helix (Fig. 1C). Between the PD and TM helix is a ferridoxin-like fold domain; we refer to this as the neck domain (residues 616 to 726) (Fig. 1C and fig. S4). Homodimerization is entirely mediated by ACE2, which is sandwiched by B0AT1. Both the PD and neck domains contribute to dimerization, whereas each B0AT1 interacts with the neck and TM helix in the adjacent ACE2 (Fig. 1C). The extracellular region is highly glycosylated, with seven and five glycosylation sites on each ACE2 and B0AT1 monomer, respectively.

SARS-CoV-2 S protein binds to the PD of ACE2 with a dissociation constant (Kd) of ~15 nM (Fig. 4). Although ACE2 is hijacked by some coronaviruses, its primary physiological role is in the maturation of angiotensin (Ang), a peptide hormone that controls vasconstriction and blood pressure. ACE2 is a type I membrane protein expressed in lungs, heart, kidneys, and intestine (15–17). Decreased expression of ACE2 is associated with cardiovascular diseases (18–20). Full-length ACE2 consists of an N-terminal PD and a C-terminal collectrin-like domain (CLD) that ends with a single transmembrane helix and a ~40-residue intracellular segment (15, 21). The PD of ACE2 cleaves Ang I to produce Ang-(1-9), which is then processed by other enzymes to become Ang-(1-7). ACE2 can also directly process Ang II to give Ang-(1-7) (15, 22).

Structures of the claw-like ACE2-PD alone and in complex with the RBD or the S protein of SARS-CoV-2 have revealed the molecular details of the interaction between the RBD of the S protein and PD of ACE2 (7, 8, 23, 24). Structural information on ACE2 is limited to the PD domain. The single transmembrane (TM) helix of ACE2 makes it challenging to determine the structure of the full-length protein. ACE2 also functions as the chaperone for membrane trafficking of the amino acid transporter B0AT1, also known as SLC6A19 (25), which mediates uptake of neutral amino acids into intestinal cells in a sodium-dependent manner. Mutations in B0AT1 may cause Hartnup disorder, an inherited disease with symptoms such as pellagra, cerebellar ataxia, and psychosis (26–28). Structures have been determined for the SLC6 family members dDAT (Drosophila dopamine transporter) and human SERT (serotonin transporter, SLC6A4) (29, 30). It is unclear how ACE2 interacts with B0AT1.

The membrane trafficking mechanism for ACE2 and B0AT1 is similar to that of the LAT1-4F2hc complex, a large neutral–amino acid transporter complex that requires 4F2hc for its plasma membrane localization (31). Our structure of LAT1-4F2hc shows that the cargo LAT1 and chaperon 4F2hc interact through both extracellular and transmembrane domains (32). We reasoned that the structure of full-length ACE2 may be revealed in the presence of B0AT1.

Here, we report cryo-electron microscopy (cryo-EM) structures of the full-length human ACE2-B0AT1 complex at an overall resolution of 2.9 Å and a complex between the RBD of SARS-CoV-2 and the ACE2-B0AT1 complex, also with an overall resolution of 2.9 Å and with 3.5-Å local resolution at the ACE2-RBD interface. The ACE2-B0AT1 complex exists as a dimer of heterodimers. Structural alignment of the RBD-ACE2-B0AT1 ternary complex with the S protein of SARS-CoV-2 suggests that two S protein trimers can simultaneously bind to an ACE2 homodimer.
During classification, another subset with 143,857 particles was processed to an overall resolution of 4.5 Å. Whereas the neck domain still dimerizes, the PDs are separated from each other in this reconstruction (Fig. 1D and fig. S1, H to K). We therefore define the two classes as the open and closed conformations. Structural comparison shows that the conformational changes are achieved through rotation of the PD domains, with the rest of the complex left nearly unchanged (movie S1).

**Homodimer interface of ACE2**

Dimerization of ACE2 is mainly mediated by the neck domain, with the PD contributing a minor interface (Fig. 2A). The two ACE2 protomers are hereafter referred to as A and B, with residues in protomer B followed by a prime symbol. Extensive polar interactions are mapped to the interface between the second (residues 636 to 658) and fourth (residues 708 to 717) helices of the neck domain (Fig. 2B). Arg652 and Arg710 in ACE2-A form cation-p interactions with Tyr641′ and Tyr633′ in ACE2-B. Meanwhile, Arg652 and Arg710 are respectively hydrogen-bonded (H-bonded) to Asn638′ and Glu639′, which also interact with Gln653, as does Asn636′. Ser709 and Asp713 from ACE2-A are H-bonded to Arg716′. This extensive network of polar interactions indicates stable dimer formation.

The PD dimer interface appears much weaker, with only one pair of interactions between Gln139 and Gln175′ (Fig. 2C). Gln139 is in a loop that is stabilized by a disulfide bond between Cys133 and Cys141 as well as multiple intraloop polar interactions (Fig. 2C). The weak interaction is consistent with the ability to transition to the open conformation, in which the interface between the neck domains remains the same while the PDs are separated from each other by ~25 Å (Fig. 2D and movie S1).

**Overall structure of the RBD-ACE2-B0AT1 complex**

To gain insight into the interaction between ACE2 and SARS-CoV-2, we purchased 0.2 mg of recombinantly expressed and purified RBD-mFc of SARS-CoV-2 (for simplicity, hereafter referred to as RBD; mFc, mouse Fc tag) from Sino Biological Inc., mixed it with our purified ACE2-B0AT1 complex at a stoichiometric ratio of ~1.1 to 1, and proceeded with cryo-grid preparation and imaging. Finally, a 3D EM reconstruction of the ternary complex was obtained.

In contrast to the ACE2-B0AT1 complex—which has two conformations, open and closed—only the closed state of ACE2 was observed in the dataset for the RBD-ACE2-B0AT1 ternary complex. The structure of the ternary complex was determined to an overall resolution of 2.9 Å from 527,017 selected particles. However, the resolution for the ACE2-B0AT1 complex was substantially higher than that for the RBDs, which are at the periphery of the complex (Fig. 3A). To improve the local resolution, focused refinement was applied; this allowed us to reach a resolution of 3.5 Å for the RBD, supporting reliable modeling and analysis of the interface (Fig. 3, figs. S5 to S7, and table S1).

**Interface between the RBD and ACE2**

As expected, each PD accommodates one RBD (Fig. 3B). The overall interface is similar to that between SARS-CoV and ACE2 (7, 8), mediated mainly through polar interactions (Fig. 4A). An extended loop region of the RBD spans the arch-shaped α1 helix of the ACE2-PD like a bridge. The α2 helix and a loop that connects the β3 and β4 antiparallel strands, referred
to as loop 3-4, of the PD also make limited contributions to the coordination of the RBD.

The contact can be divided into three clusters. The two ends of the bridge interact with the N and C termini of the $\alpha_1$ helix as well as small areas on the $\alpha_2$ helix and loop 3-4. The middle segment of $\alpha_1$ reinforces the interaction by engaging two polar residues (Fig. 4A). At the N terminus of $\alpha_1$, Gln$^{497}$, Thr$^{500}$, and Asn$^{501}$ of the RBD form a network of H-bonds with Tyr$^{41}$, Gln$^{42}$, Lys$^{353}$, and Arg$^{357}$ from ACE2 (Fig. 4B). In the middle of the bridge, Lys$^{417}$ and Tyr$^{453}$ of the RBD interact with Asp$^{30}$ and His$^{34}$ of ACE2, respectively (Fig. 4C). At the C terminus of $\alpha_1$, Gln$^{474}$ of the RBD is H-bonded to Gln$^{24}$ of ACE2, whereas Phe$^{486}$ of the RBD interacts with Met$^{82}$ of ACE2 through van der Waals forces (Fig. 4D).

Comparing the SARS-CoV-2 and SARS-CoV interfaces with ACE2

Superimposition of the RBD in the complex of SARS-CoV (SARS-CoV-RBD) and ACE2-PD [Protein Data Bank (PDB) 2AJF] with the RBD in our ternary complex shows that the SARS-CoV-2 RBD (SARS-CoV-2-RBD) is similar to SARS-CoV-RBD with a root mean square deviation (RMSD) of 0.68 Å over 139 pairs of C$\alpha$ atoms (Fig. 5A) ($^8$). Despite the overall similarity, a number of sequence variations and conformational deviations are found in their respective interfaces with ACE2 (Fig. 5 and fig. S8). At the N terminus of $\alpha_1$, the variations Arg$^{426}$ → Asn$^{439}$, Tyr$^{484}$ → Gln$^{498}$, and Thr$^{487}$ → Asn$^{501}$ at equivalent positions are observed between SARS-CoV-RBD and SARS-CoV-2-RBD (Fig. 5B). More variations are observed in the middle of the bridge. The most prominent alteration is the substitution of Val$^{404}$ in the SARS-CoV-RBD with Lys$^{417}$ in the SARS-CoV-2-RBD. In addition, from SARS-CoV-RBD to SARS-CoV-2-RBD, the substitution of interface residues Tyr$^{442}$ → Leu$^{455}$, Leu$^{443}$ → Phe$^{456}$, Phe$^{460}$ → Tyr$^{473}$, and Asn$^{479}$ → Gln$^{493}$ may also change the affinity

Fig. 2. Dimerization interface of ACE2. (A) ACE2 dimerizes through two interfaces, the PD and the neck domain. The regions enclosed by the cyan and red dashed lines are illustrated in detail in (B) and (C), respectively. (B) The primary dimeric interface is through the neck domain in ACE2. Polar interactions are represented by red dashed lines. (C) A weaker interface between PDs of ACE2. The only interaction is between Gln$^{139}$ and Gln$^{175}$′, which are highlighted as spheres. The polar residues that may contribute to the stabilization of Gln$^{139}$ are shown as sticks. (D) The PDs no longer contact each other in the open state. Single-letter abbreviations for the amino acid residues used in the figures are as follows: C, Cys; D, Asp; E, Glu; F, Phe; H, His; K, Lys; L, Leu; M, Met; N, Asn; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.

Fig. 3. Overall structure of the RBD-ACE2-B0AT1 complex. (A) Cryo-EM map of the RBD-ACE2-B0AT1 complex. The overall reconstruction of the ternary complex at 2.9 Å is shown on the left. The inset shows the focused refined map of RBD. The color scheme is the same as that in Fig. 1B, with the addition of red and gold, which represent RBD protomers. (B) Overall structure of the RBD-ACE2-B0AT1 complex. The color scheme is the same as that in Fig. 1C. The glycosylation moieties are shown as sticks.
for ACE2 (Fig. 5C). At the C terminus of α1, Leu472 in the SARS-CoV-RBD is replaced by Phe486 in the SARS-CoV-2-RBD (Fig. 5D).

Discussion
Although ACE2 is a chaperone for B⁰AT1, our focus is on ACE2 in this study. With the stabilization by B⁰AT1, we elucidated the structure of full-length ACE2. B⁰AT1 is not involved in dimerization, suggesting that ACE2 may be a homodimer even in the absence of B⁰AT1. Further examination suggests that a dimeric ACE2 can accommodate two S protein trimers, each through a monomer of ACE2 (fig. S9). The trimeric structure of the S protein of SARS-CoV-2 was recently reported, with one RBD in an up conformation and two in down conformations (PDB 6VSB) (14). The PD clashes with the rest of the S protein when the ternary complex is aligned to the RBD in the up conformation. There is no clash when the complex is superimposed on RDB in the up conformation, with a RMSD of 0.98 Å over 126 pairs of Cα atoms, confirming that an up conformation of RDB is required to bind to the receptor (fig. S9) (14).

Cleavage of the S protein of SARS-CoV is facilitated by cathepsin L in endosomes, indicating a mechanism of receptor-mediated endocytosis (10). Further characterization is required to examine the interactions between ACE2 and the viral particle as well as the effect of cofactors on this process (25, 33). It remains to be investigated whether there is clustering between the dimeric ACE2 and trimeric S proteins, which may be important for invagination of the membrane and endocytosis of the viral particle, a process similar to other types of receptor-mediated endocytosis.

Cleavage of the C-terminal segment, especially residues 697 to 716 (fig. S4), of ACE2 by proteases, such as transmembrane protease serine 2 (TMPRSS2), enhances the S protein–driven viral entry (34, 35). Residues 697 to 716 form the third and fourth helices in the neck domain and map to the dimeric interface of ACE2. The presence of B⁰AT1 may block the access of TMPRSS2 to the cutting site on ACE2. The expression distribution of ACE2 is broader than that of B⁰AT1. In addition to kidneys and intestine, where B⁰AT1 is mainly expressed, ACE2 is also expressed in lungs and heart (27). It remains to be tested whether B⁰AT1 can suppress SARS-CoV-2 infection by blocking ACE2 cleavage. Enteric infections have been reported for SARS-CoV, and possibly also for SARS-CoV-2 (36, 37). B⁰AT1 has also been shown to interact with another coronavirus receptor, aminopeptidase N (APN or CD13) (38). These findings suggest that B⁰AT1 may play a regulatory role for the enteric infections of some coronaviruses.

Comparing the interaction interfaces of SARS-CoV-2-RBD and SARS-CoV-RBD with ACE2...
reveals some variations that may strengthen the interactions between SARS-CoV-2-RBD and ACE2 and other variations that are likely to reduce the affinity compared with SARS-CoV-RBD and ACE2. For instance, the change from Val^{164} to Lys^{377} may result in a tighter association because of the salt bridge formation between Lys^{377} and Asp^{363} of ACE2 (Figs. 4C and 5C). The change from Leu^{772} to Phe^{782} may also result in a stronger van der Waals contact with Met^{782} (Fig. 5D). However, replacement of Arg^{235} with Asn^{249} appears to weaken the interaction by eliminating one important salt bridge with Asp^{293} on ACE2 (Fig. 5B).

Our structural work reveals the high-resolution structure of full-length ACE2 in a dimeric assembly. Docking the S protein trimer onto the structure of the ACE2-CBD with the RBD of the S protein bound suggests simultaneous binding of two S protein trimers to an ACE2 dimer. Structure-based rational design of binders with enhanced affinities to either ACE2 or the S protein of the coronaviruses may facilitate development of decoy ligands or neutralizing antibodies for suppression of viral infection.

REFERENCES AND NOTES

1. N. Zhu et al., N. Engl. J. Med. 382, 727–733 (2020).
2. P. Zhou et al., Nature (2020).
3. T. M. Gallagher, M. J. Buchmeier, Virology 279, 371–374 (2000).
4. G. Simmons, P. Zsibra, S. Giezen, A. Heurich, S. Pöhlmann, Antiviral Res. 100, 605–614 (2013).
5. S. Beiluazard, V. C. Chu, G. R. Whittaker, Proc. Natl. Acad. Sci. U.S.A. 106, 5871–5876 (2009).
6. G. Simmons et al., Proc. Natl. Acad. Sci. U.S.A. 101, 4240–4245 (2004).
7. W. Song, M. Gui, X. Wang, Y. Xiang, PLoS Pathog. 14, e1007375 (2018).
8. F. Li, W. Li, M. Fanaz, S. C. Harrison, Science 309, 1864–1868 (2005).
9. J. K. Miller, G. R. Whittaker, Virus Res. 202, 120–134 (2015).
10. G. Simmons et al., Proc. Natl. Acad. Sci. U.S.A. 102, 11876–11881 (2005).
11. M. Hoffmann et al., bioRxiv 2020.01.31.929042 [Preprint].
12. L. Wu et al., Nature 426, 450–454 (2003).
13. K. Kuba et al., Nat. Med. 11, 875–879 (2005).
14. D. Wrapp et al., Science eabb2507 (2020).
15. M. Donoghue et al., Circ. Res. 87, E1–E9 (2000).
16. H. Zhang et al., bioRxiv 2020.01.30.927806 [Preprint].
17. K. K. Lei et al., J. Med. Virol. 82, 882–887 (2010).
18. L. S. Zisman et al., Circulation 108, 1707–1712 (2003).
19. M. K. Raizada, A. J. Ferreira, J. Cardiobioe. Pharmacol. 50, 112–119 (2007).
20. H. Zhang et al., J. Biol. Chem. 276, 1712–1739 (2001).
21. I. Hamming et al., J. Pathol. 212, 1–11 (2007).
22. R. N. Kirchdoerfer et al., Sci. Rep. 8, 15701 (2018).
23. P. Towler et al., J. Biol. Chem. 279, 17966–17970 (2004).
24. J. K. Miller et al., J. Biol. Chem. 279, 2896–2897 (2004).
25. H. F. Seow et al., Proc. Natl. Acad. Sci. U.S.A. 102, 21291–21296 (2005).
26. R. Kleta et al., Nat. Genet. 36, 999–1002 (2004).
27. R. A. Brier et al., J. Biol. Chem. 279, 24467–24476 (2004).
28. A. Perinat et al., Cell 103, 503 (2001).
29. J. I. S. Zisman et al., Circulation 108, 1707–1712 (2003).
30. G. Simmons et al., Proc. Natl. Acad. Sci. U.S.A. 101, 4240–4245 (2004).
31. L. Mastroberardino et al., Nature 426, 450–454 (2003).
32. L. Zhang et al., Proc. Natl. Acad. Sci. U.S.A. 102, 11876–11881 (2005).
33. J. K. Miller, G. R. Whittaker, Virus Res. 202, 120–134 (2015).
34. S. Belouzard et al., J. Virol. 88, 1293–1307 (2014).
35. C. Drosten et al., N. Engl. J. Med. 348, 1967–1976 (2003).
36. J. Jando, S. M. R. Camargo, B. Herzog, F. Verrey, PLOS ONE 12, e0184845 (2017).

ACKNOWLEDGMENTS

We thank the Cryo-EM Facility and Supercomputer Center of Westlake University for providing cryo-EM and computation support, respectively. This work was funded by the Natural Science Foundation of China (projects 31971123, 81902080, and 31903009), the Key R&D Program of Zhejiang Province (2020C04010), and the SARS-CoV-2 emergency project of the Science and Technology Department of Zhejiang Province (2020C03129). Author contributions: Q. Z. and R. Y. conceived the project, Q. Z. and R. Y. designed the experiments, and all authors performed the experiments. Q. Z., R. Y., Y. L., and Y. L. contributed to data analysis. Q. Z. and R. Y. wrote the manuscript. Competing interests: The authors declare no competing interests. Data and materials availability: Atomic coordinates and cryo EM maps for the ACE2-20F1 complex of closed conformation (whole structure and map, PDB 6M1D and EMD-30040; extracellular region map, EMD-30041; and TM region map, EMD-30045), the ACE2-B0AT1 complex of open conformation (PDB 6M1D and EMD-30041), and the complex of the RBD of SARS-CoV-2 with the ACE2-B0AT1 complex (whole structure and map, PDB 6M17 and EMD-30039; extracellular region map, EMD-30042; TM region map, EMD-30043; and ACE2-RBD interface map, EMD-30046) have been deposited in the Protein Data Bank (www.rcsb.org) and the Electron Microscopy Data Bank (www.ebi.ac.uk/pdbe/emdb/). Correspondence and requests for materials should be addressed to corresponding author Q. Z. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6485/1444/suppl/DC1

Materials and Methods: Figs. S1 to S9

Additional references (39–52)

MDAR Reproducibility Checklist

View request a protocol for this paper from Bio-protocol.