Indian J. Applied & Pure Bio. Vol. 38(3), 1341-1345 (2023). **ISSN: 0970-2091**
A web of Science Journal

Conformational analysis of SARS Covid using protein database

***C. Thurigha and S. Arul Mugilan**

P.G. & Research Department of Physics,
Kamarajar Government Arts College, Surandai-627859, Tamilnadu, (India),
Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamilnadu, (India)
*Register Number : 21211062132001
*e-mail: thuriakash@gmail.com

Abstract

Globally, 10 million people have been infected by the COVID-19-causing severe acute respiratory syndrome coronavirus 2 (SARSCoV-2). Due to the coronavirus's fast spread, its DNA has undergone significant changes. Biologists can considerably benefit from the secondary structure of a protein sequence in the creation and testing of hypotheses. Proteins are crucial for the proper operation of living things. Using bioinformatics databases, structural biologists may extrapolate secondary structural components from amino acid sequences. It has been demonstrated that the relationship between bond length and angle created with an adjacent bond in simple molecules is inverse, with shorter bond lengths resulting from bigger bond angles. Bond order is the whole variety of bonds between two atoms. When Bond Length is the measurement of the space between two atoms within a molecule. It is closely associated with stability. In this project, the conformational analysis of amino acids and sequential pair should be derived. With the use of Python and RasMol programming, calculations of SARS Covid Spike protein can be determined. α - helix, which makes up 30% of typical globular proteins, which are the most prevalent secondary structure. Various interactions (α helix & β sheet) of Protein folding result from unpaired interactions between α- helix and β- sheets. This research gives in-depth insights into the structure and dynamics of the SARS-CoV-2 protein that might be used to develop strong and specific inhibitors that target the membrane protein and are used in the process of creating drugs.

Key words : Protein Data Bank, RASMOL, Python Programming

The severe acute respiratory syndrome coronavirus 2 (SARSCoV-2) that causes coronavirus disease 2019 (COVID-19) has infected ten million people worldwide since 2022, and the rapid spread of this novel coronavirus has resulted in significant mutations in the virus genome[5]. Protein structure is generally described at a primary or three-dimensional level and also by hydrophobic interactions among the amino acids and other intramolecular forces (Vander Waals interactions) these interactions affect secondary, tertiary and quaternary structure.[4] Aminoacid sequence influences the conformation of proteins and affects their biological function. It has been demonstrated that the relationship between various interactions created with an adjacent bond in simple molecules is inverse, with shorter bond lengths resulting from bigger bond angles. Proteins interact with other molecules to handle particular processes. In these circumstances, the protein residues that physically bind to the ligands are referred to as the binding sites. Ligands are tiny chemicals that join with proteins to generate a complex that performs a biological function.[2]

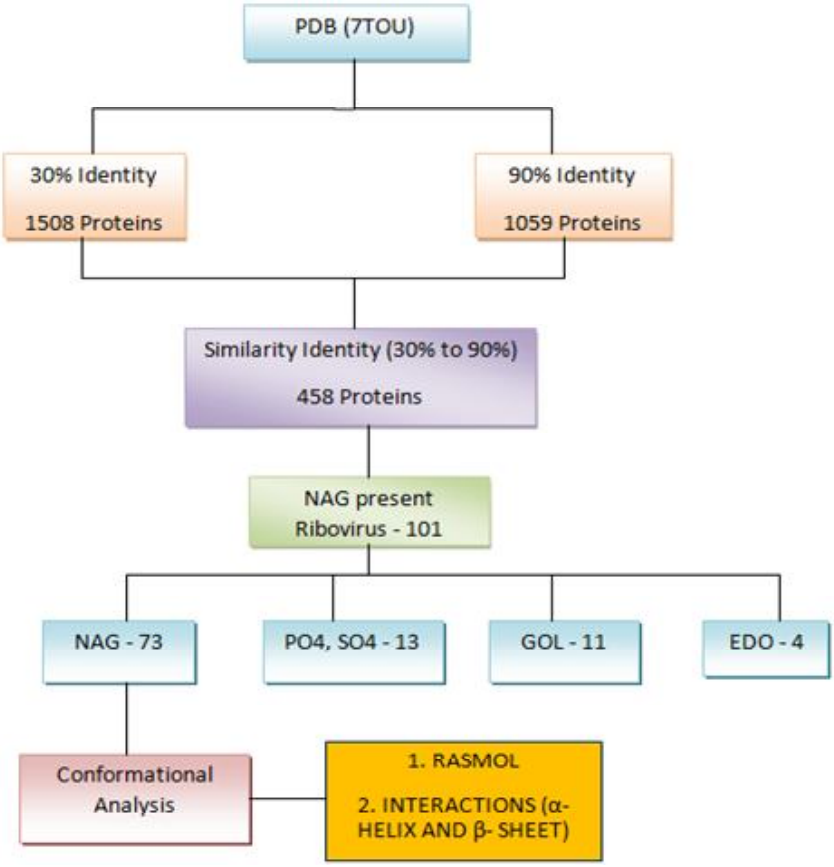

Figure 1. Flow Chart

Protein data bank :

For 26 years, the PDB has provided support to a global community of researchers, teachers, and students working in a wide range of scientific professions. Every day, scientists from all across the world add new structures to the PDB. This community's unifying focus is the need for information that can link the three-dimensional structures of macromolecules to their biological functions. There are various instances where the PDB is playing an increasingly significant role in molecular biology, medicine, and drug development.[6]

Through modifications to the receptor binding domain (RBD) and the presentation of neutralizing antibody peptides, spike conformation plays a crucial part in SARS-CoV-2 development by influencing the virus's transmissibility and immune evasion. For my project study, I used the protein structure database of the risky Delta (B.1.617.2) SARS-CoV-2 variant spike protein.

Using the PDB website I collected similar proteins 30% of identity proteins are 1508 and 90% of identity proteins are 1059 downloaded from PDB. Below 30% of identities and above 90% of identities bias the results. So 30% to 90% ranges of similar proteins are taken for my analysis. These 30% to 90% of identity proteins are non-homologous proteins. Only 458 proteins are available in the 30% to 90% range. From the above database, I have taken the ligand (NAG) 3 proteins structural database for our further study.

Python programming:

Python is an object-oriented, interpretive programming language. By simply copying the program's source code, users may execute Python-written programs on a variety of operating systems, including Microsoft Windows, Linux, and Unix systems like Mac OS X, with almost full support for both the built-in and third-party libraries. All the computations were done using our own Python programming[3].

Rasmol :

The most widely used graphical tool for the presentation of macromolecules (such as proteins and nucleic acid structures) and tiny molecules is called RasMol, and it is accessed every day by many individuals all over the world. The Protein Data Bank (PDB), run by the Research Collaboratory for Structural Bioinformatics, uses it frequently to help process macromolecules[1].

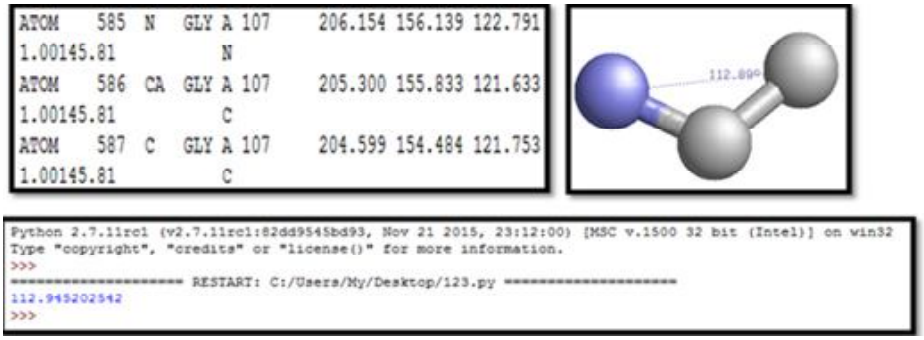

Figure 2. Results from Python and Rasmol

Implementing the Python computer language to determine the angle of the bonds and length for the atoms of various protein entities for conformational analysis.

Identity similarities of 30% to 90% are examined. Then, for more than 90% of identical mutations, it may differ, and for less than 30%, it may differ. There may be a combination of homologous and no homologous proteins in proteins with a similarity of 30% to 90%. Although several proteins interact with amino acids, the majority of them are built on the NAG spike protein.

The sequence of amino acid pairings for the protein 7TOU is depicted in this figure (3). The principal peaks and greatest peaks are built (Q) between glutamine (G) and glycine (G). It is possible to build up to six pairs. There are four possible combinations of the amino acid pairs isoleucine (I) and serine (S). The pair of cysteine and cysteine has a wide range. The possible values of Serine (S) and Histidine (H) are four. There is a three-part coupling between phenylalanine (F) and threonine (T). This graph illustrates other interactions, such as hydrogen bonds, side chains, main chains, and interactions between the α-helix, β -sheet, and other molecules.

The interactions between the α -helixes of 7TOU are represented by the triplet code shown in this graph. The lowest peak arises between Serine (S) and Arginine (R), with additional peaks developing between them, due to steric hindrance caused by its side chain. The side group of the other amino acids is oriented downward and outward concerning the helix. Globular proteins

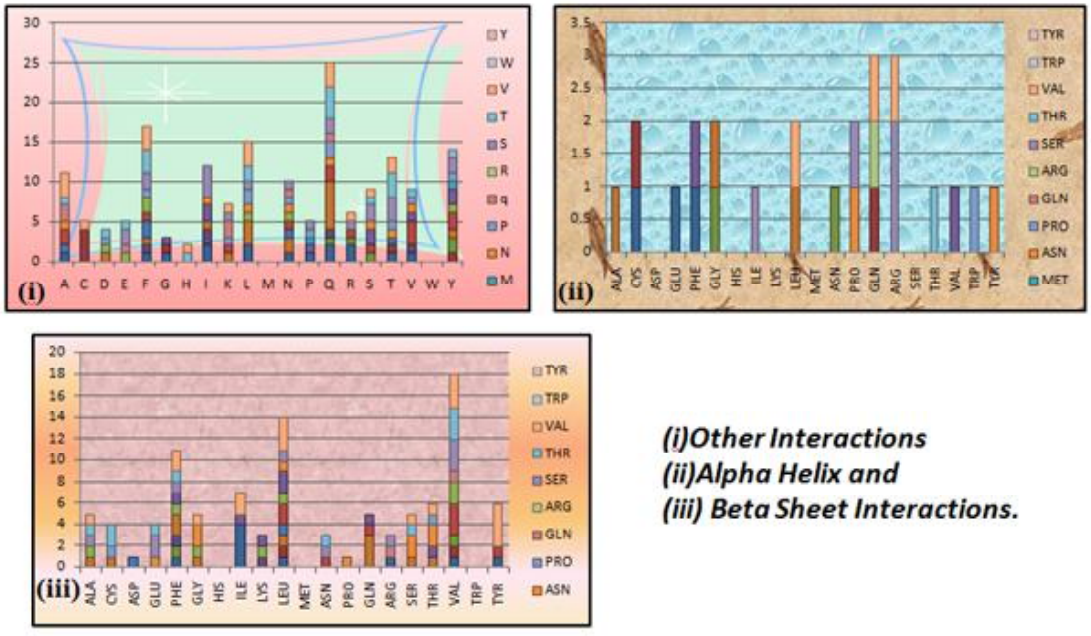

Figure 3. Result Analysis of SARS COVID

frequently have α -helix on their surface. Alanine (A) and cysteine (C) both have variations of one. Serine and isoleucine have been coupled with valine and leucine, respectively.

The triplet code sequence of the 7TOU β -sheet interactions is shown. The principal peaks and the highest mountain (I) are created between Alanine (A) and Isoleucine. Up to four may be present. There are four possible tyrosine and valine pair numbers. There are three different combinations of the amino acids glycine (G) and glutamine (Q). It has proven possible to couple serine (S) and valine (V) three times in its entirety.

Tyrosine is a long way from the NAG ligand and is the target location for drug design. It is common practice to often alter the spike protein of a certain SARS Covid or any other virus to generate drugs and a vaccine.

Hydrogen connections between the residues on two separate strands form a pattern in sheets, which is what gives them their characteristic structure. Both globular and fibrous proteins have an α-helix and a pleated sheet. As a result, α- helices, which make up about 30% of the typical globular protein's structure, are the secondary structure that occurs the most frequently. To visualize the relationships between α- helix and β -sheet structures, a graph should be drawn between the singlet code and the amino acid sequential pair. Protein surfaces tend to include globular proteins' helix structures. The NAG-Ligand that is shown and obtained from short-range interaction. This approach might be improved for use in molecular modeling and drug development.

References :

1. Bernstein, F.C. *et al*., (1977) *J. Mol. Biol. 112:* 535–542. Also published in *Eur. J. Biochem. 80:* 319–324, *Arch. Biochem. Biophys. 185:* 584–591.
2. Chen, Peng; Hu, ShanShan; Zhang, Jun; Gao, Xin; Li, Jinyan; Xia, Junfeng; Wang, Bing (2015). *A* sequence-based dynamic ensemble learning system for protein ligand-binding site prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 1–1.
3. Mészárosová,e., (2015). *International Journal of Information and Communication Technologies in Education, 4*(2): 5–14.
4. Rittle J, MJ Field, MT Green, and FA. Tezcan (2019). *Nat Chem. May; 11*(5): 434-441.
5. R. Lu, X. Zhao, J. Li, *et al*., (2020) *Lancet 395:* 565-574.
6. Sussman, J. L., D Lin, J Jiang, N.O Manning, J Prilusky, O Ritter, E. E Abola, (1998). *Acta Crystallographica Section D Biological Crystallography, 54*(6): 1078–1084.