

Genetic Uniformity of a Specific Region in SARS-CoV-2 Genome and *In-Silico* Target-Oriented Repurposing of N-Acetyl-D- Glucosamine

Ömür Baysal^{1*}¶, Ragıp Soner Silme^{2¶}, Çağatay Karaaslan³

¹ Department of Molecular Biology and Genetics, Faculty of Science, Muğla Sıtkı Koçman University, Menteşe-Muğla, Turkey

² Center for Research and Practice in Biotechnology and Genetic Engineering, Istanbul University, Istanbul, Turkey

³ Department of Molecular Biology, Faculty of Science, Hacettepe University, Beytepe Campus, Ankara, Turkey

*Corresponding Author

E-mail: omurbaysal@mu.edu.tr (Ö.B)

¶The authors have equal contribution on paper.

Abstract

Causative agent of the viral pneumonia outbreak in the World identified as SARS-CoV2 leads to a severe respiratory illness like SARS and MERS. The pathogen spreading has turned into a pandemic dissemination and increased the mortality rate. Therefore, useful information is urgently necessary for effective control of the disease. Our study shows the existence of unvarying sequence with no mutation, including ORF1ab regions in 134 high-quality filtered genome sequences of SARS-CoV2 downloaded from the GISAID database. We have detected this sequence stability by using MAUVE analysis and pairwise alignment with Global Needleman Wunsch algorithm for each two different sequences, reciprocally. They also confirmed all these results were also with the Clustal W analysis. The first 6500 bp including ORF1ab region is an unvarying sequence. According to the highest TM-score of predicted

protein structure analysis, the results showed it is very similar to spike protein of feline infectious peritonitis virus strain UU4 (PDB 6JX7) depending on amino acid sequences encoded by this unvarying region, and N-acetyl-D-glucosamine is the ligand of this protein. These results have confirmed that N-acetyl-D-glucosamine could play an important role in controlling of SARS-CoV-2. Also, our molecular docking analysis data supports a strong protein-ligand interaction of N-acetyl-D-glucosamine with spike receptor-binding domain bound with ACE2 (PDB 6M0J) and RNA-binding domain of nucleocapsid phosphoprotein (PDB 6WKP) from SARS CoV-2. Therefore, binding of N-acetyl-D-glucosamine to these proteins could inhibit SARS CoV-2's replication. In the present work, we have suggested providing a repurposing compound for further *in vitro* and *in vivo* studies and new insights for ongoing clinical treatments as a new strategy to control of SARS-CoV-2 infections.

Keywords: SARS-CoV-2; Drug repurposing, N-acetyl-D-glucosamine, ORF1ab, Biodata mining, Protein modelling

Introduction

Coronaviruses (CoVs) are positive-strand RNA viruses belonging to the order of Nidovirales including three families Arteriviridae, Coronaviridae, and Roniviridae [1]. Relied on the genetic studies, they classify CoVs into four genera as alpha, beta, gamma, and delta CoVs. The diameter of CoVs is between 80 to 120 nm and their shapes are spherical. The fundamental structural proteins of CoVs are envelope (E), membrane (M), nucleocapsid (N), and spike (S). Its RNA genome composes of six to ten open reading frames (ORFs).

The SARS-CoV-2 outbreak started from a local seafood market in Huanan. Even initial reports showed that human-to-human transmission of the virus was not or limited, now human-to-human transmission exists [2]. It passes from person to person by respiratory droplets and also spreads through contact and fomites [3]. Coronaviruses have error-prone RNA-dependent RNA polymerases, mutations and recombination events occur that is concerned with rapidly strengthening and it increases its capacity to cause disease, which enhances also virulence [4]. ACE2, the receptor protein is present in humans in the epithelia of the lung and small intestine [5], and coronavirus binds to this receptor to enter into cell and infects the upper respiratory and gastrointestinal tract of mammals [6]. ORF1ab is a genomic region coding the putative replicase polyprotein. In other coronaviruses has also reported this

polyprotein has also as proteases encoded by this ORF region that causes expression of 10 proteins encoding important enzymes, which is essential for the survival of the viruses. ORF1a is the longest part of the RNA encoding replicases and ORF1b expresses for two large polyproteins including pp1a and pp1ab. The expression of pp1ab polyprotein is playing a role for programmed ribosomal frame shifting of signal conducting a bridge between ORF1a and ORF1ab [7]. This frame shifting signal leads to the expression of RNA-dependent RNA polymerase (RdRP) that is required for the coronavirus replication [8]. The increasing epidemiological and clinical evidence indicates that the SARS-CoV-2 has a stronger transmissibility than SARS-CoV [9]. But the exact mechanism of SARS-CoV-2 is unclear [10]. As a result of a unique mechanism of viral replication, they have a high frequency of recombination [11, 12, 13, 14]. As a strategy, DNA sequence comparisons using single nucleotide polymorphisms (SNPs) are often followed for evolutionary studies to recognize the mutated coronavirus genomes where high mutations occurs due to an error-prone RNA-dependent RNA polymerase in genome replication [15, 16].

However, to our best knowledge, there is no detailed study comparing whole-genome sequences using algorithmic fragmentation programs such as MAUVE to see the stability of sequences among genomic pairs. Whole-genome comparison using MAUVE can be an efficient manner for aligning multiple nucleotide or protein sequences. Because MAUVE analysis is a method based on constructing multiple genome alignments with large-scale, which shows evolutionary changes and re-arrangements of inversion cases in genomes. We believe that any information resulted from the protein modelling can be also beneficial for drug designing. As an alternative approach, *in silico* analysis can accelerate discovering novel therapeutics for the prevention and treatment.

The mutations in the viral genome can be important for adaptation to host conditions, but the mechanism of these changes remains unclear [17]. Therefore, recent studies will fill the knowledge gaps to reveal how the virus is evolving and adapting to new conditions and which parts of the genome have stability than the other regions of genetic structure. Related information on its genetic stability can help us for treatment of SAR-CoV2.

In this study, we have investigated unvarying regions with less mutation than other parts of the genome on 134 different genome sequences of the GISAID database from distinct parts of the World. Our study aims to show stable regions in the viral genome, to do prediction on protein structure and docking analysis to find an effective molecule interacting with proteins of SARS CoV-2 to control its replication.

Material and Methods

Homology genome blast and genomes information

We retrieve total 134 complete genome sequences from the GISAID database [18] as of April 19, 2020. Only the complete genomes of high-coverage are included in the dataset. The complete genomes of the countries and territories infected by SARS-CoV-2 and shared are given in supplementary material (S1 File).

Phylogenetic analysis

To analyse the obtained SARS-CoV-2 genomes, sequence alignment was performed using Multiple Sequence Comparison by MAUVE and Clustal W of MegAlign from DNASTAR software [19]. The phylogenetic tree was conducted by a maximum likelihood using for the tree topology estimated with 1,000 bootstrap replicates. The maximum likelihood phylogenetic tree was constructed.

Nucleotide and amino acid sequence alignment and analysis

Nucleotide sequence editing and alignment were conducted using MAUVE and Clustal W of MegAlignPro, DNASTAR software [19]. The evolutionary history was inferred using the Neighbor-Joining method in MegAlignPro software. The sequences were analysed and common regions of all genomes were detected using MAUVE from pairwise alignment results obtained with Global Needleman-Wunsch algorithm [20]. Each unvarying genomic region was excised of whole sequences and subjected to protein similarity program of NCBI database using BlastX. This obtained FASTA sequence was converted to protein sequence using ExPASy proteomics server (<https://web.expasy.org/translate/>) [21] then loaded to I-

TASSER (Iterative Threading ASSEmbly Refinement) server of Michigan University, US (<https://zhanglab.ccmb.med.umich.edu/I-TASSER>) for prediction of proteins [22].

Homology modelling and protein prediction

Corresponding homology models predicted by I-TASSER server System for each target protein were downloaded from Protein Data Bank (PDB) (www.rcsb.org). Alignment of the protein sequences and subsequent homology modeling were done using ExpASy proteomics server [21] to study on the protein sequence and further structural details.

Ligand retrieval

The structure of NAGD was retrieved from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>). This structure was used for docking calculations. The selected 3D structure of the ligands was retrieved from PubChem Compound database in SDF format followed by conversion in the PDB format. The ligand parameters were analyzed using PRODRG online server (<http://prodr2.dyndns.org/cgi-bin/prodr2.cgi>) [23]. Further shape complementarity principle was applied with clustering RMSD 4.0 for docking calculations.

Molecular docking studies

Homology modeling and protein prediction analysis have directed us to test of protein receptors of SARS-CoV-2 with our ligand. Later on PatchDock server (<http://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>), a geometry based molecular docking algorithm was used for docking analysis using cluster RMSD at default value of 4.0 and protein-small ligand complex type as the analysis parameters [24, 25, 26, 27]. Analysis on PatchDock yielded results for geometric shape complementarity score (GSC score) and approximate interface area (AI area). Additionally CHIMERA, an AutoDock software Vina based tool for molecular docking was used. The flexible docking study was carried out using Autodock v 4.0 [28]. The interaction analysis of protein- ligand complexes and their amino acid position with bond distances were calculated and visualized through the PyMol. Molecular docking simulations were confirmed by Protein dock server SWISSDOCK (<http://www.swissdock.ch/docking>) within protein receptors and ligand [29]. Later, Pymol software has been used to get insight into their all binding preferences within the active site of these receptors.

Results

Phylogenetic tree

The maximum likelihood phylogenetic tree shows main clades containing several clusters and the viral genome sequences show genetic diversity according to Mauve ve Clustal W analysis, respectively (S2a and S2b Files).

Nucleotide and amino acid sequence alignment and analysis

Our results showed high mutational changes in whole genomes except for the first 6500 bp which is constantly unvarying part in whole sequences. MAUVE results have confirmed the Clustal W analyses, reciprocally (Fig 1).

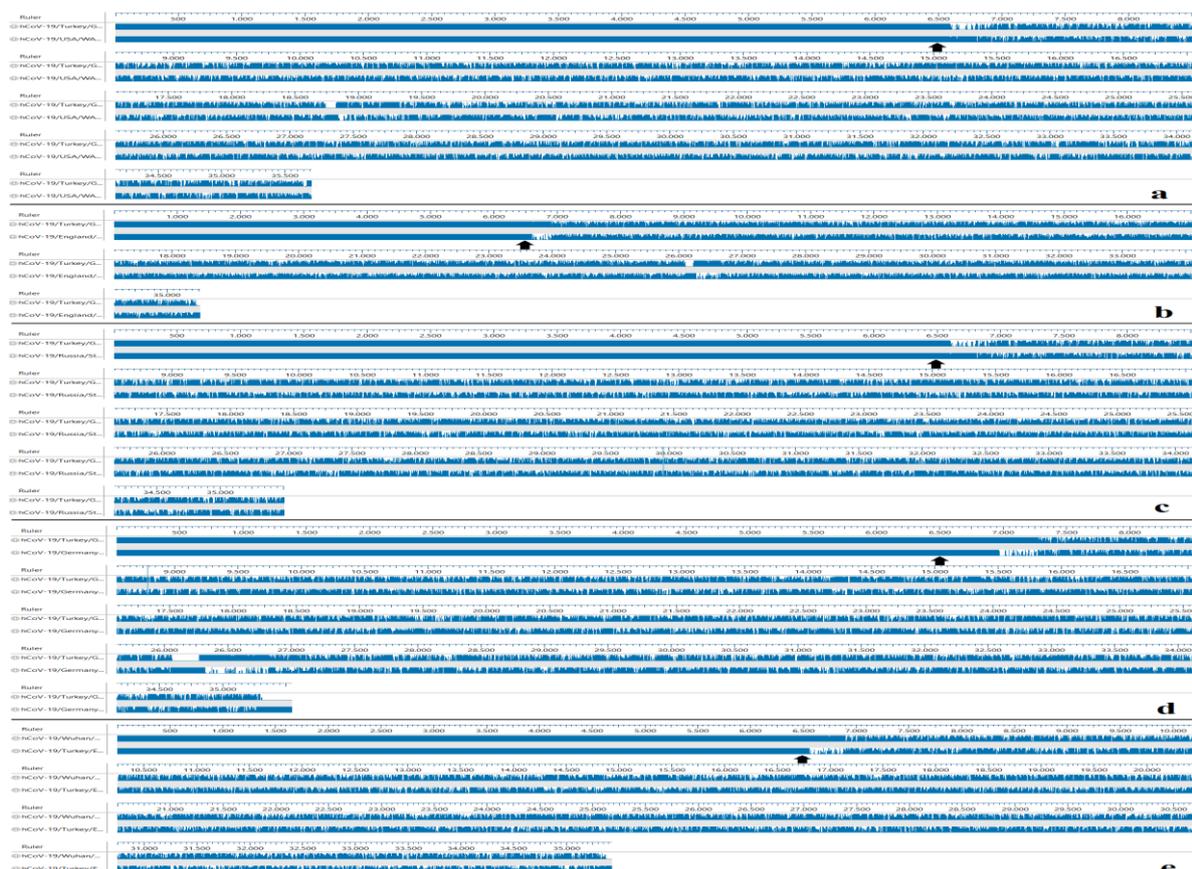


Fig 1. Sequence comparison show genomic differences of SARS-CoV2 samples belonging to a) US, b) UK, c) Russia, d) Germany and e) China with Turkey. The analysis using Global Needleman Wunsch DNA fragment algorithm clearly shows genetic stability up to first 6500 bp. after alignment using MAUVE.

Homology modeling and protein prediction

The excised uniform regions of sequences subjected to alignment for protein similarity indicated that 6500 bp region including ORF1ab consists of constantly unvarying sequences (Fig 1). These stable sequences were selected as a template for further protein structural predictions (Fig 2). The results were structurally very close to 6JX7 (Cryo-EM structure of spike protein of feline infectious peritonitis virus strain UU4) as a target protein according to I-TASSER analysis and the ligand was determined as N-acetyl-D-glucosamine (NAG), a small molecule showing interaction with these stable sequences (S3, S4 File). Therefore, our further study is based on testing the same protein-ligand interaction possibilities on spike receptor-binding domain bound with ACE2 (6M0J) and RNA-binding domain of nucleocapsid phosphoprotein (6WKP) with our ligand NAG.

```
>QJA17583.1:1-2042 ORF1a polyprotein [Severe acute respiratory syndrome coronavirus 2] >protein
MSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECNLLLQYGSFCTQLNRALTGIAVEQDK
NTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIARDLCAQKFN
GLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKI
QDLSSTASALGKLQDVVNQNAQALNVLKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLI
RAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLSFPQSAPHGVVFLHVTYVPAQEKNTTAPAICHGDKAHFP
REGVFVSNQTHWFVTQRNFYEPQIITDNTFVSGNCDVVIGVNNVTYDPLQPELDSFKEELDKYFKNHTSPDVLGDI
SGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCLKGCC
```

Fig 2. Submitted protein sequence for predictional analysis. Asparagine regions (N) are marked bold.

Protein Docking

For docking analysis of NAG with 6M0J and 6WKP, the ligand structure of NAG retrieved from PubChem database was analysed (Fig 3a, 3b), using PatchDock server and confirmed by visualization of the docked complexes by CHIMERA software. The prominent binding sites were also predicted through MetaPocket 2.0 server. Docking of 6M0J and 6WKP with the target molecule NAG was studied with respect to following parameters: (a) interacting amino acids (b) ligand and protein atoms involved in hydrogen bonding (S5 File).

The results of SWISSDOCK server confirmed our results obtained with CHIMERA and Vina softwares calculation. Furthermore binding possibilities of ligand on protein surface have been confirmed with results of SWISSDOCK (Fig 4, 5, S6, S7 Files).

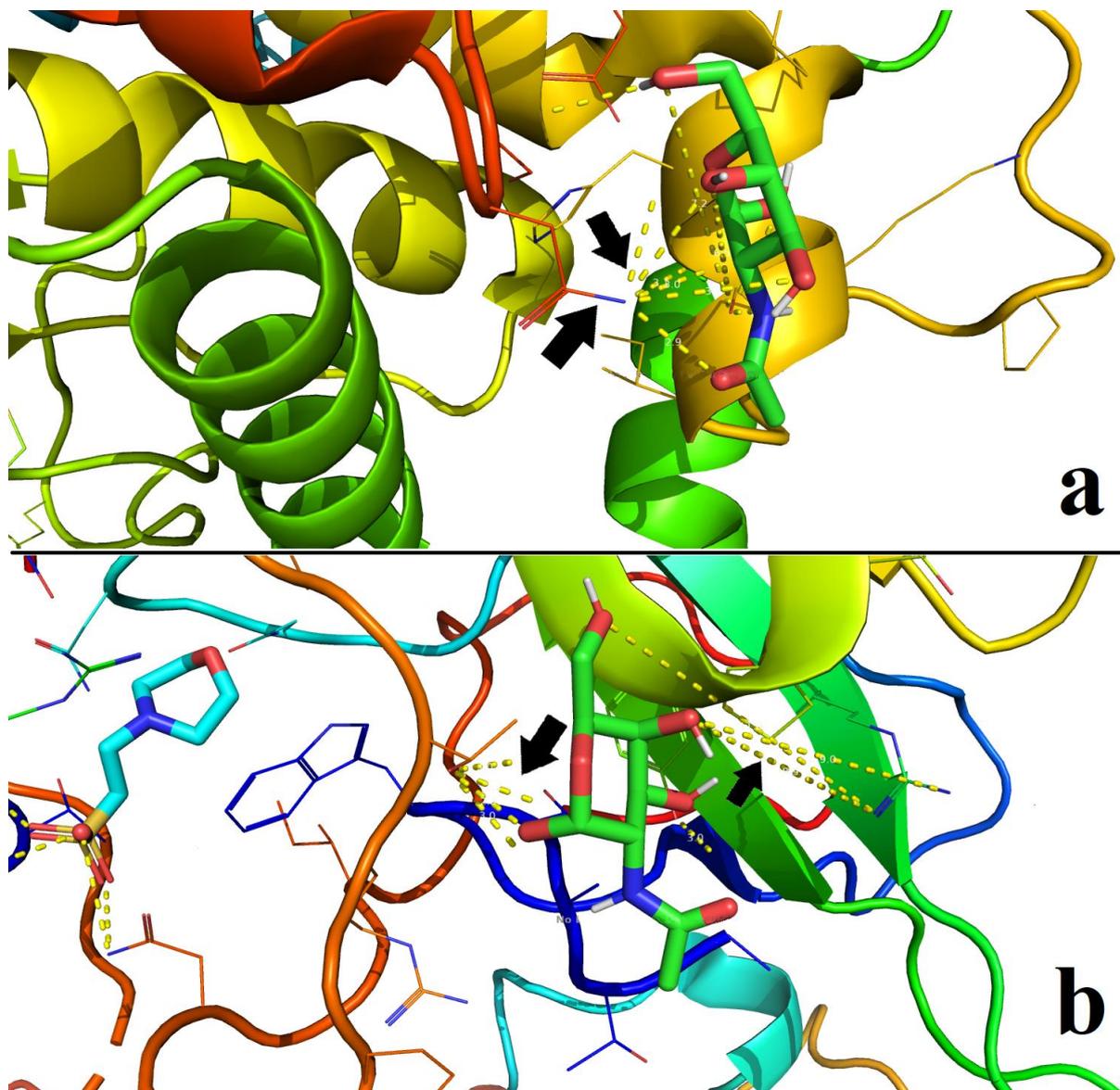


Fig.3 Arrow shows binding possibilities (Chimera, Autodock results) and measurements of NAGD to a) SARS-CoV-2 spike receptor-binding domain bound with ACE2 (6M0J), b) RNA-binding domain of nucleocapsid phosphoprotein from SARS CoV-2 (6WKP) by Pymol software.

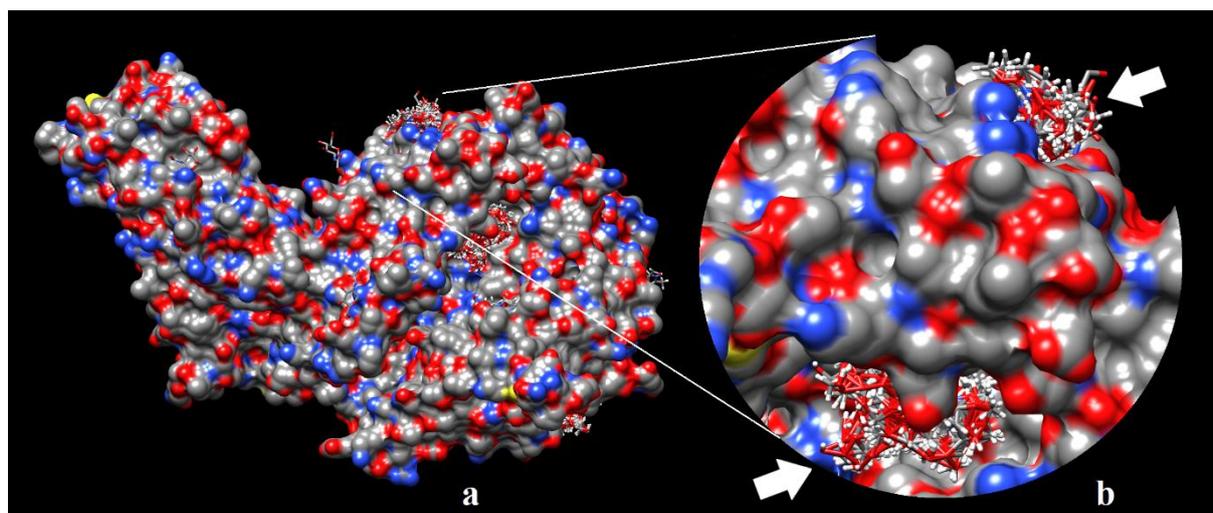


Fig.4 a) Protein docking analysis and binding possibilities (SWISSDOCK results) of protein 6M0J with N-acetyl D-glucosamine, b) closer view of binding possibilities marked with arrow.

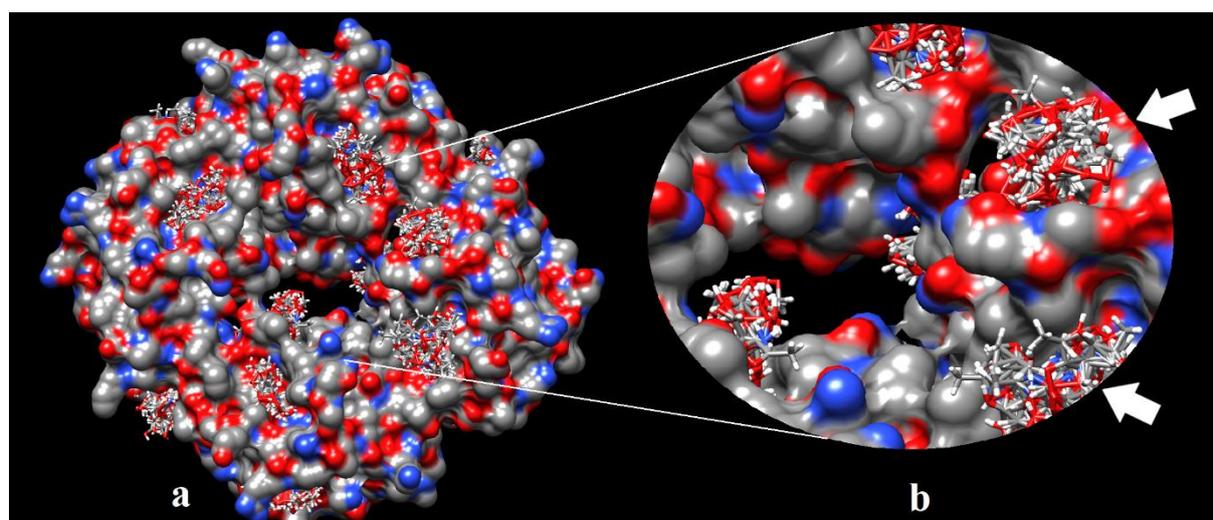


Fig. 5 a) Protein docking analysis and binding possibilities (SWISDOCK results) of protein 6WKP with N-acetyl D-glucosamine, b) closer view of binding possibilities marked with arrows.

Discussion

We have clearly shown the existence of the genetical specific, unvarying region in whole SARS-COV-2 genomes. After alignment of all sequences by MAUVE, we have seen this

uniformity in all different 134 sequences. In analysis, we have detected common sequences showing no mutational differentiations (Fig 1). We have determined that MAUVE is the most effective method for genome comparisons. Pairwise alignment by Global Needleman-Wunsch algorithm has shown this uniform sequence with no mutations in all paired sequences up to 6500 bp by MAUVE and also confirmed with Clustal W. Only does the first 6500 bp seem unvarying region rather than remaining part of the viral genome. The determined unvarying part in viral genome has special characteristic properties to encompass further immunologic studies. We suggest Global Needleman Wunsch pairwise alignment analysis for observing uniformity of genome sequences as an effective method.

We have reported our detected unvarying sequences including ORF1ab to be an important region responsible for the putative replicase polyprotein of proteases secretions [30]. Correspondently, we observed sequence variation with high ratio of genomes (except for 6500 bp fragment) (Fig 1) seems not very convenient as a target point for drug discovery. Hence, it can be hypothesised according to our findings that during the transmission and evolutionary processes the first genetically stable 6500 bp could be an appropriate target for antiSARS drugs. Also, our data showed N-acetyl-D-glucosamine interacts with proteins encoded by ORF1ab region (S3 File, I-TASSER analysis data). Previous studies have also reported the effectiveness of NADG against influenza [31]. In another study, glucosamine has been reported to influence on replication of hepatitis B virus by in-vitro and in-vivo experiments [32]. Therefore, NADG can also be suggested as an antiviral drug for SARS-Cov2.

As known glycosylation is a major process which affects the binding of monoclonal antibodies to the coated virus in the vaccine development process but deglycosylation reduces binding of the antibodies in *vice versa*. Therefore, binding of neutralising monoclonal bodies to virus protein depends on glycosylation of the virus [33]. N-linked glycans on an immune cell's surface will help for the migration pattern of the cell with specific glycosylations [34]. These patterns on the various immunoglobulins give specific shape and unique effector properties for affinities of immune receptors. It could also involve glycans in "self" and "non-self" discrimination, which could apply to response against virus as previously reported on the various autoimmune diseases [34]. Glycans consist of different derivatives of NADG and suggests having an important role in the immune system.

Moreover, Pant et al. reported asparagine supply is a critical barrier and limiting factor for replication of virus proteins to development of antiviral drugs [35]. We assume that virus

prefers glutamine to glucose for efficient replication, and the viral replication reduces in glutamine-free medium. Asparagine supplementation compensates for glutamine depletion, for viral replication. Asparagine-linked glycosylation is an enzyme-catalyzed, co-translational protein modification which influences either the protein folding process or the stability of the native conjugated glycoprotein form [36]. In our study, we have found NADG interacts with proteins encoded by ORF1ab region; we suggest binding of NADG to asparagine and inhibition of virus replication as reported by Pant et al. [35]. We have observed by docking analysis that binding of NADG to asparagine is also possible. In unvarying sequences, we have detected 34 asparagine amino acid residues that can be a target point for binding of NADG as ligand molecule (Fig 2). Particularly, the effect of NADG has also been tested against HIV1 with different concentrations (0.25 mM, 1 mM, 4 mM, and 16 mM) [37]. We can suggest the same mode of action to HIV1 [35, 37]. Our predictional protein structure and docking analysis showed N-acetyl D glucosamine is a major compound showing high interaction possibility, which can interact with our tested proteins 6M0J and 6WKP of SARS-CoV-2. A previous study has reported seven glycosylation sites playing a role on the S protein, which is critical for DC/L-SIGN-mediated virus attachment to asparagine residues at amino acid positions that are distinct from residues of the ACE2-binding domain [38]. Defects in secretion and infectivity of several flaviviruses concerned with blocking of the N-linked oligosaccharides have confirmed the role of glycosylation [39]. A previous study has reported to have the effect of removing the terminal glucose residues on the N-linked glycans for altering the mechanism of controlling protein folding mediated by ER chaperones for virus replication [40]. The results of another study were evidence that some viruses (some members of the NCLDV, such as Chlorella viruses) use the host ER/Golgi system for their glycoprotein production, which is the machinery required for the glycosylation of its structural proteins [41]. A virus-encoded uridine diphosphate-N-acetylglucosamine pathway associated with NAG is a ubiquitous sugar which represents a fundamental process for virus. Therefore, NADG can be a substitute of NAG and could convert all processes to support immunity defence mechanism [42]. Our molecular docking analysis on NADG, which could mimic NAG, could keep the cell of the SARS-CoV-2's viral integration into ER and Golgi system. Our results also show that NADG has an interaction with 6M0J and/or 6WKP (Fig 4, 5).

However, the role of asparagine availability in virus replication remained unclear up to now [36]. The influences of NAG on cell surface signalling proteins alter signal transduction depending on the degree of branching of N-linked glycans [43]. Therefore, this signal

transduction could change in immunity system's favour by NAGD mimicing the same role in signalling instead of NAG. We have found binding possibilities of NADG with 6M0J and 6WKP could defect the attachment on human cell and replication mechanism of the virus.

Our study has indicated that highly frequent mutations in all SARS-CoV-2 genomes except for first 6500 bp regions (Fig 1). The highly frequent SNP mutations discovered with pairwise alignment using comparative computational analysis, our results show correspondence with other studies reporting the changes in transmissibility and virulence of the virus [44]. Therefore, the high-frequency SNP mutations are important limiting factors for vaccine development and preventing of SARS-CoV-2 infections. For effective drug treating, the rapid way is to find potential molecules to SARS-CoV-2. Once the efficacy has determined, it can be tested for the clinical treatment of patients. A recent study has reported ORF1ab region polyprotein belonging to a part of nonstructural protein 1 (nsp1) with the high antigenicity residues in a glycine-proline or hydrophobic amino acid-rich domain. Nsp1 is a virulence factor and crucial agent in spreading of the virus among the society can be a potential target for the future epidemiology, drug, and vaccine studies [45]. Our suggested successful construction of the 3-D structure model with docking analyses, the preliminary function predicted showing stable expression of proteins including ORF1ab established the foundation for the further exploration of its biological process and contributed to the search for antiSARS-CoV-2 drugs.

Our study predicts a repurposing compound that has a high potential for inhibiting of the virus and provides information to scientists on this compound. Subsequent validation of anti-viral effects in vitro and in vivo will be useful data for clinical treatment of SARS-CoV-2. Our results of the entire article based on in-silico screening shows prediction on the effect of a molecule showing interaction with proteins of the virus. Unfortunately, we have not conducted further in-vivo and in-vitro experiments yet. But we want to share our results with scientific area of anti-SARS-CoV-2 research groups since we have thought that it could be an effective molecule. In previous studies, even genetic structure, mutation, the protein structure of SARS-Cov-2 have been explored in details, there is no similar study relied on an effective potential molecule as we have suggested. Given cost- and time- effective strategy, computational methods are useful to those who wish to understand essential information about SARS-CoV-2 for subsequent analyses. We stress in-silico studies are important tools for the elucidation of major effective compounds interacting with the virus. We purpose, the recent advances in drug discovery by in-silico screening [46, 47] give scientists an opportunity for rapid detection of efficient molecules target-oriented on SARS-CoV-2 [48].

Conclusion

The SARS-CoV-2 epidemic gave rise to substantial health emergency and economic drawbacks in the world. Hence, understanding the nature of this virus and to monitor its spreading in the epidemic are critical in disease control. Potential importance in targeting ORF1ab region should draw the attention of researchers for future preventive strategies in pharmaceutical and vaccine development studies. Given attention to the finding of new targets to effectively treat SARS-CoV-2, understanding the molecular effects of repurposed compounds can be in prioritizing pharmacological strategies. Our suggested approach can be drastically helpful for the clinical inefficacy of common antiviral drugs [49, 50].

We strongly suggest testing different concentrations of NADG to SARS-CoV2, considering interaction with proteins including ORF1ab region showing constantly unvarying piece of sequence in whole paired genome sequences. Our results are likely to increase the underpinning data for drug repurposing in the therapeutic options against SARS- CoV in the future.

Acknowledgement

The authors wish to thank Dr. Venkatesh Balakrishnan for DNASTAR MegAlignPro software. We sincerely appreciate the researchers worldwide who sequenced and shared the complete genome data of SARS-CoV-2 in GISAID database (<https://www.gisaid.org/>). This research relied on their precious submitted data. The genome sequences downloaded from GISAID are given in the supplementary material (S1 File). The authors also thank to anonymous reviewers for their insightful suggestions.

This article is dedicated to the heroic medical workers fighting in the front line of anti-epidemic and made great sacrifices all around World.

Competing interests: The authors have declared that no competing interests exist.

References

1. Chan JF, To KK, Tse H, Jin DY, Yuen KY. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. *Trends Microbiol.* 2013;10: 544-555. doi: 10.1016/j.tim.2013.05.005.
2. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol.* 2019;17: 181-192. doi: 10.1038/s41579-018-0118-9.
3. Booth TF1, Kournikakis B, Bastien N, Ho J, Kobasa D, Stadnyk L, et al. Detection of airborne severe acute respiratory syndrome (SARS) coronavirus and environmental contamination in 20 SARS outbreak units. *J Infect Dis.* 2005;191: 1472–1477. doi: 10.1086/429634.
4. Cui J, Li F, Shi Z. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol.* 2019;17: 181–192. doi: 10.1038/s41579-018-0118-9.

5. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol.* 2004;203: 631-637. doi: 10.1002/path.1570.
6. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci.* 2020;12: 1-5. doi: 10.1038/s41368-020-0074-x.
7. Plant EP, Sims AC, Baric RS, Dinman JD, Taylor DR. Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses.* 2013;5: 279-294. doi: 10.3390/v5010279.
8. Plant EP, Dinman JD. The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci.* 2008;13: 4873-4881. doi: 10.2741/3046.
9. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* 2020;382: 1708-1720. doi: 10.1056/NEJMoa2002032.
10. Ziebuhr J. The coronavirus replicase. *Curr Top Microbiol Immunol.* 2005;287: 57-94. doi:10.1007/3-540-26765-4_3.
11. Lai MM, Baric RS, Makino S, Keck JG, Egbert J, Leibowitz JL, et al. Recombination between nonsegmented RNA genomes of murine coronaviruses. *J Virol.* 1985;56: 449-456.
12. Makino S, Keck JG, Stohlman SA, Lai MM. High-frequency RNA recombination of murine coronaviruses. *J Virol.* 1986;57: 729-737.
13. Kusters JG, Jager EJ, Niesters HG, van der Zeijst BA. Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. *Vaccine.* 1990;8: 605-608. doi: 10.1016/0264-410x(90)90018-h.
14. Lee CW, Jackwood MW. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Arch Virol.* 2000;145: 2135-2148. doi: 10.1007/s007050070044.
15. Domingo E, Holland J. RNA virus mutations and fitness for survival. *Annu Rev Microbiol.* 1997;51: 151-178. doi: 10.1146/annurev.micro.51.1.151.

16. Holmes EC, Rambaut A. Viral evolution and the emergence of SARS coronavirus. *Philos Trans R Soc Lond B Biol Sci.* 2004;359: 1059-1065. doi: 10.1098/rstb.2004.1478.
17. Cherry JD, Krogstad P. SARS: the first pandemic of the 21st century. *Pediatr Res.* 2004;56: 1-5. doi: 10.1203/01.PDR.0000129184.87042.FC.
18. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *EuroSurveillance.* 2017; 22(13). doi: 10.2807/1560-7917.ES.2017.22.13.30494.
19. Burland TG. DNASTAR's Lasergene Sequence Analysis Software. *Methods Mol Biol.* 2000;132: 71-91. doi: 10.1385/1-59259-192-2:71.
20. Needleman, SB, Wunsch, CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* 1970;48: 443-453. doi: 10.1016/0022-2836(70)90057-4.
21. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003;31: 3784-3788. doi: 10.1093/nar/gkg563.
22. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. *Nat Meth.* 2015;12: 7-8. doi: 10.1038/nmeth.3213.
23. Schüttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* 2004;D60: 1355-1363. doi: 10.1107/S09074444904011679.
24. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. In Gusfield et al. editors. *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI) Lecture Notes in Computer Science.* Rome, Italy: Springer Verlag; 2002. pp. 185-200. doi: 10.1007/3-540-45784-4_14.
25. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, et al. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins.* 2003;52: 107-12. doi: 10.1002/prot.10397.
26. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucl Acids Res.* 2005;33: W363-367. doi: 10.1093/nar/gki481.

27. Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An integrated suite of fast docking algorithms. *Proteins*. 2010;78: 3197-3204. doi: 10.1002/prot.22790.
28. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31: 455-461. doi: 10.1002/jcc.21334.
29. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res*. 2011;39 suppl_2: W270–W277. doi: 10.1093/nar/gkr366.
30. Woo PC, Huang Y, Lau SK, Tsoi HW, Yuen KY. In Silico Analysis of ORF1ab in Coronavirus HKU1 genome reveals a unique putative cleavage site of Coronavirus HKU1 3C-like protease. *Microbiol Immunol*. 2005;49: 899-908. doi: 10.1111/j.1348-0421.2005.tb03681.x.
31. Chen CA, Fang JM. Synthesis of oseltamivir and tamiphosphor from N-acetyl-D-glucosamine. *Org Biomol Chem*. 2013;11: 7687-7699. doi: 10.1039/c3ob41622d.
32. Lin Y, Wu C, Wang X, Liu S, Zhao K, Kemper T, et al. Glucosamine promotes hepatitis B virus replication through its dual effects in suppressing autophagic degradation and inhibiting MTORC1 signaling. *Autophagy*. 2020;16: 548-561. doi: 10.1080/15548627.2019.1632104.
33. Koch G, Kant A. Binding of antibodies that strongly neutralise infectious bronchitis virus is dependent on the glycosylation of the viral peplomer protein. In: Cavanagh D, Brown TDK, editors. *Coronaviruses and their Diseases. Advances in Experimental Medicine and Biology*. Boston, MA: Springer; 1990. Pp. 143-150. doi: 10.1007/978-1-4684-5823-7_21.
34. Maverakis E, Kim K, Shimoda M, Gershwin ME, Patel F, Wilken R, et al. Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *J Autoimmun*. 2015;57: 1–13. doi: 10.1016/j.jaut.2014.12.002.
35. Pant A, Cao S, Yang Z. Asparagine is a critical limiting metabolite for Vaccinia virus protein synthesis during glutamine deprivation. *J Virol*. 2019; 93(13) e01834-18; doi: 10.1128/JVI.01834-18.

36. Imperiali B, O'Connor SE. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol.* 1999;3: 643-649. doi: 10.1016/S1367-5931(99)00021-6.
37. Jochmann R, Thureau M, Jung S, Hofmann C, Naschberger E, Kremmer E, et al. O-Linked N-Acetylglucosaminylation of Sp1 inhibits the Human Immunodeficiency Virus Type 1 promoter. *J Virol.* 2009;83: 3704-3718. doi: 10.1128/JVI.01384-08.
38. Han DP, Lohani M, Cho MW. Specific Asparagine-Linked glycosylation sites are critical for DC-SIGN- and L-SIGN-Mediated Severe Acute Respiratory Syndrome Coronavirus entry. *J Virol.* 2007;81: 12029-12039. doi: 10.1128/JVI.00315-07.
39. Mondotte JA, Lozach PY, Amara A, Gamarnik AV. Essential role of Dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation. *J Virol.* 2007;81: 7136-7148. doi: 10.1128/JVI.00116-07.
40. Parodi AJ. Protein glucosylation and its role in protein folding. *Annu Rev Biochem.* 2000. 69:69-93. doi: 10.1146/annurev.biochem.69.1.69.
41. Van Etten JL, Gurnon JR, Yanai-Balser GM, Dunigan DD, Graves MV. Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. *Biochim Biophys Acta.* 2010;1800: 152-159. doi: 10.1016/j.bbagen.2009.07.024.
42. Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. *Glycobiology.* 2014;24: 51–61. doi: 10.1093/glycob/cwt089.
43. Konopka JB. N-acetylglucosamine (GlcNAc) functions in cell signaling. *Scientifica (Cairo).* 2012;1: pii: 489208. doi.org/10.6064/2012/489208.
44. Tilocca B, Soggiu A, Sanguinetti M, Musella V, Britti D, Bonizzi L, et al. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes and Infection.* 2020; Forthcoming. doi: 10.1016/j.micinf.2020.04.002.
45. Zolfaghari Emameh R, Nosrati H, Taheri RA. Combination of biodata mining and computational modelling in identification and characterization of ORF1ab polyprotein of

SARS-CoV-2 isolated from Oronasopharynx of an Iranian patient. *Biol Proced Online*. 2020; 22, 8. doi: 10.1186/s12575-020-00121-9.

46. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov*. 2020;6; 14. doi: 10.1038/s41421-020-0153-3.

47. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein J Org Chem*. 2016;12: 2694-2718. doi: 10.3762/bjoc.12.267.

48. Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat Rev Drug Discov*. 2020;19: 149-150. doi: 10.1038/d41573-020-00016-0.

49. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet*. 2020;395: 1054-1062. doi: 10.1016/S0140-6736(20)30566-3.

50. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. *N Engl J Med*. 2020; 382: 1787-1799. doi: 10.1056/NEJMoa2001282.

Supporting information

S1 File. The genome sequences downloaded from GISAID.

S2a File. MAUVE genetic phylogenetic analysis.

S2b File. Clustal W genetic phylogenetic analysis.

S3 File. I-TASSER analysis results.

S4 File. RCSB 6JX7 PDB results.

S5 File. Torsion Tables. a) 6M0J Vina Autodock results, b) 6WKP Vina Autodock results.

S6 File. 6M0J SWISSDOCK results.

S7 File. 6WKP SWISSDOCK results.