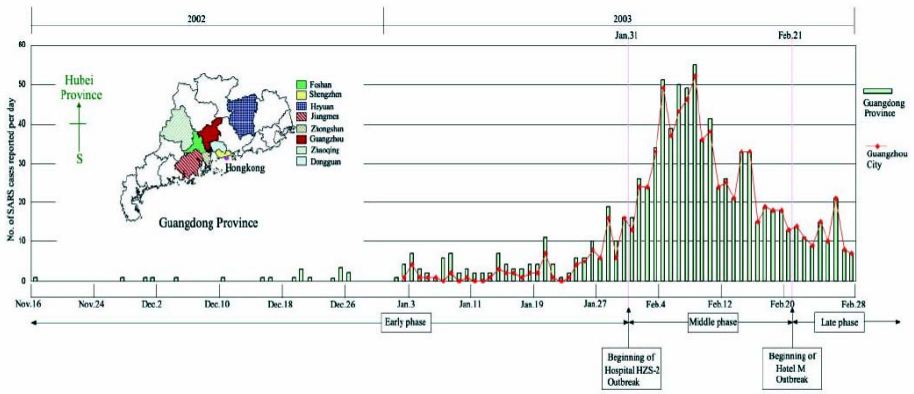
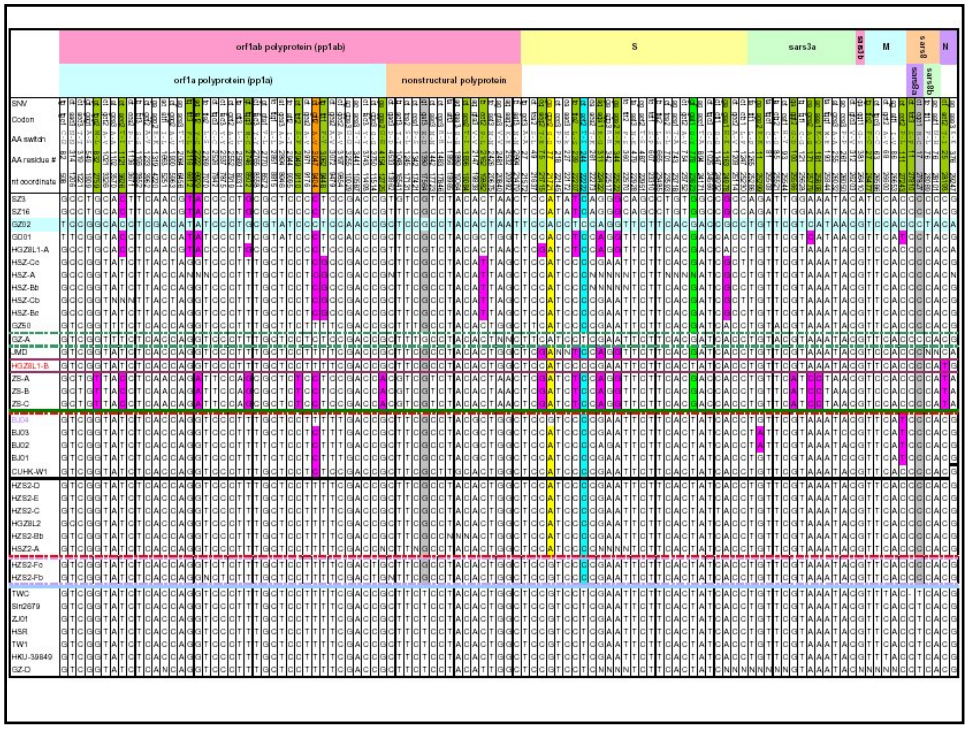


Fig. 1. The triphasic SARS epidemic in Guangdong Province, China. Shown are daily numbers of SARS cases reported in Guangdong Province, in particular the city of Guangzhou. The early, middle, and late phases of the epidemic are defined in the text. The map shows the geographical distribution of cases belonging to the early phase by administrative districts of Guangdong Province. The detailed data for individual cities are presented in fig. S1.

C. List of GenBank accession numbers for sequences mentioned in the text and SOM:

Sequences generated by this study	GenBank accession number	Sequences previously available	GenBank accession number
GD03T13 (S gene)	AY525636	SZ16 (palm civet)	AY304488
		SZ3 (palm civet)	AY304486
GZ02	AY390556	GD01 (GZ01)	AY278489
HGZ8L1-A	AY394981		
HSZ-A	AY394984		
HSZ-B (b, c)	AY394985, AY394994		
HSZ-C (b, c)	AY394986, AY394995		
ZS-A	AY394997	gz43 (S gene)	AY304490
ZS-B	AY394996	gz60 (S gene)	AY304491
ZS-C	AY395003		
GZ-A	AY394977		
JMD	AY394988		
HGZ8L1-B	AY394982		
HZS2-A	AY394983	CUHK-W1	AY278554
HZS2-Bb	AY395004	BJ04	AY279354
HZS2-C	AY394992	BJ01	AY278488
HZS2-D	AY394989	BJ02	AY278487
HZS2-E	AY394990	BJ03	AY278490
HGZ8L-2	AY394993		
HZS2-Fc	AY394991		
HZS2-Fb	AY394987		
CUHK-LC1	AY394998		
GZ-B	AY394978	TOR2	AY274119
GZ-C	AY394979	ZJ01	AY297028
GZ-D	AY394980		
CUHK-LC2	AY394999	CUHK-AG01	AY345986
CUHK-LC3	AY395000	CUHK-AG02	AY345987
CUHK-LC4	AY395001		
CUHK-LC5	AY395002		


orf1ab polyprotein (pp1ab)
S
sars3a
sars3b
M
s sars9
N
orf1a polyprotein (pp1a)
nonstructural polyprotein
sars3b
s sars9
SNV
Codon
AA switch
AA residue #
nt coordinate
SZ3
SZ16
SZ02
SD01
HGZ8L1-A
HSZ-Cc
HSZ-A
HSZ-Bb
HSZ-Cb
HSZ-Bc
SZ50
GZ-A
JMD
HGZ8L1-B
ZS-A
ZS-B
ZS-C
BJ04
BJ03
BJ02
BJ01
CUHK-W1
HZS2-D
HZS2-E
HZS2-C
HGZ8L2
HZS2-Bb
HSZ2-A
HZS2-Fc
HZS2-Fb
TWC
Btr2679
ZJ01
HSR
TW1
HKU-39849
SZ-D

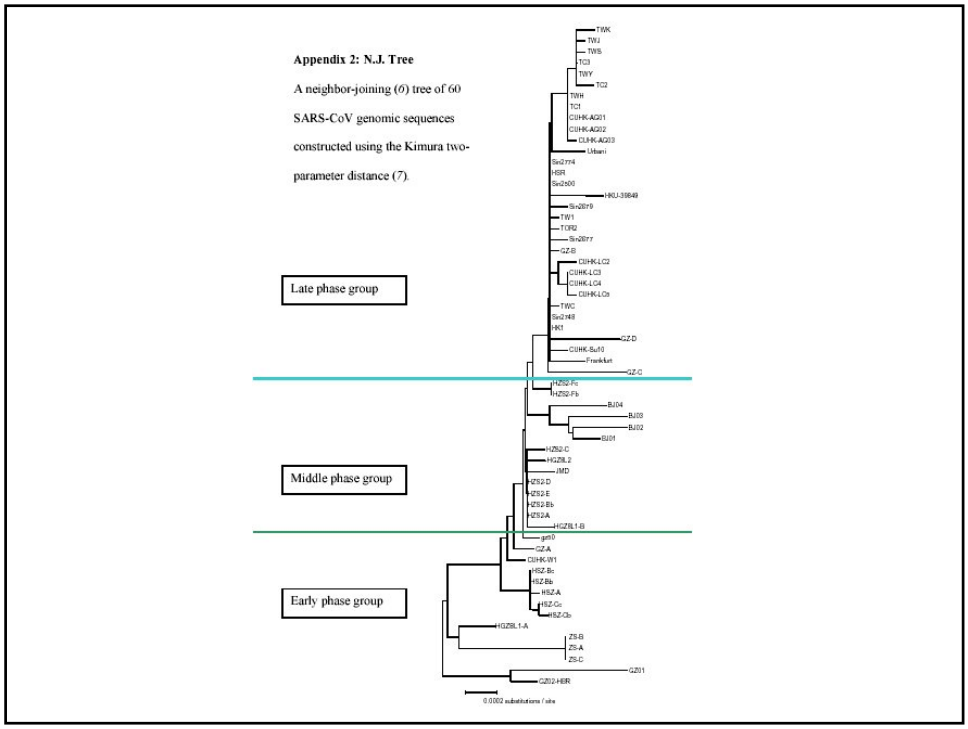

Appendix 2: N.J. Tree
A neighbor-joining (6) tree of 60 SARS-CoV genomic sequences constructed using the Kimura two-parameter distance (7).
Late phase group
Middle phase group
Early phase group
TWK
TWJ
TWB
TC3
TWY
TC2
TWH
TC1
CUHK-AG01
CUHK-AG02
CUHK-AG03
Urbani
Sin2774
HSR
Sin2500
HKU-39849
Sin281
TW1
TOR2
Sin2677
GZ-B
CUHK-LC2
CUHK-LC3
CUHK-LC4
CUHK-LC5
TWC
Sin2748
HK1
GZ-D
CUHK-Su10
Frankfurt
GZ-C
HZS2-Fc
HZS2-Fb
BJ04
BJ03
BJ02
BJ01
HZS2-C
HGZ8L2
JMD
HZS2-D
HZS2-E
HZS2-Bb
HZS2-A
HGZ8L1-B
gd01
GZ-A
CUHK-W1
HSZ-Bc
HSZ-Bb
HSZ-A
HSZ-Cc
HSZ-Cb
HGZ8L1-A
ZS-B
ZS-A
ZS-C
GZ01
GZ02-HSR
0.0002 substitutions / site

Table 1. Statistical analysis for the change of Ka/Ks ratios for different coding regions of the SARS-CoV sequences during the different epidemic phases.

Proteins	Epidemic phases	$\overline{Ks}(10^{-3})$	$\overline{Ka}(10^{-3})$	$\overline{Ka/Ks}$	$s.e.(\overline{Ka/Ks})$	H_1*	P-value†
Spike	early	2.113	2.647	1.321	0.0609	Ka/Ks (early) > Ka/Ks (middle)	0.007
	middle	1.556	0.950	0.731	0.1882		
	late	1.355	0.258	0.219	0.0430	Ka/Ks (middle) > Ka/Ks (late)	0.0135
Orf1b	early	1.233	0.678	0.800	0.1095	Ka/Ks (early) > Ka/Ks (late)	2.5×10^{-4}
	middle	1.124	0.208	0.213	0.0527		
	late	0.577	0.159	0.344	0.0476		
Orf1a	early	1.167	0.953	0.936	0.0821	Ka/Ks (early) > Ka/Ks (late)	$<1 \times 10^{-5}$
	middle	0.434	0.637	1.859	0.2519		
	late	0.557	0.139	0.369	0.0601		

*H_1 means the alternative hypothesis.

†One-sided unpaired two-sample t-test was used.

Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China

The Chinese SARS Molecular Epidemiology Consortium*

Sixty-one SARS coronavirus genomic sequences derived from the early, middle, and late phases of the severe acute respiratory syndrome (SARS) epidemic were analyzed together with two viral sequences from palm civets. Genotypes characteristic of each phase were discovered, and the earliest genotypes were similar to the animal SARS-like coronaviruses. Major deletions were observed in the Orf8 region of the genome, both at the start and the end of the epidemic. The neutral mutation rate of the viral genome was constant but the amino acid substitution rate of the coding sequences slowed during the course of the epidemic. The spike protein showed the strongest initial responses to positive selection pressures, followed by subsequent purifying selection and eventual stabilization.

Severe acute respiratory syndrome (SARS) first emerged in Guangdong Province, China. Subsequently, the SARS coronavirus (SARS-CoV) was identified as the causative agent (*1–5*). It remains a challenge to establish the relationship between observed genomic variations and the biology of SARS (*4–8*). Recent molecular epidemiological studies have identified characteristic variant sequences in SARS-CoV for tracking disease transmission (*7, 9–11*). Evidence suggests that SARS-CoV emerged from nonhuman sources (*8, 12*). In this study, we sought epidemiological and genetic evidence for viral adaptation to human beings through molecular investigations of the characteristic viral lineages found in China (*13*).

On the basis of epidemiological investigations (*14*), we divided the course of the epidemic into early, middle, and late phases (Fig. 1). The early phase is defined as the period from the first emergence of SARS to the first documented superspreader event (SSE) (*13*). The middle phase refers to the ensuing events up to the first cluster of SARS cases in a hotel (Hotel M) in Hong Kong (*15*). Cases following this cluster fall the retrospectively identified SARS index patient from the city of Foshan (onset date, 16 November 2002) (*13*) through to an index patient from the city of Dongguan (onset date, 10 March 2003). All of these cases were confined to regions directly west of Guangzhou, the capital city of Guangdong Province, and to the city of Shenzhen in the south, with no cases being reported to the north or east of Guangzhou (Fig. 1) (fig. S1). This region, the Pearl River Delta, has enjoyed rapid economic development since the late 1970s, leading to the adoption of culinary habits requiring exotic animals. Seven of these 11 cases had documented contact with wild animals. In contrast to the apparently independent seeding of the earliest cases, the rest of the epidemic was characterized by SSEs and clusters of cases that were epidemiologically linked (Fig. 1) (fig. S1) (*10, 11, 13, 15, 16*).

The first major SARS outbreak occurred in a hospital, HZS-2, in the city of Guangzhou, beginning on 31 January 2003 where an SSE was identified to be associated with more than 130 primary and secondary infections, of which 106 were hospital-acquired cases. Doctor A, a nephrologist who worked in this hospital, visited Hong Kong and stayed in Hotel M on 21 February 2003. Other visitors to the hotel later became infected with SARS-CoV (*13, 15*). This led to the transmission of SARS to Vietnam, Canada, Singapore, and the United States (*17*) with two further SSEs in Hong Kong, each resulting in the virus being transmitted to >100 contacts (*10, 16*).

Genomic sequence data for SARS-CoV were largely derived from isolates linked to the Hotel M cluster (*6*), hence they were predominantly from the late phase of the epidemic. We determined 29 SARS-CoV genomic sequences obtained from 22 patients from Guangdong Province with disease onset

*Epidemiology group: Jian-Feng He,[1]† Guo-Wen Peng,[1] Jun Min,[2] De-Wen Yu,[1] Wen-Jia Liang,[1] Shu-Yi Zhang,[2] Rui-Heng Xu[1]‡ Virology group: Huan-Ying Zheng,[1]† Xin-Wei Wu,[4]† Jun Xu,[5] Zhan-Hui Wang,[6] Ling Fang,[1] Xin Zhang,[1] Hui Li,[1] Xin-Ge Yan,[1] Jia-Hai Lu,[1] Zhi-Hong Hu,[7] Ji-Cheng Huang,[1] Zhuo-Yue Wan,[1]‡ Jin-Lin Hou,[5]‡ Jin-Yan Lin[1]‡ Molecular biology group: Huai-Dong Song,[4]† Sheng-Yue Wang,[8]† Xiang-Jun Zhou,[10]† Guo-Wei Zhang,[4] Bo-Wei Gu,[9] Hua-Jun Zheng,[9] Xiang-Lin Zhang,[9] Mei He,[10] Kui Zheng,[1] Bo-Fei Wang,[4] Gang Fu,[4] Xiao-Ning Wang,[4] Sai-Juan Chen,[6] Zhu Chen[6]‡ Data analysis group: Pei Hao,[11,12]† Hua Tang,[12]† Shuang-Xi Ren,[8]† Yang Zhong,[12,14]† Zong-Ming Guo,[12] Qi Liu,[11] You-Gang Miao,[11] Xiang-Yin Kong,[11] Wei-Zhong He,[12] Yi-Xue Li,[11,12] Chung-I Wu,[13]‡ Guo-Ping Zhao[6,11]§ Chinese University of Hong Kong group: Rossa W. K. Chiu,[15]† Stephen S. C. Chim,[15] Yu-kwan Tong,[15] Paul K. S. Chan,[16] John S. Tam,[16] Y. M. Dennis Lo[15]‡

Science (2004)

Available online at www.sciencedirect.com
SCIENCE DIRECT
Computational Biology and Chemistry
ELSEVIER
Computational Biology and Chemistry 29 (2005) 254–257
www.elsevier.com/locate/compbiolchem

Short communication

A molecular docking model of SARS-CoV S1 protein in complex with its receptor, human ACE2

Yuan Zhang [a], Nan Zheng [b], Pei Hao [c], Ying Cao [d,e], Yang Zhong [a,*]

[a] School of Life Sciences, Fudan University, Shanghai 200433, China
[b] Institute of Viral Disease Control and Prevention, Chinese CDC, Beijing 100052, China
[c] Shanghai Center for Bioinformation Technology, Shanghai 200235, China
[d] The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan
[e] Department of Biosystems Science, The Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa 240-0193, Japan

Received 17 April 2005; accepted 18 April 2005

Abstract

The exact residues within severe acute respiratory syndrome coronavirus (SARS-CoV) S1 protein and its receptor, human ACE2, involved in their interaction still remain largely undetermined. Identification of exact amino acid residues that are crucial for the interaction of S1 with ACE2 could provide working hypotheses for experimental studies and might be helpful for the development of antiviral inhibitor. In this paper, a molecular docking model of SARS-CoV S1 protein in complex with human ACE2 was constructed. The interacting residue pairs within this complex model and their contact types were also identified. Our model, supported by significant biochemical evidence, suggested receptor-binding residues were concentrated in two segments of S1 protein. In contrast, the interfacial residues in ACE2, though close to each other in tertiary structure, were found to be widely scattered in the primary sequence. In particular, the S1 residue ARG453 and ACE2 residue LYS341 might be the key residues in the complex formation.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Severe acute respiratory syndrome coronavirus; Spike protein; Angiotensin-converting enzyme 2; Receptor binding; Protein docking

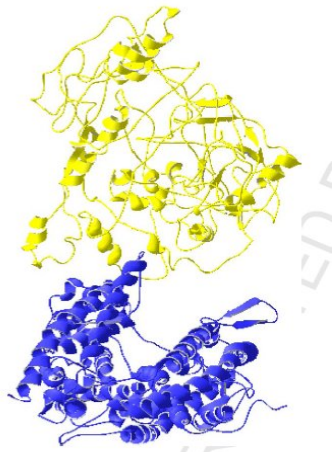


Fig. 1. Ribbon diagram of the SARS-CoV S1 (yellow)/ACE2 (blue) complex model (PDB code: 1XJF). The theoretical model of S1 domain (PDB code: 1Q4Z) and the native crystal structure of the human ACE2 extracellular domain (PDB code: 1R42) were downloaded from the protein data bank (PDB). This model was generated by the fully automatic ZDOCK protein–protein docking server and manually selected on the basis of structural biology knowledge.

Table 1
The interacting residue pairs in the S1/ACE2 complex model

S1 residue	ACE2 residue	Contact type
VAL307	VAL298	Hydrophobic interaction
VAL308	VAL298	
VAL308	VAL364	
ASP312	THR334	
PRO450	LYS341	
ARG453	LYS341	
ARG449	GLU57	Electrostatic interaction (attractive)
ARG453	GLU56	
ASP454	LYS341	
ASP312	ASP335	Electrostatic interaction (repulsive)
ARG453	LYS341	Hydrogen bond
ASP312	THR334	

leting *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

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-

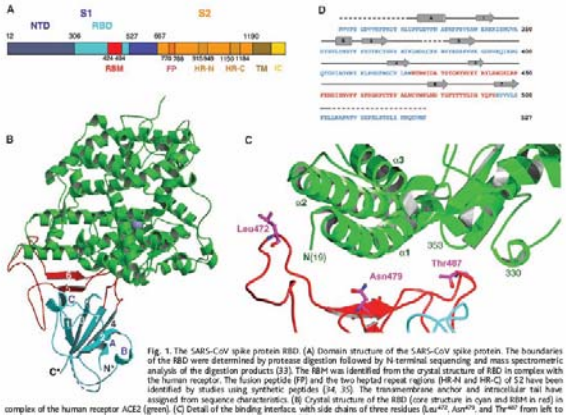


Fig. 1. The SARS-CoV spike protein RBD. (A) Domain structure of the SARS-CoV spike protein. The boundaries of the RBD were determined by protease digestion followed by N-terminal sequencing and mass spectrometric analysis of the digestion products (*33*). The RBM was identified from the crystal structure 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

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