

Understanding the SARS Genome/Proteome Using DS Gene, DS GeneAtlas, and DS AtlasStore

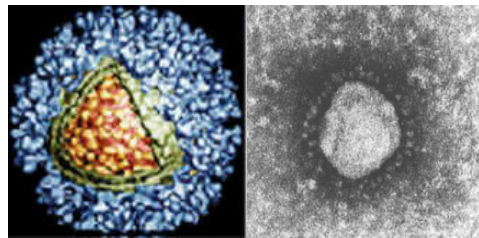
Publication by Lisa Yan, Mikhail Velikanov, Paul Flook, Wenjin Zheng, Sándor Szalma, and Scott Kahn, “Assessment of Putative Protein Targets Derived from the SARS Genome,” *FEBS Letters* **2003**, 554, 257-263.

Rapidly and reliably developing hypotheses on the function of newly discovered protein sequences requires systematic and comprehensive analysis. Such an analysis, embodied within the Discovery Studio® GeneAtlas™ pipeline, has been used to critically evaluate the SARS genome with the goal of identifying new potential targets for therapeutic intervention. This study discusses several new functional hypotheses on the roles played by the constituent gene products of SARS, and serves as an example of how such assignments can be developed or extended on other systems of interest. The goal of this study is to identify new potential targets for SARS therapeutic intervention.

Understanding SARS

SARS stands for Severe Acute Respiratory Syndrome, and was first identified in Guangdong Province, China in November 2002 during an outbreak. Since then, the outbreak has spread to several countries and has had significant health and economic impact.

- Introduced into Hong Kong in February 2003 by visiting health care worker
- Rapidly spread to 30 countries since then
- 916 deaths as of September 2003



(Left) Schematic view of a crystallized virus particle. (Right) Electron microscopy image of the SARS virus (<http://www.scienceinAfrica.co.za/2003/june/sars.htm>)

The SARS Virus

The SARS virus is a large, enveloped, RNA virus with club-like projections that give them a halo appearance. The SARS virus is a Corona viruses which are ubiquitous viruses that cause cold infections, acute respiratory distress (~30% of infections), sore throat, malaise, fever, headache and cough, mild pneumonia, and acute gastroenteritis

The virus infects the epithelium of both the respiratory and digestive tracts. In the digestive tract it reduces absorptive capacity and causes diarrhea, dehydration, and electrolyte imbalances. In addition, the virus is not handled well by vaccines.

It is suspected that the SARS virus can be transmitted by airborne transmission (surviving in dry air/surfaces for up to 3 hours in a crystallized form).

The SARS Genome

The SARS-CoV genome contains ~30k bases (Tor2 strain) and contains five major open reading frames (ORFs) that encode the following proteins:

Replicase polyprotein (cleaved into replicase 1a and 1b)

- orf1a (polypeptide 1a, pp1a) and orf1b (polypeptide 1b, pp1b) are believed to be polyprotein constructs that are cleaved to form the constituent proteins post transcription

Nucleocapsid protein (N)

- Binds and packages RNA genomes

Spike (S), envelope (E), and membrane (M) glycoproteins

- The S protein binds the host cell receptors and induces fusion, which is important for determining virulence and species
- The M and E proteins interact with the N protein and control envelope function

BJ01	Tor2	CDCP	Start	Stop	Frame	Actual Start	Actual Stop	AA Length
1a	1a	1a	265	13398	1	265	13413	4382
1b	1b	1b	13398	21485	3	13398	21485	2695
S	S	S	21492	25259	3	21492	25259	1255
1	3	X1	25268	26092	2	25268	26092	274
2	4	X2	25689	26153	3	25689	26153	154
E	E	E	26117	26347	2	26117	26347	76
M	M	M	26398	27063	1	26398	27063	221
3	7	X3	27074	27265	2	27074	27265	63
4	8	X4	27273	27641	3	27273	27641	122
N/A	9	N/A	27638	27772	2	27638	27772	44
N/A	10	N/A	27779	27898	2	27779	27898	39
N/A	11	X5	27864	28118	3	27864	28118	84
N	N	N	28120	29388	1	28120	29388	422
5	13	N/A	28130	28426	2	28130	28426	98
N/A	14	N/A	28583	28795	2	28583	28795	70
N/A	s2m motif	N/A	29590	29621	N/A	29590	29621	N/A

This table shows the ORF information for 3 SARS strains: BJ01, Tor2, and CDCP.

Methods

Overview of Steps

1. Analyzed different strains variations, ensuring that all non-synonymous polymorphisms were consistent with the assignments being proposed using DS Gene
2. Predicted protein sequences from the different strains based on putative cleavage sites
3. Genome data and protein transcripts were extracted from GenBank for all the known isolates of SARS virus. In many cases the assignments of all resulting protein sequences contained inferences to structural homologies.
4. Protein products were submitted to DS GeneAtlas, a high-throughput functional proteomics annotation pipeline
5. Results were stored and analyzed in DS AtlasStore

Polyprotein Sequences Analyzed in DS GeneAtlas

Polyprotein products from pp1ab were produced by cleavage with enzymes 3CL main protease (Mpro) and the papain-like accessory proteinase (PLP).

Mpro

- Mpro cleavage sites were identified using [GAVSTP]XLQ[SAGN] motif, with cleavage occurring immediately after Gln residue. Yielded 11 cleavage sites

PLP

- PLP cleavage sites were identified only in the first (N-terminal) product of Mpro cleavage (residues 1-3240) using the motif [RK]XXXG[GA]. Yielded 6 cleavage sites.

We used the polyprotein pp1a and pp1ab (includes pp1b) as well as the cleaved products defined by the putative cleavage site as input to DS GeneAtlas.

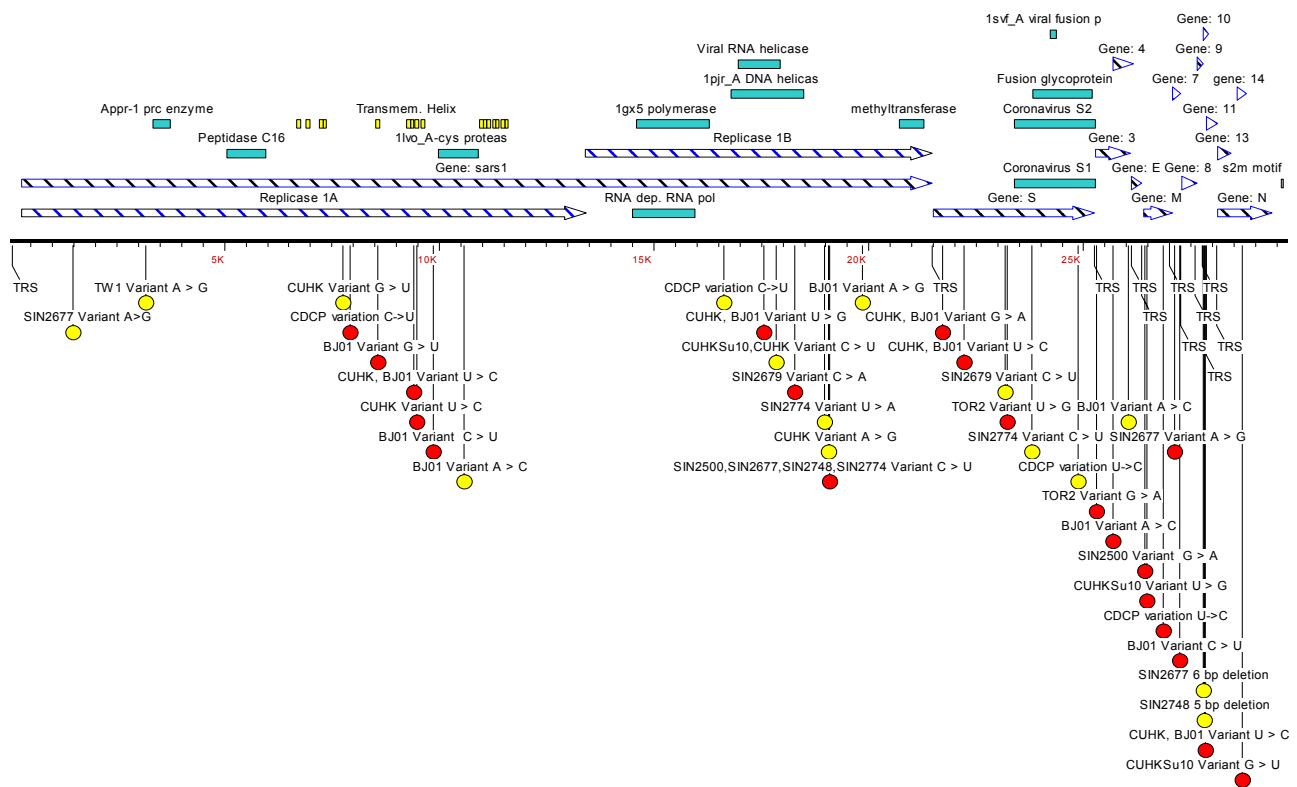
Protein Transcripts of SARS Isolates and their Genomic Coordinates

- 11 protein sequences from BJ01 and CDCP strains
- 15 protein sequences of the Tor2 strain

SARS Genome

All predicted structural domains and functional domains mapped to the genome sequence of Tor2 strain and displayed below using DS Gene

- DS GeneAtlas structural and functional annotations (in **green**)
- Putative protein transcripts (in **blue arrow**)
- Non-synonymous Single Nucleotide Polymorphisms (SNPs) (in **red dots**) give rise to amino acid residue changes
- Synonymous SNPs (in **yellow dots**)



Structural and Functional Annotations Obtained by DS GeneAtlas

Structural and functional domains annotated for three sequences: polyprotein1a (pp1a), polyprotein 1b (pp1b), and S protein

Domain	Methods	Template and function	Scores	Confidence
pp1a 3241-3543	Structure	1lvo_A cysteine like protease (TGEV)	PSI-BLAST E-value = 0 Model Score = 0.96; Seq-ID% = 43.9%	Consensus_Score = 1.76 high
pp1a 1026-1154	HMMer/Pfam	Appr-1"-p processing enzyme family	E-value = 1.1e-20 Bit score = 78.3; Noise cutoff = -21.5	high
pp1a 1598-1893	HMMer/Pfam	Peptidase C16 family	E-value = 0.043 Bit score = -87.5; Noise cutoff = -95.7	medium
pp1b 4780-5334 ^c	Structure	1gx5 RNA dependent RNA-polymerase (HCV)	PSI-BLAST E-value = 9.2e-25 Model Score = -0.10; Seq-ID% = 10.2%	Consensus score = 0.5 Medium
pp1b 4770-5249 ^d	Structure	1gx5 RNA dependent RNA-polymerase (HCV)	PSI-BLAST E-value = 1.4e-49 Model Score = -0.21; Seq-ID% = 10.4%	Consensus score = 0.41 low
pp1b 5512-6066	Structure	1pjr_A DNA helicase	PSI-BLAST E-value = 3.4e-78 Model Score = -0.23; Seq-ID% = 8.2%	Consensus score = 0.23 low
pp1b 4747-5219	HMMer/Pfam	RNA dependent RNA polymerase	E-value = 0.093 Bit score = -194.6; Noise cutoff = -130.1	low
pp1b 5569-5887	HMMer/Pfam	Viral (Superfamily 1) RNA helicase	E-value = 0.0058 Bit score = -49.6; Noise cutoff = -30.1	low
pp1b 6815-6998	HMMer/Pfam	FtsJ-like methyltransferase	E-value = 0.0044 Bit score = -51.6; Noise cutoff = -53.1	medium
S protein 910-949	Structure	1svf_A viral fusion protein core	PSI-BLAST E-value = 8e-05 Model Score = -0.29; Seq-ID% = 17.5%	Consensus score = 0.29 low
S protein 56-619	HMMer/Pfam	Coronavirus S1 glycoprotein	E-value = 0.87 Bit score = -283.4; Noise cutoff = -273.1	low
S protein 631-1255	HMMer/Pfam	Coronavirus S2 glycoprotein	E-value = 1.3e-132 Bit score = 450.2; Noise cutoff = -460.8	High
S protein 777-1231	HMMer/Pfam	Fusion glycoprotein F0	E-value = 0.031 Bit score = -276.3; Noise cutoff = -236.1	Low

Summary of protein domain, annotation methods, template and or functional annotation, method type, and noise cutoff, and confidence level (consensus score).

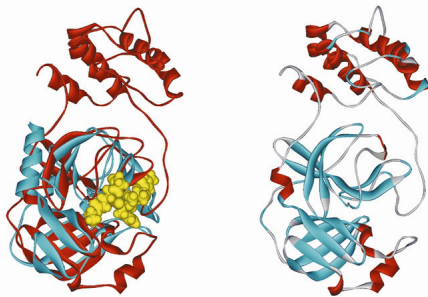
3D Model of SARS Polypeptide 1a

(3241 to 3543 Residue Domain)

PP1a is homologous to coronavirus main cysteine proteinase (Mpro, 1lvo). Another coronavirus is porcine transmissible gastroenteritis virus (TGEV).

The 3D model is built based on the template 1lvo (chain A)

- Shares a common fold with 1cqq chain A - the human rhinoviral protease, a 3C cysteine protease (3Cpro)
- The Mpro (1lvo) has an additional helical domain at C-terminus
- Mpro and 3Cpro have very low sequence similarity given the common fold; less than 10% sequence identity given the common fold, their ligand binding pockets are located at same position in the cleft between the two beta domains and may bind to similar ligand



Left: Superimposed crystal structures of Mpro (1lvo_A in red) and 3Cpro (1cqq_A in cyan) with ligand AG7088 drug candidate (in yellow)
Right: Model structure of SARS Mpro (residues 3241-3543) predicted using DS GeneAtlas

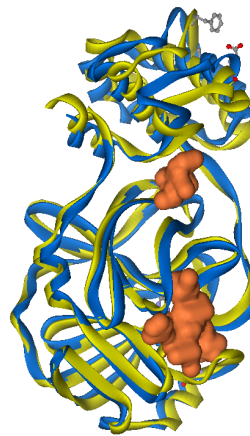
Crystal Structure of SARS Main Protease

After our calculation was completed, two new similar structures of the main protease (Mpro) were solved by ray crystallography.

- Human coronavirus 229E (HCoV): sequence identity 41% between SARS and HCoV Mpro
- Porcine transmissible gastroenteritis virus (TGEV) with bound ligand: RMSD < 1.0 Å without major conformational changes around the ligand-binding site, sequence identity between 44% between SARS Mpro and TGEV Mpro

Subsequently, the crystal structure of the SARS Mpro protein was determined (1q2w)

- RMSD 2.45 Å over 295 residues
- RMSD 1.32 Å over 220 residues in ligand binding site

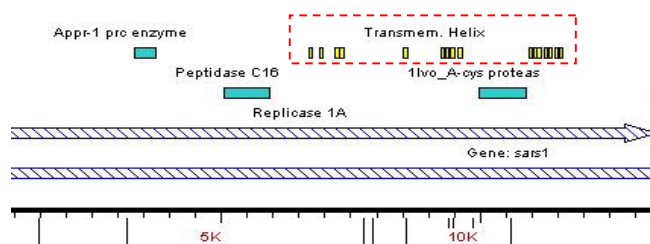


Actual crystal structure of SARS Mpro (1q2w, blue) superimposed with DS GeneAtlas model with 3D annotation (yellow). Model binding pocket predictions displayed as (orange) surface displays.

X-

Novel Transmembrane Domain in SARS Polypeptide 1a

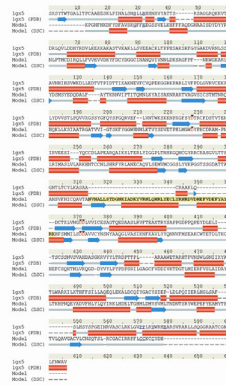
A large transmembrane (TM) domain from residues 3561 to 3774 was found, based on the predicted cleavage site of the main proteases. This TM domain is part of the cleaved product from residues 3547 to 3919 and may represent a 7 TM protein motif if the seven transmembrane helices are indeed distinct as currently assigned.



SARS Polypeptide 1b

A new functional assignment was made to residues 6815 to 6998. As FtsJ-like methyltransferase Pfam analysis. Further analysis of this domain is in progress.

DS GeneAtlas identified several known RNA dependent RNA polymerases from hepatitis C virus (HCV) as templates for domain 4780-5334. The sequence alignment was between template (1gx5) and the model of RNA dependent RNA-polymerase.



The catalytic residues are annotated with carrot (in red) and the residues in surface pocket is annotated with underline (in blue) The long insertion in the model sequence where the structure is uncertain is colored with yellow background

RNA dependent RNA-polymerase Model of SARS PP1b Domain 4780-5334

The model created based on the 1gx5 template is the longest model with the most complete structure. The catalytic triad Asp220, Asp318, Asp319 in the template is conserved in the model-template alignment.

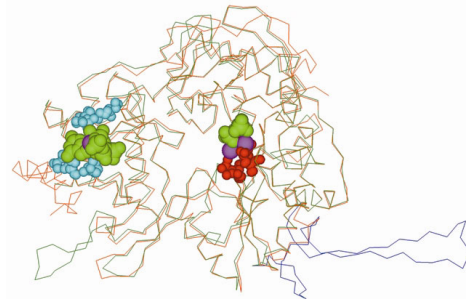
The crystal structure of the HCV RNA polymerase has two ligand binding sites:

Catalytic reaction center

- No large gaps in this region
- More conserved than surface location

Surface location

- About 30 Å from the catalytic center
- Few large insertions in model
- Not conserved between template and model
- Suggests that the SARS virus replication is likely to be activated by other means or by a different type of ligand

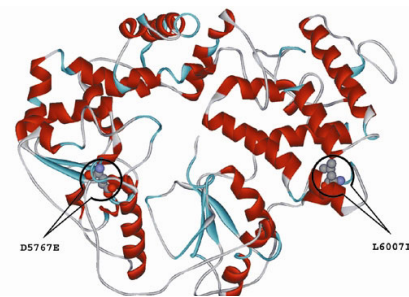


Model structure (in dark green) of RNA dependent RNA-polymerase domain superimposed with template structure 1gx5 (in brown) The rGTP (riboguanosine triphosphate) ligands from the template are shown in green and the Mn ions are in purple The catalytic triad is in red and the other ligand binding site is in cyan. The long insertion in the model where the structure is uncertain is colored dark blue.

Non-synonymous SNPs Mapped to DNA Helicase Structure

Using DS Gene, we found 35 single nucleotide polymorphisms (SNPs) for SARS virus by comparing the genome sequences from different SARS strains in the GenBank database. Twenty of them are non-synonymous mutations resulting in amino acid residue changes.

- Three non-synonymous SNPs V2770L, V3047A, and V3072A are inside the predicted TM helices or near the end of the TM helix of pp1a
- All conserved mutations which are unlikely to change the property of the TM domain
- Two additional SNPs, D5767E and L6007I are predicted in the DNA helicase domain 5512-6066 of pp1b
- Other SNPs are not mapped to any structural or functional domains that are predicted



Non-synonymous SNPs mapped to DNA helicase model based on 1pjr_A

SARS Spike (S) Protein

The domain from residues 910-949 of S protein is found to have medium similarity to the viral fusion protein. A model was built based on template 1svf chain A. The template adopts a coiled-coil fold with four helices from four monomers and is a repeat of two different monomers, chain A and chain B. We did not find any domain that was homologous to chain B, and this can be attributed to the short length of chain B with only 38 residues. This is consistent with the knowledge that it is difficult to find sequence matches to short segment using PSI-BLAST program. The similarity to viral fusion protein was confirmed by a Pfam search, even though the confidence level of the HMMer score was low. It is known that the C-terminal domain of coronavirus glycoprotein forms coiled-coil structure during the cell membrane fusion process

Residues 1195-1217 were predicted to be a transmembrane helix. This differs notably from the prediction by Rota *et al.*, who predicted a longer 37 residue transmembrane domain. Our predicted TM domain is within the TM domain reported by Rota *et al.* with a few polar residues from the N-terminal side and the cysteine rich segment from the C-terminal side removed.

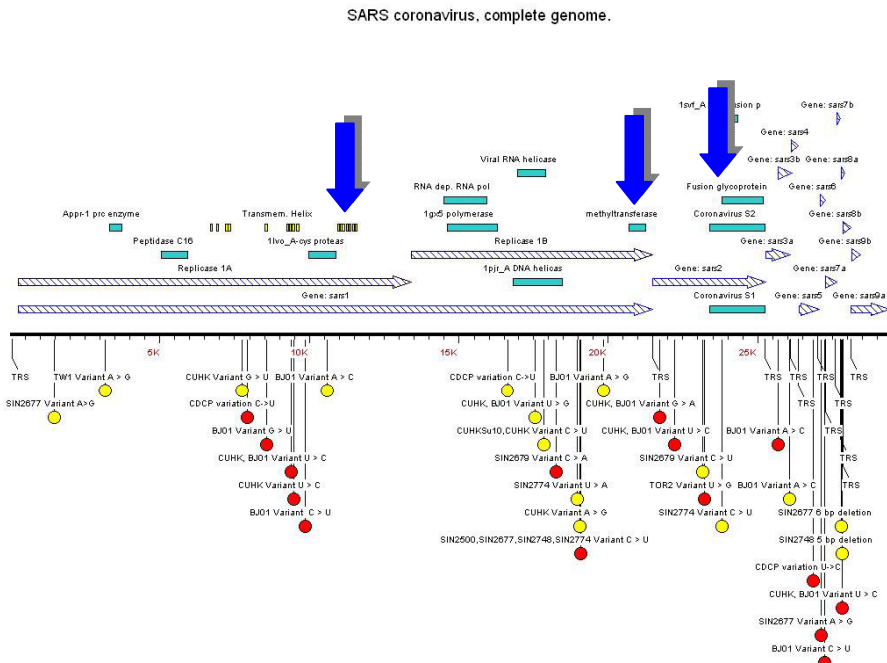
Summary of the SARS Genome Study

Comprehensive analysis of the SARS genome proposes several novel hypotheses regarding the function of the gene products of the SARS genome using an automated assignment pipeline.

Most of the hypotheses are supported by multiple methods of assignment in DS GeneAtlas. In summary:

- We confirmed the existing annotations including 3D annotation
- We proposed 3 new targets for drug discovery (blue arrows): 7 Transmembrane protein, FtsJ methyltransferase, and Fusion protein
- Found that DS Gene and DS GeneAtlas were robust for studying polymorphisms due to SNPs (e.g. DNA helicase)

We propose that such automated development of functional hypotheses successfully extends what is known about SARS, and in most cases suggests follow-on experiments to refine these hypotheses further.



References

1. Drosten, C., *et al.*, (2003) N Engl J Med 348(20): 1967-76.
2. Ksiazek, T.G., *et al.*, (2003) N Engl J Med 348(20): 1953-66
3. Marra, M., *et al.*, (2003) Science 300:1399-1404.
4. Rota, P. A., *et al.*, (2003) Science 300:1394-1399
5. Lisa Yan, Mikhail Velikanov, Paul Flook, Wenjin Zheng, Sándor Szalma, and Scott Kahn. Assessment of Putative Protein Targets Derived from the SARS Genome, FEBS Letters, 554 (2003), 257-263.
6. Kitson, D. H., *et al.*, (2002) Briefings in Bioinformatics., 3:32-44.
7. Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., (2000) J. Gen. Virol., 81:853
8. Hegyi, A., Ziebuhr, J., (2002) J. Gen. Virol., 83:595
9. Hughes, S.A., Bonilla, P.J., Weiss, S.R., (1995) J. Virol., 69:809
10. Ziebuhr, J., Thiel, V., Gorbalenya, A.E., (2001) J. Biol. Chem., 276:33220
11. Sequences from NCBI: <http://www.ncbi.nlm.nih.gov/>
12. Kitson, D., *et al.* Briefings in Bioinform.,3:1-13, (2002)
13. Anand, K., *et al.*, (2003) Science, 300:1763-1767
14. Figures were created using DS Modeling 1.1 software from Accelrys, Inc.
15. Anand, K., *et al.*, (2003) Science, 300:1763-1767
16. GenBank database <http://www.ncbi.nlm.nih.gov/Genbank/index.html>
17. Rota, P. A., *et al.*, (2003) Science 300:1394-1399

Snapshots of DS GeneAtlas SARS analysis results

Sequence Annotation List - SARS

Sequence Annotation List

1-17 of 17 Sequences

Sequence ID	Name	Function of the Best Hit	Hits	Domains	Best Hit	AS
SARS000001	AAP30028	HYDROLASE 12-MAY-03 1P95 CORONAVIRUS MAIN PROTEINASE (3CLPROV) STRUCTURE: BASIS FOR DESIGN OF ANTI-SARS DRUGS MOL_ID: 1; MOLECULE: REPLICASE POLYPROTEIN 1AB; CHAIN: A, B; FRAGMENT: RESIDUE 2966-3265,...	66/50	123	HTM and PB90	+
SARS000002	AAP30029	HYDROLASE 12-MAY-03 1P95 CORONAVIRUS MAIN PROTEINASE (3CLPROV) STRUCTURE: BASIS FOR DESIGN OF ANTI-SARS DRUGS MOL_ID: 1; MOLECULE: REPLICASE POLYPROTEIN 1AB; CHAIN: A, B; FRAGMENT: RESIDUE 2966-3265,...	54/29	65	HTM and PB90	+
SARS000003	AAP30030	SIGNALING PROTEIN 07-JUL-99 1QU7 FOUR HELICAL-BUNDLE STRUCTURE OF THE CYTOSOLIC DOMAIN OF A SERINE CHEMOTAXIS RECEPTOR MOL_ID: 1; MOLECULE: METHYL-ACCEPTING CHEMOTAXIS PROTEIN I; CHAIN: A, B; FRAGME...	17/9	4	PB90 only	
SARS000004	AAP30031	gbl:gi 31416293 gb AAP51228.1 BGI-PUP1 [SARS coronavirus GZ01]	7	5	0	
SARS000005	AAP30032	CELL CYCLE 19-APR-00 1EVO SOLUTION STRUCTURE OF THE MINE TOPOLOGICAL SPECIFICITY DOMAIN MOL_ID: 1; MOLECULE: MINE; CHAIN: A, B; FRAGMENT: MINE TOPOLOGICAL SPECIFICITY DOMAIN; SYNONYM: CELL DIVISION TO...	3/2	3	HTM only	
SARS000006	AAP30033	gbl:gi 31416298 gb AAP51230.1 envelope protein E [SARS coronavirus GZ01]	2	1	4.8e-26	
SARS000007	AAP30034	CALCIUM-BINDING PROTEIN 29-OCT-93 1REC RECOVERIN (CALCIUM SENSOR IN VISION)	2/1	1	PB90 only	
SARS000008	AAP30035	OXIDOREDUCTASE 29-NOV-99 1D1I CRYSTAL STRUCTURE OF P-CRESOL METHYLHYDROXYLASE AT 2.5 Å RESOLUTION MOL_ID: 1; MOLECULE: P-CRESOL METHYLHYDROXYLASE; CHAIN: A, B; FRAGMENT: FLAVOPROTEIN SUBUNIT; SYNONYM:...	1/1	0	HTM only	
SARS000009	AAP30036	TRANSPORT (THYROXINE) 05-JAN-96 1TFP TRANSTHYRETIN (FORMERLY KNOWN AS PREALBUMIN) MOL_ID: 1; MOLECULE: TRANSTHYRETIN; CHAIN: A, B; SYNONYM: PREALBUMIN TRANSPORT (THYROXINE), ALBUMIN, RETINOL-BINDING	1/1	2	HTM only	
SARS000010	AAP30037	COMPLEX (HYDROLASE/IMMUNOGLOBULIN) 06-MAY-94 1NMA MOL_ID: 1; MOLECULE: N9 NEURAMINIDASE; CHAIN: N; EC: 3.2.1.18; MUTATION: WILD TYPE; MOL_ID: 2; MOLECULE: FAB NGL1; CHAIN: L; N; OTHER DETAILS: RESULT...	1/1	1	HTM only	
SARS000011	AAP30038	gbl:gi 29836502 refnp_828859.1 unknown [sars9b] [SARS coronavirus] gi 30173392 gp PS9636 YS_CVHSA Hypothetical protein 5 gi 3023961 gb AAP13574.1 unknown [SARS coronavirus CUHK-W1] gi 30275677 gb AAP30038.1 putative uncharacterized protein 5 [SARS coronavirus B01] gi 3069837 gb AAP37025.1 unknown [SARS coronavirus TW1] gi 30795156 gb AAP1048.1 Orf13 [SARS coronavirus Tor2] gi 31416303 gb AAP51235.1 BGI-PUP1 [SARS coronavirus GZ01] gi 31458154 gb AAP3708.1 Orf9b [SARS coronavirus Frankfurt 1] gi 31747858 gb AAP6959.1 uncharacterized protein 9b [SARS coronavirus ZJ-HZ01] gi 32187356 gb AAP2985.1 Orf9b [SARS coronavirus HSR 1]	2	2	1.2e-48	
SARS000012	input.seq	TOXIN 14-SEP-01 1JZA CRYSTAL STRUCTURE OF VARIANT 2 SCORPION TOXIN FROM CENTRUROIDES SCULPTURATUS EWING MOL_ID: 1; MOLECULE: NEUROTOXIN 2; CHAIN: A, B; CRYSTAL STRUCTURE, SCORPION TOXIN, NONCRYSTALLOR...	3/2	3	HTM only	
SARS000013	input.seq	OXIDOREDUCTASE 28-MAY-98 2OCC BOVINE HEART CYTOCHROME C OXIDASE AT THE FULLY OXIDIZED STATE MOL_ID: 1; MOLECULE: CYTOCHROME C OXIDASE; CHAIN: A, B, C, D, E, F, G, H, J, K, L, M, N, O, P, Q, R, S, T,...	4/2	4	HTM only	
SARS000014	input.seq	STRUCTURAL GENOMICS, UNKNOWN FUNCTION 28-JUN-02 1M3S CRYSTAL STRUCTURE OF YCKF FROM BACILLUS SUBTILIS MOL_ID: 1; MOLECULE: HYPOTHETICAL PROTEIN YCKF; CHAIN: A, B; ENGINEERED: YES STRUCTURAL GENOMICS	6/1	1	PB90 only	+
SARS000015	input.seq	gbl:gi 31416293 gb AAP51225.1 orf1ab [SARS coronavirus GZ01]	23	5	0	
SARS000016	input.seq	FATTY ACID-BINDING 20-FEB-99 1A57 THE THREE-DIMENSIONAL STRUCTURE OF A HELIX-LESS VARIANT OF INTESTINAL FATTY ACID BINDING PROTEIN, NMR, 20 STRUCTURES MOL_ID: 1; MOLECULE: INTESTINAL FATTY ACID-BINDIN...	1/1	4	HTM only	
SARS000017	input.seq	MEMBRANE PROTEIN, PROTON TRANSPORT 21-AUG-02 1MHS MODEL OF NEUROSPORA CRASSA PROTON ATPASE MOL_ID: 1; MOLECULE: PLASMA MEMBRANE ATPASE; CHAIN: A, B; SYNONYM: PROTON PUMP; EC: 3.6.3.6 ION TRANSPORT, PR...	25/12	32	PB90 only	+

Legend:

- Blue rows: Detected homology with known structures
- Green rows: Detected homology with other sequences in NRDB database
- Red rows: Sequences with unassigned functions

HTM: Show HTM/AS regions for structure homology or ASHT for sequence homology
 Best Hits: Show best hit method for structure homology or best BLAST e-value for sequence homology
 AS: Show presence of active site annotation. The plus sign refers to hits other than the best hit for given target sequence. The numerical indices pertain to the best hit for given target sequence. The 1st number is the number of active site residues that are identical between the target and template. The 2nd number is those residues from the active site that are found in the alignment frame. The 3rd number is the total

Sequence Annotation Map - SARS000002

Sequence View

TransMem

DSC

Domains

Hits

Attribute Value

Start Position	3241
End Position	3543
Length	303
Method	HTM and PB90
Template	1lvo
Function	HYDROLASE 29-MAY-02 1LVO STRUCTURE OF CORONAVIRUS MAIN PROTEINASE REVEALS COMBINATION OF A CHYMOTRYPSIN FOLD WITH AN EXTRA ALPHA-HELICAL DOMAIN MOL_ID: 1; MOLECULE: REPLICASE, HYDROLASE DOMAIN; CHAIN...
E-value	0
Active Site	
Binding Site	2/175
Identities	133/303 (44%)
Positives	185/303 (61%)
Verify Score	0.77
PMF Score	1.00
Combination	
Verify Score	0.96

Annotated Alignment

```

1lvoA      : 1  SGLEMAQPSGLVEPCVVP
EP/1lvoA/1lvoA_1 :  SG REMA PSG VE c+*
                :  RAQP
CSC/1lvoA/STO0007_1 :
SARS000002_2 : 3241  SGFRKMAFP SGKVEGCMVQVTCGTTTLNGLWLDLT
1lvoA      : 51  -INTEYMSVRLNMFQSF
                :  NYS+  H+ F V
CSC/1lvoA/STO0007_1 :
  
```

3D structure shown in lower right.

Results for Mpro model. 3D structure shown in lower right.

