International Journal of Education and Psychological Research (IJEPR) Volume 10, Issue 3, September 2021
Print - ISSN: 2349 - 0853, e-ISSN: 2279 - 0179 Available online at www.ijepr.org
Orignal Research Article Open Access

Molecular Docking of Spike Protein of COVID-19 Using Selective Ligands

Mayana Ibtisam Khan [1]

Abstract

The latest coronavirus outbreak poses an unprecedented threat to health of human beings and society around the world. The development of appropriate therapies is critical in this situation. The viral spike protein and the cognate host cell receptor ACE2 can be used as therapeutic targets. The current computational study shows that fisetin not only has a high affinity for the viral s protein and the host receptor ACE2, but also for their complex RBD of the spike protein of SARS-CoV-2 and ACE2; RBD/ACE2-complex. AutoDock Vina was used for docking fisetin with ACE2 receptor and S protein of SARS-CoV-2. The binding affinities values are -2.56 and -1.9 kcal/mol for the S protein and ACE2, respectively. Fisetin binds directly to the viral S protein RBD. Fisetin binds to the interface of the RBD/ACE2-complex, causing the alpha helices and beta strands of the protein complex to fluctuate. The results of the docking showed that the binding pocket involves the amino acid residues Thr24, Thr25, Thr26, Leu27, His41, Thr45, Ser46, Met49, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, Met165, Glu143, Ser144, Cys145, His163, Met165, Glu166, His172, Val3, Leu4, Leu167,Pro168,Ala2. Protein- protein interactions in the presence of fisetin corroborate the above findings, indicating that this flavonoid is effective in preventing the formation of the S protein-ACE2 complex. Finally, for the first time, this computational study predicts the possibility of flavonoid as a therapeutic strategy against SARS-CoV-2.

Keywords: *Fisetin, binding energy, ACE2 receptor, S protein, molecular docking.*

Article History: *Submitted: 20th August, 2021, Revised: 04th September, 2021, Accepted: 15th September, 2021, Published: 30th September, 2021.*

I. INTRODUCTION

SARS-CoV-2, an enveloped single-stranded positive sense RNA virus from the coronaviridae family, is a global concern[1, 2]. The world health organization declared the outbreak a pandemic due to its high rate of transmission, rapid spread across several countries, and the lack of specific vaccines or medications to treat it[3] . SARS-CoV-2 is phylogenetically related to the order Nidovirales 4 and is classified as a Beta coronavirus, with a genome size of 30 kb that codes for various structural and accessory proteins [4,5]. Coronaviruses have a variety of structural proteins, including the Spike protein (S), Envelope protein (E), Membrane protein (M) and Nucleocapsid protein (N)[6]. Coronavirus enters human cells by binding its distinct surface spike protein to a receptor protein on the cell membrane. The spike protein is a transmembrane protein with – exo and C-endo terminals. The N S1 subunit contains RBD, whereas C S2 subunit promotes membrane fusion[7]. The S1 of viral protein binds to S to human cell receptor results in the virus- human cell fusion [7, 8]. However, following virus endocytosis, the S2 subunit, which is distinguished by Heptad repeats regions that assemble into an intra-hairpin helical structure with a six-helical bundle, promotes membrane fusion within the host cell [9, 10] . Recently, it was discovered that the RBD of SARS-CoV-2 S protein is more or less similar to that of SARS-CoV, despite amino acid differences at some key residues [2]. This suggests that the virus can also target ACE2; a human cell monomeric membrane bound protein [11, 12] . As a result, it is assumed that ACE2, the coronavirus cognate receptor present on the cell membrane of host cells, can also be specific target to prevent viral entry[13]. Computational approaches (Molecular docking and stimulation) are the first and foremost choice of scientists to predict apparent binding modesand affinities of ligands for macromolecules prior to expensive and time-consuming experimental studies [23]. The current study incorporates the findings of molecular docking of fisetin with the S protein of COVID-19 as well as ACE2 of host cells, a cognate receptor for viral S protein. Furthermore, binding affinities of fisetin with the 'RBD/ACE2-complex' reveal that flavonoids significantly alter the complex's structure [24] .

II. MATERIALS AND METHODS

Sequence of spike protein was retrieved from UniProtKB/Swiss-Prot which is freely accessible public database that is manually annotated and mentioned in Table 1. It can be accessed directly from the URL http://www.uniprot.org/. Protein sequence was retrieved in FASTA format which was further used for docking studies.

Table 1. Protein sequence used in the present study

Protein	Unique Accession No.	Length	Gene	Function	Description
S protein	P0DTC2	1273	S	Attaches the virion to the cell membrane by interacting with host receptor	A glycoprotein that protrudes from the envelope of some viruses and facilitates entry of the virion

[1]Student, Department of biotechnology and bio informatics, Yogi Vemana University, Kadapa, Email: mayanaibtisam22@gmail.com

2.1 Physico-chemical characterization:

Computational analysis for the study of the physico-chemical parameters of the S protein was performed using the EXPASY World Wide Web server. It has several online tools that predicts certain properties of proteins like pI, M.wt, peptide mass, amino acid composition, EC, atomic composition, II, GRAVY, +R and –R etc. FASTA format of the protein sequence were used for the physico-chemical analysis using ProtParam tool which is mentioned in Table 2.

Table 2. Parameters of protein computed using ProtParam tool

S.No	Physico-chemical properties	Results
1	Number of amino acid residues	1273
2	Molecular weight	141178.47
3	Theoritical pI	6.24
4	Total number of positively charged residues	105
5	Total number of negatively charged residues	110
6	Formula	$C_{6636}H_{9770}N_{1656}O_{1894}S_{54}$
7	Total number of atoms	19710
8	Instability Index	33.10
9	Aliphatic Index	84.67
10	Grand Average of Hydropathy	-0.079183

2.2 Functional characterization

SOSUI server was used to classify and predicts the secondary structure of an S protein. SOSUI searches for a (alpha) helix and predicts the transmembrane region of the protein. SOSUI signal was used to identify the transmembrane protein, and the predicted transmembrane region of the protein is represented in Table 3 (Figure 1). For the determination of functional linkages, disulphide bonds are important [13] . Tool DiANNA 1.1 (http://clavius.bc.edu/~clotelab/DiANNA/) (DiAmino acid Neural Network Application) web server was used to identify the disulphide bond in a protein sequence data. DiANNA is a web server for the prediction of cysteine classification and disulphide connectivity [14] . Hydrophobic residues (LVM) are represented as a blue square, AGPY and HKR in black octagons.

S.No	N-terminal	Transmembrane region	C-terminal	Type	Length
1	1	MFVFLVLLPLVSSQCVNLTTRT	22	secondary	22
2	1223	GLIAIVMVTIMLCCMTSCCSCLK	1245	primary	23

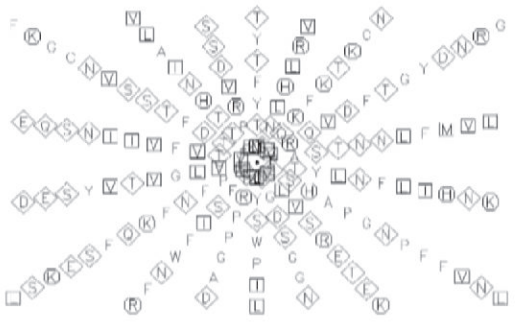

Figure 1. Helical wheel representation of predicted transmembrane range

2.3 Prediction of secondary structure of protein:

The secondary structure of protein was predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi- bin/npsa_automat.pl?page=npsa_sopma.html) that uses self-optimized prediction method (SOPM) and neural network method 16 .

Table 4. Calculated secondary structure by SOPMA

Alpha helix (Hh)	364(28.59%)
3_{10} helix (Gg)	0(0.00%)
Pi helix (Ii)	0(0.00%)
Beta bridge (Bb)	0(0.00%)
Extended strand (Ee)	296(23.25%)
Beta turn (Tt)	43(3.38%)
Bend region (Ss)	0(0.00%)
Random coil (Cc)	570(44.78%)
Ambiguous states (?)	0(0.00%)
Other states	0(0.00%)

2.4 Phylogenetic analysis:

Multiple sequence alignment for individual profiles was performed using phylogenetic analysis using MEGA X software 21 . The discovered motifs were further used to search their protein family using Pfam at the DDBJ MOTIF server (http://www.genome.jp/tools/motif/).

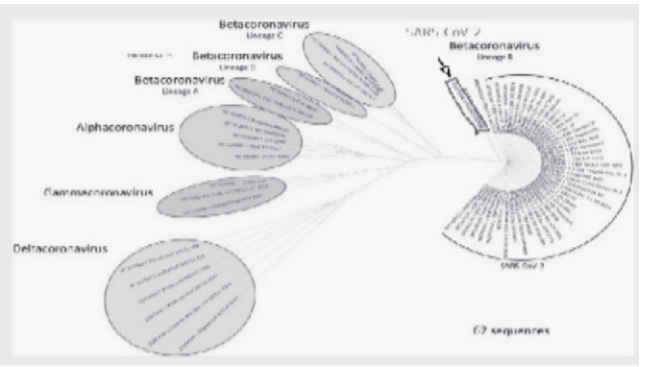

Figure 2. SARS-CoV-2 and SARS-CoV were evaluated by MEGA X software.

It was observed through structural alignment studies that SARS-CoV-2 belongs to beta lineage of SARS-CoV-2. It was apparent that SARS-CoV is an ancestor of the newly upsurging virus SARS-CoV-2.

2.5 Network interaction:

STRING [25] was used to identify protein-protein interaction. At present, string database covers up to 5,214,234 proteins from 1133 organisms.

2.6 Determination of active site analysis:

The 3D structure of S protein, the probable binding sites of the protein was searched based on the structural association of template and the model construct with CASTp [26] server. CASTp was used to recognize and determine the binding sites, surface structural pockets, active sites, area, shape and volume of every pocket and internal cavities of protein. It could also be used to calculate the number, boundary of mouth openings of every pocket, molecular reachable surface and area [27] . Active sites analysis provide a significant insight of the docking study.

2.7 Docking protocol:

Molecular docking protocol is widely used for predicting the binding affinities for a number of ligands. In current work, our aim was to examine the possibility of an existing relationship between the experimental bioactivities of the inhibitors under study and the docking scores. In order to get accurate results, all the docking experiments were performed with the default parameters. The time to dock one compound was approx. 10-20 mins. Docking with AutoDock/Vina was performed on a Linux workstation (openSUSE11.4) with an Intel Pentium D Processor (3.0 GHz) and 1 GB of RAM whereas FlexX was run on windows 7 equipped with an Intel ® Atom ™ processor (1.67 GHz) and 1 GB RAM.

III. RESULTS

Detailed information about the S protein of COVID-19 was discussed in Table 1. Physico- chemical parameters were calculated using ProtParam analysis, a tool of EXPASY server and data was represented in Table 2. pI was calculated which is an important parameter at which the protein is stable and compact, means with least solubility and zero mobility in an isoelectric focusing system. The evaluated pI value of S protein was acidic in nature. The wavelength of 280nm is most favored for protein because it shows absorbance maxima at this wavelength. The results are basic because of the absorbance of cysteine (Cys)[21] . The extinction coefficients for the S protein were ranging from 146460 to 148960M -1 cm -1 in respect to Cys. This evaluation of extinction coefficients benefit in the study of protein-ligand and protein-protein interaction[22]. In this method the occurrence of certain di-peptides is unstable which is compared to stable ones, and provides the weight value of instability. By using weight value, it becomes possible to estimate instability index. Protein with instability index value less than 40 is said to be stable and those with more than 40 are supposed to be unstable. The instability index value is 32.99 for the S protein of COVID-19. This analysis concluded that S protein is a stable protein. Aliphatic index suggests the thermostability of a protein [23] . It is well defined as the relative volume of a protein occupied by aliphatic side chains. The value of aliphatic index for the S protein of COVID-19 varies from the range of 84.61. S protein shows very high aliphatic index that means they are stable at a very wide range of temperature. When the summation of hydropathy value is divided by the total number of residues in a protein sequence, it provides the GRAVY value [12] . Least GRAVY index of about -0.079183 suggests that the protein may have better interaction with the water molecules. The detailed information about the transmembrane region, length and types was tabulated in Table 2. To determine the functional linkages, disulphide bonds are important because the thermostability of a protein depends on it. Amino acid residues were involved in the secondary structure mainly three states- alpha helix (28.59%), extended strands (23.25%) and random coils (44.78%). Random coils of S protein show dominating role towards the extended strand and alpha helix whereas in 6LU7 alpha helix dominates the secondary structure.

Main clusters are highlighted in different colors. The filled circles represent the main supported clusters bootstrap support values are indicated at the level of nodes. Bootstrap analysis was calculated and analysis values were shown next to the branch nodes. The red arrow shows SARS-CoV-2 reference sequence. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The network analyzer tool in STRING database that focus on gathering and combining the data, by merging the known and predicted protein-protein association information for a huge number of micro-organisms [24] . Since the specific protein which relates to specific disease is chosen and gathered and stored in the STRING database. In STRING database updates its periodically [25] . The power of protein-protein interactions can only set for the network construction [26]. SARS-CoV-2 was retrieved from its crystal structure joint together to form spike protein, and ligands selected based on action against ligands. The default probe radius is set to be constant at 1.4 A°

.Figure 3. Determination of active sites present in Fisetin molecule

In order to recognize an accurate docking routine for carrying out molecular docking studies of ASMT protein as well as to identify the potent inhibitors against the protein, four widely used docking routines (AutoDock/Vina) are compared in this work [28] . Each docking routine returned top ten ranked docked poses for each ligand. Among all the studied docking routines, AutoDock/Vina was found to be the best for carrying out blind docking and in generating poses that bind best deep inside the 5 A o of the binding pocket. It generated 50-70% good poses.

Overall results showed relatively poor performance ranging from 30 to 70%. These percentages were surprisingly low, indicating the docking programs often failed to find the correct binding mode [29]. The average time required for carrying out the docking calculation using different scoring functions and docking routines of a single ligand for AutoDock/Vina with an average time of 20-30 mins and approx. 1 hr respectively. Interestingly, important interactions can be found between these atoms and the residues Thr24, Thr25, Thr26, Leu27, His41, Thr45, Ser46, Met49, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, Met165, Glu143, Ser144, Cys145, His163, Met165, Glu166, His172, Val3, Leu4, Leu167, Pro168, and Ala2 which directly participate in the catalytic mechanism of this enzyme. The ligand-enzyme complex is stabilized mainly by hydrogen bonds and hydrophobic interactions. The top-ranked pose with lowest docked binding affinities and high docking scores are generally used as a standard selection in most of the docking programs [30] . Different sets of hydrogen bonding interactions with polar side chain residues of

Arg280, Ser104, Thr100 and Tyr108 and with phenol of Tyr327 are observed at distances within 4A o . An ionic interaction with the side chain residue of Arg280 is also observed. The indole ring and other side chain carbons formed strong hydrophobic interactions with non-polar residues Leu160, Tyr327, Tyr108, Leu326, Trp117 and Phe26 are suggested to increase binding affinity. Hydrogen bonds with a backbone and side chain residues of Arg210, Lys223 and Thr207, backbone residues of Val21, Phe212 and Glu224 are observed. The strong hydrophobic interactions were Thr195, Lys223, Leu228 and Arg169, pyrrole ring of His209, Val211 and aromatic ring of Phe212. Ionic interactions were Arg210 and Glu224. The strong hydrogen bonding interactions are done with phenolic, amide and side chain residues which occur by electronegative ligands. It forms hydrogen bonds, hydrophobic and ionic interactions with the important residues of the binding pocket of ASMT thus stabilizing the structure of target receptor. The dock poses with least binding energy has the highest affinity and hence is the best docked confirmation 31 .

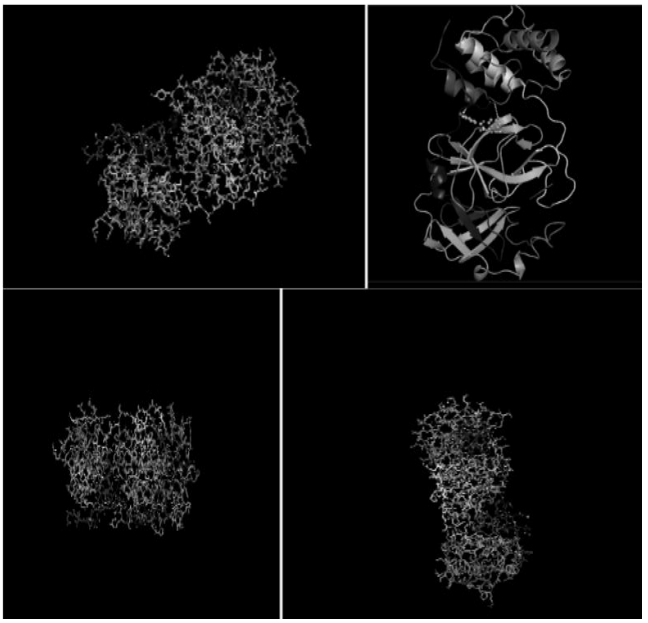

Figure 4. Best docked poses for fisetin molecule

IV. CONCLUSION

Docking and scoring have evolved significantly over the past years. It has become valuable tool in drug discovery process. Our goal of this study was to explore the feasibility of AutoDock/Vina docking approaches, for our target ASMT and to find out the lead compound. We compared the predictive power of each docking and scoring function. Our results suggest that all docking programs studied here to do a reasonable job in docking and should aid significantly the drug discovery process. However, Auto Dock/Vina consistently outperformed and was found to be relatively more useful in blind docking pose prediction. Moreover, analysis of the docked ligands with the protein brought into focus some important interactions operating at the molecular level.

The results of the ligand docking showed that the binding pocket involves the amino acid residues Thr24, Thr25, Thr26, Leu27, His41, Thr45, Ser46, Met49, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, Met165, Glu143, Ser144, Cys145, His163, Met165, Glu166, His172, Val3, Leu4, Leu167, Pro168, and Ala2. The important hydrogen bond forming amino acid residues was Arg280, Thr100, Tyr108, Asn330, Trp117 and Tyr327. In conclusion, we have discovered a highly potent lead compound which will be useful for the design of novel less toxic and highly efficient drug for the treatment of COVID-19.

V. REFERENCES

1. Abdellah El Aissouq, Oussama chedadi, Mohammed Bouachrine and Abdelkrim Ouammou, (2021). Identification of novel SARS-CoV-2 inhibitors: A structure-based virtual screening approach, Journal of chemistry, Vol 2020, pages 7.
2. Abimeal Gonzalez- Hernandez, Jair Lozano-Cuenca, Bruno A. Marichal-Cancino, Antoinette Maassen VanDenBrink, and Carlos M.Villalon,(2018). Dihydroergotamine inhibits the vasodepressor sensory cgrypergic outflow by prejunctional activation of 2-adrenoceptors and 5-ht 1 receptors, Journal of headache pain, Vol 19, pages 40.
3. Aditya ganeshpurkar, and Ajay K. Saluja,(2017). The pharmacological potential of rutin, Saudi Pharmaceutical Journal, Vol 25, pages 149-164.
4. Alex Bateman, Maria-Jesus Martin, Sandra orchard, Michele Magrane, (2021). Uniprot: the universal protein knowledgebase in 2021, Nucleic acids research, Vol. 49, pages 480-487.
5. Alex bateman, Niko Beerenwinkel, Ceciua arigui, Alexandre Almeida, (2021). Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and corona virus research, Briefings in Bioinformatics, Vol 22, Issue 2, pages 642-643.
6. Amit lakhani, Ena Sharma, Kirti gupta, Savita kapila and Shivangi gupta,(2020). Corona virus (COVID-19) and its impact on health care workers, Journal of the association of physicians of India, Vol 68.
7. Anshumali mittal, Kavyasree manjunath, Rajesh kumar Ranjan, Sandeep Kaushik, Sujeet kumar, and Vikash Verma, (2020). COVID-19 pandemic: Insights into structure, function and hACE2 receptor recognition by SARS-CoV-2, plos pathogens.
8. Atala B.Jena, Namrata Kanungo, Vinayak Nayak, G.B.N. Chainy, Jagneshwar Dandapat, (2021). Catechin and Curcumin interact with S protein of SARS-CoV-2 and ACE2 of human cell membrane: insights from computational studies, Scientific reports, Vol 11, Article Number: 2043.
9. Barabasi BA-L and Bonabeau E,(2003). Scale-free networks,Scientific American, Vol 288, pages 9-50.
10. Ben Hu, Hua Guo, Peng Zhou and Zheng-Li shi, (2021). Characteristics of SARS-CoV-2 and COVID-19, Nature reviews, Vol 19, pages 141-154.
11. Bindea G., Galon J. and Mlecnik B, (2013). CluePedia cytoscape plugin: pathway insights using integrated experimental and in silico data, Bioinformatics, Vol 29, pages 3- 661.

12. Brown J., Pirrung M. and McCue L.A., (2017). FQC dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool, Bioinformatics, Vol 33, pages 3137-3139.
13. Buddhisha Udugama, Pranav kadhiresan, Hannah N. Kozlowski, Ayden Malekjahani, Matthew Osborne, Vanessa Y.C.Li, Hongmin chen,(2020). Diagnosing COVID-19:The disease and tools for detection, ACS publications, Vol 14, pages 3822-3835.
14. Cascella M, Rajnik M, Abdul Aleem, Scott C. Dulebohn, and Raffaela Di Napoli, (2021). Features,evaluation and treatment of corona virus (COVID-19), StatPearls Publishing.
15. Channing J.Paller and Emmaneul S. Antonarakis,(2011). Cabazitaxel: a novel second- line treatment for metastatic castration resistant prostate cancer, Drug Design, Development and Therapy, Vol 5.
16. Chittaranjan Baruah, Papari Devi and Dhirendra K.Sharma, (2020). Sequence analysis and structure prediction of SARS-CoV-2 accessory proteins 9b and ORF14: Evolutionary analysis indicates close relatedness to bat corona virus, Biomedical research international, Vol 2020, pages 13.
17. Claudia dimatteo, Giovanna D'Andrea, Gennaro Vecchione and Oriana Paoletti et al., (2016). Pharmacogenetics of dabigatran etexilate interindividual variability, Thromb research, Vol 144, pages 1-5.
18. Corona virus disease (COVID-19) advice for the public, (2021). WHO.
19. Corona virus, Wikipedia.
20. Damian Szlarczyk, John H Morris, Helen cook and Michael Kuhn et al.,(2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic acids research, Vol 45, pages 362-368.
21. David E.Gordon, Gwen down M.Jang, Nevan J.Krogan, (2020). A SARS-CoV-2 protein interaction map reveals target for drug repurposing, Nature publishings, Vol. 583, pages 459-468.
22. Edison Ong, Mei U Wong, Anthony Huffman and Yongqun He, (2020). COVID-19 corona virus vaccine design using reverse vaccinology and Machine learning, Frontiers in immunology, Vol 11, Article number: 1581.
23. Elahe seyed Hosseini, Narjes Riahi Kashani and Hamed Haddad Kashani,(2020). The novel corona virus disease-2019 (COVID-19): mechanism of action, detection and recent therapeutic strategies, Virology, Vol 551, pages 1-9.
24. Elijah Kolawole Oladipo, Ayodeji Folorunsho Ajayi, Olumuyiwa Elijah Ariyo, Samson Olugbenga Onile, (2020). Exploration of surface glycoprotein to design multi-epitope vaccine for the prevention of COVID-19, Informatics in Medicine Unlocked, Vol 21.
25. F.Ferre and P.Clote,(2005). DiANNA: a web server for disulfide connectivity prediction, Nucleic Acids Research, Vol 33, pages 230-232.
26. Garrett M.Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K.Belew (2009). AutoDock4 and AutoDock Tools 4: Automated Docking with selective receptor flexibility, Journal of computational chemistry, Vol 30, No. 16, pages 2787-2791.
27. Gasteiger E., Gattiker A., Hoogland C., and Ivanyi I. et al., (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids research, Vol 31, pages 3784-3788.
28. Georgina Cox, Linda Ejim, J Stoigos peter and Kalinka koteva et al.,(2018). Plazomicin retains antibiotic activity against most aminoglycoside modifying enzymes, ACS Infectious Diseases, Vol 4, pages 980-987.
29. Gustavo R Villas-Boas, Vanessa C Rescia, Marina M Paes, and Stefania N Lavorato et al., (2020). The new corona virus (SARS-CoV-2): A Comprehensive review on immunity and the application of bioinformatics and molecular modeling to the discovery of potential Anti-SARS-CoV-2 agents, Molecules, Vol 25, pages 4086.
30. Martin M. (2011). Cutadpt removes adapter sequences from high-throughput sequencing reads, European Molecular biology network, Vol 17, pages 10-12.
31. Seemann T.,(2014).Prokka: rapid prokaryotic genome annotation, Bioinformatics, Vol 30, pages 2068-2069.