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

Renhong Yan1,2, Yuanyuan Zhang1,2, Yaning Li2,*, Lu Xia1,2, Yingying Guo1,2, Qiang Zhou1,2†

1Key Laboratory of Structural Biology of Zhejiang Province, Institute of Biology, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China. 2School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China. †Corresponding author. Email: zhouqiang@westlake.edu.cn

Angiotensin-converting enzyme 2 (ACE2) is the cellular receptor for SARS coronavirus (SARS-CoV) and the new coronavirus (SARS-CoV-2) that is causing the serious epidemic COVID-19. Here we present cryo-EM structures of full-length human ACE2, in the presence of a 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 Å, with a local resolution of 3.5 Å at the ACE2-RBD interface. The ACE2-B0AT1 complex is assembled as a dimer of heterodimers, with the Collectrin-like domain (CLD) of ACE2 mediating homo-dimerization. The RBD is recognized by the extracellular peptidase domain (PD) of ACE2 mainly through polar residues. These findings provide important insights to the molecular basis for coronavirus recognition and infection.

The SARS-CoV-2 is a positive-strand RNA virus that causes severe respiratory syndrome in human. The resulting outbreak of the coronavirus disease 2019 (COVID-19) has emerged as a severe epidemic, claiming more than two thousand lives worldwide between Dec 2019 and Feb 2020 (1, 2). The genome of SARS-CoV-2 shares about 80% identity with that of SARS-CoV and is about 96% identical to the bat coronavirus BatCoV RaTG13 (2).

In the case of SARS-CoV, the spike glycoprotein (S protein) on the virion surface mediates receptor recognition and membrane fusion (3, 4). During viral infection, the trimeric S protein is cleaved into S1 and S2 subunits and S1 subunits are released in the transition to the post-fusion conformation (4–7). S1 contains the receptor binding domain (RBD), which directly binds to the peptidase domain (PD) of ACE2 (8), while S2 is responsible for membrane fusion. When S1 binds to the host receptor ACE2, another cleavage site on S2 is exposed and is cleaved by host proteases, a process that is critical for viral infection (5, 9, 10). The S protein of SARS-CoV-2 may also exploit ACE2 for host infection (2, 11–13). A recent publication reported the structure of the S protein of SARS-CoV-2 and showed that the ectodomain of the SARS-CoV-2 S protein binds to the PD of ACE2 with a K_{d} of ~ 15 nM (14).

Although ACE2 is hijacked by some coronaviruses, its primary physiological role is in the maturation of angiotensin, a peptide hormone that controls vasoconstriction 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 angiotensin (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 have revealed the molecular details of the interaction between the RBD of S protein and PD of ACE2 (7, 8, 23, 24). Structural information on ACE2 is limited to the PD domain. The single-TM of ACE2 makes it challenging to determine the structure of the full-length protein.

ACE2 also functions as the chaperone for the 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 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 the chaperone 4F2hc interact through both extracellular and transmembrane domains (32). We reasoned that the structure of the full-length ACE2 may be revealed in the presence of B0AT1.

Here we report the cryo-EM structure of the full-length ACE2...
human ACE2-B0AT1 complex at an overall resolution of 2.9 Å and a complex between the RBD of SARS-CoV-2 and 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 exists as a dimer of heterodimers. Structural alignment of the RBD-ACE2-B0AT1 ternary complex with the S protein of the SARS-CoV-2 suggests that two S protein trimers can simultaneously bind to an ACE2 homodimer.

**Structural determination of ACE2-B0AT1 complex**

The full-length human ACE2 and B0AT1, with Strept and FLAG tags on their respective N terminus, were co-expressed in HEK293F cells and purified through tandem affinity resin and size exclusion chromatography. The complex was eluted in a single mono-disperse 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 material. A 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 extracellular domains improved to 2.7 Å, while 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 19-768 that contain the PD (residues 19-615) and the CLD (residues 616-768) that consists of a small extracellular domain, a long linker, and the single TM helix (Fig. 1C). Between the PD and the TM is a ferredoxin-like fold domain, we referred to this as the Neck domain (residues 616-726) (Fig. 1C and fig. S4). Homo-dimerization is entirely mediated by ACE2, which is sandwiched by B0AT1. Both the PD and Neck domain contribute to dimerization, while each B0AT1 interacts with the Neck and the TM helix in the adjacent ACE2 (Fig. 1C). The extracellular domain is highly glycosylated, with 7 and 5 glycosylation sites on each ACE2 and B0AT1 monomer, respectively.

During classification, another subset with 143,857 particles was processed to an overall resolution of 4.5 Å. While 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 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). We will refer to the two ACE2 protomers as A and B, with residues in protomer B labeled as residue'. Extensive polar interactions are mapped to the interface between the 2nd (residues 636-658) and the 4th (residues 708-717) helices of the Neck domain (Fig. 2B). Arg652 and Arg710 in ACE2-A form cation-π interactions with Tyr641 and Tyr633' in ACE2-B. Meanwhile, Arg652 and Arg710 are respectively hydrogen-bonded (H-bond) 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 which is stabilized by a disulfide bond between Cys133 and Cys141 as well as multiple intra-loop 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 the SARS-CoV-2, we purchased 0.2 mg recombinantly expressed and purified RBD-mFc of the SARS-CoV-2 (for simplicity, we will refer it as RBD) 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, which allowed us to reach a resolution of 3.5 Å for the RBD, supporting reliable modelling and analysis of the interface (Fig. 3, figs. S5 to S7, and table S1).

**Interface between RBD and ACE2**

As expected, each PD accommodates one RBD (Fig. 3B). The overall interface is similar to that between the 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 RBD.

The contact can be divided into three clusters. The two
ends of the “bridge” interact with the amino (N) and carboxyl (C) termini of the α1 helix as well as small areas on the α2 helix and Loop 3-4. The middle segment of α1 reinforces the interaction by engaging two polar residues (Fig. 4A). At the N terminus of α1, Gln498, Thr500, and Asn501 of the RBD form a network of hydrogen bonds (H-bonds) with Tyr41, Glu42, Lys353, and Arg357 from ACE2 (Fig. 4B). In the middle of the “bridge”, Lys417 and Tyr453 of the RBD interact with Asp30 and His34 of ACE2, respectively (Fig. 4C). At the C terminus of α1, Gln474 of RBD is H-bonded to Gln24 of ACE2, while Phe486 of RBD interacts with Met82 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 ACE2-2PD (PDB code: 2A9F) 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 squared deviation (RMSD) of 0.68 Å over 139 pairs of Cα 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 α1, the following variations, Arg426→Asn439, Tyr484→Gln498, and Thr487→Asn501, 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 Val404 in the SARS-CoV-RBD with Lys417 in the SARS-CoV-2-RBD. In addition, from SARS-CoV-RBD to SARS-CoV-2-RBD, the substitution of interface residues, Tyr442→Leu455, Leu443→Phe456, Phe460→Tyr473, and Asn479→Gln493, may also change the affinity 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 chaperon for B^4AT1, our focus is on ACE2 in the absence of B^4AT1. With the stabilization by B^4AT1, we elucidated the structure of the full-length ACE2. B^4AT1 is not involved in dimerization suggesting that ACE2 may be a homodimer even in the absence of B^4AT1. Further examination suggests 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 “up” and two in “down” conformations (PDB code: 6VSB) (14). PD clashes with the rest of the S protein when the ternary complex is aligned to the RBD of the “down” conformation. There is no clash when the complex is superimposed to the “up” RBD, with RMSD of 0.98 Å over 126 pairs of Cα atoms, confirming that an “up” conformation of RBD 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 co-factors 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 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 697-716 form the 3rd and 4th helices in the Neck domain and map to the dimeric interface of ACE2. The presence of B^4AT1 may block the access of TMPRSS2 to the cutting site on ACE2. The expression distribution of ACE2 is broader than B^4AT1. In addition to kidneys and intestine, where B^4AT1 is mainly expressed, ACE2 is also expressed in lungs and heart (27). It remains to be tested whether B^4AT1 can suppress SARS-CoV-2 infection by blocking ACE2 cleavage. Enteric infections have been reported for the SARS-CoV, and possibly also for the SARS-CoV-2 (36, 37). B^4AT1 has also been shown to interact with another coronavirus receptor, aminopeptidase N (APN or CD13) (38). These findings suggest that B^4AT1 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 others likely to reduce the affinity compared to SARS-CoV-RBD and ACE2. For instance, the change from Val404 to Lys317 may result in tighter association because of the salt bridge formation between Lys317 and Asp30 of ACE2 (Figs. 4C and 5C). Change of Leu472 to Phe486 may also make stronger van der Waals contact with Met82 (Fig. 5D). However, replacement of Arg426 to Asn439 appears to weaken the interaction by losing one important salt bridge with Asp329 on ACE2 (Fig. 5B).

In summary, our structural work reveals the high-resolution structures of full-length ACE2 in a dimeric assembly. Docking the S protein trimer onto the structure of the ACE2 dimer with RBD of 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**


**Fig. 1. Overall structure of the ACE2-B^0AT1 complex.** (A) Representative SEC purification profile of the full-length human ACE2 in complex with B^0AT1. (B) Cryo-EM map of the ACE2-B^0AT1 complex. The map is generated by merging the focused refined maps shown in fig. S2. (C) Cartoon representation of the atomic model of the ACE2-B^0AT1 complex. The glycosylation moieties are shown as sticks. The complex is colored by subunits, with the protease domain (PD) and the Collectrin-like domain (CLD) in one ACE2 protomer colored cyan and blue, respectively. (D) An open conformation of the ACE2-B^0AT1 complex. The two PDs, which contact each other in the “closed” conformation, are separated in the “open” conformation.
**Fig. 2. Dimerization interface of ACE2.** (A) ACE2 dimerizes via two interfaces, the PD and the Neck domain. (B) The primary dimeric interface is between the Neck domain in ACE2. Polar interactions are represented by dashed, red lines. (C) A weaker interface between PDs of ACE2. The only interaction is between Gln139 and Gln175′, which are highlighted as spheres. The polar residues that may contribute to the stabilization of Gln139 are shown as sticks. (D) The PDs no longer contact each other in the open state.

**Fig. 3. Overall structure of the RBD-ACE2-B⁰AT1 complex.** (A) Cryo-EM map of the RBD-ACE2-B⁰AT1 complex. Left: Overall reconstruction of the ternary complex at 2.9 Å. Inset: focused refined map of RBD. (B) Overall structure of the RBD-ACE2-B⁰AT1 complex. The complex is colored by subunits, with the protease domain (PD) and the Collectrin-like domain (CLD) colored cyan and blue in one of the ACE2 protomers, respectively. The glycosylation moieties are shown as sticks.
Fig. 4. Interactions between SARS-CoV-2-RBD and ACE2. (A) The protease domain (PD) of ACE2 mainly engages the α1 helix in the recognition of the RBD. The α2 helix and the linker between β3 and β4 also contribute to the interaction. Only one RBD-ACE2 is shown. (B to D) Detailed analysis of the interface between SARS-CoV-2-RBD and ACE2. Polar interactions are indicated by red, dashed lines.
Fig. 5. Interface comparison between SARS-CoV-2-RBD and SARS-CoV-RBD with ACE2. (A) Structural alignment for the SARS-CoV-2-RBD and SARS-CoV-RBD. The complex structure of the ACE2 PD and the SARS-CoV-RBD (PDB code: 2AJF) is superimposed to our cryo-EM structure of the ternary complex relative to the RBDs. The boxed regions are illustrated in detail in (B) to (D). SARS-CoV-2-RBD and the PD in our cryo-EM structure are colored orange and cyan, respectively; SARS-CoV-RBD and its complexed PD are colored green and gold, respectively. (B to D) Variation of the interface residues between SARS-CoV-2-RBD (labeled brown) and SARS-CoV-RBD (labeled green).
Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2
Renhong Yan, Yuanyuan Zhang, Yaning Li, Lu Xia, Yingying Guo and Qiang Zhou

published online March 4, 2020

ARTICLE TOOLS
http://science.sciencemag.org/content/early/2020/03/03/science.abb2762

SUPPLEMENTARY MATERIALS
http://science.sciencemag.org/content/suppl/2020/03/03/science.abb2762.DC1

REFERENCES
This article cites 52 articles, 14 of which you can access for free
http://science.sciencemag.org/content/early/2020/03/03/science.abb2762#BIBL

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service