SARS-CoV-2 Pocketome: Severe acute respiratory syndrome coronavirus 2, pockets identification for antiviral & antimicrobial phytomolecules and drug repurposing.

Akhil Kumar [a], Ajit K Shasany* [a]

[a] Plant Biotechnology Division, CSIR-Central Institute of Medicinal and Aromatic Plants, P.O. CIMAP, Lucknow – 226015, India

* Corresponding author: email address –ak.shasany@cimap.res.in (Ajit Kumar Shasany)

Abstract

According to WHO, the current Coronavirus disease situation is 8,506,107 confirmed and 455,231 death cases in approx 216 Countries, areas, or territories. For the treatment of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) drug repurposing seems to be an effective strategy as it could shorten the time and reduce the cost compared to de novo drug discovery. For that, we need to identify target binding sites. Thus, we have reported for the first time structural druggability assessment for SARS-CoV-2 proteome a pan-druggability prediction based on the open-source pocket detection code fpocket and rank them on the basis of druggability score. We have identified in a total of 433 pockets on the SARS-CoV-2 proteome and characterized by physicochemical descriptors. In which, 47 pockets identified as druggable and 71 as potential drug-binding pockets. Further, Docking of antiviral & antimicrobial phytomolecules against druggable pockets identifies potential SARS-CoV-2 inhibitors eg. Theaflavin, Baicalin, Menthol, Eugenol, Catechin.

Keywords: **SARS-CoV-2 Pocketome, drug repurposing, Virtual screening, Target-based drug discovery, druggability**

1. Introduction

Corona Virus Disease 2019 (COVID-19) has become a global pandemic. This COVID-19 disease caused by a novel new member of the betacorona virus genus and is closely related to severe acute respiratory syndrome coronavirus (SARS-CoV) [1–3]. This is a positive-sense, single-stranded RNA, named as a SARS-CoV-2 [4]. People get easily infected with this virus in close contact with infected person through droplet infection produced from coughs or sneezes that resulted in millions of people demise around the world. So there is an urgent need to identify molecules that inhibit or kill the SARS-CoV-2 via targeting essential target of the virus life cycle.

However, experimental approaches for drug repurposing are costly and time-consuming. Computational approaches such as structure based virtual screening (SBVS) offer novel testable hypotheses for systematic drug repositioning. SBVS, specifically based on the identification of candidate pockets in protein structures [5]. However, limited information is available on SARS-CoV-2 structure and their binding sites and limited structures are available in PDB (SF:1). In this work, a combination of *in silico* tools and an in-house script was primarily used to predict the binding pockets of SARS-CoV-2 proteome and further rank them on the basis of druggability score (DS). Druggability is the probability of small drug-like molecules binding to a given target protein with high affinity. We report for the first time the structural druggability assessment for SARS-CoV-2 proteome a pan-druggability prediction based on the open source pocket detection code "fpocket"[6]. This method

calculate several physicochemical descriptors like hydrophobicity score (HS), polarity score(PS), total, polar and apolar solvent accessible surface area (SASA) to characterize each pocket. These physicochemical properties will guide more focused screening against the SARS-CoV-2. SARS-CoV-2 Pocketome (A pocketome representing the entire druggable/binding sites of an organism) can be efficiently utilized for drug repurposing and inhibitor screening to accelerate drug design processes. To show the applicability of this work in SARS-CoV-2 inhibitors identification, we have docked known antiviral phytomolecuels to these 118 pockets.

2. Materials and Methods

2.1. Modelling of SARS-CoV-2 proteome

SARS-CoV-2 proteome was modelled at SWISS-MODEL based on the NCBI reference sequence NC_045512 as SARS-CoV-2 proteome as repository (https://swissmodel.expasy.org/repository/species/2697049) [7]. In this study, SARS-CoV-2 proteome divided into two part, i) Model that have generated from experimental structure of SARS-CoV-2 present in PDB and ii) part in which template structure were not from the SARS-CoV-2, model was generated from remote homologue.

2.2. Pocket prediction of SARS-CoV-2 proteome

Fpocket is an open source pocket detection package based on Voronoi tessellation and alpha spheres [6]. As an output fpocket will return all atoms used for pocket detection from the input PDB file. A python script was written to retrieve the 3D coordinate of amino acid from the atom file and write into pdb format with their pocket number. Physicochemical properties for each pocket plotted as a bar plot to compare and understand the nature of the pockets.

2.3. Docking of antiviral drug and natural antiviral molecules

We have selected phytomolecules reported against influenza virus and other essential oil individual component (ST:1) from selected Umbelliferae and Labiatae plants that have antimicrobial and antiviral activity. These molecules were docked against 118 pockets that have been selected on the basis of DS (≥ 0.1) and Volume ($\geq 400\text{Å}$) with the help of Autodock4.2. We have considered pockets with volume $\geq 400\text{Å}$ because this has been suggested that size of drug molecules shows that the median number of ''heavy'' (i.e., nonhydrogen) atoms is 24 with a typical envelope of around 400 Å [8]. Grid was place at the centre of the each pocket and the grid size was set to 60ÅX60ÅX60Å. Other parameters were set to default.

3. Result & Discussion

3.1. SARS-CoV-2 proteome and pockets

This study emphasize on comprehensive analysis of the pocketome of SARS-CoV-2. The properties like DS, HS, PS, volume and SASA of identified pockets indicate that particular pocket able to accommodate what kind of a ligand i.e. size of a ligand, polarity, charge and hydrophobicity. Pockets with good druggability are identified and selected for docking with natural antiviral and antimicrobial activity.

3.2. Papain-like proteinase (PL-PRO) or Nsp3

PL-PRO is responsible for the cleavages of N-terminus of the replicasepolyprotein to release Nsp1, Nsp2 and Nsp3, which is essential for correcting virus replication [9]. In total we have identified 25 pockets (SF:2a) in Nsp3. Among these, only pocket 1 have DS of 0.201 and rest showed less than <0.06 score.

Pocket 1 has largest volume (1146.59Å) and highest PS with charge score of – 2. (SF:2b) Polar SASA of pocket 1 (113.16Å) is greater than the Apolar sasa (59.17Å) indicates the polar nature of the pocket 1. The amino acid residues in Nsp3 pocket 1 are ALA991, ASN1012, ASP909, ASP1047, GLN1014, GLU906, GLU912, GLY908, GLY1011, LEU907, LYS902, MET953, PRO992, PRO993, SER990, THR1046, TYR1009, TYR1013 and TYR1018. Four pockets have volume ≥ 400Å, pocket 3, pocket 7, pocket 9 and pocket 21.

3.3. 3C-like proteinase (3CL-PRO) or M^{pro}

3C-like proteinase also called a viral main protease (M^{pro})/ Nsp 5. The M^{pro}, is first automatically cleaved from poly-proteins to produce mature enzymes, then further cleaved downstream Nsps at 11 sites to release Nsp4–Nsp16[10]. M^{pro} exists as monomer and homodimer but only the homodimer shows catalytic activity and essential in the life cycle of the virus. M^{pro} predicted to have 13 pockets (Fig:1). Pocket 2, predicted to has the highest DS of 0.187, highest PS 12 and volume 1094.7Å. Other pockets have DS less than 0.01(SF:2a;2b). Pocket 2, of M^{pro} is polar in nature and 2 charge score similar to papain pocket 1. The amino acid residue involve in M^{pro} Pocket 2 are ALA255, ALA260, ASP263, GLY251, LEU262, SER 254, THR 224 , VAL247, VAL261 . In this case Fpocket correctly predicted the ligand binding pocket e.g. when we aligned pocket 2 on 6lu7, bound ligand N3, sits in pocket 2. Pocket7 is the only pocket has volume greater than 400Å.


SARS-CoV-2 Mpro Pockets

Druggable Pocket

Mpro Pocket2
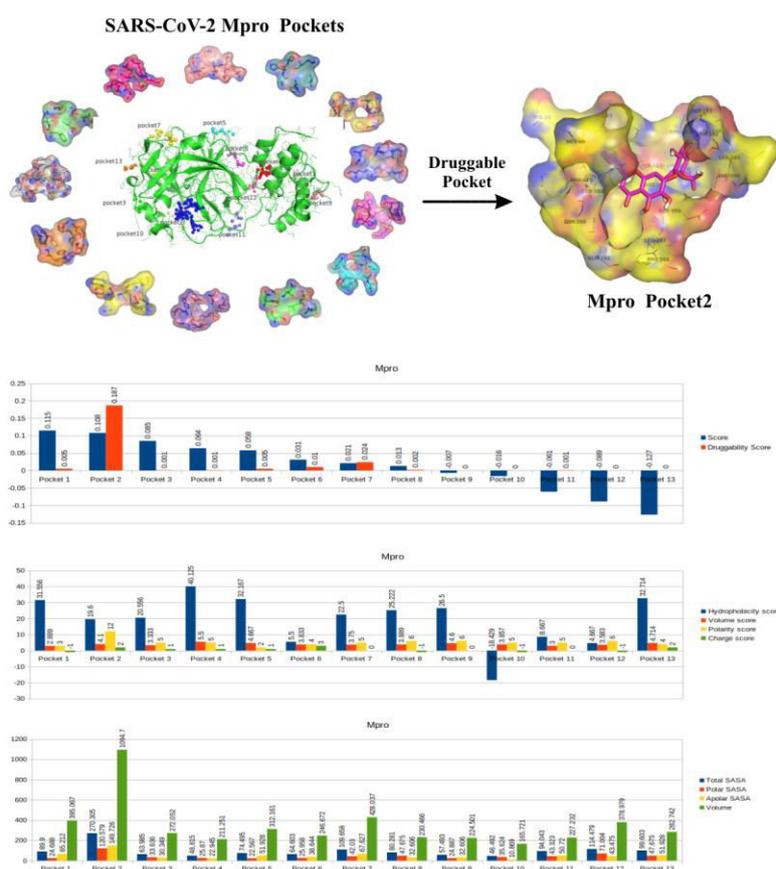
Mpro

Mpro

Mpro

Fig:1 Identified SARS-CoV-2 M^{pro} pockets visualization, bar chart showed properties for each pocket and one of the druggable M^{pro} pocket, pocket 2 interaction with theaflavin (ball & stick).

3.4. Non-structural protein 7 (Nsp7)

Nsp7 interact with Nsp8 and may participate in viral replication by acting as a primase and might help in activate RNA-synthesizing activity of Pol [11]. Five pockets were identified in Nsp7. Pocket 2 has highest DS (0.632) followed by Pocket 4 (0.178) and pocket 5 (0.109) (SF:3a; 3b). Other pockets have DS less than 0.1. Pocket 2 has volume 243.26Å, apolar sasa 85.74, and polar sasa 24.64, HS of 63 and one negative charge score. Amino acid residues of pocket 2 are ASP67, GLN63, ILE68, LEU56, LEU59, LEU60, LEU71, MET62, and VAL66. Pocket 2 seems to have smaller volume but have largest total SASA make it easily exposed and accessible.

Pocket 4 has highest HS 72.16 but its volume is smaller than pocket 2 and have zero score for charge and polar sasa. Result showed that Pocket 4 have small hydrophobic pocket with no polar residue. Amino acid residues in pocket 4 are THR9, VAL12, LEU13, VAL16, MET52, LEU56. Among druggable pocket 2, 4 and 5, pocket 5 has highest volume (320.39 Å) and residues ALA 30, GLN31, GLN34, LEU35, VAL58 with low hydrophobicity, polar and charge score. Given properties indicated these pockets are smaller in nature and will not able to accommodate larger ligands.

3.5. Non-structural protein 8 (Nsp8)

Similar to Nsp7, Nsp8 forms a hexadecamer with Nsp7 (8 subunits of each) that may participate in viral replication by acting as a primase. Alternatively, might be able tosynthesize substantially longer products than oligonucleotide primers. The Nsp7_Nsp8 complex increases the binding of Nsp12 to RNA and enhances the RdRps enzyme activity of Nsp12 [11].

Four pockets were identified in Nsp8 (SF:4a) out of which, only pocket 1 showed DS of 0.279, have highest HS (53.778) and zero charge score (SF:4b). The residue involve in pocket 1 are ASN140, LEU128, MET129, PHE147, THR141, THR148, TYR149, VAL130, and VAL131. Polar and apolar sasa is equally distributed in pocket 1 and its volume is 303.62Å. These results indicate it can accommodate only small sized hydrophobic ligand.

3.6. Non-structural protein 9 (Nsp9)

Nsp9 may be act as an important target being involved in viral replication as an ssRNA-binding protein. Total 5 pockets were identified in Nsp9, pocket 1 and pocket 2 have high DS of 0.581 and 0.482 respectively (SF:5a). Pocket 1, is largest pocket (volume 587.33Å) with and total SASA 158.37 in which major portion is apolar sasa 111.10 and has HS of 28.73 (SF:5b). Amino acid residues involve in this pocket are ARG39, ASP60, GLY37, GLY38, ILE65, ILE91, LYS58, MET12, PHE40, PHE56, PRO57, SER59, THR64, THR67, VAL41. Pocket 2 is second highest druggable pocket, volume of 546.17Å and has larger hydrophobic area with highest HS of 78.11 and large total sasa (178.34). Pocket 2 is consist of amino acid residues CYS73, GLN113, LEU88, LEU103, LEU106, LEU112, PHE75, PHE90, VAL76.

3.7. Non-structural protein 10 (Nsp10)

Nsp10, plays a pivotal role in viral transcription by stimulating both Nsp14 3'-5' exoribonuclease and Nsp16 2'-O-methyltransferase activities. Therefore plays an essential role in viral mRNAs cap methylation. Nsp10 form hetero-oligomeric complexes with both Nsp14 and Nsp16.

Total 6 pockets were identified in Nsp10 (SF:6a), in which pocket 5 and pocket 1 has DS 0.573 and 0.14 respectively. Pocket 5 has highest volume (528.67Å), total sasa (169.97) and HS (39.33) among other pockets with one positive charge score (SF:6b). Amino Acid in pocket 5 are ALA23, ALA26, ARG78, CYS79, GLN36, ILE38, LEU75, PRO37, THR39,TYR27, TYR30, VAL21.Pocket 1 has small (volume 271.39Å) hydrophobic pocket (HS 37) and zero charge score. Amino acids in this pocket are ASN114, ASP91, ILE55, LEU92, THR111, THR115, TRP123, and VAL116. Pocket 2 is the biggest pocket (volume 656.83Å) in Nsp10 however, the DS(0.002) is very low.

3.8. RNA-directed RNA polymerase (Pol/RdRp)

Nsp12, a conserved protein in coronavirus, is an RNA-dependent RNA polymerase (RdRp) and the vital enzyme of coronavirus replication/transcription complex [11]. In the study of SARS-CoV and MERS-CoV inhibitors, Nsp12-RdRp has been used as a very important drug target. In principle, targeted inhibition of Nsp12-RdRp could not cause significant toxicity and side effects on host cells [12]. Responsible for replication and transcription of the viral RNA genome. Fpocket predicted total 63 pockets in RdRp (SF:7a). Pocket 1, 2, 3, 6, 9, 17 and 47 showed DS of 0.867, 0.799, 0.13, 0.189, 0.314, 0.156, and 0.76 respectively. Physicochemical properties for these pockets are motioned in (SF:7b). Other than these high druggable pockets, there are fifteen pockets, which have volume ≥ 400Å. These pockets are 1, 4, 10, 11, 16, 18, 22, 30, 46, 47, 55 and pocket 60(SF:7b).

3.9. Uridylate-specific endoribonuclease (NendoU/Nsp15)

Mn(2+)-dependent, uridylate-specific enzyme, which leaves 2'-3'-cyclic phosphates 5' to the cleaved bond. In Nsp15, pockets 3, pocket 1 and pocket 10 showed DS of 0.702 0.624 and 0.271 respectively among twenty-one pockets identified (SF:8a).

Pocket 3 comprises of amino acid LYS46, THR47,THR 48, LEU49, LYS89, ARG90, ASP91, ALA92, PRO93, HIS95, ILE96, and PRO270. Pocket 3 has a volume of 482.97Å with low HS (5.33), high PS (7), and 3 charge score, showed the polar nature of the pocket. Total sasa of pocket 3 is 161.07 suggested that pocket is exposed.

Pocket 1 is largest pocket (volume 667.65Å) in Nsp15 and also had large total sasa indicate that it also exposed to solvent with larger polar sasa (SF:8b). Pocket 1 also showed low HS (10) and highest PS (16). These results indicate pocket 1 is most polar pocket in Nsp 15 with zero charge score. Amino acid residues in pocket 1 are ARG198, ASN199, ASP267, ASP272, ASP296, GLN201, GLU68, LEU200, LEU251, LEU265, LYS70, LYS89, LYS276, MET271, SER197, THR166, TYR278, VAL294.

Pocket 10 has a volume 402.83Å, HS 15.41, PS 8 and 3 charge score. Amino acid involve in pocket 10 are ASP239, GLN244, GLY246, GLY247, HIS234, HIS242, HIS249, LEU245, LYS289, THR340, TYR342, VAL291.

Other than pocket 1, 3 and 10 there are three pockets, have volume greater then 400Å and may accommodate drug like molecules. These pockets are Pokcet 2, Pocket 14 and pocket 20 and have a volume 509.81Å, 482.61Å, 637.76Å respectively.

3.10. 2'-O-ribose methyltransferase (Nsp16)

Functional annotation of Nsp16 says it's a methyltransferase that mediates mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs. N7-methyl guanosine cap is a prerequisite for binding of Nsp16. Therefore plays an essential role in viral mRNAs cap methylation which is essential to evade immune system." So targeting pockets of Nsp16 might be very useful. Nsp16 also forms a hetero-oligomeric complex with Nsp10.

Total 20 pockets were identified in Nsp16 (SF:9a). Among these, pocket 4, pocket 3 and pocket 7 showed highest DS of 0.174, 0.147 and 0.125 respectively. Amino acid involve in pocket 4 are ASN101, ASP 99, ASP114, ASP130, ASP133, CYS115, GLY71, GLY73, GLY113, LEU100, LYS135, MET131, PHE70, PHE149, PRO134, SER98, TYR132. Pocket 4 has low HS (9.47), high PS (8) and negative charge score of -3, indicating that its large (665.31Å), polar negatively charge pocket (SF:9b).

Pocket 7 has 2nd largest in size and has volume 977.82Å with a HS 20.23, PS 13 and charge score -4 and has largest total sasa, indicating that pocket is exposed, both hydrophobic and polar residues are present in the pocket. The residues in pocket 7 are ALA121, ARG66, ARG283, ARG287, ASN122, ASN286, GLN158, GLU264, GLU284, GLY213, ILE267, ILE282, LEU163, LEU212, LEU262, LYS123, LYS160, LYS214, SER261, TYR211, VAL289.

Pocket 13 smaller in size as compared to pocket 4 and 7. It has volume of 299.68Å, HS 49.39, PS 2 and zero charge score. It a small hydrophobic pocket surrounded by ALA45, GLN87, ILE40, LEU244, MET247, PHE245, PRO37, THR48, VAL44 amino acid residue.
Other than pocket 4, 7 and 13 there are three pockets, have volume greater then 400Å and may accommodate drug like molecules. Pocket 1 biggest pocket in Nsp16 with a volume of 1090.08Å. Pocket 11 and pocket 19 has a volume of 415.88Å and 478.67Å respectively.

3.11. Protein 7a

Fpocket identified four pockets in protein 7a (SF:10a). Highest DS (0.275) was observed in Pocket 1 which is surrounded by amino acid residues GLN21, GLN76, HIS19, LEU31, LEU77, LYS32, TYR20, TYR75. Pocket 1 also showed highest HS (35.62), PS (6) and charge score (2) in comparison to other pockets. However, pocket 1 occupy small volume of 242.43Å and total sasa of 74.31(SF:10b). Protein 7a is Non-structural protein and dispensable for virus replication in cell culture.

3.12. Nucleoprotein (NC)

Uniport functional annotation says, NC "packages the positive strand viral genome RNA into a helical ribonucleocapsid (RNP) and plays a fundamental role during virion assembly through its interactions with the viral genome and membrane protein M and also plays an important role in enhancing the efficiency of subgenomic viral RNA transcription as well as viral replication".

In total 8 pockets were identified (SF:11a). Pocket 5 showed highest DS with negative HS - 4.16, PS 3, charge score 2 and volume 302.77Å (SF:11b). These properties of pocket 5 exhibit its highly polar nature. Amino acid residues involve in pocket 5 are LYS65, PHE66, PRO67, PRO168, THR166, LYS169. Two other pockets, Pocket 1 and 2 have volume greater than 400Å.

3.13. Helicase

Helicase (Nsp13), a multi-functional protein. N-terminal structure contains 26 cysteine residues to form a Zn2+ binding domain and helicase domain with a conserved motif at the C-terminus. Importantly, it has been reported that the SARS-Nsp13 sequence is conserved and indispensable, and is a necessary component for the replication of coronavirus. Therefore, it has been identified as a target for anti-viral drug discovery, but there are few reports about Nsp13 inhibitors [12].

Total 39 pockets (SF:12a) were identified in Nsp13 in which only pocket 32 showed DS greater than 0.1. Volume (2029.96) and total sasa (630.63) suggested that pocket 32 is biggest accessible pocket in Nsp13 with major apolar residues and few polar residues (SF:12b). The residues are ALA509, ALA522, ARG502, ARG507, ARG594, ARG595, ASN503, ASN596, CYS471, GLU498, GLU591, GLY527, ILE493, ILE525, ILE545, ILE575, ILE592, LEU500, LEU526, LEU28, LEU573, LEU590, LYS473, MET474, PHE475, PHE499, PHE511, PRO504, PRO593, THR501, THR547, TRP506, TYR476, TYR543, VAL495, and VAL496. Pocket 25, 30, 33, 35, 36, 37, 38, 39 has volume ≥ 400Å.

3.14. Spike protein

Spike protein S1 attaches the virion to the cell membrane by interacting with host receptor, initiating the infection. 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. Total 91 pockets were identified, pocket 1, 6, 12, 19, 36 showed drugaability score greater than 0.10 (SF:13a). Pocket 5, 10, 12, 19, 26, 52, 54, 57, 58, 76, 82, 85, 87, 90 and 91 has volume ≥ 400Å. Physicochemical properties of these pockets are listed in (SF:13b)

Up till know we have discussed the pockets of SARS-CoV-2 for that partial or whole structure is available in PDB. Now we have predicted the pocket form the SARS-CoV-2 model structure for those experimental structures still not deposited into the PDB and some of them are model with their remote homologues. These are low quality models. So we have treated as a separate category.

3.15. Host translation Inhibitor Nsp1 (Nsp1)

Inhibits host translation by interacting with the 40S ribosomal subunit by suppressing host gene expression, Nsp1 facilitates efficient viral gene expression in infected cells and evasion from host immune response. For Nsp1, total 11 pockets were identified in which two pockets, pocket 1 and 2 have DS of 0.474 and 0.201 respectively (SF:14a). Apart from these pockets, pocket 7 and pocket 10 has volume above 400Å (SF:14b).

3.16. Non-structural protein 2 (Nsp2)

May play a role in the modulation of host cell survival signaling pathway by interacting with host PHB and PHB2 proteins. In total five pockets predicted (SF:15a). Out of which only pocket 3 has DS of 0.346, HS 48.4 and zero value for polar and charge score. Indicated that pocket 3 have hydrophobic in nature however, pocket size is very low 151.314Å. There is another pocket, pocket 1 predicted to have largest volume 948.51Å (SF:15b).

3.17. Non-structural protein 4 (Nsp4)

Participates in the assembly of virally-induced cytoplasmic double-membrane vesicles necessary for viral replication. Total 10 pockets were identified in Nsp4 (SF:16a). Pocket 2 has DS 0.432, volume 560.76Å (SF:16b) and polar in nature as the pocket have high PS and low HS. Pocket 4 is largest pocket (volume 714.07Å) in Nsp4.

3.18. Non-structural protein 6 (Nsp6)

Only remote homologues were identified as potential template structures. No pockets were identified.

3.19. Envelope small membrane protein (E protein)

Interacts with the accessory proteins 3a and 7a and plays a central role in virus morphogenesis and assembly, induction of apoptosis. Acts as a viroporin and self-assembles in host membranes forming pentameric protein-lipid pores that allow ion transport. In E protein total 40 pockets were identified in which pocket 2, 7, 20 , 23, 29 37, 39 and 40 predicted to have very good DS (SF:17a;17b). Apart from these pocket, Pocket 24, 25, 32, 33 and 34 have volume greater than 400Å might accommodate drug like molecules.

3.20. Membrane Protein (M protein)

Component of the viral envelope that plays a central role in virus morphogenesis and assembly via its interactions with other viral proteins. Only remote homologues were identified as potential template structures. Only 1 pocket identified with a DS 0.365 with HS 33.6, PS 2, charge score zero and has volume 382.33Å (SF:18a; 18b).

3.21. Non-structural protein 6 (Ns6)

Uniport annotation suggests that it could be a determinant of virus virulence and seems to stimulate cellular DNA synthesis in vitro. Only remote homologues were identified as potential template structures. Only 2 pockets were identified in which pocket 1 has DS of 0.256 and pocket 2 has zero DS. However, the pocket is relatively small with volume 161.86Å and 234Å respectively (SF:19a; 19b).

3.22. Protein non-structural 7b (Ns7b)

No homologues were identified as potential template structures for modelling.

3.23. Non-structural protein 8 (Ns8)

It might play a role in host-virus interaction. Only remote homologues were identified as potential template structures. Total eight pockets predicted out of which pocket 4 has good

DS of 0.692, volume 529.66Å, HS of 38.12, PS 8 and 1 charge score (SF:20a; 20b). Pocket 3 has biggest pocket in Ns8 with volume of 956.52Å.

3.24. ORF9b protein

Thirteen pockets were identified in ORF9b in which pocket 1 is the biggest pocket (volume 1011.11Å) with highest DS score of 0.98. HS, PS, charge score, and total sasa are 56.94, 4, 2 and 166.85 respectively (SF:21a). Pocket 2, pocket 9 and pocket 13 have volume greater than 400Å so selected as potential druggable pockets (SF:21b).

3.25. Uncharacterized protein 14: May play a role in host-virus interaction

No homologues were identified as potential template structures for modelling.

3.26. ORF10 protein

No homologues were identified as potential template structures for modelling.

3.27. Docking of antiviral and antimicrobial phytomolecules against 118

Tinosporinone, Theaflavin, Menthol, Eugenol, Curcumin, Coumarin, Catechin, Carvacrol, Baicalin, Andrographolide were showed high binding affinity with multiple pockets (SF:22-30). In which Theaflavin shown highest docking binding affinity towards multiple targets Fig: 2. Docking results showed that these compounds have high binding affinity towards spike protein, Helicase, M^{pro}, Nsp14, Nsp3 pockets. Theaflavin, and Baicalin showed highest binding affinity with spike and M^{pro}, Eugenol, menthol, ajoene, allicin, carvacrol, and coumarin are predicted to bind well into spike protein pocket1 (Table:1) and Ursolic and Andrographolide showed highest biding affinity towards Nsp14 pocket2. Tinosporinone, catechin and curcumin showed highest binding affinity towards Nsp3, M^{pro}, helicase respectively. Top five pockets from the docking of were listed in ST:1. Top pockets for rest of the phytomolecules listed in ST:2.
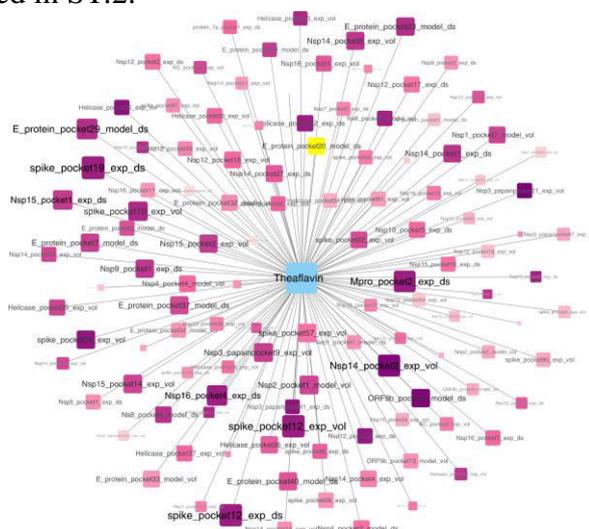
Fig: 2 The binding energy interaction network for the SARS-CoV-2 Pockets with theaflavin (centre). The target node size and colour depend on the binding energy scores, ie. greater the affinity, larger the node size; darker the colour.

Table1: Results of molecular docking of phytomolecules against 118 pockets. Phytomolecules against influenza virus showed better binding affinity towards **SARS-CoV-2** pockets. Best pocket against top hits molecules listed with their docking score.

	Natural Compunds	**Pockets**	**Binding energy kcal/mol**
1.	Theaflavin	spike_pocket19	-16.51
2.	Baicalin	Mpro_pocket2	-15.02
3.	Andrographolide	Nsp14_pocket2	-12.42
4.	Curcumin	Helicase_pocket32	-12.56
5.	UrsolicAcid	Nsp14_pocket2	-11.85
6.	Catechin	Mpro_pocket2	-11.19
7.	Tinosporinone	Nsp3 pocket21	-10.3
8.	Geranly ester	Helicase Pocket 32	-9.51
9.	Farnesol	Spike Pocket 10	-8.46
10.	Linalyl Oxide	Spike Pocket 10	-8.31
11.	Eugenol	spike_pocket10	-7.71
12.	Menthol	spike_pocket10	-7.09

4. Conclusion

This study gives us the first look at the all binding site for small ligands in the SARS-CoV-2 Structural proteome. We have identified 433 pockets in total and for these pockets hydrophobicity, polarity, solvent accessible surface area (SASA), charge, and volume were calculated that will suggest the nature of pockets and may facilitate the better computational screening of compounds against SARS-CoV-2. Further, we have docked antiviral/antimicrobial phytomolecules to 118 druggable pockets to identify SARS-CoV-2 natural inhibitors. Finally, for the first time, we have done the comprehensive druggability assessment of SARS-CoV-2 for better and effective screening of compounds as well as for target identification active drug/compounds.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Reference

[1] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet. 395 (2020) 507–513. https://www.sciencedirect.com/science/article/pii/S0140673620302117?via%3Dihub.

[2] J.F.-W. Chan, S. Yuan, K.-H. Kok, K.K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.C.-Y. Yip, R.W.-S. Poon, H.-W. Tsoi, S.K.-F. Lo, K.-H. Chan, V.K.-M. Poon, W.-M. Chan,

J.D. Ip, J.-P. Cai, V.C.-C. Cheng, H. Chen, C.K.-M. Hui, K.-Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet. 395 (2020) 514–523. https://www.sciencedirect.com/science/article/pii/S0140673620301549?via%3Dihub.

[3] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, Nat Microbiol. 5 (2020) 536–544. https://www.nature.com/articles/s41564-020-0695-z.

[4] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature. 579 (2020) 265–269. https://www.nature.com/articles/s41586-020-2008-3.

[5] P. Bongini, A. Trezza, M. Bianchini, O. Spiga, N. Niccolai, A possible strategy to fight COVID-19: Interfering with spike glycoprotein trimerization, Biochemical and Biophysical Research Communications. (2020) S0006291X2030704X. https://www.sciencedirect.com/science/article/pii/S0006291X2030704X?via%3Dihub.

[6] Fpocket: An Open Source Platform for Ligand Pocket Detection | BMC Bioinformatics | Full Text, n.d. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-168.

[7] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (2018) W296–W303. https://academic.oup.com/nar/article/46/W1/W296/5000024.

[8] The Human Pocketome - De Novo Molecular Design - Wiley Online Library, n.d. https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527677016.ch3.

[9] B.H. Harcourt, D. Jukneliene, A. Kanjanahaluethai, J. Bechill, K.M. Severson, C.M. Smith, P.A. Rota, S.C. Baker, Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity, J. Virol. 78 (2004) 13600–13612. https://jvi.asm.org/content/78/24/13600.

[10] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of wide-spectrum inhibitors targeting coronavirus main proteases, PLoS Biol. 3 (2005) e324. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030324.

[11] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus, Science. 368 (2020) 779–782. https://science.sciencemag.org/content/368/6492/779.

[12] C.K. Chu, S. Gadthula, X. Chen, H. Choo, S. Olgen, D.L. Barnard, R.W. Sidwell, Antiviral activity of nucleoside analogues against SARS-coronavirus (SARS-coV), Antivir. Chem. Chemother. 17 (2006) 285–289. https://journals.sagepub.com/doi/10.1177/095632020601700506.
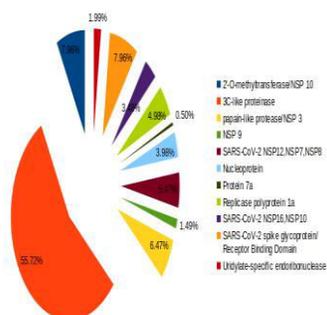
SF: 1. Total representation of SARS CoV-2 structure in PDB. More than half of the structures are of SARS CoV-2 main Protease (M^{pro}) .

## Papain-like proteinase (PL-PRO) or Nsp3
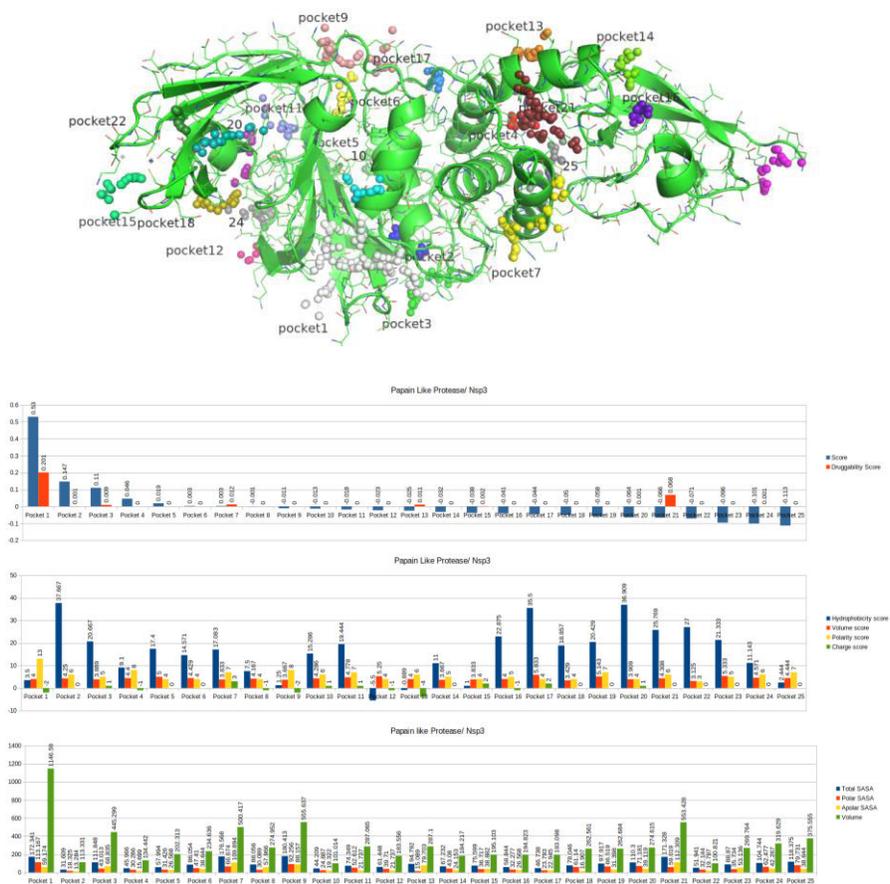
SF: 2. SARS-CoV-2 Papain-like proteinase (PL-PRO) or Nsp3 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Non-structural protein 7 (Nsp7)
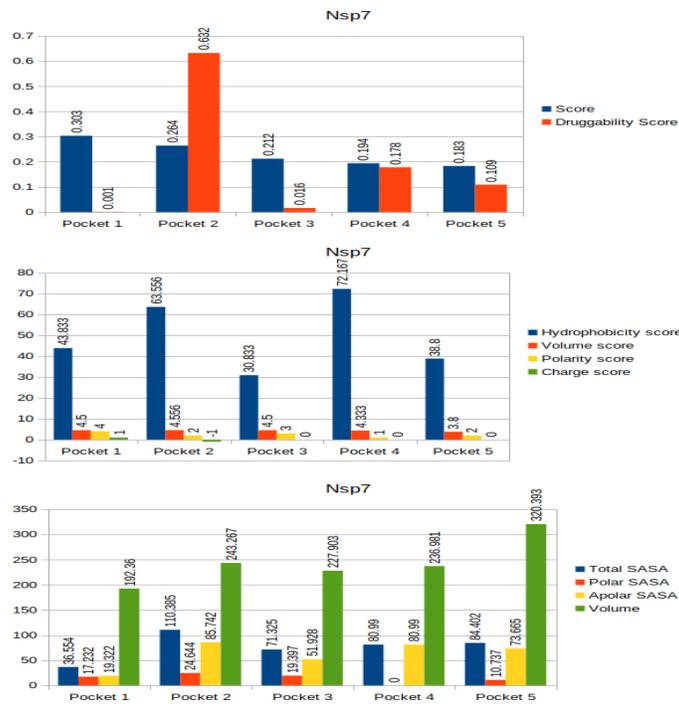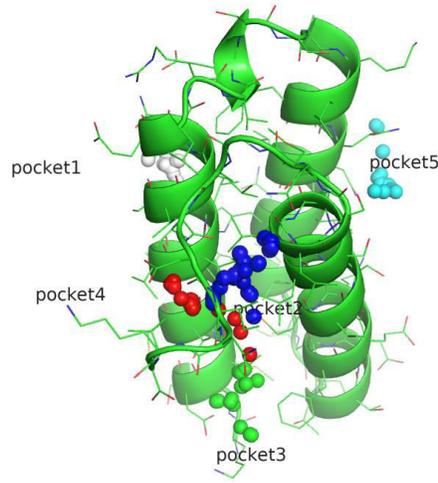

SF: 3. SARS-CoV-2 Nsp7 s pocket visualization in different colour and bar chart showed physicochemical properties for each pocket.

Non-structural protein 8 (Nsp8)
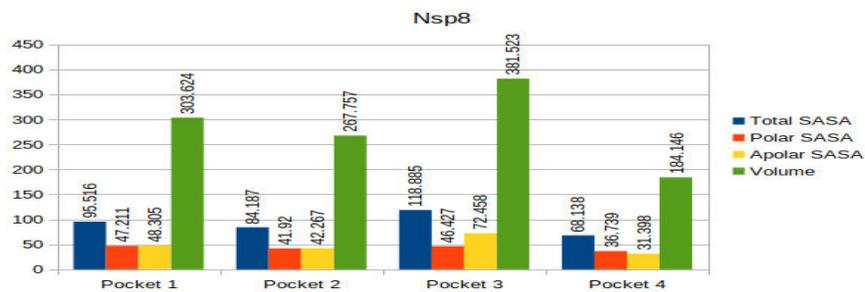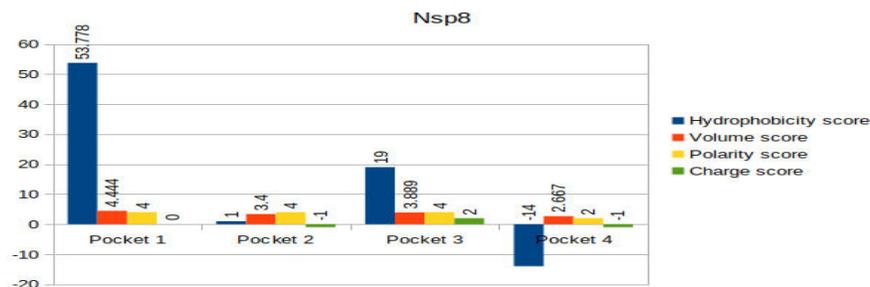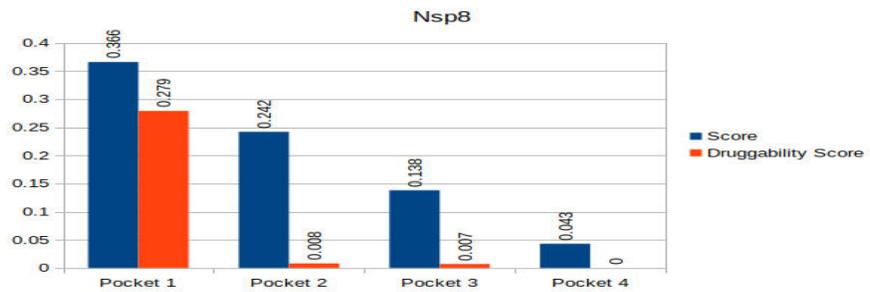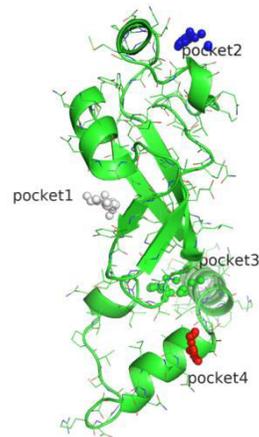
SF: 4. SARS-CoV-2 Nsp8 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Non-structural protein 9 (Nsp9)
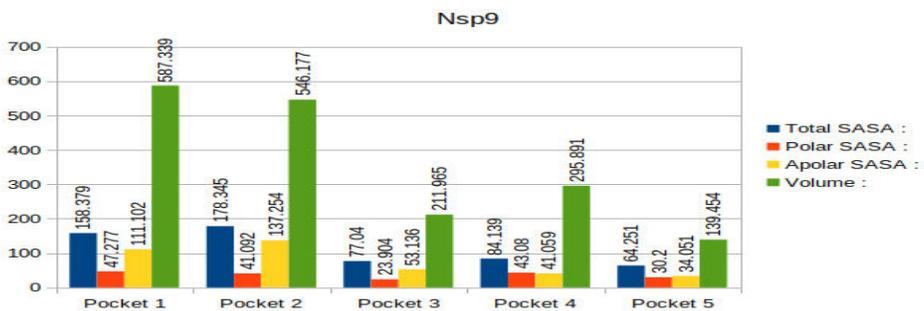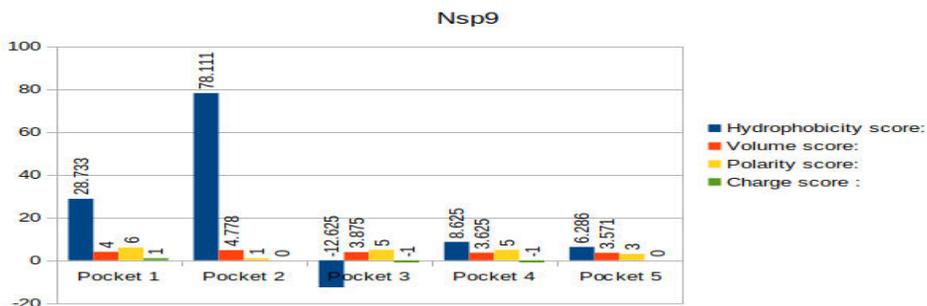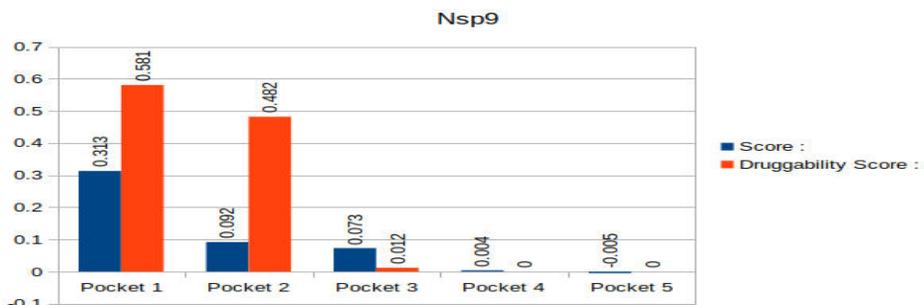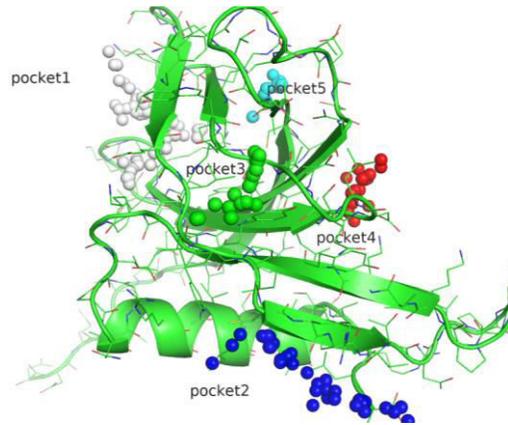
**SF: 5**. SARS-CoV-2 Nsp9 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Non-structural protein 10 (Nsp10)
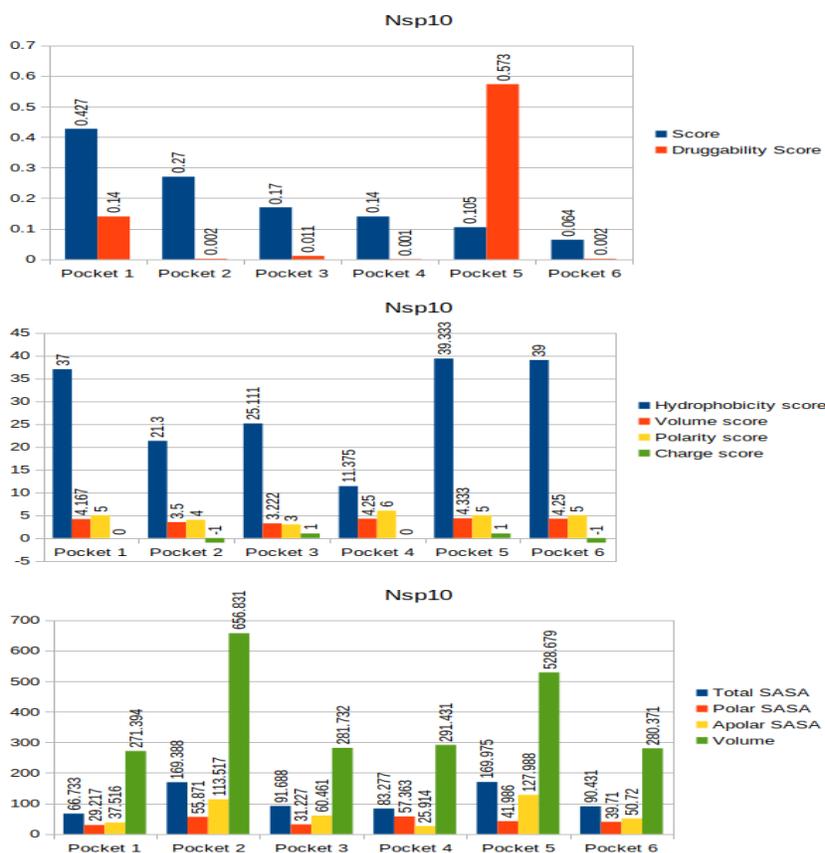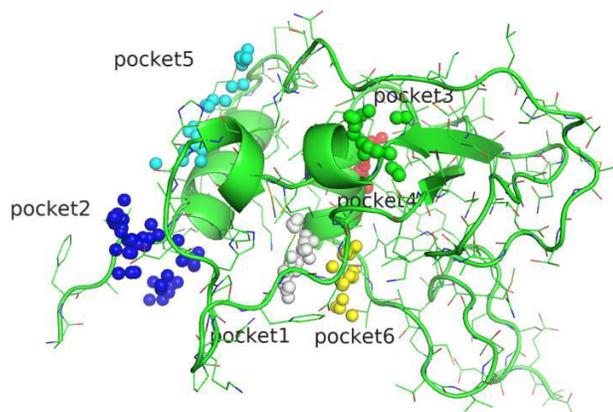
**SF: 6.** SARS-CoV-2 Nsp10 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

RNA-directed RNA polymerase (Pol/RdRp)
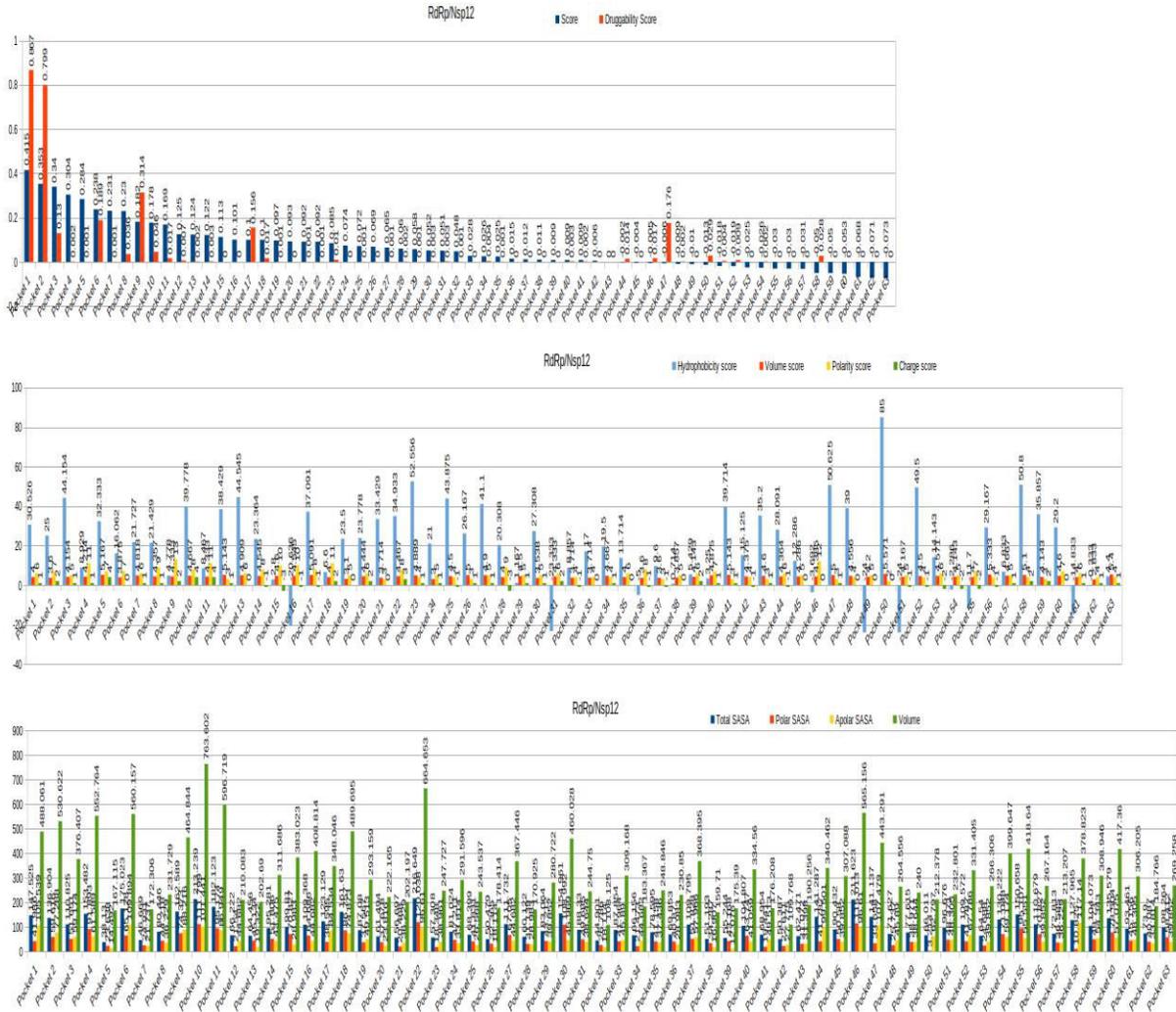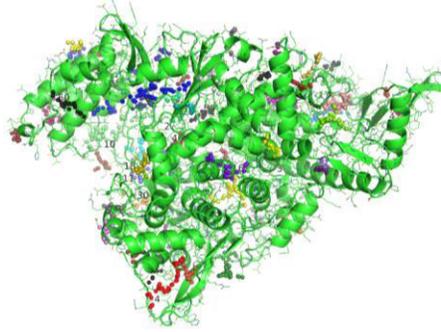
SF: 7. SARS-CoV-2 RdRp pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Uridylate-specific endoribonuclease (NendoU/Nsp15)
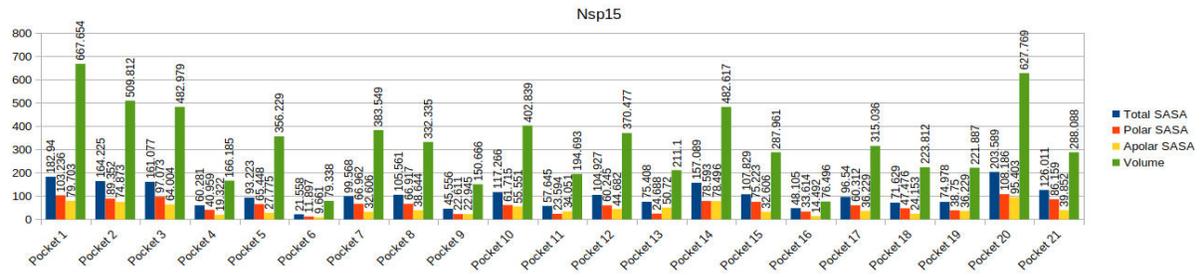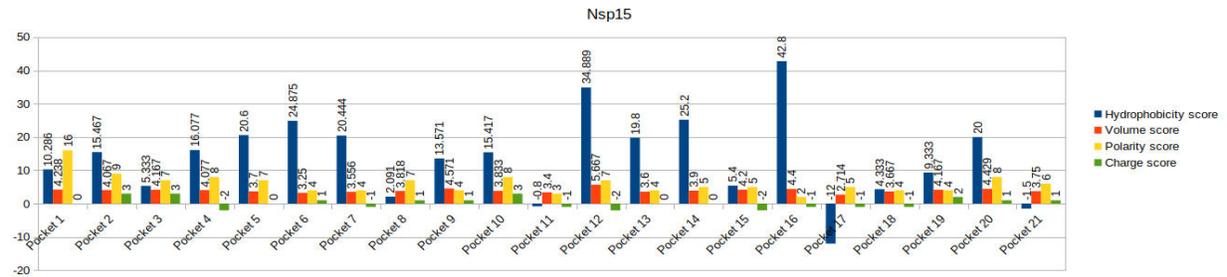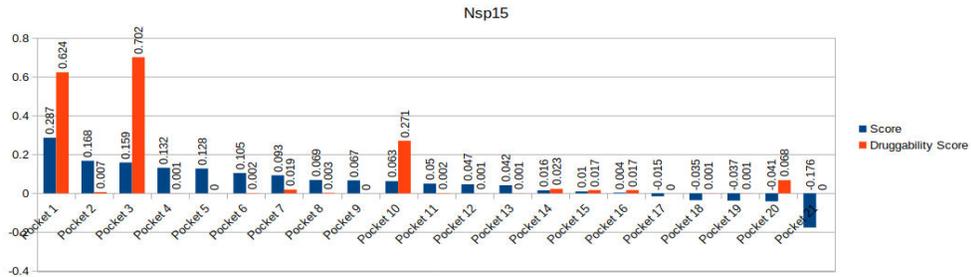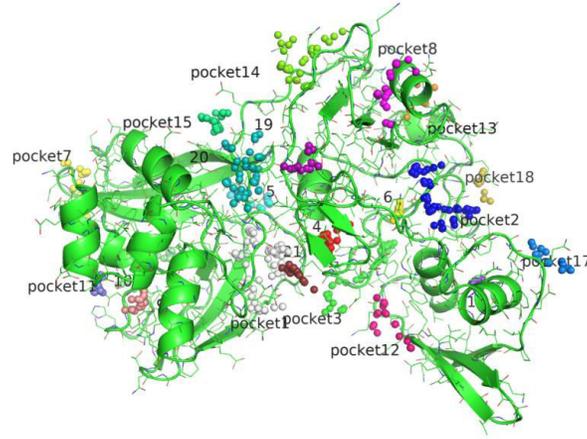
SF: 8. SARS-CoV-2 Nsp15 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

2'-O-ribose methyltransferase (Nsp16)
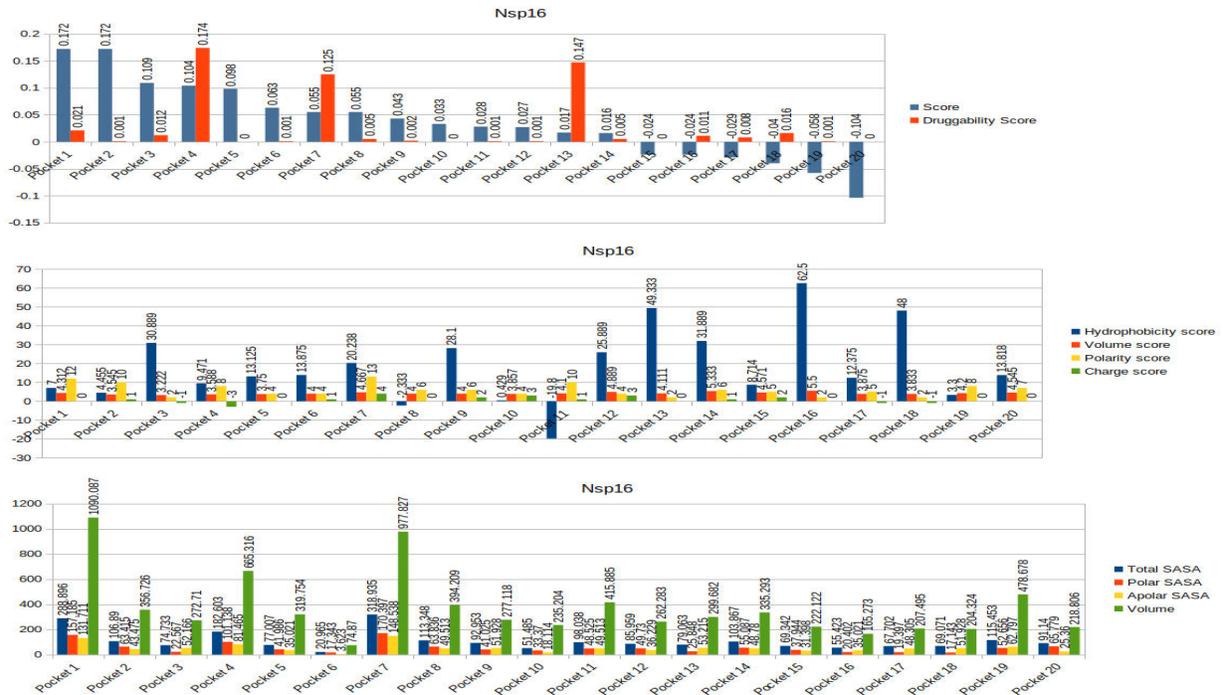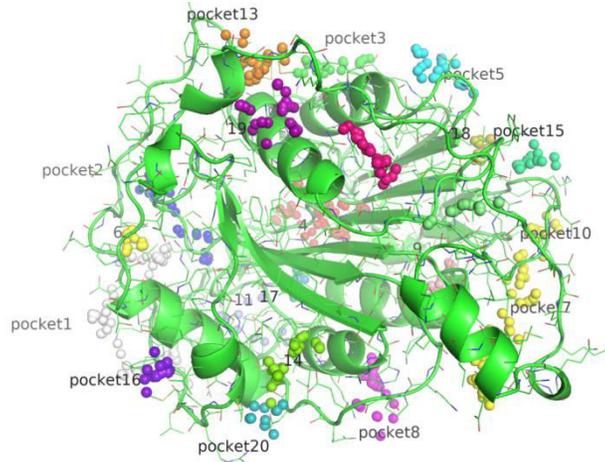
SF: 9. SARS-CoV-2 Nsp16 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Protein 7a:
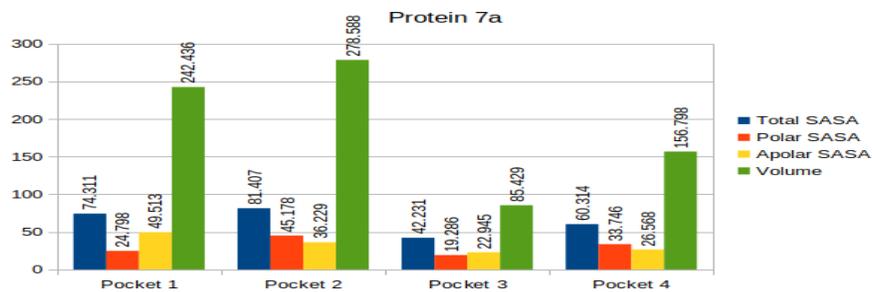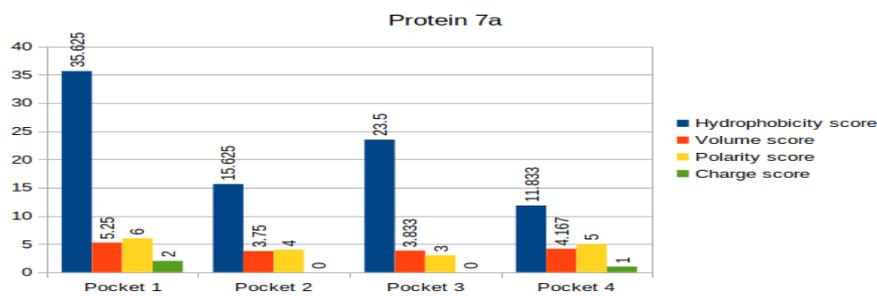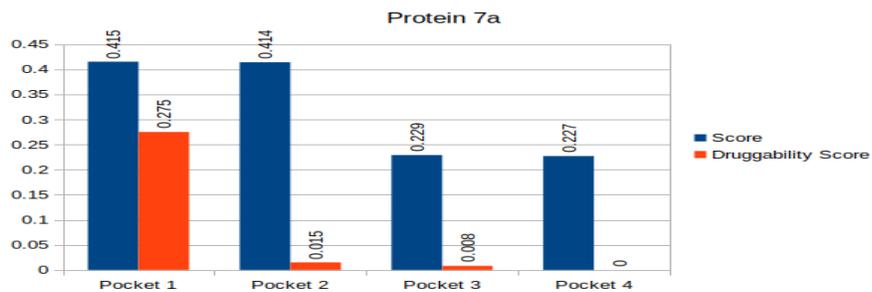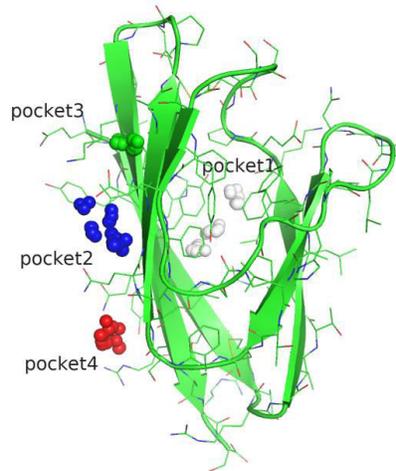


SF: 10. SARS-CoV-2 protein 7a pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Nucleoprotein (NC):
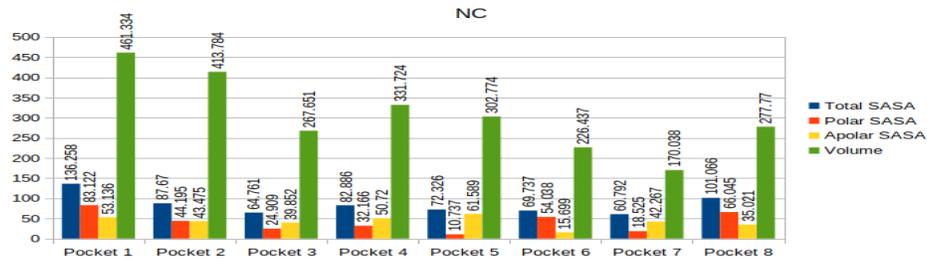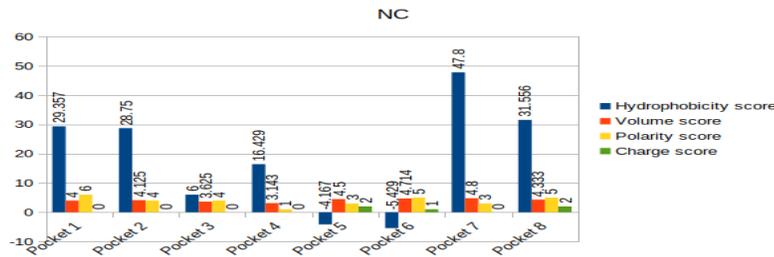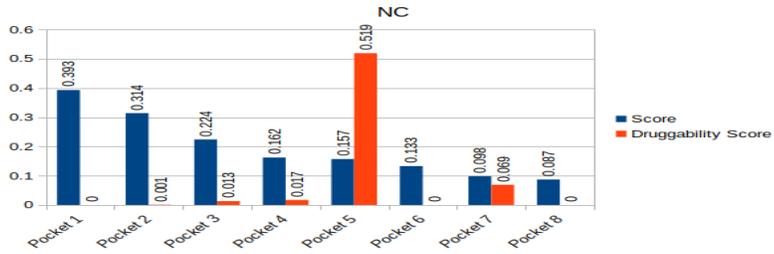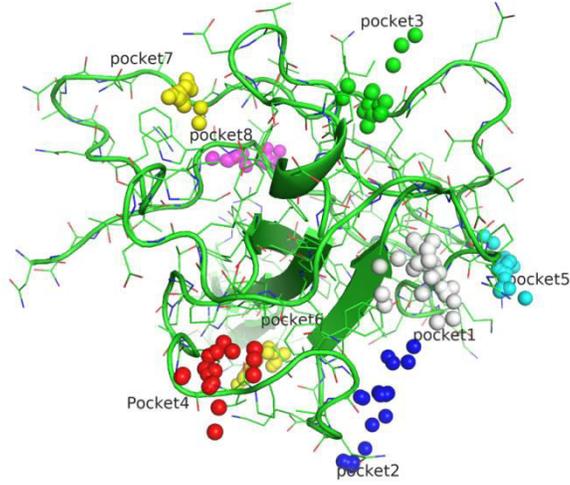
SF: 11. SARS-CoV-2 Nsp16 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Helicase
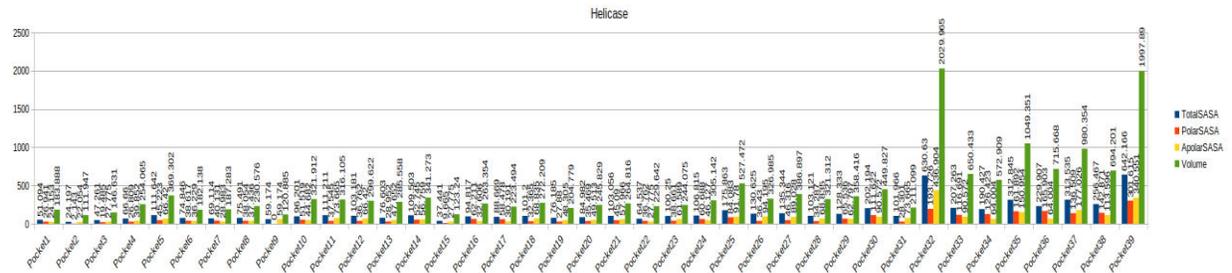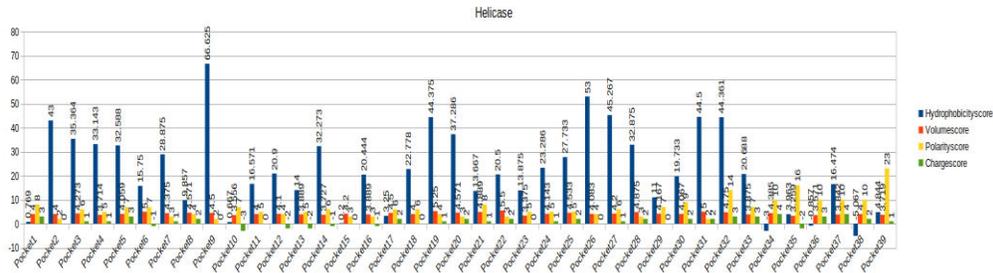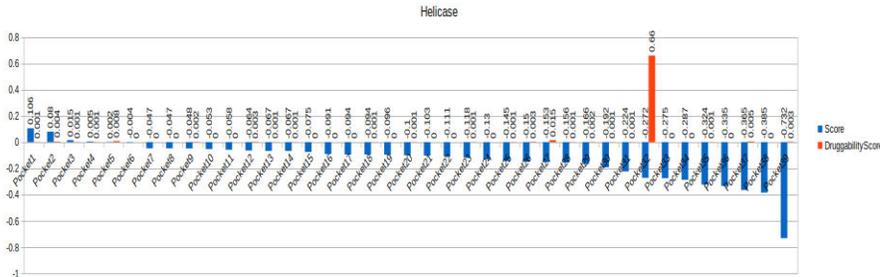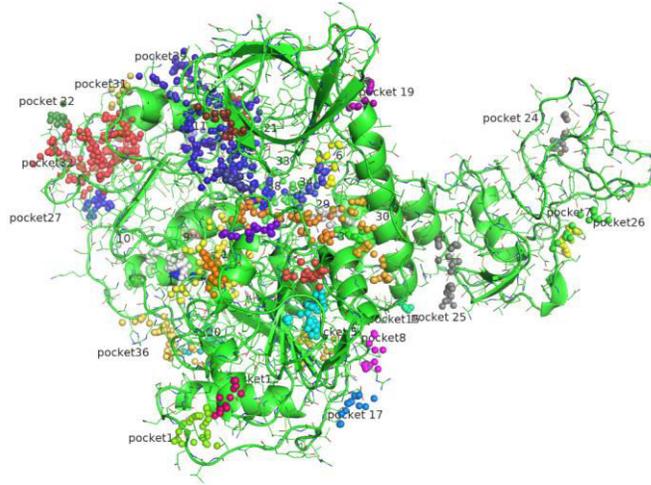
SF: 12. SARS-CoV-2 Nsp16 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Spike protein
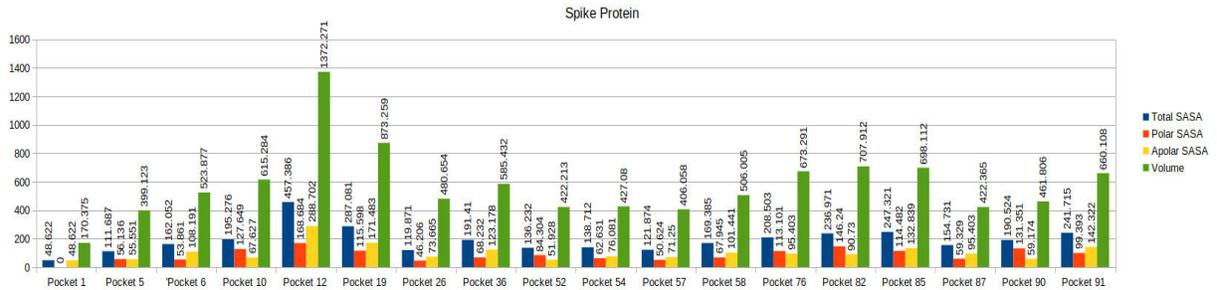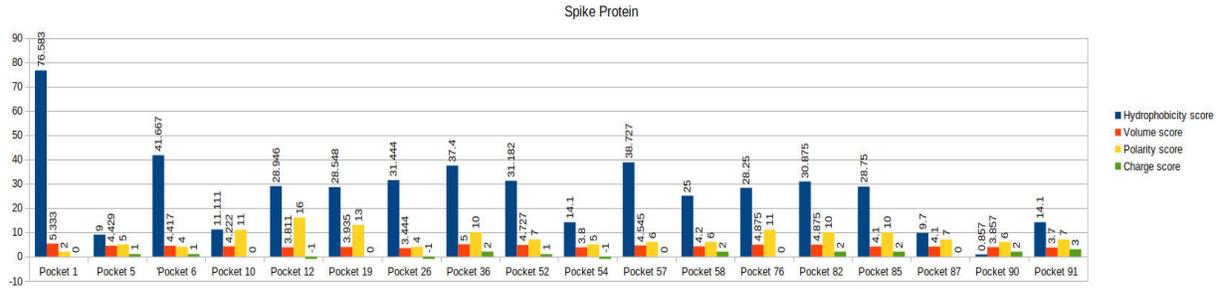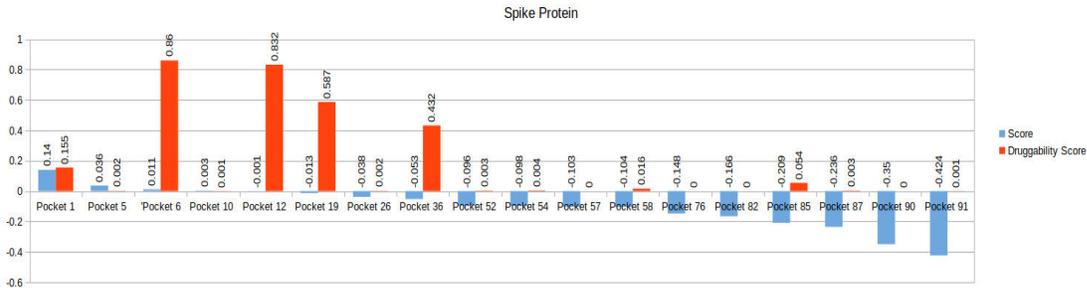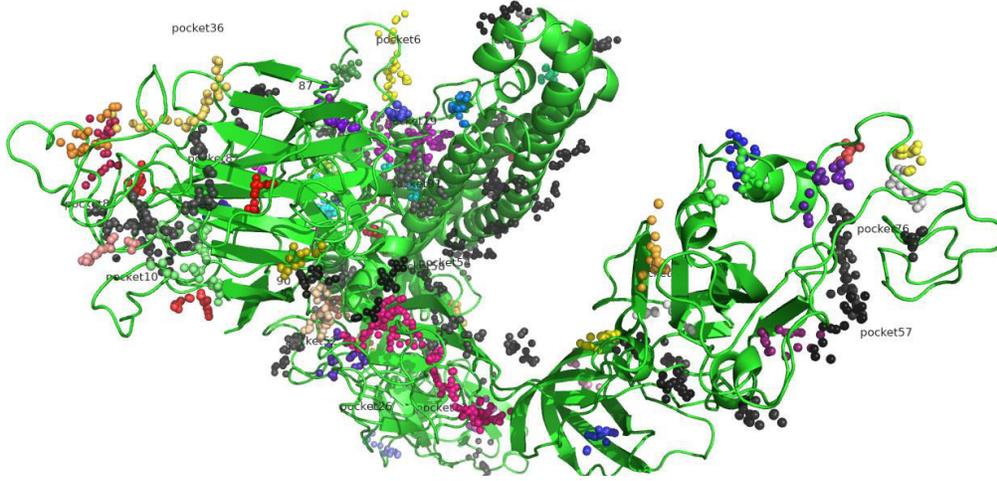
SF: 13. SARS-CoV-2 spike protein pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

Host translation Inhibitor Nsp1 (Nsp1):
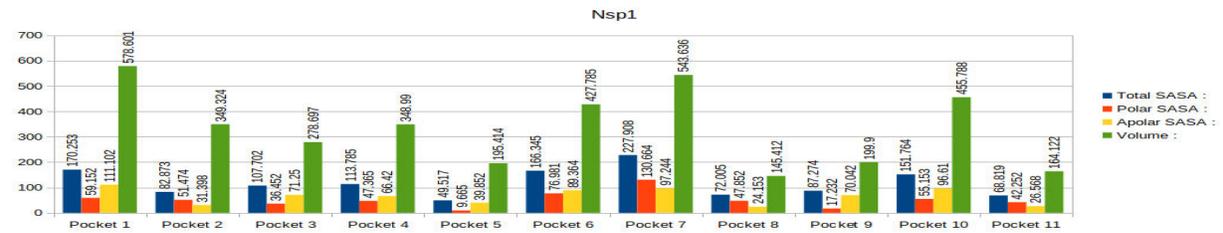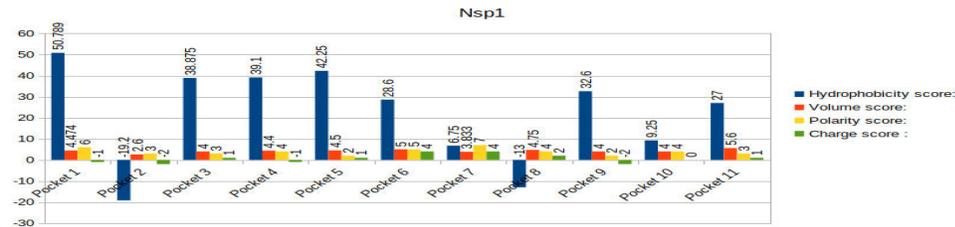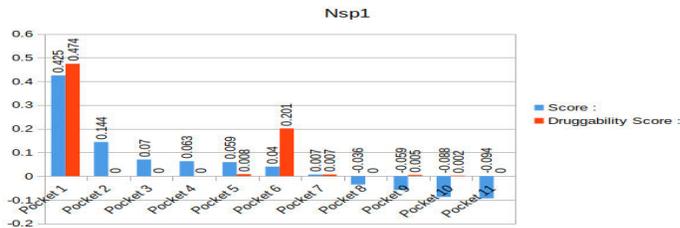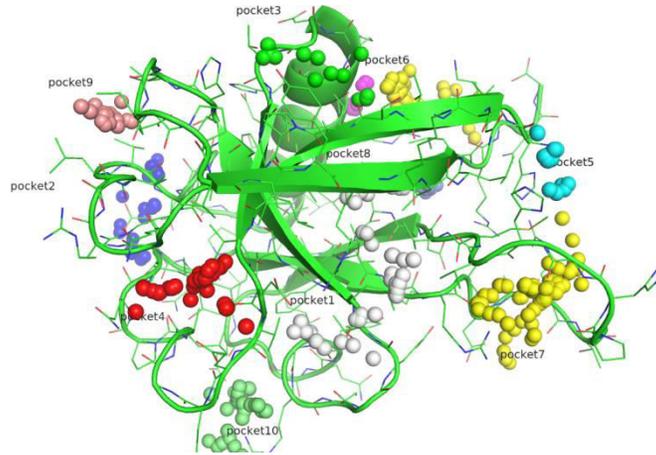
SF: 14. SARS-CoV-2 Nsp1 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

Non-structural protein 2 (Nsp2)
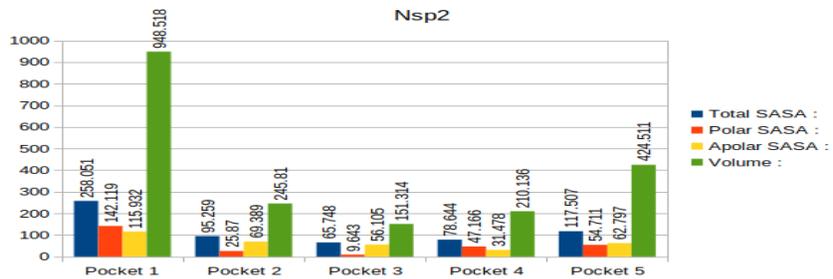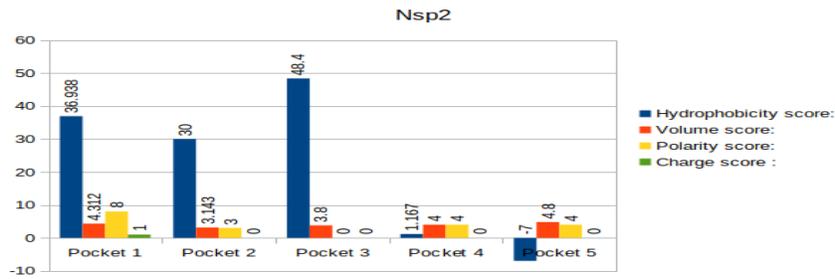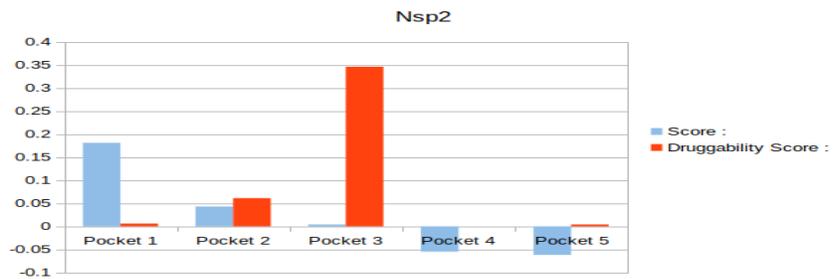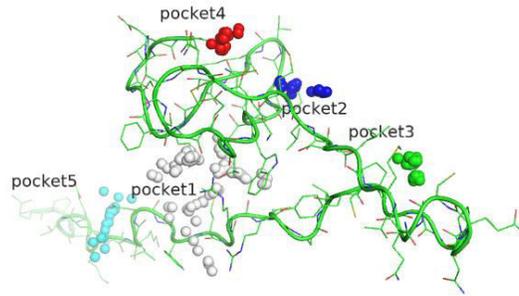
**SF: 15.** SARS-CoV-2 Nsp1 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

Non-structural protein 4 (Nsp4)
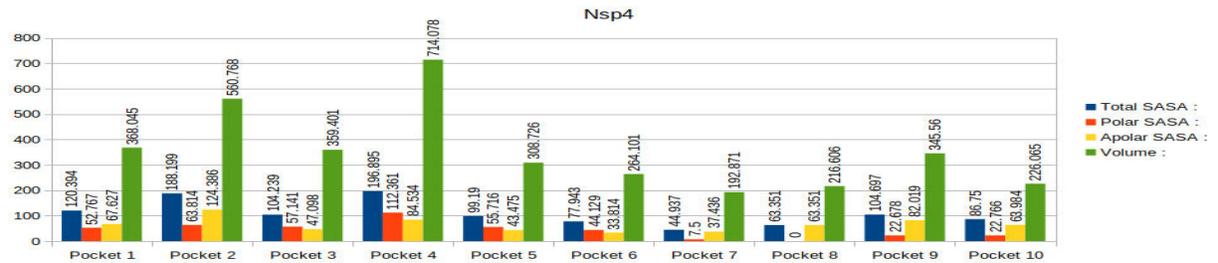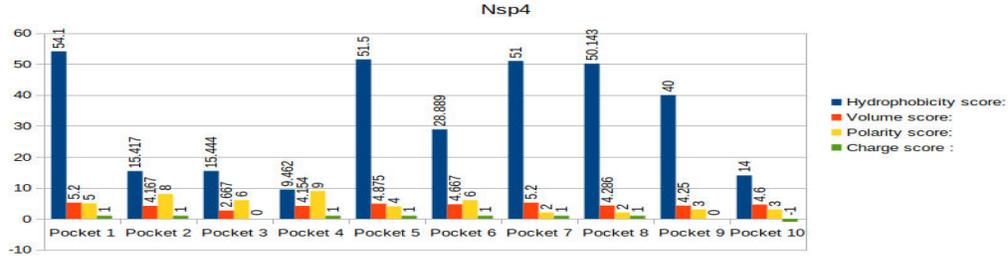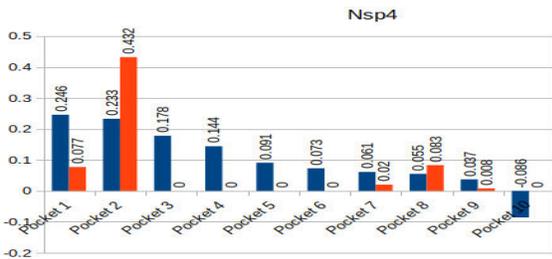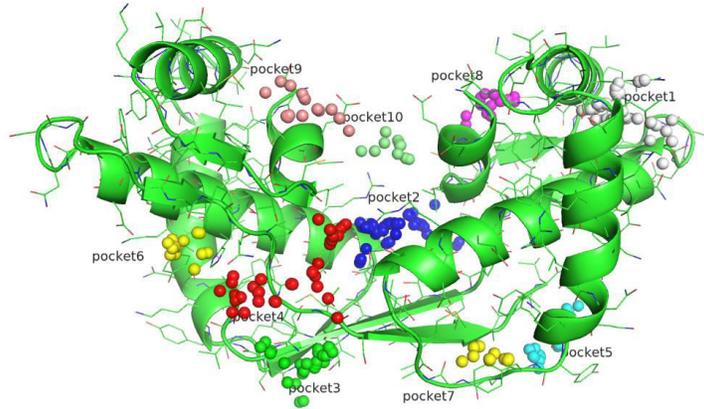
SF: 16. SARS-CoV-2 Nsp4 pockets visualization in different colour and bar chart showed physicochemical properties for each pocket.

Envelope small membrane protein (E protein)
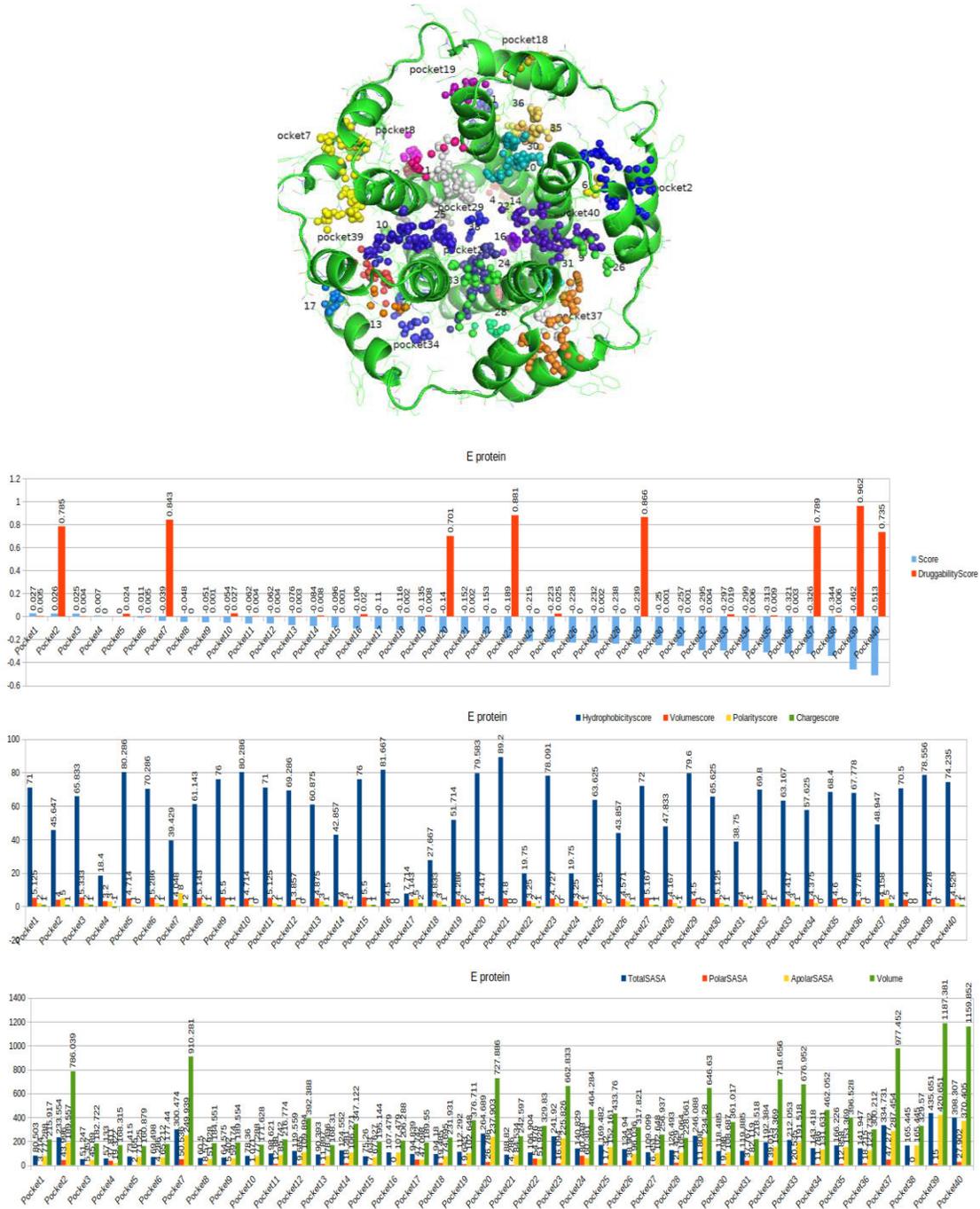
SF: 17. SARS-CoV-2 Envelope small membrane protein (E protein) pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

Membrane Protein (M protein)
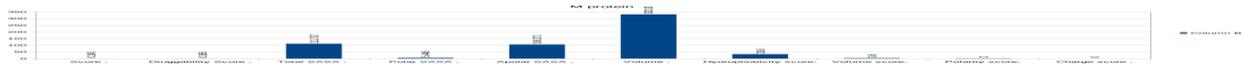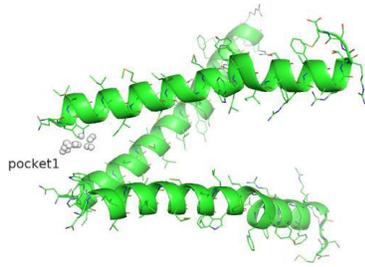
SF: 18. SARS-CoV-2 Membrane Protein (M protein) pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

Non-structural protein 6 (Ns6)
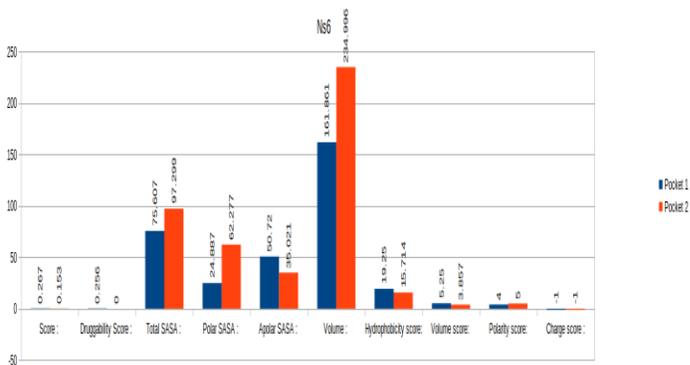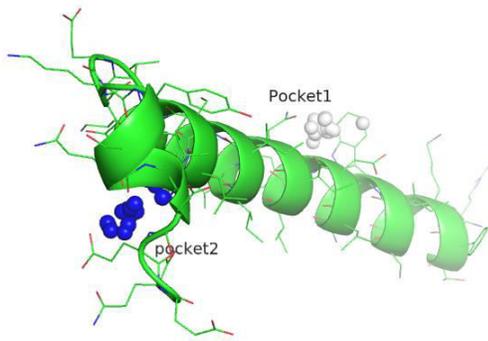
SF: 19. SARS-CoV-2 Membrane Protein (M protein) pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

Non-structural 8 (Ns8)
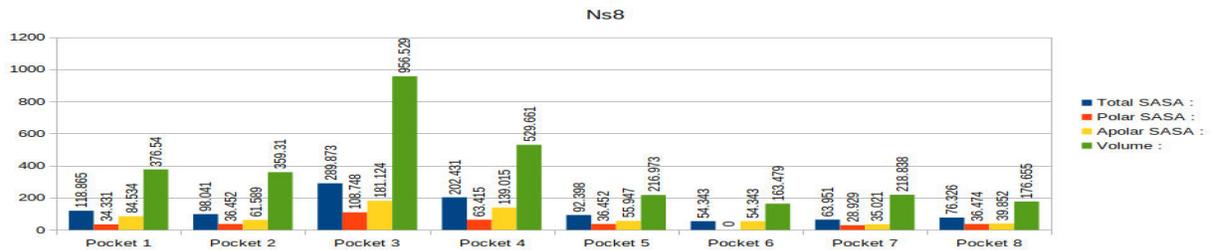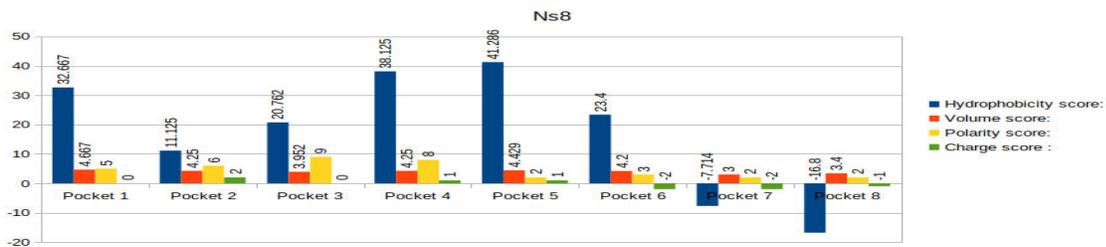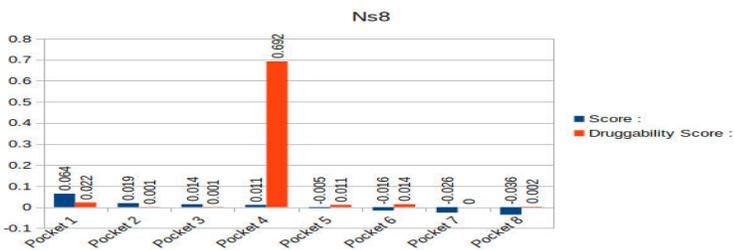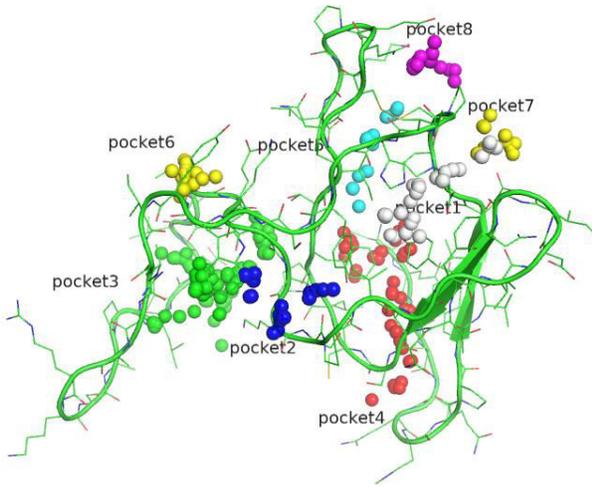
**SF: 20.** SARS-CoV-2 Membrane Protein (M protein) pockets visualization in different colour and bar chart showed physicochemical properties for each pocket

ORF9b protein
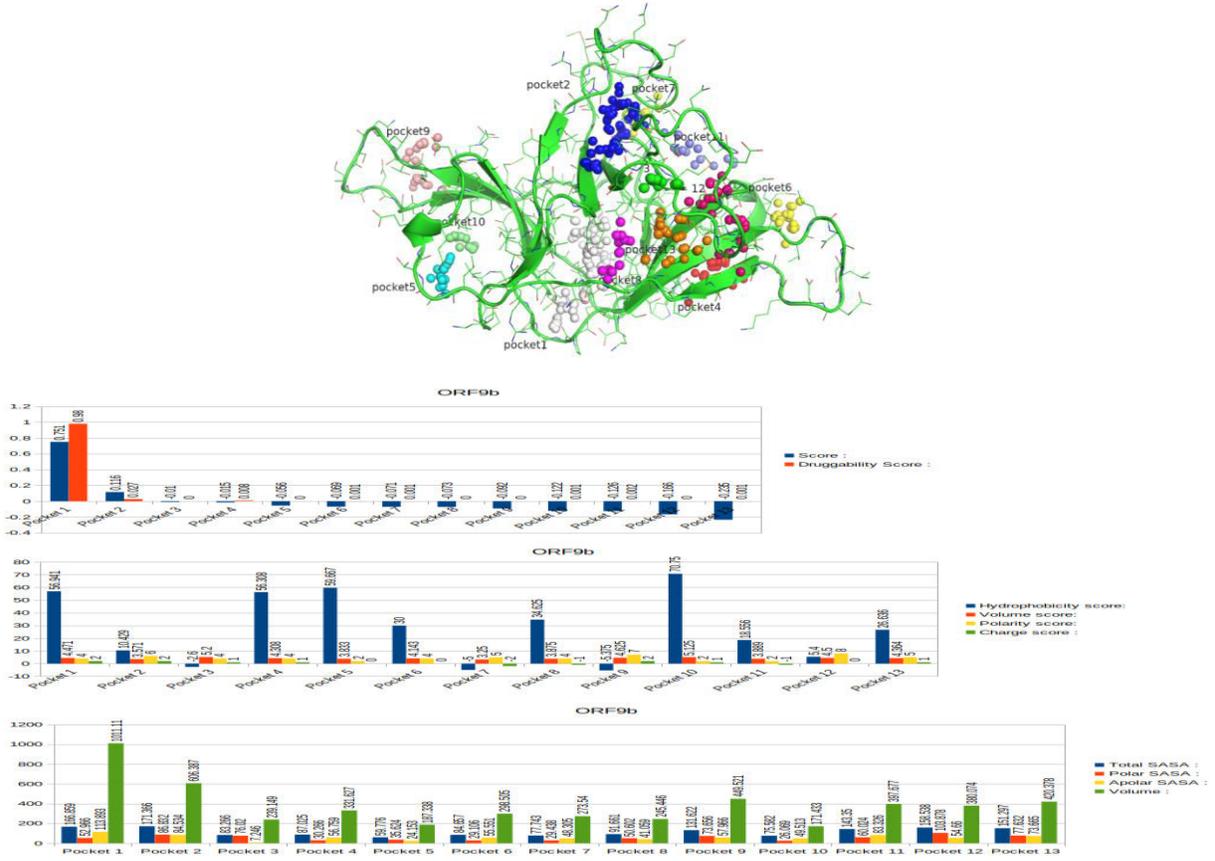
SF: 21. SARS-CoV-2 ORF9b protein pockets visualization in different colour and bar chart showed physicochemical properties for each pocket
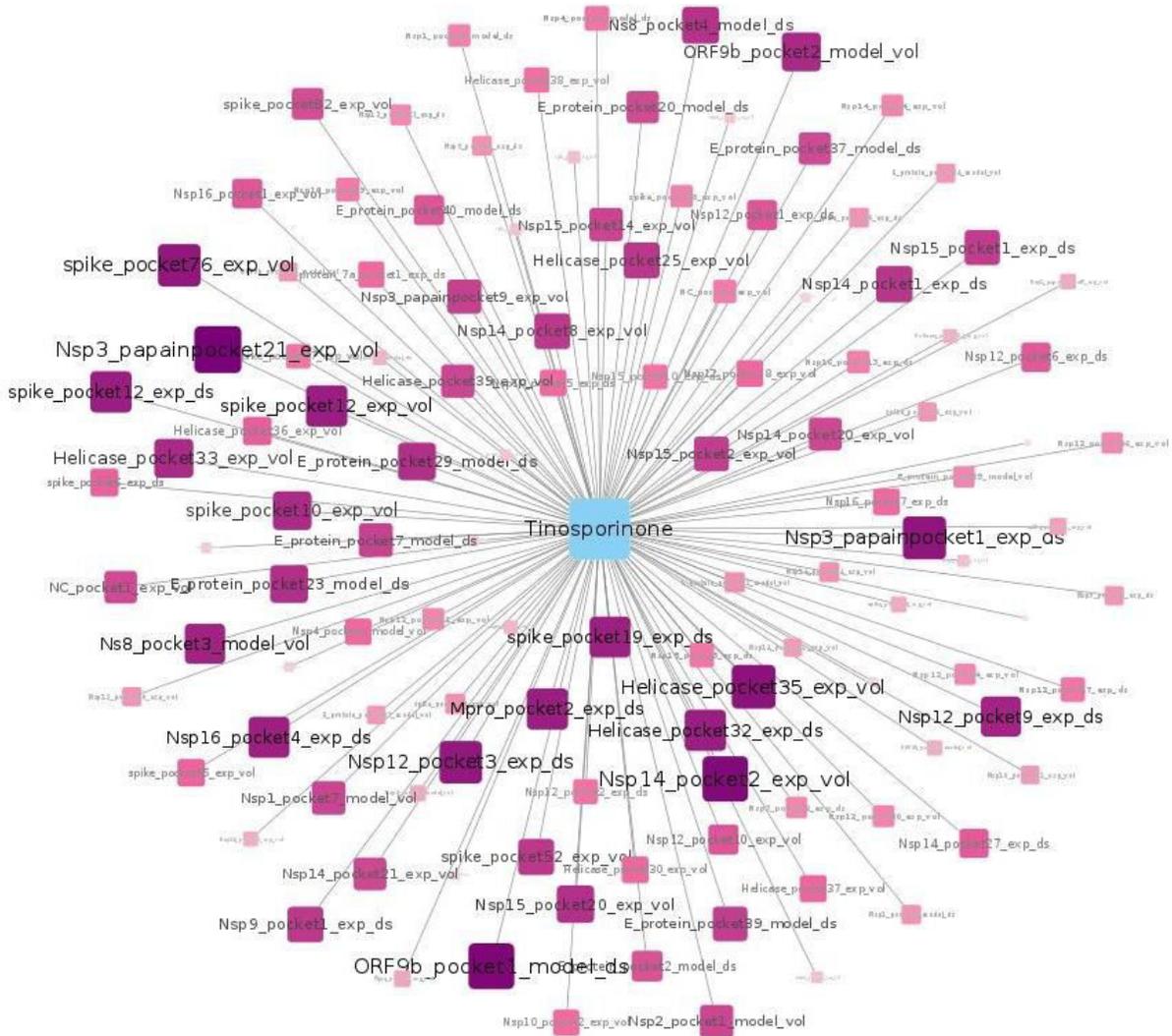
SF: 22. The binding energy interaction network for the SARS-CoV-2 pockets with Tinosporinone (centre). The target node size and colour depend on the binding energy scores, i.e. greater the affinity, larger the node size; darker the colour.
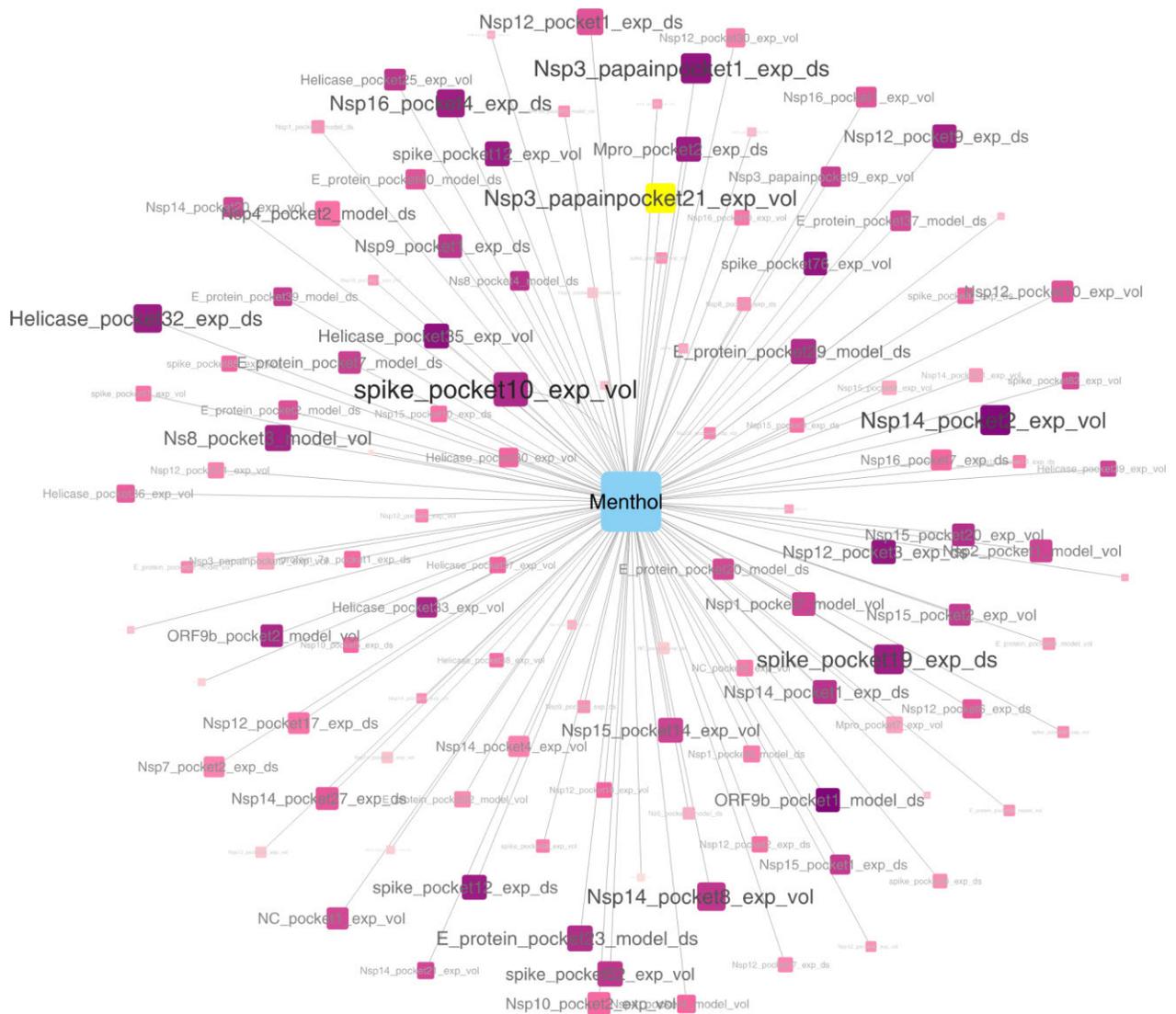
SF: 23. The binding energy interaction network for the SARS-CoV-2 Pockets with Menthol (centre). The target node size and colour depend on the binding energy scores, i.e. greater the affinity, larger the node size; darker the colour.

SF: 24. The binding energy interaction network for the SARS-CoV-2 Pockets with Eugenol (centre). The target node size and colour depend on the binding energy scores, i.e. greater the affinity, larger the node size; darker the colour.
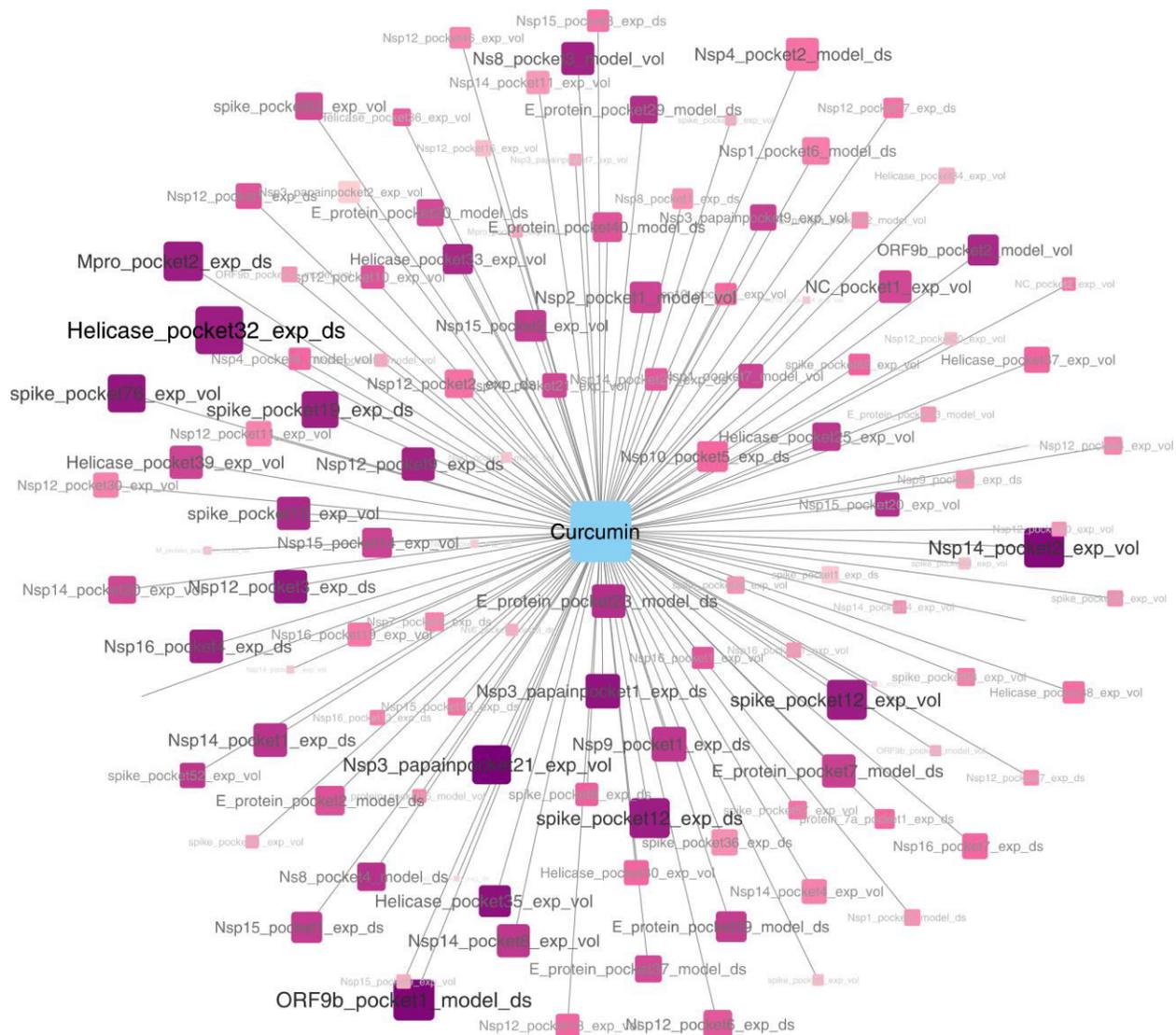
SF: 25. The binding energy interaction network for the SARS-CoV-2 Pockets with Curcumin (centre). The target node size and colour depend on the binding energy scores, i.e. greater the affinity, larger the node size; darker the colour.

SF: 26. The binding energy interaction network for the SARS-CoV-2 Pockets with Coumarin (centre). The target node size and colour depend on the binding energy scores, i.e. greater the affinity, larger the node size; darker the colour.

SF: 27. The binding energy interaction network for the SARS-CoV-2 Pockets with Catechin (centre). The target node size and colour depend on the binding energy scores, ie. greater the affinity, larger the node size; darker the colour.
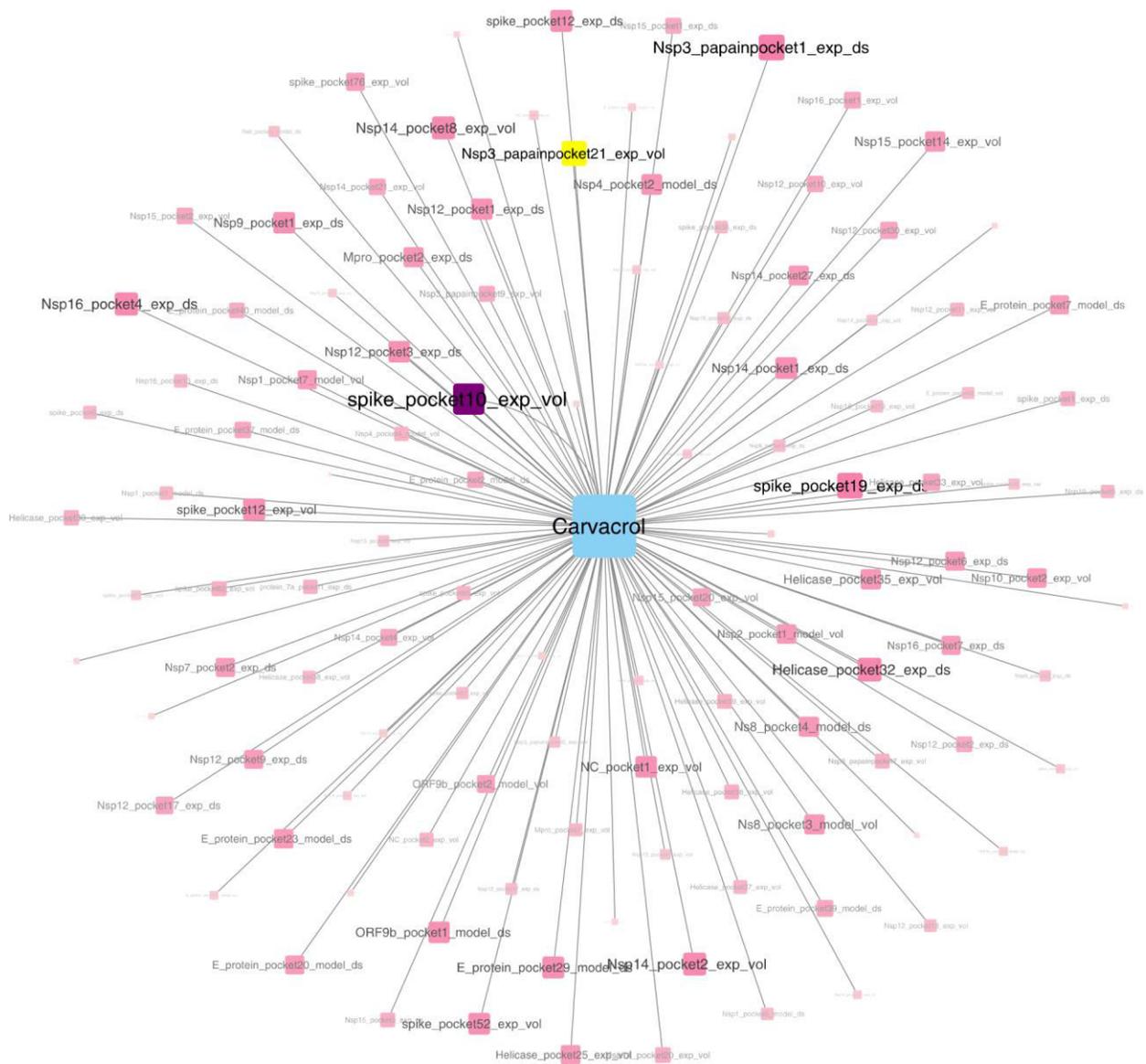
SF: 28. The binding energy interaction network for the SARS-CoV-2 Pockets with Carvacrol (centre). The target node size and colour depend on the binding energy scores, ie. greater the affinity, larger the node size; darker the colour.
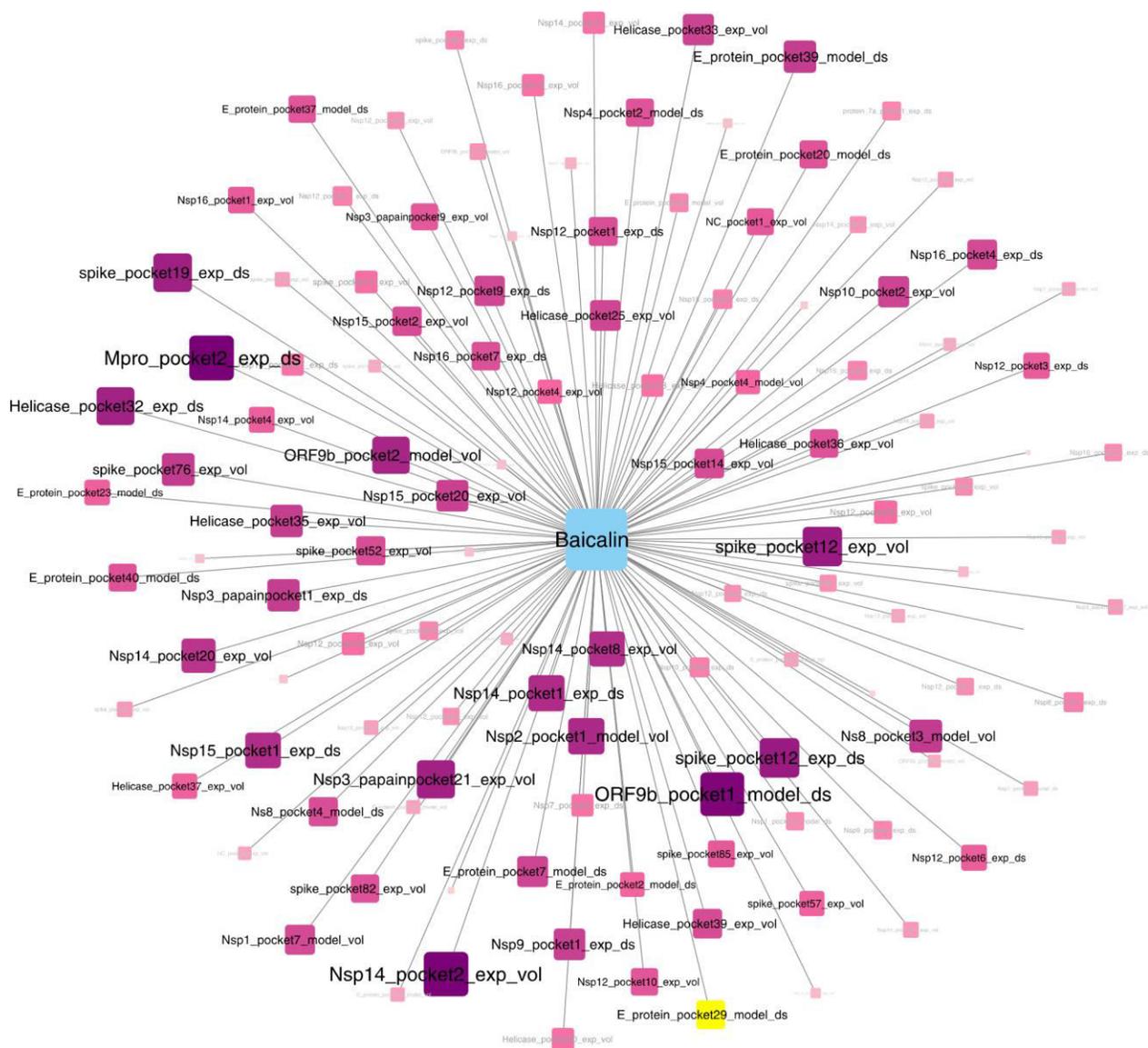
SF: 29. The binding energy interaction network for the SARS-CoV-2 Pockets with Baicalin (centre). The target node size and colour depend on the binding energy scores, ie. greater the affinity, larger the node size; darker the colour.

SF: 30. The binding energy interaction network for the SARS-CoV-2 Pockets with Andrographolide (centre). The target node size and colour depend on the binding energy scores, ie. greater the affinity, larger the node size; darker the colour.

Supplementary Tables

Table ST:1 Top five pockets identified for the phytomolecules reported against influenza virus are listed here with binding energy

	Pockets	Id	Phytomolecules	Binding Energy (kcal/mol)
1	Spike_Pocket10_Exp_Vol	5386591	Ajoene	-6.89
2	Nsp3_Papainpocket1_Exp_Ds	5386591	Ajoene	-6.25
3	Nsp12_Pocket3_Exp_Ds	5386591	Ajoene	-6.21
4	Nsp14_Pocket2_Exp_Vol	5386591	Ajoene	-6.13
5	Nsp3_Papainpocket21_Exp_Vol	5386591	Ajoene	-5.94
6	Spike_Pocket10_Exp_Vol	65036	Allicin	-5.75
7	Nsp3_Papainpocket21_Exp_Vol	65036	Allicin	-4.94
8	Nsp14_Pocket2_Exp_Vol	65036	Allicin	-4.78
9	Nsp3_Papainpocket1_Exp_Ds	65036	Allicin	-4.76
10	Spike_Pocket19_Exp_Ds	65036	Allicin	-4.7
11	Nsp14_Pocket2_Exp_Vol	5318517	Andrographolide	-12.42
12	Helicase_Pocket32_Exp_Ds	5318517	Andrographolide	-11.73
13	Spike_Pocket12_Exp_Ds	5318517	Andrographolide	-11.63
14	Mpro_Pocket2_Exp_Ds	5318517	Andrographolide	-11.62
15	Orf9b_Pocket1_Model_Ds	5318517	Andrographolide	-11.5
16	Mpro_Pocket2_Exp_Ds	64982	Baicalin	-15.02
17	Nsp14_Pocket2_Exp_Vol	64982	Baicalin	-14.97
18	Orf9b_Pocket1_Model_Ds	64982	Baicalin	-14.85
19	Spike_Pocket12_Exp_Ds	64982	Baicalin	-14
20	Spike_Pocket19_Exp_Ds	64982	Baicalin	-13.81
21	Spike_Pocket10_Exp_Vol	10364	Carvacrol	-7.27
22	Nsp3_Papainpocket1_Exp_Ds	10364	Carvacrol	-6.47

23	Spike_Pocket19_Exp_Ds	10364	Carvacrol	-6.43
24	Nsp3_Papainpocket21_Exp_Vol	10364	Carvacrol	-6.35
25	Helicase_Pocket32_Exp_Ds	10364	Carvacrol	-6.17
26	Mpro_Pocket2_Exp_Ds	9064	Catechin	-11.19
27	Nsp14_Pocket2_Exp_Vol	9064	Catechin	-11.1
28	Nsp12_Pocket1_Exp_Ds	9064	Catechin	-10.65
29	Nsp3_Papainpocket1_Exp_Ds	9064	Catechin	-10.52
30	Helicase_Pocket25_Exp_Vol	9064	Catechin	-10.43
31	Spike_Pocket10_Exp_Vol	323	Coumarin	-7.42
32	Nsp3_Papainpocket21_Exp_Vol	323	Coumarin	-6.64
33	Nsp16_Pocket4_Exp_Ds	323	Coumarin	-6.48
34	Spike_Pocket19_Exp_Ds	323	Coumarin	-6.47
35	Nsp3_Papainpocket1_Exp_Ds	323	Coumarin	-6.3
36	Helicase_Pocket32_Exp_Ds	969516	Curcumin	-12.56
37	Orf9b_Pocket1_Model_Ds	969516	Curcumin	-11.33
38	Spike_Pocket12_Exp_Ds	969516	Curcumin	-11.12
39	Spike_Pocket12_Exp_Vol	969516	Curcumin	-11.12
40	Mpro_Pocket2_Exp_Ds	969516	Curcumin	-11.06
41	Spike_Pocket10_Exp_Vol	3314	Eugenol	-7.71
42	Nsp3_Papainpocket21_Exp_Vol	3314	Eugenol	-6.99
43	Nsp14_Pocket2_Exp_Vol	3314	Eugenol	-6.52
44	Nsp16_Pocket4_Exp_Ds	3314	Eugenol	-6.52
45	Nsp14_Pocket8_Exp_Vol	3314	Eugenol	-6.47
46	Spike_Pocket10_Exp_Vol	1254	Menthol	-7.09
47	Nsp3_Papainpocket1_Exp_Ds	1254	Menthol	-6.56
48	Nsp14_Pocket2_Exp_Vol	1254	Menthol	-6.53

49	Nsp3_Papainpocket21_Exp_Vol	1254	Menthol	-6.51
50	Spike_Pocket19_Exp_Ds	1254	Menthol	-6.48
51	Spike_Pocket19_Exp_Ds	1.35e+08	Theaflavin	-16.51
52	Spike_Pocket12_Exp_Ds	1.35e+08	Theaflavin	-16.22
53	Nsp14_Pocket2_Exp_Vol	1.35e+08	Theaflavin	-15.97
54	Mpro_Pocket2_Exp_Ds	1.35e+08	Theaflavin	-15.74
55	E_Protein_Pocket29_Model_Ds	1.35e+08	Theaflavin	-15.22
56	Nsp3_Papainpocket21_Exp_Vol	42607646	Tinosporinone	-10.3
57	Orf9b_Pocket1_Model_Ds	42607646	Tinosporinone	-10.13
58	Nsp14_Pocket2_Exp_Vol	42607646	Tinosporinone	-10.06
59	Helicase_Pocket35_Exp_Vol	42607646	Tinosporinone	-9.81
60	Nsp3_Papainpocket1_Exp_Ds	42607646	Tinosporinone	-9.7
61	Nsp14_Pocket2_Exp_Vol	64945	Ursolicacid	-11.85
62	Nsp16_Pocket4_Exp_Ds	64945	Ursolicacid	-10.77
63	Mpro_Pocket2_Exp_Ds	64945	Ursolicacid	-10.57
64	Nsp3_Papainpocket21_Exp_Vol	64945	Ursolicacid	-10.56
65	Spike_Pocket12_Exp_Ds	64945	Ursolicacid	-10.03

Table ST:2. Top pockets identified for the essential oil individual component from selected Umbelliferae and Labiatae plants compounds are listed here with binding energy

	Compunds	Pockets	Binding Energy (kcal/mol)
1	(-)-isopulegol	spike_pocket10_exp_vol	-6.87
2	2-decanol	spike_pocket10_exp_vol	-6.29
3	2-heptanol	spike_pocket10_exp_vol	-5.46
4	Allylanisole	spike_pocket10_exp_vol	-6.56
5	Alpha-pinene	Nsp14_pocket2_exp_vol	-6.26
6	Anethole	spike_pocket10_exp_vol	-6.79
7	Beta-pinene	Nsp14_pocket2_exp_vol	-6.29
8	Borneol	Nsp14_pocket2_exp_vol	-6.82
9	Bornylacetate	Nsp14_pocket2_exp_vol	-7.34
10	Camphene	Nsp14_pocket2_exp_vol	-6.2
11	Carvone	spike_pocket10_exp_vol	-7.42
12	Citral	spike_pocket10_exp_vol	-7.01
13	Citronellal	spike_pocket10_exp_vol	-6.81
14	Citronellol	spike_pocket10_exp_vol	-6.85
15	Cuminol	spike_pocket10_exp_vol	-7.35
16	Cuminyl_alcohol	spike_pocket10_exp_vol	-7.35
17	Cuminyl_aldehyde	spike_pocket10_exp_vol	-7.18
18	D-limonene	spike_pocket10_exp_vol	-6.97
19	Dihydrocarvone	spike_pocket10_exp_vol	-7.42
20	Farnesol	spike_pocket10_exp_vol	-8.46
21	Gamma-terpinene	spike_pocket10_exp_vol	-7.12
22	Geraniol	spike_pocket10_exp_vol	-6.62
23	Geranyl ester	Helicase_pocket32_exp_ds	-9.51
24	Isoborneol	Nsp14_pocket2_exp_vol	-6.82
25	Isomenthone	Nsp16_pocket4_exp_ds	-6.37
26	Linalyl oxide	spike_pocket10_exp_vol	-8.31
27	Methylheptane	spike_pocket10_exp_vol	-5.16
28	Nerol	spike_pocket10_exp_vol	-7.01
29	Piperitone	Nsp16_pocket4_exp_ds	-6.54
30	Vanillin	spike_pocket10_exp_vol	-7.21

	Pockets	Volume	Hydrophobicity Score	Polarity Score	Charge Score	Total Sasa
1	Pocket 1	488.06	30.52	6	30.52	147.52
2	Pocket 2	530.62	25.00	7	25.00	136.90
3	Pocket 3	376.40	44.15	5	44.15	110.82
4	Pocket6	560.15	16.06	7	16.06	175.02
5	Pocket 9	464.84	4.7	13	4.7	162.58
6	Pocket 17	348.04	37.09	8	37.09	123.12
7	Pocket47	443.29	50.62	3	50.62	141.13