

Title: The structure and function analysis of ACE2--the receptor of SARS coronavirus (SARS-CoV2)

Nuerye Ainiwan

Department of Biochemistry and Molecular Biology, George Washington University Medical Center, Washington, DC 20037

nainiwan72@gwu.edu

ABSTRACT

Objective: angiotensin-converting enzyme 2 (ACE2), a member of the RAS (renin-angiotensin system), was found to be a receptor for SARS coronavirus (SARS-CoV2). The study of the structure, function, and spike glycoprotein binding site of ACE2 is particularly important for the development of vaccines and the design of treatment plans. Method: ProtParam, Blast, mEMBOSS, MEGA X, ClustalW, and other bioinformatics methods and tools were performed to predict the sequence characteristics, molecular structure, physical and chemical properties of ACE2 genes and their encoded proteins. The results showed that the ACE2 gene locates on X chromosomes (Xp22.2: 15494520-15602069), and the CDS is 3325bp. Its protein contained 805 amino acid residues. The secondary structure was calculated on the mEMBOSS Garnier program. The residue totals of helix, sheet, turn, and coil are 338, 172, 144, and 151, respectively, and residue percents are 42.8%, 21.8%, 18.3%, 19.1%. The tertiary structure of the ACE2 protein was calculated by SWISS-MODEL systems. ACE2 protein regions (positions 30-41, 82-84, 353-357), which interacted with SARS-CoV spike glycoprotein. Molecular phylogenetic trees of 22 proteins (manually reviewed UniProtKB / Swiss-Prot from different species) were constructed using the MEGA X tool. Sequence Alignments of ACE2 (Accession: Q9BY1), SARS CoV2 (2019-nCoV) receptor-binding domain as well as Crystal structure of SARS CoV2 (2019-nCoV) spike receptor-binding domain which bound with ACE2 were found. Molecular docking analysis of ACE2 and spike glycoprotein were analyzed by PATCHDOCK and ClusPro. VMD and APBS were used to create surface shape and surface charge distribution of ACE2 protein. These studies can provide theoretical references for further research on the properties and mechanisms of human ACE2 protein.

1.INTRODUCTION

At present, the outbreak of new coronavirus disease (COVID-19) in 2019 caused by SARS-CoV2 (Severe Acute Respiratory Syndrome Coronavirus 2) has spread rapidly in China and has developed into an international public health emergency. Since 2003, angiotensin-converting enzyme 2 (ACE2), a member of the RAS (renin-angiotensin system), has received widespread attention again because ACE2 was found to be a receptor for SARS coronavirus (SARS-CoV2). [1][2] Hoffmann M et al. 2020 states that the surface unit (S1) of the spike (S) protein of SARS engages the angiotensin-converting enzyme 2 (ACE2) as the entry receptor and then uses the host serine protease TMPRSS2 for S priming, allowing the fusion of viral and cellular membranes and viral entry into the cell. Researchers have now examined how the S protein from SARS-CoV2 facilitates viral entry into target host cells and compares the process to that used by SARS. [3]At the same time, some studies have pointed out that as a treatment for hypertension, diabetes, and heart failure, angiotensin-converting enzyme (ACE) inhibitors results in an upregulation of ACE2 in some organs.[4] ACE2 is expressed by epithelial cells of the lung, intestine, kidney, and blood vessels. This increased expression of ACE2 may facilitate infection with SARS-CoV or SARS-CoV2. [5] The study of the structure, function, spike glycoprotein binding site of ACE2, and proteins that can direct (physical) or indirect (functional) associate with ACE2 is particularly crucial for the development of vaccines and the design of treatment plans.

2.MATERIALS AND METHODS

The data were derived from the NCBI Nucleotide and Protein Database (GenBank), the nucleotide sequence of the human ACE2 gene, and its corresponding amino acid sequence in manually curated UniProtKB / Swiss-Prot: ACE2 (Accession: Q9BYF1). Prediction and analysis were studied by using ProtParam, BLAST, mEMBOSS, MEGA X, ClustalW, PATCHDOCK, ClusPro, VMD, and other software and online analysis tools provided by www.ncbi.nlm.nih.gov, www.expasy.org, and other websites. The molecular structure and physical and chemical properties of the nucleotide and amino acid sequences were analyzed using ProtParam. Based on the amino acid sequence of ACE2 in human and other species, the molecular phylogenetic tree of ACE2 protein was constructed using the MEGA X tool. Protein sequences were aligned by ClustalW, and the Neighbor-Joining (NJ) method was used. [5][6]The prediction and modeling of the secondary and tertiary structure of the ACE2 protein were calculated on the mEMBOSS and SWISS-MODEL systems, respectively. STRING (Protein-protein interaction

databases) was used to analyze ACE2 protein interactions with other proteins. Molecular docking analysis of ACE2 and spike glycoprotein were analyzed by PATCHDOCK and ClusPro. VMD and APBS were used to create surface shape, and surface charge distribution of ACE2 protein(PDB Entry:1R42), spike glycoprotein (PDB Entry:6LVN). pdb files were downloaded from the Protein data bank.

RESULTS AND DISCUSSION

Functional analysis

The human ACE2 gene is located on X chromosomes (Xp22.2: 15494520-15602069), and the CDS is 3325bp. There are 21 exons in the genome. The protein encoded by this gene belongs to the angiotensin-converting enzyme family of dipeptidyl carboxypeptidases and has considerable homology to human angiotensin 1 converting enzyme. The sequence ACE2_HUMAN consists of 805 amino acids. This secreted protein catalyzes the cleavage of angiotensin I into angiotensin 1-9[7] and angiotensin II into the vasodilator angiotensin 1-7. The organ- and cell-specific expression of this gene suggests that it may play a role in the regulation of cardiovascular and renal function. [8][9] It is a functional receptor for the spike glycoprotein of the human coronavirus HCoV-NL63 and the human severe acute respiratory syndrome coronaviruses--SARS-CoV [10] and SARS-CoV2. [11] STRING, which is a database of known and predicted protein-protein interactions. It revealed that 21 proteins could direct (physical) or indirect (functional) associate with ACE2. These proteins involved in the regulation of systemic arterial blood pressure, inflammatory response, platelet activation, viral entry into the host cell, etc.(Figure 1). ACE was found in 21 proteins that interact with ACE2. According to studies, ACE2 expression was significantly increased in patients with hypertension and type 1 or type 2 diabetes treated with ACE inhibitors and angiotensin type II receptor blockers (ARB). Lei Fang et al. (2020) hypothesized that the use of ACE2 stimulants to treat diabetes and hypertension increases the risk of developing severe and fatal COVID-19.[12]

To date, there is no reliable animal model of SARS-Cov2. Indirect studies of animal models can consciously change factors that are impossible or difficult to exclude under natural conditions. It can also help to more easily and effectively recognize the law of occurrence and development of human diseases, research, and prevention measures. Phylogenies are important for addressing relationships among species or genes, and the origin and spread of viral infection. In this paper, human ACE2 (Accession: Q9BYF1) BLAST against UniProtKB, and 22 manually reviewed proteins were found. The MEGA X tools, ClustalW and Neighbor-Joining (NJ) method were

used to construct a molecular phylogenetic tree of 22 proteins. The phylogenetic tree shows that human ACE2 (Accession: Q9BYF1) and Pongo abelii ACE2 (Accession: Q5RFN1) share more monophyletic groups (Figure 2). In summary, understanding the protein properties, prediction of their possible interacting proteins, and molecular phylogenetics are all essential for the research and treatment of COVID-19.

Structural analysis

The structure of protein shows the first step in a viral infection, which can help discover treatment and lead to the development of antibodies that block this interaction. According to Wrapp D and others, the SARS-CoV2 spike protein has two vital elements for infecting human cells. A string of amino acids in the S1 subunit directly binds to a protein-cleaving portion of ACE2 called a peptidase domain. The S2 subunit of spike protein helps the virus fuse to human cells. [13] [14] In this paper, the prediction and modeling of secondary objects were calculated according to the mEMBOSS Garnier program (Table 1). The total residues of the helix, sheet, turn, and the coil were 338, 172, 144, and 151, respectively, and the residue percentages were 42.8%, 21.8%, 18.3%, and 19.1%. The tertiary structure of the ACE2 protein was calculated using the SWISS-MODEL system. ACE2 protein and regions interacted with SARS-CoV spike glycoprotein (positions 30-41, 82-84, 353-357) (Table 2). Sequence alignment of angiotensin-converting enzyme 2 (Accession: Q9BYF1) and SARS CoV2 (2019-nCoV) chimeric receptor-binding domain (Figure 3), as well as experimental crystal structure of the SARS CoV2 (2019-nCoV) spike receptor-binding domain that binds to ACE2 were found (Figure 4).

The surface properties of proteins are one of the important components of studying protein structure. These include surface shape, electrostatic potential, and solvent accessibility. That is, the degree of residue contact with the solvent, where is buried inside, where is exposed on the surface, where is between buried and Intermediate state between exposures. Interactions of viral attachment proteins with protein receptor molecules are mostly determined by complementarity in surface charge distribution, hydrophobic interactions, and geometry; typically, carbohydrates are excluded from the binding sites.[15] Pravakaran et al. state, “a prominent feature of the model is a deep channel on the top of the molecule that contains the catalytic site. Negatively charged ridges surrounding the channel may provide a possible binding site for the positively charged receptor-binding domain (RBD) of the S-glycoprotein.” [16] In an earlier report, Han et al. performed a study to identify the critical determinants on hACE2 for SARS-CoV entry, and they found that two natural peptides from hACE2 (a.a. 22-44 and 22-57)

exhibited a modest antiviral activity to inhibit the binding of SARS-CoV RBD to hACE2.[17] In this paper, VMD and APBS are used to create surface shape and electrostatic potential of ACE2 protein (PDB Entry:1R42) Figure(5).

Furthermore, a molecular docking analysis of ACE2 and spike glycoprotein was generated. Usually, protein-protein molecule docking attempts all possible combinations and ranks each form according to a scoring function. In the process of docking, the following factors will be considered: complementary shape, hydrophobicity, and surface charge distribution. There are usually two types of protein-protein docking, one is rigid docking, and the other is flexible docking. Most of the software currently available is rigid docking. The output values are mostly in multiple docking states, sorted according to energy level, and the low energy is ranked first. [18] In this paper, PATCHDOCK software was used. This software is a molecular docking algorithm based on shape complementarity principles. As a result, molecular docking analysis of ACE2 protein (PDB Entry:1R42) and spike glycoprotein (PDB Entry:6LVN), was generated and the results were shown by using VMD Figure (6A.B). The ClusPro program was also used for the computational docking of protein structures. The coordinate files of ACE2 protein (PDB Entry:1R42) and spike glycoprotein (PDB Entry:6LVN) were uploaded to the ClusPro. The docking algorithms evaluated putative complexes, retaining a preset number with favorable surface complementarities. A filtering method was then applied to this set of structures, selecting those with good electrostatic and desolvation free energies for further clustering. The program output was a shortlist of putative complexes ranked according to their clustering properties. ClusPro has four types of models using the scoring schemes called (1) balanced, (2) electrostatic-favored, (3) hydrophobic-favored, and (4) van der Waals + electrostatics.[19] The top four protein interactions were listed by using these four types of models. (Table 3). The top four protein interactions were shown in Figure 7A.B.C.D. Based on that analysis, it is possible to develop new therapeutics to block SARS-CoV-2 from binding to ACE2.

In summary, this study analyzed the function, structure, mechanism, and phylogenetic tree of human ACE2 protein. However, many issues remain unresolved. First, do zinc metalloproteinase receptors generally have some advantages in allowing coronaviruses to enter cells? Second, does ACE2 cleave the SARSCoV2 protein on the surface of the virion and may produce bioactive peptides that contribute to the pathogenesis of SARS-CoV2? Third, is ACE2 the only functional receptor that allows SARSCoV2 to enter, or does it involve co-receptors? Fourth, does the inflammatory response to SARS-CoV2 cause the upregulation of ACE2 expression in lung tissue? Therefore, further research on SARS Cov2 is still needed.

References:

1. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. (March 2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Science China. Life Sciences*. 63 (3): 457–460. doi:10.1007/s11427-020-1637-5.
2. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veersler D (2020). Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. *bioRxiv*: 2020.02.19.956581. doi:10.1101/2020.02.19.956581.
3. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S.,... (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020 Mar 5;doi.org/10.1016/j.cell.2020.02.052
4. Li XC, Zhang J, Zhuo JL (2017). The vasoprotective axes of the renin-angiotensin system: Physiological relevance and therapeutic implications in cardiovascular, hypertensive and kidney diseases. *Pharmacological Research*. 125 (Pt A): 21–38. doi:10.1016/j.phrs.2017.06.005.
5. Fang L, Karakiulakis G, Roth M (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?. *The Lancet. Respiratory Medicine*. doi:10.1016/S2213-2600(20)30116-8.
6. Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.
7. Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.
8. Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, et al. (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. *Circulation Research*. 87 (5): E1-9. doi:10.1161/01.res.87.5.e1.
9. Keidar S, Kaplan M, Gamliel-Lazarovich A (2007). ACE2 of the heart: From angiotensin I to angiotensin (1-7). *Cardiovascular Research*. 73 (3): 463–9. doi:10.1016/j.cardiores.2006.09.006.
10. Wang W, McKinnie SM, Farhan M, Paul M, McDonald T, McLean B, et al. (2016). Angiotensin-Converting Enzyme 2 Metabolizes and Partially Inactivates Pyr-Apelin-13 and Apelin-17: Physiological Effects in the Cardiovascular System. *Hypertension*. 68 (2): 365–77. doi:10.1161/HYPERTENSIONAHA.115.06892.

11. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., ... & Choe, H. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965), 450-454.
12. Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?. *The Lancet Respiratory Medicine*. doi.org/10.1016/S2213-2600(20)30116-8
13. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020;S0092-8674(20)30262-2. doi:10.1016/j.cell.2020.02.058
14. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263.
15. D.S. Dimitrov, Virus entry: molecular mechanisms and biomedical applications, Nat. Rev. Microbiol., 2004 (in press)
16. Prabakaran, P., Xiao, X., & Dimitrov, D. S. (2004). A model of the ACE2 structure and function as a SARS-CoV receptor. *Biochemical and biophysical research communications*, 314(1), 235–241.
17. Han, D. P.; Penn-Nicholson, A.; Cho, M. W., Identification of critical determinants on ACE2 for SARS-CoV entry and development of a potent entry inhibitor. *Virology* 2006, 350, 15-25.
18. Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An integrated suite of fast docking algorithms. *Proteins*. 2010 Nov 15;78(15):3197-204.
19. Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Kozakov D. New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*. 2017 Mar; 85(3):435-444.

Figure Legends

Figure 1. The network of the interactions of secreted proteins. It reveals that 21 proteins can direct(physical) or indirect(functional) associate with ACE2. The multiplicity of lines connecting individual proteins indicates the level of evidence for the interactions. A green line - neighborhood evidence; a blue line - concurrence evidence; a purple line - experimental evidence; a light blue line - database evidence; a black line – co-expression evidence

Figure 2. Molecular phylogenetic trees of 22 proteins. Human ACE2 (Accession: Q9BYF1) and Pongo abelii ACE2 (Accession: Q5RFN1) share more monophyletic groups

Figure 3. Sequence Alignments of angiotensin-converting enzyme 2 Q9BYF1 and SARS CoV2(2019-nCoV)receptor-binding domain.

Figure 4 Crystal structure of SARS CoV2(2019-nCoV) spike receptor-binding domain bound with ACE2

Figure 5. Surface shape and electrostatic potential of ACE2 protein(PDB Entry 1R42). It was created by using VMD and APBS.

Figure 6(A.B). Molecular docking analysis of ACE2 protein(PDB Entry 6LVN) and spike glycoprotein (PDB Entry:6LVN), and the results are viewed using VMD. A:Material(Opaque)
B: Material(Transparent)

Figure 7(A.B.C.D). The top 4 protein interactions are shown by using ClusPro. (A) balanced, (B) electrostatic-favored, (C) hydrophobic-favored, and (D) van der Waals + electrostatics.

Table

Table1: The ACE2 secondary structure calculated on the mEMBOSS garnier program

Residue	Helix(n)	Sheet(n)	Turn(n)	Coil(n)
Residue totals	338	172	144	151
Residue percent(%)	42.8	21.8	18.3	19.1

Table2: Features table for ACE2 protein and regions which interaction with SARS-CoV spike glycoprotein

Feature key	Position(s)	Function and description
Signal	1-17	Signal
Chain	18-805	Angiotensin-converting enzyme 2
Chain	18-708	Processed angiotensin-converting enzyme 2
Tope_dom	18-740	Extracellular
Transmem	741-761	Helical
Tope_dom	762-805	Cytoplasmic
Region	30-41	Interaction with SARS-CoV spike glycoprotein
Region	82-84	Interaction with SARS-CoV spike glycoprotein
Region	353-357	Interaction with SARS-CoV spike glycoprotein
Region	652-359	Essential for cleavage by ADAM17
Region	697-716	Essential for cleavage by TMPRSS11D and TMPRSS2

Table3: The top 4 proteins interactions were list by four types of models using the scoring schemes called (1) balanced, (2) electrostatic-favored, (3) hydrophobic-favored, and (4) van der Waals + electrostatics.

	Balanced		Electrostatic-favored		Hydrophobic-favored		<u>VdW+Elec</u>	
	Members	Weighted Score(lowest Energy)	Members	Weighted Score(lowest Energy)	Members	Weighted Score(lowest Energy)	Members	Weighted Score(lowest Energy)
Cluster0	58	-696.1	74	-688.2	75	-898.2	63	-179.4
Cluster1	56	-635.6	59	-705.7	66	-739.9	44	-176.6
Cluster2	45	-747.0	41	-689.4	50	-751.7	43	-199.5
Cluster3	35	-649.9	40	-609.5	47	-782.5	42	-180.3

Figure 1

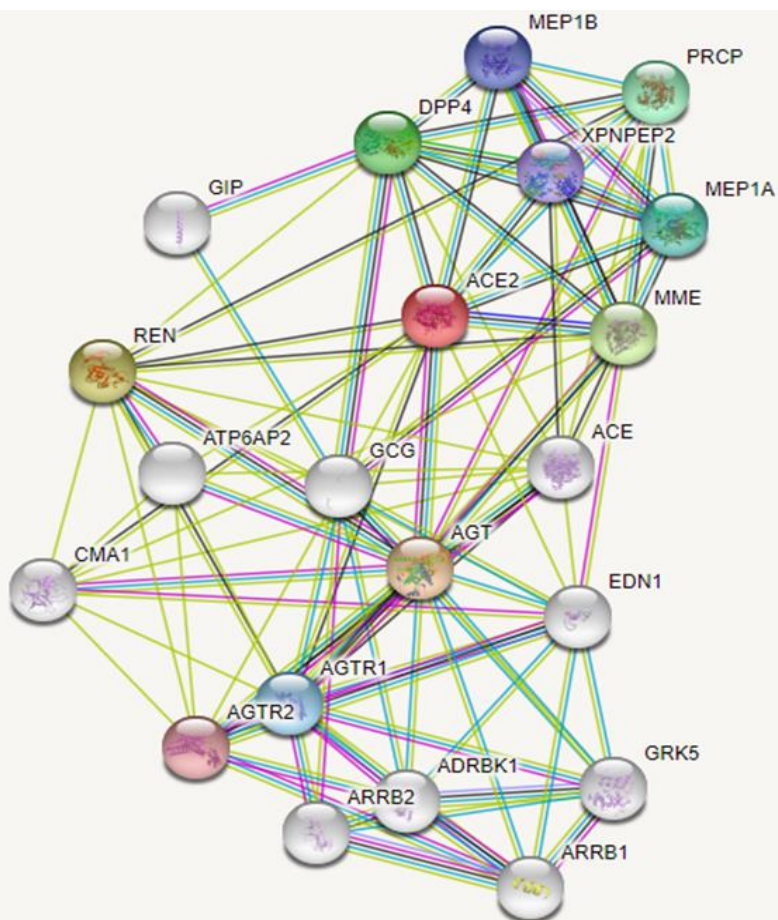


Figure 2

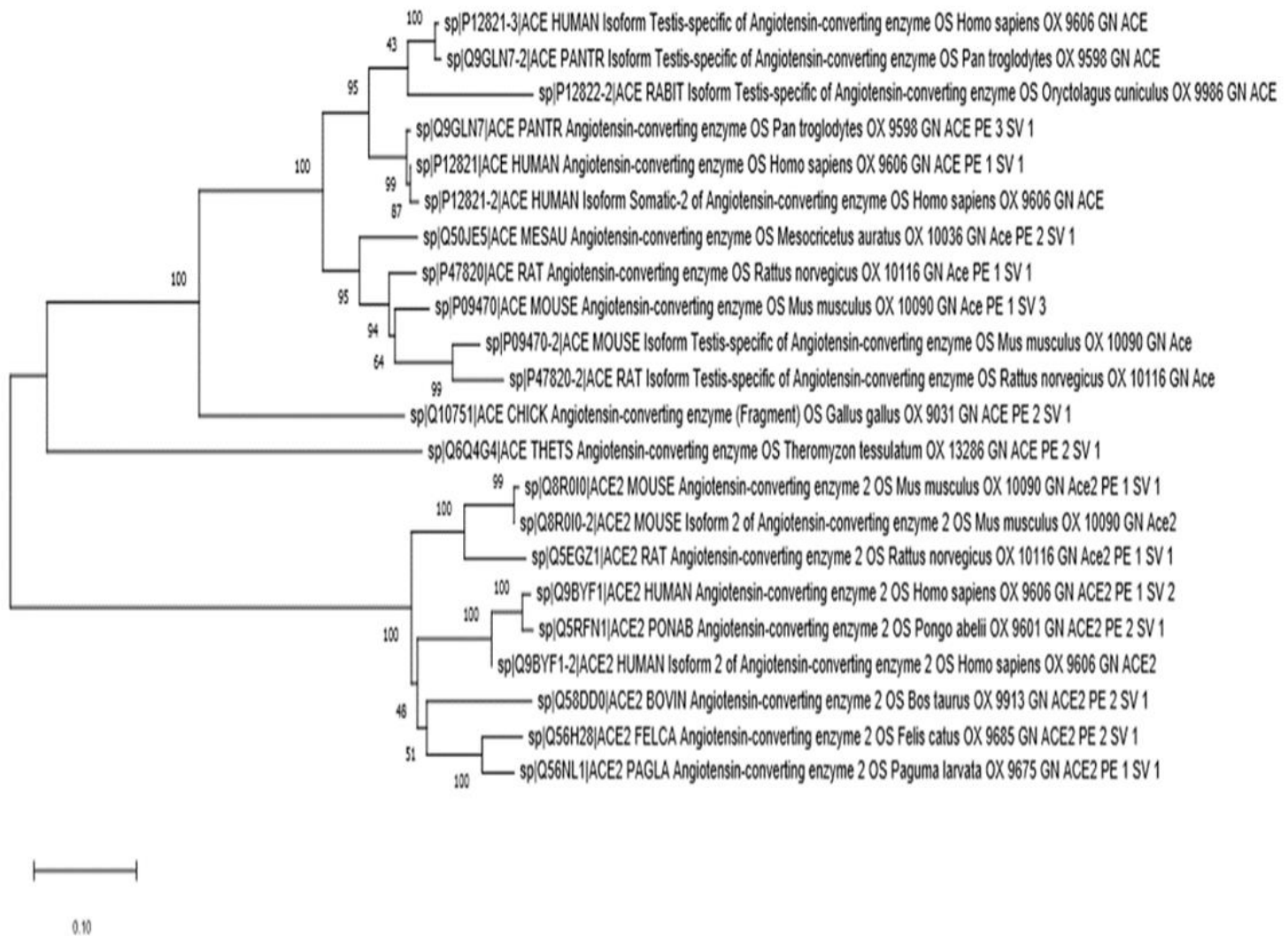


Figure 3

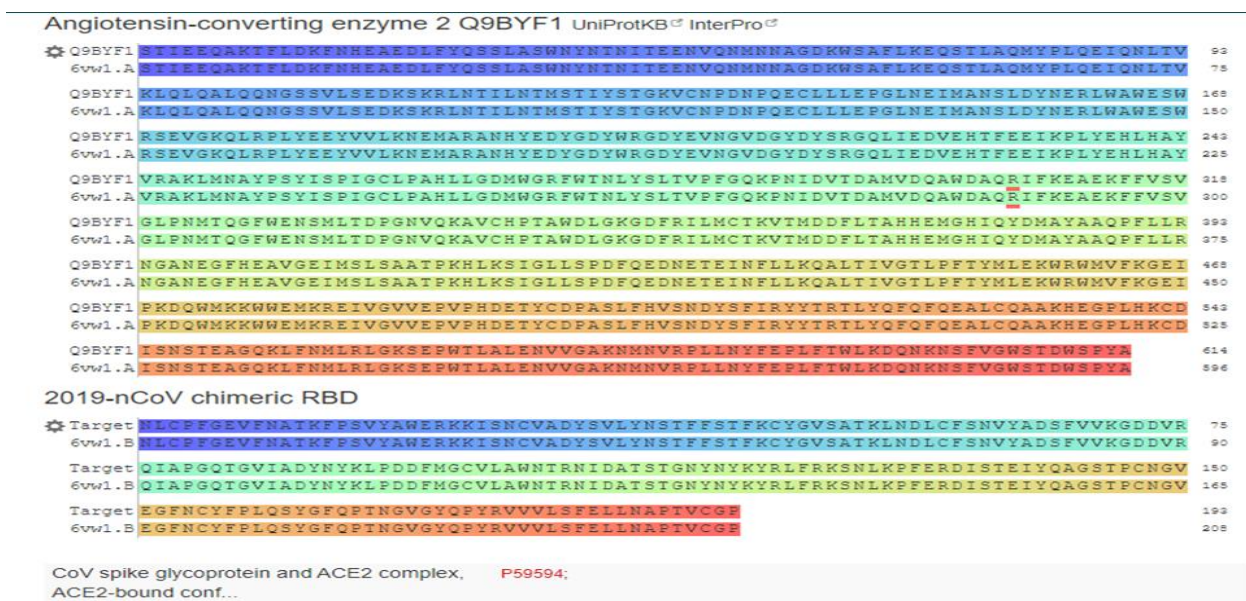


Figure 4

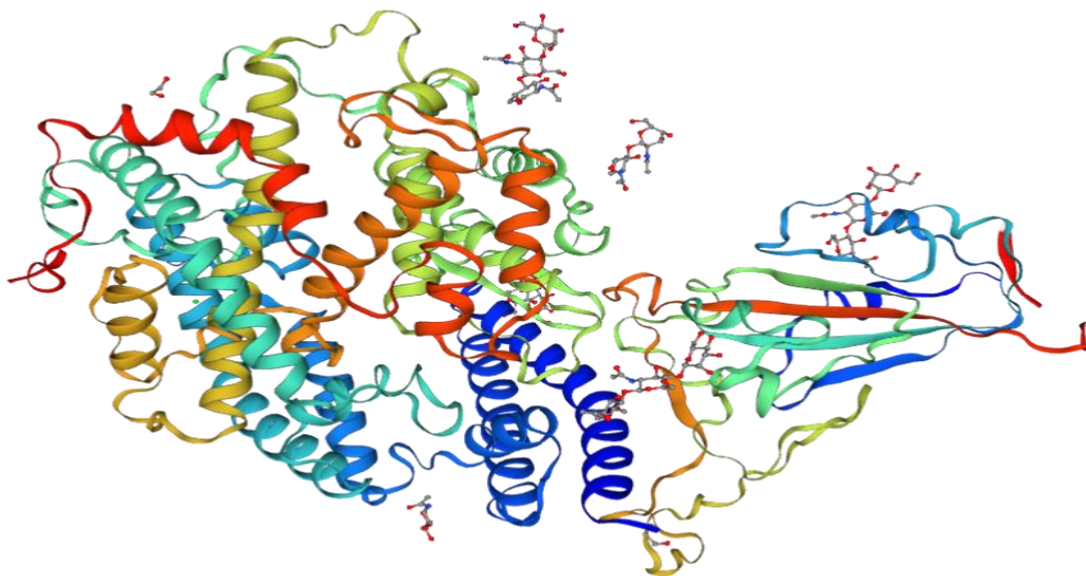


Figure 5

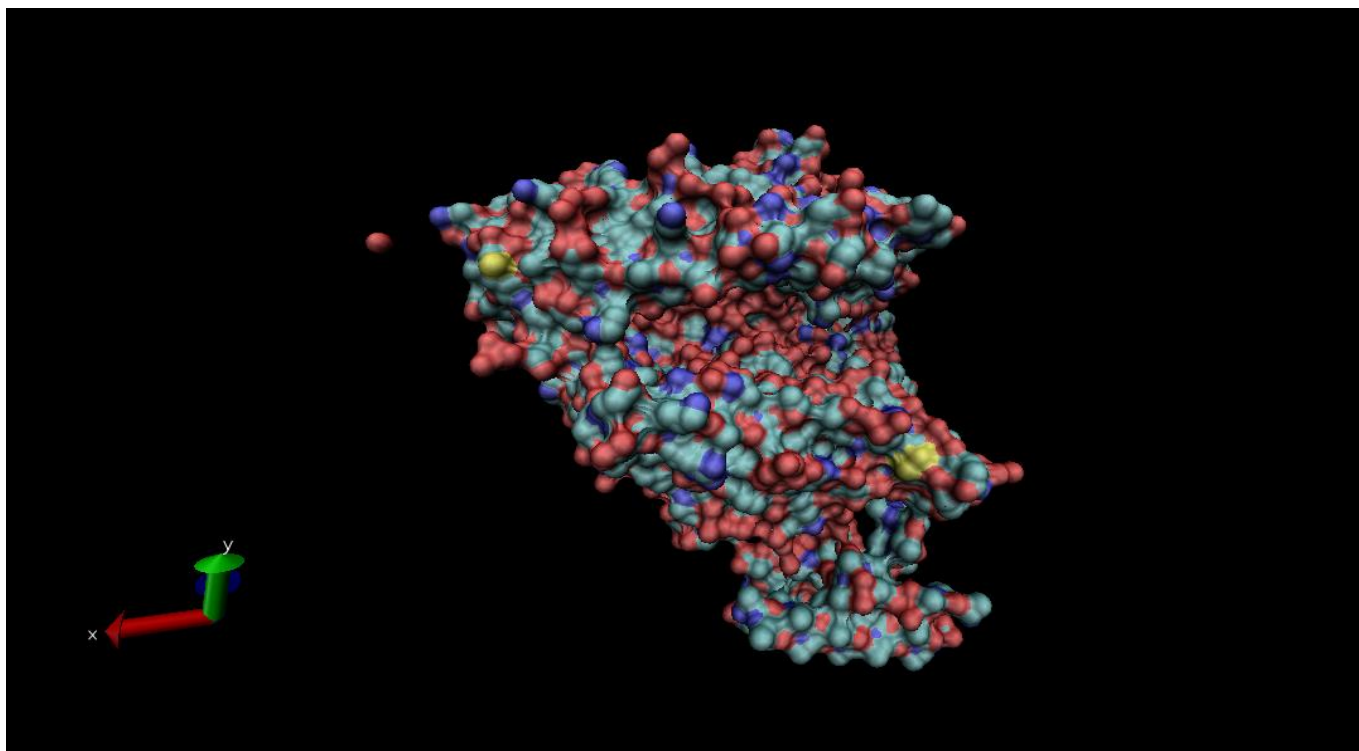
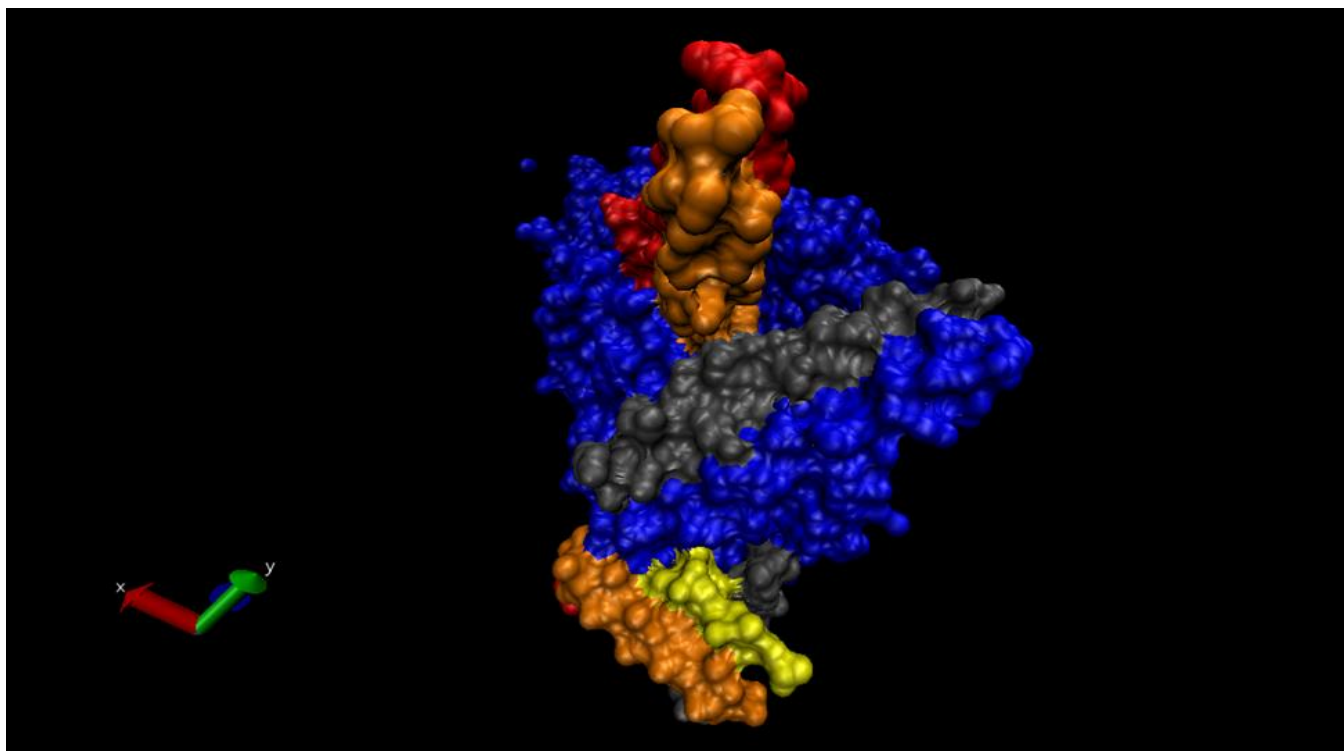


Figure 6

A:



B:

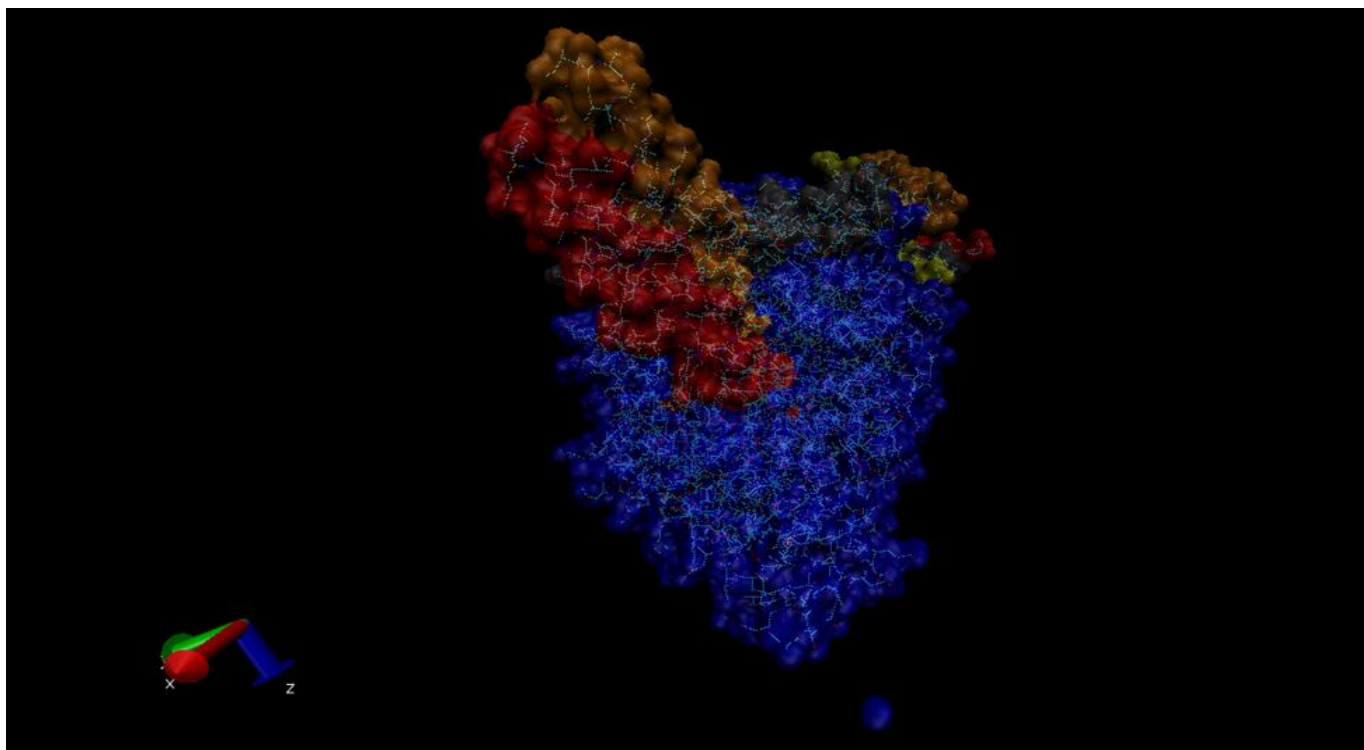
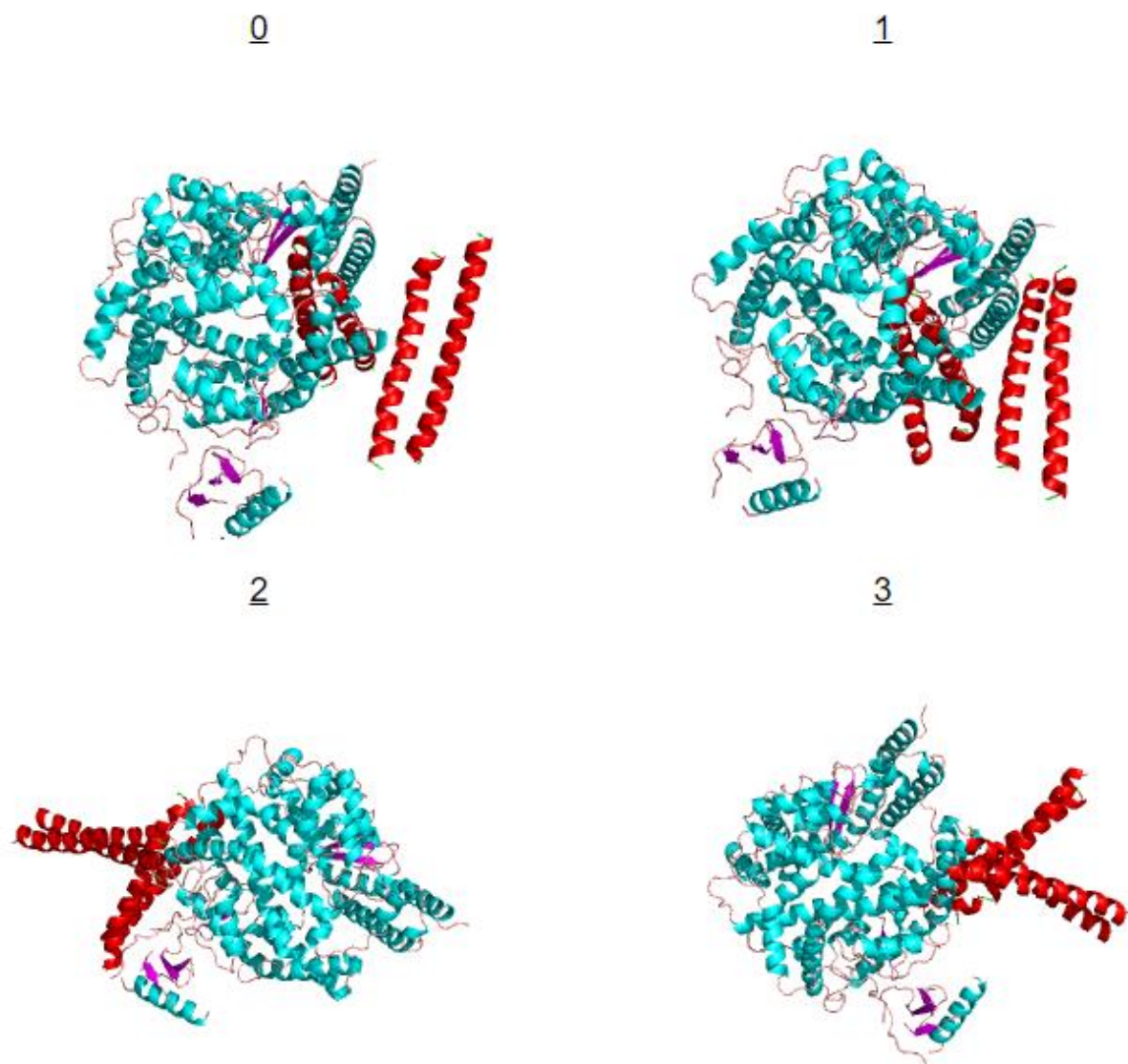
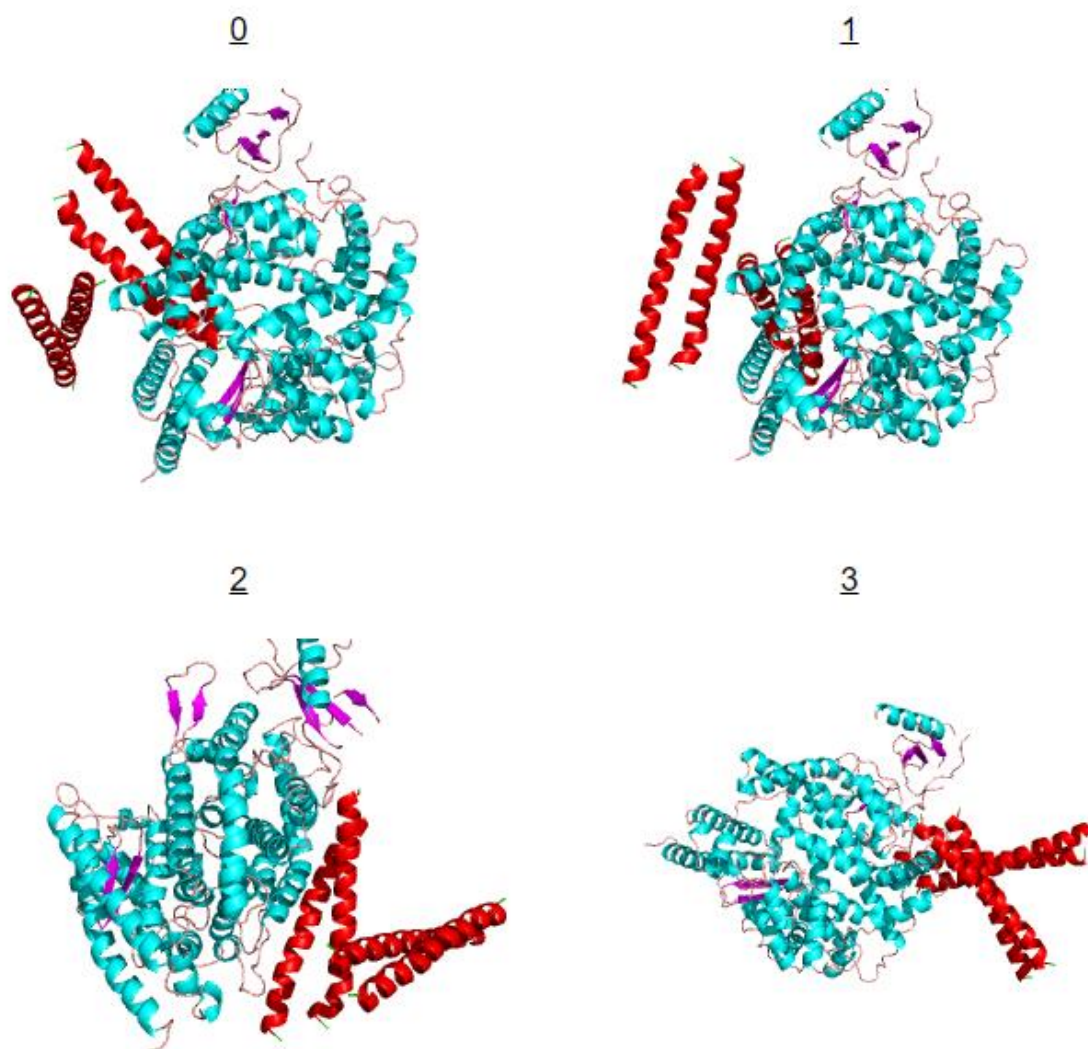


Figure 6 (A) balanced

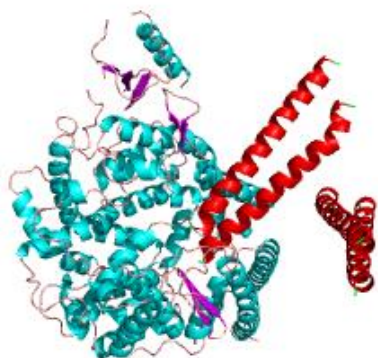


(B) electrostatic-favored

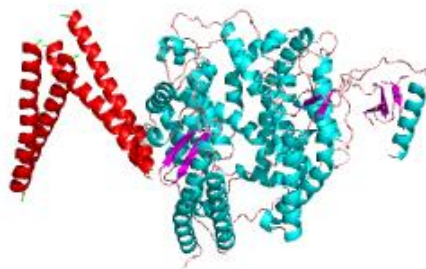


(C) hydrophobic-favored

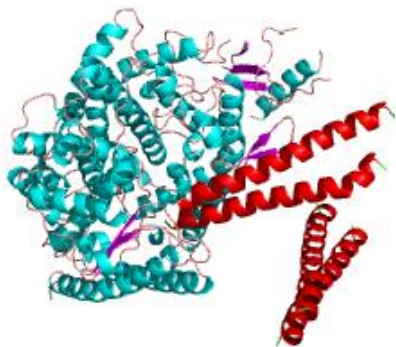
0



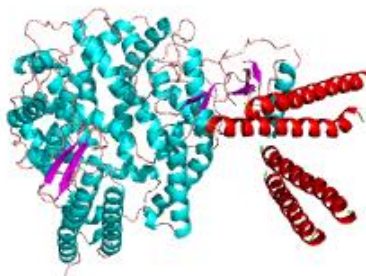
1



2



3

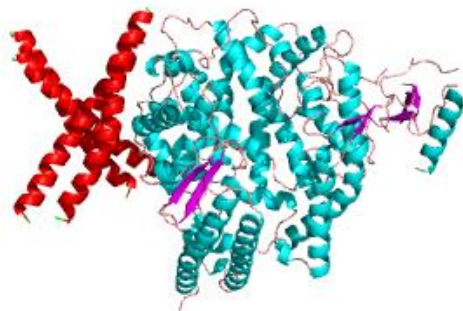


(D)) van der Waals + electrostatics

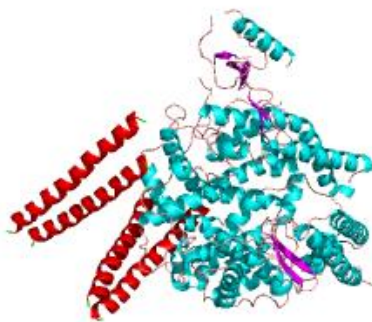
0



1



2



3

