TOOLS OF BIOINFORMATICS FOR COVID-19 RESEARCH

by

Dr. Shasank Sekhar Swain

ICMR-Young Scientist Fellow

ICMR-Regional Medical Research Centre,
Bhubaneswar-23, India
Email: swain.shasanksekhar86@gmail.com

WHAT IS TOOL..?

Definition: A device/ technique/ instrument that solves a problem by providing extra advantage in order to do some useful work.

Examples: Hammer, Screwdriver, drilling machine, Xerox machine, Google, etc.,

MEDICAL/ BIOLOGY

Stethoscope

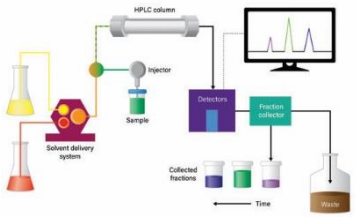
PCR machine

Glucometer

Genetic engineering

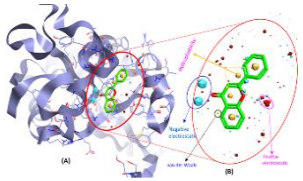
CHEMISTRY/ PHARMACY

New synthesis/ drug coating protocol



New instruments analysis

COMP. BIOLOGY/ BIOINFORMATICS



Molecular docking , networking analysis

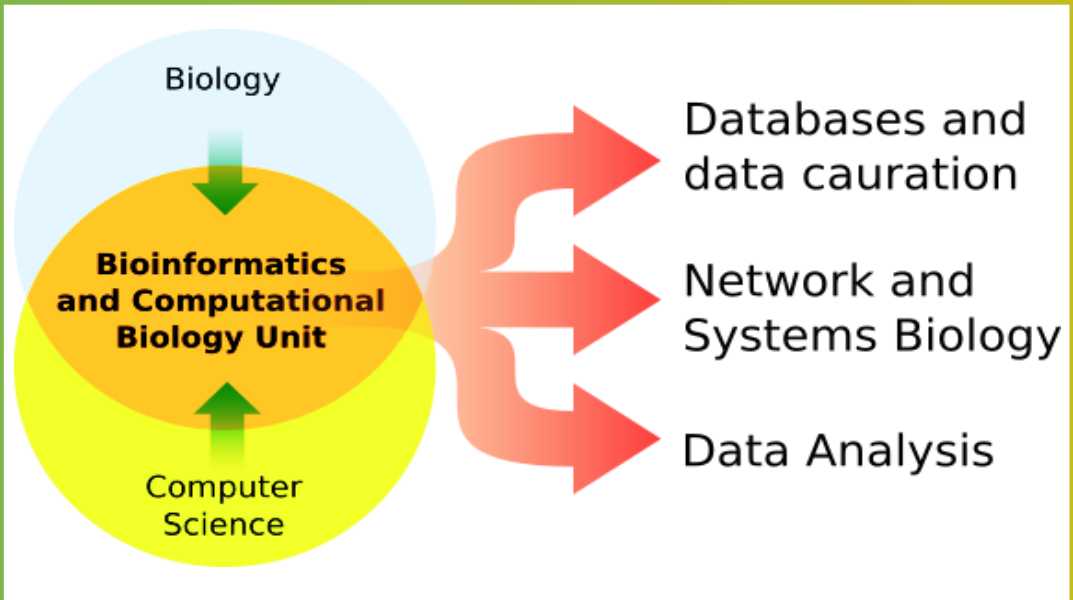
Software, program, server

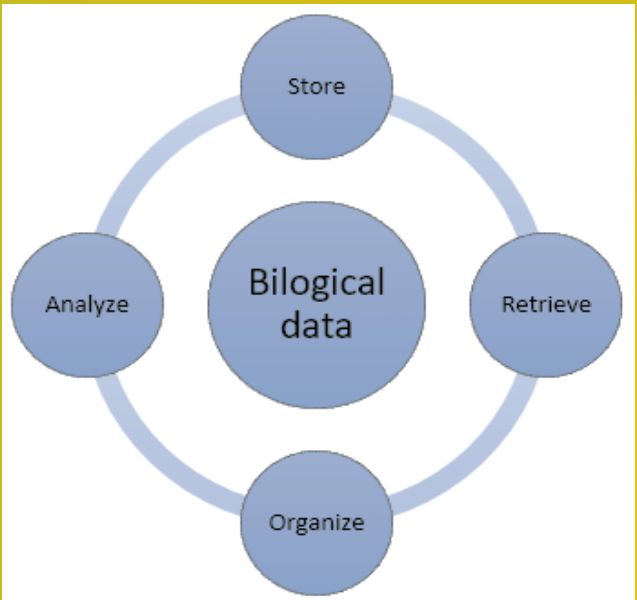
DEFINITION OF BIOINFORMATICS

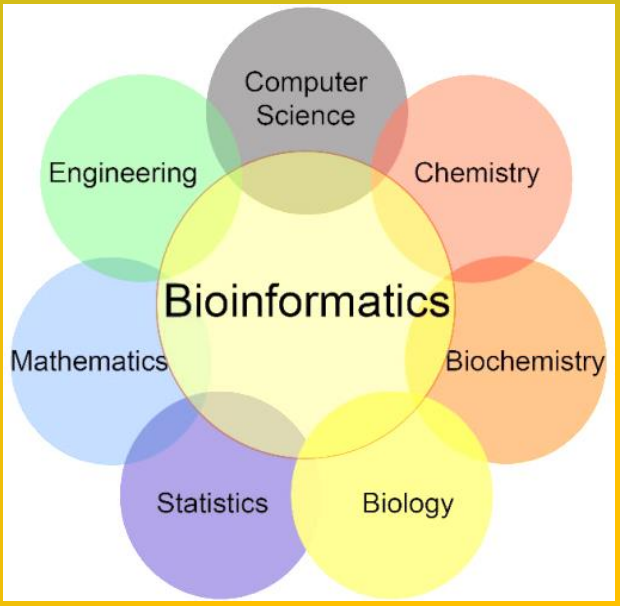
BIO → **Molecule**
(biological/ chemical/ natural)

INFORMATICS → **Techniques**
(software/ tools/ database)

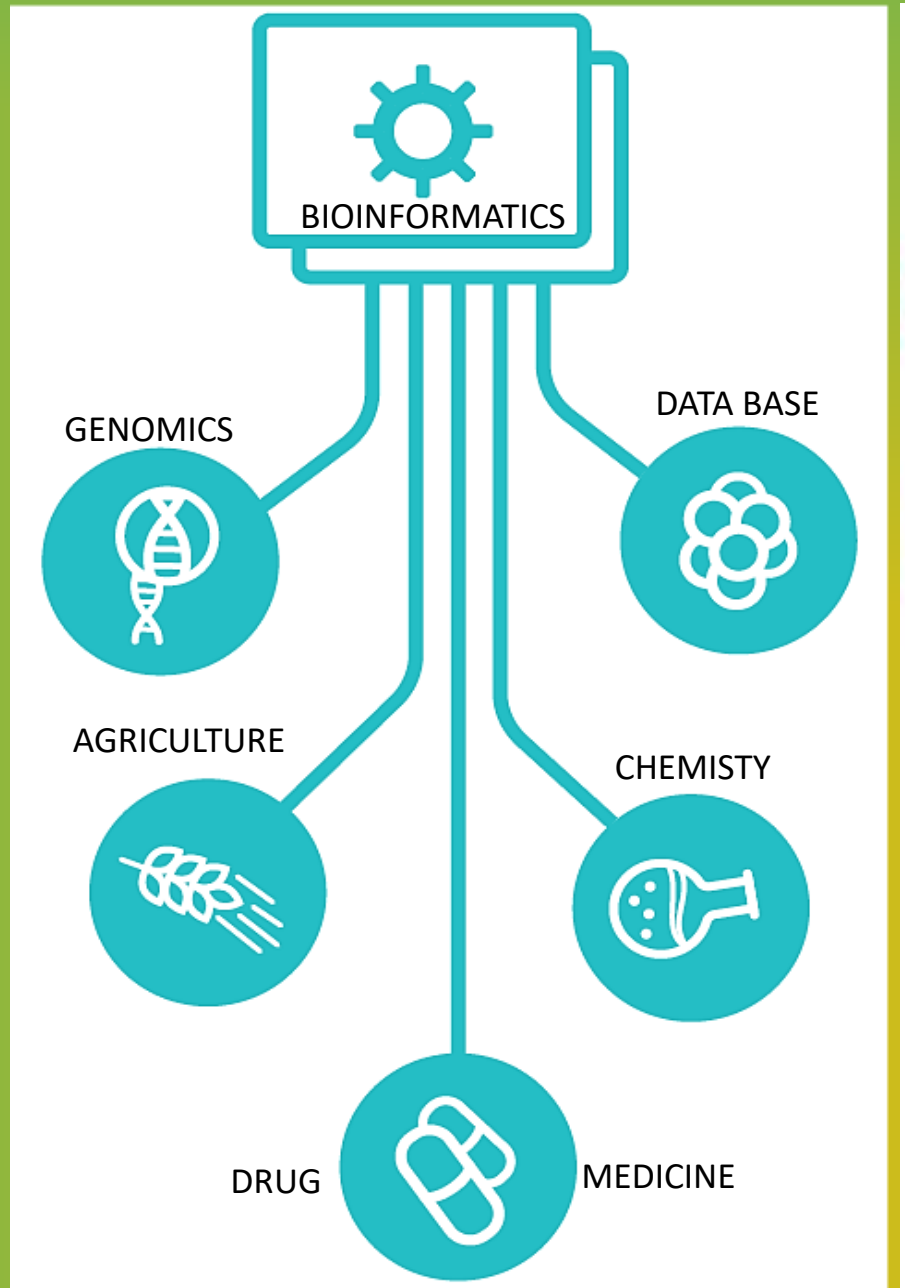
Application of information technology including statistics, mathematics to simplifying the storage, retrieval, analysis of biological dates



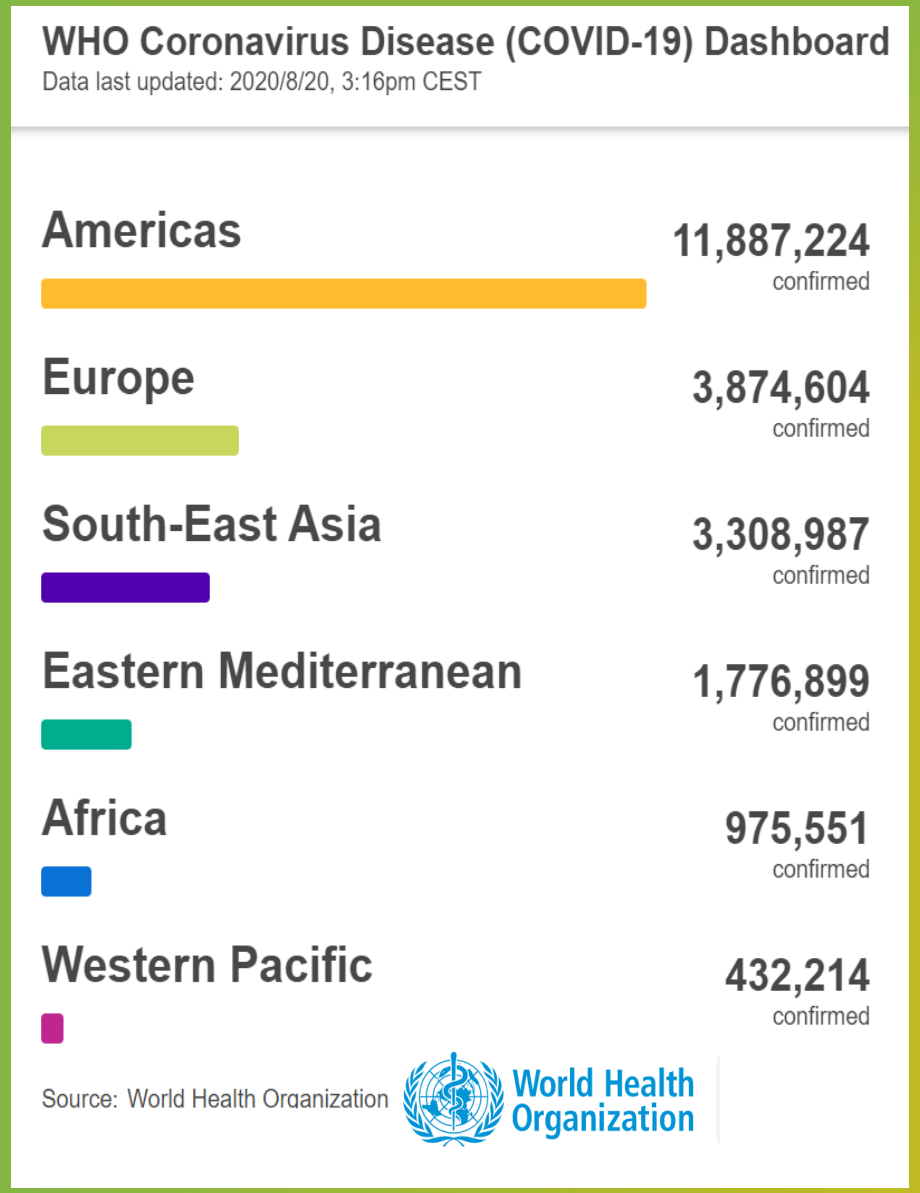


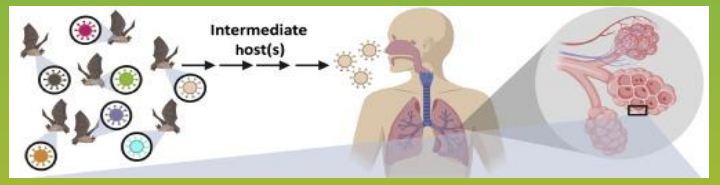


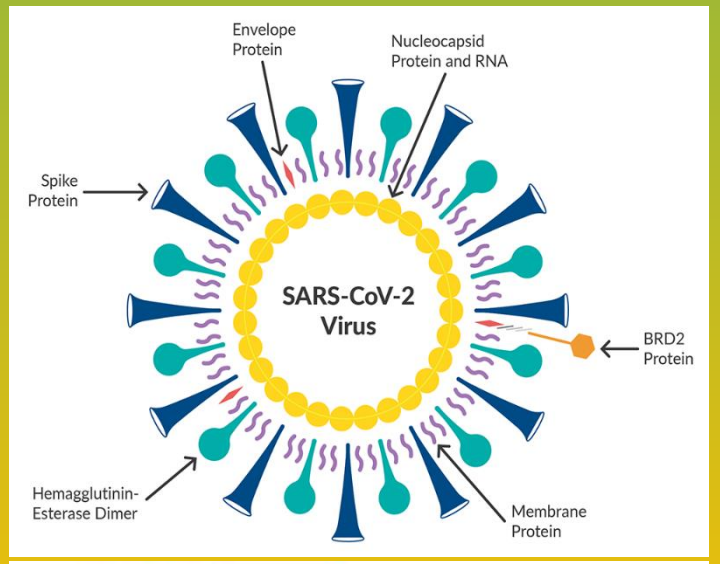
BIOINFORMATICS AREA OF APPLICATION

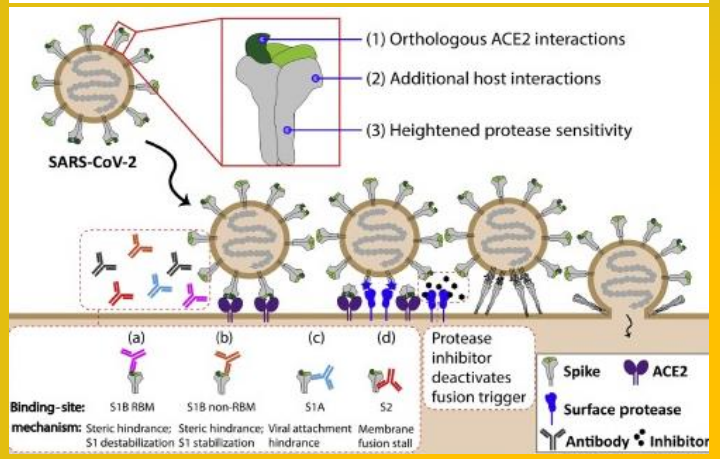


Coronavirus disease 2019 (COVID-19)









A novel **coronavirus** (nCoV) is a new strain that has not been previously identified in humans

TOOLS OF BIOINFORMATICS FOR COVID-19 RESEARCH

SEQUENCE LEVEL

1. Physicochemical properties
2. Secondary structure analysis
3. Conserve and mutation analysis
4. Phylogenetic tree analysis
5. 3-D structure prediction

STRUCTURE LEVEL

1. Structural composition
2. Therapeutic agent identification
3. Binding / active site prediction
4. Structural similarity analysis
5. Structural stability with drug

PHYSICOCHEMICAL PROPERTY PREDICTION

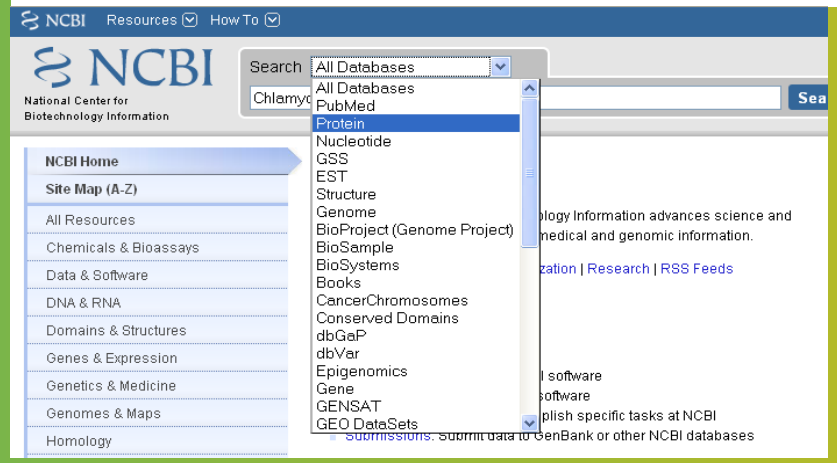
DNA

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUN/COV0425/2020, complete genome

GenBank: MT499219.1

>MT499219.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUN/COV0425/2020, complete genome

ACTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCA
CTTAGTGCACTCACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCT
CTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTTTGTCCGG
GTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTT
CCTGTTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGA
CACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTT
ACAGCCCTATGTGTTCATCAAACGTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGC
GTAGCAGAACTCGAAGGCATTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGT
GCGAAATACCAGTGGCTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCAT

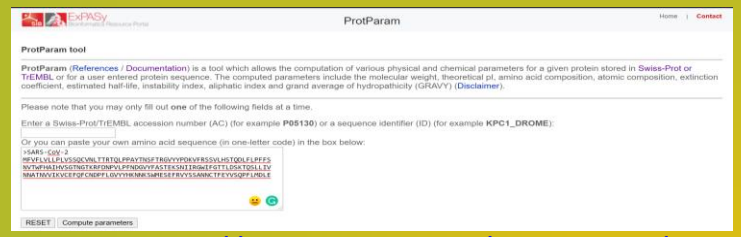


https://www.ncbi.nlm.nih.gov/

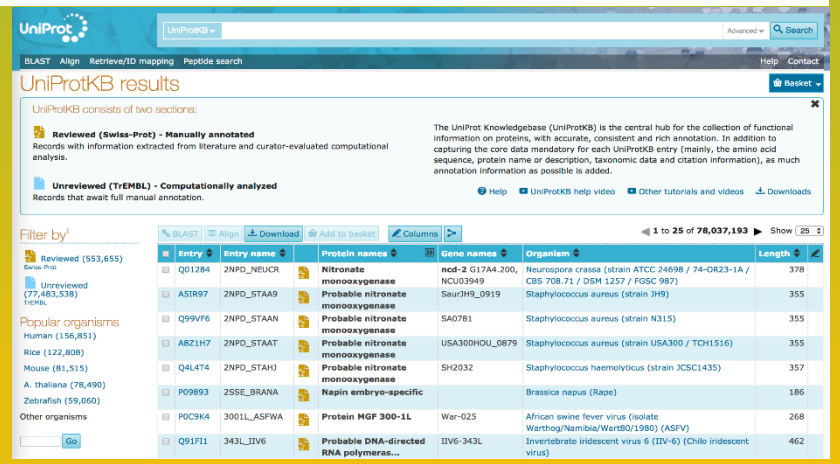
PROTEIN

>sp|P0DTC2|SPIKE_SARS2 Spike glycoprotein OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=S PE=1 SV=1

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFH
AIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCE
FQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNID
GYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAA
YYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFP
NITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVY
ADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLK
PFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPK
KSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGG
VSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
NGVEGFTESNKKFL



https://web.expasy.org/protparam/



https://www.uniprot.org/

- Molecular weight
- Total amino acids composition
- Total number of atoms
- Negatively charged residues
- Ext. coefficient
- Positively charged residues
- Grand average of hydropathicity
- Aliphatic index
- Protein as stable profile
- Estimated half-life
- Molecular formula
- Theoretical isoelectric point

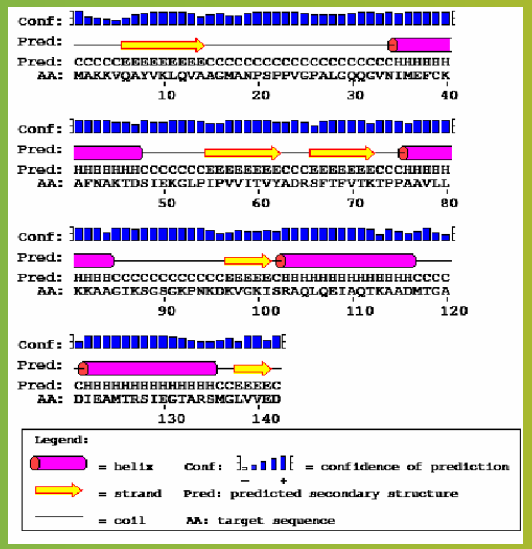
SECONDARY STRUCTURE PREDICTION

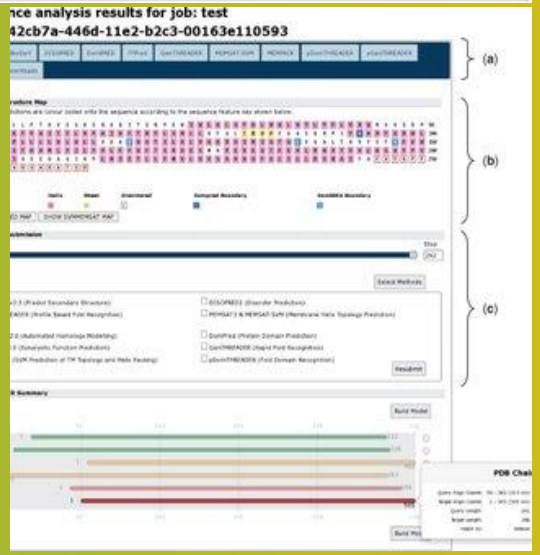
- Accurate prediction of the exact elements of protein 3D structure is essential for any research targeting a protein.

- Predicting the formation of protein structures such as alpha helices and beta strands, while for nucleic acids, it means predicting the formation of nucleic acid structures like helixes and stem-loop structures.

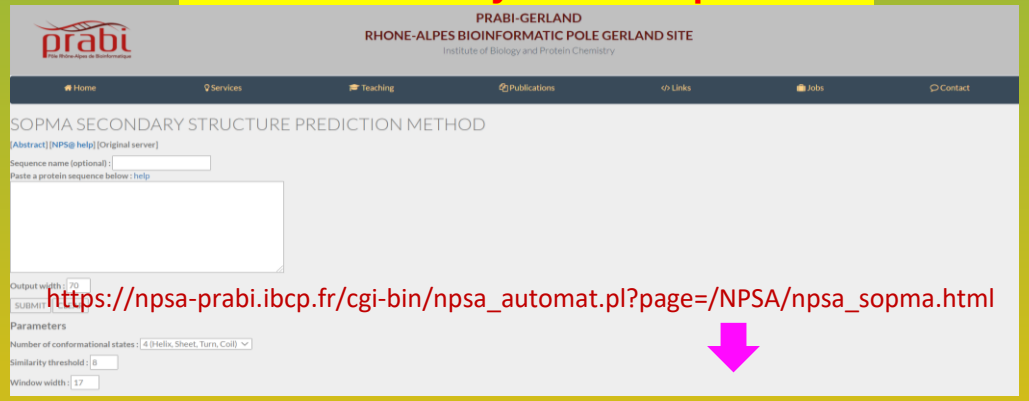
PSIPRED 4.0 (Predict Secondary Structure)

http://bioinf.cs.ucl.ac.uk/psipred/

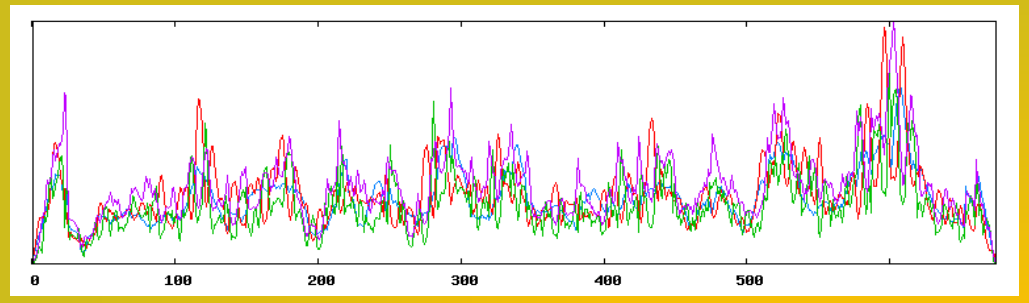


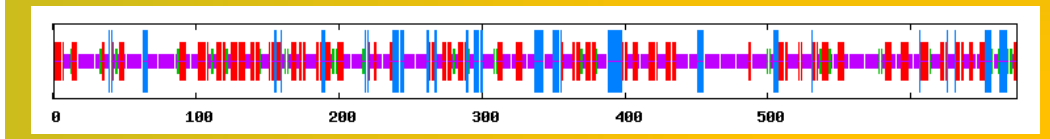


SOPMA secondary structure prediction



https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html



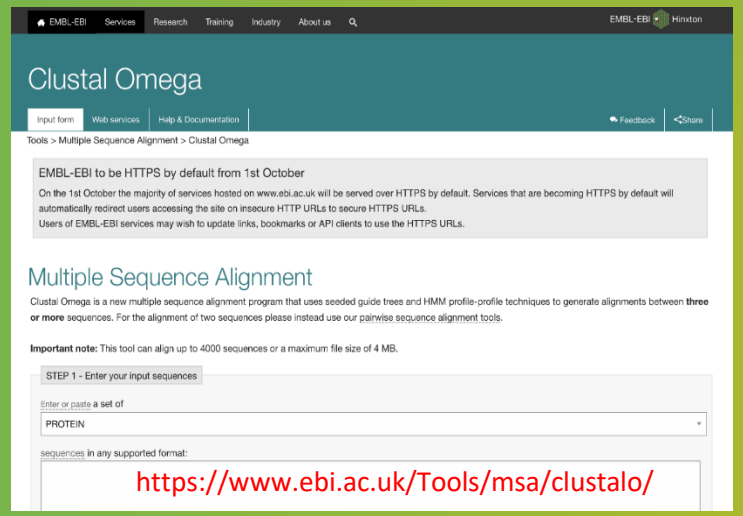


Alpha helix (Hh) : 84 is 12.46% 3_{10} helix (Gg) : 0 is 0.00% Pi helix (Ii) : 0 is 0.00% Beta bridge (Bb) : 0 is 0.00% Extended strand (Ee) : 196 is 29.08% Beta turn (Tt) : 41 is 6.08% Bend region (Ss) : 0 is 0.00% Random coil (Cc) : 353 is 52.37% Ambiguous states (?) : 0 is 0.00% Other states : 0 is 0.00%

https://List_of_protein_secondary_structure_prediction_programs

SEQUENCE ANALYSIS (MUTATION OR CONSERVED)

- A gene mutation is a permanent alteration in the DNA, protein sequence that makes up a gene, such that the sequence differs from what is found in most people/ region; In simply, sometimes our DNA sequence gets altered; this is called a mutation.



https://www.ebi.ac.uk/Tools/msa/clustalo/

aa POSITION	Kolkata
271	Q
408	R
614	G
723	I (1) (4T)
930	A
1124	V (2) (3G)

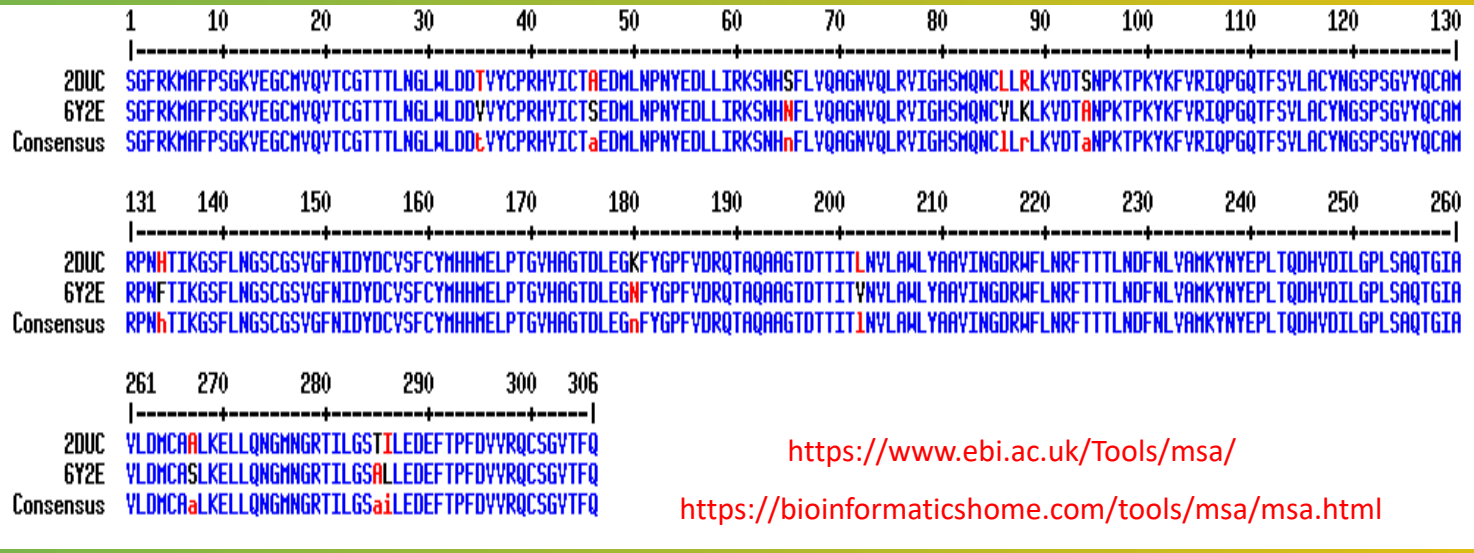
	STATES			
aa POSITION	Kolkata	Gujarat	Kerala 29	Kerala 166
271	Q	R	Q	Q
408	R	R	I	R
614	G	G	D	D
930	A	A	A	V
1124	V	G	G	G

Figure 2: Mutation analysis of isolates from Kolkata, Gujarat and Kerala

A. Multiple sequence alignment of Spike protein sequence of Kolkata isolate with sequences obtained from other parts of India. Sites of mutation are showed in Red

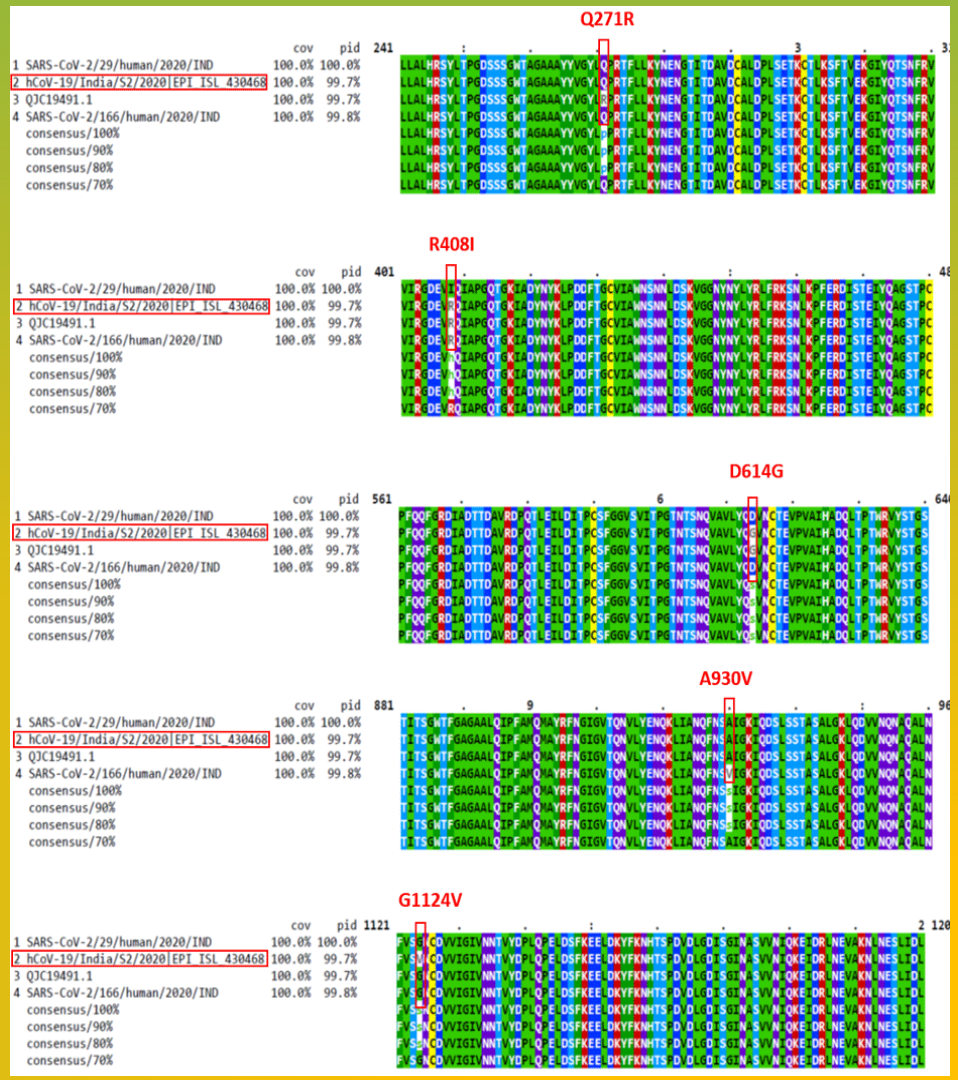
B. Tabulation of amino acid mutations among isolates from Kolkata. Mutations are shown in red. Number/s in parenthesis show number of isolates that showed the amino acid type.

C. Tabulation of amino acid present at the points of mutation for isolates from different parts of India.



https://www.ebi.ac.uk/Tools/msa/

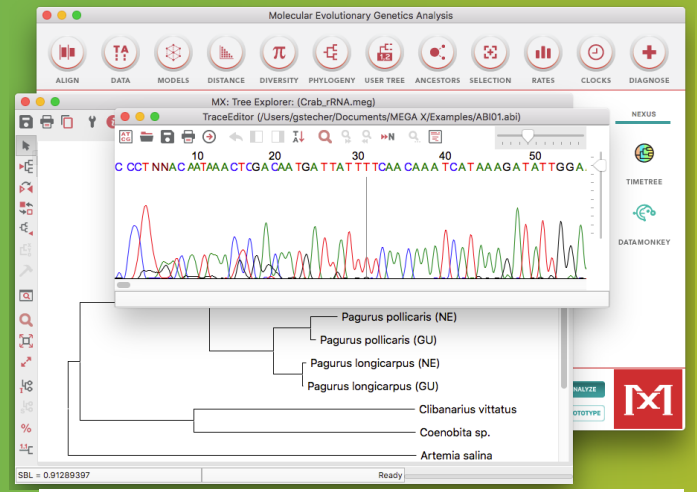
https://bioinformaticshome.com/tools/msa/msa.html

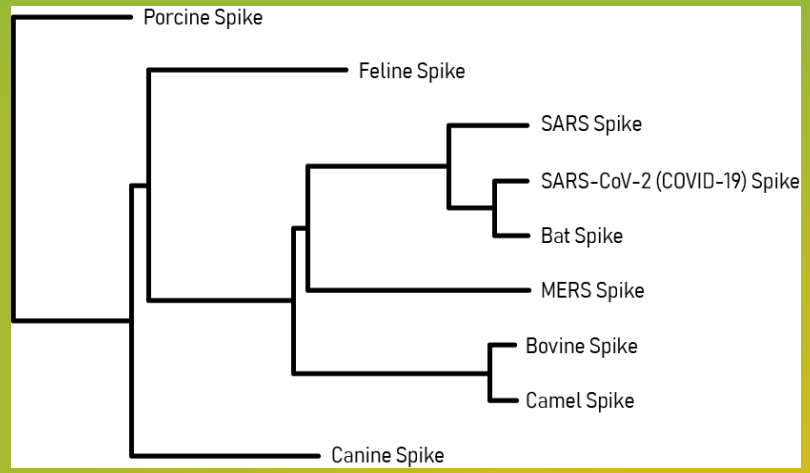


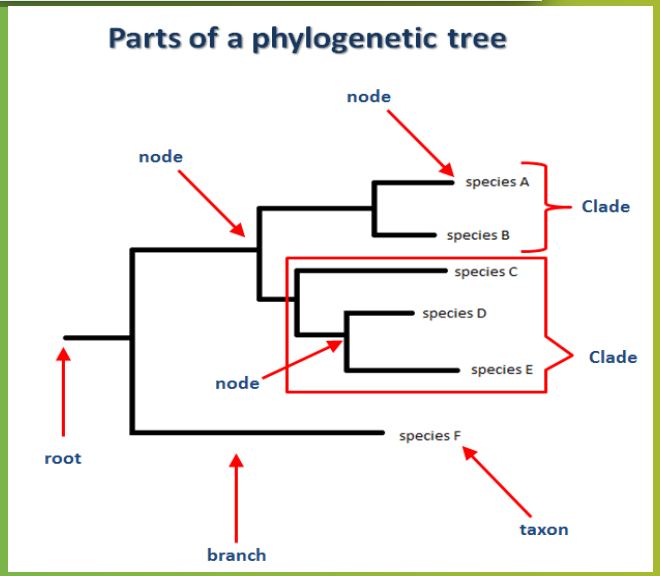
PHYLOGENETIC TREE ANALYSIS/ BUILDING

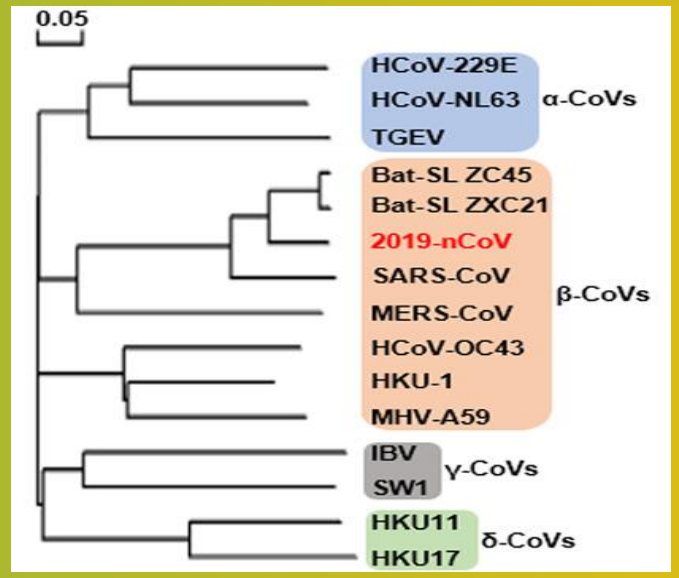
Phylogenetics is the study of evolutionary relationships among biological entities

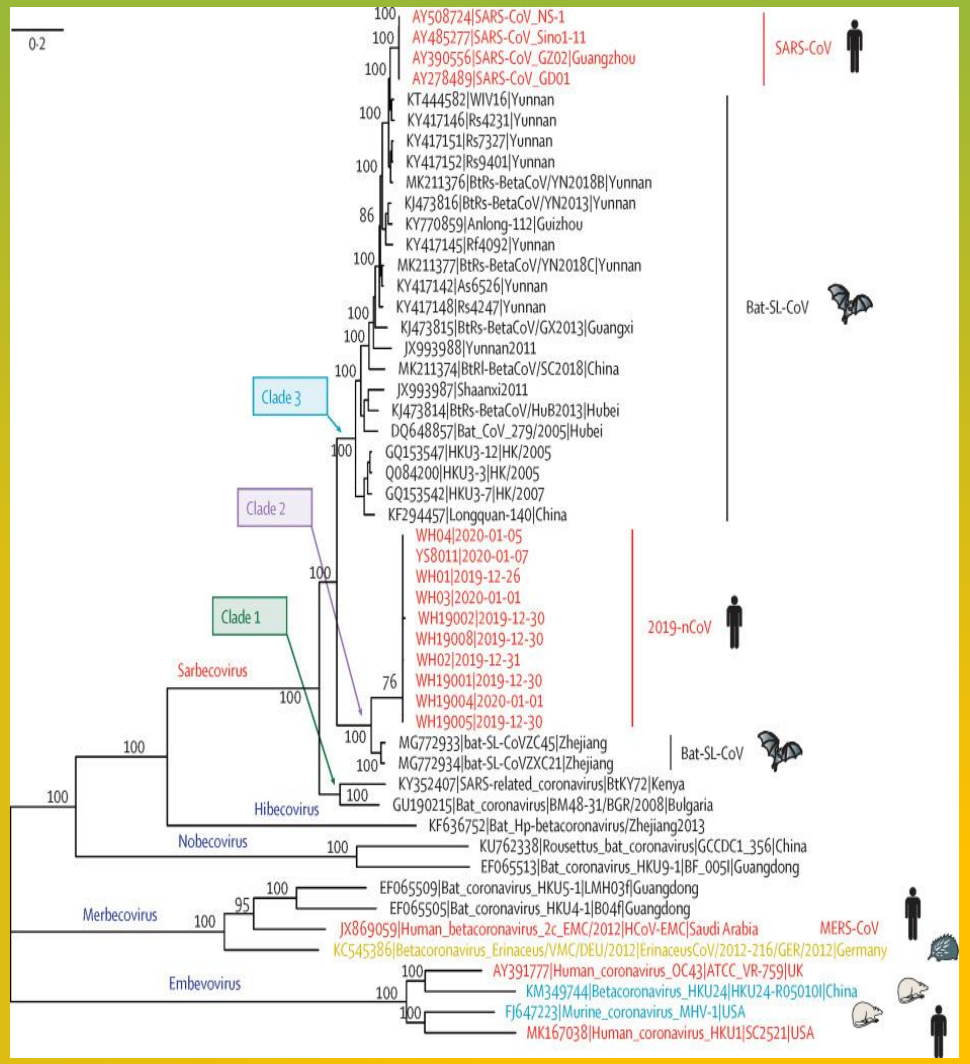
Most usable software's: MEGA, Dendroscope, FigTree, Phylotree, ggtree









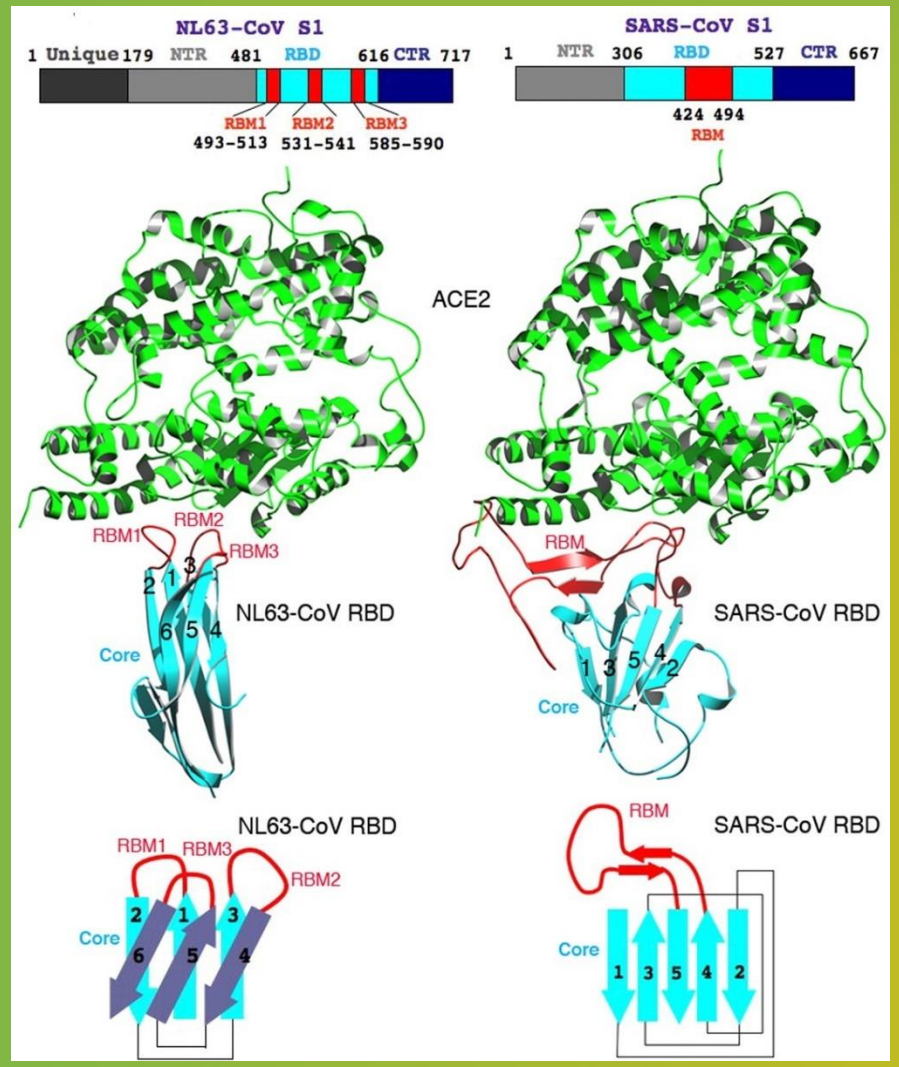


https://molbiol-tools.ca/Phylogeny.htm

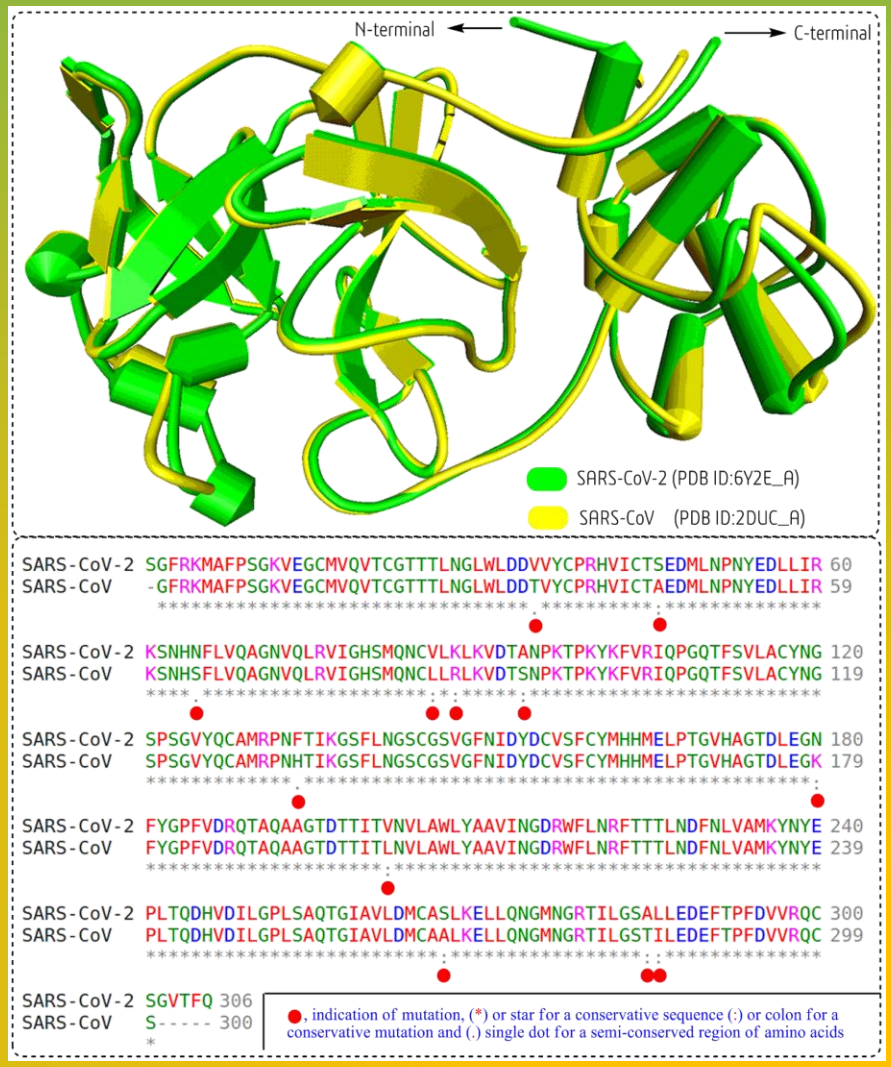
https://en.wikipedia.org/wiki/List_of_phylogenetics_software

STRUCTURAL SIMILARITY ANALYSIS

- An image quality metric that assesses the visual impact of a protein structure characteristics/ building blocks

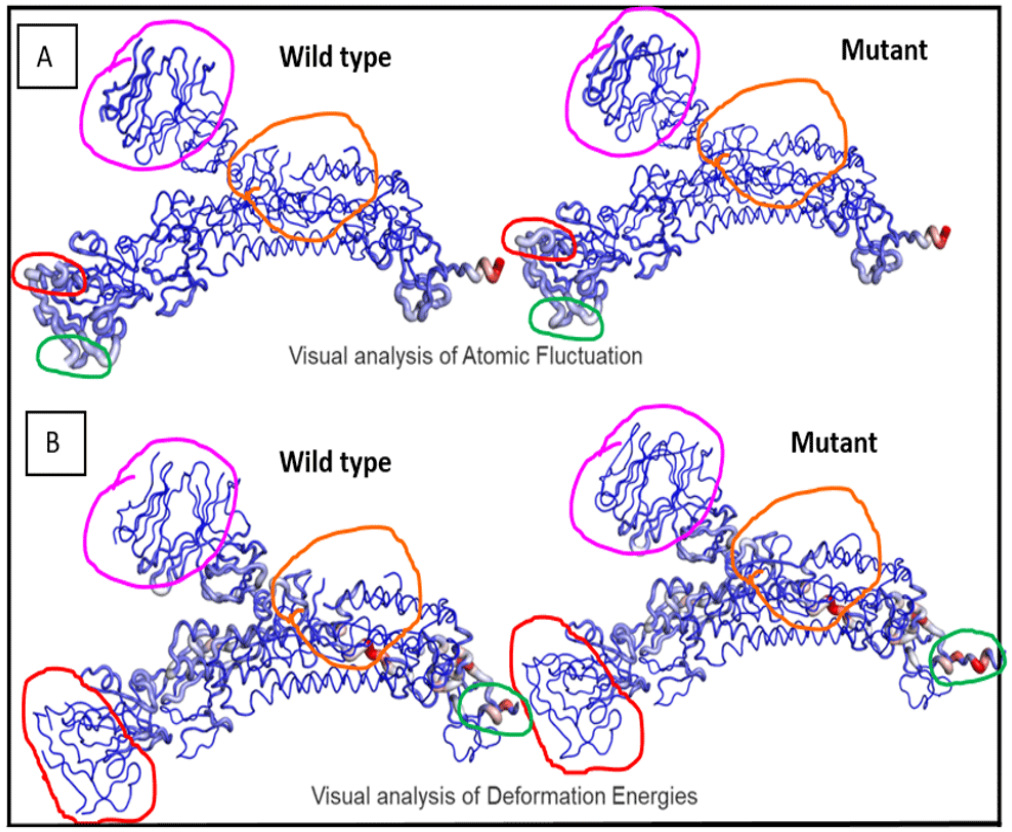


SOFTWARE'S

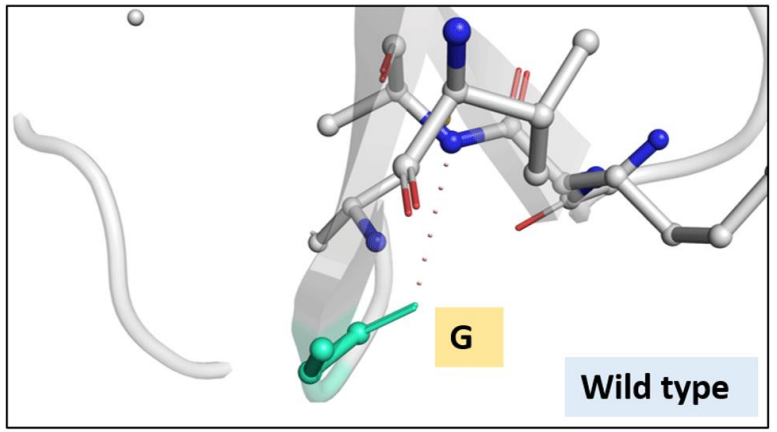


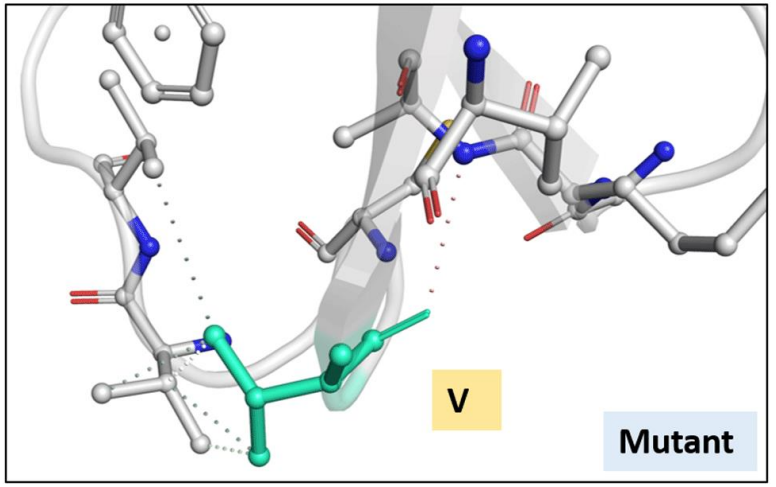
STRUCTURAL STABILITY ANALYSIS IN MUTANT PROTEIN

- Proteins are highly dynamic molecules, whose function is intrinsically linked to their molecular motions (analysis carried out by the tool, DynaMut).
- Despite the pivotal role of protein dynamics has led to most structure-based approaches for assessing the impact of mutations on protein structure and function relying upon static structures.



Magnitude of (**A**) atomic fluctuation and (**B**) deformation has been shown using thin to thick lines coloured blue (low), white (moderate) and red (high).

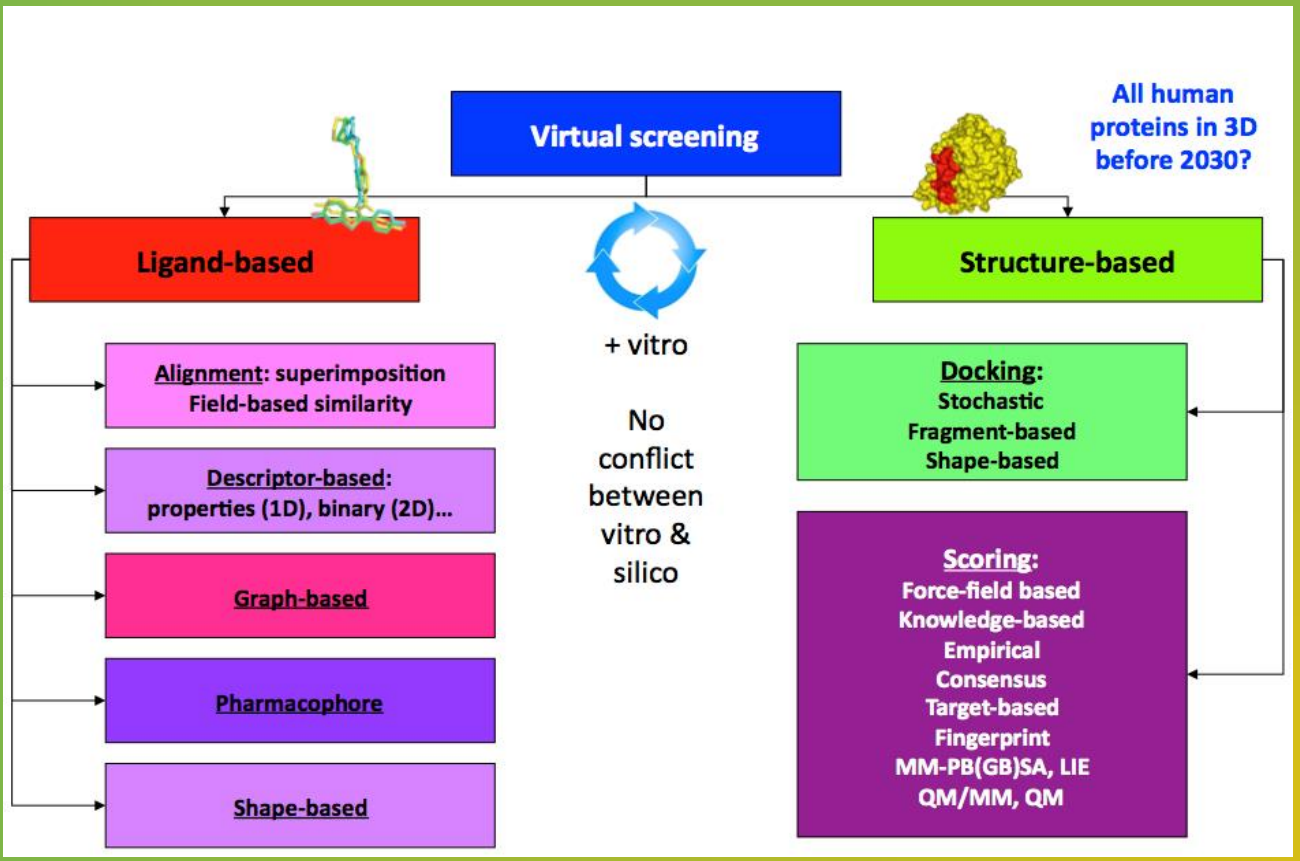


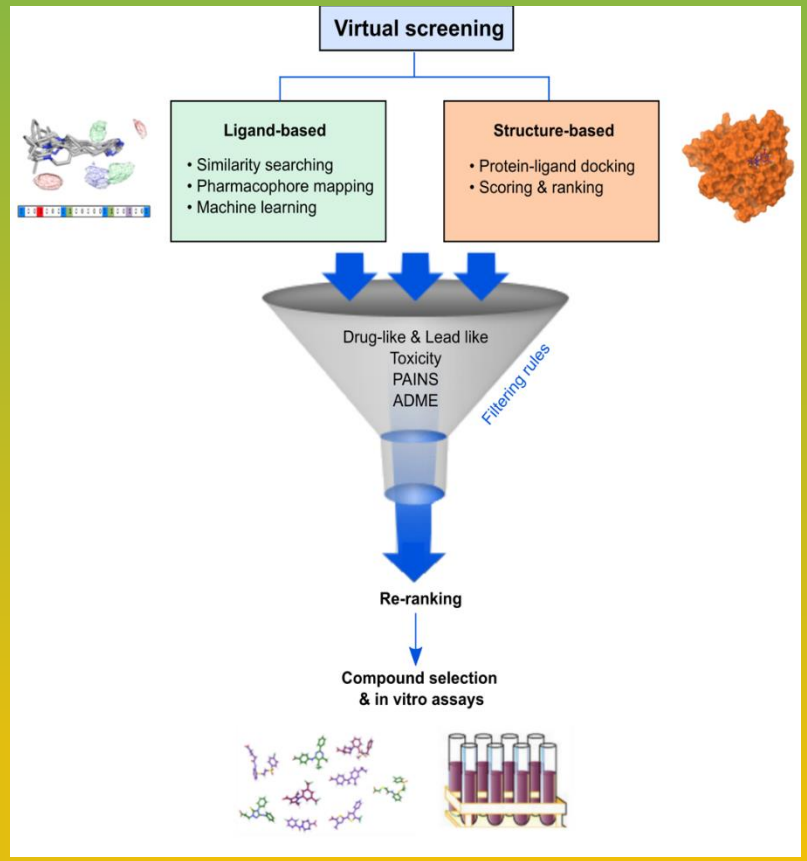


Wild-type and mutant residues are coloured in light-green and are also represented as sticks alongside with the surrounding residues which are involved on any type of interactions.

To correlate if changes in secondary structure are also reflected in the dynamics of the protein in its tertiary structure, performed normal mode analyses and studied protein stability and flexibility. Change in vibrational entropy energy (ΔΔSVib ENCoM) between the wild type Wuhan isolate and the West Bengal isolate was -4.445 kcal.mol^{-1}.K^{-1}. The ΔΔG was 0.905 kcal/mol and the ΔΔG ENCoM was 4.756 kcal/mol. All these suggested a stabilizing mutation in this type of spike.

DRUG DISCOVERY USING BIOINFORMATICS TOOLS

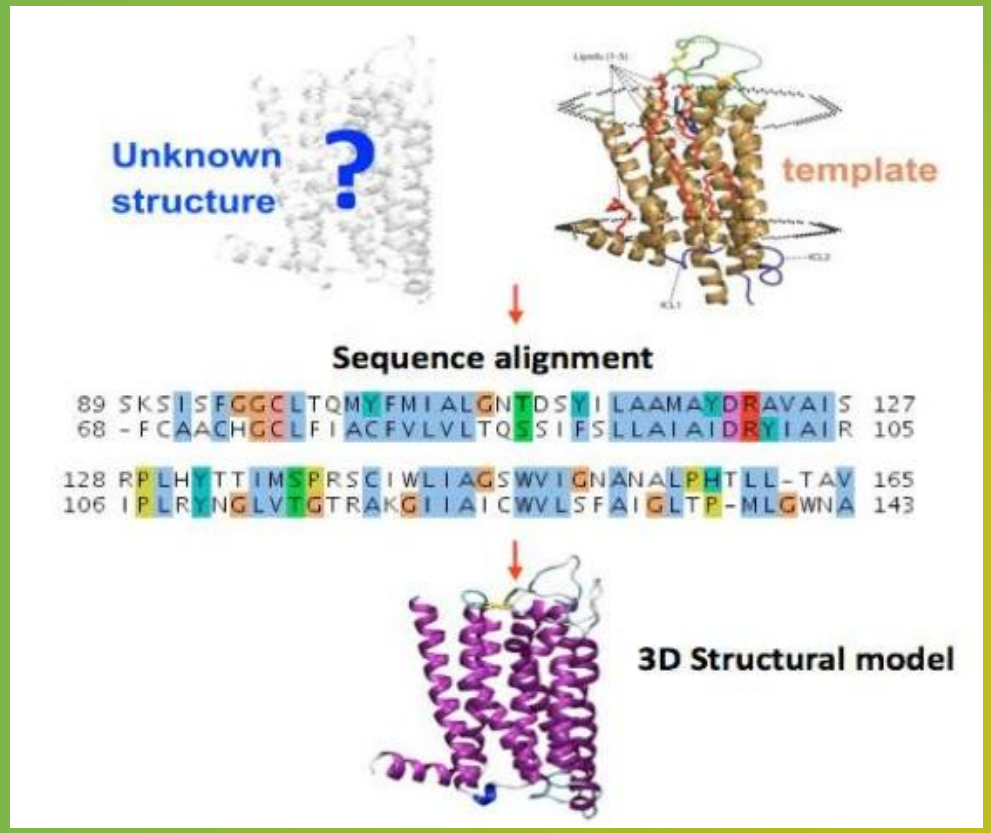


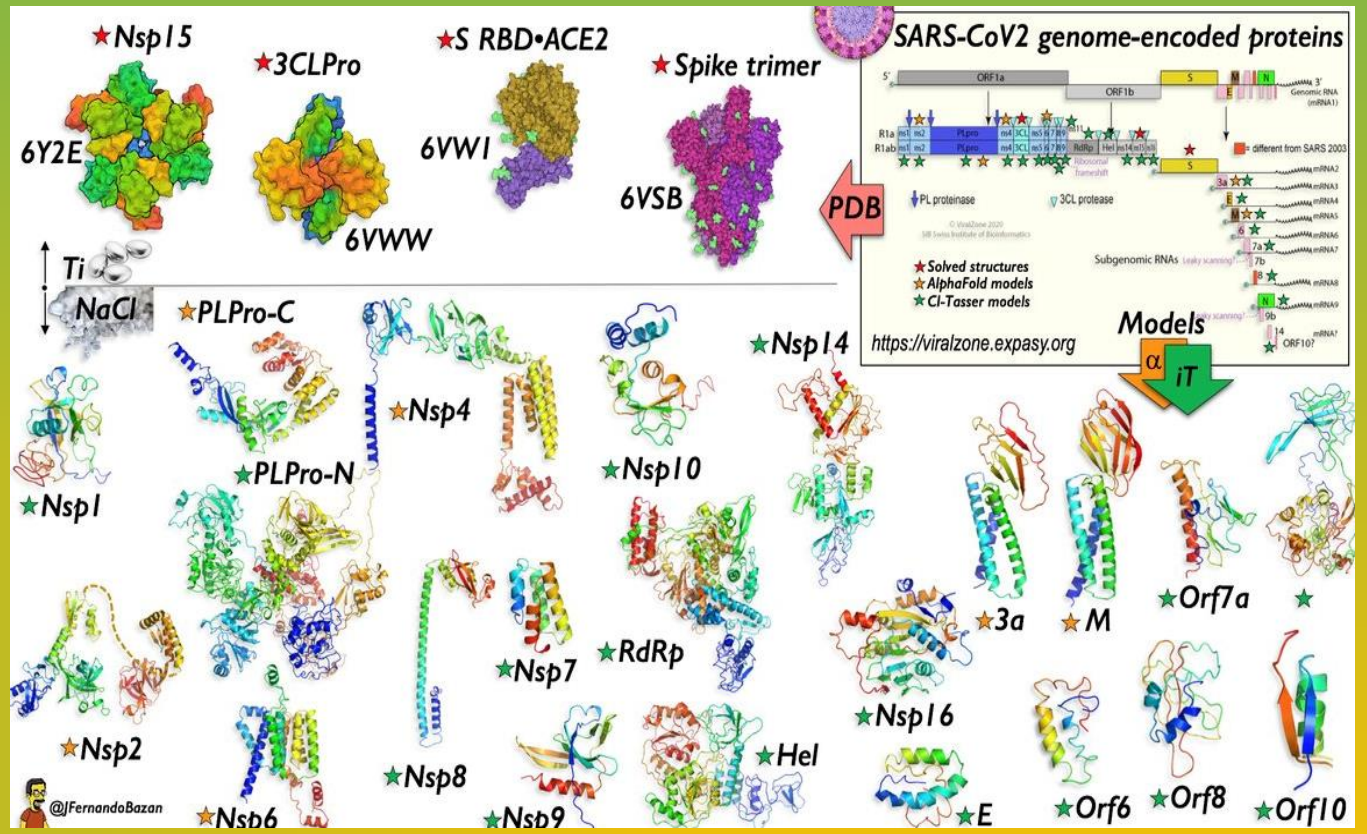


LIGAND BASED DRUG DESIGN: It is otherwise known as indirect drug design. It trusts on the awareness of different new ligand molecules that bind with the target protein molecule. (**Known ligand with unknown receptor**).

STRUCTURE BASED DRUG DESIGN: It depends on the wisdom of three-dimensional structure of the protein molecule. Practically the structure was initially identified by X-ray crystallography which improves the aptitude to produce new drugs that fight against diseases. (**Known receptor with unknown ligand**).

3D PROTEIN STRUCTRE MODELLING & RETRIEVAL

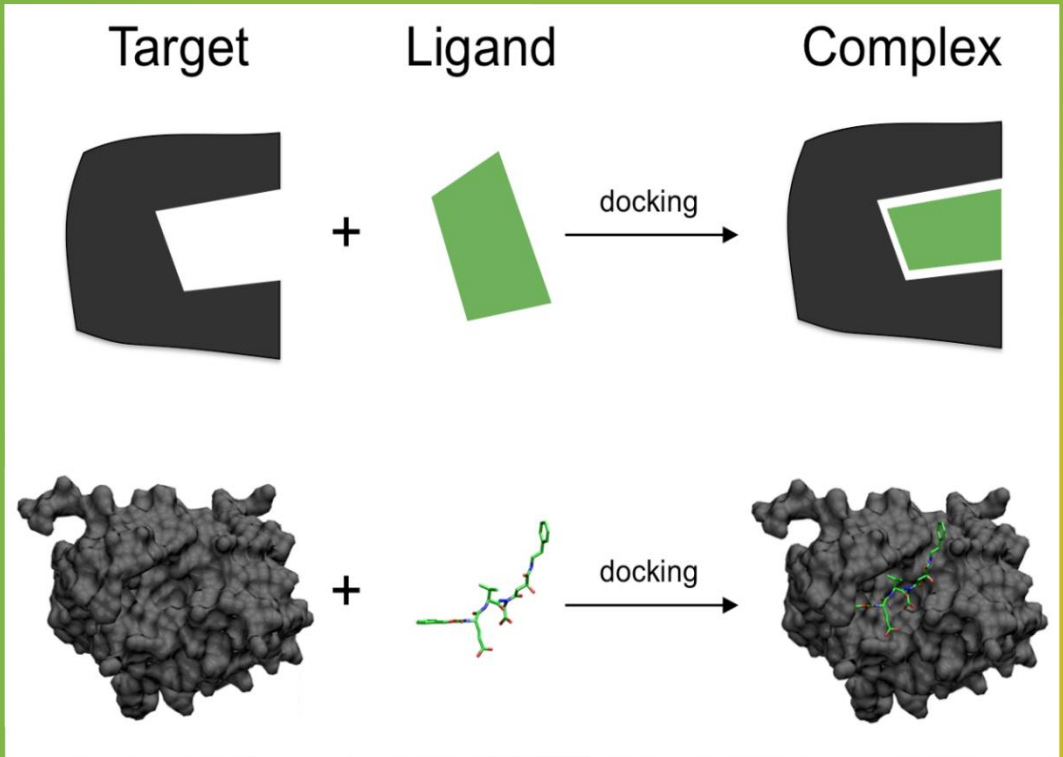


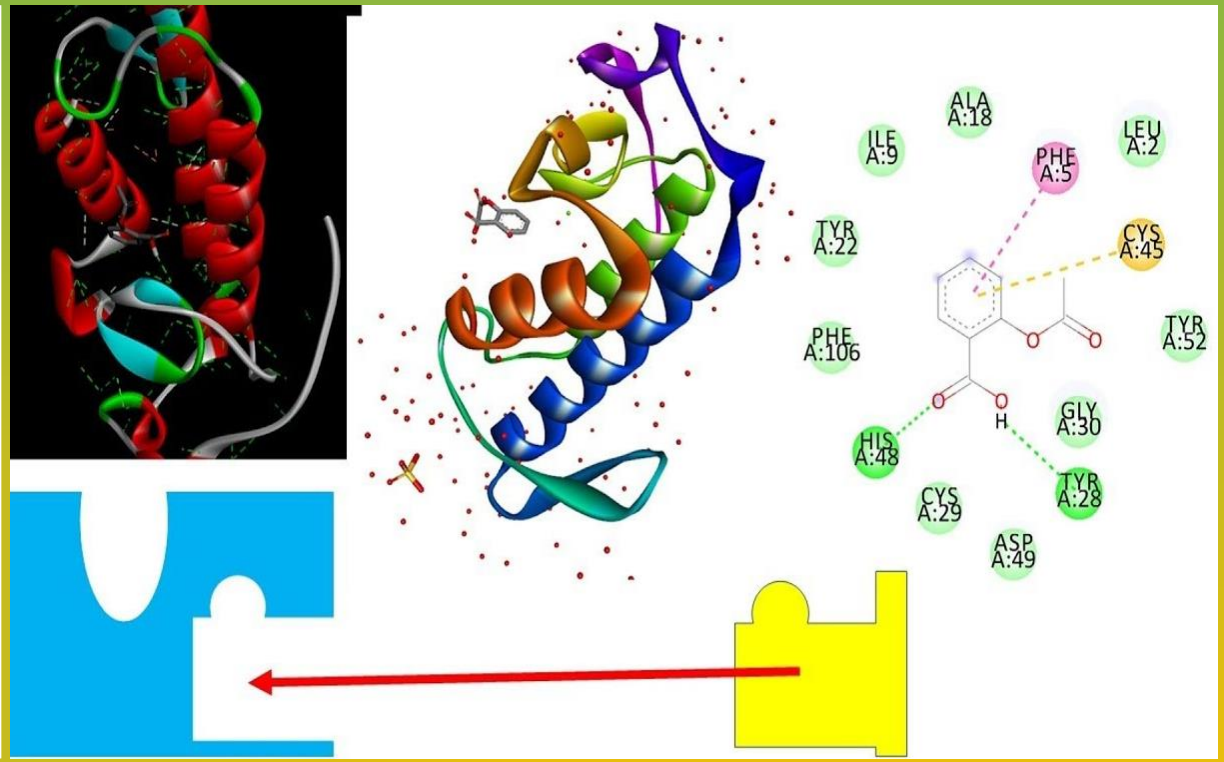


- The term "homology modeling“ or template-based modeling, refers to modeling a protein 3D structure using a known experimental structure (the template).
- Structural information is always of great assistance in the study of protein function, dynamics, interactions with ligands and other proteins.

- Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, (http://www.wwpdb.org/).
- The structural data typically obtained by X-ray crystallography, NMR spectroscopy, cryo-electron microscopy by world wide biologists and biochemists and are freely accessible.

CONCEPT OF MOLECULAR DOCKING



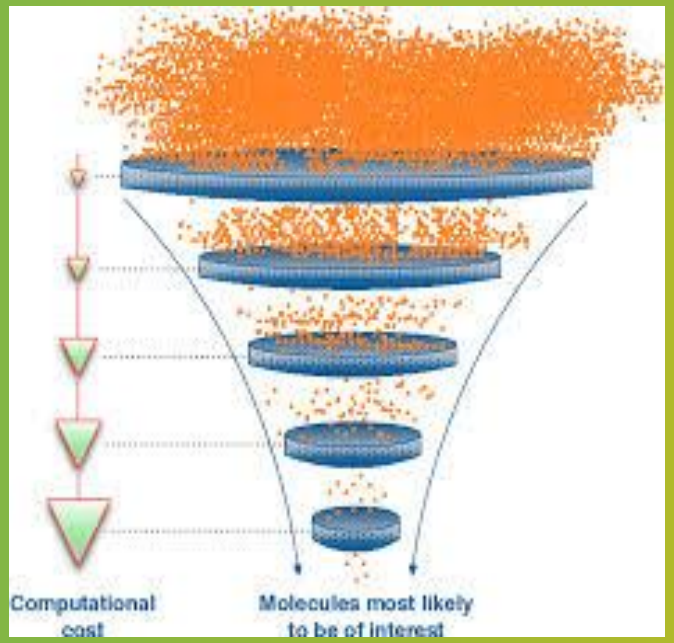


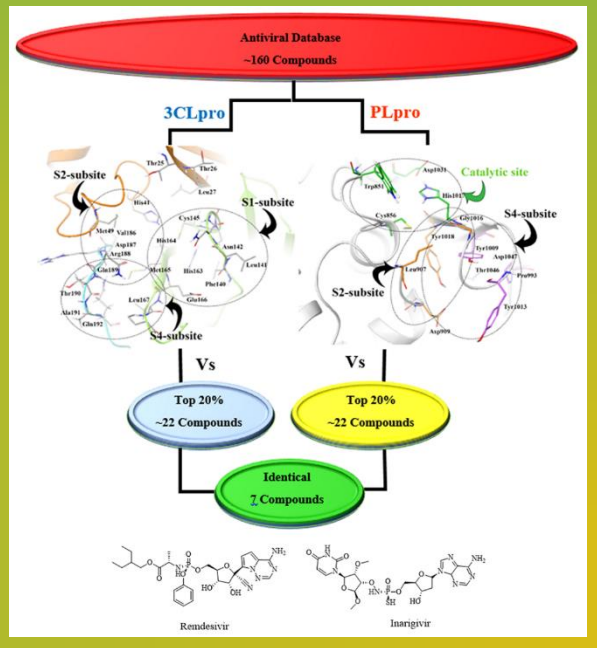
- Molecular docking, which predicts interaction patterns based on scoring function between proteins and small molecules as well as proteins and proteins, to evaluate the binding between two molecules is widely used in the field of drug screening and design.

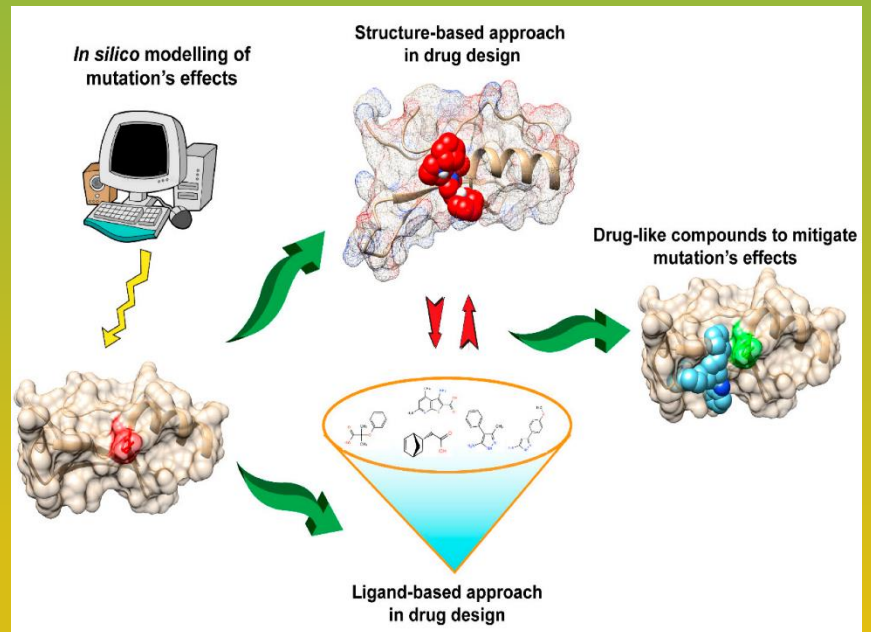
- It is currently used as a standard computational tool in drug design for lead compound optimisation and in virtual screening studies to find novel biologically active molecules.

HIGH-THROUGHPUT VIRTUAL SCREENING

Virtual screening is an important part of computer-aided drug design methods. It may be the cheapest way to identify potential lead compounds, and many successful cases have proven successful using this technology







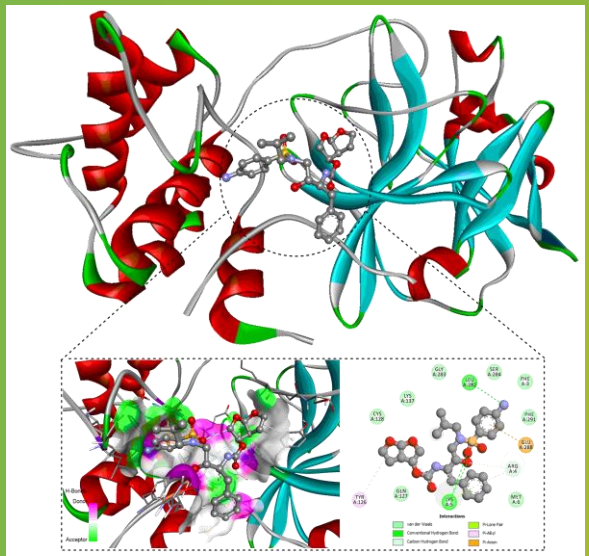
	Sections	Type of modeling	License	Reference
GOLD	Molecular docking and quantum mechanics, a combined mechanistic tool, Modeling ligand–protein interactions in drug design, Solvent effect as an important parameter	Protein–ligand	Commercial	Verdonk et al[5]
Gaussian (ONIOM)	Molecular docking and quantum mechanics, a combined mechanistic tool	QM/MM	Commercial	Gaussian[6]
AUTODOCK	Modeling ligand–protein interactions in drug design, Solvent effect as an important parameter	Protein–ligand	Open	Morris et al[7]
GLIDE	Modeling ligand–protein interactions in drug design, Solvent effect as an important parameter	Protein–ligand	Commercial	Friesner et al[8]
RosettaDock	Biological complexes and quaternary structures, the protein–protein docking approach	Protein–protein	Open	Lyskov and Gray[9]
pyDOCK	Biological complexes and quaternary structures, the protein–protein docking approach	Protein–protein	Open	Jimenez-Garcia et al[10]
AquaSol	Solvent effect as an important parameter	Solvent effect	Open	Koehl and Delarue[11]

Abbreviations: ONIOM, our own N-layered integrated molecular orbital and molecular mechanics; QM, quantum mechanics; MM, molecular mechanics.

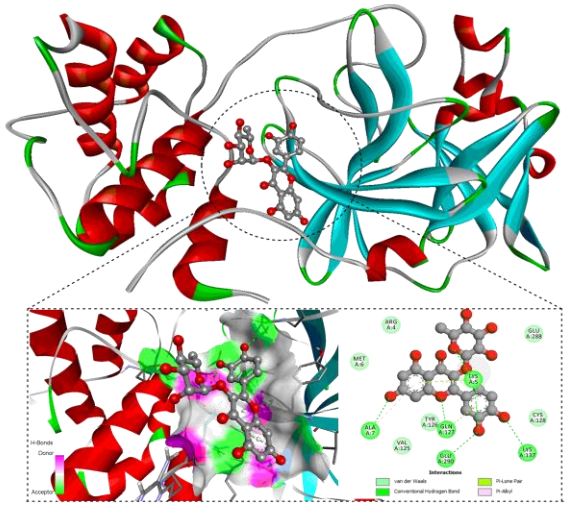
Docking approach	Examples
Matching of descriptors	DOCK, QSDOCK, SLIDE
Incremental construction	FlexX, Hammerhead
Monte Carlo Simulated Annealing	AutoDock, MCDOCK
Monte Carlo Minimization	ICM, QXP
Molecular Dynamics	MDD
Genetic Algorithms	GOLD, AutoDock3

https://en.wikipedia.org/wiki/List_of_protein-ligand_docking_software

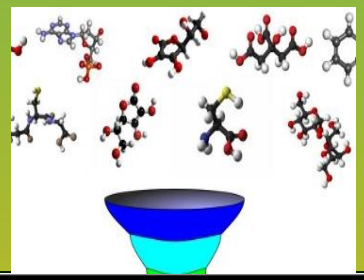
PROTEIN-LIGAND INTERACTION

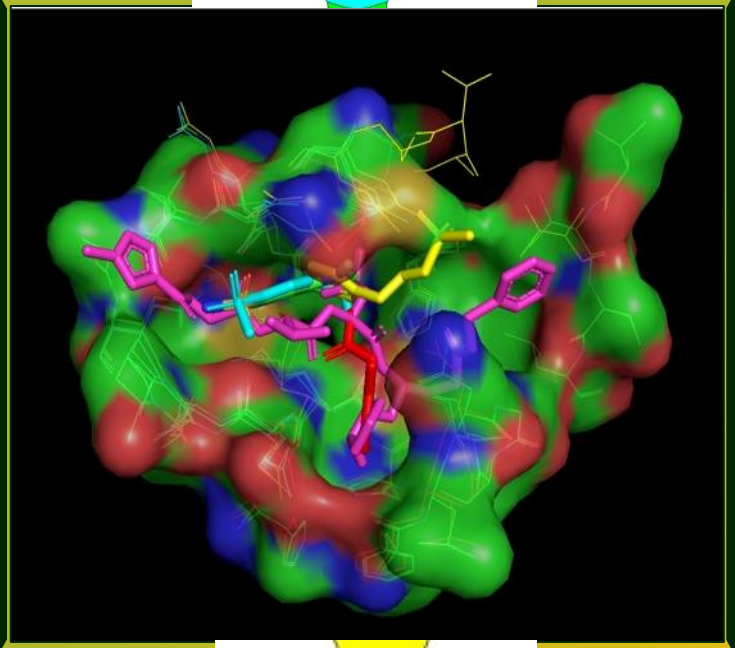


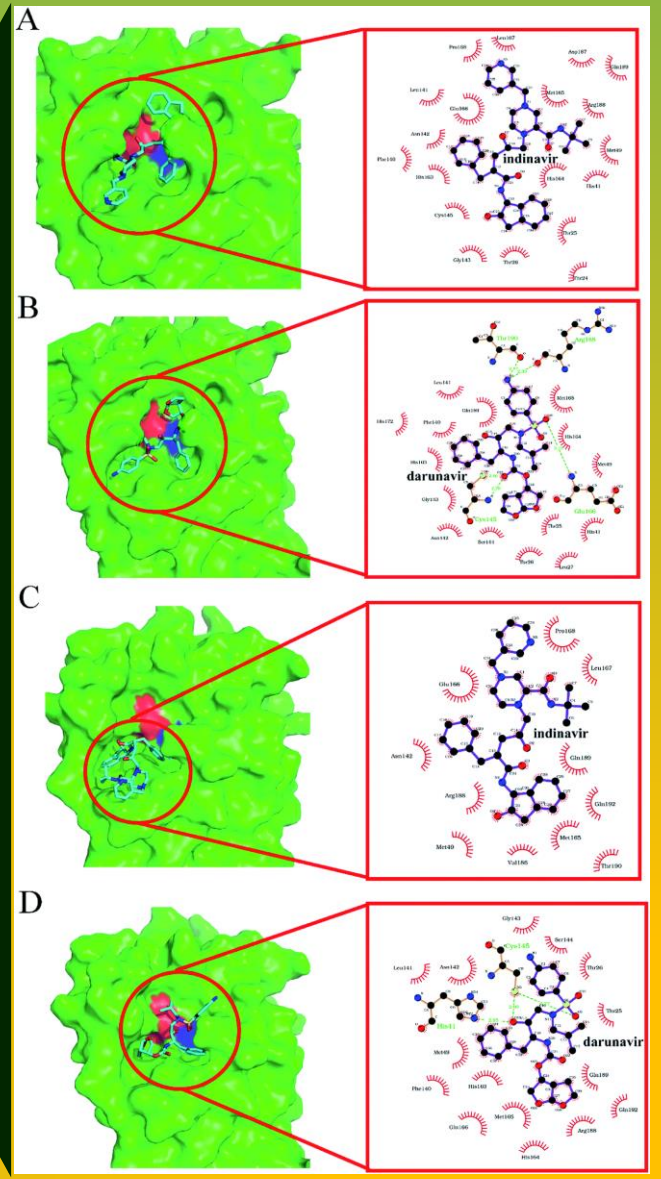
SARS-CoV-Mpro (PDB ID: 6Y2E)-Darunavir



SARS-CoV-Mpro -Quercetin-3-rhamnoside

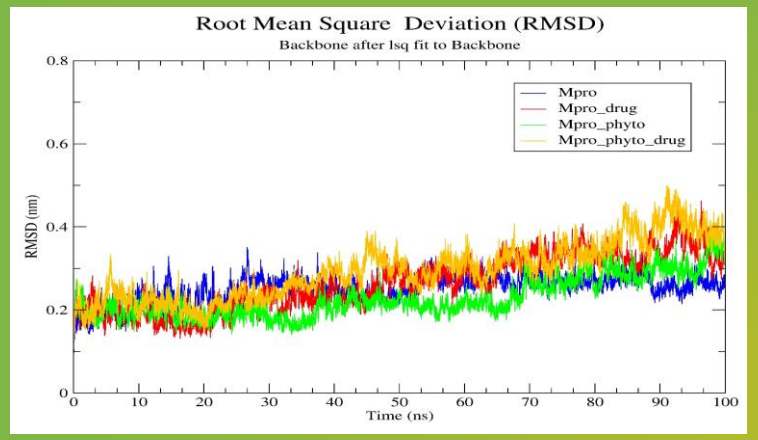


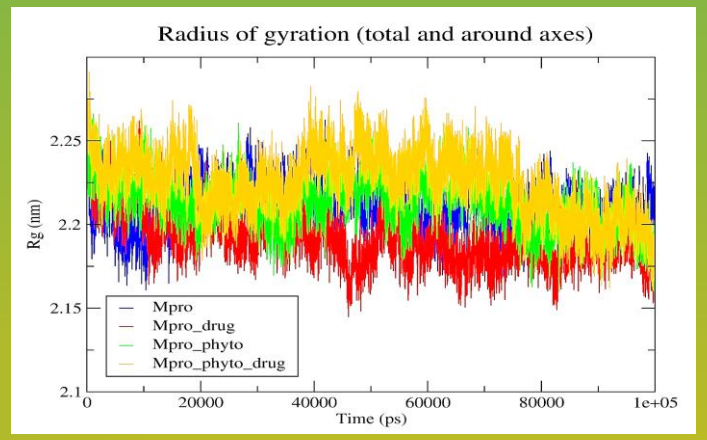


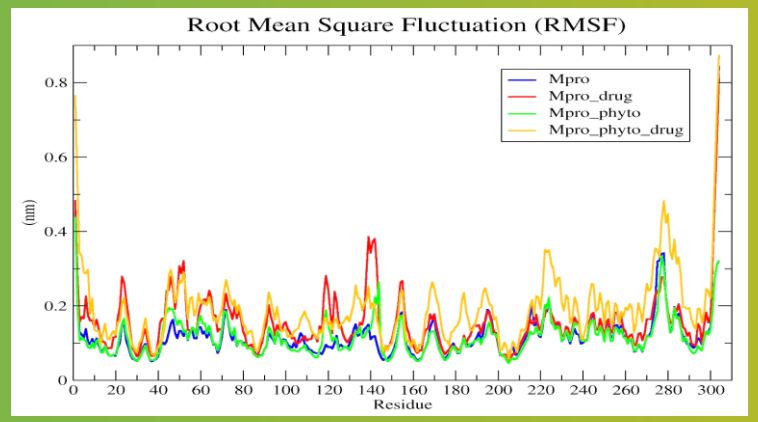


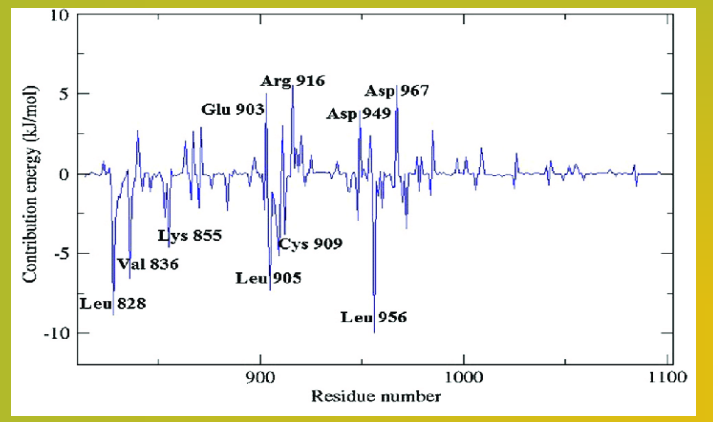
ADVANCED BIOINFORMATICS TOOLS IN DRUG DISCOVERY

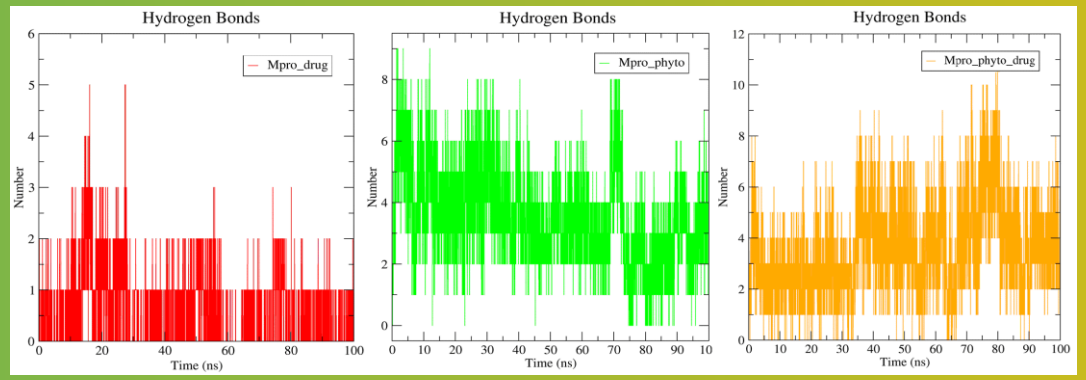
MD-SIMULATION

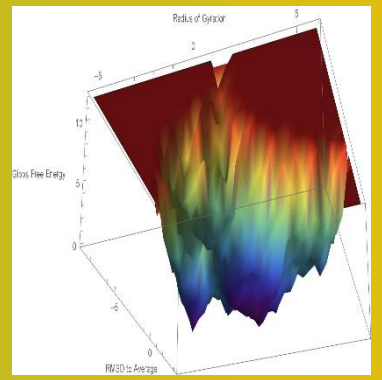


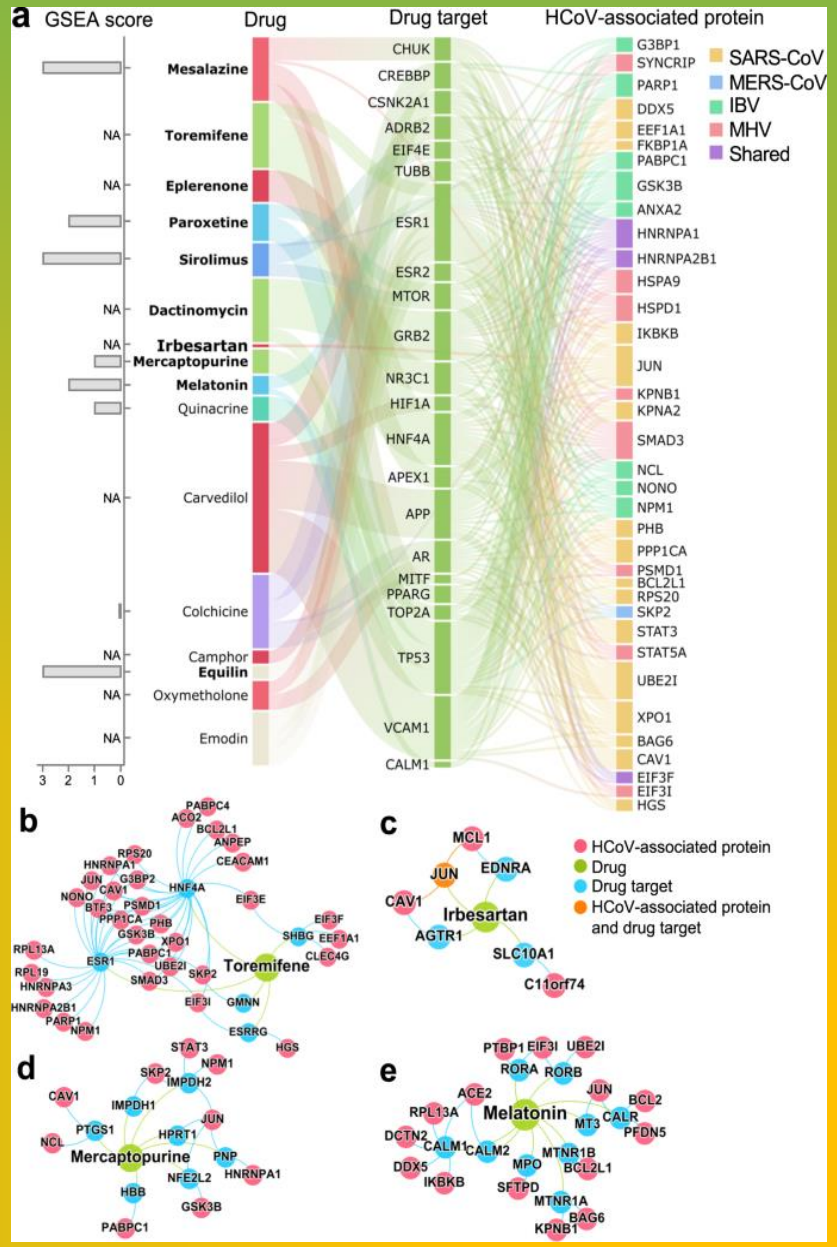












GENE NETWORKING

CONCLUDING REMARKS

Various common advantages of computational method drug design as follows

- To reduce the complexity
- Time consuming
- Accurate results
- Reproducibility
- Lower cost
- Novel target identification

Major advantages of computation in the drug design process as follows

- Virtual screening and de novo drug design
- *In silico* pharmacokinetic properties prediction
- Improved methods for to determine protein-ligand binding.

Thank you for your patience

STAY HOME ..STAY HEALTHY FROM