SARS-CoV-2 Disulfide-Bond Stabilized S Protein

Thinh B. Vo, John Bowry, Giuliana Gordon, Daanish Fiaz, Nathaniel Smith, Owais A. Shahzada

[1] Center For The Study of Biological Complexity, Virginia Commonwealth University, Richmond VA, United States of America

Abstract

The emergence of SARS-CoV-2 has resulted in over 800,000 infections and over 40,000 deaths. Coronavirus enters cells through its homo-trimeric surface spike (S) glycoprotein making it the main target for antibodies. Vaccines are the current medical approach to preventing the spread of the Coronavirus. However, vaccines have shown to have poor thermostability. The exposure of vaccines in elevated temperatures have shown to lose its potency. The loss of potency can lead to the large number of infections and ultimately deaths in the population of developing countries. The formation of disulphide bonds in proteins can lead to a more stable protein which will be resistant to thermal-denaturation. We identify and examine potential sites within SARS-CoV-2's S protein complex to construct a disulphide bridge between two peptides. The sites of interest were determined by our program Cysteine Disulfide Substitution Tool (CDSTool) for the best possible position to substitute a cysteine residue. The parameters given to the program predicts the sites that should not have a conformational drastic conformational change of the S protein. The identification of potential sites can serve as a basis to help develop vaccines for long term storage and transportation.

Keywords: S protein, Thermostability, Cysteine, Disulfide-Bond, CDSTool

1. Introduction

1.1 Protein Stability

Proteins that are used in therapeutics are often exposed to environments that may cause protein denaturation, resulting in ineffective therapy. A major issue with protein therapeutics, and a primary cause of under vaccination is inadequate thermal storage (McColloster et al, 2014).

Many common vaccines must be refrigerated between 2°C to 8 °C, and if exposed to elevated temperatures, the vaccine may be less effective, or completely ineffective (Chojnacky et al, 2010). This is a very common occurrence in remote areas, where poor accessibility, limited healthcare, and frequent power outages cause under vaccination and the loss of millions of dollars in vaccine development.

1.2 Engineering Stabilized Proteins

One way to combat this is to re-engineer the protein structure to allow for more stability, resulting in a protein with an increased resistance to denaturation (Marshall et al, 2003). This has been seen in the Fusion protein of Respiratory syncytial virus (RSV) (Sung-Youl et al, 2014), the Fusion protein of Measles virus (Lee et al, 2007), and the Hemagglutinin (HA) protein of Influenza H3N2 virus, where an 11°C increase in melting temperature was observed versus the wildtype using disulfide bond engineering (Lee et al, 2015).

Disulfide bond engineering is one common approach to stabilizing proteins, and involves introducing a pair of cysteines into the amino acid sequence of the protein, creating a disulfide bond. Disulfide bonds are formed between the thiols of two cysteines, and thermodynamically stabilize proteins by decreasing the entropy, according to traditional

polymer theory (Lodge et al, 1996). By engineering more stable proteins, more efficient vaccines may be produced, resulting in higher vaccination rates, and potentially lower rates of disease.

1.3 Stabilized SARS-CoV-2 Spike Protein

The recent pandemic of SARS-CoV-2, and the necessity of an effective vaccine for COVID19 is of utmost concern to the global community. One primary target for vaccine development is the Coronavirus Spike protein, or S protein.

The S protein is the Coronavirus surface glycoprotein responsible for host cell recognition, attachment, and entry. The protein exists as a homo-trimer, meaning the protein structure consists of three identical peptide chains that are linked to one another. There are two subunits of the S protein: the S1 head region, the area of the S protein that binds to the host cell receptor (ACE2), and the S2 stalk region, the area of the S protein that fuses the host and viral membranes together for host cell entry (Li et al, 2016). The function of the S protein is critical in the viral infection cycle, making it a perfect target for vaccines, which often target receptor-binding surface proteins (Li et al, 2016). The objective of this paper is to use bioinformatics tools such as BioPython to develop a program that identifies sites for disulfide bond engineering in the SARS-CoV-2 S protein *in silico* (Cock et al, 2009).

2. Methods

2.1 Overview

The PDBTool provides a method to extract, analyze, and return locations of possible cysteine-cysteine disulfide bridge sites found on protein structures given a PDB file. A key aspect of this tool is that it aims to help researchers in the field of vaccine development to obtain fast and accurate results by leveraging the power of an in-silico based approach.

Briefly, the tool parses out information based on set parameters passed through each function in the selected amino acid chains found in the PDB file. The source code of the tool is written in Python and is available in the CDSTool'sGitHub repository [link to repo]. For further instruction and guide on how to use the module, documentation and API guide can be found on our website https://cdstools.netlify.app/api/. Given a PDB file, the functions of the tool will parse out the information found in the file using the BioPython library and return key information; such as, the atom, amino acid, residue position, distance of amino acids on different/same chains in angstrom (Å), and distance from the residue site to the epitope site (angstrom). The tool can be leveraged for downstream analysis by the end-user integrating the package and functions in Python.

2.2 Compare Function

Chain comparison is implemented through a series of conditionals. The algorithm takes a file from the Protein Data Bank (PDB) as input in PDB format. The pairwise comparison of chains iterates through the first chain's structure in each pair until the target atom, Cβ, is identified in amino acid. The target atom's position is marked in the first chain. The second chain in the pair is similarly iterated through until a Cβ is identified. Distance is then calculated between the two atoms using the distance function. If the distance between two atoms is less than 6.5 plus the resolution of the crystal structure being examined, Angstroms. Analysis of the potential site continues. The amino acid type is checked to avoid conformation change. If the amino acid type is Serine, Threonine, or Alanine amino acids which exhibit the best possible chance of being substituted with cysteine, the closest Cα is searched for in relation to Cβ identified previously in each chain. Distance is calculated in a similar fashion to Cβ, for the identified Cα atoms. The distance between Cα atoms is limited to 4.5 angstroms plus the resolution of the crystal structure. Pairs with a distance shorter than the specified limit are then recorded as a candidate for a potential disulfide bond site. The function continues down the chain until the end of the first chain is reached.

2.3 Distance Function

The purpose of the distance function is given a chain(s), the algorithm will calculate the distance from the epitopes found in the chain(s) to the possible residue site at which the cysteine-cysteine disulfide bridge can be formed. The function is configurable allowing the end-user to decide their parameters. Given a pair of atoms from a PDB file, the function utilizes BioPython to determine the Euclidean distance between the atoms. The function returns a Boolean based on the comparison of calculated distance vs maximum distance, true if the distance is less than or equal to the maximum distance specified and false if greater than the maximum distance. If the chain is being compared to itself an additional check is completed to ensure the two-parent residues, in which the Cβ has been identified, are far enough apart in the sequence of the chain to interact properly, the default value is a gap of ten residues.

3. Results

3.1 Validation

The first validation was performed using data from an article that studied the Influenza virus surface protein, hemagglutinin. This article was also trying to form a disulfide bond and find the ideal site for the bond. The resulting cysteine pair from A/Hong Kong/1/1968 (H3N2) influenza virus hemagglutinin Lee et al (2015) were ~4.4 Å apart (PDB code 4FNK).

When we ran our program with the parameters of a distance <4.5 Å for this viral strain, 3 results were found (**Table 1**). Two between chain A and B . Only one, THR30-GLN47was on the chains examined in the article, chains B and C. This result is identical to the sites found in the article.

We next performed more validation using RSV Fusion Glycoprotein F2 From the work of Mclellan et al (2013). The goal for this article was also the stabilization of the RSV Fusion Glycoprotein F2. In the article, the sites S155C-S290C, and S403C-T420C were used to stabilize the protein(PDB code 4JHW). We ran our program with the self stabilizing setting, and found seven potential sites for a disulfide bond (**Table 2**).

3.2 Disulfide Bridge Analysis in SARS-CoV-2

Structure of the Spike-protein in the SARS-CoV-2 was determined by Cryo-EM Wrapp et al (2020) (PDB code 6VSB). The results in (**Table 3**) are the suggested replacement sites generated from CDSTool with chain A, chain B, and chain C.

The Mean Distance function of CDSTool was then run with the epitope from the work of Yuan et al (2020). A visualization of chain A in SARS-CoV-2 S-Protein with all the suggested sites and their mean distance to the epitope region can be found in (**Figure 1**).

3.3 Negative Control

For the negative control, a model of Influenza A M2 protein transmembrane domain from the work of Kreitler et al (2019) (**Figure 2**) was used. The self stabilizing setting was used. The program failed to suggest a site for cysteine substitution.

4. Discussion

Our results indicate that our program identifies potential amino acid sites to substitute for cysteine to create a disulfide bridge within the surface proteins of SARs-CoV-2 and the influenza virus. There was an optimal distance less than 8.5 Angstroms that we used that was used in the design of a stabilized influenza steam domain (Lu, Welsh, and Swartz, 2013). Ideally, these substitutions should be in positions that are distant from other areas that are important for antigenicity (Lu et al., 2013). The regions of functional importance are the epitopes contained within the S1 head region of SARs-CoV-2. We do have potential sites that are contained within the S1 region of the spike glycoprotein. These results do lie within our expectations, we wanted to identify potential sites to introduce these disulfide bonds that lie within our criteria given including amino acid differences, regions of these potential sites, and then cross validate these results with the works of authors including Lee et al. and Mcllelan et al (2013). There are sites that we identified that are flawed. For example, many of the sites lie within the S1 region of the spike glycoprotein, and some are too far away from the epitope. For this reason, a filtering of sites too far or too close to the epitopes was done. It is important for stabilization design of a surface protein to keep these criteria in mind. In addition, even when the protein of interest is a homotrimer, their PDB file may contain differences between polypeptide chains. In our experience, all the combinations of nearby chains should be analyzed (in this case: AB, AC, BC) to maximize the result.

All in all, our results do partially agree with the works of Lee et al., Mcllelan et al., and Yuan et al. What they add onto their research is that in stabilization design, it is important to keep significant criteria in mind. For example, one can see from our results of SARs-CoV-2 that it can be easy to misidentify potential sites and that these sites must be filtered out. These results should be taken into account when considering how to stabilize surface proteins of viruses like SARs-CoV-2 and influenza. With the overall goal to aid in the design of a potential SARs-CoV-2 vaccine in order to help reduce pandemic potential to prevent crises like the one we are in right now.

From this study, it can be concluded that in theory, there are regions of SARs-CoV-2 S protein in which a disulfide bridge can be constructed to successfully increase thermostability of a vaccine. However, we did not carry out this methodology in a lab-based setting. In order to confirm whether our program truly serves as a viable option for identifying sites, in vivo or in vitro analysis would have to be performed. Our results are valid nonetheless because it was demonstrated that these sites can be identified through bioinformatics techniques and then those results were validated through the works of Lee et al.(2015) and McClellan et al (2013). In terms of our methodology, there is a key point that we would like to point out. For each site, we calculate the mean distance from there to the epitope to ensure its validity. However, the epitope can end up being closer than the mean distance suggested due to the fact that mean is very sensitive to outliers . In other words, the mean distance assumes the epitope is clustered together. We also provide the users with a min and max distance function to calculate how far the substitution site from the interest site.

Further research is needed to establish whether or not these disulfide bridges can be constructed in vivo. Overall, this disulfide bridge examination *in silico* is performed in hopes of being able to increase the thermostability of the vaccine to make it more feasible to transport and store in countries around the world.

References

[1] Chojnacky M, Miller W, Strouse G. Thermal analysis of refrigeration systems used for vaccine storage. National Institute of Standards and Technology Report NISTIR 7753. September 2010. Available from: http://www.nist.gov/pml/div685/grp01/upload/NISTIR7753-Thermal-Analysis-of-a-Dual-zone-Refrigerator-and-Pharmaceutical-Refrigerator-for-Vaccine-Storage.pdf

[2] Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics (Oxford, England), 25(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

[3] Kreitler DF, Yao Z, Steinkruger JD, Mortenson DE, Huang L, Mittal R, Travis BR, Forest KT, Gellman SH. A Hendecad Motif Is Preferred for Heterochiral Coiled-Coil Formation. J Am Chem Soc. 2019 Jan 30;141(4):1583-1592. doi: 10.1021/jacs.8b11246. Epub 2019 Jan 15. PubMed PMID: 30645104; PubMed Central PMCID: PMC6500429.

[4] Lee, J. K., Prussia, A., Snyder, J. P., & Plemper, R. K. (2007). Reversible inhibition of the fusion activity of measles virus F protein by an engineered intersubunit disulfide bridge. Journal of virology, 81(16), 8821–8826. https://doi.org/10.1128/JVI.00754-07

[5] Lee PS, Zhu X, Yu W, Wilson IA. Design and Structure of an Engineered Disulfide-Stabilized Influenza Virus Hemagglutinin Trimer. J Virol. 2015 Jul;89(14):7417-20. doi: 10.1128/JVI.00808-15. Epub 2015 Apr 29. PMID: 25926650; PMCID: PMC4473583.

[6] Li, Fang. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annual Review of Virology, Vol 3, pp237-261, 2016, www.annualreviews.org/doi/full/10.1146/annurev-virology-110615-042301.

[7] Lodge, T. P., Muthukumar, M. (1996). Physical Chemistry of Polymers: Entropy, Interactions, and Dynamics. Journal of Physical Chemistry 100, 13275-13292.

[8] Marshall S, Lazar G, Chrino A, Desjarlais J. (2003). Rational design and engineering of therapeutic proteins. Therapeutic Focus DDT Vol. 8, No. 5.

[9] McColloster, P. J., & Martin-de-Nicolas, A. (2014). Vaccine refrigeration: thinking outside of the box. Human Vaccines & Immunotherapeutics, 10(4):1126-1128. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4896515/

[10] McLellan JS, Chen M, Joyce MG, Sastry M, Stewart-Jones GB, Yang Y, Zhang B, Chen L, Srivatsan S, Zheng A, Zhou T, Graepel KW, Kumar A, Moin S, Boyington JC, Chuang GY, Soto C, Baxa U, Bakker AQ, Spits H, Beaumont T, Zheng Z, Xia N, Ko SY, Todd JP, Rao S, Graham BS, Kwong PD. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. Science. 2013 Nov 1;342(6158):592-8. doi: 10.1126/science.1243283. PubMed PMID: 24179220; PubMed Central PMCID: PMC4461862.

[11] Sung-Youl Ko, John-Paul Todd, Srinivas Rao, Barney S. Graham and Peter D. KwongBoyington, Gwo-Yu Chuang, Cinque Soto, Ulrich Baxa, Arjen Q. Bakker, Hergen Spits, Tim Beaumont, Zizheng Zheng,Zhang, Lei Chen, Sanjay Srivatsan, Anqi Zheng, Tongqing Zhou, Kevin W. Graepel, Azad Kumar, Syed Moin, Jeffrey C. Jason S. McLellan, Man Chen, M. Gordon Joyce, Mallika Sastry, Guillaume B. E. Stewart-Jones, Yongping Yang, Baoshan (2014). Structure-Based Design of a Fusion Glycoprotein Vaccine for Respiratory Syncytial Virus. Science 342 (6158), 592-598. DOI: 10.1126/science.1243283

[12] UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. Protein Sci. 2018 Jan;27(1):14-25.

[13] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260-1263. doi: 10.1126/science.abb2507. Epub 2020 Feb 19. PubMed PMID: 32075877.

[14] Yuan M, Wu NC, Zhu X, Lee CD, So RTY, Lv H, Mok CKP, Wilson IA. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. Science. 2020 Apr 3;. doi: 10.1126/science.abb7269. [Epub ahead of print] PubMed PMID: 32245784.

1)THR28, GLN105,4.361Å	2)THR12, GLN27, 4.726Å	3)THR30-GLN47, 4.468Å

Table 1. Output of CDSTool compared to the work of Lee et al (2015) on A/Hong Kong/1/1968 (H3N2) influenza virus hemagglutinin apart (PDB code 4FNK) *The highlighted show a match to the regions found in the article

1)SER46-THR311, 4.395Å	2)SER41-SER409, 4.011Å	3)SER155-SER290, 4.393Å
4)THR423-SER451, 4.406Å	5)SER425-THR449, 4.249Å	6)SER403-THR420, 3.954Å
7)SER38-THR318, 4.227Å	---	---

Table 2. Output of CDSTool compared to the work of Mclellan et al (2013) on RSV Fusion Glycoprotein F2 protein (PDB code 4JHW) *The highlighted show a match to the regions found in the article

1)THR547-SER982, 7.153Å	2)ALA570-SER967, 6.010Å	3)ALA668-THR866, 5.496Å

4)THR961-SER758, 5.779Å	5)ALA1070-ALA892, 6.354Å	---

Table 3. Output of CDSTool, suggested cysteine substitution site for Spike-protein in the SARS-CoV-2 (PDB code 6VSB)

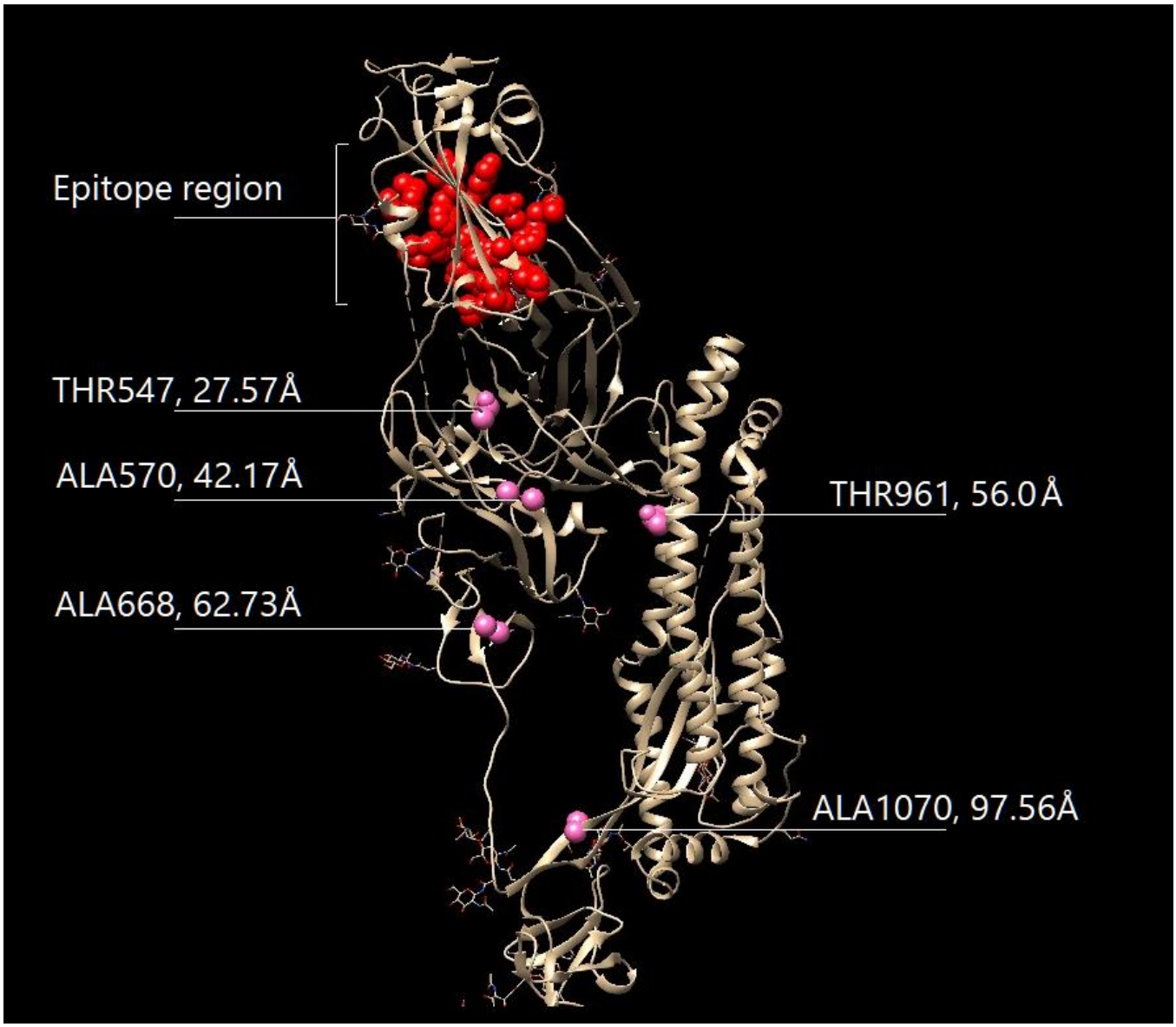

Figure 1. A. SARS-CoV-2 chain A (chain B and C are not shown) is shown here (PDB code 6VSB) Epitope residues are in red (Y369, N370, A372, F374, T376, F377, K378, V382, P384, T385, K386, D389, L390, F392, D428, F429, T430, F515, H517, H519) from Yuan et al (2020) suggested sites for cysteine substitution in pink and their respective mean distances from the epitope region in angstrom (**Å**).

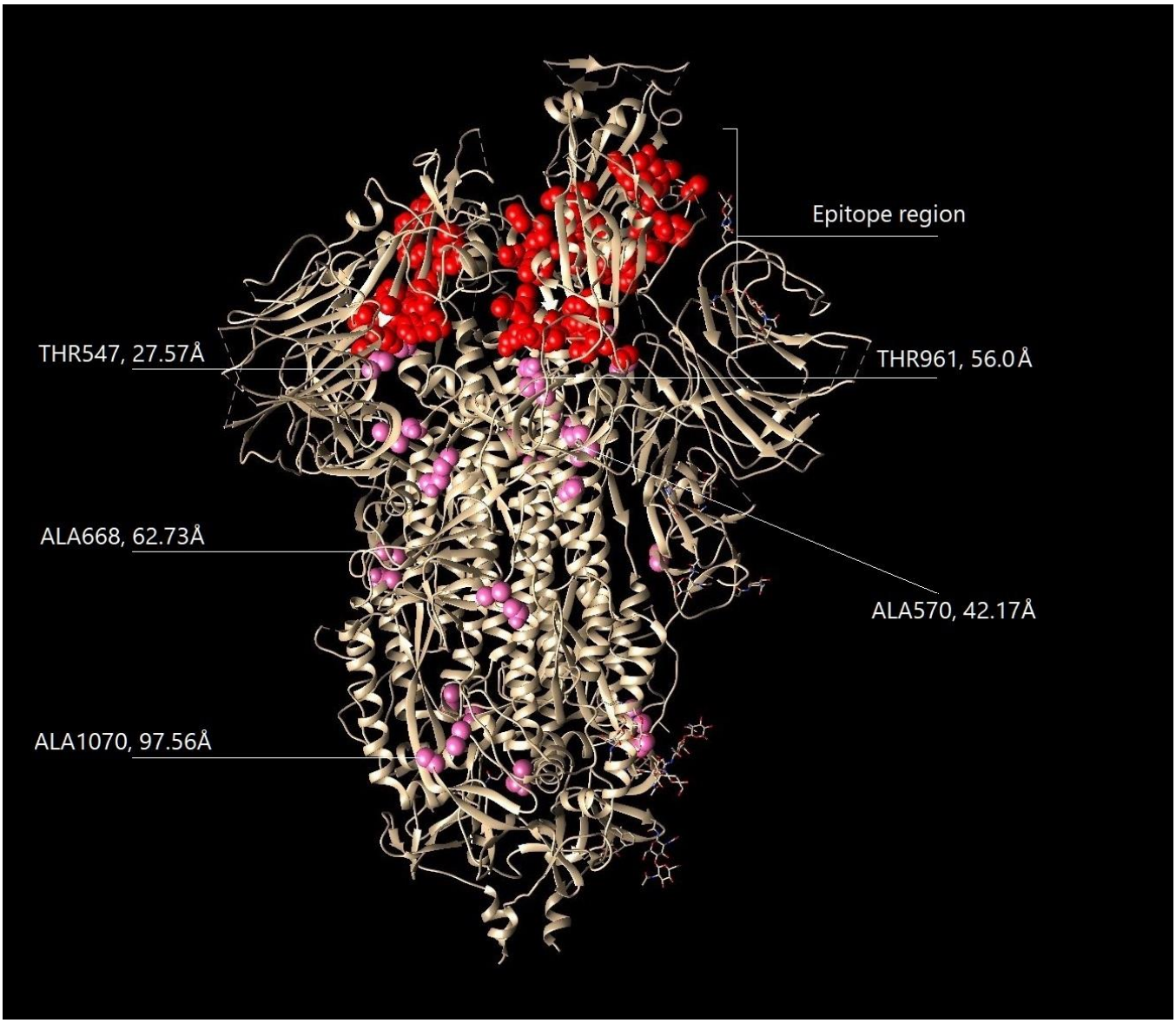

Figure 1.B. SARS-CoV-2 is shown here (PDB code 6VSB) Epitope residues are in red (Y369, N370, A372, F374, T376, F377, K378, V382, P384, T385, K386, D389, L390, F392, D428, F429, T430, F515, H517, H519) from Yuan et al (2020) suggested sites for cysteine substitution in pink and their respective mean distances from the epitope region in angstrom (**Å).**

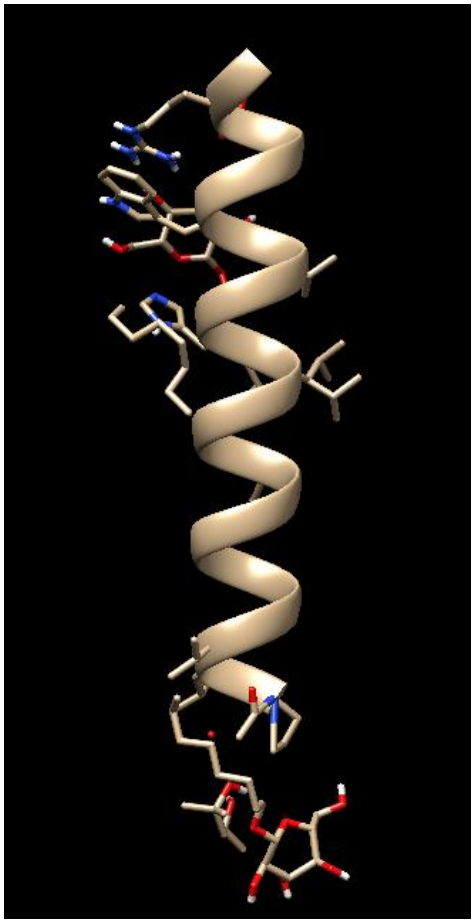

Figure 2. Model of the transmembrane domain region of Influenza A M2 protein (PDB code 6MPL) was chosen to be the negative control. It only shows one chain with a straight helix.