EXPLORING THE FUNCTIONS, REACTIONS, PATHWAYS AND BIOLOGICAL PROCESSES OF SPIKE GLYCOPROTEIN USING VARIOUS DATABASES

Project

Submitted in Subject of Practice School (BP808T)

B. Pharmacy VIII Semester

Submitted by

Siddharth Chanana

[Roll No. 170121250047]

Mentor

Dr Samridhi

Assistant Professor

DEPARTMENT OF PHARMACEUTICAL SCIENCES

GURU JAMBHESHWAR UNIVERSITY OF SCIENCE & TECHNOLOGY, HISAR-125001, HARYANA (INDIA)

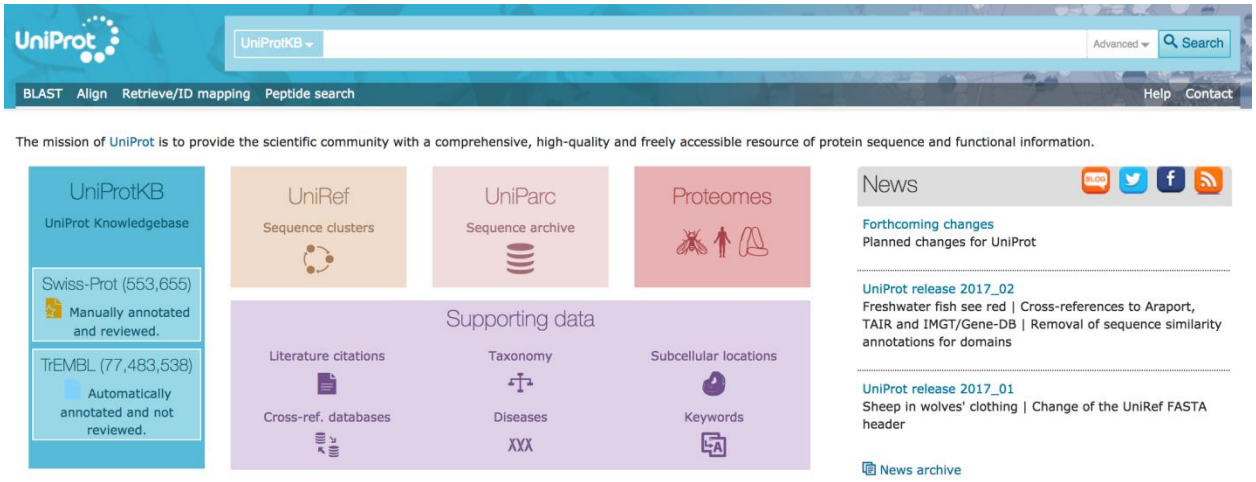
('A' GRADE NAAC ACCREDITED UNIVERSITY)

Index

The Universal Protein Resource (UniProt)

Introduction

The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc). The UniProt consortium and host institutions EMBL-EBI, SIB and PIR are committed to the long-term preservation of the UniProt databases.



Previously, Swiss-Prot + TrEMBL (Boeckmann B et al., 2003) and PIR-PSD (Wu,C.H. et al. 2003) coexisted as protein databases with differing sequence coverage and annotation priorities. In 2002, the Swiss-Prot + TrEMBL groups at the SIB (Swiss Institute of Bioinformatics) and EBI (European Bioinformatics Institute) and the PIR (Protein Information Resource) group at Georgetown University Medical Center and National Biomedical Research Foundation joined forces as the UniProt consortium (Apweiler,R. et al. 2004) .

The UniProt consortium maintains three database layers:

(i) The UniProt Archive (UniParc) provides a stable, comprehensive, non-redundant sequence collection by storing the complete body of publicly available protein sequence data.

(ii) The UniProt Knowledgebase (UniProt) provides the central database of protein sequences with accurate, consistent and rich sequence and functional annotation.

(iii) The UniProt Reference (UniRef) databases provide non-redundant data collections based on the UniProt Knowledgebase and UniParc in order to obtain complete coverage of sequence space at several resolutions.

THE UniProt ARCHIVE:

Although most protein sequence data are derived from the translation of DDBJ/EMBL/GenBank (Kulikova,T. et al. 2004) sequences, primary protein sequence data are also submitted directly to UniProt or appear in patent applications or in entries from the Protein Data Bank (PDB) (Westbrook,J. et al. 2003). The UniParc (Leinonen,R., et al. 2004) is designed to capture all available protein sequence data—not just from the aforementioned databases, but also from sources such as Ensembl (Hubbard,T. et al. 2004), the International Protein Index (IPI) (Kersey,P.J. et al. 2004), RefSeq (Pruitt,K et al.2001), FlyBase (FlyBase Consortium 2003) and WormBase (Harris,T. et al. 2003).

UniParc cross-references the promotion numbers of the source databases, utilizing banners to demonstrate the status of the passage in the original source database, with 'active' demonstrating that the entry is still display within the source database and 'obsolete' indicating that the entry not exists within the source database.

THE UniProt KNOWLEDGEBASE:

The UniProt Knowledgebase merges Swiss-Prot, TrEMBL and PIR-PSD to provide a central database of protein sequences with annotations and functional information.

The UniProt Knowledgebase has two parts: a section of fully, manually annotated records resulting from literature information extraction and curator-evaluated computational analysis, and a section with computationally analyzed records awaiting full manual annotation.The two sections are referred to as 'UniProt/Swiss-Prot' (158,337 records in UniProt release 2.6 from September 2004) and 'UniProt/TrEMBL' (1,400,776 records in UniProt release 2.6 from September 2004).

THE UniProt REFERENCE DATABASES:

Automatic procedures have been developed to create three UniRef databases, such as UniRef100, UniRef90 and UniRef50, from the UniProt Knowledgebase and UniParc as representative protein sequence databases with high information content. The databases provide complete coverage of sequence space while hiding redundant sequences from view.

UniRef90 and UniRef50 are built from UniRef100 using the CD-HIT algorithm (Li.W. et al. 2002) to provide non-redundant sequence collections for the scientific user community to perform faster homology searches. All records from all source organisms with mutual sequence identity of >90% or >50%, respectively, are merged into a single record that links to the corresponding UniProt Knowledgebase records. UniRef90 and UniRef50 yield a size reduction of ~ 40 and 65%, respectively.

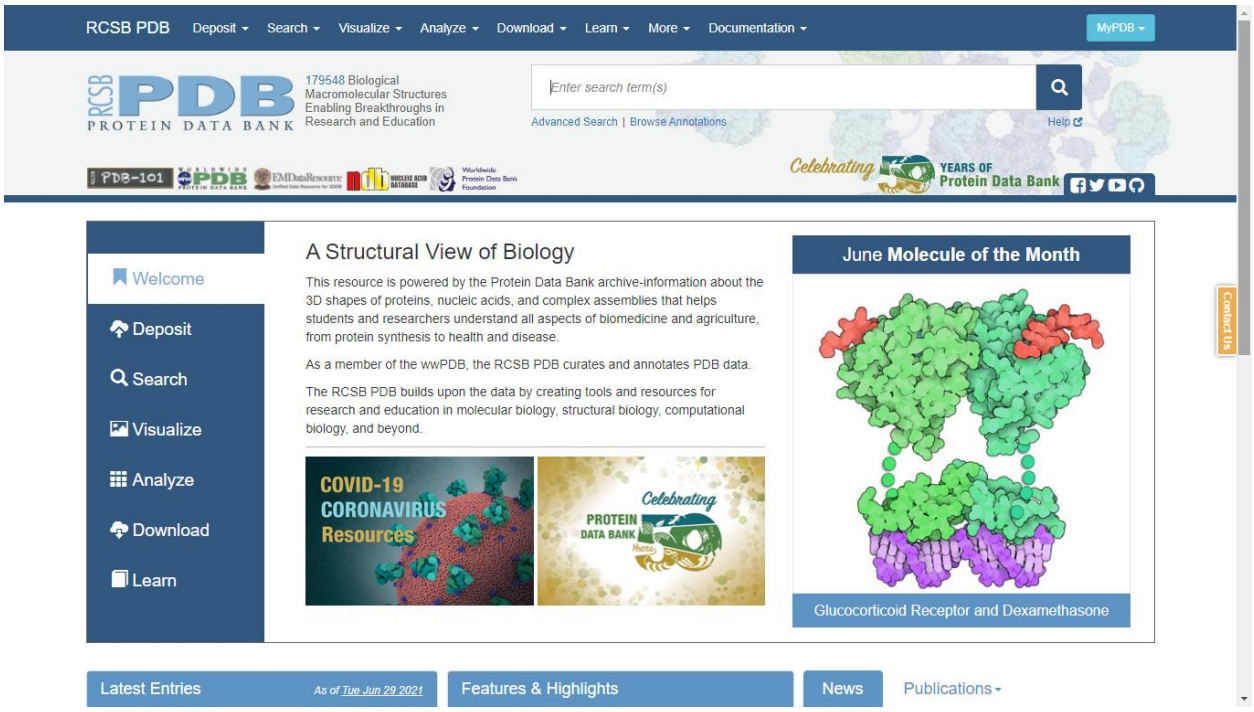
Applications of UniProt:

- It provides an up-to-date, comprehensive body of protein information at a single site.
- It aids scientific discovery by collecting, interpreting and organising this information so that it is easy to access and use.
- It saves researchers countless hours of work in monitoring and collecting this information themselves.
- It provides tools to help with protein sequence analysis.
- It provides links to related information in more than 150 other biological databases to help you access additional information in more specialised collections.

Protein Data Bank

Introduction:

The Protein Data Bank (PDB) was established at Brookhaven National Laboratories (BNL) (Bernstein FC, et al. 1977) in 1971 as an archive for biological macromolecular crystal structures. In the beginning the archive held seven structures, and with each year a handful more were deposited. Through an internet information portal and downloadable data archive, the PDB provides access to 3D structure data for large biological molecules (proteins, DNA, and RNA). These are the molecules of life, found in all organisms on the planet.



In the 1980s the number of deposited structures began to increase dramatically. This was due to the improved technology for all aspects of the crystallographic process, the addition of structures determined by nuclear magnetic resonance (NMR) methods, and changes in the community views about data sharing. By the early 1990s the majority of journals required a PDB accession code and at least one funding agency (National Institute of General Medical Sciences) adopted

the guidelines published by the International Union of Crystallography (IUCr) requiring data deposition for all structures.

Initial use of the PDB had been limited to a small group of experts involved in structural research. Today depositors to the PDB have varying expertise in the techniques of X-ray crystal structure determination, NMR, cryoelectron microscopy and theoretical modeling. Users are a very diverse group of researchers in biology, chemistry and computer scientists, educators, and students at all levels. The tremendous influx of data soon to be fueled by the structural genomics initiative, and the increased recognition of the value of the data toward understanding biological function, demand new ways to collect, organize and distribute the data.

In October 1998, the management of the PDB became the responsibility of the Research Collaboratory for Structural Bioinformatics (RCSB). In general terms, the vision of the RCSB is to create a resource based on the most modern technology that facilitates the use and analysis of structural data and thus creates an enabling resource for biological research. Specifically in this paper, we describe the current procedures for data deposition, data processing and data distribution of PDB data by the RCSB. In addition, we address the issues of data uniformity. We conclude with some current developments of the PDB.

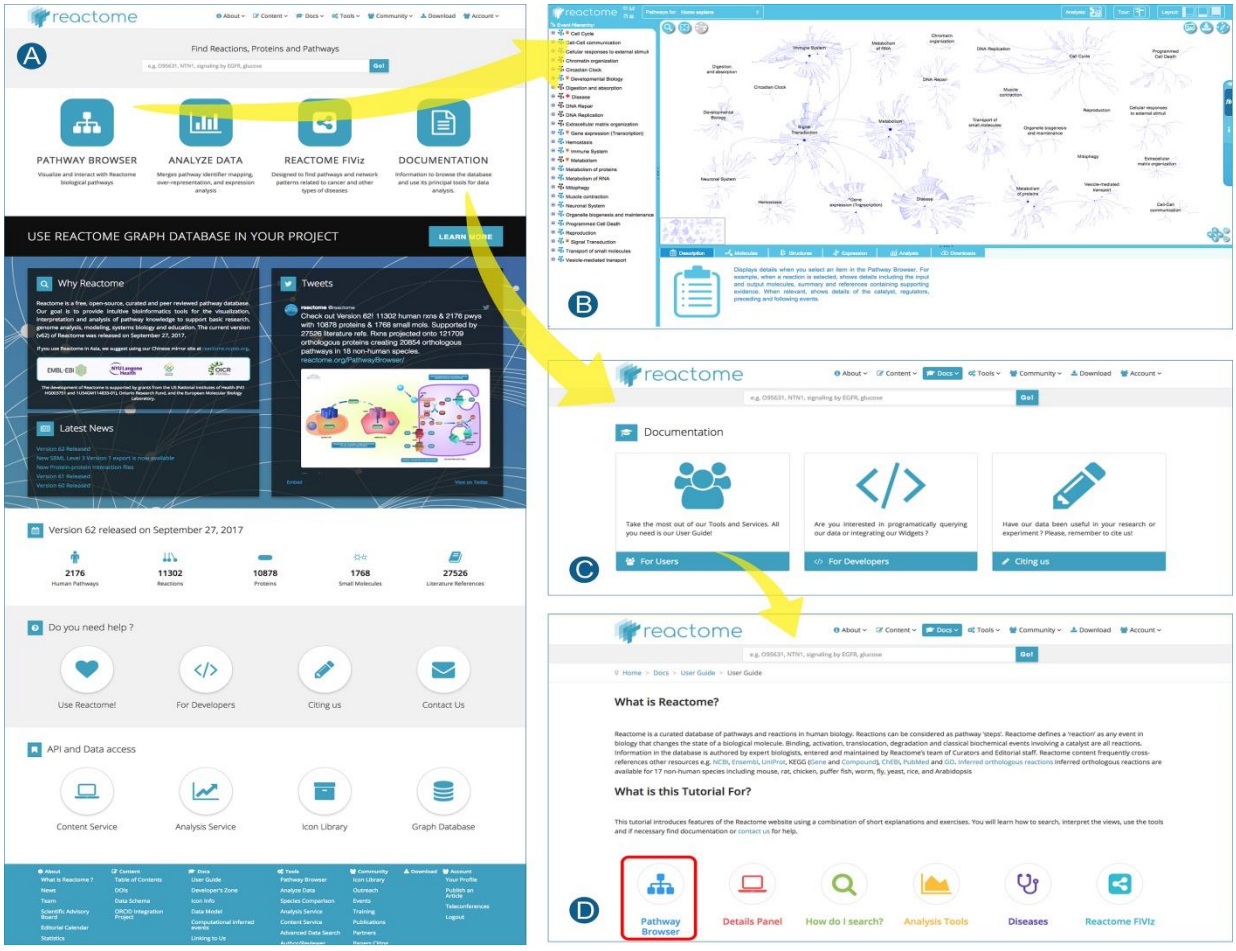
Applications of Protein Data bank:

- Innovations that can lead to new product development and company formation.
- It uses in discovery of lifesaving drugs.
- STEAM education: PDB-101 provides curricula and online tools for teachers and students.
- It is today a leading global resource for experimental data central to scientific discovery. Through an internet information portal and downloadable data archive, the PDB provides access to 3D structure data for large biological molecules (proteins, DNA, and RNA).

The Reactome Pathway Database

Introduction

REACTOME is an open-source, open access, manually curated and peer-reviewed pathway database. The goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic and clinical research, genome analysis, modeling, systems biology and education. Founded in 2003, the Reactome project is led by Lincoln Stein of OISR, Peter D'Eustachio of NYULMC, Henning Hermjakob of EMBL-EBI, and Guanming Wu of OSHU.



At the cellular level, life is a network of molecular reactions that include signal transduction, transport, DNA replication, protein synthesis, and intermediary metabolism. A variety of online resources capture aspects of this information at the level of individual reactions such as Rhea (Morgat,A. et al. 2017)or at the level of reaction sequences spanning various domains of biology such as KEGG (Kanehisa,M. et al.2017), MetaCyc (Caspi,R. et al. 2016) or PANTHER (Mi,H. et al. 2017).

The Reactome Knowledgebase is distinctive in focusing its manual annotation effort on a single species, Homo sapiens, and applying a single consistent data model across all of these domains of biology. Processes are systematically described in molecular detail to generate an ordered network of molecular transformations, resulting in an extended version of a classic metabolic map (Fabregat,A. et al. 2016). The Reactome Knowledgebase systematically links human proteins to their molecular functions, providing a resource that functions both as an archive of biological processes and as a tool for discovering unexpected functional relationships in data such as gene expression surveys or catalogs of somatic mutations in tumor cells. Reactome (version 62—September 2017) has entries for 10,719 human genes, 53% of the 20,338 predicted human protein-coding genes supporting the annotation of 24,704 specific forms of proteins distinguished by co- and post-translational modifications and subcellular localizations. These function with 1768 small molecules as substrates, catalysts, and regulators in 11,302 reactions annotated on the basis of data from 27,526 literature references. These tallies include 1334 mutant variants and their post-translationally modified forms derived from 285 gene products, used to annotate 906 disease-specific reactions, tagged with 294 Disease Ontology terms (Kibbe,W.A. et al. 2015). These reactions form 2102 pathways (e.g. Interleukin-15 signaling; phosphatidylinositol phosphate metabolism; receptormediated mitophagy) grouped into 26 superpathways that correspond to domains of biology such as metabolism and signal transduction. Reactome's dataset continues to grow briskly, with 74 new human pathways added in the first three quarters of 2017. Notable additions include extensive new annotations of cytokine signaling, including a comprehensive catalog of known interleukin signaling pathways. It have revised and supplemented existing pathways, continuing to build our catalogs of signaling processes mediated by G protein-coupled receptors, of transport processes, and of metabolism. Notably, where our initial annotations in these domains centered on the most extensively studied, 'textbook' versions of pathways and molecules, It has systematically adding

proteins whose properties indicate closely-related biological roles, to increase the density and connectivity of the reaction network in Reactome that is available for visualization and computational analysis. This growth has come at a cost. Our SBGN-based (Le Novere,N. et al. 2009) scheme for representing pathways, implemented eight years ago, yields pathway diagrams that become cluttered and difficult for biologist users to navigate as we achieve more nearly complete annotations of the participants in a process and their functions, and our approximately 120 diagrams of super pathways are essentially lists of the names of component pathways, functional but uninformative and unappealing to biologist users. Meanwhile, this growth in the number and complexity of our annotations has made the relational data structure slow and unwieldy for handling complex queries and large scale data analyses. To improve usability the developed Enhanced High Level Diagrams (EHLDs) that combine an iconography familiar from textbooks and review articles with web functionality, to represent superpathways, and have added features to pathway diagrams to improve their legibility. A redesigned web site with adaptive technology is accessible from tablets and mobile devices, and supports more intuitive navigation.

Applications of Reactome:

- It is used as a intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic and clinical research, genome analysis, modeling, systems biology and education.
- It providing a resource that functions both as an archive of biological processes and as a tool for discovering unexpected functional relationships in data such as gene expression surveys or catalogs of somatic mutations in tumor cells.
- Reactome is used by clinicians, geneticists, genomics researchers, and molecular biologists to interpret the results of high-throughput experimental studies, by bioinformaticians seeking to develop novel algorithms for mining knowledge from genomic studies, and by systems biologists building predictive models of normal and disease variant pathways.
- It provide more efficient access and that allow new forms of analysis that were not possible with information stored in the traditional printed literature.

Objectives

To Delve into the structure, location, functions, reactions, pathways and biological processes of Spike glycoprotein using various databases like The universal protein resource (UniProt), Protein Data Bank (PDB), The Reactome Pathway Database.

Methodology Adopted

Various databases are used for expanding the universe of protein information

1) Type of gene, organism, General and molecular functions of spike glycoprotein, spike glycoprotein involved in biological processes, taxonomy, location of spike glycoprotein, interactions with other proteins, length and sequence of amino acids, similar protein like spike glycoprotein data are retrieved from The universal protein resource (UniProt) (https://www.uniprot.org/).

2) The front, side and top view of 3D protein structure of spike glycoprotein are fetched from RCSB Protein Data Bank (PDB) (https://www.rcsb.org/).

3) The attachment and entry pathway of spike glycoprotein and drugs acting on this pathway was determined by The Reactome Pathway Database (https://reactome.org/).

Results and Discussion

P59594 (SPIKE_SARS)

1.) Protein: Spike glycoprotein.

2.) Gene: S

3.) Organism: Severe acute respiratory syndrome coronavirus (SARS-CoV)

4.) Functions:

(1) Spike Glycoprotein:

May down-regulate host tetherin (BST2) by lysosomal degradation, thereby counteracting its antiviral activity. (Wang S.M, et al. 2019)

(2) Spike Protein S1 :

→Attaches the virion to the cell membrane by interacting with host receptor, initiating the infection (By similarity).

Binding to human ACE2 and CLEC4M/DC-SIGNR receptors and internalization of the virus into the endosomes of the host cell induces conformational changes in the S glycoprotein. Proteolysis by cathepsin CTSL may unmask the fusion peptide of S2 and activate membranes fusion within endosomes. (Wong S.K., et al. J. 2004) (Jeffers S.A., et al. 2004).

(3) Spike Protein S2:

→Mediates fusion of the virion and cellular membranes by acting as a class I viral fusion protein. Under the current model, the protein has at least three conformational states: pre-fusion native state, pre-hairpin intermediate state, and post-fusion hairpin state. During viral and target cell membrane fusion, the coiled coil regions (heptad repeats) assume a trimer-of-hairpins structure, positioning the fusion peptide in close proximity to the C-terminal region of the ectodomain. The formation of this structure appears to drive apposition and subsequent fusion of viral and target cell membranes.

(4) Spike Protein S2':

→Acts as a viral fusion peptide which is unmasked following S2 cleavage occurring upon virus endocytosis. (Belouzard S. et al. 2009)

Molecular functions:

1. Host cell surface receptor binding.

2. Identical protein binding.

Biological function:

1. Endocytosis involved in viral entry into host cell.

2. Fusion of virus membrane with host endosome membrane.

3. Fusion of virus membrane with host plasma membrane.

4. Pathogenesis.

5. Receptor-mediated virion attachment to host cell.

6. Suppression by virus of host tetherin activity.

7. Suppression by virus of host type I interferon-mediated signaling pathway.

8. Viral protein processing.

9. Viral translation.

5.) Name and Taxonomy:

(1) Protein names: Spike glycoprotein. (S glycoprotein)

Alternative name: E2 and Peplomer protein.

Cleaved into the following parts: spike protein S1, spike protein S2 and spike protein S2'.

(2) Gene names: S

(3) Organism: Severe acute respiratory syndrome coronavirus (SARS-CoV).

(4) Taxonomic identifier: 694009.

(5) Taxonomic Lineage: Viruses > Riboviria > Orthornavirae > Pisuviricota > Pisoniviricetes > Nibovirales > Cornidovirineae > Coronaviridae > Orthocoronavirinae > Betacoronavirus > Sarbecovirus.

(6) Virus host: Homo sapiens (Human) ,Paguna Lrvata (Masked palm civet).

6.) Subcellular location:

- Virion membrane; Single-pass type 1 membrane protein (Nal B., et al. 2005).
- Host endoplasmic reticulum-Golgi intermediate compartment membrane. (Teoh K.T., et al. 2010); Single-pass type 1 membrane protein.(Nal B., et al. 2005).
- Host cell membrane; Single-pass type 1 membrane protein (Nal B., et al. 2005).

Note: Accumulates in the endoplasmic reticulum-Golgi intermediate compartment, where it participates in virus particle assembly. Colocalizes with S in the host endoplasmic reticulum-

Golgi intermediate compartment. Some S oligomers are transported to the host plasma membrane, where they may mediate cell-cell fusion. (Teoh K.T., et al. 2010)

7.) Interaction:

Interacts with the accessory proteins 3a and 7a.

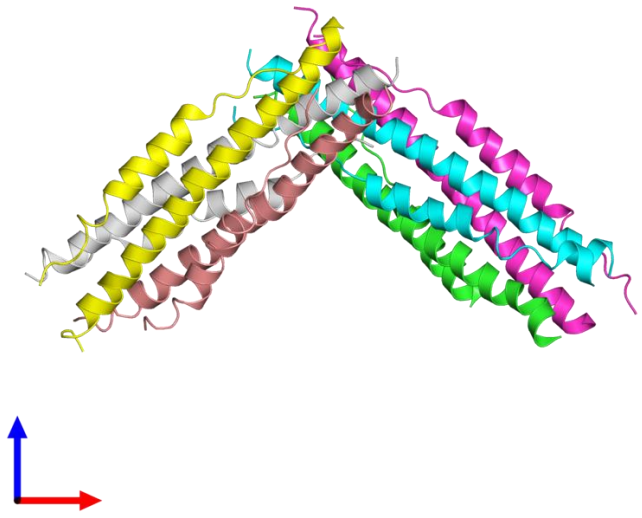
P59594 has binary interactions with 6 proteins.

- SPIKE_SARS
- NS7A_SARS
- ACE2_HUMAN
- ACE2_PAGLA
- EIF3F_HUMAN
- ACE2_RAT

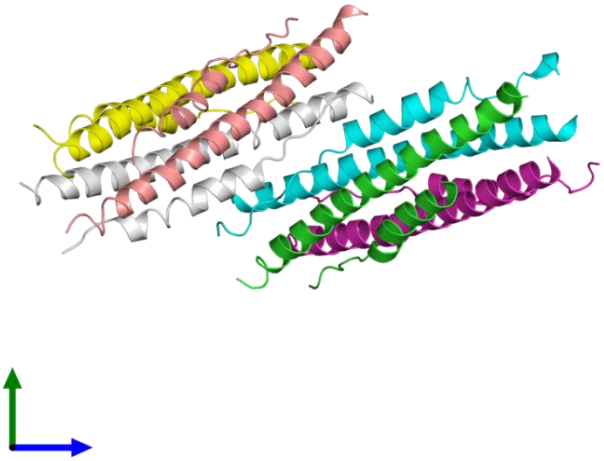
8.) Structure:

- 1 WNC:

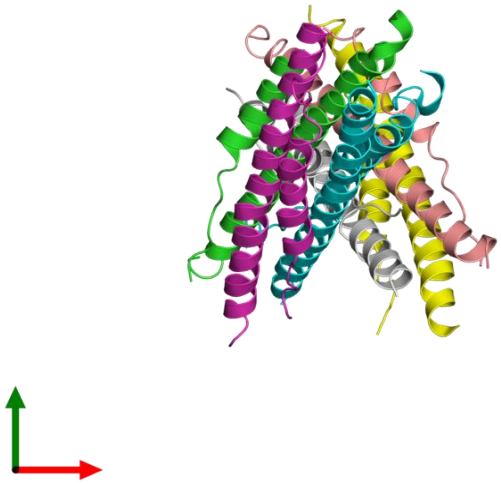
(1) **1WNC coloured by chain and viewed from the front.**



(2) **1WNC coloured by chain and viewed from the side.**



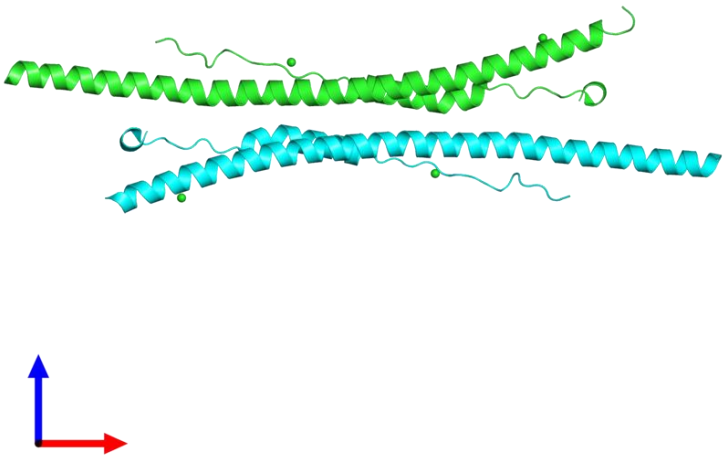
(3) **1WNC coloured by chain and viewed from the top.**



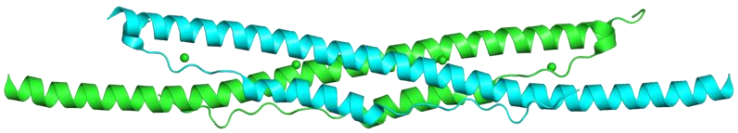
(Xu Y, et al. 2004)

- **1WYY:**

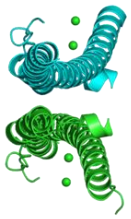
(1) **1WYY coloured by chain and viewed from the front.**



(2) **1WYY coloured by chain and viewed from the side.**



(3)1WYY coloured by chain and viewed from the top.



(Duquerroy S., et al. 2005)

9.) Sequence:

10 MFIFLLFLTL	20 TSGSDLDRCT	30 TFDDVQAPNY	40 TQHTSSMRGV	50 YYPDEIFRSD
60 TLYLTQDLFL	70 PFYSNVTGFH	80 TINHTFGNPV	90 IPFKDGIYFA	100 ATEKSNVVRG
110 WVFGSTMNNK	120 SQSVIIINNS	130 TNVVIRACNF	140 ELCDNPFFAV	150 SKPMGTQTHT
160 MIFDNAFNCT	170 FEYISDAFSL	180 DVSEKSGNFK	190 HLREFVFKNK	200 DGFLYVKGY
210	220	230	240	250

QPIDVVRDLP	SGFNTLKPIF	KLPLGINITN	FRAILTAFSP	AQDIWGTSAA
260 AYFVGYLKPT	270 TFMLKDENG	280 TITDAVDCSQ	290 NPLAELKCSV	300 KSFEIDLGIY
310 QTSNERVVPS	320 GDVVRFPNIT	330 NLCPFGEVFN	340 ATKFPSVYAW	350 ERKKISNCVA
360 DYSVLYNSTF	370 FSTEKCYGVS	380 ATKLNDLCFS	390 NVTADSFVVK	400 GDDVRQIAPG
410 QTGVIADYNY	420 KLPDDFMGCV	430 LAWNTRNIDA	440 TSTGNYNYKY	450 RYLRHGKLRP

460 FERDISNVPF	470 SPDGKTCTPT	480 FERDISNVPF	490 YGFYTTTGIG	500 YQPYRVVVLS
510 FELLNAPATV	520 CGPKLSTDLI	530 KNQCVNFNFN	540 GLTGTGVLTP	550 SSKRFQPFQQ
560 FGRDVSDFTD	570 SVRDPKTSEI	580 LDISPCSFGG	590 VSVITPGTNA	600 SSEVAVLYQD
610 VNCTDVSTAI	620 HADQLTPAWR	630 IYSTGNNVFQ	640 TQAGCLIGAE	650 HVDTSYECDI

660	670	680	690	700
PIGAGICASY	HTVSLLRSTS	QKSIVAYTMS	LGADSSIAYS	NNTIAIPTNF
710	720	730	740	750
SISITTEVMP	VSMAKTSVDC	NMYICGDSTE	CANLLLQYGS	FCTQLNRALS
760	770	780	790	800
GIAAEQDRNT	REVFAQVKQM	YKTPTLKYFG	GFNFSQILPD	PLKPTKRSFI
810	820	830	840	850
EDLLFNKVTL	ADAGFMKQYG	ECLGDINARD	LICAQKFNGL	TVLPPLLTDD
860	870	880	890	900
MIAAYTAALV	SGTATAGWTF	GAGAALQIPF	AMQMAYRFNG	IGVTQNVLYE

910	920	930	940	950
NQKQIANQFN	KAISQIQESL	TTTSTALGKL	QDVVNQNAQA	LNTLVKQLSS
960	970	980	990	1000
NFGAISSVLN	DILSRLDKVE	AEVQIDRLIT	GRLQSLQTYV	TQQLIRAAEI
1010	1020	1030	1040	1050
RASANLAATK	MSECVLGQSK	RVDFCGKGYH	LMSFPQAAPH	GVVFLHVTYV
1060	1070	1080	1090	1100
PSQERNFTTA	PAICHEGKAY	FPREGVFVFN	GTSWFITQRN	FFSPQIITTD

1110 NTFVSGNCDV	1120 VIGIINNTVY	1130 DPLQPELDSF	1140 KEELDKYFKN	1150 HTSPDVDLGD
1160 ISGINASVVN	1170 IQKEIDRLNE	1180 VAKNLNESLI	1190 DLQELGKYEQ	1200 YIKWPWYVWL
1210 GFIAGLIAIV	1220 MVTILLCCMT	1230 SCCSCLKGAC	1240 SCGSCCKFDE	1250 DDSEPVLKGV

10.) Similar Protein:

- Clusters with 50% identity.

Sr.No.	Entry Name	Protein Names	Organism	Organism Ids
1.	SPIKE_SARS2	Spike Glycoprotein	Severe acute respiratory syndrome coronavirus 2 (2019 –nCov)(SARS-CoV-2)	2697049
2.	SPIKE_BCHK3	Spike Glycoprotein	Bat coronavirus HKU3 (BtCoV) (SARS like coronavirus HKU3)	442736

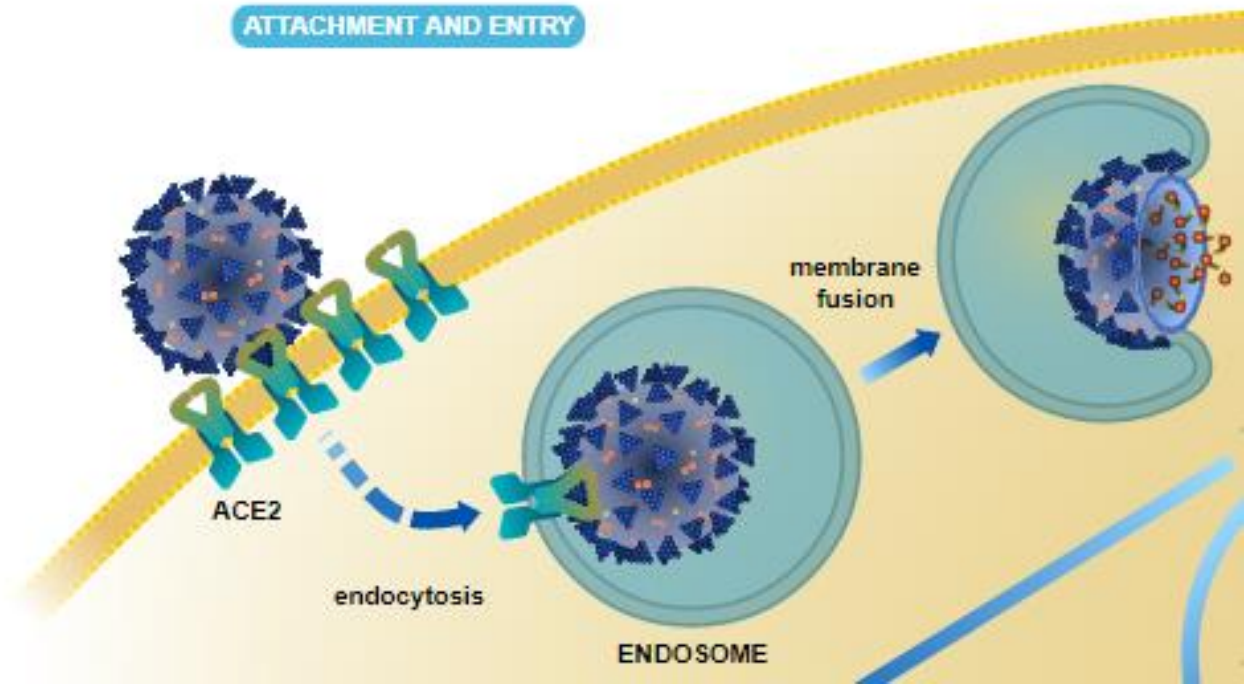
- Clusters with 90% identity.

Sr.No.	Entry Name	Protein Names	Organism	Organism IDs
1.	Q6RCW5_SARS	Spike Glycoprotein	SARS coronavirus TW9	258972
2.	Q6RCZ8_SARS	Spike Glycoprotein	SARS coronavirus TW6	258969

- Clusters with 100% identity.

Sr.No.	Entry Name	Protein Names	Organism	Organism IDs
1.	D3KDQ3_SARS	Spike Glycoprotein	SARS coronavirus HKU-39849	228404
2.	D2E1D2_SARS	Spike Glycoprotein	SARS coronavirus ExoN1	627440

11.) Pathway and Enzyme:



Coronavirus replication is initiated by the binding of S protein to the cell surface receptor(s). The S protein consists of two functional domains, S1 (bulb) which mediates receptor binding and S2 (stalk) which mediates membrane fusion. Specific interaction between S1 and therefore the cognate receptor triggers a drastic conformational change in S2, resulting in fusion between the virus envelope and therefore the cellular membrane and release of the viral nucleocapsid into the host cell cytosol. Receptor binding is that the major determinant of the host range and tissue tropism for a coronavirus. Some human coronaviruses (HCoVs) have adopted cell surface enzymes as receptors, angiotensin converting enzyme 2 (ACE2) for SARS-CoV-1 and HCoV NL63. The receptor-bound S protein is activated by cleavage into S1 and S2, mediated by one among two of two host proteases, the endosomal cysteine protease cathepsin L and another trypsin like serine protease. Type II transmembrane serine proteases TMPRSS2 and TMPRSS11D have also been implicated within the activation of S protein of SARS-CoV-1. Host factors may play additional roles in viral entry (not annotated here). Valosin containing protein (VCP) contributes by a poorly understood mechanism to the discharge of coronavirus from early endosomes. Host factors can also restrict the attachment and entry of HCoV. Some interferon inducible transmembrane proteins (IFITMs) exhibited broad spectrum antiviral

functions against various RNA viruses including SARS-CoV-1 while others may facilitate HCoV entry into host cells (Fung & Liu 2019).

Drug molecules:

1) Cepharanthine.
2) Pheneticillin.
3) Isoniazid.
4) Bacampicillin.
5) Latamoxef.
6) Mezlocillin.
7) Camostat.
8) Cefazolin.
9) Cefoxitin.
10) Cefuroxime.
11) Chloroquine.
12) Chloroquine (2+).
13) Clofazimine.
14) Denopamine.
15) Eptifibatide.
16) Hydroxychloroquine.
17) Hydroxychloroquine (2+).
18) I432.
19) Mefloquine.
20) Mepacrine.
21) Meropenem.
22) Mycophenolic acid.
23) Nafamostat.
24) Otamixaban.
25) Relacatib.
26) Resveratol
27) Rifampicin.
28) Rotigaptide.
29) Teicoplanin.
30) Ticarcillin.

Conclusion and outcomes

References

References of Uniprot:

1. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. et al. (2003) . Nucleic Acids Res., 31, 365–370.
2. Wu,C.H., Yeh,L.-S.L., Huang,H., Arminski,L., Castro-Alvear,J., Chen,Y., Hu,Z., Kourtesis,P., Ledley,R.S., Suzek,B.E. et al. (2003) The Protein Information Resource. Nucleic Acids Res., 31, 345–347.
3. Apweiler,R., Bairoch,A., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R., Magrane,M. et al. (2004) UniProt: the Universal Protein knowledgebase. Nucleic Acids Res., 32, D115–D119.
4. Kulikova,T., Aldebert,P., Althorpe,N., Baker,W., Bates,K., Browne,P., van den Broek,A., Cochrane,G., Duggan,K., Eberhardt,R. et al. (2004) The EMBL Nucleotide Sequence Database. Nucleic Acids Res., 32, D27–D30.
5. Westbrook,J., Feng,Z., Chen,L., Yang,H. and Berman,H. (2003) The Protein Data Bank and structural genomics. Nucleic Acids Res., 31, 489–491.
6. Leinonen,R., Diez,F.G., Binns,D., Fleischmann,W., Lopez,R. and Apweiler,R. (2004) UniProt Archive. Bioinformatics, 20, 3236–3237.
7. Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T. et al. (2002) The Ensembl genome database project. Nucleic Acids Res., 30, 38–41.
8. Kersey,P.J., Duarte,J., Williams,A., Karavidopoulou,Y., Birney,E. and Apweiler,R. (2004) The International Protein Index: an integrated database for proteomics experiments. Proteomics, 4, 1985–1988.
9. Pruitt,K. and Maglott,D. (2001) RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res., 29, 137–140.
10. FlyBase Consortium (2003) The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Res., 31, 172–175.

11. Harris,T., Lee,R., Schwarz,E., Bradnam,K., Lawson,D., Chen,W., Blasier,D., Kenny,E., Cunningham,F., Kishore,R. et al. (2003) WormBase: a cross-species database for comparative genomics. Nucleic Acids Res., 31, 133–137.
12. Li,W., Jaroszewski,L. and Godzik,A. (2002) Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics, 18, 77–82.

References of Protein data bank:

1. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M J Mol Biol. 1977 May 25; 112(3):535-42.

References of Reactome:

1. Morgat,A., Lombardot,T., Axelsen,K.B., Aimo,L., Niknejad,A., Hyka-Nouspikel,N., Coudert,E., Pozzato,M., Pagni,M., Moretti,S. et al. (2017) Updates in Rhea – an expert curated resource of biochemical reactions. Nucleic Acids Res., 45, D415–D418.
2. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res., 45, D353–D361.
3. Caspi,R., Billington,R., Ferrer,L., Foerster,H., Fulcher,C.A., Keseler,I.M., Kothari,A., Krummenacker,M., Latendresse,M., Mueller,L.A. et al. (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res., 44, D471–D480
4. Mi,H., Huang,X., Muruganujan,A., Tang,H., Mills,C., Kang,D. and Thomas,P.D. (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res., 45, D183–D189.
5. Fabregat,A., Sidiropoulos,K., Garapati,P., Gillespie,M., Hausmann,K., Haw,R., Jassal,B., Jupe,S., Korninger,F., McKay,S. et al. (2016) The Reactome Pathway Knowledgebase. Nucleic Acids Res., 44, D481–D487.

6. Kibbe,W.A., Arze,C., Felix,V., Mitraka,E., Bolton,E., Fu,G., Mungall,C.J., Binder,J.X., Malone,J., Vasant,D. et al. (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic Acids Res., 43, D1071–D1078.
7. Le Novere,N., Hucka,M., Mi,H., Moodie,S., Schreiber,F., ` Sorokin,A., Demir,E., Wegner,K., Aladjem,M.I., Wimalaratne,S.M. et al. (2009) The systems biology graphical notation. Nat. Biotechnol., 27, 735–741.

References of p59594 (Spike Glycoprotein):

1. Wang S.M., Huang K.J., Wang C.T.J. Med. Virol. 91:1743-1750(2019).
2. Wong S.K., Li W., Moore M.J., Choe H., Farzan M.J. Biol. Chem. 279:3197-3201(2004).
3. Jeffers S.A., Tusell S.M., Gillim-Ross L., Hemmila E.M., Achenbach J.E., Babcock G.J., Thomas W.D. Jr., Thackray L.B., Young M.D., Mason R.J., Ambrosino D.M., Wentworth D.E., Demartini J.C., Holmes K.V.Proc. Natl. Acad. Sci. U.S.A. 101:15748-15753(2004).
4. Belouzard S., Chu V.C., Whittaker G.R.Proc. Natl. Acad. Sci. U.S.A. 106:5871-5876(2009).
5. Nal B., Chan C., Kien F., Siu L., Tse J., Chu K., Kam J., Staropoli I., Crescenzo Chaigne B., Escriou N., van der Werf S., Yuen K.Y., Altmeyer R.J. Gen. Virol. 86:1423-1434(2005).
6. Teoh K.T., Siu Y.L., Chan W.L., Schlueter M.A., Liu C.J., Peiris J.S., Bruzzone R., Margolis B., Nal B.Mol. Biol. Cell 21:3838-3852(2010).
7. Li W., Moore M.J., Vasilieva N., Sui J., Wong S.-K., Berne M.A., Somasundaran M., Sullivan J.L., Luzuriaga K., Greenough T.C., Choe H., Farzan M. Nature 426:450-454(2003).
8. Tan Y.-J., Teng E., Shen S., Tan T.H.P., Goh P.-Y., Fielding B.C., Ooi E.-E., Tan H.-C., Lim S.G., Hong W.J. Virol. 78:6723-6734(2004)
9. Huang C., Ito N., Tseng C.-T.K., Makino S.J. Virol. 80:7287-7294(2006).
10. Li F., Li W., Farzan M., Harrison S.C.Science 309:1864-1868(2005).

11. Xu Y, Lou Z, Liu Y, Pang H, Tien P, Gao GF, Rao Z J. Biol. Chem. 279 49414-9 (2004)

12. Duquerroy S, Vigouroux A, Rottier PJ, Rey FA, Bosch BJ Virology 335 276-85 (2005).

13. Fung, T. S., & Liu, D. X. (2019). Human Coronavirus: Host-Pathogen Interaction. Annual Review of Microbiology, 73(1).