PNAS RESEARCH ARTICLE | MICROBIOLOGY OPEN ACCESS

Structural basis for mouse receptor recognition by bat SARS2-like coronaviruses

Wei Zhang[a,b,1], Ke Shi[c,1], Fu-Chun Hsueh[a,b], Alise Mendoza[a,b], Gang Ye[a,b], Linfen Huang[a,b], Stanley Perlman[d], Hideki Aihara[c,2], and Fang Li[a,b,2]

Affiliations are included on p. 8.

Edited by Brenda G. Hogue, Arizona State University, Tempe, AZ; received December 21, 2023; accepted June 11, 2024 by Editorial Board Member Peter Palese

The animal origin of SARS-CoV-2 remains elusive, lacking a plausible evolutionary narrative that may account for its emergence. Its spike protein resembles certain segments of BANAL-236 and RaTG13, two bat coronaviruses considered possible progenitors of SARS-CoV-2. Additionally, its spike contains a furin motif, a common feature of rodent coronaviruses. To explore the possible involvement of rodents in the emergence of SARS-CoV-2 spike, we examined the crystal structures of the spike receptor-binding domains (RBDs) of BANAL-236 and RaTG13 each complexed with mouse receptor ACE2. Both RBDs have residues at positions 493 and 498 that align well with two virus-binding hotspots on mouse ACE2. Our biochemical evidence supports that both BANAL-236 and RaTG13 spikes can use mouse ACE2 as their entry receptor. These findings point to a scenario in which these bat coronaviruses may have coinfected rodents, leading to a recombination of their spike genes and a subsequent acquisition of a furin motif in rodents, culminating in the emergence of SARS-CoV-2.

SARS-CoV-2 | spike protein | receptor-binding domain (RBD) | angiotensin-converting enzyme 2 (ACE2) | X-ray crystallography

Significance

Clarifying the evolutionary origins of SARS-CoV-2 is critical for understanding the recent COVID-19 pandemic and for preventing future coronavirus outbreaks. Up to now, there has been no plausible evolutionary narrative that may account for the emergence of SARS-CoV-2. This study provides structural and biochemical evidence that two bat SARS2-like coronaviruses, BANAL-236 and RaTG13, can use mouse ACE2 as their entry receptor. Moreover, the furin motif is a common feature of rodent coronaviruses. Therefore, it is possible that bat coronaviruses may have coinfected rodents, where their spike genes recombined and subsequently evolved the furin motif, leading to the spike protein observed in SARS-CoV-2. Our research provides the foundational structural and biochemical understanding for further studies into the animal origins of SARS-CoV-2.

Tracing the animal origin of SARS-CoV-2, which caused the COVID-19 pandemic, is critical for comprehending the recent COVID-19 pandemic and for preventing future coronavirus pandemics. This task has been hindered by difficulties in obtaining wildlife surveillance data, especially from the prepandemic era. Up to now, there has been no reasonable evolutionary narrative that may account for the emergence of SARS-CoV-2. The structural interactions between coronavirus spike proteins and their host receptors have been one of the best platforms for understanding animal origin and host range of coronaviruses (1–3). This approach was instrumental in tracing the animal origin of SARS-CoV-1 (3–6), which caused the 2002–2003 SARS epidemic and is related to SARS-CoV-2 (7, 8), and in tracking the possible animal origin of the SARS-CoV-2 Omicron variant (9, 10), which abruptly emerged in the middle of the pandemic with many mutations (11, 12). In this study, we have applied this "structure-guided virus hunting" approach to investigate the potential animal origin of the prototypic SARS-CoV-2 strain.

The SARS-CoV-2 spike protein guides viral entry into host cells (2). It contains a receptor-binding subunit S1 and a membrane-fusion subunit S2. S1 contains a receptor-binding domain (RBD), an N-terminal domain (NTD), and two subdomains (SD1 and SD2) (Fig. 1 *A* and *B*). During cell entry, the RBD binds to its host receptor, the spike is cleaved at the S1/S2 boundary by a host protease, and then, S1 dissociates and S2 undergoes a dramatic structural change to fuse viral and host membranes. Both SARS-CoV-2 and SARS-CoV-1 RBDs use human ACE2 as their receptor (5, 13–16). Their RBDs both contain a core structure and a receptor-binding motif (RBM) that directly interacts with ACE2. We previously identified several molecular features of SARS-CoV-2 spike that contribute to the virus' global spread (15, 17, 18). Two of these features are i) SARS-CoV-2 RBD binds to human ACE2 more strongly than SARS-CoV-1 RBD does; ii) SARS-CoV-2 spike, but not SARS-CoV-1 spike, contains a furin motif at the S1/S2 boundary that allows itself to be cleaved during viral maturation. The distinct molecular features of SARS-CoV-2 spike imply that SARS-CoV-1 unlikely is a direct progenitor of SARS-CoV-2. This inference is due to the considerable differences in the RBMs of the two spikes and the absence of the furin motif in SARS-CoV-1 spike.

Bats are known to be primary reservoirs of coronaviruses, along with many other virus families, that have spilled over to humans and other species (19, 20). In Southeastern Asia, numerous SARS2-like coronaviruses have been identified in bats. Of these, BANAL-236 and RaTG13 have spike proteins that closely resemble SARS-CoV-2 spike (21, 22). Particularly, BANAL-236 and SARS-CoV-2 RBDs show remarkable similarity,

Author contributions: W.Z., H.A., and F.L. designed research; W.Z., K.S., F.-C.H., A.M., G.Y., and L.H. performed research; W.Z., K.S., F.-C.H., A.M., G.Y., L.H., S.P., H.A., and F.L. analyzed data; and F.L. wrote the paper.

The authors declare no competing interest.

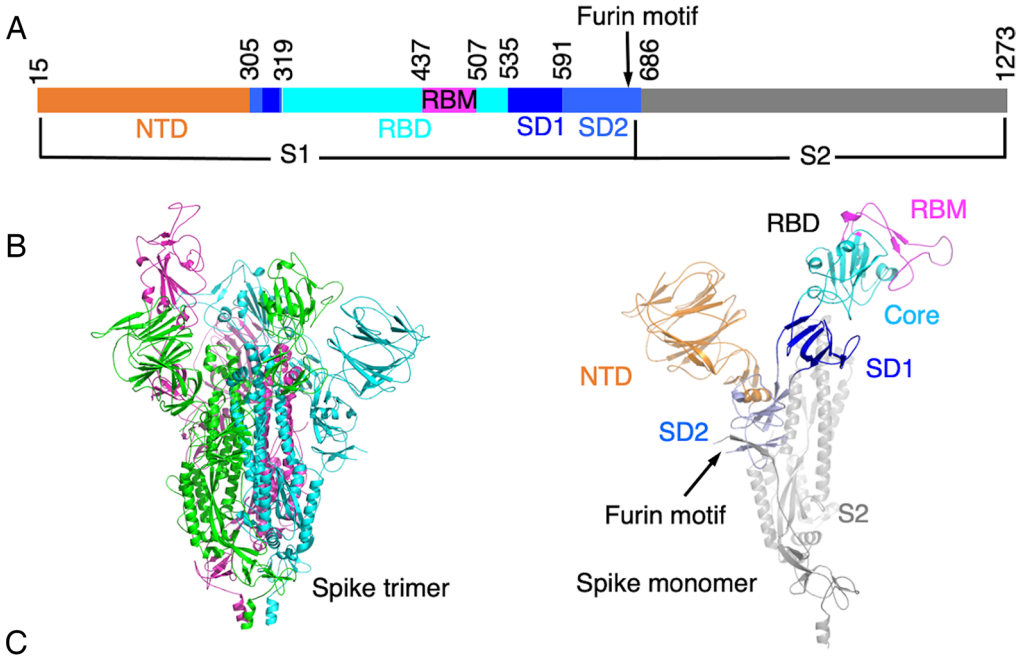
This article is a PNAS Direct Submission. B.G.H. is a guest editor invited by the Editorial Board.

[1]W.Z. and K.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: aihar001@umn.edu or lifang@umn.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2322600121/-/DCSupplemental.

Published July 31, 2024.



C

Sequence identity with SARS-CoV-2	RBM	RBD	NTD	S1	S2	Spike
BANAL-236	**97.2**	**96.8**	66.7	82.8	99.8	90.7
RaTG13	76.1	89.9	**98.7**	95.5	99.8	97.5
Recombination product	97.2	96.8	98.7	98.2	99.8	**98.7**

D

```
                437  441         451         461
SARS-CoV-2      NSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST
BANAL-236       NSNN LDSKVGGNYN YLYRLFRKSN LKPFERDIST
RaTG13          NSKH IDAKEGGNFN YLYRLFRKAN LKPFERDIST
                **:: :*:* ***:* *********:* **********

                471         481         491         501
SARS-CoV-2      EIYQAGSTPC NGVEGFNCYF PLQSYGFQPT NGVGYQP
BANAL-236       EIYQAGSTPC NGVEGFNCYF PLKSYGFHPT NGVGYQP
RaTG13          EIYQAGSKPC NGQTGLNCYY PLYRYGFYPT DGVGHQP
                *******.** **  *:***: **  *** ** :***:**
```

Fig. 1. Comparison of spike protein sequences from SARS-CoV-2 and two bat SARS2-like coronaviruses, BANAL-236 and RaTG13. (*A*) Schematic drawing of SARS-CoV-2 spike ectodomain (prototypic strain). S1: receptor-binding subunit. S2: membrane-fusion subunit. NTD: N-terminal domain of S1. RBD: receptor-binding domain of S1. RBM: receptor-binding motif of RBD. SD1: subdomain 1 of S1. SD2: subdomain 2 of S1. Furin motif: cleavage site for the furin protease. (*B*) *Left*: Structure of SARS-CoV-2 spike trimer. PDB ID: 7TGX. The three spike subunits are colored in green, magenta, and cyan, respectively. *Right*: Structure of SARS-CoV-2 spike monomer. PDB ID: 7TGX. Different regions of the spike monomer are colored in the same way in panel (*A*). (*C*) Sequence identities between SARS-CoV-2 spike and BANAL-236 spike and between SARS-CoV-2 spike and RaTG13 spike in different spike regions. (*D*) Sequence alignment of the RBMs from SARS-CoV-2, BANAL-236, and RaTG13 spikes. The residues involved in binding to human ACE2 are highlighted in blue, with the two key residues, 493 and 498, emphasized in red for this study. For clarity and consistency throughout this manuscript, the numbering of the RBD residues in BANAL-236 and RaTG13 has been adjusted to align with the numbering used for SARS-CoV-2 RBD.

providing overwhelming evidence that SARS-CoV-2 RBD has a bat origin (22). However, both BANAL-236 and RaTG13 spikes contain regions that are significantly different from SARS-CoV-2. Moreover, the furin motif, a feature not found in BANAL-236 or RaTG13 spike, is rare among bat coronaviruses (23). As for the whole genomes, both BANAL-236 and RaTG13 are closely related to SARS-CoV-2 outside of their spike genes (21, 22). These observations position BANAL-236 and RaTG13 as potential, albeit not direct, progenitors of SARS-CoV-2.

Rodents, like bats, are major reservoirs of viruses because of their vast diversity and clustering habitats (20, 24). Historically, rodents have been responsible for transmitting many viruses, coronaviruses included, to humans and other animals. For instance, the human coronavirus OC43 is believed to have come from rodents (20). Additionally, coronaviruses typically found in bats have also been identified in rodents (25, 26). A particularly concerning finding from a 2022 study showed that 16.5% of rats in New York City had been infected by SARS-CoV-2 (27). The furin motif in the spikes is a common feature of rodent coronaviruses as a majority of them contain a predicted furin motif (23). Moreover, genomic recombination of different coronaviruses coinfecting a host is ubiquitous (28), particularly in rodents (29, 30). The possibility of BANAL-236 and RaTG13 coinfecting rodents could provide insights into the origins of SARS-CoV-2. Previous

research has established receptor recognition as a critical determinant of the host ranges of coronaviruses (1, 2). This study examined the structural interactions between the bat SARS2-like coronavirus spikes and their mouse receptor ACE2. Understanding these interactions is key to unraveling how bat coronaviruses have the potential to coinfect rodents.

Results

To understand the relationships between the spike sequences of SARS-CoV-2 and bat SARS2-like coronaviruses, we compared their spike sequences in different segments. Our analysis shows that BANAL-236 and SARS-CoV-2 spikes are similar across all segments except for the NTD (Fig. 1*C*). Notably, within the RBM, they differ by only two amino acids at positions 493 and 498 (Fig. 1*D*). In contrast, RaTG13 and SARS-CoV-2 spikes are similar across all segments except for the RBD (Fig. 1*C*). Among the ACE2-binding residues, six differ between RaTG13 and SARS-CoV-2, including residues 493 and 498 (Fig. 1*D*). Consequently, our study has concentrated on how the variations at positions 493 and 498 in the RBMs of BANAL-236 and RaTG13 affect their binding with mouse ACE2.

We evaluated the capabilities of mouse ACE2 to serve as the entry receptor for bat SARS2-like coronaviruses, using human ACE2 as a comparison. We first measured the binding affinity between BANAL-236 RBD and mouse ACE2 using surface plasmon resonance (SPR) (Fig. 2*A* and *SI Appendix*, Table S1). The result showed that mouse ACE2 binds to BANAL-236 RBD less tightly than human ACE2 does, but mouse ACE2 is still a good receptor for BANAL-236 RBD with a K_d of 5.5×10^{-8} M (Fig. 2*A* and *SI Appendix*, Table S1). We then determined the binding affinity between RaTG13 spike and mouse ACE2 using SPR (Fig. 2*B* and *SI Appendix*, Table S1). The result showed that mouse ACE2 binds to RaTG13 spike more tightly than human ACE2 does. We also determined the efficiency of BANAL-236 and RaTG13 spikes in mediating viral entry into mouse ACE2-expressing cells using pseudovirus entry assays (Fig. 2*C*). Our result showed that both BANAL-236 and RaTG13 spikes mediate viral entry into mouse ACE2-expressing cells more efficiently than they do human ACE2-expressing cells (Fig. 2*C*). Moreover, only BANAL-236 spike, but not RaTG13 spike, mediates viral entry into human ACE2-expressing cells efficiently (Fig. 2*C*). Together, these biochemical data demonstrated that mouse ACE2 is an effective receptor for both BANAL-236 and RaTG13 spikes, whereas

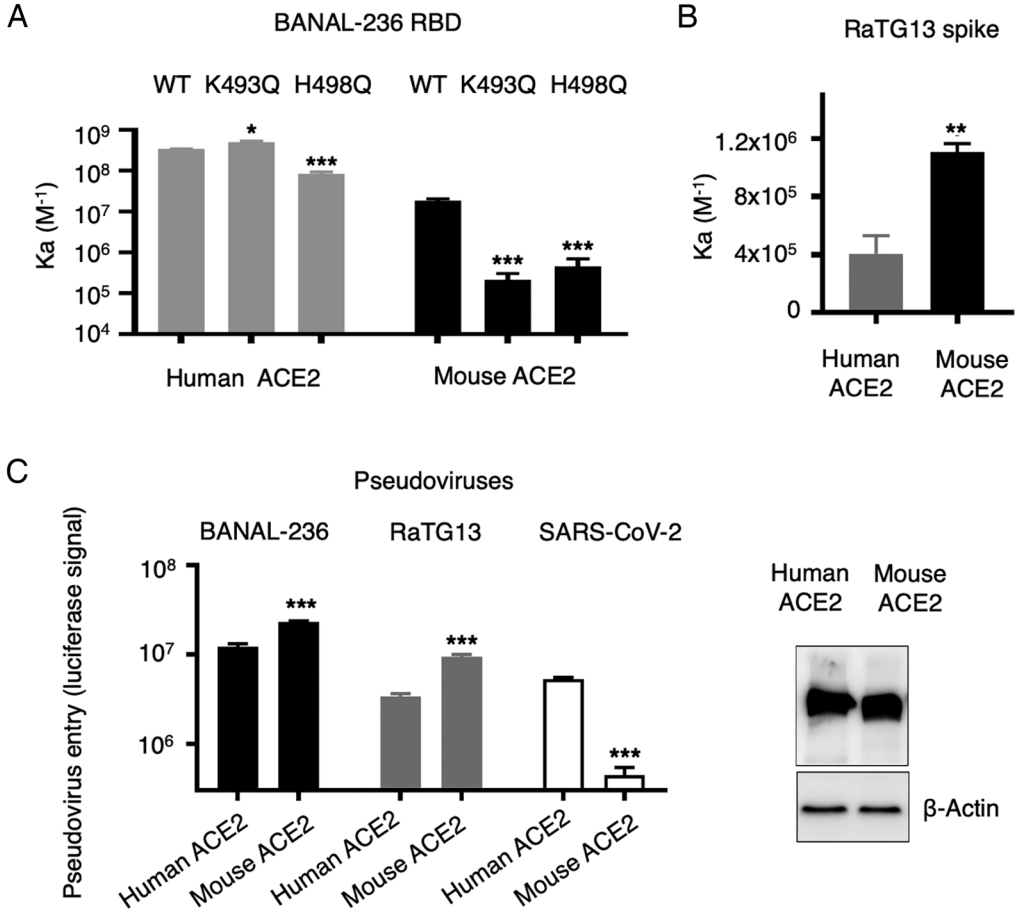


Fig. 2. Interactions between the spikes from SARS-CoV-2, BANAL-236, and RaTG13 and ACE2 from human and mouse. (*A*) SPR assay for the binding of BANAL-236 RBD (and its mutants) to ACE2 (from human or mouse). ACE2-Fc was coated to a protein A chip in a fixed direction, and individual RBDs flowed through. The data are presented as mean ± SEM (n = 3). A Student's two-tailed *t* test was performed to analyze the statistical difference between wild-type (WT) BANAL-236 RBD on the left and each of the mutant RBDs against the same ACE2; the results were labeled on top of each bar. ***$P < 0.001$; *$P < 0.05$. (*B*) SPR assay for the binding of RaTG13 spike ectodomain to ACE2 (from human or mouse). The assay was performed in the same way as in (*A*). The data are presented as mean ± SEM (n = 3). A Student's two-tailed *t* test was performed to analyze the statistical difference between human ACE2 and mouse ACE2. **$P < 0.01$. (*C*) Coronavirus pseudovirus entry into human ACE2 or mouse ACE2-expressing cells. Retroviruses pseudotyped with coronavirus spike (SARS-CoV-2, BANAL-236, or RaTG13) were used to infect HEK293T cells expressing human ACE2 or mouse ACE2. Pseudovirus entry efficiencies were characterized as luciferase signal accompanying entry and were normalized against the expression level of human ACE2 and mouse ACE2 (*Right*). The data are presented as mean ± SEM (n = 3). A Student's two-tailed *t* test was performed to analyze the statistical difference between human ACE2 and mouse ACE2 for each coronavirus pseudovirus; the results were labeled on top of each bar. ***$P < 0.001$.

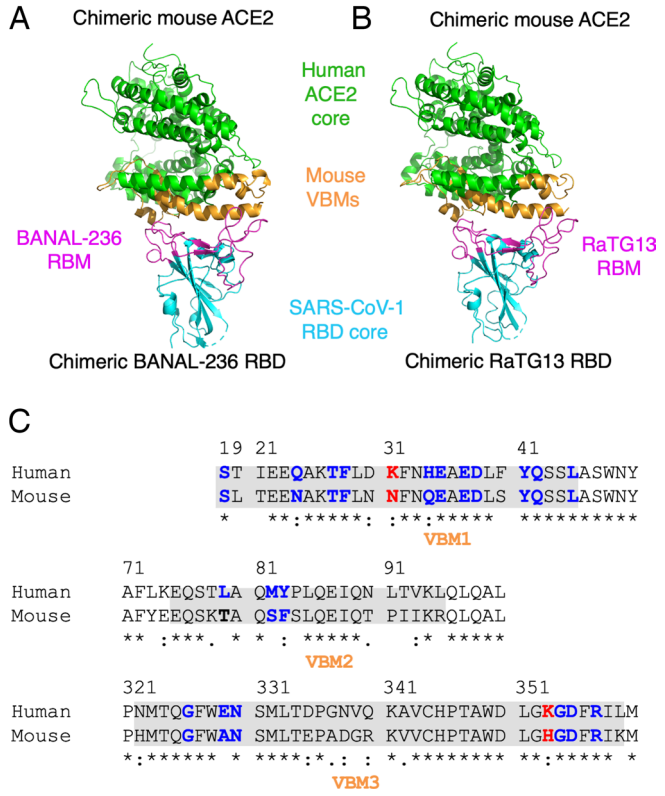


Fig. 3. Overall structures of chimeric bat SARS2-like coronavirus RBDs complexed with chimeric mouse ACE2. (*A*) Overall structure of chimeric BANAL-236 RBD complexed with chimeric mouse ACE2. The chimeric BANAL-236 RBD contains the core structure (in cyan) from SARS-CoV-1 RBD and RBM (in magenta) from BANAL-236 RBD. The chimeric mouse ACE2 contains the core structure (in green) from human ACE2 and three VBMs (in orange) from mouse ACE2. (*B*) Overall structure of chimeric RaTG13 RBD complexed with chimeric mouse ACE2. (*C*) Sequence alignment between human ACE2 and mouse ACE2 in three VBMs (shaded). The centers for hotspot-31 and hotspot-353 (i.e., residues 31 and 353, respectively) are both labeled in red. Residues that directly contact SARS-CoV-2 RBD are labeled in blue. Asterisks indicate positions that have a single, fully conserved residue. Colons indicate positions that have strongly conserved residues. Periods indicate positions that have weakly conserved residues.

human ACE2 is an effective receptor for BANAL-236 spike, but not for RaTG13 spike.

To grasp how mouse ACE2 acts as an effective receptor for BANAL-236 and RaTG13, we determined the crystal structures of BANAL-236 and RaTG13 RBDs, each complexed with mouse ACE2 (Fig. 3 *A* and *B*). We previously established a convenient structural platform for studying detailed interactions between the RBDs from SARS-CoV-2, SARS-CoV-1, and their related coronaviruses and ACE2 from human and other animal species. This platform combines chimeric RBDs (containing the core structure from SARS-CoV-1 and RBM from other related coronaviruses) with chimeric ACE2 (containing the core structure from human ACE2 and virus-binding motifs (VBMs) from other ACE2 molecules) (Figs. 1*D* and 3*C*). The complexes of these chimeric RBDs and chimeric ACE2 crystallize in the same way as the complex of SARS-CoV-1 RBD and human ACE2 but reveal structural interactions between other related coronaviruses and other ACE2 molecules. Using this strategy, we previously determined the structural interfaces between various RBDs and various ACE2 molecules (6, 9, 10, 15, 31). Importantly, we demonstrated that the structural interface of the chimeric SARS-CoV-2 RBD and human ACE2 mirrors that of the WT SARS-CoV-2 RBD and human ACE2, confirming the reliability and precision of our chimeric protein structural platform (15). In this study, we determined the structural interfaces between BANAL-236 RBD and mouse ACE2 and between RaTG13 RBD and mouse ACE2 (Fig. 3 *A* and *B* and *SI Appendix*, Table S2). The electron density maps of these structures are well-defined (*SI Appendix*, Fig. S1), which verifies the accuracy of the structural data. It should be highlighted that this study is the first to describe the structural interfaces between bat coronavirus RBDs and mouse ACE2, setting it apart from previous research that described the structural interface between RaTG13 RBD and human ACE2 (32).

We previously identified three virus-binding hotspots at the RBD/ACE2 interfaces that are critical for the binding of both SARS-CoV-2 and SARS-CoV-1 (6, 9, 10, 15, 31). "Hotspot-31" centers around Lys31 from human ACE2, "hotspot-353" centers around Lys353 from human ACE2, and "hotspot-ridge" centers around a receptor-binding ridge on the RBD. These three hotspots largely determine the host ranges of SARS-CoV-2 and SARS-CoV-1. Interestingly, the previously discussed two residues (493 and 498) in BANAL-236 and RaTG13 RBMs are located at hotspot-31 and hotspot-353, respectively. Consequently, we explored how mutations at these two RBM residues influence the stability of the two hotspots.

At hotspot-31, ACE2 from human and many other host species all contain a lysine at the 31 position in their VBMs (Fig. 3*C*) (10). Coronavirus RBD binding to these ACE2 molecules mainly concerns how to accommodate $Lys31^{VBM}$ in a relatively hydrophobic environment at the RBD/ACE2 interface. SARS-CoV-2 RBM contains a glutamine at the 493 position, which stabilizes the SARS-CoV-2/human ACE2 interface by forming two hydrogen bonds with $Lys31^{VBM}$ and $Glu35^{VBM}$ (Fig. 4*A*). BANAL-236 RBM contains a lysine at the 493 position, which forms a salt bridge with $Glu35^{VBM}$ ($Lys31^{VBM}$ also forms a salt bridge with $Glu35^{VBM}$) (Fig. 4*B*). The K493Q mutation introduced to BANAL-236 RBD slightly increased its binding affinity for human ACE2 (Fig. 2*A*), suggesting that $Gln493^{VBM}$ stabilizes the BANAL-236/human

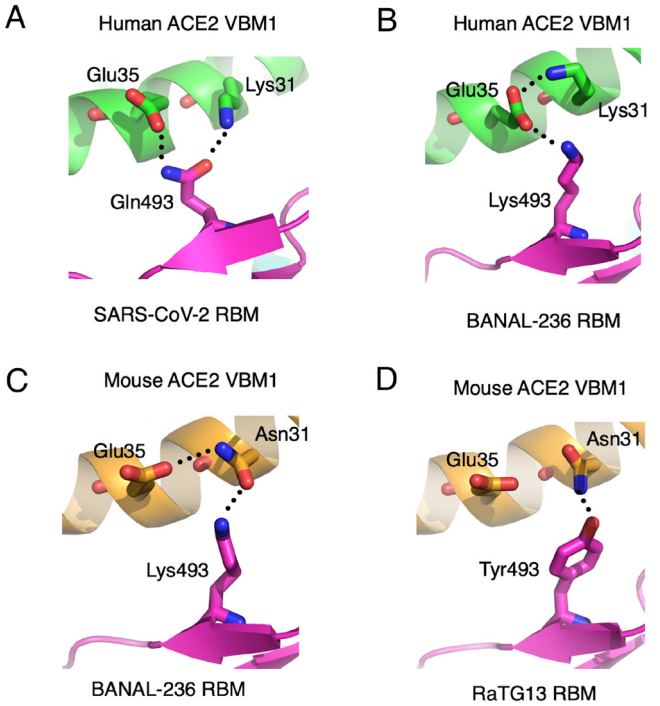


Fig. 4. Structural details at hotspot-31. (*A*) Interface between SARS-CoV-2 RBM and human VBM1. PDB ID: 6VW1. (*B*) Interface between BANAL-236 RBM and human ACE2 VBM1. PDB ID: 7PKI. (*C*) Interface between BANAL-236 RBM and mouse ACE2 VBM1. (*D*) Interface between RaTG13 RBM and mouse ACE2 VBM1. Dotted lines indicate hydrogen bonds or salt bridges.

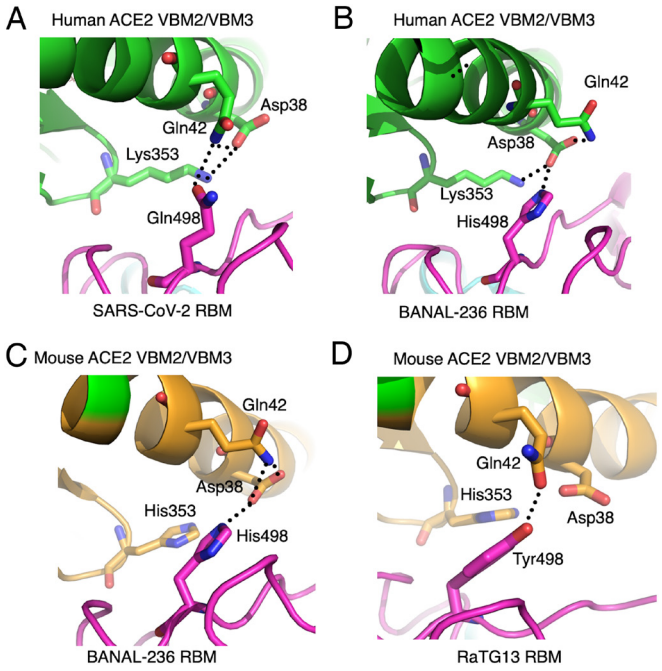


Fig. 5. Structural details at hotspot-353. (*A*) Interface between SARS-CoV-2 RBM and human ACE2 VBM2/VBM3. PDB ID: 6VW1. (*B*) Interface between BANAL-236 RBM and human ACE2 VBM2/VBM3. PDB ID: 7PKI. (*C*) Interface between BANAL-236 RBM and mouse ACE2 VBM2/VBM3. (*D*) Interface between RaTG13 RBM and mouse ACE2 VBM2/VBM3. Dotted lines indicate hydrogen bonds or salt bridges.

ACE2 interface slightly better than Lys493^{VBM} does. Mice are one of a few species whose ACE2 contain an asparagine at the 31 position (10), which forms a hydrogen bond with Lys493^{VBM} from BANAL-236 (Fig. 4*C*). The K493Q mutation introduced to BANAL-236 RBD dramatically reduced its binding affinity for mouse ACE2 (Fig. 2*A*), suggesting that Lys493^{VBM} stabilizes the BANAL-236/mouse ACE2 interface much better than Gln493^{VBM} does. RaTG13 RBM contains a tyrosine at the 493 position, which stabilizes the RaTG13/mouse ACE2 interface by forming a hydrogen bond with Asn31^{VBM} (Fig. 4*D*). The Y493Q mutation introduced to RaTG13 RBD dramatically reduced its binding affinity for mouse ACE2 (32), suggesting that Tyr493^{VBM} stabilizes the RaTG13/mouse ACE2 interface much better than Gln493^{VBM} does. Collectively, these data reveal that the 493 residues in both BANAL-236 and RaTG13 RBMs are well suited to interact with mouse ACE2 at hotspot-31.

At hotspot-353, ACE2 from human and many other host species all contain a lysine at the 353 position in their VBMs as well (Fig. 3*C*) (10). Lys353^{VBM} is buried in a hydrophobic tunnel and at the end of the tunnel it forms a salt bridge with Asp38^{VBM} (Fig. 5*A*). Asp38^{VBM} is further stabilized by a hydrogen bond network involving Gln42^{VBM}. SARS-CoV-2 RBM contains a glutamine at the 498 position, which stabilizes the SARS-CoV-2/human ACE2 interface by being part of the above hydrogen bond network. BANAL-236 RBM contains a histidine at the 498 position, which also stabilizes the BANAL-236/human ACE2 interface by forming a hydrogen bond with Asp38^{VBM} (Fig. 5*B*). The H498Q mutation introduced to BANAL-236 RBD reduced its binding affinity for human ACE2, suggesting that His498^{VBM} stabilizes the BANAL-236/human interface better than Gln498^{VBM} does (Fig. 2*A*). Mice are one of a few species whose ACE2 contain a histidine at the 353 position that cannot directly interact with Asp38^{VBM}, and hence, mouse ACE2 is an ineffective receptor for prototypic SARS-CoV-2 (10). However, His498^{VBM} can stabilize the BANAL-236/mouse ACE2 interface by forming a hydrogen bond with Asp38^{VBM} (Fig. 5*C*). The H498Q mutation introduced to BANAL-236 RBD significantly reduced its binding affinity for mouse ACE2 (Fig. 2*A*), suggesting that His498^{VBM} stabilizes the BANAL-236/mouse ACE2 interface better than Gln498^{VBM} does. This explains why mouse ACE2 is a much more effective receptor for BANAL-236 RBD (which contains His498^{VBM}) than for SARS-CoV-2 RBD (which contains Gln498^{VBM}). Residue 498 in RaTG13 RBM is a tyrosine, which stabilizes the RaTG13/mouse ACE2 interface by forming a hydrogen bond with Gln42^{VBM} (Fig. 5*D*). The Y498Q mutation introduced to RaTG13 RBD dramatically reduced its binding affinity for mouse ACE2 (32), suggesting that Tyr498^{VBM} stabilizes the RaTG13/mouse ACE2 interface much better than Gln498^{VBM} does. Collectively, these data reveal that the 498 residues in both BANAL-236 and RaTG13 RBMs are well suited to interact with mouse ACE2 at hotspot-353.

Our above structural and biochemical analyses revealed that both BANAL-236 and RaTG13 are capable of using mouse ACE2 as their receptor for potential coinfection of mouse cells. To understand how BANAL-236 and RaTG13 spike genes could recombine to produce SARS-CoV-2 spike gene, we investigated potential recombination sites between BANAL-236 and RaTG13 spike genes using a variety of analyses including Simplot 3.5.1, RDP (7.782×10^{-60}), GENECONV (3.720×10^{-66}), BootScan (4.094×10^{-60}), MaxChi (2.527×10^{-27}), Chimaera (9.018×10^{-22}), and 3Seq (1.110×10^{-16}) (*SI Appendix*, Table S3). Data from all of these analyses consistently showed that potential recombinant events between BANAL-236 and RaTG13 spike genes approximately occurred at two positions, one before the spike gene region and the other at the junction of the NTD- and RBD-encoding regions (Fig. 6 *A* and *B* and *SI Appendix*, Table S3). Furthermore, the BANAL-236 and RaTG13 spike genes were detected as the major and minor parents, respectively. Hence, the BANAL-236 spike gene could have kept its RBD and S2 regions while swapping its NTD region with that of the RaTG13 spike gene, producing the SARS-CoV-2 spike gene.

Discussion

The COVID-19 pandemic has had devastating effects on global health and economies. Understanding the animal origin of SARS-CoV-2 is crucial for the scientific community to unravel the mysteries of the pandemic and to prevent future coronavirus outbreaks. However, this is a difficult task due to the limited prepandemic data on coronavirus infections in animals. Additionally, the evolutionary events leading to SARS-CoV-2 may have been sporadic and transient, making it possible that evidence of these events could not be found years later. To address the challenge of limited animal surveillance data, we have been advancing a structure-guided virus hunting approach to explore how coronaviruses cross species barriers (1–3, 5, 6, 10, 15). The current study presents several lines of scientific evidence that suggest a scenario where two bat coronaviruses may have coinfected rodents and the subsequent evolutionary events in rodents might have given rise to SARS-CoV-2.

Initially, two mysteries surrounded the origin of SARS-CoV-2: the strong affinity of its RBD for human ACE2 and the presence of a furin motif in its spike protein (13, 15, 17). The discovery of BANAL-236 in bats resolved the first mystery, showing that BANAL-236 RBD is remarkably similar to that of prototypic SARS-CoV-2 and providing overwhelming evidence that SARS-CoV-2 RBD originated from bats (22). Yet, this finding

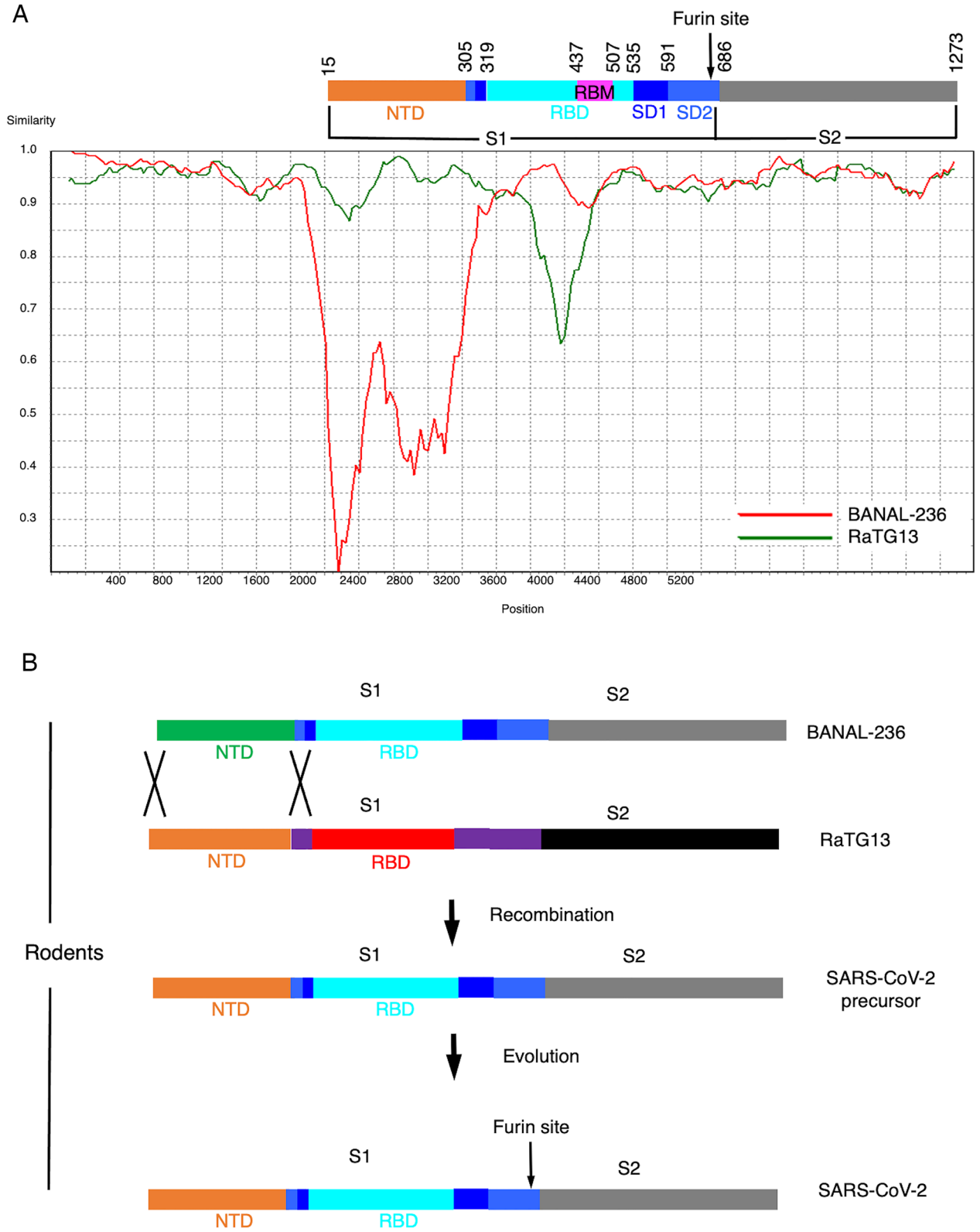


Fig. 6. Recombinant analyses of SARS-CoV-2 and bat SARS2-like coronaviruses. (*A*) Analysis of the recombination events among coronaviruses based on the K2P model calculated using Simplot. The query is prototypic SARS-CoV-2. The window size and step were set to 200 bp and 20 bp, respectively, with *P*-value equal to 0.05 and the gapstrip on. (*B*) The schema of the recombinant and evolutionary progress of SARS-CoV-2.

introduced another conundrum as BANAL-236 and SARS-CoV-2 spikes differ in their NTD, indicating that BANAL-236 is not the direct progenitor of SARS-CoV-2. Another bat coronavirus, RaTG13, has a spike resembling that of SARS-CoV-2 in the NTD but not in the RBD (21). Neither bat virus has the furin motif found in SARS-CoV-2. One possible scenario is that BANAL-236 and RaTG13 might have coinfected a host, leading to a recombination of their spike genes to form a precursor to SARS-CoV-2 spike, which subsequently acquired the furin motif. To be considered a potential originator of SARS-CoV-2, a host species would

need to meet several key conditions: its ACE2 should be able to act as an effective receptor for both BANAL-236 and RaTG13, it should facilitate the recombination of coronavirus genomes, and it should aid in the evolution of coronavirus spikes to acquire the furin motif. RaTG13 does not efficiently use human ACE2 as a receptor and bat coronaviruses rarely possess the furin motif (Fig. 2) (15, 23), suggesting that neither humans nor bats are probable candidates for the initial host of SARS-CoV-2. Consequently, this research explores the possibility that rodents could have been the host where SARS-CoV-2 originated.

This study assessed the capability of mouse ACE2 to act as a receptor for BANAL-236 and RaTG13 spikes. Our previous research identified three virus-binding hotspots at the RBD/ACE2 interface (6, 9, 10, 15, 31), essential for determining the host range of both SARS-CoV-2 and SARS-CoV-1. Two of the hotspots are centered around residues 31 and 353 on ACE2. Uniquely, mouse ACE2 contains an asparagine at position 31 and a histidine at position 353, contrasting with the lysines found at these positions in humans and many other species. Coronaviruses that recognize ACE2 as their receptor, including SARS-CoV-2, SARS-CoV-1, and NL63-CoV, have evolved structural adaptations to engage with Lys31 and Lys353 on human ACE2 (6, 9, 10, 15, 31, 33, 34). However, the presence of Asn31 and His353 in mouse ACE2 allows for greater flexibility in binding coronavirus RBDs. Consequently, BANAL-236 spike, which differs from the SARS-CoV-2 RBM by just two residues, can use mouse ACE2 for entry. Similarly, RaTG13 spike, with more significant differences from the SARS-CoV-2 RBM, can also engage mouse ACE2, despite its inability to use the human ACE2 receptor effectively. Given that receptor binding is a critical factor in determining the host range of coronaviruses (1–3), our findings indicate that both BANAL-236 and RaTG13 have the potential to infect mice. Indeed, infection with BANAL-236 was confirmed in B6.Cg-Tg(K18-ACE2)2Prlmn/J transgenic mice, BALB/cJRj mice, and also in macaques (35), though the infectivity of RaTG13 in mice remains unverified due to the unavailability of the live virus—its genomic sequence is all that is published (21). It is worth noting that Asn31 and His353 are features of rodent ACE2: Some other rodent species, such as rats, contain either Asn31 or His353, whereas only mouse ACE2 contains both (10). Overall, this study suggests the possibility of coinfection by BANAL-236 and RaTG13 in mice or some other rodents.

Should mice become coinfected with BANAL-236 and RaTG13, the frequent recombination seen in coronaviruses (28), especially in those infecting rodents (29, 30), may occur between the two viruses. Our analysis indicates that a recombination of the spike genes from BANAL-236 and RaTG13 could produce a spike protein closely resembling that of SARS-CoV-2, yet lacking the furin cleavage site. This hypothetical recombination would involve the NTD of RaTG13 supplanting that of BANAL-236 (Fig. 6). The resulting recombination product would share a striking 98.7% sequence identity with the SARS-CoV-2 spike across all regions (Fig. 1*C*). Subsequently this SARS-CoV-2 spike precursor could have evolved the furin motif in rodents. It is unclear why a majority of rodent coronaviruses contain a furin motif in their spikes, unlike bat coronaviruses, which rarely possess this motif (23). Nevertheless, the prevalence of the furin motif in rodent coronaviruses suggests that SARS-CoV-2 had a greater opportunity to acquire this feature in rodents than in other species.

It is important to recognize that our analysis presents just one possible scenario for the origin of SARS-CoV-2. There are, certainly, other possible scenarios. For instance, there could be an unidentified bat coronavirus that closely matches SARS-CoV-2 in every region of its spike, potentially rendering the need for genomic recombination of multiple viruses unnecessary. Alternatively, a different host species, other than rodents, may have been the site for the coinfection by BANAL-236 and RaTG13, providing the conditions for their genomic recombination and the subsequent rise of SARS-CoV-2. While rodents have a known propensity for generating coronaviruses with the furin motif, this does not preclude other species from doing so. Furthermore, even if rodents were involved in the initial development of SARS-CoV-2, the virus could have been transmitted to humans via one or more intermediate hosts. The precise duration that SARS-CoV-2 circulated in nature before being detected in humans is still unknown. Nevertheless, the scenario outlined in the current study aligns with the scientific data currently available.

In summary, our research presents structural and biochemical evidence that mouse ACE2 is a viable receptor for both BANAL-236 and RaTG13. The data we have gathered outline a potential evolutionary path where these two bat SARS2-like coronaviruses might have coinfected mice or another rodent, leading to a recombination of their spike genes and the acquisition of a furin motif, ultimately resulting in SARS-CoV-2 spike. Supporting this, a surveillance study recently found that 16.5% of rats in New York City tested positive for SARS-CoV-2 (27), indicating that rodents are efficient carriers of the virus. However, to truly comprehend the origins of SARS-CoV-2, surveillance should focus on rodents in the regions where the virus or its closely related bat coronaviruses were first detected. Our research provides the foundational structural and biochemical understanding for further studies into the animal origins of SARS-CoV-2.

Materials and Methods

Cell Lines and Plasmids. HEK293T cells [American Type Culture Collection (ATCC)] were cultured in Dulbecco's modified Eagle medium (DMEM) (containing 10% fetal bovine serum, 2 mM L-glutamine, 100 units/mL penicillin, and 100 μg/mL streptomycin). 293F cells (ThermoFisher) were cultured in FreeStyle 293 Expression Medium (ThermoFisher). Both cells were authenticated by ATCC using STR profiling and were also tested for mycoplasma contamination. No commonly misidentified cell lines were used.

The genes encoding the spike proteins of SARS-CoV-2 prototypic strain (GenBank: QHD43416.1), BANAL-236 (GenBank: UAY13253.1), and RaTG13 (GenBank: QHR63300.2) and the ACE2 protein of human (GenBank: NM_021804) and mouse (GenBank: NM_027286) were synthesized (GenScript Biotech).

For biochemical assays, the genes encoding the BANAL-236 RBD (residues 319 to 535) and its mutants, RaTG13 spike ectodomain (residues 1 to 1,208), human ACE2 (residues 1 to 615), and mouse ACE2 (residues 1 to 615) were constructed from the above full-length genes, respectively. BANAL-236 RBD was subcloned into pLenti-transfer vector (Addgene) with an N-terminal tissue plasminogen activator (tPA) signal peptide and a C-terminal His tag. Human ACE2 and mouse ACE2 were subcloned into the same vector except that a C-terminal human IgG4 Fc region replaced the His tag. RaTG13 spike ectodomain was also subcloned into this vector with both a C-terminal foldon trimerization tag and a C-terminal His tag.

For crystallography studies, the genes encoding the chimeric RBDs of BANAL-236 and RaTG13 (residues 319 to 535) were constructed by site-directed mutagenesis of the gene encoding the chimeric prototypic RBD (9, 10, 15, 31). They were then subcloned into pFastBac I vector (Life Technologies) with an N-terminal honeybee melittin signal peptide and a C-terminal His tag. Human ACE2 and mouse ACE2 (residues 1 to 615) were subcloned in the same way as the chimeric RBDs.

Protein Expression and Purification. For biochemical assays, BANAL-236 RBD and its mutants, RaTG13 spike ectodomain, human ACE2, and mouse ACE2 were prepared from 293F mammalian cells (36). Briefly, lentiviral particles were packaged for construction of stable cell lines expressing one of the above proteins. Stable cell lines expressing one of the above proteins were selected in the presence of Puromycin (Gibco). Each of the proteins was collected from

cell culture medium, purified on Ni-NTA column for His-tagged proteins or on Protein A column for Fc-tagged proteins, and purified further on Superdex200 gel filtration column (Cytiva).

For crystallography studies, chimeric BANAL-236 RBD, chimeric RaTG13 RBD, and mouse ACE2 were prepared from sf9 insect cells using the Bac-to-Bac system (Life Technologies) (15). Subsequently, the His-tagged proteins were harvested from cell culture medium, purified on Ni-NTA column, and purified further on Superdex200 gel filtration column (Cytiva).

SPR Assay. Binding interactions between mouse ACE2 and BANAL-236 RBD (and its mutants) and between mouse ACE2 and RaTG13 spike ectodomain were measured by SPR using a Biacore S200 system (Cytiva) (37). Briefly, Fc-tagged mouse ACE2 was immobilized to a protein A sensor chip (Cytiva). Serial dilutions of purified recombinant RBD or spike ectodomain were injected at different concentrations. Binding kinetics were calculated using Biacore Evaluation Software (Cytiva).

Pseudovirus Entry. Pseudoviruses were packaged as previously described (38). Briefly, pcDNA3.1(+) plasmid, which encodes the full-length spikes of SARS-CoV-2, BANAL-236, or RaTG13, was cotransfected into HEK293T cells along with helper plasmid psPAX2 and reporter plasmid plenti-CMV-luc, using a molar ratio of 1:1:1 and Lipofectamine 3000 (Life Technologies). Seventy-two hours posttransfection, the produced pseudoviruses were harvested and then used to enter HEK293T cells stably expressing human ACE2 or mouse ACE2. Following a 5-h incubation at 37 °C, the medium was refreshed and the cells were incubated for an additional 48 h. Subsequently, the cells were washed with PBS and lysed. The cell lysate aliquots were dispensed into an Optiplate-96 (PerkinElmer) and treated with a luciferase substrate. The resulting relative light units (RLUs) were measured using an EnSpire plate reader (PerkinElmer). Concurrently, the quantity of spikes packaged within the pseudoviruses was determined through Western blot analysis employing an anti-C9-tag antibody and quantified with Fiji (https://imagej.net/). The RLUs were then normalized against the amounts of pseudovirus-packaged spikes.

Crystallization and Structure Determination. The complexes of the chimeric BANAL-236 RBD and chimeric mouse ACE2 and of the chimeric RaTG13 RBD and chimeric mouse ACE2 were each purified using gel filtration chromatography. Crystals of each of the complexes were grown at room temperature over wells containing 100 mM Tris (pH 8.5), 18 to 24% PEG 6000, and 100 mM NaCl. X-ray diffraction data were collected at NECAT (24-IDC and 24-IDE) beamlines. HKL2000 was used for data processing (39). Both of the structures were determined by molecular replacement using the structure of prototypic chimeric RBD complexed with human ACE2 as the search model (PDB ID 6VW1). PHENIX and CCP4 were used for molecular replacement and model refinement (40, 41). COOT was used for model building (42). PYMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) was used for making structural figures. Structure data and refinement statistics are shown in *SI Appendix*, Table S2.

Phylogenetic and Recombinant Analyses. Multiple sequence alignments of SARS-CoV-2 (GISAID: EPI_ISL_402124), BANAL-236 (GenBank: MZ937003.2) and RaTG13 (GenBank: MN996532.2) were constructed using Clustal W in the Molecular Evolutionary Genetics Analysis (MEGA) software version 11 (43). Following this, we employed SimPlot software version 3.5.1 to identify potential recombinant regions (44) and the Recombination Detection Program (RDP) version 4.101 to determine precise breakpoints (45). Three additional coronaviruses, BANAL-20-52 (GenBank: MZ937000.1), RmYN02 (GISAID: EPI_ISL_412977), and RpYN06 (GenBank: MZ081381.1) were included as references in the RDP analysis. Sequences that showed high similarity were aligned with automatic sequence masking as recommended by the guidelines. Six algorithms in RDP4.101, including RDP, GENECONV, BootScan, MaxChi, Chimaera, and 3Seq (46–48), were used to calculate the recombinant breakpoints, both major and minor parents, and recombination events shown via average *P* value.

Data, Materials, and Software Availability. Coordinates and structure factors have been deposited in the Protein Data Bank: Accession No. 8UZE for chimeric BANAL-236 RBD complexed with chimeric mouse ACE2 (49) and Accession No. 8UZF for chimeric RaTG13 RBD complexed with chimeric mouse ACE2 (50).

ACKNOWLEDGMENTS. This study was supported by NIH Grants R01AI089728 and R01AI110700 (to F.L.) and R35GM118047 (to H.A.). The Biacore S200 system was supported by NIH ORIP Grant 1S10OD021539. X-ray diffraction data were collected at the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, which are funded by NIH (NIGMS P30 GM124165).

Author affiliations: [a]Department of Pharmacology, University of Minnesota Medical School, Minneapolis, MN 55455; [b]Center for Emerging Viruses, University of Minnesota, Minneapolis, MN 55455; [c]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455; and [d]Department of Microbiology and Immunology, University of Iowa, Iowa City, IA 52242

1. F. Li, Receptor recognition mechanisms of coronaviruses: A decade of structural studies. *J. Virol.* **89**, 1954–1964 (2015).
2. F. Li, Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **3**, 237–261 (2016).
3. F. Li, Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res.* **100**, 246–254 (2013).
4. F. Li, Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J. Virol.* **82**, 6984–6991 (2008).
5. F. Li, W. Li, M. Farzan, S. C. Harrison, Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
6. K. Wu, G. Peng, M. Wilken, R. J. Geraghty, F. Li, Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* **287**, 8904–8911 (2012).
7. N. Lee *et al.*, A major outbreak of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **348**, 1986–1994 (2003).
8. J. S. M. Peiris *et al.*, Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325 (2003).
9. W. Zhang *et al.*, Structural evolution of SARS-CoV-2 omicron in human receptor recognition. *J. Virol.* **97**, e0082223 (2023).
10. W. Zhang *et al.*, Structural basis for mouse receptor recognition by SARS-CoV-2 omicron variant. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2206509119 (2022).
11. S. S. A. Karim, Q. A. Karim, Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic. *Lancet* **398**, 2126–2128 (2021).
12. C. Maslo *et al.*, Characteristics and outcomes of hospitalized patients in South Africa during the COVID-19 Omicron wave compared with previous waves. *JAMA* **327**, 583–584 (2022).
13. J. Lan *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
14. W. H. Li *et al.*, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
15. J. Shang *et al.*, Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
16. Y. Wan, J. Shang, R. Graham, R. S. Baric, F. Li, Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* **94**, e0012720 (2020), 10.1128/jvi.00127-00120.
17. J. Shang *et al.*, Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11727–11734 (2020).
18. Q. Geng *et al.*, Lys417 acts as a molecular switch that regulates the conformation of SARS-CoV-2 spike protein. *eLife* **12**, e74060 (2023).
19. A. C. P. Wong, X. Li, S. K. P. Lau, P. C. Y. Woo, Global epidemiology of bat coronaviruses. *Viruses* **11**, 174 (2019).
20. J. Cui, F. Li, Z.-L. Shi, Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
21. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020), 10.1038/s41586-020-2012-7.
22. S. Temmam *et al.*, Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
23. A. E. Stout, J. K. Millet, M. J. Stanhope, G. R. Whittaker, Furin cleavage sites in the spike proteins of bat and rodent coronaviruses: Implications for virus evolution and zoonotic transfer from rodent species. *One Health* **13**, 100282 (2021).
24. A. D. Luis *et al.*, A comparison of bats and rodents as reservoirs of zoonotic viruses: Are bats special? *Proc. Biol. Sci.* **280**, 20122753 (2013).
25. S. K. Lau *et al.*, Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of Betacoronavirus 1 and has implications for the ancestor of Betacoronavirus lineage A. *J. Virol.* **89**, 3076–3092 (2015).
26. X. Y. Ge *et al.*, Detection of alpha- and betacoronaviruses in rodents from Yunnan, China. *Virol. J.* **14**, 98 (2017).
27. Y. Wang *et al.*, SARS-CoV-2 exposure in Norway Rats (*Rattus norvegicus*) from New York City. *mBio* **14**, e0362122 (2023).
28. R. L. Graham, R. S. Baric, Recombination, reservoirs, and the modular spike: Mechanisms of coronavirus cross-species transmission. *J. Virol.* **84**, 3134–3146 (2010).
29. S. Makino, J. G. Keck, S. A. Stohlman, M. M. Lai, High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**, 729–737 (1986).
30. J. G. Keck, L. H. Soe, S. Makino, S. A. Stohlman, M. M. Lai, RNA recombination of murine coronaviruses: Recombination between fusion-positive mouse hepatitis virus A59 and fusion-negative mouse hepatitis virus 2. *J. Virol.* **62**, 1989–1998 (1988).
31. Q. Geng *et al.*, Structural basis for human receptor recognition by SARS-CoV-2 Omicron Variant BA.1. *J. Virol.* **96**, e0024922 (2022).

32. K. Liu *et al.*, Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. *Cell* **184**, 3438–3451.e3410 (2021).
33. K. Wu *et al.*, A virus-binding hot spot on human angiotensin-converting enzyme 2 is critical for binding of two different coronaviruses. *J. Virol.* **85**, 5331–5337 (2011).
34. K. Wu, W. Li, G. Peng, F. Li, Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19970–19974 (2009).
35. S. Temmam *et al.*, SARS-CoV-2-related bat virus behavior in human-relevant models sheds light on the origin of COVID-19. *EMBO Rep.* **24**, e56055 (2023).
36. Q. Geng *et al.*, Novel virus-like nanoparticle vaccine effectively protects animal model from SARS-CoV-2 infection. *PLoS Pathog.* **17**, e1009897 (2021).
37. Q. Geng *et al.*, Structural basis for human receptor recognition by SARS-CoV-2 Omicron Variant BA.1. *J. Virol.* **96**, e0024922 (2022), 10.1128/jvi.00249-22.
38. G. Peng *et al.*, Structural and molecular evidence suggesting coronavirus-driven evolution of mouse receptor. *J. Biol. Chem.* **292**, 2174–2181 (2017).
39. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr., Pt A.* **276**, 307–326 (1997).
40. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr., Sect. D, Struct. Biol.* **75**, 861–877 (2019).
41. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D, Biol. Crystallogr.* **67**, 235–242 (2011).
42. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr., Sect. D-Biol. Crystallogr.* **60**, 2126–2132 (2004).
43. K. Tamura, G. Stecher, S. Kumar, MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
44. K. S. Lole *et al.*, Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**, 152–160 (1999).
45. D. P. Martin *et al.*, RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **7**, veaa087 (2021).
46. G. A. McVean *et al.*, The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
47. D. P. Martin, D. Posada, K. A. Crandall, C. Williamson, A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retroviruses* **21**, 98–102 (2005).
48. D. Posada, K. A. Crandall, Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13757–13762 (2001).
49. W. Zhang *et al.*, Crystal structure of chimeric bat coronavirus BANAL-236 RBD complexed with chimeric mouse ACE2. Protein Data Bank. https://www.rcsb.org/structure/8uze. Deposited 15 November 2023.
50. W. Zhang *et al.*, Crystal structure of chimeric RaTG13 RBD complexed with chimeric mouse ACE2. Protein Data Bank. https://www.rcsb.org/structure/8uzf. Deposited 15 November 2023.