fivefold increase in recombination (Fig. 3C). This elevated recombination was only slightly reduced by CR. Finally, we observed a highly significant negative correlation between life span and rDNA recombination rate (fig. S3). Although these data do not exclude the possibility that CR may mediate yeast life span independently of its effects on the rDNA, these data provide strong evidence that CR extends life span by suppressing rDNA recombination irrespective of whether *SIR2* is present or absent. They also demonstrate that in a *sir2*Δ *fob1*Δ strain, Hst2 is critical for maintaining rDNA stability.

Although the deletion of *HST2* blocked the ability of CR to extend life span in the *sir2*Δ *fob1*Δ strain, it was formally possible that this was caused by toxic levels of ERCs in the strain, precluding alternative CR pathways from taking effect. Therefore, we determined whether *HST2* could increase life span when overexpressed in order to test whether *HST2* is a bona fide longevity gene (*9*). Consistent with the ability of *HST2* to increase rDNA silencing and decrease rDNA recombination (Fig. 1 and fig. S1), overexpression of *HST2* in W303AR5 *sir2*Δ *fob1*Δ extended life span to the same extent as CR in this strain background (Fig. 4A), as well as in a wild-type strain (fig. S4). No additive effect of *HST2* overexpression and CR was observed, indicating that *HST2* and CR extend life span of *sir2*Δ *fob1*Δ mutants through the same pathway (*28*).

Next, we investigated whether the residual life-span extension seen for the *hxk2*Δ mutant (a mimic of intense CR) lacking *SIR2* and *HST2* (Fig. 2C) was due to the activity of another sirtuin. As previously reported (*16*), deletion of *HST1* markedly increased rDNA recombination in a wild-type strain (Fig. 4B). Although deleting *HST3* and *HST4* together has been shown to decrease chromosomal stability and increase mitotic recombination (*29*), we did not observe increased rDNA recombination in a W303AR5 *hst3*Δ *hst4*Δ strain, although recombination in an *hst4*Δ single mutant is about twice as high as that in the wild type. Because deletion of *HST1* had the greatest effect on rDNA recombination, we suspected that Hst1 might be the factor responsible for the residual life-span extension. This hypothesis was consistent with our finding that the general sirtuin inhibitor NAM completely blocked the life-span extension of a *sir2*Δ *fob1*Δ strain by *hxk2*Δ (Fig. 1D) and a recent report that Hst1 functions in the nucleus with Hst2 in gene silencing (*23*). Whereas deletion of either *HST3* or *HST4* in this strain did not affect the ability of *hxk2*Δ to extend life span (fig. S5), deletion of *HST1* completely eliminated the residual life-span extension provided by *hxk2*Δ in the BY4742 *sir2*Δ *fob1*Δ *hst2*Δ strain (Fig. 4C).

In a previous study, the life span of a *sir2*Δ *fob1*Δ *hst1*Δ strain was extended by CR (*19*), leading the authors to conclude that *HST1* plays no role in CR. Indeed, in agreement with this finding, we find that CR is effective in suppressing recombination of such a mutant (Fig. 4D). However, this study implies that *HST2* underlies the CR-mediated life-span extension of this strain and that *HST1* plays a minor role that is observed only in the absence of *SIR2* and *HST2*.

Our results show that *HST2* is responsible for Sir2-independent life-span extension by CR and that it does so by suppressing rDNA recombination, the same mechanism by which *SIR2* extends life span. These findings highlight the importance of genomic stability as a determinant of yeast life span and raise the likelihood that multiple members of the sirtuin family in higher organisms also play critical roles in maintaining genomic stability and possibly in extending life span during times of adversity.

References and Notes

1. A. A. Falcon, J. P. Aris, *J. Biol. Chem.* **278**, 41607 (2003).
2. D. A. Sinclair, L. Guarente, *Cell* **91**, 1033 (1997).
3. T. Kobayashi, D. J. Heck, M. Nomura, T. Horiuchi, *Genes Dev.* **12**, 3821 (1998).
4. P. A. Defossez *et al.*, *Mol. Cell* **3**, 447 (1999).
5. J. S. Smith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6658 (2000).
6. S. Imai, C. M. Armstrong, M. Kaeberlein, L. Guarente, *Nature* **403**, 795 (2000).
7. K. G. Tanner, J. Landry, R. Sternglanz, J. M. Denu, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14178 (2000).
8. J. C. Tanny, D. Moazed, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 415 (2001).
9. M. Kaeberlein, M. McVey, L. Guarente, *Genes Dev.* **13**, 2570 (1999).
10. E. J. Masoro, *Exp. Gerontol.* **35**, 299 (2000).
11. S. J. Lin, P. A. Defossez, L. Guarente, *Science* **289**, 2126 (2000).
12. R. M. Anderson, K. J. Bitterman, J. G. Wood, O. Medvedik, D. A. Sinclair, *Nature* **423**, 181 (2003).
13. S. J. Lin *et al.*, *Nature* **418**, 344 (2002).
14. B. Rogina, S. L. Helfand, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15998 (2004).
15. J. G. Wood *et al.*, *Nature* **430**, 686 (2004).
16. S. Perrod *et al.*, *EMBO J.* **20**, 197 (2001).
17. K. Houthoofd, B. P. Braeckman, T. E. Johnson, J. R. Vanfleteren, *Exp. Gerontol.* **38**, 947 (2003).
18. J. C. Jiang, J. Wawryn, H. M. Shantha Kumara, S. M. Jazwinski, *Exp. Gerontol.* **37**, 1023 (2002).
19. M. Kaeberlein, K. T. Kirkland, S. Fields, B. K. Kennedy, *PLoS Biol.* **2**, E296 (2004).
20. D. A. Sinclair, J. Wood, data not shown.
21. M. Kaeberlein, K. T. Kirkland, S. Fields, B. K. Kennedy, *Mech. Ageing Dev.* **126**, 491 (2005).
22. B. K. Kennedy, N. R. Austriaco Jr., J. Zhang, L. Guarente, *Cell* **80**, 485 (1995).
23. A. Halme, S. Bumgarner, C. Styles, G. R. Fink, *Cell* **116**, 405 (2004).
24. S. M. Gasser, M. M. Cockell, *Gene* **279**, 1 (2001).
25. K. J. Bitterman, R. M. Anderson, H. Y. Cohen, M. Latorre-Esteves, D. A. Sinclair, *J. Biol. Chem.* **277**, 45099 (2002).
26. J. Landry, J. T. Slama, R. Sternglanz, *Biochem. Biophys. Res. Commun.* **278**, 685 (2000).
27. C. M. Gallo, D. L. Smith Jr., J. S. Smith, *Mol. Cell. Biol.* **24**, 1301 (2004).
28. D. W. Lamming *et al.*, data not shown.
29. C. B. Brachmann *et al.*, *Genes Dev.* **9**, 2888 (1995).
30. K. T. Howitz *et al.*, *Nature* **425**, 191 (2003).
31. M. S. Longtine *et al.*, *Yeast* **14**, 953 (1998).
32. A. L. Goldstein, J. H. McCusker, *Yeast* **15**, 1541 (1999).

33. We thank members of the Lin and Sinclair labs for valuable insights and J. Wood for critical reading of the manuscript. This work was supported by the National Institute of General Medical Sciences, the National Institute on Aging, the Harvard-Armenise Foundation, and The Paul F. Glenn Laboratories for the Biological Mechanisms of Aging at Harvard. S.-J.L. and D.A.S. are Ellison Medical Research Foundation New Research Scholars. D.W.L. is supported by a National Eye Institute training grant. D.A.S. is a cofounder and board member of, and has equity in, Sirtris Pharmaceuticals, a company whose goal is to discover sirtuin-modulating drugs.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1113611/DC1
Materials and Methods
Figs. S1 to S5
References and Notes

15 April 2005; accepted 8 July 2005
Published online 28 July 2005;
10.1126/science.1113611
Include this information when citing this paper.

Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor

Fang Li,[1] Wenhui Li,[3] Michael Farzan,[3] Stephen C. Harrison[1,2]*

The spike protein (S) of SARS coronavirus (SARS-CoV) attaches the virus to its cellular receptor, angiotensin-converting enzyme 2 (ACE2). A defined receptor-binding domain (RBD) on S mediates this interaction. The crystal structure at 2.9 angstrom resolution of the RBD bound with the peptidase domain of human ACE2 shows that the RBD presents a gently concave surface, which cradles the N-terminal lobe of the peptidase. The atomic details at the interface between the two proteins clarify the importance of residue changes that facilitate efficient cross-species infection and human-to-human transmission. The structure of the RBD suggests ways to make truncated disulfide-stabilized RBD variants for use in the design of coronavirus vaccines.

The SARS coronavirus (SARS-CoV) is the agent of severe acute respiratory syndrome, which emerged as a serious epidemic in 2002 to 2003, with over 8,000 infected cases and a fatality rate of ~10% (*1*–*4*). Coronaviruses, which are large, enveloped, positive-strand RNA viruses, infect a variety of mammalian and avian species and can cause upper res-

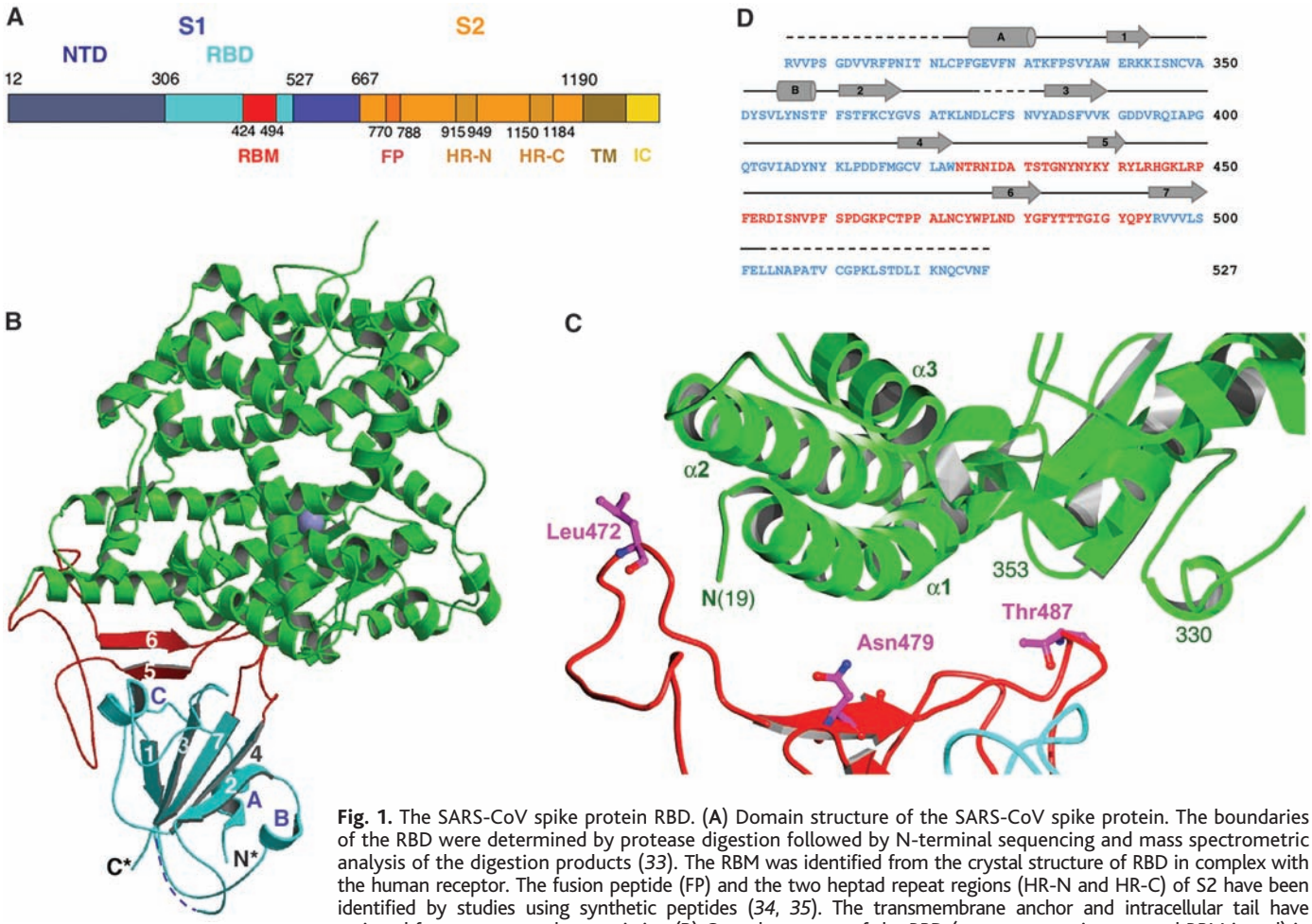

Fig. 1. The SARS-CoV spike protein RBD. (**A**) Domain structure of the SARS-CoV spike protein. The boundaries of the RBD were determined by protease digestion followed by N-terminal sequencing and mass spectrometric analysis of the digestion products (*33*). The RBM was identified from the crystal structure of RBD in complex with the human receptor. The fusion peptide (FP) and the two heptad repeat regions (HR-N and HR-C) of S2 have been identified by studies using synthetic peptides (*34*, *35*). The transmembrane anchor and intracellular tail have assigned from sequence characteristics. (**B**) Crystal structure of the RBD (core structure in cyan and RBM in red) in complex of the human receptor ACE2 (green). (**C**) Detail of the binding interface, with side chains of three residues (Leu^{472}, Asn^{479}, and Thr^{487} from left to right) critical for cross-species and human-to-human transmission of SARS-CoV. (**D**) Sequence and secondary structures of the RBD. Helices are drawn as cylinders, and strands are drawn as arrows. The RBM is in red; the remainder of the RBD is in cyan. Disordered regions are shown as dashed lines (*36*).

piratory, gastrointestinal, and central nervous system diseases (*5*). The large spike protein (S) on the virion surface mediates both cell attachment and membrane fusion (*5*). In the case of several avian and mammalian coronaviruses, S is cleaved by furin or a related protease into S1 and S2; the former bears the receptor attachment site; the latter, the fusion activity. The structures of refolded heptad-repeat fragments of S2 from the mouse hepatitis coronavirus (MHV) and from SARS-CoV (*6*–*8*) confirm earlier predictions (*4*) that the postfusion conformation has the trimer-of-hairpins organization characteristic of "class 1" fusion proteins, such as those of HIV, influenza virus, and Ebola virus (*9*). S on mature SARS-CoV virions does not appear to be cleaved, and the sequence that aligns with the MHV cleavage site lacks the essential residues for furin susceptibility (*3*, *4*, *10*, *11*). We therefore refer to the S1 and S2 "regions" (*12*), which contain 666 and 583 amino acid residues, respectively (Fig. 1A).

Coronaviruses exploit a wide variety of cellular receptors (*5*). SARS-CoV and another human coronavirus, HCoV-NL63, both use as their receptor a cell-surface zinc peptidase, angiotensin-converting enzyme 2 (ACE2) (*13*, *14*). The crystal structure of the ACE2 ectodomain (*15*) shows a claw-like N-terminal peptidase domain, with the active site at the base of a deep groove, and a C-terminal "collectrin" domain. A fragment of the S1 region, residues 318 to 510, is sufficient for tight binding to the peptidase domain of ACE2 (*11*, *16*, *17*). This fragment, the receptor-binding domain (RBD), is the critical determinant of virus-receptor interaction and thus of viral host range and tropism (*18*). SARS-CoV isolated from patients during the 2002–2003 epidemic, and also from milder sporadic cases in 2003 to 2004, appears to derive from a nearly identical virus circulating in palm civets and raccoon dogs (*19*, *20*). Changes in just a few residues in the RBD can lead to efficient cross-species transmission (*18*, *20*). The RBD also includes important viral-neutralizing epitopes (*21*–*23*), and it may be sufficient to raise a protective antibody response in inoculated animals.

We expressed the SARS-CoV spike protein RBD, residues 306 to 575, in Sf9 cells and purified the fragment (*24*). Brief treatment with chymotrypsin yielded a shorter fragment, residues 306 to 527. Soluble ACE2, residues 19 to 615, was expressed in Sf9 cells and purified as described in (*24*). The two components were mixed, and the complex was purified by size-exclusion chromatography on Superdex 200 (Amersham Biosciences, Piscataway, NJ). Crystals in space group $P21$, a = 82.3 Å, b = 119.4 Å, c = 113.2 Å, β = 91.2°, with two complexes per asymmetric unit, were grown at room temperature from a mother liquor containing 24% polyethylene glycol 6000,

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School and Laboratory of Molecular Medicine, and [2]Howard Hughes Medical Institute, Children's Hospital, 320 Longwood Avenue, Boston, MA 02115, USA. [3]Department of Microbiology and Molecular Genetics, Harvard Medical School, New England Primate Research Center, Southborough, MA 01772, USA.

*To whom correspondence should be addressed. E-mail: harrison@crystal.harvard.edu

150 mM NaCl, 100 mM Tris at pH 8.2, and 10% ethylene glycol. We determined the structure of the ACE2/SARS-CoV/RBD complex by molecular replacement with ACE2 as the search model, and we refined it at 2.9 Å resolution (*24*). The final model contains residues 19 to 615 of the N-terminal peptidase domain of human ACE2 and residues 323 to 502 (except for 376 to 381) of the RBD; as well as glycans N-linked to ACE2 residues 53, 90, 322, and 546 and to RBD residue 330; and 65 solvent molecules. The R_{free} is 27.5% and R_{work} is 22.1% (see table S1 for definitions).

The ACE2 peptidase domain has two lobes that close toward each other after substrate engagement (*15*). In one of the two complexes in the asymmetric unit of our crystals, ACE2 is fully open; in the other, it is slightly closed (fig. S1). The SARS-CoV S protein contacts the tip of one lobe of ACE2 (Fig. 1). It does not contact the other lobe, nor does it occlude the peptidase active site. Binding of the spike protein to ACE2 is not altered by the addition of a specific ACE2 inhibitor, which is expected to favor the closed state (*18*). Thus, both structural and biochemical data indicate that viral attachment is unaffected by the open-to-closed transition.

The RBD contains two subdomains (Fig. 1): a core and an extended loop. The core is a five-stranded anti-parallel β sheet (β1 to β4 and β7), with three short connecting α helices (αA to αC). There are nine cysteines in the chymotryptic fragment. Disulfide bonds connect cysteines 323 to 348, 366 to 419, and 467 to 474. The remaining cysteines are disordered but two (378 and 511) are in the same neighborhood and could form a disulfide in the recombinant fragment, even if they have other partners in the intact S protein. The extended loop subdomain lies at one edge of the core; it presents a gently concave outer surface formed by a two-stranded β sheet (β5 and β6). The base of this concavity cradles the N-terminal helix of ACE2; a ridge to one side of it, which is reinforced by the Cys^{467}–Cys^{474} disulfide bridge, contacts the loops between ACE2 helices α2 and α3; a ridge to the other side inserts between a short ACE2 helix (residues 329 to 333) and a β hairpin at ACE2 residue 353 (Fig. 1C). Residues 445 to 460 of the RBD anchor the entire receptor-binding loop to the core of the RBD. We refer to this loop (residues 424 to 494), which makes all the contacts with ACE2, as the receptor-binding motif (RBM).

The RBM surface is complementary to the receptor tip, with about 1700 Å^2 of buried surface at the interface (Fig. 2A and fig. S2), consistent with their high affinity (dissociation constant $K_d \sim 10^{-8}$) (*18*, *21*). A total of 18 residues of the receptor contact 14 residues of the viral spike protein (Table 1). Networks of hydrophilic interactions, which occur largely among amino acid side chains, predominate.

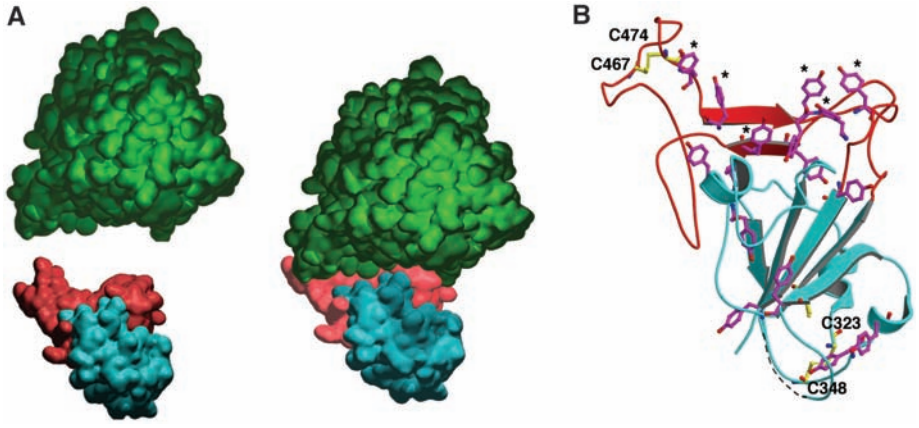

Fig. 2. Features contributing to specific recognition of ACE2 by the SARS-CoV RBD. (**A**) Surface complementarity, Space-filling representation of ACE2 (in green), RBD (core structure in cyan and RBM in red), and the complex of ACE2 and RBD are shown. The complex buries 1700 Å^2 at the binding interface. (**B**) Distribution of tyrosines (magenta) and cysteines (yellow) on the RBD. The RBM is particularly tyrosine-rich. The six tyrosines that contact ACE2 are accompanied by an asterisk. The three disulfide bonds link C323 to C348, C366 to C419, and C467 to C474; two are labeled, and the third is partly concealed by the lower corner of the β sheet.

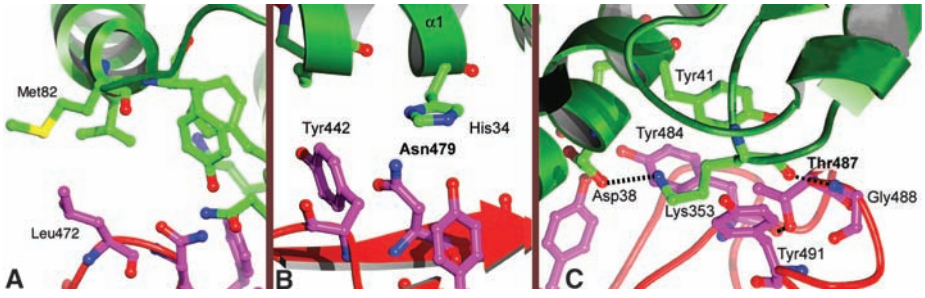

Fig. 3. Residues important for species specificities of SARS-CoV. (**A**) Met^{82} of human ACE2 is asparagine in rat ACE2, introducing a glycan that appears to interfere with infection of rat cells. (**B**) Asn^{479} (boldface) is present in most SARS-CoV sequences from human specimens. Lys^{479}, which is found in most sequences from palm-civet specimens, would have steric and electrostatic interference from residues (e.g., His^{34}) on the N-terminal helix of human ACE2. (**C**) Thr^{487} (boldface) appears to enhance human-to-human transmission of SARS-CoV. The methyl group of Thr^{487} lies in a hydrophobic pocket at the ACE2/RBD interface. On rat and mouse ACE2, residue 353 is histidine, disfavoring viral binding. The dashed black lines indicate hydrogen bonds.

Six RBM residues at this interface are tyrosines, which present both a polar hydroxyl group and a hydrophobic aromatic ring (Fig. 2B).

Coronaviruses are classified in three groups (*5*); SARS-CoV belongs to group 2 (fig. S3). Spike-protein sequences from several members of group 2 lead us to expect that all have rather similar structures, including the RBD core (fig. S3). The SARS-CoV RBM is substantially shorter than are the corresponding regions in several other group-2 viral spike proteins, however, and it has no evident sequence similarity to the others (fig. S3). Thus, this extended loop is probably a hypervariable decoration of an otherwise-conserved domain. In the case of MHV, the receptor (murine carcinoembryonic antigen cell adhesion molecule 1a, or CEACAM1a) (*25*, *26*) makes contact not with the extended-loop subdomain (nor, indeed, with any part of the domain homologous to the SARS-CoV RBD), but rather with structures in the N-terminal region of the spike protein (*27*). Receptors and receptor-binding regions of other group-2 coronaviruses have not been identified. The group-1 human coronavirus 229E receptor is aminopeptidase N; the corresponding RBD on its spike protein is known (*28*).

The SARS-CoV appears to derive from a cross-species infection with a coronavirus isolated from palm civets (*19*, *20*). S-gene sequences from civet and human specimens obtained during the 2002-to-2003 epidemic show that their RBDs differ at only four positions, residues 344, 360, 479, and 487, but the human viral spike protein binds the human receptor 10^3 to 10^4 times more tightly than does its civet spike counterpart (*18*). Residues 344 and 360 are far from the binding interface in the complex described here, and mutation to the corresponding civet CoV residues does not affect affinity or infectivity (*18*). The critical changes are therefore at positions 479 and 487, both of which lie in the RBD-receptor contact (Figs. 1 and 3 and Table 1).

Table 1. Contacts between ACE2 and SARS-CoV RBD. Residues in ACE2 that contact the RBD are listed by their position (numbers across the top of each column) and by their single-letter identity (*36*) in the palm-civet, mouse, rat, and human receptors. The residues they contact in the structure described here and their position numbers in the spike proteins from human isolates are shown at the bottom of each column.

24	27	31	34	37	38	41	42	45	79	82	83	90	325	329	330	353	354	
L	T	T	Y	Q	E	Y	Q	V	L	T	Y	D	Q	E	N	K	G	civet ACE2
N	T	N	Q	E	D	Y	Q	L	T	S	F	T	Q	A	N	H	G	mouse ACE2
K	S	K	Q	E	D	Y	Q	L	I	N	F	N	P	T	N	H	G	rat ACE2
Q	T	K	H	E	D	Y	Q	L	L	M	Y	N	Q	E	N	K	G	human ACE2
N473	Y475	Y475 Y442	Y440 N479	Y491	Y436	Y484 T486 T487	Y436 Y484	Y484	L472	L472	N473 Y475	T402	R426	R426	T486	G488 T487 Y491	Y491 G488	human SARS

The changes at these two positions are relatively subtle. In most viral sequences from palm-civet specimens, residue 479 is lysine and 487 is serine, whereas in SARS-CoV sequences from the 2002–2003 epidemic, these residues are asparagine and threonine, respectively. The presence of lysine at 479 reduces affinity for human but not for civet ACE2; serine at 487 reduces affinity for both receptors (*18*). Position 479 lies opposite the ACE2 N-terminal helix (α1), on which several residues differ in identity between civet and human (Table 1). Some civet coronavirus sequences have asparagine at position 479, and the difference does not appear to be critical for binding to the civet receptor (*18*). At position 487 in the spike protein, replacing threonine (SARS-CoV) with serine (civet viral sequences) would remove the threonine methyl group, which lies in a hydrophobic pocket bounded by atoms in the side chains of Tyr^{41} and Lys^{353} on the receptor and Tyr^{484} in the RBM (Fig. 3C). This pocket appears to be relatively inflexible. A main-chain hydrogen bond (carbonyl of ACE2 Lys^{353} to amide of RBD Gly^{488}) fixes the relative positions of receptor and spike protein quite precisely. Moreover, the Thr^{487} rotamer is determined by a hydrogen bond from Oγ to the main-chain carbonyl of Tyr^{484}; the aliphatic part of the Lys^{353} side chain is sandwiched between the rings of ACE2 Tyr^{41} and RBD Tyr^{491}, and the $\varepsilon\text{-}NH_3^+$ is neutralized by ACE2 Asp^{38}. Mutation to serine would thus leave a hard-to-fill van der Waals hole; indeed, a mutation in which Thr^{487} is replaced by Ser in the human RBD decreases affinity for human ACE2 by more than 20-fold (*18*). Civet ACE2 is essentially identical to human ACE2 at all the relevant positions in the vicinity of this interaction; like the human receptor, it appears to bind RBDs with threonine at 487 more tightly than those with serine (*18*). All of the more than 100 S-protein sequences obtained during the 2002–2003 SARS epidemic have threonine at this position, whereas all 14 such sequences from palm-civet and raccoon-dog isolates have serine (*29*, *30*).

Viruses from sporadic SARS cases during 2003 to 2004, each of which was an independent cross-species event from which no human-to-human transmission occurred, all had asparagine at 479 and serine at 487 (*29*, *30*). It is therefore plausible that a key factor determining severity (and possibly human-to-human transmission) is the presence or absence of a γ-methyl group on the 487 side chain. The 2003–2004 sequences differed, however, at two other RBD positions from those sequences obtained during the epidemic of the previous winter: Leu^{472} had changed to proline and Asp^{480} to glycine. Inspection of the model suggests that the leucine-to-proline change might have contributed to attenuation, by reducing the spike-receptor contact surface (Fig. 3A). A similar rationale is harder to find for the aspartate-to-glycine substitution, because the aspartyl side chain projects into solution, and mutation of this residue to alanine has no effect on RBD binding to ACE2 (*16*).

Two other species differences are worth noting. Rat ACE2 does not support infection by SARS-CoV, and mouse ACE2 does so only inefficiently (*30*). At position 82, where the human receptor has a methionine, the rat protein has a glycosylated asparagine; the glycan would disrupt by steric interference a hydrophobic contact between Met^{82} and Leu^{472} in the RBM (Fig. 3A). At position 353, where the human receptor has a lysine critical for the contact with Thr^{487} in the RBM (Fig. 3B), the rat receptor has histidine. Mouse ACE2 also has histine at 353, but it does not have a glycosylation site at 82. It thus bears one but not both of the differences that render rat ACE2 inactive as a receptor, and mutation of His^{353} to lysine in mouse ACE2 allows high-level infection of murine cells by SARS-CoV (*30*).

The residues singled out for description in the preceding paragraphs are not, of course, the only ones critical for the tight complementarity of the SARS-CoV RBD and human (or palm civet) ACE2. They are simply the positions at which there are differences among isolates and receptors important for binding and entry. Other species might in principle harbor variants of the same virus that would require changes at different positions to be able to infect human cells, and other changes in the civet virus might permit cross-species infection even in the absence of the serine-to-threonine mutation at position 487. The structure might allow one to recognize such changes in future animal isolates. For example, the human receptor (but not the civet receptor) bears an N-linked glycan at position 90. Mutation of Asn^{90} to eliminate the glycan enhances S-protein–mediated binding and infection of human cells by pseudotyped lentiviruses (*18*). The glycan faces a loop in the RBD containing residues 399 to 412. Changes in this loop that reduce likely interference with the glycan might have the same enhancing effects as does elimination of the glycan on the receptor or mutation of Ser^{487} to threonine on the S protein.

Neutralizing antibodies against SARS-CoV recognize epitopes in the RBD (*21*–*23*). For example, a high-affinity recombinant human monoclonal antibody, 80R, which is sensitive to mutation within the RBM, inhibits viral entry by blocking association of virus and receptor (*21*, *31*). The soluble SARS-CoV RBD is therefore of potential use as an immunogen (*23*, *32*). In the structure described here, the interface of the RBD with the receptor is very well defined, but the opposite face of the RBD is more disordered. The latter surface would interact with the rest of the spike protein, and it indeed contains the N and C termini of the RBD fragment as well as the disordered loop, residues 376 and 381. Thus, this face of the protein could be modified in various ways in the molecular engineering of a candidate vaccine. The loop from 376 to 381 could probably be shortened and the disordered cysteines removed; other disulfides could be introduced to add stability; and the C-terminal segment could be used to link the RBD to an oligomeric core. Of the 23 glycosylation sites on S, three are in the RBD. Only one (Asn^{330}) is sufficiently ordered in our structure to show even a single sugar, and all are well separated from the RBM. Glycosylation is therefore unlikely to interfere with potential neutralizing epitopes within the RBD; introduction of new glycosylation sites could in principle "focus" the antigenicity of a candidate immunogen.

References and Notes

1. T. G. Ksiazek *et al.*, *N. Engl. J. Med.* **348**, 1953 (2003).
2. J. S. Peiris *et al.*, *Lancet* **361**, 1319 (2003).
3. M. A. Marra *et al.*, *Science* **300**, 1399 (2003).
4. P. A. Rota *et al.*, *Science* **300**, 1394 (2003).

5. M. M. C. Lai, K. V. Holmes, in *Fields' Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott. Williams. and Wilkins, Philadelphia, PA, 2001).
6. Y. Xu *et al.*, *J. Biol. Chem.* **279**, 49414 (2004).
7. Y. Xu *et al.*, *J. Biol. Chem.* **279**, 30514 (2004).
8. V. M. Supekar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17958 (2004).
9. W. Weissenhorn *et al.*, *Mol. Membr. Biol.* **16**, 3 (1999).
10. M. J. Moore *et al.*, *J. Virol.* **78**, 10628 (2004).
11. X. Xiao, S. Chakraborti, A. S. Dimitrov, K. Gramatikoff, D. S. Dimitrov, *Biochem. Biophys. Res. Commun.* **312**, 1159 (2003).
12. T. M. Gallagher, M. J. Buchmeier, *Virology* **279**, 371 (2001).
13. W. Li *et al.*, *Nature* **426**, 450 (2003).
14. H. Hofmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7988 (2005).
15. P. Towler *et al.*, *J. Biol. Chem.* **279**, 17996 (2004).
16. S. K. Wong, W. Li, M. J. Moore, H. Choe, M. Farzan, *J. Biol. Chem.* **279**, 3197 (2004).
17. G. J. Babcock, D. J. Esshaki, W. D. Thomas Jr., D. M. Ambrosino, *J. Virol.* **78**, 4552 (2004).
18. W. Li *et al.*, *EMBO J.* **24**, 1634 (2005).
19. Y. Guan *et al.*, *Science* **302**, 276 (2003).
20. H. D. Song *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2430 (2005).
21. J. Sui *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2536 (2004).
22. E. N. van den Brink *et al.*, *J. Virol.* **79**, 1635 (2005).
23. Y. He, H. Lu, P. Siddiqui, Y. Zhou, S. Jiang, *J. Immunol.* **174**, 4908 (2005).
24. Materials and methods are available as supporting materials on *Science* Online.
25. R. K. Williams, G. S. Jiang, K. V. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5533 (1991).
26. G. S. Dveksler *et al.*, *J. Virol.* **65**, 6881 (1991).
27. H. Kubo, Y. K. Yamada, F. Taguchi, *J. Virol.* **68**, 5403 (1994).
28. A. Bonavia, B. D. Zelus, D. E. Wentworth, P. J. Talbot, K. V. Holmes, *J. Virol.* **77**, 2530 (2003).
29. Chinese SARS Molecular Epidemiology Consortium, *Science* **303**, 1666 (2004).
30. W. Li *et al.*, *J. Virol.* **78**, 11429 (2004).
31. J. Sui *et al.*, *J. Virol.* **79**, 5900 (2005).
32. Y. He *et al.*, *Biochem. Biophys. Res. Commun.* **324**, 773 (2004).
33. F. Li, W. Li, M. Farzan, S. C. Harrison, data not shown.
34. B. Sainz Jr., J. M. Rausch, W. R. Gallaher, R. F. Garry, W. C. Wimley, *J. Virol.* **79**, 7195 (2005).
35. Y. Xu *et al.*, *Biochemistry* **43**, 14064 (2004).

36. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

37. We thank staff at the Advanced Light Source beamlines 8.2.1 and 8.2.2 for assistance and M. Berardi and E. Settembre for discussions. This work was supported by NIH grants CA13202 (to S.C.H.) and AI061601 (to M.R.F.). S.C.H. is an Investigator in the Howard Hughes Medical Institute. Coordinates and structure factors have been submitted to the Protein Data Bank with accession number 2AJF.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5742/1864/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

22 June 2005; accepted 11 August 2005
10.1126/science.1116480

Toward High-Resolution de Novo Structure Prediction for Small Proteins

Philip Bradley, Kira M. S. Misura, David Baker*

The prediction of protein structure from amino acid sequence is a grand challenge of computational molecular biology. By using a combination of improved low- and high-resolution conformational sampling methods, improved atomically detailed potential functions that capture the jigsaw puzzle–like packing of protein cores, and high-performance computing, high-resolution structure prediction (<1.5 angstroms) can be achieved for small protein domains (<85 residues). The primary bottleneck to consistent high-resolution prediction appears to be conformational sampling.

It has been known for more than 40 years that the three-dimensional structures of proteins are completely determined by their amino acid sequences (*1*), and the prediction of protein structure from amino acid sequence—the "de novo" structure prediction problem—is a long-standing challenge in computational biology and chemistry. Although there are notable exceptions, the majority of protein structures are likely to be at global free-energy minima for their amino acid sequences. The de novo protein structure prediction problem hence is to find the lowest free-energy structure for a specified amino acid sequence. The problem is challenging because the size of the conformational space to be searched is vast (*2*) and because the accurate calculation of the free energies of protein conformations in solvent is difficult.

Although there has been considerable progress in low-resolution de novo protein structure prediction (*3*), both the accuracy and the reliability of the structural models produced by these methods is fairly low: C_α-RMSDs (root mean square deviation of alpha-carbon coordinates after optimal superposition) of ~4 Å with incorrect packing of the amino acid side chains. Achieving higher resolution requires both more physically realistic energy functions and better conformational searching; the problem is difficult because the more realistic the energy function, the more rugged the landscape, and thus the more difficult it is to search. Here, we show that high-resolution de novo structure prediction can be achieved by generating structurally diverse populations of low-resolution models and refining these structures in the context of a physically realistic all-atom energy function.

Critical to high-resolution structure prediction is a force field for which native structures are low in free energy compared with non-native structures and a refinement protocol that can efficiently navigate the corresponding free-energy landscape. We have developed an all-atom force field (*4*) that focuses on short-range interactions—primarily van der Waals packing, hydrogen bonding, and desolvation—while neglecting long-range electrostatics. The high-resolution refinement protocol (*5*, *6*) is designed to search in the local neighborhood of a starting model for low-energy structures. The protocol consists of multiple rounds of Metropolis Monte Carlo with minimization (*7*); each trial consists of a random perturbation of one or several backbone torsion angles, fast side-chain optimization using a rotamer representation (*8*, *9*), and a gradient-based minimization of the energy function with respect to backbone and side-chain torsion angles. In this way, the continuous space of backbone conformations and the discrete set of side-chain packing arrangements are searched simultaneously. Details on the energy function and methods are provided in (*10*).

Figure 1 and fig. S1 illustrate the challenge of high-resolution de novo structure prediction. All-atom refinement trajectories begun at the native state produce models (refined natives) that sample a deep near-native free-energy basin. Although these structures typically have lower all-atom energies than do non-native structures, Rosetta de novo models—built from an extended-chain starting conformation—do not sample close enough to the native structure to fall into this narrow energy well during all-atom refinement. The narrow widths of the native basins reflect the fact that nativelike side-chain packing can be disrupted by even relatively small backbone perturbations. Thus, the critical step in high-resolution structure prediction is generating low-resolution models that are within the "radius of convergence" of the native free-energy minimum using the all-atom refinement protocol. This is challenging, because the low-resolution search integrates out the side-chain degrees of freedom to smooth the energy landscape and hence lacks the detail necessary to reliably discriminate nativelike models, leading to false minima. We attempt to overcome this problem by generating low-resolution models for a large number of se-

University of Washington, Department of Biochemistry, and Howard Hughes Medical Institute, Box 357350, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu